FIELDS

Warren Siegel

C. N. Yang Institute for Theoretical Physics
State University of New York at Stony Brook
Stony Brook, New York 11794-3840 USA

mailto:siegel@insti.physics.sunysb.edu
http://insti.physics.sunysb.edu/~siegel/plan.html

CONTENTS

PART ONE: SYMMETRY

PART TWO: QUANTA

PART THREE: HIGHER SPIN

OUTLINE

In this Outline we give a brief description of each item listed in the Contents. While the Contents and Index are quick ways to search, or learn the general layout of the book, the Outline gives more detail for the uninitiated. (The PDF version also allows use of Acrobat Reader's "Find" command.)

PART ONE: SYMMETRY

Relativistic quantum mechanics and classical field theory. Poincaré group = special relativity. Enlarged spacetime symmetries: conformal and supersymmetry. Equations of motion and actions for particles and fields/wave functions. Internal symmetries: global (classifying particles), local (field interactions).

PART TWO: QUANTA

Quantum aspects of field theory. Perturbation theory: expansions in loops, helicity, and internal symmetry. Although some have conjectured that nonperturbative approaches might solve renormalization difficulties found in perturbation, all evidence indicates these problems worsen instead in complete theory.

PART THREE: HIGHER SPIN

General spins. Spin 2 must be included in any complete theory of nature. Higher spins are observed experimentally for bound states, but may be required also as fundamental fields.

IX. General relativity

Treatment closely related to that applied to Yang-Mills, super Yang-Mills, and supergravity. Based on methods that can be applied directly to spinors, and therefore to supergravity and superstrings. Somewhat new, but simplest, methods of calculating curvatures for purposes of solving the classical field equations.

XI. Strings

Approach to studying most important yet least understood property of QCD: confinement. Other proposed methods have achieved explicit results for only low hadron energy. String theory is also useful for field theory in general.

PREFACE

Scientific method

Although there are many fine textbooks on quantum field theory, they all have various shortcomings. *Instinct* is claimed as a basis for most discussions of quantum field theory, though clearly this topic is too recent to affect evolution. Their subjectivity more accurately identifies this as *fashion*: (1) The *old-fashioned* approach justifies itself with the instinct of *intuition*. However, anyone who remembers when they first learned quantum mechanics or special relativity knows they are counter-intuitive; quantum field theory is the synthesis of those two topics. Thus, the intuition in this case is probably just *habit*: Such an approach is actually *historical* or *traditional*, recounting the chronological development of the subject. Generally the first half (or volume) is devoted to quantum electrodynamics, treated in the way it was viewed in the 1950's, while the second half tells the story of quantum chromodynamics, as it was understood in the 1970's. Such a "dualistic" approach is necessarily redundant, e.g., using canonical quantization for QED but path-integral quantization for QCD, contrary to scientific principles, which advocate applying the same "unified" methods to all theories. While some teachers may feel more comfortable by beginning a topic the way they first learned it, students may wonder why the course didn't begin with the approach that they will wind up using in the end. Topics that are unfamiliar to the author's intuition are often labeled as "formal" (lacking substance) or even "mathematical" (devoid of physics). Recent topics are usually treated there as advanced: The opposite is often true, since explanations simplify with time, as the topic is better understood. On the positive side, this approach generally presents topics with better experimental verification.

(2) In contrast, the *fashionable* approach is described as being based on the instinct of *beauty*. But this subjective beauty of *art* is not the instinctive beauty of nature, and in science it is merely a consolation. Treatments based on this approach are usually found in review articles rather than textbooks, due to the shorter life expectancy of the latest fashion. On the other hand, this approach has more imagination than the traditional one, and attempts to capture the future of the subject.

A related issue in the treatment of field theory is the relative importance of *concepts* vs. *calculations*: (1) Some texts emphasize the concepts, including those which have not proven of practical value, but were considered motivational historically (in the traditional approach) or currently (in the artistic approach). However, many approaches that were once considered at the forefront of research have faded into oblivion not because they were proven wrong by experimental evidence or lacked conceptual attractiveness, but because they were too complex for calculation, or so vague they lacked predicitive ability. Some methods claimed total generality, which they used to prove theorems (though sometimes without examples); but ultimately the only useful proofs of theorems are by construction. Often a dualistic, two-volume approach is again advocated (and frequently the author writes only one of the two volumes): Like the traditional approach of QED volume + QCD volume, some prefer concept volume + calculation volume. Generally, this means that gauge theory S-matrix calculations are omitted from the conceptual field theory course, and left for a "particle physics" course, or perhaps an "advanced field theory" course. Unfortunately, the particle physics course will find the specialized techniques of gauge theory too technical to cover, while the advanced field theory course will frighten away many students by its title alone.

(2) On the other hand, some authors express a desire to introduce Feynman graphs as quickly as possible: This suggests a lack of appreciation of field theory outside of diagrammatics. Many essential aspects of field theory (such as symmetry breaking and the Higgs effect) can be seen only from the action, and its analysis also leads to better methods of applying perturbation theory than those obtained from a fixed set of rules. Also, functional equations are often simpler than pictorial ones, especially when they are nonlinear in the fields. The result of over-emphasizing the calculations is a cookbook, of the kind familiar from some lower-division undergraduate courses intended for physics majors but designed for engineers.

The best explanation of a theory is the one that fits the principles of *scientific method*: simplicity, generality, and experimental verification. In this text we thus take a more *economical* or *pragmatic* approach, with methods based on efficiency and power. Unattractiveness or counter-intuitiveness of such methods become advantages, because they force one to accept new and better ways of thinking about the subject: The efficiency of the method directs one to the underlying idea. For example, although some consider Einstein's original explanation of special relativity in terms of relativistic trains and Lorentz transformations with square roots as being more physical, the concept of Minkowski space gave a much simpler explanation and deeper understanding that proved more useful and led to generalization. Many theories have "miraculous cancelations" when traditional methods are used, which led to new methods (background field gauge, supergraphs, spacecone, etc.) that not only incorporate the cancelations automatically (so that the "zeros" need not be calculated), but are built on the principles that explain them. We place an emphasis on such new concepts, as well as the calculational methods that allow them to be compared with nature. It is important not to neglect one for the sake of the other, artificial and misleading to try to separate them.

As a result, many of our explanations of the standard topics are new to textbooks, and some are completely new. For example:

(1) We derive the Foldy-Wouthuysen transformation by dimensional reduction from an analogous one for the massless case (subsections IIB3,5).
(2) We derive the Feynman rules in terms of background fields rather than sources (subsection VC1); this avoids the need for amputation of external lines for S-matrices or effective actions, and is more useful for background-field gauges.
(3) We obtain the nonrelativistic QED effective action, used in modern treatments of the Lamb shift (because it makes perturbation easier than the older Bethe-Salpeter methods), by field redefinition of the relativistic effective action (subsection VIIIB6), rather than fitting parameters by comparing Feynman diagrams from the relativistic and nonrelativistic actions. (In general, manipulations in the action are easier than in diagrams.)
(4) We present two somewhat new methods for solving for the covariant derivatives and curvature in general relativity that are slightly easier than all previous methods (subsections IXA2,A7,C5).

There are also some completely new topics, like:
(1) the anti-Gervais-Neveu gauge, where spin in U(N) Yang-Mills is treated in almost the same way as internal symmetry — with Chan-Paton factors (subsection VIB4);
(2) the superspacecone gauge, the simplest gauge for QCD (subsection VIB7); and
(3) a new "(almost-)first-order" superspace action for supergravity, analogous to the one for super Yang-Mills (subsection XB1).

We try to give the simplest possible calculational tools, not only for the above reasons, but also so group theory (internal and spacetime) and integrals can be performed with the least effort and memory. Some traditionalists may claim that the old methods are easy enough, but their arguments are less convincing when the order of perturbation is increased. Even computer calculations are more efficient when left as a last resort; and you can't see what's going on when the computer's doing the calculating, so you don't gain any new understanding. We give examples of (and exercises on) these methods, but not exhaustively. We also include more recent topics (or those more recently appreciated in the particle physics community) that might be deemed non-introductory, but are commonly used, and are simple and important enough to include at the earliest level. For example, the related topics of (unitary) lightcone gauge, twistors, and spinor helicity are absent from all field theory texts, and as a result no such text performs the calculation of as basic a diagram as the 4-gluon tree amplitude. Another missing topic is the relation of QCD to strings through the random worldsheet lattice and large-color (1/N) expansion, which is the only known method that might quantitatively describe its high-energy nonperturbative behavior (bound states of arbitrarily large mass).

This text is meant to cover all the field theory every high energy theorist should know, but not all that any particular theorist might need to know. It is not meant as an introduction to research, but as a preliminary to such courses: We try to fill in the cracks that often lie between standard field theory courses and advanced specialized courses. For example, we have some discussion of string theory, but it is more oriented toward the strong interactions, where it has some experimental justification, rather than quantum gravity and unification, where its usefulness is still under investigation. We do not mention statistical mechanics, although many of the field theory methods we discuss are useful there. Also, we do not discuss any experimental results in detail; phenomenology and analysis of experiments deserve their own text. We give and apply the methods of calculation and discuss the qualitative features of the results, but do not make a numerical comparison to nature: We concentrate more on the "forest" than the "trees".

Unfortunately, our discussions of the (somewhat related) topics of infrared-divergence cancelation, Lamb shift, and the parton model are sketchy, due to our inability to give fully satisfying treatments — but maybe in a later edition?

Unlike all previous texts on quantum field theory, this one is available for free over the Internet (as usual, from arXiv.org and its mirrors), and may be periodically updated. Errata, additions, and other changes will be posted on my web page at http://insti.physics.sunysb.edu/˜siegel/plan.html until enough are accumulated for a new edition. Electronic distribution has many advantages:

- It's free.
- It's available quickly and easily. You can download it from the arXive.org or its mirrors, just like preprints, without a trip to the library (where it may be checked out) or bookstore or waiting for an order from the publisher. (If your connection is slow, download overnight.) And it won't go "out of print".
- Download it at work, home, etc. (or carry it on a CD), rather than carrying a book or printing multiple copies.
- Get updates just as quickly, rather than printing yet again.
- It has the usual Web links, so you can get the referenced papers just as easily.
- It has a separate "outline" window containing a table of contents on which you can click to take the main window to that item.
- You can electronically search (do a "find" on) the text.
- Easier to read on the computer screen (arbitrary magnification, etc.)
- Save trees, ink, and space.
- Theft is not a problem.
- No wear or tear.
- No paper cuts.

- If you have Adobe Acrobat (not just Reader) you can even add notes far bigger than would fit in the margin.

Highlights

The preceding Table of Contents lists the three parts of the text: Symmetry, Quanta, and Higher Spin. Each part is divided into four chapters, each of which has three sections, divided further into subsections. Each section is followed by references to reviews and original papers. Exercises appear throughout the text, immediately following the items they test: This purposely disrupts the flow of the text, forcing the reader to stop and think about what he has just learned. These exercises are interesting in their own right, and not just examples or memory tests. This is not a crime for homeworks and exams, which at least by graduate school should be about more than just grades.

This text also differs from others in most of the following ways:

(1) We place a greater emphasis on *mechanics* in introducing some of the more elementary physical concepts of field theory:

   (a) Some basic ideas, such as antiparticles, can be more simply understood already with classical mechanics.

   (b) Some interactions can also be treated through first-quantization: This is sufficient for evaluating certain tree and one-loop graphs as particles in external fields. Also, Schwinger parameters can be understood from first-quantization: They are useful for performing momentum integrals (reducing them to Gaussians), studying the high-energy behavior of Feynman graphs, and finding their singularities in a way that exposes their classical mechanics interpretation.

   (c) Quantum mechanics is very similar to free classical field theory, by the usual "semiclassical" correspondence ("duality") between particles (mechanics) and waves (fields). They use the same wave equations, since the mechanics Hamiltonian or Becchi-Rouet-Stora-Tyutin operator is the kinetic operator of the corresponding classical field theory, so the free theories are equivalent. In particular, (relativistic) quantum mechanical BRST provides a simple explanation of the off-shell degrees of freedom of general gauge theories, and introduces concepts useful in string theory. As in the nonrelativistic case, this treatment starts directly with quantum mechanics, rather than by (first-)quantization of a classical mechanical system. Since supersymmetry and strings are so important in present theoretical research, it is useful to have a text that includes the field theory concepts that are prerequisites to a course on these topics. (For the same reason, and because it can be treated so similarly to Yang-Mills, we also discuss general relativity.)

(2) We also emphasize *conformal invariance*. Although a badly broken symmetry, the fact that it is larger than Poincaré invariance makes it useful in many ways:
   (a) General classical theories can be described most simply by first analyzing conformal theories, and then introducing mass scales by various techniques. This is particularly useful for the general analysis of free theories, for finding solutions in gravity theories, and for constructing actions for supergravity theories.
   (b) Quantum theories that are well-defined within perturbation theory are conformal ("scaling") at high energies. (A possible exception is string theories, but the supposedly well understood string theories that are finite perturbatively have been discovered to be hard-to-quantize membranes in disguise nonperturbatively.) This makes methods based on conformal invariance useful for finding classical solutions, as well as studying the high-energy behavior of the quantum theory, and simplifying the calculation of amplitudes.
   (c) Theories whose conformal invariance is not (further) broken by quantum corrections avoid certain problems at the nonperturbative level. Thus conformal theories ultimately may be required for an unambiguous description of high-energy physics.

(3) We make extensive use of *two-component (chiral) spinors*, which are ubiquitous in particle physics:
   (a) The method of twistors (more recently dubbed "spinor helicity") greatly simplifies the Lorentz algebra in Feynman diagrams for massless (or high-energy) particles with spin, and it's now a standard in QCD. (Twistors are also related to conformal invariance and self-duality.) On the other hand, most texts still struggle with 4-component Dirac (rather than 2-component Weyl) spinor notation, which requires gamma-matrix and Fierz identities, when discussing QCD calculations.
   (b) Chirality and duality are important concepts in all the interactions: Two-component spinors were first found useful for weak interactions in the days of 4-fermion interactions. Chiral symmetry in strong interactions has been important since the early days of pion physics; the related topic of instantons (self-dual solutions) is simplified by two-component notation, and general self-dual solutions are expressed in terms of twistors. Duality is simplest in two-component spinor notation, even when applied to just the electromagnetic field.

   (c) Supersymmetry still has no convincing experimental verification (at least not at the moment I'm typing this), but its theoretical properties promise to solve many of the fundamental problems of quantum field theory. It is an element of most of the proposed generalizations of the Standard Model. Chiral symmetry is built into supersymmetry, making two-component spinors unavoidable.

(4) The topics are *ordered* in a more pedagogical manner:

   (a) Abelian and nonabelian gauge theories are treated together using modern techniques. (Classical gravity is treated with the same methods.)

   (b) Classical Yang-Mills theory is discussed before any quantum field theory. This allows much of the physics, such as the Standard Model (which may appeal to a wider audience), of which Yang-Mills is an essential part, to be introduced earlier. In particular, symmetries and mass generation in the Standard Model appear already at the classical level, and can be seen more easily from the action (classically) or effective action (quantum) than from diagrams.

   (c) Only the method of path integrals is used for second-quantization. Canonical quantization is more cumbersome and hides Lorentz invariance, as has been emphasized even by Feynman when he introduced his diagrams. We thus avoid such spurious concepts as the "Dirac sea", which supposedly explains positrons while being totally inapplicable to bosons. However, for quantum physics of general systems or single particles, operator methods are more powerful than any type of first-quantization of a classical system, and path integrals are mainly of pedagogical interest. We therefore "review" quantum physics first, discussing various properties (path integrals, S-matrices, unitarity, BRST, etc.) in a general (but simpler) framework, so that these properties need not be rederived for the special case of quantum field theory, for which path-integral methods are then sufficient as well as preferable.

(5) *Gauge fixing* is discussed in a way more general and efficient than older methods:

   (a) The best gauge for studying unitarity is the (unitary) lightcone gauge. This rarely appears in field theory texts, or is treated only half way, missing the important explicit elimination of all unphysical degrees of freedom.

   (b) Ghosts are introduced by BRST symmetry, which proves unitarity by showing equivalence of convenient and manifestly covariant gauges to the manifestly unitary lightcone gauge. It can be applied directly to the classical action, avoiding the explicit use of functional determinants of the older Faddeev-Popov method. It also allows direct introduction of more general gauges (again at the classical level) through the use of Nakanishi-Lautrup fields (which are omitted in older treatments of BRST), rather than the functional averaging over Landau gauges required by the Faddeev-Popov method.

(c) For nonabelian gauge theories the background field gauge is a must. It makes the effective action gauge invariant, so Slavnov-Taylor identities need not be applied to it. Beta functions can be found from just propagator corrections.

(6) *Dimensional regularization* is used exclusively (with the exception of one-loop axial anomaly calculations):

(a) It is the only one that preserves all possible symmetries, as well as being the only one practical enough for higher-loop calculations.

(b) We also use it exclusively for infrared regularization, allowing all divergences to be regularized with a single regulator (in contrast, e.g., to the *three* regulators used for the standard treatment of Lamb shift).

(c) It is good not only for regularization, but renormalization ("dimensional renormalization"). For example, the renormalization group is most simply described using dimensional regularization methods. More importantly, renormalization itself is performed most simply by a minimal prescription implied by dimensional regularization. Unfortunately, many books, even among those that use dimensional regularization, apply more complicated renormalization procedures that require additional, finite renormalizations as prescribed by Slavnov-Taylor identities. This is a needless duplication of effort that ignores the manifest gauge invariance whose preservation led to the choice of dimensional regularization in the first place. By using dimensional renormalization, gauge theories are as easy to treat as scalar theories: BRST does not have to be applied to amplitudes explicitly, since the dimensional regularization and renormalization procedure preserves it.

(7) Perhaps the most fundamental omission in most field theory texts is the expansion of QCD in the inverse of the *number of colors*:

(a) It provides a gauge-invariant organization of graphs into subsets, allowing simplifications of calculations at intermediate stages, and is commonly used in QCD today.

(b) It is useful as a perturbation expansion, whose experimental basis is the Okubo-Zweig-Iizuka rule.

(c) At the nonperturbative level, it leads to a resummation of diagrams in a way that can be associated with strings, suggesting an explanation of confinement.

(8) Our treatment of gravity is closely related to that applied to Yang-Mills theory, and differs from that of most texts on gravity:

(a) We emphasize the action for deriving field equations for gravity (and matter), rather than treating it as an afterthought.

(b) We make use of local (Weyl) scale invariance for cosmological and spherically symmetric solutions, gauge fixing, field redefinitions, and studying conformal properties. In particular, other texts neglect the (unphysical) dilaton, which is crucial in such treatments (especially for generalization to supergravity and strings).

(c) While most gravity texts leave spinors till the end, and treat them briefly, our discussion of gravity is based on methods that can be applied directly to spinors, and therefore to supergravity and superstrings.

(d) Our methods of calculating curvatures for purposes of solving the classical field equations are somewhat new, but probably the simplest, and are directly related to the simplest methods for super Yang-Mills theory and supergravity.

Notes for instructors

This text is intended for reference and as the basis for a full-year course on relativistic quantum field theory for second-year graduate students. A preliminary version of the first two parts was used for a one-year course I taught at Stony Brook, and more recently the same with Version 1. The chapter on gravity and pieces of early chapters cover a one-semester graduate relativity course I gave several times here — I used most of the following: IA, IB3, IC2, IIA, IIIA-C5, VIB1, IX, XIA2-4, XIB4-5. The prerequisites (for the quantum field theory course) are the usual first-year courses in classical mechanics, classical electrodynamics, and quantum mechanics. For example, the student should be familiar with Hamiltonians and Lagrangians, Lorentz transformations for particles and electromagnetism, Green functions for wave equations, SU(2) and spin, and Hilbert space. Unfortunately, I find that many second-year graduate students (especially many who got their undergraduate training in the USA) still have only an undergraduate level of understanding of the prerequisite topics, lacking a working knowledge of action principles, commutators, creation and annihilation operators, etc. While most such topics are briefly reviewed here, they should be learned elsewhere.

There is far more material here than can be covered comfortably in one year, mostly because of included material that should be covered earlier, but rarely is. Ideally, a modern curriculum for field theory students should include:

(1) courses on classical mechanics, nonrelativistic quantum mechanics, and classical electrodynamics in the first semester of graduate study, without overly reviewing aspects that should have been covered in undergraduate study (and in particular avoiding the enormous overlap of the last two subjects due to both covering primarily the solution of wave equations);

(2) in the second semester, statistical mechanics as the sequel to classical, relativistic quantum mechanics as the sequel to nonrelativistic, and classical nonabelian field theory (Yang-Mills and gravity) as the sequel to classical electrodynamics;

(3) in the second year, one year of quantum field theory, and at least one semester on "phenomenology" (model-building and direct comparison with observations, including those for general relativity and cosmology); and
(4) in the third year, more specialized courses, such as a semester on supersymmetry and strings.

Unfortunately, in practice little of relativistic quantum mechanics and classical field theory (other than electromagnetism) will have been covered previously, which means they will comprise half of the "quantum" field theory course, while the true quantum field theory will be squeezed into the last half.

An alternative is to start field theory one semester earlier: The courses of the first year, second semester are not really needed to start field theory (although some of the topics of second-semester quantum mechanics, like scattering theory, might be useful for second-semester field theory). Then field theory can be extended into a 3-semester course (or the third semester can be replaced with other advanced courses).

One way to cut the material to fit a one-year course is to omit Part Three, which can be left for a third semester on "advanced quantum field theory"; then the first semester (Part One) is classical while the second (Part Two) is quantum. Furthermore, the ordering of the chapters is somewhat flexible: The "flow" is indicated by the following "3D" plot:

		classical		$\rightarrow$	**quantum**	
		symmetry	*fields*		*quantize*	*loop*
lower spin	**Bose**	I	III		V	VII
$\searrow\swarrow$	$\downarrow$			IX	XI	
higher spin				X	XII	
	Fermi	II	IV		VI	VIII

where the 3 dimensions are spin ("j"), quantization ("$\hbar$"), and statistics ("s"): The three independent flows are down the page, to the right, and into the page. (The third dimension has been represented as perpendicular to the page, with "higher spin" in smaller type to indicate perspective, for legibility.) To present these chapters in the 1 dimension of time we have classified them as $j\hbar s$, but other orderings are possible:

$j\hbar s$: I II III IV V VI VII VIII IX X XI XII
$js\hbar$: I III V VII II IV VI VIII IX XI X XII
$\hbar js$: I II III IV IX X V VI XI XII VII VIII
$\hbar sj$: I II III IX IV X V XI VI XII VII VIII
$sj\hbar$: I III V VII IX XI II IV VI VIII X XII
$s\hbar j$: I III IX V XI VII II IV X VI XII VIII

(However, the spinor notation of II is used for discussing instantons in III, so some rearrangement would be required, except in the $j\hbar s$, $\hbar js$, and $\hbar sj$ cases.) For example, the first half of the course can cover all of the classical, and the second quantum, dividing Part Three between them ($\hbar js$ or $\hbar sj$). Another alternative ($js\hbar$) is a one-semester course on quantum field theory, followed by a semester on the Standard Model, and finishing with supergravity and strings. Although some of these (especially the first two) allow division of the course into one-semester courses, this should not be used as an excuse to treat such courses as complete: Any particle physics student who was content to sit through another entire year of quantum mechanics in graduate school should be prepared to take at least a year of field theory.

Notes for students

Field theory is a hard course. (If you don't think so, name me a harder one at this level.) But you knew as an undergraduate that physics was a hard major. Students who plan to do research in field theory will find the topic challenging; those with less enthusiasm for the topic may find it overwhelming. The main difference between field theory and lower courses is that it is not set in stone: There is much more variation in style and content among field theory courses than, e.g., quantum mechanics courses, since quantum mechanics (to the extent taught in courses) was pretty much finished in the 1920's, while field theory is still an active research topic, even though it has had many experimentally confirmed results since the 1940's. As a result, a field theory course has the flavor of research: There is no set of mathematically rigorous rules to solve any problem. Answers are not final, and should be treated as questions: One should not be satisfied with the solution of a problem, but consider it as a first step toward generalization. The student should not expect to capture all the details of field theory the first time through, since many of them are not yet fully understood by people who work in the area. (It is far more likely that instead you will discover details that you missed in earlier courses.) And one reminder: The *only* reason for lectures (including seminars and conferences) is for the attendees to ask questions (and not just in private), and there are no stupid questions (except for the infamous "How many questions are on the exam?"). Only half of teaching is the responsibility of the instructor.

Some students who have a good undergraduate background may want to begin graduate school taking field theory. That can be difficult, so you should be sure you have a good understanding of most of the following topics:

(1) Classical mechanics: Hamiltonians, Lagrangians, actions; Lorentz transformations; Poisson brackets

(2) Classical electrodynamics: Lagrangian for electromagnetism; Lorentz transformations for electromagnetic fields, 4-vector potential, 4-vector Lorentz force law; Green functions

(3) Quantum mechanics: coupling to electromagnetism; spin, SU(2), symmetries; Green functions for Schrödinger equation; Hilbert space, commutators, Heisenberg and Schrödinger pictures; creation and annihilation operators, statistics (bosons and fermions); JWKB expansion

It is not necessary to be familiar with all these topics, and most will be briefly reviewed, but if most of these topics are not familiar then there will not be enough time to catch up. A standard undergraduate education in these three courses is *not* enough.

Acknowledgments

I thank everyone with whom I have discussed field theory, especially Gordon Chalmers, Marc Grisaru, Marcelo Leite, Martin Roček, Jack Smith, George Sterman, and Peter van Nieuwenhuizen. More generally, I thank the human race, without whom this work would have been neither possible nor necessary.

December 20, 1999

Version 2

This update contains no new topics or subsections, but many small changes (amounting to a 10% increase in size): corrections, improved explanations, examples, (20% more) exercises, figures, references, cosmetics (including more color), and an expanded Outline and AfterMath. There are also a few small additions, such as a more fundamental explanation of causality and unitarity in quantum mechanics, and the use of Weyl scaling as a general method for spherical (as well as cosmological) solutions to Einstein's equations. It now TeX's with either ordinary TeX or pdftex. (Pdf figures can be created from ps with ghostscript; also available at my web site.)

September 19, 2002

SOME FIELD THEORY TEXTS

Traditional, leaning toward concepts

Canonically quantize QED and calculate, then introduce path integrals

1 S. Weinberg, *The quantum theory of fields*, 3 v. (Cambridge University, 1995, 1996, 2000) 609+489+419 pp.:
First volume QED; second volume contains many interesting topics; third volume supersymmetry. By one of the developers of the Standard Model.

2 M. Kaku, *Quantum field theory: a modern introduction* (Oxford University, 1993) 785 pp.:
Includes introduction to supergravity and superstrings.

3 C. Itzykson and J.-B. Zuber, *Quantum field theory* (McGraw-Hill, 1980) 705 pp. (but with lots of small print):
Emphasis on QED.

4 N.N. Bogoliubov and D.V. Shirkov, *Introduction to the theory of quantized fields*, 3rd ed. (Wiley, 1980) 620 pp.:
Ahead of its time (1st English ed. 1959): early treatments of path integrals, causality, background fields, and renormalization of general field theories; but before Yang-Mills and Higgs.

Traditional, leaning toward calculations

Emphasis on Feynman diagrams

5 M.E. Peskin and D.V. Schroeder, *An introduction to quantum field theory* (Perseus, 1995) 842 pp.:
Comprehensive; style similar to Bjorken and Drell.

6 B. de Wit and J. Smith, *Field theory in particle physics*, v. 1 (Elsevier Science, 1986) 490 pp.:
No Yang-Mills or Higgs (but wait till v. 2, due any day now...).

7 A.I. Akhiezer and V.B. Berestetskii, *Quantum electrodynamics* (Wiley, 1965) 868 pp.:
Numerous examples of QED calculations.

8 R.P. Feynman, *Quantum electrodynamics: a lecture note and reprint volume* (Perseus, 1961) 198 pp.:
Original treatment of quantum field theory as we know it today, but from mechanics; includes reprints of original articles (1949).

Modern, but somewhat specialized

Basics, plus thorough treatment of an advanced topic

9 J. Zinn-Justin, *Quantum field theory and critical phenomena*, 4th ed. (Clarendon, 2002) 1074 pp.:
First 1/2 is basic text, with interesting treatments of many topics, but no S-matrix examples or discussion of cross sections; second 1/2 is statistical mechanics.

10 G. Sterman, *An introduction to quantum field theory* (Cambridge University, 1993) 572 pp.:
First 3/4 can be used as basic text, including S-matrix examples; last 1/4 has extensive treatment of perturbative QCD, emphasizing factorization.

Modern, but basic: few S-matrix examples

Should be supplemented with a "QED/particle physics text"

11 L.H. Ryder, *Quantum field theory*, 2nd ed. (Cambridge University, 1996) 487 pp.:
Includes introduction to supersymmetry.

12 D. Bailin and A. Love, *Introduction to gauge field theory*, 2nd ed. (Institute of Physics, 1993) 364 pp.:
All the fundamentals.

13 P. Ramond, *Field theory: a modern primer*, 2nd ed. (Perseus, 1989) 329 pp.:
Short text on QCD: no weak interactions or Higgs.

Advanced topics

For further reading; including brief reviews of some standard topics

14 Theoretical Advanced Study Institute in Elementary Particle Physics (TASI) proceedings, University of Colorado, Boulder, CO (World Scientific):
Annual collection of summer school lectures on recent research topics.

15 W. Siegel, *Introduction to string field theory* (World Scientific, 1988), hep-th/0107094, 244 pp.:
Reviews lightcone, BRST, gravity, first-quantization, spinors, twistors, strings; besides, I like the author.

16 S.J. Gates, Jr., M.T. Grisaru, M. Roček, and W. Siegel, *Superspace: or one thousand and one lessons in supersymmetry* (Benjamin/Cummings, 1983), hep-th/0108200, 548 pp.:
Covers supersymmetry, spinor notation, lightcone, Stückelberg fields, gravity, Weyl scale, gauge fixing, background-field method, regularization, and anomalies; same author as previous, plus three other guys whose names sound familiar.

PART ONE: SYMMETRY

The first four chapters present a one-semester course on "classical field theory". Perhaps a more accurate description would be "everything you should know before learning quantum field theory". This is basically a study of global and local symmetries: Classical dynamics represents only a certain limit of quantum dynamics, and not the one usually emphasized, but most of the symmetries of classical physics survive quantization. The phenomenon of symmetry breaking, and the related mechanisms of mass generation, can also be seen at the classical level. In perturbative quantum field theory, classical field theory is simply the leading term in the perturbation expansion.

Continuous symmetry is one of the most fundamental and important concepts of physics. In the framework of an action principle (which is required in quantum physics), it is equivalent to conservation laws, which have been a cornerstone of physics since Newton. From a practical viewpoint, it simplifies calculations by relating different solutions to equations of motion, and allowing these equations to be written more concisely by treating independent degrees of freedom as a single entity. In particular, local ("gauge") symmetries, which allow independent transformations at each coordinate point, are basic to all the fundamental interactions: All the fundamental forces are mediated by particles described by Yang-Mills theory and its generalizations.

Symmetries are the result of a redundant, but useful, description of a theory. (Note that here we refer to symmetries of a theory, not of a solution to the theory.) For example, translation invariance says that only differences in position are measurable, not absolute position: We can't measure the position of the "origin". There are two ways to deal with this: (1) Choose an origin; i.e., make a "choice of coordinates". For example, place an object at the origin; i.e., choose the position of an object at a certain time to be the origin. (2) Work only in terms of differences of coordinates, which are "translationally invariant". Although the latter choice is more physical, the former is usually more convenient: The use of redundant variables, together with symmetry, often gives a simpler description of a theory. Another example is quantum mechanics, where the arbitrariness of the phase of the wave function can be considered a symmetry: Although quantum mechanics can be reformulated in terms of phase-invariant probabilities, currents, or density matrices instead of wave functions, and this can be useful for some purposes of exposing physical properties, formulating and solving the Schrödinger equation is simpler in terms of the wave function. The same applies to "local" symmetries, where there is an independent symmetry at each point of space and time: For example, quarks and gluons have a local "color" symmetry, and are not (yet) observed independently in nature, but are simpler objects in terms of which to describe strong interactions than the observed hadrons (protons, neutrons, etc.), which are described by color-invariant products of quark/gluon wave functions, in the same way that probabilities are phase-invariant products of wave functions. (Note that in quantum mechanics there is a subtle distinction between observed and observer that can obscure this symmetry if the observer is not invariant under it. This can always be avoided by choosing to define the observer as invariant: For example, the detection apparatus can be included as part of the quantum mechanical system, while the observer can be defined as some "remote" recorder, who may be abstracted as even being translationally invariant. In practice we are less precise, and abstract even the detection apparatus to be invariant: For example, we describe the scattering of particles in terms of the coordinates of only the particles, and deal with the origin problem as above in terms of just those coordinates.)

Note that "global" (time-, and usually space-independent) symmetries can eliminate a variable, but not its time derivative. For example, translation invariance allows us to fix (i.e., eliminate) the position of the center of mass of a system at some initial time, but not its time derivative, which is just the total momentum, whose conservation is a consequence of that same symmetry. A local symmetry, being time dependent, may allow the elimination of a variable at all times: The existence of this possibility depends on the dynamics, and will be discussed later.

Of particular interest are ways in which symmetries can be made manifest. Frequently in the literature "manifest" is used vacuously; a "manifest symmetry" is an obvious one: If you know the group, the representation under consideration doesn't need to be stated, but can be seen from just the notation. (In fact, one of the main uses of index notation is just to manifest the symmetry.) Formulations where global and local symmetries are manifest simplify calculations and their results, as well as clarifying their meaning.

One of the main uses of manifest symmetry is rarely needing to explicitly perform a specific symmetry transformation. For example, one might need to examine a relativistic problem in different Lorentz frames. Rather than starting with a description of the problem in one frame, and then explicitly transforming to another, it is much simpler to start with a manifestly covariant description, make one choice of frame, then make another choice of frame. One then never uses the messy square roots of the familiar Lorentz contraction factors (although they may appear at the end from kinematic constraints). A more extreme example is the corresponding situation for local symmetries, where such transformations are intractable in general, and one always starts with the manifestly covariant form.

I. GLOBAL

In the first chapter we study symmetry in general, concentrating primarily on spacetime symmetries, but also discussing general properties that will have other applications in the following chapter.

A. COORDINATES

In this section we discuss the Poincaré (and conformal) group as coordinate transformations. This is the simplest way to represent it on the physical world. In later sections we find general representations by adding spin.

1. Nonrelativity

We begin by reviewing some general properties of symmetries, including as an example the symmetry group of nonrelativistic physics. In the Hamiltonian approach to mechanics, both symmetries and dynamics can be expressed conveniently in terms of a "bracket": the Poisson bracket for classical mechanics, the commutator for quantum mechanics. In this formulation, the fundamental variables (operators) are some set of coordinates and their canonically conjugate momenta, as functions of time. The (Heisenberg) operator approach to quantum mechanics then is related to classical mechanics by identifying the semiclassical limit of the commutator as the Poisson bracket: For any functions A and B of p and q, the quantum mechanical commutator is

$$AB - BA = -i\hbar \left(\frac{\partial A}{\partial p_m} \frac{\partial B}{\partial q^m} - \frac{\partial B}{\partial p_m} \frac{\partial A}{\partial q^m} \right) + \mathcal{O}(\hbar^2)$$

In other words, the true classical limit of $AB - BA$ is zero, since classically functions commute; thus the semiclassical limit is defined by

$$\lim_{\hbar \to 0} \left[\frac{1}{\hbar}(AB - BA) \right]$$

(which is really a derivative with respect to $\hbar$). We therefore define the bracket for the two cases by

$$[A, B] \equiv \begin{cases} -i \left(\dfrac{\partial A}{\partial p_m} \dfrac{\partial B}{\partial q^m} - \dfrac{\partial B}{\partial p_m} \dfrac{\partial A}{\partial q^m} \right) & \text{semiclassically} \\ AB - BA & \text{quantum mechanically} \end{cases}$$

The semiclassical definition of the bracket then can be applied to classical physics (where it was originally discovered). Classically A and B are two arbitrary functions of the coordinates q and momenta p; in quantum mechanics they can be arbitrary operators. We have included an "i" in the classical normalization so the two agree in the semiclassical limit. We generally use (natural/Planck) units $\hbar = 1$, so mass is measured as inverse length, etc.; when we do use an explicit $\hbar$, it is a dimensionless parameter, and appears only for defining Jeffries-Wentzel-Kramers-Brillouin (JWKB) expansions or (semi)classical limits.

Our indices may appear either as subscripts or superscripts, with preferences to be explained later: For nonrelativistic purposes we treat them the same. We also use the Einstein summation convention, that any repeated index in a product is summed over ("contracted"); usually we contract a superscript with a subscript:

$$A^m B_m \equiv \sum_m A^m B_m$$

The definition of the bracket is equivalent to using

$$[p_m, q^n] = -i\delta_m^n$$

(where δ_m^n is the "Kronecker delta function": 1 if $m = n$, 0 if $m \neq n$) together with the general properties of the bracket

$$[A, B] = -[B, A], \qquad [A, B]^\dagger = -[A^\dagger, B^\dagger]$$

$$[[A, B], C] + [[B, C], A] + [[C, A], B] = 0$$

$$[A, BC] = [A, B]C + B[A, C]$$

The first set of identities exhibit the antisymmetry of the bracket; next are the "Jacobi identities". In the last identity the ordering is important only in the quantum mechanical case: In general, the difference between classical and quantum mechanics comes from the fact that in the quantum case operator reordering after taking the commutator results in multiple commutators.

Infinitesimal symmetry transformations are then written as

$$\delta A = i[G, A], \qquad A' = A + \delta A$$

where G is the "generator" of the transformation. More explicitly, infinitesimal generators will contain infinitesimal parameters: For example, for translations we have

$$G = \epsilon^i p_i \quad \Rightarrow \quad \delta x^i = i[G, x^i] = \epsilon^i, \quad \delta p_i = 0$$

where ϵ^i are infinitesimal numbers.

The most evident physical symmetries are those involving spacetime. For nonrelativistic particles, these symmetries form the "Galilean group": For the free particle, those infinitesimal transformations are linear combinations of

$$M = m, \quad P_i = p_i, \quad J_{ij} = x_{[i}p_{j]} \equiv x_i p_j - x_j p_i, \quad E = H = \frac{p_i^2}{2m}, \quad V_i = mx_i - p_i t$$

in terms of the position x^i ($i = 1,2,3$), momenta p_i, and (nonvanishing) mass m, where $[ij]$ means to antisymmetrize in those indices, by summing over all permutations (just two in this case), with plus signs for even permutations and minus for odd. (In three spatial dimensions, one often writes $J_i = \frac{1}{2}\epsilon_{ijk}J_{jk}$ to make J into a vector. This is a peculiarity of three dimensions, and will lose its utility once we consider relativity in four spacetime dimensions.) These transformations are the space translations (momentum) P, rotations (angular momentum — just orbital for the spinless case) J, time translations (energy) E, and velocity transformations ("Galilean boosts") V. (The mass M is not normally associated with a symmetry, and is not conserved relativistically.)

Exercise IA1.1

Let's examine the Galilean group more closely. Using just the relations for $[x,p]$ and $[A,BC]$ (and the antisymmetry of the bracket):

a Find the action on x_i of each kind of infinitesimal Galilean transformation.

b Show that the nonvanishing commutation relations for the generators are

$$[J_{ij}, P_k] = i\delta_{k[i}P_{j]}, \qquad [J_{ij}, V_k] = i\delta_{k[i}V_{j]}, \qquad [J_{ij}, J^{kl}] = i\delta_{[i}^{[k}J_{j]}{}^{l]}$$

$$[P_i, V_j] = -i\delta_{ij}M, \qquad [H, V_i] = -iP_i$$

For more than one free particle, we introduce an m, x^i, and p_i for each particle (but the same t), and the generators are the sums over all particles of the above expressions. If the particles interact with each other the expression for H is modified, in such a way as to preserve the commutation relations. If the particles also interact with dynamical fields, field-dependent terms must be added to the generators. (External, nondynamical fields break the invariance. For example, a particle in a Coulomb potential is not translation invariant since the potential is centered about some point.) Note that for N particles there are 3N coordinates describing the particles, but still only 3 translations: The particles interact in the same 3-dimensional space. We can use translational invariance to fix the position of any one particle at a given time, but not the rest: The differences in position are translationally invariant. On the other hand, it is often useful not to fix the position of any particle, since keeping this invariance (and the corresponding redundant variables) allows all particles to be treated equally. We might also consider using the differences of positions themselves as the variables, allowing a symmetric treatment of the particles in terms of translationally invariant variables: However, this would require applying constraints on the variables, since there are 3N(N−1)/2 differences, of which only 3(N−1) are independent. We will find similar features later for "local" invariances: In general, the most convenient description of a theory is with the invariance; the invariance can then be fixed, or invariant combinations of variables used, appropriately for the particular application.

The rotations (or at least their "orbital" parts) and space translations are examples of coordinate transformations. In general, generators of coordinate transformations are of the form

$$G = \lambda^i(x)p_i \quad \Rightarrow \quad \delta\phi(x) = i[G,\phi] = \lambda^i \partial_i \phi$$

where $\partial_i = \partial/\partial x^i$ and $\phi(x)$ is a "scalar field" (or "spin-0 wave function"), a function of only the coordinates.

In classical mechanics, or quantum mechanics in the Heisenberg picture, time development also can be expressed in terms of the Hamiltonian using the bracket:

$$\frac{d}{dt}A = \left[\frac{\partial}{\partial t} + iH, A\right] = \frac{\partial}{\partial t}A + i[H,A]$$

(The middle expression with the commutator of $\partial/\partial t$ makes sense only in the quantum case, and is not defined for the Poisson bracket.) Again, this general relation is equivalent to the special cases, which in the classical limit are Hamilton's equations of motion:

$$\frac{dq^m}{dt} = i[H,q^m] = \frac{\partial H}{\partial p_m}, \qquad \frac{dp_m}{dt} = i[H,p_m] = -\frac{\partial H}{\partial q^m}$$

The Hamiltonian has no explicit time dependence in the absence of time-dependent nondyamical fields (external potentials whose time dependence is fixed by hand, rather than by introducing the fields and their conjugate variables into the Hamiltonian). Consequently, time development is itself a symmetry: Time translations are generated by the Hamiltonian; the $\partial/\partial t$ term in d/dt term can be dropped when acting on operators without explicit time dependence.

Invariance of the theory under a symmetry means that the equations of motion are unchanged under the transformation:

$$\left(\frac{dA}{dt}\right)' = \frac{dA'}{dt}$$

To apply our above translation of infinitesimal transformations into bracket language, we define $\delta(d/dt)$ by

$$\delta\left(\frac{d}{dt}A\right) = \delta\left(\frac{d}{dt}\right)A + \frac{d}{dt}\delta A$$

In the quantum case we can write

$$\delta\left(\frac{d}{dt}\right) = \left[\left[iG, \frac{\partial}{\partial t} + iH\right], \quad \right]$$

which follows from the Jacobi identity using $B = iG$ and $C = \partial/\partial t + iH$, and inserting A into the blank spaces of the commutators above. (The classical case can be treated similarly, except that the time derivatives are not written as brackets.) We then find that the generator of a symmetry transformation is conserved (constant), since

$$0 = \delta\left(\frac{d}{dt}\right) = \left[-i\frac{\partial G}{\partial t} - [G, H], \quad \right] = -i\left[\frac{dG}{dt}, \quad \right]$$

Exercise IA1.2

Show that the generators of the Galilean group are conserved:

a Use the relation $d/dt = \partial/\partial t + i[H, \ ]$ for the hamiltonian H of a free particle.

b Solve the equations of motion for $x(t)$ and $p(t)$ in terms of initial conditions, and substitute into the expression for the generators to give an independent derivation of their time independence.

Note that in the case where the Galilean symmetry persists for interacting multiparticle systems, (total) mass is conserved. In particular, invariance under translations and velocity transformations implies mass conservation.

In the cases where time dependence is not involved, symmetries can be treated in almost exactly the same way either classically or quantum mechanically using the corresponding bracket (Poisson or commutator), by using the properties that they have in common. In particular, the fact that a symmetry generator $G = \lambda^m(x)p_m$ is conserved means that we can solve for a component of p in terms of the constant G, and substitute the result into the remaining equations of motion, and that the conjugate to that component doesn't appear in H. For example, translation invariance of a potential in a particular direction means that component of the momentum is a constant ($dp_1/dt = -\partial H/\partial q^1 = 0$), rotational invariance about some axis means that component of angular momentum is a constant ($dJ/dt = -\partial H/\partial\theta = 0$), etc.

2. Fermions

As we learned in our quantum mechanics course, two particles of the same type are indistinguishable. Furthermore, while an arbitrary number of *bosons* (particles satisfying Bose-Einstein statistics) can each exist in the same one-particle state, only one (or zero) *fermions* can exist in the same one-particle state. (For example, we can have a state consisting of 17 photons each of the same momentum and each of the same polarization, and we can't tell which is which, but we can only have 1 electron in such a state.) In terms of wave functions, e.g., a 2-particle wave function, made from 1-particle wave functions of the form $\psi_i(x)$ (where x labels the spatial position and i other properties), we conveniently define

$$\begin{aligned} bosons: &\quad \psi_{ii'}(x,x') = +\psi_{i'i}(x',x) \\ fermions: &\quad \psi_{ii'}(x,x') = -\psi_{i'i}(x',x) \end{aligned}$$

For $x = x'$ and $i = i'$ the signs (which could be phases, but are chosen real for convenience) are chosen so $\psi_{ii}(x,x)$ vanishes for fermions but not necessarily for bosons, so no 2 fermions are in the same state. For other cases the relation avoids double counting for the 2 particles being switched; the signs are arbitrary, but are chosen consistently with the previous case so that the relation is local. The symmetry of wave functions for bosons and antisymmetry for fermions corresponds to operators that commute for bosons and anticommute for fermions (or for properties associated with fermions).

As we know experimentally, and we will see follows from relativistic field theory, particles with half-integral spins obey Fermi-Dirac statistics. Let's therefore consider the classical limit of fermions: This will lead to generalizations of the concepts of brackets and coordinates. Bosons (more generally, bosonic operators) obey commutation relations, such as $[x,p] = i\hbar$; in the classical limit they just commute. Fermions obey anticommutation relations, such as $\{\zeta, \zeta^\dagger\} = \hbar$ for a single fermionic harmonic oscillator, where

$$\{A, B\} = AB + BA$$

is the anticommutator. So, in the truly classical (not semiclassical) limit they anticommute, $\zeta\zeta^\dagger + \zeta^\dagger\zeta = 0$. Actually, the simplest case is a single real (hermitian) fermion: Quantum mechanically, or semiclassically, we have

$$\hbar = \{\xi, \xi\} = 2\xi^2$$

while classically $\xi^2 = 0$. There is no analog for a single boson: $[x,x] = x^2 - x^2 = 0$. This means that classical fermionic fields must be "anticommuting": Two such objects get a minus sign when pushed past each other. As a result, the product of two fermionic quantities is bosonic, while fermionic times bosonic gives fermionic.

Exercise IA2.1

Show

$$[B,C] = [A,D] = 0 \quad \Rightarrow \quad [AB,CD] = \tfrac{1}{2}\{A,C\}[B,D] + \tfrac{1}{2}[A,C]\{B,D\}$$

To work with wave functions that are functions of anticommuting numbers, we first must understand how to define general properties of functions of anticommuting variables. For instance, given a single anticommuting variable ψ, we need to be able to Taylor expand functions in ψ, e.g., to find a basis for the states. We therefore have an anticommuting derivative $\partial/\partial\psi$, satisfying

$$\left(\frac{\partial}{\partial\psi}\right)^2 = 0$$

from either anticommutativity or the fact functions of ψ terminate at first order in ψ. We also need a ψ integral to define the inner product; indefinite integration turns out to be enough. The most important property of the integral is integration by parts; then, when acting on any function of ψ,

$$\int d\psi\, \frac{\partial}{\partial\psi} = 0 \quad \Rightarrow \quad \int d\psi = \frac{\partial}{\partial\psi}$$

where the normalization is fixed for convenience. This also implies a definition of the "(anticommuting) Dirac delta function",

$$\delta(\psi) = \psi$$

which satisfies

$$\int d\psi'\ \delta(\psi' - \psi) f(\psi') = f(\psi)$$

for any function f.

Exercise IA2.2

Prove this is the most general possibility for anticommuting integration by considering action of integration and differentiation on the most general function of ψ (which has only two terms).

In general, when Taylor expanding a function of anticommuting variables we must preserve the statistics: If we Taylor expand a quantity that is defined to be commuting (bosonic), then the coefficients of even powers of anticommuting variables will also be commuting, while the coefficients of odd powers will be anticommuting (fermionic), to maintain the commuting nature of that term (the product of the variables and coefficient). Similarly, when expanding an anticommuting quantity the coefficients of even powers will also be anticommuting, while for odd powers it will be commuting.

We can now consider operators that depend on both commuting (ϕ^m) and anticommuting (ψ^μ) classical variables,

$$\Phi^M = (\phi^m, \psi^\mu)$$

Classically they satisfy the "graded" commutation relations (anticommutation if both elements are fermionic, commutation otherwise), not to be confused with the Poisson bracket,

$$\textit{classically} \qquad [\Phi^M, \Phi^N\} = 0 \ : \qquad \phi^m\phi^n - \phi^n\phi^m = \phi^m\psi^\nu - \psi^\nu\phi^m = \psi^\mu\psi^\nu + \psi^\nu\psi^\mu = 0$$

This relation is then generalized to the graded quantum mechanical commutator or Poisson bracket by

$$[\Phi^M, \Phi^N\} = \hbar\Omega^{MN}, \qquad \Omega^{MN}\Omega_{PN} = \delta_P^M$$

where Ω is constant, hermitian, and "graded antisymmetric":

$$\Omega_{(MN]} = 0 : \qquad \Omega_{(mn)} = \Omega_{[\mu\nu]} = \Omega_{m\nu} + \Omega_{\nu m} = 0$$

where $[\mu\nu]$ is the difference of the two orderings, as above, while $(\mu\nu)$ is the sum. For the standard normalization of canonically conjugate pairs of bosons

$$\phi^m = \phi^{i\alpha} = (q^i, p^i)$$

and self-conjugate fermions, we choose

$$\Omega^{\mu\nu} = \delta^{\mu\nu}; \qquad \Omega^{i\alpha,j\beta} = \delta^{ij}C^{\alpha\beta}, \qquad C^{\alpha\beta} = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$$

Because of signs resulting from ordering anticommuting quantities, we define derivatives unambiguously by their action from the left:

$$\frac{\partial}{\partial\Phi^M}\Phi^N = \delta_M^N$$

The general Poisson bracket then can be written as

$$\textit{semiclassically} \qquad [A, B\} \equiv -A\frac{\overleftarrow{\partial}}{\partial\Phi^M}\Omega^{NM}\frac{\partial}{\partial\Phi^N}B$$

Since derivatives are normally defined to act from the left, there is a minus sign from pushing the first derivative to the left if A and that particular component of $\partial/\partial\Phi^M$ are both fermionic.

Exercise IA2.3

Let's examine some properties of fermionic oscillators:

a For a single set of harmonic oscillators we have

$$\{a, a^\dagger\} = 1, \qquad \{a, a\} = \{a^\dagger, a^\dagger\} = 0$$

Show that the "number operator" $a^\dagger a$ has the property

$$\{a, e^{i\pi a^\dagger a}\} = 0$$

(Hint: Since this system has only 2 states, the easiest way is to check the action on those states.)

b Define eigenstates of the annihilation operator ("coherent states") by

$$a|\zeta\rangle = \zeta|\zeta\rangle$$

where ζ is anticommuting. Show that this implies

$$a^\dagger|\zeta\rangle = -\frac{\partial}{\partial\zeta}|\zeta\rangle, \quad |\zeta\rangle = e^{\zeta a^\dagger}|0\rangle, \quad e^{\zeta' a^\dagger}|\zeta\rangle = |\zeta+\zeta'\rangle, \quad x^{a^\dagger a}|\zeta\rangle = |x\zeta\rangle,$$

$$\langle\zeta|\zeta'\rangle = e^{\zeta^*\zeta'}, \quad 1 = \int d\zeta^* d\zeta \; e^{-\zeta^*\zeta}|\zeta\rangle\langle\zeta|$$

Define wave functions in this space, $\Psi(\zeta) = \langle\zeta|\Psi\rangle$. Taylor expand them in ζ, and compare this to the usual two-component representation using $|0\rangle$ and $a^\dagger|0\rangle$ as a basis.

c Define the "supertrace" by

$$str(A) = \int d\zeta^* d\zeta \; e^{-\zeta^*\zeta}\langle\zeta|A|\zeta\rangle$$

Find the relation between any operator in this space and a 2×2 matrix, and find the expression for the supertrace in terms of this matrix.

d For *two* sets of fermionic oscillators, we define

$$\{a_1, a^\dagger{}_1\} = \{a_2, a^\dagger{}_2\} = 1, \qquad other\ \{\ ,\ \} = 0$$

Show that the new operators

$$\tilde{a}_1 = a_1, \qquad \tilde{a}_2 = e^{i\pi a^\dagger{}_1 a_1} a_2$$

(and their Hermitian conjugates) are equivalent to the original ones except that one set of the new oscillators *commutes* (not anticommutes) with the other ($[\tilde{a}_1, \tilde{a}^\dagger{}_2] = 0$, etc.), even though each set satisfies the same anticommutation relations with itself ($\{\tilde{a}_1, \tilde{a}^\dagger{}_1\} = 1$, etc.). Thus, choice of statistics is relevant only for particles in the same state: at most one fermion, but unlimited bosons. (This change of oscillator basis is called a "Klein transformation". It can be useful for discrete sets of oscillators, but not for those labeled by a continuous parameter, because of the discontinuity in the commutation relations when the two labels are equal.)

Exercise IA2.4

Repeat exercise IA2.3 for the bosonic oscillator ($[a, a^\dagger] = 1$), where the Hilbert space is infinite-dimensional, paying attention to signs, interchanging commutators with anticommutators where necessary, etc. Show that the analog of part **c** defines the ordinary trace.

3. Lie algebra

Since the same symmetries can be expressed in terms of different kinds of brackets for classical and quantum theories, it can be useful to work with just those properties that the Poisson bracket and commutator have in common, i.e., those that involve only the bracket of two operators, not just their ordinary product:

$$[\alpha A + \beta B, C] = \alpha[A, C] + \beta[B, C] \quad \text{for numbers } \alpha, \beta \qquad \text{(distributivity)}$$

$$[A, B] = -[B, A] \qquad \text{(antisymmetry)}$$

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0 \qquad \text{(Jacobi identity)}$$

with similar expressions (differing only by signs) for anticommutators or mixed commutators and anticommutators.

Exercise IA3.1

Find the generalizations of the Jacobi identity using also anticommutators, corresponding to the cases where 2 or 3 of the objects involved are considered as fermionic instead of bosonic.

These properties also give an abstract definition of a form of multiplication, the "Lie bracket", which defines a "Lie algebra". (The first property is true of algebras in general.) Other Lie brackets include those defined by another, associative, form of multiplication, such as matrix multiplication, or operator (infinite matrix) multiplication as in quantum mechanics: In those cases we can write $[A, B] = AB - BA$, and use the usual properties of multiplication (distributivity and associativity) to derive the properties of the Lie bracket. (Another familiar example in physics is the "cross" product for three-vectors; however, this can also be expressed in terms of matrix multiplication.) The most important use of Lie algebras for physics is for describing (continuous) infinitesimal transformations, especially those describing symmetries.

Exercise IA3.2

Using only the commutation relations of the generators of the Galilean group (exercise IA1.1), check all the Jacobi identities.

For describing transformations, we can also think of the bracket as a derivative: The "Lie derivative" of B with respect to A is defined as

$$\mathcal{L}_A B = [A, B]$$

As a consequence of the properties of the Lie bracket, this derivative satisfies the usual properties of a derivative, including the Leibniz rule. (In fact, for coordinate transformations the Lie derivative is really a derivative with respect to the coordinates.)

We can now define finite transformations by exponentiating infinitesimal ones:

$$A' \approx (1 + i\epsilon\mathcal{L}_G)A \quad \Rightarrow \quad A' = \lim_{\epsilon\to 0}(1 + i\epsilon\mathcal{L}_G)^{1/\epsilon}A = e^{i\mathcal{L}_G}A$$

In cases where we have $[A, B] = AB - BA$, we can also write

$$e^{i\mathcal{L}_G}A = e^{iG}Ae^{-iG}$$

This follows from replacing G on both sides with αG and taking the derivative with respect to α, to see that both satisfy the same differential equation with the same initial condition. We then can recognize this as the way transformations are performed in quantum mechanics: A linear transformation that preserves the Hilbert-space inner product must be unitary, which means it can be written as the exponential of an antihermitian operator.

Just as infinitesimal transformations define a Lie algebra with elements A, finite ones define a "Lie group" with elements

$$g = e^{iG}$$

The multiplication law of two group elements follows from the fact the product of two exponentials can be expressed in terms of multiple commutators:

$$e^A e^B = e^{A+B+\frac{1}{2}[A,B]+...}$$

We now have the mathematical properties that define a group, namely:

(1) a product, so that for two group elements g_1 and g_2, we can define $g_1 g_2$, which is another element of the group (closure),

(2) an identity element, so $gI = Ig = g$,

(3) an inverse, where $gg^{-1} = g^{-1}g = I$, and

(4) associativity, $g_1(g_2 g_3) = (g_1 g_2)g_3$.

In this case the identity is $1 = e^0$, while the inverse is $(e^A)^{-1} = e^{-A}$.

Since the elements of a Lie algebra form a vector space (we can add them and multiply by numbers), it's useful to define a basis:

$$G = \alpha^i G_i \quad \Rightarrow \quad g = e^{i\alpha^i G_i}$$

The parameters α^i then also give a set of (redundant) coordinates for the Lie group. (Previously they were required to be infinitesimal, for infinitesimal transformations; now they are finite, but may be periodic, as determined by topological considerations that we will mostly ignore.) Now the multiplication rules for both the algebra and the group are given by those of the basis:

$$[G_i, G_j] = -if_{ij}{}^k G_k$$

for the ("structure") constants $f_{ij}{}^k = -f_{ji}{}^k$, which define the algebra/group (but are ambiguous up to a change of basis). They satisfy the Jacobi identity

$$[[G_{[i}, G_j], G_{k]}] = 0 \quad \Rightarrow \quad f_{[ij}{}^l f_{k]l}{}^m = 0$$

A familiar example is SO(3) (SU(2)), 3D rotations, where $f_{ij}{}^k = \epsilon_{ijk}$ if we use $G_i = \frac{1}{2}\epsilon_{ijk}J_{jk}$.

Another useful concept is a "subgroup": If some subset of the elements of a group also form a group, that is called a "subgroup" of the original group. In particular, for a Lie group the basis of that subgroup will be a subset of some basis for the original group. For example, for the Galilean group J_{ij} generate the rotation subgroup.

Exercise IA3.3

Let's examine the subgroup of the Galilean group describing (spatial) coordinate transformations — rotations and spatial translations:

a Show that the infinitesimal transformations are given by

$$\delta x^i = x^j \epsilon_j{}^i + \hat{\epsilon}^i, \qquad \epsilon_{ij} = -\epsilon_{ji}$$

where the ϵ's are constants.

b Exponentiate to find the finite transformations

$$x'^i = x^j \Lambda_j{}^i + \hat{\Lambda}^i$$

c Show that $\Lambda_i{}^j$ must satisfy

$$\Lambda_i{}^k \Lambda_j{}^l \delta_{kl} = \delta_{ij}$$

both to preserve the scalar product, and as a consequence of exponentiating. (Hint: Use matrix notation, and find the equivalent relation between Λ and Λ^{-1}.)

d Show that the last equation implies $det\ \Lambda = \pm 1$, while exponentiating can give only $det\ \Lambda = 1$ (since $+1$ can't change continuously to -1). What is the physical interpretation of a transformation with $det\ \Lambda = -1$? (Hint: Consider a simple example.)

These results can be generalized to include anticommutators: When some of the basis elements G_i are fermionic, the corresponding parameters α^i are anticommuting numbers, the structure constants are defined by $[G_i, G_j\}$, etc.. Then $G = \alpha^i G_i$ is bosonic term by term, as is g, so bosons transform into bosons and fermions into fermions, but Taylor expansion in the α's will have both bosonic and fermionic coefficients. (For example, for $\delta A = \epsilon B$, if A is bosonic, then so is ϵB, but if also ϵ is fermionic, then B will also be fermionic.)

For some purposes it is more convenient to absorb the "i" in the infinitesimal transformation into the definition of the generator:

$$G \to -iG \quad \Rightarrow \quad \delta A = [G, A] = \mathcal{L}_G A, \quad g = e^G, \quad [G_i, G_j] = f_{ij}{}^k G_k$$

This affects the reality properties of G: In particular, if g is unitary ($gg^\dagger = I$), as usually required in quantum mechanics, $g = e^{iG}$ makes G hermitian ($G = G^\dagger$), while $g = e^G$ makes G antihermitian ($G = -G^\dagger$). In some cases anithermiticity can be an advantage: For example, for translations we would then have $P_i = \partial_i$ and for rotations $J_{ij} = x_{[i}\partial_{j]}$, which is more convenient since we know the i's in these (and any) coordinate transformations must cancel anyway. On the other hand, the U(1) transformations of electrodynamics (on the wave function for a charged particle) are just phase transformations $g = e^{i\theta}$ (where θ is a real number), so clearly we want the explicit i; then the only generator has the representation $G_i = 1$. In general we'll find that for our purposes absorbing the i's into the generators is more convenient for just spacetime symmetries, while explicit i's are more convenient for internal symmetries.

4. Relativity

The Hamiltonian approach singles out the time coordinate. In relativistic theories time can be treated on equal footing with space, and it is useful to take advantage of this fact, so that the full Poincaré invariance is manifest. So, we treat the time t and spatial position x^i together as a four-vector (or D-vector in D$-$1 space and 1 time dimension)

$$x^m = (x^0, x^i) = (t, x^i)$$

where $m = 0, 1, ..., 3$ (or D$-$1), $i = 1, 2, 3$. Since the energy E and three-momentum p^i are canonically conjugate to them,

$$[p^i, x^j] = -i\delta^{ij}, \qquad [E, t] = +i$$

we define the 4-momentum as

$$p^m = (E, p^i) = \eta^{mn} p_n, \qquad p_m = \eta_{mn} p^n; \qquad [p^m, x^n] = -i\eta^{mn}, \qquad [p_m, x^n] = -i\delta_m^n$$

where we raise and lower indices with the "Minkowski metric", in an "orthonormal basis",

$$\eta_{mn} = \begin{array}{c c} & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \qquad \Rightarrow \qquad p_0 = -p^0 = -E$$

in four spacetime dimensions, with obvious generalizations to higher dimensions. (Sometimes the metric with signs $+ - --$ is used; we prefer $- + ++$ because it is more convenient for quantum calculations. The numbers of positive and negative eigenvalues of an invertible matrix is known as its "signature".) Therefore, we now distinguish upper and lower indices in general: At least for position and momentum, the upper-indexed x^m and p^m have the usual physical interpretation (so x_m and p_m have extra signs). This is consistent with our previous nonrelativistic notation, since 3-vector indices do not change sign upon raising or lowering.

Of course, we could have done that much nonrelativistically. Relativity is a symmetry of kinematics and dynamics: In particular, a free, spinless, relativistic particle is completely described by the constraint

$$p^2 + m^2 = 0$$

where we define the covariant square

$$p^2 = p^m p_m = p^m p^n \eta_{mn} = -(p^0)^2 + (p^1)^2 + (p^2)^2 + (p^3)^2$$

(The square of p on the left should not be confused with the second component of p on the right.) Our relativistic symmetry must leave this constraint invariant: Thus the metric defines the norm of a vector (and an invariant inner product). Therefore, to preserve Lorentz invariance it is important that we contract only an upper index with a lower index. For similar reasons, we have

$$\partial_m = \frac{\partial}{\partial x^m}, \qquad \partial_m x^n = \delta_m^n$$

so quantum mechanically $p_m = -i\partial_m$.

Unlike the positive-definite nonrelativistic norm of a 3-vector V^i, for an arbitrary 4-vector V^m we can have

$$V^2 \left\{ \begin{matrix} < \\ = \\ > \end{matrix} \right\} 0 : \qquad \begin{cases} timelike \\ lightlike/null \\ spacelike \end{cases}$$

In particular, the 4-momentum is timelike for massive particles ($m^2 > 0$) and lightlike for massless ones (while "tachyons", with spacelike momenta and $m^2 < 0$, do not exist, for reasons that are most clear from quantum field theory).

The quantum mechanics will be described later, but the result is that this constraint can be used as the wave equation. The main qualitative distinction from the nonrelativistic case in the constraint

$$\begin{aligned} nonrelativistic: &\quad -2mE + \vec{p}^{\,2} = 0 \\ relativistic: &\quad -E^2 + m^2 + \vec{p}^{\,2} = 0 \end{aligned}$$

is that the equation for the energy $E \equiv p^0$ is now quadratic, and thus has two solutions:

$$p^0 = \pm\omega, \qquad \omega = \sqrt{(p^i)^2 + m^2}$$

Later we'll see how the second solution is interpreted as an "antiparticle". We also use (natural/Planck) units $c = 1$, so length and duration are measured in the same units; c then appears only as a parameter for defining nonrelativistic expansions and limits.

The translations and Lorentz transformations make up the Poincaré group, the symmetry that defines special relativity. (The Lorentz group in D−1 space and 1 time dimension is the "orthogonal" group "O(D−1,1)". The "proper" Lorentz group "SO(D−1,1)", where the "S" is for "special", transforms the coordinates by a matrix whose determinant is 1. The Poincaré group is ISO(D−1,1), where the "I" stands for "inhomogeneous".) For the spinless particle they are generated by coordinate transformations $G_I = (P_a, J_{ab})$:

$$P_a = p_a, \quad J_{ab} = x_{[a}p_{b]}$$

(where also $a, b = 0, ..., 3$). Then the fact that the physics of the free particle is invariant under Poincaré transformations is expressed as

$$[P_a, p^2 + m^2] = [J_{ab}, p^2 + m^2] = 0$$

Writing an arbitrary infinitesimal transformation as a linear combination of the generators, we find

$$\delta x^m = x^n \epsilon_n{}^m + \hat{\epsilon}^m, \qquad \epsilon_{mn} = -\epsilon_{nm}$$

where the ϵ's are constants. Note that antisymmetry of ϵ_{mn} does not imply antisymmetry of $\epsilon_m{}^n = \epsilon_{mp}\eta^{pn}$, because of additional signs. (Similar remarks apply to J_{ab}.) Exponentiating to find the finite transformations, we have

$$x'^m = x^n \Lambda_n{}^m + \hat{\Lambda}^m, \qquad \Lambda_m{}^p \Lambda_n{}^q \eta_{pq} = \eta_{mn}$$

The same Lorentz transformations apply to p^m, but the translations do not affect it. The condition on Λ follows from preservation of the Minkowski norm (or inner product), but it is equivalent to the antisymmetry of $\epsilon_m{}^n$ by exponentiating $\Lambda = e^\epsilon$ (compare exercise IA3.3).

Since $dx^a p_a$ is invariant under the coordinate transformations defined by the Poisson bracket (the chain rule, since effectively $p_a \sim \partial_a$), it follows that the Poincaré invariance of p^2 is equivalent to the invariance of the line element

$$ds^2 = -dx^m dx^n \eta_{mn}$$

which defines the "proper time" s. Spacetime with this indefinite metric is called "Minkowski space", in contrast to the "Euclidean space" with positive definite metric used to describe nonrelativistic length measured in just the three spatial dimensions. (The signature of the metric is thus the numbers of space and time dimensions.)

Exercise IA4.1

For general variables (q^m, p_m) and generator G, show from the definition of the Poisson bracket that

$$\delta(dq^m p_m) = -d\left(G - p_m \frac{\partial G}{\partial p_m}\right)$$

and that this vanishes for any coordinate transformation.

For the massive case, we also have

$$p^a = m\frac{dx^a}{ds}$$

For the massless case $ds = 0$: Massless particles travel along lightlike lines. However, we can define a new parameter τ such that

$$p^a = \frac{dx^a}{d\tau}$$

is well-defined in the massless case. In general, we then have

$$s = m\tau$$

While this fixes $\tau = s/m$ in the massive case, in the massless case it instead restricts $s = 0$. Thus, proper time does not provide a useful parametrization of the world line of a classical massless particle, while τ does: For any piece of such a line, $d\tau$ is given in terms of (any component of) p^a and dx^a. Later we'll see how this parameter appears in relativistic classical mechanics, and is useful for quantum mechanics and field theory.

Exercise IA4.2

Starting from the usual Lorentz force law for a *massive* particle in terms of proper time s (which doesn't apply to $m = 0$), rewrite it in terms of τ to find a form which can apply to $m = 0$.

Exercise IA4.3

The relation between x and p is closely related to the Poincaré conservation laws:

a Show that

$$dP_a = dJ_{ab} = 0 \quad \Rightarrow \quad p_{[a}dx_{b]} = 0$$

and use this to prove that conservation of P and J imply the existence of a parameter τ such that $p^a = dx^a/d\tau$.

b Consider a multiparticle system (but still without spin) where some of the particles can interact only when at the same point (i.e., by collision; they act as free particles otherwise). Define $P_a = \sum_I p_a^I$ and $J_{ab} = \sum_I x_{[a}^I p_{b]}^I$ as the sum of the individual momenta and angular momenta (where we label the particle with "I"). Show that momentum conservation implies angular momentum conservation,

$$\Delta P_a = 0 \quad \Rightarrow \quad \Delta J_{ab} = 0$$

where "Δ" refers to the change from before to after the collision(s).

Special relativity can also be stated as the fact that the only physically observable quantities are those that are Poincaré invariant. (Other objects, such as vectors, depend on the choice of reference frame.) For example, consider two spinless particles that interact by collision, producing two spinless particles (which may differ from the originals). Consider just the momenta. (Quantum mechanically, this is a complete description.) All invariants can be expressed in terms of the masses and the "Mandelstam variables" (not to be confused with time and proper time)

$$s = -(p_1 + p_2)^2, \qquad t = -(p_1 - p_3)^2, \qquad u = -(p_1 - p_4)^2$$

where we have used momentum conservation, which shows that even these three quantities are not independent:

$$p_I^2 = -m_I^2, \quad p_1 + p_2 = p_3 + p_4 \quad \Rightarrow \quad s + t + u = \sum_{I=1}^{4} m_I^2$$

(The explicit index now labels the particle, for the process 1+2→3+4.) The simplest reference frame to describe this interaction is the center-of-mass frame (actually the center of momentum, where the two 3-momenta cancel). In that Lorentz frame, using also rotational invariance, momentum conservation, and the mass-shell conditions, the momenta can be written in terms of these invariants as

$$\begin{aligned}
p_1 &= \tfrac{1}{\sqrt{s}}\left(\tfrac{1}{2}(s + m_1^2 - m_2^2), \lambda_{12}, 0, 0\right) \\
p_2 &= \tfrac{1}{\sqrt{s}}\left(\tfrac{1}{2}(s + m_2^2 - m_1^2), -\lambda_{12}, 0, 0\right) \\
p_3 &= \tfrac{1}{\sqrt{s}}\left(\tfrac{1}{2}(s + m_3^2 - m_4^2), \lambda_{34}\, cos\,\theta, \lambda_{34}\, sin\,\theta, 0\right) \\
p_4 &= \tfrac{1}{\sqrt{s}}\left(\tfrac{1}{2}(s + m_4^2 - m_3^2), -\lambda_{34}\, cos\,\theta, -\lambda_{34}\, sin\,\theta, 0\right)
\end{aligned}$$

$$cos\ \theta = \frac{s^2 + 2st - (\sum m_I^2)s + (m_1^2 - m_2^2)(m_3^2 - m_4^2)}{4\lambda_{12}\lambda_{34}}$$

$$\lambda_{IJ}^2 = \tfrac{1}{4}[s - (m_I + m_J)^2][s - (m_I - m_J)^2]$$

The "physical region" of momentum space is then given by $s \geq (m_1 + m_2)^2$ and $(m_3 + m_4)^2$, and $|cos\ \theta| \leq 1$.

Exercise IA4.4

Derive the above expressions for the momenta in terms of invariants in the center-of-mass frame.

Exercise IA4.5

Find the conditions on s, t and u that define the physical region in the case where all masses are equal.

For some purposes it will prove more convenient to use a "lightcone basis"

$$p^\pm = \tfrac{1}{\sqrt{2}}(p^0 \pm p^1) \quad \Rightarrow \quad \eta_{mn} = \begin{array}{c c} & \begin{array}{cccc} + & - & 2 & 3 \end{array} \\ \begin{array}{c} + \\ - \\ 2 \\ 3 \end{array} & \begin{pmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}, \quad p^2 = -2p^+p^- + (p^2)^2 + (p^3)^2$$

and similarly for the "lightcone coordinates" $(x^\pm, x^2, x^3)$. ("Lightcone" is an unfortunate but common misnomer, having nothing to do with cones in most usages.) In this basis the solution to the mass-shell condition $p^2 + m^2 = 0$ can be written as

$$p^\pm = -p_\mp = \frac{(p^i)^2 + m^2}{2p^\mp}$$

(where now $i = 2, 3$), which more closely resembles the nonrelativistic expression. (Note the change on indices $+ \leftrightarrow -$ upon raising and lowering.) A special lightcone basis is the "null basis",

$$p^\pm = \tfrac{1}{\sqrt{2}}(p^0 \pm p^1), \quad p^t = \tfrac{1}{\sqrt{2}}(p^2 - ip^3), \quad \bar{p}^t = \tfrac{1}{\sqrt{2}}(p^2 + ip^3)$$

$$\Rightarrow \quad \eta_{mn} = \begin{array}{c c} & \begin{array}{cccc} + & - & t & \bar{t} \end{array} \\ \begin{array}{c} + \\ - \\ t \\ \bar{t} \end{array} & \begin{pmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array}, \qquad p^2 = -2p^+p^- + 2p^t\bar{p}^t$$

where the square of a vector is linear in each component. (We often use "$\overline{}$" to indicate complex conjugation.)

Exercise IA4.6

Show that for $p^2 + m^2 = 0$ ($m^2 \geq 0$, $p^a \neq 0$), the signs of p^+ and p^- are always the same as the sign of the canonical energy p^0.

Exercise IA4.7

Consider the Poincaré group in 1 extra space dimension (D space, 1 time) for a massless particle. Interpret p^+ as the mass, and p^- as the energy.

a Show that the constraint $p^2 = 0$ gives the usual *nonrelativistic* expression for the energy.

b Show that the subgroup of the Poincaré group generated by all generators that commute with p^+ is the Galilean group (in D$-$1 space and 1 time dimensions). Now nonrelativistic mass conservation is part of momentum conservation, and all the Galilean transformations are coordinate transformations. Also, positivity of the mass is related to positivity of the energy (see exercise IA4.4).

5. Discrete: C, P, T

By considering only symmetries than can be obtained continuously from the identity (Lie groups), we have missed some important symmetries: those that reflect some of the coordinates. It's sufficient to consider a single reflection of a spacelike axis, and one of a timelike axis; all other reflections can be obtained by combining these with the continuous ("proper, orthochronous") Lorentz transformations. (Spacelike and timelike vectors can't be Lorentz transformed into each other, and reflection of a lightlike axis won't preserve $p^2 + m^2$.) Also, the reflection of one spatial axis can be combined with a π rotation about that axis, resulting in reflection of all three spatial coordinates. (Similar generalizations hold for higher dimensions. Note that the product of an even number of reflections about different axes is a proper rotation; thus, for even numbers of spatial dimensions reflections of all spatial coordinates are proper rotations, even though the reflection of a single axis is not.) The reversal of the spatial coordinates is called "parity (P)", while that of the time coordinate is called "time reversal" ("T"; actually, for historical reasons, to be explained shortly, this is usually labeled "CT".) These transformations have the same effect on the momentum, so that the definition of the Poisson bracket is also preserved. These "discrete" transformations, unlike the proper ones, are not symmetries of nature (except in certain approximations): The only exception is the transformation that reflects all axes ("CPT").

While the metric η_{mn} is invariant under all Lorentz transformations (by definition), the "Levi-Civita tensor"

$$\epsilon_{mnpq} \; totally\ antisymmetric, \qquad \epsilon_{0123} = -\epsilon^{0123} = 1$$

is invariant under only proper Lorentz transformations: It has an odd number of space indices and of time indices, so it changes sign under parity or time reversal. (More precisely, under P or T the Levi-Civita tensor does not suffer the expected sign change, since it's constant, so there is an "extra" sign compared to the one expected for a tensor.) Consequently, we can use it to define "pseudotensors": Given "polar vectors", whose signs change as position or momentum under improper Lorentz transformations, and scalars, which are invariant, we can define "axial vectors" and "pseudoscalars" as

$$V_a = \epsilon_{abcd} B^b C^c D^d, \qquad \phi = \epsilon_{abcd} A^a B^b C^c D^d$$

which get an extra sign change under such transformations (P or CT, but not CPT).

There is another such "discrete" transformation that is defined on phase space, but which does not affect spacetime. It changes the sign of all components of the momentum, while leaving the spacetime coordinates unchanged. This transformation is called "charge conjugation (C)", and is also only an approximate symmetry in nature. (Quantum mechanically, complex conjugation of the position-space wave function changes the sign of the momentum.) Furthermore, it does not preserve the Poisson bracket, but changes it by an overall sign. (The misnomer "CT" for time reversal follows historically from the fact that the combination of reversing the time axis and charge conjugation preserves the sign of the energy.) The physical meaning of this transformation is clear from the spacetime-momentum relation of relativistic classical mechanics $p = m\ dx/ds$: It is proper-time reversal, changing the sign of s. The relation to charge follows from "minimal coupling": The "covariant momentum" $m\ dx/ds = p + qA$ (for charge q) appears in the constraint $(p + qA)^2 + m^2 = 0$ in an electromagnetic background; $p \to -p$ then has the same effect as $q \to -q$.

In the previous subsection, we mentioned how negative energies were associated with "antiparticles". Now we can better see the relation in terms of charge conjugation. Note that charge conjugation, since it only changes the sign of τ but does not effect the coordinates, does not change the path of the particle, but only how it is parametrized. This is also true in terms of momentum, since the velocity is given by p^i/p^0. Thus, the only observable property that is changed is charge; spacetime properties (path, velocity, mass; also spin, as we'll see later) remain the same. Another way to say this is that charge conjugation commutes with the Poincaré group. One way to identify an antiparticle is that it has all the same kinematical properties (mass, spin) as the corresponding particle, but opposite sign for internal quantum numbers (like charge). (Another way is pair creation and annihilation: See subsection IIIB5 below.)

All these transformations are summarized in the table:

	C	CT	P	T	CP	PT	CPT
s	$-$	$+$	$+$	$-$	$-$	$-$	$+$
t	$+$	$-$	$+$	$-$	$+$	$-$	$-$
$\vec{x}$	$+$	$+$	$-$	$+$	$-$	$-$	$-$
E	$-$	$-$	$+$	$+$	$-$	$+$	$-$
$\vec{p}$	$-$	$+$	$-$	$-$	$+$	$+$	$-$

(The upper-left 3×3 matrix contains the definitions, the rest is implied.) In terms of complex wave functions, we see that C is just complex conjugation (no effect on coordinates, but momentum and energy change sign because of the "i" in the Fourier transform). On the other hand, for CT and P there is no complex conjugation, but changes in sign of the coordinates that are arguments of the wave functions, and also on the corresponding indices — the "orbital" and "spin" parts of these discrete transformations. (For example, derivatives ∂_a have sign changes because x^a does, so a vector wave function ψ^a must have the same sign changes on its indices for $\partial_a\psi^a$ to transform as a scalar.) The other transformations follow as products of these.

Exercise IA5.1

Find the effect of each of these 7 transformations on wave functions that are: **a** scalars, **b** pseudoscalars, **c** vectors, **d** axial vectors.

However, from the point of view of the "particle" there *is* some kind of kinematic change, since the proper time has changed sign: If we think of the mechanics of a particle as a one-dimensional theory in τ space (the worldline), where $x(\tau)$ (as well as any such variables describing spin or internal symmetry) is a wave function or field on that space, then $\tau \to -\tau$ is T on that one-dimensional space. (The fact we don't get CT can be seen when we add additional variables: For example, if we describe internal U(N) symmetry in terms of creation and annihilation operators $a^{\dagger i}$ and a_i, then C mixes them on both the worldline and spacetime. So, on the worldline we have the "pure" worldline geometric symmetry CT times C = T.) Thus, in terms of "zeroth quantization",

$$worldline\ T \leftrightarrow spacetime\ C$$

On the other hand, *spacetime* P and CT are simply internal symmetries with respect to the worldline (as are proper, orthochronous Poincaré transformations).

Quantum mechanically, there is a good reason for particles of negative energy: They appear in complex-conjugate wave functions, since $(e^{-i\omega t})^* = e^{+i\omega t}$. Since we always evaluate expressions of the form $\langle f|i\rangle$, it is natural for energies of both signs to appear.

In classical field theory, we can identify a particle with its antiparticle by requiring the field to be invariant under charge conjugation: For example, for a scalar field (spinless particle), we have the reality condition

$$\phi(x) = \phi^*(x)$$

or in momentum space, by Fourier transformation,

$$\tilde{\phi}(p) = [\tilde{\phi}(-p)]^*$$

which implies the particle has charge zero (neutral).

6. Conformal

Poincaré transformations are the most general coordinate transformations that preserve the mass condition $p^2 + m^2 = 0$, but there is a larger group, the "conformal group", that preserves this constraint in the massless case. Although conformal symmetry is not observed in nature, it is important in all approaches to field theory:

(1) First of all, it is useful in the construction of free theories (see subsections IIB1-4 below). All massive fields can be described consistently in quantum field theory in terms of coupling massless fields. Massless theories are a subset of conformal theories, and some conditions on massless theories can be found more easily by finding the appropriate subset of those on conformal theories. This is related to the fact that the conformal group, unlike the Poincaré group, is "simple": It has no nontrivial subgroup that transforms into itself under the rest of the group (like the way translations transform into themselves under Lorentz transformations).

(2) In interacting theories at the classical level, conformal symmetry is also important in finding and classifying solutions, since at least some parts of the action are conformally invariant, so corresponding solutions are related by conformal transformations (see subsections IIIC5-7). Furthermore, it is often convenient to treat arbitrary theories as broken conformal theories, introducing fields with which the breaking is associated, and analyze the conformal and conformal-breaking fields separately. This is particularly true for the case of gravity (see subsections IXA7,B5,C2-3,XA3-4,B5-7).

(3) Within quantum field theory at the perturbative level, the only physical quantum field theories are ones that are conformal at high energies (see subsection VIIIC1). The quantum corrections to conformal invariance at high energy are relatively simple.

(4) Beyond perturbation theory, the only quantum theories that are well defined may be just the ones whose breaking of conformal invariance at low energy is only classical (see subsections VIIC2-3,VIIIA5-6). Furthermore, the largest possible symmetry of a nontrivial S-matrix is conformal symmetry (or superconformal symmetry if we include fermionic generators).

(5) Self-duality (a generalization of a condition that equates electric and magnetic fields) is useful for finding solutions to classical field equations as well as simplifying perturbation theory, and is closely related to "twistors" (see subsections IIB6-7,C5,IIIC4-7). In general, self-duality is related to conformal invariance: For

example, it can be shown that the free conformal theories in arbitrary even dimensions are just those with (on-mass-shell) field strengths on which self-duality can be imposed. (In arbitrary odd dimensions the free conformal theories are just the scalar and spinor.)

Transformations λ that satisfy

$$[\lambda^a(x)p_a, p^2] = \zeta(x)p^2$$

for some ζ also preserve $p^2 = 0$, although they don't leave p^2 invariant. Equivalently, we can look for coordinate transformations that scale

$$dx'^2 = \xi(x)dx^2$$

Exercise IA6.1

Find the conformal group explicitly in two dimensions, and show it's infinite dimensional (not just the SO(2,2) described below). (Hint: Use lightcone coordinates.)

This symmetry can be made manifest by starting with a space with one extra space and time dimension:

$$y^A = (y^a, y^+, y^-) \quad \Rightarrow \quad y^2 = y^A y^B \eta_{AB} = (y^a)^2 - 2y^+y^-$$

where $(y^a)^2 = y^a y^b \eta_{ab}$ uses the usual D-dimensional Minkowski-space metric η_{ab}, and the two additional dimensions have been written in a lightcone basis (not to be confused for the similar basis that can be used for the Minkowski metric itself). With respect to this metric, the original SO(D$-$1,1) Lorentz symmetry has been enlarged to SO(D,2). This is the conformal group in D dimensions. However, rather than also preserving (D+2)-dimensional translation invariance, we instead impose the constraint and invariance

$$y^2 = 0, \quad \delta y^A = \zeta(y)y^A$$

This reduces the original space to the "projective" (invariant under the ζ scaling) lightcone (which in this case really is a cone).

These two conditions can be solved by

$$y^A = ew^A, \quad w^A = (x^a, 1, \tfrac{1}{2}x^a x_a)$$

Projective invariance then means independence from e (y^+), while the lightcone condition has determined y^-. $y^2 = 0$ implies $y \cdot dy = 0$, so the simplest conformal invariant is

$$dy^2 = (edw + wde)^2 = e^2 dw^2 = e^2 dx^2$$

where we have used $w^2 = 0 \Rightarrow w \cdot dw = 0$. This means any SO(D,2) transformation on y^A will simply scale dx^2, and scale e^2 in the opposite way:

$$dx'^2 = \left(\frac{e^2}{e'^2}\right) dx^2$$

in agreement with the previous definition of the conformal group.

The explicit form of conformal transformations on $x^a = y^a/y^+$ now follows from their linear form on y^A, using the generators

$$G^{AB} = y^{[A}r^{B]}, \qquad [r_A, y^B] = -i\delta_A^B$$

of SO(D,2) in terms of the momentum r_A conjugate to y^A. (These are defined the same way as the Lorentz generators $J^{ab} = x^{[a}p^{b]}$.) For example, G^{+-} just scales x^a. (Scale transformations are also known as "dilatations".) We can also recognize G^{+a} as generating translations on x^a. The only complicated transformations are generated by G^{-a}, known as "conformal boosts" (acceleration transformations). Since they commute with each other (like translations), it's easy to exponentiate to find the finite transformations:

$$y' = e^G y, \qquad G = v_a y^{[-}\partial^{a]}$$

for some constant D-vector v^a (where $\partial_A \equiv \partial/\partial y^A$). Since the conformal boosts act as "lowering operators" for scale weight ($+ \to a \to -$), only the first three terms in the exponential survive:

$$Gy^- = 0, \qquad Gy^a = v^a y^-, \qquad Gy^+ = v^a y_a \quad \Rightarrow$$

$$y'^- = y^-,\ y'^a = y^a + v^a y^-,\ y'^+ = y^+ + v^a y_a + \tfrac{1}{2}v^2 y^- \quad \Rightarrow$$

$$x'^a = \frac{x^a + \frac{1}{2}v^a x^2}{1 + v \cdot x + \frac{1}{4}v^2 x^2}$$

using $x^a = y^a/y^+$, $y^-/y^+ = \frac{1}{2}x^2$.

Exercise IA6.2

Make the change of variables to $x^a = y^a/y^+$, $e = y^+$, $z = \frac{1}{2}y^2$. Express r_A in terms of the momenta (p_a, n, s) conjugate to (x^a, e, z). Show that the conditions $y^2 = y^A r_A = r^2 = 0$ become $z = en = p^2 = 0$ in terms of the new variables.

Exercise IA6.3

Find the generator of infinitesimal conformal boosts in terms of x^a and p_a.

We actually have the full O(D,2) symmetry: Besides the continuous symmetries, and the discrete ones of SO(D−1,1), we have a second "time" reversal (from our second time dimension):

$$y^+ \leftrightarrow -y^- \quad \Rightarrow \quad x^a \leftrightarrow -\frac{x^a}{\frac{1}{2}x^2}$$

This transformation is called an "inversion".

Exercise IA6.4

Show that a finite conformal boost can be obtained by performing a translation sandwiched between two inversions.

Exercise IA6.5

The conformal group for Euclidean space (or any spacetime signature) can be obtained by the same construction. Consider the special case of D=2 for these SO(D+1,1) transformations. (This is a subgroup of the 2D superconformal group: See exercise IA6.1.) Use complex coordinates for the two "physical" dimensions:

$$z = \tfrac{1}{\sqrt{2}}(x^1 + ix^2)$$

a Show that the inversion is

$$z \leftrightarrow -\frac{1}{z*}$$

b Show that the conformal boost is (using a complex number also for the boost vector)

$$z \to \frac{z}{1 + v^* z}$$

Exercise IA6.6

Any parity transformation (reflection in a spatial axis) can be obtained from any other by a rotation of the spatial coordinates. Similarly, when there is more than one time dimension, any time reversal can be obtained from another (but time reversal can't be rotated into parity, since a timelike vector can't be rotated into a spacelike one). Thus, the complete orthogonal group O(m,n) can be obtained from those transformations that are continuous from the identity by combining them with 1 parity transformation and 1 time reversal transformation (for mn≠0).

a For the conformal group, find the rotation (in terms of an angle) that rotates between the two time directions, and express its action on x^a.

b Show that for angle π it produces a transformation that is the product of time reversal and inversion.

c Use this to show that inversion is related to time reversal by finding the continuum of conformal transformations that connect them.

REFERENCES

1 F.A. Berezin, *The method of second quantization* (Academic, 1966): calculus with anticommuting numbers.
2 P.A.M. Dirac, *Proc. Roy. Soc.* **A126** (1930) 360: antiparticles.
3 E.C.G. Stückelberg, *Helv. Phys. Acta* **14** (1941) 588, **15** (1942) 23; J.A. Wheeler, 1940, unpublished: the relation of antiparticles to proper time.
4 S. Mandelstam, *Phys. Rev.* **112** (1958) 1344.
5 H.W. Brinkmann, *Proc. Nat. Acad. Sci. (USA)* **9** (1923) 1: projective lightcone as conformal to flat space.
6 P.A.M. Dirac, *Ann. Math.* **37** (1936) 429; H.A. Kastrup, *Phys. Rev.* **150** (1966) 1186; G. Mack and A. Salam, *Ann. Phys.* **53** (1969) 174; S. Adler, *Phys. Rev.* **D6** (1972) 3445; R. Marnelius and B. Nilsson, *Phys. Rev.* **D22** (1980) 830: conformal symmetry.
7 S. Coleman and J. Mandula, *Phys. Rev.* **159** (1967) 1251: conformal symmetry as the largest (bosonic) symmetry of the S-matrix.
8 W. Siegel, *Int. J. Mod. Phys. A* **4** (1989) 2015: equivalence between conformal invariance and self-duality in all dimensions.

B. INDICES

In the previous section we saw various spacetime groups (Galilean, Poincaré, conformal) in terms of how they acted on coordinates. This not only gave them a simple physical interpretation, but also allowed a direct relation between classical and quantum theories. However, as we know from studying rotations in quantum theory in terms of spin, we will often need to study symmetries of quantum theories for which the classical analog is not so useful or perhaps even nonexistent.

We therefore now consider some general results of group theory, mostly for continuous groups. We use tensor methods, rather than the slightly more powerful but greatly less convenient Cartan-Weyl-Dynkin methods. Much of this section should be review, but is included here for completeness; it is not intended as a substitute for a group theory course, but as a summary of those results commonly useful in field theory.

1. Matrices

Matrices are defined by the way they act on some vector space; an n×n matrix takes one n-component vector to another. Given some group, and its multiplication table (which defines the group completely), there is more than one way to represent it by matrices. Any set of matrices we find that has the same multiplication table as the group elements is called a "representation" of that group, and the vector space on which those matrices act is called the "representation space." The representation of the algebra or group in terms of explicit matrices is given by choosing a basis for the vector space. If we include infinite-dimensional representations, then a representation of a group is simply a way to write its transformations that is linear: $\psi' = M\psi$ is linear in ψ. More generally, we can also have a "realization" of a group, where the transformations can be nonlinear. These tend to be more cumbersome, so we usually try to make redefinitions of the variables that make the realization linear. A precise definition of "manifest symmetry" is that all the realizations used are linear. (One possible exception is "affine" or "inhomogeneous" transformations $\psi' = M_1\psi + M_2$, such as the usual coordinate representation of Poincaré transformations, since these transformations are still very simple, because they are really still linear, though not homogeneous.)

Exercise IB1.1

Consider a general real affine transformation $\psi' = M\psi + V$ on an n-component vector ψ for arbitrary real $n \times n$ matrices M and real n-vectors V. A general group element is thus (M, V).

a Perform 2 such transformations consecutively, and give the resulting "group multiplication" rule for (M_1, V_1) "$\times$" $(M_2, V_2) = (M_3, V_3)$.

b Find the infinitesimal form of this transformation. Define the n^2+n generators as operators on ψ, in terms of ψ^a and $\partial/\partial\psi^a$.

c Find the commutation relations of these generators.

d Compare all the above with (nonrelativistic) rotations and translations.

Exercise IB1.2

Show $(AB)^{-1} = B^{-1}A^{-1}$ for matrices A and B that have inverses but don't necessarily commute with each other. Use it to show that

$$\frac{1}{A+B} = \frac{1}{A} - \frac{1}{A}B\frac{1}{A} + \frac{1}{A}B\frac{1}{A}B\frac{1}{A} - ...$$

(These may be different A and B. There may be other assumptions; ignore convergence questions.)

For convenience, we write matrices with a Hilbert-space-like notation, but unlike Hilbert space we don't necessarily associate bras directly with kets by Hermitian conjugation, or even transposition. In general, the two spaces can even be different sizes, to describe matrices that are not square; however, for group theory we are interested only in matrices that take us from some vector space into itself, so they are square. Bras have an inner product with kets, but neither necessarily has a norm (inner product with itself): In general, if we start with some vector space, written as kets, we can always define the "dual" space, written as bras, by defining such an inner product. In our case, we may start with some representation of a group, in terms of some vector space, and that will give us directly the dual representation. (If the representation is in terms of unitary matrices, we have a Hilbert space, and the dual representation is just the complex conjugate.)

So, we define column vectors $|\psi\rangle$ with a basis $|^I\rangle$, and row vectors $\langle\psi|$ with a basis $\langle_I|$, where $I = 1,...,$n to describe n×n matrices. The two bases have a relative normalization defined so that the inner product gives the usual component sum:

$$|\psi\rangle = |^I\rangle\psi_I,\ \langle\chi| = \chi^I\langle_I|; \quad \langle_I|^J\rangle = \delta_I^J \quad \Rightarrow \quad \langle\chi|\psi\rangle = \chi^I\psi_I; \quad \langle_I|\psi\rangle = \psi_I,\ \langle\chi|^I\rangle = \chi^I$$

These bases then define not only the components of vectors, but also matrices:

$$M = |^I\rangle M_I{}^J\langle_J|, \quad \langle_I|M|^J\rangle = M_I{}^J$$

where as usual the I on the component (matrix element) $M_I{}^J$ labels the row of the matrix M, and J the column. This implies the usual matrix multiplication rules, inserting the identity in terms of the basis,

$$I = |^K\rangle\langle_K| \quad \Rightarrow \quad (MN)_I{}^J = \langle_I|M|^K\rangle\langle_K|N|^J\rangle = M_I{}^K N_K{}^J$$

Closely related is the definition of the trace,

$$tr\ M = \langle_I|M|^I\rangle = M_I{}^I \quad \Rightarrow \quad tr(MN) = tr(NM)$$

(We'll discuss the determinant later.)

The bra-ket notation is really just matrix notation written in a way to clearly distinguish column vectors, row vectors, and matrices. We can, of course, also use the usual pictorial notation

$$|\psi\rangle = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \end{pmatrix}, \qquad \langle\chi| = (\chi^1\ \chi^2\ \ldots)$$

$$M = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ I \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & \ldots & J & \ldots \end{array} \\ \begin{pmatrix} M_1{}^1 & M_1{}^2 & \ldots & M_1{}^J & \ldots \\ M_2{}^1 & M_2{}^2 & \ldots & M_2{}^J & \ldots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ M_I{}^1 & M_I{}^2 & \ldots & M_I{}^J & \ldots \\ \vdots & \vdots & \ddots & \vdots & \ddots \end{pmatrix} \end{array}$$

This is useful only when listing individual components.

We can easily translate transformation laws from matrix notation into index notation just by using a basis for the representation space. We now write g and G to refer to either matrix representations of the group and algebra elements, or to the abstract elements: i.e., either to a specific representation, or the most general one. Again writing $g = e^{iG}$,

$$g|^I\rangle = |^J\rangle g_J{}^I, \quad G|^I\rangle = |^J\rangle G_J{}^I$$

$$G = \alpha^i G_i, \quad \delta|\psi\rangle = iG|\psi\rangle = |^I\rangle i\alpha^i (G_i)_I{}^J \psi_J \quad \Rightarrow \quad \delta\psi_I = i\alpha^i (G_i)_I{}^J \psi_J$$

The dual space isn't needed for this purpose. However, for any representation of a group, the transpose

$$(M^T)^I{}_J = M_J{}^I$$

of the inverse of those matrices also gives a representation of the group, since

$$g_1 g_2 = g_3 \quad \Rightarrow \quad (g_1)^{T-1}(g_2)^{T-1} = (g_3)^{T-1}$$

$$[G_1, G_2] = G_3 \quad \Rightarrow \quad [-G_1^T, -G_2^T] = -G_3^T$$

This is the dual representation, which follows from defining the above inner product to be invariant under the group:

$$\delta\langle\psi|\chi\rangle = 0 \quad \Rightarrow \quad \delta\psi^I = -i\psi^J\alpha^i(G_i)_J{}^I$$

The complex conjugate of a complex representation is also a representation, since

$$g_1 g_2 = g_3 \quad \Rightarrow \quad g_1{}^* g_2{}^* = g_3{}^*$$

$$[G_1, G_2] = G_3 \quad \Rightarrow \quad [G_1{}^*, G_2{}^*] = G_3{}^*$$

From any given representation, we can thus find three others from taking the dual and the conjugate: In matrix and index notation,

$$\begin{aligned} \psi' = g\psi : &\qquad \psi'_I = g_I{}^J\psi_J \\ \psi' = (g^{-1})^T\psi : &\qquad \psi'^I = g^{-1}{}_J{}^I\psi^J \\ \psi' = g^*\psi : &\qquad \psi'_{\dot{I}} = g^*{}_{\dot{I}}{}^{\dot{J}}\psi_{\dot{J}} \\ \psi' = (g^{-1})^\dagger\psi : &\qquad \psi'^{\dot{I}} = g^{*-1}{}_{\dot{J}}{}^{\dot{I}}\psi^{\dot{J}} \end{aligned}$$

since $(g^{-1})^T$, g^*, and $(g^{-1})^\dagger$ (but not g^T, etc.) satisfy the same multiplication algebra as g, including ordering. We use up/down and dotted/undotted indices to denote the transformation law of each type of index; contracting undotted up indices with undotted down indices preserves the transformation law as indicated by the remaining indices, and similarly for dotted indices. These four representations are not necessarily independent: Imposing relations among them is how the classical groups are defined (see subsections IB4-5 below).

2. Representations

For example, we always have the "adjoint" representation of a Lie group/algebra, which is how the algebra acts on its own generators:

(1) adjoint as operator: $G = \alpha^i G_i, \quad A = \beta^i G_i \quad \Rightarrow \quad \delta A = i[G, A] = \beta^j\alpha^i f_{ij}{}^k G_k$

$$\Rightarrow \quad \delta\beta^i = -i\beta^k\alpha^j(G_j)_k{}^i, \quad (G_i)_j{}^k = if_{ij}{}^k$$

This gives us two ways to represent the adjoint representation space: as either the usual vector space, or in terms of the generators. Thus, we either use the matrix

$A = \beta^i G_i$ (for arbitrary representation of the matrices G_i, or treating G_i as just abstract generators), or we can write A as a row vector:

$$(2)\ \text{adjoint as vector:} \quad \langle A| = \beta^i \langle_i| \quad \Rightarrow \quad \delta\langle A| = -i\langle A|G$$

$$\Rightarrow \quad \delta\beta^i\langle_i| = -i\beta^k\alpha^j (G_j)_k{}^i\langle_i|$$

The adjoint representation also provides a convenient way to define a (symmetric) group metric invariant under the group, the "Cartan metric":

$$\eta_{ij} = tr_A(G_i G_j) = -f_{ik}{}^l f_{jl}{}^k$$

(tr_A refers to the trace taken with respect to the representation A; equivalently, we could take the G's inside the trace to be in the A representation.) For "Abelian" groups the structure constants vanish, and thus so does this metric. "Semisimple" groups are those where the metric is invertible (no vanishing eigenvalues). A "simple" group has no nontrivial subgroup that transforms into itself under the rest of the group: Semisimple groups can be written as "products" of simple groups. "Compact" groups are those where it is positive definite (all eigenvalues positive); they are also those for which the invariant volume of the group space is finite. For simple, compact groups it's convenient to choose a basis where

$$\eta_{ij} = c_A \delta_{ij}$$

for some constant c_A (the "Dynkin index" for the adjoint representation). For some general irreducible representation R of such a group the normalization of the trace is

$$tr_R(G_i G_j) = c_R \delta_{ij} = \frac{c_R}{c_A}\eta_{ij}$$

Now the proportionality constant c_R/c_A is fixed by the choice of R (only), since we have already fixed the normalization of our basis.

Exercise IB2.1

What is c_R for an Abelian group? (Hint: not just 1.)

In general, the cyclicity property of the trace implies, for any representation, that

$$0 = tr([G_i, G_j]) = -if_{ij}{}^k tr(G_k)$$

so $tr(G_i) = 0$ for semisimple groups. Similarly, we find

$$f_{ijk} \equiv f_{ij}{}^l \eta_{lk} = i\ tr_A([G_i, G_j]G_k)$$

is totally antisymmetric: For semisimple groups, this implies the total antisymmetry of the structure constants $f_{ij}{}^k$, up to factors (which are absent for compact groups in a basis where $\eta_{ij} \sim \delta_{ij}$). This also means the adjoint representation is its own dual. (For example, for the compact group SO(3), we have $\eta_{ij} = -\epsilon_{ikl}\epsilon_{jlk} = 2\delta_{ij}$.) Thus, we can write A in a third way, as a column vector

$$\text{(3) adjoint as dual vector:} \quad |A\rangle = |^i\rangle\beta_i \equiv |^i\rangle\beta^j\eta_{ji} \quad \Rightarrow \quad \delta|A\rangle = iG|A\rangle$$

We can also do this for Abelian groups, by defining an invertible metric unrelated to the Cartan metric: This is trivial for Abelian groups, since the generators themselves are invariant, and thus so is *any* metric on them.

An identity related to the trace one is the normalization of the value k_R of the "Casimir operator" for any particular representation,

$$\eta^{ij}G_iG_j = k_R I$$

Its proportionality to the identity follows from the fact that it commutes with each generator:

$$[\eta^{jk}G_jG_k, G_i] = -if^j{}_i{}^k\{G_j, G_k\} = 0$$

using the antisymmetry of the structure constants. (Thus it takes the same value on any component of an irreducible representation, since they are all related by group transformations.) By tracing this identity, and contracting the trace identity,

$$\frac{c_R}{c_A}d_A = tr_R(\eta^{ij}G_iG_j) = k_R d_R$$

$$\Rightarrow \quad k_R = \frac{c_R d_A}{c_A d_R}$$

where $d_R \equiv tr_R(I)$ is the dimension of that representation.

Although quantum mechanics is defined on Hilbert space, which is a kind of complex vector space, more generally we want to consider real objects, like spacetime vectors. This restricts the form of linear transformations: Specifically, if we absorb i's as $g = e^G$, then in such representations G itself must be real. These representations are then called "real representations", while a "complex representation" is one whose representation isn't real in any basis. A complex representation space can have a real representation, but a real representation space can't have a complex representation. In particular, coordinate transformations (of real coordinates) have only real representations, which is why absorbing the i's into the generators is a useful convention there. For semisimple unitary groups, hermiticity of the generators of the adjoint representation implies (using total antisymmetry of the structure constants and reality of the Cartan metric) that the structure constants are real, and thus the adjoint representation is a real representation. More generally, any real unitary representation will have antisymmetric generators ($G = G^* = -G^\dagger \Rightarrow G = -G^T$). If the complex conjugate representation is the same as the original (same matrices up to a similarity transformation $g^* = MgM^{-1}$), but the representation is not real, then it is called "pseudoreal". (An example is the spinor of SU(2), to be described in the next section.)

For any representation g of the group, a transformation $g \to g_0 g g_0^{-1}$ on every group element g for some particular group element g_0 clearly maps the algebra to itself, and preserves the multiplication rules. (Similar remarks apply to applying the transformation to the generators.) However, the same is true for complex conjugation, $g \to g^*$: Not only are the multiplication rules preserved, but for any element g of that representation of the group, g^* is also an element. (This can be shown, e.g., by defining representations in terms of the values of all the Casimir operators, contructed from various powers of the generators.) In quantum mechanics (where the representations are unitary), the latter is called an "antiunitary transformation". Although this is a symmetry of the group, it cannot be reproduced by a unitary transformation, except when the representation is (pseudo)real.

Exercise IB2.2

Show how this works for the Abelian group U(1). Explain this antiunitary transformation in terms of two-dimensional rotations O(2). (U(1)=SO(2), the "proper rotations" obtained continuously from the identity.)

A very simple way to build a representation from others is by "direct sum". If we have two representations of a group, on two different spaces, then we can take their direct sum by just putting one column vector on top of the other, creating a bigger vector whose size ("dimension") is the sum of that of the original two. Explicitly, if we start with the basis $|^\iota\rangle$ for the first representation and $|^{\iota'}\rangle$ for the second, then the union $(|^\iota\rangle, |^{\iota'}\rangle)$ is the basis for the direct sum. (We can also write $|^I\rangle = (|^\iota\rangle, |^{\iota'}\rangle)$, where $\iota = 1, ..., m$; $\iota' = 1, ..., n$; $I = 1, ..., m, m+1, ..., m+n$.) The group then acts on each part of the new vector in the obvious way:

$$\psi = |^\iota\rangle\psi_\iota, \quad \chi = |^{\iota'}\rangle\chi_{\iota'}; \qquad g|^\iota\rangle = |^\kappa\rangle g_\kappa{}^\iota, \quad g|^{\iota'}\rangle = |^{\kappa'}\rangle g_{\kappa'}{}^{\iota'}$$

$$\Rightarrow \quad |\Psi\rangle = |^\iota\rangle\psi_\iota \oplus |^{\iota'}\rangle\chi_{\iota'} = |\psi\rangle \oplus |\chi\rangle \quad or \quad (\Psi) = \begin{pmatrix} \psi \\ \chi \end{pmatrix}$$

$$g|\Psi\rangle = |^\kappa\rangle g_\kappa{}^\iota\psi_\iota \oplus |^{\kappa'}\rangle g_{\kappa'}{}^{\iota'}\chi_{\iota'} \quad or \quad (g) = \begin{pmatrix} g_\iota{}^\kappa & 0 \\ 0 & g_{\iota'}{}^{\kappa'} \end{pmatrix}$$

(We can replace the $\oplus$ with an ordinary $+$ if we understand the basis vectors to be now in a bigger space, where the elements of the first basis have zeros for the new components on the bottom while those of the second have zeros for the new components on top.) The important point is that no group element mixes the two spaces: The group representation is block diagonal. Any representation that can be written as a direct sum (after an appropriate choice of basis) is called "reducible". For example, we can build a reducible real representation from an irreducible complex one by just taking the direct sum of this complex representation with the complex conjugate representation. Similarly, we can take direct sums of more than two representations.

A more useful way to build representations is by "direct product". The idea there is to take a colummn vector and a row vector and use them to construct a matrix, where the group element acts simultaneously on rows according to one representation and columns according to the other. If the two original bases are again $|^{\iota}\rangle$ and $|^{\iota'}\rangle$, the new basis can also be written as $|^{I}\rangle = |^{\iota\iota'}\rangle$ $(I = 1, ..., mn)$. Explicitly,

$$|\psi\rangle = |^{\iota}\rangle \otimes |^{\iota'}\rangle\psi_{\iota\iota'}, \quad g(|^{\iota}\rangle \otimes |^{\iota'}\rangle) = |^{\kappa}\rangle \otimes |^{\kappa'}\rangle g_{\kappa}{}^{\iota} g_{\kappa'}{}^{\iota'} \quad \Rightarrow \quad g_{\iota\iota'}{}^{\kappa\kappa'} = g_{\iota}{}^{\kappa} g_{\iota'}{}^{\kappa'}$$

or in terms of the algebra

$$G_{\iota\iota'}{}^{\kappa\kappa'} = G_{\iota}{}^{\kappa}\delta_{\iota'}{}^{\kappa'} + \delta_{\iota}{}^{\kappa}G_{\iota'}{}^{\kappa'}$$

A familar example from quantum mechanics is rotations (or Lorentz transformations), where the first space is position space (so ι is the continuous index x), acted on by the orbital part of the generators, while the second space is finite-dimensional, and is acted on by the spin part of the generators. Direct product representations are usually reducible: They then can be written also as direct sums, in a way that depends on the particulars of the group and the representations.

Consider a representation constructed by direct product: In matrix notation

$$\hat{G}_i = G_i \otimes I' + I \otimes G'_i$$

Using $tr(A \otimes B) = tr(A)tr(B)$, and assuming $tr(G_i) = tr(G'_i) = 0$, we have

$$tr(\hat{G}_i\hat{G}_j) = tr(I')tr(G_iG_j) + tr(I)tr(G'_iG'_j)$$

For example, for SU(N) (see subsection IB4 below) we can construct the adjoint representation from the direct product of the N-dimensional, "defining" representation and its complex conjugate. (We also get a singlet, but it will not affect the result for the adjoint.) In that case we find

$$tr_A(G_iG_j) = 2N\, tr_D(G_iG_j) \quad \Rightarrow \quad \frac{c_D}{c_A} = \frac{1}{2N}$$

For most purposes, we use $tr_D(G_iG_j) = \delta_{ij}$ $(c_D = 1)$ for SU(N), so $c_A = 2N$.

3. Determinants

We now "review" some properties of determinants that will prove useful for the group analysis of the following subsections. Determinants can be defined in terms of the Levi-Civita tensor ϵ. As a consequence of its antisymmetry,

$$\epsilon \textit{ totally antisymmetric}, \quad \epsilon_{12...n} = \epsilon^{12...n} = 1 \quad \Rightarrow \quad \epsilon_{J_1...J_n}\epsilon^{I_1...I_n} = \delta^{I_1}_{[J_1} \cdots \delta^{I_n}_{J_n]}$$

since each possible numerical index value appears once in each ϵ, so they can be matched up with δ's. By similar reasoning,

$$\tfrac{1}{m!}\epsilon_{K_1...K_m J_1...J_{n-m}}\epsilon^{K_1...K_m I_1...I_{n-m}} = \delta^{I_1}_{[J_1} \cdots \delta^{I_{n-m}}_{J_{n-m}]}$$

where the normalization compensates for the number of terms in the summation.

Exercise IB3.1

Apply these identities to rotations in three dimensions:

a Given only the commutation relations $[J_{ij}, J^{kl}] = i\delta_{[i}^{[k} J_{j]}{}^{l]}$ and the definition $G_i \equiv \frac{1}{2}\epsilon_{ijk}J_{jk}$, derive $f_{ij}{}^k = \epsilon_{ijk}$.

b Show the Jacobi identity $\epsilon_{[ij}{}^l\epsilon_{k]l}{}^m = 0$ by explicit evaluation.

c Find the Cartan metric, and thus the value of c_A.

This tensor is used to define the determinant:

$$det\ M_I{}^J = \tfrac{1}{n!}\epsilon_{J_1...J_n}\epsilon^{I_1...I_n}M_{I_1}{}^{J_1}\cdots M_{I_n}{}^{J_n} \quad \Rightarrow \quad \epsilon_{J_1...J_n}M_{I_1}{}^{J_1}\cdots M_{I_n}{}^{J_n} = \epsilon_{I_1...I_n} det\ M$$

since anything totally antisymmetric in n indices must be proportional to the ϵ tensor. This yields an explicit expression for the inverse:

$$(M^{-1})_{J_1}{}^{I_1} = \tfrac{1}{(n-1)!}\epsilon_{J_1...J_n}\epsilon^{I_1...I_n}M_{I_2}{}^{J_2}\cdots M_{I_n}{}^{J_n}(det\ M)^{-1}$$

From this follows a useful expression for the variation of the determinant:

$$\frac{\partial}{\partial M_I{}^J} det\ M = (M^{-1})_J{}^I\, det\ M$$

which is equivalent to

$$\delta\ ln\ det\ M = tr(M^{-1}\delta M)$$

Replacing M with e^M gives the often-used identity

$$\delta\ ln\ det\ e^M = tr(e^{-M}\delta e^M) = tr\ \delta M \quad \Rightarrow \quad det\ e^M = e^{tr\ M}$$

where we have used the boundary condition for $M = 0$. Finally, replacing M in the last identity with $ln(1+L)$ and expanding both sides to order L^n gives general expressions for determinants of $n \times n$ matrices in terms of traces:

$$det(1+L) = e^{tr\, ln(1+L)} \quad \Rightarrow \quad det\ L = \tfrac{1}{n!}(tr\ L)^n - \tfrac{1}{2(n-2)!}(tr\ L^2)(tr\ L)^{n-2} + \cdots$$

Exercise IB3.1

Use the definition of the determinant (and not its relation to the trace) to show

$$det(AB) = det(A)det(B)$$

These identities can also be derived by defining the determinant in terms of a Gaussian integral. We first collect some general properties of (indefinite) Gaussian integrals. The simplest such integral is

$$\int \frac{d^2x}{2\pi}\ e^{-x^2/2} = \int_0^{2\pi} \frac{d\theta}{2\pi} \int_0^\infty dr\ re^{-r^2/2} = \int_0^\infty du\ e^{-u} = 1$$

$$\Rightarrow \quad \int \frac{d^Dx}{(2\pi)^{D/2}}\ e^{-x^2/2} = \left(\int \frac{dx}{\sqrt{2\pi}}\ e^{-x^2/2}\right)^D = \left(\int \frac{d^2x}{2\pi}\ e^{-x^2/2}\right)^{D/2} = 1$$

The complex form of this integral is

$$\int \frac{d^Dz^*\ d^Dz}{(2\pi i)^D}\ e^{-|z|^2} = 1$$

by reducing to real parameters as $z = (x+iy)/\sqrt{2}$. These generalize to integrals involving a real, symmetric matrix S or a Hermitian matrix H as

$$\int \frac{d^Dx}{(2\pi)^{D/2}}\ e^{-x^TSx/2} = (det\ S)^{-1/2}, \qquad \int \frac{d^Dz^*\ d^Dz}{(2\pi i)^D}\ e^{-z^\dagger Hz} = (det\ H)^{-1}$$

by diagonalizing the matrices, making appropriate redefinitions of the integration variables, and identifying the determinant of a diagonal matrix. Alternatively, we can use these integrals to *define* the determinant, and derive the previous definition. The relation for the symmetric matrix follows from that for the Hermitian one by separating z into its real and imaginary parts for the special case $H = S$. If we treat z and z^* as independent variables, the determinant can also be understood as the Jacobian for the (dummy) variable change $z \to H^{-1}z$, $z^* \to z^*$. More generally, if we define the integral by an appropriate limiting procedure or analytic continuation (for convergence), we can choose z and z^* to be unrelated (or even separate real variables), and S and H to be complex.

Exercise IB3.2

Other properties of determinants can also be derived directly from the integral definition:

a Find an integral expression for the inverse of a (complex) matrix M by using the identity

$$0 = \int \frac{\partial}{\partial z_I}(z_J\, e^{-z^\dagger M z})$$

b Derive the identity $\delta\ ln\ det\ M = tr(M^{-1}\delta M)$ by varying the Gaussian definition of the (complex) determinant with respect to M.

An even better definition of the determinant is in terms of an *anticommuting* integral (see subsection IA2), since anticommutativity automatically gives the antisymmetry of the Levi-Civita tensor, and we don't have to worry about convergence. We then have, for *any* matrix M,

$$\int d^D\zeta^\dagger\ d^D\zeta\ e^{-\zeta^\dagger M\zeta} = det\ M$$

where $\zeta^\dagger$ can be chosen as the Hermitian conjugate of ζ or as an independent variable, whichever is convenient. From the definition of anticommuting integration, the only terms in the Taylor expansion of the exponential that contribute are those with the product of one of each anticommuting variable. Total antisymmetry in ζ and in $\zeta^\dagger$ then yields the determinant; we define "$d^D\zeta^\dagger\ d^D\zeta$" to give the correct normalization. (The normalization is ambiguous anyway because of the signs in ordering the $d\zeta$'s.) This determinant can also be considered a Jacobian, but the inverse of the commuting result follows from the fact that the integrals are now really derivatives.

Exercise IB3.3

Divide up the range of a square matrix into two (not necessarily equal) parts: In block form,

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

and do the same for the (commuting or anticommuting) variables used in defining its determinant. Show that

$$det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = det\ D \cdot det(A - BD^{-1}C) = det\ A \cdot det(D - CA^{-1}B)$$

a by integrating over one part of the variables first (this requires off-diagonal changes of variables of the form $y \to y + \mathcal{O}x$, which have unit Jacobian), or

b by first proving the identity

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} I & 0 \\ D^{-1}C & I \end{pmatrix}$$

We then have, for any *anti*symmetric (even-dimensional) matrix A,

$$\int d^{2D}\xi\ e^{-\xi^T A\xi/2} = Pf\ A, \qquad (Pf\ A)^2 = det\ A$$

by the same method as the commuting case (again with appropriate definition of the normalization of $d^{2D}\xi$; the determinant of an odd-dimensional antisymmetric matrix vanishes, since $det\ M = det\ M^T$). However, there is now an important difference: The "Pfaffian" is not merely the square root of the determinant, but itself a polynomial, since we can evaluate it also by Taylor expansion:

$$Pf\ A_{IJ} = \tfrac{1}{D!2^D}\epsilon^{I_1...I_{2D}} A_{I_1I_2}\cdots A_{I_{2D-1}I_{2D}}$$

which can be used as an alternate definition. (Normalization can be checked by examining a special case; the overall sign is part of the normalization convention.)

4. Classical groups

The rotation group in three dimensions can be expressed most simply in terms of 2×2 matrices. This description is the most convenient for not only spin 1/2, but all spins. This result can be extended to orthogonal groups (such as the rotation, Lorentz, and conformal groups) in other low dimensions, including all those relevant to spacetime symmetries in four dimensions.

There are an infinite number of Lie groups. Of the compact ones, all but a finite number are among the "classical" Lie groups. These classical groups can be defined easily in terms of (real or complex) matrices satisfying a few simple constraints. (The remaining "exceptional" compact groups can be defined in a similar way with a little extra effort, but they are of rather specialized interest, so we won't cover them here.) These matrices are thus called the "defining" representation of the group. (Sometimes this representation is also called the "fundamental" representation; however, this term has been used in slightly different ways in the literature, so we will avoid it.) These constraints are a subset of:

$$\begin{array}{llrll}
\text{volume:} & & \textbf{S}\text{pecial:} & det(g) = 1 & \\
\text{metric:} & \left\{\begin{array}{l}\text{hermitian:} \\ \text{(anti)symmetric:}\end{array}\right. & \begin{array}{r}\textbf{U}\text{nitary:} \\ \left\{\begin{array}{r}\textbf{O}\text{rthogonal:} \\ \textbf{S}\text{ym}\textbf{p}\text{lectic:}\end{array}\right.\end{array} & \begin{array}{l} g^\dagger \Upsilon g = \Upsilon \\ g^T\eta g = \eta \\ g^T\Omega g = \Omega\end{array} & \begin{array}{l}(\Upsilon^\dagger = \Upsilon) \\ (\eta^T = \eta) \\ (\Omega^T = -\Omega)\end{array} \\
\text{reality:} & & \left\{\begin{array}{r}\textbf{R}\text{eal:} \\ \text{pseudoreal (*):}\end{array}\right. & \begin{array}{l} g^* = \eta g\eta^{-1} \\ g^* = \Omega g\Omega^{-1}\end{array} &
\end{array}$$

where g is any matrix in the defining representation of the group, while Υ, η, Ω are group "metrics", defining inner products (while the determinant defines the volume, as in the Jacobian). For the compact cases Υ and η can be chosen to be the identity, but we will also consider some noncompact cases. (There are also some uninteresting variations of "Special" for complex matrices, setting the determinant to be real or its magnitude to be 1.)

Exercise IB4.1

Write all the defining constraints of the classical groups (S, U, O, Sp, R, pseudoreal) in terms of the algebra rather than the group.

Note the modified definition of unitarity, etc. Such things are also encountered in quantum mechanics with ghosts, since the resulting Hilbert space can have an indefinite metric. For example, if we have a finite-dimensional Hilbert space where the inner product is represented in terms of matrices as

$$\langle\psi|\chi\rangle = \psi^\dagger \Upsilon \chi$$

then "observables" satisfy a "pseudohermiticity" condition

$$\langle\psi|H\chi\rangle = \langle H\psi|\chi\rangle \quad \Rightarrow \quad \Upsilon H = H^\dagger \Upsilon$$

and unitarity generalizes to

$$\langle U\psi|U\chi\rangle = \langle\psi|\chi\rangle \quad \Rightarrow \quad U^\dagger \Upsilon U = \Upsilon$$

Similar remarks apply when replacing the Hilbert-space "sesquilinear" (vector times complex conjugate of vector) inner product with a symmetric (orthogonal) or antisymmetric (symplectic) bilinear inner product. An important example is when the wave function carries a Lorentz vector index, as expected for a relativistic description of spin 1; then clearly the time component is unphysical.

The groups of matrices that can be constructed from these conditions are then:

GL(n,C) [SL(n,C)]

U: [S]U(n_+,n_-)
O: [S]O(n,C)
Sp: Sp(2n,C)
R: GL(n) [SL(n)]
: [S]U(2n)

U	R	*
O	[S]O(n_+,n_-)	SO*(2n)
Sp	Sp(2n)	USp($2n_+$,$2n_-$)

Of the non-determinant constraints, in the first column we applied none ("GL" means "general linear", and "C" refers to the complex numbers; the real numbers "R" are implicit); in the second column we applied one; in the third column we applied three, since two of the three types (unitarity, symmetry, reality) imply the third. (The corresponding groups with unit determinant, when distinct, are given in brackets.) These square matrices are of size n, n_++n_-, 2n, or 2n_++2n_-, as indicated. n_+ and n_- refer to the number of positive and negative eigenvalues of the metric Υ or η. O(n) differs from SO(n) by including "parity"-type transformations, which can't be obtained continuously from the identity. (SSp(2n) is the same as Sp(2n).) For this reason, and also for studying "topological" properties, for finite transformations it is sometimes more useful to work directly with the group elements g, rather than parametrizing them in terms of algebra elements as $g = e^{iG}$. U(n) differs from SU(n) (and similarly for GL(n) vs. SL(n)) only by including a U(1) group that commutes with the SU(n): Although U(1) is noncompact (it consists of just phase transformations), a compact form of it can be used by requiring that all "charges" are integers (i.e., all representations transform as $\psi' = e^{iq\theta}\psi$ for group parameter θ, where q is an integer defining the representation).

Of these groups, the compact ones are just SU(n), SO(n) (and O(n)), and USp(2n) (all with n_-=0). The compact groups have an interesting interpretation in terms of various number systems: SO(n) is the unitary group of n×n matrices over the real numbers, SU(n) is the same for the complex numbers, and USp(2n) is the same for the quaternions. (Similar interpretations can be made for some of the noncompact groups.) The remaining compact Lie groups that we didn't discuss, the "exceptional" groups, can be interpreted as unitary groups over the octonions. (Unlike the classical groups, which form infinite series, there are only five exceptional compact groups, because of the restrictions following from the nonassociativity of octonions.)

5. Tensor notation

Although historically group representations have usually been taught in the notation where an m-component representation of a group defined by n×n matrices is represented by an m-component vector, carrying a single index with values 1 to m, a much more convenient and transparent method is "tensor notation", where a general representation carries many indices ranging from 1 to n, with certain symmetries (and perhaps tracelessness) imposed on them. (Tensor notation for a covering group is generally known as "spinor notation" for the corresponding orthogonal group: See subsection IC5.) This notation takes advantage of the property described above for expressing arbitrary representations in terms of direct products of vectors. In terms of transformation laws, it means we need to know only the defining representation, since the transformation of this representation is applied to each index. There are at most four vector representations, by taking the dual and complex conjugate; we use the corresponding index notation. Then the group constraints simply state the invariance of the group metrics (and their complex conjugates and inverses), which thus can be used to raise, lower, and contract indices:

$$\begin{array}{llrl}
\text{volume:} & & \textbf{S}\text{pecial:} & \epsilon^{I_1 \dots I_n} \\
\text{metric:} & \left\{ \begin{array}{l} \text{hermitian:} \\ \text{(anti)symmetric:} \left\{ \begin{array}{r} \\ \\ \end{array} \right. \end{array} \right. & \begin{array}{r} \textbf{U}\text{nitary:} \\ \textbf{O}\text{rthogonal:} \\ \textbf{S}\text{ym}\textbf{p}\text{lectic:} \end{array} & \begin{array}{l} \Upsilon^{\dot I J} \\ \eta^{IJ} \\ \Omega^{IJ} \end{array} \\
\text{reality:} & & \left\{ \begin{array}{r} \textbf{R}\text{eal:} \\ \text{pseudoreal (*):} \end{array} \right. & \begin{array}{l} \eta_{\dot I}{}^J \\ \Omega_{\dot I}{}^J \end{array}
\end{array}$$

As a result, we have relations such as

$$\langle{}^I|{}^J\rangle = \eta^{IJ} \quad or \quad \Omega^{IJ}, \qquad \langle{}^{\dot I}|{}^J\rangle = \Upsilon^{\dot I J}$$

We also define inverse metrics satisfying

$$\eta^{KI}\eta_{KJ} = \Omega^{KI}\Omega_{KJ} = \Upsilon^{\dot K I}\Upsilon_{\dot K J} = \delta_J^I$$

(and similarly for contracting the second index of each pair). Therefore, with unitarity/(pseudo)reality we can ignore complex conjugate representations (and dotted indices), converting them into unconjugated ones with the metric, while for orthogonality/symplecticity we can do the same with respect to raising/lowering indices:

$$\begin{array}{rl}
\textbf{U}\text{nitary:} & \psi^{\dot I} = \Upsilon^{\dot I J}\psi_J \\
\textbf{O}\text{rthogonal:} & \psi^I = \eta^{IJ}\psi_J \\
\textbf{S}\text{ym}\textbf{p}\text{lectic:} & \psi^I = \Omega^{IJ}\psi_J \\
\textbf{R}\text{eal:} & \psi_{\dot I} = \eta_{\dot I}{}^J\psi_J \\
\text{pseudoreal (*):} & \psi_{\dot I} = \Omega_{\dot I}{}^J\psi_J
\end{array}$$

For the real groups there is also the constraint of reality on the defining representation:

$$\bar\psi_{\dot I} \equiv (\psi_I)^* = \psi_{\dot I} \equiv \eta_{\dot I}{}^J\psi_J$$

Exercise IB5.1

As an example of the advantages of index notation, show that SSp is the same as Sp. (Hint: Write one ϵ in the definition of the determinant in terms of Ω's by total antisymmetrization, which then can be dropped because it is enforced by the other ϵ. One can ignore normalization by just showing $det\ M = det\ I$.)

For SO(n_+,n_-), there is a slight modification of a sign convention: Since then indices can be raised and lowered with the metric, $\epsilon^{I\cdots}$ is usually defined to be the result of raising indices on $\epsilon_{I...}$, which means

$$\epsilon_{12...n} = 1 \quad \Rightarrow \quad \epsilon^{12...n} = det\ \eta = (-1)^{n_-}$$

Then $\epsilon^{I\cdots}$ should be replaced with $(-1)^{n_-}\epsilon^{I\cdots}$ in the equations of subsection IB3: For example,

$$\epsilon_{J_1...J_n}\epsilon^{I_1...I_n} = (-1)^{n_-}\delta^{I_1}_{[J_1}\cdots\delta^{I_n}_{J_n]}$$

We now give the simplest explicit forms for the defining representations of the classical groups. The most convenient notation is to label the generators by a pair of fundamental indices, since the adjoint representation is obtained from the direct product of the fundamental representation and its dual (i.e., as a matrix labeled by row and column). The simplest example is GL(n), since the generators are arbitrary matrices. We therefore choose as a basis matrices with a 1 as one entry and 0's everywhere else, and label that generator by the row and column where the 1 appears. Explicitly,

$$GL(n): \quad (G_I{}^J)_K{}^L = \delta^L_I\delta^J_K \quad \Rightarrow \quad G_I{}^J = |^J\rangle\langle_I|$$

This basis applies for GL(n,C) as well, the only difference being that the coefficients α in $G = \alpha_I{}^J G_J{}^I$ are complex instead of real. The next simplest case is U(n): We can again use this basis, although the matrices $G_I{}^J$ are not all hermitian, by requiring that $\alpha_I{}^J$ be a hermitian matrix. This turns out to be more convenient in practice than using a hermitian basis for the generators. A well known example is SU(2), where the two generators with the 1 as an off-diagonal element (and 0's elsewhere) are known as the "raising and lowering operators" $J_\pm$, and are more convenient than their hermitian parts for purposes of contructing representations. (This generalizes to other unitary groups, where all the generators on one side of the diagonal are raising, all those on the other side are lowering, and those along the diagonal give the maximal Abelian subalgebra, or "Cartan subalgebra".)

Representations for the other classical groups follow from applying their definitions to the GL(n) basis. We thus find

$$SL(n): \quad (G_I{}^J)_K{}^L = \delta^L_I\delta^J_K - \tfrac{1}{n}\delta^J_I\delta^L_K \quad \Rightarrow \quad G_I{}^J = |^J\rangle\langle_I| - \tfrac{1}{n}\delta^J_I|^K\rangle\langle_K|$$
$$SO(n): \quad (G_{IJ})^{KL} = \delta^K_{[I}\delta^L_{J]} \quad \Rightarrow \quad G_{IJ} = |_{[I}\rangle\langle_{J]}|$$
$$Sp(n): \quad (G_{IJ})^{KL} = \delta^K_{(I}\delta^L_{J)} \quad \Rightarrow \quad G_{IJ} = |_{(I}\rangle\langle_{J)}|$$

As before, SL(n,C) and SU(n) use the same basis as SL(n), etc. For SO(n) and Sp(n) we have raised and lowered indices with the appropriate metric (so SO(n) includes

SO(n_+,n_-)). For some purposes (especially for SL(n)), it's more convenient to impose tracelessness or (anti)symmetry on the matrix α, and use the simpler GL(n) basis.

Exercise IB5.2

Our normalization for the generators of the classical groups is the simplest, and independent of n (except for subtracting out traces):

a Find the commutation relations of the generators (structure constants) for the defining representation of GL(n) as given in the text. Note that the values of all the structure constants are 0, $\pm i$. Show that

$$c_D = 1$$

(see subsection IB2).

b Consider the GL(m) subgroup of GL(n) (m<n) found by restricting the range of the index of the above defining representation. Show the structure constants are the same as those given by starting with the above representation of GL(m).

c Find the structure constants for SO(n) and Sp(n).

d Directly evaluate $k_D c_A$ $(= \delta^{ij} G_i G_j)$ for SL(n), SO(n), and Sp(n), and compare with $c_D d_A / d_D$.

Exercise IB5.3

A tensor that pops up in various contexts is

$$d_{ijk} = tr(G_i \{G_j, G_k\})$$

It takes a very simple form in terms of defining indices:

a Show that for SU(n) this tensor is determined to be, up to an overall normalization (that depends on the representation),

$$tr\left(G_{I_1}{}^{J_1}\left\{G_{I_2}{}^{J_2}, G_{I_3}{}^{J_3}\right\}\right) \sim [(231)+(312)] - \frac{2}{n}[(132)+(213)+(321)] + \frac{4}{n^2}(123)$$

$$(abc) \equiv \delta_{I_a}^{J_1} \delta_{I_b}^{J_2} \delta_{I_c}^{J_3}$$

(where a, b, c are some permutation of $1, 2, 3$) from just the total symmetry of d_{ijk} (and $G_I{}^I = 0$), since the only invariant tensor available is δ_I^J. (If $\epsilon_{IJ...}$ were used, $\epsilon^{IJ\cdots}$ would also be required, to balance the number of subscripts and superscripts; but their product can be expressed in terms of just δ's also.)

b Check this result by using the explicit G's for the defining representation, and determine the proportionality constant for that representation.

With the exception of the "spinor" representations of SO(n) (to be discussed in subsection IC5, section IIA, and subsection XC1), general representations can be obtained by reducing direct products of the defining representations. This means they can be described by objects with multiple indices (up/down, dotted/undotted), where each index is that of a defining representation, and satisfying various (anti)symmetry and tracelessness conditions on the indices.

Exercise IB5.4

Consider the representations of SU(n) obtained from the symmetric and antisymmetric part of the direct product of two defining representations. For simplicity, one can work with the U(n) generators, since the U(1) pieces will appear in a simple way.

a Using tensor notation for the generators $(G_I{}^J)_{KL}{}^{MN}$, find their explicit representation for these two representations.

b By evaluating the trace, show that the Dynkin index for the two cases is

$$c_a = n - 2, \qquad c_s = n + 2$$

c Show the sum of these two is consistent with the argument at the end of subsection IB2. Show each case is consistent with n=2, and the antisymmetric case with n=3, by relating those cases to the singlet, defining, and adjoint representations.

REFERENCES

1 H. Georgi, *Lie algebras in particle physics: from isospin to unified theories*, 2nd ed. (Perseus, 1999):
best book on Lie groups; unlike other texts, covers not only more powerful Cartan-Weyl methods, but also more useful tensor methods; also has useful applications to nonrelativistic quark model.

2 S. Helgason, *Differential geometry and symmetric spaces* (Academic, 1962);
R. Gilmore, *Lie groups, Lie algebras, and some of their applications* (Wiley, 1974):
noncompact classical Lie groups.

C. REPRESENTATIONS

We now consider some of the more useful representations, as explicit examples of the results of the previous section. In particular, we consider symmetries of the quark model.

1. More coordinates

We began our "review" of group theory by looking at how symmetries were represented on coordinates. We now return to coordinates as a special case (particular representation) of the general results of the previous section. The idea is that the coordinates themselves are already a representation of the group, and the wave functions are functions of these coordinates. For example, for ordinary rotations we use wave functions that depend on position or momentum, which transforms as a vector. (This is not always the case: For example, in our description of the conformal group the usual space and time coordinates transformed nonlinearly, and not just by multiplication by constant matrices unless the extra two coordinates were introduced.) This is the basic distinction between classical mechanics and classical field theory: Mechanics uses the coordinates themselves as the basic variables, while field theory uses functions of the coordinates. (Similarly, in quantum mechanics the wave functions are functions of the coordinates, while in quantum field theory the wave functions are "functionals" of functions of the coordinates.)

In general, the construction of such a "coordinate representation" starts with a given matrix representation (usually finite dimensional) $(G_i)_I{}^J$ and then defines a new representation

$$\hat{G}_i = q^I (G_i)_I{}^J p_J; \qquad [p_I, q^J\} = \delta_I^J, \quad [q,q\} = [p,p\} = 0$$

for some objects q and p, which are interpreted as either coordinates and their conjugate momenta (up to a factor of i), or as creation and annihilation operators: The latter nomenclature is used when the boundary conditions allow the existence of a state $|0\rangle$ called the "vacuum", satisfying $p|0\rangle = 0$, so we can define the other states as functions of q acting on $|0\rangle$. (If the coordinates are fermionic, the distinction is moot, since by the usual Taylor expansion the Hilbert space is finite dimensional. See exercise IA2.3.) It is easy to check that $\hat{G}_i$ satisfy the same commutation relations as G_i. In particular, if the matrices are in the adjoint representation, q^i can be interpreted as the group coordinates themselves: This follows from considering the action of an infinitesimal transformation on the group element $g(q) = e^{iq^i G_i}$ (or just the Lie algebra element $G(q) = q^i G_i$).

If we write these results in bra/ket notation, since

$$\hat{G}_i q^I = q^J (G_i)_J{}^I, \quad \hat{G}_i p_I = -(G_i)_I{}^J p_J$$

it is more natural to look at the action on bras:

$$\langle q| = q^I \langle_I|, \quad |p\rangle = |^I\rangle p_I \quad \Rightarrow \quad \hat{G}_i \langle q| = \langle q|G_i, \quad \hat{G}_i|p\rangle = -G_i|p\rangle$$

Note that this vector space is coordinate space itself, not the space of functions of the coordinates; it is the same space on which G_i is defined. (Of course, $\hat{G}_i$ is defined on arbitrary functions of the coordinates; it has a reducible representation bigger than $(G_i)_I{}^J$. Effectively, $(G_i)_I{}^J$ is represented on the space of functions linear in the coordinates.) Then, for example

$$\hat{G}_1 \hat{G}_2 \langle q| = \hat{G}_1 \langle q|G_2 = \langle q|G_1 G_2$$

is obviously equivalent, while (ignoring any extra signs for fermions)

$$\hat{G}_1 \hat{G}_2 |p\rangle = -\hat{G}_1 G_2 |p\rangle = -G_2 \hat{G}_1 |p\rangle = G_2 G_1 |p\rangle$$

at least gives an equivalent result for the commutator algebra $[\hat{G}_1, \hat{G}_2]$. This is the expected result for the dual representation $G_i \to -G_i^T$.

Interesting examples are given by using the defining representation for G. For example, the commonly used oscillator representation for U(n) is

$$U(n): \quad \hat{G}_I{}^J = a^{\dagger J} a_I, \quad [a_I, a^{\dagger J}\} = \delta_I^J$$

where the oscillators can be bosonic or fermionic. For the SO and Sp cases, because we can raise and lower indices, and because of the (anti)symmetry on the indices, the interesting possibility arises to identify the coordinates with their momenta, with the statistics appropriate to the symmetry:

$$Sp(n): \quad \hat{G}_{IJ} = \tfrac{1}{2} z_{(I} z_{J)}, \quad [z_I, z_J] = \Omega_{IJ}$$

$$SO(n): \quad \hat{G}_{IJ} = \tfrac{1}{2} \gamma_{[I} \gamma_{J]}, \quad \{\gamma_I, \gamma_J\} = \eta_{IJ}$$

For SO(n) the representation is finite dimensional because of the Fermi-Dirac statistics, and is called a "Dirac spinor" (and γ the "Dirac matrices"). If the opposite statistics are chosen, the coordinates and momenta can't be identified: For example, bosonic coordinates for SO(n) give the usual spatial rotation generators $\hat{G}_{IJ} = x_{[I} \partial_{J]}$.

Exercise IC1.1

Use this bosonic oscillator representation for U(2)=SU(2)⊗U(1), and use the SU(2) subgroup to describe spin.

a Show that the spin s (the integer or half-integer number that defines the representation) itself has a very simple expression in terms of the U(1) generator. Show this holds in the quantum mechanical case (by interpreting the bracket as the quantum commutator), giving the usual $s(s+1)$ for the sum of the squares of the generators (with appropriate normalization).

b Use this result to show that these oscillators, acting on the vacuum state, can be used to construct the usual states of arbitrary spin s.

Exercise IC1.2

Considering SO(2n), divide up γ_I into pairs of canonical (and complex) conjugates $a_1 = (\gamma_1 + i\gamma_2)/\sqrt{2}$, etc., so $\{a, a^\dagger\} = 1$.

a Write the SO(2n) generators in terms of aa, $a^\dagger a^\dagger$, and $a^\dagger a$. Show that the $a^\dagger a$'s by themselves generate a U(n) subgroup.

b Decompose the Dirac spinor into U(n) representations. Show that the product of all the γ's is related to the U(1) generator, and commutes with all the SO(2n) generators. Show that the states created by even or odd numbers of $a^\dagger$'s on the vacuum don't mix with each other under SO(2n), so the Dirac spinor is reducible into two "Weyl spinors".

2. Coordinate tensors

We have just seen how groups can be represented on coordinates. Depending on the choice of coordinates, the coordinates may transform nonlinearly (i.e., as a realization, not a representation), as for the D-dimensional conformal group in terms of D (not D+2) coordinates. However, given the nonlinear transformation of the coordinates, there are always representations other than the defining one (scalar field) that we can immediately write down (such as the adjoint). We now consider such representations: These are useful not only for the spacetime symmetries we have already considered, but also for general relativity, where the symmetry group consists of arbitrary coordinate transformations. Furthermore, these considerations are useful for describing coordinate transformations that are not symmetries, such as the change from Cartesian to polar coordinates in nonrelativistic theories.

When applied to quantum mechanics, we write the action of a symmetry on a state as $\delta\psi = iG\psi$ (or $\psi' = e^{iG}\psi$), but on an operator as $\delta A = i[G, A]$ (or $A' = e^{iG}Ae^{-iG}$). In classical mechanics, we always write $\delta A = i[G, A]$ (since classical objects are identified with quantum operators, not states). However, if $G = \lambda^m \partial_m$ is a coordinate transformation (e.g., a rotation) and ϕ is a scalar field, then in quantum notation we can write

$$\delta\phi(x) = [G, \phi] = G\phi = \lambda^m \partial_m \phi \qquad (\phi' = e^G \phi e^{-G} = e^G \phi)$$

since the derivatives in G just differentiate ϕ. (For this discussion of coordinate transformations we switch to absorbing the i's into the generators.) The coordinate transformation G has the usual properties of a derivative:

$$[G, f(x)] = Gf \quad \Rightarrow \quad Gf_1 f_2 = [G, f_1 f_2] = (Gf_1)f_2 + f_1 G f_2$$

$$e^G f_1 f_2 = e^G f_1 f_2 e^{-G} = (e^G f_1 e^{-G})(e^G f_2 e^{-G}) = (e^G f_1)(e^G f_2)$$

and similarly for products of more functions.

The adjoint representation of coordinate transformations is a "vector field" (in the sense of a spatial vector), a function that has general dependence on the coordinates (like a scalar field) but is also linear in the momenta (as are the Poincaré generators):

$$G = \lambda^m(x)\partial_m, \quad V = V^m(x)\partial_m \quad \Rightarrow \quad \delta V = [G, V] = (\lambda^m \partial_m V^n - V^m \partial_m \lambda^n)\partial_n$$

$$\Rightarrow \quad \delta V^m = \lambda^n \partial_n V^m - V^n \partial_n \lambda^m$$

The same result follows if we use the Poisson bracket instead of the quantum mechanical commutator, replacing ∂_m with ip_m in both G and V.

Finite transformations can also be expressed in terms of transformed coordinates themselves, instead of the transformation parameter:

$$\phi(x) = e^{-\lambda^m \partial_m} \phi'(x) = \phi'(e^{-\lambda^m \partial_m} x)$$

as seen, for example, from a Taylor expansion of ϕ', using $e^{-G}\phi' = e^{-G}\phi' e^G$. We then define

$$\phi'(x') = \phi(x) \quad \Rightarrow \quad x' = e^{-\lambda^m \partial_m} x$$

This is essentially the statement that the active and passive transformations cancel. However, in general this method of defining coordinate transformations is not convenient for applications: When we make a coordinate transformation, we want to know $\phi'(x)$. Working with the "inverse" transformation on the coordinates, i.e., our original e^{+G},

$$\tilde{x} \equiv e^{+\lambda^m \partial_m} x \quad \Rightarrow \quad \phi'(x) = e^G \phi(x) = \phi(\tilde{x}(x))$$

So, for finite transformations, we work directly in terms of $\tilde{x}(x)$, and simply plug this into ϕ in place of x $(x \to \tilde{x}(x))$ to find ϕ' as a function of x.

Similar remarks apply for the vector, and for derivatives in general. We then use

$$x' = e^{-G}x \quad \Rightarrow \quad \partial' = e^{-G}\partial e^{G}$$

where $\partial' = \partial/\partial x'$, since $\partial' x' = \partial x = \delta$. This tells us

$$V^m(x)\partial_m = e^{-G}V'^m(x)\partial_m e^G = V'^m(x')\partial'_m$$

or $V'(x') = V(x)$. Acting with both sides on x'^m,

$$V'^m(x') = V^n(x)\frac{\partial x'^m}{\partial x^n}$$

On the other hand, working in terms of $\tilde{x}$ is again more convenient: Changing the transformation for the vector operator in the same way as the scalar

$$V'(x) = V(\tilde{x})$$

$$\Rightarrow \quad V'^m(x)\partial_m = V^m(\tilde{x})\tilde{\partial}_m$$

$$\Rightarrow \quad V'^m(x) = V^n(\tilde{x})\frac{\partial x^m}{\partial \tilde{x}^n}$$

where $\tilde{\partial} = \partial/\partial\tilde{x}$ and as usual

$$(\tilde{\partial}_m x^n)(\partial_n \tilde{x}^p) = \delta_m^p \quad \Rightarrow \quad \frac{\partial x^n}{\partial \tilde{x}^m} = \left[\left(\frac{\partial \tilde{x}(x)}{\partial x}\right)^{-1}\right]_m{}^n$$

We can also use

$$V'(x) = e^G V(x) e^{-G}$$

$$\Rightarrow \quad V'^m(x)\partial_m = (e^G V^m(x) e^{-G})(e^G \partial_m e^{-G}) = V^m(\tilde{x})\tilde{\partial}_m$$

A "differential form" is defined as an infinitesimal $W = dx^m W_m(x)$. Its transformation law under coordinate transformations, like that of scalar and vector fields, is defined by $W'(x') = W(x)$. For any vector field $V = V^m(x)\partial_m$, $V^m W_m$ transforms as a scalar, as follows from the "chain rule" $d = dx'^m\partial'_m = dx^m\partial_m$. Explicitly,

$$W'_m(x') = W_n(x)\frac{\partial x^n}{\partial x'^m}$$

or in infinitesimal form

$$\delta W_m = \lambda^n \partial_n W_m + W_n \partial_m \lambda^n$$

Thus a differential form is dual to a vector, at least as far as the matrix part of coordinate transformations is concerned. They transform the same way under rotations, because rotations are orthogonal; however, more generally they transform differently, and in the absence of a metric there is not even a way to relate the two by raising or lowering indices.

Higher-rank differential forms can be defined by antisymmetric products of the above "one-forms". These are useful for integration: Just as the line integral $\int W = \int dx^m W_m$ is invariant under coordinate transformations by definition (as long as we choose the curve along which the integral is performed in a coordinate-independent way), so is a totally antisymmetric Nth-rank tensor ("N-form") $W_{m_1 \cdots m_N}$ integrated on an N-dimensional subspace as

$$\int dx^{m_1} \cdots dx^{m_N} W_{m_1 \cdots m_N} : \qquad W'_{m_1 \cdots m_N}(x') = W_{p_1 \cdots p_N}(x) \frac{\partial x^{p_1}}{\partial x'^{m_1}} \cdots \frac{\partial x^{p_N}}{\partial x'^{m_N}}$$

where the surface element $dx^{m_1} \cdots dx^{m_N}$ is interpreted as antisymmetric. (The signs come from switching initial and final limits of integration, as prescribed by the "orientation" of the hypersurface.) This is clear if we rewrite the integral more explicitly in terms of coordinates σ^i for the subspace: Then

$$\int dx^{m_1} \cdots dx^{m_N} W_{m_1 \cdots m_N}(x) = \int d\sigma^{i_1} \cdots d\sigma^{i_N} \widehat{W}_{i_1 \cdots i_N}(\sigma) = \int d^N \sigma \; \epsilon^{i_1 \cdots i_N} \widehat{W}_{i_1 \cdots i_N}(\sigma)$$

where

$$\widehat{W}_{i_1 \cdots i_N}(\sigma) = \frac{\partial x^{m_1}}{\partial \sigma^{i_1}} \cdots \frac{\partial x^{m_N}}{\partial \sigma^{i_N}} W_{m_1 \cdots m_N}(x)$$

is the result of a coordinate transformation that converts N of the x's to σ's, an interpretation of the functions $x(\sigma)$ that define the surface. Then any coordinate transformation on $x \to x'$ (not on σ) will leave $\widehat{W}(\sigma)$ invariant. In particular, if the subspace is the full space, so we can look directly at $\int d^N x \; \epsilon^{m_1 \cdots m_N} W_{m_1 \cdots m_N}$, we see that a coordinate transformation generates from W an N-dimensional determinant exactly canceling the Jacobian resulting from changing the integration measure $d^N x$.

Exercise IC2.1

For all of the following, use the exponential form of the finite coordinate transformation:

a Show that any (local) function of a scalar field (without explicit x dependence additional to that in the field) is also a scalar field (i.e., satisfies the same coordinate transformation law).

b Show that the transformation law of a vector field or differential form remains the same when multiplied by a scalar field (at the same x).

c Show that $V\phi = V^m \partial_m \phi$ is a scalar field for any scalar field ϕ and vector field V.

d Show that $[V, W]$ is a vector field for any vector fields V and W.

Exercise IC2.2

Examine finite coordinate transformations for integrals of differential forms in terms of $\tilde{x}$ rather than x'. Find the explicit expression for $W'(x)$ in terms of $W(\tilde{x}(x))$, etc., and use this to show invariance:

$$\int dx^{m_1}\cdots dx^{m_N}W'_{m_1\cdots m_N}(x) = \int d\tilde{x}^{m_1}\cdots d\tilde{x}^{m_N}W_{m_1\cdots m_N}(\tilde{x})$$

$$= \int dx^{m_1}\cdots dx^{m_N}W_{m_1\cdots m_N}(x)$$

where in the last step we have simply substituted $\tilde{x}\to x$ as a change of integration variables. Note that, using the $\tilde{x}$ form of the transformation rather than x', the transformation generates the needed Jacobian, rather than canceling one.

From the above transformation law, we see that the curl of a differential form is also a differential form:

$$\partial'_{[m_1}W'_{m_2\cdots m_N]}(x') = \partial'_{[m_1}(\partial'_{m_2}x^{p_2})\cdots(\partial'_{m_N]}x^{p_N})W_{p_2\cdots p_N}(x)$$

$$= [\partial_{[p_1}W_{p_2\cdots p_N]}(x)](\partial'_{m_1}x^{p_1})\cdots(\partial'_{m_N}x^{p_N})$$

because the curl kills $\partial'\partial' x$ terms that would appear if there were no antisymmetrization. Objects that transform "covariantly" under coordinate transformations, without such higher derivatives of x (or λ in the other notation), like scalars, vectors, differential forms and their products, are called (coordinate) "tensors". Getting derivatives of tensors to come out covariant in general requires special fields, and will be discussed in chapter IX. An important application of the covariance of the curl of differential forms is the generalized Stokes' theorem (which includes the usual Stokes' theorem and Gauss' law as special cases):

$$\int dx^{m_1}\cdots dx^{m_{N+1}}\tfrac{1}{(N+1)!}\partial_{[m_1}W_{m_2\cdots m_{N+1}]} = \oint dx^{m_1}\cdots dx^{m_N}W_{m_1\cdots m_N}$$

where the second integral is over the boundary of the space over which the first is integrated. (We use the symbol "$\oint$" to refer to boundary integrals, including those over contours, which are closed boundaries of 2D surfaces.) It is basically just the fundamental theorem of (integral) calculus ($\int_a^b dx\ f'(x) = f(b) - f(a)$), as is clear from choosing a coordinate system where the boundary is at a fixed value of just one coordinate (at least in patches). (A standard example is a pair of infinite constant-time surfaces, neglecting the boundaries that connect them at spatial infinity.)

3. Young tableaux

We now return to our discussion of finite-dimensional representations. In the previous section we gave the machinery for describing them using index notation, but examined only the defining representation in detail. Now we analyze general irreducible representations.

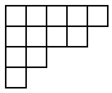
All the irreducible finite-dimensional representations of the groups SU(N) can be described by tensors with lower N-valued indices with various (anti)symmetrizations. (An upper index can be replaced with N−1 lower indices by using the Levi-Civita tensor.) Although detailed calculations require explicit use of these indices, three properties can be more conveniently discussed pictorially:

(1) the (anti)symmetries of the indices,

(2) the dimension (number of independent components) of the representation, and

(3) the reduction of the direct product of two representations (which irreducible representations result, and how many of each).

A "Young tableau" is a picture representing an irreducible representation in terms of boxes arranged in a regular grid into rows and columns, such that the columns are aligned at the top, and their depths are nonincreasing to the right: for example,



Each box represents an index, with antisymmetry among indices in any column, and symmetry among indices in any row. More precisely, since one can't simultaneously have these symmetries and antisymmetries, it corresponds to the result of taking any arbitrary tensor with that many indices, first symmetrizing the indices in each row, and then antisymmetrizing the indices in each column (or vice versa; symmetrizing and then antisymmetrizing and then symmetrizing again gives the same result as skipping the first symmetrization, etc.). This gives a simple way to classify and symbolize each representation. (We can denote the singlet representation, which has no boxes, by a dot.) Note that the deepest column should have no more than N−1 boxes for SU(N) because of the antisymmetry.

To calculate the dimension of the representation for a given tableau, we use the "factors over hooks" rule:

(1) Write an "N" in the box in the upper-left corner, and fill the rest of the boxes with numbers that decrease by 1 for each step down and increase by 1 for each step to the right.

(2) Draw (or picture in your mind) a "hook" for each box — a "Γ" with its corner in the box and lines extending right and down out of the tableau.

(3) The dimension is then given by the formula

$$\text{dimension} = \prod_{\text{each box}} \frac{\text{integer written there}}{\#\text{ boxes intersected by its hook}}$$

For the previous example, we find (listing boxes first down and then to the right)

$$\frac{N}{8}\cdot\frac{N-1}{6}\cdot\frac{N-2}{3}\cdot\frac{N-3}{1}\cdot\frac{N+1}{6}\cdot\frac{N}{4}\cdot\frac{N-1}{1}\cdot\frac{N+2}{4}\cdot\frac{N+1}{2}\cdot\frac{N+3}{3}\cdot\frac{N+2}{1}\cdot\frac{N+4}{1}$$

The direct product of two Young tableaux A⊗B is analyzed by the following rules: First, label all the boxes in B by putting an "a" in each box in the top row, "b" in the second row, etc. Then, take the following steps in all possible ways to find the Young tableaux resulting from the direct product:

(1) Add all the "a" boxes from B to the right side and bottom of A, then "b" to the right and bottom of that, etc., to make a new Young tableaux. Any two tableaux constructed in this way with the same arrangement of boxes but different assignment of letters are considered distinct, i.e., multiple occurences of the same representation in the direct product.

(2) No more than 1 "a" can be in any column, and similarly for the other letters.

(3) Reading from right to left, and then from top to bottom (i.e., like Hebrew/Arabic), the number of a's read should always be $\geq$ the number of b's, b's $\geq$ c's, etc.

For example,

$$\ydiagram{1} \otimes \ytableaushort{aa,b} = \ytableaushort{{}aa,b} \oplus \ytableaushort{{}a,ab} \oplus \ytableaushort{{}a,a,b}$$

Note that A⊗B always gives the same result as B⊗A, but one way may be simpler than the other. For a given value of N, a column of N boxes is equivalent to none (again by antisymmetry), while more than N boxes in a column gives a vanishing tableau.

Exercise IC3.1

Calculate

$$\ydiagram{2,1} \otimes \ydiagram{2,1}$$

Check the result by finding the dimensions of all the representations and adding them up.

These SU(N) tableaux also apply to SL(N): Only the reality properties are different. Similar methods can be applied to USp(2N) (or Sp(2N)), but tracelessness (with respect to the symplectic metric) must be imposed in antisymmetrized indices, so these trace pieces must be separated out when considering the above rules. (I.e., consider USp(2N)⊂SU(2N).) Similar remarks apply to SO(N), which has a symmetric metric, but there are also "spinor" representations (see below). The additional irreducible representations then can be constructed from taking direct products of the above with the smallest spinors, and removing the "gamma-matrix" traces. Furthermore, using the Levi-Civita tensor, all columns can be reduced to no more than N/2 in height.

4. Color and flavor

We now consider the application of these methods to "internal symmetries" (those that don't act on the coordinates) in particle physics. The symmetries with experimental confirmation involve only the unitary groups (U and SU) of small dimension. However, we will find later that larger unitary groups can be useful for approximation schemes. (Also, larger unitary and other groups continue to be investigated for unification and other purposes, which we consider in later chapters.)

The "Standard Model" describes all of particle physics that is well confirmed experimentally (except gravity, which is not understood at the quantum level). It includes as its "fundamental" particles:

(1) the spin-1/2 quarks that make up the observed strongly interacting particles, but do not exist as asymptotic states,

(2) the weakly interacting spin-1/2 leptons,

(3) the spin-1 gluons that bind the quarks together, which couple to the charges associated with SU(3) "color" symmetry, but also are not asymptotic,

(4) the spin-1 particles that mediate the weak and electromagnetic interactions, which couple to SU(2)⊗U(1) "flavor", and

(5) the yet unobserved spin-0 Higgs particles that are responsible for all the masses of these weakly interacting particles.

(However, quarks and gluons are temporarily free at high energy, eventually recombining to give rise to "jets", clusters of resulting hadrons.) These particles, along with their masses (in GeV) and (electromagnetic) charges ($Q = \bar{Q} + \Delta Q$), are:

$s=\frac{1}{2}$		
color: $\to$ flavor (ΔQ)	quark (3) ($\bar{Q}=\frac{1}{6}$)	lepton (1) ($\bar{Q}=-\frac{1}{2}$)
$-\frac{1}{2}$ $+\frac{1}{2}$	d (.007) u (.003)	e (.00051099890) ν_e ($< 3\cdot 10^{-9}$)
$-\frac{1}{2}$ $+\frac{1}{2}$	s (.12) c (1.2)	μ (.105658357) ν_μ (<.00019)
$-\frac{1}{2}$ $+\frac{1}{2}$	b (4.2) t (174)	τ (1.7770) ν_τ (<.0182)

$s=1$		
color: $\to$ flavor (Q)	gluon (8)	electroweak (1)
0 0 ± 1	g (0)	γ ($< 2\cdot 10^{-25}$) Z (91.188) W (80.42)

$s=0$		
($Q=0$)	H	(>114.3)

The quark masses we have listed are the "current quark masses", the effective masses when the quarks are relativistic with respect to their hadron (at least for the lighter quarks), and act as almost free. But since they are not free, their masses are ambiguous and energy dependent, and defined by some convenient conventions. Nonrelativistic quark models use instead the "constituent quark masses", which include potential energy from the gluons. This extra potential energy is about .30 GeV per quark in the lightest mesons, .35 GeV in the lightest baryons; there is also a contribution to the binding energy from spin-spin interaction. Unlike electrodynamics, where the potential energy is negative because the electrons are free at large distances, where the potential levels off (the top of the "well"), in chromodynamics the potential energy is positive because the quarks are free at high energies (short distances, the bottom of the well), and the potential is infinitely rising. Masslessness of the gluons is implied by the fact that no colorful asymptotic states have ever been observed. We have divided the spin-1/2 particles into 3 "families" with the same quantum numbers (but different masses). Within each family, the quarks are similar to the leptons, except that:

(1) the masses and average charges ($\bar{Q}$) are different,

(2) the quarks come in 3 colors, while the leptons are colorless, and

(3) the neutrinos, to within experimental error, are massless, so they have half as many components as the massive fermions (1 helicity state each, instead of 2 spin states each).

This means that each lepton family has 1 SU(2) doublet and 1 SU(2) singlet. For symmetry (and better, quantum mechanical, reasons to be explained later), we also assume the quarks have 1 SU(2) doublet, but therefore 2 SU(2) singlets. (Some experiments have indicated small masses for neutrinos: This would require generalization of the Standard Model, such as models with parity broken by interactions. Some examples of such theories will be discussed in subsection IVB4.)

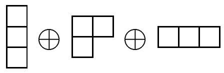
We first look at the color group theory of the physical states, which are color singlets. The fundamental unobserved particles are the spin-1 "gluons", described by the Yang-Mills gauge fields, and the spin-1/2 quarks. Suppressing all but color indices, we denote the quark states by q^i, and the antiquarks by $q^\dagger{}_i$, where the indices are those of the defining representation of SU(n), and its complex conjugate. The quarks also carry a representation of a "flavor" group, unlike the gluons. The simplest flavorful states are those made up of only (anti)quarks, with indices completely contracted by one factor of an SU(n) group metric: From the "U" of SU(n), we can contract defining indices with their complex conjugates, giving the "mesons", described by $q^\dagger{}_i q^i$ (quark-antiquark), which are their own antiparticles. From the "S" of SU(n), we have the "baryons", described by $\epsilon_{i_1 \dots i_n} q^{i_1} \dots q^{i_n}$ (n-quark), and the antibaryons, described by the complex conjugate fields. All other colorless states made of just (anti)quarks can be written as products of these fields, and therefore considered as describing composites of them. Thus, we can approximate the ground states of the mesons by $q^\dagger{}_i(x) q^i(x)$, which describe spins 0 and 1 because of the various combinations of spins (from $\frac{1}{2} \otimes \frac{1}{2} = 0 \oplus 1$). The first excited level will then be described by $q^\dagger \overset{\leftrightarrow}{\partial} q$ (where

$$A \overset{\leftrightarrow}{\partial} B \equiv A \partial B - (\partial A) B$$

and picks out the relative momentum of the two quarks): It includes spins 0, 1, and 2, etc., where each derivative introduces orbital angular momentum 1. (Similar remarks apply to baryons.) We can also have flavorless states made from just gluons, called "glueballs": The ground states can be described by $F_i{}^j F_j{}^i$, where each F is a gluon state (in the adjoint representation of SU(n)), and includes spins 0 and 2 (from the symmetric part of $1 \otimes 1$). Because of their flavor multiplets and (electroweak) interactions, many mesons and baryons corresponding to such ground and excited states have been experimentally identified, while the glueballs' existence is still uncertain. Actually, quarks and gluons can almost be observed independently at high energies, where the "strong" interaction is weak: The energetic particle appears as a "jet" — a particle of high energy accompanied by particles of much lower energy (perhaps too small to detect) in color-singlet combinations. (Depending on the available decay modes, the jet might not be observed until after decaying, but still within a small angle of spread.)

We now look at the flavor group theory of the physical hadronic states. In contrast to the previous paragraph, we now suppress all but the flavor indices. Mesons $M_i{}^j =$

$q^\dagger{}_i q^j$ are thus in the adjoint representation of flavor U(m) ($m \otimes \bar{m}$, where m is the defining representation and $\bar{m}$ its complex conjugate), for both the spin-0 and the spin-1 ground states. The baryons are more complicated: For simplicity we consider SU(3) color, which accurately describes physics at observed energies. Then the color structure described above results in total symmetry in combined flavor and Lorentz indices (from the antisymmetry in the color indices, and the overall antisymmetry for Fermi-Dirac statistics). Thus, for the 3-quark baryons, the Young tableaux



for SU(m) flavor are accompanied by the same Young tableaux for spin indices: In nonrelativistic notation, the first tableau, being totally antisymmetric in flavor indices, is also totally antisymmetric in the three two-valued spinor indices, and thus vanishes. Similarly, the last tableau describes spin 3/2 (total symmetry in both types of indices), while the middle one describes spin 1/2. Since only 3 flavors of quarks have small masses compared to the hadronic mass scale, hadrons can be most conveniently grouped into flavor multiplets for SU(3) flavor: The ground states are then, in terms of SU(3) flavor multiplets, 8⊕1 for the pseudoscalars, 8⊕1 for the vectors, 8 for spin 1/2, and 10 for spin 3/2.

Exercise IC4.1

What SU(flavor) Young tableaux, corresponding to what spins, would we have for mesons and baryons if there were

a 2 colors?

b 4 colors?

However, the differing masses of the different flavors of quarks break the SU(3) flavor symmetry (as does the weak interaction). In particular, the mass eigenstates tend to be pure states of the various combinations of the different flavors of quarks, rather than the linear combinations expected from the flavor symmetry. Specifically, the linear combinations predicted by an 8⊕1 separation for mesons (trace and traceless pieces of a 3⊗3 matrix) are replaced with particles that are more accurately described by a particular flavor of quark bound to a particular flavor of antiquark. (This is known as "ideal mixing".) The one exception is the lighest mesons (pseudoscalars), which are more accurately described by the 8⊕1 split, for this restriction to the 3 lighter flavors of quarks, but the mass of the singlet differs from that naively expected from group theory or nonrelativistic quark models. (This is known as the "U(1) problem".) The solution is probably that the singlet mixes strongly with the lightest psuedoscalar glueball (described by $tr\ \epsilon^{abcd} F_{ab} F_{cd}$); the mass eigenstates are linear combinations of these two fields with the same quantum numbers. In any case, the most convenient notation for labeling the entries of the matrix $M_i{}^j$ representing the various meson states for any particular spin and angular momentum of the quark-antiquark combination is that corresponding to the choice we gave earlier for the generators of U(n): Label each entry by a separate name, where the complex conjugate appears reflected across the diagonal. These directly correspond to the combination of a particular quark with a particular antiquark, and to the mass eigenstates, with the possible exception of the entries along the diagonal for the 3 lightest flavors, where the mass eigenstates are various linear combinations. (However, the SU(2) of the 2 lightest flavors is only slightly broken by the quark masses, so in that case the combinations are very close to the 3⊕1 split of SU(2).)

For example, for the lightest multiplet of mesons (spin 0, and relative angular momentum 0 for the quark and antiquark, but not all of which have yet been observed), we can write the U(6) matrix (for the 6 flavors of the 3 known families)

$$M_i{}^j = \begin{pmatrix} \bar{d}d & \bar{d}u & \bar{d}s & \bar{d}c & \bar{d}b & \bar{d}t \\ \bar{u}d & \bar{u}u & \bar{u}s & \bar{u}c & \bar{u}b & \bar{u}t \\ \bar{s}d & \bar{s}u & \bar{s}s & \bar{s}c & \bar{s}b & \bar{s}t \\ \bar{c}d & \bar{c}u & \bar{c}s & \bar{c}c & \bar{c}b & \bar{c}t \\ \bar{b}d & \bar{b}u & \bar{b}s & \bar{b}c & \bar{b}b & \bar{b}t \\ \bar{t}d & \bar{t}u & \bar{t}s & \bar{t}c & \bar{t}b & \bar{t}t \end{pmatrix}$$

$$= \begin{array}{c} \\ \bar{d} \\ \bar{u} \\ \bar{s} \\ \bar{c} \\ \bar{b} \\ \bar{t} \end{array} \begin{array}{c} \begin{array}{llllll} d & u & s & c & b & t \end{array} \\ \begin{pmatrix} \eta_d & \pi^+(.1395702) & \bar{K}^0(.49767) & D^+(1.8693) & \bar{B}^0(5.2794) & T^+ \\ \pi^-(") & \eta_u & K^-(.49368) & D^0(1.8645) & B^-(5.2790) & T^0 \\ K^0(") & K^+(") & \eta_s & D_s^+(1.9685) & \bar{B}_s^0(5.370) & T_s^+ \\ D^-(") & \bar{D}^0(") & D_s^-(") & \eta_c(2.980) & B_c^-(6.4) & T_c^0 \\ B^0(") & B^+(") & B_s^0(") & B_c^+(") & \eta_b & T_b^+ \\ T^- & \bar{T}^0 & T_s^- & \bar{T}_c^0 & T_b^- & \eta_t \end{pmatrix} \end{array}$$

where (approximately)

$$\eta_u = \tfrac{1}{\sqrt{2}}\pi^0(.1349766) + \tfrac{1}{2}[\eta'(.9578) + \eta(.5473)]$$

$$\eta_d = -\tfrac{1}{\sqrt{2}}\pi^0 + \tfrac{1}{2}(\eta' + \eta), \quad \eta_s = \tfrac{1}{\sqrt{2}}(\eta' - \eta)$$

in terms of the mass eigenstates (observed particles), with masses again in GeV, and ditto marks refer to the transposed entry. (We have neglected the important contribution from the glueball.) For the corresponding spin-1 multiplet,

$$\widetilde{M}_i{}^j = \begin{array}{c} \\ \bar{d} \\ \bar{u} \\ \bar{s} \\ \bar{c} \\ \bar{b} \\ \bar{t} \end{array} \!\!\! \begin{array}{c} \begin{array}{llllll} \quad d & \qquad u & \qquad s & \qquad c & \qquad b & \qquad t \end{array} \\ \begin{pmatrix} \omega_d & \rho^+(.7711) & \bar{K}^{*0}(.8961) & D^{*+}(2.0100) & \bar{B}^{*0}(5.3250) & T^{*+} \\ \rho^-(") & \omega_u & K^{*-}(.8917) & D^{*0}(2.0067) & B^{*-}(5.3250) & T^{*0} \\ K^{*0}(") & K^{*+}(") & \phi(1.01946) & D^{*+}_s(2.1124) & \bar{B}^{*0}_s & T^{*+}_s \\ D^{*-}(") & \bar{D}^{*0}(") & D^{*-}_s(") & J/\psi(3.09687) & B^{*-}_c & T^{*0}_c \\ B^{*0}(") & B^{*+}(") & B^{*0}_s & B^{*+}_c & \Upsilon(9.4603) & T^{*+}_b \\ T^{*-} & \bar{T}^{*0} & T^{*-}_s & \bar{T}^{*0}_c & T^{*-}_b & \theta \end{pmatrix} \end{array}$$

where

$$\omega_u = \tfrac{1}{\sqrt{2}}[\omega(.7826) + \rho^0(.7711)], \quad \omega_d = \tfrac{1}{\sqrt{2}}(\omega - \rho^0)$$

(with $\bar{s}s = \phi$, ideal mixing, also approximate).

5. Covering groups

The orthogonal groups O(n_+,n_-) are of obvious interest for describing Lorentz symmetry in spacetimes with n_+ space and n_- time dimensions, or conformal symmetry in spacetimes with n_+-1 space and n_--1 time dimensions. This means we should be interested in O(n) for n≤6, and their "Wick rotations": transformations that put in extra factors of i to change some signs on the metric. Coincidentally, these are just the cases where the Lie algebras of the orthogonal groups are equivalent to those of some algebras for smaller matrices. The smaller representation then can be identified as the "spinor" representation of that orthogonal group. Since the "vector", or defining representation space of the orthogonal group, itself is represented as a matrix with respect to the other group (i.e., the state carries two spinor indices), the other group may include certain phase transformations (such as -1) that cancel in the transformation of the vector. The other group is then called the "covering" group for that orthogonal group, since it includes those missing transformations in its defining representation. (As a result, its group space also has a more interesting topology, which we won't discuss here.)

One way to discover these covering groups is to first count generators, then try to construct explicitly the orthogonal metric on matrices. SO(n) has n(n−1)/2 generators (antisymmetric matrices), Sp(n) has n(n+1)/2 (symmetric), and SU(n) has n^2-1 (traceless). (These are hermitian generators, since we applied reality or hermiticity.) So, for some group SO(n), we look for another group that has the same number of generators. Then, if the new group is defined on m×m matrices, we look for conditions to impose on an m×m matrix (not necessarily the adjoint) to get an n-component representation. This is easy to do by inspection for small n; for large n it's easy to see that it can't work, since m will be of the order of n, and the simple constraints will give of the order of n^2 components instead of n. We then construct the norm of this matrix M as $tr(M^\dagger M)$, which is just the sum of the absolute value squared of the components, for SO(n), and the other orthogonal groups by Wick rotation. (Wick rotation affects mainly the reality conditions on M.)

The identifications for the Lie algebras are then:

SO(2) = U(1), SO(1,1) = GL(1)
SO(3) = SU(2) = SU*(2) = USp(2), SO(2,1) = SU(1,1) = SL(2) = Sp(2)
SO(4) = SU(2)⊗SU(2), SO(3,1) = SL(2,C) = Sp(2,C), SO(2,2) = SL(2)⊗SL(2)
SO(5) = USp(4), SO(4,1) = USp(2,2), SO(3,2) = Sp(4)
SO(6) = SU(4), SO(5,1) = SU*(4), SO(4,2) = SU(2,2), SO(3,3) = SL(4)

Note that the Euclidean cases are all unitary, while the ones with (almost) equal numbers of space and time dimensions are all real. There are also some similar relations for the pseudoreal orthogonal groups:

SO*(2) = U(1), SO*(4) = SU(2)⊗SL(2), SO*(6) = SU(3,1), SO*(8) = SO(6,2)

The norm and conditions for an m-spinor of SO(n_+,n_-) are:

			$n_- \Rightarrow$	0	1	2	3
m	n	$norm$	$symmetry: z^T =$	$reality: \; z^* =$			
1	2	$z'z$		z'	$z\ (z'^* = z')$		
2	3	$z^{\alpha\beta}z^{\gamma\delta}\epsilon_{\gamma\alpha}\epsilon_{\delta\beta}$	z	$-\epsilon z\epsilon$	z		
	4	$z^{\alpha\beta'}z^{\gamma\delta'}\epsilon_{\gamma\alpha}\epsilon_{\delta'\beta'}$		$-\epsilon z\epsilon$	z^T	z	
4	5	$z^{\alpha\beta}z^{\gamma\delta}\epsilon_{\delta\gamma\beta\alpha}$	$-z\ (z^{\alpha\beta}\Omega_{\beta\alpha} = 0)$	$\frac{1}{2}\epsilon z$	$\frac{1}{2}\epsilon(\Upsilon z\Upsilon)$	z	
	6	$z^{\alpha\beta}z^{\gamma\delta}\epsilon_{\delta\gamma\beta\alpha}$	$-z$	$\frac{1}{2}\epsilon z$	$-\Omega z\Omega$	$\frac{1}{2}\epsilon(\Upsilon z\Upsilon)$	z

Note that in all but the 2D cases the norms are associated with determinants: For D=3 and 4 the norm is given by the determinant, while for D=5 and 6 we use the fact that the determinant of an antisymmetric matrix is the square of the Pfaffian.

Exercise IC5.1

Show that for D=5 $zz\epsilon$ and $zz\Omega\Omega$ give the same norm. (Hint: Consider $\Omega_{[\alpha\beta}\Omega_{\gamma\delta]}$.)

Unfortunately, for SO(n) for larger n, the spinor is as least as large as, and usually larger than, the vector. In general, the spinor is like the "square root" of the vector, in that the vector can be found by taking the direct product of two spinors.

It is impossible to find the spinor representation by taking direct products of vectors. This situation occurs only for orthogonal groups: In all other classical groups, all (finite-dimensional) representations are among those obtained from multiple direct products of vectors. Furthermore, in those cases the "irreducible" representations (those that can't be divided into smaller representations) can be picked out by (anti)symmetrization, and by separating trace and traceless pieces (where traces are taken with the group metrics). Fortunately, for the above cases of orthogonal groups, we can perform the same construction starting with the spinor representations, since those are the "vectors" of non-orthogonal groups.

REFERENCES

1 J. Schwinger, On angular momentum, *Quantum theory of angular momentum: a collection of reprints and original papers*, eds. L.C. Biedenharn and H. Van Dam (Academic, 1965) p. 229:
spin using spinor oscillators.

2 Georgi, *loc. cit.* (IB).

3 P.A.M. Dirac, *Proc. Roy. Soc.* **A117** (1928) 610.

4 H. Weyl, *Z. Phys.* **56** (1929) 330.

5 M. Hamermesh, *Group theory and its application to physical problems* (Dover, 1962): detailed discussion of Young tableaux.

6 M. Gell-Mann, *Phys. Lett.* **8** (1964) 214;
G. Zweig, preprints CERN-TH-401 and 412 (1964); Fractionally charged particles and SU_6, in *Symmetries and elementary particle physics*, proc. Int. School of Physics "Ettore Majorana", Erice, Italy, Aug.-Sept., 1964, ed. A. Zichichi (Academic, 1965) p. 192:
quarks.

7 Particle Data Group (in 2002, K. Hagiwara, K. Hikasa, K. Nakamura, M. Tanabashi, M. Aguilar-Benitez, C. Amsler, R.M. Barnett, P.R. Burchat, C.D. Carone, C. Caso, G. Conforto, O. Dahl, M. Doser, S. Eidelman, J.L. Feng, L. Gibbons, M. Goodman, C. Grab, D.E. Groom, A. Gurtu, K.G. Hayes, J.J. Hernandez-Rey, K. Honscheid, C. Kolda, M.L. Mangano, D.M. Manley, A.V. Manohar, J. March-Russell, A. Masoni, R. Miquel, K. Monig, H. Murayama, S. Navas, K.A. Olive, L. Pape, C. Patrignani, A. Piepke, M. Roos, J. Terning, N.A. Tornqvist, T.G. Trippe, P. Vogel, C.G. Wohl, R.L. Workman, W.-M. Yao, B. Armstrong, P.S. Gee, K.S. Lugovsky, S.B. Lugovsky, V.S. Lugovsky, M. Artuso, D. Asner, K.S. Babu, E. Barberio, M. Battaglia, H. Bichsel, O. Biebel, P. Bloch, R.N. Cahn, A. Cattai, R.S. Chivukula, R.D. Cousins, G. Cowan, T. Damour, K. Desler, R.J. Donahue, D.A. Edwards, V.D. Elvira, J. Erler, V.V. Ezhela, A. Fasso, W. Fetscher, B.D. Fields, B. Foster, D. Froidevaux, M. Fukugita, T.K. Gaisser, L. Garren, H.-J. Gerber, F.J. Gilman, H.E. Haber, C. Hagmann, J. Hewett, I. Hinchliffe, C.J. Hogan, G. Hohler, P. Igo-Kemenes, J.D. Jackson, K.F. Johnson, D. Karlen, B. Kayser, S.R. Klein, K. Kleinknecht, I.G. Knowles, P. Kreitz, Yu.V. Kuyanov, R. Landua, P. Langacker, L. Littenberg, A.D. Martin, T. Nakada, M. Narain, P. Nason, J.A. Peacock, H.R. Quinn, S. Raby, G. Raffelt, E.A. Razuvaev, B. Renk, L. Rolandi, M.T. Ronan, L.J. Rosenberg, C.T. Sachrajda, A.I. Sanda, S. Sarkar, M. Schmitt, O. Schneider, D. Scott, W.G. Seligman, M.H. Shaevitz, T. Sjostrand, G.F. Smoot, S. Spanier, H. Spieler, N.J.C. Spooner, M. Srednicki, A. Stahl, T. Stanev, M. Suzuki, N.P. Tkachenko, G. Valencia,

K. van Bibber, M.G. Vincter, D.R. Ward, B.R. Webber, M. Whalley, L. Wolfenstein, J. Womersley, C.L. Woody, and O.V. Zenin), http://pdg.lbl.gov/pdg.html (published biannually, including *Phys. Rev.* **D66** (2002) 010001):
Review of Particle Physics, a.k.a. Review of Particle Properties, a.k.a. Rosenfeld tables; tables of masses, decay rates, etc., of all known particles, plus useful brief reviews.

II. SPIN

Special relativity is simply the statement that the laws of nature are symmetric under the Poincaré group. Free relativistic quantum mechanics or field theory is then equivalent to a study of the representations of the Poincaré group. Since the conformal group is a classical group, while its subgroup the Poincaré group is not, it is easier to first study the conformal group, which is sufficient for finding the massless representations of the Poincaré group. The massive ones then can be found by dimensional reduction, which gives them in the same form as occurs in interacting field theories. In four spacetime dimensions we use the covering group of the conformal group, which is the easiest way to include spinors. These methods extend straightforwardly to supersymmetry, a symmetry between fermions and bosons that includes the Poincaré group.

A. TWO COMPONENTS

Although we have already specialized to spacetime symmetries, we have considered arbitrary spacetime dimensions. We have also noted that many of the lower-dimensional Lie groups have special properties, especially with regard to covering groups. In this section we will take advantage of those features; specifically, we examine the physical case D=4, where the rotation group is SO(3)=SU(2), the Lorentz group is SO(3,1)=SL(2,C), and the conformal group is SO(4,2)=SU(2,2).

1. 3-vectors

The most important nontrivial Lie group in physics is the rotations in three dimensions. It is also the simplest nontrivial example of a Lie group. This makes it the ideal example to illustrate the properties discussed in the previous chapter, as well as lay the groundwork for later discussions. We have already mentioned the orbital part of rotations, i.e., the representation of rotations on spatial coordinates. In this chapter we discuss the spin part; this is really the same as finding all (finite dimensional, unitary) representations.

A useful way to understand spin, or general representations of the rotation group in three dimensions of space, is to consider properties of 2×2 matrices. (This way also generalizes in a very simple way to relativity, in three space and one time dimensions.) Consider such matrices to be hermitian, which is natural from the quantum mechanical point of view. Then they have four real components, one too many for a three-vector (but just right for a relativistic four-vector), so we restrict them to also be traceless:

$$V = V^\dagger, \quad tr\ V = 0$$

The simplest way to get a single number out of a matrix, besides taking the trace, is to take the determinant. By expanding a general matrix identity to quadratic order we find an identity for 2×2 matrices

$$det(I+M) = e^{tr\, ln(I+M)} \quad \Rightarrow \quad -2\ det\ M = tr(M^2) - (tr\ M)^2$$

It is then clear that in our case $-det\ V$ is positive definite, as well as quadratic, so we can define the norm of this 3-vector as

$$|V|^2 = -2\ det\ V = tr(V^2)$$

This can be compared easily with conventional notation by picking a basis:

$$V = \frac{1}{\sqrt{2}}\begin{pmatrix} V^1 & V^2 - iV^3 \\ V^2 + iV^3 & -V^1 \end{pmatrix} = \vec{V}\cdot\vec{\sigma} \quad \Rightarrow \quad det\ V = -\tfrac{1}{2}(V^i)^2$$

where $\vec{\sigma}$ are the Pauli σ matrices, up to normalization. As usual, the inner product follows from the norm:

$$|V+W|^2 = |V|^2 + |W|^2 + 2V\cdot W$$

$$\Rightarrow \quad V\cdot W = det\ V + det\ W - det(V+W) = tr(VW)$$

Applying our previous identities for determinants to 2×2 matrices, we have

$$MCM^TC = I\ det\ M, \qquad M^{-1} = CM^TC(det\ M)^{-1}$$

where we now use the imaginary, hermitian matrix

$$C = \left(\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right)$$

If we make the replacement $M \to e^M$ and expand to linear order in M, we find

$$M + CM^TC = I\ tr\ M$$

This implies

$$tr\ V = 0 \quad \Leftrightarrow \quad (VC)^T = VC$$

i.e., the tracelessness of V is equivalent to the symmetry of VC. Furthermore, the combination of the trace and determinant identities tell us

$$M^2 = M\ tr\ M - I\ det\ M \quad \Rightarrow \quad V^2 = -I\ det\ V = I\tfrac{1}{2}|V|^2$$

Here by "V^2" we mean the square of the matrix, while "$|V|^2$"$= (V^i)^2$ is the square of the norm (neither of which should be confused with the component $V^2 = V^i\delta_i^2$.)

Again expressing the inner product in terms of the norm, we then find

$$\{V, W\} = (V \cdot W)I$$

Also, since the commutator of two finite matrices is traceless, and picks up a minus sign under hermitian conjugation, we can define an outer product (vector×vector = vector) by

$$[V, W] = \sqrt{2} iV \times W$$

Combining these two results,

$$VW = \tfrac{1}{2}(V \cdot W)I + \tfrac{1}{\sqrt{2}} iV \times W$$

In other words, the product of two traceless hermitian 2×2 matrices gives a real trace piece, symmetric in the two matrices, plus an antihermitian traceless piece, antisymmetric in the two. Thus, we have a simple relation between the matrix product, the inner ("dot") product and the outer ("cross") product. Therefore, the cross product is a special case of the Lie bracket, or commutator. This way of treating vectors (except for factors of "i") is basically Hamilton's "quaternions".

Exercise IIA1.1

Check this result in two ways:

a Show the normalization agrees with the usual outer product. Using only the above definition of $V \times W$, along with $\{V, W\} = (V \cdot W)I$, show

$$-I|V \times W|^2 = ([V, W])^2 = -I[|V|^2|W|^2 - (V \cdot W)^2]$$

b Use components, with the above basis.

Exercise IIA1.2

Write an arbitrary two-dimensional vector in terms of a complex number as $V = \frac{1}{\sqrt{2}}(v_x - iv_y)$.

a Show that the phase (U(1)) transformation $V' = Ve^{i\theta}$ generates the usual rotation. Show that for any two vectors V_1 and V_2, $V_1{}^*V_2$ is invariant, and identify its real and imaginary parts in terms of well known vector products. What kind of transformation is $V \to V^*$, and how does it affect these products?

b Consider two-dimensional functions in terms of $z = \frac{1}{\sqrt{2}}(x + iy)$ and $z^* = \frac{1}{\sqrt{2}}(x - iy)$. Show by the chain rule that $\partial_z = \frac{1}{\sqrt{2}}(\partial_x - i\partial_y)$. Write the real and imaginary parts of the equation $\partial_{z^*}V = 0$ in terms of the divergence and curl. (Then V is a function of just z.)

c Consider the complex integral

$$\oint \frac{dz}{2\pi i}\, V$$

where "$\oint$" is a "contour integral": an integral over a closed path in the complex plane defined by parametrizing $dz = du(dz/du)$ in terms of some real parameter u. This is useful if V can be Laurent expanded as $V(z) = \sum_{n=-\infty}^{\infty} c_n(z - z_0)^n$ inside the contour about a point z_0 there, since by considering circles $z = z_0 + re^{i\theta}$ we find only the $1/(z - z_0)$ term contributes. Show that this integral contains as its real and imaginary parts the usual line integral and "surface" integral. (In two dimensions a surface element differs from a line element only by its direction.) Use this fact to solve Gauss' law in two dimensions for a unit point charge as $E = 1/4\pi z$.

Exercise IIA1.3

Consider electromagnetism in 2×2 matrix notation: Define the field strength as a *complex* vector $F = \sqrt{2}(E + iB)$. Write partial derivatives as the sum of a (rotational) scalar plus a (3-)vector as $\partial = \frac{1}{\sqrt{2}}I\partial_t + \nabla$, where $\partial_t = \partial/\partial t$ is the time derivative and ∇ is the partial space derivatives written as a traceless matrix. Do the same for the charge density ρ and (3-)current j as $J = -\frac{1}{\sqrt{2}}I\rho + j$. Using the definition of dot and cross products in terms of matrix multiplication as discussed in this section, show that the simple matrix equation $\partial F = -J$, when separated into its trace and traceless pieces, and its hermitian and antihermitian pieces, gives the usual Maxwell equations

$$\nabla \cdot B = 0, \quad \nabla \cdot E = \rho, \quad \nabla \times E + \partial_t B = 0, \quad \nabla \times B - \partial_t E = j$$

(Note: Avoid the Pauli σ-matrices and explicit components.)

2. Rotations

One convenience of representing three-vectors as 2×2 instead of 3×1 is that rotations are easier to write. Since vectors are hermitian, we expect their transformations to be unitary:

$$V' = UVU^\dagger, \qquad U^\dagger = U^{-1}$$

It is easily checked that this preserves the properties of these matrices:

$$(V')^\dagger = (UVU^\dagger)^\dagger = V', \qquad tr(V') = tr(UVU^{-1}) = tr(U^{-1}UV) = tr(V) = 0$$

Furthermore, it also preserves the norm (and thus the inner product):

$$det(V') = det(UVU^{-1}) = det(U)det(V)(det\ U)^{-1} = det\ V$$

Unitary 2×2 matrices have 4 parameters; however, we can elimimate one by the condition

$$det\ U = 1$$

This eliminates only the phase factor in U, which cancels out in the transformation law anyway. Taking the product of two rotations now involves multiplying only 2×2 matrices, and not 3×3 matrices.

We can also write U in exponential notation, which is useful for going to the infinitesimal limit:

$$U = e^{iG} \quad \Rightarrow \quad G^\dagger = G, \quad tr\ G = 0$$

This means that G itself can be considered a vector. Rotations can be parametrized by a vector whose direction is the axis of rotation, and whose magnitude is $(1/\sqrt{2}\times)$ the angle of rotation:

$$V' = e^{iG} V e^{-iG} \quad \Rightarrow \quad \delta V = i[G, V] = -\sqrt{2} G \times V$$

We also now see that the Lie bracket we previously identified as the cross product is the bracket for the rotation group.

Exercise IIA2.1

Evaluate the elements of the matrix e^{iG} in closed form for a diagonal generator G. Generalize this result to arbitrary G. (Hint: Use rotational invariance.)

The hermiticity condition on V can also be expressed as a reality condition:

$$V = V^\dagger \quad and \quad tr\ V = 0 \quad \Rightarrow \quad V* = -CVC, \quad (VC)* = C(VC)C$$

where " * " is the usual complex conjugate. A similar condition for U is

$$U^\dagger = U^{-1} \quad and \quad det\ U = 1 \quad \Rightarrow \quad U* = CUC$$

which is also a consequence of the fact that we can write U in terms of a vector as $U = e^{iV}$. As a result, the transformation law for the vector can be written in terms of VC in a simple way, which manifestly preserves its symmetry:

$$(VC)' = UVU^{-1}C = U(VC)U^T$$

Exercise IIA2.2

We return to our example of D=2:

a Write an arbitrary rotation in two dimensions in terms of the *slope* (dy/dx) of the rotation (the slope to which the x-axis is rotated) rather than the angle. (This is actually more convenient to measure if you happen to have a ruler, which you need to measure lengths anyway, but not a protractor.) This avoids trigonometry, but introduces ugly square roots: Compare Lorentz transformations. Also note that this square root form covers only half of the available angles.

b Show that the square roots can be eliminated by using the slope of *half* the angle of transformation as the variable. Show the relation to the variables used in writing 3D rotations in terms of 2×2 matrices, i.e., the use of complex variables, as in exercise IIA1.2a. (Hint: Consider U and VC diagonal.)

3. Spinors

Note that the mapping of SU(2) to SO(3) is two-to-one: This follows from the fact $V' = V$ when U is a phase factor. We eliminated continuous phase factors from U by the condition $det\ U = 1$, which restricts U(2) to SU(2). However,

$$det(Ie^{i\theta}) = e^{2i\theta} = 1 \quad \Rightarrow \quad e^{i\theta} = \pm 1$$

for 2×2 matrices. More generally, for any SU(2) element U, $-U$ is also an element of SU(2), but acts the same way on a vector; i.e., these two SU(2) transformations give the same SO(3) transformation. Thus SU(2) is called a "double covering" of SO(3). However, this second transformation is not redundant, because it acts differently on half-integral spins, which we discuss in the following subsections.

The other convenience of using 2×2 matrices is that it makes obvious how to introduce spinors — Since a vector already transforms with two factors of U, we define a "square root" of a vector that transforms with just one U:

$$\psi' = U\psi \quad \Rightarrow \quad \psi^{\dagger\prime} = \psi^\dagger U^{-1}$$

where ψ is a two-component "vector", i.e., a 2×1 matrix. The complex conjugate of a spinor then transforms in essentially the same way:

$$(C\psi^*)' = CU^*\psi^* = U(C\psi^*)$$

Note that the antisymmetry of C implies that ψ must be complex: We might think that, since $C\psi^*$ transforms in the same way as ψ, we can identify the two consistently with the transformation law. But then we would have

$$\psi = C\psi^* = C(C\psi^*)^* = CC^*\psi = -\psi$$

Thus the representation is pseudoreal. The fact that $C\psi^*$ transforms the same way under rotations as ψ leads us to consider the transformation

$$\psi' = C\psi^*$$

Since a vector transforms the same way under rotations as $\psi\psi^\dagger$, under this transformation we have

$$V' = CV^*C = -V$$

which identifies it as a reflection.

Another useful way to write rotations on ψ (like looking at VC instead of V) is

$$(\psi^T C)' = (\psi^T C)U^{-1}$$

This tells us how to take an invariant inner product of spinors:

$$\psi' = U\psi, \quad \chi' = U\chi \quad \Rightarrow \quad (\psi^T C\chi)' = (\psi^T C\chi)$$

In other words, C is the "metric" in the space of spinors. An important difference of this inner product from the familiar one for three-vectors is that it is antisymmetric. Thus, if ψ and χ are anticommuting spinors,

$$\psi^T C\chi = -\chi^T C^T\psi = \chi^T C\psi$$

where one minus sign comes from anticommutativity and another is from the antisymmetry of C. Thus, it makes sense to take the norm of an anticommuting spinor as $\psi^T C\psi$, which would vanish if ψ were commuting. Of course, since rotations are unitary, we also have the usual $\psi^\dagger\psi$ as an invariant, positive definite, inner product.

Exercise IIA3.1

Consider a hermitian but *not* traceless 2×2 matrix M ($M = M^\dagger$, $tr\ M \neq 0$).

a Show

$$det\ M = 0 \quad \Rightarrow \quad M = \pm\psi\otimes\psi^\dagger$$

for some *commuting* spinor (column vector) ψ (and some sign $\pm$).

b Define a vector by

$$V = \sqrt{2}(M - \tfrac{1}{2}I\ tr\ M)$$

Show $|V|$ (*not* $|V|^2$) is simply $\psi^\dagger\psi$.

4. Indices

The best way to discuss general spins is to use index notation, rather than matrix notation. Then a spinor rotates as

$$\psi'_\alpha = U_\alpha{}^\beta \psi_\beta$$

with two-valued indices

$$\boxed{\alpha = \oplus, \ominus}$$

The inner product is *defined* by

$$\boxed{\psi^\alpha \chi_\alpha = \psi^\alpha C_{\beta\alpha} \chi^\beta = -\psi_\alpha \chi^\alpha}$$

where we have *defined* raising and lowering of indices by

$$\boxed{\psi_\alpha = \psi^\beta C_{\beta\alpha}, \quad \psi^\alpha = C^{\alpha\beta}\psi_\beta}$$

$$\boxed{C_{\alpha\beta} = -C_{\beta\alpha} = -C^{\alpha\beta} = C^{\beta\alpha} = \left(\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right)}$$

paying careful attention to signs. (In general, we fix signs by using a convention of contracting indices from upper-left to lower-right.) Then objects with many indices transform as the product of spinors:

$$A'_{\alpha\beta...\gamma} = U_\alpha{}^\delta U_\beta{}^\epsilon ... U_\gamma{}^\zeta A_{\delta\epsilon...\zeta}$$

An infinitesimal transformation is then a sum:

$$-i\delta A_{\alpha\beta...\gamma} = G_\alpha{}^\delta A_{\delta\beta...\gamma} + G_\beta{}^\delta A_{\alpha\delta...\gamma} + ... + G_\gamma{}^\delta A_{\alpha\beta...\delta}$$

This is also true for $C_{\alpha\beta}$, even though it is an invariant constant:

$$C'_{\alpha\beta} = U_\alpha{}^\gamma U_\beta{}^\delta C_{\gamma\delta} = C_{\alpha\beta} \ det\ U = C_{\alpha\beta}$$

A more interesting case is the vector: The transformation law is

$$V'_{\alpha\beta} = U_\alpha{}^\gamma U_\beta{}^\delta V_{\gamma\delta}$$

where $V_{\alpha\beta}$ is the symmetric VC considered earlier (in contrast to the antisymmetric C).

There is basically only one identity in index notation, namely

$$\boxed{0 = \tfrac{1}{2} C_{[\alpha\beta} C_{\gamma]\delta} = C_{\alpha\beta} C_{\gamma\delta} + C_{\beta\gamma} C_{\alpha\delta} + C_{\gamma\alpha} C_{\beta\delta}}$$

The expression vanishes because it is antisymmetric in those indices, and thus the indices must all have different values, but there are three two-valued indices. Another way to write this identity is to use the definition of $C^{\alpha\beta}$ as the inverse of $C_{\alpha\beta}$:

$$\boxed{C_{\alpha\gamma}C^{\beta\gamma} = \delta_\alpha^\beta \quad \Rightarrow \quad C_{\alpha\beta}C^{\gamma\delta} = \delta_{[\alpha}^\gamma \delta_{\beta]}^\delta \equiv \delta_\alpha^\gamma \delta_\beta^\delta - \delta_\beta^\gamma \delta_\alpha^\delta}$$

This tells us that antisymmetrizing in any pair of indices automatically contracts (sums over) them: Contracting this identity with an arbitrary tensor $A_{\gamma\delta}$,

$$\boxed{A_{[\alpha\beta]} = C_{\alpha\beta}C^{\gamma\delta}A_{\gamma\delta} = -C_{\alpha\beta}A^\gamma{}_\gamma}$$

That means that we need to consider only objects that are totally symmetric in their free indices. This gives all spins: Such a field with 2s indices describes spin s; we have already seen spins 0, 1/2, and 1.

We have defined the transformation law of all fields with lower indices by considering the direct product of spinors. Transformations for upper indices follow from multiplication with $C^{\alpha\beta}$: They all follow from

$$\psi'^\alpha = \psi^\beta (U^{-1})_\beta{}^\alpha$$

Since the vertical position of the index indicates the form of the transformation law, we define

$$\bar{\psi}_\alpha \equiv (\psi^\alpha)^*$$

where the "$\overline{}$" indicates complex conjugation. Thus, a hermitian matrix is written as

$$M_\alpha{}^\beta = (M^\dagger)_\alpha{}^\beta \equiv (M_\beta{}^\alpha)^* \equiv \overline{M}^\beta{}_\alpha \quad \Rightarrow \quad M_{\alpha\beta} = \overline{M}_{\beta\alpha}$$

So, for a vector we have

$$V_{\alpha\beta} = \overline{V}_{\alpha\beta} = V_{\beta\alpha}$$

Spin s is usually formulated in terms of a (2s+1)-component "vector". Then one needs to calculate Clebsch-Gordan-Wigner coefficients to construct Hamiltonians relating different spins. For example, to couple two spin-1/2 objects to a spin-1 object, one might write something like $\vec{V} \cdot \psi^\dagger \vec{\sigma} \chi$. The matrix elements of the Pauli matrices $\vec{\sigma}$ are the CGW coefficients for the spin-1 piece of $\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0$. This method gets progressively messier for higher spins. On the other hand, in spinor notation such a term would be simply $V^{\alpha\beta}\bar{\psi}_\alpha \chi_\beta$; no special coefficients are necessary, only contraction of indices. Similarly the decomposition of products of spins involves only the picking out of the various symmetric and antisymmetric pieces: For example, for $\frac{1}{2} \otimes \frac{1}{2}$,

$$\psi_\alpha \chi_\beta = \tfrac{1}{2}(\psi_{(\alpha}\chi_{\beta)} + \psi_{[\alpha}\chi_{\beta]}) = \tfrac{1}{2}\psi_{(\alpha}\chi_{\beta)} - C_{\alpha\beta}\psi^\gamma \chi_\gamma = V_{\alpha\beta} + C_{\alpha\beta}S$$

where $(\alpha\beta)$ means to symmetrize in those indices, by adding all permutations with plus signs. We have thus explicitly separated out the spin-1 and spin-0 parts V and S of the product. The square roots of various integers that appear in the CGW coefficients come from permutation factors that appear in the normalizations of the various fields/wave functions that appear in the products: For example,

$$A^{\alpha\beta\gamma}\bar{A}_{\alpha\beta\gamma} = |A^{\oplus\oplus\oplus}|^2 + 3|A^{\oplus\oplus\ominus}|^2 + 3|A^{\oplus\ominus\ominus}|^2 + |A^{\ominus\ominus\ominus}|^2$$

In the spinor index method, the square roots never appear explicitly, only their squares appear in normalizations: For example, in calculating a probability for $A \otimes B \to C$, we evaluate

$$\frac{\langle A \otimes B|C\rangle\langle C|A \otimes B\rangle}{\langle A|A\rangle\langle B|B\rangle\langle C|C\rangle}$$

where A, B, and C each have 2s indices for spin s, and $\langle|\rangle$ means contracting all indices (with the usual complex conjugation). (Normalizing states to other than 1 is often convenient and sometimes necessary: For example, plane waves are normalized with δ functions.)

Exercise IIA4.1

Redo exercise IIA3.1 in index notation: For $\psi_\alpha \chi_\beta$ above (both now bosonic), show $V^{\alpha\beta}V_{\alpha\beta} = -2S^2$.

5. Lorentz

Consider now a 2×2 matrix, whose elements we label as

$$(V)^{\alpha\dot{\beta}} = \begin{pmatrix} V^{\oplus\dot{\oplus}} & V^{\oplus\dot{\ominus}} \\ V^{\ominus\dot{\oplus}} & V^{\ominus\dot{\ominus}} \end{pmatrix} = \begin{pmatrix} V^{+} & V^{t*} \\ V^{t} & V^{-} \end{pmatrix}$$

$$= \tfrac{1}{\sqrt{2}} \begin{pmatrix} V^0 + V^1 & V^2 + iV^3 \\ V^2 - iV^3 & V^0 - V^1 \end{pmatrix} = V^a(\sigma_a)^{\alpha\dot{\beta}}$$

which we choose to be hermitian,

$$V = V^\dagger \quad \Rightarrow \quad V^{\alpha\dot{\beta}} = (V^\dagger)^{\alpha\dot{\beta}} \equiv (V^{\beta\dot{\alpha}})^*$$

where we distinguish the right spinor index by a dot because it will be chosen to transform differently from the left one: According to our discussion of subsection IB5, this is the general labeling consistent with hermiticity, i.e., $V' = gVg^\dagger$ (but without the extra restriction of group unitarity of the previous subsections). For comparison, lowering both spinor indices with the matrix C as for SU(2), and the vector indices with the Minkowski metric (in either the orthonormal or null basis, as appropriate — see subsection IA4), we find another hermitian matrix

$$(V)_{\alpha\dot{\beta}} = \begin{pmatrix} V_+ & V_t{}^* \\ V_t & V_- \end{pmatrix} = \tfrac{1}{\sqrt{2}} \begin{pmatrix} V_0 + V_1 & V_2 - iV_3 \\ V_2 + iV_3 & V_0 - V_1 \end{pmatrix} = V_a (\sigma^a)_{\alpha\dot{\beta}}$$

In the orthonormal basis, σ_a are the Pauli matrices and the identity, up to normalization. They are also the Clebsch-Gordan-Wigner coefficients for spinor⊗spinor = vector. In the null basis, they are completely trivial: 1 for one element, 0 for the rest, the usual basis for matrices. In other words, they are simply an arbitrary way (according to choice of basis) to translate a 2×2 (hermitian) matrix into a 4-component vector. We will sometimes treat a vector index "a" as an abbreviation for a spinor index pair "$\alpha\dot{\alpha}$":

$$V^a = V^{\alpha\dot{\alpha}}, \quad a = \alpha\dot{\alpha} = (\oplus\dot{\oplus}, \oplus\dot{\ominus}, \ominus\dot{\oplus}, \ominus\dot{\ominus}) = (+, \bar{t}, t, -)$$

where α and $\dot{\alpha}$ are understood to be independent indices ($\oplus \neq \dot{\oplus}$, etc.).

Examining the determinant of (either version of) V, we find the correct Minkowski norms:

$$-2\ det\ V = -2V^+V^- + 2V^tV^{t*} = -(V^0)^2 + (V^1)^2 + (V^2)^2 + (V^3)^2 = V^2$$

Thus Lorentz transformations will be those that preserve the hermiticity of this matrix and leave its determinant invariant:

$$V' = gVg^\dagger, \quad det\ g = 1$$

($det\ g$ could also have a phase, but that would cancel in the transformation.) Thus g is an element of SL(2,C). In terms of the representation of the Lie algebra,

$$g = e^G, \quad tr\ G = 0$$

Thus the group space is 6-dimensional (G has three independent complex components), the same as SO(3,1) (where $g\eta g^T = \eta \Rightarrow (G\eta)^T = -G\eta$).

Exercise IIA5.1

SL(2,C) also can be seen (less conveniently) from vector notation:

a Consider the generators

$$J_{ab}^{(\pm)} = \tfrac{1}{2}(J_{ab} \pm i\tfrac{1}{2}\epsilon_{abcd}J^{cd})$$

of SO(3,1). Find their commutation relations, and in particular show $[J^{(+)}, J^{(-)}] = 0$. Express $J_{0i}^{(\pm)}$ in terms of $J_{ij}^{(\pm)}$. Show $J_{ij}^{(\pm)}$ have the same commutation relations as J_{ij}. Finally, take a general infinitesimal Lorentz transformation in terms of J_{ab} and rewrite it in terms of $J^{(\pm)}_{ij}$, paying special attention to the reality properties of the coefficients. This demonstrates that the algebra of SO(3,1) is the same as that of SU(2)⊗SU(2), but Wick rotated to SL(2,C).

b Apply the same procedure to SO(4) and SO(2,2) to derive their covering groups.

Exercise IIA5.2

Consider relativity in two dimensions (one space, one time):

a Show that SO(1,1) is represented in lightcone coordinates by

$$x'^{+} = \Lambda x^{+}, \qquad x'^{-} = \Lambda^{-1} x^{-}$$

for some (nonvanishing) real number Λ, and therefore SO(1,1) = GL(1). Write this one Lorentz transformation, in analogy to exercise IIA1.2a on rotations in two space dimensions, in terms of an analog of the angle ("rapidity") for those transformations that can be obtained continuously from the identity. Do the relativistic analog of exercise IIA2.2.

b Still using lightcone coordinates, find the parity and time reversal transformations. Note that writing Λ as an exponential, so it can be obtained continuously from the identity, restricts it to be positive, yielding a subgroup of GL(1). Explicitly, what are the transformations of O(1,1) missing from this subgroup? Which of P, (C)T, and (C)PT are missing from these transformations, and which are missing from GL(1) itself?

In index notation, we write for this vector

$$V'_{\alpha\dot\beta} = g_\alpha{}^\gamma g^*{}_{\dot\beta}{}^{\dot\delta} V_{\gamma\dot\delta}$$

while for a ("Weyl") spinor we have

$$\psi'_\alpha = g_\alpha{}^\beta \psi_\beta$$

The metric of the group SL(2,C) is the two-index antisymmetric symbol, which is also the metric for Sp(2,C): In our conventions,

$$\boxed{C_{\alpha\beta} = -C_{\beta\alpha} = -C^{\alpha\beta} = C_{\dot\alpha\dot\beta} = \left(\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right)}$$

We also have the identities

$$det\ L_\alpha{}^\beta = \tfrac{1}{2} C^{\alpha\beta} C_{\gamma\delta} L_\alpha{}^\gamma L_\beta{}^\delta = \tfrac{1}{2}(tr\ L)^2 - \tfrac{1}{2} tr(L^2), \quad (L^{-1})_\delta{}^\beta = C^{\alpha\beta} C_{\gamma\delta} L_\alpha{}^\gamma (det\ L)^{-1}$$

$$\boxed{A_{[\alpha\beta]} = C_{\alpha\beta}C^{\gamma\delta}A_{\gamma\delta}, \quad A_{[\alpha\beta\gamma]} = 0}$$

discussed earlier in this section. As there, we use the metric to raise, lower, and contract indices:

$$\boxed{\begin{array}{c} \psi_\alpha = \psi^\beta C_{\beta\alpha}, \quad \psi_{\dot\alpha} = \psi^{\dot\beta}C_{\dot\beta\dot\alpha} \\ V\cdot W = V^{\alpha\dot\beta}W_{\alpha\dot\beta} \end{array}}$$

These results for SO(3,1) = SL(2,C) generalize to SO(4) = SU(2)⊗SU(2) (relevant to the Standard Model: see subsection IVB2) and SO(2,2) = SL(2)⊗SL(2). As described earlier, the reality conditions change, so now

$$SO(4): \quad (V^{\alpha\beta'})^* = V^{\gamma\delta'}C_{\gamma\alpha}C_{\delta'\beta'}, \qquad SO(2,2): \quad (V^{\alpha\beta'})^* = V^{\gamma\delta'}$$

consistent with the (pseudo)reality properties of spinors for SU(2) and SL(2), where we now use unprimed and primed indices for the two independent group factors ($V \to gVg'$).

Exercise IIA5.3

Take the explicit 2×2 representation for a vector given above, change the factors of i to satisfy the new reality conditions for SO(4) and SO(2,2), and show the determinant gives the right signatures for the metrics.

A common example of index manipulation is to use antisymmetry whenever possible to give vector products. For example, from the fact that $V^{\alpha\dot\beta}V^{\gamma\dot\delta}C_{\dot\delta\dot\beta}$ is antisymmetric in $\alpha\gamma$ we have that

$$V^{\alpha\dot\beta}V_{\gamma\dot\beta} = \tfrac{1}{2}\delta_\gamma^\alpha V^2$$

where the normalization follows from tracing both sides. Similarly,

$$V^{\alpha\dot\beta}W_{\gamma\dot\beta} + W^{\alpha\dot\beta}V_{\gamma\dot\beta} = \delta_\gamma^\alpha V\cdot W$$

It then follows that

$$V^{\alpha\dot\beta}W_{\gamma\dot\beta}V^{\gamma\dot\delta} = (\delta_\gamma^\alpha V\cdot W - W^{\alpha\dot\beta}V_{\gamma\dot\beta})V^{\gamma\dot\delta} = V\cdot W V^{\alpha\dot\delta} - \tfrac{1}{2}V^2 W^{\alpha\dot\delta}$$

Antisymmetry in vector indices also implies some antisymmetry in spinor indices. For example, the antisymmetric Maxwell field strength $F_{ab} = -F_{ba}$, after translating vector indices into spinor, can be separated into its parts symmetric and antisymmetric in undotted indices; antisymmetry in vector indices (now spinor index pairs) then implies the opposite symmetry in dotted indices:

$$F_{\alpha\dot\gamma,\beta\dot\delta} = -F_{\beta\dot\delta,\alpha\dot\gamma} = \tfrac{1}{4}(F_{(\alpha\beta)[\dot\gamma\dot\delta]} + F_{[\alpha\beta](\dot\gamma\dot\delta)}) = \bar C_{\dot\gamma\dot\delta}f_{\alpha\beta} + C_{\alpha\beta}\bar f_{\dot\gamma\dot\delta}, \qquad f_{\alpha\beta} = \tfrac{1}{2}F_{\alpha\dot\gamma,\beta}{}^{\dot\gamma}$$

Thus, an antisymmetric tensor also can be written in terms of a (complex) 2×2 matrix. (However, our normalization of tensor vs. symmetric spinor matrix will vary according to application.)

We also need to define complex (hermitian) conjugates carefully because C is imaginary, and uses indices consistent with transformation properties:

$$\boxed{\bar{\psi}^{\dot{\alpha}} \equiv (\psi^{\alpha})^* \quad \Rightarrow \quad \bar{\psi}_{\dot{\alpha}} = -(\psi_{\alpha})^*, \quad (\psi^{\alpha}\psi_{\alpha})^{\dagger} = \bar{\psi}^{\dot{\alpha}}\bar{\psi}_{\dot{\alpha}}}$$

$$\overline{V}^{\alpha\dot{\beta}} \equiv (V^{\dagger})^{\alpha\dot{\beta}} \equiv (V^{\beta\dot{\alpha}})^* \quad \Rightarrow \quad \bar{x}^{\alpha\dot{\beta}} = x^{\alpha\dot{\beta}}$$

where we assume the spinor is fermionic (when re-ordering for hermitian conjugation), and we have used the spacetime coordinates as an example of a real vector (hermitian 2×2 matrix). (Sometimes we will drop the "$^-$" on $\bar{\psi}^{\dot{\alpha}}$, since it is redundant to the "$^{\bullet}$". Note that, unlike SU(2), $\bar{\psi}_{\alpha} \neq (\psi^{\alpha})^*$.) In general, hermitian conjugation properties for any Lorentz representation are defined by the corresponding product of spinors: For example,

$$(\psi^{(\alpha}\chi^{\beta)})^{\dagger} = \bar{\chi}^{(\dot{\alpha}}\bar{\psi}^{\dot{\beta})} = -\bar{\psi}^{(\dot{\alpha}}\bar{\chi}^{\dot{\beta})} \quad \Rightarrow \quad \bar{f}^{\dot{\alpha}\dot{\beta}} \equiv -(f^{\alpha\beta})^*$$

More generally, we find

$$(T^{(\alpha_1\ldots\alpha_j)(\dot{\beta}_1\ldots\dot{\beta}_k)})^{\dagger} \equiv (-1)^{j(j-1)/2+k(k-1)/2}\bar{T}^{(\beta_1\ldots\beta_k)(\dot{\alpha}_1\ldots\dot{\alpha}_j)}$$

As we'll see later, most spinor algebra involves, besides spinors, just vectors and antisymmetric tensors, which carry only two spinor indices, so matrix algebra is often useful. When using bra-ket notation for 2-component spinors, it is often convenient to distinguish undotted and dotted spinors. Furthermore, since spinor indices can be raised and lowered, we can always choose the bras to carry upper indices and the kets lower, consistent with our index-contraction conventions, to avoid extra signs and factors of C. We therefore define (see subsection IB1)

$$\boxed{\begin{array}{c} \langle\psi| = \psi^{\alpha}\langle_{\alpha}|, \qquad |\psi\rangle = |^{\alpha}\rangle\psi_{\alpha}; \qquad [\psi| = \psi^{\dot{\alpha}}[_{\dot{\alpha}}|, \qquad |\psi] = |^{\dot{\alpha}}]\psi_{\dot{\alpha}} \\ V = |^{\alpha}\rangle V_{\alpha}{}^{\dot{\beta}}[_{\dot{\beta}}| \quad \Rightarrow \quad V^* = -|^{\dot{\alpha}}]V^{\beta}{}_{\dot{\alpha}}\langle_{\beta}|; \quad f = |^{\alpha}\rangle f_{\alpha}{}^{\beta}\langle_{\beta}| \quad \Rightarrow \quad f^* = |^{\dot{\alpha}}]f_{\dot{\alpha}}{}^{\dot{\beta}}[_{\dot{\beta}}| \end{array}}$$

As a result, we also have

$$\boxed{\begin{array}{c} \langle\psi\chi\rangle = \langle\chi\psi\rangle = \psi^{\alpha}\chi_{\alpha}, \quad [\psi\chi] = \psi^{\dot{\alpha}}\chi_{\dot{\alpha}}; \qquad \langle\psi\chi\rangle^{\dagger} = [\psi\chi] \\ \langle\psi|V|\chi] = \psi^{\alpha}V_{\alpha}{}^{\dot{\beta}}\chi_{\dot{\beta}}, \qquad \langle\psi|f|\chi\rangle = \psi^{\alpha}f_{\alpha}{}^{\beta}\chi_{\beta} \\ VW^* + WV^* = (V\cdot W)I \end{array}}$$

where we have used the anticommutativity of the spinor fields. From now on, we use this notation for the matrix representing a vector V $(V_\alpha{}^{\dot\beta})$, rather than the one with which we started $(V^{\alpha\dot\beta})$.

Exercise IIA5.4

Consider the generators

$$G_\alpha{}^\beta = x^{\beta\dot\gamma}\partial_{\alpha\dot\gamma} + |^\beta\rangle\langle_\alpha|$$

and their Hermitian conjugates, where $\partial_{\alpha\dot\beta} = \partial/\partial x^{\alpha\dot\beta}$. Show their algebra closes. What group do they generate? Find a subset of these generators that can be identified with (a representation of) the Lorentz group.

Since we have exhausted all possible linear transformations on spinors (except for scale, which relates to conformal transformations), the only way to represent discrete Lorentz transformations is as antilinear ones:

$$\psi'_\alpha = \sqrt{2}n_\alpha{}^{\dot\beta}\bar\psi_{\dot\beta} \qquad (\psi' = -\sqrt{2}n\psi^*)$$

From its index structure we see that n is a vector, representing the direction of the reflection. The product of two identical reflections is then, in matrix notation

$$\psi'' = 2n(n\psi^*)^* = n^2\psi \quad \Rightarrow \quad n^2 = \pm 1$$

where we have required closure on an SL(2,C) transformation (± 1). Thus n is a unit vector, either spacelike or timelike. Applying the same transformation to a vector, where $V^{\alpha\dot\beta}$ transforms like $\psi^\alpha\chi^{\dot\beta}$, we write in matrix notation

$$V' = -2nV^*n = n^2V - 2(n\cdot V)n$$

(The overall sign is ambiguous, and depends on whether it is a polar or axial vector.) This transformation thus describes parity (actually CP, because of the complex conjugation). In particular, to describe purely CP without any additional rotation (i.e., exactly reflection of the 3 spatial axes), in our basis we must choose a unit vector in the time direction,

$$\sqrt{2}n^{\alpha\dot\beta} = \delta^{\alpha\dot\beta} \quad \Rightarrow \quad \psi'^\alpha = \bar\psi_{\dot\alpha} \qquad (\bar\psi'^{\dot\alpha} = -\psi_\alpha)$$

$$\Rightarrow \quad V'^{\alpha\dot\beta} = -V_{\beta\dot\alpha}$$

which corresponds to the usual in vector notation, since in our basis

$$\sigma_a^{\alpha\dot\beta} = \sigma^a_{\beta\dot\alpha}$$

To describe time reversal, we need a transformation that does not preserve the complex conjugation properties of spinors: For example, CPT is

$$\psi'^{\alpha} = \psi^{\alpha}, \quad \bar{\psi}'^{\dot{\beta}} = -\bar{\psi}^{\dot{\beta}} \quad \Rightarrow \quad V' = -V$$

(The overall sign on V is unambiguous.)

In principle, whenever we work on a problem with both spinors and vectors we could use a mixed vector-spinor notation, converting between the usual basis for vectors and the spinor-index basis with identities such as

$$\sigma^a_{\alpha\dot{\beta}}\sigma_b^{\alpha\dot{\beta}} = \delta_a^b, \qquad \sigma^a_{\alpha\dot{\beta}}\sigma_a^{\gamma\dot{\delta}} = \delta_\alpha^\gamma \delta_{\dot{\beta}}^{\dot{\delta}}$$

However, in practice it's much simpler to use spinor indices exclusively, since then one needs no σ-matrix identities at all, but only the trivial identities for the matrix C that follow from its antisymmetry. For example, converting the vector index on the σ matrices themselves into spinor indices ($a \to \alpha\dot{\beta}$), they become trivial:

$$(\sigma^{\alpha\dot{\beta}})_{\gamma\dot{\delta}} = \delta_\gamma^\alpha \delta_{\dot{\delta}}^{\dot{\beta}}$$

(This is the same as saying an orthonormal basis of vectors has the components $(V^a)_b = \delta_b^a$ when the components are defined with respect to the same basis.)

Thus, the most general irreducible (finite-dimensional) representation of SL(2,C) (and thus SO(3,1)) has an arbitrary number of dotted and undotted indices, and is totally symmetric in each: $A_{(\alpha_1 ... \alpha_m)(\dot{\beta}_1 ... \dot{\beta}_n)}$. Treating a vector index directly as a dotted-undotted pair of indices (e.g., $a = \alpha\dot{\alpha}$, which is just a funny way of labeling a 4-valued index), we can translate into spinor notation the two constant tensors of SO(3,1): Since the only constant tensor of SL(2,C) is the antisymmetric symbol, they can be expressed in terms of it:

$$\boxed{\eta_{\alpha\dot{\alpha},\beta\dot{\beta}} = C_{\alpha\beta}C_{\dot{\alpha}\dot{\beta}}, \qquad \epsilon_{\alpha\dot{\alpha},\beta\dot{\beta},\gamma\dot{\gamma},\delta\dot{\delta}} = i(C_{\alpha\beta}C_{\gamma\delta}C_{\dot{\alpha}\dot{\delta}}C_{\dot{\beta}\dot{\gamma}} - C_{\alpha\delta}C_{\beta\gamma}C_{\dot{\alpha}\dot{\beta}}C_{\dot{\gamma}\dot{\delta}})}$$

When we work with just vectors, these can be expressed in matrix language:

$$\boxed{\begin{gathered} V \cdot W = tr(VW^*) \\ \epsilon_{abcd}V^aW^bX^cY^d \equiv \epsilon(V,W,X,Y) = i\ tr(VW^*XY^* - Y^*XW^*V) \end{gathered}}$$

(We have assumed real vectors; for complex vectors we should really write $V \cdot W^* = ...$, etc.)

Exercise IIA5.5

Prove this expression for the ϵ tensor (in either index or matrix version) agrees with that defined in subsection IB3 (as modified in subsection IB5) by (1) showing total antisymmetry, (2) explicitly evaluating a nonvanishing component.

6. Dirac

The Dirac spinor we encountered in subsection IC1 is a 4-component reducible representation in D=4: in terms of two ("left" and "right") two-component spinors,

$$\Psi = \begin{pmatrix} \psi_{L\alpha} \\ \bar{\psi}_{R\dot{\alpha}} \end{pmatrix}$$

The Hermitian metric Υ that defines the (Lorentz-invariant) Dirac spinor inner product

$$\bar{\Psi}\Psi = \Psi^\dagger \Upsilon \Psi = \psi_L^\alpha \psi_{R\alpha} + h.c., \qquad \bar{\Psi} \equiv \Psi^\dagger \Upsilon = (\psi_R^\alpha \; \bar{\psi}_L^{\dot{\alpha}})$$

takes the simple form

$$\Upsilon = \begin{pmatrix} 0 & \bar{C}^{\dot{\alpha}\dot{\beta}} \\ C^{\alpha\beta} & 0 \end{pmatrix} = \sqrt{2}\gamma_0$$

The Dirac matrices are given by

$$\not{V} \equiv \gamma \cdot V = \begin{pmatrix} 0 & V_\alpha{}^{\dot{\beta}} \\ V^\beta{}_{\dot{\alpha}} & 0 \end{pmatrix} = \begin{pmatrix} 0 & V \\ -V* & 0 \end{pmatrix}$$

where the indices have been chosen to insure that the γ matrices always take a Dirac spinor to the same type of spinor. Since $\{\not{V}, \not{W}\} = -V \cdot W$, the γ matrices satisfy

$$\{\gamma^a, \gamma^b\} = -\eta^{ab}$$

The extra sign is the result of normalizing the γ's to be pseudohermitian with respect to the metric: $\Upsilon \gamma^\dagger \Upsilon^{-1} = +\gamma$. This Dirac spinor can be made irreducible by imposing a reality condition that relates ψ_L and ψ_R: The resulting "Majorana spinor" is then

$$\Psi = \tfrac{1}{\sqrt{2}} \begin{pmatrix} \psi_\alpha \\ \bar{\psi}_{\dot{\alpha}} \end{pmatrix}$$

The product of all the γ's is a pseudoscalar, and an additional γ-matrix:

$$\gamma_{-1} = \tfrac{2\sqrt{2}}{4!}\epsilon_{abcd}\gamma^a\gamma^b\gamma^c\gamma^d = \tfrac{1}{\sqrt{2}} \begin{pmatrix} -i\delta_\alpha^\beta & 0 \\ 0 & i\delta_{\dot{\alpha}}^{\dot{\beta}} \end{pmatrix}$$

$$\Rightarrow \quad \{\gamma_{-1}, \gamma_a\} = 0, \quad \{\gamma_{-1}, \gamma_{-1}\} = -1$$

(This is usually called "γ_5" in the literature for $D = 4$, or "γ_D" for $D \neq 4$. We have renamed it for consistency with dimensional reduction.) It can be used to project a Dirac or Majorana spinor onto its two two-component spinors:

$$\Pi_\pm = \tfrac{1}{2}(I \pm \sqrt{2}i\gamma_{-1}) = \left(\begin{smallmatrix} I & 0 \\ 0 & 0 \end{smallmatrix}\right), \; \left(\begin{smallmatrix} 0 & 0 \\ 0 & I \end{smallmatrix}\right)$$

Various identities for these matrices can be derived directly from the anticommutation relations: For example,

$$\gamma^a \gamma_a = -2, \qquad \gamma^a \not{a} \gamma_a = \not{a}, \qquad \gamma^a \not{a}\not{b}\gamma_a = a \cdot b, \qquad \gamma^a \not{a}\not{b}\not{c}\gamma_a = \not{c}\not{b}\not{a}$$

$$tr(I) = 4, \qquad tr(\not{a}\not{b}) = -2a \cdot b, \qquad tr(\not{a}\not{b}\not{c}\not{d}) = a \cdot b \; c \cdot d + a \cdot d \; b \cdot c - a \cdot c \; b \cdot d$$

The trace identities follow from the fact that the only way to get a nonvanishing trace out of a product of γ matrices is when there are terms proportional to the identity; since $\{\gamma^a, \gamma^b\} = -\eta^{ab}$, this only happens when the indices are pairwise identical. The above results then follow from examination of relevant special cases. (Traces of odd numbers of γ matrices vanish. An exception is γ_{-1}, until it is rewritten in terms of its definition as the product of the other γ-matrices.)

Although use of the anticommutation relations is convenient for generalization of such identities to arbitrary dimensions, 2-spinor bra-ket notation is easier for deriving 4D identities. Since a Dirac spinor is the direct sum of a Weyl spinor and its complex conjugate, we write

$$\Psi = |^{\alpha}\rangle \psi_{L\alpha} + |^{\dot{\alpha}}]\bar{\psi}_{R\dot{\alpha}}, \qquad \bar{\Psi} = \psi_R^{\alpha}\langle_{\alpha}| + \bar{\psi}_L^{\dot{\alpha}}[_{\dot{\alpha}}|$$

In this notation, there is no need to use a spinor metric Υ, just as in Minkowski 4-vector bra-ket notation there is no need for an explicit matrix to represent the Minkowski metric: It is included implicitly in the definition of the inner product for the basis elements ($\langle_a|_b\rangle = \eta_{ab}$ or $\langle_\alpha|_\beta\rangle = C_{\alpha\beta}$). Thus hermitian conjugation is automatically pseudohermitian conjugation, etc.: $\bar{\Psi}$ *is* $\Psi^\dagger$, from the effect of hermitian conjugating the basis vectors along with the components of the spinors they multiply. (See subsections IB4-5.) We then have simply

$$\gamma_{\alpha\dot{\beta}} = -|_\alpha\rangle[_{\dot{\beta}}| - |_{\dot{\beta}}]\langle_\alpha|; \qquad \Pi_+ = |^\alpha\rangle\langle_\alpha|, \qquad \Pi_- = |^{\dot{\alpha}}][_{\dot{\alpha}}|$$

where we have replaced the vector index $a \to \alpha\dot{\beta}$ on γ_a.

Exercise IIA6.1

Use this representation for the γ matrices and projection operators $\Pi_\pm$ for all of the following:

a Derive

$$\gamma^{\alpha\dot{\beta}} \not{a}_1 \cdots \not{a}_{2n+1} \gamma_{\alpha\dot{\beta}} = \not{a}_{2n+1} \cdots \not{a}_1$$

$$\gamma^{\alpha\dot{\beta}} \not{a}_1 \cdots \not{a}_{2n} \gamma_{\alpha\dot{\beta}} = -\tfrac{1}{2} I \; tr(\not{a}_1 \cdots \not{a}_{2n}) - \gamma_{-1} \; tr(\gamma_{-1}\not{a}_1 \cdots \not{a}_{2n})$$

b Rederive the trace identities above. (Hint: For the last identity, use the identity $C_{[\alpha\beta}C_{\gamma]\delta} = 0$ repeatedly.)

c Show that

$$tr[(\varPi_+ - \varPi_-)\gamma_{\alpha\dot\alpha}\gamma_{\beta\dot\beta}\gamma_{\gamma\dot\gamma}\gamma_{\delta\dot\delta}] = -i\epsilon_{\alpha\dot\alpha,\beta\dot\beta,\gamma\dot\gamma,\delta\dot\delta}$$

by comparison with the expression of the previous subsection for ϵ.

Exercise IIA6.2

Again using this representation:

a Show that $\gamma_{[a}\gamma_{b]}$ (up to a proportionality constant) generates the usual Lorentz transformations of SL(2,C) on the 2 2-component spinors in the Dirac spinor.

b Relate this representation of the γ matrices to the defining representation of Sp(4) as given in subsection IB5, noting that Sp(4) is the covering group of SO(3,2) (subsection IC5).

7. Chirality/duality

$\varPi_\pm$ are often called "chiral projectors"; 2-component spinors (not paired into Dirac spinors) are often called "chiral spinors", and appear in "chiral theories"; the two 2-component spinors of a Dirac spinor are often labeled as having left and right "chirality"; etc. When these two halves decouple, a theory can have a "chiral symmetry"

$$\psi'_\alpha = e^{i\theta}\psi_\alpha$$

Since chirality is closely related to parity (chiral spinors can represent CP, but need to be doubled to allow C, and thus P), Dirac spinors are often used to describe theories where parity is preserved, or softly broken, or to analyze parity violation specifically, using γ_{-1} to identify it.

A similar feature appears in electrodynamics. We first translate the theory into spinor notation: The Maxwell field strength F_{ab} is expressed in terms of the vector potential ("gauge field") A_a, with a "gauge invariance" in terms of a "gauge parameter" λ with spacetime dependence. The gauge transformation $\delta A_a = -\partial_a\lambda$ becomes

$$A'_{\alpha\dot\beta} = A_{\alpha\dot\beta} - \partial_{\alpha\dot\beta}\lambda$$

where $\partial_{\alpha\dot\beta} = \partial/\partial x^{\alpha\dot\beta}$. It leaves invariant the field strength $F_{ab} = \partial_{[a}A_{b]}$:

$$F_{\alpha\dot\gamma,\beta\dot\delta} = \partial_{\alpha\dot\gamma}A_{\beta\dot\delta} - \partial_{\beta\dot\delta}A_{\alpha\dot\gamma} = \tfrac{1}{4}(F_{(\alpha\beta)[\dot\gamma\dot\delta]} + F_{[\alpha\beta](\dot\gamma\dot\delta)})$$

$$= \bar{C}_{\dot\gamma\dot\delta}f_{\alpha\beta} + C_{\alpha\beta}\bar{f}_{\dot\gamma\dot\delta}, \quad f_{\alpha\beta} = \tfrac{1}{2}\partial_{(\alpha\dot\gamma}A_{\beta)}{}^{\dot\gamma}$$

Maxwell's equations are

$$\partial^\beta{}_{\dot\gamma}f_{\beta\alpha} \sim J_{\alpha\dot\gamma}$$

They include both the field equations (the hermitian part) and the "Bianchi identities" (the antihermitian part).

Exercise IIA7.1

We already saw $VW^* + WV^*$ gave the dot product; show how $VW^* - WV^*$ is related to the "cross product" $V_{[a}W_{b]}$.

Exercise IIA7.2

Write Maxwell's equations, and the expression for the field strength in terms of the gauge vector, in 2×2 matrix notation, without using C's. Combine them to derive the wave equation for A.

Maxwell's equations now can be easily generalized to include magnetic charge by allowing the current J to be complex. (However, the expression for F in terms of A is no longer valid.) This is because the "duality transformation" that switches electric and magnetic fields is much simpler in spinor notation: Using the expression given above for the 4D Levi-Civita tensor using spinor indices,

$$F'_{ab} = \tfrac{1}{2}\epsilon_{abcd}F^{cd} \quad \Rightarrow \quad f'_{\alpha\beta} = -if_{\alpha\beta}$$

More generally, Maxwell's equations in free space (but not the expression for F in terms of A) are invariant under the continuous duality transformation

$$f'_{\alpha\beta} = e^{i\theta}f_{\alpha\beta}$$

(and $J'_{\alpha\dot\beta} = e^{i\theta}J_{\alpha\dot\beta}$ in the presence of both electric and magnetic charges).

Exercise IIA7.3

Prove the relation between duality in vector and spinor notation. Show that $F_{ab} + i\frac{1}{2}\epsilon_{abcd}F^{cd}$ contains only $f_{\alpha\beta}$ and not $f_{\dot\alpha\dot\beta}$.

Exercise IIA7.4

How does complexifying $J_{\alpha\dot\beta}$ modify Maxwell's equations in *vector* notation?

In even time dimensions, Wick rotation kills the i (or $-i$) in the spinor-index expression for $\epsilon_{\alpha\beta',\gamma\delta',\epsilon\zeta',\eta\theta'}$. Since the (discrete and continuous) duality transformation now contains no i, we can impose self-duality or anti-self-duality; i.e., that $f_{\alpha\beta}$ or $f_{\alpha'\beta'}$ vanishes, since they are now independent and real instead of complex conjugates.

These continuous chirality and duality symmetries on the field strengths generalize to the free field equations for arbitrary massless fields in four dimensions. For reasons to be explained in the following section, they distinguish the two polarizations of the waves described by such fields. They are closely related to conformal invariance: In higher dimensions, where not all free, massless theories are conformal (even on the mass shell), these symmetries exist exactly for those that are conformal.

REFERENCES

1 L.D. Landau and E.M. Lifshitz, *Quantum mechanics, non-relativistic theory*, 2nd ed. (Pergamon, 1965) ch. VIII:
review of SU(2) spinor notation.

2 B. van der Waerden, *Göttinger Nachrichten* (1929) 100:
SL(2,C) spinor notation.

3 E. Majorana, *Nuo. Cim.* **14** (1937) 171.

4 A. Salam, *Nuo. Cim.* **5** (1957) 299;
L. Landau, *JETP* **32** (1957) 407, *Nucl. Phys.* **3** (1957) 27;
T.D. Lee and C.N. Yang, *Phys. Rev.* **105** (1957) 1671:
identification of neutrino with Weyl spinor.

B. POINCARÉ

The general procedure for finding arbitrary representations of the Poincaré group relevant to physics is to:

(1) Describe spin 0. As we have seen, this means starting with the coordinate representation, which is reducible, and apply the constraint $p^2 + m^2 = 0$ to get an irreducible one.

(2) Find arbitrary, finite-dimensional, irreducible representations of the Lorentz group. This we have done in the previous section.

(3) Take the direct product of these two representations of the Poincaré group, which give the orbital and spin parts of the generators. (The spin part of translations vanishes.) We then need a further constraint to pick out an irreducible unitary piece of this product, which is the subject of this section.

1. Field equations

We have already constrained the momentum: The equation

$$\boxed{p^2 + m^2 = 0}$$

as an operator equation acting on a field or wave function is the "Klein-Gordon (or relativistic Schrödinger) equation". States or fields that satisfy their field equations are called "on-(mass-)shell", while those that don't (or for which the equations haven't been imposed) are "off-shell".

The next step is to constrain the "spin" (actually, its Lorentz generalization). The basic idea of the extra constraint is very simple: The Lorentz group introduces states of negative probability, since the Minkowski space metric is indefinite. For example, if we write the naive Lorentz invariant Hilbert-space norm for a vector wave function, the time component will have negative probability. (Similar remarks apply to spinors, e.g., for the metric $\Upsilon \sim \gamma^0$ for the Dirac spinor.) The solution to this problem, in first-quantized operator language, is to constrain the spin to eliminate the negative-metric component, in analogy to the way we have already constrained the momentum by the Klein-Gordon equation. We thus impose

$$\boxed{S_a{}^b p_b + w p_a = 0}$$

to kill the part of the Lorentz generators in the direction of the momentum, where "w" is a constant to be determined. (Its term can be attributed to ordering ambiguities.)

This equation is the general field equation for all spins (acting on the field strength), in addition to the Klein-Gordon equation (which is redundant except for spin 0).

We will see that this constraint is appropriate for massless particles. Massive particles then follow from dimensional reduction: adding a further spatial dimension and fixing its component of momentum to a constant, the mass, so $p^2 \to p^2 + m^2$.

Before examining this constraint, we first give an alternative "derivation" based on the conformal group. The earlier derivation of the massless particle from the conformal particle for spin 0 can be generalized to all "spins", i.e., all representations of the Poincaré group in arbitrary dimensions. There is a way to do this in terms of classical mechanics for all representations of the conformal group, by generalizing the description of the classical spinning particle. However, by analyzing the conformal particle quantum mechanically instead, applying a set of constraints, it will be clear how to generalize from conformal particles to general massless particles by weakening the constraints. The general idea is that the symmetry group for massive particles is the Poincaré group, while that for massless particles includes also scale transformations, and finally conformal particles have also conformal boosts. So, starting with the conformal group and dropping anything to do with conformal boosts will give massless particles.

We begin with a general representation of the conformal group SO(D,2) in terms of generators G_{AB}, where A, B are D+2-component vector indices. We then impose constraints that are the conformally covariant form of $p^2 = 0$: Identifying

$$(G^{+a}, G^{ab}, G^{+-}, G^{-a}) = (P^a, J^{ab}, \varDelta, K^a)$$

(where $A = (\pm, a)$) as the generators for translations, Lorentz transformations, dilatations, and conformal boosts, we see that

$$\mathcal{G}^{AB} = \tfrac{1}{2} G^{C(A} G_C{}^{B)} - \tfrac{1}{D+2} \eta^{AB} G^{CD} G_{CD} = 0$$

is an irreducible piece of the product GG (symmetric and traceless) and includes:

$$(\mathcal{G}^{++}, \mathcal{G}^{+a}, \mathcal{G}^{ab}, \mathcal{G}^{+-}, \mathcal{G}^{-a}, \mathcal{G}^{--}) = (P^2, \tfrac{1}{2}\{J^{ab}, P_b\} + \tfrac{1}{2}\{\varDelta, P^a\}, ...)$$

where "..." all have terms containing K^a.

Exercise IIB1.1

Work out all the $\mathcal{G}$'s in terms of P, J, $\varDelta$, and K.

In general theories, even massless ones, it is not always possible to have invariance under conformal boosts. (We'll see examples of this insubsection IXA7.) However, all massless theories are scale invariant, at least at the free level. (In D=4, free massless theories can always be made conformal on shell. However, the fact that even these theories can have actions that are not invariant under conformal boosts proves that it is sufficient to add just dilatations to the Poincaré group. Furthermore, the fact that conformal boosts are not always an invariance in D>4 means that dropping them will give results in a dimension-independent form.) Therefore, only $\mathcal{G}^{++}$ and $\mathcal{G}^{+a}$ can be defined in general massless theories, but we'll see that these are sufficient to define the kinematics. The former is just the masslessness condition, which we used to pick the constraints in the first place.

As we saw earlier, Δ just scales x^a: We can therefore write the relevant generators as

$$P^a = \partial^a, \qquad J^{ab} = x^{[a}\partial^{b]} + S^{ab}, \qquad \Delta = \tfrac{1}{2}\{x^a, \partial_a\} + w - 1 = x^a\partial_a + w + \tfrac{D-2}{2}$$

(We have used the antihermitian form of the generators.) The "scale weight" $w + \frac{D-2}{2}$ is the real "spin" part of Δ, just as S_{ab} is the spin part of the angular momentum J_{ab}. To preserve the algebra it must commute with everything, and thus we can set it equal to a constant on an irreducible representation. We'll see shortly that its value is actually determined by the spin S_{ab}. It is the engineering dimension of the corresponding field. It has been normalized for later convenience; the value of w depends on the representation of S^{ab}, but is independent of D. The dilatation generator Δ is not exactly antihermitian because the integration measure d^Dx isn't invariant under scaling. This is another reason w is determined, by the free action. The form we have given preserves reality of fields. The commutation relations for the spin parts, and the total generators, are the same as those for the orbital parts; e.g.,

$$[S_{ab}, S^{cd}] = -\delta_{[a}^{[c} S_{b]}{}^{d]}$$

(A convenient mnemonic for evaluating this commutator in general is to use $S^{ab} \to x^{[a}\partial^{b]}$ instead.)

Exercise IIB1.2

We can also use this method to find the stronger conditions for the fully conformal case:

a Find an expression for K^a in terms of x, ∂, S, and w that preserves the commutation relations.

b Evaluate all the constraints $\mathcal{G}$, and express the independent ones in terms of just ∂, S, and w (no x).

Substituting the explicit representation of the generators into the constraint $\mathcal{G}^{+a}$, and using the former constraint $P^2 = 0$ (when acting on wave functions on the right), we find that all x dependence drops out, leaving for $\mathcal{G}^{+a}$ the condition

$$S_a{}^b \partial_b + w\partial_a = 0$$

(paying careful attention to quantum mechanical ordering).

Exercise IIB1.3

Define spin for the conformal group by starting in D+2 dimensions: In terms of the (D+2)-dimensional coordinates y^A and their derivatives ∂_A,

$$G_{AB} = y_{[A}\partial_{B]} + S_{AB}$$

Besides the previous conditions

$$y^2 = \partial^2 = \{y^A, \partial_A\} = 0$$

impose the constraints, in analogy to the D-dimensional field equations, and taking into account the symmetry between y and ∂,

$$S_A{}^B y_B + w y_A = S_A{}^B \partial_B + w\partial_A = 0$$

a Show that the algebra of constraints closes, if we include the additional constraint

$$\tfrac{1}{2} S_{(A}{}^C S_{B)C} + w(w + \tfrac{D}{2})\eta_{AB} = 0$$

b Solve all the constraints with explicit y's for everything with an upper "−" index, reducing the manifest symmetry to SO(D−1,1), in analogy to the way $y^2 = 0$ was solved to find y^-.

c Write all the conformal generators in terms of x_a, ∂_a, S_{ab}, and w.

2. Examples

We now examine the constraints $S_a{}^b\partial_b + w\partial_a = 0$ in more detail. We begin by looking at some simple (but useful) examples. The simplest case is spin 0:

$$S_{ab} = 0 \quad \Rightarrow \quad w = 0$$

The next simplest case (for arbitrary dimension) is the Dirac spinor (see subsections IC1 and IIA6):

$$S_{ab} = -\tfrac{1}{2}[\gamma_a, \gamma_b] \quad \Rightarrow \quad S_a{}^b\partial_b + w\partial_a = -\gamma_a\gamma^b\partial_b + (w - \tfrac{1}{2})\partial_a$$

$$\Rightarrow \quad \gamma^a \partial_a = 0, \quad w = \tfrac{1}{2}$$

where we have separated out the pieces of the constraint that are irreducible with respect to the Lorentz group (e.g., by multiplying on the left with γ^a). This gives the (massless) "Dirac equation" $\partial\!\!\!/ \Psi = 0$. The next case is the vector: In terms of the basis $|V\rangle = V^a|_a\rangle$, the spin is (see subsection IB5)

$$S_{ab} = |_{[a}\rangle\langle_{b]}|$$

However, the vector yields just another description of the scalar:

Exercise IIB2.1

Apply the field equations for general field strengths to the case of a vector field strength.

a Find the independent field equations (assuming the field strength is not just a constant)

$$\partial_{[a} F_{b]} = 0, \qquad \partial^a F_a = 0, \qquad w = 1$$

Note that solving the first equation determines the vector in terms of a scalar, while the second then gives the Klein-Gordon equation for that scalar, and the third fixes the weight of the scalar to be the same as that found by starting with a scalar field strength.

b Lorentz covariantly solve the second equation first to find a gauge field that is not a scalar.

All other representations can be built up from the spinor and vector. As our final example, we consider the case where the field is a 2nd-rank antisymmetric tensor: Using the direct product representation (applied as in subsection IB2 given the vector representation)

$$F = F^{ab}|_a\rangle \otimes |_b\rangle, \quad S_{ab}(|_c\rangle \otimes |_d\rangle) = (S_{ab}|_c\rangle) \otimes |_d\rangle + |_c\rangle \otimes S_{ab}|_d\rangle$$

$$\Rightarrow \quad (S_{ab}F)^{cd} = \delta_{[a}^{[c} F_{b]}{}^{d]}$$

we find the equations

$$(S_a{}^b \partial_b + w\partial_a) F_{cd} = \tfrac{1}{2}\partial_{[a} F_{cd]} - \eta_{a[c}\partial^b F_{d]b} + (w-1)\partial_a F_{cd}$$

$$\Rightarrow \quad \partial_{[a} F_{bc]} = \partial^b F_{ab} = 0, \qquad w = 1$$

which are Maxwell's equations, again separating out irreducible pieces (e.g., by tracing and antisymmetrizing).

Exercise IIB2.2

Verify the representation of Lorentz spin given above for F_{ab} by finding the commutation relations implied by this representation.

Exercise IIB2.3

Use the definition of the action of the Lorentz generators on a vector in vector and spinor notations,

$$S_{ab} = |_{[a}\rangle\langle_{b]}|, \quad |_a\rangle = |_\alpha\rangle \otimes |_{\dot\alpha}\rangle$$

$$S_{\alpha\beta} = |_{(\alpha}\rangle\langle_{\beta)}|, \quad S_{\dot\alpha\dot\beta} = |_{(\dot\alpha}\rangle\langle_{\dot\beta)}|,$$

to derive

$$S_{\alpha\dot\alpha,\beta\dot\beta} = -\tfrac{1}{2}(C_{\alpha\beta}\bar S_{\dot\alpha\dot\beta} + \bar C_{\dot\alpha\dot\beta}S_{\alpha\beta})$$

Exercise IIB2.4

Consider the field equations in 4D spinor notation for a general field strength, totally symmetric in its m undotted indices and n dotted indices,

$$S_\alpha{}^\beta \partial_{\beta\dot\gamma} - m\partial_{\alpha\dot\gamma} = \bar S_{\dot\alpha}{}^{\dot\beta}\partial_{\gamma\dot\beta} - n\partial_{\gamma\dot\alpha} = 0, \qquad w = \tfrac{1}{2}(m+n)$$

a Show this implies

$$\partial^{\alpha\dot\gamma}\psi_{\alpha\ldots\dot\beta\ldots} = \partial^{\gamma\dot\beta}\psi_{\alpha\ldots\dot\beta\ldots} = 0$$

b Translate the field equations into vector notation (in terms of S_{ab}), finding $S_a{}^b\partial_b + w\partial_a = 0$ and an axial vector equation.

c Show that the two equations are equivalent by deriving the equations of part **a** from $S_a{}^b\partial_b + w\partial_a = 0$ alone, and from the axial equation alone (except that the axial equation doesn't work for the cases $m = n$).

In each case, choosing the wrong scale weight w would imply the field was constant. Note that we chose the field *strength* F_{ab} to describe electromagnetism: The arguments we used to derive field equations were based on physical degrees of freedom, and did not take gauge invariance into account. In chapter XII we use more powerful methods to find the gauge covariant field equations for the gauge fields, and their actions.

3. Solution

Free field equations can be solved easily in momentum space. Then the simplest way to do the algebra is in the "lightcone frame". This is a reference frame, obtained by a Lorentz transformation, where a massless momentum takes the simple form

$$p^a = \delta^a_+ p^+$$

(using only rotations), or the even simpler form $p^a = \pm\delta^a_+$ (using also a Lorentz boost), where again $\pm$ is the sign of the energy. In that frame the general field equation $S_a{}^b \partial_b + w\partial_a = 0$ reduces to

$$S^{-i} = 0, \qquad w = S^{+-}$$

The constraint $S^{-i} = 0$ determines S^{+-} to take its maximum possible value within that irreducible representation, since S^{-i} is the raising operators for S^{+-}: For any eigenstate of S^{+-},

$$S^{+-}|h\rangle = h|h\rangle \quad \Rightarrow \quad S^{+-}(S^{-i}|h\rangle) = (S^{-i}S^{+-} + [S^{+-}, S^{-i}])|h\rangle = (h+1)(S^{-i}|h\rangle)$$

The remaining constraint then determines w: It is the maximum value of S^{+-} for that representation. By parity $(+ \leftrightarrow -)$, $-w$ is the minimum, so

$$w \geq 0; \qquad w = 0 \quad \Leftrightarrow \quad S^{ab} = 0$$

since if $S^{+-} = 0$ for all states then $S^{ab} = 0$ by Lorentz transformation. As we have seen by other methods (but can easily be derived by this method), $w = \frac{1}{2}$ for the Dirac spinor and $w = 1$ for the vector; since general representations can be built from reducing direct products of these, we see that w is an integer for bosons and half-integer for fermions. If we describe a general irreducible representation by a Young tableau for SO(D−1,1) (with tracelessness imposed), or a Young tableau times a spinor (with also γ-tracelessness $\gamma^a\psi_{a...b} = 0$), then it is easy to see from the results for the spinor and vector, and antisymmetry in rows, that w is simply the number of columns of the tableau (its "width"), counting a spinor index as half a column: S^{+-} just counts the maximum number of "−" indices that can be stuck in the boxes describing the basis elements. (In fact, Dirac spinor $\otimes$ Dirac spinor gives just all possible 1-column representations.)

This leaves undetermined only S^{ij} and S^{+i}. However, S^{+i} ("creation operator") is canonically conjugate to S^{-i} ("annihilation operator"), so its action has also been fixed:

$$[S^{-i}, S^{+j}] = \delta^{ij}S^{+-} + S^{ij}$$

(S^{ij} vanishes for $i=j$, so S^{+i} and S^{-i} are conjugate, though not "orthonormal". The constant S^{+-} was fixed above to be nonvanishing, except for the trivial case of spin 0.) Equivalently, S^{ij} preserves $S^{-i}=0$, while S^{+i} doesn't: S^{ij} are the only nontrivial spin operators acting within the subspace satisfying the constraint.

Thus only the "little group" SO(D−2) spin S^{ij} remains nontrivial: The original irreducible representation of SO(D−1,1) Lorentz spin S^{ab} was a reducible representation of SO(D−2) spin S^{ij}; the irreducible SO(D−2) representation with the highest value of S^{+-} is picked out of this SO(D−1,1) representation. This solution also gives the field strength in terms of the gauge field: Working with just the highest-S^{+-}-weight states is equivalent to working with the gauge field, up to factors of ∂^+.

As an explicit example, for spin 1/2 we have simply $\gamma^-\Psi=0$, which kills half the components, leaving the half given by $\gamma^+\Psi$. For spin 1, we find

$$\left.\begin{array}{rcl} p^b F_{ab}=0 & \Rightarrow & F^{-a}=0 \\ p^{[a}F^{bc]}=0 & \Rightarrow & only\ F^{+a}\neq 0 \end{array}\right\} \quad \Rightarrow \quad only\ F^{+i}\neq 0$$

In the "lightcone gauge" $A^+=0$, we have $F^{+i}=\partial^+A^i$, so the highest-weight part of F^{ab} is the transverse part of the gauge field. The general pattern, in terms of field strengths, is then to keep only pieces with as many as possible upper + indices and no upper − indices (and thus highest S^{+-} weight). In terms of the vector potential, we have

$$F^{ab}\sim p^{[a}A^{b]} \quad \Rightarrow \quad only\ A^i\neq 0$$

The general rule for the gauge field is to drop ± indices, so the field becomes an irreducible representation of SO(D−2). All + indices on the field strength are picked up by the momenta, which also account for the scale weight of the field strength: All gauge fields have $w=0$ for bosons and $w=\frac{1}{2}$ for fermions.

Exercise IIB3.1

Using only the anticommutation relations $\{\gamma^a,\gamma^b\}=-\eta^{ab}$, construct projection operators from $\gamma^\pm$: These are operators Π_I that satisfy

$$\Pi_I\Pi_J=\delta_{IJ}\Pi_I\ \left(no\ \sum\right), \quad \sum \Pi_I=1$$

Because of time reversal symmetry $\gamma^+\leftrightarrow-\gamma^-$ (or parity $\gamma^+\leftrightarrow\gamma^-$), these project onto two subspaces equal in size.

A method equivalent to using the lightcone frame is to perform a unitary transformation U on the spin that is the inverse of the transformation on the coordinates/momentum that would take us to the lightcone frame: We want a Lorentz transformation $\Lambda_a{}^b$ on the field equations, which are of the form

$$\mathcal{O}_a{}^b p_b=0, \qquad \mathcal{O}_a{}^b=S_a{}^b+w\delta_a^b$$

that has the effect

$$U\mathcal{O}_a{}^b U^{-1} = \Lambda_a{}^c \mathcal{O}_c{}^d \Lambda^b{}_d, \qquad \Lambda^b{}_a p_b = p'_a, \qquad p'^a = \delta^a_+ p^+ \quad \Rightarrow$$

$$0 = U\mathcal{O}_a{}^b p_b U^{-1} = \Lambda_a{}^c \mathcal{O}_c{}^d \Lambda^b{}_d p_b = \Lambda_a{}^c \mathcal{O}_c{}^d p'_d \quad \Rightarrow \quad \mathcal{O}_a{}^b p'_b = 0$$

If $|\psi\rangle$ satisfies the original constraint, then $U|\psi\rangle$ will satisfy the new one. If we like, we can always transform back at the end. This is equivalent to a gauge transformation in the field theory.

It is easy to check that the appropriate operator is

$$U = e^{S^{+i}p^i/p^+}$$

Any operator V^a that transforms as a vector under S^{ab},

$$[S^{ab}, V^c] = V^{[a}\eta^{b]c}$$

but commutes with p, is transformed by U into $UVU^{-1} = V'$ as

$$V'^+ = V^+, \qquad V'^i = V^i + V^+ \frac{p^i}{p^+}, \qquad V'^- = V^- + V^i \frac{p^i}{p^+} + V^+ \frac{(p^i)^2}{2(p^+)^2}$$

as follows from explicit Taylor expansion, which terminates because S^{+i} act as lowering operators (as for conformal boosts in subsection IA6). This yields the desired result

$$V'^a p_a = V^a p'_a + \frac{V^+}{2p^+} p^2$$

when we impose the field equation $p^2 = 0$.

Exercise IIB3.2

Check this result by performing the transformation explicitly on the constraint. Before the transformation, the lightcone decomposition of the constraint is

$$(-S^{+-} + w)p^+ + S^{+i}p^i = 0$$

$$-S^{i-}p^+ + S^{ij}p^j + wp^i - S^{i+}p^- = 0$$

$$S^{-i}p^i + (-S^{-+} + w)p^- = 0$$

Show that after this transformation, the constraint becomes

$$(-S^{+-} + w)p^+ = 0$$

$$-S^{i-}p^+ + (-S^{+-} + w)p^i - \tfrac{1}{2} S^{+i} \frac{p^2}{p^+} = 0$$

$$S^{-i}p^i + (-S^{+-} + w)p^- - S^{+-}\frac{p^2}{p^+} - \tfrac{1}{2}S^{+i}p^i\frac{p^2}{p^{+2}} = 0$$

Clearly these imply

$$w = S^{+-}, \qquad S^{-i} = 0$$

with $p^2 = 0$.

On the other hand, if instead of using the lightcone identification of x^+ as "time", we choose to use the usual x^0 for purposes of finding the evolution of the system, then we want to consider transformations that do not involve p^0, instead of not involving the "energy" p^-. Thus, by p^0-independent rotations alone, the best we can do is to choose

$$p^i = 0, \qquad p^1 = \omega$$

i.e., we can fix the value of the spatial momentum, but not in a way that relates to the sign of the energy. The result is then

$$p^0 > 0: \quad p^a = \delta^a_+ p^+$$
$$p^0 < 0: \quad p^a = \delta^a_- p^-$$

The result is similar to before, but now the positive and negative energy solutions are separated: In this frame the field equations reduce to

$$p^0 > 0: \quad S^{-i} = 0, \quad S^{+-} = w$$
$$p^0 < 0: \quad S^{+i} = 0, \quad S^{+-} = -w$$

Thus, while w takes the same value as before, now the positive-energy states are associated with the highest weight of S^{+-}, while the negative-energy ones go with the lowest weight (and nothing between). The unitary transformation that achieves this result is a spin rotation that rotates S^{ab} in the field equations with the same effect as an orbital transformation that would rotate $(p^1, p^i) \to (\omega, 0)$. By looking at the special case $D = 3$ (where there is only one rotation generator), we easily find the explicit transformation

$$U = exp\left[tan^{-1}\left(\frac{|p^i|}{p^1}\right) S^{1i}\frac{p^i}{|p^i|}\right]$$

Exercise IIB3.3

Perform this transformation:

a Find the action of the above transformation on an arbitrary vector V^a. (Hint: Look at $D = 3$ to get the transformation on the "longitudinal" part of the vector.) In particular, show that

$$V'^a p_a = V^a p'_a, \qquad p'^a = \delta^a_0 p^0 + \delta^a_1 \omega$$

b Show the field equations are transformed as

$$S^{0a}p_a + wp^0 \quad \Rightarrow \quad -\omega S^{10} + wp^0 = p^0(w - \tfrac{p^0}{\omega}S^{10}) - \tfrac{1}{\omega}S^{10}p^2$$

$$S^{1a}p_a + wp^1 \quad \Rightarrow \quad \tfrac{1}{\omega}p^i(\omega S^{1i} - p^0 S^{0i}) + p^1(w - \tfrac{p^0}{\omega}S^{10})$$

$$S^{ia}p_a + wp^i \quad \Rightarrow \quad -[\delta^{ij} - \tfrac{1}{\omega(\omega+p^1)}p^i p^j](\omega S^{1j} - p^0 S^{0j}) + p^i(w - \tfrac{p^0}{\omega}S^{10})$$

Note that the first equation gives the time-dependent Schrödinger equation, with Hamiltonian

$$H = \tfrac{1}{w}(S^{10}p^1 - S^{0i}p^i) \;\to\; \tfrac{1}{w}S^{10}\omega$$

This diagonalizes the Hamiltonian H (in a representation where S^{10} is diagonal). Thus the only independent equations are

$$p^2 = 0, \qquad S^{10} = \epsilon(p^0)w, \qquad S^{1i} - \epsilon(p^0)S^{0i} = 0$$

leading to the advertised result.

c Find the transformation that rotates to the p^i direction instead of the 1 direction, so

$$H \to -\tfrac{1}{w}S^{0i}\frac{p_i}{|p^j|}\omega$$

4. Mass

So far we have considered only massless theories. We now introduce masses by "dimensional reduction", identifying mass with the component of momentum in an extra dimension. As with the extra dimensions used for describing conformal symmetry, this extra dimension is just a mathematical construct used to give a simple derivation. (Theories have been postulated with extra, unseen dimensions that are hidden by "compactification": Space curls up in those directions to a size too small to detect with present experiments. However, no compelling reason has been given for why the extra dimensions should want to compactify.)

The method is to:

(1) extend the range of vector indices by one additional spatial direction, which we call "-1";

(2) set the corresponding component of momentum to equal the mass,

$$p_{-1} = m$$

and

(3) introduce extra factors of i to restore reality, since $\partial_{-1} = ip_{-1} = im$, by a unitary transformation.

Since all representations can be constructed by direct products of the vector and spinor, it's sufficient to define this last step on them. For the scalar this method is trivial, since then simply $p^2 \to p^2 + m^2$. Except for the last step, the other constraint becomes

$$\boxed{S_a{}^b\partial_b + S_{a,-1}im + w\partial_a = 0, \quad S_{-1}{}^a\partial_a + wim = 0}$$

For the spinor, since any transformation on the spinor index can be written in terms of the gamma matrices, and the transformation must affect only the -1 direction, we can use only γ_{-1}. (For even dimensions, we can identify the γ_{-1} of dimensional reduction with the one coming from the product of all the other γ's, since in odd dimensions the product of all the γ's is proportional to the identity.) We find

$$U = exp(-\pi\gamma_{-1}/2\sqrt{2}) : \qquad \gamma_{-1} \to \gamma_{-1}, \quad \gamma_a \to -\sqrt{2}\gamma_{-1}\gamma_a$$

We perform this transformation directly on the spin operators appearing in the constraints, or the inverse transformation on the states. Dimensional reduction, followed by this transformation, then modifies the massless equation of motion as

$$i\partial\!\!\!/ \to i\partial\!\!\!/ - m\gamma_{-1} \to -\sqrt{2}\gamma_{-1}(i\partial\!\!\!/ + \tfrac{m}{\sqrt{2}})$$

so $i\partial\!\!\!/\Psi = 0 \to (i\partial\!\!\!/ + \frac{m}{\sqrt{2}})\Psi = 0$.

The prescription for the vector is

$$U = exp(\tfrac{1}{2}i\pi|^{-1}\rangle\langle^{-1}|) : \qquad |^{-1}\rangle \to i|^{-1}\rangle, \quad \langle^{-1}| \to -i\langle^{-1}| \qquad (\langle^{-1}|^{-1}\rangle = 1)$$

with the other basis states unchanged. This has the effect of giving each field a $-i$ for each (-1)-index. For example, for Maxwell's equations

$$\partial_{[a}F_{bc]} \to \begin{cases} \partial_{[a}F_{bc]} \\ \partial_{[a}F_{b]-1} + imF_{ab} \end{cases} \to \begin{cases} \partial_{[a}F_{bc]} \qquad \text{(redundant)} \\ -i(\partial_{[a}F_{b]-1} - mF_{ab}) \end{cases}$$

$$\partial^b F_{ab} \to \begin{cases} \partial^b F_{ab} + imF_{a-1} \\ \partial^a F_{-1a} \end{cases} \to \begin{cases} \partial^b F_{ab} + mF_{a-1} \\ -i\partial^a F_{-1a} \qquad \text{(redundant)} \end{cases}$$

Note that only the mass-independent equations are redundant. Also, F_{a-1} appears explicitly as the potential for F_{ab}, but without gauge invariance. Alternatively, we can keep the gauge potential:

$$F_{ab} = \partial_{[a}A_{b]} \to \begin{cases} F_{ab} = \partial_{[a}A_{b]} \\ F_{a-1} = \partial_a A_{-1} - imA_a \end{cases} \to \begin{cases} F_{ab} = \partial_{[a}A_{b]} \\ -iF_{a-1} = -i(\partial_a A_{-1} + mA_a) \end{cases}$$

This is known as the "Stückelberg formalism" for a massive vector, which maintains gauge invariance by having a scalar A_{-1} in addition to the vector: The gauge transformations are now

$$\delta A_a = -\partial_a \lambda \to \begin{cases} \delta A_a = -\partial_a \lambda \\ \delta A_{-1} = -im\lambda \end{cases} \to \begin{cases} \delta A_a = -\partial_a \lambda \\ -i\delta A_{-1} = -im\lambda \end{cases}$$

Exercise IIB4.1

Consider the general massive field equations that follow from the general massless ones by dimensional reduction. One of these is

$$S_{-1}{}^a \partial_a + wim = 0$$

(before restoring reality). This scalar equation alone gives the complete field equations for w=1/2 and 1 (antisymmetric tensors), 0 being trivial.

a Show that for w=1/2 it gives the (massive) Dirac equation.

b Expanding the state over explicit fields, find the covariant field equations it implies for w=1. Show these are sufficient to describe spins 0 (vector field strength: see exercise IIB2.1) and 1 (F_{ab} and F_{a-1}). Note that S^{-1a} act as generalized γ matrices (the Dirac matrices for spin 1/2, the "Duffin-Kemmer matrices" for w=1), where

$$S^{ab} = -[S^{-1a}, S^{-1b}]$$

c Show that these covariant field equations imply the Klein-Gordon equation for arbitrary antisymmetric tensors. Show that in D=4 all antisymmetric tensors (coming from 0-5 indices in D=5) are equivalent to either spin 0 or spin 1, or trivial. (Hint: Use ϵ_{abcd}.)

d Consider the *reducible* representation coming from the direct product of two Dirac spinors, and represent the wave function itself as a matrix:

$$S^{ij}\Psi = \tilde{S}^{ij}\Psi + \Psi\tilde{S}^{ij}$$

where $i = (-1, a)$ and $\tilde{S}^{ij}$ is the usual Dirac-spinor representation. Using the fact that any 4×4 (in D=4) matrix can be written as a linear combination of products of γ-matrices (antisymmetric products, since symmetrization yields anticommutators), find the irreducible representations of SO(4,1) in Ψ, and relate to part **c**.

Exercise IIB4.2

Solve the field equations for massive spins 1/2 and 1 in momentum space by going to the rest frame.

The solution to the general massive field equations can also be found by going to the rest frame ($p^0 = m$): The combination of that and dimensional reduction is, in terms of the massive analog of lightcone components,

$$p^+ = \tfrac{1}{\sqrt{2}}(p^0 + p^{-1}) = \sqrt{2}m, \qquad p^- = \tfrac{1}{\sqrt{2}}(p^0 - p^{-1}) = 0, \qquad p^i = 0$$

where p^i are now the other D$-$1 (spatial) components. This fixing of the momentum is the same as the lightcone frame except that p^1 has been replaced by p^{-1}, and thus p^i now has D$-$1 components instead of D$-$2. The solution to the constraints is thus also the same, except that we are left with an irreducible representation of the "little group" SO(D$-$1) as found in the rest frame for the massive particle, vs. one of SO(D$-$2) found in the lightcone frame for the massless case.

5. Foldy-Wouthuysen

The other frame we used for the massless analysis, which involved only energy-independent rotations, can also be applied to the massive case by dimensional reduction. The result is known as the "Foldy-Wouthuysen transformation", and is useful for analyzing *interacting* massive field equations in the nonrelativistic limit. Replacing $p^1 \to p^{-1} = m$ in our previous result, we have for the free case

$$U = exp\left[tan^{-1}\left(\frac{|\vec{p}\,|}{m}\right) S^{-1i}\frac{p^i}{|\vec{p}\,|}\right], \qquad UHU^{-1} = \tfrac{1}{w}S^{-10}\omega$$

For purposes of generalization to interactions, it was important that in the free transformation (1) we used only the spin part of a rotation, since the orbital part could introduce explicit x dependence, and (2) we used only rotations, since a Lorentz boost would introduce p^0 dependence in the "parameters" of the transformation, which could generate additional p^0 (time derivative) terms in the field equation.

Exercise IIB5.1

Perform this transformation for the Dirac spinor, and then apply the reality-restoring transformation to obtain

$$H \;\to\; \sqrt{2}\gamma_0\omega$$

We then can use the diagonal representation $\gamma_0 = \left(\begin{smallmatrix} I & 0 \\ 0 & -I \end{smallmatrix}\right)/\sqrt{2}$. (We can define this representation, up to phases, by switching γ_0 and γ_{-1} of the usual representation.) In general the reality-restoring transformation will be unnecessary for any spin, since applying the field equation $S^{-10} = \pm w$ picks out a representation of the "little group" SO(D$-$1).

In the interacting case the result generally can't be obtained in closed form, so it is derived perturbatively in $1/m$. The goal is again a Hamiltonian diagonal with respect to S^{-10}, to preserve the separation of positive and negative energies; we then can set $S^{-10} = w$ to describe just positive energies. We thus choose the transformation to cancel any terms in H that are off-diagonal, which come from odd total numbers of "-1" and "0" indices from the spin factors in any term: i.e., odd numbers of S^{0i} and S^{-1i} (e.g., the $S^{0i}p_i$ term in the original H). For example, for coupling to an electromagnetic field, the exponent of U is generalized by covariantizing derivatives (minimal coupling $\partial \to \nabla = \partial + iA$), but also requires field-strength (E and B) terms to cancel certain ones of those generated from commutators of these derivatives in the transformation:

$$\nabla^a = \partial^a + iA^a \quad \Rightarrow \quad [\nabla^a, \nabla^b] = iF^{ab}$$

Before performing this transformation explicitly for the first few orders, we consider some general properties that will allow us to collect similar terms in advance. (Few duplicate terms would appear to the order we consider, but they breed like rabbits at higher orders.) We start with a field equation $\mathcal{F}$ that can be separated into "even" terms $\mathcal{E}$ and "odd" ones $\mathcal{O}$, each of which can be expanded in powers of $1/m$:

$$\mathcal{F} = \mathcal{E} + \mathcal{O}: \qquad \mathcal{E} = \sum_{n=-1}^{\infty} m^{-n}\mathcal{E}_n, \qquad \mathcal{O} = \sum_{n=0}^{\infty} m^{-n}\mathcal{O}_n$$

Note that the leading (m^{+1}) term is even; thus we choose only odd generators to transform away the odd terms in $\mathcal{F}$, perturbatively from this leading term:

$$\mathcal{F}' = e^G \mathcal{F} e^{-G}, \qquad G = \sum_{n=1}^{\infty} m^{-n} G_n$$

Since $\mathcal{F}'$ is even while G is odd, we can separate this equation into its even and odd parts as

$$\begin{aligned} \mathcal{F}' &= cosh(\mathcal{L}_G)\mathcal{E} + sinh(\mathcal{L}_G)\mathcal{O} \\ 0 &= sinh(\mathcal{L}_G)\mathcal{E} + cosh(\mathcal{L}_G)\mathcal{O} \end{aligned}$$

(with $\mathcal{L}_G = [G, \ ]$ as in subsection IA3). Since we can perturbatively invert any Taylor-expandable function of $\mathcal{L}_G$ that begins with 1, we can use the second equation to give a recursion relation for G_n: Separating the leading term of $\mathcal{F}$,

$$\mathcal{E} = m\mathcal{E}_{-1} + \Delta\mathcal{E}, \qquad -m[G, \mathcal{E}_{-1}] = [G, \Delta\mathcal{E}] + \mathcal{L}_G\, coth(\mathcal{L}_G)\mathcal{O}$$

which we can expand in $1/m$ [after Taylor expanding $\mathcal{L}_G\, coth(\mathcal{L}_G)$] to give an expression for $[G_n, \mathcal{E}_{-1}]$ to solve for G_n. We can also use the implicit solution for $[G, \mathcal{E}]$ directly to simplify the expression for $\mathcal{F}'$:

$$\mathcal{F}' = \mathcal{E} + tanh(\tfrac{1}{2}\mathcal{L}_G)\mathcal{O}$$

For example, to order $1/m^2$ we have for $\mathcal{F}'$

$$\mathcal{F}'_{-1} = \mathcal{E}_{-1}, \qquad \mathcal{F}'_0 = \mathcal{E}_0, \qquad \mathcal{F}'_1 = \mathcal{E}_1 + \tfrac{1}{2}[G_1, \mathcal{O}_0]$$

$$\mathcal{F}'_2 = \mathcal{E}_2 + \tfrac{1}{2}[G_2, \mathcal{O}_0] + \tfrac{1}{2}[G_1, \mathcal{O}_1]$$

To this order we therefore need to solve

$$-[G_1, \mathcal{E}_{-1}] = \mathcal{O}_0, \qquad -[G_2, \mathcal{E}_{-1}] = \mathcal{O}_1 + [G_1, \mathcal{E}_0]$$

For our applications we will always have

$$\mathcal{E}_{-1} = -\tfrac{1}{w} S^{-10}$$

unchanged by interactions. We have oversimplified things a bit in the above derivation: For general spin we need to consider more than just even and odd terms; we need to consider all eigenvalues of S^{-10}:

$$[S^{-10}, \mathcal{F}_s] = s\mathcal{F}_s$$

and find the transformation that makes $\mathcal{F}'$ commute with it ($s = 0$). The procedure is to first divide into even and odd values of s, as above, then to divide the remaining even terms in $\mathcal{F}'$ into twice even values of s (multiples of 4) as the new $\mathcal{E}'$ and twice odd as the new $\mathcal{O}'$, which are transformed away with the new twice odd G', and so on. This very rapidly removes the lower nonzero values of $|s|$ ($1 \to 2 \to 4 \to ...$), which has a maximum value of $2w$ (from the operators that mix the maximum value $S^{-10} = w$ with the minimum $S^{-10} = -w$). For example, for the case of most interest, the Dirac spinor, the only eigenvalues (for operators) are 0 and ± 1, so the original even part does commute with S^{-10}, and the procedure need be applied only once. Furthermore, terms in $\mathcal{F}$ of eigenvalue s can be generated only at order m^{1-s} or higher; so at any given order the procedure rapidly removes all undesired terms for any spin.

Since the terms we want to cancel are exactly the ones with nonvanishing eigenvalues of S^{-10}, they can always be written as $[G, S^{-10}]$ for some G, so we can always find a transformation to eliminate them:

$$[S^{-10}, G_{sn}] = sG_{sn} \quad \Rightarrow \quad G_{sn} = -\tfrac{w}{s}\{[G, \Delta\mathcal{E}] + \mathcal{L}_G\, coth(\mathcal{L}_G)\mathcal{O}\}_{sn}$$

(This is just diagonalization of a Hermitian matrix in operator language.) In particular for the Dirac spinor, since $\mathcal{E}_{-1}$ has only ± 1 eigenvalues, it's easy to see that not only do all even operators commute with it, but all odd operators anticommute with it. (Consider the diagonal representation of $\mathcal{E}_{-1}$: $\{\left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right), \left(\begin{smallmatrix}0 & b\\ a & 0\end{smallmatrix}\right)\} = 0$.) We then have simply

$$w = \tfrac{1}{2} \quad \Rightarrow \quad (\mathcal{E}_{-1})^2 = 1 \quad \Rightarrow \quad [\mathcal{E}_{-1}, \varDelta\mathcal{E}] = \{\mathcal{E}_{-1}, \mathcal{O}\} = \{\mathcal{E}_{-1}, G\} = 0$$

$$\Rightarrow \quad mG = -\tfrac{1}{2}\{[G, \varDelta\mathcal{E}] + \mathcal{L}_G\, coth(\mathcal{L}_G)\mathcal{O}\}\mathcal{E}_{-1}$$

As a final step, we can apply the usual transformation

$$U_0 = e^{imtS^{-10}/w}$$

which commutes with all but the p^0 term in $\mathcal{E}_0$ to have the sole effect of canceling $\mathcal{E}_{-1}$, eliminating the rest-mass term from the nonrelativistic-style expression for the energy.

For the minimal electromagnetic coupling described above, we have besides $\mathcal{E}_{-1}$

$$\mathcal{E}_0 = \pi^0, \qquad \mathcal{O}_0 = \tfrac{1}{w}S^{0i}\pi^i$$

where we have written $\pi^a = p^a + A^a$ (instead of $\pi^a = -i\nabla^a$, to save some i's). There are no additional terms in $\mathcal{F}$ for minimal coupling for spin 1/2, but later we'll need to include nonminimal effective couplings coming from quantum (field theoretic) effects. There are also extra terms for spins 0 and 1 because the field strength is not the same as the fundamental field, so we'll treat only spin 1/2 here, but we'll continue to use the general notation to illustrate the procedure. Using the above results, we find to order $1/m^2$ for $\mathcal{F}'$

$$G_1 = S^{-1i}\pi^i, \qquad G_2 = wS^{0i}iF^{0i}$$

in agreement with with the free case up to field strength terms. The diagonalized Schrödinger equation is then to this order, including the effect of U_0,

$$\mathcal{F}'_{-1} = 0, \qquad \mathcal{F}'_0 = \pi^0, \qquad \mathcal{F}'_1 = -\tfrac{1}{2w}[\tfrac{1}{2}\{S^{-1i}, S^{0j}\}iF^{ij} + S^{-10}(\pi^i)^2]$$

$$\mathcal{F}'_2 = -\tfrac{1}{4}[\{S^{0i}, S^{0j}\}(\partial^i F^{0j}) - S^{ij}\{iF^{0i}, \pi^j\}]$$

For spin 1/2 we are done, but for other spins we would need a further transformation (before U_0) to pick out the part of $\mathcal{F}'_2$ that commutes with S^{-10} (by eliminating the twice odd part); the final result is

$$\mathcal{F}'_2 = \tfrac{1}{4}[\tfrac{1}{2}(\{S^{-1i}, S^{-1j}\} - \{S^{0i}, S^{0j}\})(\partial^i F^{0j}) + S^{ij}\{iF^{0i}, \pi^j\}]$$

It can also be convenient to translate into $\pm$ notation (as for the massless case, but with index $1 \to -1$): We then write

$$\mathcal{F}'_{-1} = 0, \qquad \mathcal{F}'_0 = \pi^0, \qquad \mathcal{F}'_1 = -\tfrac{1}{2w}[\tfrac{1}{2}\{S^{+i}, S^{-j}\}iF^{ij} + S^{+-}(\pi^i)^2]$$

$$\mathcal{F}'_2 = -\tfrac{1}{4}[\tfrac{1}{2}\{S^{+(i}, S^{-j)}\}(\partial^i F^{0j}) - S^{ij}\{iF^{0i}, \pi^j\}]$$

In this notation the eigenvalue of $S^{+-} = S^{-10}$ for any combination of spin operators can be simply read off as the number of $-$ indices minus the number of $+$.

Exercise IIB5.2

Find the Hamiltonian for spin 1/2 in background electromagnetism, expanded nonrelativistically to this order, by substituting the appropriate expressions for the spin operators in terms of γ matrices, and applying $S^{+-} = \pm w$ on the right for positive/negative energy. (Ignore the reality-restoring transformation.) γ-matrix algebra can be performed directly with the spin operators: For the Dirac spinor we have the identities

$$S_{(a}{}^{(b}S_{c)}{}^{d)} = \tfrac{1}{2}\delta^b_{(a}\delta^d_{c)} - \eta_{ac}\eta^{bd} \quad \Rightarrow \quad \{S^{+i}, S^{-j}\} = \tfrac{1}{2}\delta^{ij} - 2S^{ij}S^{+-}$$

6. Twistors

Besides describing spin 1/2, spinors provide a convenient way to solve the condition $p^2 = 0$ covariantly: Any hermitian matrix with vanishing determinant must have a zero eigenvalue (consider the diagonalized matrix), and so such a 2×2 matrix can be simply expressed in terms of its other eigenvector. Absorbing all but the sign of the nontrivial eigenvalue into the normalization of the eigenvector, we have

$$p^2 = 0 \quad \Rightarrow \quad p^{\alpha\dot{\beta}} = \pm p^\alpha p^{\dot{\beta}}$$

for some spinor p^α (where $p^{\dot{\alpha}} \equiv (p^\alpha)^*$). Since p^0 is the (canonical) energy, the $\pm$ is the sign of the energy. This explains why time reversal (actually CT in the usual terminology) is not a linear transformation. Note that p^α is a commuting object, while most spinors are fermionic, and thus anticommuting (at least in quantum theory). Such commuting spinors are called "twistors".

Exercise IIB6.1

Show that, in terms of its energy E and the angular direction (θ, ϕ) (with respect to the "1" axis) of its velocity, a massless particle is described by the twistor

$$p^\alpha = 2^{1/4}\sqrt{|E|}(cos\tfrac{\theta}{2}e^{i\phi/2}, sin\tfrac{\theta}{2}e^{-i\phi/2})$$

One useful way to think of twistors is in terms of the lightcone frame. In spinor notation, the momentum is

$$p^{\alpha\dot\beta} = \pm\left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)$$

If we write an arbitrary massless momentum as a Lorentz transformation from this lightcone frame, then the twistor is just the part of the SL(2,C) matrix that contributes:

$$p'^{\alpha\dot\beta} = p^{\gamma\dot\delta}g_\gamma{}^\alpha \bar g_{\dot\delta}{}^{\dot\beta} = \pm\delta^\gamma_\oplus\delta^{\dot\delta}_{\dot\oplus}g_\gamma{}^\alpha\bar g_{\dot\delta}{}^{\dot\beta} = \pm g_\oplus{}^\alpha\bar g_{\dot\oplus}{}^{\dot\beta} = \pm p^\alpha p^{\dot\beta}$$

For this reason, the twistor formalism can be understood as a Lorentz covariant form of the lightcone formalism.

The twistor construction thus gives a covariant way of constructing wave functions satisfying the mass-shell condition (Klein-Gordon equation) for the massless case, $\Box\psi = 0$, where $\Box = \partial^2 = -p^2$. We simply Fourier transform, and use the twistor expression for the momentum, writing the momentum-space wave function in terms of twistor variables ("Penrose transform"):

$$\psi(x) = \int d^2p_\alpha d^2\bar p_{\dot\alpha}[exp(ix^{\alpha\dot\beta}p_\alpha\bar p_{\dot\beta})\chi_+(p_\alpha,p_{\dot\alpha}) + exp(-ix^{\alpha\dot\beta}p_\alpha\bar p_{\dot\beta})\chi_-(p_\alpha,\bar p_{\dot\alpha})]$$

where $\chi_\pm$ describe the positive- and negative-energy states, respectively. (The integral over $\bar p_{\dot\alpha}$ can be performed also, effectively taking the Fourier transform with respect to that variable only, treating $\pm x^{\alpha\dot\beta}p_\alpha$ as the conjugate.)

We can extend the matrix notation of subsection IIA5-6 to twistors:

$$\boxed{\begin{gathered}\langle p| = p^\alpha\langle{}_\alpha|, \qquad |p\rangle = |{}^\alpha\rangle p_\alpha; \qquad [p| = p^{\dot\alpha}[{}_{\dot\alpha}|, \qquad |p] = |{}^{\dot\alpha}]p_{\dot\alpha}\\ P = |p\rangle[p|, \qquad\quad -P^* = |p]\langle p|\end{gathered}}$$

As a result, we also have for twistors

$$\boxed{\begin{gathered}\langle pq\rangle = -\langle qp\rangle, \qquad [pq] = -[qp]; \qquad\quad \langle pq\rangle^* = [qp]\\ \langle pq\rangle\langle rs\rangle + \langle qr\rangle\langle ps\rangle + \langle rp\rangle\langle qs\rangle = 0\end{gathered}}$$

These properties do not apply to physical, anticommuting spinors, where $\langle\psi\chi\rangle = +\langle\chi\psi\rangle$, and $\langle\psi\psi\rangle \neq 0$.

Another natural way to understand twistors is through the conformal group. We have already seen that the conformal group in D dimensions is SO(D,2). Since this group in four dimensions is the same as SU(2,2), it's simpler to describe its general representations (and in particular spinors) in SU(2,2) spinor notation. Then the simplest way to generate representations of this group is to use spinor coordinates: We therefore write the generators as (see subsection IC1)

$$G_{\mathcal A}{}^{\mathcal B} = \zeta^{\mathcal B}\bar\zeta_{\mathcal A} - \tfrac14\delta^{\mathcal B}_{\mathcal A}\zeta^{\mathcal C}\bar\zeta_{\mathcal C}$$

where we have subtracted out the trace piece to reduce U(2,2) to SU(2,2) and, consistently with the group transformation properties under complex conjugation, we have chosen the complex conjugate of the spinor to also be the canonical conjugate: The Poisson bracket is defined by

$$[\bar{\zeta}_{\mathcal{A}}, \zeta^{\mathcal{B}}] = \delta_{\mathcal{A}}^{\mathcal{B}}$$

To compare with four-dimensional notation, we reduce this four-component spinor by recognizing it as a particular use of the Dirac spinor. Using the same representation as in subsection IIA6, we write

$$\zeta^{\mathcal{A}} = (p^\alpha, \bar{\omega}^{\dot{\alpha}}), \qquad \bar{\zeta}_{\mathcal{A}} = (\omega_\alpha, \bar{p}_{\dot{\alpha}}); \qquad \Upsilon^{\dot{\mathcal{A}}\mathcal{B}} = \begin{pmatrix} 0 & \bar{C}^{\dot{\alpha}\dot{\beta}} \\ C^{\alpha\beta} & 0 \end{pmatrix}$$

Now the Poisson brackets are

$$[\omega_\alpha, p^\beta] = \delta_\alpha^\beta, \quad [\bar{\omega}_{\dot{\alpha}}, \bar{p}^{\dot{\beta}}] = \delta_{\dot{\alpha}}^{\dot{\beta}}$$

The group generators themselves reduce to

$$p_\alpha \bar{p}_{\dot{\beta}}, \quad \omega_\alpha \bar{\omega}_{\dot{\beta}}, \quad p_{(\alpha}\omega_{\beta)}, \quad \bar{p}_{(\dot{\alpha}}\bar{\omega}_{\dot{\beta})}, \quad p^\alpha \omega_\alpha + \bar{p}^{\dot{\alpha}}\bar{\omega}_{\dot{\alpha}} - 2$$

(for $E > 0$, with an overall $-$ for $E < 0$), which are translations, conformal boosts, SL(2,C) generators and their complex conjugates, and dilatations.

Another kind of twistor, related to position space instead of momentum space, follows from this (D+2)-coordinate description of conformal symmetry for D=4 (see subsection IA6). In practice, it's more convenient to work with invariances than constraints. In this case, we can solve the lightcone constraint on Wick-rotated D=3+3 or 5+1 space, replacing 6-component conformal vector indices with 4-component conformal spinor indices, with a position-space twistor:

$$y^2 = \tfrac{1}{4}\epsilon_{\mathcal{ABCD}} y^{\mathcal{AB}} y^{\mathcal{CD}} = 0 \quad \Rightarrow \quad y^{\mathcal{AB}} = z^{\mathcal{A}\alpha} z^{\mathcal{B}}{}_\alpha$$

where $\mathcal{A}$ is an SL(4) (or SU*(4)) index and α is an SL(2) (or SU(2)) index, and $z^{\mathcal{A}\alpha}$ is real (with either two real or two pseudoreal indices). (Here the SL groups apply to 3+3 dimensions, the SU groups to 5+1.) Whereas y had $6 - 1 = 5$ components due to the constraint, z has $4 \cdot 2 - 3 = 5$ components due to the SL(2) (SU(2)) gauge invariance of the above relation to y. These coordinates reduce to the usual by an SL(2) transformation:

$$\mathcal{A} = (\mu, \dot{\mu}), \quad z^{\mathcal{A}}{}_\alpha = \lambda_\alpha{}^\nu (\delta_\nu^\mu, x_\nu{}^{\dot{\mu}}) \quad \Rightarrow \quad SL(2)\ gauge\ \lambda_\alpha{}^\nu = \lambda \delta_\alpha^\nu$$

where $e = \lambda^2$.

Exercise IIB6.2

Substitute this spinor-notation $z(\lambda, x)$ into $y \sim z^2$ and compare with the vector-notation $y(e, x)$ of subsection IA6.

7. Helicity

A sometimes-useful way to treat the transverse spin operators S^{ij} is in terms of

$$W_{abc} = \tfrac{1}{2}P_{[a}J_{bc]} = \tfrac{1}{2}P_{[a}S_{bc]}$$

which reduces to S^{ij} in the rest frame, and (like the field equations) can be written in terms of just the Poincaré generators. This is the part of S_{ab} whose commutator with the field equations is proportional to the field equations (i.e., it preserves the constraints). In D=4 this is the "Pauli-Lubański (axial) vector"

$$W_a = \tfrac{1}{6}W^{bcd}\epsilon_{bcda}$$

We can choose our states to be eigenstates of a component of it: For example, for massless states W^0/P^0 is called the "helicity". For massive states the helicity is defined as $W^0/|\vec{P}|$, but is less useful, especially since it is undefined (0/0) in the rest frame. In that case one instead chooses a component in terms of a (momentum-dependent axial) vector s^a as $s^a W_a$, where $s^a P_a = 0$ and $s^2 = 1/m^2$.

Exercise IIB7.1

Show in both the massless and massive cases that W_{abc} reduces to the little group generators on shell by going to the appropriate reference frame.

The twistor representation of the conformal group does not give the most general representation, but it does give all the (free) massless ones. The reason it gives massless ones is that this representation satisfies the constraint (see subsection IIB1)

$$\mathcal{G}_{[\mathcal{AB}]}{}^{[\mathcal{CD}]} = G_{[\mathcal{A}}{}^{[\mathcal{C}}G_{\mathcal{B}]}{}^{\mathcal{D}]} - traces = 0$$

which includes $p^2 = 0$ as well as all the equations that follow from $p^2 = 0$ by conformal transformations. As a consequence, this representation also satisfies

$$G_{\mathcal{A}}{}^{\mathcal{C}}G_{\mathcal{C}}{}^{\mathcal{B}} - trace = hG_{\mathcal{A}}{}^{\mathcal{B}}$$

where h is the helicity. This equation may be more recognizable in SO(4,2) notation, as

$$\tfrac{1}{8}\epsilon^{ABCDEF}G_{CD}G_{EF} = ihG^{AB}$$

This equation includes, as its lowest mass-dimension part (as defined by dilatations), the Pauli-Lubański vector

$$W^a = \tfrac{1}{2}\epsilon^{bcda}P_b J_{cd} = ihP^a$$

(The "i" appears in the last two equations only when we use the antihermitian form of the generators G_{AB} and J_{ab}.) Although any massless representation of the conformal group satisfies the above conditions (see exercise IIB2.3), the twistor representation satisfies the unusual property that helicity is realized as a linear transformation on the coordinates: For the twistors the implicit definition of helicity can be solved explicitly to give

$$h = \tfrac{1}{4}\{\bar{\zeta}^{\mathcal{A}}, \zeta_{\mathcal{A}}\} = \tfrac{1}{2}\bar{\zeta}^{\mathcal{A}}\zeta_{\mathcal{A}} + 1 = \tfrac{1}{2}(p^{\alpha}\omega_{\alpha} - \bar{p}^{\dot{\alpha}}\bar{\omega}_{\dot{\alpha}})$$

(also for $E > 0$), which is exactly the U(1) transformation of U(2,2)=SU(2,2)⊗U(1). (This is similar to SU(2) in terms of "twistors": See exercise IC1.1.) On functions of p_{α} and $\bar{p}_{\dot{\alpha}}$, it effectively just counts half the number of p_{α}'s minus $\bar{p}_{\dot{\alpha}}$'s.

Exercise IIB7.2

These results are pretty clear from symmetry, but we should do some algebra to check coefficients: Express J_{ab} and P_a in terms of the twistors $p_{\alpha}, \bar{p}_{\dot{\alpha}}, \omega_{\alpha}, \bar{\omega}_{\dot{\alpha}}$ (see also exercise IIB2.3 for normalization), and plug into $\epsilon PJ = ihP$ to derive the above expression of h in terms of twistors.

The simple form of the helicity in the twistor formalism is another consequence of it being a covariantized lightcone formalism. In the lightcone frame, there is still a residual Lorentz invariance; in particular, a rotation about the spatial direction in which the momentum points leaves the momentum invariant. This is another definition of the helicity, as the part of the angular momentum performing that rotation. (Only spin contributes, since by definition the momentum is not rotated.) Since the product of two Lorentz transformations is another one, this rotation can be interpreted as a transformation acting on the Lorentz transformation to the lightcone frame, i.e., on the twistor, such that the momentum is invariant. This is simply a phase transformation:

$$g'_{\alpha}{}^{\beta} = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}_{\alpha}{}^{\gamma} g_{\gamma}{}^{\beta} \quad \Rightarrow \quad p'^{\alpha} = e^{i\theta}p^{\alpha}$$

We can generalize the Penrose transform in a simple way to wave functions carrying indices to describe spin:

$$\psi_{\alpha_1 \ldots \alpha_m \dot{\beta}_1 \ldots \dot{\beta}_n}(x) = \int d^2p_{\alpha} d^2\bar{p}_{\dot{\alpha}}\ p_{\alpha_1} \cdots p_{\alpha_m} \bar{p}_{\dot{\beta}_1} \cdots \bar{p}_{\dot{\beta}_n}$$
$$\times \left[exp(ix^{\alpha\dot{\beta}}p_{\alpha}\bar{p}_{\dot{\beta}})\chi_{+}(p_{\alpha}, p_{\dot{\alpha}}) + exp(-ix^{\alpha\dot{\beta}}p_{\alpha}\bar{p}_{\dot{\beta}})\chi_{-}(p_{\alpha}, p_{\dot{\alpha}})\right]$$

For the integral to give a nonvanishing result, the integrand must be invariant under the U(1) transformation generated by the helicity operator h: In other words, $\chi_{\pm}$ must have a transformation under h, i.e., a certain helicity, that is exactly the opposite that of the explicit p factors that carry the external indices to give a contribution to the integral, since otherwise integrating over the phase of p_α would average it to zero. (Explicitly, if we derive the helicity by acting on the Penrose transform, this minus sign comes from integration by parts.) This means that $\psi(x)$ automatically has a certain helicity, half the number of dotted minus undotted indices:

$$h = \tfrac{1}{2}(n-m) \qquad [w = \tfrac{1}{2}(m+n)]$$

(also for $E > 0$), as given by the above twistor operator expression acting on $\chi_\pm$. (Alternatively, comparing the x-space form of the Pauli-Lubański vector, its action plus that of the twistor-space one must vanish on $|^\alpha\rangle p_\alpha$, so the helicity is again minus the twistor-space helicity operator acting on the prefactor.)

Since, after restricting to the appropriate helicity, the integral over this phase is trivial, we can also eliminate it by replacing the "volume" integral over the twistor or its complex conjugate (but not both) with a "surface" (boundary) integral:

$$\int d^2 p_\alpha \to \oint p^\alpha dp_\alpha$$

(Alternatively, we can insert a δ-function in the helicity.) The result is equivalent to the usual integral over the three independent components of the momentum.

This generalization of the Penrose transform implies that $\psi(x)$ satisfies some equations of motion besides $p^2 = 0$, namely

$$p^{\alpha\dot\alpha}\psi_{\alpha\ldots\beta\dot\gamma\ldots\dot\delta} = p^{\gamma\dot\gamma}\psi_{\alpha\ldots\beta\dot\gamma\ldots\dot\delta} = 0$$

which are also implied by $S_a{}^b\partial_b + w\partial_a = 0$ (see exercise IIB2.3). Besides Poincaré invariance, these equations are invariant under the phase transformation

$$\psi'_{\alpha\ldots\beta\dot\gamma\ldots\dot\delta} = e^{i2h\theta}\psi_{\alpha\ldots\beta\dot\gamma\ldots\dot\delta}$$

that generalizes duality and chiral transformations. We also see that (anti-)self-duality and chirality are related to helicity. Another way to understand the twistor result is to remember its interpretation as a Lorentz transformation from the light cone: In the light cone frame, where $p^{+\dot+}$ is the only nonvanishing component of $p^{\alpha\dot\alpha}$, the above equations of motion imply the only nonvanishing component of $\psi^{\alpha_1\ldots\alpha_m\dot\beta_1\ldots\dot\beta_n}$ is $\psi^{+\ldots+\dot+\ldots\dot+}$, which can be identified with χ_+ (for $p^{+\dot+} > 0$) or χ_- (for $p^{+\dot+} < 0$).

REFERENCES

1 E. Wigner, *Ann. Math.* **40** (1939) 149;
V. Bargmann and E.P. Wigner, *Proc. Nat. Acad. Sci. US* **34** (1946) 211:

little group; more general discussion of Poincaré representations and relativistic wave equations for D=4.

2 A.J. Bracken, *Lett. Nuo. Cim.* **2** (1971) 574;
A.J. Bracken and B. Jessup, *J. Math. Phys.* **23** (1982) 1925;
W. Siegel, *Nucl. Phys.* **B263** (1986) 93:
conformal constraints.

3 E.C.G. Stückelberg, *Helv. Phys. Acta* **11** (1938) 299.

4 P.A.M. Dirac, *Rev. Mod. Phys.* **21** (1949) 392:
lightcone gauge.

5 G. Nordström, *Phys. Z.* **15** (1914) 504:
dimensional reduction.

6 O. Klein, *Z. Phys.* **37** (1926) 895;
V. Fock, *Z. Phys.* **39** (1927) 226:
mass from dimensional reduction.

7 A.S. Eddington, *Proc. Roy. Soc. A* **121** (1928) 524:
γ_{-1} ("γ_5").

8 R.J. Duffin, *Phys. Rev.* **54** (1938) 1114;
N. Kemmer, *Proc. Roy. Soc.* **A173** (1939) 91.

9 M.H.L. Pryce, *Proc. Roy. Soc.* **A195** (1948) 62;
S. Tani, *Soryushiron Kenkyu* **1** (1949) 15 (in Japanese):
free version of Foldy-Wouthuysen.

10 L.L. Foldy and S.A. Wouthuysen, *Phys. Rev.* **78** (1950) 29.

11 M. Cini and B. Touschek, *Nuo. Cim.* **7** (1958) 422;
S.K. Bose, A. Gamba, and E.C.G. Sudarshan, *Phys. Rev.* **113** (1959) 1661:
"ultrarelativistic" version of Foldy-Wouthuysen transformation.

12 R. Penrose, *J. Math. Phys.* **8** (1967) 345, *Int. J. Theor. Phys.* **1** (1968) 61;
M.A.H. MacCallum and R. Penrose, *Phys. Rep.* **6** (1973) 241:
twistors.

13 W. Pauli, unpublished;
J.K. Lubański, *Physica* **IX** (1942) 310, 325.

C. SUPERSYMMETRY

Supersymmetry is a symmetry that relates fermions to bosons. It includes the Poincaré group as a subgroup. We'll see later that quantum field theory requires particles with integer spin to be bosons, and those with half-integer spin to be fermions. This means that any symmetry that relates bosonic wave functions/fields to fermionic ones must be generated by operators with half-integer spin. The simplest (but also the most general, at least of those that preserve the vacuum) is spin 1/2. In this section we look at representations, generalizing the results of the previous sections for Poincaré symmetry.

Although supersymmetry has not been experimentally verified yet, it is a major ingredient in the most promising generalizations of the Standard Model:

(1) The fact that it enlarges the symmetry of nature means that it further restricts the allowed models, and thus makes stronger predictions.

(2) The greater symmetry also simplifies quantum calculations in many ways, especially through the use of the concept of "superspace". The results of these calculations are also often simplified.

(3) Because supersymmetric calculations are simpler, they can be used to simplify nonsupersymmetric calculations, at both the classical and quantum levels.

(4) This simplification in quantum rules results in improved high-energy behavior. In some cases it even results in the absence of the infinities in momentum integration that occur in all nonsupersymmetric theories. Although these infinities can be removed in perturbation theory, their effects reappear upon summation of the expansion.

(5) This improvement at high energies also improves the experimental agreement of Grand Unified Theories of the strong, electromagnetic, and weak interactions.

1. Algebra

From quantum mechanics we know that for any operator A

$$\begin{aligned}\langle\psi|\{A,A^\dagger\}|\psi\rangle &= \sum_n (\langle\psi|A|n\rangle\langle n|A^\dagger|\psi\rangle + \langle\psi|A^\dagger|n\rangle\langle n|A|\psi\rangle) \\ &= \sum_n (|\langle n|A^\dagger|\psi\rangle|^2 + |\langle n|A|\psi\rangle|^2) \geq 0\end{aligned}$$

from inserting a complete set of states. In particular,

$$\{A,A^\dagger\} = 0 \quad \Rightarrow \quad A = 0$$

from examining the matrix element for all states $|\psi\rangle$. This means the anticommutation relations of the supersymmetry generators must be nontrivial.

We are then led to anticommutation relations of the form, in Dirac (Majorana) notation,

$$\{q, \bar{q}\} = \not{p} \qquad or \qquad \{q, q^\dagger\} = p^a \gamma_a \sqrt{2}\gamma_0$$

(We use translations instead of internal symmetry or Lorentz generators because of dimensional analysis: Bosonic fields differ in dimension from fermionic ones by half integers.) Note that this implies the positivity of the energy:

$$tr\{q, q^\dagger\} = \sqrt{2}p^a\, tr(\gamma_a\gamma_0) = \sqrt{2}p^a\, tr(\tfrac{1}{2}\{\gamma_a, \gamma_0\}) = p^0 \tfrac{1}{\sqrt{2}}\, tr\ I$$

Similar arguments imply that the supersymmetry generators are constrained, just as the momentum is constrained by the mass-shell condition. For example, in the massless case,

$$\{\not{p}q, \bar{q}\not{p}\} = \not{p}\not{p}\not{p} = -\tfrac{1}{2}p^2\not{p} = 0 \quad\Rightarrow\quad \not{p}q = 0$$

In four dimensions the commutation relations can be written in terms of irreducible spinors as

$$\{q^\alpha, \bar{q}^{\dot\beta}\} = p^{\alpha\dot\beta}, \qquad \{q, q\} = \{\bar{q}, \bar{q}\} = 0$$

This generalizes straightforwardly to more than one spinor, carrying a U(N) index:

$$\{q_i{}^\alpha, \bar{q}^{j\dot\beta}\} = \delta_i^j p^{\alpha\dot\beta}$$

Exercise IIC1.1

Show positivity of energy in 2-component spinor notation for 4D U(N) supersymmetry.

2. Supercoordinates

Since the momentum is usually represented as coordinate derivatives, we naturally look for a similar representation for supersymmetry. We therefore introduce an anticommuting spinor coordinate θ^α. Because of the anticommutation relations q can't be simply $\partial/\partial\theta$, but the modification is obvious:

$$q_\alpha = -i\frac{\partial}{\partial\theta^\alpha} + \tfrac{1}{2}\bar{\theta}^{\dot\beta}\frac{\partial}{\partial x^{\alpha\dot\beta}}, \qquad \bar{q}_{\dot\alpha} = -i\frac{\partial}{\partial\bar{\theta}^{\dot\alpha}} + \tfrac{1}{2}\theta^\beta\frac{\partial}{\partial x^{\beta\dot\alpha}}$$

We can also express supersymmetry in terms of its action on the "supercoordinates": Using the hermitian infinitesimal generator $\epsilon^\alpha q_\alpha + \bar{\epsilon}^{\dot\alpha} q_{\dot\alpha}$,

$$\delta\theta^\alpha = \epsilon^\alpha, \qquad \delta\bar{\theta}^{\dot\alpha} = \bar{\epsilon}^{\dot\alpha}, \qquad \delta x^{\alpha\dot\beta} = \tfrac{1}{2}i(\epsilon^\alpha\bar{\theta}^{\dot\beta} + \bar{\epsilon}^{\dot\beta}\theta^\alpha)$$

Note that $(q^\alpha)^\dagger = \bar{q}^{\dot\alpha}$, $(q_\alpha)^\dagger = -\bar{q}_{\dot\alpha}$.

We can also define "covariant derivatives": derivatives that (anti)commute with (are invariant under) supersymmetry. These are easily found to be

$$d_\alpha = \frac{\partial}{\partial\theta^\alpha} + \tfrac{1}{2}\bar\theta^{\dot\beta} p_{\alpha\dot\beta}, \qquad \bar{d}_{\dot\alpha} = \frac{\partial}{\partial\bar\theta^{\dot\alpha}} + \tfrac{1}{2}\theta^\beta p_{\beta\dot\alpha}$$

Besides overall normalization factors of i, leading to the opposite hermiticity condition $(d^\alpha)^\dagger = -\bar{d}^{\dot\alpha}$, these differ from the q's by the relative sign of the two terms. These changes combine to preserve

$$\{d^\alpha, \bar{d}^{\dot\beta}\} = p^{\alpha\dot\beta}, \qquad \{d,d\} = \{\bar d, \bar d\} = 0$$

as a result of which p is also a covariant derivative as well as being a symmetry generator (as for the Poincaré group), but now $(d_\alpha)^\dagger = +\bar{d}_{\dot\alpha}$.

In classical mechanics, the fact that $\partial/\partial x$ commutes with translations is "dual" to the fact that the infinitesimal change dx, or the finite change $x - x'$, is also invariant under translations. Furthermore, the d'Alembertian $\square = (\partial/\partial x)^2$ being Poincaré invariant is dual to the line element $ds^2 = -(dx)^2$ being invariant. This allows the construction of the action from $\dot{x}^2$. In the supersymmetric case the infinitesimal invariants under the q's (and therefore p) are

$$d\theta^\alpha, \qquad d\bar\theta^{\dot\alpha}, \qquad dx^{\alpha\dot\beta} + \tfrac{1}{2}i(d\theta^\alpha)\bar\theta^{\dot\beta} + \tfrac{1}{2}i(d\bar\theta^{\dot\beta})\theta^\alpha$$

and the corresponding finite ones (by integration) are

$$\theta^\alpha - \theta'^\alpha, \qquad \bar\theta^{\dot\alpha} - \bar\theta'^{\dot\alpha}, \qquad x^{\alpha\dot\beta} - x'^{\alpha\dot\beta} + \tfrac{1}{2}i\theta^\alpha\bar\theta'^{\dot\beta} + \tfrac{1}{2}i\bar\theta^{\dot\beta}\theta'^\alpha$$

Although these can be used to construct classical mechanics actions, their quantization is rather complicated. Just as for particles of one particular spin, direct treatment of the quantum mechanics has proven much simpler than deriving it by quantization of a classical system.

Exercise IIC2.1

Check explicitly the invariance of the above infinitesimal and finite differences under supersymmetry.

Now that we have a (super)coordinate representation of the supersymmetry generators, we can examine the wave functions/fields that carry this representation. Such "superfields" can be Taylor expanded in the θ's with a finite number of terms, with ordinary fields as the coefficients. For example, if we expand a real (hermitian) scalar superfield

$$\Phi(x,\theta,\bar\theta) = \phi(x) + \theta^\alpha\psi_\alpha(x) + \bar\theta^{\dot\alpha}\bar\psi_{\dot\alpha}(x) + ...$$

and also expand its supersymmetry transformation

$$\delta\Phi = \epsilon^\alpha \psi_\alpha + \bar{\epsilon}^{\dot{\alpha}} \bar{\psi}_{\dot{\alpha}} + \tfrac{1}{2} i \epsilon^\alpha \bar{\theta}^{\dot{\beta}} \partial_{\alpha\dot{\beta}} \phi + \tfrac{1}{2} i \bar{\epsilon}^{\dot{\alpha}} \theta^\beta \partial_{\beta\dot{\alpha}} \phi + ...$$

we find the component field transformations

$$\delta\phi = \epsilon^\alpha \psi_\alpha + \bar{\epsilon}^{\dot{\alpha}} \bar{\psi}_{\dot{\alpha}}, \qquad \delta\psi_\alpha = -\tfrac{1}{2} i \bar{\epsilon}^{\dot{\beta}} \partial_{\alpha\dot{\beta}} \phi + ..., \qquad \delta\bar{\psi}_{\dot{\alpha}} = -\tfrac{1}{2} i \epsilon^\beta \partial_{\beta\dot{\alpha}} \phi + ..., \qquad ...$$

which mix the different spins.

An alternative, and more convenient, way to define the θ expansion is by use of the covariant derivatives. Using "$|$" to mean "$|_{\theta=0}$", we can define

$$\phi = \Phi|, \qquad \psi_\alpha = (d_\alpha \Phi)|, \qquad \bar{\psi}_{\dot{\alpha}} = (\bar{d}_{\dot{\alpha}} \Phi)|, \qquad ...$$

There is some ambiguity at higher orders in θ because the d's don't anticommute, and this can be resolved according to whatever is convenient for the particular problem, avoiding field redefinitions in terms of fields appearing at lower order in θ: Since the field equations must be covariant under supersymmetry (otherwise there is no advantage to using superfields), they must be written with the covariant derivatives. Then one defines the component expansions by choosing the same ordering of d's as appear in the field equations (where relevant), which gives the component expansion of the field equations the simplest form. It also gives a convenient method for deriving supersymmetry transformations, since the d's anticommute with the q's:

$$\delta[(d...d\Phi)|] = [d...d(\delta\Phi)]| = [d...d(i\epsilon q\Phi)]| = [(i\epsilon q)d...d\Phi]| = [(\epsilon d)d...d\Phi]|$$

where we have used the fact that $q = -id + \theta$-stuff, where the θ-stuff is killed by evaluating at $\theta = 0$, once it has been pulled in front of all the θ-derivatives. Covariant derivatives can also be used for integration, since $\int d\theta = \partial/\partial\theta = d$ up to an x-derivative, which can be dropped when also integrating $\int dx$.

3. Supergroups

We saw certain relations between the lower-dimensional classical groups that turned out to be useful for just the cases of physical interest of rotational (SO(D−1)), Lorentz (SO(D−1,1)), and conformal (SO(D,2)) groups. In particular, the Poincaré group, though not a classical group, is a certain limit ("contraction") of the groups SO(D,1) and SO(D−1,2), and a subgroup of the conformal group. Similar remarks apply to supersymmetry, but because of its relation to spinors, these classical "supergroups" (or "graded" classical groups) exist only for certain lower dimensions, the same as those where covering groups for the orthogonal groups exist. In higher dimensions the supergroups do not correspond to supersymmetry, at least not in any way that can be represented on physical states.

We'll consider only the graded generalization of the classical groups that appear in the bosonic case. The basic idea is then to take the group metrics and combine them in ways that take into account the difference in symmetry between bosons and fermions:

$$\begin{array}{rl} \textbf{U}\text{nitary:} & \Upsilon^{\dot{A}B} \\ \textbf{O}\text{rtho}\textbf{S}\text{ym}\textbf{p}\text{lectic:} & M^{AB} \\ \textbf{R}\text{eal:} & \eta_{\dot{A}}{}^{B} \\ \text{pseudoreal (*):} & \Omega_{\dot{A}}{}^{B} \end{array}$$

where η is symmetric and Ω antisymmetric, as before, while M is graded symmetric: For $A = (a, \alpha)$ with bosonic indices a and fermionic ones α,

$$M^{[AB)} = 0 : \quad M^{ab} - M^{ba} = M^{a\beta} - M^{\beta a} = M^{\alpha\beta} + M^{\beta\alpha} = 0$$

Again we have inverse metrics, e.g.,

$$M^{KI} M_{KJ} = \delta_J^I$$

With respect to the usual index-contraction convention (no extra grading signs when superscript is contracted with subscript immediately following), we should take the ordering of indices on δ as $\delta_J{}^I$.

There is no analog of the ϵ tensor, at least for finite-dimensional groups, since it would have an infinite number of indices when totally symmetric. However, "special" supergroups can still be defined by generalizing the definition of trace and determinant to supermatrices. One convenient way to do this is by using Gaussian integrals, since this is a common way that such expressions will arise. As a generalization of the bosonic and fermionic identities we therefore define the "superdeterminant"

$$(sdet\ M)^{-1} = N \int dz^\dagger\ dz\ e^{-z^\dagger M z}$$

where "N" is a normalization factor defined so $sdet\ I = 1$. By explicitly evaluating the integral, separating out the commuting and anticommuting parts, we find (see exercise IB3.3)

$$sdet \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \frac{det\ A}{det(D - CA^{-1}B)} = \frac{det(A - BD^{-1}C)}{det\ D}$$

where here A and D contain only bosonic elements, while B and C are completely fermionic.

Exercise IIC3.1

Generalize exercise IB3.3 to superdeterminants: Divide up the range of a square matrix into two (not necessarily equal) parts, where each of the two parts may include indices of both fermionic and bosonic grading, so the four resulting blocks in the matrix may each include both commuting and anti-commuting elements. Show that

$$sdet \begin{pmatrix} A & B \\ C & D \end{pmatrix} = sdet\ D \cdot sdet(A - BD^{-1}C) = sdet\ A \cdot sdet(D - CA^{-1}B)$$

by integration.

The "supertrace" (see also exercise IA2.3c) then can be defined by generalizing the bosonic identity $det(e^M) = e^{tr\, M}$:

$$sdet(e^M) = e^{str\, M}$$

$$str(M_A{}^B) = (-1)^A M_A{}^A = M_a{}^a - M_\alpha{}^\alpha = tr\ A - tr\ D$$

follows, as in the bosonic case, from $\delta\ ln\ sdet\ M = str(M^{-1}\delta M)$, which is derived by varying the Gaussian definition.

Exercise IIC3.2

Show that for graded matrices we need to use *str* (and not *tr*) for the identity

$$str(MN) = str(NM)$$

A useful identity for superdeterminants can be derived by starting with the following identity for the inverse of a matrix for which the range of the indices has been divided into two pieces:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} (a - bd^{-1}c)^{-1} & (c - db^{-1}a)^{-1} \\ (b - ac^{-1}d)^{-1} & (d - ca^{-1}b)^{-1} \end{pmatrix}$$

We have assumed all the submatrices are square and invertible; equivalent expressions, which are more useful in other cases, can be derived easily by multiplying and dividing by the submatrices: For example,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} (a - bd^{-1}c)^{-1} & -a^{-1}b(d - ca^{-1}b)^{-1} \\ -d^{-1}c(a - bd^{-1}c)^{-1} & (d - ca^{-1}b)^{-1} \end{pmatrix}$$

From either of these we immediately see

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} \tilde{A} & \tilde{B} \\ \tilde{C} & \tilde{D} \end{pmatrix} \quad \Rightarrow \quad sdet \begin{pmatrix} A & B \\ C & D \end{pmatrix} = det\ A\ det\ \tilde{D} = \frac{1}{det\ D\ det\ \tilde{A}}$$

The graded generalizations of the classical groups are then

GL(m|n,C) [SL(m|n,C),SSL(n|n,C)]

U: [S]U(m_+,m_-|n) [SSU(n_+,n_-|n_++n_-)]
OSp: OSp(m|2n,C)
R: GL(m|n) [SL(m|n),SSL(n|n)]
: [S]U(2m|2n) [SSU*(2n|2n)]

U & OSp
R: OSp(m_+,m_-|2n)
: OSp(2m|2n)

where "$(m|n)$" refers to m bosonic and n fermionic indices, or vice versa. In the matrices of the defining representation, the elements with one bosonic index and one fermionic are anticommuting numbers, while those with both indices of the same kind are commuting. In particular, the commuting parts give the bosonic subgroups:

$$
\begin{array}{rcl}
\mathrm{GL(m|n,C)} & \supset & \mathrm{GL(m,C)\otimes GL(n,C)} \\
\mathrm{SL(m|n,C)} & \supset & \mathrm{GL(m,C)\otimes SL(n,C)} \\
\mathrm{SSL(n|n,C)} & \supset & \mathrm{SL(n,C)\otimes SL(n,C)} \\
\mathrm{U(m_+,m_-|n)} & \supset & \mathrm{U(m_+,m_-)\otimes U(n)} \\
\mathrm{SU(m_+,m_-|n)} & \supset & \mathrm{U(m_+,m_-)\otimes SU(n)} \\
\mathrm{SSU(n_+,n_-|n_+{+}n_-)} & \supset & \mathrm{SU(n_+,n_-)\otimes SU(n_+{+}n_-)} \\
\mathrm{OSp(m|2n,C)} & \supset & \mathrm{SO(m,C)\otimes Sp(2n,C)} \\
\mathrm{GL(m|n)} & \supset & \mathrm{GL(m)\otimes GL(n)} \\
\mathrm{SL(m|n)} & \supset & \mathrm{GL(m)\otimes SL(n)} \\
\mathrm{SSL(n|n)} & \supset & \mathrm{SL(n)\otimes SL(n)} \\
\mathrm{U^*(2m|2n)} & \supset & \mathrm{U^*(2m)\otimes U^*(2n)} \\
\mathrm{SU^*(2m|2n)} & \supset & \mathrm{U^*(2m)\otimes SU^*(2n)} \\
\mathrm{SSU^*(2n|2n)} & \supset & \mathrm{SU^*(2n)\otimes SU^*(2n)} \\
\mathrm{OSp(m_+,m_-|2n)} & \supset & \mathrm{SO(m_+,m_-)\otimes Sp(2n)} \\
\mathrm{OSp^*(2m|2n)} & \supset & \mathrm{SO^*(2m)\otimes USp(2n)}
\end{array}
$$

When the commuting and anticommuting dimensions are equal, we can impose tracelessness conditions on both bosonic parts of the generators separately ("SS": $tr\ A = tr\ D = 0$). This is related to the fact $str(I) = 0$ in such cases.

4. Superconformal

Since the conformal group is a classical group, its supersymmetric generalization should be a classical supergroup. Because the fermionic generators must include the supersymmetry generators, which are spinors, the representation of the conformal group that appears in the defining representation of the supergroup must be the spinor representation. However, we have seen that only for n≤6 (where covering groups exist) and n=8 (where the spinor of SO(8) is another of its defining representations) can the spinor representation of SO(n) be defined by classical group restrictions. This implies that the superconformal group exists only in D≤4 and D=6.

The relevant supergroups can be identified easily by looking at the bosonic subgroups:

$$\begin{aligned} D = 3: &\quad \mathrm{OSp(N|4)} \\ 4: &\quad \mathrm{SU(2,2|N)\ (or\ SSU(2,2|4))} \\ 6: &\quad \mathrm{OSp^*(8|2N)} \end{aligned}$$

(We consider only D>2, since the conformal group is infinite-dimensional in D≤2.) These three cases of D=3,4,6 are special for a number of reasons: In particular, these three supergroups can be related to SU(N|4) over the division algebras: the real numbers, complex numbers, and quaternions, respectively. (Similar remarks apply to their important classical bosonic subgroups: the conformal, Lorentz, and rotation groups. Attempts have been made to extend these results to the octonions for D=10, but with less success, and there seems to be no superconformal group for that case.) However, just as in the case of the Hilbert space of quantum mechanics, the complex numbers seems to be the best of these "division algebras", having the analytic properties the real numbers lack, while avoiding the noncommutativity of the quaternions. We'll see later that nontrivial interacting (local, classical) conformal field theories exist only in D≤4.

For example, for D=4 we find that the bosonic generators are the conformal group and the internal symmetry group U(N) (or SU(4) for N=4), while the fermionic generators include supersymmetry (N spinors) and its fraternal twin, "S-supersymmetry". As supersymmetry is the "square root" of translations, so S-supersymmetry is the square-root of conformal boosts.

Exercise IIC4.1

For D=4, write the (graded) commutation relations of the superconformal generators. Decompose them into representations of the Lorentz group, and find their commutation relations.

5. Supertwistors

We saw that a simple way to find representations of SO(4,2) was to use the coordinate representation for SU(2,2): The resulting twistors gave all massless representations ($p^2 = 0$ for all helicities). This method generalizes straightforwardly to the superconformal groups: The generators are

$$G_A{}^B = \zeta^B \bar{\zeta}_A$$

(For the SU case we should also subtract out the trace, but that generator commutes with the rest anyway.) The coordinates and their conjugate momenta satisfy

$$[\bar{\zeta}_A, \zeta^B\} = \delta_A^B$$

ζ^A is then in the defining representation of the supergroup, while the wave function, which is a function of ζ, contains more general representations.

For D=3, the reality condition sets $\zeta = \bar{\zeta}$, so the ζ's are the graded generalization of Dirac γ matrices. In fact, the anticommuting ζ's are the γ matrices of the SO(N) subgroup of the OSp(N|4). On the other hand, the commuting ζ's carry the index of the defining representation of Sp(4), so they are a spinor of SO(3,2), the 3D conformal group: They are the bosonic twistor, and can be used in a similar way to the 4D twistors discussed earlier.

For D=4, there is a U(1) symmetry acting on ζ under which $G_A{}^B$ is invariant, generated by $(-1)^A G_A{}^A$, as in the bosonic case: This is the "superhelicity".

For D=6, ζ is pseudoreal. In general, for pseudoreal representations of groups it is often convenient to introduce a new SU(2) under which the pseudoreal representation ζ^A and its equivalent complex conjugate representation $\bar{\zeta}^{\dot{B}} \Omega_{\dot{B}}{}^A$ transform as a doublet (SU(2) spinor). This is also obvious from construction, since half of the components are related to the complex conjugate of the other half. We then can write

$$\zeta^{Ai} = (\zeta^A, i\bar{\zeta}^{\dot{B}} \Omega_{\dot{B}}{}^A) = \bar{\zeta}^{\dot{B}\dot{k}} \Omega_{\dot{B}\dot{k}}{}^{Ai}; \qquad \Omega_{\dot{B}\dot{k}}{}^{Ai} = \Omega_{\dot{B}}{}^A C^{ki}, \quad M^{Ai,Bk} = M^{AB} C^{ik}$$

$$G_A{}^B = \zeta^{Bi} \zeta_{Ai}$$

(Thus, OSp*(2m|2n)⊂OSp(4n|4m), and SO*(2m)⊂Sp(4m), USp(2n)⊂SO(4n).) This means there is now an SU(2) symmetry on ζ, generated by

$$G_{ij} = \zeta^A{}_{(i} \zeta_{Aj)}$$

under which $G_A{}^B$ is invariant. This is the 6D version of superhelicity. In the D=6 light cone, the manifest part of Lorentz invariance is SO(D−2)=SO(4)=SU(2)⊗SU(2). This is one of those SU(2)'s.

We now concentrate on D=4 (although our methods generalize straightforwardly to D=3 and 6). The simplest way to find (massless) representations of 4D supersymmetry is to generalize the Penrose transform. Just as twistors automatically satisfy the massless field equations in D=4, supertwistors automatically satisfy their supersymmetric generalization. The supertwistor is the defining representation of SU(2,2|N). The SU(2,2) part is the usual twistor, while the SU(N) part is the usual fermionic creation and annihilation operators for SU(N). Thus, to relate superspace to supertwistors, we write

$$p_{\alpha\dot\beta} = -i\partial_{\alpha\dot\beta} \to \pm p_\alpha \bar p_{\dot\beta}$$

$$q_{i\alpha} = -i\partial_{i\alpha} + \tfrac{1}{2}\bar\theta_i^{\dot\beta}\partial_{\alpha\dot\beta} \to \pm a_i p_\alpha, \qquad \bar q^i_{\dot\alpha} = -i\bar\partial^i_{\dot\alpha} + \tfrac{1}{2}\theta^{i\beta}\partial_{\beta\dot\alpha} \to \pm a^{\dagger i}\bar p_{\dot\alpha}$$

This determines the Penrose transform from superspace to supertwistors:

$$\phi_{\alpha\ldots\dot\beta\ldots}(x,\theta,\bar\theta) = \int d^2p_\alpha d^2\bar p_{\dot\alpha} d^N a_i\ p_\alpha \cdots \bar p_{\dot\beta}\cdots [e^{i\varphi}\chi_+(p_\alpha,\bar p_{\dot\alpha},a_i) + e^{-i\varphi}\chi_-(p_\alpha,\bar p_{\dot\alpha},a_i)]$$

$$\varphi = (x^{\alpha\dot\beta} - i\tfrac{1}{2}\theta^{i\alpha}\bar\theta_i^{\dot\beta})p_\alpha\bar p_{\dot\beta} + \theta^{i\alpha}a_i p_\alpha$$

where we have used "chiral superfields" (trivial dependence on $\bar\theta$, via the constraint $\bar d^i_{\dot\alpha}\phi = 0$) without loss of generality. (Instead of treating a_i as coordinates to be integrated, we can also treat them as operators; we then make the χ's functions of $a^\dagger$, and replace the integration with vacuum evaluation $\langle 0|\ |0\rangle$.) As for ordinary twistors, this result can be related to the lightcone: For given momentum, we can choose the lightcone frame $p^\alpha = \delta^\alpha_\oplus$; then $q_i^\oplus = \pm a_i$, while $q_i^\ominus = 0$ is a result of the supertwistor formalism automatically incorporating $\not p q = 0$.

Exercise IIC5.1

Find the Penrose transform for D=3. (Warning: The anticommuting part of the twistor is now like Dirac matrices rather than creation/annihilation operators.)

Taylor expanding in a_i (and thus $\theta^{i\alpha}$, producing terms antisymmetric in $i...j$ and symmetric in $\alpha...\beta$), the states then carry the index structure $\phi, \phi_i, \phi_{ij}, ..., \tilde\phi^i, \tilde\phi$, totally antisymmetric, and terminating with another singlet, where

$$\tilde\phi = \tfrac{1}{N!}\epsilon^{i_1\cdots i_N}\phi_{i_1\cdots i_N}, \qquad \tilde\phi^{i_1} = \tfrac{1}{(N-1)!}\epsilon^{i_1\cdots i_N}\phi_{i_2\cdots i_N}, \qquad \ldots$$

From our discussion of helicity in subsection IIB7, we see that the states also decrease in helicity by 1/2 for each a (i.e., ignoring $\bar\theta$, each $\theta^{i\alpha}$ comes with a p_α, simply because it adds an undotted index). Taking the direct product with any helicity (coming from the explicit p_α's and $\bar p_{\dot\alpha}$'s carrying the external Lorentz indices), we see that the states have helicity $h, h-1/2, h-1, ..., h-N/2$, with multiplicty $\binom{N}{n}$ for helicity $h-n/2$:

state	helicity (Poincaré)	multiplicity [SU(N)]
ϕ	h	1
ϕ_i	$h - \frac{1}{2}$	N
ϕ_{ij}	$h - 1$	$\frac{N(N-1)}{2}$
$\vdots$	$\vdots$	$\vdots$
$\phi_{i_1 \cdots i_n}$	$h - \frac{n}{2}$	$\binom{N}{n}$
$\vdots$	$\vdots$	$\vdots$
$\tilde{\phi}^i$	$h - \frac{N}{2} + \frac{1}{2}$	N
$\tilde{\phi}$	$h - \frac{N}{2}$	1

This multiplet structure is carried separately by χ_+ and by χ_-, which are related by charge (complex) conjugation, one describing the antiparticles of the other, as for ordinary twistors. (The existence of both multiplets also follows from CPT invariance, which is required for local actions, to be discussed in subsection IVB1. Here we generalized from the Penrose transform, which contained both terms as a consequence of being the most general solution to $S^{ab}p_b + wp^a = 0$, which is CPT invariant.) Because of the values of the helicities, we can impose a reality condition, identifying all states with helicity j as the complex conjugates of those with $-j$, only for $-h = h - N/2 \to h = N/4$, when N is a multiple of 4. We can also get larger representations by taking the direct product of these smallest representations of supersymmetry with representations of U(N), in which case the fields will carry those additional SU(N) indices.

Exercise IIC5.2

Let's examine these 4D multiplets in detail:

- **a** List the SU(N) representations for each allowed value of N for each of the cases where the helicity $|h| \leq 1/2$ ("scalar multiplets"), 1 ("vector multiplets"), 3/2 ("gravitino multiplets"), or 2 ("graviton multiplets"), assuming the maximum-helicity state is a singlet.
- **b** Show that "supergravity" (graviton multiplets) can exist only for N≤8. Show that the relevant representation for N=8, if real, is the same as the one (complex) for N=7.
- **c** Find the analogous statements for "super Yang-Mills" (vector multiplets).

The explicit form of the reality condition is somewhat complicated in terms of the chiral superfields, because they are really field strengths of real gauge fields.

(Consider, for example, expressing reality of $A_{\alpha\dot{\beta}}$ in terms of $f_{\alpha\beta}$ in the case of electromagnetism.) However, in terms of the twistor variables, charge conjugation can be expressed as

$$C : \qquad \chi_{\pm} \to \chi_{\mp}{}^{*}, \qquad a_i \to (a_i)^{\dagger}$$

where the transformation on a_i is required because it carries the SU(N) "charge". Since this violates "chirality" in these variables (dependence on a and not $a^{\dagger}$), it is accomplished by Fourier transformation:

$$C : \qquad \chi_{\pm}(a_i) \to C \int d\tilde{a}^{\dagger i}\ e^{\tilde{a}^{\dagger i} a_i}[\chi_{\mp}(\tilde{a}_i)]^{*} C^{-1}$$

for some "charge conjugation matrix" C (in case the field carries an additional index).

REFERENCES

1 P. Ramond, *Phys. Rev.* **D3** (1971) 86;
A. Neveu and J.H. Schwarz, *Nucl. Phys.* **B31** (1971) 86, *Phys. Rev.* **D4** (1971) 1109:
2D supersymmetry in strings.

2 Yu.A. Gol'fand and E.P. Likhtman, *JETP Lett.* **13** (1971) 323;
D.V. Volkov and V.P. Akulov, *Phys. Lett.* **46B** (1973) 109;
J. Wess and B. Zumino, *Nucl. Phys.* **B70** (1974) 39:
4D supersymmetry.

3 A. Salam and J. Strathdee, *Nucl. Phys.* **B76** (1974) 477;
S. Ferrara, B. Zumino, and J. Wess, *Phys. Lett.* **51B** (1974) 239:
superspace.

4 Gates, Grisaru, Roček, and Siegel, *loc. cit.*;
J. Wess and J. Bagger, *Supersymmetry and supergravity*, 2nd ed. (Princeton University, 1992);
P. West, *Introduction to supersymmetry and supergravity*, 2nd ed. (World-Scientific, 1990);
I.L. Buchbinder and S.M. Kuzenko, *Ideas and methods of supersymmetry and supergravity, or a walk through superspace* (Institute of Physics, 1995).

5 Berezin, *loc. cit.* (IA):
superdeterminant.

6 P. Ramond, *Physica* **15D** (1985) 25:
supergroups and their relation to supersymmetry.

7 J. Wess and B. Zumino, *Nucl. Phys.* **B70** (1974) 39:
superconformal group in D=4.

8 R. Haag, J.T. Lopuszański, and M. Sohnius, *Nucl. Phys.* **B88** (1975) 257:
superconformal symmetry as the largest possible symmetry of the S-matrix, most general supersymmetries in D=4.

9 A. Ferber, *Nucl. Phys.* **B132** (1978) 55:
supertwistors.

10 T. Kugo and P. Townsend, *Nucl. Phys.* **B221** (1983) 357;
A. Sudbery, *J. Phys.* **A17** (1984) 939;
K.-W. Chung and A. Sudbery, *Phys. Lett.* **198B** (1987) 161:
spacetime symmetries and division algebras.

III. LOCAL

In the previous chapters we considered symmetries acting on coordinates or wave functions. For the most part, the transformations we considered had constant parameters: They were "global" transformations. In this chapter we will consider mostly field theory. Since fields are functions of spacetime, it will be natural to consider transformations whose parameters are also functions of spacetime, especially those that are localized in some small region. Such "local" or "gauge" transformations are fundamental in defining the theories that describe the fundamental interactions.

A. ACTIONS

A fundamental concept in physics, of as great importance as symmetry, is the action principle. In quantum physics the dynamics is necessarily formulated in terms of an action (in the path-integral approach), or an equivalent Hamiltonian (in the Heisenberg and Schrödinger approaches). Action principles are also convenient and powerful for classical physics, allowing all field equations to be derived from a single function, and making symmetries simpler to check.

1. General

We begin with some general properties of actions. (For this subsection we'll restrict ourselves to bosonic variables; however, in the following subsection we'll find that the only modification for fermions is a more careful treatment of signs.) Generally, equations of motion are derived from actions by setting their variation with respect to their arguments to vanish:

$$\delta S[\phi] \equiv S[\phi + \delta\phi] - S[\phi] = 0$$

The solutions to this equation (find ϕ, given S) are "extrema" of the action; generally we want them to be minima, corresponding to minima of the energy, so that they will be stable under small perturbations.

Exercise IIIA1.1

Often continuous coordinates are replaced with discrete ones, for calculational or conceptual purposes. Consider

$$S = -\sum_{n=-\infty}^{\infty} \tfrac{1}{2}(q_{n+1} - q_n)^2$$

The integer n is interpreted as a discrete time, in terms of some "small" unit.

a Show that

$$\delta S = 0 \quad \Rightarrow \quad q_{n+1} - 2q_n + q_{n-1} = 0$$

b Examine the continuum limit of the action and equations of motion: Introduce appropriate factors of ϵ, with $t = n\epsilon$, and take the limit $\epsilon \to 0$.

Now we take the variables ϕ to be functions of time; thus, S is a function of functions, a "functional". It just means that S is a function of an infinite set of variables. We can generalize properties of ordinary functions (derivatives, etc.) as usual by considering discrete time and taking a continuum limit:

$$\begin{aligned} i = 1, 2, ... \quad &\to \quad t \in [-\infty, \infty] \\ \phi_i \quad &\to \quad \phi(t) \\ \sum_i \quad &\to \quad \int dt \\ \delta_{ij} \quad &\to \quad \delta(t - t') \\ \frac{\partial}{\partial \phi_i} \quad &\to \quad \frac{\delta}{\delta \phi(t)} \\ \int d\phi_i \quad &\to \quad \int D\phi(t) \end{aligned}$$

(the last, a "functional integral", will appear in quantum theory) where δ_{ij} is the usual Kronecker delta function, while $\delta(t - t')$ is the "Dirac delta function". It's not really a function, since it takes only the values 0 or ∞, but a "distribution", meaning it's defined only by integration:

$$\int dt' \ f(t')\delta(t - t') \equiv f(t)$$

Of course, the variable $\phi(t)$ can also carry an index (or indices). In field theory, it will also be a function of more coordinates, those of space.

For example, making these substitutions into the definition of a (partial) derivative to get a "functional derivative",

$$\frac{\partial f(\phi_i)}{\partial \phi_j} = \lim_{\epsilon \to 0} \frac{f(\phi_i + \epsilon \delta_{ij}) - f(\phi_i)}{\epsilon} \quad \Rightarrow \quad \frac{\delta f[\phi(t)]}{\delta \phi(t')} = \lim_{\epsilon \to 0} \frac{f[\phi(t) + \epsilon \delta(t - t')] - f[\phi(t)]}{\epsilon}$$

Sometimes the functional derivative is defined in terms of that of the variable itself:

$$\frac{\delta \phi(t)}{\delta \phi(t')} = \delta(t - t')$$

If we apply this definition of the Dirac δ to $\delta\phi/\delta\phi$, we obtain the previous definition of the functional derivative. (Consider, e.g., varying $S = \int dt \ f(t)\phi(t)$ for a fixed function f.) However, in practice we never need to use these definitions of the functional derivative: The only thing for which we need a functional derivative is the action, whose functional derivative is defined by its variation,

$$\delta S[\phi] \equiv S[\phi + \delta\phi] - S[\phi] \equiv \int dt\ \delta\phi(t) \frac{\delta S}{\delta\phi(t)}$$

(The fact that the variation can always be written in this form is just the statement that it is linear in $\delta\phi$, since $\delta\phi$ is "infinitesimal".)

A general principle of mechanics is "locality", that events at one time directly affect only those events an infinitesimal time away. (In field theory these events can be also only an infinitesimal distance away in space.) This means that the action can be expressed in terms of a Lagrangian:

$$S[\phi] = \int dt\ L[\phi(t)]$$

where L at time t is a function of only $\phi(t)$ and a finite number of its derivatives. For more subtle reasons, this number of time derivatives is restricted to be no more than two for any term in L; after integration by parts, each derivative acts on a different factor of ϕ. The general form of the action is then

$$L(\phi) = -\tfrac{1}{2}\dot{\phi}^m \dot{\phi}^n g_{mn}(\phi) + \dot{\phi}^m A_m(\phi) + U(\phi)$$

where " $\dot{}$ " means $\partial/\partial t$, and the "metric" g, "vector potential" A, and "scalar potential" U are not to be varied independently when deriving the equations of motion. (Specifically, $\delta U = (\delta\phi^m)(\partial U/\partial\phi^m)$, etc. Note that our definition of the Lagrangian differs in sign from the usual.) The equations of motion following from varying an action that can be written in terms of a Lagrangian are

$$0 = \delta S \equiv \int dt\ \delta\phi^m \frac{\delta S}{\delta\phi^m} \quad \Rightarrow \quad \frac{\delta S}{\delta\phi^m} = 0$$

where we have eliminated $\delta\dot{\phi}^m$ terms by integration by parts (assuming $\delta\phi = 0$ at the boundaries in t), and used the fact that $\delta\phi(t)$ is arbitrary at each value of t. For example,

$$S = -\int dt\ \tfrac{1}{2}\dot{q}^2 \quad \Rightarrow \quad 0 = \delta S = -\int dt\ \dot{q}\delta\dot{q} = \int dt\ (\delta q)\ddot{q} \quad \Rightarrow \quad \frac{\delta S}{\delta q} = \ddot{q} = 0$$

Exercise IIIA1.2

Find the equations of motion for ϕ^m from the above general action in terms of the external fields g, A, and U (and their partial derivatives with respect to ϕ).

Locality applies only to the classical action; in quantum field theory we will also find "effective actions" that include nonlocal contributions from quantum effects. Similar effects can appear in classical theories; for example, electrodynamics in the Coulomb gauge includes a (spatially) nonlocal Coulomb interaction term. The interpretation is always that some quantity has been eliminated, which would return locality (e.g., the "longitudinal photon" in the Coulomb gauge). Such actions can still be varied by the same methods as above. However, one should always avoid the rule $\partial L/\partial\phi = \partial_t(\partial L/\partial_t\phi)$, since (1) it applies only to actions that can be expressed in terms of just ϕ and $\dot{\phi}$ (and not higher derivatives nor nonlocalities), and (2) it arbitrarily separates terms into two sets.

Such actions can be reduced to ones that are only linear in time derivatives by introducing additional variables. First, separate out the subspace where g is invertible, with coordinates q ($\phi^m = (q^i, r^\mu)$); the Lagrangian is then written as

$$L(q,r) = -\tfrac{1}{2}\dot{q}^i\dot{q}^j g_{ij}(q,r) + \dot{q}^i A_i(q,r) + \dot{r}^\mu A_\mu(q,r) + U(q,r)$$

This Lagrangian gives equivalent equations of motion to

$$L'(q,p,r) = [-\dot{q}^i p_i + \dot{r}^\mu A_\mu] + [\tfrac{1}{2}g^{ij}(p_i + A_i)(p_j + A_j) + U]$$

where g^{ij} is the inverse of g_{ij}. (Many other forms are possible by redefinitions of p.) Eliminating the new variables p by their equations of motion gives back $L(q,r)$. Note that this works only because p's equations of motion are algebraic: For example, eliminating x from the Lagrangian $-\dot{x}p + \frac{1}{2}p^2$ by the equation of motion $\dot{x} = p$ is illegal (it would give the trivial action $S = -\int dt\ \frac{1}{2}p^2$), since it would require solving for the time dependence of x. On the other hand, p is given explicitly in terms of the other variables by its equations of motion without inverting time derivatives, so eliminating it does not lose any of the dynamics. (It is an "auxiliary variable".)

The result is a Hamiltonian form of the Lagrangian:

$$L_H(\Phi) = i\dot{\Phi}^M \mathcal{A}_M(\Phi) + H(\Phi)$$

in terms of the Hamiltonian H, where $\Phi = (q,p,r)$. It has the "gauge invariance"

$$\delta\mathcal{A}_M = \partial_M \Lambda(\Phi)$$

(where $\partial_M = \partial/\partial\Phi^M$), since that adds only a total derivative term $i\dot{\Lambda}$. Clearly $\mathcal{A}$ will introduce a modification of the Poisson bracket if it is not linear in Φ (e.g., as when we make independent nonlinear redefinitions of coordinates and momenta on the usual form of the Lagrangian). To determine this modification we compare the equation of motion as defined by a Poisson bracket,

$$\dot{\Phi}^M = -i[\Phi^M, H] = -i[\Phi^M, \Phi^N]\partial_N H$$

with that following from varying the action,

$$-i\dot{\Phi}^N \mathcal{F}_{NM} + \partial_M H = 0, \qquad \mathcal{F}_{MN} = \partial_{[M}\mathcal{A}_{N]}$$

to find

$$[\Phi^M, \Phi^N] = (\mathcal{F}^{-1})^{NM}$$

where "$\mathcal{F}^{-1}$" is the inverse on the maximal subspace where $\mathcal{F}$ is invertible. The variables in the directions where $\mathcal{F}$ vanishes are "auxiliary", since they appear without time derivatives: Their equations of motion are not described by the Poisson bracket. In particular, if they appear linearly in H they are "Lagrange multipliers", whose variation imposes algebraic constraints on the rest of Φ.

Finally, we can make redefinitions of the part of Φ describing the invertible subspace so that $\mathcal{A}$ is linear:

$$\mathcal{A}_M = \tfrac{1}{2}\Phi^N \Omega_{NM} \quad \Rightarrow \quad L_H(\Phi) = \tfrac{1}{2}i\dot{\Phi}^M \Phi^N \Omega_{NM} + H(\Phi)$$

where Ω is a constant, hermitian, antisymmetric (and thus imaginary) matrix. For some purposes it is more convenient to assume this Hamiltonian form of the action as a starting point. We now have the canonical commutation relations as

$$[\Phi^M, \Phi^N] = \Omega^{MN}$$

where Ω^{MN} is the inverse of Ω_{NM} on the maximal subspace:

$$\Omega^{MN}\Omega_{PN} = \Pi_P{}^M$$

for the projection operator Π for that subspace.

Exercise IIIA1.3

For electromagnetism, define $\vec{\psi} = \vec{E} + i\vec{B}$.

a Show that Maxwell's equations (in empty space) can be written as two equations in terms of $\vec{\psi}$.

b Interpret the equation involving the time derivative as a Schrödinger equation for the wave function $\vec{\psi}$, and find the Hamiltonian operator.

c Define the obvious inner product $\int d^3x \;\; \vec{\psi}^* \cdot \vec{\psi}$: What physical conserved quantity does this represent? (Note that, unlike electrons, the number of photons is not conserved.)

Note that the requirement of the existence of a Hamiltonian formulation determines that the kinetic term for a particle in the Lagrangian formulation go as $\dot{x}^2$ and not $x\ddot{x}$. Although such terms give the same equations of motion, they are not equivalent quantum mechanically, where boundary terms (dropped when using integration by parts for deriving the equations of motion) contribute. Furthermore, the Hamiltonian form of the action

$$S = \int dt\ H - dx^i p_i$$

shows that the energy H relates to the time in the same way the momentum relates to the coordinates, except for an interesting minus sign that is explained only by special relativity.

2. Fermions

In nonrelativistic quantum mechanics, spin is usually treated as a quantum effect, rather than being derived from classical mechanics. Although it is possible to derive spin from classical mechanics, in general it is rather cumbersome, and involves first introducing a large number of spins and then constraining away all the undesired ones, whereas in the quantum mechanics one can just directly introduce some particular representation of the spin angular momentum operators. The one nontrivial exception is spin 1/2.

We know from quantum mechanics that the spin variables for spin 1/2 are described by the Pauli σ matrices. Since they satisfy anticommutation relations, and are represented by finite-dimensional matrices, they are interpreted as fermionic. We have already seen that classical fermions are described by anticommuting numbers, so we begin by considering general quantization of such objects.

We can now consider actions that depend on both commuting and anticommuting classical variables, $\Phi^M = (\phi^m, \psi^\mu)$, where now ϕ refers to the bosonic variables and ψ to the fermionic ones. The Hamiltonian form of the Lagrangian can again be written as

$$L_H(\Phi) = \tfrac{1}{2} i \dot{\Phi}^M \Phi^N \Omega_{NM} + H(\Phi)$$

When Ω is invertible, the *graded* bracket is defined by (see subsection IA2)

$$[\Phi^M, \Phi^N\} = \hbar \Omega^{MN}, \qquad \Omega^{MN}\Omega_{PN} = \delta_P^M$$

To describe spin 1/2, we therefore look for particle actions of the form

$$S_H = \int dt[-\dot{x}^i p_i + \tfrac{1}{2} i \dot{\psi}^i \psi_i + H(x, p, \psi)]$$

This corresponds to using

$$\Phi^M = (\phi^m; \psi^\mu) = (\phi^{i\alpha}; \psi^i) = (x^i, p^i; \psi^i)$$

$$\Omega_{mn} = \Omega_{i\alpha,j\beta} = \delta_{ij} C_{\alpha\beta}, \qquad \Omega_{\mu\nu} = \Omega_{ij} = \delta_{ij}, \qquad \Omega_{m\nu} = \Omega_{\mu n} = 0$$

The fundamental commutation relations are then

$$[x^i, p_j] = i\hbar\delta_j^i, \qquad \{\psi^i, \psi^j\} = \hbar\delta^{ij} \qquad ([x,x] = [p,p] = [x,\psi] = [p,\psi] = 0)$$

We recognize ψ^i as the Pauli σ matrices (the Dirac matrices of subsection IC1 for the special case of SO(3), normalized as in subsection IIA1), $\psi^i = \sqrt{\hbar}\sigma^i$. The free Hamiltonian is just

$$H = \frac{p^2}{2m}$$

as for spin 0: Spin does not affect the motion of free particles.

A more interesting case is coupling to electromagnetism: Quantum mechanically, the Hamiltonian can be written in the simple form

$$H = \frac{\{\psi^i[p_i + qA_i(x)]\}^2}{m\hbar} - qA^0(x)$$

in terms of the vector and scalar potentials A^i and A^0. The classical expression is not as simple, because the commutation relations must be used to cancel the $1/\hbar$ before taking the classical limit. This is an example of "minimal coupling",

$$H(p_i) \to H(p_i + qA_i) - qA^0$$

However, this prescription works only if H for spin 1/2 is written in the above form: Using the commutation relations before or after minimal coupling gives different results. The form we have used is justified only by considering the nonrelativistic limit of the relativistic theory.

Exercise IIIA2.1

Use the multipication rules of the σ matrices to show that the quantum mechanical Hamiltonian for spin 1/2 in an electromagnetic field can be written as a spin-independent piece, identical to the spin-0 Hamiltonian, plus a term coupling the spin to the magnetic field.

3. Fields

Actions for field theories are just a special case (*not* a generalization) of the actions we have just considered: We just treat spatial coordinates $\vec{x}$ as part of the indices carried by the variables appearing in the action. In the notation used above,

$$M \to (i, \vec{x})$$
$$\Phi^M(t) \to \Phi^i(t, \vec{x})$$

Then spatial derivatives are just certain matrices with respect to the M index, $\int d\vec{x}$ comes from summation over M, etc.

The field equations for all field theories (e.g., electromagnetism) are wave equations. Wave equations also follow from mechanics upon quantization. Although classical field theory and quantum mechanics are not equivalent in their physical interpretation, they are mathematically equivalent in that they have identical wave equations. This is true not only for the free theories, but also for particles in external fields, and without direct self-interactions. This is no accident: Classical field theory and classical mechanics are two different limits of quantum field theory. They are both called classical limits, and written as $\hbar \to 0$, but since $\hbar$ is really 1, this limit depends on how one inserts $\hbar$'s into the quantum field theory action.

The wave equation in quantum mechanics is the Schrödinger equation. The corresponding field theory action is then simply the one that gives this wave equation as the equation of motion, where the wave function is replaced with the field:

$$S_{ft} = \int d^4x \ \psi^*(-i\partial_t + H)\psi$$

As usual (cf. electromagnetism), the field is a function of space and time; thus, we integrate $d^4x = dt \ d^3x$ over the three space and one time dimensions. The Hamiltonian is some function of coordinates and momenta, with the replacement $p_i \to -i\partial_i$, where $\partial_i = \partial/\partial x^i$ are the space derivatives and $\partial_t = \partial/\partial t$ is the time derivative.

The Hamiltonian can contain coupling to other fields. For a general Hamiltonian quadratic in momenta, in a notation implied by the corresponding Lagrangian quadratic in time derivatives,

$$H = \tfrac{1}{2} g^{ij}(-i\partial_i + A_i)(-i\partial_j + A_j) + U$$

where g^{ij}, A_i, and U are now interpreted as fields, and thus depend on both x^i and t, as does ψ. In the case $g^{ij} = \delta^{ij}$, we can identify A_i and U as the three-vector and scalar potentials of electromagnetism, and we can add the usual action for electromagnetism to the action for ψ. The action then can be varied also with respect to A and U to obtain Maxwell's equations with a current in terms of ψ and ψ^*. We can also treat g^{ij} as a field, in which case it and parts of A and U are the components of the gravitational field.

Field theory actions can be quantized in the same ways as mechanics ones. In this case, we recognize the $\psi^*\dot{\psi}$ term as a special case of the $\dot{\Phi}\Phi\Omega$ term in the generic Hamiltonian form of the action discussed earlier. Thus, $\psi(x^i)$ and $\psi^*(x_i)$ have replaced x^i and p_i as the variables; x^i is now just an index (label) on ψ and ψ^*, just as i was an index on x^i and p^i. The field-theory Hamiltonian is then identified as

$$H_{ft}[\psi,\psi^*] = \int d^3x\ \mathcal{H}, \qquad \mathcal{H} = \psi^* H \psi$$

In field theory the Hamiltonian will always be a space integral of a "Hamiltonian density" $\mathcal{H}$.

The classical limit of a quantum theory defined by a classical action S can be defined as follows: Introduce $\hbar$ into the theory by replacing

$$S \to \hbar^{-1} S$$

This has no effect on the classical equations of motion, but it introduces $\hbar$ into the Poisson bracket:

$$-\dot{q}p + H \quad \to \quad -\frac{1}{\hbar}\dot{q}p + \frac{1}{\hbar}H$$
$$\Rightarrow \quad \frac{1}{\hbar}[q,p] = i, \quad \frac{d}{dt} = \frac{\partial}{\partial t} + i\frac{1}{\hbar}[H,\ ]$$

We can then recognize the limit $\hbar \to 0$ as the classical limit. In the quantum theory, it is equivalent to replacing

$$p \to -i\hbar\partial_q, \qquad i\partial_t - H \to i\hbar\partial_t - H$$

i.e., all derivatives get a factor of $\hbar$. (More details will be possible when we consider quantization in subsection VA2.) However, a quantum theory can often be described by more than one classical action: This is known as "duality" (between any two such actions).

In particular, any free quantum field theory, and many interacting ones, can be described by both a classical mechanics action and by a classical field theory action: This is the well-known "wave-particle duality". We have just seen the standard nonrelativistic example. Furthermore, since we know the direct relation between the two actions in terms of the mechanics Hamiltonian H, we can describe both classical limits directly in terms of just the field theory action. The classical field theory limit is defined by inserting $\hbar$ as

$$S_{ft} \to \hbar_{ft}^{-1} S_{ft}$$

On the other hand, if we put in $\hbar$'s as

$$\partial_i \to \hbar_m \partial_i, \qquad \partial_t \to \hbar_m \partial_t$$

which gives the usual $\hbar$ dependence associated with the Schrödinger equation, then the classical limit $\hbar_m \to 0$ gives classical mechanics. This defines classical mechanics as the macroscopic limit, the limit of large distances and times.

A convenient way to implement this limit is to introduce the mechanics action $S = \int dt(-\dot{x}^i p_i + H)$ into the field theory, and then take the limit $\hbar_m \to 0$ after the replacement

$$S \to \hbar_m^{-1} S$$

on the mechanics action instead of on the derivatives. The mechanics action can be introduced when solving the field equations: The solution to the wave equation can be expressed in terms of the propagator, which in turn can be written in terms of the mechanics action or Hamiltonian.

More generally, we can define actions that are not restricted to be quadratic in any field. The Hamiltonian density $\mathcal{H}(t,x^i)$ or Lagrangian density $\mathcal{L}(t,x^i)$,

$$S[\phi] = \int dt\ d^3x\ \mathcal{L}[\phi(t,x^i)]$$

should be a function of fields at that point, with only a finite number (usually no more than two) spacetime derivatives. This is the definition of locality used for general quantum systems in subsection IIIA1, but extended from derivatives in time to also those in space. Although this condition is not always used in nonrelativistic field theory (for example, when long-range interactions, such as Coulomb or gravitational, are described without attributing them to fields), it is crucial in relativistic field theory. For example, global symmetries lead by locality to local (current) conservation laws. Locality is also the reason that spacetime coordinates are so important: Translation invariance says that the position of the origin is an unphysical, redundant variable; however, locality is most easily used with this redundancy.

Field equations are derived by the straightforward generalization of the variation of actions defined in subsection IIIA1: As follows from treating the spatial coordinates in the same way as discrete indices,

$$\delta S \equiv \int dt\ d^3x\ \delta\phi^m(t,x^i)\frac{\delta S}{\delta\phi^m(t,x^i)}$$

For example,

$$S = -\int dt\ d^3x\ \tfrac{1}{2}\dot{\phi}^2 \quad \Rightarrow \quad \frac{\delta S}{\delta \phi} = \ddot{\phi}$$

Exercise IIIA3.1

Consider the action

$$S[\phi] = \int dt\ d^{D-1}x\ [-\tfrac{1}{2}\dot{\phi}^2 + V(\phi)]$$

for potential $V(\phi)$ (a function, not a functional).

a Find the field equations.

b Assume $V(\phi) = \lambda\phi^n$ for some positive integer n and constant, *dimensionless* λ, in units $\hbar = c = 1$. Use dimensional analysis to relate n and D (of course, also a positive integer), and list all paired possibilities of (n, D).

4. Relativity

Generalization to relativistic theories is straightforward, except for the fact that the Klein-Gordon equation is second-order in time derivatives; however, we are familiar with such actions from nonrelativistic quantum mechanics. As usual, we need to check the sign of the terms in the action: Checking the positivity of the Hamiltonian (i.e., the energy), we see from the general relation between the Lagrangian and Hamiltonian (subsection IIIA1) that the terms without time derivatives must be positive; the time-derivative terms are then determined by Lorentz covariance.

At this point we introduce some normalizations and conventions that will prove convenient for Fourier transformation and other reasons to be explained later. Whenever D-dimensional integrations are involved (as should be clear from context), we use

$$\int dx \equiv \int \frac{d^Dx}{(2\pi)^{D/2}}, \qquad \int dp \equiv \int \frac{d^Dp}{(2\pi)^{D/2}}$$

$$\delta(x-x') \equiv (2\pi)^{D/2}\delta^D(x-x'), \qquad \delta(p-p') \equiv (2\pi)^{D/2}\delta^D(p-p')$$

In particular, this normalization will be used in Green functions and actions. For example, these implicit 2π's appear in functional variations:

$$\delta S \equiv \int dx\ \delta\phi\,\frac{\delta S}{\delta\phi} \quad \Rightarrow \quad \frac{\delta}{\delta\phi(x)}\phi(x') = \delta(x-x')$$

The action for a real scalar is then

$$S = \int dx\ L, \qquad L = \tfrac{1}{4}(\partial\phi)^2 + V(\phi)$$

where $V(\phi) \geq 0$, and we now write L for the Lagrange density. In particular, $V = \frac{1}{4}m^2\phi^2$ for the free theory. The free field equation is then $p^2 + m^2 = -\Box + m^2 = 0$, replacing the nonrelativistic $-i\partial_t + H = 0$. For a complex scalar, we replace $\frac{1}{2}\phi\phi \to \bar{\chi}\chi$ in both terms.

We know from previous considerations (subsection IIB2) that the field equation for a free, massless, Dirac spinor is $\gamma \cdot \partial \Psi = 0$. The generalization to the massive case (subsection IIB4) is obvious from various considerations, e.g., dimensional analysis; the action is

$$S = \int dx\ \bar{\Psi}(i\partial\!\!\!/ + \tfrac{m}{\sqrt{2}})\Psi$$

in arbitrary dimensions, again using the notation $\partial\!\!\!/ = \gamma \cdot \partial$. In four dimensions, we can decompose the Dirac spinor into its two Weyl spinors (see subsection IIA6):

$$L = \bar{\Psi}(i\partial\!\!\!/ + \tfrac{m}{\sqrt{2}})\Psi = (\bar{\psi}_L^{\dot{\alpha}} i\partial^\alpha{}_{\dot{\alpha}}\psi_{L\alpha} + \bar{\psi}_R^{\dot{\alpha}} i\partial^\alpha{}_{\dot{\alpha}}\psi_{R\alpha}) + \tfrac{m}{\sqrt{2}}(\psi_L^\alpha \psi_{R\alpha} + \bar{\psi}_L^{\dot{\alpha}}\bar{\psi}_{R\dot{\alpha}})$$

For the case of the Majorana spinor, the 4D action reduces to that for a single Weyl spinor,

$$S = \int dx\ [-i\bar{\psi}^{\dot{\beta}}\partial_{\alpha\dot{\beta}}\psi^\alpha + \tfrac{m}{\sqrt{2}}\tfrac{1}{2}(\psi^\alpha\psi_\alpha + \bar{\psi}^{\dot{\alpha}}\bar{\psi}_{\dot{\alpha}})]$$

Note that in our conventions $\sigma^0_{\alpha\dot{\beta}} = \frac{1}{\sqrt{2}}\delta_{\alpha\beta}$ (and similarly for the opposite indices, since $\sigma^a_{\alpha\dot{\beta}} = \sigma_a^{\beta\dot{\alpha}}$), so that the time derivative term is always proportional to $\psi^\dagger(-i\partial_0)\psi$, as nonrelativistically (previous subsection).

A scalar field must be complex to be charged (i.e., a representation of U(1)): From the gauge transformation

$$\chi' = e^{i\lambda}\chi$$

we find the minimal coupling (for $q = 1$)

$$S_\chi = \int dx\ [\tfrac{1}{2}|(\partial + iA)\chi|^2 + \tfrac{1}{2}m^2|\chi|^2]$$

This action is also invariant under charge conjugation

$$C: \chi \to \chi^*, \qquad A \to -A$$

which changes the sign of the charge, since $\chi^{*\prime} = e^{-i\lambda}\chi^*$.

Exercise IIIA4.1

Let's consider the semiclassical interpretation of a charged particle as described by a complex scalar field ψ, with Lagrangian

$$L = \tfrac{1}{2}(|\nabla\psi|^2 + m^2|\psi|^2)$$

a Use the semiclassical expansion in $\hbar$ defined by

$$\nabla \to \hbar\partial + iqA, \qquad \psi \to \sqrt{\rho}e^{-i\mathcal{S}/\hbar}$$

Find the Lagrangian in terms of ρ and $\mathcal{S}$ (and the background field A), order-by-order in $\hbar$ (in this case, just $\hbar^0$ and $\hbar^2$).

b Take the semiclassical limit by dropping the $\hbar^2$ term in L, to find

$$L \to \rho\tfrac{1}{2}[(-\partial\mathcal{S} + qA)^2 + m^2]$$

Vary with respect to $\mathcal{S}$ and ρ to find the equations of motion. Defining

$$p \equiv -\partial\mathcal{S}$$

show that these field equations can be interpreted as the mass-shell condition and current conservation. Show that A couples to this current by varying L with respect to A.

The spinor field also needs doubling for charge. (Actually, the doubling can be avoided in the massless case; however, problems show up at the quantum level, related to the fact that there is no charge conjugation transformation without doubling.) The gauge transformations are similar to the scalar case, and the action again follows from minimal coupling, to an action that has the global invariance (λ = constant in the absence of A):

$$\psi'^{\alpha}_{L} = e^{i\lambda}\psi^{\alpha}_{L}, \quad \psi'^{\alpha}_{R} = e^{-i\lambda}\psi^{\alpha}_{R}$$

$$S_e = \int dx\ [\bar{\psi}^{\dot{\beta}}_{L}(-i\partial_{\alpha\dot{\beta}} + A_{\alpha\dot{\beta}})\psi^{\alpha}_{L} + \bar{\psi}^{\dot{\beta}}_{R}(-i\partial_{\alpha\dot{\beta}} - A_{\alpha\dot{\beta}})\psi^{\alpha}_{R} + \tfrac{m}{\sqrt{2}}(\psi^{\alpha}_{L}\psi_{R\alpha} + \bar{\psi}^{\dot{\alpha}}_{L}\bar{\psi}_{R\dot{\alpha}})]$$

The current is found from varying with respect to A:

$$J^{\alpha\dot{\beta}} = \bar{\psi}^{\dot{\beta}}_{L}\psi^{\alpha}_{L} - \bar{\psi}^{\dot{\beta}}_{R}\psi^{\alpha}_{R}$$

Charge conjugation

$$C : \psi^{\alpha}_{L} \leftrightarrow \psi^{\alpha}_{R}, \qquad A \to -A$$

(which commutes with Poincaré transformations) changes the sign of the charge and current.

Exercise IIIA4.2

Show that this action can be rewritten in Dirac notation as

$$S_e = \int dx\ \bar{\Psi}(i\partial\!\!\!/ - A\!\!\!/ + \tfrac{m}{\sqrt{2}})\Psi$$

and find the action of the gauge transformation and charge conjugation on the Dirac spinor.

As a last example, we consider the action for electromagnetism itself. As before, we have the gauge invariance and field strength

$$A'_{\alpha\dot{\beta}} = A_{\alpha\dot{\beta}} - \partial_{\alpha\dot{\beta}}\lambda$$

$$F_{\alpha\dot{\gamma},\beta\dot{\delta}} = \partial_{\alpha\dot{\gamma}}A_{\beta\dot{\delta}} - \partial_{\beta\dot{\delta}}A_{\alpha\dot{\gamma}} = C_{\alpha\beta}\bar{f}_{\dot{\gamma}\dot{\delta}} + \bar{C}_{\dot{\gamma}\dot{\delta}}f_{\alpha\beta}, \quad f_{\alpha\beta} = \tfrac{1}{2}\partial_{(\alpha\dot{\gamma}}A_{\beta)}{}^{\dot{\gamma}}$$

We can write the action for pure electromagnetism as

$$S_A = \int dx\ \tfrac{1}{2e^2}f^{\alpha\beta}f_{\alpha\beta} = \int dx\ \tfrac{1}{2e^2}\bar{f}^{\dot{\alpha}\dot{\beta}}\bar{f}_{\dot{\alpha}\dot{\beta}} = \int dx\ \tfrac{1}{8e^2}F^{ab}F_{ab}$$

dropping boundary terms, with the overall sign again determined by positivity of the Hamiltonian, where e is the electromagnetic coupling constant, i.e., the charge of the proton. (Other normalizations can be used by rescaling $A_{\alpha\dot{\beta}}$.) Maxwell's equations follow from varying the action with a source term added:

$$S = S_A + \int dx\ A^{\alpha\dot{\beta}}J_{\alpha\dot{\beta}} \quad \Rightarrow \quad \tfrac{1}{e^2}\partial^{\beta}{}_{\dot{\gamma}}f_{\beta\alpha} = J_{\alpha\dot{\gamma}}$$

Exercise IIIA4.3

By plugging in the appropriate expressions in terms of A_a (and repeatedly integrating by parts), show that all of the above expressions for the electromagnetism action can be written as

$$S_A = -\int dx\ \tfrac{1}{4e^2}[A\cdot\Box A + (\partial\cdot A)^2]$$

Exercise IIIA4.4

Find all the field equations for all the fields, found from adding to S_A all the minimally coupled matter actions above.

Having seen many of the standard examples of relativistic field theory actions, we now introduce one of the most important principles in field theory; unfortunately, it can be justified only at the quantum level (see chapter VII):

Good ultraviolet behavior: All quantum field theories should have only couplings with nonnegative mass (engineering) dimension.

Exercise IIIA4.5

Show in D=4 using dimensional analysis that this restriction on bosons ϕ and fermions ψ restricts terms in the action to be of the form

$$\phi, \phi^2, \phi^3, \phi^4, \phi\partial\phi, \phi^2\partial\phi, \phi\partial\partial\phi; \psi^2, \psi\partial\psi; \phi\psi^2$$

and find the dimensions of all the corresponding coupling constants.

The energy-momentum tensor for electromagnetism is much simpler in this spinor notation, and follows (up to normalization) from gauge invariance, dimensional analysis, Lorentz invariance, and the vanishing of its trace. It has a form similar to that of the current in electrodynamics:

$$T_{\alpha\beta\dot{\gamma}\dot{\delta}} = -\tfrac{1}{e^2} f_{\alpha\beta} \bar{f}_{\dot{\gamma}\dot{\delta}}$$

Note that it is invariant under the duality transformations of subsection IIA7 (as is the electrodynamic current under chirality).

We have used conventions where e appears multiplying only the action S_A, and not in the "covariant derivative"

$$\nabla = \partial + iqA$$

where q is the charge in units of e: e.g., $q = 1$ for the proton, $q = -1$ for the electron. Alternatively, we can scale A, as a field redefinition, to produce the opposite situation:

$$A \to eA : \qquad S_A \to \int dx\ \tfrac{1}{8} F^2, \qquad \nabla \to \partial + iqeA$$

The former form, which we use unless noted otherwise, has the advantage that the coupling appears only in the one term S_A, while the latter has the advantage that the kinetic (free) term for A is normalized the same way as for scalars. The former form has the further advantage that e appears in the gauge transformations of none of the fields, making it clear that the group theory does not depend on the value of e. (This will be more important when generalizing to nonabelian groups in section IIIC.)

Note that the massless part of the kinetic (free) terms in these actions are scale invariant (in arbitrary dimensions, when the dimension-independent forms are used), when the fields are assigned the scale weights found from conformal arguments in subsection IIB2.

Exercise IIIA4.5

Using vector notation, minimal coupling, and dimensional analysis, find the mass dimensions of the electric charge e in arbitrary spacetime dimensions, and show it is dimensionless only in $D = 4$.

An interesting distinction between gravity and electromagnetism is that static bodies always attract gravitationally, whereas electrically they repel if they are like and attract if they are opposite. This is a direct consequence of the fact that the graviton has spin 2 while the photon has spin 1: The Lagrangian for a field of integer spin s coupled to a current, in an appropriate gauge and the weak-field approximation, is

$$-\tfrac{1}{4s!} \phi^{a_1 \ldots a_s} \Box \phi_{a_1 \ldots a_s} + \tfrac{1}{s!} \phi_{a_1 \ldots a_s} J^{a_1 \ldots a_s}$$

where the sign of the first term is fixed by unitarity in quantum field theory. (Classically the sign can also be related to positivity of the energy.) From a scalar field in the semiclassical approximation (see exercise IIIA4.1 above), starting with

$$J^{a_1 \dots a_s} \sim \psi^* \overleftrightarrow{\partial}^{a_1} \cdots \overleftrightarrow{\partial}^{a_s} \psi$$

where "$A\overleftrightarrow{\partial}B$" means "$A\partial B - (\partial A)B$", we see that the current will be of the form

$$J^{a_1 \dots a_s} \sim \rho p^{a_1} \cdots p^{a_s}$$

for a scalar particle, for some ρ. (The same follows from comparing the expressions for currents and energy-momentum tensors for particles as in subsection IIIB4 below. The only way to get vector indices out of a scalar particle, to couple to the vector indices for the spin of the force field, is from momentum.) In the static approximation, only time components contribute: We then can write this Lagrangian as, taking into account $\eta_{00} = -1$,

$$-(-1)^s \tfrac{1}{4s!} \phi_{0 \dots 0} \Box \phi_{0 \dots 0} + \tfrac{1}{s!} \phi_{0 \dots 0} \rho (p^0)^s$$

where $E = p^0 > 0$ for a particle and < 0 for an antiparticle. Thus the spin-dependence of the potential/force between two particles goes as $(-E_1 E_2)^s$. It then follows that all particles attract by forces mediated by even-spin particles, and a particle and its antiparticle attract under all forces, while repulsion will occur for odd-spin forces between two identical particles. (We can substitute "particles of the same sign charge" for "identical particles", and "particles of opposite sign charge" for "particle and its antiparticle", where the charge is the coupling constant appropriate for that force.)

Exercise IIIA4.6

Show that the above current is conserved,

$$\partial_{a_1} J^{a_1 \cdots a_s} = 0$$

(and the same for the other indices, by symmetry) if ψ satisfies the free Klein-Gordon equation (massless or massive).

5. Constrained systems

Constraints not only frequently appear in nonrelativistic physics, but are a general feature of relativistic particles, so we now give a brief description of how they are incorporated into actions. Consider a general action, with constraints, in Hamiltonian form:

$$S = \int dt (-\dot{q}^m p_m + H), \quad H = H_{gi}(q,p) + \lambda^i G_i(q,p)$$

(For simplicity, we consider all physical variables to be bosonic for this subsection, but the method generalizes straightforwardly paying careful attention to signs.) This action is a functional of q^m, p_m, λ^i, which are in turn functions of t, where m and i run over any number of values. We can think of this as describing a nonrelativistic particle with coordinates q and momenta p in terms of time t, but the form is general enough to apply to relativistic theories. The $\dot{q}p$ term tells us p is canonically conjugate to q; the rest of the action gives the Hamiltonian, usually quadratic in momenta. The variables λ^i are "Lagrange multipliers", whose variation in the action implies the constraints $G_i = 0$. We then can interpret H_{gi} as the usual ("gauge invariant") Hamiltonian. We also require that the transformations generated by the constraints close, and that the Hamiltonian be invariant:

$$[G_i, G_j] = -if_{ij}{}^k G_k, \quad [G_i, H_{gi}] = 0$$

(More generally, we can allow $[G_i, H_{gi}] = -if_i{}^j G_j$.) This says that the constraints don't imply any new constraints that we might have missed, and that the "energy" represented by H_{gi} is invariant under these transformations.

We then find that the action is invariant under the canonical transformations

$$\delta(q,p) = i[\zeta^i G_i, (q,p)] \quad \Rightarrow \quad \delta q^m = \zeta^i \frac{\partial G_i}{\partial p_m}, \quad \delta p_m = -\zeta^i \frac{\partial G_i}{\partial q^m}$$

$$0 = \delta\left(\frac{d}{dt}\right) = \delta\left(\frac{\partial}{\partial t} + iH\right) = i(\delta\lambda^i)G_i - i\dot{\zeta}^i G_i + [\lambda^j G_j, \zeta^i G_i]$$

$$\Rightarrow \quad \delta\lambda^i = \dot{\zeta}^i + \zeta^j \lambda^k f_{kj}{}^i$$

(with $\delta(d/dt)$ defined as in subsection IA1), where $\partial/\partial t$ acts on the "explicit" t dependence (that in everything except q and p): For general expressions, the total time derivative and total variation are given by commutators as

$$\frac{d}{dt}A = \frac{\partial}{\partial t}A + i[H, A], \quad \delta A = \delta_0 A + i[\zeta^i G_i, A]$$

where δ_0 acts on everything except q and p. The action then varies under these transformations as the integral of a total derivative, which vanishes under appropriate boundary conditions:

$$\delta S_H = \int dt \frac{d}{dt}[-(\delta q^m)p_m + \zeta^i G_i] = 0$$

The simplest example is the case with one constraint, which is linear in the variables: If the constraint is p, the gauge transformation is $\delta q = \lambda$, so we gauge $q = 0$ and use the constraint $p = 0$. In general, this means that for every degree of freedom we can gauge away, the conjugate variable can be fixed by the constraint. Thus, for each constraint we eliminate 3 variables: the variable fixed by the constraint, its conjugate, and the Lagrange multiplier that enforced the constraint, which has no conjugate. (In the Lagrangian form of the action the conjugate may not appear explicitly, so only 2 variables are eliminated.) As another example, for a nonrelativistic particle constrained to a sphere, $G = (x^i)^2 - 1$, we can change to spherical coordinates, apply the constraint to eliminate the radial coordinate, and use the gauge invariance to eliminate the radial component of the momentum, leaving an unconstrained theory in terms of angles and their conjugates. In most cases in field theory a similar procedure can be applied: The result is called a "unitary gauge".

Exercise IIIA5.1

Let's look closer at this example:

a Perform quantization of a nonrelativistic particle on a sphere ($G = (x^i)^2 - 1$ for $i = 1, 2, 3$), reducing to an action in terms of just the angles θ and ϕ, and their conjugates.

b Repeat this procedure using instead the constraint $G' = x^i p_i$, and compare.

The standard example of a relativistic constrained system is in field theory — electromagnetism. Its action can be written in "first-order (in derivatives) formalism" by introducing an auxiliary field G_{ab}:

$$F^2 \to F^2 - G^2 \to F^2 - (G - F)^2 = 2GF - G^2$$

where in the first step we added a trivial term for G and in the second step made a trivial redefinition of G, so elimination of G by its algebraic equation of motion returns the original Lagrangian. The Hamiltonian form comes from eliminating only G_{ij} by its field equation, since only F_{0i} contains time derivatives:

$$2GF - G^2 \to (F_{ij})^2 - 4G_{0i}F_{0i} + 2(G_{0i})^2$$

$$= -4\dot{A}_i G_{0i} + [2(G_{0i})^2 + (F_{ij})^2] - 4A_0 \partial_i G_{0i}$$

which we recognize as the three generic terms for the action in Hamiltonian form, with G_{0i} as the canonical momenta for A_i, and A_0 as the Lagrange multiplier. The constraint is Gauss' law, and it generates the usual gauge transformations.

Thus λ^i are also gauge fields for the gauge (time-dependent) transformations $\zeta^i(t)$. They allow construction of the gauge-covariant time derivative

$$\nabla = \partial_t + i\lambda^i G_i, \qquad \frac{d}{dt} = \nabla + iH_{gi} \quad \Rightarrow \quad \delta\nabla = i[\zeta^i G_i, \nabla]$$

It is convenient to transform the gauge fields away using these gauge transformations, so $H = H_{gi}$. However, with the usual boundary conditions $\int_{-\infty}^{\infty} dt\ \lambda^i$ is gauge invariant under the linearized transformations, so the most we could expect is to gauge λ^i to constants. More precisely, the group element

$$\mathcal{T}\left[exp\left(-i\int_{-\infty}^{\infty} dt\ \lambda^i(t)G_i\right)\right]$$

is gauge invariant, where "$\mathcal{T}$" is time ordering, meaning we write the exponential of the integral as the product of exponentials of infinitesimal integrals, and order them with respect to time, later time intervals going to the left of earlier ones. (We treat G_i quantum mechanically or use Poisson brackets when combining the exponentials.) This is the quantum mechanical version of the time development resulting from the corresponding term in the classical action. It is also the phase factor coming from the infinite limit of the covariant time translation

$$e^{-k\nabla(t)} = \mathcal{T}\left[exp\left(-i\int_{t-k}^{t} dt'\ \lambda^i(t')G_i\right)\right]e^{-k\partial_t}$$

as seen from reordering the time derivatives when writing $e^{-k\nabla(t)}$ as the product of exponentials of infinitesimal exponents. This allows us to write the explicit gauge transformation

$$e^{-i\Lambda(t)} = \mathcal{T}\left[exp\left(-i\int_{t_0}^{t} dt'\ \lambda^i(t')G_i\right)\right] = e^{-(\Delta t)\nabla(t)}e^{(\Delta t)\partial_t}, \qquad \Delta t = t - t_0$$

$$\Rightarrow \quad \nabla'(t) = e^{i\Lambda(t)}\nabla(t)e^{-i\Lambda(t)} = e^{-(\Delta t)\partial_t}\nabla(t)e^{(\Delta t)\partial_t} = \nabla(t_0) = \partial_t + i\lambda^i(t_0)G_i$$

(where we define ∂_t to vary t while keeping $t - t_0$ fixed). Thus, we can gauge λ to its value at a fixed time t_0. Another way to see this is that varying λ^i in the action at a fixed time gives $G_i = 0$ at that time, but the remaining field equations imply $\dot{G}_i = 0$, so $G_i = 0$ always, and λ^i is redundant at other times. This means that if we carelessly impose $\lambda^i = 0$ at all times, we must also impose $G_i = 0$ at some fixed time.

Note that this special gauge transformation itself has a very simple gauge transformation: Transforming the λ in Λ by an arbitrary finite transformation $\zeta^i(t)$,

$$e^{-i\Lambda'(t)} = e^{-i\zeta^i(t)G_i}e^{-i\Lambda(t)}e^{i\zeta^i(t_0)G_i}$$

consistent with the transformation law of $\nabla'(t)$ above. Thus, applying the transformation Λ to any gauge-dependent quantity ϕ gives a gauge-independent quantity $\phi'(\phi,\lambda)$, which is invariant under the local transformations $\zeta(t)$ and transforms only under the "global" transformations $\zeta(t_0)$. Thus, fixing the gauge $\lambda(t) = 0$ is equivalent to working with gauge-invariant quantities.

Fixing an invariance of the action is not unique to gauge invariances: Global invariances also need to be fixed, although the procedure is so trivial we seldom discuss it. For example, even in nonrelativistic systems Galilean invariance needs to be fixed: When analyzing a specific problem, we often choose some object to be at rest (velocity transformations), choose another to be oriented or moving in a specific direction (rotations), and choose a specific event to happen at the origin of space and time (translations). Alternatively, we can work with Galilean invariants, just as in gauge theories we can work with gauge invariants; however, in practice, for explicit calculations (as opposed to discussing general properties), it is more convenient to fix the invariance, as this allows simplification of the equations.

REFERENCES

1 M.B. Halpern and W. Siegel, *Phys. Rev.* **D16** (1977) 2486:
Classical mechanics as a limit of quantum field theory.

2 P.A.M. Dirac, *Proc. Roy. Soc.* **A246** (1958) 326;
L.D. Faddeev, *Theo. Math. Phys.* **1** (1969) 1:
Hamiltonian formalism for constrained systems.

B. PARTICLES

The simplest relativistic actions are those for the *mechanics* (as opposed to field theory) of particles. These also give the simplest examples of gauge invariance in relativistic theories. Later we will find that various properties of the quantum mechanics of these actions help to explain some features of quantum field theory.

1. Free

For nonrelativistic mechanics, the fact that the energy is expressed as a function of the three-momentum is conjugate to the fact that the spatial coordinates are expressed as functions of the time coordinate. In the relativistic generalization, all the spacetime coordinates are expressed as functions of a parameter τ: All the points that a particle occupies in spacetime form a curve, or "worldline", and we can parametrize this curve in an arbitrary way. Such parameters generally can be useful to describe curves: A circle is better described by $x(\theta), y(\theta)$ than $y(x)$ (avoiding ambiguities in square roots), and a cycloid can be described explicitly only this way.

The action for a free, spinless particle then can be written in relativistic Hamiltonian form as

$$S_H = \int d\tau [-\dot{x}^m p_m + v\tfrac{1}{2}(p^2 + m^2)]$$

where v is a Lagrange multiplier enforcing the constraint $p^2 + m^2 = 0$. This action is very similar to nonrelativistic ones, but instead of $x^i(t), p_i(t)$ we now have $x^m(\tau), p_m(\tau), v(\tau)$ (where "$\dot{}$ " now means $d/d\tau$). The gauge invariance generated by $p^2 + m^2$ is

$$\delta x = \zeta p, \quad \delta p = 0, \quad \delta v = \dot{\zeta}$$

A more recognizable form of this invariance can be obtained by noting that any action $S(\phi^A)$ has invariances of the form

$$\delta\phi^A = \epsilon^{AB} \frac{\delta S}{\delta\phi^B}, \quad \epsilon^{AB} = -\epsilon^{BA}$$

which have no physical significance, since they vanish by the equations of motion. In this case we can add

$$\delta x = \epsilon(\dot{x} - vp), \quad \delta p = \epsilon\dot{p}, \quad \delta v = 0$$

and set $\zeta = v\epsilon$ to get

$$\delta x = \epsilon\dot{x}, \quad \delta p = \epsilon\dot{p}, \quad \delta v = (\epsilon v)^{\bullet}$$

We then can recognize this as a (infinitesimal) coordinate transformation for τ:

$$x'(\tau') = x(\tau), \quad p'(\tau') = p(\tau), \quad d\tau' v'(\tau') = d\tau\ v(\tau); \qquad \tau' = \tau - \epsilon(\tau)$$

The transformation laws for x and p identify them as "scalars" with respect to these "one-dimensional" (worldline) coordinate transformations (but they are vectors with respect to D-dimensional spacetime). On the other hand, v transforms as a "density": The "volume element" $d\tau\ v$ of the world line transforms as a scalar. This gives us a way to measure length on the worldline in a way independent of the choice of τ parametrization. Because of this geometric interpretation, we are led to constrain

$$v > 0$$

so that any segment of the worldline will have positive length. Because of this restriction, v is not a Lagrange multiplier in the usual sense. This has significant physical consequences: $p^2 + m^2$ is treated neither as a constraint nor as the Hamiltonian. While in nonrelativistic theories the Schrödinger equation is $(E - H)\psi = 0$ and $G_i\psi = 0$ is imposed on the initial states, in relativistic theories $(p^2 + m^2)\psi = 0$ is the Schrödinger equation: This is more like $H\psi = 0$, since p^2 already contains the necessary E dependence.

The Lagrangian form of the free particle action follows from eliminating p by its equation of motion $vp = \dot{x}$:

$$S_L = \int d\tau\ \tfrac{1}{2}(vm^2 - v^{-1}\dot{x}^2)$$

For $m \neq 0$, we can also eliminate v by its equation of motion $v^{-2}\dot{x}^2 + m^2 = 0$:

$$S = m \int d\tau \sqrt{-\dot{x}^2} = m \int \sqrt{-dx^2} = m \int ds = ms$$

The action then has the purely geometrical interpretation as the proper time; however, this last form of the action is awkward to use because of the square root, and doesn't apply to the massless case. Note that the v equation implies $ds = m(d\tau\ v)$, relating the "intrinsic" length of the worldline (as measured with the worldline volume element) to its "extrinsic" length (as measured by the spacetime metric). As a consequence, in the massive case we also have the usual relation between momentum and "velocity"

$$p^m = m\frac{dx^m}{ds}$$

(Note that p^0 is the energy, not p_0.)

Exercise IIIB1.1

Take the nonrelativistic limit of the Poincaré algebra:

a Insert the speed of light c in appropriate places for the structure constants of the Poincaré group (guided by dimensional analysis) and take the limit $c \to 0$ to find the algebra of the Galilean group.

b Do the same for the representation of the Poincaré group generators in terms of coordinates and momenta. In particular, take the limit of the Lorentz boosts to find the Galilean boosts.

c Take the nonrelativistic limit of the spinless particle action, in the form ms. (Note that, while the relativistic action is positive, the nonrelativistic one is negative.)

Exercise IIIB1.2

Consider the following action for a particle with additional fermionic variables γ and additional fermionic constraint $\gamma \cdot p$:

$$S_H = \int d\tau(-\dot{x}^m p_m - \tfrac{1}{2} i \dot{\gamma}^m \gamma_m + \tfrac{1}{2} v p^2 + i\lambda \gamma \cdot p)$$

where λ is also anticommuting so that each term in the action is bosonic.

a Find the algebra of the constraints, and the transformations they generate on the variables appearing in the action.

b Show that the "Dirac equation" $\gamma \cdot p|\Psi\rangle = 0$ implies $p^2|\Psi\rangle = 0$.

c Find the Lagrangian form of the action as usual by eliminating p by its equation of motion. (Note $\lambda^2 = 0$.)

Exercise IIIB1.3

Consider a "supercoordinate" X^m that is a function of both a fermionic variable ζ and the usual τ:

$$X^m(\tau, \zeta) = x^m(\tau) + i\zeta\gamma^m(\tau)$$

where the Taylor expansion in ζ terminates because $\zeta^2 = 0$. Identify x with the usual x, and γ with its fermionic partner introduced in the previous problem. In analogy to the way $\gamma \cdot p$ was the square root of the τ-translation generator $\frac{1}{2}p^2$, we can define a square root of $\partial/\partial\tau$ by the "covariant fermionic derivative"

$$D = \frac{\partial}{\partial\zeta} + i\zeta\frac{\partial}{\partial\tau} \quad \Rightarrow \quad D^2 = i\frac{\partial}{\partial\tau}$$

We also want to generalize v in the same way as x, to make the action independent of coordinate choice for both τ and ζ. This suggests defining

$$E = v^{-1} + i\zeta\lambda$$

and the gauge invariant action

$$S_L = \int d\tau d\zeta \ \tfrac{1}{2}E(D^2X^m)DX_m$$

Integrate this action over ζ, and show this agrees with the action of the previous problem after suitable redefinitions (including the normalization of $\int d\zeta$).

The (D+2)-dimensional (conformal) representation of the massless particle (subsection IA6) can be derived from the action

$$S = \int d\tau \ \tfrac{1}{2}(-\dot{y}^2 + \lambda y^2)$$

where λ is a Lagrange multiplier. This action is gauge invariant under

$$\delta y = \epsilon\dot{y} - \tfrac{1}{2}\dot{\epsilon}y, \quad \delta\lambda = \epsilon\dot{\lambda} + 2\dot{\epsilon}\lambda + \tfrac{1}{2}\dddot{\epsilon}$$

If we vary λ to eliminate it and y^- as in subsection IA6, the action becomes

$$S = -\int d\tau \ \tfrac{1}{2}e^2\dot{x}^2$$

which agrees with the previous result, identifying $v = e^{-2}$, which also guarantees $v > 0$.

Exercise IIIB1.4

Find the Hamiltonian form of the action for y: The constraints are now y^2, r^2, and $y \cdot r$, in terms of the conjugate r to y (see exercise IA6.2). Find the gauge transformations in the standard way (see subsection IIIA5). Show how the above Lagrangian form can be obtained from it, including the gauge transformations.

Using instead the corresponding twistor (subsection IIB6) to satisfy $y^2 = 0$, the massless, spinless particle now has a single term for its mechanics action:

$$S = \int d\tau \ \tfrac{1}{4}\epsilon_{\mathcal{ABCD}}\dot{z}^{\mathcal{A}\alpha}\dot{z}^{\mathcal{B}}{}_{\alpha}z^{\mathcal{C}\beta}z^{\mathcal{D}}{}_{\beta}$$

Unlike all other relativistic mechanics actions, all variables have been unified into just z, without the introduction of square roots.

Exercise IIIB1.5

Expressing z in terms of $\lambda_\alpha{}^\mu$ and $x^{\mu\dot{\mu}}$ as in subsection IIB6, show this action reduces to the previous one.

2. Gauges

Rather than use the equation of motion to eliminate v it's more convenient to use a gauge choice: The gauge $v = 1$ is called "affine parametrization" of the worldline. Note that the gauge transformation of v, $\delta v = \dot{\zeta}$, has no dependence on the coordinates x and momenta p, so that choosing the gauge $v = 1$ avoids any extraneous x or p dependence that could arise from the gauge fixing. (The appearance of such dependence will be discussed in later chapters.) Since $T = \int d\tau\ v$, the intrinsic length, is gauge invariant, that part of v still remains when the length is finite, but it can be incorporated into the limits of integration: The gauge $v = 1$ is maintained by $\dot{\zeta} = 0$, and this constant ζ can be used to gauge one limit of integration to zero, completely fixing the gauge (i.e., the choice of τ). We then integrate $\int_0^T$, where $T \geq 0$ (since originally $v > 0$), and T is a variable to vary in the action. The gauge-fixed action is then

$$S_{H,AP} = \int_0^T d\tau[-\dot{x}^m p_m + \tfrac{1}{2}(p^2 + m^2)]$$

In the massive case, we can instead choose the gauge $v = 1/m$; then the equations of motion imply that τ is the proper time. The Hamiltonian $p^2/2m$ + constant then resembles the nonrelativistic one.

Another useful gauge is the "lightcone gauge"

$$\tau = \frac{x^+}{p^+}$$

which, unlike the Poincaré covariant gauge $v = 1$, fixes τ completely; since the gauge variation $\delta(x^+/p^+) = \zeta$, we must set $\zeta = 0$ to maintain the gauge. Also, the gauge transformation is again x and p independent. In lightcone gauges we always assume $p^+ \neq 0$, since we often divide by it. This is usually not too dangerous an assumption, since we can treat $p^+ = 0$ as a limiting case (in D>2).

We saw from our study of constrained systems that, for every degree of freedom we can gauge away, the conjugate variable can be fixed by the constraint that generates that gauge invariance: In the case where the constraint is p, the gauge transformation is $\delta q = \lambda$, so we gauge $q = 0$ and use the constraint $p = 0$. In lightcone gauges the constraints are almost linear: The gauge condition is $x^+ = p^+\tau$ and the constraint is $p^- = ...$, so the Lagrange multiplier v is varied to determine p^-. On the other hand, varying p^- gives

$$\delta p^- \quad \Rightarrow \quad v = 1$$

so this gauge is a special case of the gauge $v = 1$. An important point is that we used only "auxiliary" equations of motion: those not involving time derivatives. (A slight trick involves the factor of p^+: This is a constant by the equations of motion, so we can ignore $\dot{p}^+$ terms. However, technically we should not use that equation of motion; instead, we can redefine $x^- \to x^- + ...$, which will generate terms to cancel any $\dot{p}^+$ terms.) The net result of gauge fixing and the auxiliary equation on the action is

$$S_{H,LC} = \int_{-\infty}^{\infty} d\tau [\dot{x}^- p^+ - \dot{x}^i p^i + \tfrac{1}{2}(p^{i2} + m^2)]$$

where $x^a = (x^+, x^-, x^i)$, etc. In particular, since we have fixed one more gauge degree of freedom (corresponding to constant ζ), we have also eliminated one more constraint variable (T, the constant part of v). This is one of the main advantages of lightcone gauges: They are "unitary", eliminating all unphysical degrees of freedom.

Exercise IIIB2.1

Another obvious gauge is $\tau = x^0$, which works as well as the lightcone gauge as far as eliminating worldline coordinate invariance is concerned. (The same is true for $\tau = n \cdot x$ for any constant vector n.)

a Consider the auxiliary equations of motion: Apply this gauge condition; then p^0 appears without time derivatives, so eliminate it and v by their equations of motion. Show this gauge is consistent only for $p^0 > 0$.

b The resulting square root is awkward except in the nonrelativistic limit: Take it, and compare with the usual nonrelativistic mechanics.

c A better type of gauge is $\tau = n \cdot x / n \cdot p$, what we actually used for the lightcone. Compare the value of v that results from the field equations in this case to that of the case $\tau = n \cdot x$. Discuss the consistency of this case in terms of the allowed signs of $n \cdot x$ and $n \cdot p$ vs. those of τ and v.

3. Coupling

One way to introduce external fields into the mechanics action is by considering the most general Lagrangian quadratic in τ derivatives:

$$S_L = \int d\tau [-\tfrac{1}{2} v^{-1} g_{mn}(x) \dot{x}^m \dot{x}^n + A_m(x) \dot{x}^m + v\phi(x)]$$

In the free case we have constant fields $g_{mn} = \eta_{mn}$, $A_m = 0$, and $\phi = \frac{1}{2}m^2$. The v dependence has been assigned consistent with worldline coordinate invariance. The curved-space metric tensor g_{mn} describes gravity, the D-vector potential A_m describes electromagnetism, and ϕ is a scalar field that can be used to introduce mass by interaction.

Exercise IIIB3.1

Use the method of the problem IIIB1.3 to write the *nonrelativistic* action for a spinning particle in terms of a 3-vector (or (D−1)-vector) $X^i(\tau,\zeta)$ and the fermionic derivative D. Find the coupling to a magnetic field, in terms of the 3-vector potential $A_i(X)$. Integrate the Lagrangian over ζ. Show that the quantum mechanical square of $\psi^i[p_i + A_i(x)]$ is proportional to the Hamiltonian.

Exercise IIIB3.2

Derive the relativistic Lorentz force law

$$\partial_\tau(v^{-1}\dot{x}_m) + F_{mn}\dot{x}^n = 0$$

by varying the Lagrangian form of the action for the relativistic particle, in an external electromagnetic field (but flat metric), with respect to x.

This action also has very simple transformation properties under D-dimensional gauge transformations on the external fields:

$$\delta g_{mn} = \epsilon^p\partial_p g_{mn} + g_{p(m}\partial_{n)}\epsilon^p, \quad \delta A_m = \epsilon^p\partial_p A_m + A_p\partial_m\epsilon^p - \partial_m\lambda, \quad \delta\phi = \epsilon^p\partial_p\phi$$

$$\Rightarrow \quad S_L[x] + \delta S_L[x] = S_L[x+\epsilon] - \lambda(x_f) + \lambda(x_i)$$

where we have integrated the action $\int_{\tau_i}^{\tau_f} d\tau$ and set $x(\tau_i) = x_i$, $x(\tau_f) = x_f$. These transformations have a very natural interpretation in the quantum theory, where

$$\int Dx\ e^{-iS} = \langle x_f|x_i\rangle$$

Then the λ transformation of A is canceled by the U(1) (phase) transformation

$$\psi'(x) = e^{i\lambda(x)}\psi(x)$$

in the inner product

$$\langle\psi_f|\psi_i\rangle = \int dx_f dx_i\ \langle\psi_f|x_f\rangle\langle x_f|x_i\rangle\langle x_i|\psi_i\rangle = \int dx_f dx_i\ \psi_f{}^*(x_f)\langle x_f|x_i\rangle\psi_i(x_i)$$

while the ϵ transformation associated with g_{mn} is canceled by the D-dimensional coordinate transformation

$$\psi'(x) = \psi(x+\epsilon)$$

4. Conservation

There are two types of conservation laws generally found in physics: In mechanics we usually have global conservation laws, of the form $\dot{Q} = 0$, associated with a symmetry of the Hamiltonian H generated by a conserved quantity Q:

$$0 = \delta H = i[Q, H] = -\dot{Q}$$

On the other hand, in field theory we have local conservation laws, since the action for a field is written as an integral $\int d^D x$ of a Lagrangian density that depends only on fields at x, and a finite number of their derivatives. The local conservation law implies a global one, since

$$\partial_m J^m = 0 \quad \Rightarrow \quad 0 = \int \frac{d^D x}{(2\pi)^{D/2}}\, \partial_m J^m \sim \frac{d}{dt}\int \frac{d^{D-1}x}{(2\pi)^{D/2}}\, J^0 = \dot{Q} = 0$$

where we have integrated over a volume whose boundaries in space are at infinity (where J vanishes), and whose boundaries in time are infinitesimally separated. Equivalently, the global symmetry is a special case of the local one.

A simple way to derive the local conservation laws is by coupling gauge fields: We couple the electromagnetic field A_m to arbitrary charged matter fields ϕ and demand gauge invariance of the matter part of the action, the matter-free part of the action being separately invariant. We then have

$$0 = \delta S_M = \int dx \left[(\delta A_m)\frac{\delta S_M}{\delta A_m} + (\delta\phi)\frac{\delta S_M}{\delta\phi} \right]$$

using just the definition of the functional derivative δ/δ. Applying the matter field equations $\delta S_M/\delta\phi = 0$, integration by parts, and the gauge transformation $\delta A_m = -\partial_m \lambda$, we find

$$0 = \int dx\ \lambda \left(\partial_m \frac{\delta S_M}{\delta A_m} \right) \quad \Rightarrow \quad J^m = \frac{\delta S_M}{\delta A_m}, \quad \partial_m J^m = 0$$

Similar remarks apply to gravity, but only if we evaluate the "current", in this case the energy-momentum tensor, in flat space $g_{mn} = \eta_{mn}$, since gravity is self-interacting. We then find

$$T^{mn} = -2\, \frac{\delta S_M}{\delta g_{mn}}\bigg|_{g_{mn}=\eta_{mn}}, \quad \partial_m T^{mn} = 0$$

where the normalization factor of -2 will be found later for consistency with the particle. In this case the corresponding "charge" is the D-momentum:

$$P^m = \int \frac{d^{D-1}x}{(2\pi)^{D/2}}\ T^{0m}$$

In particular we see that the condition for the energy in any region of space to be nonnegative is

$$T^{00} \geq 0$$

Exercise IIIB4.1

Show that the local conservation of the energy-momentum tensor allows definition of a conserved angular momentum

$$J^{mn} = \int \frac{d^{D-1}x}{(2\pi)^{D/2}}\ x^{[m}T^{n]0}$$

Show this result (local conservation of energy-momentum inplies local conservation of angular momentum) is the same as that of exercise IA4.1.

To apply this to the action for the particle in external fields, we must first distinguish the particle coordinates $X(\tau)$ from coordinates x for all of spacetime: The particle exists only at $x = X(\tau)$ for some τ, but the fields exist at all x. In this notation we can write the mechanics action as

$$S_L = \int dx \left[-g_{mn}(x) \int d\tau \; \delta(x-X)\tfrac{1}{2}v^{-1}\dot{X}^m\dot{X}^n \right.$$
$$\left. + A_m(x) \int d\tau \; \delta(x-X)\dot{X}^m + \phi(x) \int d\tau \; \delta(x-X)v \right]$$

using $\int dx \; \delta(x - X(\tau)) = 1$. We then have

$$J^m(x) = \int d\tau \; \delta(x-X)\dot{X}^m$$

$$T^{mn} = \int d\tau \; \delta(x-X)v^{-1}\dot{X}^m\dot{X}^n$$

Note that $T^{00} \geq 0$ (since $v > 0$). Integrating to find the charge and momentum:

$$Q = \int d\tau \; \delta(x^0 - X^0)\dot{X}^0 = \int dX^0 \; \epsilon(\dot{X}^0)\delta(x^0 - X^0) = \epsilon(p^0)$$

$$P^m = \int d\tau \; \delta(x^0 - X^0)v^{-1}\dot{X}^0\dot{X}^m = \int dX^0 \; \epsilon(\dot{X}^0)\delta(x^0 - X^0)v^{-1}\dot{X}^m = \epsilon(p^0)p^m$$

where we have used $p = v^{-1}\dot{X}$ (for the free particle), where p is the momentum conjugate to X, not to be confused with P. The factor of $\epsilon(p^0)$ ($\epsilon(u) = u/|u|$ is the sign of u) comes from the Jacobian from changing integration variables from τ to X^0.

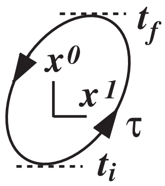
The result is that our naive expectations for the momentum and charge of the particle can differ from the correct result by a sign. In particular p^0, which semiclassically is identified with the angular frequency of the corresponding wave, can be either positive or negative, while the true energy $P^0 = |p^0|$ is always positive, as physically required. (Otherwise all states could decay into lower-energy ones: There would be no lowest-energy state, the "vacuum".) When p^0 is negative, the charge Q and $dX^0/d\tau$ are also negative. In the massive case, we also have dX^0/ds negative. This means that as the proper time s increases, X^0 decreases. Since the proper time is the time as measured in the rest frame of the particle, this means that the particle is traveling backward in time: Its clock changes in the direction opposite to that of the coordinate system x^m. Particles traveling backward in time are called "antiparticles", and have charges opposite to their corresponding particles. They have positive true energy, but the "energy" p^0 conjugate to the time is negative.

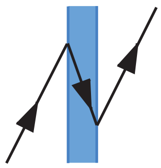
Exercise IIIB4.2

Compare these expressions for the current and energy-momentum tensor to those from the semiclassical expansion in exercise IIIA4.1. (Include the *inverse* metric to define the square of $-\partial_m \mathcal{S} + qA_m$ there.)

5. Pair creation

Free particles travel in straight lines. Nonrelativistically, external fields can alter the motion of a particle to the extent of changing the signs of spatial components of the momentum. Relativistically, we might then expect that interactions could also change the sign of the energy, or at least the canonical energy p^0. As an extreme case, consider a worldline that is a closed loop: We can pick τ as an angular coordinate around the loop. As τ increases, X^0 will either increase or decrease. For example, a circle in the x^0-x^1 plane will be viewed by the particle as repeating its history after some finite τ, moving forward with respect to time x^0 until reaching a latest time t_f, and then backward until some earliest time t_i. On the other hand, from the point of view of an observer at rest with respect to the x^m coordinate system, there are no particles until $x^0 = t_i$, at which time both a particle and an antiparticle appear at the same position in space, move away from each other, and then come back together and disappear. This process is known as "pair creation and annihilation".





Whether such a process can actually occur is determined by solving the equations of motion. A simple example is a particle in the presence of only a static electric field, produced by the time component A^0 of the potential. We consider the case of a piecewise constant potential, vanishing outside a certain region and constant inside. Then the electric field vanishes except at the boundaries, so the particle travels in straight lines except at the boundaries. For simplicity we reduce the problem to two dimensions:

$$A^0 = -V \ \ for \ 0 \leq x^1 \leq L, \quad 0 \ otherwise$$

for some constant V. The action is, in Hamiltonian form,

$$S_H = \int d\tau \ \{-\dot{x}^m p_m + v\tfrac{1}{2}[(p+A)^2 + m^2]\}$$

and the equations of motion are

$$\dot{p}_m = -v(p+A)^n \partial_m A_n \quad \Rightarrow \quad p^0 = E$$

$$(p+A)^2 = -m^2 \quad \Rightarrow \quad p^1 = \pm\sqrt{(E+A^0)^2 - m^2}$$

$$v^{-1}\dot{x} = p + A \quad \Rightarrow \quad v^{-1}\dot{x}^1 = p^1, \quad v^{-1}\dot{x}^0 = E + A^0$$

where E is a constant (the canonical energy at $x^1 = \infty$) and the equation $\dot{p}^1 = ...$ is redundant because of gauge invariance. We assume $E > 0$, so initially we have a particle and not an antiparticle.

We look only at the cases where the worldline begins at $x^0 = x^1 = -\infty$ (lower left) and continues toward the right till it reaches $x^0 = x^1 = +\infty$ (upper right), so that $p^1 = v^{-1}\dot{x}^1 > 0$ everywhere (no reflection). However, the worldline might bend backward in time ($\dot{x}^0 < 0$) inside the potential: To the outside viewer, this looks like pair creation at the right edge before the first particle reaches the left edge; the antiparticle then annihilates the original particle when it reaches the left edge, while the new particle continues on to the right. From the particle's point of view, it has simply traveled backward in time so that it exits the right of the potential before it enters the left, but it is the same particle that travels out the right as came in the left. The velocity of the particle outside and inside the potential is

$$\frac{dx^1}{dx^0} = \begin{cases} \dfrac{\sqrt{E^2 - m^2}}{E} & \text{outside} \\[2ex] \dfrac{\sqrt{(E-V)^2 - m^2}}{E - V} & \text{inside} \end{cases}$$

From the sign of the velocity we then see that we have normal transmission (no antiparticles) for $E > m + V$ and $E > m$, and pair creation/annihilation when

$$V - m > E > m \quad \Rightarrow \quad V > 2m$$

The true "kinetic" energy of the antiparticle (which appears only inside the potential) is then $-(E - V) > m$.

Exercise IIIB5.1

This solution might seem to violate causality. However, in mechanics as well as field theory, causality is related to boundary conditions at infinite times. Describe another solution to the equations of motion that would be interpreted by an outside observer as pair creation *without any initial particles*: What happens ultimately to the particle and antiparticle? What are the allowed values of their *kinetic* energies (maximum and minimum)? Since many such pairs can be created by the potential alone, it can be accidental (and not acausal) that an external particle meets up with such an antiparticle. Note that the generator of the potential, to maintain its value, continuously loses energy (and charge) by emitting these particles.

REFERENCES

1 A. Barducci, R. Casalbuoni, and L. Lusanna, *Nuo. Cim.* **35A** (1976) 377;
L. Brink, S. Deser, B. Zumino, P. DiVecchia, and P. Howe, *Phys. Lett.* **64B** (1976) 435;
P.A. Collins and R.W. Tucker, *Nucl. Phys.* **B121** (1977) 307:
world-line metric; classical mechanics for relativistic spinors.

2 R. Marnelius, *Phys. Rev.* **D20** (1979) 2091;
W. Siegel, *Int. J. Mod. Phys. A* **3** (1988) 2713:
conformal invariance in classical mechanics actions.

3 W. Siegel, hep-th/9412011, *Phys. Rev.* **D52** (1995) 1042:
twistor formulation of classical mechanics action.

4 R.P. Feynman, *Phys. Rev.* **74** (1948) 939:
classical pair creation.

C. YANG-MILLS

The concept of a "covariant derivative" allows the straightforward generalization of electromagnetism to a self-interacting theory, once U(1) has been generalized to a nonabelian group. Yang-Mills theory is an essential part of the Standard Model.

1. Nonabelian

The group U(1) of electromagnetism is Abelian: Group elements commute, which makes group multiplication equivalent to multiplication of real numbers, or addition if we write $U = e^{iG}$. The linearity of this addition is directly related to the linearity of the field equations for electromagnetism without matter. On the other hand, the nonlinearity of nonabelian groups causes the corresponding particles to interact with themselves: Photons are neutral, but "gluons" have charge and "gravitons" have weight.

In coupling electromagnetism to the particle, the relation of the canonical momentum to the velocity is modified: Classically, the covariant momentum is $dx/d\tau = p + qA$ for a particle of charge q (e.g., $q = 1$ for the proton). Quantum mechanically, the net effect is that the wave equation is modified by the replacement

$$\partial \rightarrow \nabla = \partial + iqA$$

which accounts for all dependence on A ("minimal coupling"). This "covariant derivative" has a fundamental role in the formulation of gauge theories, including gravity. Its main purpose is to preserve gauge invariance of the action that gives the wave equation, which would otherwise be spoiled by derivatives acting on the coordinate-dependent gauge parameters: In electromagnetism,

$$\psi' = e^{iq\lambda}\psi, \quad A' = A - \partial\lambda \quad \Rightarrow \quad (\nabla\psi)' = e^{iq\lambda}(\nabla\psi)$$

or more simply

$$\nabla' = e^{iq\lambda}\nabla e^{-iq\lambda}$$

(More generally, q is some Hermitian matrix when ψ is a reducible representation of U(1).)

Yang-Mills theory then can be obtained as a straightforward generalization of electromagnetism, the only difference being that the gauge transformation, and therefore the covariant derivative, now depends on the generators of some nonabelian group. We begin with the hermitian generators

$$[G_i, G_j] = -if_{ij}{}^k G_k, \qquad G_i{}^\dagger = G_i$$

and exponentiate linear combinations of them to obtain the unitary group elements

$$\mathbf{g} = e^{i\lambda}, \quad \lambda = \lambda^i G_i; \qquad \lambda^{i*} = \lambda^i \quad \Rightarrow \quad \mathbf{g}^\dagger = \mathbf{g}^{-1}$$

We then can define representations of the group (see subsection IB1)

$$\psi' = e^{i\lambda}\psi, \quad \psi^{\dagger\prime} = \psi^\dagger e^{-i\lambda}; \qquad (G_i\psi)_A = (G_i)_A{}^B\psi_B$$

For compact groups charge is quantized: For example, for SU(2) the spin (or, for internal symmetry, "isospin") is integral or half-integral. On the other hand, with Abelian groups the charge can take continuous values: For example, in principle the proton might decay into a particle of charge π and another of charge $1-\pi$. The experimental fact that charge is quantized suggests already semiclassically that all interactions should be descibed by (semi)simple groups.

If λ is coordinate dependent (a local, or "gauge" transformation), the ordinary partial derivative spoils gauge covariance, so we introduce the covariant derivative

$$\nabla_a = \partial_a + iA_a, \qquad A_a = A_a{}^i G_i$$

Thus, the covariant derivative acts on matter in a way similar to the infinitesimal gauge transformation,

$$\delta\psi_A = i\lambda^i G_{iA}{}^B\psi_B, \qquad \nabla_a\psi_A = \partial_a\psi_A + iA_a{}^i G_{iA}{}^B\psi_B$$

Gauge covariance is preserved by demanding it have a covariant transformation law

$$\nabla' = e^{i\lambda}\nabla e^{-i\lambda} \quad \Rightarrow \quad \delta A = -[\nabla,\lambda] = -\partial\lambda - i[A,\lambda]$$

The gauge covariance of the field strength follows from defining it in a manifestly covariant way:

$$[\nabla_a,\nabla_b] = iF_{ab} \quad \Rightarrow \quad F' = e^{i\lambda}Fe^{-i\lambda}, \quad F_{ab} = F_{ab}{}^i G_i = \partial_{[a}A_{b]} + i[A_a,A_b]$$

$$\Rightarrow \quad F_{ab}{}^i = \partial_{[a}A_{b]}{}^i + A_a{}^j A_b{}^k f_{jk}{}^i$$

The Jacobi identity for the covariant derivative is the Bianchi identity for the field strength:

$$0 = [\nabla_{[a},[\nabla_b,\nabla_{c]}]] = i[\nabla_{[a},F_{bc]}]$$

(If we choose instead to use antihermitian generators, all the explicit i's go away; however, with hermitian generators the i's will cancel with those from the derivatives when we Fourier transform for purposes of quantization.) Since the adjoint representation can be treated as either matrices or vectors (see subsection IB2), the covariant derivative on it can be written as either a commutator or multiplication: For example, we may write either $[\nabla, F]$ or ∇F, depending on the context.

Actions then can be constructed in a manifestly covariant way: For matter, we take a Lagrangian $L_{M,0}(\partial, \psi)$ that is invariant under global (constant) group transformations, and couple to Yang-Mills as

$$L_{M,0}(\partial, \psi) \to L_{M,A} = L_{M,0}(\nabla, \psi)$$

(This is the analog of minimal coupling in electrodynamics.) The representation we use for G_i in $\nabla_a = \partial_a + iA_a^i G_i$ is determined by how ψ represents the group. (For an Abelian group factor U(1), G is just the charge q, in multiples of the g for that factor.) For example, the Lagrangian for a massless scalar is simply

$$L_0 = \tfrac{1}{2}(\nabla^a \phi)^\dagger (\nabla_a \phi)$$

(normalized for a complex representation).

For the part of the action describing Yang-Mills itself we take (in analogy to the U(1) case)

$$L_A(A_a^i) = \tfrac{1}{8g_A^2} F^{iab} F_{ab}^j \eta_{ij}$$

where η_{ij} is the Cartan metric (see subsection IB2). This way of writing the action is independent of our choice of normalization of the structure constants, and so gives one unambiguous definition for the normalization of the coupling constant g. (It is invariant under any simultaneous redefinition of the fields and the generators that leaves the covariant derivative invariant.) Generally, for simple groups we can choose to (ortho)normalize the generators G_i with the condition (see subsection IB2)

$$\eta_{ij} = c_A \delta_{ij}$$

for some constant c_A; for groups that are products of simple groups (semisimple), we might choose different normalization factors (but, of course, also different g's) for each simple group. For Abelian groups (U(1) factors) $\eta_{ij} = 0$, but then the gauge field has no self-interactions, so the normalization of the coupling constant is defined only by matter terms in the action, and we can replace η_{ij} with δ_{ij} in the above.

Usually it will prove more convenient to use matrix notation: Choosing some convenient representation R of G_i (not necessarily the adjoint), we write

$$L_A(A_a^i) = \tfrac{1}{8g_R^2} tr_R\ F^{ab} F_{ab}$$

The normalization of the trace is determined by R, and thus so is the normalization convention for the coupling constant; a change in the representation used in the action can also be absorbed by a redefinition of the coupling. For example, comparing the defining and adjoint representations of SU(N) (see subsection IB2),

$$L_A = \tfrac{1}{8g_D^2} tr_D\ F^{ab}F_{ab} = \tfrac{1}{8g_A^2} tr_A\ F^{ab}F_{ab} \quad \Rightarrow \quad g_A^2 = 2Ng_D^2$$

In general, we specify our normalization of the structure constants by fixing c_R for some R, and our normalization of the coupling constant by specifying the choice of representation used in the trace (or use explicit adjoint indices). As a rule, we find the most convenient choices of normalization are

$$c_D = 1, \qquad g = g_D$$

(see subsection IB5).

Exercise IIIC1.1

Write the action for SU(N) Yang-Mills coupled to a massless (2-component) spinor in the defining representation. Make all (internal and Lorentz) indices explicit (no "*tr*", etc.), and use defining (N-component) indices on the Yang-Mills field.

We have chosen a normalization where the Yang-Mills coupling constant g appears only as an overall factor multiplying the F^2 term (and similarly for the electromagnetic coupling, as discussed in previous chapters). An alternative is to rescale $A \to gA$ and $F \to gF$ everywhere; then $\nabla = \partial + igA$ and $F = \partial A + ig[A, A]$, and the F^2 term has no extra factor. This allows the Yang-Mills coupling to be treated similarly to other couplings, which are usually not written multiplying kinetic terms (unless analogies to Yang-Mills are being drawn), since (almost) only for Yang-Mills is there a nonlinear symmetry relating kinetic and interaction terms.

Current conservation works a bit differently in the nonabelian case: Applying the same argument as in subsection IIIB4, but taking into account the modified (infinitesimal) gauge transformation law, we find

$$J^m = \frac{\delta S_M}{\delta A_m}, \qquad \nabla_m J^m = 0$$

Since $\partial_m J^m \neq 0$, there is no corresponding covariant conserved charge.

Exercise IIIC1.2

Let's look at the field equations:

a Using properties of the trace, show the entire covariant derivative can be integrated by parts as

$$\int dx\ tr(\mathcal{A}[\nabla, \mathcal{B}]) = -\int dx\ tr([\nabla, \mathcal{A}]\mathcal{B}), \qquad \int dx\ \psi^\dagger \nabla \chi = -\int dx\ (\nabla\psi)^\dagger \chi$$

for matrices $\mathcal{A}, \mathcal{B}$ and column vectors ψ, χ.

b Show

$$\delta F_{ab} = \nabla_{[a}\delta A_{b]}$$

c Using the definition of the current as for electromagnetism (subsection IIIB4), derive the field equations with arbitrary matter,

$$\tfrac{1}{g^2}\tfrac{1}{2}\nabla^b F_{ba} = J_a$$

d Show that gauge invariance of the action S_A implies

$$\nabla^a(\nabla^b F_{ba}) = 0$$

Also show this is true directly, using the Jacobi identity, but not the field equations. (Hint: Write the covariant derivatives as commutators.)

Exercise IIIC1.3

Expand the left-hand side of the field equation (given in excercise IIIC1.2c) in the field, as

$$\tfrac{1}{g^2}\tfrac{1}{2}\nabla^b F_{ba} = \tfrac{1}{g^2}\tfrac{1}{2}\partial^b\partial_{[b}A_{a]} - j_a$$

where j contains the quadratic and higher-order terms. Show the *noncovariant* current

$$\mathcal{J}_a = J_a + j_a$$

is conserved. The j term can be considered the gluon contribution to the current: Unlike photons, gluons are charged. Although the current is gauge dependent, and thus physically meaningless, the corresponding charge can be gauge independent under situations where the boundary conditions are suitable.

2. Lightcone

Since gauge parameters are always of the same form as the gauge field, but with one less vector index, an obvious type of gauge choice (at least from the point of view of counting components) is to require the gauge field to vanish when one vector index is fixed to a certain value. Explicitly, in terms of the covariant derivative we set

$$n\cdot\nabla = n\cdot\partial \quad \Rightarrow \quad n\cdot A = 0$$

for some constant vector n^a. We then can distinguish three types of "axial gauges":

(1) "Arnowitt-Fickler", or spacelike ($n^2 > 0$),

(2) "lightcone", or lightlike ($n^2 = 0$), and

(3) "temporal", or timelike ($n^2 < 0$).

By appropriate choice of reference frame, and with the usual notation, we can write these gauge conditions as $\nabla^1 = \partial^1$, $\nabla^+ = \partial^+$, and $\nabla^0 = \partial^0$.

One way to apply this gauge in the action is to keep the same set of fields, but have explicit n dependence. A much simpler choice is to use a gauge choice such as $A_0 = 0$ simply to eliminate A_0 explicitly from the action. For example, for Yang-Mills we find

$$A_0 = 0 \quad \Rightarrow \quad F_{0i} = \dot{A}_i \quad \Rightarrow \quad \tfrac{1}{8}(F_{ab})^2 = -\tfrac{1}{4}(\dot{A}_i)^2 + \tfrac{1}{8}(F_{ij})^2$$

where " $\dot{}$ " here refers to the time derivative. Canonical quantization is simple in this gauge, because we have the canonical time-derivative term. However, the gauge condition can't be imposed everywhere, as seen for the corresponding gauge for the one-dimensional metric in subsection IIIB2, and in our general discussion in subsection IIIA5: Here we can generalize the time-ordered integral for the temporal gauge to an integral *path-ordered* with respect to a straight-line path in the n direction:

$$e^{-kn\cdot\nabla(x)} = e^{-i\Lambda(x,x-kn)}e^{-kn\cdot\partial}, \qquad e^{-i\Lambda(x,x-kn)} = \mathcal{P}\left[exp\left(-i\int_{x-kn}^{x} dx'\cdot A(x')\right)\right]$$

Applying this gauge transformation to $n\cdot\nabla$, as in subsection IIIA5, fixes $n\cdot A$ to a constant with respect to $n\cdot\partial$; the effect on all of ∇ is:

$$\nabla'(x) = e^{i\Lambda(x,x-kn)}\nabla(x)e^{-i\Lambda(x,x-kn)} \quad \Leftrightarrow \quad \nabla'(x+kn) = e^{kn\cdot\nabla(x)}\nabla(x)e^{-kn\cdot\nabla(x)}$$

For example, for the temporal gauge, if we choose "x" to be on the initial hypersurface $x^0 = t_0$, then we can choose $k = t - t_0$ so that ∇' is evaluated at arbitrary time t:

$$\nabla'_a(x) = [(e^{k\nabla_0(x)}\nabla_a(x)e^{-k\nabla_0(x)})|_{x^0=t_0}]|_{k=x^0}$$

By Taylor expanding in k, this gives an explicit expression for A_a at all times in terms of A_a, and F_{ab} and its covariant time-derivatives, evaluated at some initial time, but with simply $A_0(t,x^i) = A_0(t_0,x^i)$. Thus, we still need to impose the A_0 field equation $[\nabla_i, \dot{A}_i] = 0$ as a constraint at some initial time.

Exercise IIIC2.1

First set $A_0 = 0$, then derive the field equations for A_i from the Yang-Mills action. Compare the results of exercise IIIC1.2c for $J = 0$. Show explicitly that these field equations imply the *time derivative* of the constraint $[\nabla_i, \dot{A}_i] = 0$. (Hint: Write everything in terms of F's and ∇'s till the end.)

In the case of the lightcone gauge we can carry this analysis one step further. In subsection IIB3 we saw that lightcone formalisms are described by massless fields with (D−2)-dimensional ("transverse") indices. In the present analysis, gauge fixing alone gives us, again for the example of pure Yang-Mills,

$$A^+ = 0 \quad \Rightarrow \quad F^{+i} = \partial^+ A^i, \quad F^{+-} = \partial^+ A^-, \quad F^{-i} = \partial^- A^i - [\nabla^i, A^-]$$

$$\Rightarrow \quad \tfrac{1}{8}(F^{ab})^2 = -\tfrac{1}{4}(\partial^+ A^-)^2 - \tfrac{1}{2}(\partial^+ A^i)(\partial^- A^i - [\nabla^i, A^-]) + \tfrac{1}{8}(F^{ij})^2$$

In the lightcone formalism ∂^- $(-\partial_+)$ is to be treated as a time derivative, while ∂^+ can be freely inverted (i.e., modes propagate to infinity in the x^+ direction, but boundary conditions set them to vanish in the x^- direction). Thus, we can treat A^- as an auxiliary field. The solution to its field equation is

$$A^- = \frac{1}{\partial^{+2}}[\nabla^i, \partial^+ A^i]$$

which can be substituted directly into the action:

$$\begin{aligned}
\tfrac{1}{8}(F^{ab})^2 &= \tfrac{1}{2}A^i \partial^+ \partial^- A^i + \tfrac{1}{8}(F^{ij})^2 - \tfrac{1}{4}[\nabla^i, \partial^+ A^i]\frac{1}{\partial^{+2}}[\nabla^j, \partial^+ A^j] \\
&= -\tfrac{1}{4}A^i \Box A^i + i\tfrac{1}{2}[A^i, A^j]\partial^i A^j + i\tfrac{1}{2}(\partial^i A^i)\frac{1}{\partial^+}[A^j, \partial^+ A^j] \\
&\quad - \tfrac{1}{8}[A^i, A^j]^2 + \tfrac{1}{4}[A^i, \partial^+ A^i]\frac{1}{\partial^{+2}}[A^j, \partial^+ A^j]
\end{aligned}$$

We can save a couple of steps in this derivation by noting that elimination of any auxiliary field, appearing quadratically (as in going from Hamiltonian to Lagrangian formalisms), has the effect

$$L = \tfrac{1}{2}ax^2 + bx + c \rightarrow -\tfrac{1}{2}ax^2|_{\partial L/\partial x = 0} + L|_{x=0}$$

In this case, the quadratic term is $(F^{-+})^2$, and we have

$$\tfrac{1}{8}(F^{ab})^2 = \tfrac{1}{8}(F^{ij})^2 - \tfrac{1}{2}F^{+i}F^{-i} - \tfrac{1}{4}(F^{+-})^2 \rightarrow \tfrac{1}{8}(F^{ij})^2 - \tfrac{1}{2}(\partial^+ A^i)(\partial^- A^i) + \tfrac{1}{4}(F^{+-})^2$$

where the last term is evaluated at

$$0 = [\nabla_a, F^{+a}] = -\partial^+ F^{+-} + [\nabla^i, F^{+i}] \quad \Rightarrow \quad F^{+-} = \frac{1}{\partial^+}[\nabla^i, F^{+i}]$$

$$\Rightarrow \quad L = \tfrac{1}{8}(F^{ij})^2 + \tfrac{1}{2}A^i \partial^+ \partial^- A^i - \tfrac{1}{4}[\nabla^i, \partial^+ A^i]\frac{1}{\partial^{+2}}[\nabla^j, \partial^+ A^j]$$

as above.

In this case, canonical quantization is even simpler, since interpreting ∂^- as the time derivative makes the action look like that for a nonrelativistic field theory, with a kinetic term linear in time derivatives (as well as interactions without them). The free part of the field equation is also simpler, since the kinetic operator is now just $\square$. (This is true in general in lightcone formalisms from the analysis of free theories in chapter XII.) In general, lightcone gauges are the simplest for analyzing physical degrees of freedom (within perturbation theory), since the maximum number of degrees of freedom is eliminated, and thus kinetic operators look like those of scalars. On the other hand, interaction terms are more complicated because of the nonlocal Coulomb-like terms involving $1/\partial^+$: The inverse of a derivative is an integral. (However, in practice we often work in momentum space, where $1/p^+$ is local, but Fourier transformation itself introduces multiple integrals.) This makes lightcone gauges useful for discussing unitarity (they are "unitary gauges"), but inconvenient for explicit calculations. However, in subsection VIB6 we'll find a slight modification of the lightcone that makes it the most convenient method for certain calculations. (In the literature, "lightcone gauge" is sometimes used to refer to an axial gauge where A^+ is set to vanish but A^- is not eliminated, and D-vector notation is still used, so unitarity is not manifest. Here we always eliminate both components and explicitly use $(D-2)$-vectors, which has distinct technical advantages.)

Although spin 1/2 has no gauge invariance, the second step of the lightcone formalism, eliminating auxiliary fields, can also be applied there: For example, for a massless spinor in D=4, identifying $\partial^{\ominus\dot\ominus} = \partial^-$ as the lightcone "time" derivative, we vary $\bar\psi^{\dot\ominus}$ (or $\psi^\ominus$) as the auxiliary field:

$$-iL = \bar\psi^{\dot\ominus}\partial^{\oplus\dot\oplus}\psi^\ominus + \bar\psi^{\dot\oplus}\partial^{\ominus\dot\ominus}\psi^\oplus - \bar\psi^{\dot\ominus}\partial^{\ominus\dot\oplus}\psi^\oplus - \bar\psi^{\dot\oplus}\partial^{\oplus\dot\ominus}\psi^\ominus$$

$$\Rightarrow \quad \psi^\ominus = \frac{1}{\partial^{\oplus\dot\oplus}}\partial^{\ominus\dot\oplus}\psi^\oplus$$

$$\Rightarrow \quad L = \bar\psi^{\dot\oplus}\frac{\frac{1}{2}\square}{i\partial^{\oplus\dot\oplus}}\psi^\oplus$$

This tells us that a 4D massless spinor, like a 4D massless vector (or a *complex* scalar) has only 1 complex (2 real) degree of freedom, describing a particle of helicity $+1/2$ and its antiparticle of helicity $-1/2$ (± 1 for the vector, 0 for the scalar), in agreement with our general discussion of helicity in subsection IIB7. On the other hand, in the massive case we can always go to a rest frame, so the analysis is in terms of spin (SU(2) for D=4) rather than helicity. For a massive Weyl spinor we can perform the same analysis as above, with the modifications

$$L \to L + \frac{im}{\sqrt{2}}(\psi^\oplus\psi^\ominus + \bar\psi^{\dot\oplus}\bar\psi^{\dot\ominus}) \quad \Rightarrow \quad L = \bar\psi^{\dot\oplus}\frac{\frac{1}{2}(\square - m^2)}{i\partial^{\oplus\dot\oplus}}\psi^\oplus$$

where we have dropped some terms that vanish upon using integration by parts and the antisymmetry of the fermions. So now we have the two states of an SU(2) spinor, but these are identified with their antiparticles. This differs from the vector: While for the spinor we have 2 states of a given energy for both the massless and massive cases, for a vector we have 2 for the massless but 3 for the massive, since for SU(2) spin s has 2s+1 states:

4D states of given E:

spin	0	$\frac{1}{2}$	1	$\frac{3}{2}$	...
$m = 0$	1	2	2	2	...
$m > 0$	1	2	3	4	...

Exercise IIIC2.2

Show that integration by parts for $1/\partial$ gives just a sign change, just as for ∂.

In general dimensions, massless particles are representations of the "little group" SO(D−2) (the helicity SO(2) in D=4), as described in subsection IIB3. Massive particles represent the little group SO(D−1), corresponding to dimensional reduction from an extra dimension, as described in subsection IIB4.

3. Plane waves

The simplest nontrivial solutions to nonabelian field equations are the generalizations of the plane wave solutions of the free theory. We begin with general, free, massless theories, as analyzed in subsection IIB3. In the lightcone frame only p^+ is nonvanishing. In position space this means the field strength depends only on x^-. This describes a wave traveling at the speed of light in the positive x^1 direction, with no other spatial dependence (i.e., a plane wave). We allow arbitrary dependence on x^-, corresponding to a superposition of waves with parallel momenta (but different values of p^+). While its dependence on only x^- solves the Klein-Gordon equation, Maxwell's equations are solved by giving the field strength as many upper + indices as possible, and no upper −'s.

Generalizing to interactions, we notice that the Yang-Mills field equations and Bianchi identities differ from Maxwell's equations only by the covariantization of the derivatives (at least for pure Yang-Mills). Because Maxwell's equations were satisfied by just restricting the index structure, we can do the same for the covariant derivatives by assuming that only ∇^+ is novanishing on the field strengths. In other words, we can solve the field equations and Bianchi identities by choosing the only nontrivial components of the gauge fields to be those in ∇^+.

The final step is to solve the relation between covariant derivative and field strength. This is simple because the index structure we found implies the only non-trvial commutators are

$$[\partial^i, \nabla^+] = iF^{i+}, \qquad [\partial^-, \nabla^+] = 0$$

In particular, this implies that the gauge fields have no x^+ dependence, and only a very simple dependence on x^i. We find directly

$$A^+ = x^i F^{i+}(x^-)$$

where $F^{i+}(x^-)$ is unrestricted (other than the explicit index structure and coordinate dependence). Of course, this result can also be used in the free theory, although it differs from the usual lightcone gauge.

Exercise IIIC3.1

Gauge transform this solution to the lightcone gauge $A^+ = 0$ in the Abelian case.

Exercise IIIC3.2

Translate the above results into spinor notation in D=4.

4. Self-duality

The simplest and most important solutions to the field equations are those that are invariant under the "duality" symmetry that relates electric and magnetic charge:

$$[\nabla_a, \nabla_b] = \pm\tfrac{1}{2}\epsilon_{abcd}[\nabla^c, \nabla^d]$$

Applying the self-duality condition twice, we find

$$\epsilon_{abef}\epsilon^{efcd} = +\delta_{[a}^c\delta_{b]}^d$$

which requires an even number of time dimensions. For example, since the action is usually Wick rotated anyway for perturbative purposes, we might assume that we should do the same for classical solutions that are not considered as "small" fluctations about the usual vacuum. (Such a Euclidean definition of field theory has been considered for a mathematically rigorous formalism, called "constructive quantum field theory", since the Gaussian path integrals for scalars and vectors are then well-defined and convergent. However, other spins, such as for fermions or gravity, are a problem in this approach.) Alternatively, we can replace ϵ_{abcd} with

$i\epsilon_{abcd}$ and complexify our fields. The self-duality condition, when combined with the Bianchi identities, implies the field equations: For Yang-Mills,

$$\nabla_{[a}F_{bc]} = 0 \quad \Rightarrow \quad 0 = \pm\tfrac{1}{2}\epsilon^{abcd}\nabla_a F_{bc} = \tfrac{1}{4}\epsilon^{abcd}\nabla_a \epsilon_{bcef}F^{ef} = \nabla_a F^{ad}$$

Since the self-duality condition is only first-order in derivatives, it's easier to solve than the usual field equations.

Plane wave solutions provide a simple example of self-duality, since the field strengths can easily be written as the sum of self-dual and anti-self-dual parts: In Minkowski space we define the self-dual part as helicity $+1$ ($f_{\dot\alpha\dot\beta}$), and anti-self-dual as -1 ($f_{\alpha\beta}$). For example, for a wave traveling in the "1" direction, the $F^{+2} \mp iF^{+3}$ components give the two self-dualities for Yang-Mills, describing helicities ± 1 (the two circular polarizations).

Exercise IIIC4.1

Generalize the results of the previous subsection to more general waves, with an A^+ which is a general function of x^- and x^i (with the other components of A still vanishing).

a Find the field strength, and show it satisfies the interacting field equations if A^+ satisfies the free Laplace equation

$$(\partial^i)^2 A^+ = 0$$

b In D=4, the solution to this equation is

$$A^+ = f(x^-, x^t) + \bar{f}(x^-, \bar{x}^t)$$

Show this decomposition describes the two separate helicities.

Before further analyzing solutions to the self-duality condition, we consider actions that use self-dual fields directly. This will allow us to describe not only theories whose only solutions are self-dual, but also more standard theories as perturbations about self-duality, and even massive theories. The most unusual feature of this approach is that complex fields are used without their complex conjugates, since this is implied in D=3+1 by self-duality. (Alternatively, we can Wick rotate to 2+2 dimensions, where all Lorentz representations are real.) There are two stages to this approach: (1) Use a first-order formalism where the auxiliary field is self-dual. The usual first-order actions for spin 1/2 (Weyl or Dirac) already can be interpreted in this way, where "self-duality" means "chirality". (2) For the massive theory, eliminate the non-self-dual field (as an auxiliary field, as allowed by the mass term), so that the dynamics is described by the self-dual field, which was formerly considered as auxiliary. The massless theory then can be treated as a limiting case.

The simplest (and perhaps most useful) example is massive spin 1/2 coupled in a real representation to Yang-Mills fields:

$$L = \psi^{T\alpha} i\nabla_\alpha{}^{\dot\alpha}\bar\psi_{\dot\alpha} + \tfrac{1}{2\sqrt{2}}m(\psi^{T\alpha}\psi_\alpha + \bar\psi^{T\dot\alpha}\bar\psi_{\dot\alpha})$$

where the transposition ("^{T}") refers to the Yang-Mills group index (with respect to which the spinors are column vectors). Note that ψ must be a real representation of this group ($A^T = -A$) for the mass term to be gauge invariant (unless the mass term includes scalars: see the following chapter). Even though ψ and $\bar\psi$ are complex conjugates, they can be treated independently as far as field equations are concerned, since they are just different linear combinations of their real and imaginary parts. (Complex conjugation can be treated as just a symmetry, related to unitarity.) Noticing that the quadratic term for $\bar\psi$ has no derivatives, we can treat it as an auxiliary field, and integrate it out (i.e., eliminate it by its equation of motion, which gives an explicit local expression for it):

$$L \to -\tfrac{\sqrt{2}}{m}[\tfrac{1}{4}\psi^{T\alpha}(\Box - m^2)\psi_\alpha + \tfrac{1}{2}\psi^{T\alpha} i f_\alpha{}^\beta\psi_\beta]$$

where we have used the identity

$$\nabla_\alpha{}^{\dot\gamma}\nabla^\beta{}_{\dot\gamma} = \tfrac{1}{2}\{\nabla_\alpha{}^{\dot\gamma}, \nabla^\beta{}_{\dot\gamma}\} + \tfrac{1}{2}[\nabla_\alpha{}^{\dot\gamma}, \nabla^\beta{}_{\dot\gamma}] = -\tfrac{1}{2}\delta_\alpha^\beta\Box - i f_\alpha{}^\beta$$

whose simplicity followed from ψ being a real representation of the Yang-Mills group. (Of course, we could have eliminated ψ instead, but not both.) For convenience we also scale ψ by a constant

$$\psi \to 2^{-1/4}\sqrt{m}\psi$$

to find the final result

$$L \to -\tfrac{1}{4}\psi^{T\alpha}(\Box - m^2)\psi_\alpha - \tfrac{1}{2}\psi^{T\alpha} i f_\alpha{}^\beta\psi_\beta$$

Now the massless limit can be taken easily. This action resembles that of a scalar, plus a "magnetic-moment coupling", which couples the "(anti-)self-dual" (chiral) spinor ψ_α to only the (anti-)self-dual part $f_{\alpha\beta}$ of the Yang-Mills field strength.

For the same reason, the kinetic operator can be written in terms of just the self-dual part $S_{\alpha\beta}$ of the spin operator:

$$L = -\tfrac{1}{4}\psi^T(\Box - m^2 - i f^{\alpha\beta}S_{\beta\alpha})\psi$$

This operator is of the same form found by squaring the Dirac operator:

$$-2\nabla\!\!\!/^2 = -2(\gamma \cdot \nabla)^2 = -(\{\gamma^a, \gamma^b\} + [\gamma^a, \gamma^b])\nabla_a \nabla_b = \Box - iF^{ab}S_{ba}$$

except for the self-duality. The simple form of this result again depends on the reality (parity invariance) of the Yang-Mills representation; although this squaring trick can be applied for complex representations (parity violating), the coupling does not simplify. This is related to the fact that real representations are required for our derivation of the self-dual form.

In the special case where the real representation is the direct sum of a complex one $\psi_{+\alpha}$ with its complex conjugate $\psi_{-\alpha}$ (as for quarks in the Standard Model, or electrons in electrodynamics), we can rewrite the Lagrangian as

$$L_c = -\tfrac{1}{2}\psi_+^{T\alpha}(\Box - m^2)\psi_{-\alpha} - \psi_+^{T\alpha} i f_\alpha{}^\beta \psi_{-\beta}$$

The method can also be generalized to the case of scalar couplings, but the action becomes nonpolynomial.

For spin 1, we start with the massless case. We can write the Lagrangian for Yang-Mills as

$$L = tr(G^{\alpha\beta} f_{\alpha\beta} - \tfrac{1}{2}g^2 G^2_{\alpha\beta})$$

where $G_{\alpha\beta}$ is a (anti-)self-dual auxiliary field. Although this action is complex, eliminating G by its algebraic field equation gives the usual Yang-Mills action up to a total derivative term ($\epsilon^{abcd}F_{ab}F_{cd}$), which can be dropped for purposes of perturbation theory. For $g = 0$, this is an action where G acts as a Lagrange multiplier, enforcing the self-duality of the Yang-Mills field strength.

If we simply add a mass term

$$L_m = \tfrac{1}{4}(\tfrac{m}{g})^2 A^2$$

then A can be eliminated by its field equation, giving a nonpolynomial action of the form

$$L + L_m \to -\tfrac{1}{2}(\partial G)[(\tfrac{m}{g})^2 + G]^{-1}(\partial G) - \tfrac{1}{2}g^2 G^2$$

Just as the spin-1/2 action contained only a 2-component spinor describing the 2 polarizations of spin 1/2, this action contains only the 3-component $G_{\alpha\beta}$, describing the 3 polarizations of (massive) spin 1.

Exercise IIIC4.2

Find the Abelian part of this action. Show the free field equation is $(\Box - m^2)G_{\alpha\beta} = 0$ (without gauge fixing).

5. Twistors

In four dimensions with an even number of time dimensions, the "Lorentz" group factorizes (into SU(2)^2 for D=4+0 and SL(2)^2 for D=2+2). This makes self-duality especially simple in spinor notation: For Yang-Mills (cf. electromagnetism in subsection IIA7),

$$[\nabla^{\alpha\beta'}, \nabla^{\gamma\delta'}] = iC^{\alpha\gamma} f^{\beta'\delta'} \qquad (f^{\alpha\beta} = 0)$$

where we have written primes instead of dots to emphasize that the two kinds of indices transform independently (instead of as complex conjugates, as in D=3+1). For purposes of analyzing self-duality within perturbation theory, we can use a lightcone method that breaks only one of the two SL(2)'s (or SU(2)'s), by separating out its indices into the $\oplus$ and $\ominus$ components:

$$[\nabla^{\oplus\alpha'}, \nabla^{\oplus\beta'}] = 0 \quad \Rightarrow \quad \nabla^{\oplus\alpha'} = \partial^{\oplus\alpha'}$$

where we have chosen a lightcone gauge: The vanishing of all field strengths for the covariant derivative $\nabla^{\oplus\alpha'}$ says that it is pure gauge (as seen by ignoring all but the $x^{\ominus\alpha'}$ coordinates). We now solve

$$[\nabla^{[\oplus\alpha'}, \nabla^{\ominus]\beta'}] = 0 \quad \Rightarrow \quad \nabla^{\ominus\alpha'} = \partial^{\ominus\alpha'} + i\partial^{\oplus\alpha'}\phi$$

i.e., $\nabla^{\ominus\alpha'} - \partial^{\ominus\alpha'}$ has vanishing curl, and is therefore a gradient. We therefore have

$$A^{\oplus\alpha'} = 0, \quad A^{\ominus\alpha'} = \partial^{\oplus\alpha'}\phi; \qquad f^{\alpha'\beta'} = -i\partial^{\oplus\alpha'}\partial^{\oplus\beta'}\phi$$

These can also be written in terms of an arbitrary constant twistor ϵ^α ($= \delta_\ominus^\alpha$ above) as

$$A^{\alpha\beta'} = \partial^{\gamma\beta'}(-i\epsilon^\alpha\epsilon_\gamma\phi), \qquad f^{\alpha'\beta'} = \partial^{\gamma\alpha'}\partial^{\delta\beta'}(i\epsilon_\gamma\epsilon_\delta\phi)$$

The final self-duality condition $[\nabla^{\ominus\alpha'}, \nabla^{\ominus\beta'}] = 0$ then gives the equation of motion

$$\tfrac{1}{2}\Box\phi + (\partial^{\oplus\alpha'}\phi)(\partial^{\oplus}{}_{\alpha'}\phi) = 0$$

Exercise IIIC5.1

Show that the sign convention for Wick rotation of the Levi-Civita tensor consistent with the above equations is

$$F_{ab} = \tfrac{1}{2}\epsilon_{abcd}F^{cd}, \qquad F_{\alpha\alpha'\beta\beta'} = C_{\alpha\beta}f_{\alpha'\beta'} \quad \Rightarrow$$

$$\epsilon_{\alpha\alpha'\beta\beta'\gamma\gamma'\delta\delta'} = C_{\alpha\beta}C_{\gamma\delta}C_{\alpha'\delta'}C_{\beta'\gamma'} - C_{\alpha\delta}C_{\beta\gamma}C_{\alpha'\beta'}C_{\gamma'\delta'}$$

Exercise IIIC5.2

Look at the action $G^{\alpha\beta}f_{\alpha\beta}$ for self-dual Yang-Mills in the lightcone gauge, using the results above. Show that this action is equivalent to the lightcone action for ordinary Yang-Mills (subsection IIIC2), with some terms in the interaction dropped.

At least for 4D Yang-Mills, advantage can be taken of the conformal invariance of the classical interacting theory by using a formalism where this invariance is manifest. We saw in subsection IA6 that classical mechanics could be made manifestly conformal by use of extra coordinates. Covariant derivatives can be defined in terms of projective lightcone coordinates, but the twistor coordinates $z^{\mathcal{A}\alpha}$ (see subsections IIB6 and IIIB1) are more useful. The self-dual covariant derivatives then satisfy

$$[\nabla_{\mathcal{A}\alpha}, \nabla_{\mathcal{B}\beta}] = iC_{\alpha\beta}f_{\mathcal{A}\mathcal{B}}$$

in direct analogy to 4D spinor notation. This equation also can be solved by the lightcone method used above, but now this method breaks only the internal SL(2) symmetry, leaving SL(4) conformal symmetry manifest. More general self-dual field strengths in this twistor space are also of the form $f_{\mathcal{A}...\mathcal{B}}$, totally symmetric in the indices. We also need to impose the constraint on the field strength

$$z^{\mathcal{A}\alpha}f_{\mathcal{A}\mathcal{B}} = 0$$

(and similarly for the more general case) to restrict the range of indices to the usual 4D spinor indices (in which the field strengths are totally symmetric). Self-duality implies the Bianchi identity

$$\nabla_{[\mathcal{A}\alpha}f_{\mathcal{B}]\mathcal{C}} = 0$$

which also generalizes to the other field strengths, and is the equivalent of the usual first-order differential equations (Dirac, Maxwell, etc.) satisfied by 4D field strengths. As usual, it in turn implies the interacting Klein-Gordon equation, which in the Yang-Mills case is

$$\tfrac{1}{2}\nabla_{[\mathcal{A}}{}^{\alpha}\nabla_{\mathcal{B}]\alpha}f_{\mathcal{C}\mathcal{D}} = -i[f_{\mathcal{C}[\mathcal{A}}, f_{\mathcal{B}]\mathcal{D}}]$$

The Bianchi together with the z index constraint imply the constraint on coordinate dependence

$$(z^{\mathcal{A}\alpha}\nabla_{\mathcal{A}\beta} + \delta_{\beta}^{\alpha})f_{\mathcal{B}\mathcal{C}} = 0$$

which eliminates dependence on all but the usual 4D coordinates. These four equations are generically satisfied by self-dual field strengths. The self-duality itself of the field strengths is a consequence of their total symmetry in their indices, and the fact that they are all lower (SL(4)) indices. (The z index constraint then reduces them to SL(2) Weyl indices all of the same chirality.)

Exercise IIIC5.3

Derive the last three equations from the previous two (self-duality and zf=0).

Exercise IIIC5.4

Show that non-self-dual Yang-Mills is conformally invariant in D=4 by extending the (4+2)-dimensional formalism of subsections IA6 and IIIB1 (especially exercise IIIB1.4): Show the field strength

$$F_{ABC} = -i\tfrac{1}{2}y_{[A}[\nabla_B, \nabla_{C]}]$$

satisfies the gauge covariances

$$\delta A_A = -[\nabla_A, \lambda] - y_A\hat{\lambda}$$

and Bianchi identities

$$y_{[A}F_{BCD]} = \nabla_{[A}F_{BCD]} = 0$$

The duality transformation

$$F_{ABC} \to \tfrac{1}{6}\epsilon_{ABCDEF}F^{DEF}$$

then suggests the field equations

$$y^A F_{ABC} = \nabla^A F_{ABC} = 0$$

in addition to the usual constraint

$$y^2 F_{ABC} = 0$$

By reducing to D=4 coordinates with the aid of the above yF conditions, show F reduces to the usual field strength, and the remaining equations reduce to the usual gauge transformation, Bianchi identity, duality, and field equation.

6. Instantons

Another interesting class of self-dual solutions to Yang-Mills theory are "instantons", so called because the field strength is maximum at points in spacetime, unlike the plane waves, whose wavefronts propagate from and toward timelike infinity. A particular subset of these can be expressed in a very simple form by the 't Hooft ansatz in terms of a scalar field: In twistor notation, choosing the Yang-Mills gauge group GL(2) (in 2+2 dimensions, or SU(2)⊗GL(1) for 4+0),

$$iA_{\mathcal{A}\alpha\iota}{}^{\kappa} = -\delta_\alpha^\kappa \partial_{\mathcal{A}\iota} ln\ \phi \quad \Rightarrow \quad iA_{\mathcal{A}\alpha\iota}{}^{\iota} = -\partial_{\mathcal{A}\alpha} ln\ \phi$$

so the GL(1) piece is pure gauge, and has been included just for convenience. Note that this ansatz ties the SL(2) twistor index with the SL(2) gauge group indices (ι, κ), but in this notation the index that carries the spacetime (conformal) symmetry is free. Imposing the self-duality condition on the field strength, and separating out the terms symmetric and antisymmetric in $\mathcal{AB}$, we find

$$if_{\mathcal{AB}\iota}{}^{\kappa} = -\tfrac{1}{2}\phi \partial_{(\mathcal{A}}{}^{\kappa}\partial_{\mathcal{B})\iota}\phi^{-1}$$

$$\phi^{-1}\partial_{\mathcal{A}}{}^{\alpha}\partial_{\mathcal{B}\alpha}\phi = 0$$

The "field equation" for ϕ is just the twistor version of the (free) Klein-Gordon equation, and its solution is the projective lightcone version of 4D point sources (see subsection IA6): Since for any two 6D lightlike vectors y and y'

$$y = e(x, 1, \tfrac{1}{2}x^2) \quad \Rightarrow \quad y \cdot y' = -\tfrac{1}{2}ee'(x - x')^2$$

we have the solution

$$\phi = \sum_{i=1}^{k+1} (y \cdot y_i)^{-1}, \quad y_i^2 = 0$$

with y given in terms of z as before, and y_i are constant null vectors. "k" is the number of instantons. (The one term for $k = 0$ is pure gauge.) The usual singularities in the Klein-Gordon equation at $y = y_i$ are killed by the extra factor of ϕ^{-1} in the field equation.

Exercise IIIC6.1

Let's check the Klein-Gordon equation for $y \neq y_i$ directly in twistor space. We will need the identity

$$y_i^2 = 0 \quad \Rightarrow \quad y_{i\mathcal{A}[\mathcal{B}}y_{i\mathcal{CD}]} = 0 \qquad \left(no \sum_i \right)$$

in the product

$$y \cdot y_i = \tfrac{1}{2} y^{\mathcal{AB}} y_{i\mathcal{AB}}$$

Prove this identity in two ways:

a Show it follows from the definition

$$y_i^2 = \tfrac{1}{4}\epsilon^{\mathcal{ABCD}} y_{i\mathcal{AB}} y_{i\mathcal{CD}}$$

b Show it follows from plugging in the solution to the lightlike condition,

$$y_{i\mathcal{AB}} = z_{i\mathcal{A}}{}^{\alpha} z_{i\mathcal{B}\alpha}$$

c Now use the identity to show the above solution satisfies its field equation by evaluating the z derivatives.

We can rewrite this in the usual 4D coordinates by transforming from $z^{\mathcal{A}\alpha}$ to $\lambda^{\alpha\mu}$ and $x^{\mu\nu'}$ as $z^{\mathcal{A}}{}_{\alpha} = \lambda_{\alpha}{}^{\nu}(\delta^{\mu}_{\nu}, x_{\nu}{}^{\mu'})$ (see subsection IIB6):

$$dz^{\mathcal{A}\alpha} A_{\mathcal{A}\alpha\iota}{}^{\kappa} = dx^{\mu\nu'} A_{\mu\nu'\iota}{}^{\kappa} + [(d\lambda^{\alpha}{}_{\nu})\lambda^{-1\nu}{}_{\beta}] z^{\mathcal{A}\beta} A_{\mathcal{A}\alpha\iota}{}^{\kappa} = dx^{\mu\nu'} A_{\mu\nu'\iota}{}^{\kappa} - i\lambda^{-1\nu}{}_{\iota} d\lambda^{\kappa}{}_{\nu}$$

where in the first step we have used the expression for z in terms of λ and x, and in the second we used the result that

$$(z^{\mathcal{A}\alpha}\partial_{\mathcal{A}\beta} + \delta^{\alpha}_{\beta})\phi = 0$$

We now recognize that the gauge transformation that gets rid of all but the "x components" of A (whose existence is guaranteed by the condition $z^{\mathcal{A}\alpha} f_{\mathcal{AB}} = 0$) uses λ itself as the gauge parameter:

$$dz^{\mathcal{A}\alpha} A_{\mathcal{A}\alpha\iota}{}^{\kappa} = -i\lambda^{-1\nu}{}_{\iota} d\lambda^{\kappa}{}_{\nu} + \lambda^{-1\nu}{}_{\iota}(dz^{\mathcal{A}\alpha} A'_{\mathcal{A}\alpha\nu}{}^{\mu})\lambda^{\kappa}{}_{\mu}$$

The net result is that A can be reduced to an ordinary 4-dimensional expression by just setting $\lambda = \delta$ in the original expression. Then

$$iA_{\mu\mu'\iota}{}^{\kappa} = -\delta^{\kappa}_{\mu}\partial_{\iota\mu'} ln\ \phi, \qquad \phi = \sum_i \frac{1}{e_i(x - x_i)^2}$$

with y in terms of x and a scale factor (worldline metric) e as in subsections IA6 and IIIB1 (and dropping an overall factor that doesn't contribute to A). Note that, unlike the expression in twistor space, where conformal invariance is manifest, here Lorentz invariance is tied to the Yang-Mills symmetry.

Exercise IIIC6.2

Show in 4D coordinates that the gauge-invariant quantity $tr(f^2)$ is finite at the points $x = x_i$, where A is singular. (This means that the gauge choice is singular, not physical quantities.)

Another important property of instantons is that they give finite contributions to the action. In vector notation, we have

$$F^{ab} = \tfrac{1}{2}\epsilon^{abcd}F_{cd} \quad \Rightarrow \quad S = \tfrac{1}{8g^2}tr\int \frac{d^4x}{(2\pi)^2}\ F^{ab}F_{ab} = \tfrac{1}{16g^2}tr\int \frac{d^4x}{(2\pi)^2}\ \epsilon^{abcd}F_{ab}F_{cd}$$

The last expression can be reduced to a boundary term, since

$$\tfrac{1}{8}tr\ F_{[ab}F_{cd]} = \tfrac{1}{6}\partial_{[a}B_{bcd]}$$

in terms of the "Chern-Simons form"

$$B_{abc} = tr(\tfrac{1}{2}A_{[a}\partial_b A_{c]} + i\tfrac{1}{3}A_{[a}A_b A_{c]})$$

Exercise IIIC6.3

Although the Chern-Simons form is not manifestly invariant, its variation is, up to a total derivative:

a Show that its general variation is

$$\delta B_{abc} = \tfrac{1}{2}\ tr[(\delta A_{[a})F_{bc]} - \partial_{[a}A_b\delta A_{c]}]$$

b Show the gauge transformation of B is

$$\delta B_{abc} = -\tfrac{1}{2}\partial_{[a}\lambda_{bc]}, \qquad \lambda_{ab} = \tfrac{1}{2}tr(\lambda\partial_{[a}A_{b]})$$

If we assume boundary conditions such that F drops off rapidly at infinity, then A must drop off to pure gauge at infinity:

$$iA_m \to \mathbf{g}^{-1}\partial_m\mathbf{g}$$

Since instantons always deal with an SU(2) subgroup of the gauge group, we'll assume now for simplicity that the whole group is itself SU(2). Then the action can be given a group theory interpretation directly, since the integral over the surface at infinity is an integral over the 3-sphere, which covers the group space of SO(3), and thus half the group space of SU(2). Explicitly,

$$S = \tfrac{1}{8\pi^2 g^2}\oint d^3\sigma_m\ \tfrac{1}{6}\epsilon^{mnpq}B_{npq}$$

$$\to \tfrac{1}{8\pi^2 g^2}\oint d^3\sigma_m\ \tfrac{1}{6}\epsilon^{mnpq}tr(\mathbf{g}^{-1}\partial_n\mathbf{g})(\mathbf{g}^{-1}\partial_p\mathbf{g})(\mathbf{g}^{-1}\partial_q\mathbf{g})$$

$$= \tfrac{1}{8\pi^2 g^2} \oint d^3x \; \tfrac{1}{6}\epsilon^{ijk} tr(\mathbf{g}^{-1}\partial_i \mathbf{g})(\mathbf{g}^{-1}\partial_j \mathbf{g})(\mathbf{g}^{-1}\partial_k \mathbf{g})$$

where in the last step we have switched to coordinates for the 3-sphere, using the fact $\int d^4x \; \epsilon^{mnpq} f_{mnpq}$ is independent of coordinate choice. In fact, in the case where $\mathbf{g}$ is a one-to-one map between the 3-sphere and the group SO(3), this last expression is just the definition of the invariant volume of the SO(3) group space. In that case, the integral gives just the volume of the 3-sphere ($2\pi^2$). In general, the map $\mathbf{g}$ will cover the SU(2) group space an integer number q of times, and thus cover the SO(3) group space $2q$ times, so the result will be

$$S = \frac{|q|}{2g^2}$$

where we have used the fact that self-dual solutions have $q > 0$ while anti-self-dual have $q < 0$.

(Anti-)self-dual solutions give relative minima of the action with respect to more general field configurations:

$$0 \leq tr \int \tfrac{1}{2}(F_{ab} \pm \tfrac{1}{2}\epsilon_{abcd}F^{cd})^2 = tr \int (F^2 \pm \tfrac{1}{2}\epsilon^{abcd}F_{ab}F_{cd})$$

$$\Rightarrow \quad S \geq \frac{|q|}{2g^2}$$

q is an integer, and thus can't be changed by continuous variations: It is a topological property of finite-action configurations. Thus the self-dual solutions give absolute minima for a given topology. (All these solutions will be given implicitly by twistor construction in the following subsection. Note that our normalization for the structure constants of SU(2) differs from the usual, since we use effectively $tr(G_i G_j) = \delta_{ij}$ instead of the more common $tr(G_i G_j) = \frac{1}{2}\delta_{ij}$, which would normalize the structure constants as in SO(3): $f_{ijk} = \epsilon_{ijk}$. The net effect is that our g^2 contains a relative extra factor of 1/2, in addition to the effective extra factors coming from our different normalization of the action.)

Exercise IIIC6.4

Explicitly evaluate the integral for the instanton number q for the solutions of the 't Hooft ansatz. Show that the asymptotic form can be expressed in terms of $(\mathbf{g})_\lambda{}^\kappa = x^{\kappa\lambda'}$. ($det\ \mathbf{g} \neq 1$ because of the GL(1) piece.) Note that there are boundary contributions not only at $x = \infty$ but also around the singular points $x = x_i$, which are of the same form but opposite sign. (Since the singular parts of A are pure gauge, they cancel in F.)

At the quantum level, instantons are important mostly because they are an example of fields that don't fall off rapidly at infinity, and thus contribute to $\int \epsilon FF$.

However, once the restriction on boundary conditions is relaxed, there can be many such field configurations. The instantons are then distinguished by the fact that they are the minimal action solutions for a given topology; this makes them important for describing low-energy behavior.

7. ADHM

Much more general solutions of this form can be constructed using twistor methods. (In fact, they can be shown to be the most general self-dual solutions that fall off fast enough at infinity in all directions.) The first step of the Atiyah-Drinfel'd-Hitchin-Manin (ADHM) construction is to introduce a scalar square matrix in a larger group space

$$U_I{}^{I'} = (u_I{}^\iota, v_{Ii\alpha}) \qquad (I' = (\iota, i\alpha))$$

The index α is the usual two-valued twistor index, for SU(2) in Euclidean space or SL(2) in 2+2 dimensions. The other indices are

ι (H)	I (G)	i
SO(N)	SO(N+4k)	GL(2k)
SU(N) (SL(N))	SU(N+2k) (SL(N+2k))	GL(k,C)
USp(2N) (Sp(2N))	USp(2N+2k) (Sp(2N+2k))	GL(k)

The index ι is for the defining representation of the Yang-Mills group H, which is any of the compact classical groups for Euclidean space, but is its real Wick rotation for 2+2 dimensions. The index I is for the defining representation of the group G, a larger version of H, where k is the instanton number. Finally, the index i is for a general linear group. We also have the matrix

$$U^I{}_{I'} = (u^I{}_\iota, v^I{}_{i'\alpha})$$

For the SO and (U)Sp cases both U matrices are real, for the SU case they are complex conjugates of each other, and for the SL case they are real and independent. We next relate the two U's by

$$u^I{}_\iota u_I{}^\kappa = \delta_\iota^\kappa, \quad u^I{}_\iota v_{Ii\alpha} = v^I{}_{i'\alpha} u_I{}^\iota = 0, \quad v^I{}_{i'\alpha} v_{Ii\beta} = C_{\beta\alpha} g_{ii'}$$

so they are almost inverses of each other, except that the "metric" g is not constrained to be a Kronecker δ. We then write the gauge field as a generalization of pure gauge:

$$iA_{\mathcal{A}\alpha\iota}{}^\kappa = u^I{}_\iota \partial_{\mathcal{A}\alpha} u_I{}^\kappa$$

(This is similar to the method used for nonlinear σ models of coset spaces G/H as discussed in subsection IVA3 below, except for g.)

Self-duality then follows from requiring a certain coordinate dependence of the U's: This is fixed by giving the explicit dependence of the v's as

$$v_{Ii\alpha} = b_{Ii\mathcal{A}} z^{\mathcal{A}}{}_{\alpha}, \quad v^I{}_{i'\alpha} = b^I{}_{i'\mathcal{A}} z^{\mathcal{A}}{}_{\alpha}$$

where the b's are constants. The orthonormality conditions on the U's then implies the constraint on the b's

$$b^I{}_{i'(\mathcal{A}} b_{Ii\mathcal{B})} = 0$$

as well as determining the u's in terms of the b's (with much messier dependence than the v's), and thus A. Note that the z dependence of u can be written in terms of just x, as follows from rewriting the uv orthogonality as (after multiplying by z)

$$u^I{}_{\iota} b_{Ii\mathcal{A}} y^{\mathcal{A}\mathcal{B}} = u_I{}^{\iota} b^I{}_{i'\mathcal{A}} y^{\mathcal{A}\mathcal{B}} = 0$$

and noting scale invariance. Then the x components of A can also be written in terms of just x. We then can check the self-duality condition by calculating f: The orthonormality condition on the U's can be written as

$$\delta_I^J = u_I{}^{\iota} u^J{}_{\iota} + v_{Ii}{}^{\alpha} g^{ii'} v^J{}_{i'\alpha}$$

where $g^{ii'}$ is the inverse of $g_{ii'}$. Then schematically we have

$$\begin{aligned} iF &= \partial iA + iAiA \\ &= (\partial\bar{u})(\partial u) - (\partial\bar{u})u\bar{u}(\partial u) \\ &= (\partial\bar{u})vg\bar{v}(\partial u) \\ &= \bar{u}(\partial v)g(\partial\bar{v})u \\ &= \bar{u}bg\bar{b}u \end{aligned}$$

or more explicitly

$$if_{\mathcal{A}\mathcal{B}\iota}{}^{\kappa} = -(u^I{}_{\iota} b_{Ii(\mathcal{A}}) g^{ii'} (u_J{}^{\kappa} b^J{}_{i'\mathcal{B})})$$

where self-duality is $F_{\mathcal{A}\alpha,\mathcal{B}\beta} = C_{\alpha\beta} f_{\mathcal{A}\mathcal{B}}$. We can also directly show $z^{\mathcal{A}\alpha} f_{\mathcal{A}\mathcal{B}} = 0$.

Exercise IIIC7.1

Solve the bb constraint for k=1 and H=SU(2), and compare to the 1-instanton solution of subsection IIIC6.

8. Monopoles

Instantons are essentially 0-dimensional objects, localized near a point in 4-dimensional spacetime (or many points for multi-instanton solutions). Another type of solution is 1-dimensional; this represents a particle (with a 1D worldline). Unlike the plane-wave solutions, which represent the massless particles already described explicitly by fields in the action, we now look for time-independent solutions, which describe massive (since they have a rest frame), bound-state particles.

Looking at time-independent solutions is similar to the dimensional reduction that we considered in subsection IIB4 to introduce masses into free theories, only (1) this mass vanishes, and (2) we reduce the time dimension, not a spatial one. In our case, the dimensional reduction of a 4-vector (the Yang-Mills potential) gives a 3-vector and a scalar, both in the adjoint representation of the group. Let's consider the reduction in Euclidean space, so the scalar kinetic term comes out with the right sign. Then the 4D Yang-Mills action reduces as

$$\tfrac{1}{8}F_{ab}^2 \to \tfrac{1}{8}F_{ij}^2 + \tfrac{1}{4}[\nabla_i, \phi]^2$$

where we have labeled the scalar $A_0 = \phi$ and by dimensional reduction $\partial_0 \to 0$. Note that this is the same action that would have been obtained by starting out with Yang-Mills coupled to an adjoint scalar in four dimensions, either Minkowski or Euclidean, and choosing the gauge $A_0 = 0$. Thus, time-independent solutions to Euclidean Yang-Mills theory are also time-independent solutions to Minkowskian Yang-Mills coupled to an adjoint scalar (although not the most general, since the gauge $A_0 = 0$ is not generally possible globally, especially when we assume time independence of even gauge-dependent quantities). In particular, this means that time-independent solutions to self-dual Yang-Mills are also solutions of Minkowskian Yang-Mills coupled to an adjoint scalar. This allows us to use the first-order differential equations and topological properties of self-dual Yang-Mills theory to find physical bound-state particles in this vector-scalar theory.

Dimensionally reducing the (Euclidean) self-duality condition, we have

$$-[\nabla_i, \phi] = \tfrac{1}{2}\epsilon_{ijk}F_{jk}$$

As for instantons, the simplest solutions are for SU(2). As for the 't Hooft ansatz, we look for a solution that is covariant under the combined SU(2) of the gauge group and 3D rotations: In SO(3) vector notation for both kinds of indices (using the SO(3) normalization of the structure constants $[iG_i, iG_j] = \epsilon_{ijk} iG_k$),

$$\phi_i = x_i \varphi(r), \qquad (A_i)_j = \epsilon_{ijk} x_k \mathcal{A}(r)$$

(We know to use an ϵ tensor in A because of covariance under parity.) The self-duality equation then reduces to two nonlinear first-order differential equations (the coefficients of δ_{ij} and $x_i x_j/r^2$):

$$-\varphi - r^2\mathcal{A}\varphi = 2\mathcal{A} + r\mathcal{A}', \quad -r\varphi' + r^2\mathcal{A}\varphi = -r\mathcal{A}' + r^2\mathcal{A}^2$$

After some massaging, we find the change of variables

$$\tilde{\varphi} = \tfrac{1}{r} + r\varphi, \qquad \tilde{\mathcal{A}} = \tfrac{1}{r} + r\mathcal{A}$$

leads to the simplification

$$\tilde{\varphi}' = -\tilde{\mathcal{A}}^2, \qquad \tilde{\mathcal{A}}' = -\tilde{\mathcal{A}}\tilde{\varphi}$$

$\tilde{\varphi}$ then can be eliminated, giving an equation for $\tilde{\mathcal{A}}$. Making a final change of variables,

$$\psi = \tilde{\mathcal{A}}^{-1} \quad \Rightarrow \quad \psi\psi'' - (\psi')^2 = -1$$

we can guess the solution (with regularity at $r = 0$)

$$\psi = k^{-1}sinh(kr) \quad \Rightarrow \quad \mathcal{A} = \frac{1}{r^2}\left(\frac{kr}{sinh(kr)} - 1\right), \quad \varphi = \frac{1}{r^2}[kr\ coth(kr) - 1]$$

Exercise IIIC8.1

Repeat this calculation in spinor notation:

a In Euclidean space we can choose $\sigma^0_{\alpha\beta'} \sim C_{\alpha\beta'}$. Show that we can then write the 4-vector potential for the monopole as

$$i(A_{\alpha\beta})^{\gamma\delta} = \delta_\alpha^\gamma x_\beta{}^\delta \mathcal{A}_+(r) + \delta_\beta^\delta x_\alpha{}^\gamma \mathcal{A}_-(r)$$

which is symmetric in neither $\alpha\beta$ nor $\gamma\delta$. (Compare the 't Hooft ansatz in subsection IIIC6.) However, $x^{\alpha\beta}$ is now symmetric from dropping x^0.

b Impose self-duality, where

$$\partial_{\alpha\beta}x^{\gamma\delta} = \tfrac{1}{2}\delta_{(\alpha}^\gamma\delta_{\beta)}^\delta = \delta_\alpha^\gamma\delta_\beta^\delta - \tfrac{1}{2}C_{\alpha\beta}C^{\gamma\delta}$$

from subtracting out the $\partial_0 x^0$ piece. Derive the resulting equations for $\mathcal{A}_\pm$, and show they agree with the above for

$$\mathcal{A}_\pm = -\tfrac{1}{2}(\mathcal{A} \pm \varphi)$$

In general, the Lagrangian of a Euclidean theory is the Hamiltonian of the Minkowskian theory (with the sign conventions we introduced in subsection IIIA1), since Wick rotation changes the sign of the kinetic energy and not the potential energy. In our case, this means the Minkowskian energy of the Yang-Mills + adjoint scalar theory can be evaluated in terms of the same topological expression we used for instantons: From the previous subsection, using $S = \int dt\ E$ and $\partial_0 = 0$, we evaluate in Euclidean space

$$E = \tfrac{1}{16\pi^2 g^2} \oint d^2\sigma_i\ \epsilon^{i0jk} B_{0jk}, \quad \epsilon^{i0jk} B_{0jk} \to -\epsilon_{ijk} tr(\phi F_{jk}) = 2\ tr(\phi[\nabla_i, \phi]) = \partial_i\ tr(\phi^2)$$

where we have used an integration by parts to simplify B. (Compare exercise IIIC6.3a.) Since at spatial infinity

$$\phi_i \to x_i \left(\frac{|k|}{r} - \frac{1}{r^2} \right), \qquad A_{ij} \to -\epsilon_{ijk} x_k \frac{1}{r^2}$$

and effectively $\oint d^2\sigma_i \to 4\pi r x_i$, we find

$$E = \frac{|k|}{2\pi g^2}$$

Also by similar arguments to those used for instantons, we see that any solutions with boundary conditions $A \to 0$, $|\phi| \to |k|$ as $r \to \infty$ have energy at least as great as this. There is also a topological interpretation to this energy: Writing it as

$$E = -\tfrac{1}{16\pi^2 g^2} \oint d^2\sigma_i\ \epsilon_{ijk} tr(\langle\phi\rangle F_{jk})$$

we see that the energy is proportional to the magnetic flux, i.e., the "magnetic charge" of the monopole. (The asymptotic value $\langle\phi\rangle$ of ϕ picks out a direction in isospace, reducing SU(2) to U(1).) As in electromagnetism, magnetic charge is quantized in terms of electric charge. However, for compact gauge groups, electric charge is also quantized. (For the usual U(1), charges are arbitrary, but for SU(2), any component of the isospin is quantized.) The energy is thus quantized in terms of k: It is a multiple of the energy we found for the single monopole above.

Exercise IIIC8.2

Perform a singular gauge transformation that makes $\langle\phi\rangle$ point in a constant (rather than radial) direction in isospin (SU(2)) space. Show that the isospin component of the asymptotic form of A describes a U(1) magnetic monopole: magnetic flux radiating outward from the origin.

REFERENCES

1. O. Klein, On the theory of charged fields, in *New theories in physics*, proc. Warsaw, May 30 - June 3, 1938 (International Institute of Intellectual Co-operation, Paris, 1939) p. 77;

W. Pauli, unpublished (1953):
early work on "Yang-Mills" theory.
2 C.N. Yang and R.L. Mills, *Phys. Rev.* **96** (1954) 191;
R. Shaw, *The problem of particle types and other contributions to the theory of elementary particles*, Cambridge University Ph.D. thesis (1955);
R. Utiyama, *Phys. Rev.* **101** (1956) 1597:
complete formulation of the (classical) theory.
3 L. O'Raifeartaigh, *The dawning of gauge theory* (Princeton University, 1997):
early history of Yang-Mills theory, including reprints and previously unpublished material.
4 R.L. Arnowitt and S.I. Fickler, *Phys. Rev.* **127** (1962) 1821;
W. Kummer, *Acta Phys. Austriaca* **14** (1961) 149.
5 S. Weinberg, *Phys. Rev.* **150** (1966) 1313;
J.B. Kogut and D.E. Soper, *Phys. Rev.* **D1** (1970) 2901:
lightcone field theory.
6 S. Coleman, *Phys. Lett.* **70B** (1977) 59:
Yang-Mills plane waves.
7 R. Feynman and M. Gell-Mann, *Phys. Rev.* **109** (1958) 193;
L.M. Brown, *Phys. Rev.* **111** (1958) 957;
M. Tonin, *Nuo. Cim.* **14** (1959) 1108;
G. Chalmers and W. Siegel, hep-ph/9708251, *Phys. Rev.* **D59** (1999) 045012;
M. Veltman, hep-th/9712216, *Acta Phys. Polon. B* **29** (1998) 783:
description of spin 1/2 with only undotted spinors.
8 A. Ashtekar, *Phys. Rev. Lett.* **57** (1986) 2244, *Phys. Rev.* **D36** (1987) 1587;
T. Jacobson and L. Smolin, *Phys. Lett.* **196B** (1987) 39, *Class. Quant. Grav.* **5** (1988) 583;
J. Samuel, *Pramana* **28** (1987) L429:
first-order actions with self-dual auxiliary fields.
9 C.N. Yang, *Phys. Rev. Lett.* **38** (1977) 1377:
reduction of self-dual Yang-Mills field to single component.
10 A.N. Leznov, *Theor. Math. Phys.* **73** (1988) 1233,
A.N. Leznov and M.A. Mukhtarov, *J. Math. Phys.* **28** (1987) 2574;
A. Parkes, hep-th/9203074, *Phys. Lett.* **286B** (1992) 265:
lightcone gauge for self-dual Yang-Mills.
11 A.A Belavin, A.M. Polyakov, A.S. Shvarts, and Yu.S. Tyupkin, *Phys. Lett.* **59B** (1975) 85:
instantons.
12 G. 't Hooft, unpublished;
R. Jackiw, C. Nohl, and C. Rebbi, *Phys. Rev.* **D15** (1977) 1642:
't Hooft ansatz for multi-instanton solutions.
13 S.-S. Chern and J. Simons, *Ann. Math.* **99** (1974) 48.
14 M.F. Atiyah and R.S. Ward, *Comm. Math. Phys.* **55** (1977) 117;
M.F. Atiyah, V.G. Drinfel'd, N.J. Hitchin, and Yu.I. Manin, *Phys. Lett.* **65A** (1978) 185;
E. Corrigan, D. Fairlie, P. Goddard, and S. Templeton, *Nucl. Phys.* **B140** (1978) 31;
N.H. Christ, E.J. Weinberg, and N.K. Stanton, *Phys. Rev.* **D18** (1978) 2013;
M.F. Atiyah, *Geometry of Yang-Mills fields* (Scuola Normale Superiore, Pisa, 1979);

V.E. Korepin and S.L. Shatashvili, *Math. USSR Izvestiya* **24** (1985) 307:
general multi-instanton construction, using twistors.

15 G. 't Hooft, *Nucl. Phys.* **B79** (1974) 276;
A.M. Polyakov, *JETP Lett.* **20** (1974) 194:
monopoles in nonabelian theories.

16 E.B. Bogomol'nyi, *Sov. J. Nucl. Phys.* **24** (1976) 449;
M.K. Prasad and C.M. Sommerfeld, *Phys. Rev. Lett.* **35** (1975) 760:
exact solution for monopole.

17 W. Nahm, *Phys. Lett.* **90B** (1980) 413:
general monopole construction based on ADHM instanton construction.

IV. MIXED

In this chapter we consider ways in which gauge symmetry combines with global symmetries for new effects. The interplay between global internal symmetries of scalar and spinor theories and local symmetries of Yang-Mills is important for understanding mass generation for all spins, and is fundamental for the Standard Model.

A. HIDDEN SYMMETRY

Symmetries, especially local ones, are clearly very important in the formulation of interactions. However, symmetries are not always apparent in nature: For example, while most symmetries prefer massless particles, of all the observed particles the only massless ones are the graviton, photon, and neutrinos. Furthermore, of the massive ones, none with different properties have the same mass, although some are close (e.g., the proton and neutron). There are three solutions to this problem:

(1) The symmetry is not a property of nature, but only an approximate symmetry. Some terms in the action are invariant under the symmetry, but other terms violate it. We can treat such "explicit symmetry breaking" by first studying the symmetry for the invariant terms, and then treating the breaking terms as a perturbation.

(2) Although the laws of physics are symmetric, nature is an asymmetric solution to them. In particular, such a solution is the "vacuum", or state of lowest energy, with respect to which all other states are defined. This case is called "spontaneous symmetry breaking".

(3) The particles in terms of which these laws are formulated are not those observed in nature. For example, the hydrogen atom is most conveniently described in terms of a proton and an electron, but in its low-energy physics only the atom itself is observed as a separate entity: The U(1) symmetry related to charge is not seen from the neutral atoms. The more extreme case where such particles always appear in bound states is known as "confinement".

Generally, such broken symmetries are at least partially restored at high energies. For example, if the symmetry breaking introduces masses, or mass differences between related particles, then the symmetry may become apparent at energies large with respect to those masses. Similarly, a hydrogen atom excited to an energy much larger than its lower energy levels will ionize to reveal its constituent particles.

It often is possible to change to a set of variables that are invariant under a local symmetry. For example, if we can define everywhere a variable that transforms as $\delta\phi(x) = \lambda(x)$, then it can be used to everywhere undo the invariance. We can choose the "gauge" $\lambda = -\phi$, transforming ϕ to 0 everywhere, leaving no residual invariance, or we can work with composite, invariant variables: E.g., $\psi' = \psi e^{i\lambda}$ is replaced (invertibly) with $\hat{\psi} = \psi e^{-i\phi}$, so $\hat{\psi}' = \hat{\psi}$.

1. Spontaneous breakdown

We first consider symmetry breaking by the vacuum, known as "spontaneous breakdown". The action is invariant under the symmetry, but the vacuum state is not: Thus, the symmetry acting on the vacuum produces other zero-energy solutions to the field equations, but this symmetry is not apparent when considering perturbation about the vacuum. In this case, although the symmetry is broken, there are obvious residual effects, particularly if the breaking can be considered as "small" with respect to some other effects.

The "Goldstone theorem" is an important statement about the effect of symmetry breakdown: If a continuous global symmetry is spontaneously broken, then there is a corresponding massless scalar. The proof is simple: Consider a (relative) minimum of the potential, as the vacuum. By definition, we have spontaneous symmetry breaking if this minimum is not invariant under the continuous symmetry: i.e., applying infinitesimal symmetry transformations gives a curve of nearby states, which have the same energy, because the transformations are a symmetry of the theory. But the mass of a scalar, by definition, is given by the quadratic term in its potential, i.e., the second derivative of the potential evaluated at the vacuum value. (The first derivative vanishes because the vacuum is a minimum.) So, if we look at the scalar defined to parametrize this curve of constant energy in field space, its mass vanishes. (This field may be a function of the given fields, such as an angle in field space.)

We can also formulate this more mathematically, for purposes of calculation: Consider a theory with potential $V(\phi^i)$. (The Lagrangian is V plus derivative terms. For simplicity we consider just scalars.) The masses of the scalars are defined by the quadratic term in the potential, expanding about a minimum, the vacuum. The statement of symmetry of the potential means that

$$symmetry \quad \delta\phi^i = \zeta^i(\phi) \quad \Rightarrow \quad 0 = \delta V = \zeta^i \partial_i V \quad for\ all\ \phi^i$$

where we allow nonlinear symmetries, and $\partial_i = \partial/\partial\phi^i$. Differentiating, and then evaluating at this minimum,

$$\langle \partial_i V \rangle = 0 \quad at\ minimum\ \phi = \langle \phi \rangle$$

$$\Rightarrow \quad 0 = \langle \partial_j(\zeta^i \partial_i V)\rangle = \langle (\partial_j \zeta^i)(\partial_i V)\rangle + \langle \zeta^i \partial_i \partial_j V\rangle = \langle \zeta^i\rangle\langle \partial_i\partial_j V\rangle$$

where here the vacuum value $\langle\ \rangle$ classically means to just evaluate at $\phi = \langle\phi\rangle$. (So classically $\langle AB\rangle = \langle A\rangle\langle B\rangle$.) Spontaneous symmetry breaking means the vacuum breaks the symmetry: If this symmetry is broken, then $\langle\zeta^i\rangle \neq 0$, so it is a nontrivial eigenvector of $\langle\partial_i\partial_j V\rangle$ (the mass matrix) with vanishing eigenvalue. So, we can write

$$\phi^i = \langle\phi^i\rangle + \chi\langle\zeta^i\rangle + ...$$

where χ is a massless field.

The simplest example is a single free, massless field, $V = 0$. Then ζ is simply a constant. The simplest choice of vacuum is just $\langle\phi\rangle = 0$, which breaks the symmetry:

$$L = \tfrac{1}{4}(\partial\phi)^2, \qquad \delta\phi = constant, \qquad \langle\phi\rangle = 0$$

Then ϕ is a "Goldstone boson".

The simplest nontrivial example, and a useful one, is a complex scalar with the potential

$$V(\phi) = \tfrac{1}{4}\lambda^2(|\phi|^2 - \tfrac{1}{2}m^2)^2$$

This is invariant under phase transformations $\delta\phi = i\zeta\phi$. There is a continuous set of minima at $|\phi| = m/\sqrt{2}$. We choose $\langle\phi\rangle = m/\sqrt{2}$; then the Goldstone theorem tells us that the imaginary part of ϕ is the Goldstone field. Explicitly, separating the field into its real and imaginary parts,

$$\phi = \tfrac{1}{\sqrt{2}}(m + \psi + i\chi) \quad \Rightarrow \quad V = \tfrac{1}{4}\lambda^2 m^2\psi^2 + \tfrac{1}{4}\lambda^2 m\psi(\psi^2 + \chi^2) + \tfrac{1}{16}\lambda^2(\psi^2 + \chi^2)^2$$

where $\langle\psi\rangle = \langle\chi\rangle = 0$. We could also use the nonlinear separation of the field into magnitude and phase, $\phi = (m + \rho)e^{i\theta}/\sqrt{2}$: Then θ drops out of the potential, and its transformation (ρ is invariant) is the same as that of the free massless scalar. If ϕ had been real, then only the discete symmetry $\phi \leftrightarrow -\phi$ would have been broken, and there would be no Goldstone boson.

Exercise IVA1.1

Write the complete action in terms of ρ and θ.

Note that this model would naively seem to have a tachyon (state with negative $(\text{mass})^2$) if we had expanded about $\langle\phi\rangle = 0$. However, since the vacuum is defined always as a minimum in the potential (or the energy), the true states always have nonnegative $(\text{mass})^2$. This is the case for positive spins for similar reasons: We saw in subsection IIB4 that free massive theories follow from massless ones by dimensional reduction from one extra spatial dimension. If we had used an extra time dimension instead, as required for the "wrong sign" for the mass term in $p^2 + m^2$, there would also be wrong signs for Lorentz indices, resulting in kinetic terms with arbitrarily negative energy.

Spontaneous symmetry breaking will also effect the actions for fields other than those getting vacuum values, that couple to them. For example, terms of the form $\psi^2 f(\phi)$ will tend to generate a mass for ψ if $\langle\phi\rangle \neq 0$ (actually $f(\langle\phi\rangle) \neq 0$). Such couplings exist for ψ of spins $0, \frac{1}{2}, 1$. Since masslessness is generally associated with symmetry (chiral symmetry for spin $\frac{1}{2}$ and gauge symmetry for spin 1), this type of mass generation implies symmetries other than just those of the scalars are broken by this mechanism (see subsections IVA4-6).

2. Sigma models

The Goldstone mechanism thus produces massive particles as well as massless ones, at least for polynomial potentials, to which we are restricted by quantum considerations, to be discussed later. We now look for approximations to polynomial scalar actions that eliminate the massive fields, but still take them into account through their equations of motion, in the limit where their masses tend to infinity. For example, in the above simple model, we can take the limit $\lambda \to \infty$, which takes the ψ mass (λm) to infinity. In this limit, the potential energy can remain finite only if it vanishes: $|\phi|^2 = \frac{1}{2}m^2$. (In quantum language, the potential's contribution to the path integral is just $\delta(\int V)$ in that limit. Alternatively, we can neglect the kinetic energy for $|\phi|$ in comparison to the mass or potential, and then eliminate $|\phi|$ through its equation of motion in this approximation.) We can also enforce this limit directly by using a Lagrange multiplier field Λ:

$$L = \tfrac{1}{2}|\partial\phi|^2 + \Lambda(|\phi|^2 - \tfrac{1}{2}m^2)$$

The solution to the constraint is $\phi = \frac{m}{\sqrt{2}}e^{i\theta}$, and the action then describes just a free, real scalar θ.

A less trivial example is a nonabelian generalization of this example: Consider ϕ as a vector of an internal SO(n) symmetry. (The previous example was the case SO(2).) The Lagrangian is then

$$L = \tfrac{1}{4}(\partial\phi)^2 + \tfrac{1}{2}\Lambda(\phi^2 - m^2)$$

The usual way to solve quadratic constraints without introducing square roots is to use the identity

$$|(1+ix)^2|^2 = (|1+ix|^2)^2 \quad \Rightarrow \quad (2x)^2 + (1-x^2)^2 = (1+x^2)^2$$

This is often used for trigonometric substitutions or simplifying integrals. For example, when an integrand has a $\sqrt{1-x^2}$, substituting $x = sin\ \theta$ eliminates the square root at the price of requiring trigonometric identities, which in turn are usually solved by making a second variable change to $y = tan(\theta/2)$. On the other hand, the above identity suggests making instead the variable change $x = 2y/(1+y^2)$, which actually gives the same result, more directly, as the previous two-step method. (This identity can also be used for finding integer solutions to the Pythagorean theorem: A right triangle with two shorter sides of integer lengths $2mn$ and m^2-n^2 has the hypotenuse m^2+n^2, where m, n are integers.)

We then can solve the constraint $\phi^2 = m^2$ with the coordinates for the sphere in terms of an SO(n−1) vector χ,

$$\phi = m\left(\frac{\chi}{1+\frac{1}{4}\chi^2}, \frac{1-\frac{1}{4}\chi^2}{1+\frac{1}{4}\chi^2}\right)$$

Then the kinetic term (now the whole action) becomes

$$\tfrac{1}{4}(\partial\phi)^2 = \tfrac{1}{4}m^2\frac{(\partial\chi)^2}{(1+\frac{1}{4}\chi^2)^2}$$

Exercise IVA2.1

For SO(3), express χ in terms of the usual spherical polar angular coordinates θ and φ, along with the inverse expressions (θ and φ in terms of χ).

Another way to obtain this result is to use the solution of subsection IA6 to the constraint

$$0 = y^2 = (y^a)^2 - 2y^+y^- \quad \Rightarrow \quad y = e(x^a, 1, \tfrac{1}{2}x^2)$$

(but now $(y^a)^2$ is positive definite). Then the desired constraint

$$(y^a)^2 + (y^1)^2 = 1$$

follows from further constraining

$$1 = y^0 = e\tfrac{1}{\sqrt{2}}(1+\tfrac{1}{2}x^2) \quad \Rightarrow \quad e = \frac{\sqrt{2}}{1+\frac{1}{2}x^2}$$

$$\Rightarrow \quad dy^2 = e^2 dx^2 = \frac{2dx^2}{(1+\frac{1}{2}x^2)^2}$$

yielding the above result for $x = \chi/\sqrt{2}$. We thus have a nonpolynomial action, each term having derivatives. The original SO(n) symmetry is nonlinearly realized on the "angle" variables χ, and the vacuum ($\langle\chi\rangle = 0$) spontaneously breaks the symmetry to SO(n−1). The constant m acts as a dimensionful coupling, as seen by scaling $\chi \to \chi/m$ to give the kinetic term the standard normalization.

A complex generalization of this model is described by the Lagrangian

$$L = \tfrac{1}{2}|\nabla\phi|^2 + \Lambda(|\phi|^2 - m^2)$$

where ϕ is now a complex n-component vector, ∇ is a U(1)-covariant derivative ($\nabla\phi = (\partial + iA)\phi$), and Λ is a Lagrange multiplier enforcing that ϕ has magnitude m. This model thus has a U(n) symmetry. Since A has no kinetic term (F^2), we can eliminate it by its algebraic field equation:

$$L \to \tfrac{1}{2}|\partial\phi|^2 + \tfrac{1}{8m^2}(\phi^\dagger \overleftrightarrow{\partial}\phi)^2 + \Lambda(|\phi|^2 - m^2)$$

where we have applied the constraint $|\phi|^2 = m^2$ (or shifted Λ to cancel terms proportional to $|\phi|^2 - m^2$). Since the U(1) gauge was not fixed yet, we still have local U(1) invariance even without an explicit gauge field. We can use this invariance to fix the phase of one component of ϕ, and use the constraint from Λ to fix its magnitude. In terms of the remaining (n−1)-component complex vector χ,

$$\phi = m\left(\frac{\chi}{1 + \frac{1}{4}|\chi|^2}, \frac{1 - \frac{1}{4}|\chi|^2}{1 + \frac{1}{4}|\chi|^2}\right)$$

$$\Rightarrow \quad L = \tfrac{1}{2}m^2 \frac{|\partial\chi|^2 + \frac{1}{4}(\chi^\dagger \overleftrightarrow{\partial}\chi)^2}{(1 + \frac{1}{4}|\chi|^2)^2}$$

(Alternatively, we can solve the constraint and fix the gauge first, then eliminate A by its field equation.) This model is known as the CP(n−1) model ("complex projective").

Another example that will prove more relevant to physics is to generalize ϕ to an n⊗n matrix: We then consider the Lagrangian

$$L = tr[\tfrac{1}{2}(\partial\phi)^\dagger \cdot (\partial\phi) + \tfrac{1}{4}\lambda^2(\phi^\dagger\phi - \tfrac{1}{2}m^2 I)^2]$$

(where I is the identity matrix). Since $\phi^\dagger\phi$ is hermitian and positive definite, the minimum of the potential is at $\phi^\dagger\phi = \frac{1}{2}m^2 I$, and we can choose

$$\langle\phi\rangle = \tfrac{m}{\sqrt{2}}I$$

using the SU(n)⊗SU(n)⊗U(1) invariance

$$\phi' = U_L \phi U_R{}^{-1}$$

(We can include the U(1) in either U_L or U_R.) The vacuum then spontaneously breaks this invariance to SU(n):

$$\langle\phi'\rangle = \langle\phi\rangle \quad \Rightarrow \quad U_L = U_R$$

In the large-mass limit, we get the constraint

$$\phi^\dagger\phi = \tfrac{1}{2}m^2 I \quad \Rightarrow \quad \phi = \tfrac{m}{\sqrt{2}}U, \quad U^\dagger U = I, \quad \langle U\rangle = I, \quad L \to \tfrac{1}{4}m^2 tr[(\partial U)^\dagger(\partial U)]$$

so the field U itself is now unitary.

3. Coset space

The appearance of the scalar fields (Goldstone bosons) as group elements can be generalized directly in terms of the effective theory, without reference to massive fields. Such a theory should be considered as a low-energy approximation to some unknown theory. Although the unknown theory may be better behaved at high energies quantum mechanically (see later), the low-energy effective theory can be determined from just (broken) symmetry. We therefore assume a symmetry group G that is broken down to a subgroup H by the vacuum. (I.e., the vacuum is invariant under the subgroup H, but not the full group G.) We are interested in only the Goldstone bosons, associated with all the generators of the group G less those of H. These fields are thus coordinates for the "coset space" G/H: They correspond to elements of the group G, but elements related by the subgroup H are identified.

Explicitly, we first write the field g as an element of the group G, either by choosing a matrix representation of the group (as in the U(N) example above), or explicitly expanding over the group generators G_I:

$$g = e^{i\phi}, \qquad \phi = \phi^I(x)G_I$$

We then "factor" out the subgroup H by introducing a gauge invariance for that subgroup:

$$g' = gh, \qquad h = e^{ih^\iota(x)H_\iota}$$

in terms of the H generators H_ι, which are a subset of G_I:

$$G_I = (H_\iota, T_i)$$

where T_i are the remaining generators, corresponding to G/H. In particular, we can choose

$$gauge\ \phi^\iota = 0 \quad \Rightarrow \quad \phi = \phi^i T_i$$

However, G should still be a global invariance of the theory, though not of the vacuum. We therefore assume the global transformation

$$g' = g_0 g$$

where g_0 is an element of the full group G, but is constant in x. The vacuum

$$\langle g \rangle = I$$

is then invariant under the global subgroup $g_0 = h^{-1}$, where thus h is constant and $g_0 \in$H (i.e., G is spontaneously broken to H).

g can be used to convert any representation of the global group G into one (but usually reducible) of the smaller local group H:

$$\psi' = g_0 \psi \quad \Rightarrow \quad \tilde{\psi} \equiv g^{-1}\psi, \quad \tilde{\psi}' = h^{-1}\tilde{\psi}$$

We can apply a similar procedure to find a field strength for g, invariant under the global group, as an element of the Lie algebra of G:

$$g^{-1}\partial_a g = \partial_a + iA_a{}^\iota H_\iota + iF_a{}^i T_i = \nabla_a + iF_a{}^i T_i$$

This can be evaluated in the ϕ parametrization as multiple commutators, as usual: A and F are both nonpolynomial functions of ϕ, but with only one derivative. We have absorbed A into a covariant derivative ∇ because of the remaining transformation law under the local group H:

$$(\nabla + iF)' = h^{-1}(\nabla + iF)h \quad \Rightarrow \quad \nabla' = h^{-1}\nabla h, \quad F' = h^{-1}Fh$$

where we have assumed $[H_\iota, T_i] \sim T_j$. (In particular, this is true for compact groups, where the structure constants are totally antisymmetric: Then $f_{\iota\kappa i} = 0 \Rightarrow f_{\iota i\kappa} = 0$.) Then the action invariant under global and local transformations can be chosen as

$$L = \tfrac{1}{4}m^2 tr(F^2)$$

For example, the real vector model we gave in the previous subsection describes the coset space SO(n)/SO(n−1), the complex vector describes SU(n)/U(n−1), and the matrix model describes U(n)⊗U(n)/U(n).

Exercise IVA3.1

Use the coset-space construction to derive the specific σ models explicitly given in the previous subsection, as just identified.

a Find the real and complex vectors by dividing up the adjoint representation into appropriate blocks.

b For the case of U(n)⊗U(n)/U(n), the direct product means we use separate group-element fields for the two global groups, with

$$g'_L = g_{L0} g_L h, \qquad g'_R = g_{R0} g_R h$$

for the same h. Find an expression for the field U of the previous subsection without breaking any global or local symmetries.

Note that the field redefinition between the G-representation matter field ψ and the H-representation matter field $\tilde{\psi}$ modifies the form of the couplings. For example, the kinetic term for ψ will have ordinary partial derivatives ∂, while that for $\tilde{\psi}$ will have covariant ones ∇. (One or the other will also have F terms.) On the other hand, a mass term for $\tilde{\psi}$ may turn into a potential/Yukawa term for ψ, since the larger group G might not allow mass terms permitted by the smaller group H. The result is that what appears as a nonderivative coupling in terms of ψ may appear as a derivative coupling in terms of $\tilde{\psi}$.

We can formulate general spontaneous breakdown in this language:

(1) Start with a polynomial action with symmetry G, including scalars ϕ that will suffer the breakdown through expectation, and other fields ψ.

(2) Introduce an appropriate g and define the new scalar fields $\tilde{\phi} \equiv g^{-1}\phi$, as well as the new matter fields $\tilde{\psi} \equiv g^{-1}\psi$. Thus $S[\phi,\psi] \to S[\tilde{\phi},\tilde{\psi},g]$. In terms of these new fields, the action has a local symmetry H, and G now acts *only* on g.

(3) (H-covariantly) constrain $\tilde{\phi}$ in such a way that g effectively replaces the missing parts. Then g describes all the Goldstone bosons, while the reduced $\tilde{\phi}$ describes the other scalars in the original ϕ (in the previous examples, the massive ones, which decoupled at low energies).

For example, for the SO(n) model, ϕ is an n-vector, while g parametrizes the coset SO(n)/SO(n−1), and thus has n(n−1)/2 − (n−1)(n−2)/2 = n−1 non-gauge components — it is an n−1-vector under H. Thus for $\tilde{\phi}$, which is n→n−1⊕1 under H, we just constrain the n−1 part to vanish.

We have described how nonpolynomial actions quadratic in derivatives can arise as a low-energy approximation to polynomial theories. Further nonpolynomial terms quartic in derivatives (but no more than quadratic in time derivatives) can be useful for certain applications, but these arise from polynomial actions quadratic in derivatives (which are preferred for quantum reasons) only by quantum effects. One use is in models which describe (pseudo)scalar mesons by fundamental fields (i.e., solutions to the free field equations, which yield interacting solutions through perturbation theory), but baryons by nonperturbative solutions to the field equations of these scalars.

Such an interpretation is suggested by an expansion in 1/N, where N is the number of colors, since a baryon is made of N quarks (whereas a meson contains just one quark and one antiquark). Such models are useful for describing static properties of baryons (masses, quantum numbers), but the complexity of such solutions to the field equations prevents their use for interactions of baryons (especially with other baryons).

4. Chiral symmetry

Later we'll examine a description of the strongly interacting particles ("hadrons") in which they are all considered as composites (bound states) of fermionic "quarks". However, this theory is extremely difficult to solve, so we first consider treating the hadrons as fundamental instead. Since there are probably an infinite number of kinds of hadrons (or at least some integer power of 10^{40}, considering the (Planck mass)^2), this would require a formulation in terms of a "string" that treated all "mesons" (bosonic hadrons) as a single entity. That possibility also will be considered later; for now, we look at the simpler possibility of studying just the low-energy physics of hadrons by using fields for just the lightest particles.

So far, the only observed scalar particles have been strongly interacting ones. Some of the scalar mesons, especially the "pions", are not only the lightest hadrons, but can be considered close to massless on the hadronic scale. We therefore look for a description of pions (and some close relatives) in the massless approximation; then mass-generating corrections can be considered.

Normally, quantum corrections can affect masses. The only way to guarantee masslessness at the quantum level is through some symmetry; we then can study this symmetry already at the classical level. We have seen that (unbroken) gauge invariance can require masslessness for all fields except the scalar and spinor. Masslessness for a spinor can be enforced by "chiral symmetry": If there is a U(1) symmetry for all irreducible spinors ψ_α, then no mass terms (bilinears $\psi_1^\alpha \psi_{2\alpha}$) can be constructed. (Generally, each spinor can have different U(1) charges, as long as no two charges add to zero. Of course, this U(1) can be a subgroup of a larger chiral symmetry group.) The only way a scalar can be guaranteed masslessness is if it is a Goldstone boson. We therefore look for a description of pions as Goldstone bosons of some spontaneously broken symmetry. (Supersymmetry is another possibility to enforce massless scalars, but only if there are also massless fermions, which is not the case for hadrons.) Furthermore, pions and the other lightest scalars are actually *pseudo*scalars: This suggests that they are the Goldstone bosons of broken chiral symmetry, which simultaneously generates masses for the fermions.

For simplicity, we consider the coupling of scalar mesons to quarks. We could instead couple mesons to "baryons" (fermionic hadrons), thus treating only hadrons, but the principles would be the same, only the indices would be messier. Combining C invariance with chiral symmetry, and including a meson potential for spontaneous symmetry breaking, we can write the action for just the quarks and scalar mesons as

$$S = \int dx \; tr \; L$$

$$L = [q^{\dagger\dot{\alpha}}_L i\partial^\beta{}_{\dot{\alpha}} q_{L\beta} + q_R^{T\beta} i\partial_\beta{}^{\dot{\alpha}} q^*{}_{R\dot{\alpha}}] + [\tfrac{1}{2}(\partial\phi)^\dagger \cdot (\partial\phi) + \tfrac{1}{4}\lambda^2(\phi^\dagger\phi - \tfrac{1}{2}m^2 I)^2]$$
$$+ \Lambda[q_L^\alpha \phi q_{R\alpha}^T + q^{*\dot{\alpha}}_R \phi^\dagger q^\dagger{}_{L\dot{\alpha}}]$$

where ϕ is an m$\otimes$m matrix (m "flavors"), q_L and q_R are n$\otimes$m matrices (n "colors"), and Λ is the "Yukawa coupling". Sometimes it will be convenient to drop Lorentz indices to emphasize internal symmetries:

$$L = (q^\dagger{}_L i\partial q_{L\beta} + q_R^T i\partial q^*{}_R) + [\tfrac{1}{2}(\partial\phi)^\dagger \cdot (\partial\phi) + \tfrac{1}{4}\lambda^2(\phi^\dagger\phi - \tfrac{1}{2}m^2 I)^2] + \Lambda(q_L\phi q_R^T + q_R{}^*\phi^\dagger q^\dagger{}_L)$$

Besides color symmetry (local if we had bothered to write in the Yang-Mills fields for the "gluons", by $\partial \to \nabla$ on the quarks), we have the (global) $\mathrm{U(m)}_L \otimes \mathrm{U(m)}_R$ chiral (flavor) symmetry

$$q_L' = q_L U_L, \quad q_R' = q_R U_R{}^*, \quad \phi' = U_L^{-1}\phi U_R$$

including the (global) U(1) "baryon number" symmetry

$$U_L = U_R = e^{i\theta} \quad \Rightarrow \quad q_L' = e^{i\theta} q_L, \quad q_R' = e^{-i\theta} q_R, \quad \phi' = \phi$$

If we think of baryon number as an SO(2) symmetry, then charge conjugation is just the reflection that completes this to an O(2) symmetry (see exercise IIA1.2):

$$C: \qquad q_L \leftrightarrow q_R, \quad \phi \to \phi^T$$

From this, and the usual CP

$$CP: \qquad q_L \to q_L{}^*, \quad q_R \to q_R{}^*, \quad \phi \to \phi^*$$

we find the parity symmetry

$$P: \qquad q_L \leftrightarrow q_R{}^*, \quad \phi \to \phi^\dagger$$

(where for CP and P we also transform the coordinates as usual).

As before, the vacuum $\langle\phi\rangle = \frac{m}{\sqrt{2}}I$ breaks the flavor symmetry to the diagonal subgroup $U_L = U_R$, which commutes with parity (and is therefore no longer "chiral"). It also gives masses to the quarks (since chiral symmetry is broken); this is a general feature of spinors coupled to scalars under spontaneous breakdown. In the limit $\lambda \to \infty$ (where the mass of all bosons but the Goldstones becomes infinite, but the quark mass $M = \Lambda m$ is fixed), the Goldstone bosons are described by the unitary matrix U, which transforms as $U' = U_L^{-1}UU_R$.

Exercise IVA4.1

Rewrite this action according to the analysis of exercise IVA3.1b:

a Separate the Goldstone bosons from the massive scalars.

b Replace the G-representation quarks with the H-representation quarks.

An interesting special case is m=1 (one flavor). The Goldstone boson of axial U(1) can be identified with the π^0. In the limit $\lambda \to \infty$, the Lagrangian becomes (with a tr no longer needed)

$$L = (q^{\dagger\dot{\alpha}}{}_L i\partial^\beta{}_{\dot{\alpha}} q_{L\beta} + q_R^{T\beta} i\partial_\beta{}^{\dot{\alpha}} q^*{}_{R\dot{\alpha}}) + \tfrac{1}{4}m^2(\partial\pi)^2 + \tfrac{M}{\sqrt{2}}(e^{i\pi}q_R^{T\alpha}q_{L\alpha} + e^{-i\pi}q^{\dagger\dot{\alpha}}{}_L q^*{}_{R\dot{\alpha}})$$

writing π for the neutral pion field. $\Lambda = M/m$ is still the coupling of the pion to the quarks, as can be seen by rescaling $\pi \to \pi/m$ to give the kinetic term the usual normalization. (The coupling m is known as the "pion decay constant", and is usually denoted f_π. If we include leptons with the quarks, then this coupling also describes the decay of the pion into two leptonic fermions.)

In this case, the (broken) axial U(1) transformations are

$$q_L' = e^{i\theta}q_L, \quad q_R' = e^{i\theta}q_R, \quad \pi' = \pi - 2\theta$$

The corresponding axial current (determined, e.g., by coupling a gauge vector) is

$$J_A^{\alpha\dot{\beta}} = (q^{\dagger\dot{\beta}}{}_L q_L^\alpha - q_R^{T\alpha} q^{*\dot{\beta}}{}_R) - m^2\partial^{\alpha\dot{\beta}}\pi$$

This current is still conserved, since the field equations aren't changed by the properties of the vacuum. The linear term is characteristic of expanding the Goldstone field about the spontaneously broken vacuum; it corresponds to the fact that that field has an inhomogeneous transformation under the broken symmetry.

However, in reality the pion is not exactly massless, so we should add to the previous action a mass term for the pion, which explicitly violates the symmetry. (It is then a "pseudogoldstone boson".) In the general chiral symmetry model, where the

Goldstone bosons are described by a unitary matrix, a simple term that gives them masses while preserving the polar (parity-preserving) diagonal symmetry $U_L = U_R$ of the vacuum is, for some constant ξ,

$$L_m = -\xi \; tr(\phi + \phi^\dagger - \sqrt{2} m I)$$

Since this explicitly breaks the axial U(m) symmetries, the corresponding currents are no longer conserved. In the U(1) case, we can also add just a mass term

$$L_m = \tfrac{1}{4}\zeta\pi^2, \qquad \zeta = m^2 m_\pi^2$$

(for some constant ζ), which is the leading contribution from the general term above. The change in the field equation for π now violates the conservation law as

$$\partial \cdot J_A = -\zeta\pi$$

This explicitly broken conservation law is known as "Partially Conserved Axial Current" (PCAC).

5. Stückelberg

By definition, only gauge-invariant variables are observable. Although in general a change of variables to gauge-invariant ones can be complicated and impractical, there are certain theories where such a procedure can be implemented very simply as part of the normal gauge-fixing. Not surprisingly, the only nonlinearity in these redefinitions involves scalars.

The simplest cases of such redefinitions are free theories, and are thus contained in our earlier discussion of general free, massive gauge theories. The simplest of these is the massive vector. As described in subsection IIB4, the Lagrangian and gauge invariance are

$$L = \tfrac{1}{8}F^2 + \tfrac{1}{4}(mA + \partial\phi)^2$$

$$\delta A = -\partial\lambda, \qquad \delta\phi = m\lambda$$

where F_{ab} is the Abelian field strength. Note that the scalar is pure gauge: It is called a "compensator" for this gauge invariance. Since it has a nonderivative gauge transformation, it can easily be gauged to zero at each point, by just choosing $\lambda = -\phi/m$. This means that without loss of generality we can consider the theory in terms of just the gauge-invariant field

$$A' = A + \tfrac{1}{m}\partial\phi$$

This "composite" field can also be considered as a field redefintion or gauge transformation on A. The lagrangian simplifies to

$$L = \tfrac{1}{8}F'^2 + \tfrac{1}{4}m^2 A'^2$$

Later we'll see that it is often more useful to keep ϕ as an independent field.

Exercise IVA5.1

Choose the gauge $A^0 = \phi$. Show that ϕ then can be eliminated by its equation of motion, leaving only the transverse 3-vector A^i, with Lagrangian $-\frac{1}{4}A^i(\Box - m^2)A^i$. Show the relation to the lightcone gauge of subsection IIIC2, using the dimensional reduction langauge of subsection IIB4.

The original Lagrangian can also be considered an unusual coupling of a massless vector to a massless scalar: Remember that the massless scalar is the simplest example of a Goldstone boson, with the spontaneously broken global symmetry

$$\delta\phi = \epsilon T\phi, \qquad T\phi = 1$$

where we have defined the symmetry generator T to act inhomogeneously on ϕ. We then couple the "photon" to this charge: After a trivial rescaling of the gauge field,

$$L = \tfrac{1}{8m^2}F^2 + \tfrac{1}{4}(\nabla\phi)^2, \qquad \nabla = \partial + AT$$

where m is the "charge" with which A couples to ϕ, which in this case happens to have dimensions of mass. The electromagnetic current in this case is simply $J = \frac{1}{2}\nabla\phi$, whose conservation is the scalar field equation $\Box\phi = 0$ (with gauge-covariantized $\Box$).

Because the spontaneously broken symmetry of the corresponding Goldstone model is now gauged, expanding about $\langle\phi\rangle = 0$ is no longer a physical statement about the vacuum, since ϕ is no longer gauge invariant. (As we saw, we can even choose $\phi = 0$ as a gauge condition.) Therefore, from now on, when we make a statement such as "$\langle\phi\rangle = 0$" in such a case, it will be understood to refer to choosing $\phi = 0$ as the value about which to perform perturbation expansions (e.g., for separating actions into kinetic terms and interactions).

Note that the Stückelberg action can be generated starting from the action with just A', and performing a gauge transformation that is not an invariance:

$$A' \to A' + \tfrac{1}{m}\partial\phi$$

Dropping the prime from A, this transformation is just the inverse of the one we used to eliminate the scalar. If we start from an action that has also a coupling of A' to matter, we see that conserved currents decouple from ϕ:

$$\int A' \cdot J \to \int A \cdot J - \tfrac{1}{m}\int \phi\partial \cdot J$$

More precisely, if the *only* term in the action for vector + matter that is not gauge invariant is the vector mass term ($\frac{1}{4}m^2A'^2$), then the above gauge transformation affects only that term.

6. Higgs

We have seen that spontaneous symmetry breakdown can generate masses for spinors. We also saw how a massless vector could become massive by "eating" a would-be Goldstone scalar, in the simplest case of a scalar without self-interactions. We'll now examine more interesting models: Yang-Mills theories, which describe self-interacting vectors, must couple to self-interacting scalars to become massive.

We can expect, by considering the linearization of any Yang-Mills theory coupled to scalars, that we will need more scalars than massive vectors, since each vector needs to eat a scalar to become massive, and some scalars will become massive and thus uneaten. (Only would-be Goldstone bosons can be eaten, as seen by linearization to the Stückelberg model.) For the simple (and most useful) example of U(n) for the gauge group, an obvious choice for the scalar "Higgs" field is an n⊗n matrix. (SU(n) can be treated as a slight modification.) The simplest such model is the one studied in subsection IVA2: We now consider one of the SU(n) symmetries (together with the U(1)) as the local "color" symmetry to which the Yang-Mills fields couple, and the other SU(n) as the global "flavor" symmetry (where we use the names "color" and "flavor" to distinguish local and global symmetries, not necessarily related to chromodynamics).

The Lagrangian for this "Gervais-Neveu model" is then

$$L = tr[\tfrac{1}{8g^2}F^2 + \tfrac{1}{2}(\nabla\phi)^\dagger\cdot(\nabla\phi) + \tfrac{1}{4}\lambda^2(\phi^\dagger\phi - \tfrac{1}{2}m^2I)^2]$$

where $\nabla = \partial + iA$, and A_a and ϕ are n⊗n matrices (but A_a are hermitian). Now $\phi^\dagger\phi$ is gauge invariant (although not invariant under the flavor group), so we still have

$$\langle\phi^\dagger\phi\rangle = \tfrac{1}{2}m^2I$$

as a gauge-invariant statement (but $\langle\phi\rangle = \frac{m}{\sqrt{2}}I$, or $\langle\phi\phi^\dagger\rangle = \frac{1}{2}m^2I$, still makes sense only for purposes of gauge-dependent perturbation expansions).

Since any complex matrix can be written as $\phi = UH/\sqrt{2}$, where U is unitary and H is hermitian, we can choose the "unitary gauge" $U = I$ (i.e., $\phi = \phi^\dagger$). As for the Stückelberg case, this is equivalent to working in terms of the gauge-invariant fields (defined by using this U as a gauge transformation)

$$A' = U^{-1}(-i\partial + A)U, \quad \phi' = \tfrac{1}{\sqrt{2}}H = U^{-1}\phi$$

where U can be defined by

$$\tfrac{1}{\sqrt{2}}H = \sqrt{\phi^\dagger\phi}, \quad U = \phi\sqrt{2}H^{-1}$$

This is well-defined as long as H is invertible, which is true for small perturbations about its vacuum value

$$\langle H\rangle = mI$$

If the perturbation is so large that H has vanishing eigenvalues, then this is equivalent to looking at states so far away from the vacuum that some of the broken symmetry is restored. Expanding about the vacuum ($H \to mI + H$), the Lagrangian is now

$$L = tr[\tfrac{1}{8g^2}F'^2 + \tfrac{1}{4}m^2A'^2 + \tfrac{1}{4}(\partial H)^2 + \tfrac{1}{4}\lambda^2m^2H^2$$

$$+A'\cdot\tfrac{1}{4}(Hi\overset{\leftrightarrow}{\partial}H) + \tfrac{1}{2}mA'^2H + \tfrac{1}{4}\lambda^2mH^3 + \tfrac{1}{4}A'^2H^2 + \tfrac{1}{16}\lambda^2H^4]$$

Thus all particles are now massive. As for the Goldstone case, we can take the limit $\lambda \to \infty$ to get rid of all the massive scalars, which in this case leaves just the massive vectors, adding only the mass term to the original Yang-Mills action. This was clear from the nonlinear σ model that resulted from that limit, by coupling that field (U) to Yang-Mills directly.

Exercise IVA6.1

Find the chiral action for this model of the type described in subsection IIIC4, where the massive vectors are described by self-dual tensors instead of vectors.

Exercise IVA6.2

Consider again this model, for the case n=2. We modify this example by dropping the U(1) gauge field, so we have just SU(2). Since SU(2) is pseudoreal, we can further restrict the Higgs field to satisfy the reality condition $\phi^* = C\phi C$. Thus, both color and flavor groups are SU(2), and ϕ is the usual matrix representation of the 4-vector of SO(4)=SU(2)⊗SU(2) (see subsection IIA5). Repeat the analysis given above.

Exercise IVA6.3

Consider again the gauge group SU(2), but now take the Higgs field in the *adjoint* representation, with no flavor group (i.e., a real 3-vector). Show that only 2 of the 3 vectors get mass, leaving a residual U(1) gauge invariance. Explain this in terms of the gauge transformations of the 3-vector. (Hint: think 3D rotations.)

REFERENCES

1 Y. Nambu, *Phys. Rev. Lett.* **4** (1960) 380;
Y. Nambu and G. Jona-Lasinio, *Phys. Rev.* **122** (1961) 345, **124** (1961) 246:
introduced into relativistic physics massless bosons associated with a broken symmetry.

2 J. Goldstone, *Nuo. Cim.* **19** (1961) 154;
J. Goldstone, A. Salam, and S. Weinberg, *Phys. Rev.* **127** (1962) 965:
found the theorem relating the two.

3 M. Gell-Mann and M. Lévy, *Nuo. Cim.* **16** (1960) 705:
σ-models.

4 S. Weinberg, *Phys. Rev. Lett.* **18** (1967) 188:
nonlinear σ-models.

5 S. Coleman, J. Wess, and B. Zumino, *Phys. Rev.* **177** (1969) 2239,
C.G. Callan, S. Coleman, J. Wess, and B. Zumino, *Phys. Rev.* **177** (1969) 2247;
A. D'Adda, M. Lüscher, and P. Di Vecchia, *Phys. Rep.* **49C** (1979) 239;
E. Cremmer and B. Julia, *Phys. Lett.* **80B** (1978) 48, *Nucl. Phys.* **B159** (1979) 141;
A.P. Balachandran, A. Stern, and G. Trahern, *Phys. Rev.* **D19** (1979) 2416:
sigma models on coset spaces.

6 H. Eichenherr, *Nucl. Phys.* **B146** (1978) 215;
V.L. Golo and A.M. Perelomov, *Lett. Math. Phys.* **2** (1978) 477, *Phys. Lett.* **79B** (1978) 112;
E. Cremmer and J. Scherk, *Phys. Lett.* **74B** (1978) 341:
classical CP(n).

7 T.H.R. Skyrme, *Proc. Roy. Soc.* **A260** (1961) 227;
J. Wess and B. Zumino, *Phys. Lett.* **37B** (1971) 95;
E. Witten, *Nucl. Phys.* **B223** (1983) 422, 433:
higher-derivative terms.

8 H. Yukawa, *Proc. Phys.-Math. Soc. Japan* **17** (1935) 48.

9 R. Marshak and E.C.G. Sudarshan, *Proc. Padua-Venice conference on mesons and recently discovered particles*, September, 1957; *Phys. Rev.* **109** (1958) 1860;
Feynman and Gell-Mann, *loc. cit.* (IIIC):
chiral symmetry in weak interactions (for coupling to vector currents).

10 Gell-Mann and Levy, *loc. cit.*;
Nambu, *loc. cit.*;
K.-C. Chou, *Sov. Phys. JETP* **12** (1961) 492:
PCAC.

11 Stückelberg, *loc. cit.* (IIB).

12 P.W. Anderson, *Phys. Rev.* **112** (1958) 1900, **130** (1963) 439:
nonrelativistic "Higgs" effect in condensed matter theory.

13 F. Englert and R. Brout, *Phys. Rev. Lett.* **13** (1964) 321;
P.W. Higgs, *Phys. Lett.* **12** (1964) 132;
G.S. Guralnik, C.R. Hagen, and T.W.B. Kibble, *Phys. Rev. Lett.* **13** (1964) 585;
A.A. Migdal and A.M. Polyakov, *Sov. Phys. JETP* **24** (1967) 91:
"Higgs" effect.

14 J.L. Gervais and A. Neveu, *Nucl. Phys.* **B46** (1972) 381.

B. STANDARD MODEL

In this section we discuss the "Standard Model", the minimal theory that describes all the observed particles and forces (except gravity). We also consider some features of "Grand Unified Theories" (GUTs), an extension of the Standard Model that uses fewer multiplets.

1. Chromodynamics

One way in which particles naively described by the action can be hidden from observation is if the force binding them is too strong to allow them to exist freely. Such a condition is often called "infrared slavery" since this alleged property of the force is a long-range feature, preventing the constituent particles from escaping to infinity. This "confinement" is not a classical phenomenon, and its occurence even at the quantum level has not yet been proven. Therefore, in this section we'll simply assume confinement, and describe the resultant symmetry properties, leaving the dynamical properties for later chapters.

The assumption of "color" confinement is that the color forces are so strong that they bind any objects of color to other such objects; thus, only "colorless" states, those that are singlets under the color gauge group, can exist freely. Composite fields that are invariant under the local group transformations can be obtained by multiplying matter fields or Yang-Mills field strengths, perhaps using also covariant derivatives, and contracting all color indices. The color gauge group is generally assumed to be SU(n): usually SU(3), but sometimes larger n for purposes of perturbation in 1/n. Larger n is also used for unification, but in that case the Higgs mechanism is used to reduce the group of the massless vectors to SU(3) (times Abelian factors).

Another feature of these confined states, to be considered later, is their geometrical structure. The observed spectrum and scattering amplitudes of the "hadrons" (strongly interacting particles) indicates a stringlike identification of at least the excited states. (The ground states may behave more like "bags".) This picture also fits in with confinement, since the spatial separation of the quarks and antiquarks in excited states would force the gluons that convey their interactions (and self-interact) to confine themselves as much as possible by collapsing into "strings" connecting the quarks. Thus, we describe a meson with an "open string", with a quark at one end and an antiquark at the other. Similarly, an excited glueball would no longer be a ball, but rather a "closed string", forming a closed loop.

We will need to reconsider also the discrete symmetries, C, P, and T, and their combinations. First of all, we note the "CPT theorem": All local, hermitian, Poincaré invariant actions are CPT invariant. This is easy to see from the fact that CPT only changes the overall sign of the coordinates, which is effectively the same as changing the sign of each derivative, as well as giving a -1 for each vector index on a field. Since CPT also gives signs to dotted spinors and not undotted ones, we also get -1's for vector combinations of indices on spinors ($\psi^\alpha \bar{\psi}^{\dot{\alpha}}$; signs cancel when contracting spinor indices on pairs of dotted or undotted spinors). Thus, all these signs cancel because Poincaré invariance requires an even number of vector indices (in even numbers of dimensions, from contracting with η_{ab} and ϵ_{abcd}). Alternatively, and even more simply, in D=4 we can attribute it to having even numbers both of undotted spinor indices and of dotted spinor indices, since we can define CPT by associating a sign with each dotted index (including those that appear as part of a vector index). Consequently, from now on we ignore T and consider only C, P, and CP.

Although we have considered C (and thus CP) in the context of electromagnetism, nonabelian gauge fields require some (simple) generalization, since they carry charge themselves. We start with the general coupling of massless fermions to nonabelian gauge fields:

$$L = \psi^{\dagger\dot{\beta}}(-i\partial_{\alpha\dot{\beta}} + A_{\alpha\dot{\beta}})\psi^\alpha$$

where ψ is a column vector with respect to the gauge group, and A a hermitian matrix. The CP transformation of the fermions then determines that of the vectors, needed for invariance:

$$CP: \qquad \psi'^\alpha = \psi^*{}_{\dot{\alpha}}, \qquad \psi'^{*\dot{\alpha}} = -\psi_\alpha, \qquad \partial'_{\alpha\dot{\beta}} = -\partial^{\beta\dot{\alpha}}, \qquad A'_{\alpha\dot{\beta}} = A^{T\beta\dot{\alpha}}$$

(remember $(\psi^\alpha)^* \equiv \bar{\psi}^{\dot{\alpha}}$, but $(\psi_\alpha)^* = -\bar{\psi}_{\dot{\alpha}}$ because of the factor of $C_{\alpha\beta}$), where we have chosen to represent parity on the coordinates as acting on the explicit ∂ rather than on the arguments of the fields. The transformation on the vector is thus parity on the vector index, combined with charge conjugation $A'_a = -A_a^T = -A_a{}^*$: The minus sign can be associated with change in sign of the coupling (as for the Abelian case), while the complex conjugation takes into account the charge of the vector fields themselves. (As discussed in subsection IB2, $G \to -G^*$ is an invariance of the algebra, where $g \to g^*$ and $g = e^{iG}$.) Although these terms, as well as the F^2 term for the vectors, are CP invariant, this invariance can be broken by coupling to scalars: The Yukawa coupling

$$L_Y = \psi^{T\alpha}\phi\psi_\alpha + h.c.$$

for some matrix ϕ of scalars, would require the CP transformation

$$\phi' = \phi^*$$

(up to perhaps some unitary transformation), but unlike the vectors there is no guarantee that under complex conjugation the matrix $\phi = \phi^i M_i$ for real scalars ϕ^i and constant matrix (Yukawa couplings) M_i will preserve this form, i.e., satisfy

$$\phi'^i M_i = \phi^i M_i{}^*$$

since the matrices M_i can be complex.

The basic assumption of "chromodynamics", or in the quantized version "quantum chromodynamics (QCD)", is that we have a nonabelian gauge theory without fundamental scalars that couple directly (but scalars will show up when we introduce electroweak interactions). Namely, we assume only Yang-Mills for the "color" gauge group SU(n), specifically n=3, with the usual action, minimally coupled to spin-1/2 "quarks" in the defining representation of the color group, which may have masses. (These masses are actually generated by weakly interacting Higgs bosons, whose coupling we consider in the next subsection; for now we include just the resulting mass terms.) Such an action is automatically invariant under CP and T. We furthermore assume invariance under charge conjugation: Just as an irreducible real scalar describes particles that are their own antiparticles, and needs doubling (or complexification) to define charge, an irreducible (massive) spinor cannot describe distinguishable particle and antiparticle. But the quarks come in the defining representation of SU(n), which is complex, and thus requires doubling to define mass terms. Therefore, for every quark field $q_{L\alpha}$ ("L" for "left") we have an "antiquark" field $q_{R\alpha}$ ("R" for "right"), and they transform into each other under charge conjugation, just as a scalar transforms into its complex conjugate. (A spinor can't transform into its complex conjugate under C, since C commutes with spacetime symmetries, like Lorentz transformations.) Besides this doubling, and the n colors of the quarks, we also assume a further multiplicity of m different "flavors" of quarks. Gauge invariance requires the quark masses be independent of color, and C invariance requires the mass terms couple q_L to q_R, but these terms violate an otherwise global U(m)⊗U(m) flavor symmetry.

The action is then of the form

$$tr[\tfrac{1}{8}F^2 + (q^\dagger{}_L i\nabla q_L + q^\dagger{}_R i\nabla q_R) + (\tfrac{M}{\sqrt{2}} q_R^T q_L + h.c.)]$$

where we have written q_L and q_R as matrices with respect to SU(n) color (U_c) and U(m)⊗U(m) flavor (U_{fL} and U_{fR}) such that they transform as

$$q'_L = U_c q_L U_{fL}, \qquad q'_R = U_c{}^* q_R U_{fR}{}^*$$

and thus the covariant derivatives can be written as

$$\nabla_a q_L = (\partial_a + iA_a)q_L, \qquad \nabla_a q_R = (\partial_a - iA_a{}^*)q_R$$

where A_a are hermitian, traceless, n×n matrices. (By definition, charge conjugation takes a representation of an internal symmetry into the complex-conjugate one.)

While the color symmetry is a local symmetry, the flavor symmetry is broken, inducing the transformation on the mass matrix

$$M' = U_{fL} M U_{fR}^{-1}$$

This transformation allows the mass matrix M to be chosen real and diagonal: Any matrix can be written as a hermitian one times a unitary one. A U_{fR} transformation, as a field redefinition, then can be made to cancel the unitary factor in M; then a unitary transformation $U_{fL} = U_{fR}$ can be made to diagonalize M (while keeping it hermitian). These diagonal elements are then simply the masses of the m different quark flavors. The most symmetric case is $M = 0$, which leaves the entire U(m)⊗U(m) chiral symmetry unbroken. (See subsection IVA4.) The least symmetric case is where all the masses are nonzero and unequal, leaving as unbroken only the subgroup U(1)^m, with $U_{fL} = U_{fR}$. (In general, $U_{fL} = U_{fR}$ if all masses are nonvanishing.)

Exercise IVB1.1

Show the most general case is the product of U(N)'s for various subspaces, with 2 U(N)'s for the massless subspace.

Since the above transformation allows M to be diagonalized, in particular it can be made symmetric, which is sufficient to define charge conjugation:

$$C: \qquad q_{L\alpha} \leftrightarrow q_{R\alpha}, \qquad A_a \to -A_a{}^*$$

Furthermore, since M can be chosen not only symmetric but real, in particular it can be made hermitian, which is enough to define parity:

$$P: \qquad q^\alpha_{L,R} \to \bar{q}_{R,L\dot{\alpha}}, \qquad A_a \to -A^a$$

The minimal form of this action, besides making CP and T automatic, also automatically extends the discrete symmetry C to an O(2) symmetry, whose "parity" transformation is C and whose continuous SO(2)=U(1) symmetry is the U(1) part of the U(m) flavor symmetry, which is not broken by the mass term. It corresponds to a charge called "baryon number": Up to an overall normalization factor, it simply counts the number of quarks $q_{L\alpha}, \bar{q}_{R\dot{\alpha}}$ (which form a Dirac spinor) minus the number of "antiquarks" $q_{R\alpha}, \bar{q}_{L\dot{\alpha}}$. However, such an O(2) symmetry can be defined separately for each flavor, since (after M has been diagonalized) the action can be written as a sum of independent terms for each flavor. In particular, each flavor has its own separately conserved quark number. These flavor conservation laws, at the classical level, are broken only by the weak interactions, which we have not included in the above action. (Gravity and electromagnetism do not violate them.)

Since confinement is a quantum effect, the details of hadronic scattering cannot be discussed within classical field theory. However, we saw that low-energy properties of mesons (and similarly for baryons) could be described by effective Lagrangians. The fact that hadrons are made of quarks can be used to obtain a bit more information even at the classical level, especially if the relevant quarks are heavy. (Heavy with respect to what is unfortunately also a question that can be answered only at the quantum level.) For example, in a nonrelativistic approximation, low-energy properties of hadrons can be found from just the quantum numbers, spin-spin interactions, and masses of the quarks, while their velocities are ignored, and the gluons are neglected altogether. In such an approximation, reasonably accurate predictions are made for the masses and magnetic moments of the ground-state hadrons.

Actually, the claim that color nonsinglet states can never be observed needs a bit of stipulation: There may be a "quark-gluon plasma" phase of hadronic matter that can exist only at extremely high temperatures or pressures. Thus, a hypothetical observer during the first moments of the universe might observe "free" quarks and gluons. Similarly, a small enough observer, living inside an individual hadron, might see individual quarks and gluons, since the size of his equipment would be much smaller than what we consider "asymptotic" distances. Conversely, we could consider the possibility of a new chromodynamic force, other than the one responsible for the hadrons of which we are composed, that has a confinement scale that is astronomical (extremely low energy), so that earthly laboratories would fit inside the new "hadrons". Thus, any statement about the observability of color must be a dynamical one, and does not follow as an automatic consequence of the appearance of a nonabelian group: Just as for the Higgs effect, confinement can be repealed under appropriate circumstances, and the observability of color depends on the details of the dynamics, and in particular on the values of the various parameters (momenta and couplings).

2. Electroweak

The weak and electromagnetic interactions are mediated by observed spin-1 particles, some of which have charge and mass. Specifically (see subsection IC4), the massive vectors form a triplet (W^+, W^-, Z), while there is only one massless vector (the photon). This suggests a gauge group of SU(2)⊗U(1), with a Higgs effect that leaves only U(1) unbroken. From the table of known fundamental fermions, we can see that they fall into doublets and singlets of this SU(2), with the U(1) charge being that of electromagnetism. (This SU(2)⊗U(1) unification of the weak and electromagnetic interactions is called the "Glashow-Salam-Weinberg" model.)

We saw in subsection IVA4 a very simple model of spontaneously broken chiral U(m)⊗U(m) symmetry where masses were generated for quarks. In subsection IVA6 we saw how the same scalars could generate masses for vectors, by coupling to one of the U(m)'s. We now combine those two models, specializing to the case m=2, but with two slight modifications: (1) Since the defining representation of SU(2) is pseudoreal, we can impose a reality condition on the Higgs field, which is in the $(\frac{1}{2},\frac{1}{2})$ representation of SU(2)⊗SU(2):

$$\phi^* = C\phi C$$

This makes it a vector of SO(4)=SU(2)⊗SU(2) (See exercises IIA5.3 and IVA6.2.) It's also the reality condition satisfied by an element of (the defining representation of) SU(2). (See subsection IIA2.) This is not surprising, since the group product $U' = U_L U U_R$ allows the interpretation of a group element itself as a representation of chiral symmetry. This is the situation described in subsection IVA2 ($\phi \to U$ in the large-mass limit), but in this case $\phi^\dagger\phi$ is automatically proportional to the identity (it gives the square of the 4-vector), so in general an SO(4) 4-vector can be written as the product of a scalar with an SU(2) element. This reality condition breaks the chiral U(1)⊗U(1) to the diagonal U(1) that leaves the Higgs invariant.

(2) The gauged SU(2) is still one of the two chiral SU(2)'s, but the gauged U(1) must now be a *subgroup* of the other SU(2), since the Higgs is now invariant under the usual U(1)'s. Thus, the ungauged SU(2) is explictly broken, and this accounts for the mass splittings in the doublets of known fundamental fermions. Remember that observables are singlets of gauged nonabelian groups (except perhaps for Abelian subgroups), so any observed internal SU(2) must be a global symmetry, even when it's broken. As described in subsection IVA6, these singlets can be constructed as composite fields resulting from the gauge transformation obtained from the SU(2) part of ϕ.

Using the electromagnetic charges of the various particles, we thus determine their SU(2)⊗U(1) representations: For spin 1, we have W=(1,0) and V=(0,0), where the first entry is the "isospin" and the second is the U(1) charge. For spin 0, we have $\phi=(\frac{1}{2},\pm\frac{1}{2})$, choosing the U(1) generator as the diagonal one from U_R. Finally, for spin 1/2, we have for the leptons $l_L=(\frac{1}{2},-\frac{1}{2})$, which combines with ϕ to produce (0,0)⊕(0,−1), and l_R=(0,1). Similarly, for the quarks we have $q_L=(\frac{1}{2},\frac{1}{6})$, and $q_R=(0,-\frac{1}{6}\pm\frac{1}{2})$. (We use for undotted spinors the convention "L" = fermion, "R" = antifermion.) The Lagrangian is then

$$L = L_1 + L_0 + L_{1/2}$$

$$L_1 = \tfrac{1}{8g'^2}F^2(V) + \tfrac{1}{8g^2}tr\ F^2(W)$$

$$L_0 = tr[\tfrac{1}{4}(\nabla\phi)^\dagger(\nabla\phi) + \tfrac{1}{4}\lambda^2(\phi^\dagger\phi - \tfrac{1}{2}m^2)^2]$$

$$L_{1/2} = tr(\psi^\dagger i\nabla\psi) + tr\left[\begin{pmatrix} \Lambda_+ & 0 \\ 0 & \Lambda_- \end{pmatrix} q_R^T q_L\phi + \Lambda l_R^T l_L\phi\begin{pmatrix}1\\0\end{pmatrix} + h.c.\right]$$

where the fermions $\psi = (q_L, q_R, l_L, l_R)$, and the SU(2)⊗U(1) covariant derivative acts as

$$\nabla\phi = \partial\phi + iW\phi - i\tfrac{1}{2}V\phi\left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right)$$

$$\nabla q_L = \partial q_L - iq_L W + i\tfrac{1}{6}Vq_L$$

$$\nabla q_R = \partial q_R + i\tfrac{1}{2}Vq_R\left[-\tfrac{1}{3}I + \left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right)\right]$$

$$\nabla l_L = \partial l_L - il_L W - i\tfrac{1}{2}Vl_L$$

$$\nabla l_R = \partial l_R + iVl_R$$

(The infinitesimal gauge transformations have the same form, dropping the derivative term and replacing the gauge field with the corresponding gauge parameter.) For simplicity we have ignored the indices for color (and its gauge fields, treated in the previous section), families (treated in the following subsection), and spin. We have also used matrix notation with respect to the local SU(2) (gauged by W) and the global SU(2) (explicitly broken in $L_{1/2}$ by the gauging of a U(1) subgroup, the Yukawa couplings, and the chirality of the massless neutrinos): Thus W is a traceless hermitian 2×2 matrix, ϕ is also 2×2 but satisfying the "reality" condition given above (traceless antihermitian plus real trace), q_L, q_R, and l_L are 2-component rows, and l_R is a single component. (By definition, the diagonal parts of W and ϕ are electromagnetically neutral.) The quark Yukawa coupling is diagonal in the broken SU(2) to preserve the local U(1) symmetry. (The tr here is trivial for the lepton Yukawa term, but we have left it for generalization to more than one family.) Explicitly, we can write

$$W = \begin{pmatrix} \frac{1}{\sqrt{2}}W^0 & W^+ \\ W^- & -\frac{1}{\sqrt{2}}W^0 \end{pmatrix}$$

and for the lightest family (see subsection IC4)

$$q_L = (d_L \; u_L), \qquad q_R = (d_R \; u_R), \qquad l_L = (e_L \; \nu), \qquad l_R = e_R$$

In the unitary gauge for the local SU(2),

$$\phi = \tfrac{1}{\sqrt{2}}\varphi I, \qquad \langle\varphi\rangle = m$$

where φ is a single real scalar, the simplifications to the Lagrangian are

$$L_0 \to \tfrac{1}{4}(\partial\varphi)^2 + \tfrac{1}{8}\varphi^2\, tr\{[W - \tfrac{1}{2}V\left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right)]^2\} + \tfrac{1}{8}\lambda^2(\varphi^2 - m^2)^2$$

$$L_{1/2} \to tr(\psi^\dagger i\nabla\psi) + \tfrac{1}{\sqrt{2}}\varphi\; tr\left[\begin{pmatrix} \Lambda_+ & 0 \\ 0 & \Lambda_- \end{pmatrix} q_R^T q_L + \Lambda l_R^T l_L \begin{pmatrix} 1 \\ 0 \end{pmatrix} + h.c.\right]$$

We then can expand φ about its vacuum value m: The lowest order terms give masses for most of the vectors and fermions: The massless fermions are the neutrinos, while the massless vector gauging the unbroken U(1) (a combination of the original U(1) with a U(1) subgroup of the SU(2)) is the photon (of electromagnetic fame). The mass of the remaining vectors accounts for the weakness and short range of the "weak" interactions.

Exercise IVB2.1

Diagonalize this Lagrangian with respect to the mass eigenstates. For convenience, normalize

$$g = \tfrac{1}{\sqrt{2}}g_0 \cos\,\theta_W, \quad g' = g_0 \sin\,\theta_W$$

where θ_W is the "weak mixing (Weinberg) angle".

a Find explicitly the masses for all the particles in the Standard Model (first family for fermions) in terms of the couplings $m, \lambda, g_0, \theta_W, \Lambda_\pm, \Lambda$. Show from the experimental values for the vector masses given in subsection IC4 that $sin^2\theta_W \approx .223$.

b Find all the other couplings of the mass eigenstates. Show that, with the conventional electric charge assignments,

$$\frac{1}{e^2} = \frac{1}{2g^2} + \frac{1}{g'^2}$$

(Hint: Rather than rescaling the vectors, note that the generated mass term, and the given couplings of V and W, suggest defining

$$W' = W - \tfrac{1}{2}V\left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right) \quad \Rightarrow \quad V = \gamma + k_1 Z, \qquad \sqrt{2}W^0 = \gamma + k_2 Z$$

in the conventions of subsection IIA1, for the new fields Z and photon γ, for appropriate constants k_i.)

Note that, unlike the strong (chromodynamic) or purely electromagnetic (or even gravitational) interactions, the weak interactions break every discrete spacetime symmetry possible. (The others break none. CP violation will be discussed in the following subsection. Of course, CPT is always preserved.) Sometimes this is attributed to the presence of a chiral symmetry, used to reduce 4-component spinors to 2-component; however, we have already seen that in general chiral and parity symmetries are unrelated. (You can have either without the other. This fact will be further discussed in subsections IVB4 and VIIIB3.) A better explanation is to attribute P and C to doubling, which converts spinors from 2-component to 4-component: 2-component spinors are the simplest description of helicity/spin $\frac{1}{2}$; 4-component spinors are useful only to manifest a larger symmetry, when it exists. The weak interactions violate parity because the neutrino is not doubled, and because the fermions that are doubled no longer have a symmetry relating their two halves.

3. Families

In the Standard Model (and its simpler generalizations) there is no explanation for the existence of more than one family of fermions. However, the existence of 3 families does have interesting consequences. Most of these follow from the form of the Yukawa couplings, and thus the fermion masses. In subsection IVB1 we considered redefinitions of the fermion fields as unitary flavor transformations. These allowed us to obtain the simplest form of the mass matrices, since they were not flavor singlets, and thus transformed. We now perform similar transformations, but only on the family indices, since transformations that don't commute with the gauge symmetries would complicate the other terms in the action. Now ignoring spin, color, and local flavor indices, and using matrix notation for the family indices, the fermions transform as

$$q_L' = q_L U_{qL}, \quad q_{R\pm}' = q_{R\pm} U_{qR\pm}{}^*, \quad l_L' = l_L U_{lL}, \quad l_R' = l_R U_{lR}{}^*$$

where q_L, $q_{R\pm}$, l_L, and l_R have m components for the m families. $q_{R\pm}$ are the 2 components of the (explicitly broken) *global* flavor doublet q_R. We thus have 5 U(m) symmetries, all broken by the Yukawa couplings: These field redefinitions induce transformations on them,

$$\Lambda'_\pm = U_{qL}\Lambda_\pm U_{qR\pm}^{-1}, \qquad \Lambda' = U_{lL}\Lambda U_{lR}^{-1}$$

As in subsection IVB1, $U_{qR\pm}$ and U_{lR} can be used to make $\Lambda_\pm$ and Λ hermitian. Then U_{lL} can be used to make Λ diagonal, also as in subsection IVB1, leaving a $\mathrm{U}(1)^m$ symmetry $U_{lL} = U_{lR}$, corresponding to separate conservation laws for electron number (including its neutrino), muon number, and tauon number for the 3 known flavors. However, the quark sector works a bit differently: We can use U_{qL} to diagonalize Λ_+ or Λ_-, but not both. This leaves another $\mathrm{U}(1)^m$ symmetry $U_{qL} = U_{qR+} = U_{qR-}$. If Λ_+ has been diagonalized, then 1 of the m U(1)'s, corresponding to total quark number (baryon number) conservation, leaves Λ_- invariant, while the remaining m−1 U(1)'s can be used to eliminate some of the phases of the complex off-diagonal components of Λ_-.

The remaining global flavor symmetries are thus m lepton U(1)'s and 1 quark U(1). The remaining Yukawa couplings are the real, diagonal Λ, describing the m masses of the massive leptons (the neutrinos remain massless), the real, diagonal Λ_+, giving the m masses of half of the quarks, and the hermitian Λ_-, consisting of m diagonal components, describing the masses of the other quarks, m(m−1)/2 magnitudes of the off-diagonal components, and (m−1)(m−2)/2 phases of the off-diagonal components. These phases violate CP invariance: CP, besides its affect on the coordinates, switches each spinor field with its complex conjugate. Since the complex conjugate term in the action uses the complex conjugates of the Λ's, this symmetry is violated whenever any of the components have imaginary parts (after taking into account all possible symmetries that could compensate for this, as we have just done). Note that CP is violated only for 3 families or more. (C and P are separately violated for any number of families by the SU(2)⊗U(1) coupling: As discussed in subsection IVB1, C invariance of the strong interactions is the symmetry $q_L \leftrightarrow q_R$.) Since we can choose to transform away the phases in the subsector of the 2 lighter quark families, the large masses of the heavier quarks suppress this effect, accounting for the smallness of CP violation.

Since observed particles are mass eigenstates, it's convenient to perform a further unitary transformation (the "Cabibbo-Kobayashi-Maskawa matrix") that diagonalizes the mass matrix. Although this is clearly possible by the arguments of subsection IVB1, it is not part of the unitary transformations considered in this subsection because it does not commute with the SU(2) gauge symmetry: After such a transformation, we find that the components of each SU(2) quark multiplet are linear superpositions of different families.

Exercise IVB3.1

Perform this diagonalization explicitly for the case m=2 (two families), using the two lightest families of quarks and leptons as listed in subsection IC4. Which particles mix? Parametrize this mixing by an angle θ_c (the "Cabibbo angle").

An important experimental result with which the Standard Model is consistent is the suppression of "flavor-changing neutral currents (FCNC)". The two electrically neutral gauge fields in this model, the Z and the photon γ, couple to currents that are neutral with respect to the U(1) symmetries associated with each of the quark (flavor) numbers. This is true by construction before the unitary CKM transformation, but this transformation also leaves these two currents invariant (the "Glashow-Iliopoulos-Maiani mechanism"). Thus, at the classical level we do not see effects such as the decay $K^0 \to Z \to \mu^+\mu^-$, which would violate this "conservation law". Furthermore, the quantum corrections are suppressed (though nonvanishing) for similar reasons: For example, the lowest-order nonvanishing quantum correction comes from replacing the Z with a W^+W^- pair. Without the CKM matrix, this contribution would vanish; treating CKM, and its resulting contribution to quark masses, as a perturbtation, the resulting contribution is suppressed by a factor of m_q^2/m_W^2. The absence of FCNC is an important constraint on generalizations of the Standard Model.

4. Grand Unified Theories

The Standard Model gives a description of the weak and electromagnetic interactions that describes the spin-1 particles in terms of gauge fields, and accounts for all masses by the Higgs effect. However, it does not give any unification, in the sense that we still have 3 groups (SU(3), SU(2), and U(1)) for 3 interactions (strong, weak, and electromagnetic), and a large variety of spin-1/2 fields that are unrelated except by color and broken SU(2) flavor. Grand Unified Theories unify this symmetry by forcing all 3 gauge groups to be subgroups of a simple group, which is broken to SU(3)⊗SU(2)⊗U(1) by Higgs (and then broken to SU(3)⊗U(1) by more Higgs). This means introducing new spin-1 particles that are unobserved so far because of their very large masses. On the other hand, the known fermions are then grouped together in a small number of multiplets without introducing new fermions (except perhaps partners for the neutrinos to allow them to have small masses). Unfortunately, this requires a more complicated (and ambiguous) Higgs sector, with separate spin-0 multiplets and couplings for first breaking to SU(3)⊗SU(2)⊗U(1) and then breaking to SU(3)⊗U(1); we won't discuss those Higgs fields here.

The simplest such model uses the group SU(5). Recall the SU(3)⊗SU(2)⊗U(1) representations of each family of fermions:

$$q_L = (3, \tfrac{1}{2}, \tfrac{1}{6}), \quad q_{R+} = (\bar{3}, 0, \tfrac{1}{3}), \quad q_{R-} = (\bar{3}, 0, -\tfrac{2}{3}), \quad l_L = (1, \tfrac{1}{2}, -\tfrac{1}{2}), \quad l_R = (1, 0, 1)$$

where the first argument is the dimension of the SU(3) representation ($\bar{3}$ being the complex conjugate of the 3), the second is the SU(2) isospin, and the third is the U(1) charge. An SU(3)⊗SU(2)⊗U(1) subgroup of SU(5) can be found easily by taking the 5-component defining representation and picking 3 components as the defining representation of SU(3) and the other 2 for that of SU(2): I.e., consider a traceless hermitian 5×5 matrix as an element of the SU(5) Lie algebra, and take

$$(\, SU(5)\,) \to \begin{pmatrix} SU(3) - \frac{1}{3}I \times U(1) & 0 \\ 0 & SU(2) + \frac{1}{2}I \times U(1) \end{pmatrix}$$

or in other words

$$5 \to (3, 0, -\tfrac{1}{3}) \oplus (1, \tfrac{1}{2}, \tfrac{1}{2})$$

From this we recognize the fermions as falling into a $\bar{5} \oplus 10$, where the 10 is the antisymmetric product of two 5's, which consists of the antisymmetric product of the two 3's (a $\bar{3}$), the antisymmetric product of the two SU(2) doublets, and the product of one of each:

$$\bar{5} \to (\bar{3}, 0, \tfrac{1}{3}) \oplus (1, \tfrac{1}{2}, -\tfrac{1}{2}) = q_{R+} \oplus l_L$$

$$10 \to (\bar{3}, 0, -\tfrac{2}{3}) \oplus (1, 0, 1) \oplus (3, \tfrac{1}{2}, \tfrac{1}{6}) = q_{R-} \oplus l_R \oplus q_L$$

Exercise IVB4.1

Find the symmetric product of 2 5's, and its decomposition into representations of SU(3)⊗SU(2)⊗U(1).

A more unifying model is based on SO(10). A U(5) subgroup can be found from the spinor representation by dividing up the set of 10 Dirac γ matrices into two halves, and taking complex combinations to get 5 sets of anticommuting creation and annihilation operators. (See exercise IC1.2.) The Dirac spinor is then

$$(1, -\tfrac{5}{2}) \oplus (5, -\tfrac{3}{2}) \oplus (10, -\tfrac{1}{2}) \oplus (\overline{10}, \tfrac{1}{2}) \oplus (\bar{5}, \tfrac{3}{2}) \oplus (1, \tfrac{5}{2})$$

in terms of the SU(5) representation and the U(1) charge. This Dirac spinor is reducible into Weyl spinors $16 \oplus \overline{16}$; in fact, $i\sqrt{2}\gamma_{-1}$ is just $(-1)^{Y+1/2}$ in terms of the

U(1) charge Y. (The SO(10) generators are even in oscillators, and thus do not mix even levels with odd.) We then have

$$16 \to (1, -\tfrac{5}{2}) \oplus (10, -\tfrac{1}{2}) \oplus (\bar{5}, \tfrac{3}{2})$$

Ignoring the U(1) charge, these are the multiplets found for each family in the SU(5) GUT, plus an extra singlet.

A simple way to understand this extra singlet is to look at a different path of breaking to SU(3)⊗SU(2)⊗U(1): Looking at the vector (defining) representation of SO(10), we can break it up as 6+4 (in the same way we broke up the 5 of SU(5) as 3+2) to get the subgroup SO(6)⊗SO(4)=SU(4)⊗SU(2)⊗SU(2). We can also see that a Dirac spinor of SO(10) ($16\oplus\overline{16}$) will be a Dirac spinor of SO(6) ($4\oplus\bar{4}$) *times* (not plus) a Dirac spinor of SO(4), while the Dirac spinor of SO(4) is a defining representation of one SU(2) (($\frac{1}{2}$,0)) *plus* a defining representation of the other SU(2) ((0,$\frac{1}{2}$)). Thus,

$$16 \to (4, \tfrac{1}{2}, 0) \oplus (\bar{4}, 0, \tfrac{1}{2})$$

where we have used the fact that γ_{-1} (used for projection to Weyl spinors) of SO(10) is proportional to the product of all the γ-matrices, and thus the product of γ_{-1}'s for SO(6) and SO(4).

Looking at this model ("Pati-Salam model") as an alternative to SU(5) (but with a semisimple, rather than simple, group, so it unifies only spin 1/2, not spin 1), we now look at breaking SU(4)→U(3)= SU(3)⊗U(1) (using 4=3+1, as we did 5=3+2 for SU(5)), and breaking one SU(2)→ U(1). We then find

$$(4, \tfrac{1}{2}, 0) \to (3, -\tfrac{1}{3}, \tfrac{1}{2}, 0) \oplus (1, 1, \tfrac{1}{2}, 0) = q_L \oplus l_L$$

$$(\bar{4}, 0, \tfrac{1}{2}) \to (\bar{3}, \tfrac{1}{3}, 0, \tfrac{1}{2}) \oplus (\bar{3}, \tfrac{1}{3}, 0, -\tfrac{1}{2}) \oplus (1, -1, 0, \tfrac{1}{2}) \oplus (1, -1, 0, -\tfrac{1}{2})$$
$$= q_{R+} \oplus q_{R-} \oplus l_R \oplus l_{R-}$$

where the arguments are the SU(3) representation, the U(1) charge from SU(4), the SU(2) isospin, and the U(1) charge from the broken SU(2). If we choose the U(1) charge of SU(3)⊗SU(2)⊗U(1) as $-1/2$ times the former of these two U(1) charges plus 1 times the latter, this agrees with the result obtained by way of SU(5). However, we now see that all the left-handed fermions are contained within one SU(4)⊗SU(2)⊗SU(2) multiplet, and the right-handed within another, but with a partner for the neutrino. Also, one of the SU(2)'s is that of SU(3)⊗SU(2)⊗U(1), while the other is the other SU(2) of the Standard Model, which was broken explicitly there to U(1), whereas here it is broken spontaneously. Thus, there is a local chiral SU(2)⊗SU(2) flavor symmetry.

$$
\begin{array}{ccc}
 & \mathrm{SO}(10) & \\
\swarrow & & \searrow \\
\mathrm{SU}(5) & & \mathrm{SU}(4)\otimes\mathrm{SU}(2)_L\otimes\mathrm{SU}(2)_R \\
\searrow & & \swarrow \quad \approx 10^{16}\ \mathrm{GeV?} \\
 & \mathrm{SU}(3)\otimes\mathrm{SU}(2)_L\otimes\mathrm{U}(1)_R & \\
 & \downarrow \quad \approx 100\ \mathrm{GeV} & \\
 & \mathrm{SU}(3)\otimes\mathrm{U}(1) &
\end{array}
$$

Furthermore, the SU(4)⊗SU(2)⊗SU(2) model is invariant under C: In general, C is just a permutation symmetry. In this case, it simply switches the two multiplets of each family,

$$C: \qquad (4, \tfrac{1}{2}, 0) \leftrightarrow (\bar{4}, 0, \tfrac{1}{2})$$

Combining with the usual CP, this model is thus also invariant under P:

$$P: \qquad (4, \tfrac{1}{2}, 0) \leftrightarrow (\bar{4}, 0, \tfrac{1}{2})^* \qquad i.e., \qquad \psi^\alpha(4, \tfrac{1}{2}, 0) \leftrightarrow \bar{\psi}_{\dot{\alpha}}(4, 0, \tfrac{1}{2})$$

But both C and P are broken spontaneously on reduction to the Standard Model. However, SO(10) lacks C and P invariance (contrary to some statements in the literature), since there is only a single complex representation for each family of fermions (and thus no nontrivial C; of course, there is still CP, at least for the vector-spinor coupling, as always). In fact, the C of SU(4)⊗SU(2)⊗SU(2) is just an SO(10) transformation: Although SO(10) is not O(10) (which is why it lacks a C), it still includes reflections in an *even* number of "axes", since reflection in any pair of axes is a π rotation (just as for SO(2)). Thus, breaking $10 \to 6 + 4$ includes not only SO(6)⊗SO(4), but also the reflection of an odd number of the "6" axes together with an odd number of the "4" axes — a combined "parity" of both SO(6) and SO(4). (They are all the same up to continuous SO(6)⊗SO(4) transformations.) This parity of the internal space is the C given above. (We saw a similar situation for O(2) in subsection IVB1.)

The identification of C is somewhat semantic in a nonabelian gauge theory (except for unbroken U(1) subgroups), since it is defined by changes in sign of unobserved charges: The C appearing above at an intermediate stage of breaking of the SO(10) GUT originates as a global symmetry of only the Higgs sector, leaving all "fundamental" particles with spin invariant. After breaking to SU(4)⊗SU(2)⊗SU(2), the vectors and the spinors are composites of the original ones and the Higgs responsible for the breaking, so they pick up this symmetry. (In the same way, the spinning particles of the Standard Model pick up the broken global SU(2) of its Higgs.)

Since GUTs unify quarks and leptons, they allow decay of the proton. However, since this requires simultaneous decay of all 3 quarks into 3 leptons, it is an extremely unlikely (i.e, slow) decay, but barely within limits of experiment, depending on the model. Proton decay is still unobserved: This eliminates the simplest version of the SU(5) GUT.

REFERENCES

1 *The rise of the standard model: Particle physics in the 1960s and 1970s*, eds. L. Hoddeson, L. Brown, M. Riordan, and M. Dresden (Cambridge University, 1997): interesting accounts of the development of the Standard Model from many of the people responsible.

2 J.J.J. Kokkedee, *The quark model* (Benjamin, 1969);
J.L. Rosner, Plenary report on recursive spectroscopy (theory), in *Proc. of the XVII Int. Conf. on High Energy Theory*, London, July, 1974, ed. J.R. Smith (Science Research Council, 1974) p. II-171;
A. De Rújula, H. Georgi, and S. Glashow, *Phys. Rev.* **D12** (1975) 147:
nonrelativistic quark model.

3 H. Fritzsch and M. Gell-Mann, *Proc. XVI International Conference on High Energy Physics*, eds. J.D. Jackson and A. Roberts, Chicago and Batavia, Sep. 6-13, 1972 (National Accelerator Laboratory, 1972) v. 2, p. 135;
H. Fritzsch, M. Gell-Mann, and H. Leutwyler, *Phys. Lett.* **47B** (1973) 365;
S. Weinberg, *Phys. Rev. Lett.* **31** (1973) 494, *Phys. Rev.* **D8** (1973) 4482:
QCD, with confinement.

4 W. Pauli, *Niels Bohr and the development of physics* (McGraw-Hill, 1955) p. 30;
G. Lüders, *Ann. Phys.* **2** (1957) 1:
CPT theorem.

5 E. Fermi, *Ric. Scientifica* **4** (1933) 491; *Nuo. Cim.* **II** (1934) 1; *Z. Phys.* **88** (1934) 161:
weak interactions, without vector bosons.

6 T.D. Lee and C.N. Yang, *Phys. Rev.* **104** (1956) 254:
parity violation in weak interactions.

7 S.L. Glashow, *Nuc. Phys.* **22** (1961) 579;
A. Salam and J.C. Ward, *Phys. Lett.* **13** (1964) 168;
S. Weinberg, *Phys. Rev. Lett.* **19** (1967) 1264:
unification of weak and electromagnetic interactions.

8 N. Cabibbo, *Phys. Rev. Lett.* **10** (1963) 531;
M. Kobayashi and K. Maskawa, *Prog. Theor. Phys.* **49** (1972) 282.

9 S.L. Glashow, J. Iliopoulos, and L. Maiani, *Phys. Rev.* **D2** (1970) 1285.

10 J. Pati and A. Salam, *Phys. Rev. Lett.* **31** (1973) 275:
earliest GUT, but semisimple.

11 H. Georgi and S. Glashow, *Phys. Rev. Lett.* **32** (1974) 438:
SU(5) GUT; earliest with simple group.

12 H. Georgi, in *Particles and Fields — 1974*, proc. AIP conference, Division of Particles and Fields, Sep. 5-7, 1974, Williamsburg, ed. E. Carlson (American Institute of Physics, 1975) p. 575;
H. Fritzsch and P. Minkowski, *Ann. Phys.* **93** (1975) 193:
SO(10) GUT.

13 G.G. Ross, *Grand unified theories* (Benjamin/Cummings, 1984).

C. SUPERSYMMETRY

In section IIC we studied some general properties of supersymmetry in arbitrary dimensions, and its representations in D=4. We now consider 4D interactions, by introducing gauge fields defined on superspace, and their actions. A complete discussion of supersymmetry would require (at least) a semester; but here we give more than just an overview, and include the basic tools with examples, which is enough for many applications. Quantum aspects of supersymmetry will be discussed in chapters VI and VIII, supergravity in chapter X, and some aspects of superstrings in chapter XI.

1. Chiral

We first consider some field equations that appear in all free, massless, supersymmetric theories. Of course, since the theory is massless it satisfies the massless Klein-Gordon equation by definition: $\Box\Phi = 0$. From our earlier discussion of general properties of supersymmetry, we also know that $p^{\alpha\dot{\beta}}\bar{q}_{\dot{\beta}}\Phi = p^{\alpha\dot{\beta}}q_\alpha\Phi = 0$. These don't look covariant, but noticing that pq differs from pd only by $\theta\Box$ terms (because of the index contraction), which already vanishes, we have the field equations

$$p^{\alpha\dot{\beta}}\bar{d}_{\dot{\beta}}\Phi = p^{\alpha\dot{\beta}}d_\alpha\Phi = 0$$

These equations imply the Klein-Gordon equation, as seen by hitting them with another d and using the anticommutation relations $\{d_\alpha, \bar{d}_{\dot{\beta}}\} = p_{\alpha\dot{\beta}}$. They imply stronger equations: By evaluating at $\theta = 0$, $d_\alpha\Phi$ yields a spinor component field ψ_α, and we find

$$p^{\alpha\dot{\beta}}\bar{\psi}_{\dot{\beta}} = p^{\alpha\dot{\beta}}\psi_\alpha = 0$$

the usual for massless spin 1/2.

Another equation that can be imposed is the "chirality" condition

$$\bar{d}_{\dot{\alpha}}\phi = 0$$

where ϕ now refers to such a "chiral superfield" (and thus $\bar{\phi}$ to an "antichiral" one, $d_\alpha\bar{\phi} = 0$). This requires that ϕ be complex, otherwise we would also have $d_\alpha\phi = 0$ and thus $p\phi = 0$ by the anticommutation relations. The component expansion is given completely by just the d's and not the $\bar{d}$'s:

$$\phi| = A, \qquad (d_\alpha\phi)| = \psi_\alpha, \qquad (d^2\phi)| = B$$

where A and B are complex scalars, and we use the normalization

$$d^2 = \tfrac{1}{2} d^\alpha d_\alpha$$

All other components are x-derivatives of these, since the $\bar{d}$'s can be pushed past the d's (producing p's) until they annihilate ϕ. Another way to state this is to use the fact

$$d_\alpha = e^{-U/2} \partial_\alpha e^{U/2}, \qquad \bar{d}_{\dot{\alpha}} = e^{U/2} \bar{\partial}_{\dot{\alpha}} e^{-U/2}; \qquad U = \theta^\alpha \bar{\theta}^{\dot{\beta}} p_{\alpha\dot{\beta}}$$

to solve the chirality constraint as

$$\phi(x, \theta, \bar{\theta}) = e^{U/2} \hat{\phi}(x, \theta)$$

where $\hat{\phi}$ is independent of $\bar{\theta}$: It is defined on "chiral superspace". (In this equation U generates a *complex* coordinate transformation.) Another way to solve the chirality constraint is to use the covariant derivatives: Since $d_\alpha d_\beta d_\gamma = 0$ by anticommutativity (and similarly for $\bar{d}$'s),

$$\bar{d}_{\dot{\alpha}} \phi = 0 \quad \Rightarrow \quad \phi = \bar{d}^2 \psi$$

where ψ is a "general" (unconstrained) complex superfield. It is the "prepotential" for the field ϕ.

Exercise IVC1.1

Let's analyze the supersymmetry generators $q_\alpha, \bar{q}_{\dot{\alpha}}$ in this case.

a Find similar expressions for $q, \bar{q}$ in terms of $e^{U/2}$.

b Find $q, \bar{q}$ on $\hat{\phi}$ in terms of just θ and $\partial/\partial\theta$ (no $\bar{\theta}$ nor $\partial/\partial\bar{\theta}$).

Exercise IVC1.2

Show that the prepotential has a gauge invariance, under which ϕ is invariant. (Hint: Use the same identity that led to the prepotential.)

From the anticommutation relations we find

$$[\bar{d}^{\dot{\alpha}}, d^2] = p^{\beta\dot{\alpha}} d_\beta$$

Since this must vanish on ϕ, we find

$$d_\alpha d^2 \phi = \bar{d}_{\dot{\alpha}} d^2 \phi = 0 \quad \Rightarrow \quad p_{\alpha\dot{\beta}} d^2 \phi = 0 \quad \Rightarrow \quad d^2 \phi = constant$$

(We can safely ignore this constant, at least when considering the free theory: It corresponds to a term in the action linear in the fields.) This field equation, together with the chirality constraint, is sufficient to determine the theory: A is the usual free (complex) scalar, ψ_α is the usual free spinor, and B is a constant.

To describe interactions of this ("scalar") multiplet, we keep the chirality condition, since that greatly simplifies the field content of the superfield. In fact, this is clearly the simplest off-shell superfield we can define, since it already has the smallest number of fermions (as do the coordinates of chiral superspace). ("Off shell" means all components less gauge degrees of freedom.) This means that the equation $d^2\phi = 0$ will be generalized, since it implies the Klein-Gordon equation. The simplest way to do this is by constructing an explicit action, our next topic.

2. Actions

The construction of actions in superspace is different from ordinary theories because the geometrically simple objects, the potentials, are constrained, while the unconstrained objects, the prepotentials, can be awkward to work with directly. (This problem is magnified with extended supersymmetry, whose actions we don't consider here.)

We start with the simplest supermultiplet, the chiral superfield. Since chiral superfields are defined on chiral superspace, a natural generalization of a potential (nonderivative) term in the action to superspace is

$$S_1 = \int dx\ d^2\theta\ f(\phi) + h.c.$$

in terms of some function (not functional) f of chiral superfields ϕ (the "superpotential"). We can ignore any $\bar{\theta}$ dependence because it contributes only total derivatives:

$$\phi = e^{U/2}\hat{\phi}(x,\theta) \quad \Rightarrow \quad f(\phi) = e^{U/2}f(\hat{\phi})$$

Integration over θ is defined as in subsection IA2; however, now we can replace partial derivatives with covariant ones, since the modification is again only by total derivatives:

$$\int dx\ d^2\theta = \int dx\ d^2$$

with an appropriate normalization. This turns out to be the most convenient one, since it allows covariant manipulations of the action, and the θ integration can be performed covariantly: Since we know that the result of θ integration gives a Lagrangian that depends only on x, up to total derivative terms, we can evaluate it as

$$\int dx\ d^2\theta\ f(\phi) = \int dx\ [d^2 f(\phi)]|$$

$$= \int dx\ [f'(\phi|)(d^2\phi)| + f''(\phi|)\tfrac{1}{2}(d^\alpha\phi)|(d_\alpha\phi)|]$$

(suppressing indices on multiple ϕ's). This gives the result directly in terms of component fields, using the covariant method of defining the component expansion: In the conventions of the previous subsection, this part of the action becomes

$$\int dx\ [f'(A)B + f''(A)\tfrac{1}{2}\psi^\alpha\psi_\alpha]$$

We now consider integration over the full superspace. As a generalization of the above, we can write

$$\int dx\ d^4\theta\ K(\phi,\bar{\phi}) = \int dx\ (d^2\bar{d}^2 K)|$$

Supersymmetric versions of nonlinear σ models can be written in this way; here we consider just the case where K is quadratic, which is the one interesting for quantum theory. Since a function of just ϕ (or just $\bar{\phi}$) will give zero in the $d^4\theta$ integral, we choose

$$K = -\bar{\phi}\phi \quad \Rightarrow \quad S_0 = -\int dx\ d^4\theta\ \bar{\phi}\phi$$

Explicitly,

$$L_0 = d^2\bar{d}^2(-\bar{\phi}\phi) = -d^2(\bar{d}^2\bar{\phi})\phi = -(\tfrac{1}{2}\Box\bar{\phi})\phi + (i\partial^{\alpha\dot{\beta}}\bar{d}_{\dot{\beta}}\bar{\phi})d_\alpha\phi - (\bar{d}^2\bar{\phi})(d^2\phi)$$

$$\to -A\tfrac{1}{2}\Box\bar{A} + \psi^\alpha i\partial_\alpha{}^{\dot{\beta}}\bar{\psi}_{\dot{\beta}} - B\bar{B}$$

where we have used the commutation relations of the covariant derivatives to push all d's past $\bar{d}$'s to hit $\bar{\phi}$. Clearly, this term by itself reproduces the results derived in the previous subsection based on kinematics, so it is the desired massless kinetic term.

We can now see the influence of adding the superpotential term to the action: The result of combining the two terms, and then eliminating the auxiliary field B by its equation of motion, is

$$S_0 + S_1 \to L = -A\tfrac{1}{2}\Box\bar{A} + \psi^\alpha i\partial_\alpha{}^{\dot{\beta}}\bar{\psi}_{\dot{\beta}} + |f'(A)|^2 + [f''(A)\tfrac{1}{2}\psi^\alpha\psi_\alpha + h.c.]$$

For example, a quadratic f gives mass to the physical scalar and spinor. This action is invariant under modified supersymmetry transformations, where the auxiliary fields are replaced by their equations of motion there also; those transformations then become nonlinear in the presence of interactions. Note that the scalar potential is positive definite; this is a consequence of supersymmetry, since it implies that the energy is always positive.

Exercise IVC2.1

These results generalize straightforwardly:

a Find the explicit form of the component-field action for arbitrary $K(\phi^i,\bar{\phi}_i)$ and $f(\phi^i)$ for an arbitrary number of chiral superfields ϕ^i, including all indices.

b Eliminate the auxiliary fields from the action, and find the modified supersymmetry transformations.

c Show by direct evaluation that the action is still invariant.

As a notational convenience, we can drop the "|" after expanding a superspace action in components: For example, we can write simply

$$\psi_\alpha = d_\alpha \phi, \qquad B = d^2\phi$$

After performing the θ-integration as above by using derivatives d and $\bar{d}$, and then "evaluating" these derivatives on ϕ by writing ψ and B, the component action is expressed completely in terms of such superfields and only *spacetime* derivatives $\partial_{\alpha\dot{\beta}}$. This component action is independent of θ (the Lagrangian is independent up to total spacetime derivatives): This is the statement of supersymmetry invariance. Thus, we can choose to evaluate at $\theta = 0$, or $\theta = \epsilon$, or whatever; it is irrelevant. It is then understood that the relation to the usual component actions is simply to treat the superfield as a component field, since the θ-derivatives (in d and $\int d\theta$) have been eliminated. From now on we will generally drop the |'s.

The above results can also be derived from the superfield equations of motion by varying the action. Since ϕ is constrained, it can't be varied arbitrarily; we vary instead the prepotential ψ ($\phi = \bar{d}^2\psi$). For example, we find $d^2\phi = 0$ (and the complex conjugate) from the free action. Effectively, since chiral superfields are essentially independent of $\bar{\theta}$, not only integration is modified, but also (functional) variation. Since a chiral superfield is (up to a transformation) an arbitrary function on chiral superspace, we define

$$\delta S[\phi] = \int dx\ d^2\theta\ (\delta\phi)\frac{\delta S}{\delta\phi}$$

for an arbitrary variation of a chiral superfield ϕ, and similarly for varying $\bar{\phi}$. In evaluating such variations, we make use of the identities

$$\int dx\ d^4\theta\ L = \int dx\ d^2\theta\ \bar{d}^2 L$$

$$\bar{d}^2 d^2\phi = \tfrac{1}{2}\Box\phi \qquad (\bar{d}^2 d^2 \bar{d}^2 = \tfrac{1}{2}\Box\bar{d}^2)$$

Thus, to vary a general action, it is convenient to first integrate over $\bar{\theta}$, and then vary in the naive way: For example,

$$S = -\int dx\ d^4\theta\ \bar{\phi}\phi + \left[\int dx\ d^2\theta\ f(\phi) + h.c.\right]$$

$$\Rightarrow \qquad 0 = \frac{\delta S}{\delta\phi} = -\bar{d}^2\bar{\phi} + f'(\phi)$$

Exercise IVC2.2

Check for this action that the component expansion of the superfield equations of motion agree with the variation of the corresponding component action.

3. Covariant derivatives

The supersymmetric generalization of nonabelian gauge theories can be derived by similar methods. We first write the supersymmetry covariant derivatives collectively as

$$d_A = (d_\alpha, \bar{d}_{\dot{\alpha}}, \partial_{\alpha\dot{\alpha}}) = E_A{}^M \partial_M$$

$$\partial_M = (\partial_\mu, \partial_{\dot{\mu}}, \partial_m) = \partial/\partial z^M, \qquad z^M = (\theta^\mu, \bar{\theta}^{\dot{\mu}}, x^m)$$

Unlike the nonsupersymmetric case, the "vielbein" $E_A{}^M$ has θ dependence even in "flat" superspace, and thus the "torsion" T is nonvanishing:

$$[d_A, d_B\} = T_{AB}{}^C d_C$$

$$T_{\alpha\dot{\beta}}{}^{\gamma\dot{\gamma}} = T_{\dot{\beta}\alpha}{}^{\gamma\dot{\gamma}} = -i\delta_\alpha^\gamma \delta_{\dot{\beta}}^{\dot{\gamma}}, \quad rest = 0$$

We now gauge-covariantize all the supersymmetry-covariant derivatives:

$$\nabla_A = d_A + iA_A$$

The covariant field strengths are then defined as

$$[\nabla_A, \nabla_B\} = T_{AB}{}^C \nabla_C + iF_{AB}$$

From our analysis of general representations of supersymmetry in D=4 in subsection IIC5, we know that the simplest supersymmetrization of Yang-Mills is to include a spinor with the vector, in terms of physical degrees of freedom. (The spinor and vector each have two physical degrees of freedom, one for each sign of the helicity.) Off shell, Fermi and Bose components must still balance, so there must also be an auxiliary scalar. From dimensional analysis, the field strengths must therefore satisfy

$$F_{\alpha\beta} = F_{\dot{\alpha}\dot{\beta}} = F_{\alpha\dot{\beta}} = 0; \qquad F_{\alpha,\beta\dot{\beta}} = -iC_{\alpha\beta}\overline{W}_{\dot{\beta}}, \quad F_{\dot{\alpha},\beta\dot{\beta}} = -iC_{\dot{\alpha}\dot{\beta}}W_\beta$$

where $W_\alpha|$ is the physical spinor.

The constant piece of the torsion implies stronger relations among the field strengths than in nonsupersymmetric theories. For super Yang-Mills we find from the Jacobi identity for the covariant derivatives the Bianchi identity for the field strengths

$$\nabla_{[A}F_{BC)} = T_{[AB|}{}^D F_{D|C)}$$

Specifically, the dimension-1 constraints above on the field strengths imply the dimension-3/2 algebraic constraint that defines W_α, as well as

$$F_{\alpha\dot{\alpha},\beta\dot{\beta}} = C_{\alpha\beta}\tfrac{1}{2}\nabla_{(\dot{\alpha}}\overline{W}_{\dot{\beta})} + C_{\dot{\alpha}\dot{\beta}}\tfrac{1}{2}\nabla_{(\alpha}W_{\beta)}$$

They also imply that W_α is covariantly chiral and satisfies a "reality" condition,

$$\overline{\nabla}_{\dot{\alpha}}W_\beta = 0, \qquad \nabla^\alpha W_\alpha + \nabla^{\dot{\alpha}}W_{\dot{\alpha}} = 0$$

The most straightforward way to derive these results is to just evaluate the Jacobi identities directly. We begin with a weaker set of conditions, both of dimension 1, that will be found (in the following subsection) to be necessary and sufficient for solving explicitly. One directly determines the vector derivative in terms of the spinor ones:

$$F_{\alpha\dot{\beta}} = 0 \quad \Rightarrow \quad -i\nabla_{\alpha\dot{\alpha}} = \{\nabla_\alpha, \overline{\nabla}_{\dot{\alpha}}\}$$

Since one could always define the vector covariant derivative this way, imposing this condition simply eliminates redundant degrees of freedom.

The remaining constraint (including its complex conjugate) allows coupling of super Yang-Mills to the chiral superfield:

$$\overline{\nabla}_{\dot{\alpha}}\phi = 0 \quad \Rightarrow \quad 0 = \{\overline{\nabla}_{\dot{\alpha}}, \overline{\nabla}_{\dot{\beta}}\}\phi = iF_{\dot{\alpha}\dot{\beta}}\phi$$

It also implies the maintenance of certain free identities, such as

$$\nabla_\alpha\nabla_\beta = \tfrac{1}{2}[\nabla_\alpha, \nabla_\beta] + \tfrac{1}{2}\{\nabla_\alpha, \nabla_\beta\} = C_{\beta\alpha}\nabla^2$$

(Such constraints appear also for first quantization, e.g., in superstring theory, whenever a supersymmetric system is put in a background of a supersymmetric gauge field of higher superspin. This should not be confused with background *field equations* imposed by any gauge system put in a background of the *same* type: see subsection VIB8.)

Thus, our minimal set of constraints can be written directly in terms of the field strengths as

$$F_{\alpha\beta} = F_{\dot{\alpha}\dot{\beta}} = F_{\alpha\dot{\beta}} = 0$$

but for our purposes it will prove more convenient to write them directly as (anti)commutators:

$$\{\nabla_\alpha, \nabla_\beta\} = \{\overline{\nabla}_{\dot{\alpha}}, \overline{\nabla}_{\dot{\beta}}\} = 0, \qquad \{\nabla_\alpha, \overline{\nabla}_{\dot{\beta}}\} = -i\nabla_{\alpha\dot{\beta}}$$

The solution to the dimension-3/2 Jacobis are then

$$[\nabla_{(\alpha}, \{\nabla_\beta, \nabla_{\gamma)}\}] = 0 \quad \Rightarrow \quad \textit{trivial}$$

$$[\nabla_{(\alpha},\{\nabla_{\beta)},\overline{\nabla}_{\dot\gamma}\}]+[\overline{\nabla}_{\dot\gamma},\{\nabla_\alpha,\nabla_\beta\}]=0 \quad\Rightarrow\quad [\nabla_\alpha,\nabla_{\beta\dot\gamma}]=C_{\alpha\beta}\overline{W}_{\dot\gamma}$$

for some field W, simply applying the constraints to drop $\{\nabla_\alpha,\nabla_\beta\}$ and replace $\{\nabla_\alpha,\overline{\nabla}_{\dot\beta}\}$ with $\nabla_{\alpha\dot\beta}$. Similarly, we find from the dimension-2 Jacobis

$$\{\nabla_{(\alpha},[\nabla_{\beta)},\nabla_{\gamma\dot\gamma}]\}+[\nabla_{\gamma\dot\gamma},\{\nabla_\alpha,\nabla_\beta\}]=0 \quad\Rightarrow\quad \nabla_\alpha\overline{W}_{\dot\gamma}=0$$

$$[\nabla_{\alpha\dot\alpha},\{\nabla_\beta,\overline{\nabla}_{\dot\gamma}\}]+\{\nabla_\beta,[\overline{\nabla}_{\dot\gamma},\nabla_{\alpha\dot\alpha}]\}+\{\overline{\nabla}_{\dot\gamma},[\nabla_\beta,\nabla_{\alpha\dot\alpha}]\}=0$$

$$\Rightarrow\quad [\nabla_{\alpha\dot\alpha},\nabla_{\beta\dot\beta}]=i(C_{\alpha\beta}\bar f_{\dot\alpha\dot\beta}+C_{\dot\alpha\dot\beta}f_{\alpha\beta}),\qquad f_{\alpha\beta}=\tfrac{1}{2}\nabla_{(\alpha}W_{\beta)},\qquad \nabla^\alpha W_\alpha+\overline{\nabla}^{\dot\alpha}\overline{W}_{\dot\alpha}=0$$

where we separated the last equation into its (Lorentz) irreducible pieces. (The dimension-5/2 and 3 identities are redundant.)

Exercise IVC3.1

Explicitly evaluate all the remaining Jacobi identities, and show that they imply no further conditions on W_α.

4. Prepotential

We saw in the previous subsection that coupling super Yang-Mills to matter gave directly one of the minimal constraints on the super Yang-Mills fields themselves. Hence, as for ordinary Yang-Mills, the definition of the gauge theory follows from considering the transformation of matter, and generalizing it to a local symmetry. As for self-dual Yang-Mills (see subsection IIIC5), the vanishing of some field strengths implies that part of the covariant derivative is pure gauge:

$$\{\nabla_\alpha,\nabla_\beta\}=0 \quad\Rightarrow\quad \nabla_\alpha=e^{-\Omega}d_\alpha e^{\Omega}$$

However, since $\{\nabla_\alpha,\overline{\nabla}_{\dot\beta}\}\neq 0$, this gauge transformation Ω ("prepotential") is complex. We therefore have the covariantly chiral superfield

$$\overline{\nabla}_{\dot\alpha}\phi=0,\quad \overline{\nabla}_{\dot\alpha}=e^{\bar\Omega}d_{\dot\alpha}e^{-\bar\Omega} \quad\Rightarrow\quad \phi=e^{\bar\Omega}\hat\phi,\quad d_{\dot\alpha}\hat\phi=0$$

Alternatively, we could combine this exponential with that already contained in the free spinor derivative:

$$\nabla_\alpha=e^{-U/2-\Omega}\partial_\alpha e^{U/2+\Omega},\quad \phi=e^{U/2+\bar\Omega}\hat\phi,\quad \partial_{\dot\alpha}\hat\phi=0$$

$U+2\Omega$ is the analog of the covariant derivative for the Yang-Mills prepotential. This is a hint at supergravity: U is just the flat piece of the supergravity prepotential. We thus see that supersymmetry automatically gives gravity the interpretation of the gauge theory of translations.

Component expansions are now defined with Yang-Mills-covariant derivatives:

$$\nabla_\alpha \phi = \psi_\alpha, \qquad \nabla^2 \phi = B$$

$$\nabla_\alpha W_\beta = f_{\alpha\beta} + iC_{\alpha\beta} D, \qquad \nabla^2 W_\alpha = -i\nabla_\alpha{}^{\dot\beta} \overline{W}_{\dot\beta}$$

where we have used the Bianchi identities for W, and $f_{\alpha\beta}$ (not to be confused with $F_{\alpha\beta}$) is the usual Yang-Mills field strength (in spinor notation). The "vector multiplet" thus consists of the component fields A_a (the gauge field whose strength is f), W_α, and D (auxiliary). (As explained earlier, we drop all |'s.)

Note that the gauge parameter is real, while the matter multiplet is (covariantly) chiral. The resolution of this apparent inconsistency is that solving the constraints introduces a new gauge invariance, just as solving the source-free half of Maxwell's equations (really just constraints, not field equations) introduces the potential and its gauge invariance:

$$\nabla'_A = e^{iK} \nabla_A e^{-iK}, \quad \nabla_\alpha = e^{-\Omega} d_\alpha e^{\Omega} \quad \Rightarrow \quad e^{\Omega'} = e^{i\bar\Lambda} e^{\Omega} e^{-iK}, \quad d_\alpha \bar\Lambda = 0$$

$$\phi' = e^{iK}\phi, \quad \phi = e^{\bar\Omega}\hat\phi \quad \Rightarrow \quad \hat\phi' = e^{i\Lambda}\hat\phi$$

This suggests the definition of a new ("chiral") representation, where we use the obvious field $\hat\phi$ and the chiral gauge parameter Λ replaces the real one K: Making a nonunitary similarity transformation,

$$\widehat\nabla_A = e^{-\bar\Omega} \nabla_A e^{\bar\Omega} \quad \Rightarrow \quad \widehat{\overline\nabla}_{\dot\alpha} = d_{\dot\alpha}, \quad \widehat\nabla_\alpha = e^{-V} d_\alpha e^V, \quad e^V = e^{\Omega} e^{\bar\Omega}$$

$$\hat\phi = e^{-\bar\Omega}\phi, \quad \hat{\bar\phi} = \bar\phi e^{\bar\Omega} \quad \Rightarrow \quad d_{\dot\alpha}\hat\phi = 0, \quad \hat{\bar\phi} = (\hat\phi)^\dagger e^V$$

$$\widehat\nabla'_A = e^{i\Lambda} \widehat\nabla_A e^{-i\Lambda}, \quad e^{V'} = e^{i\bar\Lambda} e^V e^{-i\Lambda}$$

Alternatively, we can also include U in the transformation as above; then U and V appear only in the combination $U + V$.

Exercise IVC4.1

Show that the explicit expression for the field strength W_α in terms of the prepotential V in the chiral representation is

$$W_\alpha = -i\bar d^2 (e^{-V} d_\alpha e^V)$$

Show this expression is chiral.

Exercise IVC4.2

In the Abelian case, give an explicit component expansion of the prepotential V, such that the vector potential A_a, the physical spinor W_α, and the auxiliary field D appear as independent components. Note that the other components do not appear explicitly in component expansions when gauge-covariant expansion $(\nabla ...|)$ is used. The component (nonsupersymmetric) gauge where these components are set to vanish is the "Wess-Zumino gauge", and is the θ part of the radial gauge of subsection VIB1 below.

Exercise IVC4.3

For some purposes (like quantization) we need the explicit form of an infinitesimal gauge transformation of V. Show this can be written as

$$\delta V = -i\mathcal{L}_{V/2}[(\Lambda + \bar{\Lambda}) + coth(\mathcal{L}_{V/2})(\Lambda - \bar{\Lambda})]$$

(Hint: Consider $e^{-V}\delta e^{V}$, and think of δ as an operator, as for the expansion of $\nabla_\alpha = e^{-V}d_\alpha e^{V}$. $\mathcal{L}_A$ was defined in subsection IA3.)

5. Gauge actions

Generalization of actions to super Yang-Mills theory is straightforward. Matter coupling is achieved simply by replacing the chiral superfields of the matter multiplets with Yang-Mills-covariantly chiral superfields. The coupling can be seen explicitly in the chiral representation: In the kinetic term,

$$\bar{\phi}\phi = (\hat{\phi})^{\dagger} e^{V} \hat{\phi}$$

while in the $\int d^2\theta$ term all V-dependence drops out because of gauge invariance. (The superpotential is a gauge invariant function of the ϕ's, and the transformation to the chiral representation is a complex gauge transformation. The fact that the gauge transformation is complex is irrelevant, since the superpotential depends only on ϕ and not $\bar{\phi}$.) Component expansion can be performed covariantly by replacing d's with ∇'s in the definition of θ integration: Since the Lagrangian is a gauge singlet, this is the same acting on it, although individual terms in the expansion differ because the fields are not singlets. Similarly, $\bar{d}^2$ can be replaced with $\bar{\nabla}^2$ also when performing $\bar{\theta}$ integration for purposes of varying an action with respect to a chiral superfield. This is equivalent to gauge covariantizing the functional derivative (e.g., by transforming from a chiral representation) as

$$\frac{\delta\phi(x,\theta)}{\delta\phi(x',\theta')} = \bar{\nabla}^2 \delta(x-x')\delta^4(\theta-\theta')$$

Usually we will drop the "$\hat{\ }$"'s on ϕ and $\bar{\phi}$, when the representation is clear from the context by the use of explicit V's.

The action for super Yang-Mills itself follows from dimensional analysis: Since each θ integral is really a θ derivative, $d^2\theta$ integration has mass dimension +1, the same as a spacetime derivative. Since the Lagrangian for a physical spinor, in this case W^α, has a single such derivative, dimensional analysis says the action must be

$$S_{sYM} = -\tfrac{1}{g^2} tr \int dx\ d^2\theta\ \tfrac{1}{2} W^\alpha W_\alpha$$

where the (covariant) chirality of W^α allows integration over chiral superspace. (Similar analysis applies to the matter multiplet, where $\int d^4\theta$ takes the place of a $\Box$ for the scalar ϕ.) Replacing $\int d^2\theta \to \nabla^2$, we evaluate the component expansion as

$$S_{sYM} = \tfrac{1}{g^2} tr \int dx\ (\tfrac{1}{2} f^{\alpha\beta} f_{\alpha\beta} + W^\alpha i \nabla_\alpha{}^{\dot\beta} \overline{W}_{\dot\beta} - D^2)$$

Another term we can write, for superelectromagnetism (supersymmetrization of an Abelian gauge theory) is the "Fayet-Iliopoulos term"

$$S_{FI} = \zeta \int dx\ d^4\theta\ V = \zeta \int dx\ D$$

which involves only the auxiliary field D. (The analog for the chiral scalar superfield is $\int dx\ d^2\theta\ \phi$.)

Exercise IVC5.1

Derive the supersymmetric analog of the Stückelberg model of subsection IVA5, by coupling an Abelian vector multiplet to a massless chiral scalar multiplet using the symmetry generator T defined there. ($G \to -iT$ in transformation laws, covariant derivatives, etc., on ϕ, where $T\phi = 1 \Rightarrow T^2\phi = 0$.)

a To couple the gauge field it is necessary to start, as usual, with a (quadratic) matter action that is *globally* invariant under this symmetry:

$$S_0 = \int dx\ d^4\theta\ \tfrac{1}{2}(\phi - \bar\phi)^2$$

(At this point this is the usual, since only the cross-term survives, but this will not be the case for the covariantly chiral superfields.) Find the supersymmetric gauge coupling, and express the resulting action in terms of V and $\hat\phi$.

b Use this result to find the mass term for V in the gauge $\hat\phi = 0$.

Another interesting form of the action uses a generalization of the Chern-Simons form defined in the discussion of instantons in subsection IIIC6. In superspace, the calculation of the field strength with curved indices is modified to

$$\nabla_M = E_M{}^A \nabla_A = \partial_M + iA_M, \quad -i[\nabla_M, \nabla_N\} = F_{MN} = E_M{}^A E_N{}^B F_{AB}$$

where we have left sign factors from index reordering in the last equation implicit. Although the curved-index expressions are not as useful (for example, for seeing which components vanish by constraints), we can see easily that some arguments used in nonsupersymmetric theories carry over to superspace. Thus, we can define the super Chern-Simons form by

$$\tfrac{1}{8}tr\ F_{[MN}F_{PQ)} = \tfrac{1}{6}\partial_{[M}B_{NPQ)}$$

$$B_{MNP} = tr(\tfrac{1}{2}A_{[M}\partial_N A_{P)} + i\tfrac{1}{3}A_{[M}A_N A_{P)})$$

Converting to flat superspace (again with some implicit sign factors),

$$B_{ABC} = E_A{}^M E_B{}^N E_C{}^P B_{MNP} = tr(\tfrac{1}{2}A_{[A}d_B A_{C)} - \tfrac{1}{4}A_{[A}T_{BC)}{}^D A_D + i\tfrac{1}{3}A_{[A}A_B A_{C)})$$

In terms of this expression, the super Yang-Mills action can be written simply in terms of the spinor-spinor-vector part $B_{\alpha\dot{\beta}c}$ of B_{ABC} as

$$S_{sYM,1} = \tfrac{1}{2}i\tfrac{1}{g^2}tr\int dx\ d^4\theta\ B_{\alpha,\dot{\alpha}}{}^{\alpha\dot{\alpha}}$$

Note that the fact that the curl of B is gauge invariant implies that B transforms under a gauge transformation as the curl of something, and thus the integral of any part of B is gauge invariant (up to possible torsion terms: see the exercise below). Furthermore, we can drop the $F_{\alpha\dot{\beta}} = 0$ constraint on the A in this action; it follows from variation with respect to $A_{\alpha\dot{\beta}}$. One simple way to check this action is to use the chiral representation $A_{\dot{\alpha}} = 0$: Then only the $A^{\alpha\dot{\beta}}\overleftrightarrow{d}_{\dot{\beta}}A_\alpha$ and $(A_{\alpha\dot{\beta}})^2$ terms contribute, and $A_{\alpha\dot{\beta}} = i\bar{d}_{\dot{\beta}}A_\alpha$, while $W_\alpha = \bar{d}^2 A_\alpha$, so $\int d^2\bar{\theta}$ integration gives $-\int dx\ d^2\theta\ W^2$.

Exercise IVC5.2

Derive the expression for B_{ABC} directly using only flat indices:

a Start with $F_{[AB}F_{CD)}$ expressed in terms of T and A, and write it as a total derivative plus torsion terms.

b Do the same for the gauge transformation of B. Show that the torsion terms do not contribute to $\delta B^{\alpha\dot{\alpha},}{}_{\alpha,\dot{\alpha}}$.

The multiplets and couplings we have considered are sufficient to write a supersymmetric generalization of the Standard Model. Unfortunately, supersymmetry provides no unification. To get the right symmetry breaking, it turns out to be necessary to provide a supersymmetry multiplet for each particle of the Standard Model: The spin-1 gauge bosons are accompanied by spin-1/2 "gauginos" ("gluinos", "photino", "Wino", "Zino"), the spin-1/2 leptons by spin-0 "sleptons", the quarks by "squarks", and the spin-0 Higgs' by spin-1/2 "Higgsinos". Furthermore, since a reality condition can't be imposed on chiral scalar multiplets, the Higgs scalars are themselves doubled. Ultimately, the success of supersymmetry depends on the experimental detection of these particles.

6. Breaking

The methods of the section IVA can be generalized straightforwardly to supersymmetric theories: Goldstone bosons and Higgs fields become supermultiplets, etc. However, to obtain realistic models supersymmetry itself must be broken, since fermions and bosons with similar mass and other properties are not observed in nature. More specifically, since gravity is observed, any supersymmetric theory of the world must include supergravity, and thus the breaking must be spontaneous. (Explicit breaking would violate gauge invariance.) Then the gravitino, which gauges supersymmetry, will become massive by a superhiggs mechanism, by eating a Goldstone fermion. (See subsections XB6-7. If the graviton and gravitino are treated as composites, then this fermion could also be a composite.)

We saw in subsection IIC1 that energy is always nonnegative in supersymmetric theories. In particular, from the same arguments used there we see that a state can be invariant under supersymmetry ($q|\psi\rangle = q^\dagger|\psi\rangle = 0$) if and only if it has zero energy. Any such state can be identified as the vacuum, since no state has lower energy. This means that the only way to guarantee spontaneous supersymmetry breaking is to choose a theory which has no zero-energy state. (Note that energy is uniquely defined by the supersymmetry algebra; there is no possibility of adding a constant as in nonsupersymmetric theories.) In theories with extended supersymmetry, the relation between supersymmetry and energy applies for each supersymmetry; thus supersymmetry is either completely broken spontaneously or completely unbroken. (An exception is central charges, which modify the supersymmetry algebra: See the following subsection.)

Furthermore, physical scalars appear at $\theta = 0$ in matter multiplets, while auxiliary fields appear at higher order. Since supersymmetry breaking requires θ dependence in a vacuum value of a superfield, this means an auxiliary field must get a vacuum value.

A simple example of spontaneous supersymmetry breaking is the O'Raifeartaigh model; it has the Lagrangian

$$L_{O'R} = -\int d^4\theta \sum_{i=1}^{3} \bar{\Phi}_i \Phi_i + \left[\int d^2\theta \; \lambda(\zeta\Phi_1 + m\Phi_2\Phi_3 + \Phi_1\Phi_2^2) + h.c. \right]$$

To study symmetry breaking we ignore derivative terms, since vacuum values are constants. Then the scalar field equations are:

$$\frac{\delta}{\delta B_i} \to -\bar{B}_i + \partial_i f = 0 : \quad -\bar{B}_1 + \zeta + A_2^2 = -\bar{B}_2 + mA_3 + 2A_1 A_2 = -\bar{B}_3 + mA_2 = 0$$

$$\frac{\delta}{\delta A_i} \to B_j \partial_i \partial_j f = 0 : \quad 2A_2 B_2 = mB_3 + 2A_2 B_1 + 2A_1 B_2 = mB_2 = 0$$

(where $\partial_i = \partial/\partial A_i$ on the superpotential $f(A)$). Since there is no solution for $B_i = 0$, supersymmetry breaking is required. In general, for superpotential $f(\Phi)$, the field equations for $B = 0$ are $f'(A) = 0$, so a linear term is always needed for supersymmetry breaking.

With Abelian vector multiplets, a Fayet-Iliopoulos term $\int d^4\theta\, V$ can also generate such breaking, since it also is a linear term of an auxiliary field.

Exercise IVC6.1

Evaluate the Lagrangian $-\int d^4\theta\, \bar{\phi}\phi$ for covariantly chiral ϕ by using covariant θ-integration, $\int d^4\theta = \nabla^2 \bar{\nabla}^2$. For the case of U(1) gauge theory, add the action for the gauge superfield with a Fayet-Iliopoulos term, and find the potential for the physical scalars by eliminating the auxiliary field D by its field equation.

For simplicity (as in this chapter), we may want to ignore supergravity; however, we still need to take account of its contribution to breaking global supersymmetry via the superhiggs effect. The net low-energy contribution from the supergravity fields (assuming no cosmological constant is generated) is to introduce effective explicit supersymmetry breaking: Although the original theory is locally supersymmetric, we neglect the supergravity fields but not their vacuum values (in particular, those of the auxiliary fields). In particular, if the supergravity fields are bound states, then this procedure is essentially the classical introduction of nonperturbative quantum effects.

Thus we consider adding terms to the classical action that break supersymmetry explicitly. The easiest way to do this is to introduce constant superfields ("spurions"); this allows us to continue to take advantage of the superspace formalism (at both the classical and quantum levels). Since we are neglecting (super)gravity, and in particular its nonrenormalizability (see chapter VII), we consider only terms that will preserve the quantum properties of the unbroken theories. This will clearly be the case if we consider only the usual terms, with some fields replaced by spurions: This is equivalent to using background (fixed) fields, in addition to (but in the same way as) the usual field variables, performing all (classical/quantum) calculations as usual, and then setting the background fields (specifically, the auxiliary fields, which are responsible for breaking supersymmetry) to constants.

Thus, introducing constant (in x) chiral and real spurion fields

$$\varphi = \theta^2 c, \qquad \mathcal{V} = \theta^2 \bar{\theta}^2 r$$

in terms of complex and real parameters c and r, in addition to the true fields ϕ and V, we have terms of the form

$$\int d^2\theta \; [\varphi\phi, \varphi\phi^2, \varphi\phi^3, \varphi W^2, (\bar{d}^2 d^\alpha \mathcal{V})\phi W_\alpha], \qquad \int d^2\theta \; d^2\bar{\theta} \; \mathcal{V}\bar{\phi} e^V \phi$$

(and complex conjugates). These terms can preserve the usual gauge invariances, and can be shown to also preserve the desirable quantum properties of supersymmetry: The condition is that replacing the spurion field by 1 (instead of its above value) gives either 0 or a conventional term (one with coupling constant of nonnegative mass dimension). Another way to introduce these spurions (except perhaps for the ϕV crossterm, which is less useful) is as *coupling constants*, rather than as fields: Instead of introducing new terms to the action, we generalize the old ones, so the constant part of each coupling is the usual coupling, while its θ-dependent terms produce the breaking.

Exercise IVC6.2

Find the component expansions of the above explicit breaking terms. What are the mass dimensions of the constants c and r in the various cases?

Exercise IVC6.3

Expand the Lagrangian

$$L = -\int d^4\theta \; \bar{\phi}\phi + \left[\int d^2\theta \; (\tfrac{1}{6}\phi^3 + \varphi\phi) + h.c. \right]$$

in components. Find the masses.

7. Extended

The supersymmetry we discussed earlier in this chapter, with a single spinor coordinate, is called "simple (N=1) supersymmetry"; the generalization to many spinors is called "extended (N>1) supersymmetry" (for N spinor coordinates). N=1 supersymmetric theories, at least for spins$\leq$1, are most conveniently described by superspace methods. (There are also some definite advantages for N=1 supergravity at the quantum level.) On the other hand, the technical difficulties of extended superspace often outweigh the advantages. (The main advantage of extended superspace is proving certain properties of the quantum theories. Of course, extended supersymmetric theories are complicated in any case.) Alternative formulations of extended supersymmetry are either

(1) on shell,

(2) in terms of components (ordinary spacetime, not superspace), or

(3) in simple superspace (manifesting only one of the supersymmetries).

By going half way, using N=1 superfields to describe extended supersymmetry, some of the advantages of the superspace approach can be retained. In this subsection we will list some of the extended supersymmetric actions for lower spins in N=1 superspace form. These actions can be obtained by: (1) using extended superspace to derive the component field equations (usually using dimensional reduction: see subsections XC5-6), and combining components into N=1 superfields, or (2) writing the extra supersymmetries in N=1 superspace form, and using them to determine the action.

The simplest example is N=2 supersymmetry. As for any extended supersymmetry, its algebra can be modified by including Abelian generators Z (with dimensions of mass), called "central charges":

$$\{q_{i\alpha}, \bar{q}^j_{\dot{\beta}}\} = \delta_i^j p_{\alpha\dot{\beta}}, \quad \{q_{i\alpha}, q_{j\beta}\} = C_{\alpha\beta}C_{ij}Z, \quad \{\bar{q}^i_{\dot{\alpha}}, \bar{q}^j_{\dot{\beta}}\} = C_{\dot{\alpha}\dot{\beta}}C^{ij}Z; \quad [Z,q] = [Z,\bar{q}] = 0$$

(where $i = 1, 2$). In terms of dimensional reduction (for N=2, from D=5 or 6; see subsections XC5-6), the origin of these generators can be understood as the higher-dimensional components of the momentum. N=2 supersymmetry is sometimes called "hypersymmetry", and N=2 supermultiplets, "hypermultiplets".

Our first example is the free, massive N=2 scalar multiplet: Since we already know the field content (see subsection IIC5), it's easy to write the free Lagrangian

$$L_{sm,N=2} = -\int d^4\theta\ \bar{\phi}^{i'}\phi_{i'} + \tfrac{1}{2}\left(\int d^2\theta\ m^{i'j'}\phi_{i'}\phi_{j'} + h.c.\right)$$

where the index "i'" is for an extra SU(2) (not the one acting on the supersymmetry generators), broken by the mass term, and the mass matrix $m^{i'j'}$ is symmetric while $m_{i'}{}^{j'} = C_{k'i'}m^{k'j'}$ is hermitian. In other words, it represents a 3-vector of this SU(2), and thus a generator of the preserved U(1) subgroup, which we have used to define the central charge:

$$Z\phi_{i'} = m_{i'}{}^{j'}\phi_{j'}$$

The other N=2 multiplet of low spin is the vector multiplet. It also has a simple Lagrangian,

$$L_{sYM,N=2} = -\tfrac{1}{g^2} tr \left(\int d^2\theta \ W^2 + \int d^4\theta \ \bar{\phi}\phi \right)$$

where ϕ is covariantly chiral and in the adjoint representation of the Yang-Mills gauge group. In the Abelian case, we can also add an N=2 Fayet-Iliopoulos term,

$$L_{FI,N=2} = \int d^4\theta \ \zeta_0 V + \left(\int d^2\theta \ \zeta_+ \phi + h.c. \right)$$

where $(\zeta_0, \zeta_+, \zeta_-)$ $(\zeta_- = \zeta_+{}^*)$ is a constant 3-vector of the SU(2) of the N=2 supersymmetry: The 3 scalar auxiliary fields of this N=2 multiplet form a 3-vector of the SU(2). Unlike the previous example, this multiplet has all the auxiliary fields needed for an off-shell N=2 superspace formulation: Not only do the physical components balance between bosons and fermions (4 of each), but also the auxiliary ones (also 4 of each).

These 2 N=2 multiplets can be coupled: The scalar multiplet action is modified to

$$L_{sm,N=2} = -\int d^4\theta \ \bar{\phi}^{i'} \phi_{i'} + \tfrac{1}{2} \left[\int d^2\theta \ \tau^{i'j'} \phi_{i'} (\phi + M) \phi_{j'} + h.c. \right]$$

where now $\phi_{i'}$ is also a representation of the Yang-Mills group (not necessarily adjoint), with respect to which it is covariantly chiral. However, the same SU(2) matrix τ that appears in the mass matrix $m_{i'j'} = M\tau_{i'j'}$ now also appears with the N=2 super Yang-Mills fields,

$$\nabla_A \phi_{i'} = d_A \phi_{i'} + i A_A^n G_n \tau_{i'}{}^{j'} \phi_{j'}, \qquad \phi = \phi^n G_n$$

where G_n are the usual Yang-Mills group generators. (Without loss of generality, we can choose $\tau_{i'}{}^{j'} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$; then $\phi_{+'}$ is some arbitrary representation of the Yang-Mills group, while $\phi_{-'}$ is the complex conjugate.) Note that the mass term appears in exactly the same way as an Abelian N=2 vector multiplet that has been replaced by a vacuum value for its physical scalars. This can also be seen from the commutation relations for the N=2 super Yang-Mills covariant derivatives (see below), since the scalars appear in exactly the same way as the central charge.

By our earlier helicity arguments, the only N=3 supersymmetric theory with spins ≤ 1 is N=3 super Yang-Mills. The analogous statement also holds for N=4, while no such theories exist for N>4. Since theories with N supersymmetries are a subset of those with only N−1 supersymmetries, N=3 and N=4 super Yang-Mills must be the same: Counting states of supersymmetry representations, we see that this theory is the same as N=2 super Yang-Mills coupled to one N=2 scalar multiplet in the adjoint representation (in direct analogy to N=2 super Yang-Mills in terms of N=1 multiplets). In terms of N=1 multiplets, this is super Yang-Mills plus 3 adjoint scalar multiplets. The action then follows from the above results (without central charges and Fayet-Iliopoulos terms):

$$L_{sYM,N=4} = \tfrac{1}{g^2} tr \left[-\int d^2\theta \; W^2 - \int d^4\theta \; \bar{\phi}^I \phi_I + \left(\int d^2\theta \; \tfrac{1}{6}\epsilon^{IJK}\phi_I[\phi_J,\phi_K] + h.c. \right) \right]$$

where "I" is a U(3) index. (The U(1) part of the U(3) symmetry involves also a phase transformation of the θ's.)

For comparison, here are the general (UV well-behaved) actions for all numbers of supersymmetries (in D=4):

$$L_{N=1} = -\tfrac{1}{g^2} tr \int d^2\theta \; \tfrac{1}{2}W^\alpha W_\alpha + \zeta \int d^4\theta \; V - \int d^4\theta \; \bar{\phi} e^V \phi + \left[\int d^2\theta \; f(\phi) + h.c. \right]$$

$$L_{N=2} = -\tfrac{1}{g^2} tr \left(\int d^2\theta \; W^2 + \int d^4\theta \; e^{-V}\bar{\phi} e^V \phi \right) + \int d^4\theta \; \zeta_0 V + \left(\int d^2\theta \; \zeta_+ \phi + h.c. \right)$$
$$- \int d^4\theta \; \bar{\phi}^{i'} (e^{V\tau})_{i'}{}^{j'} \phi_{j'} + \tfrac{1}{2} \left[\int d^2\theta \; \tau^{i'j'} \phi_{i'} (\phi + M) \phi_{j'} + h.c. \right]$$

$$L_{N=4} = \tfrac{1}{g^2} tr \left[-\int d^2\theta \; W^2 - \int d^4\theta \; e^{-V}\bar{\phi}^I e^V \phi_I + \left(\int d^2\theta \; \tfrac{1}{6}\epsilon^{IJK}\phi_I[\phi_J,\phi_K] + h.c. \right) \right]$$

where we now use ordinary chiral superfields, making dependence on V explicit.

As for off-shell N=1 supersymmetry, much information on extended supersymmetric gauge theories can be gained by examining the properties of the covariant derivatives and their field strengths. In fact, this is more true in the extended case, where the "obvious" constraints often imply field equations (which is more than one would want for an off-shell formulation). The empty-space covariant derivatives are the direct generalization of N=1: Introducing N θ's as $\theta^{i\alpha}$ (and complex conjugate $\bar{\theta}_i{}^{\dot{\alpha}}$), where "$i$" is an N-valued index with as much as a U(N) symmetry,

$$d_A = (d_{i\alpha}, \bar{d}^i{}_{\dot{\alpha}}, \partial_{\alpha\dot{\alpha}}); \qquad d_{i\alpha} = \partial_{i\alpha} - i\tfrac{1}{2}\bar{\theta}_i{}^{\dot{\alpha}}\partial_{\alpha\dot{\alpha}}, \quad \bar{d}^i{}_{\dot{\alpha}} = \bar{\partial}^i{}_{\dot{\alpha}} - i\tfrac{1}{2}\theta^{i\alpha}\partial_{\alpha\dot{\alpha}}$$

$$T_{i\alpha,}{}^j{}_{\dot{\beta}}{}^{\gamma\dot{\gamma}} = T^j{}_{\dot{\beta},i\alpha}{}^{\gamma\dot{\gamma}} = -i\delta_i^j \delta_\alpha^\gamma \delta_{\dot{\beta}}^{\dot{\gamma}}, \qquad rest = 0$$

Exercise IVC7.1

Find the superspace representation of the extended supersymmetry generators (which anticommute with these covariant derivatives). For N=2, include the central charge.

By definition, extended super Yang-Mills has only spins 1 and less. Dimensional analysis then gives the unique result, including physical fields only,

$$\{\nabla_{i\alpha}, \overline{\nabla}^j{}_{\dot\beta}\} = -\delta_i^j i\nabla_{\alpha\dot\beta}$$

$$\{\nabla_{i\alpha}, \nabla_{j\beta}\} = C_{\beta\alpha} i\bar\phi_{ij}$$

$$[\overline{\nabla}^i{}_{\dot\alpha}, -i\nabla_{\beta\dot\beta}] = C_{\dot\beta\dot\alpha} iW^i{}_\beta$$

$$[\nabla_{\alpha\dot\alpha}, \nabla_{\beta\dot\beta}] = C_{\alpha\beta} i\bar f_{\dot\alpha\dot\beta} + C_{\dot\alpha\dot\beta} if_{\alpha\beta}$$

(and complex conjugates of some of these equations). This corresponds directly to our discussion in subsection IIC5, where we saw that a general representation looked like antisymmetric tensors $\phi, \phi^i, \phi^{ij}, ...$ of U(N), corresponding to helicities h, h−1/2, h−1,... . In this case, h=1, and these helicities come from the surviving on-shell components of $f_{\alpha\beta}, W^i{}_\alpha, \phi^{ij}$,... . For N=4 we have self-duality with respect to charge conjugation (see also subsection IIC5),

$$\phi^{ij} = \tfrac{1}{2}\epsilon^{ijkl}\bar\phi_{kl}$$

Exercise IVC7.2

Analyze the Bianchi identities of these covariant derivatives:

a Show that for N>2 they imply the field equations.

b Find a component action that yields these field equations for N=4.

An interesting simplification of extended superspace occurs for self-duality: Constraining

$$f^{\alpha\beta} = W^{i\alpha} = \phi^{ij} = 0$$

and dropping the self-duality condition for N=4 (so $\bar\phi_{ij} \neq 0$), we find all commutators involving $\overline{\nabla}^i{}_{\dot\alpha}$ are trivial:

$$\{\nabla_{i\alpha}, \overline{\nabla}^j{}_{\dot\beta}\} = -\delta_i^j i\nabla_{\alpha\dot\beta}, \qquad \{\overline{\nabla}^i{}_{\dot\alpha}, \overline{\nabla}^j{}_{\dot\beta}\} = [\overline{\nabla}^i{}_{\dot\alpha}, \nabla_{\beta\dot\beta}] = 0$$

while all the remaining commutators have a similar form:

$$\{\nabla_{i\alpha}, \nabla_{j\beta}\} = C_{\beta\alpha} i\bar\phi_{ij}, \qquad [\nabla_{i\alpha}, -i\nabla_{\beta\dot\beta}] = C_{\beta\alpha} i\overline{W}_{i\dot\beta}, \qquad [\nabla_{\alpha\dot\alpha}, \nabla_{\beta\dot\beta}] = C_{\alpha\beta} i\bar f_{\dot\alpha\dot\beta}$$

The latter result suggests we combine the internal and dotted spinor indices as

$$\mathcal{A} = (\dot\alpha, i)$$

so that we can combine the nontrivial equations as

$$[\nabla_{\mathcal{A}\alpha}, \nabla_{\mathcal{B}\beta}\} = iC_{\alpha\beta} f_{\mathcal{A}\mathcal{B}}$$

The former equations then allow us to interpret the remaining covariant derivatives $\overline{\nabla}^i{}_{\dot\alpha}$ as a subset of the SL(2|N) generators that rotate the $\mathcal{A}$ index, which form a subgroup of the superconformal group (S)SL(4|N). We therefore restrict ourselves to the chiral superspace described by the coordinates

$$z^{\mathcal{A}\alpha} = (x^{\alpha\dot\alpha}, \theta^{i\alpha})$$

The net result is that we have a superspace with no torsion, with coordinates that represent half of the supersymmetries as translations and the other half as rotations.

By comparison with our treatment of the self-dual bosonic theory in subsections IIIC5-7, we see that we can extend trivially all our results for the bosonic case to the (extended) supersymmetric case by simply extending the range of the indices. In particular, we also have a chiral twistor superspace: Extending the range on the twistor coordinates $z^{\mathcal{A}\alpha}$ used there so $\mathcal{A}$ is now an SL(4|N) index, the superconformal group is now manifest, and all the methods and results there (e.g., the ADHM construction) apply automatically to the supersymmetric case.

REFERENCES

1 Wess and Zumino, *loc. cit.* (IIC, ref. 2):
free scalar and vector multiplets, Wess-Zumino gauge.
2 J. Wess and B. Zumino, *Phys. Lett.* **49B** (1974) 52:
interacting chiral scalar multiplets.
3 J. Wess and B. Zumino, *Nucl. Phys.* **B78** (1974) 1:
coupling Abelian vector multiplet to scalar multiplets.
4 S. Ferrara and B. Zumino, *Nucl. Phys.* **B79** (1974) 413;
A. Salam and J. Strathdee, *Phys. Lett.* **51B** (1974) 102:
nonabelian vector multiplets.
5 J. Wess and B. Zumino, *Phys. Lett.* **66B** (1977) 361;
J. Wess, Supersymmetry-supergravity, in *Topics in quantum field theory and gauge theories*, VIII GIFT Int. Seminar on Theoretical Physics, Salamanca, Spain, Jun 13-19, 1977 (Springer-Verlag, 1978) p. 81:
covariant derivatives for super Yang-Mills.
6 P. Fayet and J. Iliopoulos, *Phys. Lett.* **51B** (1974) 461.
7 S.J. Gates, Jr. and H. Nishino, *Int. J. Mod. Phys. A* **8** (1993) 3371:
Chern-Simons superform.
8 L. O'Raifeartaigh, *Nucl. Phys.* **B96** (1975) 331.
9 L. Girardello and M.T. Grisaru, *Nucl. Phys.* **B194** (1982) 65:
general soft explicit supersymmetry breaking, with spurions.
10 L.A. Avdeev, D.I. Kazakov, and I.N. Kondrashuk, hep-ph/9709397, *Nucl. Phys.* **B510** (1998) 289;
N. Arkani-Hamed, G.F. Giudice, M.A. Luty, and R. Rattazzi, hep-ph/9803290, *Phys. Rev.* **D58** (1998) 115005:
spurions in couplings.

11 H.P. Nilles, *Phys. Rep.* **110** (1984) 1:
review of supersymmetry phenomenology.
12 P. Fayet, *Nucl. Phys.* **B113** (1976) 135:
extended supersymmetry.
13 P. Fayet, *Nucl. Phys.* **B149** (1979) 137:
central charges.
14 F. Del Aguila, M. Dugan, B. Grinstein, L. Hall, G. Ross, and P. West, *Nucl. Phys.* **B250** (1985) 225;
L. Girardello, M. Porrati, and A. Zaffaroni, hep-th/9704163, *Nucl. Phys.* **B505** (1997) 272:
semi-realistic N=2 models.
15 F. Gliozzi, J. Scherk, and D.I. Olive, *Nucl. Phys.* **B122** (1977) 253;
L. Brink, J.H. Schwarz, and J. Scherk, *Nucl. Phys.* **B121** (1977) 77:
N=4 super Yang-Mills.
16 R. Grimm, M. Sohnius, and J. Wess, *Nucl. Phys.* **B133** (1978) 275;
M.F. Sohnius, *Nucl. Phys.* **B136** (1978) 461;
E. Witten, *Phys. Lett.* **77B** (1978) 394:
covariant super Yang-Mills derivatives for extended supersymmetry.
17 W. Siegel, hep-th/9205075, *Phys. Rev.* **D46** (1992) R3235; hep-th/9207043, *Phys. Rev.* **D47** (1993) 2504; *loc. cit.* (IIIB, ref. 3):
conformal chiral superspace.
18 Gates, Grisaru, Roček, and Siegel, *loc. cit.*

PART TWO: QUANTA

Many important new features show up in field theory at the quantum level. Probably the most important is "renormalizability", which states that all the parameters (masses and couplings) that appear as coefficients of terms in the action must have nonnegative mass dimension (when the massless part of the kinetic term has no dimensionful coefficient). Since the action is dimensionless, $\int d^4x$ has dimension -4, and the fields have positive dimension, this allows only a small number of terms for any given set of fields. This one condition gives relativistic quantum field theory more predictive power than any known alternative.

There are many perturbation expansions that can be applied to quantum field theory. One is the mechanical JWKB expansion, which is an expansion in derivatives. Of the inherently field theoretical expansions, the simplest is to expand directly in fields, or equivalently, in the coupling constants. This expansion is the basis of perturbative quantum field theory. However, this expansion does not preserve gauge invariance term by term. On the other hand, the terms in this expansion can be collected into small subsets that do preserve gauge invariance. There are three such regroupings, discussed in the four following chapters, and they are based on perturbation expansions:

(1) the field theoretic JWKB ("loop") expansion,

(2) expansions in spin or helicity, and

(3) expansions in internal symmetry (color or flavor).

V. QUANTIZATION

For the most part, integrals are hard to evaluate, in particular the path integrals of exponentials that appear in quantum theory. The only exponentials that are generally easy to integrate are Gaussians, and the products of them times polynomials, which can in turn be evaluated as derivatives of Gaussians. Such integrals are the basis of perturbation theory: We keep the quadratic part of the action, but Taylor expand the exponential of higher-order terms. Effectively, this means that we not only expand in orders of $\hbar$ to perturb about the classical theory, but also expand in orders of the coupling constants to perturb about the free theory. This makes particularly useful our analysis of relativistic quantum mechanics (as free field theory). The JWKB expansion for the wave function (or S-matrix) expands the exponent in powers of $\hbar$, dividing it onto three qualitatively different parts:

(1) negative powers of $\hbar$ (generally $1/\hbar$ only), which describe the classical theory (they dominate the classical limit $\hbar \to 0$), whose physical implications have been considered in previous chapters;

(2) $\hbar$-independent, where almost all of the important (perturbative) quantum features appear (including topological ones, and quantum breaking of classical symmetries); and

(3) positive powers, which give more quantum corrections, but little new physics, except when summed to all orders.

These are generally known as "trees", "one-loop", and "multiloop", because of their graphical interpretation.

A. GENERAL

In the Schrödinger approach to quantum mechanics one solves a differential equation. The Feynman approach is complementary: There one performs an integral. Integrals are solutions to differential equations (e.g., $f' = g \Rightarrow f = \int g$), but usually differential equations are easier to solve than integral equations. However, there is an important exception: Gaussian integrals are easy, and so are their boundary conditions. In field theory the most important approximation is one where the integrand is approximated as a Gaussian, and the exact integral is evaluated as a perturbation about that Gaussian. Of course, solving the corresponding differential equation is also easy, but in that case the integral is easier because it corresponds to working with the action, while the differential equation corresponds to working with the field equations.

A major advantage of Feynman's approach is that it allows space and time to be treated on an equal footing. For example, as in classical electrodynamics, we can solve the wave equation inside a spacetime volume in terms of conditions on the boundary of that volume: It is not necessary to choose the spatial boundary at infinity so that it can be ignored, and divide the temporal boundary into its "future" and "past" halves so that all conditions are "initial" ones imposed at the past boundary. It is not even necessary to distinguish between preparation ("if") and measurement ("then") when describing probabilities: We can instead ask the probability of a given wave function describing the whole boundary. This is a particular advantage for relativistic quantum field theory, where space and time are more closely related than in nonrelativistic theories. We now "review" Feynman's approach for general quantum systems, and quantum mechanics in particular, so that it can be applied without further explanation when we come to quantum field theory.

1. Path integrals

Before discussing the path integral approach to quantum mechanics, we first review some features of quantum mechanics. We can separate the fundamentals of quantum mechanics into "kinematics" and "dynamics": The kinematics are everything at a fixed time — Hilbert space, preparation/measurement, probability, observables. The dynamics are the time development. There are several ways to describe time dependence of matrix elements; we will start with a general framework, then specialize.

Time dependence may be associated with either the states (Schrödinger picture) or operators (Heisenberg picture). We will be more explicit at first, taking all the time dependence out of the states and operators and putting it into a "time development operator" $U(t,t')$ that transforms the Hilbert space from time t' (earlier) to time t (later). For example, if we want to relate an earlier state to a later one we evaluate $\langle f|U(t,t')|i\rangle$; more generally, we can look at things like

$$\langle f|...\mathcal{O}_2 U(t_2,t_1)\mathcal{O}_1 U(t_1,t_i)|i\rangle$$

which means to prepare an initial state $|i\rangle$ at time t_i, then act with an operator $\mathcal{O}_1$ at time t_1, operator $\mathcal{O}_2$ at time t_2, etc., and eventually measure the amplitude for a final state $\langle f|$.

Now the dynamics can be described entirely through the properties of U. The general physical properties it must satisfy are

$$\begin{aligned} \textit{causality (locality)}: &\qquad U(t_3,t_2)U(t_2,t_1) = U(t_3,t_1) \\ \textit{unitarity}: &\qquad U(t_2,t_1)^\dagger U(t_2,t_1) = I \end{aligned}$$

Causality tells us that things happen in chronological order: Each event is determined by those immediately preceding. It is a kind of group property; in particular, from considering $t_3 = t_2$ we find that

$$U(t,t) = I$$

We can then write

$$U(t+\epsilon,t) \approx I - i\epsilon H(t)$$

by expanding in ϵ, for some operator $H(t)$ that we call the Hamiltonian. Again applying causality, we find

$$\begin{aligned} \partial_t U(t,t') &= \lim_{\epsilon\to 0} \frac{U(t+\epsilon,t') - U(t,t')}{\epsilon} \\ &= \left(\lim_{\epsilon\to 0} \frac{U(t+\epsilon,t) - I}{\epsilon}\right) U(t,t') = -iH(t)U(t,t') \end{aligned}$$

which is the Schrödinger equation for U. Again applying causality to build up the finite U from products of the infinitesimal ones,

$$U(t_f, t_i) = e^{-i\epsilon H(t_f - \epsilon)} \cdots e^{-i\epsilon H(t_i+\epsilon)} e^{-i\epsilon H(t_i)} \equiv \mathcal{T}\left\{ exp\left[-i \int_{t_i}^{t_f} dt\ H(t) \right] \right\}$$

which defines the "time-ordered product" $\mathcal{T}$. Finally, unitarity, another group property, tells us that probability is conserved; in particular, from applying to $U(t+\epsilon, t)$,

$$H(t)^\dagger = H(t)$$

The expression of U in terms of a hermitian Hamiltonian guarantees causality and unitarity. (It "solves" those conditions.) If H is time independent and we have a (orthonormal) basis of eigenstates of H, we can write

$$H|I\rangle = E_I|I\rangle \quad \Rightarrow \quad U(t,t') = \sum_I |I\rangle\langle I| e^{-i(t-t')E_I}$$

In Feynman's path integral approach to quantum mechanics (based on an analogy of Dirac), the action is the starting point for quantization. The basic idea is to begin with the basic quantity in quantum mechanics, the transition amplitude, and write it as an integral of the action

$$\langle f|i\rangle = \int D\phi\ e^{-iS[\phi]}$$

where $\int D\phi$ is a "functional integral": Integrate over $\phi(t)$ for each t (with some appropriate normalization). The boundary conditions in t are defined by the choice of initial and final states. In this subsection we will define this integral in a more explicit way by breaking up the time interval into discrete points and taking the continuum limit; in the next subsection we will study ways to evaluate it using its general properties.

The path integral can be derived from the usual Hamiltonian operator formalism. Considering for simplicity a single coordinate q, the wave function is given in coordinate space by

$$\psi(q) = \langle q|\psi\rangle, \qquad |\psi\rangle = \int \frac{dq}{\sqrt{2\pi}}\, \psi(q)|q\rangle$$

where we use the convenient normalizations

$$\int \frac{dq}{\sqrt{2\pi}} |q\rangle\langle q| = 1 = \int \frac{dp}{\sqrt{2\pi}} |p\rangle\langle p| \quad \left[\langle q|q'\rangle = \sqrt{2\pi}\delta(q-q'),\ \langle p|p'\rangle = \sqrt{2\pi}\delta(p-p') \right]$$

for coordinate and momentum space. To describe time development, we work in the Heisenberg picture, where time dependence is in the operators (and thus their eigenstates):

$$\psi(q,t) = \langle q,t|\psi\rangle$$

Time development is then given completely by the "propagator" or "Green function"

$$G(q_f,t_f;q_i,t_i) \equiv \langle q_f,t_f|q_i,t_i\rangle \quad \Rightarrow \quad \psi(q_f,t_f) = \int \frac{dq_i}{\sqrt{2\pi}} G(q_f,t_f;q_i,t_i)\psi(q_i,t_i)$$

Exercise VA1.1

Let's review the relationship between time development in the Heisenberg and Schrödinger pictures. Using the usual relation

$$\langle\psi|Q(t)|\chi\rangle \equiv \langle\psi(t)|Q|\chi(t)\rangle$$

between the time-independent states $|\psi\rangle$ and time-dependent operators $Q(t)$ of the Heisenberg picture and the time-dependent states $|\psi(t)\rangle$ and time-independent operators Q of the Schrödinger picture, define time-dependent eigenstates in two ways:

$$Q|q\rangle = q|q\rangle \quad \Rightarrow \quad \begin{cases} \langle q(t)|\psi(t)\rangle \equiv \langle q|\psi\rangle \\ \psi(q,t) \equiv \langle q|\psi(t)\rangle \equiv \langle q,t|\psi\rangle \end{cases}$$

Given the time development of a state

$$|\psi(t)\rangle = U(t)|\psi\rangle$$

($U(t) \equiv U(t,0)$), find the development of $Q(t)$, $|q(t)\rangle$, and $|q,t\rangle$, and show in particular that $|q(t)\rangle \neq |q,t\rangle$. Which is the eigenstate of $Q(t)$?

In general, even for time-dependent Hamiltonians, we can find the infinitesimal time development explicitly from the definition of the time derivative and the time-dependent Schrödinger equation:

$$[i\partial_t - H(-i\partial_q,q,t)]\langle q,t| = 0$$

$$\Rightarrow \quad \langle q,t+\epsilon| = \langle q,t|\{1 - i\epsilon H[P(t),Q(t),t]\} \approx \langle q,t|e^{-i\epsilon H[P(t),Q(t),t]}$$

and similarly for $\langle p,t+\epsilon|$ (where P and Q are the Hilbert-space operators). To derive the path-integral formalism, we then iterate this result to obtain finite time development by inserting unity infinitely many times, alternating between coordinate and momentum,

$$\langle q_f,t_f|q_i,t_i\rangle = \int \frac{dp_0}{\sqrt{2\pi}}\frac{dq_1}{\sqrt{2\pi}}\frac{dp_1}{\sqrt{2\pi}}...\langle q_f,t_f|...$$

$$...|p_1, t_i + 3\epsilon\rangle\langle p_1, t_i + 3\epsilon|q_1, t_i + 2\epsilon\rangle\langle q_1, t_i + 2\epsilon|p_0, t_i + \epsilon\rangle\langle p_0, t_i + \epsilon|q_i, t_i\rangle$$

to obtain successive infinitesimal exponentials,

$$\int \frac{dp_0}{\sqrt{2\pi}}\frac{dq_1}{\sqrt{2\pi}}\frac{dp_1}{\sqrt{2\pi}}...\langle q_f|e^{-i\epsilon H}...e^{-i\epsilon H}|p_1\rangle\langle p_1|e^{-i\epsilon H}|q_1\rangle\langle q_1|e^{-i\epsilon H}|p_0\rangle\langle p_0|e^{-i\epsilon H}|q_i\rangle$$

where the time dependence follows from the previous equation. However, note that all the implicit time dependence of the Heisenberg picture drops out, because we extracted the $e^{-i\epsilon H}$'s, putting all the factors of each matrix element at the same time: Although each matrix element is evaluated at time ϵ earlier than the one to its immediate left, each is of the form

$$\langle a, t + \epsilon|b, t\rangle = \langle a, t|e^{-i\epsilon H[P(t),Q(t),t]}|b, t\rangle = \langle a|e^{-i\epsilon H[P,Q,t]}|b\rangle$$

(where $|b\rangle \equiv |b, t_i\rangle$, etc.), leaving only any explicit time dependence that may appear in the Hamiltonian, effectively translating the other t's $\to t_i$. Then we only need to know

$$\langle q|p\rangle = e^{ipq}, \qquad \langle p|q\rangle = e^{-ipq}$$

to evaluate the matrix elements in the path integral as

$$\int \frac{dp_0}{\sqrt{2\pi}}\frac{dq_1}{\sqrt{2\pi}}\frac{dp_1}{\sqrt{2\pi}}...$$

$$exp\{-i[q_i p_0 + \epsilon H(p_0, q_i, t_i) - q_1 p_0 + \epsilon H(p_0, q_1, t_i + \epsilon) + q_1 p_1 + \epsilon H(p_1, q_1, t_i + 2\epsilon) + ...]\}$$

More explicitly, this result is

$$\langle q_f, t_f|q_i, t_i\rangle = \int Dp\ Dq\ e^{-iS}, \qquad Dp\ Dq = \prod_{n=0}^{N-1}\frac{dp_n}{\sqrt{2\pi}}\prod_{n=1}^{N-1}\frac{dq_n}{\sqrt{2\pi}}$$

$$S = \sum_{n=0}^{N-1}\{-(q_{n+1} - q_n)p_n + \epsilon[H(p_n, q_n, t_i + 2n\epsilon) + H(p_n, q_{n+1}, t_i + (2n+1)\epsilon)]\}$$

$$q_0 = q_i, \quad q_N = q_f; \qquad t_f - t_i = 2N\epsilon$$

Note that by adding (or subtracting) a step or two we could just as well evaluate $\langle q_f, t_f|p_i, t_i\rangle$ or $\langle p_f, t_f|q_i, t_i\rangle$ or $\langle p_f, t_f|p_i, t_i\rangle$.

The classical picture is a segmented path, with the particle traveling along a straight line segment from point q_n to point q_{n+1} with momentum p_n: Each q is associated with a point, while each p is associated with the line segment connecting two consecutive points. In the "continuum" limit $\epsilon \to 0$, $N \to \infty$, $t_f - t_i$ fixed,

$$S = \int_{t_i}^{t_f} dt[-\dot{q}p + H(p, q, t)]$$

(We have dropped some terms in $\langle q|H|p\rangle$ and $\langle p|H|q\rangle$ from reordering the operators Q and P in $H(P,Q)$ to apply $P|p\rangle = p|p\rangle$ and $Q|q\rangle = q|q\rangle$. These commutator terms alternate in sign, combining to give terms of order ϵ^2, and can be dropped in the continuum limit.)

More generally, we can evaluate an arbitrary transition amplitude as

$$\mathcal{A} = \langle f|i\rangle = \int \frac{dq_f}{\sqrt{2\pi}}\frac{dq_i}{\sqrt{2\pi}}\psi_f{}^*(q_f)\langle q_f,t_f|q_i,t_i\rangle\psi_i(q_i) = \int Dp\ Dq\ \psi_f{}^*(q_f)e^{-iS}\psi_i(q_i)$$

where now

$$Dp\ Dq = \prod_{n=0}^{N-1}\frac{dp_n}{\sqrt{2\pi}}\prod_{n=0}^{N}\frac{dq_n}{\sqrt{2\pi}}$$

Note that we can combine the initial and final wave function, as

$$\mathit{\Psi}(q_i,q_f) \equiv \psi_f{}^*(q_f)\psi_i(q_i) \quad \Rightarrow \quad \mathcal{A} = \int Dp\ Dq\ \mathit{\Psi}(q_i,q_f)e^{-iS}$$

The complex conjugation of ψ_f vs. ψ_i is due to the complex conjugation involved in time reversal (as seen, e.g., when comparing an eigenstate of p at the inital time to the same eigenstate at the final time). In field theory, where the "p's and q's" are functions of space as well as time, if we choose the boundary in space also to be finite, so that the space and time boundaries form a single connected and closed boundary, then $\mathit{\Psi}$ is simply a function of the q's over all that boundary.

We now see the relationship of the path integral approach to the time development operator: From the above derivation of the path integral, by integrating back out the insertions of unity immediately after extracting the infinitesimal exponentials and translating the time of each matrix element to zero, we find

$$\langle q,t_f|q,t_i\rangle = \langle q_f|U(t_f,t_i)|q_i\rangle$$

$$U(t_f,t_i) = e^{-i\epsilon H(t_f-\epsilon)}\cdots e^{-i\epsilon H(t_i+\epsilon)}e^{-i\epsilon H(t_i)} = \mathcal{T}\left\{exp\left[-i\int_{t_i}^{t_f} dt\ H(t)\right]\right\}$$

as previously. This is effectively a Schrödinger-picture expression (all the P's and Q's are at the initial time), and can also be derived in that picture by solving for the time dependence of any state $|\psi(t)\rangle$.

2. Semiclassical expansion

The path integral formulation is especially suited for semiclassical approximations: The Bohr-Sommerfeld quantization rule follows from the fact that the functional integral is invariant under $S \to S + 2\pi n$, since S appears only as e^{-iS}; in that sense the action is more like an angle than a single-valued function. The JWKB expansion follows from $S \to S/\hbar$ and expanding in $\hbar$. This expansion can be interpreted as an expansion in (space and time) derivatives, since it leads in the usual way to the identification $p = -i\hbar\partial/\partial x$ and $E = i\hbar\partial/\partial t$.

Exercise VA2.1

For comparison, we review the Schrödinger equation approach. Consider the nonrelativistic JWKB expansion for the propagator (for an arbitrary Hamiltonian H) to the first two orders in $\hbar$, writing it as

$$G \approx \sqrt{\rho}e^{-iS/\hbar}$$

a Show the corresponding orders in the time-dependent Schrödinger equation at $t > 0$ can be written as the classical equation of motion for the action S and the (probability) current conservation law for the (probability) density ρ ("Hamilton-Jacobi equations"),

$$H = \dot{S}, \qquad \frac{\partial}{\partial q^i}\left(\rho\frac{\partial H}{\partial p_i}\right) + \dot{\rho} = 0$$

when the argument p of H is evaluated at

$$p_i = -\frac{\partial S}{\partial q^i}$$

(Assume a symmetric ordering of p's and q's in the quantum H.) Compare the *relativistic* case examined in exercise IIIA4.1.

b The propagator is expressed in terms of q and q_0, where $G(q, q_0, t) \sim \delta(q - q_0)$ at $t = 0$, so the first order in $\hbar$ is found by using the solution to the Hamilton-Jacobi equations to write the classical action in terms of the "final" position q and initial position q_0. (In principle; in general even the classical equations may be too difficult to solve analytically.) However, the Hamiltonian is given as a function of p and q. Show that the change in variables from q, p to q, q_0 gives

$$\frac{\partial H}{\partial p_i} = -(M^{-1})^{ij}\frac{\partial^2 S}{\partial q_0^j \partial t}, \qquad (M)_{ij} = \frac{\partial^2 S}{\partial q_0^i \partial q^j}$$

Show that

$$\rho = det\ (-i\tfrac{1}{\hbar}M)$$

(the "van Vleck determinant") solves the current conservation law, using the explicit expression for $(det\ M)M^{-1}$ given in subsection IB3. Check the normalization, using the initial condition for propagators (or comparing to the free case).

One way to apply the path integral is as follows: (1) Find a classical solution to the equations of motion. This gives the leading contribution in $\hbar$ ("stationary phase approximation"),

$$\int D\phi\ e^{-iS/\hbar} \approx e^{-iS_0/\hbar}$$

(The validity of such an approximation with an imaginary exponent will be discussed in subsection VA5.)

(2) Expand about the classical solution as

$$\phi = \phi_{cl} + \sqrt{\hbar}\Delta\phi$$

Expanding in $\Delta\phi$ (or $\hbar$), we have schematically

$$\hbar^{-1}S = \hbar^{-1}S_0 + \hbar^{-1/2}S_0'\Delta\phi + \tfrac{1}{2}S_0''(\Delta\phi)^2 + \hbar^{1/2}\tfrac{1}{6}S_0'''(\Delta\phi)^3 + ...$$

where "$_0$" means to evaluate at $\phi = \phi_{cl}$ and the derivatives are really functional derivatives (so there is also an integral for each derivative). The first term in the action gives the classical contribution, while the linear term vanishes by the equations of motion. The quadratic term gives an $\hbar$-independent contribution to the exponential, so the next order approximation to the functional integral comes from integrating just that: Integrating Gaussians as in subsection IB3,

$$\int D\phi\ e^{-iS/\hbar} \approx e^{-iS_0/\hbar}(det\ S_0'')^{-1/2}$$

where the determinant is now a functional one, which can be defined by performing the functional integral as in the previous section, as a series of ordinary Gaussian integrals. The boundary conditions are $\Delta\phi = 0$ at t_i and t_f (since $\phi_{cl} = \phi$ there). Normalization constants can be determined by comparing the free case, or considering the limit where the initial and final times converge.

(3) We then expand the exponential in the cubic and higher terms (positive powers of $\hbar$): The resulting functional integral is that of an $\hbar$-independent Gaussian times a polynomial with positive powers of $\hbar$. Since odd orders in $\Delta\phi$ vanish by symmetry $(\Delta\phi \to -\Delta\phi)$, only integer powers of $\hbar$ appear:

$$\int D\phi\ e^{-iS/\hbar} = e^{-iS_0/\hbar}\int D(\Delta\phi)e^{-iS_0''(\Delta\phi)^2/2}\left(1 + \sum_{n=1}^{\infty}\hbar^n f_n[\Delta\phi]\right)$$

Polynomials times Gaussians are also straightforward to integrate: The easiest way is to first evaluate integrals of Gaussians with linear terms:

$$\int \frac{d^D x}{(2\pi)^{D/2}} e^{-x^T S x/2 + j^T x} = (det\ S)^{-1/2} e^{j^T S^{-1} j/2}$$

$$\int \frac{d^D z^*\ d^D z}{(2\pi i)^D} e^{-z^\dagger H z + z^\dagger j + j^\dagger z} = (det\ H)^{-1} e^{j^\dagger H^{-1} j}$$

from shifting the integration variables ($x \to x + S^{-1}j$, etc.) to eliminate the linear terms, then using the previous results. In functions of x multiplying the Gaussian, x can be replaced with $\partial/\partial j$ (and similarly for z) and then pulled outside the integral. (If a linear term is not included, it can be introduced, and the result can be evaluated at $j = 0$.) The final result then takes the form

$$\langle q_f, t_f | q_i, t_i \rangle = e^{-iS_0/\hbar}(A + \hbar B + ...) = exp\left(-i\tfrac{1}{\hbar} \sum_{n=0}^{\infty} \hbar^n S_n \right)$$

Exercise VA2.2

Generalize the above results for integration of Gaussians with linear terms to the cases with fermionic and mixed (subsection IIC3) integration variables.

Exercise VA2.3

Evaluate

$$\int \frac{d^D x}{(2\pi)^{D/2}} e^{-x^T S x/2} x^i x^j x^k x^l$$

by taking $(\partial/\partial j_i)(\partial/\partial j_j)(\partial/\partial j_k)(\partial/\partial j_l)$ on the above result.

As an example, consider the free nonrelativistic particle. The separability of the action translates into factorization of the functional integral, so the result can be found from the one-dimensional case. As usual,

$$L = -\tfrac{1}{2} m \dot{x}^2 \quad \Rightarrow \quad x_{cl}(t) = x_i + \frac{x_f - x_i}{t_f - t_i}(t - t_i)$$

where we have written the classical solution in terms of the variables appropriate to the initial and final states, namely x_i for an initial state localized there at time t_i, and x_f, t_f for the final state. Since the classical action is itself quadratic, so is its expansion:

$$S = S_{cl} + \Delta S, \qquad S_{cl} = -\tfrac{1}{2} m \frac{(x_f - x_i)^2}{t_f - t_i}, \qquad \Delta S = -\int dt\ \tfrac{1}{2} m (\Delta\dot{x})^2$$

In general, a determinant from the ΔS integral must be evaluated (but see exercise VA2.1). In this simple case, time translation invariance, dimensional analysis, and independence from x_f, x_i are enough to determine the result of that functional integral up to a constant, fixed by the short-time limit $t_f \to t_i$. The final one-dimensional result is then

$$\langle x_f, t_f | x_i, t_i \rangle = \sqrt{\frac{-im}{t_f - t_i}}\, e^{im(x_f - x_i)^2/2(t_f - t_i)}$$

where we have used

$$\sqrt{2\pi}\delta(x) = \lim_{\epsilon\to 0} \tfrac{1}{\sqrt{\epsilon}} e^{-x^2/2\epsilon}$$

(one way of defining a Dirac δ function) to normalize

$$\langle x_f, t | x_i, t \rangle = \sqrt{2\pi}\delta(x_f - x_i)$$

The Gaussian integral for the free particle can also be performed explicitly, by using the discretized Hamiltonian path integral of the previous subsection.

Exercise VA2.3

The path integral for the free, nonrelativistic particle can be evaluated much more easily using the Hamiltonian form of the action. First consider the Gaussian integral

$$\int_{-\infty}^{\infty} dx\ e^{ipx - \epsilon x^2/2}$$

as a special case of the Gaussians already evaluated, and use it to derive the identity

$$\int_{-\infty}^{\infty} dx\ e^{ipx} = 2\pi\delta(p)$$

(The ϵ thus acts as a regulator to make the integral well defined.) Then use the discretized expression of subsection VA1, and evaluate the x integrals first. All but one of the p integrals then can be trivially evaluated, the last giving a Fourier transform.

Exercise VA2.4

Consider the one-dimensional harmonic oscillator. (The multi-dimesional case is again separable.)

a Explicitly evaluate the discretized path integral to find the result

$$\langle x_f, t_i + t | x_i, t_i \rangle = \sqrt{\frac{-im\omega}{sin\ \omega t}}\ exp\left\{ \frac{im\omega[\frac{1}{2}(x_f^2 + x_i^2)cos\ \omega t - x_f x_i]}{sin\ \omega t} \right\}$$

b Rederive the result using the result of exercise VA2.1. (Hint: First solve the classical equations of motion for $x(t)$, then rewrite it in terms of $x_i = q_0$ and $x_f = q$; plug into $S_{cl} = S$ and apply the above.)

Note that we have been sloppy about the definition of the "integration measure": In going from the Hamiltonian form of the action to the Lagrangian form, we ignored some m dependence. Specifically, if we start with the Hamiltonian form, as derived in the previous subsection, and derive the Lagrangian form by integrating out p, we find the $1/m$ in $H = p^2/2m$ leads to

$$\prod_{n=0}^{N-1} \frac{dp_n}{\sqrt{2\pi}} \prod_{n=1}^{N-1} \frac{dx_n}{\sqrt{2\pi}} \quad \to \quad m^{N/2} \prod_{n=1}^{N-1} \frac{dx_n}{\sqrt{2\pi}}$$

The $m^{(N-1)/2}$ then cancels similar factors from the $N-1$ x-integrals, while the remaining $\sqrt{m}$ is that found in the final result above.

If we had considered a more general Hamiltonian, as in subsection IIIA1, where p^2 appeared as $\frac{1}{2}g^{ij}(x)p_i p_j$, then we would have obtained a measure of the form (for $i = 1, ..., D$)

$$[det\ g(x_0) det\ g(x_N)]^{-1/4} \prod_{n=1}^{N-1} \frac{d^D x_n}{(2\pi)^{D/2}\sqrt{det\ g(x_n)}}$$

(We have averaged g as $g(x)p^2 \to \sqrt{g(x_n)g(x_{n+1})}p_n^2$, since x_n is associated with the point n while p_n is associated with the link from n to $n+1$.) Such measure factors are easy to recognize, since they are always local, without any derivatives: If we included it in the action, it would be a term proportional to

$$ln \prod_n det\ g(x_n) = \tfrac{1}{\epsilon} \sum_n \epsilon\ ln\ det\ g(x_n) \sim \delta(0) \int dt\ ln\ det\ g(x(t))$$

(The factors at x_0 and x_N are for standard normalization of the wave functions, which we can absorb by a redefinition.) In practice we just drop all such factors throughout the calculation, and fix the normalization at the end of the calculation. Since the Lagrangian form follows from the Hamiltonian form, which was properly normalized, we know such factors will cancel anyway. Auxiliary fields can require similar factors for proper normalization; then such factors are simply the Jacobians from the field redefinitions from a form where they appeared with trivial quadratic terms.

3. Propagators

The amplitude we defined by path integration in subsection VA1 is the "propagator" or "Green function" for the Schrödinger equation. Explicitly, we define

$$G(q, t; q', t') \equiv \theta(t - t')\langle q, t|q', t'\rangle$$

where we have included the "step function" $\theta(t - t')$ (1 for $t > t'$, 0 otherwise) to enforce that the final time is later than the initial time (retarded propagator). This satisfies the free case of the general defining equation of the propagator

$$[\partial_t + iH(-i\partial_q, q, t)]G(q, t; q', t') = [-\partial_{t'} + iH(i\partial_{q'}, q', t')]G(q, t; q', t')$$

$$= \sqrt{2\pi}\delta(q-q')\delta(t-t')$$

where we have used

$$\partial_t \theta(t-t') = \delta(t-t')$$

and the facts that G without the θ factor is a homogeneous solution of the Schrödinger equation (no δ's) and becomes a δ in x for small times. The propagator then gives a general solution of the Schrödinger equation as

$$\langle q,t| = \int \frac{dq'}{\sqrt{2\pi}} \langle q,t|q',t'\rangle\langle q',t'| \quad \Rightarrow \quad \psi(q,t) = \int \frac{dq'}{\sqrt{2\pi}} G(q,t;q',t')\psi(q',t')$$

In particular, for $\psi(q,t') = \sqrt{2\pi}\delta(q-q')$ at some time t' for some point q', $\psi(q,t) = G(q,t;q',t')$ at all later times. These equations are matrix elements of the corresponding operator equations; e.g.,

$$G(q,t;q',t') \equiv \langle q|U(t,t')|q'\rangle$$

$$[\partial_t + iH(t)]U(t,t') = U(t,t')[-\overleftarrow{\partial}_{t'} + iH(t')] = \delta(t-t')I$$

where we now include a step function in the definition of the time development operator U:

$$U(t,t') \equiv \theta(t-t')\mathcal{T}\left\{exp\left[-i\int_{t'}^{t} dt\ H(t)\right]\right\}$$

This solution for the propagator is not unique; as usual, a first-order differential equation needs one boundary condition. Another way to say it is that the inhomogeneous differential equation is arbitrary up to a solution of the homogeneous equation. We have eliminated the ambiguity by requiring that the propagator be retarded, as incorporated in the factor $\theta(t-t')$; using instead $-\theta(t'-t)$ would give the advanced propagator.

This has an interesting translation in terms of the Fourier transform with respect to the time, which replaces the so-called "time-dependent" Schrödinger equation with the "time-independent" one. Fourier transforms are a useful way to solve differential equations when performed with respect to variables with translational invariance, since this implies conservation of the conjugate variable: The result is elimination of the corresponding derivatives. In this case, it means the time-independent Schrödinger equation needs a time-independent Hamiltonian. For example, defining

$$\tilde{U}(E,E') \equiv \int \frac{dt}{\sqrt{2\pi}}\frac{dt'}{\sqrt{2\pi}} e^{-i(E't'-Et)}U(t,t')$$

$$\Rightarrow \quad -i(E-H)\tilde{U}(E,E') = \delta(E-E')I$$

$$\Rightarrow \quad \tilde{U}(E,E') = \frac{i}{E-H}\delta(E-E')$$

Now inverse Fourier transforming,

$$U(t,t') = \int \frac{dE}{\sqrt{2\pi}}\frac{dE'}{\sqrt{2\pi}} e^{i(E't'-Et)}\tilde{U}(E,E')$$

$$= \int \frac{dE}{2\pi} e^{-iE(t-t')}\frac{i}{E-H}$$

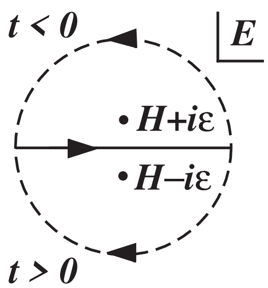
we have an ambiguity in integrating E past the pole at $E = H$. We therefore shift the pole slightly off the real axis, so we can integrate exactly on the real axis. Closing the contour by adding to the real axis a semicircle of infinite radius in either the complex upper- or lower-half-plane, wherever convergent ($\lim_{|t|\to\infty} e^{-|Et|} = 0$, but $\lim_{|t|\to\infty} e^{+|Et|} = \infty$), we find

$$\int \frac{dE}{2\pi} e^{-iEt}\frac{i}{E-H\pm i\epsilon} = \pm\theta(\pm t)e^{-iHt}$$

which gives either the retarded or advanced propagator depending on the choice of sign for the infinitesimal constant ϵ (retarded for $E - H + i\epsilon$). Remember from exercise IIA1.2 that complex integration is essentially just Gauss' law, wth poles acting as charges: The general integral result we used is

$$\oint \frac{dz'}{2\pi i} f(z')\frac{1}{z'-z} = f(z)$$

where the counterclockwise contour of integration encloses the pole at z but no singularity in f, so we can evaluate the integral by Taylor expanding f about z.



To perform the inverse Fourier transform, we note that the exponent needs an infinitesimal negative part to make the integral convergent:

$$\int dt\ e^{iEt}(\pm)\theta(\pm t)e^{-iHt\mp\epsilon t} = \frac{i}{E-H\pm i\epsilon}$$

Exercise VA3.1

Show that

$$\frac{i}{x+i\epsilon} - \frac{i}{x-i\epsilon} = 2\pi\delta(x)$$

by three methods:

a Use the above result for the Fourier transform.

b Show that this is the contour integral definition of the δ function, which is actually a distribution, by integration, multiplying by an arbitrary (nonsingular) function and integrating along the real axis. (Hint: Push the poles onto the real axis, shifting the contours along with them, to find the integral of a single function along the difference of two contours.)

c Prove the identity (checking the normalization)

$$\lim_{\epsilon\to 0} \frac{2\epsilon}{x^2+\epsilon^2} = 2\pi\delta(x)$$

For the example of the free particle in one dimension we found by various methods

$$G(x,t;x',t') = \theta(t-t')\sqrt{\frac{-im}{t-t'}}\, e^{im(x-x')^2/2(t-t')}$$

However, we could have saved the trouble if we just started in momentum space,

$$\hat{G}(p,t;p',t') \equiv \langle p|U(t,t')|p'\rangle = \langle p|\theta(t-t')e^{-i(t-t')H}|p'\rangle = \theta(t-t')e^{-i(t-t')p^2/2m}\langle p|p'\rangle$$
$$= \theta(t-t')\sqrt{2\pi}\delta(p-p')e^{-i(t-t')p^2/2m}$$

in the retarded case. If we Fourier transform p to x (the same as a change of basis from $|p\rangle$ to $|x\rangle$), the integrals are then simple Gaussians. Again, the result is simpler in p-space because p is conserved. In the relativistic case we will want to treat energy and momentum equally; doing the same here for later comparison, we define

$$\tilde{\psi}(p,E) = \int \frac{dq}{\sqrt{2\pi}}\frac{dt}{\sqrt{2\pi}} e^{-i(qx-Et)}\psi(q,t)$$

and similarly for $\tilde{G}$, and we have

$$\tilde{G}(p,E;p',E') = \frac{i}{E-p^2/2m}\sqrt{2\pi}\delta(p-p')\delta(E-E')$$

4. S-matrices

"Scattering" is defined as a process that starts with a free state and ends with a free state, with interaction (self- or with external forces) at intermediate times, e.g., particles coming in from and going out to spatial infinity and scattering from a potential of finite spatial extent. Thus, if the interaction is nonvanishing somewhere between times t_1 and t_2, where $t_f > t_2 > t_1 > t_i$, we can write

$$U(t_f,t_i) = U(t_f,t_2)U(t_2,t_1)U(t_1,t_i) = e^{-i(t_f-t_2)H_0}U(t_2,t_1)e^{-i(t_1-t_i)H_0}$$

in terms of the "free term" H_0 of the Hamiltonian $H = H_0 + H_I$, where H_I is the "interaction term". (H_I may be time dependent, but not H_0.) It is more convenient to work with a quantity that is independent of initial and final times (as long as they are outside of the interaction region t_1 to t_2). We therefore define the "S(cattering)-matrix" operator $\mathcal{S}$ as

$$\mathcal{S} \equiv \lim_{\substack{t_i\to-\infty \\ t_f\to+\infty}} e^{it_fH_0}U(t_f,t_i)e^{-it_iH_0}$$

where we have thrown in the limit because in the real world interaction doesn't just start and stop, but fades in and out. However, in our simple example above we find

$$\mathcal{S} = e^{it_2H_0}U(t_2,t_1)e^{-it_1H_0}$$

In the interacting case, the amplitude we get from the path integral is the interacting propagator. To be able to take the limit describing time development between infinite initial and final times, we need to choose boundary conditions such that the initial and final basis states have the time dependence of free particles, described by H_0, assuming that the particle behaves freely at such asymptotically large times. This is called the "interaction picture", to distinguish from the Heisenberg picture, where the states have no time dependence, and the Schrödinger picture, where the states have the complete interacting time dependence. We thus evaluate the limiting amplitude

$$\mathcal{A} = \lim_{\substack{t_i\to-\infty \\ t_f\to+\infty}} \langle\psi_f(t_f)|\psi_i(t_i)\rangle = \lim_{\substack{t_i\to-\infty \\ t_f\to+\infty}} \int \frac{dq_f}{\sqrt{2\pi}}\frac{dq_i}{\sqrt{2\pi}}\psi_f{}^*(q_f,t_f)\langle q_f,t_f|q_i,t_i\rangle\psi_i(q_i,t_i)$$

for the interaction-picture states $|\psi(t)\rangle$, relating the interaction-picture coordinate basis ${}_0\langle q,t|$ to the Heisenberg-picture basis $\langle q,t|$ (with initial conditions ${}_0\langle q,0| = \langle q,0| \equiv \langle q|$):

$${}_0\langle q,t| = \langle q|e^{-itH_0} \quad \Rightarrow \quad \langle q_f,t_f|q_i,t_i\rangle = {}_0\langle q_f,t_f|e^{it_fH_0}U(t_f,t_i)e^{-it_iH_0}|q_i,t_i\rangle_0$$

$$\psi(q,t) = {}_0\langle q,t|\psi\rangle \quad \Rightarrow \quad \mathcal{A} = \langle\psi_f|\mathcal{S}|\psi_i\rangle$$

with $\mathcal{S}$ as defined above.

The fact that time development conserves probability ($H = H^\dagger$) is reflected in the corresponding *unitarity* condition for the S-matrix:

$$\mathcal{S}^\dagger\mathcal{S} = 1$$

A more complicated condition is *causality*: The basic idea is that interactions take place in chronological order. (A stronger statement of causality will be found in the relativistic case: that any interaction should take place at a spacetime point, rather than just at a single time. It follows from this weaker one in relativistic theories, since event B is later than event A in every Lorentz frame only when B is in A's lightcone.) Causality is the condition that the Hamiltonian at any time involves only variables evaluated at that time. ($H(t)$ is a function of only $\phi(t)$, all at the same time t, where $\phi = p, q$ are the quantum variables appearing in the Hamiltonian.) A nice way to describe the interactions is by introducing a classical background as we did for the semiclassical expansion of path integrals, such as by $\phi(t) \to \phi(t) + \chi(t)$, where χ is just some function. The important point is that we have shifted $\phi(t)$ by $\chi(t)$ at the same t, so as not to disturb causality. We then consider the effect on the S-matrix of modifying the background χ by a function $\delta\chi$ localized (nonvanishing) at some particular time t, and a function $\delta\chi'$ localized at t', such that $t > t'$. Picking out the $\delta\chi$ pieces in the time-ordered product, we can therefore write

$$\begin{aligned}
\mathcal{S}[\chi + \delta\chi + \delta\chi'] &= U(f,t)\mathcal{V}(t)U(t,t')\mathcal{V}(t')U(t',i) \\
\mathcal{S}[\chi + \delta\chi] &= U(f,t)\mathcal{V}(t)U(t,t')U(t',i) \\
\mathcal{S}[\chi + \delta\chi'] &= U(f,t)U(t,t')\mathcal{V}(t')U(t',i) \\
\mathcal{S}[\chi] &= U(f,t)U(t,t')U(t',i)
\end{aligned}$$

where $U(t', i)$ is the time-development operator from time t_i to time t' (including the canceling factor with H_0), $\mathcal{V}(t')$ is the extra factor in the time development at time t' resulting from the function $\delta\chi'$ localized there, etc. Thus we replace a $\mathcal{V}$ with the identity if the corresponding $\delta\chi$ is absent. Then we easily find

$$\begin{aligned}
&\mathcal{S}[\chi + \delta\chi + \delta\chi'] = \mathcal{S}[\chi + \delta\chi]\mathcal{S}^{-1}[\chi]\mathcal{S}[\chi + \delta\chi'] \\
\Rightarrow \quad &(\mathcal{S}^{-1}[\chi + \delta\chi]\mathcal{S}[\chi + \delta\chi + \delta\chi'] - I) - (\mathcal{S}^{-1}[\chi]\mathcal{S}[\chi + \delta\chi'] - I) = 0 \\
\Rightarrow \quad &\frac{\delta}{\delta\chi(t)}\left(\mathcal{S}[\chi]^\dagger \frac{\delta}{\delta\chi(t')}\mathcal{S}[\chi]\right) = 0 \quad for\ t > t'
\end{aligned}$$

using the infinitesimal functions $\delta\chi$ and $\delta\chi'$ to define functional derivatives (as in subsection IIIA1 for the action).

In general, it is not possible to solve the Schrödinger equation for the propagator or the S-matrix exactly. One approximation scheme is the perturbation expansion in orders of the interaction:

$$\begin{aligned}
H = H_0 + V \quad \Rightarrow \quad \mathcal{T}(e^{-i\int dt\ H}) &= e^{-i(t_f - t_i)H_0} + \int_{t_i}^{t_f} dt\ e^{-i(t_f - t)H_0}[-iV(t)]e^{-i(t - t_i)H_0} \\
&+ \int_{t_i}^{t_f} dt \int_{t_i}^{t} dt'\ e^{-i(t_f - t)H_0}[-iV(t)]e^{-i(t - t')H_0}[-iV(t')]e^{-i(t' - t_i)H_0} + ...
\end{aligned}$$

$$\Rightarrow \quad \mathcal{S}_{fi} \equiv \langle f|\mathcal{S}|i\rangle = \langle f|i\rangle + \int_{-\infty}^{\infty} dt\ \langle f,t|[-iV(t)]|i,t\rangle$$
$$+ \int_{-\infty}^{\infty} dt \int_{-\infty}^{t} dt'\ \langle f,t|[-iV(t)]e^{-i(t-t')H_0}[-iV(t')]|i,t'\rangle + ...$$

The first term in $\mathcal{S}$ is just the identity (i.e., the free piece). All the other terms consist of a string of interactions ($-iV$) connected by free propagators (e^{-itH_0}, where t is the time between the interactions), with each interaction integrated over all time (subject to time-ordering of the interactions), and the initial/final state (wave function) evaluated at the initial/final interaction time.

Exercise VA4.1

Assume the initial and final states are eigenstates of the free Hamiltonian:

$$H_0|i\rangle = E_i|i\rangle, \qquad H_0|f\rangle = E_f|f\rangle$$

Assuming V has no explicit time dependence, explicitly evaluate the time integrals in the S-matrix, effectively Fourier transforming from time to energy, to find

$$\mathcal{S}_{fi} = \langle f|i\rangle - 2\pi i\delta(E_f - E_i)\langle f|(E - H_0)\frac{1}{E - H + i\epsilon}(E - H_0)|i\rangle|_{E=E_i}$$

(Hints: Redefine the integration variables to be the times between interactions. Taylor expand $1/(E - H + i\epsilon)$ in V for comparison.)

In field theory we want to express any state in terms of a basis of products of 1-particle states, so we can calculate the behavior of these specified particles. We try to do this by using field variables (the "q's" of field theory): Each field operator should produce a single particle. Unfortunately, this is not the case: An asymptotic state of given 3-momentum created by such a field operator is not necessarily an eigenstate of the energy, because such a state can be either 1-particle or n-particle, due to interactions. The propagator for the field is then of the form

$$\hat{G}(p,t;p',t') \sim \delta(p-p')\sum_I \psi*_I(p)\psi_I(p)e^{-i(t-t')E_I(p)}$$

$$E_I(p) = \sum_{i=1}^{n_I} E_{I,i}(p_i), \qquad \sum_{i=1}^{n_I} p_i = p$$

where "$E_{I,i}(p_i)$" is the energy of a 1-particle state (the $\sum_I$ will include an integral in general). However, as long as all particles have masses, such an asymptotic 1-particle state is distinguishable as that of lowest energy E_0: The higher-energy states are n-particle states to which this particle can couple. (If some of the n-particle states were lower energy, the 1-particle state could decay into them, and thus the 1-particle state would be unstable, and not asymptotic. With massless particles things are more complicated: Then 1-particle states are more difficult to define and to measure.) In principle, we could define the 1-particle states by constructing the corresponding operator, consisting of the field plus terms higher order in the fields; in practice, this is rather complicated.

A simpler way to make the asymptotic states unambiguous is by modifying the definition of the S-matrix:

$$\mathcal{S} = \lim_{\substack{t_i \to -\infty(1+i\epsilon) \\ t_f \to +\infty(1+i\epsilon)}} e^{it_f H_0}\, \mathcal{T}\left(e^{-i\int_{t_i}^{t_f} dt\ H} \right) e^{-it_i H_0}$$

introducing factors of $1 + i\epsilon$ for some positive ϵ, which may be chosen small for convenience. (Actually, we can generally replace $1 + i\epsilon$ with just i if it is not too confusing: The result is the same.) The effect is seen by considering a matrix element of particular fields that may be a superposition of different energies E in the initial state and E' in the final state, but evaluated between an initial state of energy E_i and a final state of energy E_f (which might not be equal for a time-dependent interaction, e.g., if the number of particles changes). Since $E \geq E_i$ initially and $E' \geq E_f$ finally, the time dependence of any such matrix element is proportional to

$$\mathcal{S}_{fi} \sim \lim_{t_f \to +\infty(1+i\epsilon)} e^{it_f(E'-E_f)} \lim_{t_i \to -\infty(1+i\epsilon)} e^{-it_i(E-E_i)} = \begin{cases} 1 & for\ E = E_i, E' = E_f \\ 0 & otherwise \end{cases}$$

Alternatively, we can simply impose $E = E_i, E' = E_f$ directly in the definition:

$$\mathcal{S} = \lim_{\substack{t_i \to -\infty \\ t_f \to +\infty}} e^{it_f H_0} \delta_{H(t_f),H_0}\, \mathcal{T}\left(e^{-i\int_{t_i}^{t_f} dt\ H(t)} \right) \delta_{H(t_i),H_0} e^{-it_i H_0}$$

where the free Schrödinger equation $H_0 = E_i$ or E_f defines E_i for the initial state and E_f for the final state, and δ_{H,H_0} is evaluated by examining the asymptotic time-dependence of the time-development operator with respect to t_i and t_f: Normally field theory is calculated in energy-momentum space, working with the spacetime Fourier transform of the above, where this amounts to simply comparing energies $E = E_i, E' = E_f$.

If we know some details of the interaction, this modification may be irrelevant: In particular, in local quantum field theory interactions happen at a point in space and time. For example, consider the inner product between a 1-particle state in its rest frame and a related n-particle state, which appears in the same propagator. Because of locality, the wave function for the n-particle state, when evaluated in position space (which is where the theory is local) is simply the product of n 1-particle wave functions evaluated at the same point. But we know that for small relative momenta (where a nonrelativistic approximation holds) that the individual wave functions propagate as

$$|\psi| \sim |t - t'|^{-(D-1)/2}$$

from the form of the free 1-particle propagator. (Or, we can use dimensional analysis, and consider the spread of a particle of restricted range of momenta from a confined region: Then $|\psi|^2 \sim 1/V$ and the volume $V \sim |t - t'|^{D-1}$.) This implies that the n-particle wave function will fall off as the nth power of that, so in the limit of large times the 1-particle state will dominate. In a relativistic theory the length scale associated with this fall-off will be associated with the masses involved, and thus at a subatomic scale.

5. Wick rotation

In the previous subsection we ensured convergence in the definition of the S-matrix by effectively making the "coordinate change"

$$t \to (1 - i\epsilon)t = e^{-i\epsilon}t$$

in the definition of the limit

$$(1 - i\epsilon)t \to \infty \quad \Rightarrow \quad t \to (1 + i\epsilon)\infty$$

This affected the time-development operator as

$$e^{-iHt} \to e^{-iHt - \epsilon t}$$

for $H > 0$ to pick out the ground state $H = 0$. The same effective substitution was made in subsection VA3 in defining the contour integral for the propagator:

$$\int \frac{dE}{2\pi} e^{-iE(1-i\epsilon)t} \frac{i}{E - H} = \int \frac{dE}{2\pi} e^{-iEt} \frac{i}{(1 + i\epsilon)E - H}$$

$$= \int \frac{dE}{2\pi} e^{-iEt} \frac{i}{E - (1 - i\epsilon)H} = \int \frac{dE}{2\pi} e^{-iEt} \frac{i}{E - H + i\epsilon}$$

which is the same as the substitution

$$E \to (1 + i\epsilon)E = e^{i\epsilon}E$$

(when working with the time-independent Schrödinger equation) since essentially $E = i\partial/\partial t$.

In general, having to do contour integrals and keep track of $i\epsilon$'s in propagators is inconvenient. Fortunately, there is a simple way in practical calculations to get rid of not only the $i\epsilon$'s but (almost) all the other i's as well. The method is known as "Wick rotation". The basic idea is to extend the above complex rotation from angle ϵ to angle $\pi/2$:

$$t \to -it = e^{-i\pi/2}t, \qquad E \to iE$$

pushing the contour even farther away from the singularities. Thus, the Schrödinger equation is changed to a "diffusion equation" (to describe, e.g., Brownian motion):

$$(i\partial_t - H)\psi = 0 \quad \Rightarrow \quad (\partial_t + H)\psi = 0$$

For example, for the free particle the resulting equation has no i's. The time-independent Schrödinger equation then becomes

$$(E - H)\psi = 0 \quad \Rightarrow \quad (iE - H)\psi = 0$$

The result for the propagator is then

$$\int_{-\infty}^{\infty} \frac{dE}{2\pi}\; e^{-iEt}\frac{1}{H - iE} = \theta(t)e^{-Ht}$$

Now no $i\epsilon$ prescription is needed, since the pole was moved away from the real axis. Similar remarks apply to the inverse Fourier transform

$$\int_{-\infty}^{\infty} dt\; e^{iEt}\theta(t)e^{-Ht} = \frac{1}{H - iE}$$

Exercise VA5.1

Find the Wick-rotated retarded propagator $G(x',t';x,t)$ for the free (1D) particle, satisfying

$$(\partial_t + H)G = (-\partial_{t'} + H')G = \sqrt{2\pi}\delta(x - x')\delta(t - t')$$

Furthermore, if we define the S-matrix directly in this Wick-rotated space

$$\mathcal{S} = \lim_{\substack{t_i \to -\infty \\ t_f \to +\infty}} e^{t_f H_0}\, \mathcal{T}\left(e^{-\int_{t_i}^{t_f} dt\ H}\right) e^{-t_i H_0}$$

then the limiting procedure is unambiguous even in field theory, since

$$\lim_{t_i \to -\infty} e^{t_i(E - E_i)} = \begin{cases} 1 & for\ E = E_i \\ 0 & for\ E > E_i \end{cases}$$

$$\lim_{t_f \to +\infty} e^{-t_f(E - E_f)} = \begin{cases} 1 & for\ E = E_f \\ 0 & for\ E > E_f \end{cases}$$

Another important effect is on actions. For example, in the mechanics path integral for a particle with kinetic term $T = \frac{1}{2}m\dot{x}^2$ in a potential $U(x)$, we integrated

$$e^{-iS} \quad : \quad S = \int dt(U - T)$$

Upon Wick rotation, this becomes

$$e^{-S} \quad : \quad S = \int dt(U + T)$$

The major change on the exponent $-S$ is that it is now not only real, but negative definite. (For physical purposes, we assume the potential has a lower bound, which can be defined to be nonnegative without loss of generality.) Thus, the semiclassical approximation we made earlier, called the "stationary phase" approximation, has now become the "steepest descent" approximation, namely fitting $e^{-S/\hbar}$ to a Gaussian, which is approximating the integral by the places where the integrand is largest. We thus write

$$S(x) = S(x_0) + \tfrac{1}{2}(x - x_0)^2 S''(x_0) + ..., \qquad S'(x_0) = 0, \quad S''(x_0) > 0$$

for one variable, with the obvious generalization to many variables. Explictly, we then have

$$\int \frac{dx}{\sqrt{2\pi\hbar}} e^{-S(x)/\hbar} \approx \frac{1}{\sqrt{S''(x)}} e^{-S(x)/\hbar}\bigg|_{S'(x)=0}$$

plus higher orders in $\hbar$, expressed in terms of higher derivatives of S. In the case of many variables, S'' is replaced with a determinant, as for the Gaussian integrals of subsection IB3, and for functional integrals, with a functional determinant. (But sometimes the functional determinant can be replaced with an ordinary determinant: See exercise VA2.4.)

So now we can first calculate everything in Wick-rotated spacetime, where everything is real (more precisely, classical reality properties are preserved quantum mechanically), and then Wick rotate back to find the correct result in physical spacetime. In particular, the appropriate ϵ's, still needed to correctly position the singularities in physical spacetime, can be restored by rotating back through an angle $\frac{1}{2}\pi - \epsilon$:

$$inverse\ Wick: \qquad t \to (i + \epsilon)t = e^{i(\pi/2-\epsilon)}t, \quad E \to (-i + \epsilon)E = e^{-i(\pi/2-\epsilon)}E$$

REFERENCES

1 N. Wiener, *J. Math. and Phys. Sci.* **2** (1923) 132:
Euclidean path integrals for Brownian motion.

2 P.A.M. Dirac, *Phys. Z. der Sowj.* **3** (1933) 64:
proposed path integrals for quantum mechanics.
3 R.P. Feynman, *Rev. Mod. Phys.* **20** (1948) 367:
formulated path integral approach to quantum mechanics.
4 R.P. Feynman, *Phys. Rev.* **84** (1951) 108:
path integrals in phase space.
5 R.P. Feynman and A.R. Hibbs, *Quantum mechanics and path integrals* (McGraw-Hill, 1965):
review of path integrals; interesting side stuff, like D=2.
6 R. Shankar, *Principles of quantum mechanics*, 2nd ed. (Plenum, 1994):
good quantum mechanics text that includes path integrals.
7 J.H. van Vleck, *Proc. Natl. Acad. Sci. USA* **14** (1928) 178.
8 P.A.M. Dirac, *Proc. Roy. Soc.* **A136** (1932) 453, P.A.M. Dirac, V.A. Fock, and B. Podolosky, *Phys. Z. Sowj.* **2** (1932) 468:
interaction picture.
9 J.A. Wheeler, *Phys. Rev.* **52** (1937) 1107;
W. Heisenberg, *Z. Phys.* **120** (1943) 513, 673:
S-matrix.
10 E.C.G. Stückelberg, *Helv. Phys. Acta* **19** (1946) 242;
E.C.G. Stückelberg and D. Rivier, *Helv. Phys. Acta* **24** (1949) 215;
E.C.G. Stückelberg and T. Green, *Helv. Phys. Acta* **24** (1951) 153:
causality in quantum field theory.
11 N.N. Bogoliubov, *Doklady Akad. Nauk USSR* **82** (1952) 217, **99** (1954) 225:
explicit condition of causality on S-matrix, using functionals.
12 L.S. Brown, *Quantum field theory* (Cambridge University, 1992) p. 293:
decoupling of asymptotic multiparticle states in field-theory propagators.
13 M. Kac, On some connections between probability theory and differential and integral equations, in *Proc. 2nd Berkeley Symp. Math. Stat. Probability*, ed. J. Neyman, 1950 (University of California, 1951) p. 189;
E. Nelson, *J. Math. Phys.* **5** (1964) 332:
Wick rotation of path integral in quantum mechanics.
14 G.C. Wick, *Phys. Rev.* **96** (1954) 1124:
Wick rotation in quantum field theory.

B. PROPAGATORS

Classically we distinguish between particles and waves. This can be consistent with a classical limit of a quantum theory if there is a conserved charge associated with the classical particles, with respect to which the classical waves are neutral. Such a situation is described by a field theory Lagrangian (density) of the form

$$L = \psi^\dagger \mathcal{O}(\phi)\psi + L_\phi(\phi)$$

where ψ is te field of the charged particle, and ϕ the field of the waves that carry the interaction. $\mathcal{O}$ includes the kinetic operator; a nonrelativistic example was given in subsection IIIA3. Thus $\mathcal{O}(\phi)\psi = 0$, the field equation for ψ, is also a Schrödinger equation, which we can derive from a classical mechanics action. (A zero-range interaction, as in billiard-ball collisions, is described by an L_ϕ without derivatives.) Then we can have continuous worldlines for the particles: The statement that the worldlines do not end or split is associated with charge conservation. The interaction between the particles and waves is described by ϕ dependence in the particle (mechanics) action obtained from $\mathcal{O}$ (and not the term L_ϕ for the wave fields). If we look at just the mechanics action, the modification is the same as considering external fields (like external potentials in nonrelativistic mechanics), since we are ignoring L_ϕ, which is needed for the field equations of ϕ.

L_ϕ then can be added separately. Coupling to such external fields is a simple way to study properties of particles without applying field theory. For example, in nonrelativistic mechanics it helps to explain charge and spin, which don't appear explicitly in the free Schrödinger equation.

1. Particles

All the information in quantum mechanics is contained in the propagator, which gives the general solution to the Schrödinger equation, and can be obtained by the Feynman path integral. Here we discuss the free propagator for the spinless particle (whose classical description was given in section IIIB), which is the starting point for relativistic perturbation theory.

We consider quantization first in the Lorentz covariant gauge $v = 1$. From subsection IIIB2 we have

$$S_{H,AP} = \int_0^T d\tau[-\dot{x}^m p_m + \tfrac{1}{2}(p^2 + m^2)]$$

Except for the T integration in the functional integral (in addition to the functional integration over x and p), the same methods can be applied as in the nonrelativistic case, where we had

$$S_{H,NR} = \int_{-\infty}^{\infty} dt(-\dot{x}^i p_i + \tfrac{1}{2m}p^2)$$

The simplest expression (and ultimately the most useful one) is obtained by Fourier transforming with respect to x: In comparison to the multidimensional nonrelativistic result

$$\hat{G}_{NR}(p^i,t;p'^i,t') = \delta(p-p')\theta(t-t')e^{-i(t-t')p^2/2m}$$

(where here $\delta(p-p') = (2\pi)^{(D-1)/2}\delta^{D-1}(p^i - p'^i)$ for $D-1$ spatial dimensions), the relativistic result is

$$\tilde{G}(p,p') = \int dT\ \delta(p-p')\theta(T)e^{-iT(p^2+m^2)/2}$$

(where now $\delta(p-p') = (2\pi)^{D/2}\delta^D(p^a - p'^a)$ for D spacetime dimensions).

There are several simple yet important differences from the nonrelativistic case:

(1) The dependence on the mass m is different. In particular, we can set $m=0$ only in the relativistic case.

(2) There is an additional integration $\int dT$, because the variable T, which is the remaining part of v, survives the gauge $v=1$. (It is all that remains of a would-be functional integral over v.) This is analogous to the time integral in the nonrelativistic case for $\tilde{G}(p^i,E;p'^i,E')$, if we set the energy to zero. This is as expected, since the relativistic classical mechanics differs from the nonrelativistic one mainly by constraining the "Hamiltonian" $\frac{1}{2}(p^2+m^2)$ to vanish. This interpretation also leads to the "zero-energy" version of the inhomogeneous (proper-)time-independent Schrödinger equation for this case,

$$-i\tfrac{1}{2}(\Box - m^2)G(x,x') = \delta(x-x')$$

(3) The propagator is automatically "retarded" in the "proper time" T, as a consequence of the positivity condition $v>0$, which was motivated by the geometrical interpretation of v as the worldline metric.

When used in this manner to write the propagator in terms of a Gaussian, T is known as a "Schwinger parameter".

Generally, it is convenient to remove the momentum δ-function (which resulted from translational invariance) as

$$G(x,x') = \Delta(x-x') \quad \Rightarrow \quad \tilde{G}(p,p') = \delta(p-p')\Delta(p)$$

$$\Delta(p) = \int dT\ \theta(T)e^{-iT(p^2+m^2)/2}$$

where we have simply written $\Delta(p)$ for the Fourier transform of $\Delta(x)$ (dropping the tilde). Performing the T integral, using the same methods as for the t integral in the nonrelativistic case, we have the final result

$$\Delta(p) = \frac{-i}{\frac{1}{2}(p^2 + m^2 - i\epsilon)}$$

Actually, this result is almost obvious from solving the relativistic wave equation. The only part that is not obvious is the "$i\epsilon$ prescription": how to perform the contour integration upon Fourier transformation. In the nonrelativistic case, we saw two obvious choices, corresponding to retarded or advanced propagators; the classical action did not distinguish between the two, although the retarded propagator has the obvious convenience of determining later events from earlier ones. On the other hand, in the relativistic case the choice of propagator was fixed from classical considerations. T is restricted to be positive, and the $i\epsilon$ is needed to make the T integral converge.

Exercise VB1.1

Take the nonrelativistic limit of the relativistic propagator, and compare with the propagator of nonrelativistic quantum mechanics. Explain the difference in terms of the nonrelativistic limit of the classical mechanics action.

Exercise VB1.2

Perform the analysis of exercise VA2.2 for the relativistic particle. First replace the integration over T by a sum: Instead of dividing up the time into $2N$ intervals of length ϵ and taking the limit $N \to \infty, \epsilon \to 0$, with $2N\epsilon$ fixed, sum $2\epsilon \sum_{N=0}^{\infty}$, and then take the limit $\epsilon \to 0$. ($2N\epsilon$ is now T instead of $t_f - t_i$, and we integrate over it instead of keeping it fixed.) Perform all x integrals and then all but the last p integral before summing over N. Again, the entire calculation is much easier than using the Lagrangian (second-order) form of the path integral.

To understand this point better, we examine the Fourier transformation with respect to time. In contrast to the nonrelativistic case, there are now two poles, at

$$p^0 = \pm\omega, \quad \omega = \sqrt{(p^i)^2 + m^2}: \qquad \Delta = \frac{i}{\omega}\left(\frac{1}{p^0 - (\omega - i\epsilon)} - \frac{1}{p^0 + (\omega - i\epsilon)}\right)$$

where now $p^a = (p^0, p^i)$. These are also the two classical values of the canonical energy (as opposed to the true energy, which is the absolute value), which we saw previously corresponded to particles and antiparticles. With our prescription for integrating around the poles, using the same methods as in the nonrelativistic case, we then find

$$\hat{G}(p^i,t;p'^i,t') = (2\pi)^{D/2}\delta^{D-1}(p^i-p'^i)\frac{1}{\omega}e^{-i\omega|t-t'|}$$
$$= (2\pi)^{D/2}\delta^{D-1}(p^i-p'^i)\frac{1}{\omega}[\theta(t-t')e^{-i\omega(t-t')} + \theta(t'-t)e^{i\omega(t-t')}]$$

We now see that the particles ($p^0=\omega$) have a retarded propagator, while the antiparticles ($p^0=-\omega$) have an advanced propagator. This is the quantum version of the classical result we saw earlier, that particles travel forward in time, while antiparticles travel backward. The interpretation is simple: When evaluating matrix elements of the form $\langle f|\mathcal{O}|i\rangle$, the resulting propagator ensures that the initial wave function contains only positive energies, while the final *complex conjugate* wave function contains only negative energies (i.e., the final wave function itself contains positive energies).

We next compare quantization in the lightcone gauge. Again from subsection IIIB2,

$$S_{H,LC} = \int_{-\infty}^{\infty} d\tau[\dot{x}^- p^+ - \dot{x}^i p^i + \tfrac{1}{2}(p^{i2}+m^2)]$$

Whereas in the covariant gauge the analog to the nonrelativistic time t was the "proper time" T, the analog is now the lightcone "time" τ. Since $\tau = x^+/p^+$, we have $E = p^-p^+$ ($E = i\partial/\partial\tau$ vs. $p^- = i\partial/\partial x^+$), and thus

$$\Delta(p) = \frac{i}{E - \frac{1}{2}(p^{i2}+m^2)+i\epsilon} = \frac{-i}{\frac{1}{2}(p^2+m^2-i\epsilon)}$$

as before. Note that this derivation was almost identical to the nonrelativistic one: Unlike the covariant gauge, we did not have to add in T as a separate variable of integration (but not path integration). However, this Schwinger parameter is useful for evaluating momentum integrals and analyzing momentum dependence. This is a typical characteristic of unitary gauges: They are more useful for keeping track of degrees of freedom.

2. Properties

As in electrodynamics, the free scalar satisfies a differential equation second-order in time, so the propagator is used differently from nonrelativistic quantum mechanics to give a general solution to the wave equation. We begin by considering a free "action" between two different scalar fields, written in a way where all derivatives act on just one field or just the other, i.e., where the field equation is explicit. The two forms are related by integration by parts, but now we keep boundary terms:

$$\int d^Dx\ [A(\Box - m^2)B - B(\Box - m^2)A] = \int d^Dx\ \partial\cdot(A\overleftrightarrow{\partial}B) = \oint d^{D-1}\sigma^m\ A\overleftrightarrow{\partial}_m B$$

where in the last step we have used the (generalized) Stokes' theorem (see subsection IC2); "$\oint d^{D-1}\sigma^m$" is the integral over the closed surface bounding the volume integrated over in $\int d^D x$. In practice we take the volume to encompass all spacetime in the limit, neglect the part of the boundary at spacelike infinity, and choose the parts of the boundary at timelike infinity to be surfaces at constant time, so the boundary integrals are over just space:

$$\oint d^{D-1}\sigma^m \ A\overset{\leftrightarrow}{\partial}_m B = \int d^{D-1}x \ A\overset{\leftrightarrow}{\partial}_t B|_{-\infty}^{\infty} = \int d^{D-1}x \ A\overset{\leftrightarrow}{\partial}_t B|_{\infty} - \int d^{D-1}x \ A\overset{\leftrightarrow}{\partial}_t B|_{-\infty}$$

We then have the solution for the wave function inside the volume in terms of that on the boundary:

$$(\Box - m^2)\psi = 0, \quad -i\tfrac{1}{2}(\Box - m^2)G(x,x') = -i\tfrac{1}{2}(\Box' - m^2)G(x,x') = \delta(x-x')$$

$$\Rightarrow \quad \oint \frac{d^{D-1}\sigma'^m}{(2\pi)^{D/2}} \ G(x,x')\tfrac{1}{2}i\overset{\leftrightarrow}{\partial}{}'_m\psi(x') = \psi(x)$$

where the wave equation for ψ is the Klein-Gordon equation.

Similarly, this defines a conserved current from any two wave functions

$$\partial\cdot(\psi_1{}^*\overset{\leftrightarrow}{\partial}\psi_2) = \psi_1{}^*(\Box - m^2)\psi_2 - \psi_2(\Box - m^2)\psi_1{}^* = 0$$

or, evaluating the integral over a volume infinite in space but infinitesimal in time, the conserved charge

$$\frac{d}{dt}\int d^{D-1}x \ \psi_1{}^*\overset{\leftrightarrow}{\partial}_t\psi_2 = 0$$

This leads to the covariant inner product

$$\langle 1|2\rangle = \epsilon(p^0)\int \frac{d^{D-1}x}{(2\pi)^{D/2}} \ \psi_1{}^*\tfrac{1}{2}i\overset{\leftrightarrow}{\partial}_t\psi_2$$

where the $\epsilon(p^0)$ appears because the contour integral gives a + at later times (positive energy) and a $-$ at earlier times (negative energy). Explicitly, we find for the inner product of plane waves

$$\psi_p(x) = \langle x|p\rangle = e^{ip\cdot x}$$

$$\Rightarrow \quad \langle p|p'\rangle = (2\pi)^{D/2-1}\delta^{D-1}(p^i - p'^i)\epsilon(p^0)\tfrac{1}{2}(p^0 + p'^0)$$

We have used $p^2 + m^2 = p'^2 + m^2 = 0$, which also implies that $|p^0| = |p'^0|$: Thus, the inner product vanishes if the waves have opposite-sign energy, while for the same sign $\epsilon(p^0)\frac{1}{2}(p^0 + p'^0) = |p^0|$. The result then can be written manifestly covariantly as

$$\langle p|p'\rangle = \frac{\delta(p-p')}{2\pi\delta[\frac{1}{2}(p^2+m^2)]} = \theta(p^0 p'^0)\omega(2\pi)^{D/2-1}\delta^{D-1}(p^i - p'^i)$$

Similarly, the solution for the wave function in terms of the Green function gives only positive-energy contributions from the part of the surface at earlier times, and only negative-energy contributions from the part of the surface at later times. More general on-shell wave functions, since they depend on only $D-1$ spatial momenta and the sign of the energy, can be written as a restricted Fourier transform

$$\psi(x) = \int dp\ 2\pi\delta[\tfrac{1}{2}(p^2+m^2)]e^{ip\cdot x}\tilde{\psi}(p)$$

$$\Rightarrow \quad \langle 1|2\rangle = \int dp\ 2\pi\delta[\tfrac{1}{2}(p^2+m^2)]\tilde{\psi}_1(p)^*\tilde{\psi}_2(p)$$

(Here $\tilde{\psi}(p)^*$ means to complex conjugate after Fourier transforming to p-space; otherwise, we need to change the sign of the argument.) In particular, for a plane wave we have

$$\tilde{\psi}_{p'}(p) = \frac{\delta(p-p')}{2\pi\delta[\frac{1}{2}(p^2+m^2)]}$$

It will prove useful later to have a collection of solutions to the homogeneous and inhomogeneous Klein-Gordon equations, and compare them in 4-momentum space and time-3-momentum space. Using the previous nonrelativistic and relativistic results, we find

$$\begin{array}{llclcl}
\Delta\text{:} & -i/(p^2+m^2-i\epsilon) & \Rightarrow & \theta(t)e^{-i\omega t} + \theta(-t)e^{i\omega t} & = & e^{-i\omega|t|} \\
\Delta^*\text{:} & i/(p^2+m^2+i\epsilon) & \Rightarrow & \theta(t)e^{i\omega t} + \theta(-t)e^{-i\omega t} & = & e^{i\omega|t|} \\
\Delta_R\text{:} & -i/(p^2+m^2-i\epsilon p^0) & \Rightarrow & \theta(t)e^{-i\omega t} - \theta(t)e^{i\omega t} & = & -2i\theta(t)sin(\omega t) \\
\Delta_A\text{:} & i/(p^2+m^2+i\epsilon p^0) & \Rightarrow & \theta(-t)e^{i\omega t} - \theta(-t)e^{-i\omega t} & = & 2i\theta(-t)sin(\omega t) \\
\Delta_+\text{:} & \theta(p^0)2\pi\delta(p^2+m^2) & \Rightarrow & \theta(t)e^{-i\omega t} + \theta(-t)e^{-i\omega t} & = & e^{-i\omega t} \\
\Delta_-\text{:} & \theta(-p^0)2\pi\delta(p^2+m^2) & \Rightarrow & \theta(t)e^{i\omega t} + \theta(-t)e^{i\omega t} & = & e^{i\omega t}
\end{array}$$

where we have omitted certain common factors (see above). $\Delta_\pm$ satisfy the homogeneous equation, while the rest satisfy the inhomogeneous one. (These are easily checked in the mixed space, where the Klein-Gordon operator is $-(\partial_t^2+\omega^2)$.) This table makes explicit which sign of the energy propagates in which time direction, as well as the linear relations between the momentum-space expressions. In particular, we see that Δ_+ propagates just the positive-energy states, while Δ_- propagates just the negative-energy ones.

Exercise VB2.1

The relativistic propagator uses a particular choice for integrating around the two poles in the complex energy plane, as encoded in the $i\epsilon$ prescription. If we ignored the classical determination of that prescription, there would be four simple choices, integrating either above or below the two poles.

a Show these four choices can be enforced by replacing $i\epsilon$ in $p^2 + m^2 - i\epsilon$ with

$$i\epsilon, \quad -i\epsilon, \quad i\epsilon p^0, \quad -i\epsilon p^0$$

and derive the results of the table above.

b Give explicit expressions for the four propagators in position space in four dimensions for the massless case.

We can check the propagator's behavior by explicit evaluation, using plane waves:

$$\epsilon(p^0) \int \frac{d^{D-1}x'}{(2\pi)^{D/2}}\ \Delta(x - x')\tfrac{1}{2} i \overleftrightarrow{\partial}_t{}' \psi_p(x') = \epsilon(p^0)\tfrac{1}{2}(p^0 + i\partial_t)\frac{1}{\omega} e^{-i\omega|t-t'|} e^{i\vec{p}\cdot\vec{x} - ip^0 t'}$$
$$= \theta[p^0(t - t')]\psi_p(x)$$

where we have used the previous result for $\hat{G}(\vec{p}, t; \vec{p}', t')$ (and thus $\Delta(\vec{p}, t)$). Again we see that the propagator propagates positive-energy solutions forward in time and negative-energy backward.

This propagator also applies to relativistic field theory. (See subsection IIIA3 for nonrelativistic field theory.) In comparison to the nonrelativistic case, the propagator is now $-i/\frac{1}{2}(p^2 + m^2)$ instead of $-i/(\frac{1}{2m}\vec{p}^{\,2} - E)$, and this determines the kinetic term in the field theory action:

$$S_0 = -\int dx\ \tfrac{1}{2}\phi\tfrac{1}{2}(\Box - m^2)\phi = \int dx\ \tfrac{1}{4}[(\partial\phi)^2 + m^2\phi^2]$$

To make the functional integral of e^{-iS_0} converge, we replace $m^2 \to m^2 - i\epsilon$, which is the same $i\epsilon$ prescription found in first-quantization. Note that we have used a real field $\phi^* = \phi$. (A complex field can be used by doubling $\psi = (\phi_1 + i\phi_2)/\sqrt{2}$.) This is possible only in the relativistic case because we have both positive-energy solutions e^{-iEt} as well as negative ones e^{+iEt}. (In other words, the relativistic Schrödinger equation is a second-order differential equation, so we get two i's to make the kinetic operator real.) Reality simply means identifying particles with antiparticles. (E.g., there is no "antiphoton" distinct from the photon.)

3. Generalizations

More generally, we will find propagators of the form (in momentum space)

$$\Delta = -\frac{i}{K}, \quad K = K^\dagger$$

corresponding to free actions

$$S_0 = \int dx\ \tfrac{1}{2}\phi K \phi$$

where $K = -\frac{1}{2}(\Box - m^2)$ in the case just considered. Then the inner product is defined as above in terms of the Green function by again using integration by parts,

$$i \int d^D x \; [(KA^\dagger)B - A^\dagger KB] = \epsilon(p^0) \oint d^{D-1}\sigma^m \; A^\dagger M_m B$$

to define the operator M_m, which was $\epsilon(p^0)\frac{1}{2}i\overleftrightarrow{\partial}_m$ in the previous case. (For the usual equal-time hypersurfaces, we use $M_0 = -M^0$. There may be additional implicit matrix factors in the Lorentz-invariant inner product $A^\dagger B$.) This in turn defines the inner product

$$\langle 1|2\rangle = \int \frac{d^{D-1}\sigma^m}{(2\pi)^{D/2}} \; {\psi_1}^\dagger M_m \psi_2$$

and thus

$$\psi(x) = \epsilon(p^0) \oint \frac{d^{D-1}\sigma'^m}{(2\pi)^{D/2}} \; G(x,x')M'_m \psi(x')$$

This inner product gives a nonnegative norm on physical bosonic states, but on physical fermionic states it is negative for negative energy, because ordering the initial state to the left of the final state (the wrong ordering for quantum mechanics) produces a minus sign from the anticommutativity of the fermions. (From the explicit integral, this appears because K is generally second-order in derivatives for bosons, but first-order for fermions, so M_m has one factor of p^0 for bosons and none for fermions.)

For physical fields, the (free) field equation will always imply the Klein-Gordon equation (after gauge fixing for gauge fields). Thus, the propagator can always be written as

$$\Delta = -\frac{i}{K} = -i\frac{N(p)}{\frac{1}{2}(p^2 + m^2 - i\epsilon)}$$

in terms of some matrix kinematic factor $N(p)$. Using this expression for the propagator in our above position-space inner product, this implies (e.g., using $\hat{G}$ as in the previous subsection for the denominator by using a Fourier transform)

$$\psi_p(x) = \hat{\psi}(p)e^{ip\cdot x} \quad \Rightarrow \quad N(p)M_0(p)\hat{\psi}(p) = \omega\hat{\psi}(p)$$

from integrating over a hypersurface at constant time. If we generalize to other flat hypersurfaces with timelike normals n^m (e.g., by just Lorentz transforming), we have

$$N(p)n \cdot M(p)\hat{\psi}(p) = -\epsilon(p^0)n \cdot p\hat{\psi}(p)$$

and finally, by taking linear combinations for different n's,

$$N(p)M_m(p)\hat{\psi}(p) = -\epsilon(p^0)p_m\hat{\psi}(p)$$

If we choose a basis that is orthonormalized with respect to all quantum numbers other than momenta (i.e., with repsect to spin/helicity and internal symmetry), we have

$$\langle p, i|p', j\rangle = \delta_{ij}\frac{\delta(p-p')}{2\pi\delta[\frac{1}{2}(p^2+m^2)]}$$

If we ignore coordinate/momentum dependence and focus on just these other quantum numbers, then it is clear that

$$N = [\epsilon(p^0)]^{2s}\sum_i |i\rangle\langle i|$$

where we have included an extra sign factor for negative energy and half-integer spin from the reordering of states, as explained above. In an arbitrary basis, we can generalize to

$$N(p) = [\epsilon(p^0)]^{2s}\sum_i \hat{\psi}^\dagger{}_i(p)\hat{\psi}_i(p)$$

The positive-energy propagator is then given by a sum over all positive-energy states:

$$\mathit{\Delta}_+ = N(p)\theta(p^0)2\pi\delta[\tfrac{1}{2}(p^2+m^2)]$$

The fact that K is not simply the Klein-Gordon operator is a consequence of unphysical (gauge/auxiliary) degrees of freedom appearing in the action: Then N is a projection operator that projects out the auxiliary degrees of freedom on shell, and the gauge degrees of freedom on and off (in unitary gauges), as represented above by a sum over physical states. However, more general N's are sometimes used that include unphysical degrees of freedom; these must be canceled by "ghosts", similar unphysical degrees of freedom of the opposite statistics.

Exercise VB3.1

Demonstrate all these properties for spin (helicity) $\frac{1}{2}$ (see subsections IIA6, IIB6-7, and IIIA4): Find M_m from integration by parts. Find N both from inverting to get the propagator and from summing over physical states. Show the $NM\hat{\psi}$ identity is satisfied.

a For the massless case, use twistors for the solution to the field equation to find

$$(M_{\alpha\dot{\beta}})^{\gamma\dot{\delta}} = \delta_\alpha^\gamma\delta_{\dot{\beta}}^{\dot{\delta}}\epsilon(p^0), \qquad (N)_{\alpha\dot{\beta}}(p) = \epsilon(p^0)p_\alpha\bar{p}_{\dot{\beta}} = p_{\alpha\dot{\beta}}$$

b Do the same for the massive Dirac spinor, to find

$$M_m = \gamma_m\epsilon(p^0), \qquad N = \not{p} + \tfrac{m}{\sqrt{2}}$$

(Hint: Consider the rest frames for $p^0 > 0$ and < 0.)

Exercise VB3.2

Use the construction of exercise VB1.2 to define the path integral for the spinning particle of exercise IIIB1.2. Show that in the covariant gauge $v = 1$, $\lambda =$ constant, the propagator can be written as

$$\varDelta(p) \sim \int d\xi\ dT\ \theta(T)e^{-T\frac{1}{2}p^2 - i\xi\gamma\cdot p} \sim \frac{-i\gamma\cdot p}{\frac{1}{2}p^2} = \frac{i}{\gamma\cdot p}$$

up to some arbitrary normalization factor, where $\xi = \int d\tau\ \lambda$ is the only gauge invariant part of λ (as $T = \int d\tau\ v$ is for v).

We will find that quantum corrections modify the form of the propagator. In particular, it may modify the position m^2 of the pole in $-p^2$ and its residue, as well as adding terms that are analytic near that pole. For example, consider a scalar propagator of the form

$$\varDelta(p) = -i\frac{N}{\frac{1}{2}(p^2 + m^2 - i\epsilon)} + R$$

where N is a constant and R is analytic in p. By the procedure of "renormalization", N can be set to 1, and m^2 can be set to its original value (see chapter VII). Alternatively, we can cancel N in the normalization of external states, and redefine the masses of these states to coincide with what appears in the propagator.

Exercise VB3.3

Use this propagator to define the inner product between two plane waves, and evaluate it explicitly. Show that R gives no contribution, and the plane waves need factors of $\sqrt{N}$ to maintain their normalization. (Hint: What is the wave equation corresponding to $\varDelta$, and how is it related to M_m? You can also consider the relation of N to $\varDelta_+$.)

Away from the pole, at higher values of $-p^2$ than m^2, there will also be cuts corresponding to mutiparticle states. Although these higher-energy intermediate states in the propagator will contribute to the time development even for on-shell states (those satisfying $p^2 + m^2 = 0$ asymptotically), in S-matrix elements we can ignore such contributions on external lines, using our modified definition for evaluating the asymptotic limit for the S-matrix (see subsection VA4).

Exercise VB3.4

Consider the general scalar propagator

$$\varDelta(p) = -i\int_0^\infty d\mu\ \frac{\rho(\mu)}{\frac{1}{2}(p^2 + \mu^2 - i\epsilon)}, \qquad \rho(\mu) = \delta(\mu - m) + \theta(\mu - 2m)\sigma(\mu)$$

which contains a pole at mass m and contributions from multiparticle states at mass $2m$ and higher. Fourier transform from energy to time. Use this propagator to define the time development of a momentum eigenstate satisfying the free wave equation asymptotically, using the $1 + i\epsilon$ prescription of subsection VA4 to define the asymptotic limit:

$$\psi(t) \equiv \lim_{t_i \to -\infty(1+i\epsilon)} \int G(t,t_i) \tfrac{1}{2} i \overleftrightarrow{\partial}_{t_i} \psi_0(t_i)$$

and show that $\sigma(\mu)$ does not contribute:

$$(\Box - m^2)\psi_0 = 0 \quad \Rightarrow \quad \psi(t) = \psi_0(t)$$

4. Wick rotation

As in the nonrelativistic case, the $i\epsilon$ prescription can also be fixed by the infinitesimal Wick rotation (see subsection VA5)

$$t \to (1 - i\epsilon)t, \quad E \to (1 + i\epsilon)E \quad \Rightarrow \quad \frac{1}{p^2 + m^2} \to \frac{1}{p^2 + m^2 - i\epsilon}$$

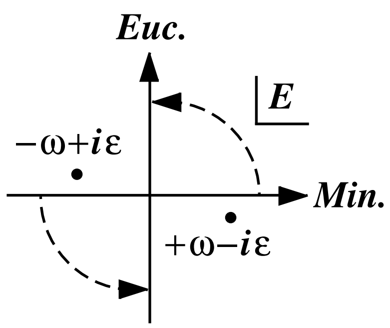
However, in the relativistic case, a finite Wick rotation gets rid of not only i's but also the annoying minus signs associated with the Minkowski metric. We now replace all timelike coordinates, including proper time, with spacelike coordinates:

$$t \to -it, \qquad \tau \to -i\tau$$

In addition, for every vector field V^a we replace

$$V^0 \to -iV^0 \qquad (V^i \to V^i)$$

and similarly for tensor fields. (Here we have defined Wick rotation in the first-quantized sense: on all explicit coordinates and momenta, as well as on explicit Lorentz indices. For example, in the field theory action we rotate the explicit derivatives and the integration measure, rather than the arguments of the fields.)



Note that E as defined in the nonrelativistic case was $-p_0$, which is the same as p^0 only in Minkowski space:

$$p^0 \to -ip^0, \qquad p_0 \to +ip_0$$

Furthermore, there is some apparent ambiguity in how to change the integration measure, corresponding to how the integration contours are rotated (i.e., changes in the limits of integration). In particular, we see from subsection VA5 that the contour rotation for E is actually in the opposite direction of that for t, consistent with Fourier transformation. (Effectively, we keep the extra i for $\int dp$ from rotating p_0, while dropping the -1 from $p_0 \leftrightarrow p^0$ because of the usual absolute value in the Jacobian in real changes of variables.) The net result is the naive change for $\int dx$, while that for $\int dp$ preserves the inverse Fourier transform:

$$\int dx \to -i\int dx, \qquad \int dp \to i\int dp; \qquad \delta(x) \to i\delta(x), \qquad \delta(p) \to -i\delta(p)$$

When manipulating explicit expressions, the factors of i on coordinates/momenta and fields can be transferred to the constant tensors contracting their indices: The net effect is that the Wick rotation is equivalent to changing just τ, the integration measures, the flat-space metric, and the Levi-Civita tensor:

$$\tau \to -i\tau, \qquad \frac{\partial}{\partial\tau} \to i\frac{\partial}{\partial\tau}$$

$$\int d\tau \to -i\int d\tau, \qquad \int dx \to -i\int dx, \qquad \int dp \to i\int dp$$

$$\eta_{mn} \to \delta_{mn}, \qquad \epsilon_{abcd} \to -i\epsilon_{abcd}$$

So now the inner product is positive definite: We have gone from Minkowski space to Euclidean space.

For example, for the relativistic particle in the gauge $v = 1$, the propagator is now

$$\varDelta(p) = \int dT\ \theta(T)e^{-T(p^2+m^2)/2} = \frac{1}{\frac{1}{2}(p^2+m^2)}$$

The integral is automatically convergent because p^2+m^2 is now positive definite. If we examine our transformation of timelike components in terms of the complex p^0 plane, we see that we have just rotated the contour from the real axis to the imaginary axis, through an angle $\pi/2$, which avoids the poles at $p^0 = \pm(\omega - i\epsilon)$. (There is actually a slight cheat in the massless case, since the two poles converge near vanishing 3-momentum, where $\omega = 0$, but this problem can be avoided by an appropriate limiting procedure.) Note that in the relativistic case the Euclidean propagator is completely real also in momentum space; even the overall $-i$ has been killed. (Compare the nonrelativistic case in subsection VA5, where $-i/(H-E) \to 1/(H-iE)$.) This follows from first Wick rotating in position space, then performing the Fourier transform as usual (avoiding an extra $-i$ from rotating the $\int dt$ in the Fourier transform).

Exercise VB4.1

Find the propagator in time-3-momentum space, $\hat{G}(p^i, t; p'^i, t')$, after Wick rotation. (I.e., Wick rotate $\Delta(p)$ first, then Fourier transform.)

Exercise VB4.2

Find the *massless* propagator in 4D Minkowski coordinate space, including the $i\epsilon$, by

a Fourier transforming the Schwinger-parametrized momentum-space propagator in Minkowski space (including the "Minkowski" τ),

b doing the same entirely in Euclidean space, and then Wick rotating the time back to Minkowski space, and

c Fourier transforming both of the above cases without using the Schwinger parameter, first doing the energy integrals as in the previous section. (Hint: Use rotational invariance to point $\vec{x}$ in a particular direction to simplify the angular integration.)

Exercise VB4.3

Although the propagator in momentum space is most useful for scattering of plane waves, its position-space dependence is more useful for bound states/scattering of localized sources:

a Evaluate the Wick-rotated propagator in arbitrary dimensions D for large $x = \sqrt{x^2}$ by

(1) Fourier transforming the Schwinger-prametrized form,

(2) performing the (Gaussian) p integration before the T, and

(3) using the steepest descent approximation on the exponential to approximate the T integral (but don't bother sticking the power of T multiplying the exponential in it as a log).

Why does this approximation correspond to large x? (It may be useful to make the redefinition $T \to (x/m)\tau$.) You should find the result

$$\Delta(x) \approx \sqrt{2\pi} m^{(D-3)/2} x^{-(D-1)/2} e^{-mx}$$

After Wick rotating back, the exponential becomes a phase inside the lightcone, while outside it gives exponential damping, as in quantum mechanical barrier penetration. (In the massless case, there is damping away from the lightcone both inside and outside, but only by powers, determined by dimensional analysis.)

b Show the above result is exact in D=1 and 3 by

(1) Fourier transforming *without* the Schwinger parameter,

(2) performing a contour integral over the magnitude of p by closing it appropriately and picking up the contributions at the poles at $p = \pm im$, and

(3) for D=3, fixing x^m in a particular direction and doing the angular part of the p integration.

(This method can also be used to obtain the exact corrections to the above in terms of elementary functions for higher odd D.) For the physical case of the potential produced by a static point source in 3 spatial dimensions, we find $\Delta(x) = \sqrt{2\pi}e^{-mx}/x$, so at short range the Coulomb potential is unmodified, while it is exponentially damped at range $1/m$.

c Check the results for **b** by taking the massless limit, and comparing to the analogous result using the method of **a**, but doing the T integral exactly for those cases. (Warning: The result is infinite in D=1, and some type of "regularization" must be used to subtract an infinite constant, leaving a finite x-dependent remainder.)

The corresponding effect on the action, where we path-integrated

$$e^{-iS} \quad : \quad S = \int d\tau \ \tfrac{1}{2}(vm^2 - v^{-1}\dot{x}^m\dot{x}^n\eta_{mn})$$

is to integrate the Wick rotated expression

$$e^{-S} \quad : \quad S = \int d\tau \ \tfrac{1}{2}(vm^2 + v^{-1}\dot{x}^m\dot{x}^n\delta_{mn})$$

This method also applies to relativistic field theory. Wick rotation $t \to -it$ of the kinetic term

$$e^{-iS_0} \quad : \quad S_0 = \int dx \ \tfrac{1}{4}[\eta^{mn}(\partial_m\phi)(\partial_n\phi) + m^2\phi^2]$$

now gives

$$e^{-S_0} \quad : \quad S_0 = \int dx \ \tfrac{1}{4}[\delta^{mn}(\partial_m\phi)(\partial_n\phi) + m^2\phi^2]$$

This has an interesting consequence in the complete action (including interactions). Positivity of the energy implied the non-time-derivative terms in the action had to be positive, but the time-derivative terms are now the same sign in the Wick-rotated action, effectively the same as adding an extra spatial dimension: The energy $T+U$ is the same as the Wick-rotated Lagrangian ($-T+U \to T+U$). So we can replace the condition of positivity of the energy with positivity of the Wick-rotated action, which we need anyway for path-integral quantization. (Note that for the particle this again requires $v \geq 0$.)

Although Wick rotation is thus very useful for application to "intermediate" states, it can never be applied to physical (i.e., initial and final) states: For example, $p^2+m^2=0$ has no solution in Euclidean space, since each term is strictly positive. Typically, this means that one performs quantum calculations first in Euclidean space, then Wick rotates back to Minkowski space before applying physical state conditions.

Unfortunately, the simple results for Wick rotation obtained here for spin 0, although they generalize to spin 1, do not work so simply for other spins. (Consider, e.g., trying it in spinor notation.) However, the method can still be applied to the coordinates, and simplifies momentum integrals, since one can avoid contours, i's, and Minkowski minus signs.

REFERENCES

1 V.A. Fock, *Phys. Z. Sowj.* **12** (1937) 404;
E.C.G. Stückelberg, *Helv. Phys. Acta* **14** (1941) 322, *loc. cit.* (IA);
Y. Nambu, *Prog. Theo. Phys.* **5** (1950) 82,
R.P. Feynman, *Phys. Rev.* **80** (1950) 440:
quantization of relativistic mechanics in terms of proper time.

2 J. Schwinger, *Phys. Rev.* **75** (1949) 651, **82** (1951) 664:
his parameters.

3 J. Polchinski, *Commun. Math. Phys.* **104** (1986) 37:
gives a more rigorous discussion of path-integral quantization of the relativistic particle.

4 Stückelberg, *loc. cit.* (IA);
R.P. Feynman, *Phys. Rev.* **76** (1949) 749, 769:
relativistic propagator.

C. S-MATRIX

A *relativisitic quantum field theory* is defined by three properties:

(1) *Poincaré invariance* ("relativistic") is the basic result of special relativity. Its consequences have been accurately observed both macroscopically and (sub)microscopically, and no violations are known.

(2) *Unitarity* ("quantum") is the main mathematical result of quantum mechanics: Any quantum theory can be considered as the corrections to the classical theory implied by unitarity. This is one way to define perturbation theory, and is equivalent to the usual (JWKB) expansion in $\hbar$. (The other major axioms of quantum mechanics are concerned with the physical interpretation of the quantities calculated, such as the preparation and measurement of states.) Quantum mechanics also has been accurately verified, with no observed violations.

(3) *Causality* ("field theory") appears in many areas of physics, formulated in many ways. The strongest way to state causality, in a way independent of special relativity and quantum mechanics, is as locality: All interactions happen at a point; there is no action at a distance. This means that any force applied by an object at one point in spacetime on another elsewhere(/when) must be mediated by yet another object that carries the effect of that force between the two. The most accurate verifications of this principle have been through the predictions of relativistic quantum field theory.

We now define the perturbation expansion of the S-matrix, and give its general properties.

1. Path integrals

Path integrals for relativistic quantum field theory in four dimensions are better in every way than canonical quantization. They are

(1) easier to learn and apply: just Gaussian integrals.

(2) more heuristic: no "Dirac sea" or harmonic oscillators (where are the springs?).

(3) less mathematical: no operators, so no time-ordered products (much less "T*" ones), Wick contraction, normal ordering, etc.

(4) more efficient: Functionals make combinatoric factors automatic.

(5) more rigorous: Constructive quantum field theory has proven the existence of certain relativistic quantum field theories in less than four dimensions (working in Euclidean space).

(6) more covariant: The action and Feynman rules are, so the middle steps should be.

As we have seen in simpler examples (subsection VA2), perturbation for path integrals is based on Taylor expansion of the exponential of the "interaction" (part of the exponent higher than quadratic in the functional integration variables). Since general transition amplitudes involve also wave functions, we also Taylor expand them: In field theory, this is an expansion in the number of particles. We therefore write

$$\Psi[\phi] = \sum_{N=0}^{\infty} \tfrac{1}{N!} \int \frac{d\sigma_1^m ... d\sigma_N^n}{(2\pi)^{ND/2}}\ \phi(x_1)...\phi(x_N) M_{1m}...M_{Nn} \psi_N(x_1,...,x_N)$$

where we have used the covariant inner product of subsection VB3. (The surfaces of integration are at $t = \pm\infty$.) We have also used the *free* N-particle wave function ψ_N,

$$K_1\psi_N = ... = K_N\psi_N = 0$$

which is sufficient to describe particles at $t = \pm\infty$. Amplitudes for such asymptotic states are elements of the S-matrix, as defined in the interaction picture (see subsection VA4). In practice we choose a particular value of N, and use a basis element for ψ_N, namely the product of N 1-particle wave functions:

$$\psi_N(x_1,...,x_N) = \prod_{i=1}^{N} \psi_{Ni}(x_i)\ +\ permutations \quad \Rightarrow \quad \Psi[\phi] = \prod_{i=1}^{N} \langle\phi|\psi_{Ni}\rangle$$

In principle, if there are bound states in the theory, we can consider similar wave functions, but besides ϕ we expand in the composite field describing the bound state. It should be possible to discover such states by looking at the properties of the amplitudes of the ϕ states. (For example, a two-particle bound state would show up in the amplitude describing the scattering of two particles.)

Note that the fields ϕ are real (or we sum over both ϕ and ϕ^*), while the wave functions (Ψ or ψ) are complex: As usual in quantum mechanics, we work in a complex Hilbert space, but often expand over a real basis. For example, $\psi(q) = \psi(0) + \psi'(0)q + \frac{1}{2}\psi''(0)q^2 + ...$ or $\binom{\psi_1}{\psi_2} = \psi_1 \binom{1}{0} + \psi_2 \binom{0}{1}$. Since in a covariant approach we treat particles and antiparticles on an equal footing, $\Psi[\phi]$ should include both the initial and final wave function: Thus, the factors in Ψ at $t = -\infty$ can be interpreted as the usual positive-energy multiparticle wave functions, while the factors at $t = +\infty$ can be interpreted as the complex conjugate of the usual positive-energy multiparticle wave functions. Since they each have one sign of the energy, they are necessarily complex. However, for a real field ϕ we can have only $\langle\phi|\psi\rangle$, while for a complex field ϕ we distinguish $\langle\phi|\psi\rangle$ as representing a particle at $t=-\infty$ or an antiparticle at $t=+\infty$, and $\langle\phi^*|\psi\rangle$ as an antiparticle at $t=-\infty$ or a particle at $t=+\infty$. (As usual, which we choose to call particle and which antiparticle is relative, because of CPT invariance.)

We therefore want to evaluate the path integral

$$\mathcal{A} = \int D\phi \; e^{-iS[\phi]}\Psi[\phi]$$

Separating out the free and interacting pieces of the (gauge-fixed) action,

$$S = S_0 + S_I = \int dx \; \tfrac{1}{2}\phi K\phi + S_I$$

and using the integration identity

$$\boxed{\int \frac{du}{\sqrt{2\pi}} \; e^{-uMu/2} f(u+v) = \int \frac{du}{\sqrt{2\pi}} \; e^{-uMu/2} e^{u\partial_v} f(v) \sim e^{\partial_v M^{-1}\partial_v/2} f(v)}$$

at $v=0$, we can evaluate the path integral as

$$\mathcal{A} = exp\left(-i\int dx \; \tfrac{1}{2}\frac{\delta}{\delta\phi}\frac{1}{K}\frac{\delta}{\delta\phi}\right) e^{-iS_I[\phi]}\Psi[\phi]\bigg|_{\phi=0}$$

$$-i\int dx \; \tfrac{1}{2}\frac{\delta}{\delta\phi}\frac{1}{K}\frac{\delta}{\delta\phi} = \int dx \; dx' \; \tfrac{1}{2}\frac{\delta}{\delta\phi(x)}\Delta(x-x')\frac{\delta}{\delta\phi(x')}$$

We have dropped the determinant factor, since in our case it will be only an overall constant coming from the kinetic operator K. We then absorbed this proportionality constant into the definition of $D\phi$, as we did for the free particle in subsection VA2. In this case, this normalization is fixed by the "free" part of the S-matrix.

It will prove convenient to distinguish the ends of propagators that attach to S_I from those that attach to Ψ, so using the identity

$$f(\partial_x)g(x,x+y) = f(\partial'_x+\partial'_y)g(x',y') \qquad (x'=x, \quad y'=y+x)$$

evaluated at $y=0$, we rewrite this expression as

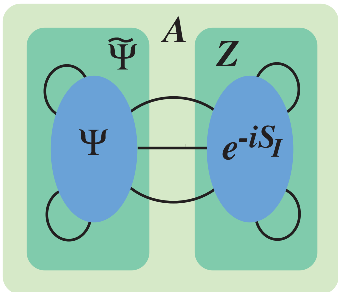
$$\mathcal{A} = exp\left(-i\int \tfrac{1}{2}\frac{\delta}{\delta\phi}\frac{1}{K}\frac{\delta}{\delta\phi} + \frac{\delta}{\delta\phi}\frac{1}{K}\frac{\delta}{\delta\varphi} + \tfrac{1}{2}\frac{\delta}{\delta\varphi}\frac{1}{K}\frac{\delta}{\delta\varphi}\right) e^{-iS_I[\phi]}\Psi[\varphi]\bigg|_{\phi=\varphi=0}$$

We then evaluate the derivatives that act on only one or only the other:

$$\mathcal{A} = exp\left(-i\int \frac{\delta}{\delta\phi}\frac{1}{K}\frac{\delta}{\delta\varphi}\right) Z[\phi]\tilde{\Psi}[\varphi]\bigg|_{\phi=\varphi=0}$$

$$\tilde{\Psi}[\phi] = \left\{exp\left(-i\int \tfrac{1}{2}\frac{\delta}{\delta\phi}\frac{1}{K}\frac{\delta}{\delta\phi}\right)\Psi[\phi]\right\}$$

$$Z[\phi] = \left\{exp\left(-i\int \tfrac{1}{2}\frac{\delta}{\delta\phi}\frac{1}{K}\frac{\delta}{\delta\phi}\right)e^{-iS_I[\phi]}\right\}$$



Then we move the differential operators into the wave functional:

$$\mathcal{A} = \hat{\Psi}\left[\frac{\delta}{\delta\phi}\right] Z[\phi]\bigg|_{\phi=0}$$

$$\hat{\Psi}\left[\frac{\delta}{\delta\phi}\right] = exp\left(-i\int \frac{\delta}{\delta\phi}\frac{1}{K}\frac{\delta}{\delta\varphi}\right)\tilde{\Psi}[\varphi]\bigg|_{\varphi=0}$$

$Z[\phi]$ (the "generating functional" for the S-matrix) contains all propagators with $S_I[\phi]$'s attached at both ends, and forms the basis of the perturbation expansion. From the integration identity above, we can also write it as

$$\boxed{Z[\varphi] = \int D\phi\ e^{-i(S_0[\phi]+S_I[\phi+\varphi])}}$$

Effectively, we have just taken the functional integral $\int D\phi\ e^{-iS[\phi]}$ and separated the field into a "quantum field" ϕ (the integration variable) and a "background field" φ, where φ includes the asymptotic states, which propagate to infinity, while ϕ vanishes at infinity (or at least goes to a constant) fast enough to allow the usual integration by parts in performing the functional integral. Thus φ gives the boundary value of the field. This is essentially the same as the general prescription for path integrals given in subsection VA2, except that we take φ to be arbitrary for convenience of functional differentiation, and we drop free φ terms, which were incorporated into $\tilde{\Psi}$ (i.e., we expand just S_I).

$\tilde{\Psi}[\phi]$ is the result of contracting some pairs of the one-particle wave functions with propagators. This gives the usual inner product of those one-particle states: Since for any one-particle wave function satisfying the free field equations we have the propagator identity

$$\boxed{\int \frac{d\sigma^m}{(2\pi)^{D/2}}\psi(x)M_m\Delta(x-x') = \psi(x')}$$

such contractions give

$$\int \frac{d\sigma^m d\sigma'^n}{(2\pi)^D}\psi(x)M_m\Delta(x-x')M'_n\psi'(x')$$

$$= \int \frac{d\sigma^m}{(2\pi)^{D/2}}\psi(x)M_m\psi'(x) = \langle\psi^*|\psi'\rangle = \langle\psi'^*|\psi\rangle$$

where the inner product vanishes unless ψ and ψ' have opposite energies (i.e., one is incoming and one is outgoing). This is the boring part of the S-matrix element: It represents the corresponding particles not interacting at all. (For example, in the free case $S_I = 0$ we have $Z = 1$, and $\mathcal{A} = \tilde{\Psi}|_{\phi=0}$ consists of only such inner products.) For most purposes we just factor out such free inner products, and consider only processes where all particles interact.

Finally, the conversion from $\tilde{\Psi}$ to $\hat{\Psi}$ replaces all the on-shell inner products with integrals over all spacetime: Using the propagator identity from above,

$$\hat{\Psi}\left[\frac{\delta}{\delta\phi}\right] = \sum_{N=0}^{\infty} \tfrac{1}{N!} \int \frac{d^{ND}x}{(2\pi)^{ND/2}} \tilde{\psi}_N(x_1, ..., x_N) \frac{\delta}{\delta\phi(x_1)} \cdots \frac{\delta}{\delta\phi(x_N)}$$

where $\tilde{\psi}_N$ appears in $\tilde{\Psi}$ exactly as ψ_N in Ψ:

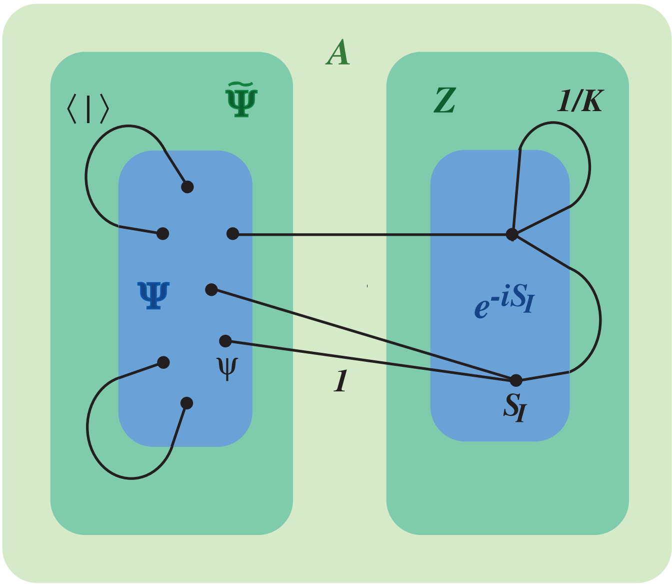
$$\tilde{\Psi}[\phi] = \sum_{N=0}^{\infty} \tfrac{1}{N!} \int \frac{d\sigma_1^m ... d\sigma_N^n}{(2\pi)^{ND/2}}\ \tilde{\psi}_N(x_1, ..., x_N) M_{1m} ... M_{Nn} \phi(x_1) ... \phi(x_N)$$

Usually we represent ψ_N as the product of single-particle wave functions,

$$\hat{\Psi}_N\left[\frac{\delta}{\delta\phi}\right] = \prod_{i=1}^{N}\left(\int \psi_{Ni}\frac{\delta}{\delta\phi}\right) \quad \Rightarrow \quad \mathcal{A}_N = \prod_{i=1}^{N}\left(\int \psi_{Ni}\frac{\delta}{\delta\phi}\right) Z[\phi]\bigg|_{\phi=0}$$

in which case $\hat{\Psi}$ simply replaces each field ϕ in Z with one of these wave functions.

$$\Delta \to \begin{cases} \Delta : S_I(\ )S_I \\ 1 : \psi(\ )S_I \\ \langle\ |\ \rangle : \psi(\ )\psi \end{cases}$$



Thus, of the three types of propagators, only the ones that connected two factors of S_I remain, all inside Z; the ones that connect the wave functions to Z have been replaced with just spacetime integrals, while those connecting the wave functions to each other have become the usual spatial integrals for the Hilbert-space inner product.

Exercise VC1.1

Show that the amplitude can also be written as

$$\mathcal{A} = \check{\Psi}\left[\frac{\delta}{\delta\phi}\right] Z[\phi]\Psi_0[\phi]\bigg|_{\phi=0}$$

$$\check{\Psi}\left[\frac{\delta}{\delta\phi}\right] = \sum_{N=0}^{\infty} \tfrac{1}{N!} \int \frac{d^{ND}x}{(2\pi)^{ND/2}} \psi_N(x_1, ..., x_N) \frac{\delta}{\delta\phi(x_1)} \cdots \frac{\delta}{\delta\phi(x_N)}$$

$$\Psi_0 = e^{-\langle\phi|\phi\rangle/2}, \qquad \langle\phi|\phi\rangle \equiv \int dp\ 2\pi\delta[\tfrac{1}{2}(p^2+m^2)]|\phi(p)|^2$$

making all inner products of non-scattering particles explicit. (Hint: Write both this and the earlier form of the amplitude completely explicitly, in terms of operators on $Z[\phi]\Psi[\varphi]$.) We can interpret Ψ_0 as the free "vacuum wave functional" and $Z\Psi_0$ as the interacting vacuum wave functional.

2. Graphs

Before giving applications of these rules, we consider a few general properties. A convenient way to describe the terms in the expansion of the two exponentials in Z is pictorially, by "Feynman diagrams/graphs". Each factor of S_I is a "vertex" of the graph, represented by a dot in the diagram; each factor of the propagator $1/K$ is a "link" in the graph, represented by a line connecting the two dots representing the two factors of S_I on which each $\delta/\delta\varphi$ acts. (Both derivatives can also act on the same factor of S_I, giving a loop.) So any term in the expansion of Z is represented by a diagram consisting of a bunch of dots (interaction vertices) connected by lines (propagators). When we want to "draw" the amplitude $\mathcal{A}$ itself, we also draw additional lines, each with one end attached to a vertex and one end unattached. These "external lines" represent the one-particle wave functions coming from Ψ_N, and not propagators ("internal lines"). While in the diagram for Z each vertex can have φ dependence, in the diagram for $\mathcal{A}$ there is none, and the number of lines (internal and external) coming from any vertex explicitly indicates the order in φ of the corresponding term in S_I.

The physical interpretation of these diagrams is simple: The lines represent the paths of the particles, where they act free, while the vertices represent their collisions, where they interact. These diagrams are generally evaluated in momentum space: We then can associate a particular momentum with each line (propagator), and momentum is conserved at each vertex. An arrow is drawn on each line to indicate the direction of "flow" of the momentum. (Otherwise there is a sign ambiguity, since complex conjugation in position space changes the sign of the momentum.) Then the sum of all momenta flowing into (or all out of) a vertex vanishes. The momentum associated with a line is then interpreted as the momentum of that particle, with the arrow indicating the direction of flow of the proper time τ. (p changes sign with

τ.) When evaluating an S-matrix element, the fact that the external-line wave functions satisfy the free wave equation means the external momenta are on-(mass-)shell ($p^2 + m^2 = 0$); on the other hand, this is not true of the momenta on the internal lines, even though those particles are treated as free.

Some graphs in the S-matrix are disconnected: They can be divided into separate parts, each with a subset of the particles that interact with each other but not with the other subsets. For convenience, we consider only the connected graphs: If we write

$$Z[\varphi] = e^{-iW[\varphi]}, \qquad \mathcal{A}_c = \hat{\Psi}\left[\frac{\delta}{\delta\varphi}\right](-i)W[\varphi]|_{\varphi=0}$$

then W is the generating functional for the connected S-matrix $\mathcal{A}_c$. To prove this relation between Z and W, we first note that it is just the combinatorics of the graphs, and has nothing to do with spacetime. Therefore, it is sufficient to consider the simple (unphysical) case where the action has no derivatives. Since the propagator is then local, connectedness is equivalent in this case to locality. We then observe that the lack of derivatives allows the functional integral to be factorized explicitly into ordinary integrals at each point in spacetime:

$$Z[\varphi] = \int D\phi\ e^{-i\int dx\ L(\phi(x),\varphi(x))} = \prod_x \int d\phi(x)\ e^{-iL(\phi(x),\varphi(x))}$$

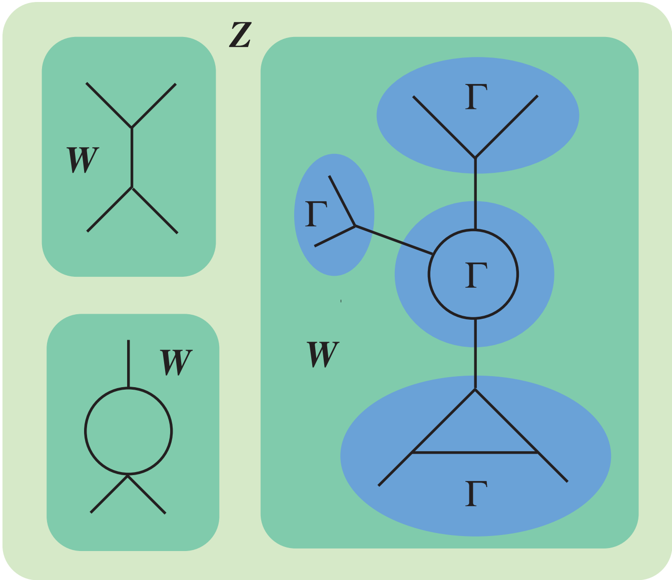
$$= \prod_x \mathcal{Z}(\varphi(x)) = e^{-i\int dx\ \mathcal{W}(\varphi(x))} = e^{-iW[\varphi]}$$

Thus, this W is local, and therefore connected; this implies W is connected in the general case. The simplest kind of connected graph is a "tree" graph, which is a graph that has no closed paths; the rest are called "loop" graphs.

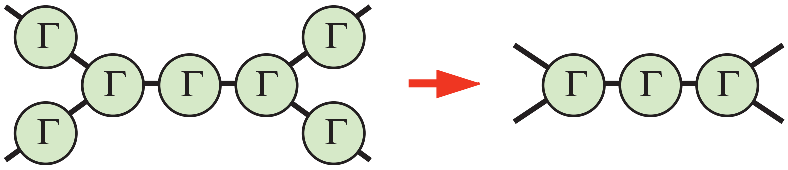
"One-particle-irreducible" (1PI) graphs are defined to be those connected graphs that can't be disconnected by severing a single propagator. It then follows that any connected graph can be represented as a generalized tree graph, whose "vertices" (*including* two-point vertices) are actually 1PI graphs. We then define the "effective action" $\Gamma[\phi]$ to be the classical action plus all 1PI loop graphs. Note that the vertices of the original action are the 1PI tree graphs; thus Γ is also the classical kinetic term plus all (tree and loop) 1PI graphs. Actually, since the 1PI tree graphs are $-iS_I$, we define the classical part of Γ to be S, but the quantum part to be the quantum 1PI part of W. Of course, the effective action is nonlocal. However, the tree graphs that follow from this action are exactly all the connected graphs of the original action: This is clear for all but the 2-point vertices from the definition. For the propagator and its relation to the 2-point 1PI loop graphs, we simply compute the expression following from Γ: Denoting the 2-point 1PI loop operator as A, the kinetic operator of Γ is $K+A$. The propagator following from Γ is then

$$\frac{1}{K+A} = \frac{1}{K} - \frac{1}{K}A\frac{1}{K} + \frac{1}{K}A\frac{1}{K}A\frac{1}{K} - ...$$

But this is exactly the result of the complete propagtor (including loop graphs) following from the original action.



This quantum modification of the propagator leads us to reanalyze our prescription for evaluating the S-matrix: For example, even in the simplest case, where this A is just a constant, the full propagator differs from the free propagator by a change in the mass. This means the mass of asymptotic states should also be changed, which invalidates part of our evaluation of the S-matrix in the previous subsection. Similar problems occur when A is proportional to K, which changes the normalization of asymptotic states. There are two ways to fix these problems: (1) We compensate by modifying the kinetic term in the classical action, replacing K with K minus such local contributions from A. Treating these new terms as part of the interaction in our derivation of the S-matrix, so our normalization and mass in the free propagator are unchanged, these "interaction" terms cancel the unwanted terms in the quantum propagator, so it then has the same residue and pole as the free one. This procedure is known as "renormalization", and will be discussed further in chapter VII, primarily for the purpose of eliminating infinities. (2) Alternatively, we modify our derivation of the S-matrix to take the full propagator into account. The easiest way to see this change is to remember that by definition the S-matrix follows from treating the effective action as a classical action (except for its nonlocality and nonhermiticity), but keeping only the "tree" graphs. Then clearly (a) the quadratic part Γ_0 of Γ is used to define the asymptotic states, and (b) instead of eliminating all free propagators except those connecting factors of S_I, we eliminate all full propagators (found from Γ_0) except those connecting factors of Γ_I, the nonquadratic part of Γ. In other words, we modify our earlier definition of the S-matrix by dropping all graphs that have *any* quantum correction to external lines. Thus, this procedure can also be applied in the case of renormalization; in fact, it should be applied in general, simply because it allows us to immediately ignore many graphs. It also allows us to avoid confusion resulting from attaching wave functions of the wrong mass to propagator corrections: E.g., in momentum space, we would have to interpret ambiguous factors such as $\delta(K)A(1/K)$..., where the factor of $\delta(K)$ comes from a plane-wave wave function.



This analysis of the quadratic part of Γ also leads us to examine the terms of lower order: constant and linear. The constant term is just a normalization, and should be dropped. (This is not true in the case of gravity, where a constant term in the Lagrangian is not gauge invariant by itself.) The linear term describes the decay of a particle into the vacuum: It implies we have the wrong vacuum. A linear term necessarily has no derivatives (otherwise it is a boundary term, which vanishes by our boundary conditions); it is part of the "effective potential" (a generalization of the potential energy, whose contribution to a classical mechanics Lagrangian contains no time derivatives; see subsection VIIB2). The existence of a linear term means that the minimum of the effective potential, i.e., the true vacuum, is not described by vanishing fields. To correct this situation we therefore apply the same procedure as for the classical action (chapter IV): (1) Shift the appropriate fields by constants, to put us at the minimum of the potential, and (2) use the new quadratic terms in the potential to determine the true masses of states defined by perturbation about this new vacuum. Again eliminating any constant terms, the resulting Γ has only quadratic and higher-order terms.

To summarize, the general procedure for calculating Feynman graphs is:

(1) Calculate the effective action, i.e., the 1PI graphs.

(2) Shift the scalars to put them at the minimum of the effective potential, dropping the resultant constant, to reveal the true masses of all particles.

(3) Calculate the S-matrix from diagrams without external-line corrections, with external wave functions whose normalization and masses are determined by the zeroes of the kinetic operators in the shifted Γ.

Another use of the effective action, besides organizing the calculation of the S-matrix, is for studying low-energy behavior: This means we apply an expansion in derivatives, as in first-quantized JWKB (see subsection VA2). Of most interest is the lowest order in the approximation, where all fields are effectively constant: This gives the effective potential. (In practice, "all fields" means just the scalars, since constant spinor fields are not generally useful, while higher spins are described by gauge fields, whose constant pieces can be set to vanish in an appropriate formulation: E.g., the constant piece of the metric tensor can be attributed to a scalar — see subsection IXA7.) However, the definition of "1PI" graphs is ambiguous, depending on how we define "particle": For example, if we include auxiliary fields in the effective action (as in supersymmetry, but also for bound-state problems: see subsections VIIB3 and 6), the result at fixed order in any expansion parameter ($\hbar$, coupling, etc.) is different, since the auxiliaries get contributions at each order, so elminating them by their effective-action field equations mixes orders. (E.g., $B^2 + \hbar B f(A) \to \hbar^2 f^2$.) This is crucial when the composite fields defined by these auxiliaries, and thus the auxiliaries themselves, obtain vacuum values. Therefore, the effective action is most useful for these purposes when, for appropriate choice of fields and definition of $\hbar$, a useful first-quantized semiclassical expansion can be found. Another important use of the effective action is that it is gauge invariant (even in the nonabelian case, when using the background-field gauge; see subsection VIB8): Sometimes simplifications due to gauge invariance are thus easier to see in the effective action than in the S-matrix.

We now consider an interesting topological property of graphs:

(1) For any graph, if we draw an extra propagator from a vertex to itself or another, that gives an extra "loop" (closed circuit) and no extra vertices.

(2) Adding a 2-point vertex to the middle of a propagator or external line gives an extra propagator and no extra loops.

(3) Adding an external line to a vertex changes nothing else.

Since any nontrivial (not a lone propagator) connected graph can be built up this way from a lone vertex, we find

$$P - V = L - 1$$

for P propagators, V vertices, and L loops (and E external lines). The same result follows from counting momentum integrals: In momentum space there is an internal momentum, and corresponding integral, for each propagator, and a momentum conservation condition, and corresponding δ function, for each vertex. The only independent momenta are the external ones (associated with each φ in $Z[\varphi]$) and one momentum vector for each loop. Thus, after integrating out all the delta functions, except for an overall momentum conservation δ function for each connected graph, we are left with integrations over only the loop momenta. So, we are again led to the above result.

Exercise VC2.1

For the figure at the beginning of this subsection, check this identity for each of the 3 connected graphs. Apply the above construction to produce each of them from a single vertex.

3. Semiclassical expansion

We can define perturbations by inserting $\hbar$'s in various ways, as discussed in subsection IIIA3. The $\hbar$ that defines classical mechanics yields an expansion in derivatives on "matter" fields (those that describe classical particles in the limit $\hbar \to 0$, as opposed to the "wave" fields). This expansion is covariant as long as the $\hbar$ multiplies covariant derivatives. However, it can't be applied to Yang-Mills fields, and it doesn't correspond to a diagrammatic expansion. On the other hand, the $\hbar$ that defines classical field theory is an expansion in the number of "loops", and allows us to group graphs in gauge-invariant sets, since gauge transformations are not $\hbar$-dependent: As in quantum mechanics, we can perform a JWKB expansion by appropriately inserting $\hbar$:

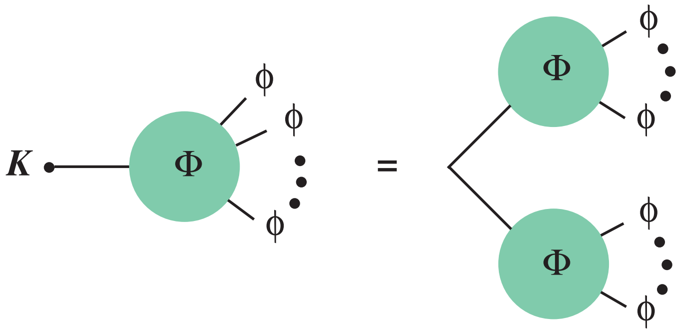
$$\begin{aligned} Z[\varphi] &= \int D\phi \; exp\left\{ -\frac{i}{\hbar}\left(S_0[\phi] + S_I[\phi+\varphi]\right)\right\} \\ &= exp\left\{ -i\hbar \int dx \; \tfrac{1}{2}\frac{\delta}{\delta\varphi}\frac{1}{K}\frac{\delta}{\delta\varphi}\right\} e^{-iS_I[\varphi]/\hbar} \end{aligned}$$

The order in $\hbar$ has a simple graphical interpretation. We see that there is a factor of $\hbar$ for each propagator and a factor of $1/\hbar$ for each vertex. Thus, by the above topological identity, for each connected graph the power in $\hbar$ is one less than the number of loops. We therefore write

$$Z[\varphi] = e^{-iW[\varphi]/\hbar}, \qquad W = \sum_{L=0}^{\infty} \hbar^L W_L$$

where W_0 generates the connected "tree" graphs, which have no loops.

We know that the leading term in the JWKB expansion is associated with the classical theory. We can make this more explicit in the field theory case by finding the general classical (perturbative) solution to the field equations from the tree graphs. Graphically the solution is very simple: We replace one φ on each tree graph with a propagator, and associate the end of the propagator with the position of the classical field $\Phi(x)$. If we then act on this Φ, which is a sum over all tree graphs, with K, it cancels the propagator, leaving a bunch of Φ's (also sums over all tree graphs) connected at x, with the appropriate vertex factor. In other words, we find $K\Phi = -\delta S_I[\Phi]/\delta\Phi$, the classical field equations.



To prove this, it's convenient to again use functionals, to automatically keep track of all combinatorics. The quantum field equations can be derived from the general identity

$$\int D\phi \; \frac{\delta}{\delta\phi} f[\phi] = 0$$

since we only integrate functionals f that are assumed to fall off fast enough as $\phi \to \infty$ to kill all boundary terms. (This follows from the perturbative definition of the functional integral.) In particular, for any action $\tilde{S}$,

$$0 = \int D\phi \; i\hbar\frac{\delta}{\delta\phi} e^{-i\tilde{S}/\hbar} = \int D\phi \; \frac{\delta\tilde{S}}{\delta\phi} e^{-i\tilde{S}/\hbar}$$

For our present purposes, we choose

$$\tilde{S} = S_0[\phi] + S_I[\phi + \varphi]$$

$$\Rightarrow \quad 0 = \int D\phi \; \left(\frac{\delta S_0[\phi]}{\delta\phi} + \frac{\delta S_I[\phi+\varphi]}{\delta\phi} \right) e^{-i\tilde{S}/\hbar} = \int D\phi \; \left(K\phi + i\hbar\frac{\delta}{\delta\varphi} \right) e^{-i\tilde{S}/\hbar}$$

$$\Rightarrow \quad K\langle\phi\rangle_\varphi + \frac{\delta W[\varphi]}{\delta\varphi} = 0$$

where "$\langle\phi\rangle_\varphi$" is the expectation value of the field in a background:

$$\langle\phi\rangle_\varphi = \frac{\int D\phi \; \phi e^{-i\tilde{S}/\hbar}}{\int D\phi \; e^{-i\tilde{S}/\hbar}}$$

(and, of course, $\int D\phi \; e^{-i\tilde{S}/\hbar} = e^{-iW/\hbar}$).

We now examine the classical limit $\hbar \to 0$ of this result by noting that if we impose the free field equation on the background

$$K\varphi = 0 \quad \Rightarrow \quad \frac{\delta \tilde{S}}{\delta \phi} = \frac{\delta S[\phi + \varphi]}{\delta \phi}$$

where $S[\phi] = S_0[\phi] + S_I[\phi]$ is the usual action: In other words, the field equations following from $\tilde{S}$ are just the usual field equations for the complete field

$$\tilde{\phi} = \phi + \varphi, \qquad \varphi = \lim_{x\to\infty} \tilde{\phi}$$

since we chose our boundary conditions so $\phi \to 0$ as $x \to \infty$ (including $|t| \to \infty$, whereas $\varphi \to 0$ only at spatial infinity). We then apply the stationary-phase approximation (or, after Wick rotation, the steepest-descent approximation)

$$\boxed{\lim_{\hbar\to 0} \int D\phi \; f[\phi] e^{-i\tilde{S}/\hbar} = \left(f[\phi] e^{-i\tilde{S}/\hbar} \right)\Big|_{\delta\tilde{S}/\delta\phi = 0}}$$

for $f = \phi$ and 1 to find

$$K\Phi + \frac{\delta W_0[\varphi]}{\delta \varphi} = 0 \quad \Rightarrow \quad \Phi = \varphi - \frac{1}{K} \frac{\delta W_0[\varphi]}{\delta \varphi}$$

$$\Phi \equiv \lim_{\hbar\to 0} \langle \tilde{\phi} \rangle_\varphi = \tilde{\phi}|_{\delta S[\tilde{\phi}]/\delta\tilde{\phi} = 0}$$

Thus, Φ is the solution to the classical field equations with boundary condition $\Phi \to \varphi$, and can be found directly from the classical (0-loop) part of W by replacing one field with a propagator.

A similar result holds for W_0 itself, by taking the classical limit as above for $f[\phi] = 1$:

$$W_0[\varphi] = S_0[\phi] + S_I[\phi + \varphi]$$

when evaluated at the result of varying the above with respect to either argument:

$$\frac{\delta}{\delta \phi} \quad \Rightarrow \quad \frac{\delta S[\phi + \varphi]}{\delta \phi} = K\varphi = 0$$

$$\frac{\delta}{\delta \varphi} - \frac{\delta}{\delta \phi} \quad \Rightarrow \quad \frac{\delta W_0[\varphi]}{\delta \varphi} = -K\phi$$

The former follows directly from the limiting procedure; the latter we have just proven equivalent by evaluating the limit for $\langle \tilde{\phi} \rangle_\varphi$. Since by definition the effective action Γ is related to W in exactly the same way that S is related to W_0 (the trees from Γ give the full W), we also have

$$W[\varphi] = \Gamma_0[\phi] + \Gamma_I[\phi + \varphi]$$

$$\frac{\delta\Gamma[\phi+\varphi]}{\delta\phi} = \tilde{K}\varphi = 0 \quad \Leftrightarrow \quad \frac{\delta W[\varphi]}{\delta\varphi} = -\tilde{K}\phi$$

where $\tilde{K}$ is the kinetic operator appearing in Γ_0, the quadratic part of Γ. (Some care must be taken for the fact that the poles and residues of $\tilde{K}$ in p^2 may differ from those of K, as discussed in the previous subsection.)

In practice, if one wants to make use of the classical field equations perturbatively, one looks at tree graphs with a specific number of external lines: For example, in a scalar theory with ϕ^3 interaction (assuming $\langle\Phi\rangle = 0$),

$$K\Phi + \tfrac{1}{2}\Phi^2 = 0, \quad \Phi = \sum_{n=1}^{\infty}\Phi_n \quad \Rightarrow \quad K\Phi_n + \tfrac{1}{2}\sum_{m=1}^{n-1}\Phi_m\Phi_{n-m} = 0$$

gives a recursion relation for the term Φ_n that is nth order in φ.

Exercise VC3.1

Consider the relativistic Schrödinger (Klein-Gordon) equation for a scalar wave function ψ in an external scalar potential ϕ:

$$(K+\phi)\psi = 0$$

(If you find it less confusing, you can consider the nonrelativistic case $K = \vec{p}^{\,2}/2m - E$, where $H = \vec{p}^{\,2}/2m + \phi$.) Find the perturbative solution for the quantum mechanical (one-particle) S-matrix for ψ (see exercise VA4.1). Show that this agrees with the contribution to the field-theoretic S-matrix for the Lagrangian

$$L(\psi,\phi) = \psi^*(K+\phi)\psi + L_\phi(\phi)$$

coming from tree graphs with an external ψ line, an external ψ^* line, and an arbitrary number of external ϕ lines.

Exercise VC3.2

Consider, instead of the background field φ, a "current" source J that attaches propagators to external lines. The current is effectively just a one-point interaction (it caps loose ends of propagators), so it can be introduced into the generating functional by the modification

$$S_I[\phi] \to S_I[\phi] + \int dx\ J\phi$$

We now have

$$e^{-iW[J]/\hbar} = \int D\phi\ exp\left\{-\frac{i}{\hbar}\left(S_0[\phi] + S_I[\phi] + \int dx\ J\phi\right)\right\}$$

$Z[J]$ is thus the Fourier transform of $e^{-iS[\phi]/\hbar}$ with respect to the conjugate variables ϕ and J.

a Derive the "Schwinger-Dyson equations"

$$\left(\left. \frac{\delta S[\phi]}{\delta \phi(x)} \right|_{\phi = i\hbar\delta/\delta J} + J(x) \right) e^{-iW[J]/\hbar} = 0$$

b Find the classical limit

$$W_0[J] = S[\phi] + \int J\phi \quad at \quad \phi = \frac{\delta W_0[J]}{\delta J} \quad \Leftrightarrow \quad J = -\frac{\delta S[\phi]}{\delta \phi}$$

(i.e., $W_0[J]$ is the "Legendre transform" of $S[\phi]$). Find the corresponding relation for $\Gamma[\phi]$ and $W[J]$.

c Show that the free part of $W[J]$ is given by

$$W_{free} = -\tfrac{1}{2} \int dx \ J \frac{1}{K} J \quad \Rightarrow \quad \varphi \equiv \phi_{free} = -\frac{1}{K} J$$

Note that J can be replaced with φ in $W[J]$ everywhere *except* the free term, since in all other terms J appears only in the combination $(1/K)J$. Show how this can be done in such a way as to reproduce the results above for the solution of the classical field equations in terms of $W[\varphi]$. Warning: Before this substitution we use $K\varphi + J = 0$, but afterwards we apply $K\varphi = 0$; also beware of integration by parts, since φ does not vanish at ∞, so the naive substitution $\phi \to \phi + \varphi$ is not very helpful. (Historically, Z was introduced as a functional of J. However, the only two applications of Feynman diagrams, the S-matrix $\mathcal{A}$ and the effective action Γ, both required that the external-line propagators resulting from the Feynman rules for $Z[J]$ be "amputated". Therefore, we use background fields exclusively. The resulting derivations, generalities, and applications are at least as simple as and often a little simpler than the corresponding ones with current sources.)

There is a major flaw in this relation between classical field theory and tree graphs, the "Klein paradox". The difference is that classical field theory uses fields everywhere, not wave functions. Thus the propagator must be real (or pure imaginary, depending on conventions) to preserve the reality (or complex conjugation) properties of the fields; usually one uses the retarded propagator. On the other hand, in quantum field theory negative-energy states must travel backward in time to preserve positivity of the true energy, so the complex Stückelberg-Feynman propagator must be used. Furthermore, in classical field theory the external line factors are the fields, which contain both positive and negative energies, the same on each line. In quantum field theory, each external line carries a different one-particle wave function, positive energy in the asymptotic past or negative energy in the asymptotic future.

4. Feynman rules

It is usually most convenient to calculate Feynman diagrams in Wick-rotated (to eliminate i's) momentum space (where massive propagators are simpler). The "Feynman rules" are then read off of the action as

$$\boxed{S = \int dx \ \tfrac{1}{2}\phi K \phi + S_I[\phi] \quad \Rightarrow \quad Z[\phi] = e^{-W[\phi]} = exp\left(\int \tfrac{1}{2} \frac{\delta}{\delta\phi} \frac{1}{K} \frac{\delta}{\delta\phi} \right) e^{-S_I[\phi]}}$$

where in $Z[\phi]$ we simply replace each field ϕ with a single-particle wave function in all possible permutations, since for the case of an N-particle amplitude we usually write the wave function as the product of N single-particle wave functions (although more generally it is a linear combination of these):

$$\boxed{\mathcal{A}_N = \hat{\Psi}_N \left[\frac{\delta}{\delta\phi} \right] Z[\phi] \ \left(\mathcal{A}_{N,c} = -\hat{\Psi}_N \left[\frac{\delta}{\delta\phi} \right] W[\phi] \right), \quad \hat{\Psi}_N \left[\frac{\delta}{\delta\phi} \right] = \prod_{i=1}^{N} \left(\int \psi_{Ni} \frac{\delta}{\delta\phi} \right)}$$

We Fourier transform as

$$\phi(x) = \int dp \ e^{ip\cdot x}\phi(p), \qquad \phi(p) = \int dx \ e^{-ip\cdot x}\phi(x)$$

(Of course, $\phi(x)$ and $\phi(p)$ are different functions, but the distinction should be clear from context.) In practice we choose the single-particle wave functions to be eigenstates of the momentum, so

$$\psi_i(x) = e^{ip_i\cdot x}\hat{\psi}_i(p_i), \qquad \psi_i(p) = \delta(p-p_i)\hat{\psi}_i(p_i) \qquad (p_i^2 + m^2 = 0)$$

where $\hat{\psi}$ is some simple factor (e.g., 1 for a scalar). Then the external line factor terms become

$$\int dx \ \psi_i(x) \frac{\delta}{\delta\phi(x)} = \hat{\psi}_i(p_i) \frac{\delta}{\delta\phi(p_i)}$$

while for propagator terms

$$\int dx \ \tfrac{1}{2} \frac{\delta}{\delta\phi(x)} \frac{1}{K(-i\partial)} \frac{\delta}{\delta\phi(x)} = \int dp \ \tfrac{1}{2} \frac{\delta}{\delta\phi(p)} \frac{1}{K(p)} \frac{\delta}{\delta\phi(-p)}$$

and for vertex terms

$$\int dx \ \phi_1(x)...\phi_n(x) = \int dp_1...dp_n \ \phi_1(p_1)...\phi_n(p_n)\delta(p_1 + ... + p_n)$$

where each of the ϕ's in the vertex may represent a field with derivatives; then we replace $-i\partial$ on $\phi(x)$ with p on $\phi(p)$. Thus, e.g., we have

$$\mathcal{A}_N = \left[\prod \hat{\psi}_{Ni}(p_i)\frac{\delta}{\delta\phi(p_i)}\right] Z[\phi]$$

Exercise VC4.1

Use the definition

$$\frac{\delta}{\delta\tilde{\phi}(p)}\tilde{\phi}(p') = \delta(p-p')$$

to show that

$$\widetilde{\left(\frac{\delta}{\delta\phi}\right)}(p) = \frac{\delta}{\delta\tilde{\phi}(-p)}$$

where we now use tildes to indicate Fourier transformation.

Note that there is some ambiguity in the normalization of external line factors, associated with the numerator factor in the propagator

$$\varDelta = \frac{1}{K} = \frac{N(p)}{\frac{1}{2}(p^2+m^2)}$$

For nonzero spin (or internal symmetry), we have already discussed the normalization analogous to that for scalars, namely

$$\sum \hat{\psi}^\dagger(p)\hat{\psi}(p) = \pm N(p)$$

(with $-$ for negative energy and half-integer spin). However, there is already some freedom with respect to coupling constants: Even for scalars, if a coupling appears in the kinetic term as a factor of $1/g^2$, then effectively the kinetic operator is $K = K_0/g^2$, where K_0 is the usual (coupling-independent) one. Thus $N = g^2 N_0$, so $\hat{\psi} = g\hat{\psi}_0$, meaning coupling dependence in external lines. Alternatively, this external-line factor of the coupling can be included in the definition of probabilities in terms of amplitudes, which already includes nontrivial factors because of the use of (non-normalizable) plane waves. Furthermore, quantum effects modify the form of the propagator: Such effects can be absorbed near the pole $p^2 = -m^2$ by a field redefinition, but often it is more convenient to leave them. Then N will again have a constant, (but more complicated) coupling-dependent factor, which must be canceled in either the external-line factors or probabilities. (However, note that these questions do not arise in calculations of the functionals $W[\phi]$ or $\varGamma[\phi]$.)

In tree graphs all momentum integrals are trivial, with the momentum conservation δ functions at each vertex, and the δ functions of the external lines, determining internal momenta in terms of external momenta. In loop graphs there is a momentum integral left for each loop, over the momentum of that loop. The amplitude will always have an overall δ function for momentum conservation for each connected piece of the graph. Since we are always interested in just the connected graphs, we pull this conservation factor off to define the "T-matrix": Including the factor of "i" from Wick rotating back to Minkowski space,

$$\mathcal{S}_{connected} = i\delta\left(\sum p\right)T$$

In general there will be combinatoric factors associated with a graph. These follow automatically from the functional expressions, but can also be seen from the symmetries of the graph. Here "symmetries" means ways in which the graph can be twisted, with external lines fixed, such that the graph looks the same, including the types of particles propagating along the lines. For example, a graph with 2 vertices that are connected by n identical propagators would get a factor of 1/n! for that symmetry. There are also sign factors from fermions: Permutation of external fermion lines gives minus signs, because it involves permutation of anticommuting fields in the functionals. Each fermion loop gets a minus sign for the same reason. (This is related to the fact that fermionic integration gives determinants instead of inverse determinants.) Explicitly, it comes from evaluating expressions of the form (ignoring momentum dependence and external fields)

$$\left(\frac{\delta}{\delta\bar\psi}\frac{\delta}{\delta\psi}\right)\cdots\left(\frac{\delta}{\delta\bar\psi}\frac{\delta}{\delta\psi}\right)(\bar\psi\psi)\cdots(\bar\psi\psi)$$

$$= \left(\frac{\delta}{\delta\bar\psi}\frac{\delta}{\delta\psi}\right)\bar\psi\left[\left(\frac{\delta}{\delta\psi}\psi\right)\left(\frac{\delta}{\delta\bar\psi}\bar\psi\right)\right]\cdots\left[\left(\frac{\delta}{\delta\psi}\psi\right)\left(\frac{\delta}{\delta\bar\psi}\bar\psi\right)\right]\psi$$

where the propagator derivatives $(\delta/\delta\bar\psi)(\delta/\delta\psi)$ give no signs connecting up successive vertex factors $\bar\psi\psi$, but the last one does in closing the loop.

The general rules for contributions to the (unrenormalized) effective action $\Gamma[\phi]$ are then:

(A1) 1PI graphs only, plus S_0 $(\int \frac{1}{2}\phi K\phi)$.
(A2) Momenta: label consistently with conservation, with $\int dp$ for each loop.
(A3) Propagators: $1/\frac{1}{2}(p^2+m^2)$, or $1/K$, for each internal line.
(A4) Vertices: read off of $-S_I$.
(A5) External lines: attach the appropriate (off-shell) fields and $\int dp$, with $\delta(\sum p)$.
(A6) Statistics: 1/n! for n-fold symmetry of internal/external lines;
-1 for fermionic loop; overall -1.

(If we want to calculate $W[\phi]$ instead, then simply replace step 1 with "Connected graphs only".) The next step is to analyze the vacuum:

(B1) Find the minimum of the effective potential (for scalars).
(B2) Shift (scalar) fields to perturb about minimum; drop constant in potential.
(B3) Find resulting masses; find wave function normalizations.

Renormalization is performed either before or after this step, depending on the scheme. Finally, the trees from Γ are identified with the complete amplitudes from S. Thus, T-matrix elements are given by:

(C1) Connected "trees" of (shifted, renormalized) Γ: (A2-4) for L=0 with $S \to \Gamma$.
(C2) "Amputate" external Γ_0-propagators.
(C3) External lines: 1, or appropriate to Γ_0 wave equation $\tilde{K}\psi = 0$ ($\sum \hat{\psi}^\dagger \hat{\psi} = \pm N$).
(C4) External-line statistics: No symmetry factors; -1 for fermion permutation.

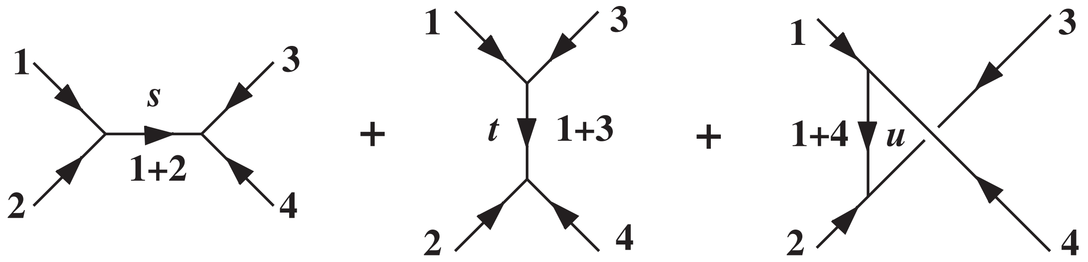
Note that Γ is usually simpler than T with respect to treatment of external lines: In T we often have contributions from graphs which are identical except for permutation of external lines from identical fields. In Γ only one such graph need be considered, since the statistics of the attached external fields automatically takes care of this symmetry. (We then also drop the 1/n!, or at least reduce it.)

We label the lines of a graph with arrows to indicate the direction of "flow" of momenta: Momentum conservation means the total momentum flowing into any vertex is equal to that flowing out, which we use to eliminate dependent momenta (integrating out δ functions). (In tree graphs all internal momenta are determined by external ones, which are constrained by conservation of total external momentum.) This "sign" of direction of the arrow is independent of the sign of the energy; we must combine the two to determine whether an external state is initial or final: An incoming external line with positive energy is initial, negative energy is final; an outgoing line is the opposite (i.e., for positive energy the arrow indicates the direction of time, but negative energy means travel backward in time). The choice of direction of arrows is arbitrary, and the convenience of any choice depends on the graph and the theory. There is no correspondence between this choice and signs of energy, since generally one wants to apply the same graph for cases with each external line with either sign, whereas internal lines in trees may have either sign depending on the external kinematics, and those in loops must be summed over both signs. Often the direction of the arrow is chosen to indicate the direction of flow of positive charge, when such a quantum number (U(1) symmetry) exists.

The simplest nontrivial tree graphs are 4-point amplitudes. We now label all momenta as incoming, which is convenient for symmetry, and corresponds naturally to using incoming (initial) states with positive energy and outgoing (final) states with negative energy (as from the complex-conjugate final wave functions). These momenta are conveniently expressed in terms of the Mandelstam variables (see subsection IA4): with these signs,

$$s = -(p_1 + p_2)^2, \qquad t = -(p_1 + p_3)^2, \qquad u = -(p_1 + p_4)^2$$

We also use the convention that s is defined in terms of the momenta of the two initial particles (and we also use this same definition when there are more than two final particles); t and u are then more or less interchangable, but if initial and final particles are pairwise related we choose t in terms of the momenta of such a pair.



For example, the simplest nontrivial theory is ϕ^3 theory, with

$$K = \tfrac{1}{2}(-\Box + m^2) \to \tfrac{1}{2}(p^2 + m^2)$$

$$S_I = \int dx\ \tfrac{1}{6} g\phi^3 = \int dp_1\ dp_2\ dp_3\ \tfrac{1}{6} g\phi(p_1)\phi(p_2)\phi(p_3)\delta\left(\sum p\right)$$

The four-point S-matrix amplitude at tree level (order g^2) then comes from using the following contributions to the factors in $\mathcal{A}$ (calculated in Euclidean space):

$$\hat{\Psi}_4 = \frac{\delta}{\delta\phi(p_1)} \frac{\delta}{\delta\phi(p_2)} \frac{\delta}{\delta\phi(p_3)} \frac{\delta}{\delta\phi(p_4)}$$

$$e^{\int (\delta/\delta\phi)(1/K)(\delta/\delta\phi)/2} = \ldots + \int dk\ \tfrac{1}{2} \frac{\delta}{\delta\phi(k)} \frac{1}{\frac{1}{2}(k^2 + m^2)} \frac{\delta}{\delta\phi(-k)} + \ldots$$

$$e^{-S_I} = \ldots + \tfrac{1}{2} S_I^2 + \ldots$$

Using $(\delta/\delta\phi(p))\phi(k) = \delta(p-k)$, keeping only the connected part, and integrating out the δ functions (except $\delta(\sum p_{external})$), we are left with

$$\mathcal{S}_c = \delta(p_1 + p_2 + p_3 + p_4) g^2 \left(\frac{1}{\frac{1}{2}(m^2 - s)} + \frac{1}{\frac{1}{2}(m^2 - t)} + \frac{1}{\frac{1}{2}(m^2 - u)} \right)$$

(There would be an extra i for the δ in Minkowski space.) Note the symmetry factor $1/3!$ for the ϕ^3 coupling, which is canceled upon taking 3 functional derivatives for the vertex factor. The T-matrix then comes from just factoring out the δ:

$$T = g^2 \left(\frac{1}{\frac{1}{2}(m^2 - s)} + \frac{1}{\frac{1}{2}(m^2 - t)} + \frac{1}{\frac{1}{2}(m^2 - u)} \right)$$

But this contribution to W is given by a single term:

$$W = -g^2 \int dp_1 dp_2 dp_3 \ \tfrac{1}{2}[\tfrac{1}{2}\phi(p_1)\phi(p_2)]\frac{1}{\frac{1}{2}(m^2 - s)}[\tfrac{1}{2}\phi(p_3)\phi(-p_1 - p_2 - p_3)]$$

or, in position space,

$$W = -g^2 \int dx \ \tfrac{1}{2}(\tfrac{1}{2}\phi^2)\frac{1}{\frac{1}{2}(m^2 - \Box)}(\tfrac{1}{2}\phi^2)$$

(The $\frac{1}{2}$'s correspond to the various symmetries: switching a pair connecting to the same vertex, or switching the two pairs.)

Exercise VC4.2

Find the 5-point tree amplitude for ϕ^3 theory. What order in g is the n-point tree?

The momentum integrals are real in Euclidean space: There are no singularities in the integrand, since $p^2 + m^2$ is always positive (although there are some subtleties in the massless case). Thus, all these integrals are most conveniently performed in Euclidean space. However, eventually the result must be analytically continued back to Minkowski space: $x^0 \to ix^0$, which means $p^0 \to ip^0$ (but $p_0 \to -ip_0$, being also careful to distinguish $\delta_m^n \to \delta_m^n$ and $\delta_{mn} \to \eta_{mn}$ for indices on fields) via a 90° rotation. This returns some of the i dependence. There are also i's associated with integration measures: Since all the momentum integrals for the S-matrix elements have already been performed, all that remains is a factor of i to go with $\delta(\sum p)$ for each connected graph. There can also be i dependence in external line factors.

Remember that negative p^0 indicates a particle traveling backward in time; the true motion of such a particle is opposite to that of the arrow indicating momentum flow. Thus, external lines with arrows pointing into the diagram and positive p^0, or out of the diagram and negative p^0, both indicate initial states, arriving from $t = -\infty$. Conversely, external lines with arrows pointing into the diagram and negative p^0, or out of the diagram and positive p^0, both indicate final states, departing to $t = +\infty$.

A related issue is particles vs. antiparticles. If a particle is described by a real field, it is identified as its own antiparticle; but if it is described by a complex field, then it is identified as a particle if it has a certain charge, and as an antiparticle if it has the opposite charge. (For example, a proton is positively charged while an antiproton is negatively.) Of course, this is convention, since a complex field can always be replaced by two real fields, and we can always relabel which is the field and which the complex conjugate; generally there should be a useful conservation law (symmetry) associated with these complex combinations (usually electric charge), and the one called "particle" is the one more common to the observer. For example, suppose we have a complex scalar external field/wave function $\phi(p)$. For $p^0 > 0$ this describes a particle propagating to $x^0 = +\infty$. Similarly, $\phi^*(p)$ for $p^0 > 0$ describes an *anti*particle propagating to $x^0 = +\infty$. On the other hand, ϕ^* for $p^0 < 0$ describes a particle propagating from $x^0 = -\infty$, while ϕ for $p^0 < 0$ describes an antiparticle propagating from $x^0 = -\infty$.

5. Semiclassical unitarity

As in nonrelativistic quantum mechanics, the only conditions for unitarity are that: (1) the metric (inner product) on the Hilbert space is positive definite (so all probabilities are nonnegative), and (2) the Hamiltonian is hermitian (so probabilities are conserved). Both of these conditions are statements about the classical action. The second is simply that the action is hermitian, which is easy to check. The first is that the kinetic (quadratic) terms in the action, which define the (free) propagators, have the right sign. This can be more subtle, since there are gauge and auxiliary degrees of freedom. Therefore, the simplest way to check is by using the lightcone formalism.

We see from the analysis of subsection IIB3 for field equations, or for actions in subsection IIIC2 (for spins ≤ 1, and more generally below in chapter XII) that after lightcone gauge fixing and elimination of auxiliary degrees of freedom the kinetic terms for physical theories always reduce to $-\frac{1}{4}\phi\Box\phi$ for bosons and $\frac{1}{4}\psi(\Box/i\partial^+)\psi$ for fermions, where there is a sum over all bosons and fermions, and each term in the sum has a field with a single hermitian component. Complex fields can multiply their complex conjugates, but these can always be separated into real and imaginary parts. There are never crossterms like $A\Box B$, since after field redefinition, i.e., diagonalization of the kinetic operator, this gives $A'\Box A' - B'\Box B'$, so one term has the wrong sign. Similar remarks apply to massive fields, but with $\Box$ replaced by $\Box - m^2$ (as seen, e.g., by dimensional reduction), or we can treat the mass term as part of the interactions.

Exercise VC5.1

What's wrong with $A\Box A + B\Box B + m^2 AB$? (Hint: something, but none of the above. Diagonalize.)

Now we only need to check that the single-component propagators of these two cases define positive-definite inner products. Since multiparticle inner products are products of uniparticle inner products, it's sufficient to look at one-particle states. We therefore examine the S-matrix defined in subsection VC1 for the special case of 1 particle at $t = -\infty$ going to 1 particle at $t = +\infty$, using the free, massless lightcone

Lagrangians given above. For the boson we found that this matrix element was simply the inner product between the two states, appearing in the form $\int d\sigma\, d\sigma'\ \psi\Delta\psi'$. This worked only because the propagator had the right sign. (It is essentially $+e^{-i\omega|t|}$.) Thus the sign we use is required for unitarity.

A simple way to treat the fermion is to use supersymmetry: Since the boson and fermion kinetic terms are spin independent in the lightcone formalism, we can look at any supersymmetric theory, and check the boson and fermion kinetic terms there. If the boson term agrees with the one we just checked, then the fermion term is OK by supersymmetry (which preserves unitarity by the lightcone-like supertwistor analysis of subsection IIC5). Alternatively, we can use the same method applied to the boson: The propagator now has an extra factor of $1/(-i\partial^+)$, or $1/p^+$ in momentum space. Since p^+ is always positive for positive energy, these states also appear with the correct-sign norm. To analyze negative energy (antiparticles), we note that in the derivation of the path integral final states always appear to the left and initial states to the right. In the fermionic case this is important because it will introduce an extra sign: Since

$$\psi_1 \frac{\Box}{\partial^+}\psi_2 = +\psi_2 \frac{\Box}{\partial^+}\psi_1$$

(with two signs canceling from reordering fermions and integration by parts), the right sign will always be produced with correct ordering of initial vs. final states (i.e., positive vs. negative energy), independent of the helicity (or whether electron vs. positron, etc.)

For similar reasons, it is clear that integral spin is always described by commuting (bosonic) fields, while half-integral spin is always described by anticommuting (fermionic) fields: The number of undotted minus dotted indices on a field is always odd for half-integer spin, even for integer, and a derivative carries one dotted and one undotted index, so contraction of all indices means an even number of derivatives for integer spin and odd for half-integer. Without loss of generality, we then can separate each field into its real and imaginary parts. Then for each real field integration by parts gives

$$\phi(-i\partial)^n\phi = (-1)^n[(-i\partial)^n\phi]\phi = (-1)^{n+\phi}\phi(-i\partial)^n\phi \quad \Rightarrow \quad (-1)^{n+\phi} = 1$$

where $(-1)^\phi$ is the statistical factor for ϕ (1 for bosons, -1 for fermions), and we have included the appropriate i's for hermiticity of the action. Thus, integer spin is associated with bosons ($(-1)^n = 1 = (-1)^\phi$), and half-integer with fermions ($(-1)^n = -1 = (-1)^\phi$). This is the "spin-statistics theorem". By using real fields with a diagonal kinetic term, we have implicitly assumed kinetic terms appear only with the correct sign: For example, for a complex bosonic field

$$\phi = A + iB \quad \Rightarrow \quad \phi^\dagger \partial \phi = A i \overleftrightarrow{\partial} B = -B i \overleftrightarrow{\partial} A$$

and thus has indefinite sign. Thus, spin and statistics follows from Poincaré invariance, locality, and unitarity. If we drop unitarity, we get "ghosts": We'll see examples of such wrong-statistics fields when quantizing gauge theory.

Note that demanding unitarity (the right sign of the kinetic term) is the same as demanding positivity of the true energy, as least as far as the kinetic term is concerned: The energy is given by the Hamiltonian of the field theory; if the kinetic term changes sign, the corresponding contribution to the Hamiltonian does also. (Compare the discussion of Wick rotation of the action in subsection VB4.)

Using anticommuting fields to describe fermions is more than a formality. In general, the significance of describing states by quantizing classical fields that commute or anticommute has two purposes: (1) to avoid multiple counting for indistinguishable particles, and (2) to insure that two identical fermions do not occupy the same state. Thus, when describing two particles in different states, the phase associated with (anti)commutation is irrelevant: A "Klein transformation" can be made that makes anticommuting quantities *commute* for different states, and anticommute (i.e., square to zero) only for the same state. However, such transformations are nonlocal, and locality is crucial in relativistic field theory. (See exercise IA2.3e.)

6. Cutting rules

For some purposes it is useful to translate the three defining properties of relativistic quantum field theory into graphical language. Poincaré invariance is trivial, since the propagators and vertices are manifestly Poincaré covariant in covariant gauges. Unitarity and causality can also be written in a simple way in functional notation. We first note that the inner product for free multiparticle wave functions can be written very simply in momentum space as

$$\langle\psi|\chi\rangle = \left(\psi^\dagger[\phi] e^{D_+} \chi[\phi]\right)\Big|_{\phi=0}, \qquad D_+ = \tfrac{1}{2}\int dp\, \frac{\overleftarrow{\delta}}{\delta\phi(p)} \Delta_+(p) \frac{\delta}{\delta\phi(-p)}$$

$$\Delta(p) = \frac{-iN(p)}{\frac{1}{2}(p^2 + m^2 - i\epsilon)} \quad \Rightarrow \quad \Delta_+(p) = \theta(p^0) 2\pi \delta[\tfrac{1}{2}(p^2 + m^2)] N(p)$$

where ψ and χ are products of positive-energy single-particle states, and Δ_+ (the "cut" propagator) projects onto the positive-energy mass shell. (The exponential takes care of the usual combinatoric factors.) We have written a generic propagator, with numerator factor N (=1 for scalars). (Without loss of generality, we have assumed a real basis for the fields, so N can be taken as real.) The S-matrix amplitude then can be written in operator language as

$$\Psi[\phi] = \psi^\dagger[\phi]\chi[\phi] \quad \Rightarrow \quad \mathcal{A} = \int D\phi\ \Psi[\phi]e^{-iS[\phi]} = \Psi\left[\frac{\delta}{\delta\phi}\right] Z[\phi] = \langle\psi|\mathcal{S}|\chi\rangle$$

We have used the fact that positive-energy states propagate forward in time and negative backwards to write the usual Hilbert-space inner product in terms of initial and final states of positive energy. The S-matrix operator $\mathcal{S}$ appears because χ and ψ satisfy the free equations of motion, and $\mathcal{S}$ performs time translation from $t = -\infty$ for χ to $t = +\infty$ for ψ to include interactions.

The unitarity condition is then (see subsection VA4)

$$\mathcal{S}^\dagger\mathcal{S} = 1 \quad \Rightarrow \quad Z[\phi]^\dagger e^{D_+} Z[\phi] = 1$$

while causality is

$$\frac{\delta}{\delta\phi(x)}\left(\mathcal{S}[\phi]^\dagger \frac{\delta}{\delta\phi(y)}\mathcal{S}[\phi]\right) = 0 \quad \Rightarrow \quad \frac{\delta}{\delta\phi(x)}\left(Z[\phi]^\dagger e^{D_+}\frac{\delta}{\delta\phi(y)}Z[\phi]\right) = 0 \quad for\ x^0 > y^0$$

This causality relation, which already holds in nonrelativistic field theory, can be strengthened by using Lorentz invariance: If $(x-y)^2 > 0$ (spacelike separation), then $x^0 < y^0$ can be Lorentz transformed to $x^0 > y^0$. Thus, the above expression vanishes everywhere except on or inside the backward lightcone with respect to $x - y$. These functional forms of unitarity and causality (and Poincaré invariance) can also be used as a basis for the derivation of the functional integral form of $Z[\phi]$ in terms of the action, rather than relying on its relation to the Hamiltonian formalism.

The fact that these conditions are satisfied by Feynman diagrams follows easily from inspection. We examine them using the explicit expression for Z following from the functional integral:

$$Z[\phi] = e^D e^{-iS_I}, \qquad Z[\phi]^\dagger = e^{D^*}e^{iS_I}$$

$$D = \tfrac{1}{2}\int dp\ \frac{\delta}{\delta\phi(p)}\Delta(p)\frac{\delta}{\delta\phi(-p)}, \qquad \Delta(p) = \frac{-iN}{\frac{1}{2}(p^2+m^2-i\epsilon)}$$

$$D^* = \tfrac{1}{2}\int dp\ \frac{\delta}{\delta\phi(p)}\Delta(p)^*\frac{\delta}{\delta\phi(-p)}, \qquad \Delta(p)^* = \frac{iN}{\frac{1}{2}(p^2+m^2+i\epsilon)}$$

These expressions can be translated straightforwardly into position space as

$$\int dp\ \frac{\delta}{\delta\phi(p)}\Delta(p)\frac{\delta}{\delta\phi(-p)} = \int dx\ dx'\ \frac{\delta}{\delta\phi(x)}\Delta(x-x')\frac{\delta}{\delta\phi(x')}$$

etc.

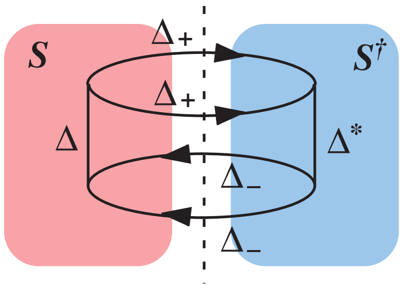
From the results at the end of subsection VB2, we see that the propagators satisfy the relations

$$\Delta_+(x) = \Delta(x) - \Delta_A(x) = \Delta^*(x) + \Delta_R(x)$$

$$\Delta_+(x) = \Delta_-(-x), \quad \Delta(x) = \Delta(-x), \quad \Delta_R(x) = \Delta_A(-x)$$

and, of course, $\Delta_R(x) = 0$ for $x^0 < 0$. We will now see that the cancelations in the unitarity and causality relations occur graph by graph: There are contributions consisting of a sum of terms represented by exactly the same diagram, with each term differing only by whether each vertex comes from Z or $Z^\dagger$. First, this affects the sign of the term, since each vertex from $Z^\dagger$ gets an extra sign from the e^{iS_I} in $Z^\dagger$, as compared to the e^{-iS_I} in Z. Second, this affects which propagators appear:

$$\begin{aligned} (x,y)\ in:\ (Z,Z) &\to \Delta(x-y) \\ (Z^\dagger,Z) &\to \Delta_+(x-y) \\ (Z,Z^\dagger) &\to \Delta_-(x-y) \\ (Z^\dagger,Z^\dagger) &\to \Delta^*(x-y) \end{aligned}$$



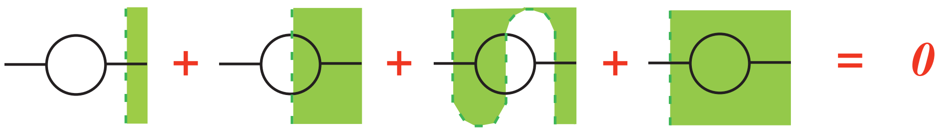
Now if we sum over two terms differing only by whether the position x of one particular vertex appears in Z or $Z^\dagger$, the result before integration is proportional to

$$\prod \Delta(x-y_i) \prod \Delta_-(x-z_j) - \prod \Delta_+(x-y_i) \prod \Delta^*(x-z_j)$$

where y_i are from Z and z_j are from $Z^\dagger$. However,

$$\Delta(x-y) - \Delta_+(x-y) = \Delta_-(x-y) - \Delta^*(x-y) = \Delta_R(y-x) = 0 \quad for \quad x^0 > y^0$$

Writing $\Delta = \Delta_+ + \Delta_R$ and $\Delta_- = \Delta^* + \Delta_R$ in the difference of the two products, each surviving term in the difference contains a Δ_R, and therefore the two products cancel if x^0 is the latest of all the vertices. We thus take any sum of the same diagram over different distributions of the vertices between Z and $Z^\dagger$ occuring in the unitarity or causality relation before integration over the coordinates of the vertices, separate the sum into pairs which are identical except for whether the latest vertex is in Z or $Z^\dagger$, and apply the above relation to show this difference vanishes. Thus, the vanishing of a sum of graphs indicated by unitarity or causality is actually satisfied by cancelation between each pair of terms before integration over coordinates. (Which pair is determined by the values of the coordinates, since we need to find the latest one; after integration, the cancelation is between the whole set of terms for the same graph.) In the unitarity relation we sum over whether a vertex occurs in Z or $Z^\dagger$ for each vertex, including the latest one, so that condition is easily satisfied. (The only diagram that survives is the one with no particles, which gives 1.) In the causality relation we perform this sum for each vertex except y (since $\delta/\delta\phi(y)$ acts only on Z), but since $y^0 < x^0$, y^0 is not the latest vertex, so again the latest one is summed over.



Exercise VC6.1

Consider an arbitrary 1-loop graph. Why would replacing all the propagators Δ with advanced propagators Δ_A (or all with retarded propagators Δ_R) all the way around the loop in the same direction give zero? Use this result, and the relation between the various propagators, to show that any one-loop diagram (with normal propagators) can be expressed as a sum of products of tree graphs, with some summations of external states ("Feynman tree theorem"). How does this differ from the cutting rule for unitarity? (Hint: Look at the signs of the energies of external states.)

There is one fine point in this construction: We may use Feynman rules from a complex action, such as those used for massive theories in subsection IIIC4, or when using complex gauges (see section VIB). In that case, since the S-matrix $\mathcal{S}$ is gauge-independent, and the original action S was real (before eliminating complex fields or choosing complex gauges), it is legal to use the *Hermitian conjugate* action $S^\dagger$ to define the Feynman rules for $\mathcal{S}^\dagger$ (and $Z^\dagger$): When multiplying $\mathcal{S}^\dagger\mathcal{S}$, we use the usual rules to find the second factor $\mathcal{S}$, and the conjugate rules to find the $\mathcal{S}$ used in the first factor $\mathcal{S}^\dagger$:

$$\mathcal{S}^\dagger\mathcal{S} = [\mathcal{S}(S^\dagger)]^\dagger\mathcal{S}(S)$$

The result of conjugating the S-matrix then will be to complex conjugate twice, and return rules identical to those used for $\mathcal{S}$, except for the differences noted above for real actions. That means that the above proof of the cutting rules goes through unmodified, where we use the same complex rules in the entire diagram, regardless of whether they are associated with Z or $Z^\dagger$. In particular, this means that vertices from the two parts of the graph will differ only by sign (conjugating just the i in e^{iS_I}, not the S_I), and propagators will differ only by their (momentum-space) denominators, not their numerators. This is particularly important for the complex fields of subsection IIIC4, since otherwise even the types of indices carried by the fields would differ.

7. Cross sections

In quantum physics, the only measurables are probabilities, the squares of absolute values of amplitudes. Since we calculate amplitudes in momentum space, probabilities are expressed in terms of scattering of plane waves. They are more naturally normalized as probabilities per unit 4-volume (or D-volume in arbitrary dimension), since plane waves are uniformly distributed throughout space. This can be seen explcitly from the amplitudes: Because of the total momentum conservation δ-function that appears with each connected S-matrix element $\mathcal{S}_{fi}$, we have for the probability P

$$\mathcal{S}_{fi} = i\delta\left(\sum p\right) T_{fi}$$

$$\Rightarrow \quad P = |\mathcal{S}_{fi}|^2 = |T_{fi}|^2 \delta\left(\sum p\right)\delta(0) = |T_{fi}|^2 \delta\left(\sum p\right)\frac{V_D}{(2\pi)^{D/2}}$$

where we have found the coordinate D-volume by the Fourier-transform definition of the δ-function:

$$\delta(0) = \int dx\ 1 = \int \frac{d^D x}{(2\pi)^{D/2}}\ 1 = \frac{V_D}{(2\pi)^{D/2}}$$

A "cross section" is defined as a probability for the scattering of two incoming particles into some number of outgoing particles. The scattering is "elastic" if the two final particles are the same as the two initial particles (they exchange only 4-momentum), "inelastic" otherwise. Generally one particle is in a beam directed at a target (at rest in the lab frame) containing the other "incoming" particle, but in some experiments two beams are directed at each other. In either case the cross section is defined by the rate at which one particle interacts divided by the flux of the other particle, where

$$flux = \frac{rate\ of\ arrival}{area} = (density) \times (relative\ velocity)$$

and thus the "differential cross section" (yet to be integrated/summed over final states) is

$$d\sigma = \frac{P}{V_D} \times \frac{1}{\rho_1 \rho_2 v_{12}}$$

The "total cross section" σ is then given by summation over all types of final states and integration over all their remaining momentum dependence.

The spatial density ρ is the integrand of the spatial integral that defines the inner product: From subsection VB2, for bosonic plane waves ($\psi_p(x) = e^{ip\cdot x}$) we have

$$\langle\psi|\psi\rangle = \epsilon(p^0)\int \frac{d^{D-1}x}{(2\pi)^{D/2}}\ \psi^* \tfrac{1}{2} i \overleftrightarrow{\partial}_t \psi = \int \frac{d^{D-1}x}{(2\pi)^{D/2}}\ \omega$$

$$\Rightarrow \quad \rho = \frac{\omega}{(2\pi)^{D/2}}$$

where $\omega = |p^0|$. (The same result can be obtained for fermions, when their external line factors are appropriately normalized.)

The expression for $d\sigma$ is actually independent of the frame, as long as the 3-momenta of the two particles are parallel. This is the case for the most frequently used reference frames, the center-of-mass frame and the "lab frame" for either particle (where that particle is at rest, as is the lab if that particle is part of a target). Then

$$(\omega_1\omega_2 v_{12})^2 = \omega_1^2\omega_2^2 \left|\frac{\vec{p}_1}{\omega_1} - \frac{\vec{p}_2}{\omega_2}\right|^2 = |\omega_2\vec{p}_1 - \omega_1\vec{p}_2|^2 = -\tfrac{1}{2}(p_{1[a}p_{2b]})^2 = \lambda_{12}^2$$

$$\boxed{\lambda_{12}^2 \equiv \tfrac{1}{4}[s-(m_1+m_2)^2][s-(m_1-m_2)^2]}$$

using again the Mandelstam variables and λ_{ij} introduced in subsection IA4. Finally, we include the "phase space" for the final states to obtain

$$d\sigma = |T_{fi}|^2 \frac{(2\pi)^D \delta^D\left(\sum p\right)}{\lambda_{12}} \prod_f \frac{d^{D-1}p}{(2\pi)^{D/2-1}\omega} \prod_{f:n\ ident} \frac{1}{n!}$$

where the first product is over all final one-particle states, and the second is over each set of n identical final particles. The normalization again follows from the inner-product for plane waves: By Fourier transformation, $d^{D-1}x\ \rho \to d^{D-1}p/(2\pi)^{D-1}\rho$. It also appears in the "cut propagator" Δ_+ used in unitarity, as in the previous subsection:

$$\int \frac{d^D p}{(2\pi)^{D/2}}\ \theta(p^0) 2\pi\delta[\tfrac{1}{2}(p^2+m^2)] = \int \frac{d^{D-1}p}{(2\pi)^{D/2-1}\omega}$$

The simplest and most important case is where two particles scatter to two particles. (This includes elastic scattering.) The "differential cross section" $d\sigma/d\Omega$, where $d\Omega$ is the angular integration element for p_3, is found by integrating $d\sigma$ over $d^{D-1}p_4$ and $d|\vec{p}_3|$. The former integration is trivial, using the δ function for $(D-1)$-momentum conservation. The latter integration is almost as trivial, integrating the remaining δ function for energy conservation:

$$d^{D-1}p_3\ \delta\left(\sum p^0\right) = d\Omega\ d|\vec{p}_3|\ (\vec{p}_3)^{D-2} \left|\frac{\partial \sum p^0}{\partial|\vec{p}_3|}\right|^{-1} \delta(|\vec{p}_3| - |\vec{p}_3|_0)$$

$$\sum p^0 = \omega_1 + \omega_2 - \sqrt{(\vec{p}_3)^2 + m_3^2} - \sqrt{(\vec{p}_1 + \vec{p}_2 - \vec{p}_3)^2 + m_4^2}$$

$$\Rightarrow \quad \frac{\partial \sum p^0}{\partial |\vec{p}_3|} = \frac{-\frac{1}{2}(s - m_3^2 - m_4^2)\omega_3 + m_3^2 \omega_4}{|\vec{p}_3|\omega_3\omega_4}$$

where $|\vec{p}_3|_0$ is $|\vec{p}_3|$ evaluated as a function of the remaining variables at $\sum p^0 = 0$. We then find

$$\frac{d\sigma}{d\Omega} = (2\pi)^2 |T_{fi}|^2 \frac{|\vec{p}_3|^{D-1}}{\lambda_{12}[\frac{1}{2}(s - m_3^2 - m_4^2)\omega_3 - m_3^2\omega_4]}$$

The center-of-mass frame (see subsection IA4) is the simplest for computations. In that frame the differential cross section simplifies to

$$\frac{d\sigma}{d\Omega} = (2\pi)^2 |T_{fi}|^2 \frac{\lambda_{34}^{D-3}}{\lambda_{12} s^{D/2-1}}$$

and in particular in D=4:

$$\frac{d\sigma}{d\Omega} = (2\pi)^2 |T_{fi}|^2 \frac{\lambda_{34}}{\lambda_{12} s}$$

Another convenient form for the differential cross section is $d\sigma/dt$, trading θ for t and integrating out the trivial dependence $\int d\phi = 2\pi$ (for D=4). In the center-of-mass frame we have

$$d\Omega = 2\pi d(\cos\,\theta) = \pi \frac{s}{\lambda_{12}\lambda_{34}} dt$$

Since t is Lorentz invariant, we therefore have in *all* frames (that conform to our earlier requirement for the λ_{12} factor in $d\sigma$)

$$\boxed{\frac{d\sigma}{dt} = \tfrac{1}{2}(2\pi)^3 |T_{fi}|^2 \frac{1}{\lambda_{12}^2}}$$

For example, for the 4-point scalar example considered in subsection VC4, we have

$$\frac{d\sigma}{dt} = \frac{(4\pi)^3 g^4}{s(s - 4m^2)} \left(\frac{1}{s - m^2} + \frac{1}{t - m^2} + \frac{1}{u - m^2} \right)^2$$

Exercise VC7.1

A simpler example than the cross section is the decay rate,

$$\frac{dP}{dt} = \sum_f \frac{P}{\rho V_D}$$

for initial density ρ (where t is now time).

a For the case of decay of a particle of mass M into 2 particles of masses m_1, m_2 in D=4, show that in the rest frame

$$\frac{dP}{dt} = 4\pi \frac{\lambda_{12}}{M^3} |T_{fi}|^2$$

(with a factor of $\frac{1}{2}$ for final identical particles). What happens when $M = m_1 + m_2$?

b For the case of coupling $S_I = \int dx\ g\phi_1\phi_2\phi_3$, evaluate the classical (tree) contribution in terms of g, M, m_1, m_2. Give the dimensional analysis of the result. Consider also the case where the final particles are massless.

Exercise VC7.2

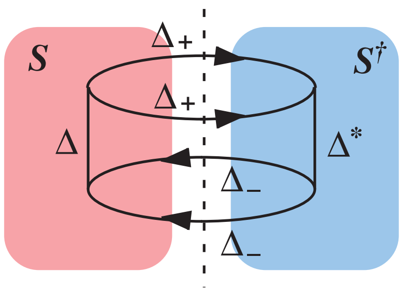
Consider the cross section for elastic scattering of two particles to lowest order (tree graphs), in four dimensions. For the following, consider the nonrelativistic limit (small velocities).

a Show from the definition that classically the "total cross section", as indicated by the name, is just the cross-sectional area with respect to the beam (assuming each "arrival" results in an "interaction").

b A ϕ^4 interaction correpsonds to a δ-function potential in classical mechanics, since it has zero range. In classical mechanics, such "billiard ball" scattering is purely geometrical, depending only on the "size" of the balls. Find the effective "radius" of these classical billiard balls in terms of their mass and coupling.

c Replace the ϕ^4 interaction with a $\phi^2\chi$ interaction, where χ is an intermediate particle with a different mass (nonzero, and also much larger than the ϕ kinetic energies). (We still consider scattering of ϕ particles.) What is the effective radius of the ϕ particles?

d In the limit where the χ mass becomes infinite, but also its coupling, the cubic interaction is effectively replaced with a ϕ^4 interaction (finite when the limit is taken appropriately). What is this coupling constant in terms of the cubic coupling and χ mass?



For some purposes (such as considerations of unitarity and causality, as in the previous subsection) it is useful to draw the Feynman diagrams for the cross section itself (or actually $|T|^2$). In such a diagram we draw one of the diagrams from $\mathcal{S}$

and one from $\mathcal{S}^\dagger$, separating the two by a line (dashed, zig-zag, or shaded on one side, according to your preference), and connecting *all* the external lines (initial and final) on one side to the corresponding ones on the other. The result is a bubble diagram with a "cut": The "cut propagators" are $\Delta_+ = 2\pi\theta(p^0)N\delta[\frac{1}{2}(p^2+m^2)]$ (or Δ_-, depending on how we label the momenta), corresponding to the propagator $\Delta = N/\frac{1}{2}(p^2+m^2)$, if we sum over all polarizations; otherwise N is replaced by a term in the sum $N = \sum \hat{\psi}^\dagger\hat{\psi}$. The momenta of the cut propagators may not be integrated over, depending on whether they represent final states whose momenta are summed over (i.e., not measured; in practice the momenta of initial particles are always measured). The only other difference in the Feynman rules from the S-matrix is that in Wick rotating back $\mathcal{S}$ gets the usual $m^2 \to m^2 - i\epsilon$ while $\mathcal{S}^\dagger$ gets $m^2 \to m^2 + i\epsilon$, and each connected graph in $\mathcal{S}$ gets an $i\delta(\sum p)$ while each in $\mathcal{S}^\dagger$ gets a $-i\delta(\sum p)$. The algebra for the cross section is thus identical to that of a vacuum bubble (although the momentum integration is not, and the cut propagators lack the usual denominators). In particular, instead of summing over just physical polarizations in a cut vector propagator, which corresponds to using a unitary gauge, we can include ghosts in the external states, and use any gauge: This follows from the cutting rules derived for unitarity.

8. Singularities

We know from free theories that any propagator has a pole at the classical value of the square of the mass. This statement can be extended to the interacting theory: The ("Landau") singularities in any Feynman diagram are exactly at classically allowed (on shell) values of the momenta.

The simplest way to see this is to write the propagators in a way reminiscent of the classical theory, where the appearance of the worldline metric in the action results in the (Wick-unrotated) Schwinger parametrization of the propagator,

$$\frac{-i}{\frac{1}{2}(p^2+m^2)} = \int_0^\infty d\tau \; e^{-i\tau(p^2+m^2)/2}$$

For simplicity we consider a scalar field theory with nonderivative self-interactions. The corresponding form of a Feynman diagram, written in momentum space by Fourier transformation, is then

$$\int dx'_i dp_{ij} d\tau_{ij} \; e^{-i\sum_{\langle ij\rangle}[\tau_{ij}(p_{ij}^2+m^2)/2-(x_i-x_j)\cdot p_{ij}]}$$

where i,j label vertices (and external endpoints), $\langle ij\rangle$ labels links (propagators), and dx' indicates integration over just vertices not attached to external lines.

Integration over these x''s produces δ functions for momentum conservation:

$$\sum_j p_{ij} = 0$$

i.e., the sum of all momenta flowing into any vertex vanishes. (Note $p_{ij} = -p_{ji}$.) This constraint can be solved by replacing momenta associated with each propagator with momenta associated with each loop (and keeping momenta associated with external lines). For example, for planar diagrams, we can write

$$p_{ij} = p_{IJ} = k_I - k_J$$

where I, J label loops: p_{IJ} labels the propagator by the two loops on either side, rather than the vertices at the ends (and similarly for $\tau_{ij} = \tau_{IJ}$). Similar remarks apply to nonplanar diagrams, but the parametrization in terms of loop momenta k_I is more complicated because of the way external momenta appear. (In the planar case, the above parametrization can also be used for external momenta, and automatically enforces overall momentum conservation.) The Feynman integral then becomes

$$\int dk'_I d\tau_{IJ}\ e^{-i\sum_{\langle IJ\rangle} \tau_{IJ}[p_{IJ}(k)^2+m^2]/2}$$

where k are loop (k') and external momenta after solving the conservation conditions.

We can now treat the exponent of the Feynman diagram in the same way as a classical mechanics action, and find the corresponding classical equations of motion by varying it. By the stationary phase approximation (or steepest descent, after Wick rotation of τ), the classical solutions give the most important contribution to the integrals (at least for weak coupling). This approximation is related to long-distance (i.e., infrared) behavior (see exercise VB4.3a). Taking account of the fact that the τ's are constrained to be positive (by treating $\tau = 0$ separately or making a temporary change of variables $\tau = \beta^2$ or e^β to an unconstrained variable), we find

$$\tau_{IJ}(p_{IJ}^2 + m^2) = 0, \qquad \sum_J \tau_{IJ} p_{IJ} = 0$$

or, in terms of the original variables,

$$\tau_{ij}(p_{ij}^2 + m^2) = 0, \qquad p_{ij} = \frac{x_i - x_j}{\tau_{ij}}, \qquad \sum_j p_{ij} = 0$$

These are known as the "Landau equations". Their correspondence with classical configurations of particles follows from treating τ as the proper time, as seen from $p = \Delta x/\Delta\tau$. (Actually, it is the generalization discussed in section IIIB of proper time to include the massless case; in the massive case $s = m\tau$.) The equation $\tau(p^2+m^2) = 0$ says that either the particle for the line is on-shell or there is no such line (it has vanishing proper length), while the equation $\sum_J \tau p = 0$ says that the sum of τp around a loop vanishes, another statement that $p\Delta\tau = \Delta x$.

Exercise VC8.1

Consider the "one-loop propagator correction": the graph where a single particle of mass M splits into two of masses m_1 and m_2, which rejoin. What is the physical condition relating these 3 masses (with all particles on shell to satisfy the above)? What happens when $M = m_1 + m_2$? What if $m_2 = 0$?

Note that these physical singularities are all in physical Minkowski space: In Euclidean space, p^2+m^2 is positive definite, so it never vanishes (except for constant, massless fields $p = m = 0$). Furthermore, in Euclidean space, one can always rotate any momentum to any direction, whereas in Minkowski space one can never Lorentz transform to or through either the forward or backward lightcone. Thus, calculating in Euclidean space makes it clear that S-matrix elements for positive-energy states are given by the same expressions as those for negative-energy states: To compare two amplitudes that are the same except for some final particles being replaced with initial antiparticles, or vice versa, we just change the sign of the energy. This is called "crossing symmetry". In the case where all particles are reversed, it is CPT invariance ("CPT theorem").

9. Group theory

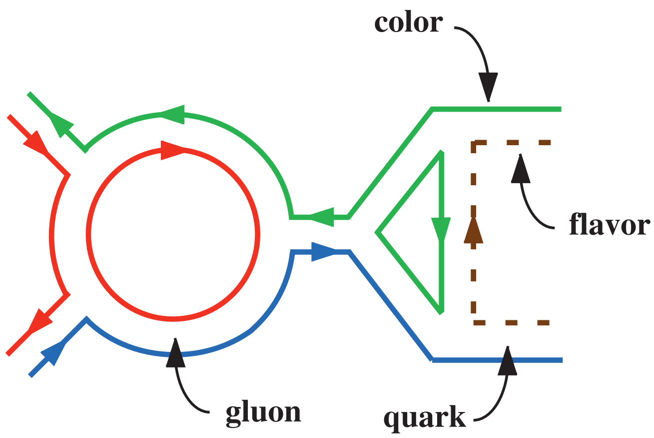
Although the manipulation of spin indices in Feynman diagrams is closely tied to momentum dependence, the group theoretic structure is completely independent, and can be handled separately. Therefore, it is sufficient to consider the simple example of scalars with a global symmetry. (The more physical case of chromodynamics will be the same with respect to group theory, but will differ in dependence on momentum and spin.) The simplest case is the U(N) family of groups. We choose an action of the form

$$L = \frac{1}{g^2} tr[\tfrac{1}{4}(\partial\phi)^2 + V(\phi)]$$

where ϕ is a hermitian N×N matrix. This action is invariant under the global U(N) symmetry

$$\phi' = U\phi U^{-1}$$

A simplification we have chosen is that the interaction has only a single trace; this is the case analogous to pure Yang-Mills theory. (We also included the coupling constant as in Yang-Mills.)



When we draw a Feynman diagram for this field theory, instead of a single line for each propagator, we draw a double (parallel) line, each line corresponding to one of the two indices on the matrix field. Because of the trace in the vertex, the propagator lines connect up there in such a way that effectively we have continuous lines that travel on through the vertices, although the two lines paired in a propagator go their separate ways at the vertex. These lines never split or join, and begin or end only on external fields. We can also draw arrows on the lines, pointing in the same direction everywhere along a single line, but pointing in opposite directions on the two lines in any propagator pair: This keeps track of the fact that ϕ appears in the trace always multipled as $\phi\phi$ and never as $\phi\phi^T$. A physical picture we can associate with this is to think of the scalar as a bound-state of a quark-antiquark pair, with one line associated with the quark and another with the antiquark; the arrows are then oriented in the direction of time of the quark (which is the opposite of the direction of time for the antiquark). The quark is thus in the defining representation of U(N) (and the antiquark in the complex conjugate representation). Group theory factors are trivial to follow in these diagrams: The same color quark continues along the extent of a "quark line"; thus, there is a Kronecker δ for the two indices appearing at the ends of the quark line (at external fields); each quark is conserved.

Exercise VC9.1

Using the quark-line notation, where the lines now represent flavor, draw all 4-point tree graphs, with 3-point vertices, representing scattering of $K^+K^- \to \pi^+\pi^-$ (see subsection IC4) via exchange of other (pseudo)scalar mesons in that U(6) multiplet. What are the intermediate states (names of mesons) in each channel? Note that the "flavor flow" of both diagrams can be represented by a single diagram, by separating all pairs of intermediate lines to leave a square gap in the middle: This "duality diagram" represents the mesons as strings; the gap between quarks represents the "worldsheet".

Exercise VC9.2

Consider quark-line notation for doing group theory in general: (Calculate using only graphs, with numerical factors — no explicit indices or Kronecker δ's, except for translating definitions.)

a Write the structure constants for U(N) as the difference of two diagrams, by considering a vertex of the form $tr(A[B,C])$.

b Find pictorially the resulting expression for the Cartan metric (see subsection IB2): Show that it is the identity times 0 for the U(1) subgroup and 2N for the SU(N) subgroup (as found previously in subsection IIIC1).

c Also use these diagrams to prove the Jacobi identity (i.e., find the resulting 6 diagrams and show they cancel pairwise).

d Derive diagrammatically the value of d_{ijk} of exercise IB5.3b for the defining representation.

Exercise VC9.3

Consider the group theory factors for the above scalar theory, with only a cubic interaction. Draw all the 1PI 1-loop diagrams with 4 external double-lines, and rewrite the corresponding factors in terms of traces and products of 4 fields, including factors of N. (Be careful to include all permutations of connecting propagators to vertices.)

In general, we find that connected Feynman diagrams may include diagrams that are disconnected with respect to the above group-theory diagrams, where we consider the two group-theory lines on an external line to be connected. Such diagrams correspond to multiple traces: There is a factor of the form $tr(G_i G_j ... G_k)$ corresponding to each connected group-theory graph, where $i, j, ..., k$ are the group-theory indices of the external lines (actually double indices in the previous notation). However, all connected trees are group-theory connected. Furthermore, they are all "planar": Any connected U(N) tree can be drawn with none of the quark lines crossing, and all the external lines on the outside of the diagram, if the external lines are "color-ordered" appropriately. Of course, one must sum over permutations of external lines in the T-matrix because of Bose symmetry. However, in calculating $W[\phi]$ one need consider only one such planar graph, with the external lines color-ordered; the Bose symmetry of the fields in W automatically incorporates the permutations. (Similar remarks apply to loop and nonplanar graphs.)

As an example, consider the U(N) generalization of the 4-point tree example of subsection VC4. The Lagrangian is now (scaling the coupling back into just the vertex)

$$L = tr\{\tfrac{1}{4}[(\partial\phi)^2 + m^2\phi^2] + \tfrac{1}{3}g\phi^3\}$$

where the interaction term now has a combinatoric factor of $\frac{1}{3}$ instead of $\frac{1}{6}$ because it is symmetric only under cyclic permutations. The result for W is now modified to

$$W = -g^2 tr \int dx\ \tfrac{1}{2}\phi^2 \frac{1}{\frac{1}{2}(m^2 - \Box)}\phi^2$$

This analysis for U(N) can be generalized to SU(N) by including extra diagrams with lines inside propagators short-circuited, representing subtraction of traces. It can also be generalized to SO(N) and USp(2N): In those cases, antisymmetry or symmetry of the matrices means the lines no longer have arrows, and we include diagrams where the lines inside the propagators have been “twisted”, with signs appropriate to symmetrization or antisymmetrization. Generalization for all the above groups to include defining representations is also straightforward: Such fields, like the true quarks in QCD, carry only a single group-theory line. For example, sticking with our simpler scalar model, we can generalize to a scalar theory with the same group theory as chromodynamics, with “free” quark fields, appearing as scalar fields in the defining representation of the group:

$$L \to L + tr[\tfrac{1}{2}(\partial\psi)^\dagger \cdot (\partial\psi) + \psi^\dagger f(\phi)\psi], \qquad \psi' = U\psi U_f^{-1}$$

where ψ is an N×M matrix with M flavors, and U_f is the flavor symmetry. (This color+flavor symmetry was treated in subsections IC4 and IVA4.) This field has a propagator with a single color line (with an arrow); however, we can also use another double-line notation, where ψ propagators carry one line for color and another for flavor. This method can be generalized to arbitrary representations obtained by direct products of defining representations, (anti)symmetrizations, and subtractions of traces, by giving the propagators the corresponding number of lines (though usually two lines are sufficient for the interesting cases).

REFERENCES

1 P.A.M. Dirac, *Proc. Roy. Soc.* **A114** (1927) 243;
W. Heisenberg and W. Pauli, *Z. Phys.* **56** (1929) 1, **59** (1930) 168;
E. Fermi, *Rev. Mod. Phys.* **4** (1932) 87:
birth of (interacting) quantum field theory (QED).

2 S. Tomonaga, *Prog. Theor. Phys.* **1** (1946) 27;
J. Schwinger, *Phys. Rev.* **74** (1948) 1439, *loc. cit.* (VB, ref. 2, first ref.), **76** (1949) 790:
gave relativistic procedure for treating quantum field theory.

3 Feynman, *loc. cit.* (VB, ref. 4):
same as above, with diagrams.
4 F.J. Dyson, *Phys. Rev.* **75** (1949) 486, 1736:
related Tomonaga-Schwinger and Feynman approaches; gave procedure to all loops.
5 J. Schwinger, *Proc. Natl. Acad. Sci. USA* **37** (1951) 452, 455:
generating functionals for field theory.
6 Feynman, *loc. cit.* (VA):
path-integrals for field theory.
7 J. Schwinger, *Quantum electrodynamics* (Dover, 1958):
reprints of most of the important quantum field theory papers of the 40's and 50's.
8 S.S. Schweber, *QED and the men who made it: Dyson, Feynman, Schwinger, and Tomonaga* (Princeton University, 1994).
9 J. Goldstone, A. Salam, and S. Weinberg, *Phys. Rev.* **127** (1962) 965;
G. Jona-Lasinio, *Nuo. Cim.* **34** (1964) 1790:
effective action.
10 Y. Nambu, *Phys. Lett.* **26B** (1968) 626:
tree graphs are classical field theory.
11 D.G. Boulware and L.S. Brown, *Phys. Rev.* **172** (1968) 1628:
expression for classical field in terms of tree graphs.
12 S. Coleman, *Secret symmetry: An introduction to spontaneous symmetry breakdown and gauge fields*, in *Laws of hadronic matter*, proc. 1973 International School of Subnuclear Physics, Erice, July 8-26, 1973, ed. A. Zichichi (Academic, 1975) part A, p. 139:
relation of Legendre transform to 1PI graphs via path integrals.
13 Dyson, *loc. cit.* (ref. 4);
Schwinger, *loc. cit.* (ref. 5).
14 W. Pauli, *Phys. Rev.* **58** (1940) 716:
spin-statistics theorem.
15 R.E. Cutkosky, *J. Math. Phys.* **1** (1960) 429:
cutting rules for unitarity.
16 T. Veltman, *Physica* **29** (1963) 186:
cutting rules for causality.
17 R.P. Feynman, *Acta Phys. Polon.* **24** (1963) 697:
tree theorem.
18 L.D. Landau, *Nucl. Phys.* **13** (1959) 181.
19 S. Coleman and R.E. Norton, *Nuo. Cim.* **38** (1965) 438:
classical interpretation of Landau singularities.
20 W. Siegel, hep-th/9601002, *Int. J. Mod. Phys. A* **13** (1998) 381:
Landau rules from classical mechanics action.
21 R.J. Eden, P.V. Landshoff, D.I. Olive, and J.C. Polkinghorne, *The analytic S-matrix* (Cambridge University, 1966):
singularities of Feynman diagrams.
22 M. Gell-Mann and M.L. Goldberger, proc. *Fourth International Conference on High Energy Physics* (University of Rochester, 1954);
J.M. Jauch and F. Rohrlich, *The theory of photons and electrons*, 1st ed. (Springer-Verlag, 1955) p. 161:
crossing symmetry.
23 H. Harari, *Phys. Rev. Lett.* **22** (1969) 562;
J. Rosner, *Phys. Rev. Lett.* **22** (1969) 689;

T. Matsuoka, K. Ninomiya, and S. Sawada, *Prog. Theor. Phys.* **42** (1969) 56:
quark lines for group theory factors.

24 J.E. Paton and Chan H.-M., *Nucl. Phys.* **B10** (1969) 516:
traces for group theory in planar graphs.

25 G. 't Hooft, *Nucl. Phys.* **B72** (1974) 461:
double-line notation in field theory.

VI. QUANTUM GAUGE THEORY

We now consider special features of spin and gauge invariance, and introduce some special methods for dealing with them. In quantum theory, gauge fixing is necessary for functional integration: Gauge invariance says that the action is independent of some variable; integration over that variable would thus give infinity when evaluating amplitudes for gauge-invariant states. Eliminating that variable from the action (a "unitary gauge") solves the problem, but not always in the most convenient way. (If it were, we wouldn't have introduced such a redundant description in the first place.) Note that such infinities already appear for global symmetries: For example, the functional integral with wave function 1 (vacuum-to-vacuum amplitude) is infinite by translation invariance. This infinity is easier to understand for a nontrivial amplitude in momentum space, as a factor of a momentum-conservation δ-function (which is either ∞ or 0, but necessarily ∞ for the vacuum amplitude, which has vanishing momentum because the vacuum is translationally invariant).

A. BECCHI-ROUET-STORA-TYUTIN

We have seen the relationship of gauge invariances to constraints in subsection IIIA5. In this section we consider the quantization of constrained systems, and its application to gauge theories. The Becchi-Rouet-Stora-Tyutin (BRST) method is not only the most powerful, but also the easiest way to gauge fix: It replaces the gauge symmetry with an unphysical, fermionic, global symmetry that acts only on unphysical degrees of freedom.

1. Hamiltonian

Physical observables commute with the constraints. Thus, time development is described by the gauge-invariant Hamiltonian H_{gi}, or we can set the gauge fields λ^i equal to some arbitrary functions f^i as a "gauge choice":

$$\lambda^i = f^i \quad \Rightarrow \quad H = H_{gi} + f^i G_i$$

In quantum mechanics, physical states should be annihilated by the constraints. However, more generally we can require only that these states satisfy the constraints through expectation values:

$$\langle \psi | G_i | \chi \rangle = 0$$

This condition is satisfied by dividing up the constraints into:

(1) a subalgebra G_0 that annihilates all physical states,

(2) complex "lowering operators" G_- that also annihilate these states, and are a representation of the subgroup generated by G_0, and

(3) their hermitian conjugate "raising operators" $G_+ = G_-{}^\dagger$.

(We treat $+$, $-$, and 0 here as multivalued indices.) Thus,

$$(G_0 - const.)|\psi\rangle = G_-|\psi\rangle = \langle\psi|(G_0 - const.) = \langle\psi|G_+ = 0$$

where we have allowed for "normal-ordering" constants to be included as eigenvalues of (an Abelian subeset of) the G_0 constraints. *In the Abelian case,* it then follows that, although G_+ do not annihilate these states, they generate gauge invariances:

$$\delta|\psi\rangle = G_+|\zeta^+\rangle, \qquad G_0|\zeta^+\rangle = G_-|\zeta^+\rangle = 0$$

preserves the inner product of such states as well as the constraints on $|\psi\rangle$.

Unfortunately, things get more complicated in the nonabelian case. For example, the gauge invariance and constraints above are no longer compatible in general:

$$0 = G_-\delta|\psi\rangle = G_-G_+|\zeta^+\rangle = [G_-, G_+]|\zeta^+\rangle = -if_{-+}{}^+G_+|\zeta^+\rangle \neq 0$$

However, one example that doesn't have this problem is the simple case where there are only 3 constraints, forming an SU(2) algebra (so $f_{-+}{}^+ = 0$): If we choose G_- to be the lowering operator and G_+ to be the raising operator, then the constant appearing for G_0 is simply the lowest eigenvalue in some irreducible representation in the Hilbert space $|\psi\rangle$, and the constraints pick out the corresponding state (or states, if there is more than one representation with that "spin").

A convenient way to deal with this problem is to replace the nonabelian algebra G_i with a single operator, which is therefore Abelian. We define a BRST operator Q that imposes all constraints G_i by adding a classical anticommuting "ghost" variable c^i, and its canonical conjugate b_i,

$$\{b_i, c^j\} = \delta_i^j$$

for each constraint:

$$Q = c^i G_i - i\tfrac{1}{2}c^i c^j f_{ji}{}^k b_k$$

The second term has been added to insure the Poisson bracket or commutator

$$\{Q, Q\} = 0$$

so that its crossterm cancels the square of the first term, while its own square vanishes by the Jacobi identity $f_{[ij}{}^l f_{k]l}{}^m = 0$. Quantum mechanically, the BRST operator is nilpotent:

$$\textit{quantum mechanically} \qquad \{Q, Q\} \equiv 2Q^2 = 0$$

We can also describe the ghost dependence by the "ghost number"

$$J = c^i b_i \quad \Rightarrow \quad [J, Q] = Q$$

(Quantum mechanically, we need to normal order these expressions for Q and J.) Each anticommuting ghost and its conjugate will serve to "cancel" each commuting constraint and its conjugate gauge degree of freedom. (Similarly, bosonic ghosts are introduced for fermionic constraints, so each term in Q is fermionic.)

The BRST operator provides a convenient method to treat more general gauges than $\lambda = f$, such as ones where the gauge fields become dynamical, which will prove useful particularly in relativistic theories. Now the original physical observables A will satisfy

$$[G_i, A] = [b_i, A] = [c^i, A] = 0 \quad \Rightarrow \quad [Q, A] = [J, A] = 0$$

and similarly the physical quantum mechanical states $|\psi\rangle$ will satisfy

$$(G_0, G_-; b_0, b_-; c^+)|\psi\rangle = 0 \quad \Rightarrow \quad Q|\psi\rangle = J|\psi\rangle = 0$$

where we have used the fact the only nonvanishing structure constants are $f_{00}{}^0$, $f_{0+}{}^+$, $f_{0-}{}^-$, $f_{++}{}^+$, $f_{--}{}^-$, and $f_{+-}{}^k$. (In the quantum case there are also some subtleties due to normal ordering.) We also have the gauge invariances

$$\delta A = \{Q, \Lambda\}, \quad \delta|\psi\rangle = Q|\lambda\rangle$$

for arbitrary operators Λ and (unrelated) states $|\lambda\rangle$, since the Q-terms won't contribute when evaluating matrix elements with states annihilated by Q:

$$(\langle\psi_1| + \langle\lambda_1|Q)(A + \{Q, \Lambda\})(|\psi_2\rangle + Q|\lambda_2\rangle) = \langle\psi_1|A|\psi_2\rangle$$

The gauge invariances are consistent with the constraints because of the nilpotence of the BRST operator. (So $Q(|\psi\rangle + Q|\lambda\rangle)$ still vanishes, etc.) States satisfying

$$Q|\psi\rangle = 0, \quad \delta|\psi\rangle = Q|\lambda\rangle$$

(i.e., we identify states that differ by Q on something) are said to be in the "cohomology" of Q ("BRST cohomology"), and operators satisfying

$$[Q, A] = 0, \quad \delta A = \{Q, \Lambda\}$$

are said to be in its "operator cohomology". (The latter cohomology also has a classical analog.)

Exercise VIA1.1

Assume that each physical state can be represented as a physical observable (Hermitian operator) acting on a ground state, which is itself physical:

$$|\psi\rangle = A|0\rangle, \qquad A = A^\dagger, \qquad Q|0\rangle = 0$$

Show how this relates the gauge parameters and cohomologies of $|\psi\rangle$ and Q.

The BRST operator incorporates the ghosts that are necessary to generalize treatment of the constraints to the nonabelian case: For example, to reproduce the gauge transformations of the Abelian case, we choose

$$|\lambda\rangle = b_+|\zeta^+\rangle, \qquad Q|\zeta^+\rangle = 0 \quad \Rightarrow \quad \delta|\psi\rangle = Q|\lambda\rangle = \hat{G}_+|\zeta^+\rangle$$

where

$$\hat{G}_i = \{Q, b_i\} = G_i - ic^j f_{ji}{}^k b_k$$

are the "gauge-fixed" constraints, which include an extra term to transform the ghosts as the adjoint representation. They reduce to just G_i in the Abelian case, but add ghost terms to the gauge transformation law otherwise.

In particular, the Hamiltonian is a physical operator describing the energy and the time development, so we can write

$$H = H_{gi} + \{Q, \Lambda\}, \quad [Q, H_{gi}] = 0 \quad \Rightarrow \quad \mathcal{T}\left(e^{-i\int dt\ H}\right) = \mathcal{T}\left(e^{-i\int dt\ H_{gi}}\right) + \{Q, \kappa\}$$

for some κ. This includes gauge fixing for the gauge $\lambda^i = f^i$ discussed above, using:

$$\Lambda = f^i b_i \quad \Rightarrow \quad H = H_{gi} + f^i\hat{G}_i, \quad L = -(\dot{q}^m p_m - i\dot{c}^i b_i) + H$$

The ghost terms in $\hat{G}_i$ only affect the time development of unphysical states in this gauge.

For example, when calculating S-matrix elements, the result is independent of the gauge choice Λ, as long as both the gauge-invariant Hamiltonian H_{gi} and the states are BRST invariant. (H_{gi} commutes with Q, the initial and final states are annihiliated by it.) It is also independent of the gauge choice $|\lambda\rangle$ for $|\psi\rangle \to |\psi\rangle + Q|\lambda\rangle$. (Such a "residual gauge invariance" persists even though the asymptotic states satisfy the free field equations.)

In the cases of interest in relativistic physics, the constraints always consist of a linear term depending only on the canonical momenta p (conjugate to the fundamental variables q), at least after some redefinitions, plus higher-order terms, which can be treated perturbatively. Therefore, as the simplest nontrivial example, we consider a model with a single variable q, with

$$H_{gi} = 0, \quad G = p \quad \Rightarrow \quad Q = cp$$

If we assume boundary conditions on the wave functions such that they can be Taylor expanded in q (they can always be expanded in c), we can write

$$\psi(q) = \langle q|\psi\rangle = \sum_{n=0}^{\infty} (\alpha_n + c\beta_n)\tfrac{1}{n!}q^n$$

and similarly for $\lambda(q)$. We then examine $\delta|\psi\rangle = Q|\lambda\rangle$, comparing terms with the same power of c and q on both sides of the equation, to find $\delta\alpha_n$ and $\delta\beta_n$. We then see that we can easily gauge $\beta_n = 0$ for all n by choosing certain coefficients in $|\lambda\rangle$ to be $-\beta_n$ (so $\beta_n' = \beta_n + \delta\beta_n = 0$). Looking at $Q|\psi\rangle = 0$, we then find that $\alpha_n = 0$ for all n except $n = 0$, so only the constant piece of ψ survives. In other words, the cohomology is given by

$$Q|\psi\rangle = 0, \quad \delta|\psi\rangle = Q|\lambda\rangle \quad \Rightarrow \quad p|\psi\rangle = b|\psi\rangle = 0$$

So, solving for the cohomology of $Q = cp$ is the same as solving the constraint $p\psi = 0$ without ghosts.

Exercise VIA1.2

Find the cohomology of the BRST operator

$$Q = ca^\dagger + c^\dagger a$$

for creation and annihilator operators satisfying

$$[a, d^\dagger] = [d, a^\dagger] = \{c, b^\dagger\} = \{b, c^\dagger\} = 1$$

(the other commutators vanishing), by expanding in creation operators $a^\dagger$, $b^\dagger$, $c^\dagger$, $d^\dagger$ about a vacuum state destroyed by the annihilation operators a, b, c, d. (This is the common alternative to the boundary conditions used for $Q = cp$ above.) Compare the method of constraints used in the Abelian case.

2. Lagrangian

To obtain more interesting gauges we need some extra bosonic variables, such as the gauge fields λ^i that we lost along the way, and their canonical conjugates ("Nakanishi-Lautrup fields") B_i,

$$[B_i, \lambda^j] = -i\delta_i^j$$

as well as their corresponding ghosts ("antighosts") $\tilde{c}_i$. We can do this in a trivial way by including constraints that set B to zero:

$$Q = c^i G_i - i\tfrac{1}{2}c^i c^j f_{ji}{}^k b_k + \tilde{b}^i B_i, \qquad J = c^i b_i - \tilde{c}_i \tilde{b}^i$$

where $\tilde{c}_i$ is conjugate to $\tilde{b}^i$,

$$\{\tilde{c}_i, \tilde{b}^j\} = \delta_i^j$$

As a simple example, consider

$$\Lambda = \lambda^i b_i \quad \Rightarrow \quad \{Q, \Lambda\} = \lambda^i \hat{G}_i - i\tilde{b}^i b_i$$

The action now includes the gauge fields and all the ghosts as dynamical variables:

$$L = -(\dot{q}^m p_m + \dot{\lambda}^i B_i - i\dot{c}^i b_i - i\dot{\tilde{b}}^i \tilde{c}_i) + H_{gi} + \{Q, \Lambda\}$$

For this gauge we can eliminate b and $\tilde{b}$ by their equations of motion; assuming G_i is only linear in p, we then can eliminate p to return completely to a Lagrangian formalism:

$$L = L_{gi}(q, \lambda) - \dot{\lambda}^i B_i - i(\nabla_t c^i)\dot{\tilde{c}}_i$$

where L_{gi} represents the original gauge-invariant action (which depended on both q and λ, including time derivatives), and ∇_t is the covariant (time) derivative:

$$\nabla_t c^i = \dot{c}^i + c^j \lambda^k f_{kj}{}^i$$

The gauge condition (from varying B) is now $\dot{\lambda} = 0$, generalizing the non-derivative gauges found without the antighosts and Nakanishi-Lautrup fields. Correspondingly, the ghost term is now second order in derivatives.

Exercise VIA2.1

Consider the general gauge choice

$$\Lambda = \lambda^i b_i + [F^i(q, p) + E^i(B)]\tilde{c}_i$$

$$\Rightarrow \quad \{Q, \Lambda\} = \lambda^i \hat{G}_i - i\tilde{b}^i b_i + (F^i + E^i)B_i + c^i[G_i, F^j]\tilde{c}_j$$

where F^i are some arbitrary functions of the original variables, and E^i are functions that effectively average over the types of gauges produced by F^i. Find the gauge-fixed Hamiltonian and Lagrangian. In the case where E is linear in B, eliminate B, b, and $\tilde{b}$ from the Lagrangian by their algebraic equations of motion.

Now that we understand the principles, all these manipulations can be performed directly in the Lagrangian formalism. This will have the advantage that in field theory the Lagrangian is manifestly Lorentz covariant, while the Hamiltonian (or the Lagrangian in the Hamiltonian form $-\dot{q}p+H$) is not, because of the way it singles out time derivatives (and not spatial ones). (Consider, e.g., electromagnetism.) Similarly, the gauge $G_i = 0$ is usually not Lorentz covariant. We can work with just the original variables q, λ plus the new variables $B, c, \tilde{c}$, and define Q by the transformation it induces (as derived from the Hamiltonian formalism):

$$Qq^m = c^i \delta_i q^m, \quad Q\lambda^i = -i(\dot{c}^i + c^j \lambda^k f_{kj}{}^i), \quad Qc^i = -i\tfrac{1}{2}c^j c^k f_{kj}{}^i, \quad Q\tilde{c}_i = B_i, \quad QB_i = 0$$

where δ_i is the gauge transformation induced by G_i ($[G_i, \ ]$ in the Hamiltonian formalism). Note that the BRST transformations of the original variables are exactly the same as the gauge transformations, with the gauge parameters replaced with the corresponding ghosts. We can also consider the Nakanishi-Lautrup fields B as original variables, with the fact that they don't occur explicitly implying they have constraints $B = 0$. Alternatively, we can treat the antighosts $\tilde{c}$ as pure gauge degrees of freedom, with their own nonderivative gauge transformation $\delta\tilde{c} = \tilde{\lambda}$ that allows them to be completely gauged away.

The Lagrangian can be gauge-fixed directly as

$$L = L_{gi} + Q\Lambda_L$$

where in the case just considered

$$\Lambda_L = -\dot{\lambda}^i \tilde{c}_i$$

gives the same L as above for the $\dot{\lambda} = 0$ gauge. In the simpler case described earlier (the gauge $\lambda =$ constant)

$$\Lambda_L = (\lambda^i - f^i)\tilde{c}_i$$

This gives the result, for the simplest choice $f = 0$,

$$L = L_{gi}(q, \lambda) + \lambda^i B_i - i(\dot{c}^i + c^j \lambda^k f_{kj}{}^i)\tilde{c}_i \quad \to \quad L_{gi}(q, 0) - i\dot{c}^i \tilde{c}_i$$

after eliminating the Lagrange multipliers B and λ by their algebraic equations of motion. Note that $\Lambda_L = \lambda^i \tilde{c}_i$ corresponds to the Hamiltonian formalism's $\Lambda = 0$. Thus, in the Hamiltonian formalism we never quantize with $H = H_{gi} + \lambda G$, but can use just H_{gi} and $\Lambda = 0$, which is equivalent to using $H_{gi} + \{Q, \Lambda\}$ for any Λ, while in the Lagrangian formalism we can never quantize with just $L_{gi}(q, \lambda)$, or even $L_{gi}(q, 0)$, and Λ_L is never zero, but must be chosen so as to break the gauge invariance. However, the extra term for $L_{gi}(q, 0)$ is just the $\dot{\tilde{c}}b$ term found from converting H_{gi} to the Lagrangian formalism.

Exercise VIA2.2

Repeat exercise VIA2.1 directly in the Lagrangian formalism. (Find Λ_L, etc.)

All our results for quantization apply equally well in the path-integral formalism, which can be applied to either the Hamiltonian or Lagrangian. (Of course, for field theory we will be interested in applying BRST to path integrals for Lagrangians.) We then evaluate matrix elements as

$$\mathcal{A} = \int D\phi\ \Psi[\phi] e^{-iS[\phi]}; \qquad S = S_{gi} + Q\Lambda, \quad \Psi = \Psi_{gi} + Q\Lambda_\Psi; \qquad QS_{gi} = Q\Psi_{gi} = 0$$

S_{gi} and Ψ_{gi} depend on just the physical fields (no ghosts), so they are gauge invariant as well as BRST invariant. For S-matrices, since Ψ is an asymptotic state, the BRST operator used for its constraint and gauge invariance can be reduced to its free part: Q then acts on only the gauge fields. The statement of gauge invariance of Ψ_{gi} is then equivalent to the requirement that gauge fields appear in it only as their Abelian field strengths. For example, the usual gauge vector A describing electromagnetism appears in single-particle factors in the wave functional ($\Psi[\phi] = \prod \Psi_1[\phi]$ as in subsection VC1) only as:

$$\Psi_1[A] = \langle A^a | \psi_a \rangle, \qquad \delta A_a = -\partial_a \lambda, \qquad \Box \psi_a = \Box \lambda = 0$$

$$\Rightarrow \quad 0 = \delta \Psi_1 = \langle \delta A^a | \psi_a \rangle = \langle \lambda | \partial^a \psi_a \rangle$$

using $\partial_0 \langle \lambda | \psi \rangle = 0$ (where the relativistic inner product $\langle \ | \ \rangle$ was defined in subsection VB2). The transversality of ψ_a is equivalent to coupling to the Abelian field strength, since

$$\partial^a \psi_a = 0 \quad \Rightarrow \quad \psi_a = \partial^b \psi_{ab} \quad \Rightarrow \quad \langle A^a | \psi_a \rangle = \tfrac{1}{2} \langle F^{ab} | \psi_{ab} \rangle$$

in terms of an antisymmetric-tensor external-line factor ψ_{ab}.

3. Particles

We have seen that the relativistic particle (with or without spin) is a simple example of a contrained system. For the simplest case, spin 0, the BRST operator follows simply from the single constraint:

$$Q = c\tfrac{1}{2}(\Box - m^2)$$

Unlike the nonrelativistic case, the relativistic "Hamiltonian" is identified with this constraint. Since we know constraints are treated by the BRST operator, we can consider writing the field theory action in terms of it:

$$S_0 = -\int dx\ dc\ \tfrac{1}{2}\Phi Q \Phi$$

Using the explicit c dependence of the field $\Phi = \phi - ic\psi$, we find the usual scalar kinetic term. ϕ is thus the usual field, while ψ is an "antifield", which has opposite statistics to ϕ (fermion instead of boson). We'll see in chapter XII that Q can be constructed straightforwardly for arbitrary spin, and has a simple expression in term of generalized spin operators. (As in nonrelativistic theories, spin is easier to treat directly in quantum mechanics rather than by first-quantization of a classical system.) The kinetic term then generally can be written as a slight modification of the above. Then the antifields will be found to play a nontrivial function, rather than just automatically dropping out as in this case.

From the constraints and their algebra for spin 1/2 (see also exercise IIIB1.2) we find the BRST and ghost-number operators:

$$Q = c\tfrac{1}{2}(\Box - m^2) - \xi(\gamma\cdot\partial - i\tfrac{m}{\sqrt{2}}) - \xi^2 b + \tilde{\xi}\mu, \qquad J = cb + \xi\zeta + \tilde{\xi}\tilde{\zeta}$$

where ξ and its conjugate ζ are bosonic ghosts, and we have added a nonminimal term with boson $\tilde{\xi}$ (conjugate $\tilde{\zeta}$) and fermion μ (conjugate κ) to allow gauges general enough for first-quantization:

$$[\zeta, \xi] = [\tilde{\zeta}, \tilde{\xi}] = \{\kappa, \mu\} = 1$$

For convenience, we also have chosen ξ (and ζ) to *anti*commute with γ,

$$\{\xi, \gamma^a\} = \{\zeta, \gamma^a\} = 0$$

to avoid having to replace $-i\frac{m}{\sqrt{2}}$ with $\gamma_{-1} im$; this has the natural interpretation of treating ξ and ζ as bosonic (ghost) components of the γ matrices (see subsections XIIA4-5,B5).

Exercise VIA3.1

Find Q and J for spin 1 as constructed from the direct product of 2 spin 1/2's (see exercise IIB4.1d).

Note that $[Q,\xi]=0$, but $\{Q,A\}\neq\xi$ for any A, so ξ is in the operator cohomology of Q. Normally, this would imply infinite copies of the physical states in the cohomology, since applying a "translation" with the ghost variable ξ gives a new state in the cohomology from any given one. The nonminimal variables allow us to avoid this problem by combining ξ with $\tilde{\xi}$ to produce harmonic oscillator creation and annihilation operators:

$$\xi=\tfrac{1}{\sqrt{2}}(a+a^\dagger),\quad \tilde{\xi}=\tfrac{1}{\sqrt{2}}i(a^\dagger-a),\quad \zeta=\tfrac{1}{\sqrt{2}}(\tilde{a}-\tilde{a}^\dagger),\quad \tilde{\zeta}=-\tfrac{1}{\sqrt{2}}i(\tilde{a}^\dagger+\tilde{a})$$

$$[a,\tilde{a}^\dagger]=[\tilde{a},a^\dagger]=1,\quad rest=0$$

This allows us to define a ground state

$$a|0\rangle=\tilde{a}|0\rangle=0$$

which breaks the translation symmetry of ξ. In chapter XII we'll show in a more general framework how the $\Phi Q\Phi$ type of action then reproduces the Dirac action.

4. Fields

As described in subsection VIA2, we can perform gauge fixing through BRST, including the introduction of ghosts, directly on the Lagrangian at the classical level. Also, the BRST transformations on the physical fields are just the gauge transformations with the gauge parameters replaced by ghosts, and the BRST transformation on the ghosts is quadratic in ghosts times the structure constants, while on the antighosts it gives the Nakanishi-Lautrup fields, and it annihilates the NL fields. In the case of Yang-Mills we then have

$$\boxed{QA_a=-[\nabla_a,C],\quad QC=iC^2,\quad Q\tilde{C}=-iB,\quad QB=0}$$

while for matter transforming as $\delta\phi=i\lambda\phi$ we have

$$\boxed{Q\phi=iC\phi}$$

where we have used matrix notation for the group algebra, as usual. There are two minor differences from the transformation rules we used in our general discussion previously: (1) We have included an extra "i" in our definition of the relativistic Q, for a convenience that will become apparent only when we relate relativistic first- and second-quantization (see chapter XII). (2) There is a relative sign difference for QC because now Q is second-quantized while G_i is still first-quantized (i.e., matrices). More explicitly, we have, e.g.,

$$C^2 = \tfrac{1}{2}\{C, C\} \quad \Rightarrow \quad QC^i = -\tfrac{1}{2}C^j C^k f_{kj}{}^i; \qquad Q\phi = iC^i G_i \phi$$

The gauge-fixed action is then the gauge-invariant action plus the BRST transformation of some function Λ:

$$S_{gf} = S_{gi} - iQ\Lambda$$

For example, consider Yang-Mills in the most common type of gauge, where some function of A is fixed:

$$\Lambda = tr \int \tfrac{1}{2}\tilde{C}[f(A) + \tfrac{1}{2}\alpha B] \quad \Rightarrow$$

$$\boxed{L_{gf} = L_{gi} - \tfrac{1}{2}B[f(A) + \tfrac{1}{2}\alpha B] - \tfrac{1}{2}i\tilde{C}\frac{\partial f}{\partial A} \cdot [\nabla, C]}$$

for some constant α. For $\alpha = 0$, B is a Lagrange multiplier, enforcing the gauge $f(A) = 0$, while for $\alpha \neq 0$, we can eliminate B by its auxiliary field equation:

$$-\tfrac{1}{2}B[f(A) + \tfrac{1}{2}\alpha B] \rightarrow \tfrac{1}{4\alpha}f^2$$

Examples will be given in the following section.

In field theory gauge-fixing functions always have linear terms, as do gauge transformations. Furthermore, there always exist "unitary gauges", where no ghosts are required. The ghost terms in general gauges serve simply to provide the appropriate Jacobian factor for the field redefinition that transforms from the general gauge to the unitary gauge, which appears at the quantum level from functionally integrating out the ghosts. The simplest example is the trivial gauge invariance that occurs in the Stückelberg model:

$$QA = -\partial C, \qquad Q\phi = C, \qquad Q\tilde{C} = -iB, \qquad QC = QB = 0$$

which we can fix with

$$-iQ(\tilde{C}\mathcal{O}\phi) = -B\mathcal{O}\phi + i\tilde{C}\mathcal{O}C$$

for some field-independent operator $\mathcal{O}$. Functionally integrating out B still sets $\phi = 0$, but produces an inverse functional determinant of $\mathcal{O}$ (from redefinition of ϕ, or from $\delta(\mathcal{O}\phi)$), canceled by that from integrating out the ghosts. The advantage of BRST is that all this can be treated at the classical level, in terms of the classical action, without regard to functional integration, while directly giving a solution that can be expressed immediately in terms of Feynman rules.

Exercise VIA4.1

Show that the gauge fixing

$$-iQ(\tilde{C}\mathcal{O}\phi + \tilde{C}\mathcal{A}B + C\mathcal{B}\phi + C\mathcal{C}B)$$

where $\mathcal{O}$, $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are field-independent operators, gives a result equivalent to the previous, by considering functional determinants or field redefinitions.

Exercise VIA4.2

Show that the Lagrangian

$$A\mathcal{A}B + C\mathcal{B}D \to A\mathcal{A}\mathcal{B}D$$

by the field redefinition

$$D \to D + \mathcal{B}^{-1}B$$

for bosons A, B, C, D and operators $\mathcal{A}$, $\mathcal{B}$. This is the classical equivalent of $det(\mathcal{A}\mathcal{B}) = det(\mathcal{A})det(\mathcal{B})$.

These methods apply straightforwardly to supersymmetric theories in superspace: From the gauge transformations of subsection IVC4,

$$Qe^V = i\bar{C}e^V - ie^V C; \qquad QC = iC^2, \quad Q\bar{C} = i\bar{C}^2$$

$$Q\tilde{C} = -iB, \quad Q\bar{\tilde{C}} = -i\bar{B}; \qquad QB = Q\bar{B} = 0$$

where C, $\tilde{C}$, and B are chiral superfields, and $\bar{C}$, $\bar{\tilde{C}}$, and $\bar{B}$ their hermitian conjugates.

In practice, this BRST approach is sufficient for gauge fixing. In particular, this is true for the fundamental fields used in the standard model (including gravity), which have spin$\leq$2. Therefore, we'll use mostly this approach in the rest of this text. However, some observed hadrons have much higher spin. The first-quantized approach of chapter XII gives a natural and direct way of understanding ghosts and BRST for the fields describing such particles, and translates directly into the treatment of Zinn-Justin, Batalin, and Vilkovisky (ZJBV) for field theory.

REFERENCES

1 Feynman, *loc. cit.* (VC, ref. 17);
B. DeWitt, *Phys. Rev.* **162** (1967) 1195, 1239;
L.D. Faddeev and V.N. Popov, *Phys. Lett.* **25B** (1967) 29;
S. Mandelstam, *Phys. Rev.* **175** (1968) 1580:
ghosts.

2 N. Nakanishi, *Prog. Theor. Phys.* **35** (1966) 1111;
B. Lautrup, *K. Dan. Vidensk. Selsk. Mat. Fys. Medd.* **34** (1967) No. 11, 1.

3 C. Becchi, A. Rouet, and R. Stora, *Phys. Lett.* **52B** (1974) 344, *Ann. Phys.* **98** (1976) 287;
I.V. Tyutin, Gauge invariance in field theory and in statistical physics in the operator formulation, Lebedev preprint FIAN No. 39 (1975), in Russian, unpublished;
M.Z. Iofa and I.V. Tyutin, *Theor. Math. Phys.* **27** (1976) 316;
T. Kugo and I. Ojima, *Phys. Lett.* **73B** (1978) 459;
L. Baulieu, *Phys. Rep.* **129** (1985) 1:
BRST.

4 E.S. Fradkin and G.A. Vilkovisky, *Phys. Lett.* **55B** (1975) 224;
I.A. Batalin and G.A. Vilkovisky, *Phys. Lett.* **69B** (1977) 309;
E.S. Fradkin and T.E. Fradkina, *Phys. Lett.* **72B** (1978) 343;
M. Henneaux, *Phys. Rep.* **126** (1985) 1:
Hamiltonian BRST.

B. GAUGES

There are two important properties of gauges we have examined: (1) Gauges which eliminate some degrees of freedom, such as lightcone or unitary gauges, are simpler classically, which makes them easier to understand physically. (2) Gauges that manifest as many global invariances as possible, such as the Fermi-Feynman gauge, will be found later to simplify quantum calcuations, because the explicit momentum dependence of the propagator or vertices is simpler, and keeping a symmetry manifest makes it unnecessary to check. In this section we'll examine these gauges in greater detail, especially as they relate to intereacting theories.

We'll study also some special gauges, with nontrivial interaction terms, that have both of these properties to some extent. In particular, they are manifestly Lorentz covariant, but avoid many of the complications associated with ghosts.

1. Radial

We know from nonrelativistic classical and quantum mechanics that the equations of motion can be solved exactly only for certain simple external field configurations. One particular case we have already emphasized is that of an action quadratic in the dynamical variables, i.e., the harmonic oscillator and its generalizations. Higher-order terms are then treated as perturbations about the exact solution. Such an expansion in the coordinates x is the particle version of the JWKB expansion in $\hbar$: Calling the "classical" part of x "y", we substitute $x \to y + \sqrt{\hbar} x$ and Taylor expand in x. (From now on we'll drop the $\hbar$'s, and just remember to perturb about the quadratic terms.) For the scalar field we write

$$\phi \to \phi + x \cdot \partial\phi + \tfrac{1}{2} x^m x^n \partial_m \partial_n \phi + ...$$

where $\partial ... \partial\phi$ is implicitly evaluated at y.

For the gauge fields we would like to be a bit more clever: For example, for the electromagnetic potential A_m we know we can always add a constant, so $A_m(y)$ is irrelevant, while for ∂A only $F_{mn} = \partial_{[m} A_{n]}$ is gauge invariant. This means we want to choose a gauge best suited to this calculation: a gauge that both eliminates as many as possible of the lower-order terms, and expresses $A(y + x)$ in terms of only $F(y)$ and its derivatives. Similarly, we should have a Taylor expansion for charged fields in terms of *covariant* derivatives. The appropriate gauge, which easily can be found explicitly, is the "radial gauge"

$$x \cdot A(y + x) = 0$$

(Note that, unlike F, A depends on x independently of y, not just as $y+x$, since the gauge condition itself is x-dependent. We write $A(y+x)$ only to indicate that A is evaluated at position $y+x$.) One way to solve this condition is to use the identity

$$x^n F_{nm} = (x\cdot\partial + 1)A_m - [\nabla_m, x\cdot A] \qquad (\partial = \partial/\partial x)$$

which follows from the definition of F. Using the gauge condition, we then can write

$$A_m = \frac{1}{x\cdot\partial + 1} x^n F_{nm}$$

Alternatively, we can replace x everywhere (including the argument $y+x$) by τx, and then identify $x\cdot\partial = \tau\partial/\partial\tau$ to find

$$\tau x^n F_{nm}(y+\tau x) = \partial_\tau \tau A_m(y+\tau x)$$

Integrating both sides over τ from 0 to 1, we find

$$A_m(y+x) = \int_0^1 d\tau \ \tau x^n F_{nm}(y+\tau x)$$

Note in particular that $A(y) = 0$.

Another way to define this gauge is to consider *gauge covariant* translation from y to $y+x$ to produce a gauge transformation from an arbitrary gauge to the radial gauge. Writing the covariant derivative at y as

$$\mathcal{D} = D + iA(y), \qquad D = \partial/\partial y$$

we know from subsection IIIC2 that

$$\psi'(y+x) \equiv e^{x\cdot\mathcal{D}}\psi(y) = e^{i\Lambda}e^{x\cdot D}\psi(y) = e^{i\Lambda}\psi(y+x)$$

so that covariant translation produces a $\psi'(y+x)$ that is the same as $\psi(y+x)$ up to a gauge transformation. Thus the gauge-transformed ψ can be written as a covariant Taylor expansion (for purposes of perturbation) about y:

$$\psi'(x+y) = \sum_{n=0}^{\infty} \tfrac{1}{n!} x^{a_1}\cdots x^{a_n}(\mathcal{D}_{a_1}\cdots\mathcal{D}_{a_n}\psi')(y)$$

In particular, $\psi'(y) = \psi(y)$.

However, we want to define a covariant derivative with respect to x (not y), so that

$$\nabla = \partial + iA'(x+y), \qquad \nabla\psi'(y+x) = (\mathcal{D}\psi)'(y+x) = e^{x\cdot\mathcal{D}}(\mathcal{D}\psi)(y)$$

Using

$$\partial\psi(y) = 0$$

we find the solution

$$\nabla = e^{x\cdot\mathcal{D}}\mathcal{D}e^{-x\cdot\mathcal{D}}\ mod\ e^{x\cdot\mathcal{D}}\partial e^{-x\cdot\mathcal{D}}$$

where the latter term vanishes on $\psi'(y+x)$, so the right amount of it can be added to the former expression to cancel any D terms. The result is

$$\nabla = e^{x\cdot\mathcal{D}}(\partial + \mathcal{D})e^{-x\cdot\mathcal{D}}$$

This implies $x \cdot A'(y+x) = 0$ directly: Contracting both sides with x, the Taylor expansion of the right-hand side terminates after the first couple of terms. Taylor expanding the uncontracted expression, we have

$$A'_a(y+x) = \sum_{n=0}^{\infty} \tfrac{1}{n!}\tfrac{1}{n+2}(x\cdot\mathcal{D})^n x^b F_{ba}(y)$$

We can also write

$$\partial + iA'(x+y) = \nabla = e^{x\cdot\mathcal{D}}e^{-x\cdot D}[\partial + iA(x+y)]e^{x\cdot D}e^{-x\cdot\mathcal{D}} = e^{i\Lambda}[\partial + iA(x+y)]e^{-i\Lambda}$$

and

$$\nabla = e^{x\cdot\mathcal{D}}e^{-y\cdot\partial}\mathcal{D}e^{y\cdot\partial}e^{-x\cdot\mathcal{D}}$$

Exercise VIB1.1

Show this Taylor expansion is equivalent to that obtained from the first method used in this section to solve the gauge condition. (Hint: Look out for hidden x and y dependence — How does $x\cdot\partial$ on ψ' or F' relate to $x\cdot\mathcal{D}$? Also beware of notation: In the first construction we did not use a gauge transformation, so no primes were used.)

Thus, to just quadratic order in x, the mechanics action for a relativistic particle in external fields (subsection IIIB3) becomes

$$S_L \approx \int d\tau\{-\tfrac{1}{2}v^{-1}\eta_{mn}\dot{x}^m\dot{x}^n + \tfrac{1}{2}\dot{x}^m x^n F_{nm}(y)$$

$$+v[\phi(y) + x^m(\partial_m\phi)(y) + \tfrac{1}{2}x^m x^n(\partial_m\partial_n\phi)(y)]\}$$

To this approximation the classical equations of motion are linear and can be solved exactly. It can also be used to find exact solutions for constant electromagnetic fields.

2. Lorentz

For purposes of explicit calculations in perturbation theory, it's more convenient to use gauges where Lorentz covariance is manifest. "Lorentz gauges" are a class of gauges using

$$f = \partial \cdot A$$

(and similarly for other gauge fields) as the gauge-fixing function. From the discussion of subsection VIA4, we have from the usual BRST as applied to Yang-Mills

$$L_{gf} = \tfrac{1}{8}F_{ab}^2 - iQ\ \tfrac{1}{2}[\tilde{C}(\partial \cdot A + \tfrac{1}{2}\alpha B)]$$

$$= -\tfrac{1}{4}A\cdot\Box A - \tfrac{1}{4}(\partial\cdot A)^2 - \tfrac{1}{2}[A^a, A^b](-i\partial_a A_b + \tfrac{1}{2}A_a A_b) - \tfrac{1}{2}B\partial\cdot A - \tfrac{1}{4}\alpha B^2 - \tfrac{1}{2}i\tilde{C}\partial\cdot[\nabla, C]$$

After eliminating B by its field equation, the kinetic terms are

$$-\tfrac{1}{4}A\cdot\Box A - \tfrac{1}{4}(\partial\cdot A)^2 + \tfrac{1}{4\alpha}(\partial\cdot A)^2 - \tfrac{1}{2}i\tilde{C}\Box C$$

In particular, for $\alpha = 1$ we have the "Fermi-Feynman" gauge, which gives the nicest propagators. (It is also the gauge that follows automatically from a first-quantized BRST construction, which will be described in chapter XII.) More generally, we find the propagator from inverting the kinetic operator: For the ghosts this is always $2/p^2$, but for A_a,

$$2[\eta^{ab}p^2 + (\tfrac{1}{\alpha} - 1)p^a p^b]^{-1} = 2\left[\frac{\eta_{ab}}{p^2} + (\alpha - 1)\frac{p_a p_b}{(p^2)^2}\right]$$

For $\alpha = 0$ this is the "Landau gauge", which has the advantage that the propagator is proportional to the transverse projection operator. (It kills terms proportional to p_a.) However, $\alpha = 1$ is clearly the simplest, and the $1/p^4$ term can cause problems in perturbation theory.

Exercise VIB2.1

In the Abelian case, consider making a gauge transformation on the *gauge-fixed* action (including matter), with $\lambda \sim \Box^{-1}B$. Show that the only effect is to change the value of the coefficient α of the B^2 term. Find a similar transformation for the form of the action where B has been eliminated. This shows explicitly the decoupling of the longitudinal mode of the photon.

Exercise VIB2.2

Show that for general $\mathcal{A}$ and $\mathcal{B}$

$$(\eta^{ab}\mathcal{A} + p^a p^b \mathcal{B})^{-1} = \frac{1}{\mathcal{A}}\left(\eta_{ab} - p_a p_b \frac{\mathcal{B}}{\mathcal{A} + p^2\mathcal{B}}\right)$$

Note that in the Abelian case the lightcone gauge is a special case of the Landau gauge. (An analogous situation occurs in the classical mechanics of the particle for the gauges of the worldline metric, as discussed in subsection IIIB2.) Here we have

$$0 = n^a(\partial^b F_{ab}) = n\cdot\partial(\partial\cdot A) - \Box(n\cdot A)$$

In the lightcone formalism, this is the field equation that comes from varying the auxiliary field. In the lightcone gauge $n\cdot A = 0$, it implies $\partial\cdot A = 0$ (and thus also $\Box A_a = 0$), since $n\cdot\partial$ is invertible.

This is particularly useful in D=4, where we can generalize from the lightcone to a Lorentz-covariant form by using twistors: From subsection IIB6,

$$p^2 = 0 \quad \Rightarrow \quad p^{\alpha\dot\alpha} = \epsilon(p^0)p^\alpha \bar p^{\dot\alpha}, \qquad n^2 = 0 \quad \Rightarrow \quad n^{\alpha\dot\alpha} = \epsilon(n^0)n^\alpha \bar n^{\dot\alpha}$$

Massless spinors are described on shell in momentum space by

$$\psi^\alpha = p^\alpha\phi, \qquad \bar\psi^{\dot\alpha} = \bar p^{\dot\alpha}\bar\phi$$

where external-line factors for Feynman diagrams are given by setting $\phi = 1$. For massless vectors, we have $p\cdot A = n\cdot A = 0$ (but $n\cdot p \neq 0$), so depending on whether the helicity is $+1$ (self-dual field strength) or -1 (anti-self-dual field strength), we find, respectively,

$$\bar f^{\dot\alpha\dot\beta} \sim \bar p^{\dot\alpha}\bar p^{\dot\beta}, \qquad f^{\alpha\beta} = 0 \qquad \Rightarrow \qquad A^{\alpha\dot\beta} = \frac{n^\alpha \bar p^{\dot\beta}}{n^\gamma p_\gamma}\bar\phi$$

$$f^{\alpha\beta} \sim p^\alpha p^\beta, \qquad \bar f^{\dot\alpha\dot\beta} = 0 \qquad \Rightarrow \qquad A^{\alpha\dot\beta} = \frac{p^\alpha \bar n^{\dot\beta}}{\bar n^{\dot\gamma}\bar p_{\dot\gamma}}\phi$$

The normalization of A has been chosen compatible with $|\phi| = 1$ and $A^a A^*{}_a = 1$ for evaluating cross sections. In a general Landau gauge the arbitrary gauge-dependent polarization spinors n^α, $\bar n^{\dot\alpha}$ can be chosen independently for each external line, since gauge invariance means independent gauge parameters for different momenta. (This method is known as "spinor helicity".) However, in a lightcone gauge the polarization spinors are constant.

The lightcone gauge condition is thus again a stronger gauge condition than Lorentz gauges, as expected from the fact that it has fewer derivatives. This difference shows itself in various ways:

(1) In perturbation theory on shell, in the lightcone frame the Landau gauge condition $0 = p\cdot A = -p^+A^-$ kills A^- but says nothing about A^+, which can be eliminated by the residual gauge invariance $\delta A^+ = p^+\lambda$ to obtain the lightcone gauge.

(2) In perturbation theory off shell, more derivatives in the gauge transformation imply more derivatives in the ghost kinetic operator. Thus, more ghost degrees of freedom are introduced to cancel the extra unphysical degrees of freedom in the gauge field.

(3) Lorentz gauges also have a nonperturbative ambiguity (the “Gribov ambiguity”) that axial gauges avoid: Nonperturbative solutions to the gauge condition can be found that differ from the perturbative one, in the nonabelian case. Specifically, it is possible to find a nontrivial gauge transformation $\mathbf{g}$ ($\nabla' = \mathbf{g}^{-1}\nabla\mathbf{g}$) such that

$$0 = \partial \cdot A' = -i\partial \cdot \mathbf{g}^{-1}(\nabla\mathbf{g}) \quad for \quad \partial \cdot A = 0$$

even when $\mathbf{g}$ is required to satisfy boundary conditions that it approach the identity at infinity (except in the Abelian case, where $\mathbf{g} = e^{i\lambda} \Rightarrow \Box\lambda = 0 \Rightarrow \lambda = 0$). This is not the case for axial gauges, where

$$0 = n \cdot A' = n \cdot \mathbf{g}^{-1}(\nabla\mathbf{g}) \quad for \quad n \cdot A = 0 \quad \Rightarrow \quad \mathbf{g}^{-1}(n \cdot \partial\mathbf{g}) = 0 \quad \Rightarrow \quad \mathbf{g} = I$$

by simply integrating from infinity.

3. Massive

In subsection IIB4 we described the introduction of mass for the vector by dimensional reduction, giving the Stückelberg formalism for a massive (Abelian) gauge field. The gauge-invariant action (subsection IVA5) and BRST transformation laws (subsection VIA4) followed from adding an extra dimension and setting the corresponding component of the momentum equal to the mass:

$$L_{gi} = \tfrac{1}{8}F_{ab}^2 + \tfrac{1}{4}(mA_a + \partial_a\phi)^2$$

$$QA_a = -\partial_a C, \quad Q\phi = mC, \quad Q\tilde{C} = -iB, \quad QB = 0$$

where the scalar is the extra component of the vector.

There are two obvious covariant gauges for such a vector: (1) The “unitary gauge”

$$f = \phi$$

simply gauges away the scalar. Since the scalar has a nonderivative gauge transformation, the ghosts do not propagate: The gauge-fixing term

$$-iQ(\tilde{C}\phi) = -B\phi + im\tilde{C}C$$

simply eliminates the scalar and ghosts as auxiliary fields. The net result is that we could have simply chosen

$$gauge \quad \phi = 0$$

and ignored ghosts because of ϕ's nonderivative transformation law. Thus the gauge-fixed Lagrangian is just the result of adding a mass term to the massless Lagrangian:

$$L_{gf} = \tfrac{1}{8}F_{ab}^2 + \tfrac{1}{4}m^2 A^2$$

But the propagator is

$$2[\eta^{ab}(p^2+m^2) - p^a p^b]^{-1} = 2\left[\frac{\eta_{ab}}{p^2+m^2} + \frac{p_a p_b}{m^2(p^2+m^2)}\right]$$

Notice that the second term is higher in derivatives than the first; this can cause some technical problems in perturbation theory.

(2) The Fermi-Feynman gauge works similarly to the massless case. We then modify the gauge-fixing function to

$$f = \partial\cdot A + m\phi$$

so

$$-iQ[\tfrac{1}{2}\tilde{C}(\partial\cdot A + m\phi + \tfrac{1}{2}B)] = -\tfrac{1}{2}B(\partial\cdot A + m\phi + \tfrac{1}{2}B) - \tfrac{1}{2}i\tilde{C}(\Box - m^2)C$$

$$\begin{aligned} \Rightarrow \quad L_{gf} &= \tfrac{1}{8}F^2 + \tfrac{1}{4}(mA+\partial\phi)^2 + \tfrac{1}{4}(\partial\cdot A + m\phi)^2 - \tfrac{1}{2}i\tilde{C}(\Box - m^2)C \\ &= -\tfrac{1}{4}A\cdot(\Box - m^2)A - \tfrac{1}{4}\phi(\Box - m^2)\phi - \tfrac{1}{2}i\tilde{C}(\Box - m^2)C \end{aligned}$$

The propagators are again simpler. The vector has D propagating components instead of just the D$-$1 physical ones; the 2 ghosts cancel ϕ and the extra component in A.

Exercise VIB3.1

Generalize the Fermi-Feynman gauge for the Stückelberg formalism to the "renormalizable gauges" with gauge-fixing function

$$f = \frac{m}{\mu}\partial\cdot A + \mu\phi$$

a Find the gauge-fixed action.

b Show that the ghosts and ϕ have mass μ, while the vector propagator has the form

$$2\left[\left(\eta_{ab} + \frac{p_a p_b}{m^2}\right)\frac{1}{p^2+m^2} - \frac{p_a p_b}{m^2}\frac{1}{p^2+\mu^2}\right]$$

This shows explicitly the second unphysical bosonic mode of mass μ to cancel the 2 ghosts, as well as the 3 transverse physical modes of mass m.

c Look at the cases

$$\mu = \begin{cases} 0 & \text{(Landau gauge)} \\ m & \text{(Fermi-Feynman gauge)} \\ \infty & \text{(unitary gauge)} \end{cases}$$

These two choices of gauge also exist for Yang-Mills theories exhibiting the Higgs mechanism, since those models give the Stückelberg model when linearized about the vacuum values of the fields. The advantages are the same: The unitary gauge eliminates as many unphysical degrees of freedom as possible (see subsection IVA6 for an example), while the Fermi-Feynman gauge gives the simplest propagators.

Exercise VIB3.2

Work out the Fermi-Feynman gauge for an arbitrary Higgs model, generalizing the analysis for the Stückelberg case.

4. Gervais-Neveu

We next consider pure Yang-Mills theory for the gauge group U(N), but use a *complex* gauge-fixing function

$$f_0 = \partial \cdot A + iA^2$$

where A_a is a vector of hermitian N×N matrices, and $A^2 \equiv A^a A_a$. (The hermitian conjugate, $i \to -i$, gives similar results.) The gauge-fixed Lagrangian (in the action $S = g^{-2} tr \int L$) is then

$$L_A = \tfrac{1}{8}F^2 + \tfrac{1}{4}f_0^2 = -\tfrac{1}{4}A \cdot \Box A - iA^a A^b \partial_b A_a - \tfrac{1}{4}A^a A^b A_a A_b$$

(where $\Box$ is the free D'Alembertian) while the ghost action can be written as

$$L_C = -\tfrac{1}{2}i\tilde{C}\nabla^2 C - \tfrac{1}{2}\tilde{C}Cf_0$$

where ∇ acts on C as if it were in the defining representation (i.e., $\nabla C = \partial C + iAC$, not $[A, C]$). This "Gervais-Neveu gauge" already has the simplification that some of the terms in the Yang-Mills self-interaction have been canceled.

Exercise VIB4.1

Consider the "anti-Gervais-Neveu gauge", where the same gauge-fixing term is added with opposite overall sign.

a Show the resulting Lagrangian can be written as

$$L_A \sim tr[(\partial\!\!\!/ A\!\!\!/ + iA\!\!\!/^2)^2]$$

where the trace is with respect to both (N×N) internal and (4×4) Dirac matrices. Thus, spin can be treated in a manner closely analogous to internal symmetry.

b Show the propagator can be written in the form of the product of 2 (massless) Dirac-spinor propagators.

c Starting with the complex first-order formulation of Yang-Mills of subsection IIIC4, show that the action can be written in a way that replaces the above 4×4 matrices with 2×2 matrices, as

$$L_A \sim tr[\hat{G}^2 + \hat{G}(\partial A^* + iAA^*)]$$

in first-order form, where now $\hat{G}$ is neither traceless nor symmetric in spinor indices (its trace is the Nakanishi-Lautrup field), or in second-order form as

$$L_A \sim tr[(\partial A^* + iAA^*)^2]$$

(Note that this differs from the above Dirac form, as expanded in 2×2 matrices, because it includes the Chern-Simons term.)

Next, consider a model where the Yang-Mills fields couple to scalars that are also represented by N×N matrices, but that are in the defining (N-component) representation of the gauge ("color") U(N), while also being in the defining representation of a second, global ("flavor") U(N). (See subsection IVA6.) This complex field thus has $2N^2$ real components compared to the N^2 gauge vectors, and the $2N^2$ ghosts. We also choose a Higgs potential such that the masses of the scalar and vector come out the same (but we can also specialize to the massless case). The scalar Lagrangian is then (again with $g^{-2}tr$ in the action)

$$L_\phi = -\tfrac{1}{2}\phi^\dagger \nabla^2 \phi + \tfrac{1}{4}R^2, \qquad R = \phi^\dagger\phi - \tfrac{1}{2}m^2$$

Finally, we modify the gauge-fixing function to

$$f = f_0 + iR$$

With this choice, the ghost terms are unmodified (R is gauge invariant), but the scalar self-interaction is completely canceled (including the mass term). The total Lagrangian is then

$$L = (L_A + \tfrac{1}{4}m^2A^2) + (-\tfrac{1}{2}\phi^\dagger\nabla^2\phi + i\tfrac{1}{2}\phi^\dagger\phi f_0) + i(-\tfrac{1}{2}\tilde{C}\nabla^2 C + i\tfrac{1}{2}\tilde{C}Cf_0)$$

Since the scalar Lagrangian is identical in form to that of the ghosts, and neither has self-interactions, functional integration over them will produce canceling functional determinants, because they have opposite statistics. This is a reflection of the fact that both sets of fields now describe unphysical polarizations, since both describe massless states in a theory where all physical states are massive (as seen, e.g., in a unitary gauge). This has the great advantage that, for this particular model, both the scalar fields and the ghosts can be dropped altogether, while the Lagrangian

$$L \to L_A + \tfrac{1}{4}m^2 A^2$$

completely describes the physical massive vector and scalar states. This was possible only because of the use of a complex gauge condition: The longitudinal component of the vector is now imaginary, which fixes the wrong sign associated with the Minkowski metric. A related characteristic of this gauge is that we nowhere needed to change the vacuum value of any field, unlike other gauges for actions where there is a Higgs effect.

We now note that this result for the massive case (and its massless limit) actually can be obtained more easily than the result for the pure Yang-Mills case: Since the final result has no ghosts, it is in a unitary gauge, where the vector not only "eats" the usual compensating scalar, but "overeats" by absorbing the physical scalar. The appropriate gauge condition is still complex and involves the scalars, but is now *linear*:

$$gauge \quad \phi = \langle\phi\rangle = \tfrac{1}{\sqrt{2}}mI$$

where $\phi^\dagger$, treated as independent, is unfixed. (As for the usual unitary gauge $Im\ \phi = 0$, i.e., $\phi = \phi^\dagger$, there are no propagating ghosts, since the gauge transformation of ϕ has no derivatives.) In this gauge the action becomes quadratic in $\phi^\dagger$:

$$L_\phi \to -\tfrac{1}{2}\phi^\dagger i(\partial\cdot A + iA^2)\tfrac{1}{\sqrt{2}}m + \tfrac{1}{4}(\phi^\dagger \tfrac{1}{\sqrt{2}}m - \tfrac{1}{2}m^2)^2$$

In fact, $\phi^\dagger$ appears as an auxiliary field (taking the place of the Nakanishi-Lautrup field), so we can eliminate it by its equation of motion:

$$\frac{\delta}{\delta\phi^\dagger} \quad \Rightarrow \quad \phi^\dagger = \tfrac{m}{\sqrt{2}} + \tfrac{\sqrt{2}}{m}if_0 \quad \Rightarrow \quad L = L_A + \tfrac{1}{4}m^2A^2$$

This procedure is analogous to that used for the lightcone gauge, where one component of the gauge field is fixed and one is eliminated as an auxiliary field: A closer analogy will be found in subsection VIB6.

Of course, such gauges generalize to other Higgs models, but results will not be as simple when the vector and scalar masses differ:

Exercise VIB4.2

Make the coefficient of the R^2 term in L_ϕ arbitrary, so the masses of the vector and scalar are unequal, but choose the same gauge $\phi = \langle\phi\rangle$. Find the propagator, and compare with that of exercise VIB3.1.

5. Super Gervais-Neveu

Nonhermitian gauges are also useful in supersymmetric theories: Here we consider the supersymmetric analog of the massive model of the previous section. Although we work in N=1 superspace, the model turns out to automatically have an N=2 supersymmetry. Just as the bosonic model ended with only a vector field describing only physical polarizations, we now want a real scalar superfield to have only physical polarizations. Since such a superfield has 8 bosonic components and 8 fermionic, while massless N=1 multiplets have 2+2 physical polarizations, we need 1 vector multiplet plus 3 scalar multiplets. Since the bosonic model had a complex scalar representation, 2 of these scalar multiplets must form the analogous defining $\otimes$ defining representation of local $\otimes$ global groups, so the last must be a real (adjoint) representation of the local group. The model is then given by (where again $S = g^{-2}\, tr \int dx\ L$)

$$L_{gi} = -\int d^2\theta\ W^2 - \int d^4\theta\ (e^{-V}\bar{\phi}_0 e^V \phi_0 + \bar{\phi}_+ e^V \phi_+ + \phi_- e^{-V}\bar{\phi}_-)$$

$$- \left[\int d^2\theta\ (\phi_+\phi_- - \tfrac{1}{4}m^2)\phi_0 + h.c.\right]$$

where we have included the only possible scale-invariant potential term, and introduced a Higgs mechanism by an N=2 Fayet-Iliopoulos term, which we chose to write in terms of the chiral scalar. (See subsection IVC7.)

The BRST transformations (which also imply the gauge transformations) for this action are (see subsection VIA4)

$$Qe^V = i\bar{C}e^V - ie^V C, \qquad Qe^{-V} = -e^{-V}(Qe^V)e^{-V} = iCe^{-V} - ie^{-V}\bar{C}$$

$$QC = iC^2, \quad Q\bar{C} = i\bar{C}^2; \qquad Q\tilde{C} = -iB, \quad Q\tilde{\bar{C}} = -i\bar{B}; \qquad QB = Q\bar{B} = 0$$

$$Q\phi_+ = iC\phi_+, \qquad Q\phi_0 = i[C, \phi_0], \qquad Q\phi_- = -i\phi_- C$$

$$Q\bar{\phi}_+ = -i\bar{\phi}_+\bar{C}, \qquad Q\bar{\phi}_0 = i[\bar{C}, \bar{\phi}_0], \qquad Q\bar{\phi}_- = i\bar{C}\bar{\phi}_-$$

Our nonhermitian choice for the BRST gauge-fixing function is

$$\Lambda = -\int d^2\theta\ \tilde{C}(\bar{d}^2 e^{-V} + \phi_0) - \int d^2\bar{\theta}\ \tilde{\bar{C}}(d^2 e^V + \bar{\phi}_0)$$

Note that e^V is an element of the algebra as well as a "nonunitary element" of the group, only because we chose the group U(N) (as was the case for A^2 in the bosonic version). The gauge-fixing and ghost terms are then

$$-iQ\Lambda = \int d^2\theta\ B(\bar{d}^2 e^{-V} + \phi_0) + \int d^2\bar{\theta}\ \bar{B}(d^2 e^V + \bar{\phi}_0)$$
$$- \int d^4\theta\ (\tilde{C} e^{-V} \bar{C} + \tilde{\bar{C}} e^V C) + \int d^2\theta\ C\tilde{C}\phi_0 + \int d^2\bar{\theta}\ \bar{C}\tilde{\bar{C}}\bar{\phi}_0$$

where we have used the field equation enforced by the Lagrange multipliers B and $\bar{B}$ (or, equivalently, made field redefinitions of the Lagrange multipliers to generate terms proportional to their constraints).

Exercise VIB5.1

Make a component analysis of this theory:

a Expand the gauge-invariant action in components.

b Do the same for the gauge-fixing terms.

c Compare the bosonic part of both the gauge-invariant and gauge-fixed actions to those of the previous subsection, after elimination of auxiliary fields.

We now see that the ghost terms are identical in form to those for $\phi_\pm$, under the identification

$$(\phi_+, \phi_-, \bar{\phi}_+, \bar{\phi}_-) \leftrightarrow (C, \tilde{C}, \tilde{\bar{C}}, \bar{C})$$

(but beware signs from ordering of ghosts). So again the ghosts cancel the (N=2) matter fields, leaving only the N=2 vector multiplet. But we can also eliminate the N=1 matter half of this N=2 multiplet using the Nakanishi-Lautrup Lagrange multiplers: The final simple result for the gauge-fixed action is thus

$$L = -\int d^2\theta\ W^2 - \int d^4\theta\ [e^{-V}(d^2 e^V) e^V \bar{d}^2 e^{-V} + \tfrac{1}{4}m^2(e^V + e^{-V})]$$

A further simplification results from the redefinition (again possible only for U(N))

$$e^V \to 1 + V$$

This also simplifies the BRST (and gauge) transformation for V:

$$QV = i(\bar{C} - C) + i(\bar{C}V - VC)$$

whose linear form resembles the bosonic case. Using the expression (see exercise IVC4.1)

$$W_\alpha = -i\bar{d}^2 e^{-V} d_\alpha e^V \to -i\bar{d}^2 \frac{1}{1+V} d_\alpha V$$

for the field strength, the Lagrangian becomes

$$L = -\int d^4\theta \ \left[-\tfrac{1}{2}\frac{1}{1+V}(d^\alpha V)\bar{d}^2\frac{1}{1+V}d_\alpha V + \frac{1}{1+V}(d^2 V)(1+V)\bar{d}^2\frac{1}{1+V} \right.$$
$$\left. +\tfrac{1}{4}m^2\left(V+\frac{1}{1+V}\right)\right]$$

Although the nonabelian vector multiplet has nonpolynomial self-interactions in any gauge, this gauge simplifies the lower-point interactions, which are the ones more frequently used for a fixed number of external lines. Expanding this action to cubic order, we use the identity

$$d^\alpha \bar{d}^2 d_\alpha = \bar{d}^{\dot\alpha} d^2 \bar{d}_{\dot\alpha} = -\tfrac{1}{2}\Box + \{d^2, \bar{d}^2\}$$

for the kinetic term, and

$$\bar{d}_{\dot\alpha} d^2 = d^2 \bar{d}_{\dot\alpha} + i\partial_{\alpha\dot\alpha} d^\alpha$$

for the gauge-fixing part of the cubic term, with integration by parts. (For the gauge-invariant term, some work can be saved by using the equivalent $\overline{W}^2$ form.) The result is

$$L = \int d^4\theta \ \{\tfrac{1}{4}V(\Box - m^2)V + [\tfrac{1}{4}m^2V^3 + (\bar{d}^{\dot\alpha}V)Vi\partial_{\alpha\dot\alpha}d^\alpha V] + \mathcal{O}(V^4)\}$$

Not only are there fewer terms than with linear gauge conditions, but these terms have fewer spinor derivatives, which yields viewer nonvanishing contributions in loops (see subsection VIC5). As for the bosonic model of the previous subsection, this analysis also applies for the unbroken case $m = 0$.

Exercise VIB5.2

Find the corresponding form of the kinetic and cubic terms without the re-definition $e^V \to 1 + V$.

Exercise VIB5.3

Gauge fix by using the unitary gauge

$$\phi_+ = \bar{\phi}_- = \tfrac{m}{\sqrt{2}}$$

to obtain the same result.

Exercise VIB5.4

Look at the super *anti*-Gervais-Neveu gauge, or super *anti*-Fermi-Feynman gauge, changing the sign of the gauge-fixing term for the vector multiplet (see exercise VIB4.1).

a Show that in the massless case the kinetic operator becomes, instead of $\Box$,

$$K \sim d^4 \equiv \tfrac{1}{4!}\epsilon^{\alpha\beta\gamma\delta}d_\alpha d_\beta d_\gamma d_\delta$$

where we now use 4-component spinor indices.

b Show that the resulting propagator is of the form, in supercoordinate space,

$$\varDelta(x,\theta;x',\theta') \sim \frac{d^4}{\Box^2}\delta^4(\theta-\theta')\delta^4(x-x') \sim ln[(x-x'-i\tfrac{1}{2}\theta\gamma\theta')^2]$$

where "$x-x'-i\frac{1}{2}\theta\gamma\theta'$" (see subsection IIC2) is the supersymmetry invariant. (Hint: Use $d^4 = \int d^4\zeta\ e^{\zeta^\alpha d_\alpha}$. Warning: If derived by Fourier transformation, the integral is infrared divergent, and requires dropping an infinite constant.)

6. Spacecone

We have just seen that gauge independence allows complex gauge conditions, which make the action complex. (In subsection IIIC4, we also used complex auxiliary fields, with a similar effect.) In this subsection we introduce a complex analog of the lightcone, the "spacecone", which will greatly simplify Feynman diagram calculations with massless fields. The spacecone gauge condition is

$$A^2 - iA^3 = 0$$

or more generally

$$n\cdot A = 0, \qquad n^2 = 0, \qquad n\cdot n^* > 0$$

(but only n^a, not n^{*a}, appears in the action). While this gauge is spacelike (in the sense that only spatial components of the gauge field are fixed), it is also null, by virtue of being complex. Thus, although algebraically like the lightcone, it allows canonical quantization with the usual time coordinate. In fact, it is just a Wick rotation of the lightcone. We then eliminate $A^2 + iA^3$ as an auxiliary field.

The spacecone is a new gauge to add to our list of axial gauges $n\cdot A = 0$ from subsection IIIC2, and the related gauges for scalars from subsections IVA5-6, VIB3-4:

axial gauges	*non-null*	*null (+ auxiliary field eq.)*
(partly) temporal	*timelike* : $A^0 = 0$	*lightcone* : $A^+ = 0,\ \delta/\delta A^-$
spacelike	*Arnowitt-Fickler* : $A^1 = 0$	*spacecone* : $A^t = 0,\ \ \delta/\delta\bar{A}^t$
scalar	*unitary* : $\phi = \phi^\dagger$	*Gervais-Neveu* : $\phi = \langle\phi\rangle,\ \delta/\delta\phi^\dagger$

In fact, at least for the free theories, the gauges for the scalars can be considered as dimensional reductions (from 1 or 2 extra dimensions) of those for the vector, as used for deriving the Stückelberg formalism in subsection IIB4, where the spacelike components of the vector associated with gauge fixing become scalars: Arnowitt-Fickler → unitary, spacecone → Gervais-Neveu.

The main advantages of the spacecone over the lightcone are special to D=4, so we now review the lightcone in a way specialized to physical spacetime. We first repeat the results of subsection IIIC2, relabeling the indices appropriately. Starting with the gauge condition (see subsections IA4 and IIA3 for notation)

$$A^t = 0$$

and eliminating $\bar{A}^t$ by its field equation, the Lagrangian for pure Yang-Mills becomes

$$L = A^+ \partial^t \bar{\partial}^t A^- - \tfrac{1}{4}(F^{+-})^2 + \tfrac{1}{4}(F^{t\bar{t}})^2$$

$$\begin{aligned} F^{+-} &= \partial^+ A^- - \partial^- A^+ + i[A^+, A^-] \\ F^{t\bar{t}} &= \partial^- A^+ + \partial^+ A^- + i\frac{1}{\partial^t}([A^+, \partial^t A^-] + [A^-, \partial^t A^+]) \end{aligned}$$

We simplify the Lagrangian by using the self-dual and anti-self-dual combinations: Dropping also the t superscripts on ∂ for simplicity,

$$\begin{aligned} \mathcal{F}^\pm &= \tfrac{1}{2}(F^{t\bar{t}} \pm F^{+-}) = \partial^\pm A^\mp + i\frac{1}{\partial}[A^\pm, \partial A^\mp] \\ L &= A^+ \partial\bar{\partial} A^- + \mathcal{F}^+ \mathcal{F}^- \\ &= A^+ \tfrac{1}{2}\Box A^- - i\left(\frac{\partial^-}{\partial} A^+\right)[A^+, \partial A^-] - i\left(\frac{\partial^+}{\partial} A^-\right)[A^-, \partial A^+] \\ &\qquad + [A^+, \partial A^-]\frac{1}{\partial^2}[A^-, \partial A^+] \end{aligned}$$

Exercise VIB6.1

Label all the fields and derivatives in the above forms of the Lagrangian, $\mathcal{F}$ in particular, in spinor notation.

Exercise VIB6.2

Show that the field redefinitions $A^\pm \to (\partial)^{\pm 1} \phi^\pm$, when applied to just the first two terms of the above Lagrangian, produce a local action describing the self-dual field equations of the lightcone formalism of subsection IIIC5 (taking into account the difference between the lightcone and spacecone). Compare the results of exercise IIIC5.2. Thus, by treating the latter two terms separately from the former two, Yang-Mills can be treated as a perturbation about self-dual Yang-Mills.

Another simplification for massless D=4, and closely related to the use of helicity, is twistors. For our Feynman diagram calculations for spins ≤ 1, almost all spinor algebra involves objects carrying at most two spinor indices (spinors, vectors, self-dual tensors), so we use the twistor matrix notation of subsection IIB6. In particular, in a general class of gauges the external line factors for Yang-Mills fields in this notation (see subsection VIB2) read

$$A = \epsilon_+ = \frac{|\epsilon\rangle[p|}{\langle\epsilon p\rangle} \quad \Rightarrow \quad f^* \sim |p][p|, \quad f = 0$$

for + helicity or

$$A = \epsilon_- = \frac{|p\rangle[\epsilon|}{[\epsilon p]} \quad \Rightarrow \quad f \sim |p\rangle\langle p|, \quad f^* = 0$$

for $-$, where $(\epsilon_\pm)^a$ are the polarization 4-vectors for helicity ± 1 in terms of a twistor $\epsilon^\alpha, \bar{\epsilon}^{\dot{\alpha}}$, which can vary from line to line, and whose choice defines the gauge, as a special case of the Landau gauge. (Positive helicity is the same as self-duality, negative is anti-self-dual. The Landau gauge condition is generally applied in arbitrary Lorentz gauges to external lines, to eliminate the redundant longitudinal degrees of freedom.)

One special case is the lightcone gauge

$$n \cdot A = 0, \quad n = |\epsilon\rangle[\epsilon|$$

in terms of an arbitrary constant lightlike vector n. A more convenient gauge is the spacecone gauge, which can be written in terms of two twistors:

$$n = |\epsilon_+\rangle[\epsilon_-|$$

These two twistors are sufficient to define a complete reference frame: We can convert all spinor indices into this basis, as

$$\psi^\alpha = \psi^\pm {\epsilon_\pm}^\alpha$$

etc. This corresponds to using two lightlike vectors to define the spacecone gauge, $n^\pm = |\epsilon_\pm\rangle[\epsilon_\pm|$. For simplicity, we write $|\epsilon_\pm\rangle = |\pm\rangle$; then a vector in this basis can be written as

$$p = p^+|+\rangle[+| + p^-|-\rangle[-| + p^t|-\rangle[+| + \bar{p}^t|+\rangle[-|$$

if we use the normalization

$$\langle +-\rangle = [-+] = 1$$

E.g., for massless momentum $p = |p\rangle[p|$,

$$p^+ = \langle p-\rangle[-p], \quad p^- = \langle +p\rangle[p+], \quad p^t = \langle +p\rangle[-p], \quad \bar{p}^t = \langle p-\rangle[p+]$$

We will also drop the superscript t in contexts where there is no ambiguity. This basis is related to our previous spinor basis up to phase factors, $|\pm\rangle \sim |_\pm\rangle$, $|\pm] \sim |_{\dot{\pm}}]$, and we assume them to be commuting (rather than anticommuting); these changes are more convenient for dealing with twistors (commuting spinors).

The advantage of the spacecone is that we can Lorentz covariantize the Feynman rules by identifying these two lightlike vectors with physical on-shell massless momenta. We need two such "reference" vectors because we are not allowed to have $n = p$ on any line. Since only $|+\rangle$ appears in the external line factors for helicity $+1$, and only $|-]$ in those for -1, the simplest choice is to pick the momentum of one external line with helicity $+1$ to define $|-]$ for all lines with helicity -1, and pick the momentum of one line with helicity -1 to define $|+\rangle$ for lines with helicity $+1$. (In the presence of massless external spinors, we can also choose a helicity $+1/2$ line to define $|-]$, etc.) Our above normalization means that we have chosen the phase $\langle +-\rangle/[-+] = 1$ as allowed by the usual ambiguity of twistor phases, while our choice of the magnitude $\langle +-\rangle[-+] = -\langle +|[+|\cdot|-\rangle|-] = 1$ is a choice of (mass) units. In explicit calculations, we restore generality (in particular, to allow momentum integration) by inserting appropriate powers of $\langle +-\rangle$ and $[-+]$ at the end of the calculations, as determined by simple dimensional and helicity analysis. (This avoids a clutter of normalization factors $\sqrt{\langle +-\rangle[-+]}$ at intermediate stages.) For example, looking at the form of the usual spinor helicity external line factors, and counting momenta in the usual Feynman rules, we see that any tree amplitude (or individual graph) in pure Yang-Mills must go as

$$\langle\ \rangle^{2-E_+}[\ ]^{2-E_-}$$

where $E_\pm$ is the number of external lines with helicity $\pm$.

We now return to external line factors. The naive factors for the above Lagrangian are 1, since the kinetic term resembles that of a scalar. However, this would lead to unusual normalization factors in probabilities, which are not obvious in this complex gauge. Therefore, we determine external line factors from the earlier spinor helicity expressions for external 4-vectors. In Lorentz gauges $(\epsilon_\pm)^a$ would be the polarization for helicity ± 1 for the complete 4-vector, but in the spacecone formalism only $A^\pm$ appear. Furthermore, in the spacecone we find

$$(\epsilon_+)^- = -|+\rangle[+|\cdot\frac{|+\rangle[p|}{\langle +p\rangle} = 0$$

$$(\epsilon_-)^+ = -|-\rangle[-|\cdot\frac{|p\rangle[-|}{[-p]} = 0$$

since by antisymmetry $\langle ++\rangle = [--] = 0$, so that A^+ carries only helicity $+1$ and A^- only -1. (This statement has literal meaning only on shell, but we can make this convenient identification more general by using it as a definition of helicity off shell.) The appropriate external line factors for these fields are thus

$$(\epsilon_+)^+ = -|-\rangle[-| \cdot \frac{|+\rangle[p|}{\langle +p\rangle} = \frac{[-p]}{\langle +p\rangle}$$

$$(\epsilon_-)^- = -|+\rangle[+| \cdot \frac{|p\rangle[-|}{[-p]} = \frac{\langle +p\rangle}{[-p]}$$

Note that these factors are inverses of each other, consistent with leaving invariant (the inner product defined by) the kinetic term.

An exception is the external line factors for the reference momenta themselves, where $|p\rangle = |\mp\rangle$ for helicity $\pm$ gives vanishing results. However, examination of the Lagrangian shows this zero can be canceled by a $1/\partial$ in a vertex, since $p = \bar{p} = 0$ for the reference momenta by definition. (Such cancelations occur automatically from field redefinitions in the lightcone formulation of the self-dual theory.) The actual expressions we want to evaluate, before choosing the reference lines, are then

$$\frac{p^-}{p}(\epsilon_+)^+ = \frac{\langle +p\rangle[p+]}{\langle +p\rangle[-p]}\,\frac{[-p]}{\langle +p\rangle} = \frac{[p+]}{\langle +p\rangle}$$

$$\frac{p^+}{p}(\epsilon_-)^- = \frac{\langle p-\rangle[-p]}{\langle +p\rangle[-p]}\,\frac{\langle +p\rangle}{[-p]} = \frac{\langle p-\rangle}{[-p]}$$

Evaluating the former at $|p\rangle = |-\rangle$ and the latter at $|p\rangle = |+\rangle$, we get 1 in both cases. In summary, for reference lines: (1) use only the 3-point vertex of the corresponding self-duality ($\pm\pm\mp$ for helicity $\pm$), and use only the term associating the singular factor with the reference line (the other term and the other vertices give vanishing contributions); (2) including the momentum factors on that line from the vertex, the external line factor is 1.

7. Superspacecone

To generalize these results to high-energy (massless) QCD, we consider supersymmetric QCD, i.e., Yang-Mills coupled to massless fermions in the adjoint representation. For tree graphs, this is equivalent to ordinary massless QCD except for group theory, which can be evaluated separately. We first apply the spacecone approach to the component action for supersymmetric QCD: The modification of this action for ordinary massless QCD is trivial (replacing the adjoint quark current with defining). From this we derive the "superspacecone" formalism, rewriting the action more simply in terms of spacecone superfields. (This form can also be derived from the usual superspace, but we will not consider that here.)

We now combine the spacecone approach to pure Yang-Mills of subsection VIB6 with the spacecone version of the lightcone treatment of the massless spinor in subsection IIIC2. We modify the lightcone to the spacecone for the quarks by instead eliminating $\psi^{\oplus}$ and $\bar{\psi}^{\dot{\ominus}}$ as auxiliary. For later convenience, we also write the remaining fermionic fields as

$$\bar{\psi}^{\dot{\oplus}} \to \psi^{+}, \qquad \psi^{\ominus} \to -\psi^{-}$$

Then we directly find the terms in the Lagrangian

$$L = A^{+}\partial\bar{\partial}A^{-} + \mathcal{F}^{+}\mathcal{F}^{-} + i\psi^{+}(\bar{\partial} - \nabla^{-}\tfrac{1}{\partial}\nabla^{+})\psi^{-}$$

$$\mathcal{F}^{\pm} = \partial^{\pm}A^{\mp} - \tfrac{1}{\partial}([A^{\pm}, -i\partial A^{\mp}] + \{\psi^{+}, \psi^{-}\})$$

where the quark term in $\mathcal{F}^{\pm}$ comes from the quark coupling to $A^{\bar{t}}$ when using its equation of motion to solve for $F^{t\bar{t}}$. Collecting terms, we have

$$L = L_2 + L_3 + L_4$$

$$L_2 = A^{+}\tfrac{1}{2}\Box A^{-} + \psi^{+}\frac{\frac{1}{2}\Box}{-i\partial}\psi^{-}$$

$$L_3 = \left(\frac{\partial^{\mp}}{\partial}A^{\pm}\right)([A^{\pm}, -i\partial A^{\mp}] + \{\psi^{+}, \psi^{-}\}) + \left(\frac{\partial^{\mp}}{\partial}\psi^{\pm}\right)[A^{\pm}, \psi^{\mp}]$$

$$L_4 = -([A^{+}, -i\partial A^{-}] + \{\psi^{+}, \psi^{-}\})\frac{1}{\partial^2}([A^{-}, -i\partial A^{+}] + \{\psi^{+}, \psi^{-}\}) - [A^{+}, \psi^{-}]\frac{1}{-i\partial}[A^{-}, \psi^{+}]$$

where terms with $\pm$ have only a single sum over it. Although this Lagrangian is much messier than the original covariant one, one again saves work by expanding terms once in the action rather than repeatedly for each Feynman diagram.

External-line factors for the spinors follow from the covariant ones of subsection VIB2 as they did for the spacecone vectors of subsection VIB6. We thus have

$$\psi^{+} = [-p], \qquad \psi^{-} = \langle +p\rangle$$

Compared with those for $A^{\pm}$, we see $\psi^{+}\psi^{-}$ has an extra factor of $p = \langle +p\rangle[-p]$ as compared with $A^{+}A^{-}$, as expected from the extra factor of $1/(-i\partial)$ in the kinetic operator. Similarly, if we choose to use external quark lines as reference lines, we use

$$\frac{p^{-}}{p}\psi^{+} = [p+], \qquad \frac{p^{+}}{p}\psi^{-} = \langle p-\rangle$$

which also reduce to 1 for the appropriate reference momenta.

Noting that the bosonic and fermionic terms are the same except for factors of $-i\partial$, we can combine them into chiral superfields that depend on only two anticommuting coordinates, θ^{+} and θ^{-} (really $\theta^{\oplus}$ and $\bar{\theta}^{\dot{\ominus}}$):

$$S = \tfrac{1}{g^2}\,tr\int dx\ d\theta^{+}d\theta^{-}\ \mathcal{L}, \qquad \int d\theta^{+}d\theta^{-} = d_{+}d_{-}\ or\ -d_{-}d_{+}, \qquad \{d_{+}, d_{-}\} = -i\partial$$

$$d_\pm \phi^\mp = 0; \qquad \phi^\pm| = A^\pm, \qquad d_\pm \phi^\pm| = \psi^\pm$$

(no sum on $\pm$). These spinor derivatives (and their corresponding supersymmetry generators) describe only spatial supersymmetry, since they contain no time derivatives. Then, using the identity

$$d_\pm d_\mp [\phi^\pm, \phi^\mp]| = [A^\pm, -i\partial A^\mp] + \{\psi^+, \psi^-\}$$

we easily combine the terms in the Lagrangian L into the superspacecone Lagrangian $\mathcal{L}$:

$$\mathcal{L} = \phi^+ \frac{\frac{1}{2}\Box}{-i\partial}\phi^- + \left(\frac{\partial^\mp}{\partial}\phi^\pm\right)[\phi^\pm, \phi^\mp] + [\phi^+, d_-\phi^-]\frac{1}{\partial^2}[\phi^-, d_+\phi^+]$$

The last term can also be written as

$$-[\phi^+, \phi^-]\frac{d_+ d_-}{\partial^2}[\phi^+, \phi^-]$$

Exercise VIB7.1

Introduce another pair of chiral superfields as auxiliary. Show the above $\mathcal{L}$ then can be rewritten in local form, with no spinor derivatives, where the kinetic term resembles the covariant one for a massless spinor, while the interaction term contains no derivatives and is only cubic. (Hint: $d_\pm d_\mp/(-i\partial)$ are projection operators for chiral superfields.) Thus this Lagrangian resembles the Chern-Simons one that appears on the 3D boundary for the topological term in Yang-Mills (see subsection IIIC6). Expand the action in components, and separate out the pure Yang-Mills part.

Exercise VIB7.2

Repeat exercise VIB6.2 to obtain the superspacecone action for selfdual super Yang-Mills, quadratic in ϕ^+ and linear in ϕ^-. Use the field redefinition

$$\phi^- \to d_+\varphi^-, \qquad d_-\varphi^- = 0$$

and integrate the action over just θ^- (by acting with d_-) to obtain a "chiral" action, with no spinor derivatives, and superfields that are functions of just θ^+ integrated over just θ^+. After further redefinitions as in VIB6.2, obtain an action *identical* to the nonsupersymmetric one obtained there, *except* for the $\int d\theta^+$. Expand in components, and relate to the nonsupersymmetric case.

8. Background-field

A more general type of gauge choice is the background field gauge. As we saw in subsection VC1, the generating functional can be written in a form where the quantum field is expanded about a background field in the interaction part of the classical action. The basic steps of the background field gauge method are:

(1) choose gauge fixing that is gauge invariant in the background gauge field,

(2) show that the quantum/background splitting of the entire gauge-invariant action is also gauge invariant in the background gauge field, and

(3) show that the effect of splitting the kinetic term in the gauge-invariant action can be neglected. (Only the interaction terms should have been split.)

The result is then that the effective action Γ, which depends only on the background fields, is gauge invariant in them. This gauge invariance is a strong condition which not only simplifies the effective action but allows a "background gauge" to be chosen for it that is independent of the "quantum gauge" applied to the path integral: The background fields and quantum fields can be in different gauges. For example, for a relativistic treatment of the quantum corrections to bound states whose constitutents are nonrelativistic (such as the hydrogen atom), it is convenient to use a Fermi-Feynman gauge (convenient for relativistic matter coupling to electromagnetism or chromodynamics) for the quantum fields and a Coulomb gauge (convenient for static or nonrelativistic matter) for the background fields.

A simple way to formulate the background expansion is in terms of the covariant derivative:

$$A \to \tilde{A} = \mathcal{A} + A \quad \Rightarrow \quad \nabla \to \mathcal{D} + iA, \qquad \mathcal{D} = \partial + i\mathcal{A}$$

where A is the quantum field (the variable of path integration) and $\mathcal{D}$ is the "background covariant derivative" in terms of the background field $\mathcal{A}$. We then find for the field strength

$$F_{ab} \to -i[\mathcal{D}_a + iA_a, \mathcal{D}_b + iA_b] = \mathcal{F}_{ab} + \mathcal{D}_{[a}A_{b]} + i[A_a, A_b]$$

and similarly for the action. Matter fields are split as usual,

$$\phi \to \tilde{\phi} = \varphi + \phi$$

We now have two gauge invariances, corresponding to the two gauge fields. Both transformations are defined to have the same, usual form on $\nabla = \mathcal{D} + iA$ (and on

$\tilde{\phi} = \varphi + \phi$), and thus both leave the action inert, but (1) the "background gauge invariance" is defined to transform the background fields covariantly

$$background: \qquad \mathcal{D}' = e^{i\lambda}\mathcal{D}e^{-i\lambda} \quad (\varphi' = e^{i\lambda}\varphi), \qquad \nabla' = e^{i\lambda}\nabla e^{-i\lambda} \qquad (\tilde{\phi}' = e^{i\lambda}\tilde{\phi})$$

$$\Rightarrow \quad A' = e^{i\lambda}Ae^{-i\lambda} \qquad (\phi' = e^{i\lambda}\phi)$$

and thus the quantum field transforms as a matter (non-gauge) field, while (2) the "quantum gauge invariance" is defined to leave the background fields inert

$$quantum: \qquad \mathcal{D}' = \mathcal{D} \quad (\varphi' = \varphi), \qquad \nabla' = e^{i\lambda}\nabla e^{-i\lambda} \qquad (\tilde{\phi}' = e^{i\lambda}\tilde{\phi})$$

$$\Rightarrow \quad A' = e^{i\lambda}[(-i\mathcal{D} + A)e^{-i\lambda}] \qquad [\phi' = e^{i\lambda}(\varphi + \phi) - \varphi]$$

The latter then determines the new BRST transformations

$$QA_a = -[\mathcal{D}_a + iA_a, C], \quad QC = iC^2, \quad Q\tilde{C} = -iB, \quad QB = 0 \qquad [Q\phi = iC(\varphi + \phi)]$$

The key to the background field gauge is to break the quantum invariance, so a propagator can be defined, but preserve the background invariance, so the path integral is gauge invariant. Since Q, the BRST operator for the *quantum* gauge invariance, is now background gauge invariant, we need only choose a gauge-fixing function Λ that is also background gauge invariant. Many gauges are possible: The basic rule is to modify any normal gauge condition simply by replacing any partial derivatives ∂ with background covariant derivatives $\mathcal{D}$. For example, for a Lorentz covariant gauge

$$\partial \cdot A \to \mathcal{D} \cdot A$$

We then gauge fix in the usual way, and now the gauge-fixing terms and the ghost terms are background gauge invariant, as long as we define the ghosts to transform covariantly:

$$background: \qquad C' = e^{i\lambda}Ce^{-i\lambda}, \qquad \tilde{C}' = e^{i\lambda}\tilde{C}e^{-i\lambda}, \qquad B' = e^{i\lambda}Be^{-i\lambda}$$

For example, for Lorentz gauges the ghost term is modified, by the modification of the gauge condition and the quantum BRST transformation, as

$$\tilde{C}\partial \cdot \nabla C \to \tilde{C}\mathcal{D}^2 C + \tilde{C}\mathcal{D} \cdot i[A, C]$$

Furthermore, even axial gauges are modified: For example, even though the gauge condition $A_0 = 0$ allows elimination of a component of the quantum field, it doesn't affect the background field, which now appears in the ghost Lagrangian

$$\tilde{C}\partial_0 C \to \tilde{C}\mathcal{D}_0 C$$

Since the S-matrix is gauge-independent (when BRST is used to perform gauge fixing, as we have), we can use the background field gauge version of the generating functional (now using ϕ to represent all quantum fields and φ all background),

$$Z[\varphi] = \int D\phi\ e^{-i\tilde{S}}, \qquad \tilde{S} = S_0[\phi] + S_I[\phi+\varphi] - iQ\Lambda[\phi,\varphi] = \hat{S} - (S_0[\phi+\varphi] - S_0[\phi])$$

$$\hat{S} = S[\phi+\varphi] - iQ\Lambda, \qquad S[\phi] = S_0[\phi] + S_I[\phi]$$

where $S[\phi]$ is the original gauge-invariant action, $Q\Lambda$ is the gauge-fixing as described above, and $\hat{S}$ is the sum of this gauge fixing and the background-expanded gauge-invariant action. We have thus separated the total action $\tilde{S}$ appearing in the background-gauge-fixed generating functional into the background-gauge-invariant part $\hat{S}$ minus the noninvariant terms $S_0[\phi+\varphi] - S_0[\phi]$.

As usual, the classical part of the effective action $\Gamma[\varphi]$ is given by adding the kinetic term $S_0[\varphi]$ of the (gauge-invariant) classical action to the 1PI tree graphs, which are just the vertices for the background fields. (The Q term doesn't contribute because it has no pure background piece.) Thus,

$$\Gamma_{class}[\varphi] = \tilde{S}|_{\phi=0} + S_0[\varphi] = \hat{S}|_{\phi=0} = S[\varphi]$$

We now note that, as far as calculating just the effective action is concerned, we can drop all terms in the gauge-fixed action independent of or linear in ϕ: Any independent term contributes only classically; any linear term will generate one-particle *reducible* graphs ("tadpoles"). This means we can drop the noninvariant terms $S_0[\phi+\varphi] - S_0[\phi]$ from $\tilde{S}$. Thus, the Feynman rules for calculating Γ are: (1) Use the classical gauge-invariant action $S[\varphi]$ for the classical contribution to Γ; and (2) for the quantum contribution, use all the 1PI loop graphs coming from $\hat{S}$. The result is background gauge invariant, since $\hat{S}$ is manifestly so.

Another important feature of the quantum-gauge-fixed background field action is that it is background-gauge-invariant order-by-order in the quantum fields. In fact, every term in the corresponding ordinary gauge action has been replaced by one (or more, if there are $\mathcal{F}$ terms) background-gauge-invariant term.

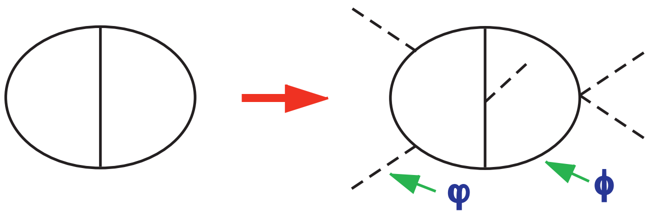
Exercise VIB8.1

Consider the Fermi-Feynman background-field gauge for the quantum field of pure Yang-Mills theory. Write all terms (both gauge-invariant and gauge-fixed) quadratic in the quantum field. Show that these combine as

$$-\tfrac{1}{4}A\cdot\Box A - i\tfrac{1}{2}A^a[\mathcal{F}_{ab}, A^b]$$

where $\Box = (\mathcal{D})^2$ (and "$\mathcal{D}A$" means "$[\mathcal{D}, A]$", etc.).

Since all external lines are associated with background fields, if we draw graphs in such a way as to exhibit only the quantum fields, they will all look like vacuum graphs: graphs with no external lines. However, any particular such vacuum graph will represent many of the original graphs, since the background lines can be attached in many ways. Furthermore, in background field gauges any such vacuum graph, considered as a contribution to the effective action, will be gauge invariant with respect to the background gauge transformations, since it results from the non-background gauge true vacuum graph by the replacement of the ordinary derivative with the background covariant derivative $\partial \to \mathcal{D}$ (plus perhaps some noniminimal $\mathcal{F}$ terms), including in the propagator. In particular, the complete one-loop contribution to Γ is given by the vacuum graph with no quantum interactions: It can be obtained from just the part of the $\hat{S}$ that is quadratic in the quantum fields.



Exercise VIB8.2

Consider an arbitrary gauge-invariant Yang-Mills action $S[\tilde{A}]$ with $\tilde{A} = \mathcal{A} + A$ in terms of the background field $\mathcal{A}$ and quantum field A. Taylor expand the action in A as

$$S[\tilde{A}] = S[\mathcal{A}] + A\frac{\delta S[\mathcal{A}]}{\delta \mathcal{A}} + ...$$

The infinitesimal *quantum* gauge transformation mixes different-order terms in the expansion. Show that the term quadratic in A is invariant under an *Abelian* quantum gauge transformation only if the background satisfies the field equations, $\delta S[\mathcal{A}]/\delta\mathcal{A} = 0$. Similar remarks apply to BRST transformations and the gauge-fixed action. (Since quadratic actions, even in background fields, yield only a propagator, they can be described by first-quantization: Thus gauge invariance implying background field equations occurs whenever a gauge field appears as both a quantum mechanical state and a background field, for example in string theory. See subsection XIIB7 for a simpler example.)

The S-matrix is then given in the usual way from $\Gamma[\varphi]$, after adding another gauge-fixing term for the *background* gauge invariance. Since the total S-matrix is given by just the trees following from treating Γ as a classical action, we need only a gauge-fixing term for the physical fields, and we can ignore background ghosts. (Of course, quantum ghosts were already used to calculate Γ.) This background gauge fixing is independent of the quantum gauge fixing. In particular, we can choose different quantum and background gauges: For example, when treating spontaneously broken gauge theories, it's often more convenient to choose a Fermi-Feynman quantum gauge and a unitary background gauge; i.e., we expand the background fields about the physical vacuum to make the physical states obvious, but leave the quantum fields unexpanded to avoid complicating the Feynman rules. This also avoids the complication of having to expand about the vacuum twice, since vacuum values get quantum corrections to those appearing in the classical action.

The gauge invariance of the effective action in the background-field formalism is a big advantage over other quantum gauges, where the effective action is only BRST invariant, since gauge invariance is a much stronger constraint than BRST invariance: Gauge symmetry is local, while BRST is only global. Thus, the background-field gauge produces a much simpler effective action. In other words, the background-field gauge produces an effective action without ghosts: Although we can drop ghost terms from the effective action in general, because there are no physical external ghost states (since we calculate only the "tree" graphs of the effective action), the result is not normally BRST invariant; but in the background-field gauge it is still BRST invariant, since it is gauge invariant. This means that the background-field gauge yields not only simpler results, but fewer calculations: Many terms can be determined by "gauge covariantization".

Exercise VIB8.3

Consider an effective action for Yang-Mills plus matter in a background-field gauge. Its gauge invariance can be used to derive "Ward-Takahashi identities". (These were originally expressed as properties of the S-matrix, but are much simpler to understand in terms of the effective action.)

a Show that the part of the effective action *quadratic* in the Yang-Mills fields, and independent of the matter fields, is invariant under the *Abelian* gauge transformations. (Hint: Taylor expand.) Thus, in such gauges the quantum correction to the gluon propagator is transverse.

b By the same method, find a relation between any quantum 3-point vertex coupling matter to Yang-Mills and the corresponding matter propagator correction. (Note a simpler case: Since the renormalization counterterms are local, gauge invariance just says that the coefficients of the two corresponding counterterms are the same, i.e., they occur in the combination $\sim \bar{\Psi}\nabla\!\!\!\!\nabla\Psi$.)

However, this does not mean we can completely ignore BRST and ghosts by using background-field gauges: Although the *effective* action is gauge invariant and ghost free, ghosts and BRST still appear in the (quantum-gauge-fixed) *classical* action. In practice this means, as far as calculating the Feynman graphs that contribute to the effective action, that in the background-field gauge calculations are about *one loop simpler* than in other gauges. For example, for one-loop graphs we effectively calculate free one-loop vacuum bubbles (including ghosts) covariantly coupled to background fields: There are fewer of the complications of nonabelian theories, since the quantum fields appear only as non-gauge fields with covariant couplings and no self-interactions. However, already at two loops we have self-interactions of the quantum fields, which include the same kinds of terms that would have appeared had we not used a background-field formalism.

Another complication is that BRST invariance is not as restrictive as gauge invariance: It can be shown that in general gauges at the quantum level BRST invariance is preserved only up to "wave-function renormalizations" (rescalings) of the quantum fields. However, in the background-field gauge wave-function renormalizations of the quantum fields can be ignored, since the quantum field is a dummy variable: There are no external quantum fields, so all such factors cancel. (Actually, we can also ignore wave function renormalization counterterms in non-background-field gauges, since when calculating S-matrix elements such divergences will be canceled by corresponding divergences in the external-line factors. In general, external-line normalization factors may be nontrivial even when wave-function renormalization is performed, depending on the renormalization scheme.)

An exception is Abelian gauge theories, such as QED: Because the gauge-invariant action for just the gauge fields is free, background field gauges are identical to ordinary gauges. Also, the ghosts decouple (for linear gauge conditions).

9. Nielsen-Kallosh

So far we have considered only gauges where the gauge-fixing term is the square of the gauge-fixing function. More generally, we'll need gauge-fixing terms of the form $f\mathcal{O}f$ for some operator $\mathcal{O}$. Straightforwardly, we can write

$$iQ\tfrac{1}{2}\tilde{C}[f(A) - \tfrac{1}{2}\alpha\mathcal{O}^{-1}B] = \tfrac{1}{2}B(f - \tfrac{1}{2}\alpha\mathcal{O}^{-1}B) + \tfrac{1}{2}i\tilde{C}\frac{\partial f}{\partial A}\cdot[\nabla, C]$$

However, B is no longer auxiliary, so we can't eliminate it by its field equation. But we can diagonalize the Lagrangian by the corresponding redefinition,

$$B \to B + \tfrac{1}{\alpha}\mathcal{O}f$$

(The Jacobian of such redefinitions is unity, the determinant of a triangular matrix of the form $\left(\begin{smallmatrix}1 & x\\ 0 & 1\end{smallmatrix}\right)$.) The gauge-fixing terms are then

$$-\tfrac{1}{4}\alpha B\mathcal{O}^{-1}B + \tfrac{1}{4\alpha}f\mathcal{O}f$$

The inverse operator is inconvenient for Feynman rules. We know that integrating out B gives a functional determinant, so $\mathcal{O}^{-1}$ can be replaced by an $\mathcal{O}$ if we change the statistics of the Nakanishi-Lautrup field. However, this is a bit formal, since technically $\mathcal{O}$ must be symmetric between the two B's, while it should be antisymmetric between two fermions.

A useful example is gauge fixing for super Yang-Mills in superspace. Gauge fixing for massless Yang-Mills is actually more difficult than for the massive (Higgs) case, considered in subsection VIB5. We'll look at the Abelian theory, to determine what kind of gauge fixing we need to define the propagator. (With slight generalization, this is also sufficient for the background-field gauge: See the following subsection.) In that case the BRST transformations are

$$QV = i(\bar{C} - C), \qquad Q\tilde{C} = -iB, \qquad Q\bar{\tilde{C}} = -i\bar{B}, \qquad QC = QB = Q\bar{C} = Q\bar{B} = 0$$

where C, $\tilde{C}$, and B are chiral. Using the result (the Abelian case of exercise IVC4.1)

$$W_\alpha = -i\bar{d}^2 d_\alpha V$$

the gauge-invariant kinetic term is (rearranging derivatives and using integration by parts; see subsection VIB5)

$$L_0 = -\int d^2\theta\ \tfrac{1}{2}W^\alpha W_\alpha = -\int d^4\theta\ \tfrac{1}{2}V d^\alpha \bar{d}^2 d_\alpha V = \int d^4\theta\ V(\tfrac{1}{4}\Box - \bar{d}^2 d^2)V$$

To gauge-fix to the Fermi-Feynman gauge we choose

$$L_1 = -iQ\int d^4\theta\ [(\tilde{C} + \bar{\tilde{C}})V + \tilde{C}(\tfrac{1}{2}\Box)^{-1}\bar{B}]$$

$$= \int d^4\theta\ [(\bar{\tilde{C}}C - \tilde{C}\bar{C}) - (B + \bar{B})V - B(\tfrac{1}{2}\Box)^{-1}\bar{B}]$$

(dropping $d^4\theta$ integrals of totally chiral or totally antichiral terms, which vanish). If we were to simply redefine B by

$$B \to B - \bar{d}^2 d^2 V, \qquad \bar{B} \to \bar{B} - d^2\bar{d}^2 V$$

the gauge-fixing terms would diagonalize as (using $\bar{d}^2 d^2 \bar{d}^2 = \frac{1}{2}\Box\bar{d}^2$)

$$-(B+\bar{B})V - B(\tfrac{1}{2}\Box)^{-1}\bar{B} \to V\bar{d}^2 d^2 V - B(\tfrac{1}{2}\Box)^{-1}\bar{B}$$

giving the desired result for V: At this stage the total result is

$$L = L_0 + L_1 \to \int d^4\theta\ [\tfrac{1}{4}V\Box V + \bar{\tilde{C}}C - \tilde{C}\bar{C} - B(\tfrac{1}{2}\Box)^{-1}\bar{B}]$$

Because B is complex, the replacement of B with a fermionic superfield can be performed classically, just like the rest of the gauge-fixing procedure. We thus introduce ghosts for a trivial gauge invariance as described in subsection VIA4:

$$QD = E, \qquad Q\bar{E} = -i\bar{D}, \qquad QE = Q\bar{D} = 0$$

We have treated the ghosts and their hermitian conjugates independently; alternatively, we can consider $\bar{D}$ and $\bar{E}$ as not being the conjugates of D and E. The gauge fixing is simply

$$L_2 = -iQ(-\bar{E}D) = \bar{D}D - i\bar{E}E$$

We next make the redefinition

$$D \to D + (\tfrac{1}{2}\Box)^{-1}\bar{d}^2\bar{B}, \qquad \bar{D} \to \bar{D} + (\tfrac{1}{2}\Box)^{-1}d^2 B$$

which has the effect

$$\int d^4\theta\ [\bar{D}D - B(\tfrac{1}{2}\Box)^{-1}\bar{B}] \to \int d^4\theta\ \bar{D}D + \int d^2\theta\ DB + \int d^2\bar{\theta}\ \bar{D}\bar{B}$$

which vanishes, after using the now-algebraic field equations from varying B and $\bar{B}$. Alternatively, we can make this field redefinition *instead* of the previous field redefinition: We then have the terms

$$\int d^4\theta\ [\bar{D}D - (B+\bar{B})V] + \int d^2\theta\ DB + \int d^2\bar{\theta}\ \bar{D}\bar{B} \to \int d^4\theta\ V\bar{d}^2 d^2 V$$

after using the still-algebraic B equations.

The net result

$$L = L_0 + L_1 + L_2 \to \int d^4\theta\ (\tfrac{1}{4}V\Box V + \bar{\tilde{C}}C - \tilde{C}\bar{C} - i\bar{E}E)$$

is that the original nonlocal B term has been replaced classically with the local $\bar{E}E$ term, which yields the same determinant upon quantization, but gives simple Feynman rules more directly. (The determinant is nontrivial in the background-field gauge. A similar procedure can be applied to gauge fixing for spin 3/2.)

Exercise VIB9.1

Perform the analogous quantization for the nonabelian case of pure super Yang-Mills (no matter), using the super Gervais-Neveu gauge. Compare with the limit $m \to 0$ of the model considered in subsection VIB5, and show the V part of the action agrees.

Exercise VIB9.2

Use this method to produce a gauge-fixing term $\alpha(n \cdot A)\Box(n \cdot A)$ for a gauge vector A in terms of a parameter α and constant vector n. Find all propagators. Look for simplifying special cases of α and n.

10. Super background-field

Although in principle the background-field formalism is the same for supersymmetric theories as nonsupersymmetric, there are some technical differences because of the nonlinearity in the prepotentials. (Similar remarks apply to nonlinear σ models.) The basic idea is that we want to expand the full covariant derivative in quantum fields about background-covariant derivatives: As for the nonsupersymmetric case, $\nabla \to \mathcal{D} + iA$, but now A is further expressed in terms of $\mathcal{D}$ and the prepotential because of the constraints. The generalization in this case (and for nonlinear σ models) is easy because the solution to the constraints makes the prepotentials appear as (complex) group elements: Because of the closure of group multiplication, we can write

$$\mathbf{g} \to \mathbf{g}_B \mathbf{g}_Q$$

in terms of quantum ($\mathbf{g}_Q$) and background ($\mathbf{g}_B$) group elements (fields). More explicitly, for our case we write (see subsection IVC4)

$$e^{\Omega} \to e^{\Omega_B} e^{\Omega_Q}$$

and thus for the covariant derivatives

$$\nabla_\alpha \to e^{-\Omega_Q} \mathcal{D}_\alpha e^{\Omega_Q}$$

absorbing the background prepotential completely into the background covariant derivative

$$\mathcal{D}_\alpha = e^{-\Omega_B} d_\alpha e^{\Omega_B}$$

In other words, as the name suggests, the full covariant derivative ∇ has been expanded about an arbitrary background, described by Ω_B. (This is even clearer in the supergravity case, where we simply replace the flat-space d_α with the curved-space

$\mathcal{D}_\alpha$, since d_α is more than a partial derivative, and already contains the flat-space part of the metric tensor.) For purposes of quantization, it is most convenient to go to a chiral representation for the quantum field. For the background field we need not be so specific, since it is hidden in the background covariant derivatives. The result is then

$$\nabla_\alpha \to e^{-V}\mathcal{D}_\alpha e^V, \qquad \overline{\nabla}_{\dot\alpha} \to \bar{\mathcal{D}}_{\dot\alpha}, \qquad \nabla_{\alpha\dot\alpha} \to i\{\bar{\mathcal{D}}_{\dot\alpha}, e^{-V}\mathcal{D}_\alpha e^V\}$$

where V is the quantum field.

Exercise VIB10.1

Solve the rest of the commutator algebra to find expressions for all the field strengths in terms of V and $\mathcal{D}_A$.

The rest of the quantization procedure then follows as for the nonsupersymmetric case, except for the Nielsen-Kallosh ghost described in the previous subsection. In particular, for the terms in the gauge-fixed classical action quadratic in the quantum field V,

$$W_\alpha \to -i\tfrac{1}{2}[\mathcal{D}^{\dot\alpha}, \{\mathcal{D}_{\dot\alpha}, e^{-V}\mathcal{D}_\alpha e^V\}] = \mathcal{W}_\alpha - i\bar{\mathcal{D}}^2\mathcal{D}_\alpha V + i\tfrac{1}{2}\bar{\mathcal{D}}^2[V,\mathcal{D}_\alpha V] + \mathcal{O}(V^3)$$

$$\Rightarrow \quad S_{2V} = \int dx\ d^4\theta\ V(-\tfrac{1}{2}\mathcal{D}^\alpha\bar{\mathcal{D}}^2\mathcal{D}_\alpha + i\tfrac{1}{2}\mathcal{W}^\alpha\mathcal{D}_\alpha + \bar{\mathcal{D}}^2\mathcal{D}^2)V$$

Pushing the $\mathcal{D}$ in the first term to the right, we find

$$-\tfrac{1}{2}\mathcal{D}^\alpha\bar{\mathcal{D}}^2\mathcal{D}_\alpha = i\tfrac{1}{4}(\mathcal{D}^{\alpha\dot\alpha}\bar{\mathcal{D}}_{\dot\alpha}\mathcal{D}_\alpha + \bar{\mathcal{D}}_{\dot\alpha}\mathcal{D}^{\alpha\dot\alpha}\mathcal{D}_\alpha) - \bar{\mathcal{D}}^2\mathcal{D}^2$$

Using integration by parts on all the derivatives in the second term so they act to the left, then switching the V's so they again act to the right,

$$\begin{aligned}\mathcal{D}^{\alpha\dot\alpha}\bar{\mathcal{D}}_{\dot\alpha}\mathcal{D}_\alpha + \bar{\mathcal{D}}_{\dot\alpha}\mathcal{D}^{\alpha\dot\alpha}\mathcal{D}_\alpha &\to \mathcal{D}^{\alpha\dot\alpha}\bar{\mathcal{D}}_{\dot\alpha}\mathcal{D}_\alpha + \mathcal{D}_\alpha\mathcal{D}^{\alpha\dot\alpha}\bar{\mathcal{D}}_{\dot\alpha} \\ &= \mathcal{D}^{\alpha\dot\alpha}\bar{\mathcal{D}}_{\dot\alpha}\mathcal{D}_\alpha + \mathcal{D}^{\alpha\dot\alpha}\mathcal{D}_\alpha\bar{\mathcal{D}}_{\dot\alpha} + [\mathcal{D}_\alpha, \mathcal{D}^{\alpha\dot\alpha}]\bar{\mathcal{D}}_{\dot\alpha} \\ &= -i\Box + 2\overline{\mathcal{W}}^{\dot\alpha}\bar{\mathcal{D}}_{\dot\alpha}\end{aligned}$$

where $\Box = \mathcal{D}^a\mathcal{D}_a$. The final result is similar to the bosonic case (exercise VIB8.1):

$$S_{2V} = \int dx\ d^4\theta\ \tfrac{1}{4}V(\Box + 2i\mathcal{W}^\alpha\mathcal{D}_\alpha + 2i\overline{\mathcal{W}}^{\dot\alpha}\bar{\mathcal{D}}_{\dot\alpha})V$$

(This result is invariant under integration by parts because of the Bianchi identity $\mathcal{D}_\alpha\mathcal{W}^\alpha + \bar{\mathcal{D}}_{\dot\alpha}\overline{\mathcal{W}}^{\dot\alpha} = 0$.)

Ghosts and matter are quantized straightforwardly: For matter we have

$$\overline{\nabla}_{\dot\alpha}\phi = \nabla_\alpha\bar\phi = 0 \quad \Rightarrow$$

$$\phi \to \varphi + \phi, \quad \bar{\phi} \to e^V(\bar{\varphi} + \bar{\phi}); \qquad \bar{\mathcal{D}}_{\dot{\alpha}}\varphi = \mathcal{D}_\alpha \bar{\varphi} = 0$$

The action thus looks the same as usual $((\bar{\varphi}+\bar{\phi})e^V(\varphi+\phi)$, etc.), except that all chiral superfields are now background-chiral. For the standard ghosts we have for the ghost action $S_C = \int dx\ d^4\theta\ L_C$ (remembering there are no background ghosts, and using the full nonlinear transformation law from exercise IVC4.3)

$$L_C = (\tilde{C} + \bar{\tilde{C}})\mathcal{L}_{V/2}[coth(\mathcal{L}_{V/2})(C - \bar{C}) + (C + \bar{C})] = (\tilde{C} + \bar{\tilde{C}})(C - \bar{C}) + \mathcal{O}(V)$$

$$\to (\bar{\tilde{C}}C - \tilde{C}\bar{C}) + \mathcal{O}(V)$$

the same as in non-background gauges, except again the ghosts are background-chiral. Now the Nielsen-Kallosh ghost of the previous subsection is nontrivial: We again have

$$L_{NK} = -i\bar{E}E$$

but these ghosts also are background-chiral. This means they contribute to the effective action *only* at one loop, through "vacuum bubbles".

REFERENCES

1 Fock, *loc. cit.* (VB);
Schwinger, *loc. cit.* (VB, ref. 2, second ref.):
radial gauge.

2 Fermi, *loc. cit.* (VC);
Feynman, *loc. cit.* (VB, ref. 4).

3 A.A. Abrikosov, I.M. Khalatnikov, and L.D. Landau, *Doklady Akad. Nauk USSR* **95** (1954) 773:
Landau gauge.

4 P. De Causmaecker, R. Gastmans, W. Troost, and T.T. Wu, *Nucl. Phys.* **B206** (1982) 53;
F.A. Berends, R. Kleiss, P. De Causmaecker, R. Gastmans, W. Troost, and T.T. Wu, *Nucl. Phys.* **B206** (1982) 61;
Z. Xu, D.-H. Zhang, and L. Chang, *Nucl. Phys.* **B291** (1987) 392;
J.F. Gunion and Z. Kunszt, *Phys. Lett.* **161** (1985) 333;
R. Kleiss and W.J. Sterling, *Nucl. Phys.* **B262** (1985) 235:
spinor helicity.

5 V.N. Gribov, *Nucl. Phys.* **B139** (1978) 1.

6 G. 't Hooft, *Nucl. Phys.* **B35** (1971) 167;
K. Fujikawa, B.W. Lee, and A.I. Sanda, *Phys. Rev.* **D6** (1972) 2923;
Y.-P. Yao, *Phys. Rev.* **D7** (1973) 1647:
renormalizable gauges.

7 Gervais and Neveu, *loc. cit.* (IVA).

8 G.F. Chew and M. Levinson, *Z. Phys. C* **20** (1983) 19:
proposed Feynman rules similar to anti-Gervais-Neveu gauge in a context related to string theory.

9 W. Siegel, hep-th/9502163, *Phys. Rev.* **D52** (1995) 1035:
super Gervais-Neveu.
10 M. Mangano and S.J. Parke, *Phys. Rep.* **200** (1991) 301:
review of modern methods for tree graphs.
11 E.T. Newman and R. Penrose, *J. Math. Phys.* **3** (1962) 566:
use of “null tetrad” in gravity, complex basis vectors in terms of constant spinors.
12 G. Chalmers and W. Siegel, hep-ph/9801220, *Phys. Rev.* **D59** (1999) 045013:
spacecone.
13 W. Siegel and S.J. Gates, Jr., *Nucl. Phys.* **B189** (1981) 295;
S. Mandelstam, *Nucl. Phys.* **B213** (1983) 149;
L. Brink, O. Lindgren, and B.E.W. Nilsson, *Nucl. Phys.* **B212** (1983) 401:
lightcone superfields.
14 Feynman, *loc. cit.* (VC, ref. 17);
DeWitt, *loc. cit.* (VIA);
J. Honerkamp, *Nucl. Phys.* **B48** (1972) 269;
G. ’t Hooft, The background field method in gauge field theories, in *Functional and probabilistic methods in quantum field theory*, proc. 12th Winter School of Theoretical Physics, Karpacz, Feb. 17-Mar. 2, 1975, v. 1, *Acta Univ. Wratislav.* **368** (1976) 345;
L.F. Abbott, *Nucl. Phys.* **B185** (1981) 189:
background-field gauge.
15 J.C. Ward, *Phys. Rev. Lett.* **78** (1950) 182;
Y. Takahashi, *Nuo. Cim.* **6** (1957) 370.
16 N.K. Nielsen, *Nucl. Phys.* **B140** (1978) 499;
R.E. Kallosh, *Nucl. Phys.* **B141** (1978) 141.
17 M.T. Grisaru, W. Siegel, and M. Roček, *Nucl. Phys.* **B159** (1979) 429:
super background-field gauge.

C. SCATTERING

We have seen how covariant expansions of the S-matrix can be based on various definitions of $\hbar$. Covariant expansions can also be based on spacetime quantum numbers: For example, we can perturb in mass; this is equivalent to adding low-energy corrections to the high-energy approximation. Also, the first-quantized version of the $\hbar$ expansion, which expands in powers of momenta, is effectively an expansion in inverse powers of mass (low-energy approximation).

The only other spacetime property of a particle is spin, or helicity for massless particles in D=4. It is possible to define expansions in terms of it by describing the leading order by a complex action. This violates semiclassical unitarity at that order; however, the loop expansion violates unitarity at tree order also, so the expansion is still useful as long as unitarity returns once the expansion has been summed. Furthermore, we have already seen that gauges where unitarity is not manifest have some advantages over unitary gauges. In particular, the Gervais-Neveu gauge uses a complex gauge condition.

1. Yang-Mills

We first consider calculations for massless theories; these are simpler than massive ones in D=4 because the little group of the Lorentz group is SO(D−2) instead of SO(D−1), and is thus Abelian: We can label the spin of a state by an integer or half-integer, the helicity, by use of the spacecone formalism. To simplify notation, we drop the transverse index ($p^t \to p$), and distinguish 4-momentum P from its transverse component p by using upper- and lower-case. We also use color ordering; i.e., we examine only planar diagrams for each permutation of external lines.

We begin by summarizing the spacecone rules for pure Yang-Mills found in subsection VIB6: The Lagrangian appearing in the action $S = g^{-2} tr \int dx\ L$, writing derivatives as momentum operators for later convenience, is

$$L = A^+(-\tfrac{1}{2}P^2)A^- + (\tfrac{p^-}{p}A^+)[A^+, pA^-] + (\tfrac{p^+}{p}A^-)[A^-, pA^+] + [A^+, pA^-]\tfrac{1}{p^2}[A^-, pA^+]$$

Twistor notation (see subsection IIB6) is used:

$$\langle pq\rangle = -\langle qp\rangle, \quad \langle pq\rangle\langle rs\rangle + \langle qr\rangle\langle ps\rangle + \langle rp\rangle\langle qs\rangle = 0, \quad \langle pq\rangle^* = [qp]$$

$$p^+ = \langle p-\rangle[-p], \quad p^- = \langle +p\rangle[p+], \quad p = \langle +p\rangle[-p], \quad \bar{p} = \langle p-\rangle[p+]$$

$$\langle +-\rangle = [-+] = 1$$

The propagator and vertices are read from L in the usual way, but in addition we have further simplification from the choice of external line factors

$$\epsilon_+ = \frac{[-p]}{\langle +p\rangle}, \qquad \epsilon_- = \frac{\langle +p\rangle}{[-p]}; \qquad \frac{p^-}{p}\epsilon_\oplus = \frac{p^+}{p}\epsilon_\ominus = 1$$

where $\oplus$ and $\ominus$ are the reference lines, with $+$ and $-$ helicity, respectively (not to be confused with the earlier notation for spinor indices $\alpha = (\oplus, \ominus)$). However, the reference *momenta* for helicities $\pm$ are taken from lines with helicities $\mp$:

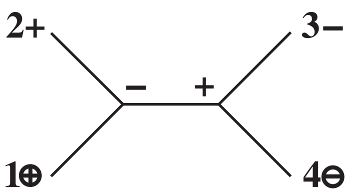
$$P_\oplus = |-\rangle[-|, \qquad P_\ominus = |+\rangle[+|$$

$$\Rightarrow \quad P_\oplus^a = \delta_-^a, \qquad P_\ominus^a = \delta_+^a$$

The reference external line factors occur only in the above combinations, because only 1 term of 1 of the 3-point vertices contributes to each.

The simplest examples are classes of diagrams that vanish by virtue of their "maximal helicity violation": By simple counting of $+$'s and $-$'s, we see that the tree graphs with the fewest external $-$'s, those with only self-dual vertices ($++-$), have a single external $-$. Thus the all $+$ amplitude vanishes automatically. Furthermore, the diagrams with a single external $-$ must have that line chosen as one of the reference lines. However, by the above rules that line can carry only the *anti*-self-dual vertex ($--+$), so those amplitudes also vanish.

The simplest nonvanishing amplitude is $++--$. We consider the case where the helicities are cyclically ordered as $++--$; we label them 1234, and choose 1 and 4 as the reference lines; this amplitude can be denoted as $\oplus+-\ominus$. ($P_4 = |+\rangle[+|$, $P_1 = |-\rangle[-|$: The positive-helicity reference line gives the reference momentum for negative helicity, and vice versa.) We label all external momenta as flowing inward. There are only three diagrams; however, the $+$ reference line uses only the $++-$ vertex, while the $-$ reference line uses only the $--+$ vertex, so the 4-point-vertex diagram vanishes, as does the diagram with both reference lines at the same vertex. Thus, we are left with only 1 graph.



Furthermore, we know that the 3-point vertices contribute only 1 term to the reference line, so this graph has only 1 term. This means we can immediately write down the answer (dropping the factors of $-g$ at each vertex):

$$\epsilon_{2+}\epsilon_{3-}p_2 p_3 \frac{1}{\frac{1}{2}(P_3+P_4)^2} = \frac{[-2]}{\langle +2\rangle}\frac{\langle +3\rangle}{[-3]}\langle +2\rangle[-2]\langle +3\rangle[-3]\frac{1}{\langle 34\rangle[34]}\frac{1}{\langle +-\rangle[-+]}$$

$$= \frac{[12]^2\langle 34\rangle}{[34][41]\langle 14\rangle}$$

where we have restored helicity and dimensions (trees go as $\langle\ \rangle^{2-E_+}[\ ]^{2-E_-}$), and used $p_1 = p_4 = 0$. We have omitted the usual group theory factor (see subsection VC9). (Note that the propagator is $-1/\frac{1}{2}P^2$, because of the signature for the $\pm$ spacecone components. This extra sign cancels that coming from the fact that one vertex has cyclic ordering and one anticyclic with respect to group theory, i.e., the commutators in the action.) Using the identities, following from overall momentum conservation,

$$(P_1+P_4)^2 = (P_2+P_3)^2 \quad \Rightarrow \quad [41]\langle 14\rangle = [23]\langle 32\rangle$$

$$\sum |p\rangle[p| = 0 \quad \Rightarrow \quad \langle 34\rangle[14] = -\langle 32\rangle[12]$$

this can be put in the standard form

$$\frac{[12]^4}{[12][23][34][41]}$$

Exercise VIC1.1

Using similar manipulations, cast it into the form

$$\frac{\langle 34\rangle^4}{\langle 12\rangle\langle 23\rangle\langle 34\rangle\langle 41\rangle}$$

Exercise VIC1.2

Repeat the calculation for the $+-+-$ (color-ordered) amplitude.

The corresponding differential cross section is very simple: Using

$$\langle pq\rangle^* = -[pq] \quad \Rightarrow \quad |\langle pq\rangle|^2 = |[pq]|^2 = -P\cdot Q$$

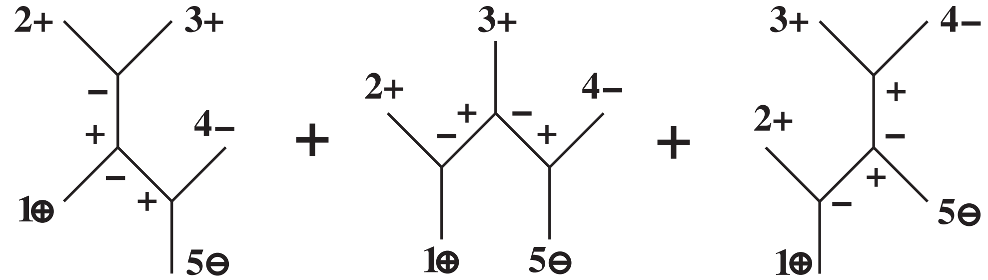
and momentum conservation, we find (after including the coupling g)

$$|T|^2 = g^4\frac{s^2}{t^2} \ \textit{or}\ g^4\frac{t^2}{s^2}$$

depending on the orientation of the diagram with respect to time, for this color-ordered contribution. (Depending on the color quantum numbers of the external states, this can be the only contribution.) Then (see subsection VC7)

$$\frac{d\sigma}{dt} = 2(2\pi)^3 g^4 \times \left(\frac{1}{t^2}\ \textit{or}\ \frac{t^2}{s^4}\right)$$

A more complicated example is the $+++--$ amplitude. Again taking color-ordered (planar) amplitudes, we choose the amplitude cyclically ordered as $+++--$ with lines labeled 12345, picking 1 and 5 as the reference lines, which we denote as $\oplus++-\ominus$. Again dropping all graphs with a reference line at a 4-point vertex or 2 references lines at a 3-point, all 5 graphs with a 4-point vertex are killed, and only 3 of the remaining 5 survive. (We also need to consider various combinations of $+$ and $-$ indices, but only 1 survives for each graph because of the chirality of 3-vertices with reference lines.)



Since 3-point vertices with (without) a reference line have 1 (2) terms, we are left with only 6 terms. The initial result for the amplitude is then

$$-\epsilon_{2+}\epsilon_{3+}\epsilon_{4-}\left[\frac{p_4^3\left(\frac{p_2^-}{p_2}-\frac{p_3^-}{p_3}\right)}{(P_2\cdot P_3)(P_4\cdot P_5)}-\frac{p_2p_4^2\left(\frac{p_2^-+1}{p_2}-\frac{p_3^-}{p_3}\right)}{(P_1\cdot P_2)(P_4\cdot P_5)}+\frac{p_2^2p_4\left(\frac{p_2^-+1}{p_2}-\frac{p_3^-}{p_3}\right)}{(P_1\cdot P_2)(P_3\cdot P_4)}\right]$$

where we have used the fact that the reference lines have trivial momenta: 1 for the component with $\pm$ index opposite to its helicity, 0 for the remaining components. The two terms for each diagram simplify to one, using

$$\frac{p^-}{p}=\frac{[p+]}{[-p]}\quad\Rightarrow\quad\frac{p_2^-}{p_2}-\frac{p_3^-}{p_3}=\frac{[2+][-3]-[3+][-2]}{[-2][-3]}=\frac{[23]}{[-2][-3]}$$

(applying the cyclic identity) with our normalization. Using this result, we find the similar result

$$\frac{p_2^-+1}{p_2}-\frac{p_3^-}{p_3}=\frac{\langle +-\rangle[-3]+\langle +2\rangle[23]}{\langle +2\rangle[-2][-3]}=\frac{\langle +4\rangle[34]}{\langle +2\rangle[-2][-3]}$$

applying momentum conservation. We next translate the momentum denominators into twistor notation, and also substitute the spacecone expressions for the polarizations and numerators. Canceling identical factors in numerator and denominator (but no further use of identities), the amplitude becomes ($+=5$, $-=1$)

$$\frac{\langle +4\rangle^3}{\langle +2\rangle\langle +3\rangle}\left(\frac{[-4]^2}{\langle 23\rangle[4+]}+\frac{[-4][34]}{\langle 2-\rangle[4+]}+\frac{\langle +2\rangle[-2]}{\langle 2-\rangle\langle 34\rangle}\right)$$

$$= \frac{\langle +4\rangle^3}{\langle +2\rangle\langle +3\rangle}\left(-\frac{\langle +2\rangle[-4]}{\langle 2-\rangle\langle 23\rangle} + \frac{\langle +2\rangle[-2]}{\langle 2-\rangle\langle 34\rangle}\right) = -\frac{\langle +4\rangle^3}{\langle 2-\rangle\langle 23\rangle\langle 34\rangle} = -\frac{\langle 45\rangle^4}{\langle 12\rangle\langle 23\rangle\langle 34\rangle\langle 45\rangle\langle 51\rangle}$$

applying momentum conservation twice, restoring normalization, and replacing the numerals for $\pm$.

Exercise VIC1.3

Using the spacecone gauge, evaluate all diagrams contributing to the six-point gluon (Yang-Mills) scattering tree amplitude (T-matrix) with color-ordered helicities $++++--$, that correspond to the symmetric diagram with a central 3-point vertex each of whose legs is connected to another 3-point vertex, each of which carries 2 of the external lines.

These results can be generalized to arbitrary (color-ordered) n-point tree amplitudes with two $-$ helicities, labeled i and j, and the rest $+$ ("Parke-Taylor amplitudes"): The result is (in an obvious notation), including now the coupling $(-g)^{n-2}$,

$$\langle +_1 \cdots +_{i-1} -_i +_{i+1} \cdots +_{j-1} -_j +_{j+1} \cdots +_n \rangle = g^{n-2}\frac{\langle ij\rangle^4}{\langle 12\rangle\langle 23\rangle\cdots\langle n-1,n\rangle\langle n1\rangle}$$

2. Recursion

A simple way to derive higher-point amplitudes is using the classical field equations. (See subsection VC3. In the literature, the field has often been mistaken for the current, since $\phi \sim \delta/\delta J$, $J \sim \delta/\delta\phi$. As usual, these are distinguished by the fact the field always has an external propagator, while the current has it amputated, since $K\phi + ... = -J$.) The steps are:

(1) Calculate the first few terms in the series (enumerated by the number of external lines).

(2) Guess the general result.

(3) Prove that it is correct by induction, using the classical field equations.

Of course, the second part is the hardest in general (at least when one simplifies the third step by using spacecone methods), and has been possible for just a couple of cases, only because the results for those cases are so simple. Since these results are for off-shell fields, and not S-matrix elements, they are gauge dependent: For example, if they are inserted into larger diagrams, the same choice of reference lines must be used.

The solution to the classical field equations is given by tree graphs with all external lines but one (the field itself) amputated and put on shell. (The usual external-line wave functions describe the asymptotic field, which is free.) The two cases with known solutions are those where all the on-shell lines have the same helicity, or one different. Note that the field $A^\pm$ has a $\mp$ associated with the opposite end of its external propagator. We then see in the former case, with all $+$'s on on-shell lines, that A^- vanishes because there are no fully-amputated diagrams, even off-shell, with only $+$'s externally (again counting $+$'s and $-$'s on vertices). Similarly, for the latter case, with only one $-$ on an on-shell line, we see that A^- has only $++-$ vertices; but setting that one on-shell $-$ to be a reference line (which by definition must be on-shell), it is not allowed such a vertex, so A^- vanishes also in this case. By similar reasoning, we see that A^+ in the former case consists entirely of $++-$ vertices; and in the latter case consists of all $++-$ except for one $--+$ (no $++--$), which must have the $-$ reference line directly attached.

The appearance of only the self-dual field (A^+) and almost only the self-dual vertex ($++-$) means that in both cases one is essentially solving equations in the self-dual theory: If we take just the kinetic term and $++-$ vertex from the action, and make the field redefinitions (see exercise VIB6.2)

$$A^+ = p\phi, \qquad A^- = p^{-1}\hat{\phi}$$

we obtain (after integration by parts and rearrangement inside the trace)

$$L_{+-,++-} = \hat{\phi}(\tfrac{1}{2}P^2\phi + [p^-\phi, p\phi])$$

These redefinitions make the $++-$ vertex local. $\hat{\phi}$ appears only as a Lagrange multiplier, and its variation gives the self-dual field equation

$$\tfrac{1}{2}\Box\phi + i(\partial^{\ominus\dot{\alpha}}\phi)(\partial^{\ominus}{}_{\dot{\alpha}}\phi) = 0$$

(which differs from the result of subsection IIIC5 by an i from the use of p instead of ∂ in the field redefinition, and $\oplus \to \ominus$ from the use of the spacecone instead of the lightcone).

We now consider in more detail the simpler (former) example (the one which does not directly give a nontrivial scattering amplitude). As a slight simplification, we look at the recursion relation for the field ϕ as defined in the self-dual theory. The recursion relation is now (see subsection VC3), scaling the coupling out of the kinetic term,

$$\phi(1,n) = -\frac{g}{\frac{1}{2}P^2(1,n)}\sum_{i=1}^{n-1}\phi(1,i)\phi(i+1,n)[p^-(1,i)p(i+1,n) - p(1,i)p^-(i+1,n)]$$

$$P(j,k) \equiv \sum_{m=j}^{k} P_m$$

where we again use color ordering, number the external lines cyclically, and $\phi(j,k)$ denotes the field with on-shell lines with momenta P_j through P_k. (Thus, on the left-hand side of the equation the field has n on-shell lines, while on the right-hand side the two fields have i and $n-i$.) Plugging in the twistor expressions for the vertex momenta, we find

$$p^-(1,i)p(i+1,n) - p(1,i)p^-(i+1,n) = \sum_{j=1}^{i}\sum_{k=i+1}^{n} \langle +j\rangle[jk]\langle +k\rangle$$

If we are clever we can guess the general result from explicit evaluation of the lower-order graphs; instead we find in the literature, after the above redefinition,

$$\phi(i,j) = (-g)^{N-1}\frac{1}{\langle +i\rangle\langle i,i+1\rangle\cdots\langle j-1,j\rangle\langle +j\rangle}$$

where N is the number of background momenta $(P_i, ..., P_j)$ for $\phi(i,j)$. For the initial-condition case $N=1$ this is simply the statement that the external line factor for ϕ is now

$$\epsilon_\phi = \frac{\epsilon_+}{p} = \frac{1}{\langle +p\rangle^2}$$

The induction hypothesis is also easy to check: The product of the two ϕ's from the induction hypothesis gives the desired result by itself up to a simple factor:

$$\phi(1,i)\phi(i+1,n) = -\tfrac{1}{g}\phi(1,n)\frac{\langle i,i+1\rangle}{\langle +i\rangle\langle +,i+1\rangle}$$

(The algebra of the color indices works as usual.) We then perform the sum over i before that over j and k (the complete sum is over all i,j,k with $1\le j\le i<k\le n$), making use of the identity

$$\frac{\langle ab\rangle}{\langle +a\rangle\langle +b\rangle} + \frac{\langle bc\rangle}{\langle +b\rangle\langle +c\rangle} = \frac{\langle ac\rangle}{\langle +a\rangle\langle +c\rangle} \quad\Rightarrow\quad \sum_{i=j}^{k-1}\frac{\langle i,i+1\rangle}{\langle +i\rangle\langle +,i+1\rangle} = \frac{\langle jk\rangle}{\langle +j\rangle\langle +k\rangle}$$

Multiplying this by the vertex momentum factor gives a sum over $j<k$ of $\langle jk\rangle[jk] = P_j\cdot P_k$, canceling the external propagator, yielding the desired result.

Exercise VIC2.1

Work out the analog of the above for the anti-selfdual case, paying careful attention to signs.

3. Fermions

We have seen in subsection VIB7 how these methods can be applied to massless spinors. Rather than applying the rules directly, in this subsection we examine the relation of the results in QCD to those in pure Yang-Mills theory. We also saw in subsection VIB7 how supersymmetry could be used to relate different QCD amplitudes. However, in practice supersymmetry relations give only a few useful relations, and only ones that can already be seen directly from the spacecone rules, which give more results than can be seen by supersymmetry alone.

The simplest relations that follow from supersymmetry are the vanishing of tree graphs with fewer than two negative helicities, which we saw in subsection VIC1 follows automatically from the spacecone rules. The remaining useful supersymmetry relation for tree graphs is the relation between Parke-Taylor amplitudes for pure Yang-Mills and those with one external line each of positive and negative helicity replaced with spinors or scalars. The easiest way to see this result is to make use of the conventions of the selfdual theory, as in the preceeding subsection. In Parke-Taylor amplitudes only one vertex is a non-selfdual vertex, which accounts for the simplicity of these amplitudes. (Tree amplitudes with only selfdual vertices vanish.) Furthermore, after transforming to the selfdual conventions, all (nonvanishing) selfdual vertices are identical — independent of spin. Furthermore, the nonselfdual 3-point vertex with one negative-helicity gluon chosen as a reference line (the only non-selfdual vertex we'll need for this relation) is independent of the spins of the remaining two lines. Consequently, the only difference between the two amplitudes we are relating comes from the difference in normalization of external line factors for gluons and quarks (and scalars).

We will not review the superspace formulation of selfdual supersymmetric theories here. The main features will be evident from the example of supersymmetric QCD that we now examine in more detail. The main result follows from treating the selfdual field of the nonsupersymmetric theory as a spacecone (or lightcone) superfield, as in subsection VIB7. Dimensional analysis then tells us that the field of helicity h has dimension $1-h$. The appropriate redefinitions of the spacecone fields are then

$$A^+ \to pA^+, \quad \psi^+ \to p\psi^+, \quad \phi \to \phi, \quad \psi^- \to \psi^-, \quad A^- \to \frac{1}{p}A^-$$

for the Yang-Mills fields $A^\pm$, spinors $\psi^\pm$, and scalars ϕ. The resulting external line factors are then simply

$$\langle +p \rangle^{-2h}$$

After these redefinitions, the kinetic terms, selfdual $(++-)$ vertices, and antiselfdual vertices for $-$ gluon reference line (referencing positive helicity) are

$$L_2 = A^+ \tfrac{1}{2} P^2 A^- + \psi^+ \tfrac{1}{2} P^2 \psi^-$$

$$L_{3,sd} = (p^- A^+)([pA^+, A^-] + \{p\psi^+, \psi^-\}) + (p^- \psi^+)[pA^+, \psi^-]$$

$$L_{3,\overline{sd},\oplus} = \left(\frac{p^+}{p^2} A^- \right) ([pA^+, A^-] + \{p\psi^+, \psi^-\})$$

for supersymmetric QCD. (In the A^3 term in the last line we have used integration by parts, and dropped a $(p^+/p)A^-$ term that vanishes for the reference line: There $(p^+/p^2)\epsilon_- = 1$ now, so $(p^+/p)\epsilon_- = 0$ vanishes for that line since $p \to 0$.)

Exercise VIC3.1

Apply these redefinitions to the full action for supersymmetric QCD given in subsection VIB7:

a Find the action and external line factors (especially for reference lines).

b Evaluate the 4-gluon tree amplitude for 2 positive and 2 negative helicities with these modified rules.

We now see easily that the terms L_2 and $L_{3,sd}$ that define the selfdual theory are independent of whether boson or fermion is chosen for the positive helicity fields and the negative helicity one (only the helicities of the fields must add up to 0 for L_2 and 1 for $L_{3,sd}$ for Lorentz invariance). Thus, supersymmetry is a much stronger restriction in a selfdual theory than a nonselfdual one. Finally, the current that couples to the reference line $(p^+/p^2)A^-$ is also the same for bosons and fermions. We therefore have, for example, the relation

$$(-, -\tfrac{1}{2}, +\tfrac{1}{2}, + \cdots +) = \frac{\langle 13 \rangle}{\langle 12 \rangle} (- - + \cdots +)$$

for the color-ordered tree amplitudes (where we have labeled helicities ± 1 by $\pm$). This follows from choosing line 1 as reference line $\ominus$ (for positive helicity, from a line with negative helicity). For example, from our result for the 4-gluon tree, we have the 2-quark, 2-gluon tree

$$(-, -\tfrac{1}{2}, +\tfrac{1}{2}, +) = \frac{\langle 12 \rangle^2 \langle 13 \rangle}{\langle 23 \rangle \langle 34 \rangle \langle 41 \rangle}$$

Exercise VIC3.2

Repeat these calculations using scalars in place of the spinors.

4. Masses

The spacecone formalism yields the simplest method for deriving S-matrix elements in massless theories (at least for trees; for loops it may be preferable to use background field gauges, with a Lorentz gauge, like Gervais-Neveu, for the quantum gauge and spacecone for the background gauge). The analogous method for the massive case is to use actions based on self-dual fields, as described in subsection IIIC4. The advantage of these two methods is that they use fields that are representations of the little group, so in the massive case fields have 2s+1 components and only undotted spinor indices (SO(3)=SU(2)), while in the massless case they have only 2 components and no indices (SO(2)=U(1)). Although the actions used are more complicated, this is just a reflection of the fact that algebra that is usually done repetitively in graphs has been performed once and for all in the action.

However, in the massive case the simplification is not as drastic as in the massless one: S-matrix elements are just simpler in massless theories, with many vanishing; the spacecone method takes advantage of this simplification in the final results by simplifying the intermediate steps. The massive examples we will consider in this subsection, taken from QED, are somewhat simple in any case, so we will stick to the older methods (although the uses of methods based on self-duality are still being explored).

The major difference in simplicity between the massless and massive cases (in any approach) is in the external line factors. The ambiguity in the explicit expressions for the external line factors is just the little group: In the massless case the solutions to the field equations (one solution and its complex conjugate) are unique up to a phase factor, which is why the twistor formalism is so useful. In the massive case the solutions (2s+1) are ambiguous up to an SU(2) transformation, which means they are messy for any choice. Just as in the massless case the twistor is part of the Lorentz transformation from an arbitrary frame to the lightcone frame, in the massive case the solutions are part of the transformation to the rest frame. In other words, the external line factors simply convert Lorentz indices to little-group indices; this makes indices trivial for the massless case (in D=4), and not as nice for the massive.

The result is that in practice whenever any of the external particles are massive their external line factors are left as implicit in S-matrix elements, and only their squares are explicitly evaluated, in cross sections. This was common in older experiments (especially QED), since recent experiments are mostly at energies so high that masses of external, stable particles are usually neglected. This adds to the algebra, since it means that Lorentz algebra is performed in each of n^2 terms in the cross section rather than n terms in the S-matrix.

Furthermore, the algebra is usually simplified by considering experiments where polarization is determined in neither the preparation of the initial states nor the measurement of the final states. This was also common in older experiments, when devices for polarization were not well developed. The result is that one averages over initial states and sums over final states, producing the same algebraic factors that appear in the propagator, as described in subsection VB3: Δ is replaced with Δ_+. One then applies the rules for Feynman diagrams for cross sections, as described in subsection VC7.

The standard S-matrices in QED are the 4-point tree graphs, with 2 3-point vertices and 1 internal propagator. There are just 2 graphs to consider, with various labelings of momenta: (1) The graph with 4 external fermions (electrons/positrons) connected by 1 internal photon describes both Møller (electron-electron) and Bhabha (electron-positron) scattering, 2 labelings each. (2) The graph with 2 external photons and 2 external fermions, as a continuous line that includes the 1 internal fermion, describes Compton (electron-photon) scattering as well as electron-positron creation/-annihilation, also 2 labelings each. In each case, the 1 S-matrix diagram results in 2 cross section diagrams, each with 2 momentum labelings (for a total of 2×2=4): 1 diagram from multiplying similar terms and 1 from cross-terms.

In Dirac spinor notation the Lagrangian for QED is (see subsection IIIA4)

$$\tfrac{1}{8}F^2 + \bar{\Psi}(i\partial\!\!\!/ - eA\!\!\!/ + \tfrac{m}{\sqrt{2}})\Psi$$

where we have scaled the "e" into the vertex. The Feynman rules are now (Fermi-Feynman gauge):

$$\begin{array}{ll} \text{Photon propagator:} & \eta_{ab}/\tfrac{1}{2}p^2 \\ \text{Fermion propagator:} & (p\!\!\!/ + \tfrac{m}{\sqrt{2}})/\tfrac{1}{2}(p^2+m^2) \\ \text{Vertex:} & e\gamma_a \end{array}$$

where we have used $p\!\!\!/^2 = -\frac{1}{2}p^2$. (We use the Fermi-Feynman-gauge propagator also for defining the cut propagator; ghosts decouple in QED.)

The cross section diagrams contain closed fermion loops, resulting in traces of products of γ matrices (with a -1 for each loop by Fermi-Dirac statistics). The algebra is manageable for the present case, using the 4D γ-matrix identities from subsection IIA6:

$$\gamma^a\gamma_a = -2, \qquad \gamma^a a\!\!\!/\gamma_a = a\!\!\!/, \qquad \gamma^a a\!\!\!/ b\!\!\!/\gamma_a = a\cdot b, \qquad \gamma^a a\!\!\!/ b\!\!\!/ c\!\!\!/\gamma_a = c\!\!\!/ b\!\!\!/ a\!\!\!/$$

$$tr(I) = 4, \qquad tr(\not{a}\not{b}) = -2a\cdot b, \qquad tr(\not{a}\not{b}\not{c}\not{d}) = a\cdot b\; c\cdot d + a\cdot d\; b\cdot c - a\cdot c\; b\cdot d$$

The traces encountered in the above processes are of the form

$$N_1 = tr(\gamma^a A\gamma^b B)tr(\gamma_a C\gamma_b D)$$

$$N_2 = tr(\gamma^a A\gamma_a B\gamma^b C\gamma_b D), \qquad N_3 = -tr(\gamma^a A\gamma^b B\gamma_a C\gamma_b D)$$

where $A = \not{a} + \frac{m}{\sqrt{2}}$, etc. Using the above identities, these are evaluated as

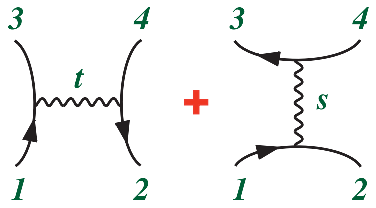
$$N_1 = 4m^4 + 2m^2(a\cdot b + c\cdot d) + 2(a\cdot c\; b\cdot d + a\cdot d\; b\cdot c)$$

$$N_2 = 4m^4 + m^2[2(a+c)\cdot(b+d) - a\cdot c - 4b\cdot d] + (a\cdot b\; c\cdot d + a\cdot d\; b\cdot c - a\cdot c\; b\cdot d)$$

$$N_3 = 2m^4 + m^2(a\cdot b + a\cdot c + a\cdot d + b\cdot c + b\cdot d + c\cdot d) + 2a\cdot c\; b\cdot d$$

Exercise VIC4.1

Generalize the above identities and expressions for the N's to arbitrary dimension D.

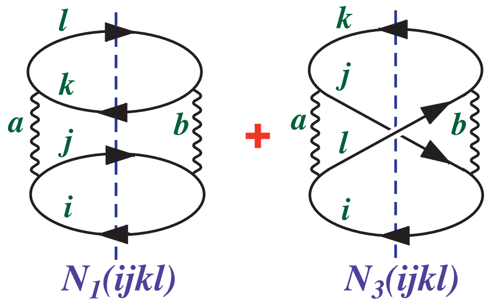


Our first example is $e^+e^- \to e^+e^-$ ("Bhabha scattering"). We have aligned all momenta to be that of the electrons (i.e., minus that of the positrons), so that all numerator factors are $\not{p} + \frac{m}{\sqrt{2}}$ without signs. Specifically, we have chosen p_1 for the (positive-energy) momentum for the incoming electron, $-p_2$ for the incoming positron, p_3 for the outgoing electron, and $-p_4$ for the outgoing positron. With these conventions,

$$-s = (p_1 - p_2)^2 = (p_3 - p_4)^2, \qquad -t = (p_1 - p_3)^2 = (p_2 - p_4)^2$$

$$-u = (p_1 + p_4)^2 = (p_2 + p_3)^2 \qquad (p_i^2 = -m^2, \qquad s + t + u = 4m^2)$$

$$\Rightarrow \quad p_1\cdot p_2 = p_3\cdot p_4 = \tfrac{1}{2}s - m^2, \quad p_1\cdot p_3 = p_2\cdot p_4 = \tfrac{1}{2}t - m^2, \quad p_1\cdot p_4 = p_2\cdot p_3 = -\tfrac{1}{2}u + m^2$$



For the squared amplitude we have for the average over initial polarizations and sum over final

$$\tfrac{1}{4}\sum_{pol}|T|^2 = \frac{N_1(1342)}{t^2} + \frac{N_1(1243)}{s^2} + \frac{N_3(1243)+N_3(1342)}{st}$$

$$= \tfrac{1}{2}\frac{f(s)+f(u)}{t^2} + \tfrac{1}{2}\frac{f(t)+f(u)}{s^2} + \frac{f(u)}{st}$$

not including the overall factor of e^4, where

$$f(x) \equiv (x-2m^2)(x-6m^2)$$

Every other N term is the result of switching $s \leftrightarrow t$ ($p_2 \leftrightarrow p_3$, or $p_1 \leftrightarrow p_4$) in the previous, since that is the relation of the 2 Feynman graphs contributing to the S-matrix. The N_1 terms are the squared diagrams, while the N_3's are the cross terms. The "$-$" in N_3 comes from Fermi-Dirac statistics, switching two fermion lines. (This is the same relative "$-$" as for 1 fermion loop vs. 2.)

Finally, from subsection VC7 we have the factors to get the differential cross section,

$$\frac{d\sigma}{dt} = \tfrac{1}{2}(2\pi)^3|T|^2\lambda_{12}^{-2}, \quad \lambda_{12}^2 = \tfrac{1}{4}[s-(m_1+m_2)^2][s-(m_1-m_2)^2]$$

so in this case

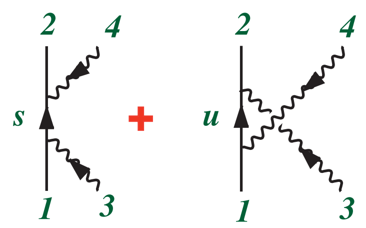
$$\left(\frac{d\sigma}{dt}\right)_{\text{Bhabha}} = \frac{(2\pi)^3e^4}{s(s-4m^2)}\left[\frac{f(s)+f(u)}{t^2} + \frac{f(t)+f(u)}{s^2} + 2\frac{f(u)}{st}\right]$$

The probabilities $|T|^2$ for $e^-e^- \to e^-e^-$ ("Møller scattering"), or $e^+e^+ \to e^+e^+$, are related by crossing symmetry $s \leftrightarrow u$ ($p_1 \leftrightarrow -p_3$ or $p_2 \leftrightarrow -p_4$):

$$\left(\frac{d\sigma}{dt}\right)_{\text{Møller}} = \frac{(2\pi)^3e^4}{s(s-4m^2)}\left[\frac{f(s)+f(u)}{t^2} + \frac{f(s)+f(t)}{u^2} + 2\frac{f(s)}{tu}\right]$$

A convenient frame for any of these cross sections is the center-of-mass frame (subsection IA4). In these cases all the external masses are equal, so the Mandelstam variables have simple expressions in terms of the energy (which is the same for all 4 particles) and the scattering angle:

$$s = 4E^2, \qquad t = -4(E^2-m^2)sin^2\tfrac{\theta}{2}, \qquad u = -4(E^2-m^2)cos^2\tfrac{\theta}{2}$$



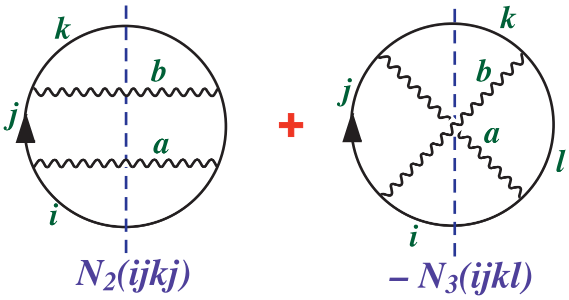
Another famous example is $e^-\gamma \to e^-\gamma$ ("Compton scattering"). Now we label p_1 for the incoming electron, p_3 for the incoming photon, p_2 for the outgoing electron, and $-p_4$ for the outgoing photon, so the Mandelstam variables are

$$-s = (p_1+p_3)^2 = (p_2-p_4)^2, \qquad -t = (p_1-p_2)^2 = (p_3+p_4)^2$$

$$-u = (p_1+p_4)^2 = (p_2-p_3)^2 \qquad (p_1^2 = p_2^2 = -m^2, \quad p_3^2 = p_4^2 = 0; \quad s+t+u = 2m^2)$$

$$\Rightarrow \quad p_1\cdot p_2 = \tfrac{1}{2}t - m^2, \qquad p_1\cdot p_3 = -\tfrac{1}{2}(s-m^2), \qquad p_1\cdot p_4 = -\tfrac{1}{2}(u-m^2)$$

$$p_2\cdot p_3 = \tfrac{1}{2}(u-m^2), \qquad p_2\cdot p_4 = \tfrac{1}{2}(s-m^2), \qquad p_3\cdot p_4 = -\tfrac{1}{2}t$$



The probability is

$$\tfrac{1}{4}\sum_{pol}|T|^2 = \frac{N_2(1,1+3,2,1+3)}{(s-m^2)^2} + \frac{N_2(1,1+4,2,1+4)}{(u-m^2)^2}$$

$$-\frac{N_3(1,1+4,2,1+3)+N_3(1,1+3,2,1+4)}{(s-m^2)(u-m^2)}$$

$$= \tfrac{1}{2}\frac{m^4+m^2(3s+u)-su}{(s-m^2)^2} + \tfrac{1}{2}\frac{m^4+m^2(3u+s)-su}{(u-m^2)^2} - \frac{m^2(t-4m^2)}{(s-m^2)(u-m^2)}$$

where now every other term comes from switching $s \leftrightarrow u$ ($p_3 \leftrightarrow p_4$), and the cross section is, after some rearrangement,

$$\left(\frac{d\sigma}{dt}\right)_{\text{Compton}} = \frac{(2\pi)^3 e^4}{(s-m^2)^2}\left[4m^4\left(\frac{1}{s-m^2}+\frac{1}{u-m^2}\right)^2 + 4m^2\left(\frac{1}{s-m^2}+\frac{1}{u-m^2}\right)\right.$$

$$\left. - \left(\frac{u-m^2}{s-m^2}+\frac{s-m^2}{u-m^2}\right)\right]$$

A useful frame is the lab frame (i.e., the rest frame of the electron), where in terms of the initial and final (positive) energies (E and E') and scattering angle of the photon we have

$$s = m^2+2mE, \quad u = m^2-2mE', \quad t = 2m(E'-E); \qquad \frac{1}{E'}-\frac{1}{E} = \frac{2\,sin^2\frac{\theta}{2}}{m}$$

By crossing symmetry, $s \leftrightarrow t$, we get $e^+e^- \to 2\gamma$ ("pair annihilation") and $2\gamma \to e^+e^-$ ("pair creation"):

$$\left(\frac{d\sigma}{dt}\right)_{\text{annihil.}} = \frac{(2\pi)^3 e^4}{s(s-4m^2)}\left[4m^4\left(\frac{1}{t-m^2}+\frac{1}{u-m^2}\right)^2 + 4m^2\left(\frac{1}{t-m^2}+\frac{1}{u-m^2}\right)\right.$$
$$\left. - \left(\frac{u-m^2}{t-m^2}+\frac{s-m^2}{t-m^2}\right)\right]$$

$$\left(\frac{d\sigma}{dt}\right)_{\text{creation}} = \frac{(2\pi)^3 e^4}{s^2}\left[4m^4\left(\frac{1}{t-m^2}+\frac{1}{u-m^2}\right)^2 + 4m^2\left(\frac{1}{t-m^2}+\frac{1}{u-m^2}\right)\right.$$
$$\left. - \left(\frac{u-m^2}{t-m^2}+\frac{s-m^2}{t-m^2}\right)\right]$$

Exercise VIC4.2

Calculate all the corresponding massless cross sections using the spacecone gauge. Show they agree with the $m = 0$ case of the above.

Exercise VIC4.3

Calculate all the above massive cross sections using scalars in place of the fermions.

Exercise VIC4.4

Calculate all the above massive cross sections replacing the photons with massless

a scalars

b pseudoscalars (with a γ_{-1} at the vertex).

5. Supergraphs

In supersymmetric theories the easiest way to calculate Feynman diagrams is in superspace. Supersymmetric cancelations are then automatic, and new special properties of supersymmetric theories are revealed. The derivation of the "supergraph" rules is similar to that of subsection VC1, except for some fine points in the treatment of chiral superfields. The path integral required the explicit evaluation of only a Gaussian for perturbation. Since we dropped proportionality constants, this was equivalent to substituting the solution to the classical, free field equations back into a quadratic action. For real scalar superfields (used for super Yang-Mills) this is trivial, but chiral scalar superfields (used for scalar multplets) satisfy the chirality constraint, and have superpotential terms: integrals over chiral superspace ($\int dx\ d^2\theta$), not the full superspace ($\int dx\ d^4\theta$).

We want to make use of the identity for evaluating the path integral (see subsection VC1)

$$\int \frac{du}{\sqrt{2\pi}}\ e^{-uMu/2} f(u+v) = \int \frac{du}{\sqrt{2\pi}}\ e^{-uMu/2} e^{u\partial_v} f(v) \sim e^{\partial_v M^{-1}\partial_v/2} f(v)$$

Then the "action" we need to integrate is

$$\tilde{S} = -\int dx\ d^4\theta\ \bar{\phi}\phi - \left[\int dx\ d^2\theta\ (\tfrac{m}{\sqrt{2}})\tfrac{1}{2}\phi^2 + h.c.\right] - \left(\int dx\ d^2\theta\ \phi\frac{\delta}{\delta\varphi} + h.c.\right)$$

consisting of the (derivative part of the) kinetic term, mass term, and (minus the) term that acts on $e^{-S_I[\varphi]}$. Solving the field equations (see subsection IVC2)

$$\bar{d}^2\bar{\phi} + \tfrac{m}{\sqrt{2}}\phi + \frac{\delta}{\delta\varphi} = d^2\phi + \tfrac{m}{\sqrt{2}}\bar{\phi} + \frac{\delta}{\delta\bar{\varphi}} = 0$$

we find

$$\phi = \frac{1}{\frac{1}{2}(-\Box + m^2)}\left(\bar{d}^2\frac{\delta}{\delta\bar{\varphi}} - \tfrac{m}{\sqrt{2}}\frac{\delta}{\delta\varphi}\right), \qquad \bar{\phi} = \frac{1}{\frac{1}{2}(-\Box + m^2)}\left(d^2\frac{\delta}{\delta\varphi} - \tfrac{m}{\sqrt{2}}\frac{\delta}{\delta\bar{\varphi}}\right)$$

The propagator exponent $\int \frac{1}{2}(\delta/\delta\varphi)(1/K)(\delta/\delta\varphi)$ thus becomes (putting back $\varphi \to \phi$)

$$\int dx\ d^4\theta\ \frac{\delta}{\delta\bar{\phi}}\left(-\frac{1}{\frac{1}{2}(-\Box + m^2)}\right)\frac{\delta}{\delta\phi} + \left[\int dx\ d^2\theta\ \tfrac{1}{2}\frac{\delta}{\delta\phi}\left(\tfrac{m}{\sqrt{2}}\frac{1}{\frac{1}{2}(-\Box + m^2)}\right)\frac{\delta}{\delta\phi} + h.c.\right]$$

Before writing the Feynman rules, we first note that functional differentiation with respect to a chiral superfield, as follows from the above variation, gives

$$\frac{\delta}{\delta\phi(x,\theta)}\phi(x',\theta') = \bar{d}^2\delta^4(\theta-\theta')\delta(x-x')$$

This means that there will be an extra $\bar{d}^2$ at the ϕ end of any chiral propagator and an extra d^2 at the $\bar{\phi}$ end. We could associate these directly with the propagator, but we will use one factor of $\bar{d}^2$ to convert a $\int d^2\theta$ into $\int d^4\theta$ at any superpotential vertex, and similarly for the complex conjugate. Therefore, we include such factors explicitly as a separate Feynman rule for the ends of chiral propagators. However, this means the $\phi\phi$ propagator (and similarly for $\bar{\phi}\bar{\phi}$) gets an extra factor of $d^2/\frac{1}{2}\Box$ to compensate for the fact that we include two $\bar{d}^2$ factors, whereas it really had only one because its integral was only $d^2\theta$. Furthermore, we Fourier transform x as usual, but not θ, basically because there is no translation invariance in θ, but also for a better reason to be explained soon. The Feynman rules of subsection VC4 are then modified as:

(A2$\frac{1}{2}$) Theta's: one for each vertex, with an $\int d^4\theta$.

(A3′) Propagators:

$$
\begin{aligned}
VV: &\qquad \frac{1}{\frac{1}{2}(p^2+m^2)}\delta^4(\theta-\theta') \\
\bar{\phi}\phi: &\qquad -\frac{1}{\frac{1}{2}(p^2+m^2)}\delta^4(\theta-\theta') \\
\phi\phi: &\qquad \frac{m}{\sqrt{2}}\left(\frac{d^2}{-\frac{1}{2}p^2}\right)\frac{1}{\frac{1}{2}(p^2+m^2)}\delta^4(\theta-\theta') \\
\bar{\phi}\bar{\phi}: &\qquad \frac{m}{\sqrt{2}}\left(\frac{\bar{d}^2}{-\frac{1}{2}p^2}\right)\frac{1}{\frac{1}{2}(p^2+m^2)}\delta^4(\theta-\theta')
\end{aligned}
$$

(A4$\frac{1}{2}$) Chiral vertex factors: $\bar{d}^2$ on the ϕ end(s) of every chiral propagator, d^2 on the $\bar{\phi}$ end(s), but drop any one such factor at superpotential vertex.

We next explain how θ integrations are performed on any connected graph. Consider any two vertices directly connected by a propagator. All the spinor derivatives acting on its $\delta^4(\theta-\theta')$ can be removed from that propagator by integration by parts. We then can use that δ function to trivially integrate over θ', removing the $\int d^4\theta'$ and that $\delta^4(\theta-\theta')$, and replacing θ' everywhere with θ. Effectively, those two vertices have been contracted to the same point in θ space, eliminating that propagator as far as θ dependence is concerned. We can repeat this procedure until *all* vertices are contracted to a single point. However, we are then left with a "tadpole" for each loop: Contracting propagators this way sequentially around a loop identifies all the vertices of that loop, and leaves the loop as a single propagator with both ends at that point. To evaluate this tadpole, we note that

$$[\bar{d}^2 d^2 \delta^4(\theta-\theta')]|_{\theta'=\theta} = 1$$

(θ' derivatives can be converted into minus θ derivatives when acting directly on the δ; this is basically integration by parts.) Fewer derivatives give 0; more can be reduced to terms of 4 or less. This completes all the evaluation in θ space, leaving an expression in terms of fields (some with d's acting on them) with different momenta, times the usual momentum factors, with the usual momentum integrals, but all at the same θ, with a single $\int d^4\theta$. This means that the generating functionals W and Γ are completely local in θ.

There is a further consequence of this evaluation. We have obtained terms with $\int d^4\theta$, but none with $\int d^2\theta$. However, to do it we had to introduce the factors $d^2/(-\frac{1}{2}p^2)$ into the $\phi\phi$ propagators. On the other hand, such a factor can easily be killed by a $\bar{d}^2$ from a vertex: We sandwich the d^2 between a $\bar{d}^2$ from each vertex, using the identity $\bar{d}^2 d^2 \bar{d}^2 = -\frac{1}{2}p^2\bar{d}^2$, and return the $\bar{d}^2$ to one vertex. The only time we can't do that everywhere is if *every* vertex is a superpotential (so every propagator in the graph is $\phi\phi$ and every external field is ϕ), since otherwise we can inductively borrow $\bar{d}^2$'s from some non-$\int d^2\theta$ vertex. Any such 1PI graph always vanishes, because there are exactly enough d's left to make the tadpoles nonvanishing, leaving an $\int d^4\theta$ of a product of ϕ's with no d's, which vanishes. On the other hand, for a tree graph there is exactly one d^2/p^2 left, which converts the $\int d^4\theta$ to an $\int d^2\theta$.

The net result is that not only are W and Γ local in θ, but only their classical parts can have $\int d^2\theta$ terms, and the spurious d^2/p^2 factors (which should not appear in massive theories) are always canceled. In particular, this implies that all UV divergences are $\int d^4\theta$ terms: All terms in the superpotential are unrenormalized (no loop corrections) to all orders in perturbation theory.

Exercise VIC5.1

Calculate all the contributions to $W[\phi,\bar{\phi}]$ from 4-point trees in massive super-ϕ^3 theory, and write the result in both p- and x-space (in analogy to the nonsupersymmetric example at the end of subsection VC4).

Improvements again result from using background-field gauges. We have already seen in subsection VIB10 the modification to the quantization for supersymmetric background-field gauges. The background-field expansion can be defined by solving the constraints on the full covariant derivatives in terms of quantum prepotentials but background *potentials* (A_A, not V), essentially by covariantizing d_A to the background ∇_A. Then ∇_α can be manipulated (integration by parts, etc.) in the graphs in the same way as d_α was, leaving only A_a (not A_α) and its derivatives (W_α, etc.) as background fields. This leads to improved power counting, and can be used to prove "superrenormalizability" (finiteness beyond one loop) for N=2 extended supersymmetric theories, and finiteness for N=4.

As for other gauges (background-)chiral superfields need special treatment, now to get the most out of background gauge invariance. Variation can be defined in the obvious way, but now we also need the covariantized identity

$$\bar{\mathcal{D}}^2\mathcal{D}^2\phi = \tfrac{1}{2}(\Box + i[\mathcal{W}^\alpha, \mathcal{D}_\alpha])\phi$$

from pushing the $\bar{\mathcal{D}}$'s to the right and using the commutation relations. The functional integral over the quantum background-chiral superfields can also be performed in the same way as for other gauges, the only modifications being background covariantization (including the above "nonminimal" term for $\Box$), and the fact that we can no longer neglect the "vacuum" contribution (one-loop diagrams with only background fields externally). Specifically, if we look at the general derivation of the Feynman rules in subsection VC1, we see it gave rules for all graphs except the one-loop vacuum bubble, since this graph has no (quantum) vertices. These rules, as adapted to superspace earlier in this subsection, are now modified only by the covariantization just discussed, which only adds background potentials (not prepotentials) and field strengths to propagators and vertices. The background-covariantized propagators then can be further expanded about the free $\square$. The net result is that in *all* diagrams *except* (perhaps) these chiral one-loop vacuum bubbles the background fields appear only in the form of potentials and field strengths. These vacuum bubbles then can be evaluated by the usual methods, since the formerly neglected Gaussian path integral for these "quantum-free" fields is just the usual one-loop path integral for a chiral superfield with external Yang-Mills superfields, only the external fields are now identified as background instead of quantum. In some cases, this last calculation can be further simplified to again yield an expression directly in terms of potentials without explicit prepotentials (see subsection VIIIA6 below).

REFERENCES

1 M.T. Grisaru, H.N. Pendleton, and P. van Nieuwenhuizen, *Phys. Rev. Lett.* **15** (1977) 996;
M.T. Grisaru and H.N. Pendleton, *Nucl. Phys.* **B124** (1977) 333:
supersymmetry identities for trees.

2 M.T. Grisaru and W. Siegel, *Phys. Lett.* **110B** (1982) 49:
generalization to one loop.

3 S.J. Parke and T. Taylor, *Nucl. Phys.* **B269** (1986) 410, *Phys. Rev. Lett.* **56** (1986) 2459;
F.A. Berends and W.T. Giele, *Nucl. Phys.* **B306** (1988) 759:
Parke-Taylor amplitudes.

4 W.A. Bardeen, *Prog. Theor. Phys. Suppl.* **123** (1996) 1;
D. Cangemi, hep-th/9605208, *Nucl. Phys.* **B484** (1997) 521;
G. Chalmers and W. Siegel, hep-th/9606061, *Phys. Rev.* **D54** (1996) 7628:
relation of Parke-Taylor amplitudes to self-dual Yang-Mills.

5 R. Brooks (May, 1993), unpublished;
N. Berkovits and W. Siegel, hep-th/9703154, *Nucl. Phys.* **B505** (1997) 139:
use of first-order actions with self-dual auxiliary fields for perturbation about self-duality.

6 Feynman, *loc. cit.* (VB, ref. 4):
Feynman diagrams for QED, from the horse's mouth. The original derivation, not counting his later-published first-quantized path-integral approach. Basically a mechanics point of view.

7 H.J. Bhabha, *Proc. Roy. Soc.* **A154** (1936) 195.

8 C. Møller, *Ann. Phys.* **14** (1932) 531.

9 O. Klein and Y. Nishina, *Z. Phys.* **52** (1929) 853:
Compton scattering.

10 A. Salam and J. Strathdee, *Phys. Rev.* **D11** (1975) 1521;
K. Fujikawa and W. Lang, *Nucl. Phys.* **B88** (1975) 61;
J. Honerkamp, M. Schlindwein, F. Krause, and M. Scheunert, *Nucl. Phys.* **B95** (1975) 397;
S. Ferrara and O. Piguet, *Nucl. Phys.* **B93** (1975) 261;
D.M. Capper, *Nuo. Cim.* **25A** (1975) 259;
R. Delbourgo, *Nuo. Cim.* **25A** (1975) 646, *J. Phys. G* **1** (1975) 800;
R. Delbourgo and M. Ramón Medrano, *Nucl. Phys.* **B110** (1976) 473;
W. Siegel, *Phys. Lett.* **84B** (1979) 193, 197:
early supergraphs.

11 Wess and Zumino, *loc. cit.* (IVC, ref. 2);
J. Iliopoulos and B. Zumino, *Nucl. Phys.* **B76** (1974) 310;
Capper, *loc. cit.*;
Delbourgo, *loc.cit.* (first ref. above);
B. Zumino, *Nucl. Phys.* **B89** (1975) 535;
P.C. West, *Nucl. Phys.* **B106** (1976) 219;
D.M. Capper and M. Ramón Medrano, *J. Phys. G* **2** (1976) 269;
S. Weinberg, *Phys. Lett.* **62B** (1976) 111:
nonrenormalization theorems for chiral superfields, from components or old-fashioned supergraphs.

12 Grisaru, Siegel, and Roček, *loc. cit.* (VB, ref. 17);
M.T. Grisaru and W. Siegel, *Nucl. Phys.* **B201** (1982) 292;
Gates, Grisaru, Roček, and Siegel, *loc. cit.*;
M.T. Grisaru and D. Zanon, *Phys. Lett.* **142B** (1984) 359, *Nucl. Phys.* **B252** (1985) 578, 591:
supergraphs as done today; more general nonrenormalization theorems.

VII. LOOPS

Although our analysis so far is sufficient to evaluate the lowest-order term in the $\hbar$ expansion ("trees"), certain new features arise at higher orders.

A. GENERAL

When infinities were first found in perturbative quantum field theory, they were thought to be a serious problem. A prescription can be given to remove these infinities, called "regularization". It was later shown that this regularization can be defined in such a way as to preserve all the desirable physical properties of the theory, called "renormalization". Unfortunately, it seems that the original infinities, exiled by renormalization from any finite order of perturbation theory, return to plague field theory when all orders of perturbation theory are summed. Therefore, renormalization should not be considered a cure to the disease of infinities, but only a treatment that allows divergent theories to be more useful.

1. Dimensional renormalization

"Perturbative renormalization" is defined to preserve the three properties that define relativistic quantum field theory (Poincaré invariance, unitarity, causality). The most general prescription is to start with a classical theory that is causal, Poincaré invariant, and satisfies the semiclassical part of unitarity (as described in subsection VC5). This gives the tree graphs of the theory. One then applies unitarity to define a perturbation expansion, determining the higher orders (loop diagrams) from the lowest (trees). Although the usual loop diagrams are divergent, there is enough ambiguity in the unitarity condition to allow removal of the divergences.

The only practical way to implement this procedure is to slightly modify the divergent graphs one obtains from the naive use of the Feynman graph rules (obtained, e.g., from path integral quantization of a classical action). The steps are: (1) "Regularize" each divergent graph by modifying the momentum integrals, introducing a parameter(s), the "regulator(s)", giving a finite result that reproduces the original divergent integral in a certain limit. (2) "Renormalize" each regularized graph by subtracting out the "divergent part" of the regularized graph, keeping only the "finite part". Once the graph has been rendered finite, the regulator can be dropped.

One then has to check that the method of removing divergences, order-by-order in the perturbation expansion, preserves the three properties of relativistic quantum field theory. The easiest way to do this is to use both a regularization scheme and a subtraction scheme that preserve these properties manifestly. The standard way for the subtraction scheme to do this is to change the coefficients of terms in the *classical* action by (real) constants that depend on both the regulator and $\hbar$. Since the classical action already satisfies Poincaré invariance, causality, and the semiclassical part of unitarity, this will automatically preserve all the desired properties. In this manner of renormalization, there thus remains only two steps to prove that a theory can be renormalized: (1) the existence of a regularization that manifestly preserves the three properties, and (2) that the modification of the action by making the coefficients regulator- and $\hbar$-dependent is sufficient to cancel all divergences that might reappear in the limit as the regulator is removed.

The latter step, discussed in subsection VIIA5 below, can be further divided into substeps, proving: (a) The ultraviolet divergence in any graph corresponding to scaling all internal (integration) momenta to infinity (the "superficial" divergence) comes from a term in that graph polynomial in external momenta, and can therefore be canceled by a local term from the action; and (b) recursively in the number L of loops, if the renormalization procedure has already been successfully applied through $L-1$ loops, the resulting modification of the action is sufficient to cancel all divergences appearing at L loops *except* the superficial ones. The former substep is the one that detemines whether the theory can be renormalized.

The former step is satisfied by dimensional regularization, the standard method of regularization in relativistic quantum field theory (and for practical purposes beyond one loop, the only one). It is defined by writing the theory under consideration in arbitrary dimensions D, and treating integrals over loop momenta as analytic functions of D. These integrals are then analytically continued from lower D, where they are (ultraviolet) convergent. The resulting expressions have pole singularities in D at integer D, so these poles can be subtracted out as the divergent parts.

There are two main reasons why dimensional regularization is so useful: (1) Most classical actions can be written as easily in arbitrary dimensions as in D=4. (The important exception is those that in some way involve the Levi-Civita tensor $\epsilon_{ab...c}$. The difficulty with such theories is not a drawback of dimensional regularization, but a general property of quantum field theory, and is related to the quantum breakdown of classical symmetries, to be discussed later.) In particular, this means it manifestly preserves gauge invariance (which is a part of unitarity), the property of relativistic quantum field theory most difficult to preserve. It is also the only workable scheme of regularization to do so.

(2) It requires only one regulator, the number of dimensions D itself. (Most other regularization schemes require at least one regulator for each loop.) This is the main reason why this scheme is the only practical method of regularization at higher loops. (An enormous number of regularization schemes have been proposed, almost all of which work well at one loop, but all of them are more difficult than dimensional regularization already at two loops, and even worse at three loops and higher.)

The renormalization scheme based on this regularization is also very simple: Defining $D = D_0 - 2\epsilon$ (where D_0 is the physical dimension, usually 4), we can Laurent expand any L-loop amplitude in ϵ, starting at $1/\epsilon^L$. These $1/\epsilon^n$ terms arise from n or more divergent D-momentum integrals. If we cancel all the negative powers of ϵ, we can take the limit $D \to D_0$ ($\epsilon \to 0$) by just dropping all the positive powers of ϵ, i.e., evaluating the remainder at $\epsilon = 0$. The procedure is to modify the coefficients of the terms in the "classical" action (couplings, masses, and field normalizations) by making them ϵ- and $\hbar$-dependent, giving them $\hbar^{L-1}/\epsilon^n$ ("counter")terms. Such terms can cancel any local divergence at L loops. One then has to show that they also cancel all nonlocal divergences at higher loops resulting from this divergence appearing in an L-loop subgraph. Thus, the procedure is recursive: (1) apply the counterterms obtained from calculations at less than L loops to cancel subdivergences; (2) cancel the remaining, local, superficial divergences by introducing new L-loop counterterms. The form of the superficial divergence can be determined by evaluating the divergence coming from the region of momentum space where all loop momenta go to infinity at the same rate. The superficial divergence is determined by $1/\epsilon$ terms of this loop and of subloop divergences; however, if the $1/\epsilon$ piece of a prospective counterterm vanishes at a certain loop order, so do all higher powers at that loop order. Thus, simple power counting (as well as global and local symmetries) is sufficient to determine what counterterms can appear at any particular loop order.

These rules are sufficient for evaluating momentum integrals to the point where renormalization can be applied. However, further simplifications are possible where spin is involved: Techniques specific to D=4 are useful to simplify algebra in general, and required to preserve manifest supersymmetry in particular. These methods treat spin indices as 4D, in contrast to the vector indices on momenta (and coordinates), which are analytically continued away from D=4 by the definition of dimensional regularization. This is natural in 4D N=1 supersymmetric theories formulated in superspace (or 4D N>1 in N=1 superspace), since there spins ≤ 1 are described by scalar prepotentials: There the simple prescription is to continue in the dimension of the commuting coordinates (x), while fixing the dimension of the anticommuting coordinates (θ). These rules, for either supersymmetric or nonsupersymmetric theories, have a simple physical interpretation for integer D<4: dimensional reduction. The reduced theories differ from those produced by simple dimensional regularization: Vectors get (4−D) extra scalars; spinors may become multiple spinors. For supersymmetric theories this is again natural, since vector and scalar multiplets remain irreducible after dimensional reduction. Unfortunately, vectors in nonsupersymmetric theories reduce to vectors plus scalars that are not related by any symmetry, so their renormalization is independent. However, the complications produced by the extra renormalizations are usually smaller than the algebraic simplifications resulting from the restriction to 4D spin algebra, especially for lower loops. Another complication is that Levi-Civita (ϵ) tensors can't be treated consistently in the dimensional reduction scheme. Although serious in principle, in practice this is not a problem as long as axial anomalies cancel, which is required anyway for unitarity. (See subsection VIIIB3. Also, axial anomalies are easier to calculate using Pauli-Villars regularization than with any form of dimensional regularization.)

Exercise VIIA1.1

Consider the identity

$$\hat{\delta}^a_{[b} p_c q_d r_e s_{f]} = 0$$

which holds in $D < 5$, where $\hat{\delta}$ is a D-dimensional Kronecker δ, as appears from momentum integrals (since momenta themselves are D-dimensional by definition), and p, q, r, s are momenta. Show by index contraction that an inconsistency arises when trying to analytic continue away from $D = 4$. This difficulty in defining totally antisymmetrized D-dimensional objects is why Levi-Civita tensors don't exist in dimensional regularization.

2. Momentum integration

The first step in performing momentum integration is to make all integrals Gaussian by exponentiating propagators using Schwinger parameters

$$\frac{1}{\frac{1}{2}(p^2+m^2)} = \int_0^\infty d\tau \ e^{-\tau(p^2+m^2)/2}$$

The general momentum integral in an arbitrary Feynman diagram is then

$$\mathcal{A} = N(p, -i\partial_x) \int \frac{d^{LD}k}{(2\pi)^{LD/2}} \int d^P\tau \ e^{-k^T\cdot A(\tau)k/2 - k^T\cdot B(\tau,p) + ik^T\cdot x - C(\tau,p,m)}\bigg|_{x=0}$$

where A, B, C are first-order in τ; B is first-order in p while C is second-order in p, m; and L is the number of loops and P the number of propagators. Also, we have used matrix notation with respect to the L-dimensional loop-space, with respect to which A is a matrix, B is a vector, and C is a scalar. Finally, N represents all "numerator" factors (propagator numerators and vertices; everything but propagator denominators) and has been brought outside the integral by Fourier transformation (i.e., introducing x to produce a generating functional for all numerators). The momentum integrals are now Gaussian, and can be evaluated by the methods of subsection IB3 (and VA2):

$$\mathcal{A} = N(p, -i\partial_x) \int d^P\tau \ (det \ A)^{-D/2} e^{(B-ix)^T A^{-1}(B-ix)/2-C}\Big|_{x=0}$$

Some of the τ integrations can be performed by various scalings of subsets of the τ's. For example, to see the superficial divergence of the graph we scale all of the τ's and insert the identity as

$$1 = \int_0^\infty d\lambda \ \delta\left(\lambda - \sum_i \tau_i\right), \qquad \tau_i = \lambda\alpha_i$$

where α are "Feynman parameters". The amplitude then becomes

$$\mathcal{A} = N(p, -i\partial_x) \int d\lambda \ \lambda^{P-1-LD/2} \ d^P\alpha \ \delta\left(1 - \sum \alpha\right) [det \ A(\alpha)]^{-D/2}$$
$$\times e^{\lambda[B(\alpha)^T A(\alpha)^{-1} B(\alpha)/2 - C(\alpha)] - iB(\alpha)^T \cdot A(\alpha)^{-1} x - x^T \cdot A(\alpha)^{-1} x/2\lambda}\Big|_{x=0}$$

(This method of introducing Feynman parameters is equivalent to directly changing variables from τ's to λ and one less α, and finding the Jacobian.)

The x derivatives in N must now be taken. For a contribution from these derivatives of order λ^{-n}, we have λ integrals of the form

$$\int d\lambda \ \lambda^{P-1-LD/2-n} e^{\lambda[B^T A^{-1} B/2 - C]} = \Gamma(P - \tfrac{1}{2}LD - n)[C - \tfrac{1}{2}B^T A^{-1} B]^{-P+LD/2+n}$$

where we have used the definition of the Γ function

$$\Gamma(z) = \int_0^\infty d\lambda \ \lambda^{z-1} e^{-\lambda}$$

This integral converges only for $Re \ z > 0$, but we can extend it to (almost) all z by analytic continuation: Using integration by parts,

$$z\Gamma(z) = \int_0^\infty d\lambda \ e^{-\lambda} \frac{d}{d\lambda} \lambda^z = (e^{-\lambda}\lambda^z)|_0^\infty + \int_0^\infty d\lambda \ \lambda^z e^{-\lambda} = \Gamma(z+1)$$

in the convergent region. Analytic continuation then says to evaluate the integral for $\Gamma(z)$ in the region $0 \geq Re \ z > -1$ as $\Gamma(z+1)/z$ in terms of the integral for $\Gamma(z+1)$, and so on:

$$\int_0^\infty d\lambda \ \lambda^{z-1} e^{-\lambda} = \left(\prod_{k=0}^{n} \frac{1}{z+k}\right) \int_0^\infty d\lambda \ \lambda^{z+n} e^{-\lambda}$$

Thus, $\Gamma(z)$ has simple poles in z at all the nonpositive integers.

Exercise VIIA2.1

Using the identity $\Gamma(z+1) = z\Gamma(z)$, derive the following special cases for nonnegative integer n:

$$\Gamma(n+1) = n!, \qquad \Gamma(n+\tfrac{1}{2}) = (n-\tfrac{1}{2})(n-\tfrac{3}{2})...\tfrac{1}{2}\sqrt{\pi} = \frac{(2n)!}{n!2^{2n}}\sqrt{\pi}$$

Feynman parameter integrations give more complicated functions. A simple but common example is the Beta function

$$B(x,y) = \int_0^1 d\alpha\ \alpha^{x-1}(1-\alpha)^{y-1} = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

The latter expression for the Beta function can itself be derived by similar methods:

Exercise VIIA2.2

Derive the following B function identities:

a Use the integral representation of the Γ function to write an expression for $\Gamma(a)\Gamma(b)$ as an integral over two Schwinger parameters, and introduce a scaling parameter (as with general two-propagator Feynman graphs, except here there is no momentum). Show the result is $\Gamma(a+b)B(a,b)$, where B is given by the integral definition, thus proving the Beta function can be expressed in terms of Gamma functions.

b Use the integral definition of B to prove

$$B(x,x) = 2^{1-2x}B(\tfrac{1}{2},x) \quad \Rightarrow \quad \frac{\Gamma(x)}{\Gamma(2x)} = \frac{2^{1-2x}\sqrt{\pi}}{\Gamma(x+\frac{1}{2})}$$

Use this result to find the expression in exercise VIIA2.1 for $\Gamma(n+\frac{1}{2})$.

c Derive the identity

$$\int_0^\infty d\tau\ \tau^{a-1}(1+\tau)^{-a-b} = B(a,b)$$

from the substitution $\tau = z/(1-z)$. Use this to show

$$\Gamma(z)\Gamma(1-z) = \pi\ csc(\pi z)$$

by changing variables $\tau = e^u$, and closing the contour in the complex plane to pick up the contributions from the poles. We thus have

$$\Gamma(z+n)\Gamma(1-z-n) = (-1)^n\Gamma(z)\Gamma(1-z)$$

(which is also seen inductively from $\Gamma(z+1) = \Gamma(z)$).

In general UV divergences will come from powers of $1/\epsilon$ in a Γ resulting from integration over some scaling parameter. We thus need an expression for the Laurent expansion of $\Gamma(z)$. This can be obtained directly from the integral expression: Using the definition of e as a limit,

$$\Gamma(z) = \lim_{n\to\infty} \int_0^n d\lambda\ \lambda^{z-1}(1-\tfrac{\lambda}{n})^n = \lim_{n\to\infty} n^z B(z,n+1) = \tfrac{1}{z} \lim_{n\to\infty} n^z \prod_{m=1}^{n} \frac{1}{1+\frac{z}{m}}$$

from the change of variables $\lambda = n\alpha$, and the above identities. Defining the "Euler-Mascheroni constant" γ by

$$\gamma = \lim_{n\to\infty}\left(-ln\ n + \sum_{m=1}^{n} \tfrac{1}{m}\right) = 0.5772156649...$$

$$\Rightarrow \qquad e^{-\gamma z} = \lim_{n\to\infty} n^z \prod_{m=1}^{n} e^{-z/m}$$

we then can write

$$\Gamma(z) = \tfrac{1}{z} e^{-\gamma z} \prod_{n=1}^{\infty} \frac{e^{z/n}}{1+\frac{z}{n}}$$

which is an alternate definition of Γ. We then have

$$ln\ \Gamma(1-z) = \gamma z + \sum_{n=1}^{\infty}[-ln(1-\tfrac{z}{n}) - \tfrac{z}{n}] \quad \Rightarrow$$

$$ln\ \Gamma(1-z) = \gamma z + \sum_{n=2}^{\infty} \frac{\zeta(n)}{n} z^n, \qquad \zeta(y) = \sum_{m=1}^{\infty} \frac{1}{m^y}$$

by Taylor expansion of the ln, where ζ is the "Riemann zeta function".

Exercise VIIA2.3

Derive the following Γ identities from the previous:

a Find the first two terms in the Laurent expansion of $\Gamma(z)$:

$$\lim_{z\to 0} \Gamma(z) = \lim_{z\to 0} \tfrac{1}{z}\Gamma(z+1) = \tfrac{1}{z} - \gamma + \mathcal{O}(z), \qquad \gamma = -\int_0^\infty d\lambda\ (ln\ \lambda) e^{-\lambda}$$

b Do the same for expansions about other integers:

$$\Gamma(n+1+\epsilon) = n!\left[1 + \epsilon\left(-\gamma + \sum_{m=1}^{n} \frac{1}{m}\right) + \mathcal{O}(\epsilon^2)\right]$$

$$\Gamma(-n+\epsilon) = (-1)^n \frac{1}{n!}\frac{1}{\epsilon}\left[1 + \epsilon\left(-\gamma + \sum_{m=1}^{n} \frac{1}{m}\right) + \mathcal{O}(\epsilon^2)\right]$$

c Using the csc relation in exercise VIIA2.2c and the above expansion of $ln\ \Gamma(1-z)$, show that

$$cot\ z = \frac{1}{z}\left(1 - 2\sum_{n=1}^{\infty} \frac{\zeta(2n)}{\pi^{2n}} z^{2n}\right)$$

and thus $\zeta(2n)$ can be written as a rational number times π^{2n}.

3. Modified subtractions

The convenient normalization for the quadratic part of the gauge-invariant action for an arbitrary field theory we use is

$$S_0 = \frac{(\frac{1}{2}\mu^2)^{(D-4)/2}}{g^2} \int \frac{d^D x}{(2\pi)^{D/2}} \ \tfrac{1}{2}\phi K \phi$$

for a real field ϕ and some coupling constant g, where for bosons

$$K = \tfrac{1}{2}(-\Box + m^2) + ...$$

The explicit factor of $1/2$ cancels the factor of 2 obtained when varying the action with respect to ϕ. (Similar permutation factors are used for interaction terms.) Equivalently, it gives the natural normalization for Gaussian functional integration over ϕ in the field theoretic path integral. For complex fields we instead have $\bar{\phi}K\phi$ without the $1/2$, since then ϕ and $\bar{\phi}$ can be varied (or integrated over) independently.

The "renormalization mass scale" μ has been introduced to preserve the mass dimension of g in arbitrary spacetime dimension; it appears naturally with the normalization $\frac{1}{2}\mu^2$ because the kinetic operator contains $\frac{1}{2}p^2$ and $\frac{1}{2}m^2$. Generally there will be more than one coupling, but only one μ. For some purposes it may be convenient to scale the fields so that the coupling and μ dependence appear only in the interaction terms. We will usually suppress the μ dependence, since it is determined by dimensional analysis, and is relevant only for quantum corrections, where $D \neq 4$ becomes important.

Note that our normalization differs from that normally chosen in the literature. It has been chosen to give the normalization appropriate for Gaussian integrals, which appear in both the first- and second-quantized theories. The first difference is the factor of $(2\pi)^{-D/2}$ for coordinate and momentum integration (rather than the usual 1 for coordinate and $(2\pi)^{-D}$ for momentum); the second is the factor of $1/2$ multiplying the kinetic operator $p^2 + m^2$ (contained in K), rather than 1. Our normalization is more natural not only for Gaussian integration and Fourier transformation, but also slightly simplifies perturbative field theory calculations, allowing one to ignore spurious factors of 2 and especially 2π. The net effect is only to change the normalization of coupling constants, since such factors can be absorbed into the $1/g^2$ sitting in front.

For example, the most accurately experimentally verified prediction of quantum theory is the anomalous magnetic moment of the electron, to be discussed later. The result for the total magnetic moment, in various normalizations, to second order in perturbation theory, is

$$\mu_{mag} = 1 + e^2 = 1 + \frac{e_m^2}{2\pi} = 1 + \frac{e_{ft}^2}{8\pi^2}$$

where "e" is for our normalization (obviously the simplest), "e_m" is the normalization you first learned in classical mechanics (the one that gives e_m^2/r^2 as the electrostratic force between two electrons), and "e_{ft}" is the one you would see in other quantum electrodynamics courses. To complicate matters, you may have also seen the definition $\alpha = e_m^2 = e_{ft}^2/4\pi$, of which the only merit is supposed to be that $1/\alpha$ is very close to the integer 137, which is silly since $1/e^2$ is even closer to the integer 861. (Actually, α is the natural expansion parameter for nonrelativistic quantum mechanics, which is basically 3D, but our e^2 is more natural for the loop expansion, which is inherently 4D.) We have also used units $\hbar = 1$; restoring it introduces the further complication $\alpha = e_m^2/\hbar_m = e_{ft}^2\hbar_{ft}/4\pi$ because of the difference in the semiclassical expansions for quantum mechanics and quantum field theory.

Furthermore, for nonabelian groups we have an extra factor of $\frac{1}{2}$ compared to the standard normalization because we normalize $tr_D(G_iG_j) = \delta_{ij}$ instead of $tr_D(G_iG_j) = \frac{1}{2}\delta_{ij}$: The latter originated from the case of SU(2), where it cancels the $\sqrt{2}$ in the diagonal generator (and in the others, if one uses hermitian ones rather than raising and lowering). Unfortunately, for SU(N) with $N \geq 3$ this historical normalization *introduces* a $\sqrt{2}$, while not canceling factors like $\sqrt{N}$ (and making $tr_D(G_iG_j)$ N-dependent would wreak havoc when considering subgroups, as well as for raising and lowering operators). Thus, in the above notation,

$$nonabelian: \qquad g^2 = \frac{g_{ft}^2}{16\pi^2}$$

Although we have chosen a normalization for the coupling constants that is natural for Gaussian momentum integration, and for symmetry with respect to Fourier transformation, in divergent integrals it has the disadvantage of having γ's (Euler-Mascheroni constant) in the finite parts. Another natural normalization that gets rid of this irrational number from all graphs is to divide out the angular part of the integrals: The volume of the unit D$-$1-dimensional sphere (the surface of unit radius in D-dimensional Euclidean space), is easily evaluated with Gaussians:

$$1 = \int \frac{d^Dk}{(2\pi)^{D/2}}\ e^{-k^2/2} = (2\pi)^{-D/2}\left(\int d^{D-1}\Omega\right)\int_0^\infty dk\ k^{D-1}e^{-k^2/2}$$
$$= \tfrac{1}{2}\Gamma(\tfrac{D}{2})\pi^{-D/2}\int d^{D-1}\Omega$$

We might therefore choose our normalization to cancel this factor in momentum integrals, along with the $(2\pi)^{-D}$ from Fourier transformation. Then the action, e.g., for a massless scalar, might be normalized as (with conventional kinetic term)

$$S_0 = \frac{\mu^{D-4}}{g^2}\int d^Dx\ \frac{(\int d^{D-1}\Omega)}{(2\pi)^D}\tfrac{1}{2}(\partial\phi)^2 = \frac{\mu^{D-4}}{g^2}\int \frac{d^Dx}{(4\pi)^{D/2}\Gamma(\frac{D}{2})}(\partial\phi)^2$$

This differs from our previous normalization by a factor of $1/\Gamma(\frac{D}{2})$, which is 1 in exactly D=4, but differs infinitesimally far away. The result is that the two schemes will differ effectively by finite renormalizations: For example, in divergent one-loops graphs the $1/\epsilon$ divergences will have the same coefficient in the two schemes, but the finite remainders will differ by constants, since effectively the coupling has been redefined by a factor of $1+\mathcal{O}(\epsilon)$. Thus, the same result can be achieved by modifiying the counterterm to be proportional to $1/\epsilon + constant$. Hence, the earlier version of dimensional regularization is called "minimal subtraction (MS)", while the modification inspired by the volume of the sphere is called "modified minimal subtraction ($\overline{\text{MS}}$)".

We now examine explicitly the difference between the two schemes. As we can see from our previous example, momentum integrals from scaling various subsets of Schwinger parameters in a multiloop diagram will produce Γ function factors of the form, with this new normalization,

$$\prod_i [\Gamma(\tfrac{D}{2})]^{L_i} \Gamma(n_i - L_i \tfrac{D}{2})$$

for some integers n_i, where $\sum_i L_i = L$. In even dimensions D_0 (especially 4), we can use $\Gamma(z+1) = z\Gamma(z)$ to write each L_i-loop factor as a rational function of $\epsilon = (D_0 - D)/2$ times

$$[\Gamma(1-\epsilon)]^{L_i} \Gamma(1+L_i\epsilon) \sim (e^{\gamma\epsilon})^{L_i} e^{-L_i\gamma\epsilon} = 1$$

where we have written only the γ dependence. Thus all γ's cancel. This modification allows some further simplification by eliminating extra terms arising at 2 loops involving $\zeta(2)$ (but this is only $\pi^2/6$; see exercise VIIA2.3). Similar results can be obtained by using $\Gamma(1-\epsilon)$ instead of $\Gamma(\frac{D}{2})$.

Another subtraction scheme, the "G scheme", is defined by normalizing momentum integration so that the coefficient of the 1-loop massless propagator correction in ϕ^3 theory in 4 dimensions (see subsection VIIB4 below) is *exactly* $1/\epsilon$ (without extra finite terms) times a power of $\frac{1}{2}p^2$, or that up to a sign in higher even dimensions. (The normalization factor must be positive, and also finite and nonvanishing as $\epsilon \to 0$.) As for $\overline{\text{MS}}$, we can also pull out rational factors to get just the $\Gamma(1+n\epsilon)$'s. The net effect of these two schemes, as compared to MS, is to modify $\hbar$, which appears as $\int dx/\hbar$ in the classical action or $\hbar \int dp$ for loop integrals, as

$$\overline{\text{MS}}: \qquad \hbar \to \Gamma(\tfrac{D}{2})\hbar \quad or\ \Gamma(1-\epsilon)\hbar$$

$$\text{G}: \qquad \hbar \to \frac{(-1)^{D_0/2}}{\epsilon\Gamma(2-\frac{D}{2})B(\frac{D}{2}-1,\frac{D}{2}-1)}\hbar \quad or\ \frac{\Gamma(1-2\epsilon)}{\Gamma(1+\epsilon)[\Gamma(1-\epsilon)]^2}\hbar$$

(If we want to be picky, we can also normalize the former forms appropriately for $D = D_0$, by including an extra factor of $1/\Gamma(\frac{D_0}{2})$ for $\overline{\mathrm{MS}}$ and $\Gamma(\frac{D_0}{2}-1)/\Gamma(D_0-2)$ for G.)

Exercise VIIA3.1

Explictly evaluate the difference between the MS, $\overline{\mathrm{MS}}$, and G schemes to order ϵ, including the picky D_0 factors.

This particular fix for eliminating irrational numbers works only for those arising at one or two loops: In general, because subdivergences produce expressions of the form $\Gamma(1+\epsilon)/\epsilon^L$ at L loops, we encounter finite terms involving $\zeta(L)$ at L loops, which is irrational (worse than just π's, e's, $\sqrt{2}$'s, etc.) for odd L. So, for example, $\zeta(3)$ appears at 3 loops, and it can be shown that $\zeta(3)$ (and higher $\zeta(n)$) can't always be canceled.

The "momentum subtraction scheme (MOM)", rather than simplifying numerically, is designed to give a more physical interpretation of the coupling constants appearing in the action: It is defined so that they take their on-shell values. Thus, it is particularly suited to low-energy calculations, which involve an expansion about the mass shell. For example, consider the quantum kinetic operator $K + \Delta K$, appearing in the quadratic part of the effective action. It depends on the momentum through one variable: p^2 or $\not{p}$, etc. We then can consider Taylor expanding it in this variable about its classical on-shell value. (For reasons to be explained later, this can be dangerous for massless fields, when it requires infrared regularization.) This is equivalent to expanding in powers of the classical kinetic operator K itself. The MOM prescription is then to use subtraction terms $-\delta K$ to cancel the terms in the quantum correction ΔK to the kinetic operator that are linear in K:

$$\Delta K = a + bK + \mathcal{O}(K^2) \quad \Rightarrow \quad \delta K = -a - bK$$

$$\Rightarrow \quad K_{ren} \equiv K + \Delta K + \delta K = K + \mathcal{O}(K^2) \quad \Rightarrow \quad \frac{1}{K_{ren}} = \frac{1}{K} + \mathcal{O}(K^0)$$

The two renormalizations are related directly to the "wave function" and mass renormalizations, one being proportional to the entire kinetic term, the other to the constant (mass) term. The result is that the renormalized propagator has the same pole and residue as the classical one.

Note that the MOM scheme, unlike the others, does not introduce an independent mass scale μ: Only physical masses set the scale. This is a consequence of the fact that the MOM scheme is designed for studying low-energy (near-mass-shell) behavior, while the others are more suited for studying high-energy behavior. This will be important for our explicit calculations later, when we see that MOM is more useful for

QED, which is better defined (and thus more useful), in terms of perturbation theory, at low energies, while QCD is better defined at high energies. More precisely, the on-shell values of QED masses and couplings are observed experimentally, whereas those of QCD are almost meaningless, since the corresponding particles are not observed as asymptotic states. On the other hand, in QCD the introduction of the arbitrary scale μ allows the definition of a more physical mass scale, and its arbitrariness can actually improve the accuracy of perturbative calculations.

4. Optical theorem

Writing the S-matrix as $\mathcal{S} = 1 + i\mathcal{T}$, unitarity can be written as

$$1 = \mathcal{S}^\dagger \mathcal{S} = (1 - i\mathcal{T}^\dagger)(1 + i\mathcal{T}) = 1 + i(\mathcal{T} - \mathcal{T}^\dagger) + \mathcal{T}^\dagger \mathcal{T} \quad \Rightarrow \quad \mathcal{T}^\dagger \mathcal{T} = i(\mathcal{T}^\dagger - \mathcal{T})$$

(Actually, the more useful statement is in terms of $\mathcal{S} = e^{i\mathcal{T}}$, since then $\mathcal{T}$ represents the connected graphs, but the result of the argument is the same.) Summation of a probability over final states then yields the "optical theorem":

$$\sum_f |\mathcal{T}_{fi}|^2 = \sum_f \langle i|\mathcal{T}^\dagger|f\rangle\langle f|\mathcal{T}|i\rangle = \langle i|\mathcal{T}^\dagger\mathcal{T}|i\rangle = 2\ Im\ \mathcal{T}_{ii}$$

Applying unitarity in terms of the cutting rules (subsection VC6), we see that this condition can be applied to $\mathcal{T}_{ii}$ diagram by diagram, using any combination of parts of $\mathcal{T}_{fi}$ and ${\mathcal{T}^\dagger}_{fi}$ that fit together to form the graph considered for $\mathcal{T}_{ii}$. (For example, the probability coming from a tree graph with two final states is the imaginary part of a one-loop graph with two intermediate states.) Separating out the momentum-conservation δ-function for a connected S-matrix element, we have finally

$$\mathcal{T} = \delta\left(\sum p\right) T \quad \Rightarrow \quad \sum_f \delta\left(\sum p\right) |T_{fi}|^2 = 2\ Im\ T_{ii}$$

The simplest example of an experimental measurement of an interaction is a decay rate. (The only particle properties contained in the free Lagrangian are mass and spin.) At the tree (classical) level, this is given directly by a cubic (for decay into two particles) or higher-order term in the Lagrangian. More generally, by the optical theorem the decay probability is given by the imaginary part of the propagator correction, evaluated on shell.

We then find the total decay probability per unit time by dividing the probability by the spatial density ρ times the spatial volume times the time duration, and summing over final states:

$$\frac{dP}{dt} = \sum_f \frac{P}{\rho V_D} = \frac{2\ Im\ T_{ii}}{\omega}$$

using the expressions for ρ, and P in terms of $|T_{fi}|^2$, given in subsection VC7. The optical theorem can be applied similarly for the total cross section:

$$\sigma = 2(Im\ T_{ii})\frac{(2\pi)^{D/2}}{\lambda_{12}}$$

The decay rate for a particle is frame dependent, but we usually pick the rest frame for massive particles, where $\omega = m$. Alternatively, we can define the total decay probability per unit *proper* time:

$$\omega = m\frac{dt}{ds} \quad \Rightarrow \quad \frac{dP}{ds} = \frac{2\ Im\ T_{ii}}{m}$$

(where s and t should not be confused with the Mandelstam variables). For the massless case, we can use instead the parameter τ, as it appears in classical mechanics in the gauge $v = 1$, or as the classical value of the Schwinger parameter from the Landau equations:

$$p^a = \frac{dx^a}{d\tau} \quad \Rightarrow \quad \frac{dP}{d\tau} = 2\ Im\ T_{ii}$$

Exercise VIIA4.1

Compare this result for dP/ds with that of exercise VC7.1a to obtain an explicit expression of $Im\ T_{ii}$ in terms of $|T_{fi}|^2$ for the decay of a particle of mass M into two particles of masses m_1, m_2. What happens to $Im\ T_{ii}$ for $M = m_1 + m_2$, and for $M < m_1 + m_2$?

Now the decay rate of a particle can also be associated with the imaginary part of the mass, since

$$M = m - ir \quad \Rightarrow \quad |\psi|^2 \sim |e^{-iMt}|^2 = e^{-2rt}$$

so the wave function for the particle at rest automatically includes a decay factor. The probability of decay, normalized by dividing by $|\psi|^2$, is thus

$$\frac{dP}{dt} = \frac{d(1-|\psi|^2)}{|\psi|^2 dt} = 2r$$

The analogous statement in momentum space is found by Fourier transforming the propagator/wave function from time to energy:

$$e^{-iMt} \to \frac{1}{E-M} = \frac{1}{E-m} + \frac{1}{E-m}(-ir)\frac{1}{E-m} + ...$$

This expansion is in terms of the free propagator $1/(E-m)$ and the connected graphs $-ir$.

We now check that this result agrees with that obtained from the effective action. The quantum propagator has a pole at $p^2 = -M^2$ for some *complex* constant M:

$$\lim_{p^2 \to -M^2} \Delta = \frac{1}{\frac{1}{2}(p^2 + M^2)}$$

We have normalized the propagator at the pole by rescaling the field; we also keep the real part of the mass the same as the classical value through renormalization of the mass term. (Only the real part can be renormalized consistently with unitarity.) The on-shell condition is then at physical (real) momentum $p^2 = -m^2$, so the kinetic operator on-shell is

$$\tfrac{1}{2}(p^2 + M^2) = -\tfrac{1}{2}m^2 + \tfrac{1}{2}(m - ir)^2 = -\tfrac{1}{2}r^2 - imr$$

Remembering that "interaction" terms from Γ contribute with a minus sign to amplitudes, we then have

$$\frac{dP}{dt} = \frac{2\ Im\ T_{ii}}{m} = \frac{2mr}{m} = 2r$$

5. Power counting

We now consider why subtracting out divergences, as poles in D, can be implemented by giving singular D-dependence to the coupling constants. This is based on dimensional analysis, which tells us how divergent a graph is at large momenta: the "ultraviolet" (UV) divergence. (There are also infrared divergences, which occur for physical reasons, and do not require renormalization. They occur only for massless particles, and will be considered later.) Consider first any 1-loop 1PI graph. In momentum space, it has an integral over the loop momentum, $\int d^D p$. It will diverge if the integrand (before introducing Schwinger parameters) goes as p^{-D} or slower to infinite momentum (UV limit). In this limit we can ignore masses. If we differentiate this graph with respect to any of the external momenta, it will become more convergent, since the power of momenta in the integrand decreases. (The numerator of the integrand is a polynomial, while each factor in the denominator depends on the loop momentum.) With enough derivatives, it becomes convergent. This means that the divergent part of the graph is a polynomial in the external momenta. Similar remarks apply to any 1PI graph, if we consider the divergence coming from letting all loop momenta go to infinity, known as the "superficial divergence". Of course, the superficial divergence is also polynomial in the coupling constants, as is the graph as a whole; but the superficial divergence is also polynomial in the masses, since differentiation with respect to them has the same effect as with respect to external momenta.

We can determine several more properties of this local, but divergent, contribution to the effective action. First of all, it is Poincaré invariant and invariant under all global symmetries of the classical action, since the effective action is invariant for all values of D, so poles in D are also. (Consider, e.g., contour integration in D to pick out the pole.) If we use the background field gauge (or consider only Abelian gauge fields), then it is also gauge invariant. (However, possible exceptions are conformal invariance and invariances involving γ_{-1}, since those are not invariances of the classical action for all D.) The other property we need is that the coefficient of the divergence is real. This follows from the fact that the S-matrix satisfies unitarity, as preserved manifestly by dimensional regularization. As discussed in the previous subsection, we have unitarity

$$\mathcal{S}^\dagger \mathcal{S} = I, \quad \mathcal{S} = I + i\mathcal{T} \quad \Rightarrow \quad i(\mathcal{T}^\dagger - \mathcal{T}) = \mathcal{T}^\dagger \mathcal{T}$$

As we saw in subsection VC6, this identity actually can be applied to a single graph, where the element of $\mathcal{T}$ on the left-hand side of the equation is that graph, while on the right-hand side the summation over intermediate states gives a sum where each term divides the graph into two parts, one for $\mathcal{T}$ and one for $\mathcal{T}^\dagger$. We then see that at any loop the imaginary part of a 1PI graph in $\mathcal{T}$ is given by "sewing" together diagrams from lower loops. This means that any new divergence at any number of loops must be real, since sewing doesn't introduce new (UV) divergences: Sewed lines are on shell, and phase space for on-shell states is always finite. (The 3-momentum of each state is bounded by the energy, and each positive energy of an outgoing state is bounded by the total energy of the system.)

Since Poincaré and gauge invariances, locality, and semiclassical unitarity were used as properties to determine the classical action, this suggests that the divergent terms in the effective action might all be of the form of terms already in the classical action. Such a property is called (perturbative) "renormalizability". When it holds, all infinities can be absorbed by a redefinition of the coupling constants (and masses) appearing in the classical action. This is physically important because all infinities are defined only up to finite pieces: For example, in dimensional regularization, we saw in subsection VIIA3 that the D-dependent normalization of the classical action is ambiguous, resulting in an ambiguity in the finite pieces left over after subtraction of $1/\epsilon$ terms. Since we now know that superficial divergences are local, we see that renormalization can produce arbitrary finite, local terms in the effective action, corresponding to the divergent terms. But if the divergent terms are all of the same form as in the classical action, all such finite terms can be absorbed by a redefintion of the coupling constants. On the other hand, if such finite terms did not already appear in the classical action, we would be forced to introduce them, to make the renormalization procedure unambiguous. (Of course, we could give an unambiguous prescription by definition, but from the point of view of another prescription this would be the same as including the extra terms in the classical action, and using the first prescription to arbitrarily fix the nonzero values of the couplings.) Thus, the condition of renormalizability is necessary to prevent the appearance of an infinite number of coupling constants, which would result in the loss of predictability. (If divergences require only a finite number of such couplings to be added, we simply include those, to obtain a renormalizable theory with a number of couplings that is finite, although larger than that with which we started.)

Since dimensional analysis determines the form of the divergent terms of any momentum integral, it also detemines which theories are renormalizable. By appropriate rescaling of fields by constants, write the classical action in a form where the derivative parts of the kinetic terms have no dependence on any couplings. Then define the couplings to be the coefficients of the interaction terms. It is easy to see that the renormalizable theories are the ones that include all terms which satisfy all the properties required of the classical action (including preservation of all appropriate symmetries that are manifestly preserved by dimensional regularization), where all couplings have engineering dimensions that are nonnegative powers of mass: Consider first the case where all couplings are dimensionless (and there are no masses, or at least we ignore them at high energies for purposes of considering UV divergences). Then the theory is renormalizable simply because there are no dimensionful parameters around, so any local term must be of the form of those originally in the classical action. If we now introduce couplings with positive mass dimension, then perturbatively they can occur only to nonnegative power in any diagram, so any divergence thus produced again has a coefficient with nonnegative mass dimension. Since the fields themselves have positive mass dimension, there are only a finite number of such terms possible. On the other hand, if we were to allow couplings with negative dimension, then terms with arbitrarily high powers of such couplings would also allow arbitrarily high powers of fields, and thus lead to nonrenormalizability. (By similar arguments, theories with only couplings of positive mass dimension, called "superrenormalizable", can have divergences only to a certain finite number of loops.)

In particular, in D=4 the derivative part of the kinetic term for bosons is of the form $\int d^4x\ \phi\partial\partial\phi$, and for fermions $\int d^4x\ \psi\partial\psi$, so bosonic fields have dimension 1 and fermions 3/2. That means that bosons can appear only quartically and fermions only quadratically. More specifically, renormalizable theories can only have terms of the form

$$\phi, \phi^2, \phi^3, \phi^4, \phi\partial\phi, \phi^2\partial\phi, \phi\partial\partial\phi; \psi^2, \psi\partial\psi; \phi\psi^2$$

(where each ϕ can be any boson with any spin, and each ψ any fermion). There can also be constant terms (field-independent), which we always drop, since they don't contribute to perturbative amplitudes (after appropriate normalization). The terms, and their relations, are restricted by Lorentz, gauge, and internal symmetries. The potential for scalar fields must also be bounded from below, to allow the existence of a vacuum (state with lowest energy); otherwise nothing would be stable, continually decaying into states of lower energy (i.e, the energy of the scalars converting to other particles): Thus, ϕ^3 terms for scalars requires also ϕ^4 terms.

Spin 1 can't couple minimally to spins >1. (One way to show this is to covariantize the general field equation of IIB1 to $S_a{}^b\nabla_b + k\nabla_a$, and show the commutator algebra of this constraint, and $\Box + ...$, doesn't close unless the spin ≤1 or the external field strength vanishes.) Furthermore, gauge invariance for spins >1 prevents them from having renormalizable gauge couplings in D=4: For example, we saw in subsection IIIA4 that spin-2 (gravity) couplings include terms of the form $\phi\partial\phi\partial\phi$. Renormalizability therefore restricts us to spins 0, 1/2, and 1. Using Poincaré and gauge invariance, the most general action is then of the form

$$L = tr\left\{\tfrac{1}{8g^2}F^2 + \psi^\alpha i\nabla_\alpha{}^{\dot\alpha}\bar\psi_{\dot\alpha} + \tfrac{1}{4}(\nabla\phi)^2 + V(\phi) + [\tfrac{1}{2}\psi^\alpha(\phi + \tfrac{m}{\sqrt{2}})\psi_\alpha + h.c.]\right\}$$

where all group matrices are implicit: They may appear in all fields, in m, and even in g, which has independent values for the different factors of the Yang-Mills gauge group. Also, the matrices may differ for the same field in different terms: A (in ∇) has different Yang-Mills representations on different scalar and spinor fields, and ϕ appears with some matrix in its Yukawa coupling $\psi\phi\psi$. (Of course, all matrices must be chosen consistently with gauge and global invariances.) The potential $V(\phi)$ is no higher than quartic. Note that the Higgs mechanism is required to give nonabelian gauge fields mass: A^2 is not gauge invariant, and $(\nabla\phi)^2$ is the only way to dress it up with scalars in a renormalizable way. (For the Abelian case we can use Stückelberg fields, with $\nabla = \partial + mAT$ as in subsection IVA5. In the nonabelian case, introducing scalars by a gauge transformation, as for Stückelberg, results in a nonrenormalizable $(e^{-i\phi}\nabla e^{i\phi})^2$ term.) If we ignore gauge invariance, the A^2 term produces unitary-gauge propagators with bad high-energy behavior (see subsection VIB3), which leads to the same nonrenormalizable behavior in the absence of a Higgs mechanism.

Exercise VIIA5.1

Show by power counting that interacting renormalizable theories with potentials that are bounded from below exist only in D≤4. Show that for D≤2 there are an infinite number of possible renormalizable terms in the action. What are the kinds of renormalizable terms possible in D=3?

Exercise VIIA5.2

Superrenormalizable theories aren't realistic, but they give oversimplified examples of many quantum features of field theory.

a What theories are superrenormalizable in D=3? Show that the only superrenormalizable interaction in D=4 is (scalar) ϕ^3.

b Let's do power counting (dimensional analysis) for 4D ϕ^3 theory. Write the action in the form

$$S = \tfrac{1}{g^2} \int dx \; [-\tfrac{1}{4}\phi(\Box - m^2)\phi + \phi^3]$$

so g^2 counts loops. ($\phi \to g\phi$ gives the form where g counts vertices.) What are the dimensionless terms

$$\mathit{\Delta} S = (g^2)^{L-1} \int dx \; \phi^n$$

for all n (including the vacuum bubbles $n = 0$; of course, $L \geq 0$)? Since superficial divergences are polynomial in everything (fields, momenta, couplings, masses), this gives the maximum number of loops $L(n)$ for a superficial divergence to appear in an n-point 1PI amplitude. Make a similar analysis for 3D ϕ^4 theory.

c Find all the divergent 1PI diagrams in 4D ϕ^3 theory. (Hint: There are 5, or 6 if we include the trivial 1-loop graph with no vertices.)

Such global and local symmetry requirements can also be applied to the effective action. In background-field gauges $\mathit{\Gamma}$ is gauge invariant (see subsection VIB8), which restricts the form of the effective potential, and even nonlocal terms. In QED charge conjugation, in addition to switching 2-spinors of opposite charge, changes the sign of the electromagnetic potential: Consequently, any pure-A term in $\mathit{\Gamma}$ must be even in A's ("Furry's theorem"). Such classical symmetries can be applied at the quantum level only in the absence of "anomalies", quantum violations (discussed in chapter VIII below). However, even the anomalies themselves are restricted by symmetries: Anomalies occur only in symmetries that can't be manifestly preserved by regularization, which means only conformal or axial symmetries. Thus, when the couplings of gauge vectors are parity invariant, the axial anomaly (which violates parity by definition) is irrelevant.

6. Infrared divergences

Although ultraviolet (UV) divergences represent a serious problem, in the sense that they strongly restrict which theories can be useful, and require renormalization, infrared (IR) divergences are merely a consequence of poor semantics: The definition of the S-matrix assumes the existence of well-defined one-particle asymptotic states. Unfortunately, these do not exist when massless particles are present, even in classical mechanics: (1) Any particle can be accompanied by an arbitrary number of massless particles with vanishing 4-momentum, and such a collection of particles can be indistinguishable from the lone particle if the (measured) quantum numbers are the same. (These are physical states, since $p^a = 0 \rightarrow p^2 = 0$.) (2) Any massless particle can be indistinguishable from an arbitrary number of massless particles, with the same total 4-momentum, and each with the same sign energy, if their 4-momenta are all proportional, since then they are all traveling in the same direction at the same speed. (This situation is not important for QED, since the photon can't decay directly into two photons.)

Experimentally, because detectors have finite accuracy, in the first case there can be such "soft" particles with total energy below some small upper limit, and in the second case there can be such "colinear" particles within some small angle of resolution. In principle this means we should change our definition of asymptotic states accordingly; in practice this is too complicated, but to any particular order in perturbation theory only a finite number of such additional massless particles will couple. The procedure is then to

(1) infrared regularize the S-matrix amplitudes (by dimensional regularization, or introducing masses for all particles, or keeping massless particles off-shell);

(2) calculate probabilities/cross sections, including contributions from soft and colinear particles, as a function of some upper limit on their energy/angle (representing the experimental accuracy); and

(3) remove the regularization.

No infrared renormalization is necessary. (Examples will be given in later sections. Of course, for total cross sections, all energies and angles are integrated over anyway.) In general, such a procedure must be applied to both initial and final states (the"Kinoshita-Lee-Nauenberg theorem"), but in QED (as opposed to QCD) it is sufficient to treat only the final ones in cases of physical interest.

The reason why infinities appear in cross sections if we ignore this careful prescription, and in S-matrix elements in any case, is the long range of forces mediated by massless particles. A cross section, although it represents a probability, is normalized in such a way that it is an area, representing the effective cross-sectional area of a particle being targeted by another particle. The range of the interaction sets the scale of this area; this is related to the mass of the particle mediating the interaction. Since massless particles produce infinite-range forces, the result is infinite cross sections. One might expect that these infinities would appear only in total cross sections, where the momenta of particles in the final state are integrated over. However, by the optical theorem, this total cross section is given by the imaginary part of an S-matrix amplitude, which thus must also have this infinity.

The fact that these infrared divergences are physical also follows from the fact that these kinematic situations occur in classical mechanics: In subsection VC8 we saw that physical singularities occur in S-matrices for momenta that are allowed classically.

Exercise VIIA6.1

Evaluate the total cross section for 2→2 scalar scattering at the tree level (as in the example of subsection VC4), with all particles massless, and show it has an infrared divergence. Relate this divergence to a classical situation.

REFERENCES

1 P.A.M. Dirac, *Proc. Cambridge Phil. Soc.* **30** (1934) 150;
W. Heisenberg, *Z. Phys.* **92** (1934) 692;
V. Weisskopf, *Kon. Dan. Vid. Sel. Mat.-fys. Medd.* **XIV** (1936) 1:
charge renormalization.

2 W. Pauli and M. Fierz, *Nuo. Cim.* **15** (1938) 167;
H. Kramers, *Nuo. Cim.* **15** (1938) 134:
mass renormalization.

3 Dyson, *loc. cit.* (VC):
renormalization to all loops.

4 A. Salam, *Phys. Rev.* **82** (1951) 217, **84** (1951) 426:
completed Dyson's proof.

5 Bogoliubov and Shirkov, *loc. cit.*:
general approach to renormalization.

6 G. 't Hooft and M. Veltman, *Nucl. Phys.* **B44** (1972) 189;
C.G. Bollini and J.J. Giambiagi, *Nuo. Cim.* **12B** (1972) 20;
J.F. Ashmore, *Lett. Nuo. Cimento* **4** (1972) 289;
G.M. Cicuta and E. Montaldi, *Lett. Nuo. Cim.* **4** (1972) 329:
dimensional regularization.

7 G. 't Hooft and M. Veltman, Diagrammar, in *Particle interactions at very high energies*, proc. 2nd Summer Institute on Particle Interactions at Very High Energies, Louvain, Belgium, Aug 12-25, 1973, eds. D. Speiser, F. Halzen, J. Weyers (Plenum, 1974) part B, p. 177; and in *Under the spell of the gauge principle*, ed. G. 't Hooft (World Scientific,

1994) p. 28:
dimensional renormalization.

8 Siegel, *loc. cit.* (VIC, ref. 10, next to last), W. Siegel, *Phys. Lett.* **94B** (1980) 37; Gates, Grisaru, Roček, and Siegel, *loc. cit.*:
regularization by dimensional reduction.

9 J. Polchinski, *Nucl. Phys.* **B231** (1984) 269:
simplest proof of renormalization; uses renormalization group, doesn't use geometry of graphs ("forests" or "skeletons", etc.) or "Weinberg's theorem"; not yet simply generalized to gauge theories.

10 Feynman, *loc. cit.* (VB, ref. 4):
Feynman parameters.

11 E.T. Whittaker and G.N. Watson, *A course of modern analysis*, 4th ed. (Cambridge University, 1927) p. 235:
detailed discussion of Γ functions.

12 A.A. Vladimirov, *Theor. Math. Phys.* **36** (1978) 732;
W.A. Bardeen, A.J. Buras, D.W. Duke, and T. Muta, *Phys. Rev.* **D18** (1978) 3998:
$\overline{\mathrm{MS}}$ scheme.

13 K.G. Chetyrkin, A.L. Kataev, and F.V. Tkachov, *Nucl. Phys.* **B174** (1980) 345:
G scheme.

14 E. Feenberg, *Phys. Rev.* **40** (1932) 40;
N. Bohr, R.E. Peierls, and G. Placzek, 1947 manuscript, in *Niels Bohr collected works*, v.9, ed. S.R. Peierls (North-Holland, 1986) p. 487:
optical theorem.

15 W.H. Furry, *Phys. Rev.* **51** (1937) 125.

16 F. Bloch and A. Nordsieck, *Phys. Rev.* **52** (1937) 54;
D.R. Yennie, S.C. Frautschi, and H. Suura, *Ann. Phys.* **13** (1961) 379:
infrared divergences.

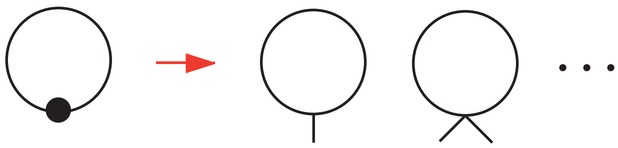
17 T. Kinoshita, *J. Math. Phys.* **3** (1962) 650;
T.D. Lee and M. Nauenberg, *Phys. Rev.* **133B** (1964) 1549.

18 R. Gastmans and R. Meuldermans, *Nucl. Phys.* **B63** (1973) 277;
W.J. Marciano and A. Sirlin, *Nucl. Phys.* **B88** (1975) 86:
earliest applications of dimensional regularization to infrared divergences.

B. EXAMPLES

We now give some explicit examples of the evaluation of S-matrices and contributions to the effective action — momentum integration, regularization, and renormalization — and some examples of their application.

1. Tadpoles



The simplest examples of dimensional regularization are one-loop "tadpoles", graphs with only one external line. By the Schwinger parameter method described in subsection VIIA2, we find

$$\mathcal{A}_1(x,m^2) = \int dk\ e^{ik\cdot x}\frac{1}{\frac{1}{2}(k^2+m^2)} = \int_0^\infty d\tau\ \tau^{-D/2}e^{-(\tau m^2+x^2/\tau)/2}$$

Further evaluation requires Taylor expansion in x (which we'll need anyway to evaluate a specific integral of $k...k/(k^2+m^2)$):

$$\mathcal{A}_1 = \sum_{n=0}^\infty \tfrac{1}{n!}(-\tfrac{1}{2}x^2)^n \frac{\Gamma(1-\frac{D}{2}-n)}{(\frac{1}{2}m^2)^{1-D/2-n}}$$

The mass dependence, as well as the argument of the Γ function, are as expected by dimensional analysis: $\int d^Dk\ k^{2n}/k^2$ is ultraviolet divergent (large k) for $D \geq 2(1-n)$, and infrared divergent (small k) in the limit $m \to 0$ for $D \leq 2(1-n)$. The ultraviolet divergence is reflected in $\Gamma(z)$, which has poles at the nonpositive integers.

To analyze the massless case, we evaluate the τ integral for $D < 2(1-n)$ and $m > 0$, where it is finite and well-defined, analytically continue to the region $Re\ D > 2(1-n)$ (but not exactly at the points where D is an even integer), take the limit $m \to 0$ there, and finally analytically continue this vanishing result to all D. Therefore, all massless tadpoles can be taken to vanish in dimensional regularization:

$$\int dk\ \frac{k_a...k_b}{\frac{1}{2}k^2} = 0$$

or more generally

$$\int dk\ \frac{k_a...k_b}{(\frac{1}{2}k^2)^a} = 0$$

This includes negative a, particularly integrals of polynomials of momenta. Such an integral can result from "measure factors", as discussed in subsections VA2 and VC1: For example, if an auxiliary field appears in the action with its quadratic term multiplied by a function of other fields, then functionally integrating it out of the action results in a functional determinant (in addition to replacing it in the classical action by the solution to its field equation). This is represented in terms of Feynman graphs as one-loop diagrams whose propagators are all those of the auxiliary field, namely 1. The result is then regularized as

$$\int dx\ 1 \sim \delta(0) \to \int dk\ 1 = 0$$

consistent with the fact that such factors would cancel corresponding factors we should include in the functional integration measure. (In other words, since we can always arrange to have all $\delta(0)$ factors cancel, we ignore them.)

On the other hand, massive tadpoles contribute both divergent and finite pieces under minimal subtraction: For example, for $D = 4 - 2\epsilon$,

$$\mathcal{A}_1(0, m^2) = \int dk\ \frac{1}{\frac{1}{2}(k^2 + m^2)} = \Gamma(1 - \tfrac{D}{2})(\tfrac{1}{2}m^2)^{D/2-1}$$
$$= -\tfrac{1}{2}m^2\{\tfrac{1}{\epsilon} + [-\gamma + 1 - ln(\tfrac{1}{2}m^2)]\}$$

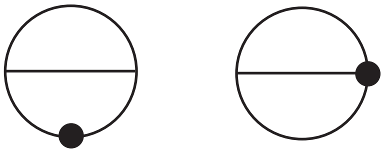
(see exercise VIIA2.3b), using $A^{-\epsilon} = e^{-\epsilon\ ln(A)}$. The γ can be killed by using an $\overline{\text{MS}}$ or G scheme (see subsection VIIA3): At 1-loop order any version of those schemes has the effect of just canceling the γ (but differences appear at 2 loops: see subsection VIIB7 below). To include the μ dependence of the coupling, we just replace everywhere (see also subsection VIIA3)

$$ln(\tfrac{1}{2}m^2) \to ln\left(\frac{m^2}{\mu^2}\right)$$

(and similarly for any momentum factors such as $ln(\frac{1}{2}p^2)$ that might appear more generally); effectively we are using units $\frac{1}{2}\mu^2 = 1$. Note that we are not allowed to Taylor expand in m: Doing so before integration would give an incorrect result; after integration it's impossible. Similar remarks apply to the exponential $e^{ik\cdot x}$ in $\mathcal{A}_1$ if we interpret it as the definition by Fourier transformation of the propagator in position space.

Exercise VIIB1.1

Find the 2D massless propagator in position space by Fourier transformation. (But *don't* Taylor expand in x.) Note that this Fourier transform is infinite, and requires "renormalization" (of a constant of integration). Compare this with the result obtained by solving the integral form of the Klein-Gordon (Laplace) equation (i.e., Gauss' law in D=2).



Two-loop tadpole integrals are not much more difficult if one line is massive, or two are massive with the same mass. (Again, tadpoles with only massless lines can be taken to vanish in dimensional regularization.) If two propagators are massless, then they can be treated first as a one-loop propagator graph: By dimensional analysis, the result of that one-loop subintegral must be a power of the momentum squared. (The explicit result will be calculated in subsection VIIB4.) We therefore consider more general one-loop tadpole integrals with more complicated propagators that may result from subintegrations in a higher-loop graph. For example, we consider

$$\hat{\mathcal{A}}_1(a,x,m^2) = \int dk\ e^{ik\cdot x}\frac{\Gamma(a)}{[\frac{1}{2}(k^2+m^2)]^a}$$

Using the definition of the Γ function, we can write

$$\frac{\Gamma(a)}{[\frac{1}{2}(k^2+m^2)]^a} = \int_0^\infty d\tau\ \tau^{a-1}e^{-\tau(k^2+m^2)/2}$$

Performing the resultant Gaussian momentum integration and Taylor expanding in x, we easily find

$$\hat{\mathcal{A}}_1(a,x,m^2) = \sum_{n=0}^\infty \tfrac{1}{n!}(-\tfrac{1}{2}x^2)^n\frac{\Gamma(a-\frac{D}{2}-n)}{(\frac{1}{2}m^2)^{a-D/2-n}}$$

A more complicated example is

$$\check{\mathcal{A}}_1(a,b,m^2) = \int dk\ \frac{\Gamma(a)}{(\frac{1}{2}k^2)^a}\frac{\Gamma(b)}{[\frac{1}{2}(k^2+m^2)]^b}$$

$$= \int_0^\infty d\tau_1 d\tau_2\ \tau_1^{a-1}\tau_2^{b-1}\int dk\ e^{-[\tau_1 k^2+\tau_2(k^2+m^2)]/2}$$

$$= \int_0^\infty d\tau_1 d\tau_2\ \tau_1^{a-1}\tau_2^{b-1}(\tau_1+\tau_2)^{-D/2}e^{-\tau_2 m^2/2}$$

We then introduce a scaling parameter λ (also described in VIIA2), scaling $\tau_i = \lambda\alpha_i$ in the insertion

$$1 = \int_0^\infty d\lambda\ \delta(\lambda-\tau_1-\tau_2) = \int_0^\infty d\lambda\ \lambda^{-1}\delta(1-\alpha_1-\alpha_2)$$

and integrating the δ over α_2 to get $\alpha_2 = 1 - \alpha_1$. This gives (with $\alpha_1 = \alpha$)

$$\check{\mathcal{A}}_1(a,b,m^2) = \int_0^1 d\alpha\ \alpha^{a-1}(1-\alpha)^{b-1}\int_0^\infty d\lambda\ \lambda^{a+b-D/2-1}e^{-\lambda(1-\alpha)m^2/2}$$

$$= \frac{\Gamma(a+b-\frac{D}{2})}{(\frac{1}{2}m^2)^{a+b-D/2}}B(a, \tfrac{D}{2}-a)$$

When two of the propagators in the two-loop tadpole graph have the same non-vanishing mass, we consider directly the two-loop integral

$$\mathcal{A}_{1,2}(a,b,c,m^2) = \int dk_1 dk_2\ \frac{\Gamma(a)}{[\frac{1}{2}(k_1+k_2)^2]^a}\frac{\Gamma(b)}{[\frac{1}{2}(k_1^2+m^2)]^b}\frac{\Gamma(c)}{[\frac{1}{2}(k_2^2+m^2)]^c}$$

(This integral also represents the physically less interesting 2-loop "vacuum bubble": no external lines, and thus field independent.) Introducing the Schwinger parameters and performing the momentum integration, we find

$$\int_0^\infty d^3\tau\ \tau_1^{a-1}\tau_2^{b-1}\tau_3^{c-1}[\tau_2\tau_3+\tau_1(\tau_2+\tau_3)]^{-D/2}e^{-(\tau_2+\tau_3)m^2/2}$$

where we have included the power of *det* A for

$$A = \begin{pmatrix} \tau_1+\tau_2 & \tau_1 \\ \tau_1 & \tau_1+\tau_3 \end{pmatrix}$$

Since τ_1 does not appear in the exponential we integrate over it first directly, using the second integral form for the Beta function, from exercise VIIA2.2c. Then τ_2 and τ_3 can be handled by introducing a scaling parameter for them only, leading to the previous types of integrals. The result is then

$$\mathcal{A}_{1,2}(a,b,c,m^2) = \frac{\Gamma(a+b+c-D)}{(\frac{1}{2}m^2)^{a+b+c-D}}B(a+b-\tfrac{D}{2}, a+c-\tfrac{D}{2})B(a,\tfrac{D}{2}-a)$$

2. Effective potential

A propagator in an external field represents a certain class of Feynman tree diagrams. Thus, some tree graphs can be described by quantum mechanics. (In principle this means we can start from classical mechanics and first-quantize, by either operator or path-integral methods. However, as we'll see in chapter XII, in practice we save some effort if we start directly with the quantum mechanics.) If we take the ends of such a propagator and sew them together, we can describe arbitrary 1PI 1-loop graphs by the background field method. While tree graphs describe classical field theory, one-loop graphs contain many of the important quantum properties, partly because they are the lowest-order quantum correction, and partly because they are associated with the functional determinant part of the (second-quantized) path integral. (In terms of the exponent, classical is the only negative power in $\hbar$, 1-loop is $\hbar$-independent, and higher loops are positive powers.)

In quantum mechanics, the expansion in $\hbar$ is an expansion in derivatives (since it appears only as $p_a = -i\hbar\partial_a$). In terms of the contribution of one-loop graphs to the effective action, this means an expansion in the number of derivatives acting on the fields. This definition can be applied in general in quantum field theory, without reference to quantum mechanics. However, the simplest one-loop calculations of this expansion are most easily expressed in quantum mechanical terms. In practice, this means expanding the external fields in x about some fixed point, expanding the exponentiated (by a Schwinger parameter) propagator about the part Gaussian in p and x, and using any of the usual methods to exactly evaluate the matrix element of a polynomial times a Gaussian.

Since we generally want arbitrary orders in a field and some of its lower derivatives for this method to have any advantage over the usual diagrammatic methods, in this approach one generally cuts off the expansion at the approximation that gives just the Gaussian. This means we can keep up to two derivatives of an external scalar, but only a constant field strength for an external gauge vector. (See subsection VIB1.) The simplest, and most useful, example is a constant scalar field. The part of the effective action that consists of only scalars without derivatives is called the "effective potential", since it generalizes the potential term of the classical action. This potential determines the quantum corrections to spontaneous symmetry breaking and the Higgs effect, and this is important for describing mass generation for all spins.

Consider a complex scalar running around a loop, under the influence of an external real scalar. The Lagrangian is

$$L = \psi^*[\tfrac{1}{2}(-\Box + m^2) + \phi]\psi + L_\phi$$

where the form of L_ϕ won't be important for calculating the ψ loop. A constant external scalar field is effectively the same as a mass term, modifying $m^2 \to m^2 + 2\phi$. Thus the effective potential in this case can be evaluated by summing tadpoles:

$$V = -\sum_{n=1}^{\infty} \tfrac{1}{n}(-1)^n \phi^n \int dp\ [\tfrac{1}{2}(p^2 + m^2)]^{-n}$$

for our complex scalar; for a real scalar running around the loop there would be an extra factor of 1/2. We can integrate before summing:

$$V = -\sum_{n=1}^{\infty}(-1)^n \phi^n \frac{\Gamma(n - \frac{D}{2})}{n!}(\tfrac{1}{2}m^2)^{-n+D/2}$$

Using the identities (from Taylor expansion in a/b, and $\Gamma(z+1) = z\Gamma(z)$)

$$(a+b)^x = \sum_{n=0}^{\infty} \binom{x}{n} a^n b^{x-n}, \quad \binom{x}{n} = \frac{\Gamma(x+1)}{n!\Gamma(x+1-n)} = (-1)^n \frac{\Gamma(n-x)}{n!\Gamma(-x)}$$

we have

$$V = -\Gamma(-\tfrac{D}{2}) \left[(\tfrac{1}{2}m^2 + \phi)^{D/2} - (\tfrac{1}{2}m^2)^{D/2} \right]$$

We can also integrate after summing: Using the identities

$$ln(a+b) - ln\ b = \int_0^a \frac{du}{u+b} = \int_0^a du\ \frac{1}{b} \sum_{n=0}^{\infty} (-1)^n \left(\frac{u}{b}\right)^n = -\sum_{n=1}^{\infty} \tfrac{1}{n}(-1)^n a^n b^{-n}$$

$$\int_0^a \frac{du}{u+b} = \int_0^a du \int_0^{\infty} d\tau\ e^{-\tau(u+b)} = -\int_0^{\infty} \frac{d\tau}{\tau} \left(e^{-\tau(a+b)} - e^{-\tau b}\right)$$

we have

$$V = -\int_0^{\infty} \frac{d\tau}{\tau} \int dp \left(e^{-\tau[\phi + (p^2+m^2)/2]} - e^{-\tau(p^2+m^2)/2} \right)$$

which gives the same result.

For D=4, we find (after subtracting divergent counterterms, and some corresponding finite pieces, corresponding to a MOM type of subtraction)

$$V = \tfrac{1}{2}(\tfrac{1}{2}m^2 + \phi)^2 ln\left(1 + \frac{2\phi}{m^2}\right)$$

Since this modifies the classical potential, it demonstrates that quantum effects can generate spontaneous symmetry breaking where there was none classically, or vice versa (the "Coleman-Weinberg mechanism").

Exercise VIIB2.1

Generalize this renormalized result to arbitrary even dimensions.

For more complicated cases we need a more general procedure: The basic idea is that any Gaussian integral gives a (inverse) determinant, of which we must take (minus) the logarithm for the effective action, and we use $ln\ det = tr\ ln$. (The trace includes integration over x or p.) After subtracting out the field-independent part (vacuum bubble), this gives an expression as above: For a general kinetic operator $H = H_0 + ...$ (generally $H_0 = \frac{1}{2}(p^2+m^2)$), we want

$$\Gamma = -\left[tr\ ln(H^{-1}) - tr\ ln(H_0^{-1})\right] = -\int_0^{\infty} \frac{d\tau}{\tau} \int dx\ \langle x|e^{-\tau H} - e^{-\tau H_0}|x\rangle$$

H (and H_0) is now treated as an operator, in terms of the coordinate operator X and momentum operator P, and $X|x\rangle = x|x\rangle$. External fields depend on X, but are Taylor expanded about x: e.g.,

$$\phi(X) = \phi(x) + (X-x)\cdot\partial\phi(x) + ...$$

We then can use translation invariance to write

$$\langle x|e^{-\tau H[P,X-x,\phi(x)]}|x\rangle = \langle 0|e^{-\tau H[P,X,\phi(x)]}|0\rangle$$

When H is quadratic in P and X, we can use (see exercise VA2.4)

$$\langle x|e^{-\tau H}|y\rangle = \sqrt{det\ \frac{\partial^2(-S)}{\partial x \partial y}}\ e^{-S}$$

where S is the classical "action" corresponding to the "Hamiltonian" H. (Further examples will be given in subsection VIIIB1.)

Note that the (one-loop) vacuum bubble, with no background fields of any kind, must always be dropped, as it is totally meaningless (although how it is subtracted may be regularization dependent): In terms of the graphs summed here, which have equal numbers n of propagators and vertices ($P - V = L - 1$ by the usual $\hbar$ counting), it is the term n=0. Thus, in a coordinate space calculation, where there are also n integrations $d^D x$, this term would have no propagators, no vertices, and no integrals (contrary to some statements in the literature, where this graph is misidentified as a one-propagator graph with one integration). All that remains is the permutation factor, $1/n$, but in this case that is an undefined $1/0$.

In actual applications, closer examination reveals the used graph to be the cut 1-loop tadpole ($P = V = L = 1$). Since the cut propagator gives a sum over states, the result is to evaluate the trace of the operator inserted at the vertex; in particular, a trivial vertex yields $str(I)$, i.e., the number of states, bosons minus fermions. Similar use can be made of the cut propagator correction ($P = V = 2$) for (super)traces of operator products or mass sum rules.

3. Dimensional transmutation

The 2D version of the CP(n) model described in subsection IVA2 is an interesting model in that it demonstrates generation of bound states at the one-loop level. Its Lagrangian is:

$$L = \tfrac{1}{2}|\nabla\phi|^2 + \Lambda(|\phi|^2 - \tfrac{1}{g^2})$$

where g is now dimensionless. For the effective potential for the Lagrange multiplier Λ from a ϕ loop, we find (modifying the calculation of the previous subsection for D=2)

$$V_1 = -\Lambda\left[ln\left(\frac{\Lambda}{\frac{1}{2}\mu^2}\right) - 1\right]$$

after including the renormalization mass scale μ to make the argument of the logarithm dimensionless, and the coupling dimensionless in all dimensions.

Now the coupling can be absorbed into the definition of this scale: Adding to the classical term $V_0 = -\Lambda/g^2$, the total effective potential for Λ up to one loop is

$$V = -\Lambda\left[\frac{1}{g^2} + ln\left(\frac{\Lambda}{\frac{1}{2}\mu^2}\right) - 1\right] = -\Lambda\left[ln\left(\frac{\Lambda}{\frac{1}{2}M^2}\right) - 1\right]$$

where M is the "renormalization group invariant mass scale":

$$M^2 = \mu^2 e^{-1/g^2}$$

Since this was the only place the coupling g appeared in the action, the mass scale M has now replaced it completely. This replacement of a dimensionless coupling (g) with a dimensionful one (M) is called "dimensional transmutation". It is also a common feature of quantum high-energy behavior (see below); its importance at low energies depends on whether the classical theory already has dimensionful parameters (like masses).

Varying the effective potential to find the minimum, which we identify as the (quantum) vacuum value of the field Λ,

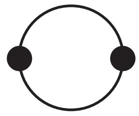
$$ln\left(\frac{\langle\Lambda\rangle}{\frac{1}{2}M^2}\right) = 0 \quad \Rightarrow \quad \langle\Lambda\rangle = \tfrac{1}{2}M^2$$

Because Λ has a vacuum value, ϕ now has a mass (as seen by expanding Λ about its vacuum value). Furthermore, since Λ now has more than just linear terms in the effective action, it is no longer a Lagrange multiplier. In fact, by calculating a massive ϕ loop with two external Λ's, we see that Λ is now a massive physical scalar also. Without a Lagrange multiplier, ϕ is now unconstrained, so it has an additional physical degree of freedom. This leads to a restoration of the spontaneously broken U(N) symmetry. This is related to ϕ gaining mass, since we no longer have Goldstone bosons associated with the symmetry breaking. Finally, if we calculate a massive ϕ loop with two external gauge vectors, we see that at low energies there is an F^2 term, so A is now a physical, massive vector instead of an auxiliary field.

Exercise VIIB3.1

Expand V about $\langle\Lambda\rangle$ to show that Λ gets a mass term. Expand $\Lambda(x)$ to quadratic order in x according to the prescription of the previous subsection to calculate the effective action in terms of Λ, $\partial\Lambda$, and $\partial\partial\Lambda$ (using the harmonic oscillator result of exercise VA2.4) to show that a $\Lambda\Box\Lambda$ term is also generated, so Λ becomes propagating.

4. Massless propagators



For the massless one-loop propagator corrections, we also introduce a scaling parameter to convert to Feynman parameters (see the examples of subsection VIIB1, or the general method in subsection VIIA2), with the result

$$\begin{aligned}\mathcal{A}_2(x,p^2) &= \int dk\ e^{ik\cdot x}\frac{1}{\frac{1}{2}(k+\frac{1}{2}p)^2\frac{1}{2}(k-\frac{1}{2}p)^2}\\ &= \int_0^\infty d\lambda\ \lambda^{1-D/2}\int_0^1 d\alpha_1 d\alpha_2\ \delta(1-\alpha_1-\alpha_2)\times\\ &\quad\times exp\{-\lambda\tfrac{1}{8}p^2[1-(\alpha_1-\alpha_2)^2]-i\tfrac{1}{2}(\alpha_1-\alpha_2)p\cdot x-\lambda^{-1}\tfrac{1}{2}x^2\}\end{aligned}$$

Making the change of variables

$$\alpha_1 = \tfrac{1}{2}(1+\beta), \qquad \alpha_2 = \tfrac{1}{2}(1-\beta)$$

the amplitude takes the form

$$\mathcal{A}_2 = \int_0^1 d\beta\ \tfrac{1}{2}(e^{i\beta p\cdot x/2}+e^{-i\beta p\cdot x/2})\int_0^\infty d\lambda\ \lambda^{1-D/2}exp[-\lambda\tfrac{1}{8}(1-\beta^2)p^2-\lambda^{-1}\tfrac{1}{2}x^2]$$

The integrals can be simplified if we make use of gauge invariance: For example, the electromagnetic current for a complex scalar is of the form $\phi^*\overleftrightarrow{\partial}\phi$, so the gauge field couples to the difference of the momenta of the two scalar lines, which is $2k$ for the above as applied to the scalar-loop correction to the photon propagator. On the other hand gauge invariance, or equivalently current conservation, says that such a vertex factor should give a vanishing contribution when contracted with the external momentum, which is p in that case. Checking this explicitly, we do in fact find

$$\int dk\ \frac{k\cdot p}{\frac{1}{2}(k+\frac{1}{2}p)^2\frac{1}{2}(k-\frac{1}{2}p)^2} = \int dk\ \left[\frac{1}{\frac{1}{2}(k-\frac{1}{2}p)^2}-\frac{1}{\frac{1}{2}(k+\frac{1}{2}p)^2}\right] = 0$$

(even with an arbitrary additional polynomial factor in the numerator), using the facts that the integral of the sum is the sum of the integrals when regularized, and that massless tadpoles vanish. (This also tells us that replacing the numerator $k\cdot p$ with $k^2+\frac{1}{4}p^2$ gives 0. Furthermore, without an extra numerator factor the integral vanishes by antisymmetry under $k\to -k$.) Thus, if x is proportional to p in A_2, the only contribution is from the $x=0$ term in the Taylor expansion. This implies that the dependence on x is only through the combination

$$u = (p\cdot x)^2 - p^2x^2$$

so we can evaluate the integral by either of the substitutions

$$x^2 \to 0, \quad p\cdot x \to \sqrt{u} \qquad or \qquad p\cdot x \to 0, \quad x^2 \to -u/p^2$$

We'll consider now the latter choice. (The former gives the same result: See the exercise below.) Again, since we need to Taylor expand in x anyway to find the result for a particular numerator, we expand and perform the λ integration:

$$\mathcal{A}_2 = \sum_{n=0}^{\infty} \tfrac{1}{n!}\left(\frac{u}{2p^2}\right)^n (\tfrac{1}{8}p^2)^{n+D/2-2}\Gamma(2-\tfrac{D}{2}-n)\int_0^1 d\beta\ (1-\beta^2)^{n+D/2-2}$$

Performing the change of variables $\beta^2=\gamma$ to convert the remaining integral to a Beta function, and using the identities

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi}, \qquad \Gamma(z)\Gamma(1-z) = \pi\ csc(\pi z)$$

(see the exercises in subsection VIIA2), the final result is

$$\mathcal{A}_2 = -\tfrac{1}{2}\pi^{3/2}csc(D\tfrac{\pi}{2})(\tfrac{1}{8}p^2)^{D/2-2}\sum_{n=0}^{\infty}\frac{1}{n!\Gamma(n+\frac{D}{2}-\frac{1}{2})}\{\tfrac{1}{16}[p^2x^2-(p\cdot x)^2]\}^n$$

From the *csc* factor we see the integral is divergent for all even D: These are ultraviolet divergences for $D \geq 4$ and infrared ones for $D \leq 4$; dimensional regularization does not carefully distinguish between the two, although the difference can usually be told by examining momentum dependence (here from the exponent $D/2-2$). Also notice that the two can be mixed up by the conversion to Feynman parameters.

Exercise VIIB4.1

Evaluate the general massless one-loop propagator correction using $x^2 \to 0$, $p\cdot x \to \sqrt{u}$.

a Show it gives the same result as $p\cdot x \to 0$, $x^2 \to -u/p^2$ by using the Γ and B identities in subsection VIIA2.

b Show it can also be written as (for convenience of expansion about D=4)

$$\mathcal{A}_2 = (\tfrac{1}{2}p^2)^{D/2-2}\Gamma(\tfrac{D}{2}-1)\Gamma(2-\tfrac{D}{2})\sum_{n=0}^{\infty}\frac{\Gamma(n+\frac{D}{2}-1)}{n!\Gamma(2n+D-2)}\{\tfrac{1}{4}[p^2x^2-(p\cdot x)^2]\}^n$$

As discussed in subsection VIIB1, sometimes certain subdiagrams of higher-loop diagrams can be evaluated explicitly, particularly propagator corrections that themselves involve only massless propagators. Furthermore, such a formula might be used recursively in appropriate diagrams. For example, a higher-loop diagram that is itself a propagator correction might reduce, as a final integration, to something of the form

$$\int dk \; \frac{\Gamma(a)}{(\frac{1}{2}k^2)^a} \frac{\Gamma(b)}{[\frac{1}{2}(k+p)^2]^b} = \frac{\Gamma(a+b-\frac{D}{2})}{(\frac{1}{2}p^2)^{a+b-D/2}} B(\tfrac{D}{2}-a, \tfrac{D}{2}-b)$$

again using the above methods, finding similar integrals to the previous.

Exercise VIIB4.2

Let's examine this integral more carefully.

a Evaluate it in two different ways: first, by the method used above; second, by Fourier transforming each factor using

$$\int dk \; e^{ik\cdot x} \frac{\Gamma(a)}{(\frac{1}{2}k^2)^a} = \frac{\Gamma(\frac{D}{2}-a)}{(\frac{1}{2}x^2)^{D/2-a}}$$

(derive this also) and its inverse, simply multiplying the resulting factors in x space, and inverse transforming.

b Show that the $\overline{\mathrm{MS}}$ scheme cancels γ's and $\zeta(2)$'s in iterated massless propagator corrections to all orders in ϵ by examining

$$\Gamma(\tfrac{D}{2}) \int dk \; \frac{1}{(\frac{1}{2}k^2)^{n_1+L_1\epsilon}[\frac{1}{2}(k+p)^2]^{n_2+L_2\epsilon}}$$

where L_i are the numbers of loops in the propagator subgraphs (show this by dimensional analysis) and n_i are other integers. Show the G scheme does the same.

Exercise VIIB4.3

Calculate the "phase space" for n massless particles

$$V_P = \int \left[\prod^n \frac{d^{D-1}p_i}{(2\pi)^{D/2-1}\omega_i} \right] (2\pi)^{D/2} \delta^D \left(p - \sum^n p_i \right)$$

where p is the total momentum of the n particles, by using the optical theorem:

a Consider the scalar graph with n massless propagators connecting 2 vertices. Show, both by induction in the number $(n-1)$ of loops, and by Fourier transformation (as in the previous problem), that this graph (for distinguishable particles) gives

$$\frac{[\Gamma(\frac{D}{2}-1)]^n}{\Gamma[n(\frac{D}{2}-1)]} \frac{\Gamma[n-(n-1)\frac{D}{2}]}{(\frac{1}{2}p^2)^{n-(n-1)D/2}}$$

b Wick rotate back to Minkowski space ($p^2 < 0$) and take (twice) the imaginary part to obtain the result for continuous real $D > 2$

$$V_P = 2\pi \frac{[\Gamma(\frac{D}{2}-1)]^n}{\Gamma[n(\frac{D}{2}-1)]\Gamma[(n-1)(\frac{D}{2}-1)]}(-\tfrac{1}{2}p^2)^{-n+(n-1)D/2}$$

which simplifies in D=4 to

$$V_P = 2\pi \frac{1}{(n-1)!(n-2)!}(-\tfrac{1}{2}p^2)^{n-2}$$

(Hint: $(\frac{1}{2}p^2 - i\epsilon)^r = (-\frac{1}{2}p^2)^r e^{-i\pi r}$.)

5. Massive propagators

Another way to distinguish infrared divergences is by introducing masses (being careful not to break any invariances, or restoring them in the massless limit). For example, we again evaluate the one-loop propagator correction, without numerator factors, but with different masses on the internal propagators. By the same steps as before, the Feynman parameter integral is

$$\hat{\mathcal{A}}_2(p^2, m_1^2, m_2^2) = \int dk \; \frac{1}{\frac{1}{2}[(k+\frac{1}{2}p)^2 + m_1^2]\frac{1}{2}[(k-\frac{1}{2}p)^2 + m_2^2]}$$

$$= \Gamma(2-\tfrac{D}{2})\tfrac{1}{2}\int_{-1}^{1} d\beta \; \mathcal{B}^{D/2-2}, \qquad \mathcal{B} = \tfrac{1}{8}p^2(1-\beta^2) + \tfrac{1}{4}(m_1^2+m_2^2) + \tfrac{1}{4}\beta(m_1^2 - m_2^2)$$

Now the β integral is harder for all D, but the masses eliminate the IR divergences (and the UV divergences are already explicit in the Γ), so we immediately expand about $D = 4 - 2\epsilon$:

$$\hat{\mathcal{A}}_2 \approx \Gamma(\epsilon)\tfrac{1}{2}\int_{-1}^{1} d\beta \; (1 - \epsilon \; ln \; \mathcal{B})$$

We then use integration by parts

$$\int_{-1}^{1} d\beta \; ln \; \mathcal{B} = (\beta \; ln \; \mathcal{B})|_{-1}^{1} - \int_{-1}^{1} d\beta \; \beta \frac{d}{d\beta} ln \; \mathcal{B}$$

$$\mathcal{B} = a\beta^2 + b\beta + c = a(\beta-\beta_+)(\beta-\beta_-), \quad \beta_\pm = \frac{-b \pm \sqrt{b^2-4ac}}{2a} = \frac{m_1^2 - m_2^2 \pm 2\lambda_{12}}{p^2}$$

$$\Rightarrow \quad \beta\frac{d}{d\beta} ln \; \mathcal{B} = \frac{\beta}{\beta-\beta_+} + \frac{\beta}{\beta-\beta_-} = 2 + \frac{\beta_+}{\beta-\beta_+} + \frac{\beta_-}{\beta-\beta_-}$$

in terms of $\lambda_{12}(s)$ of subsection IA4 for $s = -p^2$. Note that in Euclidean space

$$2\lambda_{12} = \sqrt{(p^2+m_1^2-m_2^2)^2 + 4m_2^2p^2} = \sqrt{(p^2+m_2^2-m_1^2)^2 + 4m_1^2p^2} \geq p^2 + |m_1^2 - m_2^2|$$

$$\Rightarrow \quad \pm\beta_\pm \geq 1$$

where the strict inequality holds for both masses nonvanishing. The integrals then take the simple form

$$\int_{-1}^{1} d\beta \, \left(\frac{\beta_+}{\beta - \beta_+} + \frac{\beta_-}{\beta - \beta_-}\right) = \beta_+ ln \left(\frac{\beta_+ - 1}{\beta_+ + 1}\right) + \beta_- ln \left(\frac{\beta_- - 1}{\beta_- + 1}\right)$$

Putting it all together,

$$\hat{\mathcal{A}}_2 = \Gamma(1+\epsilon)\left[\frac{1}{\epsilon} - ln(\tfrac{1}{2}m_1 m_2) + 2 + \tfrac{1}{2}\beta_+ ln\left(\frac{\beta_+ - 1}{\beta_+ + 1}\right) + \tfrac{1}{2}\beta_- ln\left(\frac{\beta_- - 1}{\beta_- + 1}\right)\right]$$

(We can cancel the $\Gamma(1+\epsilon)$ by nonminimal subtraction.) By analytic continuation from Euclidean space, taking p^2 from positive to negative along the real axis, we see there is no ambiguity at $p^2 = 0$ or $-(m_1 - m_2)^2$, and $\hat{\mathcal{A}}_2$ remains real until we reach $p^2 = -(m_1 + m_2)^2$, where it gets an imaginary part (whose sign is determined by $(m_1 + m_2)^2 \to (m_1 + m_2)^2 - i\epsilon$), corresponding to the possibility of real 2-particle intermediate states.

Exercise VIIB5.1

Let's consider some special cases:

a Show for equal masses $m_1 = m_2 = m$ that this result simplifies to

$$\hat{\mathcal{A}}_2(p^2, m^2, m^2) = \Gamma(1+\epsilon)\left[\frac{1}{\epsilon} - ln(\tfrac{1}{2}m^2) + 2 + \beta \; ln\left(\frac{\beta - 1}{\beta + 1}\right)\right]$$

$$\beta = \sqrt{\frac{p^2 + 4m^2}{p^2}}$$

b Consider the case with one internal particle massless, $m_1 = m$, $m_2 = 0$, and find

$$\hat{\mathcal{A}}_2(p^2, m^2, 0) = \Gamma(1+\epsilon)\left[\frac{1}{\epsilon} - ln(\tfrac{1}{2}m^2) + 2 - \frac{p^2 + m^2}{p^2} ln\left(\frac{p^2 + m^2}{m^2}\right)\right]$$

c Show both these results agree with the previously obtained massless result in the limit $m \to 0$. However, note that both these cases, unlike the massless case, are IR convergent at $p^2 = -(m_1 + m_2)^2$.

Exercise VIIB5.2

Find the phase space for 2 massive particles, again using the optical theorem (as in exercise VIIB4.3):

a The calculation is easier if one takes the imaginary part before performing the Feynman parameter integration: Show the result is then

$$V_P = \pi \frac{1}{\Gamma(\frac{D}{2} - 1)} \int_{\beta_+}^{\beta_-} d\beta \; (-\mathcal{B})^{D/2-2}$$

In particular, show from the explicit parameter integral expression for the propagator that the only cut is at $-p^2 \geq (m_1+m_2)^2$, as expected from the optical theorem.

b Make the change of variables

$$\beta \to \frac{\beta - \beta_+}{\beta_- - \beta_+}$$

to find the result

$$V_P = 2\pi \frac{\Gamma(\frac{D}{2}-1)}{\Gamma(D-2)}(-\tfrac{1}{2}p^2)^{1-D/2}\lambda_{12}^{D-3}$$

which simplifies in D=4 to

$$V_P = 2\pi \frac{\lambda_{12}}{-\frac{1}{2}p^2}$$

c Show this result (in all D) agrees with the result of the explicit phase space integral of subsection VC7, using the expression for $\int d^{D-2}\Omega$ from subsection VIIA3. (Hint: Use the identity from exercise VIIA2.2b.)

d Show the massless case agrees with exercise VIIB4.3.

In subsection VIIA3 we considered the application of the MOM subtraction scheme to propagator corrections. We assumed the propagator corrections were Taylor expandable in the classical kinetic operator. From the above explicit expression for the 1-loop correction in scalar theories, we see this is possible except near the branch point at $p^2 = -(m_1+m_2)^2$, i.e., when the external particle (whose propagator we're correcting) has a mass equal to the sum of the internal ones. To analyze this more carefully, let's recalculate the propagator correction, performing the Taylor expansion before evaluating the integrals. We consider the case with one vanishing mass, $m_1 = m$, $m_2 = 0$, to generate an IR divergence. Assuming the external mass is also m, we expand around the branch point in p^2+m^2. The Feynman parameter integral is then, to linear order in p^2+m^2, in terms of $\alpha = \frac{1}{2}(1+\beta)$,

$$\hat{\mathcal{A}}_2(p^2,m^2,0) = \Gamma(\epsilon)\int_0^1 d\alpha \left[\tfrac{1}{2}m^2\alpha^2\left(1+\frac{1-\alpha}{\alpha}\frac{p^2+m^2}{m^2}\right)\right]^{-\epsilon}$$

$$\approx \Gamma(1+\epsilon)(\tfrac{1}{2}m^2)^{-\epsilon}\int_0^1 d\alpha\left[\frac{1}{\epsilon}\alpha^{-2\epsilon} - (1-\alpha)\alpha^{-1-2\epsilon}\frac{p^2+m^2}{m^2}\right]$$

$$= \Gamma(1+\epsilon)(\tfrac{1}{2}m^2)^{-\epsilon}\left[\frac{1}{\epsilon}\frac{1}{1-2\epsilon} - \left(\frac{1}{-2\epsilon} - \frac{1}{1-2\epsilon}\right)\frac{p^2+m^2}{m^2}\right]$$

$$\approx \Gamma(1+\epsilon)(\tfrac{1}{2}m^2)^{-\epsilon}\left[\left(\frac{1}{\epsilon_{UV}}+2\right) + \tfrac{1}{2}\left(\frac{1}{\epsilon_{IR}}+2\right)\frac{p^2+m^2}{m^2}\right]$$

where we have distinguished the UV divergence (in the λ integral for $\epsilon \geq 0$) from the IR one (in the α integral for $\epsilon \leq 0$). After including the $(\frac{1}{2}\mu^2)^\epsilon$ in the coupling, the $(\frac{1}{2}m^2)^{-\epsilon}$ converts each $1/\epsilon$ into a $1/\epsilon - ln(m^2/\mu^2)$. (Of course, we can choose $\mu = m$ for convenience.) Note that this infrared divergence was a consequence of trying to Taylor expand about a branch point due to a massless particle.

Exercise VIIB5.3

Do MOM subtraction for external mass $M = m_1 + m_2$, with *neither* internal mass vanishing, and show there is no divergence other than the UV divergence of the minimal scheme.

Later, we will encounter propagator corrections in gauge theories with massive internal lines, and with various numerators. Here, we examine these purely from the point of view of the integrals. First, consider

$$\mathcal{A}_a = \int dk \; \frac{k_a}{\frac{1}{2}[(k-\frac{1}{2}p)^2 + m_1^2]\frac{1}{2}[(k+\frac{1}{2}p)^2 + m_2^2]}$$

Since p is the only external momentum for a propagator, by Lorentz invariance we have

$$\mathcal{A}_a = p_a \frac{1}{p^2} p \cdot \mathcal{A}$$

so it is sufficient to evaluate the integral of $p \cdot \mathcal{A}$. In analogy with the earlier massless expression, we look at

$$\frac{1}{\frac{1}{2}[(k-\frac{1}{2}p)^2 + m_1^2]} - \frac{1}{\frac{1}{2}[(k+\frac{1}{2}p)^2 + m_2^2]} = \frac{k \cdot p + \frac{1}{2}(m_2^2 - m_1^2)}{\frac{1}{2}[(k-\frac{1}{2}p)^2 + m_1^2]\frac{1}{2}[(k+\frac{1}{2}p)^2 + m_2^2]}$$

from which we find

$$\mathcal{A}_a = p_a \frac{m_1^2 - m_2^2}{2p^2}[\hat{\mathcal{A}}_2(p^2, m_1^2, m_2^2) - \hat{\mathcal{A}}_2(0, m_1^2, m_2^2)]$$

in terms of our result $\hat{\mathcal{A}}_2$ above for the integral without numerator.

As a more complicated (but important) example, we examine

$$\mathcal{A}_{ab} = \int dk \; \frac{k_a k_b}{\frac{1}{2}[(k+\frac{1}{2}p)^2 + m^2]\frac{1}{2}[(k-\frac{1}{2}p)^2 + m^2]}$$

Following our procedure of the previous example, we note

$$\int dk \; \frac{(p \cdot k)k}{\frac{1}{2}[(k+\frac{1}{2}p)^2 + m^2]\frac{1}{2}[(k-\frac{1}{2}p)^2 + m^2]}$$

$$= \int dk \; \frac{k}{\frac{1}{2}[(k-\frac{1}{2}p)^2 + m^2]} - \frac{k}{\frac{1}{2}[(k+\frac{1}{2}p)^2 + m^2]}$$

$$= \int dk \; \frac{k+\frac{1}{2}p}{\frac{1}{2}(k^2+m^2)} - \frac{k-\frac{1}{2}p}{\frac{1}{2}(k^2+m^2)} = p\int dk \; \frac{1}{\frac{1}{2}(k^2+m^2)}$$

Thus transversality again determines the amplitude in terms of a scalar:

$$\hat{\mathcal{A}}_{ab} = \int dk \; \frac{k_a k_b}{\frac{1}{2}[(k+\frac{1}{2}p)^2+m^2]\frac{1}{2}[(k-\frac{1}{2}p)^2+m^2]} - \frac{\eta_{ab}}{\frac{1}{2}(k^2+m^2)}$$

$$= (\eta_{ab}p^2 - p_a p_b)\check{\mathcal{A}}(p^2,m^2)$$

(This amplitude actually will be more useful than $\mathcal{A}_{ab}$.) We also have the identity

$$\int dk \; \frac{\frac{1}{2}(k^2+m^2)+\frac{1}{8}p^2}{\frac{1}{2}[(k+\frac{1}{2}p)^2+m^2]\frac{1}{2}[(k-\frac{1}{2}p)^2+m^2]}$$

$$= \tfrac{1}{2}\int dk \; \frac{1}{\frac{1}{2}[(k-\frac{1}{2}p)^2+m^2]} + \frac{1}{\frac{1}{2}[(k+\frac{1}{2}p)^2+m^2]} = \int dk \; \frac{1}{\frac{1}{2}(k^2+m^2)}$$

Taking the trace of the previous expression,

$$(D-1)p^2\check{\mathcal{A}}(p^2) = \int dk \; \frac{k^2}{\frac{1}{2}[(k+\frac{1}{2}p)^2+m^2]\frac{1}{2}[(k-\frac{1}{2}p)^2+m^2]} - \frac{D}{\frac{1}{2}(k^2+m^2)}$$

$$= -(\tfrac{1}{4}p^2+m^2)\int dk \; \frac{1}{\frac{1}{2}[(k+\frac{1}{2}p)^2+m^2]\frac{1}{2}[(k-\frac{1}{2}p)^2+m^2]} - (D-2)\int dk \; \frac{1}{\frac{1}{2}(k^2+m^2)}$$

$$= -(\tfrac{1}{4}p^2+m^2)\hat{\mathcal{A}}_2(p^2,m^2,m^2) - (D-2)\mathcal{A}_1(0,m^2)$$

in terms of the ϕ^3 propagator and tadpole graphs evaluated earlier. This result can be reorganized if we make use of the $p=0$ case:

$$0 = -m^2\hat{\mathcal{A}}_2(0,m^2,m^2) - (D-2)\mathcal{A}_1(0,m^2)$$

(which also follows easily from the earlier explicit expression for $\hat{\mathcal{A}}_1(a,0,m^2)$). We then find

$$\check{\mathcal{A}} = -\tfrac{1}{4(D-1)}\hat{\mathcal{A}}_2(p^2,m^2,m^2) - \tfrac{1}{D-1}m^2\frac{\hat{\mathcal{A}}_2(p^2,m^2,m^2)-\hat{\mathcal{A}}_2(0,m^2,m^2)}{p^2}$$

Exercise VIIB5.4

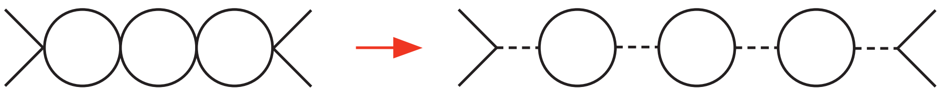
Check that these results are consistent in the massless limit with the expressions obtained in the previous subsection, by relating the first two terms in $\mathcal{A}_2(x,p^2)$ for arbitrary D.

Exercise VIIB5.5

Calculate the one-loop propagator corrections for Λ and A in the 2D CP(n) model.

6. Renormalization group

An interesting, useful, and simple application of the propagator correction is to study the high-energy behavior of coupling constants. For example, we have seen that, by a change in normalization of gauge fields $A \to A/g$, gauge couplings can be moved from the covariant derivative to the kinetic term: $\nabla = \partial + igA \to \partial + iA$, $L_0 = \frac{1}{8}(\partial A + igAA)^2 \to \frac{1}{8g^2}(\partial A + iAA)^2$. Thus, quantum corrections to gauge couplings can be found from just the propagator (kinetic-operator) correction. A simpler example is a scalar field; a ϕ^4 self-interaction has a dimensionless coupling in D=4, like Yang-Mills. However, unlike Yang-Mills, this model has no cubic coupling, and thus no 1-loop propagator correction. Furthermore, in Yang-Mills the one-loop propagator correction contribution to the effective action gives a multiloop contribution to the propagator itself, from the expansion of $1/(K+A)$. This corresponds to the graph consisting of a long string of these corrections connected by free propagators. There is a 1-loop 4-point correction in ϕ^4 theory, and this graph resembles a propagator correction, but with two external lines at each vertex instead of one. Such corrections can also be strung together, resembling the Yang-Mills string, but with no free propagators inserted. Since all the intermediate states in this graph are 2-particle, it is 1PI, so the effect of this string is not contained in just the 1-loop effective action, even though it is an iteration of a 1-loop effect.



This difficulty can be avoided by introducing the ϕ^4 interaction through an auxiliary field, just as it appears in supersymmetric theories (see subsection IVC2):

$$L = \tfrac{1}{2}\phi(-\tfrac{1}{2}\Box + \chi)\phi - \tfrac{1}{2g}\chi^2$$

where we have neglected the mass term since we will be concentrating on the high-energy behavior. Here the coupling is introduced through the *auxiliary-field* "kinetic" term. The diagrams just discussed now appear through the 1-loop correction to the auxiliary-field propagator: Since its free propagator is just a constant, it can be contracted to a point in these multiloop diagrams. The definition of 1PI graphs has now changed, since we can now cut auxiliary-field propagators, which would not exist in the usual ϕ^4 form of the action. This modification of the effective action simplifies the analysis of quantum corrections to the coupling, as well as making it more analogous to gauge theories. Note in particular the change in interpretation already at the tree level: We have used the conventional normalization of 1/n! for factors of ϕ^n in the potential, since canceling factors of n! arise upon functional differentiation. However, the result of eliminating χ from the classical action produces $\frac{1}{8}\phi^4$ instead of $\frac{1}{24}\phi^4$. The reason is that in the diagrams with χ there are 3 graphs contributing to the 4-ϕ-point tree, corresponding to χ propagators in the s, t, and u "channels". (See subsection VC4.) Although this is a trivial distinction for the trees, this is not the case for the loops, where the propagator string consists of pairs of ϕ particles running in one of these three channels.

The contribution to the 1-loop effective action for χ is then given by the above calculations, after including the factors of 1/2 for symmetries of the internal and external lines, and the usual -1 for the effective action:

$$L' = -\tfrac{1}{4}\chi\left[\Gamma(\epsilon)B(1-\epsilon,1-\epsilon)(-\tfrac{1}{2}\Box)^{-\epsilon}\right]\chi$$

where as usual $\epsilon = 2 - D/2$. Expressing the Beta function in terms of the Gamma function, and expanding as in previous subsections,

$$\Gamma(\epsilon)B(1-\epsilon,1-\epsilon)(-\tfrac{1}{2}\Box)^{-\epsilon} \approx \tfrac{1}{\epsilon} + [-\gamma + 2 - ln(-\tfrac{1}{2}\Box)]$$

Renormalizing away the constant pieces, we find for the classical action plus this part of the 1-loop effective action

$$L + L' = \tfrac{1}{2}\phi(-\tfrac{1}{2}\Box + \chi)\phi - \tfrac{1}{2}\chi\left[\tfrac{1}{g} - \tfrac{1}{2}ln\left(-\tfrac{\Box}{\mu^2}\right)\right]\chi = \tfrac{1}{2}\phi(-\tfrac{1}{2}\Box + \chi)\phi + \tfrac{1}{4}\chi\ ln\left(-\tfrac{\Box}{M^2}\right)\chi$$

where the renormalization group invariant mass scale M is given by

$$M^2 = \mu^2 e^{2/g}$$

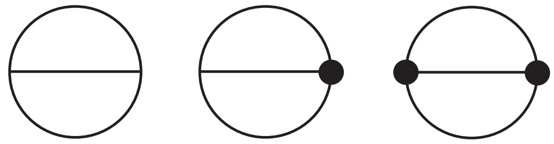
Thus, the constant coupling $1/g$ has been replaced by an effective "running coupling" $-\frac{1}{2}ln(p^2/M^2)$, with energy dependence set by the scale M. (This is sometimes called the "renormalization group", the group being related to scale invariance, which is broken by the introduction of the mass scale M.)

We saw the same dimensional transmutation occuring in the effective potential in massless theories in subsection VIIB3. The form is similar because both are related to the appearance of the renormalization mass scale μ from the breaking of scale invariance by quantum corrections, at either low or high energy: In both cases dimensional transmutation comes from a finite $ln\ \mu^2$ term arising from the infinite renormalization. The difference is that in the effective potential case we ignore higher derivatives, so the μ^2 must appear in a ratio to scalar fields, while in the high energy case we look at just the propagator correction, so it appears in the combination μ^2/p^2. (More complicated combinations will appear in more general amplitudes.)

Exercise VIIB6.1

Generalize this model to include internal symmetry: Write an analog to the scalar analog to QCD discussed in subsection VC9, where the "quark" ϕ now carries color and flavor indices, while the "gluon" χ (classically auxiliary) carries just color. Find M, especially its dependence on the numbers n of colors and m of flavors. Write the same model with the gluons replaced by "mesons" carrying just flavor indices (so that classical elimination of the auxiliary fields yields the same action), and repeat the calculation. What are the different approximation schemes relevant to the two approaches?

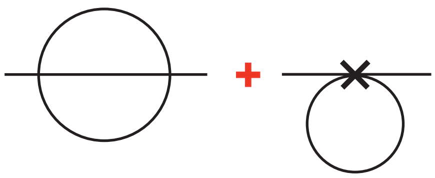
7. Overlapping divergences



We now perform some 2-loop renormalizations. Our first example is part of the propagator correction in ϕ^4 theory. By restricting ourselves to the mass renormalization (coefficient of the mass term), we need evaluate the graph only at vanishing external momentum. (It is then equivalent to a vacuum bubble in ϕ^3 theory, or a tadpole graph in the mixed theory.) Furthermore, we consider the case where some of the fields are massless. In such a theory, we encounter (a special case of) the 2-loop graph of subsection VIIB1, where 1 propagator is massless and 2 are massive. Expanding in ϵ, and keeping only the divergent terms ($1/\epsilon^2$ and $1/\epsilon$), we find (including a symmetry factor of 1/2 for the 2 massive scalar lines for real scalars)

$$\mathcal{T}_2 = \tfrac{1}{2}\frac{\Gamma(3-D)}{(\frac{1}{2}m^2)^{3-D}}B(2-\tfrac{D}{2},2-\tfrac{D}{2})B(1,\tfrac{D}{2}-1) = -\frac{(\frac{1}{2}m^2)^{1-2\epsilon}}{2(1-\epsilon)(1-2\epsilon)}[\Gamma(\epsilon)]^2$$

$$\approx \tfrac{1}{4}m^2[\Gamma(\epsilon)]^2[-1-3\epsilon+2\epsilon\, ln(\tfrac{1}{2}m^2)]$$



To this we need to add the counterterm graph, coming from inserting into the 1-loop massive tadpole $\mathcal{T}_1$ (with 2 external lines) the counterterm Δ_4 (for renormalizing the ϕ^4 term) from the 1-loop divergence in the 4-point graph with 1 massive and 1 massless propagator. (Since the massless tadpole vanishes in dimensional regularization, we need not consider the counterterm from the 4-point graph with 2 massive propagators.) From section VIIB5, we use the corresponding integral for a 1-loop propagator correction $\mathcal{A}$, which is

$$\mathcal{A} = \Gamma(\epsilon) + finite \quad \Rightarrow \quad \Delta_4 = -\Gamma(\epsilon)$$

We use a "modified minimal subtraction", using the $\Gamma(\epsilon)$ as the subtraction instead of just the $1/\epsilon$ part of $\Gamma(\epsilon) \approx 1/\epsilon - \gamma$.

The 1-loop massive tadpole without coupling is

$$\mathcal{T}_1 = \frac{\Gamma(1-\frac{D}{2})}{(\frac{1}{2}m^2)^{1-\frac{D}{2}}} = -\frac{(\frac{1}{2}m^2)^{1-\epsilon}}{1-\epsilon}\Gamma(\epsilon) \approx -\tfrac{1}{2}m^2\Gamma(\epsilon)[1+\epsilon-\epsilon\; ln(\tfrac{1}{2}m^2)]$$

Combining these results, the divergent part of the 2-loop propagator correction, with 1-loop coupling counterterm contributions included, is

$$\mathcal{T}_2 + \Delta_4\mathcal{T}_1 = \tfrac{1}{4}m^2[\Gamma(\epsilon)]^2[-1-3\epsilon+2\epsilon\; ln(\tfrac{1}{2}m^2)+2+2\epsilon-2\epsilon\; ln(\tfrac{1}{2}m^2)]$$

$$= [\Gamma(\epsilon)]^2(1-\epsilon)\tfrac{1}{4}m^2$$

Thus, the $ln\; m^2$ divergences cancel, as expected. (Divergences must be polynomial in masses as well as couplings.) The surviving divergence is the superficial divergence, to be canceled by the 2-loop mass counterterm.

Exercise VIIB7.1

Calculate the p^2 part of the 2-loop kinetic counterterm by writing the above 2-loop propagator graph with nonvanishing external momentum, introducing the Schwinger parameters, doing the loop-momentum integration, taking the derivative with respect to p^2, and then evaluating at $p=0$. Why is there no subdivergence $(1/\epsilon^2)$?

Exercise VIIB7.2

Calculate the *complete* (all graphs, infinite and finite parts of the) 2-loop propagator correction for *massless* ϕ^4. (See exercise VIIB4.3a.)

For our next example we consider massless ϕ^3 theory, and work in 6 dimensions, where the theory is renormalizable (instead of superrenormalizable, as in 4 dimensions). For the 2-loop propagator correction, there are only two graphs (plus 1-loop graphs with 1-loop counterterm insertions), one of which is simply a 1-loop propagator graph inserted into another. The other graph is

$$\mathcal{P} = \int dk\ dq\ \frac{1}{\frac{1}{2}(k+q)^2 \frac{1}{2}(k+\frac{1}{2}p)^2 \frac{1}{2}(k-\frac{1}{2}p)^2 \frac{1}{2}(q+\frac{1}{2}p)^2 \frac{1}{2}(q-\frac{1}{2}p)^2}$$

(with a symmetry factor of $\frac{1}{2}$ for real scalars). This graph can be rewritten as iterated propagator corrections by use of integration by parts in momentum space. This is legalized by dimensional regularization, since boundary terms vanish in low enough dimensions. All invariants can be expressed as linear combinations of the propagator denominators (there are 5 of each, not counting the square of the external momentum p^2), so any product of momentum times derivative acting on the integrand will give terms killing one denominator and squaring another, except for p^2 terms, which can be canceled by appropriate choice of the momentum multiplying the derivative:

$$\int dk\ dq\ \frac{\partial}{\partial k} \cdot \frac{k+q}{\frac{1}{2}(k+q)^2 \frac{1}{2}(k+\frac{1}{2}p)^2 \frac{1}{2}(k-\frac{1}{2}p)^2 \frac{1}{2}(q+\frac{1}{2}p)^2 \frac{1}{2}(q-\frac{1}{2}p)^2} = 0$$

This operation effectively gives the factor

$$\frac{\partial}{\partial k} \cdot (k+q) \to (D-4) + \frac{(q-\frac{1}{2}p)^2}{(k+\frac{1}{2}p)^2} + \frac{(q+\frac{1}{2}p)^2}{(k-\frac{1}{2}p)^2} - \frac{(k+q)^2}{(k+\frac{1}{2}p)^2} - \frac{(k+q)^2}{(k-\frac{1}{2}p)^2}$$

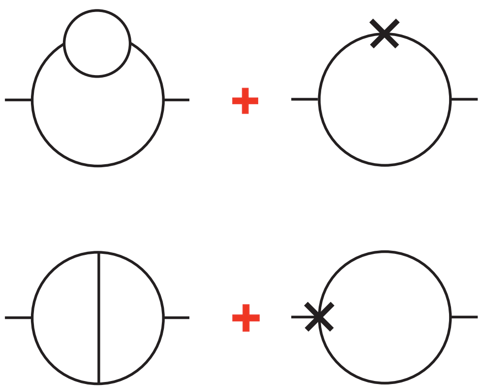
We thus have

$$(\tfrac{D}{2}-2)\mathcal{P} = \mathcal{P}_1 - \mathcal{P}_2$$

$$\mathcal{P}_1 = \int dk\ \frac{1}{[\frac{1}{2}(k+\frac{1}{2}p)^2]^2 \frac{1}{2}(k-\frac{1}{2}p)^2} \int dq\ \frac{1}{\frac{1}{2}(q+\frac{1}{2}p)^2 \frac{1}{2}(q-\frac{1}{2}p)^2}$$

$$\mathcal{P}_2 = \int dk\ \frac{1}{[\frac{1}{2}(k+\frac{1}{2}p)^2]^2 \frac{1}{2}(k-\frac{1}{2}p)^2} \int dq\ \frac{1}{\frac{1}{2}(k+q)^2 \frac{1}{2}(q+\frac{1}{2}p)^2}$$

The former term is the product of two 1-loop propagator graphs, the latter is the insertion of one 1-loop propagator graph into another.



Both graphs can be evaluated by repeated application of the generalized massless one-loop propagator correction (with arbitrary powers of free propagators) given at the end of subsection VIIB4. The result can be expressed as

$$\mathcal{P}_1 = -(D-3)\frac{(\mathcal{P}_0)^2}{\frac{1}{2}p^2}$$

$$\mathcal{P}_2 = c\mathcal{P}_1, \qquad c = \frac{[\Gamma(D-3)]^2\Gamma(5-D)}{[\Gamma(3-\frac{D}{2})]^2\Gamma(\frac{3D}{2}-5)\Gamma(\frac{D}{2}-1)}$$

in terms of the 1-loop propagator correction $\mathcal{P}_0$. We therefore modify our minimal subtraction so that $\mathcal{P}_0$ has the simplest form (G scheme):

$$\mathcal{P}_0 = -\tfrac{1}{6}\tfrac{1}{\epsilon}(\tfrac{1}{2}p^2)^{1-\epsilon}$$

where $D = 6-2\epsilon$, and we calculated the coefficient of the $1/\epsilon$ term and threw in a normalization factor that canceled the rest:

$$\hbar \to \mathcal{N}\hbar$$

$$\mathcal{N} = \frac{1}{3(D-6)\Gamma(2-\frac{D}{2})B(\frac{D}{2}-1,\frac{D}{2}-1)} = (1-\tfrac{2}{3}\epsilon)(1-2\epsilon)\frac{\Gamma(1-2\epsilon)}{\Gamma(1+\epsilon)[\Gamma(1-\epsilon)]^2}$$

Further evaluating c, we find

$$c = -\tfrac{1}{3}\epsilon\frac{1-2\epsilon}{(1-\frac{3}{2}\epsilon)(1-3\epsilon)}\frac{[\Gamma(1-2\epsilon)]^2\Gamma(1+2\epsilon)}{[\Gamma(1+\epsilon)]^2\Gamma(1-3\epsilon)\Gamma(1-\epsilon)}$$

Using the expansion of $ln\ \Gamma(1-z)$ in terms of γ and $\zeta(n)$, it is easily checked that this combination of Γ's is $1+\mathcal{O}(\epsilon^3)$, so we can just drop them. Collecting our results, we have

$$\mathcal{P} = \frac{1-c}{\frac{D}{2}-2}\mathcal{P}_1 = -\tfrac{1}{36}\tfrac{1}{\epsilon^2}\frac{3-2\epsilon}{1-\epsilon}\left[1+\tfrac{1}{3}\epsilon\frac{1-2\epsilon}{(1-\frac{3}{2}\epsilon)(1-3\epsilon)}\right](\tfrac{1}{2}p^2)^{1-2\epsilon}$$

Exercise VIIB7.3

Calculate the same graph in four dimensions. It's finite there, so no counterterms are necessary. However, in this case integration by parts gives a factor of $1/\epsilon$, and each of the two resulting graphs has an additional factor of $1/\epsilon^2$. The result then has a factor of 1 minus the previously obtained combination of Γ's, which we already saw was of order ϵ^3. The final result is thus obvious except for a factor of a rational number:

$$6\zeta(3)\frac{1}{\frac{1}{2}p^2}$$

(The on-shell infrared divergence is as expected from power counting.)

We next calculate the counterterm graphs. These are the ones that cancel the subdivergences coming from the 1-loop 3-point subgraphs. We therefore need the divergent part of this subgraph. This is easy to evaluate by our previous methods: The result of Schwinger parametrization, scaling, etc., doing all integration exactly except over the Feynman parameters is

$$\int dq\ \frac{1}{\frac{1}{2}q^2\frac{1}{2}(q+k+\frac{1}{2}p)^2\frac{1}{2}(q+k-\frac{1}{2}p)^2}$$

$$= \int_0^1 d^3\alpha\ \delta(1-\sum\alpha)\Gamma(3-\tfrac{D}{2})[\tfrac{1}{2}\alpha_+(1-\alpha_+)(k+\tfrac{1}{2}p)^2+\tfrac{1}{2}\alpha_-(1-\alpha_-)(k-\tfrac{1}{2}p)^2]^{D/2-3}$$

$$= \tfrac{1}{2}\tfrac{1}{\epsilon} + finite$$

by simply replacing the factor in brackets by 1 (since it is raised to the $-\epsilon$ power), where we have used

$$\int_0^1 d^3\alpha\ \delta(1-\sum\alpha) = \int_0^1 d\alpha_+\int_0^{1-\alpha_+} d\alpha_- = \tfrac{1}{2}$$

Since we know the divergence is momentum-independent, we can obtain the same result from a (infrared regularized) tadpole graph with its propagator raised to the third power: In the notation of subsection VIIB1,

$$\frac{1}{\Gamma(3)}\hat{\mathcal{A}}_1(3,0,m^2) = \frac{\Gamma(3-\frac{D}{2})}{\Gamma(3)}(\tfrac{1}{2}m^2)^{D/2-3} = \tfrac{1}{2}\tfrac{1}{\epsilon} + finite$$

The contribution of the 2 counterterm graphs (or one for the effective action if we drop the symmetry factor) is thus

$$2\Delta_3\mathcal{P}_0 = 2(-\tfrac{1}{2}\tfrac{1}{\epsilon})\mathcal{P}_0$$

Collecting terms, we have

$$\mathcal{P}+2\Delta_3\mathcal{P}_0 = -\tfrac{1}{12}\tfrac{1}{\epsilon^2}\frac{1-\frac{2}{3}\epsilon}{1-\epsilon}\left[1+\tfrac{1}{3}\epsilon\frac{1-2\epsilon}{(1-\frac{3}{2}\epsilon)(1-3\epsilon)}\right](\tfrac{1}{2}p^2)^{1-2\epsilon}+\tfrac{1}{6}\tfrac{1}{\epsilon^2}(\tfrac{1}{2}p^2)^{1-\epsilon}$$

After a little algebra, dropping terms that vanish as $\epsilon\to 0$, we find

$$\mathcal{P}+2\Delta_3\mathcal{P}_0 = (\tfrac{1}{2}p^2)[\tfrac{1}{12}\tfrac{1}{\epsilon^2}-\tfrac{1}{18}\tfrac{1}{\epsilon}-\tfrac{1}{12}(ln\ \tfrac{1}{2}p^2)^2+\tfrac{1}{9}ln\ \tfrac{1}{2}p^2-\tfrac{23}{216}]$$

Note that modifying minimal subtraction is equivalent to redefining $\frac{1}{2}\mu^2$, which we have set to 1, but which appears only in the ln's, as $ln(\frac{1}{2}p^2)\to ln(p^2/\mu^2)$. Thus, modifying $\mathcal{N}$, which appears only in the combination $\mathcal{N}(\frac{1}{2}p^2)^{-\epsilon}$, is the same as shifting $ln\ \frac{1}{2}p^2$:

$$\mathcal{N}\to\mathcal{N}e^{\epsilon a} \quad\Rightarrow\quad ln\ \tfrac{1}{2}p^2\to ln\ \tfrac{1}{2}p^2 - a$$

For example, choosing $a = -\frac{2}{3}$,

$$\mathcal{P} + 2\Delta_3 \mathcal{P}_0 \to (\tfrac{1}{2}p^2)[\tfrac{1}{12}\tfrac{1}{\epsilon^2} - \tfrac{1}{18}\tfrac{1}{\epsilon} - \tfrac{1}{12}(ln\ \tfrac{1}{2}p^2)^2 - \tfrac{5}{72}]$$

Only the $\mathcal{O}(\epsilon)$ part of the normalization factor affects the final result: More generally,

$$(\tfrac{1}{2}p^2)^\epsilon \to \mathcal{N}(\tfrac{1}{2}p^2)^{-\epsilon} \quad \Rightarrow \quad ln\ \tfrac{1}{2}p^2 \to ln\ \tfrac{1}{2}p^2 - \tfrac{1}{\epsilon} ln\ \mathcal{N}$$

Since after adding counterterms, which cancel nonlocal divergences arising from subdivergences, ln's appear only in finite terms, only the $\mathcal{O}(\epsilon)$ part of $ln\ \mathcal{N}$ will contribute. Thus, we can approximate any normalization factor as

$$\mathcal{N} \approx e^{\epsilon a}, \qquad a = \gamma + rational$$

as far as the renormalized results are concerned. (The γ identifies this normalization as modified minimal subtraction, as the $\overline{\text{MS}}$ or G schemes.) This is just the statement of the renormalization group, that the final result in minimal subtraction schemes depends only on the choice of scale: The complete normalization factor is really

$$\mathcal{N}_{total} = \mathcal{N}(\tfrac{1}{2}\mu^2)^\epsilon \quad \Rightarrow \quad ln\ \tfrac{1}{2}\mu^2 \to ln\ \tfrac{1}{2}\mu^2 + \tfrac{1}{\epsilon} ln\ \mathcal{N}$$

However, the higher-order terms can be convenient for intermediate stages of the calculation. In this particular case, the nonlocal divergences appearing before cancelation are of the form $(1/\epsilon) ln\ p^2$, so the $\mathcal{O}(\epsilon^2)$ part of $\mathcal{N}$ contributes at intermediate stages. For example, replacing the original $\mathcal{N}$ with

$$\mathcal{N}' = (1 - \tfrac{2}{3}\epsilon)(1 - 2\epsilon)\Gamma(1 - \epsilon)$$

would have given the same result even before cancelation, since the change is by another combination of Γ's that give $1 + \mathcal{O}(\epsilon^3)$.

Exercise VIIB7.4

Complete the 6D calculation of the exact 2-loop propagator correction in ϕ^3 theory, including the missing graph and counterterms, to find the total renormalized 2-loop propagator and its 2-loop counterterms.

REFERENCES

1 Goldstone, Salam, and Weinberg; Jona-Lasinio; *loc. cit.* (VC): effective potential.

2 S. Coleman and E. Weinberg, *Phys. Rev.* **D7** (1973) 1888.

3 A. D'Adda, M. Lüscher, and P. Di Vecchia, *Nucl. Phys.* **B146** (1978) 63; E. Witten, *Nucl. Phys.* **B149** (1979) 285: quantum CP(n).

4 L.M. Brown and R.P. Feynman, *Phys. Rev.* **85** (1952) 231;
G. Passarino and M. Veltman, *Nucl. Phys.* **B160** (1979) 151;
W.L. van Neerven and J.A.M. Vermaseren, *Phys. Lett.* **137B** (1984) 241:
reduction of 1-loop integrals with numerators to scalar 1-loop integrals.

5 A.C.T. Wu, *Mat. Fys. Medd. Dan. Vid. Selsk.* **33** (1961) 72:
scalar box (1-loop 4-point) graph.

6 R.P. Feynman, unpublished;
L.M. Brown, *Nuo. Cim.* **22** (1961) 178;
F.R. Halpern, *Phys. Rev. Lett.* **10** (1963) 310;
B. Petersson, *J. Math. Phys.* **6** (1965) 1955;
D.B. Melrose, *Nuo. Cim. A* **40** (1965) 181:
reduction of scalar 1-loop graphs to $\leq$D-point graphs.

7 G. 't Hooft and M. Veltman, *Nucl. Phys.* **B153** (1979) 365;
G.J. van Oldenborgh and J.A.M. Vermaseren, *Z. Phys. C* **46** (1990) 425;
A. Denner, U. Nierste, and R. Scharf, *Nucl. Phys.* **B367** (1991) 637:
further evaluation and simplification of 1-loop graphs.

8 K.G. Chetyrkin and F.V. Tkachov, *Nucl. Phys.* **B192** (1981) 159:
integration by parts in momentum space for exact evaluation of higher-loop massless propagators.

9 D.I. Kazakov, *Phys. Lett.* **133B** (1983) 406, *Theor. Math. Phys.* **58** (1984) 223, **62** (1985) 84; D.I. Kazakov and A.V. Kotikov, *Theor. Math. Phys.* **73** (1988) 1264:
more techniques for exact evaluation of such graphs.

10 Dyson, *loc. cit.* (VC):
treatment of lowest-order overlapping divergences.

11 Salam, *loc. cit.* (VIIA):
overlapping divergences to all loops.

C. RESUMMATION

So far we have only assumed that confinement arises nonperturbatively in 4D QCD. However, to connect with known, successful results of perturbation theory, we need to understand how the same methods used to give these perturbative results can be generalized to include the nonperturbative ones. The simplest method would be to take the perturbation expansion as is, and find a good method for evaluating (or perhaps redefining) its sum, with the hope that summation to all orders by itself would reveal features invisible at finite orders.

Besides the technical difficulties associated with such an approach, the main problem is that the summation of the perturbation expansion does not converge. Parts of this problem can be solved by appropriate redefinitions, but other parts indicate a serious problem with perturbation theory, caused by the very renormalization that was supposed to solve the main problem of finite-order perturbation theory (infinities).

1. Improved perturbation

We saw in the previous section that dimensional transmutation replaced the dimensionless coupling constant with a mass scale. In principle, we would like to explicitly make this replacement as the basis of our perturbation expansion, not only to make the perturbative parameter physical, but also to take into account the running of the original coupling. Unfortunately, this is not possible in practice; however, we can choose the arbitrary (unphysical) renormalization scale μ to be in the range of energies in the problem at hand, so that the $ln(p^2/\mu^2)$ corrections are small. A change in scale from one value of μ to another is related to a resummation of graphs: Although the one-loop term in the effective action containing $ln(p^2/\mu^2)$ comes from a single 1PI amplitude, it contributes an infinite number of terms at different loop orders to the propagator when inserted into any higher-loop 1PI graph, as $1/(K+A) = 1/K - (1/K)A(1/K) + ...$. Although $K + A$ depends only on M, K depends only on g and A depends only on μ. Thus, any redefinition of μ that leaves the physical quantity M unchanged requires a corresponding redefinition of g:

$$M^2 = \mu^2 e^{-1/g^2} \quad \Rightarrow \quad g^2(\mu^2) = \frac{1}{ln\left(\frac{\mu^2}{M^2}\right)}$$

and thus changing μ redistributes the contributions to $1/(K + A)$ (and therefore to the summation of graphs in any amplitude) over the different loop orders. For example, if the amplitude is most sensitive to the momentum in a particular propagator (independent of loop momenta), and we choose $\mu^2 \approx p^2$, then although we can't use the resummed perturbation expansion directly, we can at least push most of it into the lower orders.

Things get more complicated at higher loops: It becomes difficult to associate the running of the coupling with the resummation of a particular subset of all the graphs. However, we already know that this effect can be derived from the breaking of scale invariance by renormalization. For example, let's consider Yang-Mills theory, since gauge invariance restricts it to have only a single coupling parameter. (This makes it the simplest case conceptually, although not computationally. Here we use only the fact that it has a single coupling; its explicit renormalization constants won't be considered until the next chapter. As an alternative, we can consider the scalar QCD analog of subsections VC9 and VIIB6, a ϕ^4 theory with an auxiliary field, if we ignore mass renormalization, or arbitrarily renormalize the mass to zero.) For convenience of dimensional analysis, we use only coupling constants that are dimensionless in all dimensions, by scaling with an appropriate power of μ. (In general, we can do this even for masses.) The classical Yang-Mills action, before and after the addition of counterterms, is then

$$S_{class} = \frac{1}{g^2(\frac{1}{2}\mu^2)^\epsilon}\int dx\ tr\ \tfrac{1}{8}F^2, \qquad S_{class} + \Delta S = \frac{1}{\hat{g}^2}\int dx\ tr\ \tfrac{1}{8}F^2$$

where

$$\frac{1}{\hat{g}^2} = \frac{1}{(\frac{1}{2}\mu^2)^\epsilon}\left[\frac{1}{g^2} + \sum_{n=1}^{\infty}\frac{1}{\epsilon^n}c_n(g^2)\right], \qquad c_n(g^2) = \sum_{L=n}^{\infty}(g^2)^{L-1}c_{nL}$$

for some numerical constants c_{nL}. (We can also include $\hbar$'s as $g^2 \to g^2\hbar$.) We use $\frac{1}{2}\mu^2$ to produce the combination $(\frac{1}{2}p^2/\frac{1}{2}\mu^2)^{-\epsilon}$ in graphs. (In practice, one uses units $\frac{1}{2}\mu^2 = 1$ until the end of the calculation, and restores units.) The μ dependence is then given by varying μ for fixed $\hat{g}$:

$$\mu^2\frac{\partial}{\partial\mu^2}g^2 \equiv -\epsilon g^2 - \beta(g^2), \qquad \mu^2\frac{\partial}{\partial\mu^2}\hat{g}^2 \equiv 0$$

where the ϵg^2 term is the classical contribution, β will be found to be independent of ϵ (except indirectly through g^2), and $\hat{g}$'s independence from μ is the statement that the physics is independent of the choice of μ (i.e., $\hat{g}$ depends on only M and ϵ). We then find

$$0 = \mu^{2\epsilon}\mu^2\frac{\partial}{\partial\mu^2}\frac{1}{\hat{g}^2} \quad \Rightarrow \quad \beta = g^4\frac{\partial}{\partial g^2}(g^2c_1), \qquad \frac{\partial}{\partial g^2}(g^2c_{n+1}) = -\beta\frac{\partial}{\partial g^2}c_n$$

Thus, the coefficients of the $1/\epsilon$ terms determine those of both the higher order terms and β.

This gives us an expression for β,

$$\beta = \sum_{L=1}^{\infty} (g^2)^{L+1} \beta_L, \qquad \beta_L = L c_{1L}$$

Since g^2 is itself unphysical, the information we can get from analyzing the running of this coupling is arbitrary up to redefinitions. For example, assume that all β_L are nonvanishing, and write the definition of β as (in $D = 4 \to \epsilon = 0$)

$$\mu^2 \frac{\partial}{\partial \mu^2} \frac{1}{g^2} = f = \frac{1}{g^4}\beta = \sum_{L=0}^{\infty} (g^2)^L \beta_{L+1}$$

Then under a redefinition $g^2 \to g^2(g'^2)$ we have

$$\mu^2 \frac{\partial}{\partial \mu^2} \frac{1}{g'^2} = \left(\frac{\partial(1/g^2)}{\partial(1/g'^2)} \right)^{-1} f(g^2(g'^2)) \equiv f'(g'^2)$$

Now we consider a "perturbative" type of redefinition, as results from changing renormalization prescriptions, so g^2 gets only "$\mathcal{O}(\hbar)$" corrections: Taylor expanding

$$g^2 = g'^2 + k_1 g'^4 + +k_2 g'^6 + \mathcal{O}(g'^8)$$

$$\Rightarrow \quad \frac{1}{g^2} = \frac{1}{g'^2} + constant + \mathcal{O}(g'^2)$$

we find

$$\frac{\partial(1/g^2)}{\partial(1/g'^2)} = 1 + \mathcal{O}(g'^4), \qquad f(g^2(g'^2)) = f(g'^2) + \mathcal{O}(g'^4)$$

$$\Rightarrow \quad f'(g'^2) = f(g'^2) + \mathcal{O}(g'^4)$$

Thus, the first two coefficients of β (β_1 and β_2) are unaffected, while terms found at 3 loops and beyond can be modified arbitrarily, and even be set to vanish. In the more general case of more than 1 coupling, it is sometimes possible to eliminate also some of the 2-loop contributions.

Therefore, to consider the general behavior of the coupling as a function of energy (μ^2), it is sufficient to solve the equation

$$\mu^2 \frac{\partial}{\partial \mu^2} g^2 = -\beta_1 g^4 - \beta_2 g^6$$

(using, e.g., the change of variables $t = ln\ \mu^2$ and $u = 1/\beta_1 g^2$) as

$$\frac{\mu^2}{M^2} = e^{1/\beta_1 g^2} \left(\frac{1}{g^2} + \frac{\beta_2}{\beta_1} \right)^{-\beta_2/\beta_1^2}$$

with M^2 as the constant of integration. Using an allowed type of redefinition for g^2, and also redefining the arbitrary constant of integration M^2, we can simplify this to

$$\frac{1}{g^2} \to \frac{1}{g^2} - \frac{\beta_2}{\beta_1}, \qquad M^2 \to M^2 e^{\beta_2/\beta_1^2}$$

$$\Rightarrow \quad \beta = \frac{\beta_1 g^4}{1 - \frac{\beta_2}{\beta_1} g^2}, \qquad \frac{\mu^2}{M^2} = e^{1/\beta_1 g^2} (g^2)^{\beta_2/\beta_1^2}$$

(This redefinition changes the range of what g^2 is called negative and what positive. However, g^2 is just a parameter, not a physical coupling: As far as the unitarity of the kinetic term is concerned, only the residues near the poles of the propagator are relevant. Also, our allowed class of redefinitions do not affect behavior for small g^2, and thus perturbation theory.)

Exercise VIIC1.1

Let's analyze this solution in more detail:

a Graph the function $y(x) = e^{ax} x^b$ (or graph $ln\ y$ to make it simpler) for a and b positive, negative, and vanishing, to study the behavior of the function $\mu^2(g^2)$. The analysis can be simpliifed (and the behavior for different values of a and b related) by considering g^2 positive and negative, and the symmetries

$$a \to -a, \qquad x \to -x, \qquad y \to (-1)^b y$$

$$b \to -b, \qquad x \to -x, \qquad y \to (-1)^b \frac{1}{y}$$

Note that g^2 can be nonpositive for some values of μ^2: For example, even for $\beta_2 = 0$, we have $g^2 = 1/\beta_1 ln(\mu^2/M^2)$, which is negative for $\mu < M$ or for $\mu > M$. What happens for $\beta_2 \neq 0$?

b After applying the above redefinition, apply the second redefinition

$$\frac{1}{g} \to \frac{1}{g} + \frac{\beta_2}{\beta_1} g$$

Find the new β and $\mu^2(g^2)$. Compare to the behavior of $\mu^2(g^2)$ before this redefinition, for the cases $\beta_2/\beta_1 < 0$, noting the "duality" symmetry $g \leftrightarrow (-\beta_1/\beta_2)/g$.

Exercise VIIC1.2

Consider some theory with a single dimensionless coupling g^2, but now also a single mass m. By the above methods we find

$$\frac{\mu^2 \partial}{\partial \mu^2} m^2 = m^2[-1 - \beta_m(g^2)]$$

(The m dependence follows from dimensional analysis.) Solve for m^2 as a function of g^2, as an integral over g^2 in terms of β and β_m. Show that after an appropriate redefinition

$$\frac{\mu^2 \partial}{\partial \mu^2} m^2 = m^2(-1 - \beta_{m1} g^2)$$

for some constant β_{m1}. Solve for m^2 explicitly in terms of g^2, when we have also redefined β to $\beta_1 g^4 + \beta_2 g^6$. Then make the final redefinition $1/g^2 \to 1/g^2 - \beta_2/\beta_1$ used to simplify M^2.

For purposes of perturbation theory, it is useful to invert this: For small g^2, we have approximately

$$\frac{1}{g^2} \approx \beta_1 ln \frac{\mu^2}{M^2} + \frac{\beta_2}{\beta_1} ln \left(\beta_1 ln \frac{\mu^2}{M^2} \right)$$

This implies that the terms in the effective action that carry the M dependence are given by

$$\Gamma_M \approx tr \int dx \; \tfrac{1}{8} F \left[\beta_1 ln \frac{-\Box}{M^2} + \frac{\beta_2}{\beta_1} ln \left(\beta_1 ln \frac{-\Box}{M^2} \right) \right] F$$

(We can also replace $-\Box \to \Box$ in this limit, ignoring $i\pi$'s in comparison to ln's.)

The general class of coupling redefinitions we considered are allowed by perturbation theory: If we knew the exact solution to a field theory, we would be more restrictive, requiring invertibility. However, in perturbation theory, given two renormalization prescriptions related by some such coupling redefinition, we might know this redefinition only perturbatively, and perhaps only to a few orders. Even if we knew it exactly, and knew it to be noninvertible, it still might not be clear which of the two prescriptions was the correct one, if either. Therefore, the renormalization group alone is sufficient to draw conclusions about the behavior of a theory only at "small" ($\ll 1$) coupling.

Similar remarks apply to propagators, S-matrix elements, etc. Consider any function G_n appearing as the coefficient of n fields in a term in the effective action. Comparing the unrenormalized $\hat{G}_n$ to the renormalized G_n,

$$G_n(g^2, \mu^2) = \mathcal{Z}^{-n}(\hat{g}^2(\tfrac{1}{2}\mu^2)^{-\epsilon}, \epsilon) \hat{G}_n(\hat{g}^2, \epsilon), \qquad \frac{\mu^2 d}{d\mu^2} \hat{G} = 0$$

$$\Rightarrow \quad \left(\frac{\mu^2 \partial}{\partial \mu^2} + \beta \frac{\partial}{\partial g^2} + n\gamma \right) G = 0, \qquad \gamma = \frac{\mu^2 \partial}{\partial \mu^2} ln \; \mathcal{Z}$$

where $\mathcal{Z}$ is a wave-function renormalization factor (not required for pure Yang-Mills in the background gauge; or we can examine ratios of such quantities where the $\mathcal{Z}$'s cancel, which are more physical, such as S-matrix elements).

2. Renormalons

The perturbation expansion in general can't be resummed in the naive way because the number of diagrams increases as $n!(constant)^n$ at n loops. The simplest example of this is a self-interacting scalar in D=0:

$$Z = \int_{-\infty}^{\infty} \frac{d\phi}{\sqrt{2\pi}} e^{-\frac{1}{2}\phi^2 - \frac{1}{4}g^2\phi^4} = \sum_{n=0}^{\infty} g^{2n} Z_n$$

Since there is no momentum integration, each diagram is just 1 (times some permutation factors), so Z_n just counts the number of diagrams at n loops. We use g^2 so the coupling is similar to that in Yang-Mills: As usual, we can rescale $\phi \to \phi/g$ to recognize g^2 as $\hbar$:

$$\phi' = g\phi \quad \Rightarrow \quad \tfrac{1}{2}\phi^2 + \tfrac{1}{4}g^2\phi^4 = \tfrac{1}{g^2}(\tfrac{1}{2}\phi'^2 + \tfrac{1}{4}\phi'^4)$$

(Of course, we can be more explicit by writing $\hbar g^2$ in place of just g^2 or $\hbar$, but the effect is identical, since they both appear only in that combination.) This integral can be evaluated exactly at any order of perturbation theory:

$$Z_n = \int_{-\infty}^{\infty} \frac{d\phi}{\sqrt{2\pi}} \tfrac{1}{n!}(-\tfrac{1}{4}\phi^4)^n e^{-\frac{1}{2}\phi^2} = \tfrac{1}{n!}(-1)^n \tfrac{1}{\sqrt{\pi}}\Gamma(2n + \tfrac{1}{2}) \approx \frac{1}{\sqrt{2\pi}}(n-1)!(-4)^n$$

where we have used the Stirling approximation for $\Gamma(z)$ at large z.

Exercise VIIC2.1

Find the following properties of the Γ function for large argument:

a Derive the Stirling approximation

$$\lim_{z\to\infty} \Gamma(z) \approx \sqrt{\frac{2\pi}{z}} \left(\frac{z}{e}\right)^z$$

by applying the method of steepest descent to the integral definition of $\Gamma(z+1)$. (See subsections VA2 and VA5.)

b Use this approximation, and $lim_{z\to\infty}(1+\frac{1}{z})^z = e$, to show

$$\lim_{z\to\infty} \Gamma(az+b) \approx \sqrt{2\pi}(az)^{az+b-1/2} e^{-az}$$

Thus we might as well apply the steepest descent approximation directly to the original integral: Using also an integral for $\hbar$ ($= g^2$ in this case),

$$Z_n = \oint \frac{d\hbar}{2\pi i \hbar^{n+1}} \int D\left(\frac{\phi}{\sqrt{\hbar}}\right) e^{-S/\hbar}$$

we first apply steepest descent to the ϕ integral, yielding the usual first two terms in the JWKB expansion. Then the $\hbar$ integral can be approximated as $\Gamma(n)$ by keeping only the part of the contour on the positive real axis:

$$\oint \frac{d\hbar}{2\pi i \hbar^{n+1}} e^{-S/\hbar}\bigg|_{\delta S/\delta\phi=0} \approx \frac{1}{2\pi i}\Gamma(n)\left(\frac{1}{S}\right)^n\bigg|_{\delta S/\delta\phi=0}$$

$$\Rightarrow \quad Z_n \approx \sum_{\substack{\delta S/\delta\phi=0 \\ S\neq 0}} \frac{1}{2\pi i}\left[det\left(\frac{\delta^2 S}{\delta\phi^2}\right)\right]^{-1/2}(n-1)!\left(\frac{1}{S}\right)^n$$

($S=0$ solutions contribute only to Z_0. A similar result can be obtained by simultaneously using steepest descent for the $\hbar$ integral, yielding a "classical value" of $\hbar$ in terms of S.) In the present case, the nontrivial classical solutions are

$$S = \tfrac{1}{2}\phi^2 + \tfrac{1}{4}\phi^4 \quad \Rightarrow \quad \phi = \pm i$$

which gives the same Z_n as previously (being careful to sum the two terms for the two solutions). Thus, we see that in general we have to sum $\sum_{n=0}^{\infty} n!(\hbar/S)^n$, which does not converge. Furthermore, this divergence is associated with finite-action ("instanton") solutions to the classical equations of motion.

The simplest example of a resummation problem is the one-loop propagator correction. We have seen that the classical and one-loop kinetic terms can be combined to give a kinetic operator of the form $\beta_1 K(p^2) ln(-p^2/M^2)$ in massless theories, or at high energy in massive theories, where K is the classical kinetic operator. The free (or asymptotic) theory has solutions where this kinetic operator has a zero (the propagator blows up). Besides the classical solution at $K(p^2)=0$, there is another at $p^2=-M^2$:

$$\frac{1}{\beta_1 K(p^2)\ ln\left(-\frac{p^2}{M^2}\right)} = \frac{1}{\beta_1 K(p^2)\ ln\left(1-\frac{p^2+M^2}{M^2}\right)} \approx -\frac{M^2}{\beta_1 K(-M^2)} \times \frac{1}{p^2+M^2}$$

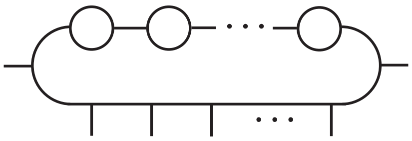
This might be expected to be a bound state, called a "renormalon" because of its relation to the renormalization group. However, the residue of this pole in the propagator can have the wrong sign, indicating the appearance of a ghost ("Landau ghost"), and thus a violation of unitarity.

Exercise VIIC2.2

The Landau ghost itself is not necessarily a problem in quantum field theory, although it indicates the possibility of such problems. Examine the behavior of this ghost after taking into account the 2-loop correction (β_2), before and after the simplifying redefinition of the previous subsection, for all the various signs of β_1 and β_2. Since the expression for $\mu^2(g^2)$ can't be inverted, use the fact that the propagator follows from the coupling $g^2(\mu^2)$ as

$$\Delta \sim \frac{g^2(p^2)}{p^2}$$

(Field redefinitions can't remove the momentum dependence of couplings.) Then new poles (or other singularities) in the propagator correspond to the limit $g^2 \to \infty$, so find $p^2(g^2)$ there.



This causes problems similar to those from instantons when the quantum propagator is inserted into another graph. We set external momenta to vanish, as an approximation for high energy for the loop momenta, or to evaluate low-energy quantities such as anomalous magnetic moments. In any one-loop 1PI graph with n 1-loop propagator insertions and l external lines, we get an integral at high energy of the form (e.g., in QCD or the scalar analog with auxiliaries of subsections VC9 and VIIB6)

$$\int d^4k\ (k^2)^{-l} \left[-\beta_1\ ln\left(\frac{k^2}{\mu^2}\right)\right]^n \sim (-\beta_1)^n \int_0^\infty du\ e^{-(l-2)u} u^n \sim n! \left(-\frac{\beta_1}{l-2}\right)^n$$

where we changed variables to $u = ln(k^2/\mu^2)$ (remembering $\Gamma(n+1) = n!$). We have used effectively an infrared cutoff by approximating the u-integral from 0 to ∞ instead of $-\infty$ to ∞. If we look instead at the low-energy (of the loop momentum) behavior, now taking $l-1$ massive classical propagators with 1 massless propagator (to insure IR convergence) with n insertions, we find

$$\int d^4k\ (k^2+m^2)^{-(l-1)}(k^2)^{-1} \left[-\beta_1\ ln\left(\frac{k^2}{\mu^2}\right)\right]^n \sim \beta_1^n \int_{-\infty}^0 du\ e^u (-u)^n = n! \beta_1^n$$

Since the former comes from UV behavior it's called a "UV renormalon", while the latter coming from IR behavior is called an "IR renormalon". The essential difference is the relative factor of $(-1)^n$. In fact, the former expression is also the high-energy limit of the latter (neglecting masses then), so the complete integral (u from $-\infty$ to ∞, so k^2 from 0 to ∞) can be approximated as the sum of the UV renormalon and IR renormalon contributions.

3. Borel

Since renormalons and instantons cause the perturbation expansion to diverge by a factor of $n!$, we look for a method to formally sum such series, by relating them to series that do converge. In general, we consider the series

$$A(\hbar) = \sum_{n=0}^{\infty} \hbar^n a_n$$

and define the "Borel transform" as:

$$\tilde{A}(z) = \int_{r-i\infty}^{r+i\infty} \frac{d(1/\hbar)}{2\pi i} e^{z/\hbar} A(\hbar)$$

(for some real number r to the right of all singularities of A) in anticipation of instanton-like contributions. The inverse is

$$A(\hbar) = \int_0^{\infty} dz\ e^{-z/\hbar} \tilde{A}(z)$$

The inverse Borel transform is related to the Laplace transform (with the variable change $x = 1/\hbar$) and the Mellin transform ($x = 1/\hbar$ and $y = e^z$). Evaluating explicitly for the above series,

$$\tilde{A}(z) = \delta(z) a_0 + \sum_{n=0}^{\infty} z^n \tfrac{1}{n!} a_{n+1}$$

So the Borel-transformed sum converges faster by a factor of $n!$, which is just what we need for perturbation theory. The idea for resumming the perturbation expansion is to first do the Borel sum, then inverse Borel transform the resulting function. Of course, this procedure does not necessarily fix the original problem, which might merely be translated into problems of convergence or ambiguity for integration of the inverse transform. In particular, we need $\tilde{A}(z)$ to be well defined along the positive real axis.

We saw that generically the sums involved were approximately of the form

$$A(\hbar) \sim \sum_{n=1}^{\infty} \hbar^n (n-1)! (-k)^n$$

In that case

$$\tilde{A}(z) \sim \sum_{n=0}^{\infty} (-1)^n z^n k^{n+1} = \frac{1}{z + \frac{1}{k}}$$

When $k < 0$, this leads to a singularity in the integral defining the inverse Borel transform. It can be "regularized" by choosing a contour that goes around the pole, but the choice of contour is ambiguous, and choosing an arbitrary linear combination of the two contours introduces a free parameter. Explicitly, we have

$$A(\hbar) = A_0(\hbar) + \zeta e^{-1/|k|\hbar}$$

where A_0 is the result of a particular prescription (e.g., principal value), and ζ is the new parameter. The ζ term is clearly nonperturbative, since each term in its Taylor expansion in $\hbar$ vanishes. This new parameter can be interpreted as a new (nonperturbative) coupling constant in the theory, just like ambiguities in renormalization of new counterterms in perturbatively nonrenormalizable theories.

Now we more carefully analyze the explicit sums we found in the previous subsection. The first example is $\sqrt{\hbar}Z$ for D=0:

$$\tilde{A}(z) = \int_{r-i\infty}^{r+i\infty} \frac{d(1/\hbar)}{2\pi i} e^{z/\hbar} \int_{-\infty}^{\infty} \frac{d\phi}{\sqrt{2\pi}} e^{-S/\hbar} = \int_{-\infty}^{\infty} \frac{d\phi}{\sqrt{2\pi}} \delta(z-S) = \frac{1}{\sqrt{2\pi}} \sum_{S=z} (S')^{-1}$$

(The contribution from $S=0$ is artificial, coming from our using $A=\sqrt{\hbar}Z$ instead of Z.) So, this integral can be explicitly evaluated. (For example, for the action we used in the previous subsection, we can explicitly solve for ϕ at $S=z$.) However, there is then a problem in inverting the Borel transform: Near $z=z_0\equiv S(\phi_0)$ for classical solutions ϕ_0, we have

$$S(\phi) \approx S(\phi_0) + \tfrac{1}{2}S''(\phi_0)(\phi-\phi_0)^2 \ \Rightarrow\ S'(\phi) \approx S''(\phi_0)(\phi-\phi_0),\ z-z_0 \approx \tfrac{1}{2}S''(\phi_0)(\phi-\phi_0)^2$$

$$\Rightarrow \quad (S')^{-1} \approx [2S''(\phi_0)]^{-1/2}(z-z_0)^{-1/2}$$

Therefore, there are cuts with branchpoints at classical values of the action, leading to ambiguities in the result for $A(\hbar)$. We thus see that new coupling constants are introduced for each solution to the classical field equation with positive action. (For our D=0 example $S<0$, and there is no problem, but more realistic examples, like Yang-Mills instantons, have $S>0$.)

Exercise VIIC3.1

Consider the D=0 action

$$S = \tfrac{1}{2}\phi^2 - \tfrac{1}{4}\phi^4$$

which differs from our previous example by the sign of the interaction. Now we have classical solutions with $S>0$. (The interaction is the wrong sign for the integral to be well defined, but the "kinetic" term is the right sign for it to be defined perturbatively.) Explicitly evaluate $\phi(S)$ (i.e., $\phi(z)$), and show it has the above behavior near $z=S(\phi_0)$.

In the case of renormalons, we see from the previous subsection that the large-n behavior gives singularities at $z = N/\beta_1$ for positive integer N. This integral also is easier to evaluate after Borel transforming: We consider a one-loop graph, but replace one internal line with the "full" quantum propagator coming from the 1-loop effective action (the same as summing a string of 1-loop propagator insertions), while using massive propagators for the remaining lines. We thus examine first the transform of the quantum propagator

$$\int \frac{d(1/\hbar)}{2\pi i} e^{z/\hbar} \frac{1}{k^2} \frac{1}{\frac{1}{\hbar} + \beta_1 \; ln(k^2/\mu^2)} = \frac{1}{k^2} \left(\frac{k^2}{\mu^2}\right)^{-\beta_1 z}$$

by closing the contour on the left. Then inserting this transformed propagator into the complete diagram,

$$\int d^4k \frac{1}{(k^2+m^2)^{l-1}} \frac{1}{k^2} \left(\frac{k^2}{\mu^2}\right)^{-\beta_1 z} \sim \left(\frac{m^2}{\mu^2}\right)^{-\beta_1 z} \Gamma(1-\beta_1 z)\Gamma(l-2+\beta_1 z)$$

using the integrals of the subsection VIIB1. This expression is the sum over n of the UV/IR renormalon example at the end of the previous subsection, except that we have done the summation over n as the first step (and used the Borel transform to assist in the evaluation). The first Γ has poles at $z = N/\beta_1$ for positive N, representing the IR renormalon, which are relevant for $\beta_1 > 0$, but the second Γ has poles at $z = -(N+l-3)/\beta_1$ for positive N (and $l \geq 3$ for the original diagram to be UV convergent), representing the UV renormalon, which are relevant for $\beta_1 < 0$. To the one-loop approximation for the β-function we have used, the singularities are just poles, but if the two-loop propagator insertions are used, these singularities become the branchpoints for cuts.

The new coupling constants that appear nonperturbatively can be given a physical interpretation in terms of vacuum values of polynomials of the fields. The basic idea is analogous to perturbative tadpoles: In that case corrections to S-matrices due to vacuum expectation values of scalar fields can be expressed by propagators that end at a "one-point vertex", whose coefficient is the vacuum value of the field:

$$\langle\phi(x)\rangle = \int D\phi \; e^{-iS}\phi(x) = c$$

in position space for some constant c, or in momentum space as

$$\int D\phi \; e^{-iS}\phi(p) = c\delta(p)$$

Similarly, we could expect graphs to have two propagators that end at a two-point vertex representing the vacuum value of the product of the two fields associated with the ends of the two propagators, and so on for higher-point vertices. For example, for a ϕ^2 vertex in a scalar theory, it would correspond to a contribution of the form

$$\int D\phi\ e^{-iS}\phi(x)\phi(y) = (c^2 + c') + ...$$

in position space, or in momentum space

$$\int D\phi\ e^{-iS}\phi(p)\phi(q) = (c^2 + c')\delta(p)\delta(q) + ...$$

where c^2 is the contribution from $\langle\phi\rangle^2$, so c' represents $(\Delta\phi)^2 = \langle\phi^2\rangle - \langle\phi\rangle^2$. Such vacuum values do not appear in perturbation theory for higher than one-point; we get only one $\delta(p)$ for each connected part of any graph. However, such contributions would be expected to give similar contributions to those we have found for renormalons: By dimensional transmutation, a contribution to an amplitude of the form $\zeta e^{-n/\beta_1\hbar}$ must appear in the combination

$$\zeta e^{-n/\beta_1\hbar} \to \zeta e^{-n/\beta_1\hbar}\left(\frac{\mu^2}{p^2}\right)^n = \zeta\left(\frac{M^2}{p^2}\right)^n$$

This is the type of contribution expected from a propagator with tadpole insertions, or in the same way from any other type of vacuum value. In particular, in QCD there are no fundamental scalar fields, but only scalar fields can get vacuum values, by Lorentz invariance. Thus, the vacuum values come from composite scalars, like $tr(F^2)$, $\bar{q}q$, etc.

Note that renormalons are a feature of renormalizable theories: They do not appear in superrenormalizable or finite theories. In particular, the path-integral methods of "constructive quantum field theory" have been used to show that certain interacting field theories in lower dimensions can be proven rigorously to exist — superrenormalizable theories with unique vacuua.

4. 1/N expansion

Perturbation theory is insufficient to evaluate all quantities in quantum physics, since

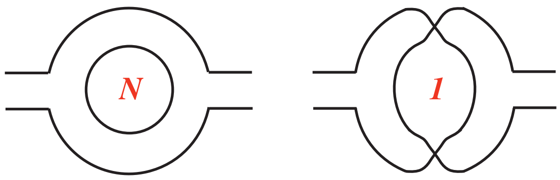
(1) such expansions don't always converge;

(2) if they do converge, they might not converge to the complete result; and

(3) even if they do give the complete answer, their summation might not be practical.

There are many perturbative expansions in quantum field theory. When we say "perturbation theory" in this context, we generally mean an expansion in the number of fields (or, in diagrammatic terms, number of vertices), since in the path integral we kept the exact quadratic part of the action but expanded in powers of the interaction terms (cubic and higher). (This is usually also an expansion in the coupling constants, depending on how we define the fields, which can be redefined by factors of the couplings.) One disadvantage of this expansion is that it violates manifest gauge invariance: Nonabelian gauge transformations are nonlinear in the fields, and thus mix diagrams with different numbers of fields. (These are the internal fields; external fields are asymptotic, and approximated as free.) Graphs that are related by gauge transformations must be added together to obtain gauge-invariant, and thus physically meaningful, expressions. Also, in practice individual graphs contain "gauge artifacts" that complicate them in certain gauges, but cancel in gauge-invariant objects, like S-matrix elements.

There can be a large number of graphs contributing to a particular physical process (given set of external states) at any particular loop order. There is another gauge-invariant expansion that can be applied to Yang-Mills theory to subdivide these sets of graphs, based on the freedom of choice of the Yang-Mills group itself: We have seen that the classical groups are defined in terms of N×N matrices, where N is arbitrary. Clearly, S-matrix elements must depend on N, even if the external states are restricted to be group singlets or representations of an N-independent subgroup, since the number of internal states increases as some polynomial in N. We now examine how this can be used to define a perturbation expansion in terms of N.



We have already seen in subsection VC9 that the group theory of any graph can be detached from the momentum and spin (so we considered there a simple model of scalars ϕ). We also saw there that the group theory of such matrices is most conveniently graphed by a double-line notation, where each line acts group-theoretically as a bound (anti)quark, reducing the group theory to trivial Kronecker δ's. We now notice that in some loop graphs, depending on how the lines are connected, some of the quark lines form closed loops. Again the group theory is trivial: There is a factor of N for each such loop, from the sum over the N colors. We can also give a physical picture to these numerical factors: Since we draw the scalar propagator as quark and antiquark lines with finite separation, think of the scalar as a (very short) string, with a quark at one end and antiquark at the other. This gives a two-dimensional structure to the diagram, by associating a surface with the area between the quarks and antiquarks (including the area at the vertices). We can extend this picture by associating a surface also with the area inside (i.e., on the other side of) each closed quark loop. In particular, for any "planar" diagram, i.e., any diagram that can be drawn on a sheet of paper without crossing any lines, and with all external lines on the outside of the diagram, the entire diagram forms an open sheet without holes, and with the topology of a disc (simply connected). It is also clear that, for a fixed number of loops and a fixed number of external lines, a planar diagram has the greatest number of factors of N, since crossing lines combines quark loops and reduces the power of N.

We can be more quantitative about this N dependence, and relate it to the topology of the graph. In subsection VC2 we saw the number of propagators, vertices, and loops were related by $P - V = L - 1$. This relation treats a Feynman diagram as just a graph, points connected by lines. We now consider a diagram as a polyhedron, with propagators as the edges, and closed quark loops as the faces, as defined by our use of matrices for fields. We then have as an additional relation for *closed* surfaces "Euler's theorem",

$$F = P - V - 2(H - 1)$$

(in terms of the "Euler number" $\chi \equiv -2(H-1) = V - P + F$), where F is the number of faces and H is the number of "handles": 0 for the sphere, 1 for the torus (doughnut), etc. This follows from the previous relation: First combining them as

$$L = F + 2H - 1$$

we note that "cutting" any handle along a loop (without separating the pieces) produces 2 faces; in other words, introducing two faces (as a "lens") into a loop that circles a handle changes the surface without changing the diagram, replacing 1 handle with 2 faces. The last relation then follows from the case with no handles, where each face gives a loop, except that the no-loop case corresponds to 1 face (or start with a less trivial case, like a cube, if that's easier to picture and count momenta for).

Using the fact that the g^2 appears in Yang-Mills the same way as $\hbar$, and that each face gets a factor of N, we find the g and N dependence of any graph is

$$(g^2)^{L-1} N^{(L-1)-2(H-1)} = (Ng^2)^{L-1} N^{-2(H-1)}$$

We thus see that effectively Ng^2 is the coupling squared suited to planar graphs, counting the number of loops, while $1/N^2$ is a new coupling squared, counting the number of handles. Therefore, we can sum over both Ng^2 and $1/N^2$: Each Feynman graph is a particular order in each of these two couplings. The sum of all graphs at fixed orders in both couplings gives a gauge-invariant subset of the graphs contributing to a particular S-matrix element. (This is sometimes called "color decomposition". Note that g^2 is the coupling normalized for matrices of the defining representation, which was required here to define the 1/N expansion, while $Ng^2 = \frac{1}{2}g_A^2$ is the coupling normalized for the adjoint: If we had used matrices for the adjoint representation, a factor $1/g_A^2$ would appear in front of the action, because of the difference in normalization of the trace of the matrices.)

Exercise VIIC4.1

Consider ϕ^4 theory in D=4, where ϕ is now an N×N hermitian matrix. Generalize the auxiliary-field propagator correction calculation of subsection VIIB6 to leading order in 1/N, showing the N-dependence at all steps. Show that now, to this leading order, both the N- and g-dependence of the effective action can be absorbed into M.

We can also consider more complicated models, such as chromodynamics, with fields appearing in the defining representation of the group, such as quarks. When a quark field makes a closed loop, it looks like a planar loop of a gluon, except that the closed quark line is missing, along with a corresponding factor of N. Thus, there is effectively a "hole" in the surface. Since only one factor of N is missing, a hole counts as half a handle. We can also draw a flavor-quark line for the quark propagator alongside the color-quark line. Since this line closes in quark-field loops, we also get a factor of M (for M flavors) for each quark loop.

The fact that the 1/N expansion is topological (the power of 1/N is the number of holes plus twice the number of handles) closely ties in with the experimental observation that hadrons (in this case, mesons) act like strings. Thus, we can expand in 1/N as well as in loops. While the leading order in the loop (Ng^2) expansion is classical (particle) field theory, the leading order in the 1/N expansion is classical open-string theory (planar graphs). However, seeing the dynamical string properties requires summing to all orders in Ng^2 for leading order in 1/N.

Thus, 1/N acts as the string coupling constant. (N appears nowhere else in the action describing string states, since they are all color singlets.) The experimental fact that the hadronic spectrum and scattering amplitudes follow so closely that of a string (more on this later) indicates that the perturbative expansion in 1/N is accurate, i.e., that quantum corrections are "small" in that sense. One application of the smallness of 1/N (largeness of N) is the "Okubo-Zweig-Iizuka rule": A planar graph describes classical scattering of open strings (mesons). It corresponds topologically to a disc, which is a sphere with one hole, and is therefore order 1/N. Compare this to two planar graphs connected by a handle. It describes classical scattering of open strings with one intermediate closed string (glueball), where the handle is a closed-string propagator connecting two otherwise-disconnected classical open-string graphs. It corresponds to a cylinder, which is a sphere with two holes, and is therefore order $1/N^2$. In terms of *flavor* lines, the latter graph differs from the former in that it has an intermediate state (the glueball) with no flavor lines. The OZI rule is that amplitudes containing an intermediate glueball are always smaller than those with an intermediate meson. This rule also has been verified experimentally, giving a further justification of the 1/N expansion (though not necessarily of string behavior).

Generalizing to groups SO(N) and USp(2N) gives more varied topologies: Since the left and right sides of propagators are no longer distinguishable, the string surface is no longer orientable (the surface no longer has two distinguishable sides), so we can also have unorientable surfaces such as Möbius strips and Klein bottles. One can also perform a separate expansion in the number M of flavors.

The fact that the leading (planar) contributions are of order $(Ng^2)^{L-1}$ requires a modification of the Borel transform of the previous subsection: We now identify

$$\hbar = Ng^2$$

instead of just $\hbar = g^2$, so we can use the 1/N expansion in conjunction with the Borel transform. In particular, this means removing the factor of N from β_1 and absorbing it into $\hbar$. The result is that the position of the renormalon singularities in the z plane is independent of N. However, the same is not true for the instantons: A one-instanton solution corresponds to choosing a single component of ϕ nonvanishing in our scalar model, so that the classical solution ϕ_0 for the action $S[\phi_0]$ has no N-dependence. (Choosing ϕ proportional to the identity matrix yields an N-instanton solution.) The analog in the Yang-Mills case is using just a (S)U(2) subgroup of the full U(N) to define the instanton. (Note that the structure constants for U(N) are N-independent for the defining representation: See exercise IB5.2.) Then $S/g^2 = NS/\hbar$. The result for the positions of the singularities in z is then at integer multiples (positive or negative, depending on considerations given in the previous subsection) of z_0, where

$$z_0 = \begin{cases} 1/\beta_1 & \text{for renormalons} \\ NS[\phi_0] & \text{for instantons} \end{cases}$$

where β_1 and the one-instanton action S are N-independent.

The net result is that instantons are unimportant for large N. Thus, if we take the 1/N approach of using a resummation to define a string theory, the instantons do not take a role in defining the string. (They might return in another form when considering classical solutions to the string theory, or their contribution might be just a small part of the total nonperturbative contribution.) On the other hand, approaches that analyze just the low-energy behavior of a theory can make use of the instantons: If the physical value of N is small, or the U(N) theory is spontaneously broken to give a small effective N at low energies (as in GUTs), then instantons may be treated as the dominant nonperturbative contribution to low-energy effects such as chiral symmetry breaking. This can be sufficient for studying low-energy bound states, but is insufficient for studying confinement, whose physical definition is the existence of bound states of very high energy.

REFERENCES

1 E.C.G. Stückelberg and A. Petermann, *Helv. Phys. Acta* **26** (1953) 499;
M. Gell-Mann and F.E. Low, *Phys. Rev.* **95** (1954) 1300;
C.G. Callan, *Phys. Rev.* **D2** (1970) 1541;
K. Symanzik, *Commun. Math. Phys.* **18** (1970) 227:
renormalization group.

2 G. 't Hooft, *Nucl. Phys.* **B61** (1973) 455:
renormalization group via dimensional regularization.

3 G. 't Hooft, Can we make sense out of "quantum chromodynamics", in *The whys of subnuclear physics*, proc. 1977 Int. School of Subnuclear Physics, Erice, ed. A. Zichichi (Plenum, 1979) p. 943:
arbitrariness of all but first two nonvanishing coefficients in β.

4 L.D. Landau and I. Pomeranchuk, *Doklady Akad. Nauk USSR* **102** (1955) 489:
Landau ghost.

5 F.J. Dyson, *Phys. Rev.* **85** (1952) 631:
divergence of perturbation expansion in quantum field theory.

6 L.N. Lipatov, *Soviet Physics JETP* **45** (1977) 216:
D=0 renormalon.

7 G. 't Hooft, *loc. cit.* (ref. 3), *Phys. Lett.* **109B** (1982) 474;
B. Lautrup, *Phys. Lett.* **69B** (1977) 109;
G. Parisi, *Phys. Lett.* **76B** (1978) 65, *Nucl. Phys.* **B150** (1979) 163;
Y. Frishman and A. White, *Nucl. Phys.* **B158** (1979) 221;
C. DeCalan and V. Rivasseau, *Comm. Math. Phys.* **82** (1981) 69;
F. David, *Nucl. Phys.* **B234** (1984) 237;
A.H. Mueller, *Nucl. Phys.* **B250** (1985) 327, *Phys. Lett.* **308B** (1993) 355;
L.S. Brown, L.G. Yaffe, and C.-X. Zhai, hep-ph/9205213, *Phys. Rev.* **D46** (1992) 4712:
renormalons.

8 A.S. Wightman, Should we believe in quantum field theory, in *The whys of subnuclear physics*, *ibid.* p. 983:
review of the relation of renormalons to constructive quantum field theory.

9 't Hooft, *loc. cit.* (VC):
1/N.
10 S. Okubo, *Phys. Lett.* **5** (1963) 165;
Zweig, *loc. cit.* (IC);
J. Iizuka, *Prog. Theor. Phys. Suppl.* **37-38** (1966) 21;
J. Iizuka, K. Okada, and O. Shito, *Prog. Theor. Phys.* **35** (1966) 1061.
11 G. Veneziano, *Phys. Lett.* **52B** (1974) 220:
expansion in (1/) the number of flavors.

VIII. GAUGE LOOPS

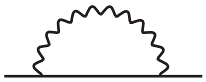
Gauge invariance plays an important role in quantum corrections. It not only simplifies their form, but leads to new effects. In particular, it not only improves high-energy behavior, but can eliminate divergences altogether, in the presence of supersymmetry.

In general, the first thing to calculate in quantum field theory is the effective action. Once this has been calculated, other properties can be determined: the vacuum, S-matrix, etc. In particular, in spontaneously broken theories, the effective action should be calculated with the symmetric (unbroken) vacuum, which has simpler Feynman rules; once the effective action has been calculated, vacuum values of the fields can be determined, and the S-matrix can be calculated as a perturbation about this quantum vacuum. (The alternative of defining Feynman rules for the classical broken vacuum and then calculating quantum corrections doubles the work in finding vacuum values.)

A. PROPAGATORS

We first consider propagator corrections in some specific theories with spin. In the following calculations we assume the gauge coupling appears only as an overall factor in the classical action: It thus also counts loops, so our 1-loop graphs are coupling-independent. All the integrals have been performed in subsections VIIB4-5; all that remains is the numerator algebra, which follows the examples of subsection VIC4. As we have seen, such corrections are important in analyzing high-energy behavior; as we'll see in the following section, they are also important for low energy. (Of course, for massless particles the two are related by conformal invariance, even when quantum corrections break it.)

1. Fermion



Our first calculation is the one-loop correction to the electron kinetic operator in QED: The S-matrix element is

$$\mathcal{A}_{2e} = \int dk \frac{\gamma^a(\not{k} + \frac{1}{2}\not{p} + \frac{m}{\sqrt{2}})\gamma_a}{\frac{1}{2}(k - \frac{1}{2}p)^2 \frac{1}{2}[(k + \frac{1}{2}p)^2 + m^2]}$$

At this loop level the only difference between using D-dimensional γ-matrix algebra (dimensional regularization) and 4-dimensional (dimensional reduction) is an unphysical finite renormalization, so for simplicity we'll use the latter method. Then the numerator is

$$\not{k} + \tfrac{1}{2}\not{p} - \sqrt{2}m$$

The result of the integral is then

$$\mathcal{A}_{2e} = -\not{p}\frac{m^2}{2p^2}[\hat{\mathcal{A}}_2(p^2,0,m^2) - \hat{\mathcal{A}}_2(0,0,m^2)] + (\tfrac{1}{2}\not{p} - \sqrt{2}m)\hat{\mathcal{A}}_2(p^2,0,m^2)$$

in the notation of subsection VIIB5. The UV divergent part follows from

$$\hat{\mathcal{A}}_2(p^2,0,m^2) = \tfrac{1}{\epsilon} + finite$$

The contribution to Γ is minus the S-matrix element, but the counterterm has a second minus sign to cancel the divergence:

$$\Delta S = \tfrac{\hbar}{\epsilon}\int dx\ \bar{\Psi}(-\tfrac{1}{2}i\not{\partial} - \sqrt{2}m)\Psi$$

The calculation for the quark self-energy in QCD is the same except for group-theory factors (see subsection VIIIA5).

Exercise VIIIA1.1

Repeat the calculation with D-dimensional γ-matrix algebra. What is the difference in the *finite* part, and why doesn't it matter?

In subsection VIIB5 we considered MOM subtraction (see subsection VIIA3) for scalar propagators. The analysis in this case is similar, but now we expand in $\not{p}$ instead of p^2:

$$\Delta K = a + b(\tfrac{m}{\sqrt{2}} - \not{p}) + \mathcal{O}[(\tfrac{m}{\sqrt{2}} - \not{p})^2]$$

However, since ΔK is normally expressed as functions of p^2 times 1 and $\not{p}$, we need to translate: Using $(\frac{m}{\sqrt{2}} + \not{p})(\frac{m}{\sqrt{2}} - \not{p}) = \frac{1}{2}(p^2 + m^2)$,

$$\begin{aligned}\Delta K &= a + b'(\tfrac{m}{\sqrt{2}} - \not{p}) + c\tfrac{1}{2}(p^2+m^2) + \mathcal{O}[(\tfrac{m}{\sqrt{2}} - \not{p})(p^2+m^2),(p^2+m^2)^2]\\ &= a + (b' + 2\tfrac{m}{\sqrt{2}}c)(\tfrac{m}{\sqrt{2}} - \not{p}) + \mathcal{O}[(\tfrac{m}{\sqrt{2}} - \not{p})^2]\end{aligned}$$

We next reevaluate the fermion propagator correction, to linear order in $\frac{m}{\sqrt{2}} - \not{p}$. Starting with

$$\tilde{\mathcal{A}}(x,p^2,m_1^2,m_2^2) = \int dk\ e^{ix\cdot k}\frac{1}{\frac{1}{2}[(k+\frac{1}{2}p)^2+m_1^2]\frac{1}{2}[(k-\frac{1}{2}p)^2+m_2^2]} = \int d^2\tau\ \lambda^{-D/2}e^{-E}$$

$$E = \tfrac{1}{2}\tfrac{1}{\lambda}x^2 + ix\cdot\tfrac{1}{2}\beta p + \tfrac{1}{8}\lambda(1-\beta^2)p^2 + \tfrac{1}{4}\lambda[(m_1^2+m_2^2) + \beta(m_1^2 - m_2^2)]$$

we keep only linear order in x and p^2+m^2, and set $m_1 = m$, $m_2 = 0$ (switching back to $\alpha = \frac{1}{2}(1+\beta)$):

$$E \approx ix\cdot(\alpha-\tfrac{1}{2})p + \tfrac{1}{2}\lambda\alpha(1-\alpha)(p^2+m^2) + \tfrac{1}{2}\lambda m^2\alpha^2$$

To clearly separate UV divergences (from $\lambda \approx 0$) and IR divergences (from $\alpha \approx 0$), we scale

$$\lambda \to \frac{\lambda}{\alpha^2} \quad \Rightarrow \quad E \approx (\alpha-\tfrac{1}{2})ix\cdot p + \tfrac{1}{2}\lambda(\tfrac{1}{\alpha}-1)(p^2+m^2) + \tfrac{1}{2}\lambda m^2$$

$$\tilde{\mathcal{A}} \approx \int_0^\infty d\lambda\ \lambda^{\epsilon-1}e^{-\lambda m^2/2}\int_0^1 d\alpha\ \alpha^{-2\epsilon}[1-(\alpha-\tfrac{1}{2})ix\cdot p][1-\tfrac{1}{2}\lambda(\tfrac{1}{\alpha}-1)(p^2+m^2)]$$

The integrals are easily performed in either order:

$$\tilde{\mathcal{A}} \approx \Gamma(1+\epsilon)(\tfrac{1}{2}m^2)^{-\epsilon}\left[\frac{1}{\epsilon_{UV}}\frac{1}{1-2\epsilon} + \tfrac{1}{2}\frac{1}{(1-\epsilon)(1-2\epsilon)}ix\cdot p\right.$$
$$\left. + \left(\tfrac{1}{2}\frac{1}{\epsilon_{IR}} + \frac{1}{1-2\epsilon}\right)\frac{p^2+m^2}{m^2} + \left(\tfrac{1}{4}\frac{1}{\epsilon_{IR}} + \tfrac{3}{2}\frac{1}{1-2\epsilon} - \tfrac{1}{2}\frac{1}{1-\epsilon}\right)ix\cdot p\,\frac{p^2+m^2}{m^2}\right]$$

and in the limit $\epsilon \to 0$,

$$\tilde{\mathcal{A}} \approx \Gamma(1+\epsilon)(\tfrac{1}{2}m^2)^{-\epsilon}\left[\frac{1}{\epsilon_{UV}} + 2 + \tfrac{1}{2}ix\cdot p + \left(\tfrac{1}{2}\frac{1}{\epsilon_{IR}}+1\right)\frac{p^2+m^2}{m^2}\right.$$
$$\left. + \left(\tfrac{1}{4}\frac{1}{\epsilon_{IR}}+1\right)ix\cdot p\,\frac{p^2+m^2}{m^2}\right]$$

The electron propagator correction to linear order in $\frac{m}{\sqrt{2}} - \not{p}$ is then

$$\mathcal{A}_{2e} \approx \Gamma(1+\epsilon)\left(\frac{m^2}{\mu^2}\right)^{-\epsilon}\left\{\left[\tfrac{1}{2} + \left(\tfrac{1}{4}\frac{1}{\epsilon_{IR}}+1\right)\frac{p^2+m^2}{m^2}\right]\not{p}\right.$$
$$\left. + \left[\frac{1}{\epsilon_{UV}} + 2 + \left(\tfrac{1}{2}\frac{1}{\epsilon_{IR}}+1\right)\frac{p^2+m^2}{m^2}\right](\tfrac{1}{2}\not{p} - 2\tfrac{m}{\sqrt{2}})\right\}$$
$$\approx \Gamma(1+\epsilon)\left(\frac{m^2}{\mu^2}\right)^{-\epsilon}(-\tfrac{1}{2})\left[\tfrac{m}{\sqrt{2}}\left(3\frac{1}{\epsilon_{UV}}+5\right) + \left(\frac{1}{\epsilon_{UV}} + 2\frac{1}{\epsilon_{IR}} + 5\right)(\tfrac{m}{\sqrt{2}} - \not{p})\right]$$

The $1/\epsilon_{UV}$ terms are the same as the $1/\epsilon$ terms obtained above for minimal subtraction. In the MOM scheme, this entire contribution ($\mathcal{O}(K^0)$ and $\mathcal{O}(K^1)$) is canceled by counterterms.

Exercise VIIIA1.2

Repeat the above calculations replacing the fermion with a scalar.

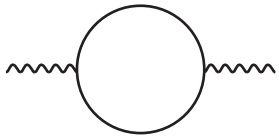
Exercise VIIIA1.3

Repeat the above calculations replacing the photon with a (massless)

a scalar

b pseudoscalar (with a γ_{-1} vertex).

2. Photon



We next calculate the spin-1/2 contribution to the photon (or gluon) self energy: The S-matrix element is

$$\int dk \frac{tr[-\gamma_a(\not{k} - \frac{1}{2}\not{p} + \frac{m}{\sqrt{2}})\gamma_b(\not{k} + \frac{1}{2}\not{p} + \frac{m}{\sqrt{2}})]}{\frac{1}{2}[(k-\frac{1}{2}p)^2 + m^2]\frac{1}{2}[(k+\frac{1}{2}p)^2 + m^2]}$$

The result of the trace (again using 4-dimensional algebra) is

$$-2[k_a k_b - \eta_{ab}\tfrac{1}{2}(k^2 + \tfrac{1}{4}p^2 + m^2)] - \tfrac{1}{2}(\eta_{ab}p^2 - p_a p_b)$$

The first part is the expression appearing in $\hat{\mathcal{A}}_{ab}$ in subsection VIIB5, once we recognize its η_{ab} terms as the average of the denominator factors, yielding tadpoles. The integral thus gives

$$(\eta_{ab}p^2 - p_a p_b)(-2\check{\mathcal{A}} - \tfrac{1}{2}\hat{\mathcal{A}}_2) \approx (\tfrac{1}{\epsilon} - ln\ p^2)(-\tfrac{1}{3})(\eta_{ab}p^2 - p_a p_b)$$

for the divergent and *high-energy* terms. Using

$$A^a(-p)(\eta_{ab}p^2 - p_a p_b)A^b(p) = \tfrac{1}{2}F^{ab}(-p)F_{ab}(p)$$

in terms of the linearized field strength F, the corresponding contributions to the unrenormalized one-loop effective action are (including a factor of $\frac{1}{2}$ for identical external lines)

$$\Gamma_1 \approx \hbar\tfrac{2}{3}\int dx\ \tfrac{1}{8}F^{ab}(\tfrac{1}{\epsilon} - ln\ \Box)F_{ab}$$

(neglecting the "-1" part of $ln(-\Box)$) and the counterterm is thus

$$\Delta S = \tfrac{\hbar}{\epsilon}(-\tfrac{2}{3})\int dx\ \tfrac{1}{8}F^{ab}F_{ab}$$

in the case of QED. For QCD, we must include the group-theory factor $tr(G_i G_j)$ multiplying $F^{iab}F^j_{ab}$. (Examples will be given in the following subsections.)

This propagator correction is easier to analyze in the MOM scheme than the electron propagator, since there are no internal massless particles, and thus no IR divergence to distinguish from the UV one. We therefore just take the explicit expressions for the integrals from subsection VIIB5 and Taylor expand in p^2 about 0 (or actually in $1/\beta$ of VIIB5.1a, substituting for p^2 only at the end). The *low-energy* part of the renormalized effective action for the photon, exhibiting the momentum dependence of the coupling, is then

$$\Gamma_{0+1,2\gamma,r} \approx \int dx\ \tfrac{1}{8}F^{ab}\left(\frac{1}{e^2}+\tfrac{2}{15}\frac{\Box}{m^2}\right)F_{ab}$$

where we have applied MOM subtraction by canceling constant (infinite and finite) constributions to the coupling.

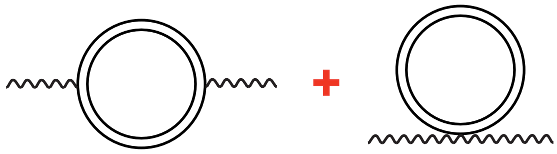
Exercise VIIIA2.1

Evaluate this contribution to the *unrenormalized* effective action to this order. Show that the constant contributions to the coupling (to be canceled by renormalization) are

$$\tfrac{1}{e^2} \to \tfrac{1}{e^2} + \tfrac{2}{3}[\tfrac{1}{\epsilon} - \gamma - ln(\tfrac{1}{2}m^2)]$$

3. Gluon

The most interesting case is the propagator of the Yang-Mills field, in a theory of Yang-Mills coupled to lower spins. There is an important simplification in this calculation in the background field gauge: Writing the classical Yang-Mills Lagrangian as $tr\ F^2/g^2$, the covariant derivative appears as $\nabla = \partial + iA$ without coupling constant, so the gauge transformation of A is coupling independent, as in general for the matter fields. (In terms of a group element $\mathbf{g}$, $\phi' = \mathbf{g}\phi$ and $\nabla' = \mathbf{g}\nabla\mathbf{g}^{-1}$.) The effective action is gauge invariant, which means the only divergent terms involving the Yang-Mills field are the gauge-covariantized kinetic (less mass) terms of the various fields. The divergences for the non-gauge fields are not so interesting, since they can be absorbed by rescaling those fields ("wave-function renormalization"), but the divergence of the $tr\ F^2/g^2$ term can be absorbed only by rescaling the coupling g itself. (On the other hand, if we use $\nabla = \partial + igA$, then renormalization of g requires the opposite renormalization of A to preserve gauge invariance.) Thus this divergence is related to the UV behavior of this coupling (as discussed in subsection VIIB6, and further later). The important point is that there is no wave-function renormalization for the Yang-Mills field (since there is no corresponding gauge-invariant counterterm), so the coupling-constant renormalization (like mass renormalizations) can be found from just the propagator correction, while in other gauges one would need also a much messier vertex (3-point) correction: BRST invariance is not enough to give the result from a single graph.



We now consider the contributions of spins 0 (including ghosts) and 1 (including gluon self-interactions), and redo the spin-1/2 contribution in a way that resembles the bosons. It is based on the observation that there is a universal form for the gauge-covariantized Klein-Gordon equation for spins 0,1/2,1, which can also be shown by supersymmetry. The kinetic operator in a background Yang-Mills field is

$$K = -\tfrac{1}{2}(\Box - iF^{ab}S_{ba})$$

where now $\Box = (\nabla)^2$ is gauge covariantized. This form is true in arbitrary dimensions. For spin 0 it is obvious. For spin 1/2, we use the fact that the one-loop contribution to the functional integral is the trace of the logarithm of the propagator, as follows from Gaussian integration,

$$\int D\psi\ D\bar{\psi}\ e^{-\bar{\psi}K\psi} = det\ K = e^{tr\ ln\ K}$$

where the trace is over all indices, including the coordinates. Then the contribution to the effective action from kinetic operator K is 1/2 the contribution from K^2. (See also exercise VIA4.2.) We then use (see subsection IIIC4)

$$-2\nabla\!\!\!/^2 = -2(\gamma\cdot\nabla)^2 = -(\{\gamma^a,\gamma^b\} + [\gamma^a,\gamma^b])\nabla_a\nabla_b = \Box + iS^{ab}F_{ab}$$

where we have used

$$S^{(1/2)}_{ab} = -\tfrac{1}{2}\gamma_{[a}\gamma_{b]}, \qquad \{\gamma_a,\gamma_b\} = -\eta_{ab}$$

In the case D=4, this is equivalent to the result obtained in subsection IIIC4 in terms of just the undotted spinor, but there the 1/2 is automatically included because there are half as many fields, so the range of the trace is half as big.

For spin 1, we use the result of the background-field version of the Fermi-Feynman gauge: At quadratic order in the quantum fields, from exercise VIB8.1 we have

$$\tfrac{1}{8}F^2 + \tfrac{1}{4}(\partial\cdot A)^2 \to \left\{\tfrac{1}{8}(\mathcal{D}_{[a}A_{b]})^2 + i\tfrac{1}{4}\mathcal{F}^{ab}[A_a,A_b]\right\} + \tfrac{1}{4}(\mathcal{D}\cdot A)^2$$

$$= -\tfrac{1}{4}A\cdot\Box A - i\tfrac{1}{2}A^a[\mathcal{F}_{ab},A^b] = -\tfrac{1}{4}A\cdot(\Box - i\mathcal{F}^{ab}S_{ba})A$$

where $\Box = (\mathcal{D})^2$ contains only the background gauge field, and in the last step we have written the quantum field A as a column vector in the group space and the background fields (like $\mathcal{F}$) as matrices for the adjoint representation (which replaces commutators with multiplication), and used the explicit expression

$$S^{(1)}_{ab} = |_{[a}\rangle\langle_{b]}|, \qquad \langle_a|_b\rangle = \eta_{ab}$$

To this order in the quantum fields, the kinetic operator for the two ghosts looks just like that for two physical scalars, but gives a contribution to the effective action of opposite sign because of statistics.

This method can be used for arbitrary one-loop graphs with external gluons, and easily generalizes to massive fields; we now specialize to propagator corrections. There are two kinds of vertices, the spin-0 kind and the vertex with the spin operator. Since $tr\ S_{ab} = 0$, we get only graphs with either 2 spin vertices or none. There is only one spin graph, with 2 internal free propagators; the 2 spinless graphs include such a graph but also a tadpole, which vanishes by dimensional regularization in the massless case. Since the spinless graphs give the complete result for internal spin-0, their sum is separately gauge invariant; the spin graph is obviously so, since it is expressed directly in terms of the field strength. (We refer here to the Abelian part of the gauge invariance, which is all you can see from just 2-point graphs.) As far as Lorentz index algebra is concerned, we need to evaluate only $tr(S_{ab}S_{cd})$. For the vector, we have

$$tr(S^{(1)}_{ab}S^{(1)}_{cd}) = 2\eta_{b[c}\eta_{d]a}$$

For spin 1/2, the traces are the same as in D=4 except for overall normalization; using earlier identities, or using the same methods for this case directly,

$$tr(S^{(1/2)}_{ab}S^{(1/2)}_{cd}) = \tfrac{1}{4}tr(I)\eta_{b[c}\eta_{d]a}$$

where $tr(I)$ is the size of the spinor.

Exercise VIIIA3.1

Let's look at other ways to interpret the last two identities:

a Use the double-line notation (subsection VC9) for the defining representation of the orthogonal group to derive the above expression for the trace of two $S^{(1)}$'s.

b Use the fermion action of IIIC4 in terms of just undotted spinors for D=4. Evaluate

$$tr(S^{(1/2)}_{\alpha\beta}S^{(1/2)}_{\gamma\delta})$$

using both bra-ket notation and double-line notation for SL(2,C). Show the result is the same as from vector notation (by relating F_{ab} and $f_{\alpha\beta}$).

All diagrams will also have a group-theory factor of $tr(G_i G_j) \sim \delta_{ij}$. We'll be interested mostly in SU(N) for Yang-Mills theory (as appropriate to describe color in the Standard Model for N=3, or arbitrary N for applying the 1/N expansion). Then the most interesting representations are the adjoint (for the gluons and their ghosts) and the defining (for the quarks). As explained in subsection IB2, or as follows from the double-line notation of subsection VC9, we use the normalization

$$tr_D(G_i G_j) = \delta_{ij} \quad \Rightarrow \quad tr_A(G_i G_j) = 2N\delta_{ij}$$

Finally, there are the momentum-space integrals, which have already been evaluated in subsection VIIB4 for the massless case (which is sufficient for determining the high-energy behavior, and thus the UV divergences) and VIIB5 for the massive case. The integral for the spin graph is the same as that for ϕ^3 theory (using the S_{ab} vertex from $-\frac{1}{2}\Box - \frac{1}{2}iF^{ab}S_{ab}$). As labeled there, the external line has momentum p and the internal lines $k \pm \frac{1}{2}p$. Then the vertex factors in the spinless graph with two propagators are both simply $-k$ (from $-\frac{1}{2}\Box = -\frac{1}{2}\partial^2 + \frac{1}{2}A\cdot(-i\partial) + \frac{1}{2}(-i\partial)\cdot A + \frac{1}{2}A^2$), giving $\mathcal{A}_{ab}$, while the addition of the tadpole, with vertex factor $-\eta$, converts it to $\hat{\mathcal{A}}_{ab}$. (By comparison, the tadpole graph that was apparently avoided in the Dirac-spinor calculation of the previous subsection appeared anyway after evaluating the trace algebra.) This contribution also gets an overall $tr(I)$ factor, simply counting the number of degrees of freedom. Note that the scalar factor $\check{\mathcal{A}}$ that appears in $\hat{\mathcal{A}}_{ab}$ is the sum of a divergent term proportional to the ϕ^3 graph and a convergent term that vanishes in the massless case.

We now combine all factors to obtain the contributions to the two-gluon part of the unrenormalized 1-loop effective action (including the -1 for getting the effective action from the S-matrix, a -1 for internal fermions, either spin $\frac{1}{2}$ or ghost, the $\frac{1}{2}$ for identical external gluon lines, the $\frac{1}{2}$ for the spinor to compensate for squaring the propagator, and yet another $\frac{1}{2}$ for identical internal lines if the group representation was real.) The result is the sum of contributions of the form

$$\Gamma_{1,2g} = \hbar\ tr \int dx\ \tfrac{1}{8}F^{ab}(\tfrac{1}{2}c_R)(-1)^{2s}[\tfrac{1}{D-1}\mathcal{B}_1(-\Box) - 4s^2\mathcal{B}_2(-\Box)]F_{ab}$$

where c_R is the group theory factor from the trace, which for the interesting cases is

$$c_R = \begin{cases} 2 & \text{for } N \oplus \bar{N} \text{ (defining)} \\ 2N & \text{for adjoint (real)} \end{cases}$$

This result applies to spins $s = 0, \frac{1}{2}, 1$, with the understanding that it is the result for two polarizations, so there is an implicit extra factor of $\frac{1}{2}$ for a single scalar, while for massive spin 1 (spontaneously broken gauge theories) the third polarization in the (background-field) Fermi-Feynman gauge is carried by a scalar field. (The result above for s=1 is the sum of the contributions from the vector field and the two fermionic ghosts.) The functions $\mathcal{B}_1$ and $\mathcal{B}_2$ are the spinless and spin contributions, related to the massive ϕ^3 propagator correction $\hat{\mathcal{A}}_2(p^2, m^2, m^2)$ as

$$\mathcal{B}_2(p^2) = \hat{\mathcal{A}}_2(p^2, m^2, m^2), \quad \mathcal{B}_1(p^2) = \mathcal{B}_2(p^2) + 4m^2 \frac{\mathcal{B}_2(p^2) - \mathcal{B}_2(0)}{p^2}$$

Note that

$$\mathcal{B}_1 \approx \mathcal{B}_2 \approx \tfrac{1}{\epsilon} - ln\ p^2$$

as far as divergent (at D=4) or high-energy (i.e., massless) terms are concerned. Also note that all contributions *exactly* cancel if all spins are in the adjoint and have the same mass, and appear in the ratio 1:4:6 for spins 1 (including ghosts), $\frac{1}{2}$, 0: For the massless case, this is N=4 super Yang-Mills, which is also the massless sector of the dimensional reduction of the open superstring from D=10. The massless sector of the reduction of the open bosonic string from D=26 yields Yang-Mills plus 22 adjoint scalars, which cancels near D=4 up to finite terms.

In examining the contribution of this term to the running of the coupling constant with energy, we see that the vectors contribute with opposite sign to lower spins. In particular, in terms of the coefficient β_1 (of subsection VIIC1), only nonabelian vectors make positive contributions (since Abelian vectors are neutral). This means that nonabelian vectors are responsible for any weakening in a coupling at high energies, known as "asymptotic freedom", an important experimental feature of the strong interactions (see section VIIIC). Note that while the sign of β_1 for ϕ^4 theory, using the method of subsection VC9, is independent of the coupling (since all 1-loop corrections are coupling-independent when the coupling appears as an overall factor in the classical action, like $\hbar$), changing the sign of the coupling changes its sign relative to β_1: The result is that this theory can be made asymptotically free only if its potential has the wrong sign (negative for large ϕ). Thus, although nonabelian vectors are required for asymptotic freedom in physical theories, "wrong-sign ϕ^4" can be used as a toy model for studying features associated with asymptotic freedom (especially resummation of the perturbation expansion: see section VIIC).

Note that for (massless) fermions that couple chirally to vectors (as in electroweak interactions), c_R consists of the contribution from a complex representation but not its complex conjugate: Only one of the two Weyl spinors of the Dirac spinor contributes. The result is that the contribution to the vector propagator is half that of the parity-invariant case. This fact follows from comparing the calculations of the chiral and nonchiral cases without the squared-propagator trick: In Dirac (4-component) spinor notation, the γ_{-1}'s drop out of the calculation; in Weyl (2-component) spinor notation, the left- and right-handed-spinor diagrams are identical except for (internal) group theory. (Things are more complicated for higher-point functions, because the group theory gives more than just $tr_R(G_i G_j)$: See subsection VIIIB3.)

Exercise VIIIA3.2

Find the conditions for exact cancelation if spins $\frac{1}{2}$ and 0 include both adjoint and defining representations. Find the weaker conditions if only the divergent (and therefore also high-energy) terms cancel.

Exercise VIIIA3.3

Use the optical theorem to find the decay rate for a massive vector (e.g., Z boson) into massive particle-antiparticle pairs of various spins.

Exercise VIIIA3.4

Find the propagator correction for internal particles of *different* masses on each of the two lines (e.g., for a W boson propagator).

In the case of QCD, with color gauge group $SU(N_c)$ and N_f flavors of quarks in the defining representation of color, the divergent and high-energy contributions to this term in the unrenormalized 1-loop effective action are

$$\Gamma_{1,2g,QCD} \approx \hbar \; tr \int dx \; \tfrac{1}{8} F^{ab} \tfrac{1}{3}(2N_f - 11N_c)(\tfrac{1}{\epsilon} - ln \; \Box) F_{ab}$$

At higher loops the effective action will still be gauge invariant in background-field gauges (for the quantum fields), so the renormalization of the Yang-Mills coupling can still be determined from just the gluon propagator correction. On the other hand, in other gauges a three-point vertex must also be calculated: It can be shown that the gauge-fixed classical action, including counterterms, is BRST invariant only up to wave-function renormalizations; i.e, the most general counterterms needed (with a BRST preserving regularization) are BRST-invariant terms with additional multiplicative renormalizations of the quantum fields. Thus, BRST invariance, unlike gauge invariance, is not strong enough to relate the gluon coupling and wave-function renormalizations. Not only does this mean evaluating many more graphs, but graphs which make the propagator correction look easy by comparison. (This is not so difficult for just the one-loop divergences we have considered, but the difficulty grows exponentially with the number of loops.)

However, in the background-field gauge the L-loop propagator correction has $(L-1)$-loop vertex subdivergences, similar to those in other gauges. The net result is: (1) We still have a BRST-invariant "classical" action, containing the same (counter)terms that appear in other gauges (including quantum ghosts), but covariantized with respect to background gauge fields (and including coupling to other background fields). However, the coefficients need be calculated only to order $L-1$ for the L-loop effective action, one loop less than in other gauges. (2) In addition, we have background-field-only terms in the classical action whose L-loop coefficients do need to be calculated, but with a relatively small amount of additional effort, due to gauge invariance. Thus renormalization consists of two steps: (1) adding BRST-invariant counterterms for the quantum fields (background covariantized) to cancel subdivergences, and (2) adding gauge-invariant counterterms for the background fields (which can be interpreted as vacuum renormalization for the quantum fields) to cancel superficial divergences. Consequently, background-field gauges save about one loop of difficulty as far as renormalization is concerned. Furthermore, similar simplifications occur for calculations of finite parts (e.g., effective potentials), because of simplifications from gauge invariance.

4. Grand Unified Theories

The best result of GUTs is their prediction that the gauge couplings of the Standard Model coincide at some high energy, as a consequence of the running of the couplings with energy. (Mixed results have been obtained for masses, arguably because renormalization group arguments are accurate only for high energies, and thus leptons with large masses. A "failed" prediction is proton decay, which has already eliminated the nonsupersymmetric SU(5) model with minimal Higgs.) The numerical details of this prediction are model dependent (and thus easy to fudge, given enough freedom in choice of nonminimal fields), but the fact that all three couplings come close together at high energies is already strong evidence in favor of unification.

Thus we make only the crudest form of this calculation, using only the one-loop results of the previous subsection. The main assumption is that there is a "desert" between the Standard Model unification scale (around the masses of the intermediate vector bosons W and Z) and the Grand Unification scale M_{GUT}, with no fundamental particles with masses in that range (although, of course, a huge number of hadrons appear there). This allows us to crudely approximate all fundamental particles below that region (i.e., those of the Standard Model) as massless, and all above as infinitely massive. In particular, in the framework of the minimal SU(5) GUT, this means *all* the fermions are treated as massless.

Therefore the calculation is to use the one-loop results to calculate the running of the couplings in the Standard Model, and use the relation of the gauge couplings in the SU(5) GUT to identify those of the Standard Model in terms of that of this GUT. From the previous subsection, the running of the couplings is given by

$$\frac{1}{g^2(-p^2)} \approx \frac{1}{g_0^2} - \beta_1 ln\left(\frac{M_{GUT}^2}{-p^2}\right), \qquad \beta_1 = \sum_{R,s} \tfrac{1}{2} c_R (-1)^{2s}(4s^2 - \tfrac{1}{3})$$

for two helicities of spin s (with an extra factor of $\frac{1}{2}$ for only 1 helicity of spin 0), where $g_0 \equiv g(M_{GUT}^2)$.

If we use g_1, g_2, g_3 to label the couplings of U(1), SU(2), and SU(3) that are identified with the single SU(5) gauge coupling at the unification scale, then their relation to those of the Standard Model (as normalized in exercise IVB2.1) is

$$\frac{1}{g_1^2} = \tfrac{6}{5}\frac{1}{g'^2} = \tfrac{6}{5}\frac{cos^2\theta_W}{e^2}, \qquad \frac{1}{g_2^2} = \frac{1}{g^2} = 2\frac{sin^2\theta_W}{e^2}, \qquad \frac{1}{g_3^2} = \frac{1}{g_s^2}$$

where g_s and g the usual SU(3) and SU(2) couplings, and the factor of $\frac{6}{5}$ is because the U(1) generator (see subsection IVB4) satisfies $tr_D(G^2) = \frac{5}{6}$ in terms of SU(5) matrices. (We generally normalize to $tr_D(G^2) = 1$ for each generator. Physical couplings are preserved if changes in normalization of generators are accompanied by changes in coupling normalization so as to preserve $g_i G_i$.)

Then the values of the β_1's for the Standard Model are

$$\beta_{1,1} = 0 - 4 - \tfrac{1}{10} = -\tfrac{41}{10}, \qquad \beta_{1,2} = \tfrac{22}{3} - 4 - \tfrac{1}{6} = \tfrac{19}{6}, \qquad \beta_{1,3} = 11 - 4 + 0 = 7$$

where we have listed the contributions from spins 1, $\frac{1}{2}$ (for 3 families), 0, respectively. (Note that the spinors contribute the same to each because they are all effectively massless: They don't notice the SU(5) breaking. Also, we can ignore SU(2)⊗U(1) breaking when calculating these β's, since we have neglected the corresponding masses.)

Exercise VIIIA4.1

Calculate the contribution of the spinors to the β_1's, in terms of both SU(5) and SU(3)⊗SU(2)⊗U(1) multiplets. (Note the chiral couplings for spinors, so for c_R a complex representation and its complex conjugate might not both contribute.)

The experimental values of the couplings (in the $\overline{\text{MS}}$ prescription) at $\mu = M_Z \approx$ 91 GeV are

$$\frac{1}{e^2} \approx 804, \qquad sin^2\theta_W \approx .231, \qquad \frac{1}{g_s^2} \approx 106$$

Unfortunately, taking any two of the equations for $1/g_i^2$ gives widely varying answers: e.g.,

$$M_{GUT} \approx 10^{15\pm2} GeV$$

Alternatively, since we have used only two parameters to fit three experimental numbers, we can try to predict the value of any one of e, θ_W, or g_s from the rest: e.g., from e and g_s we can find

$$sin^2\theta_W \approx .207$$

which shows the same disagreement (but looks better than the exponentiated error for M_{GUT}).

The result is not very accurate, since we have made many approximations, which can be improved with some effort: Two-loop corrections add $ln\ ln$ terms to the one-loop ln terms; including the mass dependence of the effective couplings also adds significant corrections. But the most important approximation assumption we made was the desert: Undiscovered particles, such as new fermions, nonminimal Higgs, or supersymmetric partners, change even the one-loop expressions β_1. Specifically, since by definition the unification scale is where the masses of all unobserved vectors reside, these new particles will all have spins 0 or $\frac{1}{2}$, and thus make the β's more negative. In particular, supersymmetrization yields a result consistent with experiment, with

$$M_{GUT} \approx 2.2 \times 10^{16} GeV$$

(This has been interpreted as the only experimental verification of supersymmetry.)

Exercise VIIIA4.2

Let's examine the effects of supersymmetry:

a Supersymmetrize the Standard Model contributions to β_1 by adding the supersymmetric partners to each spin: $1 \to 1 \oplus \frac{1}{2}$, $\frac{1}{2} \to \frac{1}{2} \oplus 0 \oplus 0$, $0 \to \frac{1}{2} \oplus 0 \oplus 0$ (where the Higgs scalars have doubled because chiral scalar superfields can't satisfy reality conditions) to find the result

$$\beta_{1,1} = 0 - 6 - \tfrac{3}{5} = -\tfrac{33}{5}, \qquad \beta_{1,2} = 6 - 6 - 1 = -1, \qquad \beta_{1,3} = 9 - 6 + 0 = 3$$

b Solve for $1/g_0^2$, $ln(M_{GUT}^2/M_Z^2)$ (and thus M_{GUT}), and $sin^2\theta_W$ in terms of $1/e^2$ and $1/g_s^2$. Then plug in to find the numerical values.

c Show the consistency condition relating the 3 couplings is

$$\frac{\beta_{1,2} - \beta_{1,3}}{g_1^2} + \frac{\beta_{1,3} - \beta_{1,1}}{g_2^2} + \frac{\beta_{1,1} - \beta_{1,2}}{g_3^2} = 0$$

and that the closest integer values for the couplings from the above data,

$$\frac{1}{g_1^2} = 742, \qquad \frac{1}{g_2^2} = 371, \qquad \frac{1}{g_3^2} = 106$$

satisfy it exactly. (OK, so this is just a numerical coincidence, considering experimental inaccuracies and theoretical approximations, but isn't it still nice?) Also, note that

$$\frac{1}{g_1^2} = 2\frac{1}{g_2^2} \quad \Rightarrow \quad sin^2\theta_W = \tfrac{3}{13}$$

d Drop the contributions of the Higgs (and its superpartners) to the β's in both the supersymmetric and nonsupersymmetric cases, and reevaluate $sin^2\theta_W$, showing both give the same (poor) value. (Thus, Higgs can make a difference.)

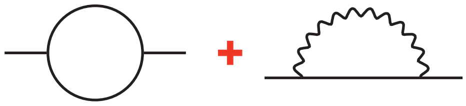
5. Supermatter

Although the problem with infrared renormalons may be only technical, the appearance of this same problem in several different approaches (including a nonperturbative one; see later) strongly suggests that the "correct" approach to quantum field theory, in the sense of a practical method for unambiguously (i.e., with predictive power) calculating perturbative and nonperturbative effects, might be to consider only theories that are perturbatively finite. In this subsection we will analyze general properties of supersymmetric field theory using superspace, and in particular improved UV behavior, concentrating on finite theories.

Finite supersymmetric theories must be in particular one-loop finite. This turns out to be enough to guarantee finiteness to all loops: Two-loop finiteness is automatic, while an appropriate renormalization prescription is required to guarantee finiteness is preserved order by order in perturbation theory. (No constraints on the coupling constants are needed beyond those found at one loop, but without the renormalization prescription infinities cancel between different loop orders.) Of course, wave-function renormalizations are gauge dependent: N=1 supersymmetric gauges eliminate some of these unphysical divergences (and gauges with higher supersymmetry more), as do background-field gauges even in nonsupersymmetric theories. So, "finite theory" in general gauges refers only to the "physical" divergences — those that affect the high-energy behavior of the theory, namely those that appear in couplings and masses.

Because of the nonrenormalization of chiral terms in the action (see subsection VIC5), it might seem that the corresponding couplings and masses are always unrenormalized. However, the kinetic terms of chiral superfields can receive quantum corrections, and the true couplings are defined by field redefinitions that eliminate these rescalings. This means that all such renormalizations are related, and given by the wave function renormalizations. The only other couplings are the Yang-Mills ones, whose renormalization is also given by kinetic terms in background-field gauges. Thus, all "physical" renormalizations in supersymmetric theories can be found from just propagator corrections. In particular, this means that if the effective action is calculated with background-field supergraphs, then it is completely finite in a finite theory.

A possible exception to our statement of all physical renormalizations coming from propagator corrections would seem to be the Fayet-Iliopoulos tadpole term $\int d^4\theta\ V$. However, massless tadpoles vanish in dimensional regularization, and massive ones require real representations, which cannot generate explicit-prepotential terms. (In particular, at more than one loop such terms never appear in the background-field gauge for any representation.)



The simplest one-loop propagator correction is to $\bar{\phi}\phi$. (The $\phi\phi$ correction vanishes, since $\int d^4\theta\ \phi^2 = 0$: See subsection VIC5.) There are two graphs to consider, one with two internal $\bar{\phi}\phi$ propagators, and one with internal $\bar{\phi}\phi$ and VV propagators. The d algebra for the two graphs is identical: Both get a d^2 and a $\bar{d}^2$ inside the loop, exactly enough to give a nonvanishing graph (using $[\bar{d}^2 d^2 \delta^4(\theta-\theta')]|_{\theta'=\theta} = 1$). There is also a symmetry factor of $\frac{1}{2}$ for the two $\bar{\phi}\phi$ propagators, and a -1 for the mixed graph because the two different types of internal propagator have opposite sign (and, as usual, an overall -1 to get Γ from the T-matrix). Thus, the supersymmetry (spin) part of the algebra is almost trivial in this case.

Exercise VIIIA5.1

Use component methods to evaluate the first graph with external fermions: the contribution of the Yukawa interaction to the fermion propagator. Show it agrees with the supergraph evaluation.

On the other hand, the internal group theory is slightly messy, so we treat the general case immediately: We take vector multiplets V^i for an arbitrary group (though we will need a semisimple group for finiteness, since Abelian groups are not even asymptotically free). Sums $\sum_G$ are over each simple subgroup (or each Abelian factor), since they can have independent coupling constants g_{RG} for representation R (especially g_{AG} for the adjoint, which we use for the pure super Yang-Mills term for definiteness; except for the Abelian factors, where a nontrivial representation should be substituted). Similarly, sums $\sum_R$ are over irreducible representations of the group;

$\Pi_{RI}{}^J$ is the corresponding projection operator. For the simple (or single-component Abelian) factors of the group

$$\eta_{Gij} = c_{AG}\Pi_{Gij}$$

is used (but again, with a different normalization for the Abelian factors). We also use the group theory identities (normalizations) from subsection IB2, now generalized to these nonsimple groups and reducible representations:

$$G_{iI}{}^K G_{jK}{}^J \Pi_{RJ}{}^I = \sum_G c_{RG}\Pi_{Gij} = \sum_G \frac{c_{RG}}{c_{AG}}\eta_{Gij}$$

$$G_{iI}{}^K G_{jK}{}^J \eta_G^{ij} = \sum_R k_{RG}\Pi_{RI}{}^J = \sum_R \frac{c_{RG}d_{AG}}{c_{AG}d_R}\Pi_{RI}{}^J$$

Then from the Lagrangian

$$L = -\int d^4\theta\ \bar\phi^I (e^V)_I{}^J\phi_J + \left(\int d^2\theta\ \tfrac{1}{6}\lambda^{IJK}\phi_I\phi_J\phi_K + h.c.\right)$$
$$-\sum_G \frac{1}{g_{AG}^2}\int d^2\theta\ \tfrac{1}{2}W^{i\alpha}W^j_\alpha \eta_{Gij}$$

(ignoring mass terms) the result is simply

$$\Gamma_{1,\phi\bar\phi} = \hbar\int d^4\theta\ \bar\phi^I M_I{}^J \hat{\mathcal{A}}_2\phi_J, \qquad M_I{}^J = \sum_{R,G} g_{AG}^2\frac{c_{RG}d_{AG}}{c_{AG}d_R}\Pi_{RI}{}^J - \tfrac{1}{2}\bar\lambda_{IKL}\lambda^{JKL}$$

where again $\hat{\mathcal{A}}_2$ is the operator representing the one-loop propagator correction (T-matrix) for self-interacting scalars. (Of course, this operator may vary depending on the internal masses; here we are concerned mostly with the divergences and leading high-energy behavior, which is mass-independent. As usual, we can rescale the gauge fields by their couplings in the Lagrangian; this moves these couplings from the propagators into the vertices, giving the same result for this term in Γ, since it has no V's.) Of course, this is the identical group theory that appears in the nonsupersymmetric case; we have been more general here because we want to consider exact cancelation, while in the nonsupersymmetric case simplicity is usually more important.

6. Supergluon

The supergluon self-energy calculation is similar to the nonsupersymmetric cases considered in subsections VIIIA2-3. Examining the Feynman rules, we see that those for the vector multiplets are similar to the nonsupersymmetric ones for vectors (as expected), while those for the scalar multiplets are similar to those for spinors: d^2 and $\bar d^2$ are analogous to (in 2×2 matrix notation) ∂ and ∂^*, etc.

There are now only two kinds of loops to consider, vector and scalar multiplets: As for the nonsupersymmetric case, ghosts in background-field gauges couple the same as matter, since at one loop the only coupling is to background fields and thus covariant, even for ghosts. For the real scalar superfield describing the quantum vector multiplet, looking at the terms in the action quadratic in V (from subsection VIB10)

$$S_{2V} = \int dx\ d^4\theta\ \tfrac{1}{4}V(\Box + 2i\mathcal{W}^\alpha \mathcal{D}_\alpha + 2i\overline{\mathcal{W}}^{\dot\alpha}\bar{\mathcal{D}}_{\dot\alpha})V$$

we see that vertices have only 1 spinor derivative at most. However, we need at least 4 spinor derivatives (2 d's and 2 $\bar{d}$'s) per loop (see subsection VIC5), since the result of reducing any loop to a point in θ space always leaves the tadpole θ-integral $[d...d\delta^4(\theta-\theta')]|_{\theta'=\theta}$, which vanishes for fewer than 4 derivatives. Thus, a V loop in a super Yang-Mills background vanishes for fewer than 4 external lines. This means the entire contribution of quantum super Yang-Mills to the supergluon propagator correction (or 3-point correction from real representations) in the background-field gauge comes from the 3 ghosts (including the Nielsen-Kallosh ghost), which couple the same as -3 scalar multiplets in the adjoint representation. Thus, for example, we see without evaluating a single graph that this correction vanishes for N=4 super Yang-Mills, which has also 3 physical adjoint scalar multiplets. (See subsection IVC7.)

For the scalar multiplets, we can find the analog of the squared-propagator trick: The easiest way is by the method of subsection IIIC4, which automatically takes care of factors of $\frac{1}{2}$, and can be applied classically, without worrying about functional determinants. This method requires we consider the massive theory at intermediate stages of the calculation, although the mass can be dropped at the end. The only resulting limitation is that we must restrict to real representations of the gauge group. (In other words, the couplings must preserve parity: For these terms, CP invariance is automatic, and reality means C invariance, so P invariance is implied.) However, this is a restriction of the usefulness of the squared-propagator trick anyway: Otherwise we get expressions like $(\partial\!\!\!/+iA\!\!\!/)(\partial\!\!\!/-iA\!\!\!/^*)$ which do not yield useful simplifications. (They require as much work as without the trick.) In such cases we are stuck with doing the calculations the hard way. This is not just a technical difficulty, it is a consequence of the final result being messier in such cases: For example, for real representations there is no possibility of anomalies. However, we can separate the generators into the real (scalar) and imaginary (pseudoscalar) ones: Then this trick simplifies the real (polar vector) couplings but not the imaginary (axial vector) couplings. (As for Pauli-Villars in subsection VIIIB2 below, but also for the physical fields before taking the mass to vanish after the trick has been applied, the mass term can be chosen to preserve the polar symmetries and thus violate the axial ones.)

However, by comparison of the propagator correction for complex and real representations without (the supersymmetric version of) the squared-propagator trick, we see that the only difference between the two is in the (Yang-Mills) group theory. Thus, we can calculate for real representations first, using the trick, and then for complex representations by simply replacing the group-theory factor in the result for the real ones.

Repeating the procedure of subsection IIIC4 with spinors replaced with chiral superfields, we begin with the Lagrangian ($S = \int dx\ L$)

$$L = -\int d^4\theta\ \bar{\phi}\phi + \tfrac{m}{\sqrt{2}}\left(\int d^2\theta\ \tfrac{1}{2}\phi^2 + \int d^2\bar{\theta}\ \tfrac{1}{2}\bar{\phi}^2\right)$$

where the chiral superfields are *covariantly* chiral (or background-covariantly chiral)

$$\overline{\nabla}_{\dot{\alpha}}\phi = \nabla_\alpha\bar{\phi} = 0$$

Treating $\bar{\phi}$ as auxiliary (the $\bar{\phi}^2$ term has no Yang-Mills coupling, as can be seen, e.g., in an "antichiral" representation), we eliminate it by its algebraic field equation

$$\bar{\phi} = \tfrac{\sqrt{2}}{m}\nabla^2\phi$$

After a trivial rescaling

$$\phi \to 2^{-1/4}\sqrt{m}\phi$$

(and using $\int d^4\theta = \int d^2\theta\ \overline{\nabla}^2$) we obtain the action

$$L_\phi = -\int d^2\theta\ \tfrac{1}{2}\phi(\overline{\nabla}^2\nabla^2 - \tfrac{1}{2}m^2)\phi = -\int d^2\theta\ \tfrac{1}{4}\phi(\Box - m^2 + i[W^\alpha, \nabla_\alpha])\phi$$

($\Box = \nabla^a\nabla_a$) using an identity from subsection VIC5.

In the chiral vacuum-bubble loop, we no longer have an explicit chiral superfield to convert $\overline{\nabla}^2\nabla^2$ to $\Box + ...$. However, using the chiral representation $\overline{\nabla}^2 = \bar{d}^2$, we can write the kinetic operator as

$$\bar{d}^2\nabla^2 = \bar{d}^2 d^2 + \bar{d}^2(\nabla^2 - d^2)$$

to separate the truly free part from the background interactions. Then quantization can be preformed as usual (see subsection VIC5): Essentially, we can now use the free $-\Box_0 + m^2 = p^2 + m^2$ as kinetic operator, since at each vertex there is a $\bar{d}^2$ to project back to chiral superfields. Of course, in general we need only one projector in any trace over a subspace: In this case that result is obtained by integrating the $\bar{d}^2$'s by parts in the loop back and forth across the free propagators, since sandwiching any $\nabla^2 - d^2$ between them produces

$$\bar{d}^2(\nabla^2 - d^2)\bar{d}^2 = \tfrac{1}{2}(\Box - \Box_0 + i[W^\alpha, \nabla_\alpha])\bar{d}^2$$

Repeating the procedure till only one $\bar{d}^2$ is left, the Feynman rules for this loop become

$$\begin{aligned} propagator: &\quad \frac{1}{\frac{1}{2}(p^2+m^2)}\delta^4(\theta-\theta') \\ one\ vertex: &\quad \bar{d}^2(\nabla^2 - d^2) \\ other\ vertices: &\quad \tfrac{1}{2}(\Box - \Box_0 + i[W^\alpha, \nabla_\alpha]) \end{aligned}$$

We thus see that one vertex has at most 3 derivatives ($\bar{d}^2 d$) while the other has at most 1 (d):

$$\bar{d}^2(\nabla^2 - d^2) = \bar{d}^2[iA^\alpha d_\alpha + \tfrac{1}{2}i(d^\alpha A_\alpha) - \tfrac{1}{2}A^\alpha A_\alpha]$$

$$\tfrac{1}{2}(\Box - \Box_0 + i[W^\alpha, \nabla_\alpha]) = iW^\alpha d_\alpha + \tfrac{1}{2}i(d^\alpha W_\alpha) - \tfrac{1}{2}[W^\alpha, A_\alpha] + iA^a\partial_a + \tfrac{1}{2}i(\partial^a A_a) - \tfrac{1}{2}A^a A_a$$

exactly the minimum needed. (Thus, there are insufficient derivatives for a tadpole contribution to the propagator.) The result for this diagram is then the same as the corresponding diagram in bosonic φ^3 theory, with a group theory factor $tr(G_i G_j)$, and replacing $\varphi(-p)\varphi(p)$ with

$$\int d^4\theta\ d^4\theta'\ [iW^{i\alpha}(-p,\theta')d'_\alpha\delta^4(\theta-\theta')][iA^{j\beta}(p,\theta)\bar{d}^2 d_\beta\delta^4(\theta-\theta')]$$

$$= \int d^4\theta\ \tfrac{1}{2}W^{i\alpha}(-p,\theta)A^j_\alpha(p,\theta) = \int d^2\theta\ \tfrac{1}{2}W^{i\alpha}(-p,\theta)W^j_\alpha(p,\theta)$$

using $d\delta = -d'\delta$, integration by parts, $[\bar{d}^2 d^2\delta^4(\theta-\theta')]|_{\theta'=\theta} = 1$, and $W_\alpha = \bar{d}^2 A_\alpha$ (chiral representation). Written in the notation of subsection VIIIA3, the 2-supergluon part of the unrenormalized 1-loop effective action is then

$$\Gamma_{1,2sg} = -\hbar\ tr \int dx\ d^2\theta\ \tfrac{1}{2}W^\alpha(\tfrac{1}{2}c_R)\hat{\mathcal{A}}_2 W_\alpha$$

for a scalar multiplet, and exactly -3 times that for a vector multiplet, including the massive case. Thus, cancelations again survive the introduction of masses. Also, if the masses of the various scalar multiplets are equal the entire propagator correction is canceled in such theories, while for unequal masses only the divergence, and the corresponding leading (logarithmic) high-energy term, is canceled.

Exercise VIIIA6.1

Show this result agrees with the restriction to N=1 supersymmetric theories of the component result of subsection VIIIA3.

Exercise VIIIA6.2

Take the result of subsection VIIIA3 literally for *all* spins s (arbitrarily large). Using the fact that multiplets with N+1 supersymmetries can be written as 2 multiplets with N supersymmetries, differing in maximum helicity by 1/2, recursively find the result for general s (now labeling maximum helicity) for all values of N$\geq$ 1, and show it vanishes for N$\geq$ 3.

Exercise VIIIA6.3

Calculate the chiral scalar contribution to the one-loop supergluon propagator correction without the squared-propagator trick. (Hint: There are 8 spinor derivatives in the loop. Integrating them by parts off one propagator produces 3 terms, since the number of d's and $\bar{d}$'s inside must be equal, because what's left is always spacetime derivatives on $[\bar{d}^2 d^2 \delta^4(\theta-\theta')]|_{\theta'=\theta}$.)

Generalizing the group theory as in the previous subsection, we have the total result

$$\Gamma_{1,VV} = -\hbar \int dx\ d^2\theta \sum_G \tfrac{1}{2} W^{i\alpha}(\tfrac{1}{2}M_G)\eta_{Gij}\hat{\mathcal{A}}_2 W^j_\alpha, \qquad M_G = \sum_R \frac{c_{RG}}{c_{AG}} - 3$$

(For Abelian factors, irrelevant for finiteness, we should take the c_{AG} factor out of η_{Gij} and put it into M_G; then $c_{AG}=0$ for Abelian groups, so $M_G \to \sum_R c_{RG} > 0$.) Therefore, combining with the results of the previous subsection, the conditions for finiteness are

$$\boxed{\sum_R \frac{c_{RG}}{c_{AG}} = 3, \qquad \sum_{R,G} g^2_{AG}\frac{c_{RG}d_{AG}}{c_{AG}d_R}\Pi_{RI}{}^J = \tfrac{1}{2}\bar{\lambda}_{IKL}\lambda^{JKL}}$$

In particular, for the case of N=4 super Yang-Mills written in terms of N=1 superfields (see subsection IVC7), we have 3 adjoint chiral scalars ϕ_I with $I = iI'$, where i is the adjoint label and $I' = 1,2,3$ (which appeared as the label I in subsection IVC7, where the adjoint label was implicit in matrix notation). Then

$$\lambda^{IJK} = g_A f^{ijk}\epsilon^{I'J'K'}$$

and the above two finiteness conditions reduce to

$$\delta^{I'}_{I'} = 3, \qquad \delta^{J'}_{I'} = \tfrac{1}{2}\epsilon_{I'K'L'}\epsilon^{J'K'L'} \qquad (\delta^j_i = f_{ikl}f^{jkl})$$

As explained in the previous subsection, in the general case the finiteness conditions may receive quantum corrections at 3 loops and beyond, depending on the model and renormalization prescription, but no new conditions are added.

Presently there is no deep understanding for the finiteness of these models (at least, not deep enough to always avoid the quantum corrections to the finiteness conditions). Note that they are finite for arbitrary values of the couplings, up to the two above restrictions: For example, we can scale all the couplings by a common factor. Thus, they are finite order-by-order in perturbation theory (loops). Nonsupersymmetric theories can also be finite, but only for specific numerical values of the coupling, i.e., not for arbitrarily small values of the coupling, and thus not order-by-order in the loop expansion; they therefore suffer from the renormalon problem. (The renormalon-like behavior of instantons is not a problem in the framework of the $1/N_c$ expansion.) The finiteness of theories with extended supersymmetry has been explained by various arguments (in particular, for N=2 there are no divergences beyond 1 loop even for theories that are just renormalizable), but none of these applies to the general case of simple supersymmetry.

To obtain more realistic models, we may want to consider adding "soft" supersymmetry breaking terms (those which have little effect on high-energy behavior), as introduced in subsection IVC6, to these finite theories. Finiteness can be maintained, but the conditions become considerably more complicated in the general case. Note that spontaneous breaking of supersymmetry is not allowed, because the first condition prohibits U(1) factors (with $c_{AG} = 0$; thus no $\int d^4\theta\ V$ terms), while the second prohibits gauge-singlet matter (with $c_{RG} = 0$; thus no $\int d^2\theta\ \phi$ terms).

7. Bosonization

A common method in field theory is to consider simpler models where calculations are easier, and see if they are analogous enough to give some insight. In particular, two-dimensional models sometimes have perturbative features that are expected only nonperturbatively in four dimensions: For example, we saw in subsection VIIB3 the generation of bound states at one loop in the 2D CP(n) model. Of course, some of the features may be misleadingly simple, and may have no analog in D=4. Two-dimensional theories, especially free, massless ones, are also useful to describe the quantum mechanics of the worldsheet in string theory (see chapter XI). In this subsection we consider free, massless 2D theories: Essentially, this means just the scalar and the spinor, since there are no transverse dimensions to give gauge fields nontrivial components.

Spinor notation is very simple in D=2, since the Lorentz group is SO(1,1)=GL(1). For that purpose it's convenient to use lightcone notation. 2D γ matrices can be chosen as

$$\gamma_+ = \begin{pmatrix} 0 & 0 \\ i & 0 \end{pmatrix}, \quad \gamma_- = \begin{pmatrix} 0 & -i \\ 0 & 0 \end{pmatrix}, \quad \gamma_{-1} = \tfrac{1}{\sqrt{2}}\begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}; \quad \varUpsilon = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \sqrt{2}\gamma_0$$

$$\varPsi = \begin{pmatrix} \psi_\oplus \\ \psi_\ominus \end{pmatrix}, \quad \bar{\varPsi} = \begin{pmatrix} i\bar{\psi}_\ominus & -i\bar{\psi}_\oplus \end{pmatrix}$$

In general, even-D γ matrices can be constructed as direct products of D/2 sets of 2D γ matrices, so $tr(I) = 2^{D/2}$. (For details, see subsection XC1.)

The Lagrangian for a massless, complex spinor can be written this way as

$$L = \bar{\varPsi} i\partial\!\!\!/ \varPsi = \bar{\psi}_\oplus(-i\partial_{\ominus\ominus})\psi_\oplus + \bar{\psi}_\ominus(-i\partial_{\oplus\oplus})\psi_\ominus$$

(This also follows from truncation of 4D spinor notation.) Note that $\psi_\oplus$ and $\psi_\ominus$ transform independently under proper Lorentz transformations, as do their real and imaginary parts. Thus, we can not only impose a reality condition, but also a chirality condition, dropping $\psi_\oplus$ or $\psi_\ominus$: A single real component is enough to not only define a spinor Lorentz representation, but also construct an action.

In position space, the propagator for a massless scalar is (see exercise VIIB1.1)

$$\frac{1}{-\frac{1}{2}\Box} 2\pi(-i)\delta^2(x-x') = -ln[(x-x')^2]$$

up to a real, dimensionful constant: We use units $\mu = 1$. The "$-i$" for the δ comes from Wick rotation back to Minkowski space. The propagator for a massless spinor is then ($\Box = -2\partial_+\partial_-$, shortening the double $\oplus\oplus$'s and $\ominus\ominus$'s from spinor to lightcone vector notation)

$$\frac{1}{-i\partial_\mp} 2\pi(-i)\delta^2(x-x') = i\partial_\pm\{-ln[(x-x')^2]\} = \frac{-i}{(x-x')^\pm}$$

which has the unusual consequence (its wave equation)

$$\partial_\pm \frac{1}{(x-x')^\mp} = 2\pi i\delta^2(x-x')$$

with an appropriate ϵ-prescription implicit:

Exercise VIIIA7.1

Show (e.g., by an infinitesimal Wick rotation) that the correct $i\epsilon$ prescription for the spinor propagator is

$$\frac{-i}{(x-x')^\pm - i\epsilon\epsilon(t-t')} = \theta(t-t')\frac{-i}{(x-x')^\pm - i\epsilon} + \theta(t'-t)\frac{-i}{(x-x')^\pm + i\epsilon}$$

and that it satisfies the wave equation. ($t-t'$ can be replaced with $(x-x')^\mp$ in the above.)

In D=2, the theory of a massless spinor field is equivalent to that of a massless scalar field ("bosonization"). The explicit correspondence between the spinor and scalar is shown by separating the scalar into its left-propagating and right-propagating parts: This can be accomplished by differentiating with respect to $x^\pm$ and then integrating back. We thus write

$$\phi = \phi_{(+)} + \phi_{(-)}$$

where $\phi_{(\pm)}$ is a function of only $x^\pm$ and not $x^\mp$ on-shell, and has propagator $-ln\ x^\pm$.

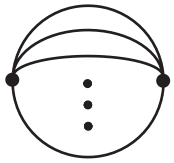
The product of two fermion propagators, for $\bar\psi(x)\psi(x')$ and $\bar\psi(x')\psi(x)$, as would follow from multiplying $\bar\psi(x)\psi(x)$ with $\bar\psi(x')\psi(x')$ in the functional integral (including the -1 for the fermion "loop", from reordering the ψ's) is

$$-\left(\frac{-i}{(x-x')^\pm}\right)\left(\frac{-i}{(x'-x)^\pm}\right) = \partial_\pm\partial'_\pm\{-ln[(x-x')^2]\}$$

This shows that $\partial_\pm\phi_{(\pm)}$ has the same propagator as $\bar\psi_\alpha\psi_\alpha$, with $\alpha = (\oplus,\ominus)$ not summed, so we can equate

$$\partial_\pm\phi_{(\pm)} = \bar\psi_\alpha\psi_\alpha$$

which gives an explicit expression for the scalar in terms of the spinor after integration over $x^\pm$.



The inverse relation is (quantum mechanically, not classically)

$$\psi_\alpha = e^{-i\phi_{(\pm)}}$$

which can be checked by a sum of multiloop diagrams: The equivalent of connecting $\psi_\alpha(x)$ to $\bar\psi_\alpha(x')$ by a single propagator is to expand $e^{-i\phi_{(\pm)}(x)}$ and $e^{i\phi_{(\pm)}(x')}$ each to nth order and connect x and x' with n $\phi_{(\pm)}$ propagators, and sum over n (with appropriate combinatoric factors). The $\phi_{(\pm)}$ propagators are

$$-ln[i(x-x')^\pm]$$

with the "i" determined by Wick rotation. The propagator of the composite fermions is then:

$$\sum_{n=0}^\infty \tfrac{1}{n!}\{-ln[i(x-x')^\pm]\}^n = \frac{-i}{(x-x')^\pm}$$

Although this gives the appearance of a scalar being the bound state of spinors, and vice versa, even in the free theory, there is a simpler interpretation, even classically: Massless particles in D=2 travel at the speed of light in one of two possible directions. Thus, a collection of free "left-(or right-)handed" massless particles travels along together, not separating, and thus acting like a bound state. (As shown in subsection VC8, singularities in perturbative quantum field theory directly correspond to configurations in classical mechanics.)

Bosonization extends to massive fermions: The "massive Thirring model"

$$L = \bar{\psi}_\oplus(-i\partial_{\ominus\ominus})\psi_\oplus + \bar{\psi}_\ominus(-i\partial_{\oplus\oplus})\psi_\ominus + \tfrac{m}{\sqrt{2}}(\bar{\psi}_\oplus\psi_\ominus + \bar{\psi}_\ominus\psi_\oplus) + g\bar{\psi}_\oplus\bar{\psi}_\ominus\psi_\oplus\psi_\ominus$$

is equivalent to the "sine-Gordon model"

$$L = \tfrac{1}{\beta^2}[\tfrac{1}{4}(\partial\phi)^2 + \tfrac{1}{2}\mu^2(1 - cos\ \phi)]$$

with the above relation between the spinor and scalar fields, and

$$\frac{1}{\beta^2} = 1 + 2g, \qquad \frac{\mu^2}{\beta^2} \sim m$$

(Note in particular the free massive fermion for $\beta = 1$.) In this case the bound states are dynamical. Note that the relation is between strong coupling in one theory and weak in the other ("duality").

8. Schwinger model

The simplest interacting model in D=2 is the "Schwinger model", massless QED. This theory is even simpler than scalar theories because its interactions occur only through a massless gauge vector, which has no physical polarizations in two dimensions (D−2=0).

The most interesting feature of the Schwinger model is that all amplitudes with external vectors can be calculated exactly. In fact, the only nonvanishing 1PI vector amplitude is the one-loop propagator correction, which gives just a mass term. In that sense the theory is trivial, and describes just a massive vector. However, the methods of calculation are instructive. We first consider some simple methods of calculation of just the propagator correction, and then show that it is the only 1PI vector graph. One method we have already considered is dimensional regularization; from subsection VIIIA2-3 we have the contribution to the effective action (correcting for the 2D normalization $tr(I) = 2$)

$$\Gamma_1 = \int dx\ F\frac{1}{\Box}F$$

where we write $F_{ab} = \epsilon_{ab}F$ in D=2. Although this calculation needs no renormalization, regularization is still necessary to allow naive manipulation of the integrand: Using dimensional regularization, we see from the result of subsection VIIIA3 that we get a factor of $\frac{1}{D-1} - 4s^2 \sim \epsilon$ in $D = 2+2\epsilon$, canceling the $1/\epsilon$ pole from the scalar integral.

It can also be calculated in position space, using the methods of the previous subsection. The Lagrangian in lightcone notation is

$$L = -\tfrac{1}{4e^2}F^2 + [\bar{\psi}_\oplus(-i\partial_{\ominus\ominus} + A_{\ominus\ominus})\psi_\oplus + \bar{\psi}_\ominus(-i\partial_{\oplus\oplus} + A_{\oplus\oplus})\psi_\ominus]$$

$$F = \partial_{\oplus\oplus}A_{\ominus\ominus} - \partial_{\ominus\ominus}A_{\oplus\oplus}$$

We can calculate separately the contributions of $\psi_\oplus$ and $\psi_\ominus$ to fermion loops. The "photon" propagator correction consists of the product of two fermion propagators, as given in the previous subsection. We then find for the effective action (including another -1 for $T \to \Gamma$ and a $\frac{1}{2}$ for identical external lines), after including a *finite* counterterm to restore gauge invariance,

$$\tfrac{1}{2}A_-(\partial_+)^2\frac{1}{-\frac{1}{2}\Box}A_- + \tfrac{1}{2}A_+(\partial_-)^2\frac{1}{-\frac{1}{2}\Box}A_+ - A_+A_- = F\frac{1}{\Box}F$$

(after integration by parts).

This same calculation also gives the "axial anomaly": Consider an axial vector gauge field B that couples to the current $\pm\bar{\psi}_\alpha\psi_\alpha$ (not summed), in addition to A's coupling to $\bar{\psi}_\alpha\psi_\alpha$. (In D=2, $W_a = \epsilon_a{}^b V_b \Rightarrow W_\pm = \pm V_\pm$.) The contribution to the 1-loop effective action with one of each vector externally is, after including a counterterm to preserve A gauge invariance (and therefore break B gauge invariance),

$$-B_-(\partial_+)^2\frac{1}{-\frac{1}{2}\Box}A_- + B_+(\partial_-)^2\frac{1}{-\frac{1}{2}\Box}A_+ + B_-A_+ - B_+A_- = -(\partial\cdot B)\frac{1}{-\frac{1}{2}\Box}F$$

The anomaly is the breaking of B gauge invariance,

$$\delta B = -\partial\lambda \quad \Rightarrow \quad \delta\Gamma = \int \lambda\partial\cdot\frac{\delta\Gamma}{\delta B} = -2\int \lambda F$$

An anomaly is by definition a quantum effect: As we have seen from the 2D axial anomaly, it is related to a divergence that violates naive classical arguments, since the regulator itself violates the symmetry. In the axial case there is no actual divergent term in the effective action, but a finite term results from a ϵ/ϵ type of cancelation. Dimensional analysis immediately reveals that the propagator correction is the only graph in D=2 that can contribute such a term from the fermion loop.

(Fermion propagators go as $1/p$, while the vertex is a constant: The electric charge has dimension in D≠4.)

The complete one-loop effective action for the vectors then follows directly from the complete anomaly for the axial current, and the vanishing of the anomaly for the polar current: By separating out the anomalous term in the effective action,

$$\Gamma = \int F\frac{1}{\Box}F + \Delta\Gamma; \qquad J = \frac{\delta\Gamma}{\delta A}, \quad \Delta J = \frac{\delta(\Delta\Gamma)}{\delta A}$$

$$\partial\cdot J = 0,\ \partial\times J = -2F \ \Rightarrow\ \partial\cdot(\Delta J) = \partial\times(\Delta J) = 0 \ \Rightarrow\ \Delta J = 0 \ \Rightarrow\ \Delta\Gamma = 0$$

(up to an irrelevant constant), where $\partial\times J = \epsilon^{ab}\partial_a J_b$ is the curl of the polar current, but also the divergence of the axial current. (There are some questions of boundary conditions in solving the divergence- and curl-free conditions as $\Delta J = 0$, but these are resolved by working in Euclidean momentum space.)

Similar remarks apply to external gravity: From a similar calculation, replacing the vector current with the energy-momentum tensor, we find

$$\partial_m T^{mn} = 0,\ \partial_m \epsilon^m{}_n T^{np} \sim \epsilon^{pm}\partial_m R \quad \Rightarrow \quad T^m{}_m \sim R,\ \Gamma \sim R\frac{1}{\Box}R$$

where R is the 2D curvature (which is just a scalar, as the vector field strength is a pseudoscalar). While in the vector case the finite local counterterm was chosen to preserve polar gauge invariance and thus violate axial, for the tensor case a term is chosen to preserve local conservation of energy-momentum and thus violate conformal invariance $T^m{}_m = 0$. (The above expressions are linearized, but the results can be generalized to fully nonlinear gravity.)

Exercise VIIIA8.1

Calculate the gravitational anomalies from a massless spinor loop in D=2, using the classical expressions (as follow from dimensional and Lorentz analysis)

$$T_{\pm\pm} \sim \tfrac{1}{2}\bar{\psi}_\alpha i\overset{\leftrightarrow}{\partial}_\pm \psi_\alpha, \qquad T_{+-} = 0$$

(If you work in terms of Γ you can define the perturbative field h_{ab} such that $\delta\Gamma/\delta h_{ab} = T^{ab}$.)

The simple form of the effective actions in the Schwinger model is a consequence of bosonization: Thus, including coupling to electromagnetism and gravity, the action for the massless spinor is equivalent to

$$L = -\tfrac{1}{4}\phi\Box\phi + (F+R)\phi$$

Integrating out the scalar generates the above effective actions *classically*.

Exercise VIIIA8.2

The above action is dual to the mass term of the Stückelberg action:

a Consider the first-order Lagrangian

$$L = -G^2 + G^a(mA_a + \partial_a \phi)$$

Eliminating the auxiliary field G_a by its field equation yields the usual mass term for the Stückelberg model. Show that if we vary ϕ instead and solve the resulting constraint on G, we obtain (the nongravitational part of) the previous action.

b Generalize this construction to D=4, where the field dual to the Stückelberg scalar is now an antisymmetric tensor gauge field. (See exercise IIB2.1.)

REFERENCES

1 A.A. Slavnov, *Theor. Math. Phys.* **10** (1972) 99;
J.C. Taylor, *Nucl. Phys.* **B33** (1971) 436:
graphical form of unitarity for nonabelian gauge theories (later simplified as BRST).

2 G. 't Hooft, *Nucl. Phys.* **B33** (1971) 173, G. 't Hooft and M. Veltman, *Nucl. Phys.* **B44** (1972) 189, **50** (1972) 318;
B.W. Lee and J. Zinn-Justin, *Phys. Rev.* **D5** (1972) 3121, 3137, 3135, **7** (1973) 1049:
renormalization proof for nonabelian gauge theories.

3 V.S. Vanyashin and M.V. Terent'ev, *Sov. Phys. JETP* **21** (1965) 375:
first 1-loop gluon coupling renormalization, but neglecting ghosts.

4 I.B. Khriplovich, *Sov. J. Nucl. Phys.* **10** (1970) 235:
first complete 1-loop gluon coupling renormalization.

5 G. 't Hooft, unpublished comment after K. Symanzik's talk at *Colloquium on renormalization of Yang-Mills fields and applications to particle physics*, Marseille, June 19 -23, 1972;
H.D. Politzer, *Phys. Rev. Lett.* **30** (1973) 1346;
D.J. Gross and F. Wilczek, *Phys. Rev. Lett.* **30** (1973) 1343:
asymptotic freedom.

6 R.J. Hughes, *Phys. Lett.* **97B** (1980) 246:
analyzed 1-loop gluon propagator corrections in terms of spinless + spin contributions.

7 T.L. Curtright, *Phys. Lett.* **102B** (1981) 17:
extended this analysis to higher spins, supersymmetry, and massless sectors of reduced strings.

8 Z. Bern, G. Chalmers, L. Dixon, and D.A. Kosower, hep-ph/9312333, *Phys. Rev. Lett.* **72** (1994) 2134;
G.D. Mahlon, hep-ph/9312276, *Phys. Rev.* **D49** (1994) 4438:
generalization of Parke-Taylor amplitudes to one loop.

9 Z. Bern, L. Dixon, D.A. Kosower, hep-ph/9602280, *Ann. Rev. Nucl. Part. Sci.* **46** (1996) 109:
review of modern methods for one-loop graphs.

10 H. Georgi, H. Quinn, and S. Weinberg, *Phys. Rev. Lett.* **33** (1974) 451:
test of running couplings in SU(5) GUT.

11 S. Dimopoulos, S. Raby, and F. Wilczek, *Phys. Rev.* **D24** (1981) 1681:
test of running couplings in supersymmetric SU(5) GUT.

12 P. Langacker and N. Polonsky, *Phys. Rev.* **D47** (1993) 4028:
a review of accurate tests of supersymmetric and nonsupersymmetric GUTs, emphasizing the improvement from supersymmetry.

13 A. Parkes and P. West, *Phys. Lett.* **138B** (1984) 99;
D.R.T. Jones and L. Mezincescu, *Phys. Lett.* **138B** (1984) 293:
one-loop finiteness implies two-loop finiteness.

14 D.I. Kazakov, *Phys. Lett.* **179B** (1986) 352;
A.V. Ermushev, D.I. Kazakov, and O.V. Tarasov, *Nucl. Phys.* **B281** (1987) 72;
O. Piguet and K. Sibold, *Phys. Lett.* **177B** (1986) 373:
one-loop finiteness implies all-loop finiteness.

15 D.R.T. Jones, L. Mezincescu, and Y.-P. Yao, *Phys. Lett.* **148B** (1984) 317:
one-loop finite theories with soft breaking.

16 I. Jack and D.R.T. Jones, hep-ph/9405233, *Phys. Lett.* **333B** (1994) 372:
one-loop finiteness of softly-broken theories implies two-loop finiteness.

17 I. Jack, D.R.T. Jones and A. Pickering, hep-ph/9712542, *Phys. Lett.* **426B** (1998) 73,
I. Jack and D.R.T. Jones, hep-ph/9907255, *Phys. Lett.* **465B** (1999) 148:
one-loop finiteness of softly-broken theories implies all-loop finiteness.

18 S. Hamidi, J. Patera, and J.H. Schwarz, *Phys. Lett.* **141B** (1984) 349, S. Hamidi and J.H. Schwarz, *Phys. Lett.* **147B** (1984) 301;
D.R.T. Jones and S. Raby, *Phys. Lett.* **143B** (1984) 137, J.E. Bjorkman, D.R.T. Jones, and S. Raby, *Nucl. Phys.* **B259** (1985) 503;
J. León, J. Perez-Mercader, M. Quiros, and J. Ramirez-Mittelbrunn, *Phys. Lett.* **156B** (1985) 66;
X.-D. Jiang and X.-J. Zhou, *Phys. Lett.* **197B** (1987) 156;
D. Kapetanakis, M. Mondragón, and G. Zoupanos, hep-ph/9210218, *Z. Phys. C* **60** (1993) 181;
D.I.Kazakov, M.Yu.Kalmykov, I.N.Kondrashuk, and A.V.Gladyshev, hep-ph/9511419, *Nucl. Phys.* **B471** (1996) 389:
realistic finite models.

19 Mandelstam; Brink, Lindgren, and Nilsson; *loc. cit.* (VIB):
finiteness proof for N=4 using lightcone superfields.

20 Grisaru and Siegel, *loc. cit.* (VIC, ref. 12):
nonrenormalization theorems for extended supersymmetry.

21 P. Howe, K.S. Stelle and P. West, *Phys. Lett.* **124B** (1983) 55:
finiteness proof for N=2 (+ matter) and N=4 using nonrenormalization theorems.

22 V.A. Novikov, M.A. Shifman, A.I. Vainshtein, and V.I. Zakharov, *Nucl. Phys.* **B229** (1983) 381:
finiteness proof for N$\geq$2 based on instantons.

23 P. Jordan, *Z. Phys.* **93** (1935) 464;
M. Born and N.S. Nagendra Nath, *Proc. Ind. Acad. Sci.* **3** (1936) 318;
A. Sokolow, *Phys. Z. Sowj.* **12** (1937) 148;
S. Tomonaga, *Prog. Theo. Phys.* **5** (1950) 544:
early attempts at bosonization.

24 W. Thirring, *Ann. Phys.* **3** (1958) 91.

25 T.H.R. Skyrme, *Proc. Roy. Soc.* **A262** (1961) 237;
D. Mattis and E. Lieb, *J. Math. Phys.* **6** (1965) 304;

B. Klaiber, The Thirring model, in *Lectures in theoretical physics*, eds. A.O. Barut and W.E. Brittin (Gordon and Breach, 1968) v. X-A, p. 141;
R.F. Streater and I.F. Wilde, *Nucl. Phys.* **B24** (1970) 561;
J. Lowenstein and J. Swieca, *Ann. Phys.* **68** (1971) 172;
K. Bardakçi and M.B. Halpern, *Phys. Rev.* **D3** (1971) 2493;
G. Dell'Antonio, Y. Frishman, and D. Zwanziger, *Phys. Rev.* **D6** (1972) 988;
A. Casher, J. Kogut, and L. Susskind, *Phys. Rev. Lett.* **31** (1973) 792, *Phys. Rev.* **D10** (1974) 732;
A. Luther and I. Peschel, *Phys. Rev.* **B9** (1974) 2911;
A. Luther and V. Emery, *Phys. Rev. Lett.* **33** (1974) 598;
S. Coleman, *Phys. Rev.* **D11** (1975) 2088;
B. Schroer, *Phys. Rep.* **23** (1976) 314;
S. Mandelstam, *Phys. Rev.* **D11** (1975) 3026;
J. Kogut and L. Susskind, *Phys. Rev.* **D11** (1975) 3594:
bosonization (mostly for sine-Gordon $\leftrightarrow$ Thirring).

26 J. Schwinger, *Phys. Rev.* **128** (1962) 2425.

B. LOW ENERGY

In general, the only loop corrections that can be evaluated exactly in terms of elementary functions are the one-loop propagator corrections. However, limiting forms of vertex corrections, for various low- or high-energy limits, explicitly yield the most important pieces for certain applications.

1. JWKB

Some low-energy contributions to the effective action can be obtained by various quantum mechanical JWKB approximations. This involves an expansion of the external field about its vacuum value in spacetime derivatives (momenta). Such an expansion makes sense if this field is massless, since then small spatial momentum means also small energy, in the relativistic sense. (Otherwise one needs to expand nonrelativistically, about $\vec{p} = 0$ but $E = m$. Such treatments were considered in subsection IIB5, and will be applied to loops in subsection VIIIB6.) It also can be useful when the mass of the external field is small compared to the mass scale relevant to the interactions, such as for chiral symmetry breaking in the low-energy description of light mesons (subsection IVA4).

On the other hand, the fields we are integrating out must be massive, with a mass greater than the energy we want to investigate: Otherwise, the internal particles would show up as poles (and cuts) in the amplitudes, where Taylor expansion in momenta would be a poor approximation. The basic principle for analyzing the behavior of such a theory in a certain energy range is thus to first find contributions to the effective action where: (1) only particles with masses of lower energy appear on external (background) lines, and (2) only particles with masses of higher energy appear on internal (quantum) lines. These contributions are approximated by Taylor expansion to finite order in external momenta, yielding a local effective action. We could then consider finishing the functional integration by integrating out the lighter particles on internal lines: However, in this approximation it would be inaccurate to consider such particles in loops, since there they would include energies above the approximation scale. Thus, the effective action obtained by integrating out just the heavier fields is useful only when the lighter fields are treated classically. We apply the same approximation scheme to the classical action: Eliminate the heavier fields by their classical equations of motion, and Taylor expand their propagators in momenta to the desired order to get a local result.

In subsection VIIB2 we saw the simplest example, the effective potential: In that case the constant background scalar field acted as just a correction to the mass. We now consider more complicated cases, where spin and gauge invariance play roles for the internal or external fields. In particular, adding coordinate dependence to the background fields means we need to consider more general propagators for quadratic kinetic operators, such as harmonic oscillators.

We saw in subsection VIB1 the most general relativistic particle action for a scalar in external fields that was quadratic in x and $\dot{x}$. We now consider such actions in more detail: They are the most general ones for which we can derive one-loop results to all orders in the external fields (i.e., without performing the JWKB expansion beyond the first quantum correction, which requires Taylor expanding the exponential in terms that are beyond quadratic, thus expanding in the number of external fields).

Without loss of generality, we can consider Lagrangians that are homogeneous of second order in x and $\dot{x}$: Terms linear in $\dot{x}$ are boundary (in τ) terms, and were already eliminated by a gauge transformation (radial gauge). Terms linear in x can be removed by a translation, in the presence of an x^2 term (which is needed to bound an x term in the potential). (Both these kinds of terms can be restored trivially at the end.) A constant term is also trivial, giving a contribution to the classical action that is just that times T (after integration $\int_0^T d\tau$), and can be treated separately. (It doesn't contribute to the equations of motion.) The remaining contribution to the mechanics action is then of the form (as usual, in the gauge $v = 1$)

$$S = \int_0^T d\tau\ \tfrac{1}{2}[-\dot{x}^2 + x\mathcal{A}\dot{x} + x\mathcal{B}x] \quad \Rightarrow \quad \ddot{x} + \mathcal{A}\dot{x} + \mathcal{B}x = 0$$

$$\Rightarrow \quad S = \int_0^T d\tau\ \tfrac{1}{2}(-\dot{x}^2 - x\ddot{x}) = -\tfrac{1}{2}(x\dot{x})|_0^T$$

where $\mathcal{A}$ is an antisymmetric matrix and $\mathcal{B}$ symmetric. The steps to this contribution to the one-loop effective field action are then:

(1) Solve the equations of motion, which are homogeneous second-order differential equations.

(2) Change variables from the two parameters used for each x to $x(0)$ and $x(T)$. (Second-order differential equations require two initial conditions, or one initial and one final.)

(3) Find $S(x(0), x(T))$, including separately the contribution from the constant term in the Lagrangian.

(4) Find the propagator for "time" T, including the e^{-iS} and the van Vleck determinant. (See exercise VA2.4.)

(5) Integrate the propagator over T to find Γ. (See subsection VIIB2.)

For example, consider in QED the contribution to Γ from a fermion loop. If we are interested in only the properties of photons, then this gives the entire contribution to the functional integral from integrating out the fermions: This contribution, plus the classical (free) Maxwell action, gives a nonlocal "classical" action of self-interacting photons, which can itself be quantized to give the exact QED result for external photons. Although this one-loop effective action is too difficult to calculate exactly, the first-quantized JWKB approximation can give an accurate description at energies small compared to the electron mass. Note that we are simultaneously approximating to the first quantum correction in JWKB expansions of both the field (second-quantized) type (one-loop) and the mechanics (first-quantized) type.

The mechanics action for a massive particle in a constant external electromagnetic field strength (the lowest nontrivial order, but also the highest that keeps the action quadratic), in the radial gauge for the background field and affine parametrization of the worldline, is (see subsection VIB1)

$$S = \int d\tau \; \tfrac{1}{2}(-\dot{x}^2 + x^a F_{ab}\dot{x}^b + M^2)$$

To include spin, we identify (see subsection VIIIA3)

$$M^2 = m^2 - iS^{ab}F_{ab}$$

Since the only appearance of spin operators in the calculation of the propagator (denominator) is this constant matrix, it commutes with everything, so we can treat it as a number till the last step. The equation of motion

$$\ddot{x} + F\dot{x} = 0$$

is easily solved in matrix notation. (Hint: Solve for $\dot{x}$ first.) Finding $x_i = x(0)$ and $x_f = x(T)$ in terms of our integration parameters and inverting, then expressing $\dot{x}(0)$ and $\dot{x}(T)$ in terms of x_i and x_f (and T and F), and making use of the antisymmetry of F, the result is

$$S = -\tfrac{1}{4}(x_f - x_i)F \; coth(\tfrac{FT}{2})(x_f - x_i) - \tfrac{1}{2}x_f F x_i + \tfrac{1}{2}M^2 T$$

The propagator is then given by (see subsections VA2 and VIIB2)

$$\langle x_f | e^{-iTH} | x_i \rangle = \sqrt{det \, \frac{\partial^2(-iS)}{\partial x_f \partial x_i}} \, e^{-iS}$$

Plugging in, and then Wick rotating $T \to -iT$, we find for the propagator with ends tied together

$$\langle x|e^{-TH}|x\rangle = \sqrt{det\,\frac{iF}{1-e^{-iFT}}}\,e^{-M^2T/2}$$

Finally, the contribution to the effective action is (see subsection VIIB2)

$$\Gamma = -c\int dx\int_0^\infty \frac{dT}{T}\,tr\left(\sqrt{det\,\frac{iF}{1-e^{-iFT}}}\,e^{-M^2T/2} - \sqrt{det\,\frac{I}{T}}\,e^{-m^2T/2}\right)$$

$$= -c\int dx\int_0^\infty dT\; T^{-D/2-1}e^{-m^2T/2}\left[\sqrt{det\,\frac{iFT}{1-e^{-iFT}}}\;tr(e^{iS\cdot FT/2}) - tr(I)\right]$$

where $c = -\frac{1}{2}$ for fermions, for statistics and squaring the propagator. (The "det" is for the vector indices on $F_a{}^b$, the "tr" is for the spin indices from powers of S^{ab} in "$S\cdot F$".)

Exercise VIIIB1.1

Explicitly evaluate the determinant and trace for D=2.

Exercise VIIIB1.2

Expand Γ in F and show the resulting F^2 terms agree with those obtained in subsection VIIIA2-3.

Exercise VIIIB1.3

Consider the quadratic action

$$S = \int d\tau\;\tfrac{1}{2}[-\dot{x}^2 + x(a-a^T)\dot{x} - xaa^Tx]$$

where the matrix a commutes with its transpose ($[a,a^T]=0$). Solve the field equations for $S(x_i,x_f;T)$. Find $\langle x|e^{-TH}|x\rangle$.

2. Axial anomaly

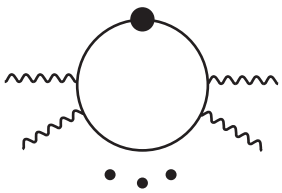
The axial anomaly comes from a finite graph, as we have already seen in subsection VIIIA8 for the case D=2. However, the naive manipulations that would show the graph to preserve gauge invariance involve evaluating the finite difference between divergent graphs, each of which needs regularization. Although in some cases the graph can be evaluated explicitly, and then shown to be anomalous, it is generally easier, and more instructive, to analyze the anomaly by itself.

The axial anomaly is associated with the use of ϵ tensors. In renormalizable theories in D=4, these occur only through γ_{-1}'s for spinors. (In nonrenormalizable theories, or in D=2, ϵ tensors can occur in scalar theories. There is also the term

$\epsilon^{abcd}F_{ab}F_{cd}$, which is a total divergence, and has no effect in perturbation theory.) In general even dimensions, the massless kinetic term for a spinor is invariant under transformations generated by γ_{-1}, but the mass term is not. Chiral symmetry is thus related to masslessness; this is also true for conformal invariance, so it's not surprising that quantum corrections can break both. (In fact, in supersymmetric theories conformal symmetry is related to a particular chiral symmetry by supersymmetry, so breaking of one requires breaking of the other if supersymmetry is to be preserved.)

Dimensional regularization manifestly preserves neither conformal nor chiral invariance; no regularization does. The existence of these anomalies proves the impossibility of such a regularization. Furthermore, dimensional reduction has difficulty dealing with γ_{-1}; it even has inconsistencies in the presence of axial anomalies. On the other hand, Pauli-Villars regularization is especially convenient for dealing with axial anomalies because it regularizes by introducing masses. Thus, it breaks chiral symmetry explicitly but softly, conveniently parametrizing the breaking by mass parameters. We therefore will use Pauli-Villars regularization for the single purpose of evaluating the axial anomaly.

The basic idea of Pauli-Villars regularization is to include massive "ghost" fields which would cancel graphs from physical fields if they had the same mass. But the masses of the ghosts are used as regulators; after subtracting local divergences, the regulator mass is taken to infinity. In our case, as we'll see by explicit evaluation, the anomaly itself is finite, so no subtraction is necessary.



The graph whose anomaly we want to evaluate is a one-loop 1PI graph with external vectors and a massless internal spinor. Of the vectors, all but one is a "polar" vector, coupling to $\bar{\psi}\gamma_a\psi$, while the last is an "axial" vector, coupling to $\bar{\psi}\gamma_{-1}\gamma_a\psi$. These are the currents associated with the symmetries $\psi' = e^{i\theta}\psi$ and $\psi' = e^{\theta\gamma_{-1}}\psi$ $(\gamma_{-1}^2 = -\frac{1}{2})$. We add to this graph a similar one, but with a massive spinor, and give the second graph an overall relative minus sign. Since the mass breaks chiral invariance, we have explicitly broken the gauge invariance of the axial vector, while preserving those of the polar vectors. Note that this is a feature of the regularization: If a regularization existed that preserved chiral symmetry, then we could freely move the γ_{-1} around the graph from one vertex to the next using the usual naive anticommutation relations, thus moving also the anomaly from one vertex to the next (i.e., violating gauge invariance in any vector we choose).

Gauge invariance is represented by vanishing divergence of the corresponding current: At each vertex we have the coupling $\int A\cdot J$, with gauge invariance $\delta A = \partial\lambda$, implying $\partial\cdot J = 0$ by integration by parts, where J may be polar or axial depending on the vertex. These currents are conserved classically. We know they are also conserved quantum mechanically in the absence of γ_{-1}'s, since dimensional regularization and renormalization preserve the gauge invariance of the effective action. In graphical terms, taking the divergence at a vertex kills a propagator (since $\partial\cdot J$ is proportional to the field equations of the internal field), and this can be shown to lead to vanishing of the graph.

However, with the Pauli-Villars regulator, the classical conservation of the axial current is explicitly broken. The result is that the complete axial anomaly can be found by looking at just the contribution coming from this explicit classical violation of current conservation (inserted into the one-loop graph). (The classically vanishing contributions are actually nonvanishing because of the anomaly, but they cancel between the physical and regulator fields, precisely because the regularization allows the naive manipulations that justify dropping them.)

We therefore want to evaluate the anomaly

$$\partial_a\mathcal{J}^a(x) \equiv \partial_a\frac{\delta\Gamma}{\delta A_a(x)}$$

where we start with a term in the classical action $\int A\cdot J$, so classically $J = \delta S/\delta A$, and then evaluate its quantum correction by looking at $\mathcal{J} \equiv \delta\Gamma/\delta A$ in terms of the one-loop part of the effective action Γ. Classically, we find a contribution from only the regulator,

$$\partial\cdot J \equiv \partial\cdot(-\sqrt{2}i\bar{\psi}\gamma_{-1}\gamma\psi) = 2m\bar{\psi}\gamma_{-1}\psi$$

So, all we need to evaluate is a one-loop diagram with the axial vector coupling to the regulator replaced with a pseudoscalar coupling $\int\phi\bar{\psi}\gamma_{-1}\psi$, and look at the graphs with one external pseudoscalar and the rest polar vectors. Clearly this is the same as coupling the pseudoscalar to the propagator of a *bosonic* spinor regulator in an external vector field:

$$\partial\cdot\mathcal{J} = 2m\ tr\left(\gamma_{-1}\frac{1}{i\nabla\!\!\!\!/ + \frac{m}{\sqrt{2}}}\right) = 2m\ tr\left[\gamma_{-1}(-i\nabla\!\!\!\!/ + \tfrac{m}{\sqrt{2}})\frac{1}{\nabla\!\!\!\!/^2 + \frac{1}{2}m^2}\right]$$

$$= \sqrt{2}m^2\ tr\left(\gamma_{-1}\frac{1}{\nabla\!\!\!\!/^2 + \frac{1}{2}m^2}\right)$$

where the trace is in the γ-matrix space. In the limit $m \to \infty$, graphs with more external lines vanish more rapidly. On the other hand, we need at least D/2 factors of S^{ab} (D γ-matrices) to give a nonvanishing γ-matrix trace. Thus, the leading contribution will be, using $-2\nabla\!\!\!/^2 = \Box + iS^{ab}F_{ab}$ from subsection IIIC4,

$$\partial \cdot \mathcal{J} = \sqrt{2}m^2 \; tr\left[\gamma_{-1}\frac{1}{\frac{1}{2}(m^2-\Box_0)}\tfrac{1}{2}iS^{ab}F_{ab}\frac{1}{\frac{1}{2}(m^2-\Box_0)}\cdots\tfrac{1}{2}iS^{ab}F_{ab}\frac{1}{\frac{1}{2}(m^2-\Box_0)}\right]$$

with D/2+1 propagators, where $\Box_0 = (\partial_a)^2$.

Thus, the only Feynman diagram we actually need to evaluate is the one-loop 1PI diagram with external and internal scalars. The limit internal $m \to \infty$ is the same as the limit external $p \to 0$. (The result does not depend on the internal momentum, which is integrated over, nor the external mass, which would appear only in external propagators.) Thus, this is just an effective potential calculation. We therefore have the integral (see subsection VIIB1)

$$\int dk \frac{1}{[\frac{1}{2}(k^2+m^2)]^{D/2+1}} = \frac{1}{\Gamma(\frac{D}{2}+1)\frac{1}{2}m^2}$$

$$\Rightarrow \quad \partial \cdot \mathcal{J} = \frac{2\sqrt{2}}{(\frac{D}{2})!}\; tr[\gamma_{-1}(\tfrac{1}{2}iS^{ab}F_{ab})^{D/2}]$$

Exercise VIIIB2.1

Check this result by using the expression from subsection VIIIB1 for the propagator in a constant external electromagnetic field (strength).

To evaluate in arbitrary even D, we note that the normalization of γ_{-1} is such that we can choose

$$(\gamma_{-1})^2 = -\tfrac{1}{2} \quad \Rightarrow \quad \gamma_{-1} = (-i)^{D/2}2^{(D-1)/2}\gamma^0\gamma^1\cdots\gamma^{D-1}$$

$$tr(I) = 2^{D/2}, \quad \epsilon^{01\cdots D-1} = -1 \quad \Rightarrow \quad tr[\gamma_{-1}(\tfrac{1}{2}iS^{ab}F_{ab})^{D/2}] = \tfrac{1}{\sqrt{2}}(\tfrac{1}{2})^{D/2}\epsilon^{ab\cdots cd}F_{ab}...F_{cd}$$

$$\Rightarrow \quad \partial \cdot \mathcal{J} = 2\frac{1}{2^{D/2}(\frac{D}{2})!}\epsilon^{ab\cdots cd}F_{ab}...F_{cd}$$

Thus, for example, for the Schwinger model (D=2) we have

$$\partial \cdot \mathcal{J} = -2F$$

in agreement with subsection VIIIA8, while for D=4

$$\partial \cdot \mathcal{J} = \tfrac{1}{4}\epsilon^{abcd}F_{ab}F_{cd}$$

3. Anomaly cancelation

When the anomaly occurs in a current that couples to a gauge field, unitarity is destroyed, since gauge invariance implies current conservation. This is a potential problem, since axial vector couplings occur in the Standard Model. (Actually, they are "V−A": (vector)−(axial vector).) The only way to avoid this problem is to have an anomaly cancelation between the different spinors: The coefficient of the anomaly is given purely by group theory, as $tr(A\{B,C\})$, where A, B, C are the matrices representing the couplings of the three vectors to all spinors, and the anticommutator comes from Bose symmetrization (from the crossed and uncrossed graph in the S-matrix, or the single contribution multiplying commuting fields in the effective action). We therefore require this trace (which represents the sum over all spinors) to vanish. (See exercises IB5.3 and VC9.2d for an example of the calculation of this trace.) The representations in the Standard Model have been chosen so this cancelation occurs in each family.

We already know in terms of Dirac notation that axial anomalies appear only in the presence of γ_{-1}'s. An absence of γ_{-1}'s is equivalent in terms of Weyl notation to the use of a (pseudo)real representation for undotted Weyl spinors. For example, consider a real representation that is reducible to a smaller (by half) representation "$\mathcal{R}$" and its complex conjugate "$\bar{\mathcal{R}}$": Then we can complex conjugate the complex-conjugate representation to produce a dotted Weyl spinor that is the same representation as the undotted spinor. The undotted and dotted spinor can then be combined, as usual, to form a Dirac spinor, which transforms as the complex representation, without γ_{-1}'s, and thus the same goes for the coupling of the gauge vector:

$$\psi_{\mathcal{R}\alpha} \oplus \psi_{\bar{\mathcal{R}}\alpha} \to \psi_{\mathcal{R}\alpha} \oplus \bar{\psi}_{\mathcal{R}\dot{\alpha}} \to \Psi_{\mathcal{R}}$$

So, in Dirac notation we can see that such representations do not contribute to anomalies because of the absence of γ_{-1}'s. Similar remarks apply to general real or pseudoreal representations: We can take an arbitrary (pseudo)real representation and make a *Majorana* spinor, as

$$\psi_{\mathcal{R}\alpha} \to \psi_{\mathcal{R}\alpha} \oplus \bar{\psi}_{\mathcal{R}\dot{\alpha}} \to \Psi_{\mathcal{R}}$$

where now $\bar{\psi}_{\mathcal{R}\dot{\alpha}}$ is simply the complex conjugate of $\psi_{\mathcal{R}\alpha}$ since $\mathcal{R} = \bar{\mathcal{R}}$.

This cancelation also can be seen directly in terms of Weyl spinors: The (pseudo)-reality of the representation is charge conjugation invariance (which is equivalent to parity invariance for spin-1 couplings to spinors, since such couplings are always CP invariant). Anomaly cancelation is then a generalization of Furry's theorem (see subsection VIIA5). Real and pseudoreal representations of the generators (including complex + complex conjugate) are antisymmetric, up to a unitary transformation, since they are hermitian:

$$G^T = G^* = -UGU^{-1}$$

(so $\delta\psi = iG\psi$ preserves reality or pseudoreality). Thus

$$tr(A\{B,C\}) = tr(-A^T\{-B^T, -C^T\}) = -tr(A\{B,C\}) \quad \Rightarrow \quad tr(A\{B,C\}) = 0$$

In particular, any mass term (without Higgs) $\psi^{T\alpha}\psi_\alpha$ requires a real representation (so its variation yields $G + G^T = 0$); a pseudoreal representation won't work because it uses an antisymmetric metric which, when combined with $C_{\alpha\beta}$, makes $\psi^{T\alpha}\psi_\alpha$ vanish by symmetry (since ψ is anticommuting). A related way to see in Weyl (or Dirac) notation that real representations are nonanomalous is to use the same squared-propagator trick we used for the propagator correction in subsection VIIIA3 (or related complex action from subsection IIIC4), which resulted in simplified Feynman rules only for real representations: With those rules, there are no potentially divergent 3-point graphs other than those that already occur for scalars (as part of the covariantization of the propagator divergence).

The absence of γ_{-1}'s is a special case of parity invariance. However, even parity invariance is not enough to enforce cancelation of anomalies, since some parity invariant theories have axial gauge vectors, which couple to axial currents $\bar{\Psi}\gamma_{-1}\gamma_a\Psi$, and the appearance of these γ_{-1}'s can be sufficient to introduce anomalies. In these anomalous cases, even if there is a C, the charge conjugation argument above does not apply because the C following from the usual CP and the obvious P does not simply replace $A \to -A^T$, but is some other permutation of similar representations. Thus, in general P (and C) invariance is unrelated to anomaly cancelation: We can have one without the other. Having real representations (i.e., no γ_{-1}'s) is a special case of both.

Exercise VIIIB3.1

Consider chiral symmetry (as in subsection IVA4 or IVB1) for a single flavor — $U(1)_L \otimes U(1)_R$. Now gauge that symmetry:

a In Weyl spinor notation, write the action for massless Weyl spinors $\psi_{L\alpha}$, $\psi_{R\alpha}$ each coupled to their own gauge vector. Clearly there is one anomaly for 3 external A_{La}'s, due to ψ_L, and another for A_{Ra}, due to ψ_R, and no mixing. Now assume the left and right coupling constants are equal (so the anomalies are equal). Write the resulting symmetry transformations on all fields under CP, C, and P.

b Rewrite this theory in Dirac notation. Using P, find the combinations of A_L and A_R that are (polar) vector and axial vector. Relate the anomaly calculations in the two notations. Show that dropping the axial vector gives (massless) QED. Find the theory that results from dropping the vector instead: Give the gauge symmetry, and show it is anomalous, and explain the anomaly (vs. the cancelation of the anomaly in QED) in both Weyl and Dirac language.

c Generalize all the above results to $\mathrm{U(n)}_L \otimes \mathrm{U(n)}_R$. (Note that C will now include complex conjugation on the hermitian matrices for the vectors, so that P won't.)

The simplest way to prove anomalies cancel in the Standard Model is to use our previous results for GUTs (subsection IVB4): (1) One way is to consider the GUT gauge group SU(4)⊗SU(2)⊗SU(2). First, we note that $tr(G_i) = 0$ because the group is semisimple, so there are no mixed anomalies. Then we see that the SU(4) couplings are the usual "color"-type couplings, without γ_{-1}'s (i.e., $4 \oplus \bar{4}$), so it has no anomalies. On the other hand, SU(2) has only (pseudo)real representations, so neither SU(2) has anomalies. Thus, anomalies cancel in the SU(4)⊗SU(2)⊗SU(2) GUT. Finally, breaking to SU(3)⊗SU(2)⊗U(1) (which also spontaneously breaks parity) leaves an extra singlet per family, which decouples, showing the cancelation for the Standard Model.

(2) Another way is to start with SO(10), which is anomaly free for any representation of fermions:

$$tr(G_{ab}, \{G_{cd}, G_{ef}\}) = 0$$

simply because there is no combination of Kronecker δ's with the appropriate symmetry (and similarly for SO(N), except for N=2 or 6, where such a term can be produced with the ϵ tensor). Breaking to the Standard Model again drops just a singlet (as does breaking to SU(5), showing its anomaly cancelation; breaking to SU(4)⊗SU(2)⊗SU(2) drops nothing, again showing its cancelation). In general, proving anomaly cancelation requires (a) using such arguments about real representations, or (b) the absence of anomalies for certain groups (namely, only SU(N) for N>2, or U(1), can have anomalies), or (c) explicitly calculating the relevant traces.

4. $\pi^0 \to 2\gamma$

When an anomalous axial symmetry appears only as a global symmetry classically, unitarity is preserved, since no gauge field couples to that current. This can be a useful way to explain approximate global symmetries. The fact that the anomaly is always a total derivative (because of the ϵ tensor and the Bianchi identity for F) means that the global symmetry is not broken perturbatively. (However, when the external vectors are nonabelian, there can be contributions from field configurations like instantons: See subsection IIIC6.) In subsection IVA4, we saw that the neutral pion (π^0), the lightest hadron, could be considered as the pseudogoldstone boson of an axial U(1) symmetry. We also want to consider the pion as a bound state of a quark and antiquark: If we knew the wave function, we could write the coupling, and calculate directly the decay of the neutral pion into two photons ($\pi^0 \to 2\gamma$) via quark-antiquark annihilation, or at least find the leading low-quark-energy contribution from the δ-function part of the wave function (in the relative coordinates of the quark and antiquark), corresponding to the coupling to $\bar{\psi}\gamma_{-1}\psi$. (An expansion of the wave function in derivatives of the δ function would give coupling to currents containing derivatives.)

Lacking such detailed information, the best we can do is extend the nonlinear σ model approach, which is to look for the terms in the phenomenological Lagrangian (expressed in terms of composite meson fields, not fundamental quark fields) with fewest derivatives (i.e., those most important at low energy), applying the condition of (approximate) chiral symmetry. Specifically, the global axial symmetry $\pi' = \pi - 2\theta$, where $A' = A$ for the photon field, along with the electromagnetic gauge invariance for A, under which the neutral pion field is invariant, would suggest couplings of pion to photon involving only $\partial\pi$ and F.

However, the anomaly allows the existence of another term: Since by definition (from considering coupling to an unphysical axial gauge field) the anomaly is given from a *local* axial transformation, while the pion field transforms in a trivial way under this transformation, we can attribute the anomaly to the pion coupling as

$$\delta\pi = -2\theta, \quad \delta\Gamma = -\int \theta\partial\cdot\mathcal{J}, \quad \Gamma = \Gamma_0 + \Delta\Gamma, \quad \delta(\Delta\Gamma) = 0$$

$$\Rightarrow \quad \Gamma_0 = \int \tfrac{1}{2}\pi\partial\cdot\mathcal{J} = \int \pi\frac{1}{2^{D/2}(\frac{D}{2})!}\epsilon^{ab\cdots cd}F_{ab}...F_{cd}$$

Thus, in four dimensions we find the contribution

$$\Gamma_0 = \int \pi\tfrac{1}{8}\epsilon^{abcd}F_{ab}F_{cd}$$

Using the abelian form of the Chern-Simons form (subsection IIIC6), we also can write this as

$$\Gamma_0 = -\int \tfrac{1}{6}\epsilon^{abcd}(\partial_a\pi)B_{bcd}$$

(In the nonabelian case, we can neglect the surface term only if the vacuum value of π has already been subtracted.) Adding this term to those found previously (the π and A kinetic terms, as well as the quark terms that define the normalization of the π field through its coupling), the decay rate for $\pi^0 \to 2\gamma$ can be calculated (including the 2 relevant flavors of quarks, and 3 colors, using the values of their electromagnetic charges), and is found to agree closely with the experimental value.

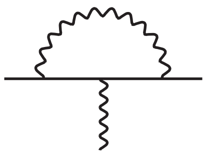
Exercise VIIIB4.1

What is Γ_0 in D=2? What is the interpretation of the pion field in terms of the fields of the Schwinger model (subsection VIIIA8)?

The global anomaly in the nonperturbative case can be applied to the strong interactions (QCD), although not as straightforwardly: Considering the external vectors to be gluons (so there is an implicit trace above over the group indices), Γ_0 gives a coupling of a neutral meson to a pseudoscalar glueball, as discussed in subsection IC4. If the vacuum gives a nontrivial value to $tr(\epsilon^{abcd}F_{ab}F_{cd})$ (as for instantons), this also leads to anomalous CP violation in the strong interactions.

5. Vertex

One-loop triangle graphs can't be evaluated in terms of elementary functions. However, in QED the most important effects are at low energy. We therefore will evaluate the effective action in the quantum mechanical version of the JWKB expansion, as an expansion in derivatives. The resulting approximation to the effective action thus will be local, but include terms of higher dimension than the classical action, whose coefficients are therefore finite and unrenormalized: By dimensional analysis, this means their coefficients will have powers of the inverse electron mass, which can be considered as the expansion parameter. (See also subsection VIIB7, where a scalar 1-loop vertex divergence was evaluated.)



The propagator corrections have been found already in subsection VIIIA1; now we calculate the vertex correction. The integral is

$$\mathcal{A}_{a,3,QED} = \int dk\ \frac{\mathcal{N}_a}{\mathcal{D}}$$

$$\mathcal{N}_a = \gamma^b(\not{k} + \not{p}' + \tfrac{m}{\sqrt{2}})\gamma_a(\not{k} + \not{p} + \tfrac{m}{\sqrt{2}})\gamma_b, \qquad \mathcal{D} = \tfrac{1}{2}k^2 \tfrac{1}{2}[(k+p')^2 + m^2]\tfrac{1}{2}[(k+p)^2 + m^2]$$

Without loss of generality, we can drop terms that vanish by the free fermion field equations; this corresponds to canceling them by fermion field redefinitions. We then evaluate the numerator by applying the identities

$$p^2 = p'^2 = -m^2, \qquad q = p' - p \quad \Rightarrow \quad (p+p')^2 = -4m^2 - q^2$$

$$\not{v}\gamma\not{v} = \tfrac{1}{2}v^2\gamma - v\not{v}$$

as well as the identities of subsection VIC4 for $\gamma^b ... \gamma_b$, and the field equations $\not{p} = \frac{m}{\sqrt{2}}$ on the far right and $\not{p}' = \frac{m}{\sqrt{2}}$ on the far left, to obtain

$$\mathcal{N} = (\not{k} + \not{p})\gamma(\not{k} + \not{p}') + \tfrac{m^2}{2}\gamma + \tfrac{m}{\sqrt{2}}(2k + p + p')$$

$$(\not{k} + \not{p})\gamma(\not{k} + \not{p}') = (\not{k} + \not{p} + \not{p}')\gamma(\not{k} + \not{p} + \not{p}') - \not{p}\gamma\not{p} - \not{p}'\gamma\not{p}' - \not{p}'\gamma\not{p} - \not{k}\gamma\not{p} - \not{p}'\gamma\not{k}$$

$$= [\tfrac{1}{2}k^2 + k\cdot(p+p') - 2m^2 - \tfrac{1}{2}q^2]\gamma - (k+p+p')(\not{k} + 2\tfrac{m}{\sqrt{2}}) + m^2\gamma + \tfrac{m}{\sqrt{2}}(p+p') - \tfrac{m^2}{2}\gamma + \tfrac{m}{\sqrt{2}}k$$

$$\Rightarrow \quad \mathcal{N} = (\tfrac{1}{2}k^2\gamma - k\not{k}) + [k\cdot(p+p')\gamma - (p+p')\not{k} + \tfrac{m}{\sqrt{2}}k] - (m^2 + \tfrac{1}{2}q^2)\gamma$$

For the momentum integral we evaluate

$$\mathcal{A}_3(x, m^2, q^2) = \int dk\ \frac{e^{ik\cdot x}}{\mathcal{D}} = \int d^3\tau\ \lambda^{-D/2} e^{-E}$$

$$E = \tfrac{1}{2}\tfrac{1}{\lambda}x^2 + ix\cdot\tfrac{1}{2}[(\alpha_1+\alpha_2)(p+p') + (\alpha_1-\alpha_2)q] + \tfrac{1}{2}\lambda[(\alpha_1+\alpha_2)^2 m^2 + \alpha_1\alpha_2 q^2]$$

again on the fermion mass shell. We also have

$$\int d^3\tau = \int_0^\infty d\lambda\ \lambda^2 \int_0^1 d^3\alpha\ \delta\left(1 - \sum \alpha\right) = \int_0^\infty d\lambda\ \lambda^2 \int_0^1 d\alpha_1 \int_0^{1-\alpha_1} d\alpha_2$$

using, e.g., the definitions

$$\int_a^b dx\ \delta(x)f(x) = \theta(-a)\theta(b)f(0), \qquad \int_{-\infty}^\infty dx\ \theta(x-a)\theta(b-x)f(x) = \int_a^b dx\ f(x)$$

As for the fermion propagator, we clearly separate UV and IR divergent integrals by the changes of variables

$$\alpha = \alpha_1 + \alpha_2, \qquad \beta = \alpha_1 - \alpha_2 \quad \Rightarrow \quad \int d^3\tau = \int_0^\infty d\lambda\ \lambda^2 \int_0^1 d\alpha\ \tfrac{1}{2}\int_{-\alpha}^{\alpha} d\beta$$

followed by

$$\lambda \to \frac{\lambda}{\alpha^2}, \qquad \beta \to \alpha\beta \quad \Rightarrow \quad \int d^3\tau \to \int_0^\infty d\lambda\ \lambda^2 \int_0^1 d\alpha\ \alpha^{-5}\ \tfrac{1}{2}\int_{-1}^1 d\beta$$

which modifies the integral to

$$\mathcal{A}_3 = \int_0^\infty d\lambda\ \lambda^\epsilon \int_0^1 d\alpha\ \alpha^{-1-2\epsilon}\ \tfrac{1}{2}\int_{-1}^1 d\beta\ e^{-E}$$

$$E = \tfrac{1}{2}\{\tfrac{1}{\lambda}\alpha^2 x^2 + i\alpha x\cdot(p+p'+\beta q) + \lambda[m^2 + \tfrac{1}{4}(1-\beta^2)q^2]\}$$

We now expand to $\mathcal{O}(x^2)$ and $\mathcal{O}(q^2)$. The β integral is then trivial (the integrand becomes quadratic in β), the λ integral gives the usual, and the α integral is similar to the case of the fermion propagator. The result is

$$\mathcal{A}_3 \approx -\tfrac{1}{2}\Gamma(1+\epsilon)(\tfrac{1}{2}m^2)^{-\epsilon}\left\{-\tfrac{1}{6}\frac{q^2}{m^2} + \left(1-\tfrac{1}{6}\frac{q^2}{m^2}\right)\left[\frac{1}{\epsilon_{IR}} + ix\cdot(p+p')\right.\right.$$
$$\left.\left. + \tfrac{1}{8}(x\cdot(p+p'))^2 + \tfrac{1}{4}m^2x^2\right] + \tfrac{1}{24}(x\cdot q)^2 + \frac{1}{\epsilon_{UV}}\tfrac{1}{4}m^2x^2\right\}$$

This leads to the expression for the vertex correction

$$\mathcal{A}_{3,QED} \approx \Gamma(1+\epsilon)(\tfrac{1}{2}m^2)^{-\epsilon}\left\{\left[\left(1+\tfrac{1}{3}\frac{q^2}{m^2}\right)\frac{1}{\epsilon_{IR}} + \tfrac{1}{2}\frac{1}{\epsilon_{UV}} + \tfrac{7}{2} + \tfrac{1}{12}\frac{q^2}{m^2}\right]\gamma\right.$$
$$\left. + \tfrac{1}{4}\left(1-\tfrac{1}{6}\frac{q^2}{m^2}\right)\frac{p+p'}{m/\sqrt{2}}\right\}$$

where we have used

$$\delta_a^a = D$$

in evaluating the contribution from the k^2 term. (Remember that all algebra from indices on the fields is done in 4 dimensions, while all algebra from indices on momenta is done in D dimensions. Since the two parts of the calculation are usually done separately, this should cause no confusion; however, the difference in evaluating δ_a^a is the main thing to watch.) Using the on-shell identity

$$4\tfrac{m}{\sqrt{2}}\gamma = \{\not p + \not p', \gamma\} + [\not q, \gamma] = -(p+p') + [\not q, \gamma]$$

we can rewrite this as (again keeping only $\mathcal{O}(q^2)$)

$$\mathcal{A}_{3,QED} \approx \Gamma(1+\epsilon)(\tfrac{1}{2}m^2)^{-\epsilon}\left\{\left[\left(\frac{1}{\epsilon_{IR}} + \tfrac{1}{2}\frac{1}{\epsilon_{UV}} + \tfrac{5}{2}\right) + \left(\tfrac{1}{3}\frac{1}{\epsilon_{IR}} + \tfrac{1}{4}\right)\frac{q^2}{m^2}\right]\gamma + \tfrac{1}{4}\frac{[\not q, \gamma]}{m/\sqrt{2}}\right\}$$

The next step is to cancel the UV divergence by adding the counterterm for electron wave-function renormalization from subsection VIIIA1:

$$\mathcal{A}_{3,QED,r} = \mathcal{A}_{3,QED} + \delta\mathcal{A}_{3,QED} \approx \left[\tfrac{1}{3}\left(\frac{1}{\epsilon_{IR}} - ln\frac{m^2}{\mu^2}\right) + \tfrac{1}{4}\right]\frac{q^2}{m^2}\gamma + \tfrac{1}{4}\frac{[\not q, \gamma]}{m/\sqrt{2}}$$

Equivalently, we can take the $q = 0$ piece of $\mathcal{A}_{3,QED}$, and note that it combines with $\mathcal{A}_{2e}$ of subsection VIIIA1 to gauge-covariantize the term proportional to $\not p \to \not p + \not A$. (The unrenormalized effective action is thus automatically gauge invariant, as is the counterterm.) At this point we can see the anomalous magnetic moment: Combining the tree and 1-loop result (including coupling), and writing as spinless + magnetic moment contributions, we have

$$\gamma + e^2 \mathcal{A}_{3,QED,r} \approx \left\{ 1 + e^2 \left[\tfrac{1}{3} \left(\frac{1}{\epsilon_{IR}} - ln\, \frac{m^2}{\mu^2} \right) + \tfrac{1}{4} \right] \frac{q^2}{m^2} \right\} (-\tfrac{1}{4}) \frac{p+p'}{m/\sqrt{2}} + (1+e^2)\tfrac{1}{4} \frac{[\not{q}, \gamma]}{m/\sqrt{2}}$$

We can translate these 1-loop corrections into a contribution to the effective action as (with the usual -1 for the effective action)

$$\Gamma_{1,3,QED,r} = -\bar{\Psi}(-p')\mathcal{A}_{3,QED,r}\Psi(p) \cdot A(q)$$

We then note, again using the spinor (free) field equations to imply $(p+p') \cdot q = 0$, to $\mathcal{O}(q^2)$,

$$-4\tfrac{m}{\sqrt{2}} q^a \gamma^b q_{[a} A_{b]} \approx q^a (p+p')^b q_{[a} A_{b]} = q^2 (p+p') \cdot A$$

The low-energy part of the renormalized effective action exhibiting up to order q^2/m^2 corrections to the coupling is then, in gauge invariant form,

$$\Gamma_{0+1,2e,r} \approx \int dx\, \overline{\Psi} \Big\{ i\not{\partial} - \not{A} + \tfrac{m}{\sqrt{2}} - \frac{e^2}{2\sqrt{2}m} i S^{ab} F_{ab}$$

$$- \frac{e^2}{m^2} \left[\tfrac{1}{3} \left(\frac{1}{\epsilon_{IR}} - ln\, \frac{m^2}{\mu^2} \right) + \tfrac{1}{4} \right] \gamma^a (\partial^b F_{ab}) \Big\} \Psi$$

Exercise VIIIB5.1

Perform the supergraph version of this calculation: a massless Abelian vector multiplet coupled to a massive chiral scalar multiplet.

6. Nonrelativistic JWKB

As for other processes, the application of quantum field theory to bound states has two steps: (1) Calculate the (gauge-invariant) effective action; (2) find solutions to the field equations following from the effective action ("on-shell" states). For bound states such solutions are nonperturbative; however, their determination is easier for nonrelativistic systems, since we can ignore production and annihilation of additional nonrelativistic (massive) particles in the second step because their effect already has been included as small corrections to the effective action.

The Lamb shift is the (field theoretic) quantum contribution to the energy levels of the hydrogen atom, which is described accurately even at one loop. The relativistic solution is found by perturbing the relativistic effective action in derivatives about the nonrelativistic one, whose solutions are the usual exact ones of the nonrelativistic

Schrödinger equation. For atoms the electron speed p/m is of the order of α $(= 2\pi e^2)$, so the loop and derivative expansions are in the same small parameter.

The effective action is more conveniently calculated with manifestly relativistic methods, since the internal ("virtual") particles can be relativistic (especially those that contribute to the UV divergences). On the other hand, the solutions to the field equations are more conveniently calculated in a representation that takes better advantage of the nonrelativistic expansion, since the external particles are nonrelativistic. Therefore, the second step begins by performing a field redefinition that converts the manifestly Lorentz invariant effective action to a form recognizable as nonrelativistic field theory with low-energy relativistic and loop corrections. (Originally the Lamb shift was calculated without this transformation. Higher-order calculations then required use of the relativistic Bethe-Salpeter equation, which made collection of terms of a given order more difficult.) In subsection IIB5 we considered the generalized Foldy-Wouthuysen transformation and its application to minimal coupling; we now apply it to the nonminimal coupling introduced by loop corrections. (In the literature this step has been performed on the Feynman diagrams themselves; however, as usual we can save some effort by working directly with the effective action.) Here the nonminimal correction to the transformation is easy, since the nonminimal terms are already near the order to which we work.

We first perform some dimensional analysis, using the fact that the leading behavior is given by the usual nonrelativistic Schrödinger equation. Then the only parameters in units $\hbar = 1$ (but there is no c in the nonrelativistic theory with just Coulomb interaction) are the mass m and *speed* e^2, so (in the notation of subsection IIB5)

$$\pi^i \sim me^2, \qquad \pi^0 \sim me^4$$

(neglecting the rest mass contribution). It is then convenient to reorganize the expansion in $1/m$ to relate to the expansion in e^2: For example, we can identify the two by choice of units

$$\tfrac{1}{m} \sim e^2 \quad \Rightarrow \quad \pi^i \sim 1, \qquad \pi^0 \sim \tfrac{1}{m}$$

along with $c = 1$ (since we will include relativistic corrections).

The relativistic form of the Schrödinger equation is obtained by multiplying $2\gamma^0$ in front of the kinetic operator of the electron in a background electromagnetic field, as obtained from the effective action. Approximating the proton as infinitely massive (for which we can partially correct by using the reduced mass for the electron), we take the electric field as described by the usual static "scalar" potential, and drop the magnetic field along with the "vector" potential.

We therefore modify the expansion of subsection IIB5 by

(1) reorganizing the $1/m$ expansion according to our dimensional analysis,

(2) using only a static electric background, and

(3) working directly in terms of γ matrices: We can either plug in the Dirac case of the spin operators into the expressions of subsection IIB5, including the reality-restoring transformation of subsection IIB4,

$$S^{ab} \to -\tfrac{1}{2}\gamma^{[a}\gamma^{b]}, \qquad S^{-1a} \to -\tfrac{1}{\sqrt{2}}\gamma^a$$

or just expand the Dirac operator directly,

$$2\gamma^0(-\not{\pi} + \tfrac{m}{\sqrt{2}}) = \pi^0 - 2\gamma^0\gamma^i\pi^i + \sqrt{2}m\gamma^0$$

(and similarly for the loop correction terms).

From the reuslts of the previous subsection, we thus choose to order $1/m^4$

$$\mathcal{E}_{-1} = \sqrt{2}\gamma^0, \qquad \mathcal{O}_0 = 2\gamma^i p^i \gamma^0, \qquad \mathcal{E}_1 = m\pi^0$$

$$\mathcal{O}_3 = -m^2 e^2 \tfrac{1}{\sqrt{2}} i\gamma^i F^{0i}, \qquad \mathcal{E}_4 = m^2 e^2 \left[\tfrac{1}{3}\left(\frac{1}{\epsilon_{IR}} - ln\,\frac{m^2}{\mu^2}\right) + \tfrac{1}{4}\right]\partial^i F^{0i}$$

(others vanishing), where we have included explicit m dependence so that the coefficients $\mathcal{E}_n$ and $\mathcal{O}_n$ are of order m^0 according to our above dimensional analysis (so our expansion in m makes sense). Using

$$tanh\ x \approx x - \tfrac{1}{3}x^3$$

the relevant commutators from IIB5 are then, for the nonvanishing generators to this order

$$mG = -\tfrac{1}{2}\{[G, \varDelta\mathcal{E}] + \mathcal{L}_G coth(\mathcal{L}_G)\mathcal{O}\}\mathcal{E}_{-1}$$

$$\Rightarrow \quad \begin{aligned} G_1 &= -\tfrac{1}{2}\mathcal{O}_0\mathcal{E}_{-1} \\ G_3 &= -\tfrac{1}{2}([G_1, \mathcal{E}_1] + \tfrac{1}{3}[G_1, [G_1, \mathcal{O}_0]])\mathcal{E}_{-1} \\ G_4 &= -\tfrac{1}{2}\mathcal{O}_3\mathcal{E}_{-1} \end{aligned}$$

and for the transformed kinetic operator

$$\mathcal{F}' = \mathcal{E} + tanh(\tfrac{1}{2}\mathcal{L}_G)\mathcal{O}$$

$$\Rightarrow \quad \begin{aligned} \mathcal{F}'_1 &= \mathcal{E}_1 + \tfrac{1}{2}[G_1, \mathcal{O}_0] \\ \mathcal{F}'_3 &= \tfrac{1}{2}[G_3, \mathcal{O}_0] - \tfrac{1}{24}[G_1, [G_1, [G_1, \mathcal{O}_0]]] \\ \mathcal{F}'_4 &= \mathcal{E}_4 + \tfrac{1}{2}[G_1, \mathcal{O}_3] + \tfrac{1}{2}[G_4, \mathcal{O}_0] \end{aligned}$$

Remembering that $\mathcal{E}_{-1}$ commutes with even and anticommutes with odd, we have identities like

$$(\mathcal{L}_{G_1})^n \mathcal{O}_0 = (-1)^{n(n-1)/2}(\mathcal{O}_0)^{n+1}(\mathcal{E}_{-1})^n, \qquad (\mathcal{E}_{-1})^2 = 1$$

Substituting for G into $\mathcal{F}'$:

$$\begin{aligned}
\mathcal{F}_1' &= \mathcal{E}_1 + \tfrac{1}{2}(\mathcal{O}_0)^2\mathcal{E}_{-1} \\
\mathcal{F}_3' &= -\tfrac{1}{8}[\mathcal{O}_0,[\mathcal{O}_0,\mathcal{E}_1]] - \tfrac{1}{8}(\mathcal{O}_0)^4\mathcal{E}_{-1} \\
\mathcal{F}_4' &= \mathcal{E}_4 + \tfrac{1}{2}\{\mathcal{O}_0,\mathcal{O}_3\}\mathcal{E}_{-1}
\end{aligned}$$

The final result is, using

$$(\mathcal{O}_0)^2 = (p^i)^2, \qquad me^2[\mathcal{O}_0,\mathcal{E}_1] = -2\mathcal{O}_3\mathcal{E}_{-1}$$

and setting $\mathcal{E}_{-1} = -1$ on the right for positive energy,

$$\begin{aligned}
\mathcal{F}_1' &= m\pi^0 - \tfrac{1}{2}(p^i)^2 \\
\mathcal{F}_3' &= -\tfrac{1}{4}m[\tfrac{1}{2}(\partial^i F^{0i}) - iS^{ij}\{F^{0i},p^j\}] + \tfrac{1}{8}(p^i)^4 \\
\mathcal{F}_4' &= m^2e^2\left[\tfrac{1}{3}\left(\frac{1}{\epsilon_{IR}} - ln\,\frac{m^2}{\mu^2}\right)\partial^i F^{0i} + i\tfrac{1}{2}S^{ij}\{F^{0i},p^j\}\right]
\end{aligned}$$

As expected from dimensional analysis, $\mathcal{F}_1'$ is the nonrelativistic result, $\mathcal{F}_3'$ is the lowest-order relativistic correction, and $\mathcal{F}_4'$ is the lowest-order part of the one-loop correction. Putting it all together, to this order we have

$$\mathcal{F}' \approx \pi^0 - \left[\frac{(p^i)^2}{2m} - \frac{(p^i)^4}{8m^3}\right] - \tfrac{1}{8m^2}\left[1 - \tfrac{8}{3}e^2\left(\frac{1}{\epsilon_{IR}} - ln\,\frac{m^2}{\mu^2}\right)\right]\partial^i F^{0i} + \frac{1+2e^2}{4m^2}iS^{ij}\{F^{0i},p^j\}$$

Exercise VIIIB6.1

Find the additional terms in $\mathcal{F}'$ to this order when the electromagnetic field is arbitrary (magnetic field, time derivatives of background), assuming the same dimensional analyis for the background.

REFERENCES

1 W. Pauli and F. Villars, *Rev. Mod. Phys.* **21** (1949) 434.
2 J. Steinberger, *Phys. Rev.* **76** (1949) 1180;
Schwinger, *loc. cit.* (VB, ref. 2, second ref.):
triangle graph, for $\pi^0 \to 2\gamma$.
3 J.S. Bell and R. Jackiw, *Nuo. Cim.* **60A** (1969) 47:
identified axial anomaly.
4 S. Adler, *Phys. Rev.* **177** (1969) 2426:
explained axial anomaly.

5 S.L. Adler and W.A. Bardeen, *Phys. Rev.* **182** (1969) 1517:
showed axial anomaly is only at 1 loop.
6 H. Georgi and S.L. Glashow, *Phys. Rev.* **D6** (1973) 429:
anomaly cancelation for general GUTs.
7 H. Bethe, *Phys. Rev.* **72** (1947) 339:
calculation of Lamb shift.
8 H. Fukuda, Y. Miyamoto, and S. Tomonaga, *Prog. Theo. Phys.* **4** (1949) 47, 121;
N.M. Kroll and W.E. Lamb, Jr., *Phys. Rev.* **75** (1949) 388;
Y. Nambu, *Prog. Theo. Phys.* **4** (1949) 82;
J.B. French and V.F. Weisskopf, *Phys. Rev.* **75** (1949) 1240:
relativistic calculation of Lamb shift.
9 W. E. Caswell and G. P. Lepage, *Phys. Lett.* **167B** (1986) 437:
nonrelativistic effective actions.
10 L.S. Brown, *loc. cit.* (VA), p. 511;
A. Pineda and J. Soto, hep-ph/9711292, *Phys. Lett.* **420B** (1998) 391:
Lamb shift with only dimensional regularization.
11 *Quantum electrodynamics*, ed. T. Kinoshita (World Scientific, 1990):
higher-loop QED.

C. HIGH ENERGY

The sign of the one-loop correction to the gauge coupling is opposite in QCD to that of QED: The photon coupling is weak at "low" energies (actually, any observable energy, since the coupling runs so slowly), while the gluon coupling is weak at high energies (with respect to the hadronic mass scale). Thus, typically perturbation in loops is used to study high-energy behavior of QCD, while the low-energy behavior awaits the discovery of a general nonperturbative approach.

1. Conformal anomaly

Symmetries of the classical action that are violated at the quantum level are called "anomalous". There are two major sources for such "anomalies" in renormalizable quantum field theory: (1) There are anomalies associated with the totally antisymmetric matrix $\epsilon_{a_1 \ldots a_D}$, called "axial" (see subsections VIIIA8 and VIIIB2-4). When they occur, they are found in graphs with at least (D+2)/2 external lines. They are associated with graphs that have no divergences, yet require regularization. (2) The existence of divergences requires the introduction of a mass scale even in theories that are classically conformal. If anywhere, these show up at least in the most divergent graphs, the propagator corrections. Normally, both kinds of anomalies will first appear at one loop.

When anomalies are associated with global symmetries, they provide a natural way to explain approximate symmetries, in the sense of the perturbative approximation. However, when they occur in local symmetries, they destroy the gauge invariance needed to prove unitarity. The latter type of theory therefore must be avoided by applying the condition of anomaly cancelation in local symmetries.

We have already seen the appearance of the conformal anomaly in our renormalization of divergent loop graphs: The introduction of a renormalization mass scale breaks the scale invariance of a classically scale-invariant theory. The simplest example, and generally the most important, is the one-loop propagator correction. If we examine only high-energy behavior, then we can neglect masses from the classical action.

Using dimensional regularization, the generic effect on the effective action of the complete one-loop propagator correction is to modify the kinetic term of an arbitrary massless theory to

$$\tfrac{1}{2}\phi K \left(\frac{1}{g^2} + \beta_1 \; ln \frac{\Box}{\mu^2} \right) \phi$$

where ϕ is an arbitrary-spin field that we have normalized $\phi \to \phi/g$ for some appropriate coupling g (like the Yang-Mills coupling if ϕ is the Yang-Mills vector), K is the classical kinetic operator, μ is the renormalization mass scale, and β_1 is a constant determined by the one-loop calculation. As long as β_1 is nonvanishing (i.e., the theory is not finite) we can rewrite this as

$$\tfrac{1}{2}\phi K \beta_1 \ ln\frac{\Box}{M^2}\phi$$

where

$$M^2 = \mu^2 e^{-1/\beta_1 g^2}$$

is a renormalization-independent mass scale: Any physical measurement will observe g and μ in only this combination. A choice of different renormalization mass scale is equivalent to a finite renormalization of g^2, such that M is unchanged. In the case where g is dimensionless (the relevant one, since we are studying the conformal anomaly), the coupling constant has undergone dimensional transmutation, being replaced with a dimensionful constant.

Exercise VIIIC1.1

Show this is the case for massless (scalar) ϕ^3 theory in D=6 from the explicit one-loop correction.

If there is more than one coupling constant, things are more complicated, but the same phenomenon occurs: One dimensionless coupling is replaced with a mass. A particularly interesting case is pure Yang-Mills theory: Then we can write an important contribution to the effective action as

$$F\beta_1 \ ln\frac{\Box}{M^2}F$$

where F is now the complete nonabelian field strength, and $\Box$ is the square of the covariant derivative. Since this contribution by itself gives the complete 1-loop conformal anomaly, the rest of the 1-loop effective action is conformally invariant. (All its M dependence cancels.)

Note that the (one-loop) anomaly itself is local: If we perform an infinitesimal conformal transformation on the one-loop part of the effective action, this variation gives a local quantity. This is clear from the way this anomaly arose in dimensional regularization: If there were no infinities, there would be no anomaly, since the naive conformal invariance of the classical theory would be preserved at each step. However, to regularize the divergence we needed to continue the theory to arbitrary dimensions, and the theory is not conformal away from 4 dimensions. The scale variation of a 4D conformal action in 4-2ϵ dimensions is proportional to ϵ times that action, as follows from dimensional analysis; this scaling can be associated with the nonvanishing (engineering) dimension of the coupling away from D=4. (Usually, we write the coupling as $g\mu^\epsilon$, where g is dimensionless. The fields have engineering dimension independent of D, defined by the value in D=4: E.g., in $\nabla = \partial + A$, A has the same dimension as ∂.) However, the one-loop effective action is coupling independent; thus, when dimensionally regularized but unrenormalized, it's scale invariant. For example, in the propagator correction discussed above, we get a regularized term $-\frac{1}{\epsilon}\beta_1\phi K(\frac{1}{2}\Box)^{-\epsilon}\phi$, which is scale invariant but divergent. On the other hand, the counterterm added to make it finite is from the 4D conformal action, and thus is not scale invariant in D≠4; so the breaking of scale invariance can be associated entirely with the counterterm. (I.e., the anomaly coming from the renormalized, nonlocal effective action is equal to that coming from the infinite, local counterterm.) Since the counterterm is local, the anomaly is local. It's also finite, since it's proportional to ϵ (from the variation) times $1/\epsilon$ (from the divergent coefficient of the counterterm). In our propagator example, we have

$$-\tfrac{1}{\epsilon}\beta_1\phi K(\tfrac{1}{2}\Box)^{-\epsilon}\phi + \tfrac{1}{\epsilon}(\tfrac{1}{2}\mu^2)^{-\epsilon}\beta_1\phi K\phi \approx \beta_1\phi K\ ln\frac{\Box}{\mu^2}\phi$$

A similar situation occurs for the axial anomaly with Pauli-Villars regularization: After regularization, the anomaly comes entirely from the regulator graph, which is not only finite by power counting, but local in the infinite-mass limit because that is the zero-momentum (effective potential or JWKB) limit.

There is a physical significance to the sign of the constant β_1. (We saw some evidence of this already in our analysis of renormalons in section VIIC.) Instead of thinking in terms of the renormalization-independent mass scale M, we can treat g as an effective energy-dependent ("running") coupling,

$$\frac{1}{g^2(p^2)} = \frac{1}{g^2} + \beta_1\ ln\frac{p^2}{\mu^2}$$

In the case $\beta_1 > 0$ the coupling gets weaker at high energy ("asymptotic freedom"), while for $\beta_1 < 0$ the coupling gets stronger at high energy (until it reaches the Landau ghost). (For low energy the situation is generally more subtle, since we usually have complications from physical masses.) For QCD, this weakening of the coupling at high energy allows the separation of an amplitude into a nonperturbative low-energy piece (describing the observed particles, the bound-state hadrons), which is determined experimentally, and a perturbative high-energy piece (describing the non-asymptotic, fundamental "partons", gluons and quarks), which can be calculated. (This sometimes goes under the somewhat misleading name of "perturbative QCD".)

This strongly contrasts with QED, where the weakening of the coupling at low energy means both fundamental particles (photons, electrons, etc.) and bound states (positronium, atoms, etc.) can be treated perturbatively, and the only experimentally determined quantities are the values of masses and the electron charge (coupling at low-energy). Thus, in QED one in principle can calculate anything, while in QCD one is restricted to parts of certain amplitudes. (Various nonperturbative methods also have been developed for QCD, but so far they have successfully calculated only a few low-energy constants, as used in σ models, i.e., masses and low-energy couplings.) Although experimental verification of these results is sufficient to confirm the QCD description of hadrons, a practical description of hadronic cross sections at all energies would seem to require a string model that can incorporate behavior attributed to both strings and partons.

2. $e^+e^- \to$ hadrons

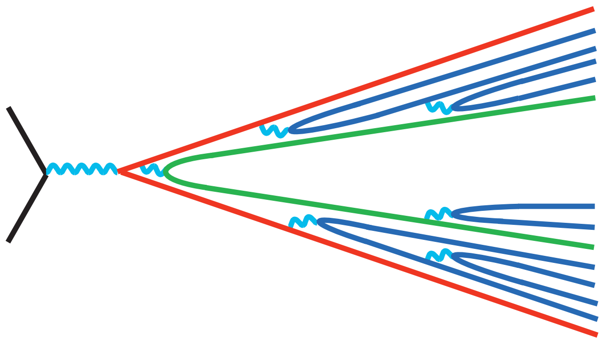
If quarks and gluons are confined, how can QCD be useful? QED is useful because the coupling is small: $e^2 \approx 1/861$ is the perturbation parameter in relativistic (quantum field theory, or 4D) calculations, $\alpha = 2\pi e^2 \approx 1/137$ in nonrelativistic (quantum mechanics, or 3D). Energy levels of the hydrogen atom can be calculated quite accurately, without the question of freely existing electrons and protons coming up. The speed of the bound electron is also α, another way to understand why pair creation/annihilation and other relativistic or multiparticle effects are small, and can be treated perturbatively.

Therefore, the real usefulness of a field theory depends not on how “physical” the choice of fields is, but how accurate the perturbation expansion is. “Nonperturbative” results may give some nice qualitative features, but they are ultimately useless unless they can be used as the basis of a new perturbation expansion. (Attempts at nonperturbative approaches to 4D quantum field theory continue, but so far the results are meager compared to perturbative results, or to nonperturbative results in quantum mechanics or 2D quantum field theory.)

The simplest application of QCD is to the production of hadrons by a photon created by the annihilation of an electron and a positron. The total cross section for such an event is given (according to the optical theorem) by the imaginary part of quark contributions to the photon propagator: Since hadrons are made up of partons (quarks and gluons), we assume a sum over hadrons can be written as a sum over partons. This assumption, that hadrons can be described by a resummation of the perturbation expansion, should be good at least at high energies, where the partons’ asymptotic freedom takes effect (and perhaps at lower energies by an appropriate extrapolation). To lowest order for the process under consideration this is a 1-loop graph, with a quark in the loop. If we compare this to the production of, e.g., muon-antimuon pairs (but not back to electron-positron pairs, because that includes the crossed diagram) by the same procedure, and we neglect masses (at high enough energies), then the only difference should be in the group theory: Hadron production should be greater by a factor of the number of colors times the sum over flavors of the square of the quark's electric charge:

$$R \equiv \frac{P(e^+e^- \to h)}{P(e^+e^- \to \mu^+\mu^-)} \approx N_c \sum_f q_f^2$$

Experimentally this relation is confirmed for $N_c = 3$, if the only flavors included in the sum are those with masses below the photon energy ($(2m_f)^2 < s$).



This result can be extended to the case where the momenta of hadrons are observed (not summed over): Although individual partons are not observed as asymptotic states, the dominant contribution to the cross section at high energies is given by the conversion of the quarks into hadrons by the creation from the vacuum of parton pairs with energies, and angular deviation from the partons created by the photon, smaller than experimental accuracy. We treat all partons as approximately massless, with respect to the energy scale of the photon. Thus, each parton created by the photon starts out initially as free, is then accompanied by parallel partons of small energy to form hadrons, and then these hadrons may further decay, but with a small angular spread with respect to the directions of each of the initial partons. Such collections of final-state hadrons are called "jets". For high-energy electron-positron annihilation, the dominant hadronic decay mode of this off-shell photon is thus into two jets. This experimental result is further verification of QCD, and in particular a jet is the most direct observation of a parton. Of course, even for asymptotic states the directness of experimental observations varies widely: For example, compare a photon or electron to a neutrino. A closer analogy is unstable particles: For example, the neutron is observed as a constituent of the nucleus (as quarks are constituents of hadrons), but eventually decays outside (as quarks "decay" into jets of hadrons).

A similar analysis can be applied to the creation of any electroweak boson by annihilation of a lepton with an antilepton.

Exercise VIIIC2.1

Find the corresponding process (particles) for positron-neutrino annihilation. Find the expected numerical value of both this and the above R in the Standard Model for energies well above the masses of all the fundamental particles.

3. Parton model

We have already seen that in quantum field theory coupling constants are usually energy-dependent. However, the dependence is only logarithmic, and thus can be treated as perturbative unless the relevant energy scale is within a few orders of magnitude of the mass scale that appears by dimensional transmutation. In QED, the value quoted for the electron charge is at the scale of the electron mass m. Using the result of subsection VIIIA2 (or VIIIA3) for the 1-loop propagator correction, we find (neglecting higher-loop corrections)

$$\frac{M_{QED}}{m} = e^{3/4e^2} = 2.837890(82) \times 10^{280} \quad \Rightarrow \quad M_{QED} = 1.450159(42) \times 10^{277} GeV$$

(where for fun we have included the 1-standard-deviation uncertainties for this 1-loop result as the figures in parentheses; the e in the exponent is the electron charge). Since the mass of the observable universe is of the order of 10^{80} GeV, and the Planck mass (beyond which a particle will gravitationally collapse from its Compton radius falling within its Schwarzschild radius) is of the order of "only" 10^{20} GeV, there is little worry of observing the QED Landau ghost, even if QED were correct to that scale.

On the other hand, the mass scale for QCD is (in the $\overline{\text{MS}}$ scheme)

$$M_{QCD} \approx .2\ GeV$$

(This result depends on renormalization scheme, and is also an effective mass in the sense that the usual experimental energy scale is among the quark masses, so the high-energy approximation of the renormalization group is inaccurate, and the full propagator correction with quark mass dependence should be used.) This indicates that perturbative QCD is inadequate to describe properties for which the energy of the quarks is low, such as hadron masses (although nonrelativistic quark models have had partial successes).

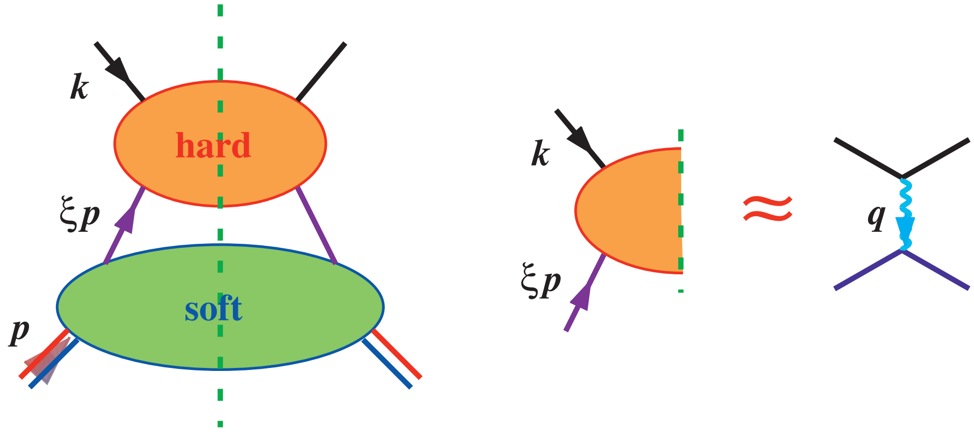
Exercise VIIIC3.1

Take masses into account in the simplest approximation: Treat particles as massless for energies above (twice) their mass, infinitely massive for energies below. Approximate the masses of Higgs and superpartners of Standard Model particles as about the mass of the Z boson. Then graph the strong coupling $1/g^2$ in the supersymmetric Standard Model (see subsection VIIIA4) as a function of the *ln* of the energy from the Grand Unification scale down to where it vanishes ($g = \infty$), M_{QCD}.

However, in certain processes a single "parton" (quark or gluon) in a hadron is given a high energy with respect to the other partons, usually a quark by electroweak interaction. In those cases, the "strong" (chromodynamic) interaction of that parton with the others in its original hadron is negligible: It has been liberated. The approach is then to factor the amplitude into a piece with the electroweak and high-energy ("hard") chromodynamic interactions of this parton, which can be calculated perturbatively, and the low-energy ("soft") chromodynamic part of the remaining partons, which is left as an unknown, to be experimentally determined. (Thus, the hard part is the easy part, while the soft part is the difficult part.) The predictive power is thus limited to the dependence of the amplitude on the energy of this parton, and on the particulars of the electroweak particles involved.

Another possible complication would be the effect of exciting many partons within a hadron, indirectly through the first parton's interactions with the rest: Then one would have several terms to sum in an amplitude, each with a different unknown soft factor, making the approach useless. Originally, it was thought that the high energy alone was enough to explain the parton acting as free once liberated from the hadron (based on "intuitive" arguments), but soon it was realized that this possibility depended totally on the high-energy behavior of the theory: It requires the decrease of the coupling with increasing energy, asymptotic freedom (or superrenormalizability, or finiteness with effective asymptotic freedom through the Higgs effect). Based on this property, one can show from the usual perturbation expansion that one soft factor (per each hadron with an excited parton) is sufficient as a leading approximation, a property known as "factorization". This feature is a consequence of the fact that the dominant contributions to Feynman graphs in this high-energy limit are those where the values of the momenta of some of the partons are those corresponding to their classical mechanics, as described in subsection VC8 and VIIA6.

This new approximation scheme is effectively a perturbation expansion in the inverse of the energy being channeled into this parton. One neglects terms that are smaller by such powers (including those from masses and renormalons), but incorporates logarithms through the renormalization group and other loop corrections to the hard factor. Since available energy scales are much nearer to M_{QCD} than to M_{QED} in QED, such an approximation scheme tends to break down around two loops, where the corrections compete with the neglected terms, ambiguities in renormalization schemes, and the relative size (convergence) of successive terms in the expansion. Although the accuracy of the predictions of this approach cannot compare numerically with those of QED, it is the only method to describe such processes that can lay claim to being a theory, and provides direct experimental evidence of the validity of QCD, both as a qualitative description of nature and as a valid perturbation scheme. (As in the previous subsection, we also have processes where all the partons appear only in intermediate states, or effectively so for final states in total cross sections via the optical theorem, so factorization is unnecessary.)



The most effective application of factorization is to "Deep(ly) Inelastic Scattering (DIS)". (An equivalent method for this process is the "operator product expansion", but unlike factorization there is no useful generalization of it to general processes.) In this process a high-energy photon (or intermediate vector boson) is exchanged between a lepton (usually an electron) and a quark. (This is the leading-electroweak-order interaction of a lepton with a hadron.) The quark and rest of the hadron do not interact again: Color singlets are obtained by the creation of soft partons from the vacuum, which split from their own singlets and eventually combine with the separated quark and hadron. For this process one calculates only the total cross section, at least as far as all the strongly interacting particles are concerned ("inclusive scattering") but again this can be generalized to the observation of jets ("exclusive scattering"). Applying the optical theorem, and ignoring the leptons, the leading contribution to this process is given by the tree graph for scattering of a vector boson off a quark, where the intermediate quark has a cut propagator. This is the perturbatively calculated hard part, which is later attached to the soft factor. Thus the hard part is the lepton-parton cross section, while the soft part is the "parton distribution", giving the probability of finding a parton in the hadron with a particular fraction ξ ($\geq 0, \leq 1$) of its momentum p. To leading order this fraction is determined by kinematics: Since the hadron and scattered parton are treated as on-shell and massless,

$$p^2 = (q + \xi p)^2 = 0 \quad \Rightarrow \quad \xi = x \equiv -\frac{q^2}{2q \cdot p}$$

so the "(Bjorken) scaling variable" x is a useful dimensionless parameter even when (at higher orders) $\xi \neq x$. The energy scale is set by the square of the momentum q of the vector boson.

There are several approximations used in this analysis, all of which can be treated as the beginnings of distinct perturbation expansions:

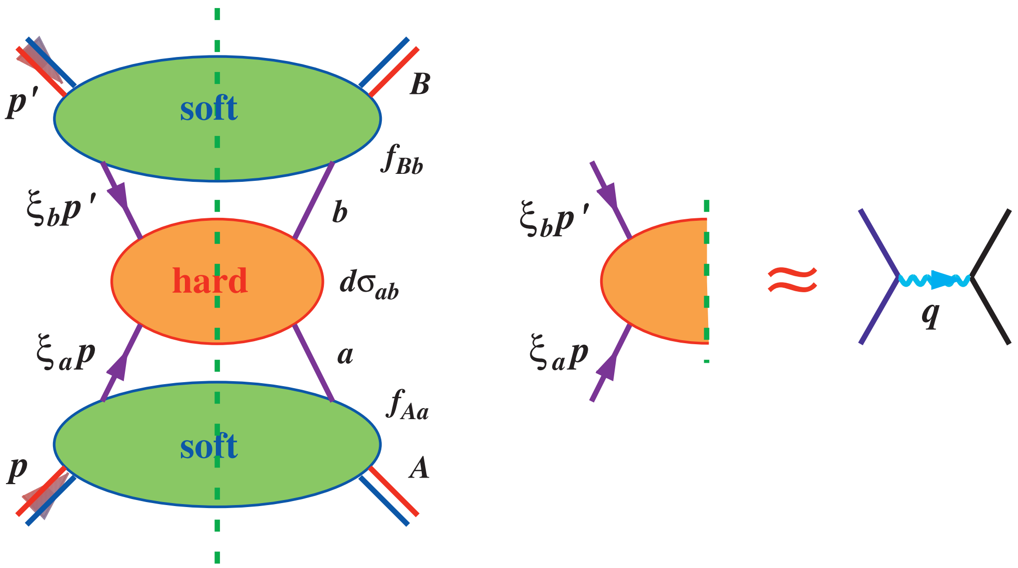
(1) The hard part is expanded in the usual (loop/coupling) perturbation expansion of field theory. The leading contribution is that of the naive (pre-QCD) parton model ("leading order"), where the quark that scatters off the photon is treated as free with respect to the strong interactions. One-loop corrections ("next-to-leading-order") introduce the running of the coupling associated with asymptotic freedom, which justifies the validity of the parton picture. This is usually the only perturbation expansion considered, because such corrections are logarithmic in the energy of the exchanged parton (rather than powers), and thus more important and easier to isolate from the data. Furthermore, by the usual renormalization group methods such logarithmic corrections can be reduced by careful choice of renormalization scale (μ^2 close to q^2 in $ln(q^2/\mu^2)$). Two-loop corrections lead to various ambiguities, and have not proven as useful yet. In particular, the β function is scheme-dependent past two loops, making the dependence on the separation between hard and soft harder to fix.

(2) In calculating the hard part "light" quarks are approximated as massless. One can rectify this by also perturbing in the masses, as a Taylor expansion in the square of each mass divided by the square of the vector boson's energy (m^2/q^2).

(3) In the explicit calculation the momentum ξp of the excited parton is assumed to be proportional to the momentum p of the initial hadron. In the rest frame of the initial hadron (which is massive in real life), this corresponds to the nonrelativistic approximation of motionless quarks; one quark is then set into relativistic motion by the photon, liberating it from the hadron. This approximation can be corrected by a JWKB expansion (expressed in operator language, the operator

product expansion), also known as an expansion in "twist" (effectively, the power of momentum transverse to p). However, this means a separate soft part for each term in the expansion: Since these are determined experimentally, such an expansion would lead to a loss of predictability. Thus generally (with few exceptions), parton model predictions are restricted to high enough energies (q^2) that such corrections can be neglected. In this sense, this approach is very similar to low-energy approaches to hadronic physics, e.g., nonlinear σ models: Useful results are obtained at lowest order for describing physics in a certain energy range, but outside that range the increasing loss of predictability, e.g., nonrenormalizability, makes the approach less and less applicable. (Another, related, similarity between this approach and nonlinear σ models is that both were originally described in the language of the operator product expansion, as applied to currents. However, this language was later replaced in both cases because of the difficulty of evaluating operator products of more than two currents.)

(4) Expansions in renormalons (see subsections VIIC2-3) introduce new coupling constants, effectively nonperturbative corrections to the otherwise perturbative hard part. Like all but the first of these expansions, this leads to correction terms that are down by powers of $1/q^2$. However, it should be possible to absorb this type of correction into the previous one, since in principle the soft parts should contain all nonperturbative corrections by definition.



The other common application of the parton model is to "Drell-Yan scattering": In this case two hadrons scatter producing, in addition to hadrons, a photon (or other electroweak boson) that decays into a lepton-antilepton pair. To lowest order, the relevant diagram is the same as for DIS (crossing some of the lines). Because both of the initial particles are hadrons, 2 soft parts are required; however, each of these is the same as that used in DIS ("universality") so they do not need to be redetermined. In fact, there is a direct progression from e^+e^- to DIS to Drell-Yan: The above diagrams are similar except for the number ($0 \to 1 \to 2$) of soft parts (corresponding to the number of initial hadrons); the leading contribution comes from the same diagram, rotated to various positions (crossing).

More generally, we can consider not only hard parts involving identified quarks in the initial state of the hard part, but also in the final state, by examining jets. Thus, for soft parts we have not only the "parton distribution functions", which are probabilities found from amplitudes for an initial hadron $\to$ parton + anything (summing over anything), we have "fragmentation functions", which are probabilities from amplitudes for parton $\to$ final hadron + anything. In principle these are related by crossing symmetry: The diagrams are similar to the previous, with the partons connecting to the hard part, but the external hadron lines may be either initial or final (and the opposite for the corresponding parton with respect to the hard subgraph). As for the parton distributions, the fragmentation function for any particular parton and hadron is measured in one particular experiment, then used universally. (The simplest is deep inelastic scattering for the parton distribution, and e^+e^- annihilation with one of the two jets $\to$ hadron + anything for fragmentation.) Then the cross section is generally of the form

$$d\sigma_{A...B} = \sum_{a...b} \int d\xi_a \cdots d\xi_b \; f_{Aa}(\xi_a) \cdots f_{Bb}(\xi_b) d\sigma_{a...b}(\xi_i)$$

where $d\sigma_{A...B}$ is the observed (differential) cross section, $a...b$ (not to be confused with vector indices) label the different partons and $A...B$ their hadrons (we leave off the labels for non-strongly interacting particles), the sum is over different kinds (and flavors) of partons (and perhaps over different hard parts, if corrections down by powers are desired), ξ_a is the momentum fraction for parton a of hadron A's momentum, f_{Aa} is either the parton distribution function for $A \to a + X$ (X = "anything") or the fragmentation function for $a \to A+X$, and $d\sigma_{a...b}$ is the hard cross section (calculated perturbatively), which is just the original with all the hadrons replaced by partons. For the parton distributions we integrate $\int_0^1 d\xi$, while for fragmentation we integrate $\int_1^\infty d\xi$, or change variables to the hadron's fraction of the parton's momentum $\zeta = 1/\xi$ and integrate $\int_0^1 d\zeta$.

Note that, while physical cross sections are independent of the renormalization mass scale μ, the same is not true of the hard cross sections calculated perturbatively in the above factorized expressions, since they are expressed in terms of unphysical quark "states". However, these hard parts satisfy renormalization group equations, as calculated in the usual perturbative way. (Of course, nontrivial contributions require calculating beyond leading order.) This implies corresponding renormalization group equations (see subsection VIIC1), the "evolution" or "Gribov-Lipatov-Dokshitzer-Altarelli-Parisi (GLDAP) equations", to be satisfied by the parton distributions, so that μ dependence cancels in the complete cross sections. This determines the energy dependence of the parton distributions. The equations take the form

$$\mu^2 \frac{d}{d\mu^2} f_{Aa}(\xi, \mu^2) = \sum_b \int_\xi^1 \frac{d\zeta}{\zeta}\ f_{Ab}(\zeta, \mu^2) P_{ba}\left(\frac{\xi}{\zeta}, g^2(\mu^2)\right)$$

where f_{Aa} describes $A \to a+X$, f_{Ab} describes $A \to b+X'$, the "splitting functions" P_{ba} describe $b \to a+X''$ ($X = X'+X''$), and the sum is over the intermediate parton b. For hadron A with momentum p, the intermediate parton b has momentum ζp, and parton a has momentum ξp, so ξ/ζ is a's fraction of b's momentum. The kinematics are such that $0 \leq x \leq \xi \leq \zeta \leq 1$ (momentum is lost to X's as $A \to b \to a$). The splitting functions are calculated perturbatively from the corresponding renormalization group equation for the hard part, since the combined μ dependence must cancel in the physical cross section.

For similar reasons, the hard cross sections are infrared divergent; the soft parts of the complete cross sections deal with low energies. This leads to complications beyond next-to-leading order, due to the fact that the renormalization group scale μ, which relates to ultraviolet divergences (high-energy behavior), and the "factorization scale", which relates to infrared divergences (it determines the division between hard and soft energies), are in principle independent scales. This allows an ambiguity in factorization prescriptions, in addition to the usual ambiguity in UV renormalization prescriptions. (In more general processes there can be other energy scales than just q^2, each with its own factorization scale, further complicating matters.)

REFERENCES

1 R.P. Feynman, *Phys. Rev. Lett.* **23** (1969) 1415;
J.D. Bjorken and E.A. Paschos, *Phys. Rev.* **185** (1969) 1975:
parton model.

2 J.C. Collins, D.E. Soper, and G. Sterman, Factorization of hard processes in QCD, in *Perturbative quantum chromodynamics*, ed. A.H. Mueller (World-Scientific, 1989) p. 1;
G. Sterman, Introduction to perturbative QCD, in *Perspectives in the standard model*, proc. TASI '91, Boulder, Colorado, June 2-28, eds. R.K. Ellis, C.T. Hill, and J.D. Lykken (World-Scientific, 1992) p. 475;
Sterman, *loc. cit.*;

G. Sterman, hep-ph/9606312, Partons, factorization and resummation, in *QCD & beyond*, proc. TASI '95, Boulder, Colorado, June 4-30, ed. D.E. Soper (World-Scientific, 1996) p. 327:
reviews of factorization.

3 G. Sterman and S. Weinberg, *Phys. Rev. Lett.* **39** (1977) 1436:
justification for jets from perturbative QCD.

4 V.N. Gribov and L.N. Lipatov, *Sov. J. Nucl. Phys.* **15** (1972) 438, 675;
Yu.L. Dokshitzer, *Sov. Phys. JETP* **46** (1977) 641;
G. Altarelli and G. Parisi, *Nucl. Phys.* **B126** (1977) 298.

PART THREE: HIGHER SPIN

Higher-spin (unstable) particles have been observed experimentally. Whether they are considered elementary depends on how their theory is formulated. In particular, a description of hadrons in terms of strings would have many advantages, such as unification of all hadrons, manifestation of duality symmetry, and calculability through an accurate perturbation scheme.

Gravity and supergravity also include higher-spin particles. String theory might also yield some solutions to some of their problems, especially renormalizability and unification of all particles. Such gravitational strings would differ from hadronic strings in their mass scale and in the appearance of massless particles, including the graviton itself (in contrast to the massive “pomeron”, the analog of the graviton in hadronic strings). Gravitational strings might require supergravity.

For these and other reasons supergravity and strings are two of the major areas of research in theoretical high energy physics today (although not the only ones). Most of the discussion of this part is introductory, and can be covered earlier, but it is not essential to the course; however, its inclusion in a field theory text is essential at least for reference.

IX. GENERAL RELATIVITY

Before discussing supergravity we need to study ordinary gravity. Both can be treated as generalizations of Yang-Mills theory. We use this approach rather than the traditional one, based on the metric, which is insufficient for describing spinors or supersymmetry: There is no useful definition of distance in anticommuting directions in curved (super)space.

Gravity is the only observed long-range (massless) force mediated by a higher-spin (2) field. It is relevant for astrophysics, cosmology, and unification, all of which have applications to the particles of lower spin.

A. ACTIONS

We begin with the general principles that define pure gravity as a nonabelian gauge theory, and use them to derive actions and couple to matter.

1. Gauge invariance

General relativity can be described by a simple extension of the methods used to describe Yang-Mills theory. The first thing to understand is the gauge group. We start with coordinate transformations, which are the local generalization of translations, since gravity is defined to be the force that couples to energy-momentum in the same way that electromagnetism couples to charge. However, these are not enough to define spinors. This is easy to see already from the linear part of coordinate transformations: Whereas SO(3,1) is the same Lie group as SL(2,C), GL(4) (a Wick rotation of U(4)) does not have a corresponding covering group; there is no way to take the square root of a vector under coordinate transformations. So we include Lorentz transformations as an additional local group. We therefore have a coordinate transformation group, which includes translations and the orbital part of Lorentz transformations, and a local Lorentz group, which includes the spin part of Lorentz transformations.

Clearly the coordinates x^m themselves, and therefore their partial derivatives ∂_m, are not affected by the (spin) Lorentz generators. We indicate this by use of "curved" vector indices $m, n, ...$. On the other hand, all spinors should be acted on by the Lorentz generators, so we give them "flat" indices $\alpha, \beta, ..$, and we also have flat vector indices $a, b, ...$ for vectors that appear by squaring spinors. Flat indices can be treated the same way as in flat space, with metrics $C_{\alpha\beta}$ and η_{ab} to raise, lower, and contract them.

Some gravity texts, particularly the more mathematical ones, emphasize the use of "index-free notation". An example of such notation is matrix notation: Matrix notation is useful only for objects with two indices or fewer, as we saw in our treatment of spinor indices in chapter II. Such mathematical texts consider the use of indices as tantamount to specifying a choice of basis; on the contrary, as we have seen in previous chapters, indices in covariant equations usually act only (1) as place holders, indicating where contractions are made and how to associate tensors on either side of equations, and (2) as mnemonics, reminding us of representations and transformation properties. Thus, the full content of the equation can be seen at a glance. In contrast, many mathematical-style equations (when indeed equal signs are actually used) say little more than "$A = B$", with the real content of the equation buried in the text of preceding paragraphs.

We therefore define the elements of the group as

$$g = e^{\lambda}, \qquad \lambda = \lambda^m \partial_m + \tfrac{1}{2}\lambda^{ab} M_{ba}$$

where ∂_m acts on all coordinates, including the arguments of the real gauge parameters λ^m and λ^{ab} and any fields. $M_{ab} = -M_{ba}$ are the *second-quantized* Lorentz generators: They act on all flat indices, including those on λ^{ab} and any fields that carry flat indices. As a shorthand notation, sometimes we will also write

$$\tfrac{1}{2}\lambda^{ab} M_{ba} = \lambda^I M_I$$

(and similarly for other appearances of the antisymmetric index pair ab). We thus have a combination of the matrix generators of section IB and the coordinate generators of subsection IC2.

Exercise IXA1.1

Sometimes it's more convenient to perform explicit finite coordinate transformations in terms of new coordinates as functions of old, as in subsection IC2. As an example for curved space we consider the sphere in arbitrary dimensions. Rather than the usual cumbersome angles, which introduce trigonometric functions into measurements of distances, we use coordinates which manifest the slightly smaller rotational invariance of the corresponding flat space, as we did for scalar fields in subsection IVA2.

a As in subsection IVA2, we can derive coordinates for the sphere by constraining flat space in Cartesian coordinates to have unit radius. Rather than looking for an explicit solution as in subsection IVA2, we can enforce the constraint by the replacement

$$x \to \frac{y}{|y|}$$

so the flat coordinate "vector" x automatically has magnitude $|x| = 1$ at the expense of introducing the scale invariance

$$y' = \lambda(y) y \qquad (x' = x)$$

Show the infinitesimal distance ds is given by

$$ds^2 = dx^2 = \frac{dy^2}{y^2} - \frac{(y \cdot dy)^2}{y^4} = \frac{(y^{[a} dy^{b]})^2}{2y^4}$$

Check scale invariance of the last form.

b Ultimately we'll need to use the scale invariance to eliminate one coordinate. Writing $y^a = (y^0, y^i)$, consider the coordinate transformation

$$y^0 = (z^0)^2 - (z^i)^2, \quad y^i = 2z^0 z^i$$

(This is just a generalization of the substitution used in subsection IVA2.) What is the interpretation in two dimensions ($y^i = y^1$) in terms of complex coordinates? (This generalizes to quaternions in four dimensions.) Show that this results in

$$ds^2 = \frac{(2z^{[0}dz^{i]})^2}{[(z^0)^2 + (z^i)^2]^2}$$

Compare the result on ds^2 of the scale gauge $y^0 = 1$ (on the previous form) to that of $z^0 = 1$ (on this form).

Exercise IXA1.2

More general scale gauges for the previous problem come from considering the kind of projections made in map making, looking at the result of shining a point light source through a transparent globe onto a plane, where the ray from the source through the center of the globe exits it at the point tangent to the plane. Instead of looking at the geometry of the rays, we consider expanding this globe of unit radius through the plane in such a way that the source remains at the same scaled position inside the globe. (The center of the sphere moves while the source and plane remain fixed, at least with respect to each other.) The globe continues to expand until it intersects a chosen point on the plane. Explicitly, in terms of coordinates y of that point with respect to the origin of the expanded globe, the distance on the original (unit-radius) globe is

$$ds^2 = dx^2 = \left(d\frac{y}{|y|}\right)^2$$

while the position of the source is

$$x_s^0 = -a, \quad x_s^i = 0 \quad \Rightarrow \quad y_s = (-a|y|, 0)$$

in terms of the constant a that defines the gauge (projection), so the condition that it hasn't moved relative to the plane is

$$y^0 + a|y| = 1 + a$$

a Show the solution is

$$y^0 = \frac{(1+b) - (1-b)\sqrt{1 + b(y^i)^2}}{2b}, \qquad b = \frac{1-a}{1+a}$$

or

$$z^0 = \sqrt{1 + b(z^i)^2}$$

b Find ds^2 in terms of both y^i and z^i for the special cases

$$\begin{array}{rl} \textit{gnomonic}: & a = 0 \quad (b = 1) \\ \textit{stereographic}: & a = 1 \quad (b = 0) \\ \textit{orthographic}: & a = \infty \quad (b = -1) \end{array}$$

In general, the relation between first- and second-quantized group generators is the same as the relation between active and passive transformations, and the relation between a matrix representation and the corresponding coordinate representation, as discussed in subsection IC1, where in this case the fields are the coordinates. In particular, the second-quantized Lorentz operators M_{ab} have the same action as the first-quantized Lorentz operators S_{ab} introduced in subsection IIB1: For any field ψ, $M\langle\psi| = \langle\psi|S$, etc. The action of the Lorentz generators on vector indices is thus given by

$$[M_{ab}, V_c] = V_{[a}\eta_{b]c} \quad \Rightarrow \quad \lambda^I [M_I, V_a] = \tfrac{1}{2}\lambda^{bc}[M_{cb}, V_a] = \lambda_a{}^b V_b$$

This implies the commutation relations

$$[M_{ab}, M^{cd}] = -\delta^{[c}_{[a} M_{b]}{}^{d]}$$

In explicit calculations, only two indices in the commutator will match, and they reduce to simple expressions such as

$$\boxed{[M_{12}, V_2] = \eta_{22} V_1, \qquad [M_{12}, M_{23}] = \eta_{22} M_{13}}$$

As for derivatives, when acting on functions instead of operators we can write the action of the Lorentz generator as simply $M_{ab}V_c$ without the commutator.

When spinors are involved in four dimensions, it's simpler to convert all flat indices to spinor indices. In that case, we can write

$$\lambda = \lambda^m \partial_m + \lambda^I M_I, \qquad \lambda^I M_I = \tfrac{1}{2}\lambda^{ab} M_{ba} = \tfrac{1}{2}\lambda^{\alpha\beta} M_{\beta\alpha} + \tfrac{1}{2}\lambda^{\dot\alpha\dot\beta} M_{\dot\beta\dot\alpha}$$

$$[M_{\alpha\beta}, \psi_\gamma] = \psi_{(\alpha} C_{\beta)\gamma} \quad \Rightarrow \quad \lambda^I [M_I, \psi_\alpha] = \tfrac{1}{2}\lambda^{\beta\gamma}[M_{\gamma\beta}, \psi_\alpha] = \lambda_\alpha{}^\beta \psi_\beta$$

$$[M_{\alpha\beta}, M^{\gamma\delta}] = \delta^{(\gamma}_{(\alpha} M_{\beta)}{}^{\delta)}$$

in terms of the SL(2,C) generators $M_{\alpha\beta} = M_{\beta\alpha}$. Note that $(M_{\alpha\beta})^\dagger = +\overline{M}_{\dot\alpha\dot\beta}$ because $(M_{ab})^\dagger = -M_{ab}$. We have used conventions consistent with OSp generators

$$\tfrac{1}{2}\lambda^{BC}[M_{CB}, \psi_A\} = \lambda_A{}^B \psi_B, \qquad \eta_{AB} = (\eta_{ab}, C_{\alpha\beta}, C_{\dot\alpha\dot\beta})$$

Relating vector to spinor indices as usual as $V_a = V_{\alpha\dot{\alpha}}$, etc., then fixes the Lorentz subgroup of the OSp group as (see exercise IIB7.2a)

$$M_{\alpha\dot{\alpha}\beta\dot{\beta}} V_{\gamma\dot{\gamma}} = V_{\alpha\dot{\alpha}} C_{\beta\gamma} C_{\dot{\beta}\dot{\gamma}} - V_{\beta\dot{\beta}} C_{\alpha\gamma} C_{\dot{\alpha}\dot{\gamma}}$$

$$= -\tfrac{1}{2}(C_{\dot{\alpha}\dot{\beta}} V_{(\alpha\dot{\gamma}} C_{\beta)\gamma} + C_{\alpha\beta} V_{\gamma(\dot{\alpha}} C_{\dot{\beta})\dot{\gamma}}) = -\tfrac{1}{2}(C_{\dot{\alpha}\dot{\beta}} M_{\alpha\beta} + C_{\alpha\beta} M_{\dot{\alpha}\dot{\beta}}) V_{\gamma\dot{\gamma}}$$

$$\Rightarrow \quad M_{\alpha\dot{\alpha}\beta\dot{\beta}} = -\tfrac{1}{2}(C_{\dot{\alpha}\dot{\beta}} M_{\alpha\beta} + C_{\alpha\beta} M_{\dot{\alpha}\dot{\beta}})$$

$$\Rightarrow \quad \lambda_{\alpha\dot{\alpha}\beta\dot{\beta}} = C_{\dot{\alpha}\dot{\beta}} \lambda_{\alpha\beta} + C_{\alpha\beta} \lambda_{\dot{\alpha}\dot{\beta}}$$

For most of the remaining discussion of gravity, we'll limit ourselves to bosonic fields in vector notation, which is easy to generalize to arbitrary dimensions. For spinors, we must either choose a dimension and use its corresponding spinor notation (for $D \leq 6$), or work in mixed spinor-vector notation (which is much messier).

Matter representations of the group work similarly to Yang-Mills. We define such fields to have only flat indices. Then their transformation law is

$$\psi' = e^{\lambda} \psi$$

where the transformation of a general Lorentz representation follows from that for a vector (or spinor, if we include them), as defined above. Alternatively, the transformation of a vector could be defined with curved indices, being the adjoint representation of the coordinate group:

$$V = V^m \partial_m \quad \Rightarrow \quad V' = V'^m \partial_m = e^{\lambda^m \partial_m} V e^{-\lambda^m \partial_m}$$

However, as in Yang-Mills theory, it is more convenient to identify only the gauge field as an operator in the group. In any case, only the adjoint representation (and direct products of it) has such a nice operator interpretation.

As an example of this algebra, we now work out the commutator of two transformations in gory detail: We first recall that the coordinate transformation commutator was already worked out in subsection IC2, using the usual quantum mechanical relations (see also subsection IA1)

$$[f, f] = [\partial, \partial] = 0, \qquad [\partial, f] = (\partial f)$$

for any function f. For the Lorentz algebra we will use the additional identities

$$[M_{ab}, \partial_m] = [M_{ab}, \lambda^m] = [M_{ab}, \lambda^{cd} M_{dc}] = 0$$

all expressing the fact the Lorentz generators commute with anything lacking free flat indices (i.e., Lorentz scalars). The commutator algebra is then

$$\begin{aligned}
[\lambda_1^m \partial_m + \tfrac{1}{2}\lambda_1^{ab} M_{ba}, &\lambda_2^n \partial_n + \tfrac{1}{2}\lambda_2^{cd} M_{dc}] \\
&= \lambda_{[1}^m [\partial_m, \lambda_{2]}^n]\partial_n + \lambda_{[1}^m [\partial_m, \tfrac{1}{2}\lambda_{2]}^{ab}] M_{ba} + \tfrac{1}{2}\lambda_1^{ab}[M_{ba}, \tfrac{1}{2}\lambda_2^{cd}] M_{dc} \\
&= (\lambda_{[1}^n \partial_n \lambda_{2]}^m)\partial_m + \tfrac{1}{2}(\lambda_{[1}^m \partial_m \lambda_{2]}^{ab} + \lambda_{[1}^{ac}\lambda_{2]c}{}^b) M_{ba}
\end{aligned}$$

One fine point to worry about: We may consider spaces with nontrivial topologies, where it is not possible to choose a single coordinate system for the entire space. For example, on a sphere spherical coordinates have singularities at the two poles, where varying the longitude gives the same point and not a line. (However, the sphere can be described by coordinates with only one singular point.) We then either treat such points by a limiting procedure, or choose different sets of nonsingular coordinates on different regions ("patches") and join them to cover the space.

2. Covariant derivatives

We can also define covariant derivatives in a manner similar to Yang-Mills theory; however, since ∂_m is now one of the generators, the "∂" term can be absorbed into the "A" term of $\nabla = \partial + A$:

$$\nabla_a = e_a{}^m \partial_m + \tfrac{1}{2}\omega_a{}^{bc} M_{cb}$$

in terms of the"vierbein (tetrad)" $e_a{}^m$ and "Lorentz connection" $\omega_a{}^{bc}$. Now the action of the covariant derivative on matter fields looks even more similar to the gauge transformations: e.g.,

$$\begin{aligned}
\delta\phi &= \lambda^m \partial_m \phi, & \nabla_a \phi &= e_a{}^m \partial_m \phi \\
\delta V_a &= \lambda^m \partial_m V_a + \lambda_a{}^b V_b, & \nabla_a V_b &= e_a{}^m \partial_m V_b + \omega_{ab}{}^c V_c \\
\delta\psi_\alpha &= \lambda^m \partial_m \psi_\alpha + \lambda_\alpha{}^\beta \psi_\beta, & \nabla_a \psi_\beta &= e_a{}^m \partial_m \psi_\beta + \omega_{a\beta}{}^\gamma \psi_\gamma
\end{aligned}$$

I.e., the covariant derivative ∇_a is essentially D elements (labeled by "a") of the gauge algebra.

Exercise IXA2.1

Write the transformation law and covariant derivative of an antisymmetric tensor in spinor notation ($f_{\alpha\beta}$), and compare to vector notation as above.

Note that the free index on the covariant derivative is flat so that it transforms nontrivially under

$$\nabla'_a = e^\lambda \nabla_a e^{-\lambda}$$

Explicitly, for an infinitesimal transformation $\delta\nabla = [\lambda, \nabla]$ we have

$$\delta e_a{}^m = (\lambda^n \partial_n e_a{}^m - e_a{}^n \partial_n \lambda^m) + \lambda_a{}^b e_b{}^m$$

$$\delta\omega_a{}^{bc} = \lambda^m \partial_m \omega_a{}^{bc} + (-e_a{}^m \partial_m \lambda^{bc} + \omega_a{}^{d[b}\lambda_d{}^{c]} + \lambda_a{}^d \omega_d{}^{bc})$$

This commutator is the same as for $[\lambda_1, \lambda_2]$ in the previous subsection, except for the two additional terms coming from the Lorentz generators acting on the free index on ∇_a. In particular, the vierbein $e_a{}^m$ transforms on its flat index as the vector (defining) representation of the local Lorentz group, and on its curved index (and argument) as the vector (adjoint) representation of the coordinate group. Also, it should be invertible, since originally we had $\nabla = \partial + A$: We want to be able to separate out the flat space part as $e_a{}^m = \delta_a^m + h_a{}^m$ for perturbation theory or weak gravitational fields. That means we can use it to convert between curved and flat indices:

$$V^m = V^a e_a{}^m \quad \Leftrightarrow \quad V^a = V^m e_m{}^a$$

where $e_m{}^a$ is the inverse of $e_a{}^m$. Furthermore, if we want to define the covariant derivative of an object with curved indices, we can simply flatten its indices, take the covariant derivative with ∇, and then unflatten its indices.

Flat indices are the natural way to describe tensors: (1) They are the *only* way to describe half-integral spin. (2) Even for integer spin, they correspond to the way components are actually measured. In fact, the above conversion of vectors from curved to flat indices is exactly the one you learned in your freshman physics course! The special cases you saw there were curvilinear coordinates (polar or spherical) for flat space. Then $e_a{}^m$ was the usual orthonormal basis. Only the notation was different: Using Gibbs' notation for the curved but not the flat indices, $\vec{V} = V^a \vec{e}_a$, where, e.g., $a = (r, \theta, \phi)$ for spherical coordinates and $\vec{e}_a = (\hat{r}, \hat{\theta}, \hat{\phi})$ are the usual orthonormal basis. Thus, you probably learned about the vierbein years before you ever saw a "metric tensor". Similarly, when you learned how to integrate over the volume element of spherical coordinates, you found it from this basis, and only learned much later (if yet) to express it in terms of the square root of the determinant of the metric. (With the orthonormal basis, there was no square root to take; the determinant came from the cross product.) You also learned how to do this for curved space: Considering again the sphere, vectors in the sphere itself can be expressed in terms of just $\hat{\theta}$ and $\hat{\phi}$. And the area element of the sphere (the volume element of this smaller space) you again found from this basis.

For example, consider nonrelativistic momentum in two flat spatial dimensions, but in polar coordinates: Now using x^m to represent just the nonrelativistic spatial coordinates,

$$x^m = (r, \theta), \qquad p^m = m\frac{dx^m}{dt} = (m\dot{r}, m\dot{\theta}) = p^a e_a{}^m$$

$$e_a{}^m = \begin{pmatrix} 1 & 0 \\ 0 & r^{-1} \end{pmatrix}, \qquad p^a = (m\dot{r}, mr\dot{\theta})$$

Then the two components of p^a (with the simplest choice of $e_a{}^m$) are the usual components of momentum in the radial and angular directions. On the other hand, one component of p^m is still the radial component of the momentum, while the other component of p^m is the *angular* momentum — a useful quantity, but not normally considered as a component along with the radial momentum, which doesn't even have the same engineering dimensions. In writing the Hamiltonian, one simply squares p^a in the naive way, whereas squaring p^m would require use of the metric.

Exercise IXA2.2

Show that the above choice of $e_a{}^m$ actually describes flat space: Use the fact that p^a transforms as a scalar under the coordinate transformations that express r and θ in terms of Cartesian coordinates x and y, and as a vector under local "Lorentz" transformations, which are in this case just 2D rotations, to transform it to the usual Cartesian $p'^a = (m\dot{x}, m\dot{y})$.

This direct conversion between curved and flat indices also leads directly to the covariant generalization of length: In terms of momentum (as would appear in the action for the classical mechanics of the particle),

$$p^m = m\frac{dx^m}{ds}, \quad -m^2 = p^2 = p^a p^b \eta_{ab} \quad \Rightarrow \quad -ds^2 = dx^m dx^n e_m{}^a e_n{}^b \eta_{ab} \equiv dx^m dx^n g_{mn}$$

Equivalently, the metric tensor g_{mn} is just the conversion of the flat-space metric η_{ab} to curved indices. Also, in terms of differential forms,

$$\Omega^a = dx^m e_m{}^a \quad \Rightarrow \quad -ds^2 = \Omega^a \Omega^b \eta_{ab}$$

The field strengths are also defined as in Yang-Mills:

$$\boxed{[\nabla_a, \nabla_b] = T_{ab}{}^c \nabla_c + \tfrac{1}{2} R_{ab}{}^{cd} M_{dc}}$$

where we have expanded the field strengths over ∇ and M rather than ∂ and M so that the "torsion" T and "curvature" R are manifestly covariant:

$$M' = e^\lambda M e^{-\lambda} = M \quad \Rightarrow \quad T' = e^\lambda T e^{-\lambda}, \quad R' = e^\lambda R e^{-\lambda}$$

The commutator can be evaluated as before, with the same change as for going from $[\lambda_1, \lambda_2]$ to $[\lambda, \nabla_a]$ (i.e., now there are two free indices on which the Lorentz generators can act), except that now we rearrange terms to convert $\partial_m \to e_a{}^m \partial_m \to \nabla_a$. Making the further definitions

$$e_a = e_a{}^m \partial_m, \quad [e_a, e_b] = c_{ab}{}^c e_c \quad \Rightarrow \quad c_{ab}{}^c = (e_{[a} e_{b]}{}^m) e_m{}^c = -e_a{}^m e_b{}^n \partial_{[m} e_{n]}{}^c$$

for the "structure functions" $c_{ab}{}^c$, we find the explicit expressions

$$T_{ab}{}^c = c_{ab}{}^c + \omega_{[ab]}{}^c = -e_a{}^m e_b{}^n (\partial_{[m} e_{n]}{}^c + e_{[m}{}^d \omega_{n]d}{}^c)$$

$$R_{ab}{}^{cd} = e_{[a} \omega_{b]}{}^{cd} - c_{ab}{}^e \omega_e{}^{cd} + \omega_{[a}{}^{ce} \omega_{b]e}{}^d = e_a{}^m e_b{}^n (\partial_{[m} \omega_{n]}{}^{cd} + \omega_{[m}{}^{ce} \omega_{n]e}{}^d)$$

If we ignore the action of ∇ on curved indices (it doesn't act on them, but alternatively we could flatten them, act, then curve them back), we can also write

$$\boxed{T_{mn}{}^a = -\nabla_{[m} e_{n]}{}^a, \qquad [\nabla_m, \nabla_n] = \tfrac{1}{2} R_{mn}{}^{ab} M_{ba}}$$

where

$$\nabla_m = e_m{}^a \nabla_a = \partial_m + \tfrac{1}{2} \omega_m{}^{ab} M_{ba}$$

is essentially a covariant derivative for the Lorentz group only.

From this expression for the torsion we find the following expressions for the curl and divergence of a vector in terms of curved indices: Defining

$$\mathbf{e} \equiv det\ e_a{}^m$$

$$\Rightarrow \quad c_{ba}{}^b = (e_{[b}{}^m \partial_m e_{a]}{}^n) e_n{}^b = (e_b{}^m \partial_m e_a{}^n) e_n{}^b - e_a{}^m [(\partial_m e_b{}^n) e_n{}^b] = \partial_m e_a{}^m - e_a{}^m \partial_m ln\ \mathbf{e}$$

we have

$$e_a{}^m e_b{}^n \partial_{[m} V_{n]} = e_{[a} (e_{b]}{}^m V_m) - c_{ab}{}^c e_c{}^m V_m = \nabla_{[a} V_{b]} - T_{ab}{}^c V_c$$

$$\mathbf{e}\, \partial_m \mathbf{e}^{-1} V^m = e_a{}^m \partial_m V^a + c_{ba}{}^b V^a = \nabla_a V^a + T_{ba}{}^b V^a$$

Exercise IXA2.3

Relate the two above identities by comparing (in D=4) $\nabla_{[a} V_{bcd]}$ (generalizing $\nabla_{[a} V_{b]}$) and $\nabla_a V^a$ for $V_{abc} = \epsilon_{abcd} V^d$.

In practice, a useful way to evaluate the commutator is to first evaluate the commutators of the Lorentz generators with the whole covariant derivative, and then subtract out the double-counted $[M, M]$ term. This is particularly convenient when considering some explicit solution to the field equations with a reduced set of components (e.g., spherically symmetric), so that explicit indices may be lost except on the Lorentz generators. Schematically, we then calculate

$$\boxed{\begin{gathered}[\nabla_1,\nabla_2]=[e_1+\omega_1,e_2+\omega_2]\\ =\{[e_1,e_2]+(e_1\omega_2)M_2-(e_2\omega_1)M_1\}+\{\omega_1[M_1,\nabla_2]-\omega_2[M_2,\nabla_1]-\omega_1\omega_2[M_1,M_2]\}\end{gathered}}$$

This method turns out to be one of the two simplest ways to calculate explicit solutions (as opposed to discussing general properties). (For examples, see subsection IXC5 below.)

The covariant derivative satisfies the Bianchi (Jacobi) identities

$$0=[\nabla_{[a},[\nabla_b,\nabla_{c]}]]=[\nabla_{[a},T_{bc]}{}^d\nabla_d+\tfrac{1}{2}R_{bc]}{}^{de}M_{ed}]$$

$$=(\nabla_{[a}T_{bc]}{}^d)\nabla_d+\tfrac{1}{2}(\nabla_{[a}R_{bc]}{}^{de})M_{ed}-T_{[ab|}{}^e(T_{e|c]}{}^f\nabla_f+\tfrac{1}{2}R_{e|c]}{}^{fg}M_{gf})-R_{[abc]}{}^d\nabla_d$$

$$\Rightarrow\quad R_{[abc]}{}^d=\nabla_{[a}T_{bc]}{}^d-T_{[ab|}{}^eT_{e|c]}{}^d,\qquad \nabla_{[a}R_{bc]}{}^{de}-T_{[ab|}{}^fR_{f|c]}{}^{de}=0$$

To make the transformation laws manifestly covariant we can define instead

$$\lambda=\lambda^a\nabla_a+\tfrac{1}{2}\lambda^{ab}M_{ba}$$

which is just a redefinition of the gauge parameters. The infinitesimal transformation law of the covariant derivative is then

$$\begin{aligned}\delta\nabla_a&=[(\delta e_a{}^m)e_m{}^b]\nabla_b+\tfrac{1}{2}(e_a{}^m\delta\omega_m{}^{bc})M_{cb}=[\lambda^b\nabla_b+\tfrac{1}{2}\lambda^{bc}M_{cb},\nabla_a]\\&=(-\nabla_a\lambda^b+\lambda^cT_{ca}{}^b+\lambda_a{}^b)\nabla_b+\tfrac{1}{2}(-\nabla_a\lambda^{bc}+\lambda^dR_{da}{}^{bc})M_{cb}\end{aligned}$$

$$\Rightarrow\quad(\delta e_a{}^m)e_m{}^b=-\nabla_a\lambda^b+\lambda^cT_{ca}{}^b+\lambda_a{}^b,\quad e_a{}^m\delta\omega_m{}^{bc}=-\nabla_a\lambda^{bc}+\lambda^dR_{da}{}^{bc}$$

Exercise IXA2.4

Show that a *finite* local Lorentz transformation takes the form

$$\nabla'_a=\Lambda_a{}^b[\nabla_b-\tfrac{1}{2}(\nabla_b\Lambda^{cf})\Lambda^d{}_fM_{dc}]$$

in the case of vanishing torsion by starting with the more general expression

$$\nabla'_a=\Lambda_a{}^b\nabla_b+\tfrac{1}{2}\Delta_a{}^{bc}M_{cb}$$

examining the commutator $[\nabla',\nabla']$, and applying $T'=0$ to determine Δ.

A “Killing vector” is a transformation that leaves the covariant derivative invariant:

$$[K,\nabla_a]=0,\qquad K=K^a\nabla_a+\tfrac{1}{2}K^{ab}M_{ba}$$

(The term is usually used to refer to just the general coordinate part K^m of the transformation, but we'll use it in a generalized sense to refer to the complete K.)

It represents a symmetry; the existence of Killing vectors depends on the particular space described by the covariant derivative. It then follows from the Jacobi identity for $[K_1,[K_2,\nabla]]$ that the Killing vectors form a group, the symmetry group of that space. Invariance of the covariant derivative requires:

$$-\nabla_a K^b + K^c T_{ca}{}^b + K_a{}^b = 0 \quad \Rightarrow \quad \nabla_{(a} K_{b)} = K^c T_{c(ab)}, \qquad K_{ab} = \tfrac{1}{2}\nabla_{[a}K_{b]} - \tfrac{1}{2}K^c T_{c[ab]}$$

$$-\nabla_a K^{bc} + K^d R_{da}{}^{bc} = 0 \quad \Rightarrow \quad \nabla_a K^{bc} = K^d R_{da}{}^{bc}$$

These equations are referred to as the "Killing equations". (Again, usually it is just the first equation, on K_a, that is called by this name, but we'll use it to refer also to the equations for K_{ab}, which are needed to describe the symmetry when acting on spinors, etc.)

Exercise IXA2.5

Express the Hamiltonian of the classical relativistic particle in terms of the vierbein:

$$e_a = e_a{}^m p_m, \quad H = \tfrac{1}{2}(\eta^{ab} e_a e_b + m^2)$$

Doing the same for general coordinate transformations $K = K^a e_a$, examine the condition for invariance $[K, H] = 0$ using the Poisson bracket. Using the commutation relations for the e_a's, show that this implies the Killing equation $\nabla_{(a} K_{b)} = K^c T_{c(ab)}$.

Exercise IXA2.6

Solve the Killing equations explicitly in the case of flat space $\nabla_a = \partial_a$. Show this gives the Poincaré group, including both orbital and spin pieces.

3. Conditions

There are two kinds of conditions we can impose to eliminate some degrees of freedom: *gauge choices* and *constraints*. Gauge choices explicitly determine degrees of freedom that drop out of the action anyway. If the gauge is not completely fixed, the form of the residual gauge transformations may change, since using particular gauge parameters to fix the gauge, rather than eliminating those parameters, may just determine them in terms of the remaining parameters: We require that the residual transformations do not violate the gauge conditions that have already been applied. Similar remarks apply to global symmetries: If they do not commute with the gauge transformations for the gauge that was fixed, then they may aquire extra gauge-transformation terms to preserve the gauge choice. On the other hand, constraints are chosen to be covariant under the transformation laws, and thus do not alter them, while eliminating degrees of freedom that might otherwise appear in the action (although not in all possible terms). Furthermore, the simplest explicit solution to constraints can itself introduce new gauge invariances. (An example of this situation is supersymmetric Yang-Mills: see subsections IVC3-4.) In this subsection this analysis will be applied to Lorentz invariance: We already saw that global Lorentz transformations are included in coordinate transformations, and that local Lorentz invariance is unnecessary when only integer spin (and in particular, pure gravity) is treated. We now examine the consequences of eliminating this useful but redundant invariance and the gauge field associated with it.

Of course, we can eliminate local Lorentz transformations by hiding flat indices: For the vierbein itself, we have the local Lorentz invariant

$$g^{mn} = \eta^{ab} e_a{}^m e_b{}^n$$

which is the inverse "metric tensor". However, we have seen that tensors with flat indices have simpler coordinate transformations, and there is no way to get rid of flat indices when spinors are involved. Furthermore, the metric has the constraint that it have Minkowski signature: This constraint is solved by expressing the metric in terms of the flat-space Minkowsi metric η and the vierbein. Thus, solving the constraint introduces local Lorentz invariance. (However, in this case the constraint does not eliminate degrees of freedom, but only limits their range.)

The Lorentz transformations in λ^{ab} are redundant to those in λ^m. The extra gauge parameters also can be fixed by an appropriate gauge choice: For example, consider the gauge

$$\delta e_a{}^m = \lambda_a{}^b e_b{}^m \quad \Rightarrow \quad \textit{Lorentz gauge } \eta_{m[a} e_{b]}{}^m = 0$$

A coordinate transformation takes us to a different Lorentz gauge, since the Lorentz gauge condition is not a scalar. This means that any coordinate transformation λ^m must be accompanied by a Lorentz transformation λ^{ab} to preserve this gauge, where this λ^{ab} is completely determined in terms of λ^m. This is easy to see perturbing $e_a{}^m$ about δ_a^m: To lowest order we have simply

$$0 = \delta(\eta_{m[a} e_{b]}{}^m) \approx -\lambda_{ab} + \partial_{[a}\lambda_{b]} \quad \Rightarrow \quad \lambda_{ab} \approx \partial_{[a}\lambda_{b]}$$

Exercise IXA3.1

Let's further analyze this gauge condition:

a By looking at the transformation of a vector, identify the specific terms in the Taylor expansions of λ^m and λ^{ab} whose coefficients can be identified with global Lorentz transformations, in the approximation used above.

b Using the same methods as exercise IVC4.3, and writing in matrix notation $e_a{}^m = (e^h)_a{}^m$ for some matrix h, solve explicitly for λ^{ab} in terms of λ^m and $e_a{}^m$ to all orders.

Similarly, the Lorentz connection $\omega_a{}^{bc}$ that gauges the λ^{ab} transformations is redundant to the vierbein that gauges λ^m: ω can be completely determined in terms of e by constraining the torsion to vanish. To see this, we first notice that in the general case the expression for the torsion in terms of the structure functions and connection can be inverted to give the connection in terms of the other two. One way to do this is to use the definition and permute the indices $a \to b \to c$ (odd permutations are redundant because of the antisymmetry of the equation in the first two indices):

$$T_{abc} = c_{abc} + \omega_{abc} - \omega_{bac}, \qquad T_{bca} = c_{bca} + \omega_{bca} - \omega_{cba}, \qquad T_{cab} = c_{cab} + \omega_{cab} - \omega_{acb}$$

Using the antisymmetry of the connection in its last two indices, we add the first and last equation and subtract the second to obtain

$$\omega_{abc} = \tfrac{1}{2}(\tilde{c}_{bca} - \tilde{c}_{a[bc]}), \qquad \tilde{c}_{abc} = c_{abc} - T_{abc}$$

Since the torsion is a covariant tensor, we can freely set it to vanish without affecting the transformation laws of the remaining objects (it's a covariant constraint, not a gauge condition):

$$T_{ab}{}^c = 0 \quad \Rightarrow \quad \omega_{abc} = \tfrac{1}{2}(c_{bca} - c_{a[bc]})$$

From now on we assume this constraint is satisfied. This simplifies the form of curls and divergences, which implies that ∇ can be integrated by parts in covariant actions (see below). However, we have already seen that the torsion is nonvanishing in superspace (subsection IVC3): In that case the symmetry on flat indices is constrained, so the connection has fewer components than the torsion, and can be determined by setting only part of the torsion to vanish. (See subsection XA1 below.)

Exercise IXA3.2

Show explicitly that when the torsion vanishes the Killing equations from $\delta e_a{}^m = 0$ imply those from $\delta\omega_m{}^{ab} = 0$:

$$\nabla_{(a}\lambda_{b)} = 0, \qquad \lambda_{ab} = \tfrac{1}{2}\nabla_{[a}\lambda_{b]} \quad \Rightarrow \quad \nabla_a\lambda^{bc} = \lambda^d R_{da}{}^{bc}$$

Exercise IXA3.3

Consider using the group GL(D) on the flat indices instead of SO(D−1,1). (This construction is not useful for fermions.) Compensate for the extra gauge invariance by replacing the Minkowski metric η_{ab} with a "flat"-index metric g_{ab} (and its inverse g^{ab}) that is *coordinate dependent*, but covariantly constant:

$$\lambda = \lambda^m \partial_m + \lambda_a{}^b G_b{}^a; \qquad \lambda_a{}^b [G_b{}^a, V_c] = \lambda_c{}^a V_a, \qquad \lambda_a{}^b [G_b{}^a, V^c] = -V^a \lambda_a{}^c$$

$$\nabla_a = e_a + \omega_{ab}{}^c G_c{}^b, \qquad [\nabla_a, \nabla_b] = T_{ab}{}^c \nabla_c + R_{abc}{}^d G_d{}^c$$

$$g^{mn} = g^{ab} e_a{}^m e_b{}^n, \qquad \nabla_a g^{bc} = 0$$

where now there is no (anti)symmetry associated with the indices on $G_a{}^b$ (or $\lambda_a{}^b$, etc.). As a result, g_{ab} transforms nontrivially under both coordinate and GL(D) transformations. Use it (in place of η) to raise and lower flat indices.

a Find the explicit expressions for the torsion and curvature in terms of the vierbein and connection. Solve these, and $\nabla g = 0$, for the connection in terms of the torsion and vierbein as

$$\omega_{abc} = \tfrac{1}{2}(\tilde{c}_{bca} - \tilde{c}_{a[bc]}) + \tfrac{1}{2}(e_c g_{ab} - e_{(a} g_{b)c}), \qquad \tilde{c}_{abc} = c_{abc} - T_{abc}$$

Show that there exists a GL(D) *gauge*

$$g_{ab} = \eta_{ab}$$

(*assuming* g_{ab} has the right signature), that gauge has as a residual flat-index invariance SO(D−1,1), and the resulting covariant derivative is identical to that used earlier in this subsection.

b Show that one can instead choose a GL(D) gauge

$$e_a{}^m = \delta_a^m \quad \Rightarrow \quad g^{mn} = g^{ab}$$

and that this completely fixes the GL(D) invariance. Since the vierbein has a curved index, the covariant derivatives are no longer covariant: Unlike the previous gauge, to maintain this gauge any coordinate transformation must be accompanied by a GL(D) transformation whose parameter is determined by the coordinate transformation parameter. Find the solution for $\lambda_a{}^b$ in terms of λ^m in the infinitesimal case. Compare with the transformation law for curved indices (see subsection IC2). In this gauge the connection is known as the "Christoffel symbols".

The vanishing of the torsion simplifies the Bianchi identities on the curvature:

$$T_{ab}{}^c = 0 \quad \Rightarrow \quad R_{[abc]d} = \nabla_{[a} R_{bc]de} = 0 \quad \Rightarrow \quad R_{abcd} = R_{cdab}$$

In terms of SU(N)-like Young tableaux, this means the curvature is of the form $\begin{array}{|c|c|}\hline \ & \ \\ \hline \ & \ \\ \hline\end{array}$. For SO(N) Young tableaux, we subtract out the trace pieces:

$$R_{abcd} \to \begin{array}{|c|c|}\hline \ & \ \\ \hline \ & \ \\ \hline\end{array} \oplus \begin{array}{|c|c|}\hline \ & \ \\ \hline\end{array} \oplus \bullet$$

where the first term is the "Weyl tensor" W_{abcd} (traceless), the last two terms combine to give the "Ricci tensor" $R_{ab} \equiv R_a{}^c{}_{bc}$, and the last (singlet) term is the "Ricci scalar" $R \equiv R^a{}_a = R^{ab}{}_{ab}$. They're simpler in spinor notation in $D = 4$: Since $[ab] \to (\alpha\beta)$ and $(\dot\alpha\dot\beta)$,

$$R_{abcd} \to R_{(\alpha\beta)(\dot\alpha\dot\beta)} \oplus (R_{(\alpha\beta)(\gamma\delta)} = W_{(\alpha\beta\gamma\delta)} + C_{(\alpha(\gamma}C_{\delta)\beta)}R)$$

in terms of Weyl $W_{(\alpha\beta\gamma\delta)}$, the traceless part of Ricci $R_{(\alpha\beta)(\dot\alpha\dot\beta)}$, and the Ricci scalar R. Later we'll see that the Ricci tensor is fixed exactly by the equations of motion. That leaves the Weyl tensor as the on-shell field strength. As explained in subsection IIB7, it describes helicity ± 2.

Exercise IXA3.4

Prove that $R_{abcd} = R_{cdab}$ follows from the Bianchi identity $R_{[abc]d} = 0$ and the antisymmetry of R_{abcd} in both ab and cd.

4. Integration

The antihermitian form of the group generators was a convenient choice because partial derivatives are antihermitian, and the generators of the Lorentz group (which is real and orthogonal) are antisymmetric in the vector representation. Thus, the generators are real. However, the group elements are not unitary, since hermitian conjugation reorders λ^m with respect to ∂_m. The fix comes from noticing that

$$\mathbf{e} = det\ e_a{}^m \quad \Rightarrow \quad \delta\ ln\ \mathbf{e} = e_m{}^a \delta e_a{}^m = \lambda^m \partial_m ln\ \mathbf{e} - \partial_m \lambda^m$$

$$\Rightarrow \quad \delta(\mathbf{e}^{-1}) = \mathbf{e}^{-1}\overleftarrow{\lambda} = \mathbf{e}^{-1}\lambda^m \overleftarrow{\partial}_m \quad \Rightarrow \quad (\mathbf{e}^{-1})' = \mathbf{e}^{-1}e^{\overleftarrow{\lambda}}$$

where the derivatives $\overleftarrow{\partial}$ act on everything to the left, λ now includes just coordinate transformations, and we have exponentiated by the same method as for Lie groups in subsection IA3. (Note that if we expand the exponential in a Taylor series such derivatives in all but the first factor will hit λ's, just as for those in e^λ acting to the right.) Any function that transforms in this way is known as a "density" (see subsection IIIB1 for the 1D case). We can easily see from the infinitesimal transformation that a density times any scalar is also a density. This allows invariant actions to be constructed as

$$S = \int dx\ \mathbf{e}^{-1}L$$

for any scalar L. For cases without spinors we can also use

$$g \equiv det\ g_{mn} = -\mathbf{e}^{-2} \quad \Rightarrow \quad \mathbf{e}^{-1} = \sqrt{-g}$$

where g_{mn} is the inverse of g^{mn}. (In spaces of general signature, i.e., arbitrary numbers of time dimensions, we should write $\sqrt{|g|}$ so, e.g., in Euclidean space we actually use $\sqrt{g}$. If we were even more general, and used $|det\ \eta| \neq 1$, then it would also appear.) This can also be understood in terms of differential forms, since

$$\Omega^a = dx^m e_m{}^a \quad \Rightarrow \quad \Omega^4 = dx^m dx^n dx^p dx^q e_m{}^0 e_n{}^1 e_p{}^2 e_q{}^3 = d^4x\ \mathbf{e}^{-1}$$

$$\Omega'^a(x') = \Omega^a(x) \quad \Rightarrow \quad \left(\int \Omega^4 L\right)' = \int \Omega^4 L$$

under coordinate transformations.

Exercise IXA4.1

Let's look at some properties of transformations acting backwards:

a Show that for any function f

$$\lambda f = [\lambda, f] = [f, \overleftarrow{\lambda}] \quad \Rightarrow \quad e^{\lambda} f = e^{\lambda} f e^{-\lambda} = e^{-\overleftarrow{\lambda}} f e^{\overleftarrow{\lambda}}$$

and use it to show that the product of $\mathbf{e}^{-1}$ with any scalar transforms the same way as $\mathbf{e}^{-1}$ (i.e., is a density) under a finite coordinate transformation.

b Derive

$$f e^{\overleftarrow{\lambda}} = \left(1 \cdot e^{\overleftarrow{\lambda}}\right)(e^{\lambda} f)$$

(where the derivatives in each factor of $\overleftarrow{\lambda}$ act on everything to the left, but vanish on "1").

Exercise IXA4.2

We now examine finite transformations in terms of transformed coordinates (see subsection IC2):

b Show that

$$det\left(\frac{\partial x'}{\partial x}\right) = 1 \cdot e^{-\overleftarrow{\lambda}}$$

by evaluating

$$\int dx\ \mathbf{e}^{-1}(x) = \int dx'\ \mathbf{e}'^{-1}(x'), \qquad dx' = dx\ det\left(\frac{\partial x'}{\partial x}\right)$$

c Show that

$$det\left(\frac{\partial \tilde{x}}{\partial x}\right) = 1 \cdot e^{\overleftarrow{\lambda}}$$

by similarly evaluating

$$\int dx\ \mathbf{e}^{-1}(x) = \int d\tilde{x}\ \mathbf{e}^{-1}(\tilde{x}) = \int dx\ \mathbf{e}'^{-1}(x)$$

From the results of subsection IXA2, we then have that covariant derivatives can be integrated by parts in such actions, since

$$T_{ba}{}^{b} = 0 \quad \Rightarrow \quad \int dx\ \mathbf{e}^{-1}\nabla_a V^a = \int dx\ \partial_m \mathbf{e}^{-1} V^m$$

Exercise IXA4.3

Derive the expression for the covariant divergence in terms of $\mathbf{e}$ and the partial divergence by assuming integration by parts:

$$\int \mathbf{e}^{-1}\phi\nabla_a V^a = -\int \mathbf{e}^{-1} V^a \nabla_a \phi$$

Use this to find a simple form for the covariant d'Alembertian on a scalar:

$$\Box\phi \equiv \nabla^2\phi = \frac{1}{\sqrt{-g}}\partial_m\sqrt{-g}g^{mn}\partial_n\phi$$

Actions for matter are constructed in a similar way to Yang-Mills: Starting with the flat-space action, replace ordinary derivatives with covariant derivatives. The new ingredient is the extra factor of $\mathbf{e}^{-1}$. This prescription, as for Yang-Mills, is unambiguous up to only field-strength (curvature) terms, which can usually be eliminated by symmetry requirements and dimensional analysis. (At least for low energies, we want terms of the lowest mass dimension.) This uniqueness (at low energies or long distances) is known as the "equivalence principle": Inertial "mass" (really energy, but also momentum), as determined by the kinetic term, is the same as gravitational mass, as determined by the coupling of the gravitational field.

A simple example of matter is a real scalar field:

$$S = \int \mathbf{e}^{-1}\tfrac{1}{4}[(\nabla\chi)^2 + m^2\chi^2 + aR\chi^2]$$

The constant a can sometimes be fixed by symmetry: In the massless case, to preserve the global symmetry $\delta\chi = \epsilon$, we must have $a = 0$. (For self-interacting scalars, this generalizes to a global nonabelian symmetry.) To preserve conformal symmetry (see subsection IXA7), also for the massless case, we need $a = \frac{1}{4}\frac{D-2}{D-1}$.

This form of actions in terms of scalar Lagrangians also suggests we modify the definition of functional variation for convenience and covariance:

$$\delta S = \int dx\ \mathbf{e}^{-1}(\delta\phi)\frac{\delta S}{\delta\phi}$$

or, equivalently, we use the covariant form of the δ function,

$$\frac{\delta\phi(x)}{\delta\phi(x')} = \mathbf{e}(x)\delta(x-x')$$

As in flat space, the action for electromagnetism follows from gauge invariance:

$$S = \tfrac{1}{8e^2}\int \mathbf{e}^{-1}F_{ab}^2 = \tfrac{1}{8e^2}\int \sqrt{-g}g^{mn}g^{pq}F_{mp}F_{nq}$$

where $F_{mn} = \partial_{[m}A_{n]}$. Integration by parts then gives a simple form for Maxwell's equations. Such simple covariant equations of motion that don't require explicit expressions for the Lorentz connection appear only for antisymmetric tensors (which in practice means just spin 0 and 1 in 4D).

Exercise IXA4.4

Methods related to differential forms can be applied to these special cases:

a Rewrite the above action for electromagnetism in terms of A_a and *covariant* derivatives. Find the field equations following from both forms of the action, and use this to find a simple expression for the covariant divergence of an antisymmetric tensor with curved indices using just the metric. Compare the results of the previous exercise.

b By converting flat indices on the covariant tensor ϵ_{abcd} to curved, show that

$$\sqrt{-g}\epsilon_{mnpq} \quad \textit{and} \quad \frac{1}{\sqrt{-g}}\epsilon^{mnpq}$$

are also covariant tensors. Use these, and the covariance of the curl (see subsection IC2), to arrive at the same expression for the covariant divergence of an antisymmetric tensor.

Another example is a Dirac spinor:

$$S = \int \mathbf{e}^{-1}\bar{\Psi}(\gamma^a i\nabla_a + \tfrac{m}{\sqrt{2}})\Psi$$

where γ^a are the usual constant Dirac matrices, in terms of which the spin operator appearing in ∇ is the usual $-M_{ab} \to S_{ab} = -\frac{1}{2}\gamma_{[a}\gamma_{b]}$. In 4D, we can rewrite this in spinor notation by simply replacing $\partial_{\alpha\dot\beta} \to \nabla_{\alpha\dot\beta}$ in the flat-space expressions given in subsection IIIA4, and replacing $M_{ab} \to M_{\alpha\beta}$ as described in subsection IXA1, as well as $\int d^4x \to \int d^4x\ \mathbf{e}^{-1}$.

5. Gravity

The Einstein-Hilbert action for gravity follows from choosing the only available scalar second-order in derivatives, the Ricci scalar:

$$L_G = -\tfrac{1}{4}R = -\tfrac{1}{4}R_{ab}{}^{ab}$$

This action normally has a coefficient of $1/\kappa^2$ (compare Yang-Mills), but we'll generally use (natural/Planck) units $\kappa = 1$; then κ is used only to parametrize expansion about the vacuum and define the weak-field limit. (Actually, Planck units normally use $G = 1$, whereas in our conventions $\kappa = 1 \to G = \pi$.) In any case, the κ's can always be absorbed (unlike Yang-Mills) by a field redefinition of $e_a{}^m$, and then appear only in the definition of the "vacuum" (perturbative ground state, or solution that defines the boundary conditions at infinity):

$$\langle e_a{}^m \rangle = \kappa^{2/(D-2)} \delta_a{}^m$$

This makes $e_a{}^m \partial_m$, and thus $dx^m e_m{}^a$ and ds^2, dimensionless. In this sense, gravity is a theory with "spontaneous breakdown" of conformal invariance: Coordinate transformations include conformal transformations, but this invariance is broken by the vacuum, which introduces a length scale (κ).

Exercise IXA5.1

Consider the covariant derivative for nonvanishing torsion. By solving for the Lorentz connection in terms of the structure functions and torsion, express the covariant derivative in terms of the torsion-free covariant derivative $\overset{\circ}{\nabla}$ and the torsion. Thus, any action in terms of ∇ can be rewritten in terms of $\overset{\circ}{\nabla}$ and T, so any theory with a nonvanishing torsion is equivalent to a similar one with vanishing torsion (assuming the action is only second-order in derivatives of the vierbein, and thus algebraic in the torsion). Take the commutator of two ∇'s to find the curvature in terms of the torsion-free curvature $\overset{\circ}{R}_{abcd}$. Write the Einstein-Hilbert action with nonvanishing torsion in terms of $\overset{\circ}{R}$, $\overset{\circ}{\nabla}$, and T, to find:

$$R = \overset{\circ}{R} - (T_{ab}{}^b)^2 - \tfrac{1}{2}T^{abc}T_{bca} + \tfrac{1}{4}T^{abc}T_{abc} - 2\overset{\circ}{\nabla}{}^a T_{ab}{}^b$$

Since the last term vanishes upon integration, T appears as an auxiliary field, so R is equivalent to just $\overset{\circ}{R}$.

Exercise IXA5.2

For some general applications, where the form of the vierbein is not specified, it is useful to have a more explicit expression for the action in terms of the vierbein. We found in subsection IXA2 that for vanishing torsion

$$\nabla_a V^a = \mathbf{e}\,\partial_m(\mathbf{e}^{-1} e_a{}^m V^a) = e_a V^a - c_{ab}{}^b V^a$$

Use this to show

$$R = (c_{ab}{}^b)^2 + \tfrac{1}{2}c^{abc}c_{bca} - \tfrac{1}{4}c^{abc}c_{abc} - 2\mathbf{e}\,\partial_m[e^{am}\partial_n(e_a{}^n\mathbf{e}^{-1})]$$

We can drop the last term in the action integral under appropriate boundary conditions. (Hint: Use the result of the previous exercise for $\omega = 0$.)

Exercise IXA5.3

In two dimensions there is a single Lorentz generator,

$$M_{ab} = \epsilon_{ab}M \quad \Rightarrow \quad \nabla_a = e_a + \omega_a M, \qquad [\nabla_a, \nabla_b] = -\tfrac{1}{2}\epsilon_{ab}RM$$

a Show that the connection and the only surviving part of the curvature then take the simple forms

$$\omega_a = -\epsilon_{ab}\mathbf{e}\partial_m\mathbf{e}^{-1}e^{bm}, \qquad R = -2\mathbf{e}\partial_m[e^{am}\partial_n(\mathbf{e}^{-1}e_a{}^n)] = -2\mathbf{e}^{-1}\overleftarrow{e}^a\overleftarrow{e}_a\mathbf{e}$$

b Derive, for the sphere in spherical coordinates,

$$e_a{}^m = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{sin\ \theta} \end{pmatrix}$$

(Hint: First use $ds^2 = dx^m dx^n g_{mn}$ in 3D flat space.) Then show the covariant derivative is

$$\nabla_\theta = \partial_\theta, \qquad \nabla_\phi = \frac{1}{sin\ \theta}\partial_\phi + cot\ \theta\ M_{\theta\phi}$$

c Use these results to calculate $\int dx\ \mathbf{e}^{-1}R$ for the sphere in two ways: (1) by showing R is a constant and pulling it out of the integral, and (2) by converting it into a boundary term, where the "boundary" consists of infinitesimal circles around the coordinate singularities at the poles. (In general, even for spaces without true boundaries, one has to treat the boundaries of patches as such.)

It's also possible to add a "cosmological term" to the gravitational action:

$$S_{cos} = \Lambda \int dx\ \mathbf{e}^{-1}$$

with the "cosmological constant" Λ. This term has no derivatives, and is thus analogous to a mass term. However, it only contributes to the nonpropagating spin-0 mode of the vierbein (see later), so it doesn't give a physical mass, but does modify the vacuum.

Exercise IXA5.4

Show that the action for gravity can be made *polynomial* in $e_a{}^m$ by a field redefinition (rescaling) of the form

$$e_a{}^m \to \mathbf{e}^k e_a{}^m$$

when k takes the values

$$k = -\frac{n+1}{D-2}, \qquad n = 2, 3, 4, ...$$

and that the resulting action is order $Dn + 2$ in the field. In what cases (of n and D) is the cosmological term also polynomial?

The variation of the curvature can be obtained directly by varying its definition in terms of $[\nabla, \nabla]$. We start with the definition

$$\delta e_a{}^m \equiv \zeta_a{}^b e_b{}^m \quad \Leftrightarrow \quad \zeta_a{}^b \equiv (\delta e_a{}^m) e_m{}^b$$

and work in terms of the flattened object ζ_{ab}. Then we drop its Lorentz piece, choosing $\zeta_{ab} = \zeta_{ba}$. We find:

$$\delta \nabla_a = \zeta_a{}^b \nabla_b + \tfrac{1}{2}\zeta_a{}^{bc} M_{cb}$$

$$\Rightarrow \tfrac{1}{2}(\delta R_{ab}{}^{cd}) M_{dc} = [\nabla_{[a}, \delta\nabla_{b]}] = (\nabla_{[a}\zeta_{b]}{}^c)\nabla_c - \zeta_{[a}{}^c \tfrac{1}{2} R_{b]c}{}^{de} M_{ed} + \tfrac{1}{2}\nabla_{[a}\zeta_{b]}{}^{cd} M_{dc} + \zeta_{[ab]}{}^c \nabla_c$$

$$\Rightarrow \quad \nabla_{[a}\zeta_{b]}{}^c + \zeta_{[ab]}{}^c = 0, \qquad \delta R_{ab}{}^{cd} = \nabla_{[a}\zeta_{b]}{}^{cd} - \zeta_{[a}{}^e R_{b]e}{}^{cd}$$

$$\Rightarrow \quad \zeta_{abc} = \nabla_{[b}\zeta_{c]a}$$

$$\Rightarrow \quad \delta R_{ab}{}^{cd} = \tfrac{1}{2}\{\nabla_{[a}, \nabla^{[c}\}\zeta_{b]}{}^{d]} - \tfrac{1}{2}(\zeta_{[a}{}^e R_{b]e}{}^{cd} + {}_{ab} \leftrightarrow {}^{cd})$$

$$\delta \mathbf{e}^{-1} = -\mathbf{e}^{-1}\delta \ ln\ \mathbf{e} = -\mathbf{e}^{-1} e_m{}^a \delta e_a{}^m = -\mathbf{e}^{-1}\zeta^a{}_a$$

$$\Rightarrow \quad \delta(\mathbf{e}^{-1}R) = 2\mathbf{e}^{-1}[(\eta^{ab}\Box - \nabla^a\nabla^b) + (R^{ab} - \tfrac{1}{2}\eta^{ab}R)]\zeta_{ab}$$

where $\Box \equiv \nabla^a\nabla_a$. Thus for pure gravity we have the field equations

$$\delta S_G = 0 \quad \Rightarrow \quad R_{ab} - \tfrac{1}{2}\eta_{ab}R = 0 \quad \Rightarrow \quad R_{ab} = R = 0$$

while with a cosmological constant we have

$$\delta S_G + \delta S_{cos} = 0 \quad \Rightarrow \quad R_{ab} - \tfrac{1}{2}\eta_{ab}(R - 4\Lambda) = 0 \quad \Rightarrow \quad R_{ab} - \tfrac{1}{D}\eta_{ab}R = 0, \quad R = 4\tfrac{D}{D-2}\Lambda$$

Note that calculating a variation is the same as performing a perturbation to lowest order: We will use this result in subsection IXB1.

Exercise IXA5.5

For gravity, a first-order formalism follows from not imposing the torsion constraint (see exercise IXA5.1), so either the torsion or the Lorentz connection can be treated as the auxiliary variable.

a Find a first-order action for gravity (in all D) by treating ${e_m}^a$ and ${\omega_m}^{ab}$ as the independent variables. In D=4, using ϵ^{mnpq}, write this action as polynomial in these variables, eliminating the explicit $\mathbf{e}$, to obtain

$$S_G = \int d^4x \ \tfrac{1}{16}\epsilon^{mnpq}\epsilon_{abcd}{e_m}^a{e_n}^b{R_{pq}}^{cd}$$

with ${R_{pq}}^{cd}$ in terms of just ω.

b Vary this action with respect to both e and ω (independently) to find the field equations, expressed in terms of torsion and curvature, using $\delta[\nabla_m, \nabla_n]$ to find the variation of ${R_{pq}}^{cd}$ (see subsection IXA2).

Exercise IXA5.6

As discussed in subsection IIIC4 for Yang-Mills, in four dimensions we can write a complex first-order action for gravity that yields the usual gravity action up to a surface term. For Yang-Mills, the complex action was obtained by starting with a normal first-order formalism and replacing the auxiliary field with its self-dual part.

a Starting with the first-order action of the previous problem, find the analog for gravity by keeping just the part of ${\omega_m}^{ab}$ self-dual in ab, using spinor notation.

b Associate the coupling κ with the term quadratic in ω (analogously to the Yang-Mills case). As for Yang-Mills, associate the self-dual theory with the limit $\kappa \to 0$. Find the equation for ${e_m}^a$ that follows from varying ω in this case, and show that it is equivalent to setting the self-dual part of ${\mathring{\omega}_m}{}^{ab}$ to zero, where $\mathring{\omega}$ is the usual torsion-free connection. Show this is equivalent to setting the self-dual part of the curvature ${R_{mn}}^{ab}$ to vanish, in an appropriate gauge. (Technically, this means we must either complexify the fields, or Wick rotate to 4+0 or 2+2 space+time dimensions, where the Lorentz group factorizes.)

6. Energy-momentum

In subsection IIIB4 we saw that in the same way as a current in electrodynamics or Yang-Mills is defined as the matter contribution to the gauge field's equation of motion, $\delta S_M/\delta A_a = J^a$ (in that case S_M excludes only the pure Yang-Mills action), the "energy-momentum tensor" is defined as the matter contribution to the gravitational field equation (in this case S_M excludes only the pure gravity action):

$$\delta S_M = \int \mathbf{e}^{-1}\zeta^{ab}T_{ab} = \tfrac{1}{2}\int \sqrt{-g}(\delta g^{mn})T_{mn} = -\tfrac{1}{2}\int \sqrt{-g}(\delta g_{mn})T^{mn}$$

The case where ζ_{ab} represents the invariances of the action implies restrictions on this tensor: Using the separate gauge invariance of the matter action $\delta_{gauge}S_M = 0$ and the matter field equations $\delta S_M/\delta(matter) = 0$ (as for the Yang-Mills case), gauge variation of the gravity fields in S_M implies

$$\zeta_{ab} = \begin{cases} \lambda_{ab} = -\lambda_{ba} & \Rightarrow \quad T^{[ab]} = 0 : \quad \text{Lorentz} \\ -\frac{1}{2}\nabla_{(a}\lambda_{b)} & \Rightarrow \quad \nabla_a T^{ab} = 0 : \quad \text{coordinate} \end{cases}$$

so coordinate invariance of the action implies local conservation of energy-momentum.

For example, for a real scalar field:

$$S = \int \mathbf{e}^{-1}\tfrac{1}{4}[(\nabla\chi)^2 + m^2\chi^2 + aR\chi^2]$$

$$\Rightarrow \quad 2T_{ab} = (\nabla_a\chi)(\nabla_b\chi) - \tfrac{1}{2}\eta_{ab}[(\nabla\chi)^2 + m^2\chi^2] + a[(\eta_{ab}\Box - \nabla_a\nabla_b) + (R_{ab} - \tfrac{1}{2}\eta_{ab}R)]\chi^2$$

Notice that for $a \neq 0$, the energy-momentum tensor gets extra total-derivative terms which are separately conserved in flat space (since they come from the $R\chi^2$ term, which is separately covariant).

Exercise IXA6.1

Using the action given in subsection IXA4 and the variation of the covariant derivative from subsection IXA5, find the energy-momentum tensor for the Dirac spinor, and use its field equations to show this tensor is conserved.

Note that this is not the same as ordinary conservation $\partial_m T^{mn} = 0$: $\int \sqrt{-g}T^{0n}$ does not define a conserved total energy-momentum. This is in contrast with the conserved current in electrodynamics, since we then can derive the usual global conservation law

$$0 = \int d^Dx\ \mathbf{e}^{-1}\nabla_a J^a = \int d^Dx\ \partial_m \mathbf{e}^{-1}J^m \sim \frac{d}{dt}\int d^{D-1}x\ \mathbf{e}^{-1}J^0$$

On the other hand, it's closely related to Yang-Mills, where $\delta A_a = -\nabla_a\lambda$ leads to $\nabla_a J^a = 0$ in terms of the derivative ∇ covariantized with respect to the Yang-Mills field (as well as gravity, if in curved space), so $\partial_m \mathbf{e}^{-1}J^m = -\mathbf{e}^{-1}[iA_m, J^m] \neq 0$ (see subsection IIIC1).

However, if there is a Killing vector K_a, then the component of momentum in that direction is conserved:

$$J^a \equiv K_b T^{ba} \quad \Rightarrow \quad \nabla_a J^a = (\nabla_a K_b)T^{ba} + K_b(\nabla_a T^{ba}) = 0$$

(Remember $\nabla_{(a}K_{b)} = 0$.) Some simple examples of this in flat space are $(K_a)^b = \delta_a^b$ (translational invariance), for which the corresponding "charge" is the total momentum, and $(K_a)^{bc} = \delta_a^{[b}x^{c]}$ (Lorentz invariance), for which the charge is the total angular momentum.

Including the variation of the gravitational action, we get the gravitational field equations

$$R_{ab} - \tfrac{1}{2}\eta_{ab}R = 2T_{ab}$$

Coordinate invariance of S_G implies $\nabla_a(R^{ab} - \frac{1}{2}\eta^{ab}R) = 0$, which also follows from the Bianchi identities: In that sense gauge invariance is said to be "dual" to Bianchi identities, one implying the other through variation of the action: In general, for any gauge field ϕ with gauge parameter λ

$$\delta\phi = \mathcal{O}\lambda, \qquad 0 = \delta S = \int dx\ (\mathcal{O}\lambda)\frac{\delta S}{\delta\phi} \quad \Leftrightarrow \quad \mathcal{O}^T\frac{\delta S}{\delta\phi} = 0$$

where the "transpose" $\mathcal{O}^T$ is defined by integration by parts. Positivity of the energy (contained in any infinitesimal volume) is the condition $T^{00} \geq 0$. The addition of the cosmological term modifies the left-hand side of the above equation of motion by adding a term $2\eta_{ab}\Lambda$.

Although there is no covariant definition of total energy-momentum, in the case where spacetime is asymptotically flat (the metric falls off to the flat metric sufficiently fast at infinity), one can define a noncovariant energy-momentum tensor t_{ab} for gravity itself which is covariant with respect to coordinate transformations that themselves fall off at infinity. (See exercise IIIC1.2 for the analogous Yang-Mills case.) This tensor satisfies $\partial_m(T^{mn} + t^{mn}) = 0$ (where T^{mn} is the usual tensor for matter), so the usual conservation laws can be derived for the total energy-momentum coming from integrating $T + t$. Many equivalent expressions exist for t. One way to derive it is to expand the field equations order-by-order in h as

$$\tfrac{1}{2}(R_{ab} - \tfrac{1}{2}\eta_{ab}R) \equiv L_{ab} - t_{ab}$$

where L_{ab} is the linearized part of the field equations (see subsection IXB1) and $-t_{ab}$ is the quadratic and higher-order parts. By the linearized Bianchi identities, we know

$$0 = \partial_a L^{ab} \equiv \partial_a(\tfrac{1}{2}R^{ab} - \tfrac{1}{4}\eta^{ab}R + t^{ab}) = \partial_a(T^{ab} + t^{ab})$$

where we used the field equations in the last step. Note that there is a great deal of ambiguity here: We could have linearized by expanding the metric around its flat space value instead of the vierbein, or by expanding R_{mn} or R^{mn} instead of R_{ab}, etc.

Because of the expression in terms of $L_{ab} \sim \partial\partial h$, the integral of $T+t$, which gives the total energy-momentum vector, can be expressed as a surface term, just as Gauss' law in electrodynamics. Since space was assumed to be asymptotically flat, only the quadratic part of t contributes in the surface integral, which is why there is so much freedom in the definition of t. Since t is not covariant, the energy-momentum of the gravitational field is not localized (coordinate transformations shift it around). However, since the total energy-momentum is invariant, one can ask questions about how much energy is radiated to infinity, etc.

7. Weyl scale

The simplest way to describe conformal transformations in field theory is as a local scale transformation. If the theory is not coupled to gravity, we couple it to gravity as in Yang-Mills theory by replacing a Poincaré invariant Lagrangian $L(\partial, \psi)$ with $L(\nabla, \psi)$ (where all fields ψ have flat indices), but also including the $\mathbf{e}^{-1}$ factor in the action. We then transform the fields as

$$e_a{}^m \to \Phi e_a{}^m, \qquad \psi \to \Phi^{w+(D-2)/2}\psi$$

where Φ is the gauge parameter and $w + \frac{D-2}{2}$ is the engineering dimension (scale weight) of the field ψ. (See subsection IIB1.) Effectively, $e_a{}^m$ has dimension 1, since it's the only field with curved indices, and thus any derivative must appear in the combination $e_a{}^m\partial_m$, while the measure appears as $dx\ \mathbf{e}^{-1}$. Of course, the action won't be locally scale invariant unless it is globally scale invariant, i.e., has only dimensionless coupling constants (and thus no masses).

If the gravity-coupled theory is invariant under this local scale transformation, then the theory will be conformally invariant after decoupling gravity. This follows from the fact that the most general combined coordinate and local scale and Lorentz transformation that preserves the flat-space vierbein $e_a{}^m = \delta_a^m$ is exactly a conformal transformation. This is equivalent to our previous definition in terms of the scaling of the flat-space ds^2 under conformal transformations, since $dx'^m dx'^n g'_{mn}(x') = dx^m dx^n g_{mn}(x)$ under coordinate transformations.

Exercise IXA7.1

Derive the usual conformal transformations by finding the most general local scale + Lorentz + coordinate transformation that preserves the flat-space vierbein.

A simple example is Yang-Mills theory. We look at the Yang-Mills field with curved index, since its gauge transformation does not depend on the vierbein. ($\delta A_m =$

$-\partial_m\lambda + ...$ vs. $\delta A_a = -e_a{}^m\partial_m\lambda + ...$) To avoid interference with the Yang-Mills gauge transformation, the Yang-Mills field with curved index must be scale invariant. Then the action

$$S = \tfrac{1}{8e^2}\int \mathbf{e}^{-1}e^{am}e^{bn}e_a{}^p e_b{}^q F_{mn}F_{pq}$$

transforms with a factor Φ^{4-D}, and so is invariant in $D = 4$ only.

Exercise IXA7.2

Consider a more general gauge field A and field strength F defined by

$$\delta A_{m_1\cdots m_N} = -\tfrac{1}{(N-1)!}\partial_{[m_1}\lambda_{m_2\cdots m_N]}, \qquad F_{m_1\cdots m_{N+1}} = \tfrac{1}{N!}\partial_{[m_1}A_{m_2\cdots m_{N+1}]}$$

where A is totally antisymmetric in its N indices. (Such theories were encountered in exercise IIB2.1b.)

a Define an action in terms of F^2. In what dimension $D(N)$ is it conformally invariant?

b Show that this theory is related by a "duality transformation" (switching Bianchi identities and field equations) to the theory with N' indices on a new A, where $N' = D - 2 - N$, and $D(N') = D(N)$.

c Examine the cases $N = D, D-1, D-2$. Note that the scalar obtained by duality does not have an $R\phi^2$ term in its action, and thus is conformal only in $D = 2$.

Gravity is not scale invariant, but it will prove useful to examine its scale breaking explicitly. To preserve gauge covariance and dimensional analysis, the scale transformation law of the covariant derivative must take the form

$$\nabla'_a = \Phi\nabla_a + k(\nabla^b\Phi)M_{ab}$$

where the Φ scaling of $e_a{}^m$ was defined above, and the linearity of $\delta\omega$ in Φ follows from the homogeneity of ∇ in e. (Alternatively, we could put in something more arbitrary, but it would be eliminated by the rest of the procedure anyway.) From the variation of commutation relations we then find

$$\tfrac{1}{2}R'_{ab}{}^{cd}M_{dc} = [\nabla'_a, \nabla'_b]$$

$$= \Phi^2[\nabla_a, \nabla_b] + (1-k)\Phi(\nabla_{[a}\Phi)\nabla_{b]} + k\Phi(\nabla_{[a}\nabla^c\Phi)M_{b]c} + k^2(\nabla\Phi)^2 M_{ab}$$

$$\Rightarrow \quad k = 1, \qquad R'_{ab}{}^{cd} = \Phi^2 R_{ab}{}^{cd} + \Phi\delta_{[a}^{[c}\nabla_{b]}\nabla^{d]}\Phi - \delta_{[a}^c\delta_{b]}^d(\nabla\Phi)^2$$

If we make the redefinition (at least for Φ positive)

$$\Phi = \phi^{-2/(D-2)}$$

then we find the very simple scaling law for the integrand of the Einstein-Hilbert action:

$$(\mathbf{e}^{-1}R)' = \mathbf{e}^{-1}(\phi^2 R - 4\tfrac{D-1}{D-2}\phi\Box\phi)$$

Exercise IXA7.3

Consider a scale factor that is invariant under a Killing vector (see subection IXA2).

a Show the Killing vector survives the scale transformation; i.e.,

$$\nabla'_a = \Phi\nabla_a + (\nabla^b\Phi)M_{ab}, \qquad [K,\nabla_a] = [K,\Phi] = 0 \quad \Rightarrow \quad [K,\nabla'_a] = 0$$

directly using commutators (rather than the Killing equations).

b Although the operator K is the same, the Killing vector is different:

$$K = K^a\nabla_a + \tfrac{1}{2}K^{ab}M_{ba} = K'^a\nabla'_a + \tfrac{1}{2}K'^{ab}M_{ba}$$

Find K'^a and K'^{ab} in terms of K^a and K^{ab}.

Exercise IXA7.4

Covariant derivatives for flat space in spherical coordinates can be obtained from those of Cartesian coordinates by a combination of coordinate and local Lorentz (rotation) transformations. However, there are simpler methods, using a combination of transformations of a single coordinate and Weyl scale transformations:

a Take the direct product of a sphere with metric $d\Omega^2$ (in arbitrary coordinates) and a line as

$$d\mathring{s}^2 = d(ln\ r)^2 + d\Omega^2$$

Then derive flat space in spherical coordinates by making a scale transformation

$$\Phi = \tfrac{1}{r}$$

to yield the metric

$$ds^2 = dr^2 + r^2 d\Omega^2$$

Show that the resulting covariant derivatives are

$$\nabla_r = \partial_r, \qquad \nabla_i = \tfrac{1}{r}(\mathring{\nabla}_i + M_{ri})$$

where $\mathring{\nabla}_i$ are the covariant derivatives on the sphere corresponding to the metric $d\Omega^2$.

b Find ∇ in terms of (r, θ, ϕ) using the result of exercise IXA5.3. Find ∇ in terms of r and conformally flat coordinates x^i for the sphere by first deriving

$$\overset{\circ}{\nabla}_i = (1 + \tfrac{1}{4}x^2)\partial_i + \tfrac{1}{2}x^j M_{ij}$$

from flat 2D space by another Weyl scaling.

Many special cases of covariant derivatives can be derived completely by Weyl scalings. This includes the most commonly used ones, for cosmology and for static spherical sources. The general procedure uses the following facts in the following order:

(1) In a space of one dimension, we can choose

$$D = 1 \quad \Rightarrow \quad \nabla = \partial$$

(There is no curvature in $D = 1$.)

(2) For a direct product space, i.e., where the metric ds^2 can be written as the sum of the metrics of two (or more) spaces, the problem for solving for the covariant derivatives is separable. We can divide up the components into the covariant derivative for one space and that of the other, each using only its own coordinates and flat indices (and thus Lorentz generators):

$$ds^2 = ds_1^2 + ds_2^2 \quad \Rightarrow \quad \nabla = (\nabla_1, \nabla_2)$$

and similarly for the curvature.

(3) Under a coordinate transformation, each component of the covariant derivative (and of the curvature) transforms as a scalar. We need only apply the redefinitions of the coordinates, including those that appear in the partial derivatives:

$$\nabla_a(x) \to \nabla_a(x')$$

(This actually applies the alternative $\tilde{x}$ definition of coordinate transformation of subsection IC2.)

(4) Under a Weyl scale transformation,

$$ds'^2 = \Phi^{-2}ds^2 \quad \Rightarrow \quad \nabla'_a = \Phi\nabla_a + (\nabla^b\Phi)M_{ab},$$

$$R'_{ab}{}^{cd} = \Phi^2 R_{ab}{}^{cd} + \Phi\delta_{[a}^{[c}\nabla_{b]}\nabla^{d]}\Phi - \delta_{[a}^{c}\delta_{b]}^{d}(\nabla\Phi)^2$$

These steps can then be repeated as necessary. (The first two steps alone lead to Cartesian coordinates for flat space.)

Exercise IXA7.5

Use this method (as opposed to that of exercise IXA5.3) to derive the covariant derivatives for the sphere in the usual spherical coordinates:

a Use steps (1) and (2) to find ∇ for the flat space with metric

$$ds^2 = du^2 + d\phi^2$$

b For step (3), apply the transformation

$$du = \frac{d\theta}{sin\ \theta} \qquad (u = ln\ tan\ \tfrac{\theta}{2})$$

c For step (4), use

$$\Phi = \frac{1}{sin\ \theta}$$

to get the usual metric and covariant derivatives for the (2-)sphere. We also note that exercise IXA7.4a is just a repetition of these steps, for a new 1D coordinate v which is redefined as $v = ln\ r$, with a new $\Phi = 1/r$.

Consider a field theory without gravity that has a conformally invariant action. Spontaneous breakdown of scale invariance produces a Goldstone boson for that symmetry, the "dilaton". Any theory can be made globally conformally invariant trivially by performing a local scale transformation and making the parameter the dilaton field.

The dilaton can also act as a Higgs field: If we couple the dilaton to conformal gravity (gravity with local Weyl scale invariance), the Higgs effect reduces conformal gravity to ordinary (Einstein) gravity. For example, if we introduce the dilaton into pure gravity by the local scale transformation above (in analogy to the Stückelberg model),

$$S_G = 4\tfrac{D-1}{D-2} \int dx\ \mathbf{e}^{-1} \tfrac{1}{4}\phi(\Box - \tfrac{1}{4}\tfrac{D-2}{D-1}R)\phi$$

Up to an (important) overall negative factor, this is the action for a conformal scalar. The dilaton field ϕ is a compensator for local scale transformations, and acts as a Higgs field for this gauge symmetry: By gauging it to its vacuum value $\langle\phi\rangle = \frac{1}{\kappa}$, we regain the usual form of the gravity action. (Alternatively, we can set $\langle\phi\rangle = 1$, and introduce κ through the proportionality constant in $\langle e_a{}^m\rangle \sim \delta_a{}^m$.) In this formalism, where we require the action to be locally scale invariant, the terms which were conformally invariant before coupling to gravity are easy to recognize: They're just the ones which have no ϕ-dependence. (This may require some field redefinition: typically rescaling the matter fields according to their weight as above.) The cosmological term becomes $S_{cos} = \int \mathbf{e}^{-1}\Lambda\phi^{2D/(D-2)}$, which is a conformal self-interaction term for a scalar.

Because what was the vierbein now appears only in the combination $e_a{}^m \to \phi^{-2/(D-2)} e_a{}^m$, there is now the local scale invariance

$$e_a{}^m \to \Phi e_a{}^m, \quad \phi \to \Phi^{(D-2)/2}\phi$$

since this transformation leaves the combination invariant. Gauge invariance of the matter action is then (using the infinitesimal parameter $\Phi = 1 + \zeta$):

$$0 = \delta S_M \sim e_a{}^m \frac{\delta S_M}{\delta e_a{}^m} + \tfrac{D-2}{2}\phi \frac{\delta S_M}{\delta \phi}$$

$$\Rightarrow \quad T^a{}_a = -\tfrac{D-2}{2}\phi \frac{\delta S_M}{\delta \phi}$$

Thus, conformal matter has vanishing $T^a{}_a$, since it decouples from ϕ. (Actually, we also need to scale the matter as above to achieve this decoupling, and there is a corresponding $\delta S_M/\delta\psi$ term in the above derivation, so the trace may vanish only after applying the matter field equations, as in the derivation of $\nabla_a T^{ab} = 0$ from coordinate invariance in the previous subsection.) In particular, this is easy to check for the massless point particle, where $T^a{}_a \sim \dot{X}^m \dot{X}^n g_{mn} = 0$.

An interesting effect is obtained by eliminating the compensator by its field equation. (We'll consider just the classical theory here: In the quantum case, integrating out this field produces an additional 1-loop contribution to the effective action.) Because this manipulation involves integration by parts, we first expand the compensator about its vacuum (asymptotic) value:

$$\phi = 1 + \tfrac{1}{2}\chi \quad \Rightarrow \quad L = \tfrac{1}{4}[\chi(\tfrac{D-1}{D-2}\Box - \tfrac{1}{4}R)\chi - R\chi - R]$$

Then eliminating χ by its field equation,

$$L \to \tfrac{1}{4}\left(R \frac{1}{R - 4\frac{D-1}{D-2}\Box} R - R \right)$$

This action still describes Einstein gravity, but is locally scale invariant (though not globally, because of the extraction of the vacuum value, and the way boundary terms were neglected). Of course, it is nonlocal, and the nonlocality becomes more complicated if nonconformal matter is included. Such terms also appear quantum mechanically: In two dimensions, dimensionally regularizing D=2+2ϵ, in a Weyl scale invariant theory we can get a divergent, yet still Weyl scale invariant, contribution to the effective action proportional to

$$\frac{1}{\epsilon}\left(R \frac{1}{R - 4\frac{D-1}{D-2}\Box} R - R \right) \approx -\frac{1}{\epsilon} R - \tfrac{1}{2} R \frac{1}{\Box} R$$

After renormalizing the divergent term, which is topological and thus locally scale invariant in exactly D=2, but not in D=2+2ϵ, the remaining finite term contributes a conformal anomaly (see subsections VIIIA8 and C1).

Exercise IXA7.6

The statement that the R term is topological in D=2 neglects boundaries. In general the topological invariant (the "Euler number") is (the "Gauss-Bonet theorem")

$$\chi = \int \frac{d^2x}{2\pi}\ \tfrac{1}{2}\mathbf{e}^{-1}R + \oint \frac{1}{2\pi}\frac{\epsilon_{ab}t^a Dt^b}{\eta_{ab}t^a t^b}$$

where t^a is a tangent vector to the boundary $X^m(\tau)$, as for the worldline of the particle, and D is the covariant differential (as for the particle equation of motion and the radial gauge; see subsections IXB2 and 4 below):

$$t^m = v^{-1}\dot{X}^m, \qquad Dt^a = dX^m e_m{}^b \nabla_b t^a = d\tau\ vt\cdot\nabla t^a = d\tau(\dot{t}^a - vt^b t^c \omega_{bc}{}^a)$$

(We have used the usual counterclockwise contour, and our convention $\epsilon_{01} = 1$, or $\epsilon_{xy} = 1$ in Euclidean space.) The additional term in χ is the angle subtended by the boundary with respect to the surface $(/2\pi)$, as obtained from the cross product of t and $t + Dt$. We have written it in a form that is manifestly invariant under the reparametrization of τ, and the v's cancel. (Of course, it is also manifestly coordinate invariant.)

a Prove that it is also scale invariant by showing that the connection part of the D exactly cancels the contribution of R to the boundaries, leaving

$$\chi = \left(\int \frac{d^2x}{2\pi}\ \tfrac{1}{2}\mathbf{e}^{-1}R\right)\Bigg|_{patch\ boundaries} + \oint \frac{1}{2\pi}\frac{\epsilon_{ab}t^a dt^b}{\eta_{ab}t^a t^b}$$

where we have turned the R term into a boundary term, and its remaining contribution is from the fake boundaries at the borders of patches (or surrounding singularities; $R = \partial\omega$ because the 2D Lorentz group is Abelian: see exercise IXA5.3).

b Note that the dt term doesn't contribute if we choose a gauge where

$$t^a = \delta_1^a$$

(i.e., $t^m = e_1{}^m$). Demonstrate this by evaluating χ in polar coordinates for a disk, and in spherical coordinates for the half-sphere. Show the result is half that for a whole sphere (exercise IXA5.3). Repeat the calculation for the disk in Cartesian coordinates (so then *only* the dt term contributes).

REFERENCES

1 A. Einstein, *Sitz. Preuss. Akad. Wiss. Berlin, Math.-phys. Kl.* (1914) 1030, (1915) 778, 799, 831, 844, *Ann. der Phys.* **49** (1916) 769:
general relativity.

2 É. Cartan, *Leçons sur la géométrie des espaces de Riemann* (Gauthier-Villars, 1928):
general relativity in terms of vierbein and (GL(D) or Lorentz) connection.

3 Weyl, *loc. cit.* (IC):
covariant derivatives on spinors.

4 L. Infeld and B.L. van der Waerden, *Sitz. Preuss. Akad. Wiss. Berlin, Math.-phys. Kl.* (1933) 380:
general relativity in two-component spinor notation.

5 H. Weyl, *Mat. Z.* **2** (1918) 384:
Weyl tensor.

6 D. Hilbert, *Nachrichten Königl. Ges. Wiss. Göttingen, Math.-phys. Kl.* (1915) 395:
found the action for Einstein's equations slightly earlier than Einstein.

7 A. Palatini, *Rend. Circ. Mat. Palermo* **43** (1919) 203:
first first-order action for gravity (in terms of metric and Christoffel symbols).

8 H. Weyl, *Sitz. Preuss. Akad. Wiss. Berlin, Math.-phys. Kl.* (1918) 465,
Raum·zeit·materie (Springer, 1919) p. 246 [English: *Space-time-matter* (Dover, 1952)]:
Weyl scale.

9 B. Zumino, Effective Lagrangians and broken symmetries, in *Lectures on elementary particles and quantum field theory*, proc. 1970 Brandeis University Summer Institute in Theoretical Physics, eds. S. Deser, M. Grisaru, and H. Pendleton (MIT, 1970) v. 2, p. 437:
dilaton.

10 E.S. Fradkin and V.I. Vilkovisky, *Phys. Lett.* **73B** (1978) 209:
nonlocal action.

11 C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, 1970), 1279 pp.:
introductory, long-winded.

12 S.W. Hawking and G.F.R. Ellis, *The large-scale structure of spacetime* (Cambridge University, 1973), 400 pp.:
mathematical; emphasis on singularity theorems and global properties (e.g., Penrose diagrams).

13 R.M. Wald, *General relativity* (University of Chicago, 1984), 492 pp.:
intermediate between the above two.

14 S. Weinberg, *Gravitation and cosmology* (Wiley & Sons, 1972), 657 pp.:
old-fashioned, but with more detail on astrophysics and cosmology.

B. GAUGES

We now consider various gauge choices for coordinate, Lorentz, and scale transformations.

1. Lorentz

We begin with gauges that preserve global Lorentz invariance, which are useful for perturbation theory. Therefore, we look first at perturbation by finding the kinetic term, which is sufficent for finding linear gauge conditions. (It can also be derived from general principles, as will be shown in subsection XIIA5.) We expand the vierbein about its flat value,

$$e_a{}^m = \delta_a{}^m + h_a{}^m$$

At the linearized level, local Lorentz invariance implies that only the symmetric part of the field, $\frac{1}{2}h_{(ab)}$, appears in the curvature and the action; we will denote this by h_{ab} to simplify notation. (In other words, the linearized curvature is invariant under the linearized local Lorentz transformations, which gauge away the antisymmetric part of the field. This is equivalent to working directly with the metric.) We then can find the linearized curvature, e.g., from the results of subsection IXA5 for the variation of the curvature, by considering variation about flat space: i.e., replacing $\zeta_{ab} \to h_{ab}$ and $\nabla_a \to \partial_a$. The result is

$$R_{ab}{}^{cd} \approx \partial_{[a}\partial^{[c}h_{b]}{}^{d]}$$

$$\Rightarrow \quad R_{ab} - \tfrac{1}{2}\eta_{ab}R \approx \Box h_{ab} + \partial_a\partial_b h^c{}_c - \partial_{(a}\partial^c h_{b)c} - \eta_{ab}(\Box h^c{}_c - \partial^c\partial^d h_{cd})$$

Since this comes from varying the action, the quadratic part of the gauge-invariant action must be

$$S_G \approx -\int \tfrac{1}{4}[h^{ab}\Box h_{ab} + 2(\partial^b h_{ab})^2 - h^a{}_a\Box h^b{}_b + 2h^a{}_a\partial^b\partial^c h_{bc}]$$

This part of the action, and the linearized curvature, are invariant under the linearized gauge transformations $\delta h_{ab} = -\partial_{(a}\lambda_{b)}$.

Exercise IXB1.1

Take the Newtonian (weak-field, nonrelativistic) limit of gravity: (1) Linearize the action by perturbing about flat space ($e_a{}^m = \delta_a^m + h_a{}^m$). Keep just the part of the pure gravity action quadratic in the perturbation, the part of the matter coupling linear in it, and the complete flat-space matter action. (2) Assume small velocities. Now consider the problem of a massive point particle in the field of a much more massive point particle (or spherical body); in the above approximations:

a Show the effect of the gravitational field generated by the heavier particle on the lighter particle is given by the action for the lighter particle (in the gauge $x^0 \equiv t = \tau$)

$$S = ms \approx \int dt \ (m - \tfrac{1}{2} m\dot{x}_i^2 + mh_{00})$$

b Show this field is given by solving Laplace's equation

$$R_{00} \approx \Delta h_{00} = T_{00}$$

c Show that, with our conventions for normalizing functional differentiation, a point mass M in D=4 generates

$$T_{00} = M(2\pi)^2 \delta^3(x) \quad \Rightarrow \quad h_{00} = -\frac{M\pi}{r}$$

using the usual solution to Laplace's equation for a point source. Combining these results, we see that the potential energy for the particle is

$$U = mh_{00} = -\frac{Mm\pi}{r}$$

which agrees with Newtonian gravity if we identify $G = \pi$. (If we restore units, this becomes $G = \kappa^2\pi$.)

The BRST transformations (see subsection VIA4) for gravity again follow from the gauge transformations:

$$Qe_a{}^m = C^n \partial_n e_a{}^m - e_a{}^n \partial_n C^m + C_a{}^b e_b{}^m$$

$$QC^m = C^n \partial_n C^m, \quad QC_{ab} = C^n \partial_n C_{ab} + C_a{}^c C_{cb}$$

$$Q\tilde{C}^m = -iB^m, \quad Q\tilde{C}_{ab} = -iB_{ab}$$

(Other forms follow from different parametrizations of the gauge transformations, and are equivalent to field redefinitions. For theories without spinors, we can work in terms of the metric, and avoid Lorentz gauge fixing.)

Lorentz gauges for coordinate invariance are similar to Yang-Mills. For gravity, the gauge-fixing function is

$$f_a = \partial^b h_{ab} - \tfrac{1}{2}\partial_a h^b{}_b$$

The BRST procedure works similarly to Yang-Mills. Looking at just the graviton kinetic term, the gauge-fixed quadratic Lagrangian for gravity is then, in the Fermi-Feynman gauge,

$$L_G \ \to \ L_{G,FF} = L_G + \tfrac{1}{2}(\partial^b h_{ab} - \tfrac{1}{2}\partial_a h^b{}_b)^2 = -\tfrac{1}{4} h^{ab} \Box h_{ab} + \tfrac{1}{8} h^a{}_a \Box h^b{}_b$$

plus ghost terms. Note that the trace part of h appears with opposite sign to the traceless part. This prevents any redefinition which would allow rewriting the Lagrangian in the simple form $-\frac{1}{4}h^{ab}\Box h_{ab}$. However, all derivatives have been absorbed into $\Box$'s, which makes the linearized field equation a simple Klein-Gordon equation.

There are various generalizations of this gauge condition to include nonlinear terms, such as the "de Donder (harmonic) gauge", which uses the gauge-fixing function

$$f^n = \tfrac{1}{2}\partial_m(\sqrt{-g}g^{mn})$$

For example, this allows the field equation for a scalar to be written with only terms with both partial derivatives acting on the scalar.

2. Geodesics

Consider the field equations for coupling gravity and electromagnetism to a scalar particle: From subsection IIIB3, the action for a particle in external fields, rewritten in Hamiltonian form, is

$$S_H = \int d\tau\{-\dot{x}^m e_m{}^a(x)[\pi_a - A_a(x)] + vH\}, \qquad H = \tfrac{1}{2}\pi^2 + \phi(x)$$

where we have pulled the v out of H for convenience, and use the "covariant momentum"

$$\pi_a = e_a{}^m p_m + A_a(x) = e_a{}^m(p_m + A_m)$$

in place of p_m (the canonical conjugate to x^m) for covariance. All the equations of motion except the Lagrange-multiplier constraint

$$\tfrac{1}{2}\pi^2 + \phi = 0$$

follow from the usual Poisson-bracket relation

$$v^{-1}\dot{\mathcal{O}} = i[H, \mathcal{O}]$$

which can be evaluated by using the canonical commutation relations (following from the simpler ones for p_m)

$$i[\pi_a, x^m] = e_a{}^m, \qquad i[\pi_a, \pi_b] = c_{ab}{}^c\pi_c + F_{ab}, \qquad [x, x] = 0$$

Thus, π_a acts effectively like $-ie_a + A_a$, which is the covariant derivative for gravity and electromagnetism, less the Lorentz term. The $\dot{x}$ equation is the obvious

$$v\pi^a = \dot{x}^m e_m{}^a$$

that follows from varying S_H with respect to π_a, while the equation of motion for π is

$$v^{-1}\dot{\pi}_a = -c_a{}^{bc}\pi_b\pi_c - F_a{}^b\pi_b - \nabla_a\phi$$

Using the relation

$$0 = T_{abc} = c_{abc} + \omega_{[ab]c} \quad \Rightarrow \quad c_{a(bc)} + \omega_{(bc)a} = 0$$

we find

$$v^{-1}\dot{\pi}^a - \pi^b\pi^c\omega_{bc}{}^a + F^a{}_b\pi^b + \nabla^a\phi = 0$$

This is the coordinate-covariant form of the Lorentz force law (plus scalar field). With only the gravitational effects we have the covariantization of the free particle equation,

$$Dp^a \equiv \dot{p}^a - vp^bp^c\omega_{bc}{}^a = 0$$

where "D" is understood as a covariantized τ derivative (along a worldline with metric v).

It's useful to consider a continuum of particles ("dust") moving under the influence of these fields, such that any two infinitesimally close particles have infinitesimally different velocities, and only one particle passes through any particular point in spacetime (at least within some small region of spacetime). We then can treat π_a (or p_m) as a field defined for all x: Choosing a point x also chooses a curve $X(\tau)$ for which $x = X(\tau)$ for some τ, so we can write $\pi(x)$ in place of $\pi(\tau)$. Specifying the field π also determines this family of curves, since the tangent to any curve is given by the X equation of motion

$$\dot{X}^m = v\pi^a e_a{}^m$$

(To determine the τ parametrization, we also specify v, and the hypersurface given by the collection of points $X(0)$ from each curve.) Then we can express the τ derivative in terms of x derivatives:

$$\frac{d}{d\tau} = \dot{X}^m\partial_m = v\pi^a e_a$$

which gives the manifestly covariant form of the equation of motion

$$\pi^b\nabla_b\pi_a + F_a{}^b\pi_b + \nabla_a\phi = 0$$

For vanishing F (and thus A) and constant ϕ $(= \frac{1}{2}m^2)$, this equation

$$p^b\nabla_b p_a = 0$$

describes "geodesics", which are curves of extremal length, since the action is

$$S = ms, \qquad -ds^2 = dx^m dx^n g_{mn}$$

for massive particles. These are the analogs of straight lines in flat space. (For positive-definite metric, they are shortest lines. Because of the indefinite signature of the Minkowski metric, the worldlines of massive particles are actually longest lines, while massless particles travel along lines with no length.)

Exercise IXB2.1

Uniform circular motion in 2D flat Euclidean space, constant $d\theta/dt$ in polar coordinates, is associated with acceleration of constant magnitude. (Or, without time, we can say that a circle has constant "extrinsic" curvature with respect to the 2D space.) Show that an analogous situation in 2D Minkowski space can be obtained by Wick rotation:

a Starting with the metric for 2D flat Euclidean space in polar coordinates, Wick rotate θ to make it a time coordinate ("Rindler coordinates"). Show by a transformation to "Cartesian" coordinates that this describes 2D flat Minkowski space.

b Show that any curve described by constant r describes acceleration of constant magnitude, by evaluating $(d^2x(\theta)/ds^2)^2$ in "Cartesian" coordinates. Note that the direction of this 2-vector is fixed to be orthogonal to dx/ds (since $(dx/ds)^2 = 1$ by definition), so this is just the acceleration as measured in the rest frame.

c Define the acceleration in arbitrary curved coordinates (in terms of $p\cdot\nabla p$) and evaluate it in Wick-rotated polar coordinates, to obtain the same result as in Cartesian coordinates. (Use the covariant derivative of exercise IXA7.4a.)

Exercise IXB2.2

Equations of motion for particles can be derived from conservation laws. We know this already nonrelativistically, for a particle in a potential using energy conservation. Now consider a dust with $T^{mn} = \rho\pi^m\pi^n$ and current $J^m = \rho\pi^m$. (Compare subsection IIIB4. We could use those single-particle expressions here, but using dust instead avoids integration. Note that using π or p allows us to describe also massless particles. The existence of a conserved current corresponds to a complex field with a global U(1) symmetry.)

a In the case with no external fields except gravity, show that the geodesic equation follows from covariant conservation of both of these quantities. (Of course, in flat space this gives the usual free particle result.)

b Generalize to the case of external fields by adding to T^{mn} that of the external fields themselves. When taking the divergence of those terms, use appropriate source terms to the field equations in terms of the particle variables ρ and π. (In the case of a single nonrelativistic particle in a static electric field, this is the usual derivation of the force on a particle from the electric field's pressure.)

For some purposes we need a weaker (but equivalent) form of the geodesic equation: If for some scalar f and vector n^a

$$(n\cdot\nabla)n^a = fn^a \quad \Rightarrow \quad p^a = un^a, \qquad (p\cdot\nabla)p^a = 0, \qquad f = -(n\cdot\nabla)ln\ u$$

for some scalar u, which we can determine by integrating f. In particular, we can identify

$$u = v^{-1} \quad \Rightarrow \quad n^m = \dot{X}^m$$

Thus, the more general geodesic equation allows arbitrary parametrization of the geodesics, while the stricter version ($f = 0$) corresponds to affine parametrization ($v = 1$) if we still want to identify p with $\dot{X}$. (Remember, as with all constrained systems, the equations of motion $p\cdot\nabla p = 0$ imply $(d/d\tau)p^2 = 0$, so any geodesic satisfying the stricter equation will have some fixed mass along that particular curve.)

Exercise IXB2.3

Show that in D=2 (one space dimension, one time) *any* lightlike curve is a geodesic, using the weaker form of the geodesic equation. (Find f.) This is a consequence of the fact that it is impossible to change direction in D=2 without slowing down.

The particle (geodesic) version of the conservation of momentum in the direction of a Killing vector is

$$p\cdot\nabla\ p_a = 0 \quad \Rightarrow \quad p\cdot\nabla\ K\cdot p = 0 \quad \Rightarrow \quad \frac{d}{d\tau}K\cdot p = 0$$

where covariant conservation $p\cdot\nabla$ has become ordinary conservation $d/d\tau$ (no connection term) because $K\cdot p$ is a scalar. (See also exercise IXA2.4.) This is the same as for the conserved current $J^a = K_b T^{ba}$ (subsection IXA6).

3. Axial

The definition of axial gauges in terms of the covariant derivative is the same as for Yang-Mills ($n\cdot\nabla = n\cdot\partial$). In terms of the explicit gravity fields,

$$n\cdot\nabla = n\cdot\partial \quad \Rightarrow \quad n^m \equiv n^a e_a{}^m = n^a\delta_a^m, \quad n^a\omega_a{}^{bc} = 0$$

In the case of gravity, this implies that lines in the n^a direction are geodesics (see previous subsection):

$$\partial_m n^a = 0 \quad \Rightarrow \quad (n\cdot\nabla)n = (n\cdot\partial)n = 0$$

To analyze the consequences of axial gauge conditions for the metric, we need a further identity: For any vector field n^a, consider the action of $n\cdot\nabla$ on $n_m = e_m{}^a n_a$, treating it as a scalar; in this calculation we ignore any indirect action of ∇ on curved indices. Then

$$(n\cdot\partial)e_m{}^a n_a = (n\cdot\nabla)e_m{}^a n_a = e_m{}^a(n\cdot\nabla)n_a + n_a(n\cdot\nabla)e_m{}^a$$

The last term simplifies for vanishing torsion, since:

$$\begin{aligned} n^n\nabla_n e_m{}^a &= n^n\nabla_{[n}e_{m]}{}^a + n^n\nabla_m e_n{}^a = -n^n T_{nm}{}^a + n^n\nabla_m e_n{}^a = \nabla_m n^a - e_n{}^a\nabla_m n^n \\ &= \nabla_m n^a - e_n{}^a\partial_m n^n \end{aligned}$$

We thus have

$$(n\cdot\partial)e_m{}^a n_a = e_m{}^a(n\cdot\nabla)n_a + \partial_m(\tfrac{1}{2}n_a^2) - (e_n{}^a n_a)\partial_m n^n$$

Applying this identity to the axial gauge condition, we find

$$n\cdot\nabla = n\cdot\partial, \qquad \partial_m n^a = 0 \quad \Rightarrow \quad (n\cdot\partial)e_m{}^a n_a = 0 \quad \Rightarrow \quad n_m \equiv e_m{}^a n_a = \delta_m^a n_a$$

by choosing the appropriate constants of integration. (This amounts to fixing a residual gauge invariance.) We can now determine the form of the gauge condition on the metric:

$$n^m = n^a\delta_a^m, \qquad n_m = \delta_m^a n_a \quad \Rightarrow \quad n_m \equiv n^n g_{nm} = n^n\eta_{nm}$$

Applying these results to perturbation theory, as

$$e_a{}^m = \delta_a^m + h_a{}^b\delta_b^m$$

we then have

$$n^a e_a{}^m = n^a\delta_a^m \quad \Rightarrow \quad n_b h^{ba} = 0$$

$$e_m{}^a n_a = \delta_m^a n_a \quad \Rightarrow \quad e_a{}^m\delta_m^b n_b = n_a \quad \Rightarrow \quad n_b h^{ab} = 0$$

and thus $n_b h^{(ba)} = 0$, so we can again work with just the symmetrized h.

The lightcone gauge is again useful for eliminating unphysical degrees of freedom. The lightcone gauge conditions are

$$n^a = \delta_-^a \quad \Rightarrow \quad h^{+a} = h^{a+} = h^{(+a)} = 0$$

For the rest of this discussion we work with just the symmetrized h. Separating out the trace part as $h_{ij} = h^T_{ij} + \delta_{ij} h$, where h^T_{ij} is traceless, we find for the linearized gauge-fixed action

$$\begin{aligned} L'_G &\approx -\tfrac{1}{4} h^{ab} \Box h_{ab} + \tfrac{1}{8} h^a{}_a \Box h^b{}_b - \tfrac{1}{2} f^2 \\ &= -\tfrac{1}{4} h^{Tij} \Box h^{Tij} - \tfrac{1}{2} (h'^{-i})^2 + \tfrac{D-2}{2} h h'^{--} \end{aligned}$$

$$\begin{aligned} h'^{-i} &\equiv f^i = -\partial^+ h^{-i} + \partial^j h^{Tij} - \tfrac{D-4}{2} \partial^i h \\ h'^{--} &\equiv \partial^+ f^- + \tfrac{D-4}{4} \Box h = -\partial^{+2} h^{--} + \partial^+ \partial^i h^{-i} - \tfrac{D-2}{2} \partial^+ \partial^- h + \tfrac{D-4}{4} \Box h \\ \tfrac{D-2}{2} h &= -\frac{1}{\partial^+} f^+ \end{aligned}$$

where we have simplified some algebra by writing the gauge-invariant action as the Lorentz gauge one *minus* its gauge-fixing term $\frac{1}{2} f^2$. (There is some ambiguity in that we can shift h'^{-i} by a $\partial^i h$ term, and absorb the generated terms into h'^{--}.) We see that all but h^T_{ij} are auxiliary fields (we redefined h^{-i} and h^{--} by just shifting and applying ∂^+), and can be eliminated (but watch out if there are matter couplings, when eliminating them gives Coulomb-like interactions).

The temporal gauge (known also as "Gaussian normal coordinates") is used when treating time and space separately: In this case we have for the metric

$$n^m = \delta_0^m \quad \Rightarrow \quad g_{0m} = \eta_{0m}$$

An alternate way of defining the temporal gauge is to start with a spatial hypersurface, and determine the geodesics normal to this hypersurface ($g_{0i} = 0$), where the positions on this hypersurface define x^i, constant along the geodesics, and the proper times along the geodesics define x^0 ($g_{00} = -1$), with $x^0 = 0$ at the hypersurface. The fact that these are geodesics guarantees that the hypersurfaces of fixed, but nonvanishing, (proper) time are still orthogonal to the geodesics (g_{0i} stays zero):

$$n \cdot V = 0, \qquad (n \cdot \nabla) n = 0 \quad \Rightarrow \quad n \cdot [(n \cdot \nabla) V] = 0$$

Equivalently, we can consider a dust of massive particles and choose an initial hypersurface orthogonal to their (timelike) geodesics to define $x^0 = s = 0$. This coordinate system is thus the "rest frame" of the dust; all the information about the geometry of the space is contained in the time dependence of the spatial separation of the particles (g_{ij}). There is still the residual coordinate ambiguity of how to assign x^i on the initial hypersurface.

Gaussian normal coordinates thus can be useful for studying the dynamics of particles: For example, we can study a gravitational field of distant, unknown (or ignored) origin (i.e., the curvature of spacetime) by watching the relative motion of two nearby particles of such a dust, neglecting the gravitational force/curvature effect acting between the two particles themselves. If the two particles start out relatively at rest at some initial time (which is well-defined only if they are close and relatively slow), then in the temporal gauge the paths of both particles are described by fixed x^i, independent of x^0, since their geodesics are simply lines in the time ($n^a = \delta_0^a$) direction, and the proper times of both particles are the same as the time x^0. Then the distance between the particles at any given time is given by the magnitude of $dx^m e_m{}^a$, with $dx^0 = 0$ and dx^i their infinitesimal separation. Thus, since the x^i's, and thus dx^i, are fixed, we want to study the change in $e_m{}^a$ (really just $e_i{}^a$; $e_0{}^a = \delta_0^a$) with time. Using our evaluation of $(n \cdot \nabla)e_m{}^a$ from above, we find

$$(n \cdot \nabla)^2 e_m{}^a = (n \cdot \nabla)\nabla_m n^a = [n \cdot \nabla, \nabla_m]n^a = n^n[\nabla_n, \nabla_m]n^a = -n^b n^c R_{bdc}{}^a e_m{}^d$$

using the axial gauge condition $n \cdot \nabla = n \cdot \partial$. For the Gaussian case $n^a = \delta_0^a$, we then have

$$\ddot{e}_m{}^a = -R_{0b0}{}^a e_m{}^b$$

(Of course, vanishing curvature implies geodesics that start parallel remain that way, because the space is then flat.) By observing different sets of particles initially at rest with respect to each other, we can choose different timelike directions n, and determine all the curvature components from their linear combinations.

Exercise IXB3.1

Let's examine some 2D examples of axial gauges in spaces with positive-definite metric:

a Gaussian normal coordinates need not be Cartesian in flat space. Show that polar coordinates for the plane define an axial gauge. What is the coordinate in the "n^a" direction? Give the geodesic interpretation.

b Repeat the above for a curved space — the (2D) sphere in spherical coordinates.

c Apply the above "equation of motion" ($\ddot{e} = -Re$) to the sphere. (See exercise IXA5.3.) Show its solution agrees with the obvious.

4. Radial

Another useful gauge similar to the axial gauge is the radial gauge ("Riemann normal coordinates"), discussed for Yang-Mills in subsection VIB1. In this case we have

$$n^m = x^m \quad \Rightarrow \quad (n \cdot \nabla) n^a = x^m \partial_m x^n \delta_n^a = n^a$$

a case of the more general form of the geodesic equation. Applying the same identity as for the axial, we again have

$$n\cdot\nabla = n\cdot\partial, \qquad \partial_m n^a = \delta_m^a \quad \Rightarrow \quad (n\cdot\partial) e_m{}^a n_a = (n\cdot\partial)\delta_m{}^a n_a \quad \Rightarrow \quad e_m{}^a n_a = \delta_m^a n_a$$

but now the boundary condition is already implied by the gauge condition near the origin: For any infinitesimal $x^m = \epsilon^m$,

$$\epsilon^m e_m{}^a(0) = \epsilon^m \delta_m^a \quad \Rightarrow \quad e_m{}^a(0) = \delta_m^a$$
$$\epsilon^m \omega_m{}^{ab}(0) = 0 \quad \Rightarrow \quad \omega_m{}^{ab}(0) = 0$$

Thus, there is no residual gauge invariance, unlike axial gauges (where the coordinates of the initial hypersurface need additional determination). Any reference frame satisfying these conditions at the origin is called a "local inertial frame", and is the most natural for an observer at that point in spacetime. (In flat space, this yields Cartesian coordinates.)

Exercise IXB4.1

We can think of Gaussian normal coordinates as defined by a dust of particles with affine parametrization $v = 1$ and unit mass $m = 1$, with $\tau = s = x^0$ and x^i constant for any particle ($\dot{X} = p = n$). For Riemann normal coordinates we can think of particles radiating out from the origin $x^m = 0$ in all possible directions in space and time, but then some must be antiparticles (traveling backward in time), some must be massless (for the lightlike geodesics), and some must be tachyons, with $m^2 < 0$ (for the spacelike geodesics). However, as for the Gaussian case, we can still identify

$$n^m = \dot{X}^m$$

Using the radial gauge condition, show that these can be chosen as geodesics with

$$v = e^\tau, \qquad X(\tau) = e^\tau X(0), \qquad p = X(0)$$

so all particles start at the origin at $\tau = -\infty$, and their position at $\tau = 0$ is determined by their initial (constant) momentum. (Thus particles with proportional momenta travel the same path, but arrive at different points at $\tau = 0$; however, in this case τ is neither the time x^0 nor the proper time s, but just an arbitrary parameter.)

As we saw in subsection VIB1, the radial gauge is related to gauge-covariant translation (in general relativity, "parallel transport") as, for any tensor ψ,

$$\tilde{\psi}(\tilde{y}) = e^{x^a(y)\mathcal{D}_a}\psi(y) = e^{\Lambda}\psi(\tilde{y}), \qquad \tilde{y}^m = e^{x^a(y)\mathcal{E}_a{}^m(y)D_m}y^m$$

where y is the "origin", $\Lambda = \Lambda^I M_I$ is just a Lorentz transformation, and $\mathcal{D}$ is the covariant derivative acting at y:

$$\mathcal{D}_a = \mathcal{E}_a{}^m(y)D_m + \varpi_a{}^I(y)M_I, \qquad D_m = \frac{\partial}{\partial y^m}; \qquad [\mathcal{D}_a, \mathcal{D}_b] = \mathcal{T}_{ab}{}^c\mathcal{D}_c + \mathcal{R}_{ab}{}^I M_I$$

As in general for coordination transformation parameters λ^a, x^a now transforms under local Lorentz transformations. (In background field language, this "quantum field" transforms under the "background" Lorentz transformations.) Thus, x^a is now a function of y; it cannot be made even covariantly constant in general:

$$\mathcal{D}_a x^b = 0 \quad \Rightarrow \quad 0 = [\mathcal{D}_a, \mathcal{D}_b]x^c = -x^d\mathcal{R}_{abd}{}^c$$

(For more practical reasons, if we defined it to be invariant or constant, the manipulations that follow would break down.) At this point we have only made a Lorentz transformation on ψ, since it and $\tilde{\psi}$ are evaluated at the same point $\tilde{y}$. However, as for Yang-Mills in subsection VIB1, for the next step we want to identify x^a as the new coordinate:

$$\psi'(x^a) = \tilde{\psi}(\tilde{y}^m(y^m, x^a)) = e^{x^a\mathcal{D}_a}\psi(y)$$

where ψ' has implicit dependence on y, since in radial gauges the choice of origin is gauge parameters that define the gauge. (The coordinates are defined as radial with respect to the origin y.) Thus, we have made a Lorentz transformation $\psi \to \tilde{\psi}$ followed by a coordinate transformation $\tilde{\psi} \to \psi'$.

We also want to define a covariant derivative for x by

$$\nabla\psi' = (\mathcal{D}\psi)' = e^{x^a\mathcal{D}_a}\mathcal{D}_a\psi(y)$$

where ∇ (as for Yang-Mills) contains only $\partial_a = \partial/\partial x^a$ and not D_m:

$$\nabla_a = e_a{}^b(x)\partial_b + \omega_a{}^I(x)M_I$$

At this point we no longer distinguish flat and curved indices, since the Lorentz gauge has been fixed. We have then transformed

$$y, \psi, \mathcal{D} \to x, \psi', \nabla$$

Note that the y-coordinate tensors are the x-coordinate tensors evaluated at the origin:

$$\psi'(0) = \psi(y), \qquad (\nabla\psi')(0) = (\mathcal{D}\psi)(y)$$

We can identify this as the radial gauge when $x(y)$ satisfies the geodesic condition, since then

$$(x\cdot\mathcal{D})x = 0 \quad \Rightarrow \quad x' \equiv e^{x\cdot\mathcal{D}}x = x$$

$$\Rightarrow \quad x\cdot\nabla\psi' = x\cdot(\mathcal{D}\psi)' = (x\cdot\mathcal{D}\psi)' = x\cdot\mathcal{D}\psi' = x\cdot\partial\psi'$$

making use of $\psi'(x) = e^{x\cdot\mathcal{D}}\psi(y)$.

Unfortunately, it is somewhat difficult to continue this construction in terms of the covariant derivative, but simpler in terms of the "dual" differential forms. We therefore define the (Lorentz-covariantized) Lie derivative as

$$\mathcal{L}_{x\cdot\mathcal{D}}\hat{\psi} = x\cdot\mathcal{D}\hat{\psi}, \qquad \mathcal{L}_{x\cdot\mathcal{D}}\hat{\mathcal{D}} = [x\cdot\mathcal{D}, \hat{\mathcal{D}}]$$

for any "tensor" (object carrying only flat indices) $\hat{\psi}$ and any "covariant derivative" (object with a flat vector index free, but multiplying partial derivatives and Lorentz generators) $\hat{\mathcal{D}}$. We generalize to evaluate on not only ψ and $\mathcal{D}$, but to apply the Lie derivative also as part of the transformation $exp(\mathcal{L}_{x\cdot\mathcal{D}})$. For that reason, for the remainder of this section we will abbreviate $\mathcal{L}_{x\cdot\mathcal{D}}$ as just $\mathcal{L}$. We then have

$$[x\cdot\mathcal{D}, \mathcal{D}_a] = -(\mathcal{D}_a x^b)\mathcal{D}_b + x^b(\mathcal{T}_{ba}{}^c\mathcal{D}_c + \mathcal{R}_{ba}{}^I M_I)$$

Defining Lie derivatives to satisfy the usual Leibniz and distributive rules like any derivative (since we use them as infinitesimal transformations), we then find

$$(\mathcal{L}\mathcal{E}_a{}^m)\mathcal{E}_m{}^b = -\mathcal{D}_a x^b + x^c\mathcal{T}_{ca}{}^b, \qquad \mathcal{E}_a{}^m\mathcal{L}\varpi_m{}^I = x^b\mathcal{R}_{ba}{}^I$$

In terms of differential forms, defined as

$$\mathcal{E}^a = dy^m\mathcal{E}_m{}^a, \qquad \varpi^I = dy^m\varpi_m{}^I$$

$$D = \mathcal{E}^a\mathcal{D}_a = d + \varpi^I M_I \quad \Rightarrow \quad Dx^a = dx^a - x^b\varpi_b{}^a$$

we then have

$$\mathcal{L}x^a = 0, \qquad \mathcal{L}(\mathcal{T},\mathcal{R}) = x\cdot\mathcal{D}(\mathcal{T},\mathcal{R})$$

$$\mathcal{L}\mathcal{E}^a = Dx^a - \mathcal{E}^b x^c\mathcal{T}_{cb}{}^a$$

$$\mathcal{L}\varpi^I = \mathcal{E}^a x^b\mathcal{R}_{ba}{}^I \quad \Rightarrow \quad \mathcal{L}(Dx^a) = -x^b\mathcal{E}^c x^d\mathcal{R}_{dcb}{}^a$$

which covers all the quantities that appear in evaluating $e^{\mathcal{L}}$ on $\mathcal{E}^a$ and ϖ^I. The geodesic condition prevents higher derivatives of x^a from appearing in the transformation law, and allows us to freely reorder all the x's to the left at the end of the calculation for identifying the coefficients of the Taylor expansion, at which point we can forget that x depends on y. Thus, these few equations for the action of $\mathcal{L}$ allow any transformed quantity to be evaluated straightforwardly by iteration, Taylor expanding $e^{\mathcal{L}}$ in powers of $\mathcal{L}$.

The important distinction between the transformation laws for $\mathcal{D}_a$ and $\mathcal{E}^a$ is that for $\mathcal{E}^a$ the derivatives of x appear only in the combination dx, which makes changing coordinates from y (or $\tilde{y}$) to x easier. Specifically, by iterating the above Lie derivatives, we find a solution of the form

$$\mathcal{E}'^a = e^{\mathcal{L}}\mathcal{E}^a = \mathcal{E}^b\mathcal{A}_b{}^a + (Dx^b)\mathcal{B}_b{}^a, \qquad \varpi'^I = e^{\mathcal{L}}\varpi^I = \mathcal{E}^a\mathcal{A}_a{}^I + (Dx^a)\mathcal{B}_a{}^I$$

where $\mathcal{A}_b{}^a$, $\mathcal{B}_b{}^a$, $\mathcal{A}_a{}^I$, $\mathcal{B}_a{}^I$ are functions of x and of tensors evaluated at the "origin" $((\mathcal{D}\cdots\mathcal{D}\mathcal{T})(y)$, $(\mathcal{D}\cdots\mathcal{D}\mathcal{R})(y))$. For Riemann normal coordinates, we want to fix y (e.g., $y = 0$), so we evaluate the above at $dy = 0$. Furthermore, we can choose the gauge $\omega(0) = 0$ (at least for vanishing torsion), so also $Dx \to dx$. Then the solution is

$$E^a = dx^b\mathcal{B}_b{}^a, \qquad \omega^I = dx^a\mathcal{B}_a{}^I$$

Thus, $\mathcal{B}_b{}^a$ and $\mathcal{B}_a{}^I$ are the inverse vierbein $e_m{}^a$ and Lorentz connection $\omega_m{}^I$ for the new coordinate system,

$$\nabla_a = (\mathcal{B}^{-1})_a{}^b(\partial_b + \mathcal{B}_b{}^I M_I)$$

written explicitly as a Taylor expansion in x by the above method, all of whose coefficients are tensors (torsions and curvatures and their derivatives) evaluated at the origin.

However, we can also use these results for first-quantization (where actions are expressed in terms of, e.g., $\mathcal{E}'^a/d\tau$ for the particle) in background field gauges, by choosing y as the background and x as the quantum coordinate (see subsection VIB1); then we keep both the dy and Dx terms.

Exercise IXB4.2

Find the first few orders of this expansion.

a Using the above method, show that for vanishing torsion

$$E^a = dx^b(\delta_b^a - \tfrac{1}{6}x^c x^d \mathcal{R}_{cbd}{}^a + ...), \qquad \omega^I = \tfrac{1}{2}dx^a(x^b\mathcal{R}_{ba}{}^I + ...)$$

b Check the validity of this result by evaluating $[\nabla, \nabla]$ to this order from the ∇ given by this E and ω.

c Use instead the covariant-derivative method of subsection VIB1. In this case, we find

$$\nabla_a = e^{x\cdot\tilde{\mathcal{D}}}\tilde{\mathcal{D}}_a e^{-x\cdot\tilde{\mathcal{D}}} + h_a{}^b e^{x\cdot\tilde{\mathcal{D}}}\partial_b e^{-x\cdot\tilde{\mathcal{D}}}$$

where $h_a{}^b$ is chosen to cancel all D_m terms in ∇, and we have defined

$$\tilde{\mathcal{D}}_a = \mathcal{E}_a{}^m(y)D_m + \varpi_a{}^I(y)\widetilde{M}_I$$

where now x^a is "constant", so D_m and $\widetilde{M}$ do not act on it. (Otherwise, in this approach, we would be stuck with tons of $\mathcal{D}\cdots\mathcal{D}x$ terms.) In terms of the previously defined Lorentz generators,

$$M_{ab} = \widetilde{M}_{ab} + x_{[a}\partial_{b]}$$

Find h to this order, and use it to obtain

$$\nabla_a = (\delta_a^b + \tfrac{1}{6}x^c x^d \mathcal{R}_{cad}{}^b + ...)\partial_b + \tfrac{1}{2}(\tfrac{1}{2}x^b\mathcal{R}_{ba}{}^{cd} + ...)M_{dc}$$

restoring $\widetilde{M}$ to M.

5. Weyl scale

The gauge-fixed kinetic term can be simplified by including the conformal compensator (see subsection IXA7). The quadratic part of the gauge-invariant Lagrangian is then ($\phi = 1 + \frac{1}{2}\chi$)

$$L_0 = -\tfrac{1}{4}\mathbf{e}^{-1}\phi(R - 4\tfrac{D-1}{D-2}\Box)\phi$$

$$\approx -\tfrac{1}{4}[h^{ab}\Box h_{ab} + 2(\partial^b h_{ab})^2 - h^a{}_a\Box h^b{}_b + 2h^a{}_a\partial^b\partial^c h_{bc}] - \tfrac{1}{2}\chi(\Box h^a{}_a - \partial^a\partial^b h_{ab}) + \tfrac{1}{4}\tfrac{D-1}{D-2}\chi\Box\chi$$

The nicest (globally Lorentz) covariant gauge comes from choosing the coordinate and scale gauge-fixing functions

$$f_a = \partial^b h_{ab} - \tfrac{1}{2}\partial_a h^b{}_b + \tfrac{1}{2}\partial_a\chi, \qquad f = \chi - h^a{}_a$$

We use these to obtain the gauge-fixed Lagrangian (see subsection VIB9)

$$\begin{aligned} L &= L_0 + \tfrac{1}{2}(\partial^b h_{ab} - \tfrac{1}{2}\partial_a h^b{}_b + \tfrac{1}{2}\partial_a\chi)^2 - \tfrac{1}{8}(\chi - h^a{}_a)\Box(\chi - h^b{}_b) \\ &= -\tfrac{1}{4}h^{ab}\Box h_{ab} + \tfrac{1}{4}\tfrac{1}{D-2}\chi\Box\chi \end{aligned}$$

plus ghost terms. Now the h kinetic term is simpler. Also, remember that χ decouples from conformal matter. These features of gauge fixing make this formalism closely analogous to the Stückelberg formalism for the massive vector. We can also define nonlinear versions of these gauge-fixing functions, such as $\partial_m(\phi\,\mathbf{e}^{-1/2}e_a{}^m)$ or $\partial_m(\phi^2\sqrt{-g}g^{mn})$ for the coordinate gauge, and $\phi\,\mathbf{e}^{-1/2}$ or $\phi^2\sqrt{-g}$ for the scale.

Exercise IXB5.1

Find the ghost terms for linearized gravity in the Fermi-Feynman gauge, and its simplification with the compensator.

The scale gauge can also be fixed in terms of the vierbein/metric alone: For example, we can fix the gauge

$$\mathbf{e} = 1$$

in which case ϕ acts simply as a renaming of $\mathbf{e}$. A more unusual gauge is

$$R = 0$$

This is not a restriction on the geometry, since the physical Ricci scalar is effectively replaced by its scale transform

$$R' = \phi^{-(D+2)/(D-2)}(R - 4\tfrac{D-1}{D-2}\Box)\phi$$

which is scale invariant. In the gauge $\phi = 1$, $R' = R$, but in the gauge $R = 0$ it is proportional to $\Box\phi$.

Exercise IXB5.2

Show that the ghosts for scale transformations propagate in the gauge $R = 0$: Find their contribution to the action.

More general gauges are possible when matter fields appear. For example, consider coupling gravity, with compensator, to a physical conformal scalar ψ. With appropriate normalization of the compensator and physical scalar, the kinetic terms for the two fields are identical except for sign: There is a manifest O(1,1) symmetry. We can take advantage of this by using a "lightcone" basis for these fields: Defining $\phi_\pm = \phi \pm \psi$, the full nonlinear (in gravity) Lagrangian L becomes ($S = \int dx\ \mathbf{e}^{-1}L$)

$$L = \phi_+(\tfrac{D-1}{D-2}\Box - \tfrac{1}{4}R)\phi_-$$

The overall normalization is arbitrary, including sign, since we can rescale either field by a constant. Many Weyl scale gauges are possible, and somewhat more transparent than making field redefinitions on the corresponding action without compensator. Effectively, we can redefine the fields $\phi_\pm$ arbitrarily as long as we don't fix ϕ_+/ϕ_- to a constant, since that combination is scale invariant. (I.e., ϕ_+/ϕ_- can be redefined, but not fixed.)

Some of the more interesting choices are:

$$\begin{aligned} \phi_\pm = 1 \pm \varphi \quad &\Rightarrow \quad L = -\tfrac{1}{4}R - \varphi(\tfrac{D-1}{D-2}\Box - \tfrac{1}{4}R)\varphi \\ \phi_\pm = e^{\pm\varphi} \quad &\Rightarrow \quad L = -\tfrac{1}{4}R - \tfrac{D-1}{D-2}\varphi\Box\varphi \\ \phi_\pm = \varphi^{1\pm a} \quad &\Rightarrow \quad L = \varphi[(1-a^2)\tfrac{D-1}{D-2}\Box - \tfrac{1}{4}R]\varphi \\ \phi_+ = \varphi, \quad \phi_- = 1 \quad &\Rightarrow \quad L = -\tfrac{1}{4}R\varphi \end{aligned}$$

We can also have any of these gauge-fixed Lagrangians with opposite overall sign, simply by changing the choice of either ϕ_+ or ϕ_- by a sign. The first two choices are useful because they put the action in standard form, as the usual gravity action plus a physical scalar kinetic term. (Thus, coupling a massless scalar to gravity either conformally or minimally is equivalent, and the two cases are distinguished only by interactions.) In fact, the first choice, or "temporal gauge" $\phi_+ + \phi_- = constant$, just returns us to the form without compensator, $\phi = 1$. On the other hand, changing the sign of ϕ_- yields the "axial gauge" $\phi_+ - \phi_- = constant$, which is fixing the *physical* scalar as $\psi = 1$. The overall sign of the action changes because the physical scalar is traded for the compensator, or the corresponding part of the metric. This gauge is closely related to the "string gauge": In our third choice above the gravity action is invisible until the surviving scalar has been expanded about its vacuum value. The constant a is arbitrary except that it must not vanish (so that ϕ_+/ϕ_- is not a constant). In particular, this action appears in string theory, with the choice

$$a = \frac{1}{\sqrt{D-1}} \quad \Rightarrow \quad L = \varphi(\Box - \tfrac{1}{4}R)\varphi$$

which eliminates explicit D-dependence. Again the scalar appears with the wrong-sign kinetic term, but R appears with the right sign (or vice versa), because of more complicated redefinitions. The sign of the $\Box$ changes back to the usual for $|a| > 1$. However, for $|a| = 1$, it disappears completely. A similar result occurs for the last choice, or "lightcone gauge" $\phi_- = 1$.

Exercise IXB5.3

The property that distinguishes this kinetic term for a scalar coupled to gravity is the O(1,1) symmetry:

a Before fixing the Weyl scale gauge, the continuous SO(1,1) subgroup of this symmetry is just the scaling $\phi_\pm \to \Lambda^{\pm 1}\phi_\pm$. After gauge fixing, this transformation may change the gauge, and thus may need to be combined with a constant Weyl scale transformation to preserve the gauge. In that case the vierbein will also transform under the resulting modified SO(1,1) transformation. Find the SO(1,1) transformations for φ and $e_a{}^m$ in the above 4 gauges.

b There is also the "parity" transformation of this O(1,1), $\phi_+ \leftrightarrow \phi_-$. Find the modified form of this transformation for φ and $e_a{}^m$.

Although all these choices are equivalent in perturbation theory (though the physical scalar may require a nonvanishing vacuum value), they aren't necessarily so nonperturbatively, depending on the ranges of the various scalars. Unfortunately, nonperturbative gravity is not understood well enough (even classically) to make such distinctions, even though they may be important physically. The above considerations generalize straightforwardly to the case with many physical scalars, where we may consider symmetry groups such as O(n,1). If the physical scalars form a nonlinear σ model, the compensator may join in to make the σ-model groups noncompact: Examples of this appear in supergravity and strings (see below).

The appearance of a physical scalar can also affect the way scale gauges are chosen in conjunction with coordinate gauges. For example, a result similar to the one found at the beginning of this subsection can be obtained from the (linearized) action with both compensator and physical scalar ψ (where $\langle\psi\rangle = 0$),

$$L \approx L_0 - \tfrac{1}{4}\psi\Box\psi$$

choosing the same χ-dependent coordinate-fixing term $(f_a)^2$, but imposing the scale gauge

$$\psi = \tfrac{1}{\sqrt{2}}(\chi - h^a{}_a)$$

The result is identical to the one given at the beginning of this subsection, except that now no scale ghosts appear: The scalar that appears as $h^a{}_a$ is now physical, and no longer needs a ghost to cancel it. This is the perturbative "string gauge" for scale invariance, which appears automatically in covariantly gauge-fixed string theory.

Exercise IXB5.4

Let's investigate such gauge choices further:

a Starting with the Fermi-Feynman-gauge-fixed linearized gravity action of subsection IXB1, add the physical-scalar kinetic term $-\frac{1}{4}\psi\Box\psi$. Separate the traceless and trace pieces of h_{ab}. Show that the string-gauge action (i.e, the one given at the beginning of this subsection if we ignore ghosts) follows from simply switching

$$\psi \leftrightarrow \tfrac{1}{\sqrt{D}}h^a{}_a$$

and then identifiying the new ψ with $\sqrt{2}\chi/(D-2)$.

b The way the physical scalar of string theory appears in the gauge-invariant and gauge-fixed action is slightly more clever than as described above. (See subsections XIB5-6 below.) The kinetic term (already in the string gauge for scale invariance) is

$$S = \int dx\ \Phi(\Box - \tfrac{1}{4}R)\Phi$$

where the missing $\mathbf{e}$ has been absorbed into Φ by a field redefinition. (Since Φ is thus not a scalar, we define $\Box\Phi$ by $\mathbf{e}^{-1/2}\Box\mathbf{e}^{1/2}\Phi$, since $\mathbf{e}^{1/2}\Phi$ is a scalar.) Expanding $\Phi = 1 + \chi$, the linearized gauge fixing is now simply

$$L \to L + \tfrac{1}{2}(\partial^b h_{ab} + \partial_a \chi)^2$$

(or we can use the nonlinear gauge-fixing function $\partial_m(\Phi e_a{}^m)$). Show the result is the same as above.

REFERENCES

1 T. de Donder, *La gravifique einsteinienne* (Paris, 1921).

2 J. Scherk and J.H. Schwarz, *Gen. Rel. and Grav.* **6** (1975) 517;
M. Kaku, *Nucl. Phys.* **B91** (1975) 99;
M. Goroff and J.H. Schwarz, *Phys. Lett.* **127B** (1983) 61:
lightcone gauge for gravity.

3 Cartan, *loc. cit.* (IXA), ch. X;
M. Spivak, *A comprehensive introduction to differential geometry* (Publish or Perish Inc., 1979) v. II, p. 299:
Riemann normal coordinate expansion.

4 U. Müller, C. Schubert, and A.E.M. van de Ven, A closed formula for the Riemann normal coordinate expansion, gr-qc/9712092.

5 W. Siegel, *Phys. Lett.* **149B** (1984) 162, **151B** (1985) 396, **211B** (1988) 55:
string gauge.

C. CURVED SPACES

There are some important solutions of general relativity that have no close analog in Yang-Mills. Here we consider the ones relevant to the only experimental verifications of this theory: Solutions outside approximately spherical matter distributions (like the Sun and Earth), and those describing the Universe itself.

1. Self-duality

Plane wave solutions can be constructed for gravity in the same way as for Yang-Mills (see subsection IIIC3): A little more work (solving the torsion constraint, or using the result of the free theory) gives

$$\nabla^+ = \partial^+ - \tfrac{1}{2}x^i x^j R^{+i+j}(x^-)\partial^- - x^i R^{+i+j}(x^-)M^{-j} \qquad (\nabla^- = \partial^-, \quad \nabla^i = \partial^i)$$

where R^{+i+j} is an arbitrary function of x^-, but symmetric in ij, and the empty-space field equations imply it is also traceless:

$$R^{+i+i} = 0$$

If we want to couple Yang-Mills to gravity, then we can still write exact solutions as long as both waves are parallel; then

$$R^{+i+i} = 2T^{++} = \tfrac{1}{g^2}tr(F^{+i}F^{+i})$$

where here g^2 refers to the Yang-Mills coupling. (Similarly, we can add in other fields, such as massless, neutral scalars or particles.)

Exercise IXC1.1

Check that the gravitational plane wave solution satisfies the field equations and torsion constraint. Show that we can also find more-special solutions of this form satisfying

$$g_{mn} = \eta_{mn} + \tfrac{1}{g^2}tr(A_m A_n), \qquad R_{mnpq} = \tfrac{1}{g^2}tr(F_{mn}F_{pq})$$

This has the interpretation that the "graviton" is the bound-state of two "gluons". However, it is only a kinematic effect, since the two gluons happen to be traveling in the same direction at the same speed. (We saw in subsection VIIIA7 that a similar effect always occurs in D=2, since there only two spatial directions exist.)

Self-duality for Yang-Mills was discussed in subsection IIIC4. Similar remarks apply to gravity: We again impose

$$[\nabla_a, \nabla_b] = \pm\tfrac{1}{2}\epsilon_{abcd}[\nabla^c, \nabla^d]$$

Self-duality again implies the field equations, by dualizing the Bianchi identities: For gravity

$$R_{[abc]d} = 0 \quad \Rightarrow \quad 0 = \pm\tfrac{1}{2}\epsilon^{abcd}R_{bcde} = \tfrac{1}{4}\epsilon^{abcd}\epsilon_{bcfg}R^{fg}{}_{de} = R^a{}_e$$

While it might appear that the self-duality condition is still second-order because solving the torsion constraint makes the Lorentz connection the derivative of the vierbein, the self-duality allows the gauge where the connection is also self-dual, and this condition effectively becomes a first-order field equation:

$$R_{abcd} = R_{cdab} \quad \Rightarrow \quad R_{abcd} = \pm\tfrac{1}{2}\epsilon_{cdef}R_{ab}{}^{ef} \quad \Rightarrow \quad \omega_{abc} = \pm\tfrac{1}{2}\epsilon_{bcde}\omega_a{}^{de}$$

Exercise IXC1.2

Apply exercises IIIC3.2 and IIIC4.1 to gravity:

a Rewrite all the above results of this subsection in spinor notation for D=4.

b For arbitrary dimension D, generalize $e_-{}^+$ to an arbitrary function of x^-, x^i, find the covariant derivative and curvature in terms of it, show the source-free Einstein's equations imply it satisfies

$$(\partial^i)^2 e_-{}^+ = 0$$

and in D=4 identify the pieces analytic and anti-analytic in x^t with the two polarizations.

In four dimensions (2 space + 2 time), lightcone methods can again be applied (see subsection IIIC5): Now

$$[\nabla^{\alpha\beta'}, \nabla^{\gamma\delta'}] = C^{\alpha\gamma}\tfrac{1}{2}R^{\beta'\delta'\epsilon'\zeta'}M_{\epsilon'\zeta'} \qquad (\omega_{\alpha\beta'}{}^{\gamma\delta} = 0)$$

The fact that $[\nabla^{(\oplus\alpha'}, \nabla^{\ominus)\beta'}]$ has only an $M_{\alpha'\beta'}$ term poses an additional constraint; the full solution is then

$$\nabla^{\oplus\alpha'} = \partial^{\oplus\alpha'}, \quad \nabla^{\ominus\alpha'} = \partial^{\ominus\alpha'} + (\partial^{\oplus\alpha'}\partial^{\oplus\beta'}\phi)\partial^{\oplus}{}_{\beta'} + \tfrac{1}{2}(\partial^{\oplus\alpha'}\partial^{\oplus\beta'}\partial^{\oplus\gamma'}\phi)M_{\beta'\gamma'}$$

$$R^{\alpha'\beta'\gamma'\delta'} = -i\partial^{\oplus\alpha'}\partial^{\oplus\beta'}\partial^{\oplus\gamma'}\partial^{\oplus\delta'}\phi$$

(In this case, the existence of covariantly constant spinors is a consequence of self-duality.) The equation of motion that follows from the final condition is now

$$\Box\phi - i(\partial^{\oplus\alpha'}\partial^{\oplus\beta'}\phi)(\partial^{\oplus}{}_{\alpha'}\partial^{\oplus}{}_{\beta'}\phi) = 0$$

2. De Sitter

The simplest spaces are those where the Ricci scalar is constant, and the other parts of the curvature vanish:

$$R_{ab}{}^{cd} = k\delta_{[a}^c\delta_{b]}^d$$

These are special solutions of the field equations without matter, but with a cosmological term, where there are no physical gravitons (the Weyl tensor vanishes), and thus represent the vacuum. Since there are no physical degrees of freedom, we can represent this space by just the conformal compensator: i.e. the vierbein (metric) is just the flat one up to a local Weyl scale transformation. We thus have (from subsection IXA7)

$$R_{ab}{}^{cd} \to \Phi\delta_{[a}^{[c}\partial_{b]}\partial^{d]}\Phi - \delta_{[a}^c\delta_{b]}^d(\partial\Phi)^2 = k\delta_{[a}^c\delta_{b]}^d$$

where we have written the curvature as a scale transformation of flat space $R_{ab}{}^{cd} = 0$, $\nabla = \partial$: The space is "conformally flat". Separating this equation into its irreducible parts with respect to the Lorentz group, the Weyl tensor part vanishes identically, leaving

$$2\Phi\Box\Phi - D(\partial\Phi)^2 = Dk, \quad D\partial_a\partial_b\Phi = \eta_{ab}\Box\Phi$$

(The latter equation isn't implied in $D = 2$, where the global conformal group is larger, and more general coordinate choices are possible for this solution. However, we can still use it consistently.) The latter equation can be solved easily: Looking at $a \neq b$, we see that Φ is a sum of functions of one variable. Then looking at $a = b$ tells us that these functions are quadratic and have the same quadratic coefficient, while the former equation gives k:

$$\Phi = A + B^a x_a + C\tfrac{1}{2}x^a x_a, \qquad k = 2AC - B^2$$

We can choose any A, B^a, and C that give the desired value of k: For example, we can choose the solution $\Phi = 1 + \frac{1}{4}kx^2$ (giving the usual flat-space coordinates for $k = 0$), or $\Phi = B^a x_a$ (choosing the direction of B^a as appropriate to $k = -B^2$ — spacelike, lightlike, or timelike).

Exercise IXC2.1

Show that after a Weyl scale transformation the action for gravity including a cosmological term is (up to a sign) that of a conformal self-interacting scalar coupled to gravity. Use this to show that the de Sitter space solution (in the R=0 gauge) yields an "instanton" for this scalar theory, and compare with the

Yang-Mills instanton of subsection IIIC6. Show that similar solutions exist for massless scalars in arbitrary dimensions with potentials $\sim \phi^n$ for arbitrary n (but then $k = 0$).

The geometry of this space can be understood most easily as that of a D-dimensional hyperboloid embedded in a flat (D+2)-dimensional space, where we add one space and one time dimension: Again using the methods of subsections IA6 and IVA2, we now supplement the constraint

$$y^2 = 0 \quad \Rightarrow \quad y^A = ew^A, \qquad (w^+, w^-, w^a) = (1, \tfrac{1}{2}x^a x_a, x^a)$$

with the additional constraint

$$n^A y_A = 1 \quad \Rightarrow \quad e = \frac{1}{n^A w_A} = \frac{1}{-n^- - n^+ \frac{1}{2}x^a x_a + n^a x_a}$$

where n^A is a (D+2)-vector, yielding the intersection of a cone and plane. In particular, for $n^2 \neq 0$ we can write the metric on the space whose coordinates are all but $n \cdot y$:

$$y^A = (|n^2|^{-1/2}, z^{\mathcal{A}}) \quad \Rightarrow \quad z^2 + n^{-2} = 0, \qquad -ds^2 = dz^2$$

which is the definition of a hyperboloid. Comparing the metric, we find the previous result:

$$-ds^2 = dy^2 = e^2 dx^a dx_a, \qquad e = \Phi^{-1} \quad \Rightarrow \quad k = -n^2$$

Of course, by appropriate choice of the original flat space, we can choose a space of any signature. In particular, we see that for a unit sphere k is normalized to 1. Thus, with our conventions we have in that case

$$unit\ sphere: \quad R_{ab}{}^{cd} = \delta^c_{[a}\delta^d_{b]}$$

(but the constant value of the Ricci scalar will depend on the dimension).

This gives the most general coordinate system for de Sitter space as a local scale of flat space, since conformal transformations are the most general coordinate transformations that will just replace this scale factor with another, and they just rotate n^A. The symmetry group of the D-dimensional subspace that satisfies these two constraints is as big as the Poincaré group, namely SO(D,1), ISO(D−1,1), or SO(D−1, 2), depending on whether n^2<, =, or >0: The former constraint preserves the conformal group, while the latter kills a timelike, lightlike, or spacelike coordinate.

Exercise IXC2.2

We can also start instead with a D+1-dimensional space, which is a natural choice for the symmetry group of de Sitter space: Consider the metric and constraint

$$-k\ ds^2 = dz^2 = k\ dz^a dz^b \eta_{ab} + dz_{D+1}^2, \qquad 1 = z^2 = k\ z^a z^b \eta_{ab} + z_{D+1}^2$$

Both equations have the same global symmetry group, determined by the sign of k; $k = 0$, flat space, can be considered as a limiting case of the others.

a Solve the constraint $y^2 = 1 \to z(x)$ as in subsection IVA2 for $\phi^2 = m^2 \to \phi(\chi)$, and substitute to find the metric in terms of x.

b Find the conformal transformation on x^a that relates this coordinate system to the more general one above. (Hint: Use z of the D+2-dimensional construction.)

3. Cosmology

To a good approximation the universe can be described by a spacetime which is (spatially) rotationally invariant ("isotropic") with respect to a preferred time direction. Furthermore, it should be (spatially) translationally invariant ("homogeneous"), so the metric should depend only on that time coordinate. This means that the 3D subspace at any fixed time should be 3D spherical, flat, or de Sitter space, up to an overall time-dependent scale factor:

$$-ds^2 = -d\tau^2 + \phi^2(\tau)\Upsilon$$

where Υ is the de Sitter metric for the 3 other dimensions for $k = 1, 0, -1$ (given, e.g., by the coordinates in the previous subsection.) By a simple redefinition of the time coordinate, this can be put in a form which is conformal to a static space:

$$ds^2 = \phi^2(t)\widehat{ds}^2, \quad \widehat{ds}^2 = -dt^2 + \Upsilon$$

where by "$\phi(t)$" we really mean "$\phi(\tau(t))$", and the two time coordinates are related by

$$d\tau = dt\ \phi \quad \Rightarrow \quad t = \int d\tau \frac{1}{\phi(\tau)} \quad or \quad \tau = \int dt\ \phi(\tau(t))$$

Using previous results for 3D de Sitter space, we find $\widehat{ds}^2$ has curvature

$$\hat{R}_{ij}{}^{kl} = k\delta_{[i}^k \delta_{j]}^l, \quad rest = 0$$

where $k = 1, 0, -1$.

To a good approximation the matter in the universe can be approximated as a "dust", a collection of noninteracting particles. It should also be rotationally invariant with respect to the preferred time direction, so the momenta of the particles should be aligned in that time direction. (Really it is this matter that defines the time direction, since it generates the curvature of spacetime.) Furthermore, the dust should be translationally invariant, so the energy-momentum tensor should depend only on that time coordinate. We then can write (compare exercises IIIB4.2 and IXB2.2)

$$T_M^{ab} = \rho_M(t) u^a u^b, \quad u^a = \delta_0^a$$

where ρ_M is just the spatial density of particles in the "rest" frame. One way to derive the ϕ dependence of ρ_M that generalizes straightforwardly to other cases is by using conservation laws: By considering particles all of the same mass in units $m = 1$, or by considering J and T for each individual particle (since in this case we neglect interactions), we have from current conservation

$$J^a = \rho_M u^a \quad \Rightarrow \quad 0 = \nabla_a J^a = \mathbf{e}\partial_m \mathbf{e}^{-1} J^a e_a{}^m = (\phi^{-4})\partial_0(\phi^4)(\rho_M \phi^{-1})$$

$$\Rightarrow \quad \rho_M = 3a\phi^{-3}$$

for some nonnegative constant $3a$, and using (covariant) energy-momentum conservation as a check,

$$u^a \nabla_a u_b = \nabla_0 u_b = \phi^{-1}\partial_0 u_b = 0 \quad (geodesic) \quad \Rightarrow \quad \nabla_a T_M^{ab} = u^b \nabla_a J^a + J^a \nabla_a u^b = 0$$

where we have used (from the result of subsection IXA7 for scaling covariant derivatives)

$$\widehat{\nabla}_0 = \partial_0 \quad \Rightarrow \quad \nabla_0 = \phi^{-1}\partial_0$$

(Note, however, that ∇_i has Lorentz pieces, and $M_{0i}J_i \sim J_0 \neq 0$ even though $J_i = 0$.)

Exercise IXC3.1

Find completely explicit expressions for the covariant derivatives in this case (choosing some coordinates for Υ for $k = 0, \pm 1$) using just the Weyl transformation method of subsection IXA7.

For radiation, the momenta of the photons can't be timelike (they're lightlike, of course), but we can still use rotational and translational invariance, together with the fact that the trace of the energy-momentum tensor vanishes (from scale invariance: see subsection IXA7). Then

$$T_R^{ab} = \rho_R(t)\tfrac{1}{3}(4u^a u^b + \eta^{ab})$$

There is no conserved current, but energy-momentum conservation alone determines

$$0 = \nabla_a T_R^{ab} = \tfrac{4}{3}\rho_M u^a u^b \nabla_a \frac{\rho_R}{\rho_M} + \nabla^b \tfrac{1}{3}\rho_R = \phi^{-1}\delta_0^b\left(\tfrac{4}{3}\rho_M \partial_0 \frac{\rho_R}{\rho_M} - \tfrac{1}{3}\partial_0 \rho_R\right)$$

$$= \phi^{-1}\delta_0^b \rho_M^{4/3}\partial_0 \rho_M^{-4/3}\rho_R \quad \Rightarrow \quad \rho_R = \tfrac{3}{2}b\phi^{-4}$$

for some nonnegative constant $\frac{3}{2}b$.

Writing the vierbein as a scale-transformation of the constant curvature space discussed above (de Sitter in spatial directions, flat in other directions), the gravitational field equations with matter and radiation become (using results from subsection IXA6 or 7):

$$6[(\widehat{\nabla}_a\phi)(\widehat{\nabla}_b\phi) - \tfrac{1}{2}\eta_{ab}(\widehat{\nabla}\phi)^2] + [(\eta_{ab}\widehat{\Box} - \widehat{\nabla}_a\widehat{\nabla}_b) + (\hat{R}_{ab} - \tfrac{1}{2}\eta_{ab}\hat{R})]\phi^2 = 2(T_{Mab} + T_{Rab})\phi^4$$

The only independent components of this equation are the 00-component and trace, which are, after multiplying by an appropriate power of ϕ:

$$\tfrac{1}{2}\dot{\phi}^2 + \tfrac{1}{2}k\phi^2 = a\phi + \tfrac{1}{2}b, \qquad \ddot{\phi} + k\phi = a$$

For $k = 1$, these are just energy conservation and the equation of motion for a harmonic oscillator (centered at $\phi = a$). The 00 equation gave energy conservation because T_{00} is the energy density. The trace equation gave the field equation for ϕ, which is proportional to the time derivative of the 00 equation, due to the relation of $T^a{}_a$ to $\delta S/\delta\phi$ given earlier. These equations are easily solved: Imposing the initial condition $\phi(0) = 0$ (i.e., we set the "Big Bang", when curvatures and energy density were infinite, to be $t = 0$) and $\dot{\phi}(0) > 0$ (so $\phi \geq 0$),

$$k = \begin{Bmatrix} 1 \\ 0 \\ -1 \end{Bmatrix}: \quad \phi = a\begin{Bmatrix} 1 - \cos t \\ \frac{1}{2}t^2 \\ \cosh t - 1 \end{Bmatrix} + \sqrt{b}\begin{Bmatrix} \sin t \\ t \\ \sinh t \end{Bmatrix}$$

The "physical" time coordinate is then $\tau = \int_0 dt\ \phi$. In general ϕ can't be expressed directly in terms of τ, so we use the expressions for both in terms of t. For example, for $k = 1$ and $b = 0$ (just matter), we get a cycloid, which has only such a parametric expression. Explicit expressions can be found for $a = 0$ (just radiation): $\phi(\tau)$ is then a circle, parabola, or hyperbola for k=1, 0, -1. Also, for $k = 0$ and $b = 0$, $\phi \sim \tau^{2/3}$ (vs. $\sqrt{\tau}$ for $a = 0$).

Exercise IXC3.2

Find the modification to the equations of motion when a cosmological term is included.

For the case of pure matter ($b = 0$), the energy conservation equation written in terms of the τ coordinate becomes, using $d\tau = \phi\ dt$,

$$\tfrac{1}{2}\left(\frac{d\phi}{d\tau}\right)^2 - \frac{a}{\phi} = -\frac{k}{2}$$

This is the same as the Newtonian equation for the radial motion of a particle under the influence of a fixed point mass (or the relative motion of 2 point particles).

4. Red shift

The most obvious effect of the cosmological expansion is the cosmological "red shift". The expansion of the universe causes photons to lose energy, including those of the black-body radiation of the universe as well as those emitted long ago from distant sources. The easiest way to see this is to consider Killing vectors. Since the cosmological solutions are related to static, isotropic, homogeneous spaces by a time dependent (but space independent) scale transformation, the symmetries of this space are just in the spatial directions, and are basically the same as before the scale transformation. Specifically, the Killing vectors that survive the scale transformation $\nabla_a = \varPhi\widehat{\nabla}_a + (\widehat{\nabla}^b\varPhi)M_{ab}$ satisfy (see excercise IXA7.3)

$$\widehat{K}\cdot\widehat{\nabla}\varPhi = 0 \quad \Rightarrow \quad K = \widehat{K} \quad \Rightarrow \quad K_a = \varPhi^{-1}\widehat{K}_a$$

We then find for conserved momenta $K^a p_a \sim \varPhi^{-1}p_a$. Since the K's which survive are just the spatial ones, we at first find only the spatial components of $\varPhi^{-1}p_a$ conserved, but the conservation of the time component follows from $p^a p_a = 0$ for photons. Thus, $p_a \sim \varPhi \sim \phi^{-1}$. Since p_a is what an observer measures as the components of momentum (in his "local inertial frame", a gauge where at his location the metric is flat and its first derivative vanishes), observers measure the photon's energy, frequency, and corresponding black-body radiation (whose distribution depends only on E/T) as having time dependence $\sim \phi^{-1}$ (and wavelength as ϕ). The spectrum of radiation emitted by a distant object is then shifted by this energy loss, so the amount of shift determines how long ago it was emitted, and thus the distance of the emitter.

Exercise IXC4.1

Using this result for the ϕ dependence of the momenta of individual particles, we can now rederive the ϕ dependence of ρ's of the previous subsection directly from the explicit expressions for J and T of the point particle.

a Rederive J and T in curved space as in subsection IIIB4 and show that

$$J^m\delta_m^0 = \epsilon(p^0)\mathbf{e}(2\pi)^2\delta^3(x-X), \qquad T^{mn} = J^m p^n = J^n p^m$$

b From Killing vectors we just saw that

$$p^a\delta_a^0 \sim \begin{cases} \phi^0 & (m \neq 0) \\ \phi^{-1} & (m = 0) \end{cases}$$

where the massive particles are at rest ($p^a = m\delta_0^a$). Combine these results to find

$$J^a\delta_a^0 \sim \phi^{-3}, \qquad \rho \sim T^{ab}\delta_a^0\delta_b^0 \sim \begin{cases} \phi^{-3} & (m \neq 0) \\ \phi^{-4} & (m = 0) \end{cases}$$

c Find the factors multiplying the ρ's in T^{ab} for the two cases of dust and radiation from the explicit expression for T^{mn}. For the massive case (dust) all particles can be taken at rest, but for the massless case the particles travel at the speeed of light, so average over particles traveling in the three spatial directions. (ρ is a continuous function obtained by summing the δ functions of all the particles. However, for the above results it is sufficient to consider each individual particle for the massive case, and 3 particles at the same point going in orthogonal directions for the massless case.)

Astronomers use 3 parameters which are more directly observable. The "size" of the Universe ϕ is difficult to measure, but we can measure the change in time of this scale through red shifts: Comparing lengths at different times, we measure $\phi(\tau_2)/\phi(\tau_1)$, more conveniently represented in terms of the difference of the ln: In terms of the derivative, we have

$$ln\left(\frac{\phi(\tau_2)}{\phi(\tau_1)}\right) \equiv \int_{\tau_1}^{\tau_2} d\tau\ H(\tau)$$

The "Hubble constant" H measures the expansion rate, and gives an inverse length (time) scale. Its own rate of change can be defined in terms of a dimensionless quantity by comparing its inverse with the true time:

$$q \equiv \frac{d(H^{-1})}{d\tau} - 1$$

The "deceleration parameter" q tells how fast the expansion rate is slowing down. Defining another dimensionless parameter to measure energy density by using H^{-1} as a length scale, we have the 3 parameters

$$H \equiv \frac{d\phi/d\tau}{\phi}, \quad q \equiv -\frac{\phi\ d^2\phi/d\tau^2}{(d\phi/d\tau)^2}, \quad \sigma \equiv \frac{\rho}{3H^2}$$

where $\rho = T^{00} = \rho_M + \rho_R$. Note that in our conventions spatial integrals are weighted as $\int d^{D-1}x/(2\pi)^{D/2}$; thus the relation of our density to the more standard one is (see, e.g., exercise IXC4.1a)

$$\rho = (2\pi)^2 \rho_{usual} \quad \Rightarrow \quad \sigma = \frac{\frac{4\pi}{3} G\rho_{usual}}{H^2}$$

where $G = \pi$ in our conventions (but sometimes $G = 1$ is useful, especially for solutions describing stars and planets. See exercise IXB1.1.) The "density parameter" σ measures matter density with respect to the amount needed to close the universe. In the case of pure matter, and with vanishing cosmological constant, $\sigma = q$. Then the "critical" value is $q = \frac{1}{2}$, for which $k = 0$: For $q > \frac{1}{2}$, $k = 1$, while for $q < \frac{1}{2}$,

$k = -1$. In this case we also see that for a given value of H the critical value of the matter density is $\rho_c = \frac{3}{2}H^2$. If the matter of the universe has this density, we have $k = 0$, and spacetime is conformally flat. If it has greater density, we have $k > 1$, and space is closed.

Exercise IXC4.2

Solve for σ and q in terms of just a, b, k, and ϕ (but no time derivatives). In particular, show

$$b = 0 \quad \Rightarrow \quad \sigma = q = (2 - \tfrac{k}{a}\phi)^{-1}$$
$$a = 0 \quad \Rightarrow \quad 2\sigma = q = (1 - \tfrac{k}{b}\phi^2)^{-1}$$

These quantities are difficult to measure. Some recent estimates for their present values are:

$$H^{-1} = 14(2) \times 10^9 \ yrs., \quad .02 < q < 2$$

Even the gravitational constant is not so easy to measure: Its presently accepted value is

$$G = 6.67259(85) \times 10^{-11} m^3 kg^{-1} s^{-2}$$

which is accurate to only a few parts per 10,000, compared to the standard atomic and nuclear constants, which are known to better than 1 part per million. In "natural (Planck) units," $c = G/\pi = \hbar = 1$, $H^{-1} = 1.5(2) \times 10^{61}$.

5. Schwarzschild

All gravitational experiments outside of cosmology are based on the "Schwarzschild solution", which describes spherical symmetry outside the region with matter. Assuming also time independence, which is a consequence of spherical symmetry (Birkhoff's theorem), we look for a metric of the form

$$-ds^2 = -A^{-2}(r)dt^2 + B^{-2}(r)dr^2 + r^2(d\theta^2 + sin^2\theta \ d\phi^2)$$

(Other coordinate choices are possible, e.g. $-A^{-2}(r)dt^2 + B^{-2}(r)[dr^2 + r^2(d\theta^2 + sin^2\theta \ d\phi^2)]$.) The first step is the choice of a vierbein: The simplest choice following from this metric is

$$e_t = A\partial_t, \quad e_r = B\partial_r, \quad e_\theta = \frac{1}{r}\partial_\theta, \quad e_\phi = \frac{1}{r \ sin \ \theta}\partial_\phi$$

(This can also be used as a starting point in place of the metric.) The next step is to find the commutators of the e's, which tells us what ω terms the ∇'s must have to cancel these $c_{ab}{}^c$'s (vanishing torsion):

$$\begin{array}{lclcl}
[e_\theta, e_\phi] = -r^{-1} cot\ \theta\ e_\phi & & \nabla_\phi\ has\ M_{\theta\phi} & & \nabla_r = B\partial_r \\
[e_r, e_t] = B(ln\ A)' e_t & \Rightarrow & \nabla_t\ has\ M_{tr} & \Rightarrow & \nabla_t = A\partial_t + \alpha M_{tr} \\
[e_r, e_\theta] = -r^{-1} B e_\theta & & \nabla_\theta\ has\ M_{r\theta} & & \nabla_\theta = r^{-1}\partial_\theta + \beta M_{r\theta} \\
[e_r, e_\phi] = -r^{-1} B e_\phi & & \nabla_\phi\ has\ M_{r\phi} & & \nabla_\phi = (r\ sin\ \theta)^{-1}\partial_\phi + \gamma M_{r\phi} + \delta M_{\theta\phi}
\end{array}$$

where α, β, and γ depend only on r, while δ depends also on θ. (Their explicit forms are already clear at this point, but we'll collect the results below.)

We can now determine these Lorentz connections and compute the curvatures by calculating the ∇ commutators. Since we now use explicit functions for the vierbein and connections, we use the method described in subsection IXA2 for this situation: Using the identities

$$[M_{12}, V_2] = \eta_{22} V_1, \qquad [M_{12}, M_{23}] = \eta_{22} M_{13}$$

$$[\nabla_1, \nabla_2] = [e_1 + \omega_1, e_2 + \omega_2]$$

$$= \{[e_1, e_2] + (e_1\omega_2)M_2 - (e_2\omega_1)M_1\} + \{\omega_1[M_1, \nabla_2] - \omega_2[M_2, \nabla_1] - \omega_1\omega_2[M_1, M_2]\}$$

we then find:

$$\begin{aligned}
[\nabla_t, \nabla_\theta] &= -\alpha\beta M_{t\theta} \quad \Rightarrow \quad R_{t\theta t\theta} = -\alpha\beta \\
[\nabla_t, \nabla_\phi] &= -\alpha\gamma M_{t\phi} \quad \Rightarrow \quad R_{t\phi t\phi} = -\alpha\gamma \\
[\nabla_t, \nabla_r] &= -B(ln\ A)' e_t - B\alpha' M_{tr} + \alpha\nabla_t \quad = [\alpha - B(ln\ A)'] e_t + (\alpha^2 - B\alpha') M_{tr} \\
&\Rightarrow \quad \alpha = B(ln\ A)', \qquad R_{trtr} = \alpha^2 - B\alpha' \\
[\nabla_r, \nabla_\theta] &= -\frac{B}{r} e_\theta + B\beta' M_{r\theta} + \beta\nabla_\theta \quad = (\beta - \frac{B}{r}) e_\theta + (\beta^2 + B\beta') M_{r\theta} \\
&\Rightarrow \quad \beta = \frac{B}{r}, \qquad R_{r\theta r\theta} = -(\beta^2 + B\beta') \\
[\nabla_r, \nabla_\phi] &= -\frac{B}{r} e_\phi + B\gamma' M_{r\phi} + B\delta' M_{\theta\phi} + \gamma\nabla_\phi \\
&= (\gamma - \frac{B}{r}) e_\phi + (\gamma^2 + B\gamma') M_{r\phi} + (\gamma\delta + B\delta') M_{\theta\phi} \\
&\Rightarrow \quad \gamma = \frac{B}{r}, \qquad R_{r\phi r\phi} = -(\gamma^2 + B\gamma'), \qquad R_{r\phi\theta\phi} = -(\gamma\delta + B\delta') \\
[\nabla_\theta, \nabla_\phi] &= -\frac{cot\ \theta}{r} e_\phi + \frac{1}{r}(\partial_\theta\delta) M_{\theta\phi} + \delta\nabla_\phi + \beta\gamma M_{\theta\phi} - \beta\delta M_{r\phi} \\
&= (\delta - \frac{cot\ \theta}{r}) e_\phi + (\gamma - \beta)\delta M_{r\phi} + (\delta^2 + \beta\gamma + \frac{1}{r}\partial_\theta\delta) M_{\theta\phi} \\
&\Rightarrow \quad \delta = \frac{cot\ \theta}{r}, \qquad R_{\theta\phi\theta\phi} = -(\delta^2 + \beta\gamma + \frac{1}{r}\partial_\theta\delta), \qquad R_{r\phi\theta\phi} = 0
\end{aligned}$$

Collecting the results:

$$\nabla_t = A\partial_t + B(ln\ A)' M_{tr} \qquad R_{trtr} = BA[B(A^{-1})']'$$

$$\nabla_r = B\partial_r \qquad R_{t\theta t\theta} = R_{t\phi t\phi} = -\frac{B^2}{r}(ln\ A)'$$

$$\nabla_\theta = \frac{1}{r}\partial_\theta + \frac{B}{r}M_{r\theta} \qquad R_{r\theta r\theta} = R_{r\phi r\phi} = -\frac{BB'}{r}$$

$$\nabla_\phi = \frac{1}{r\ sin\ \theta}\partial_\phi + \frac{cot\ \theta}{r}M_{\theta\phi} + \frac{B}{r}M_{r\phi} \qquad R_{\theta\phi\theta\phi} = \frac{1-B^2}{r^2}$$

Exercise IXC5.1

Find the covariant derivative for the 2-sphere in spherical coordinates

$$ds^2 = d\theta^2 + sin^2\theta\ d\phi^2$$

in terms of the single SO(2) generator $M_{ab} = \epsilon_{ab}M$ by the above methods (and not that of exercises IXA5.3 nor IXA7.5). Calculate the curvature. Find the three Killing vectors. (Hint: What is the symmetry of the sphere?)

A simpler method of finding covariant derivatives and curvatures in this case is the Weyl scale method of subsection IXA7. (We already applied this method to the much simpler example of cosmology in subsection IXC3.) We start with the trivial covariant derivatives for the 2D metric

$$-ds^2 = -dt^2 + dr^2$$

which are just partial derivatives (zero curvature). Then we make the coordinate transformation

$$dr \to \frac{A(r)}{B(r)}dr$$

(explicit integration of this expression isn't needed in either the metric or the covariant derivatives), which modifies one of the covariant derivatives,

$$-ds^2 \to -dt^2 + \frac{A^2}{B^2}dr^2; \qquad \nabla_r \to \frac{B}{A}\partial_r, \qquad \nabla_t \to \partial_t$$

while the curvature still vanishes. (We have only chosen non-Cartesian coordinates for flat space.) Next, we make the scale transformation

$$\Phi = rA$$

to obtain the metric

$$-ds^2 = -(rA)^{-2}dt^2 + (rB)^{-2}dr^2$$

Applying the general formula

$$ds'^2 = \Phi^{-2}ds^2 \quad \Rightarrow \quad \nabla'_a = \Phi\nabla_a + (\nabla^b\Phi)M_{ab},$$

we find

$$\nabla_r \to rB\partial_r, \qquad \nabla_t \to rA\partial_t + \frac{B}{A}(rA)' M_{tr}$$

Since the space is only 2D, the general equation

$$R'_{ab}{}^{cd} = \Phi^2 R_{ab}{}^{cd} + \Phi \delta_{[a}^{[c} \nabla_{b]} \nabla^{d]} \Phi - \delta_{[a}^{c} \delta_{b]}^{d} (\nabla \Phi)^2$$

simplifies to

$$R'_{ab}{}^{cd} = \tfrac{1}{2} R' \delta_{[a}^{c} \delta_{b]}^{d}, \qquad \tfrac{1}{2} R' = \Phi^2 (\tfrac{1}{2} R + \Box \; ln \; \Phi)$$

so at this stage we have

$$\tfrac{1}{2} R \to (rA)^2 \frac{B}{A} \left[\frac{B}{A} (ln \; rA)' \right]'$$

Any 2D space can be expressed as a scale transformation of flat space locally, essentially because the curvature has only one component. (Globally this is not true, since the integral of the curvature, which is scale invariant in D=2, is different for different topologies. This is related to the fact that for nontrivial topologies more than one coordinate patch is needed; the missing part of the integral can be hidden in the boundaries of the patches: see exercise IXA7.6.)

Now we should repeat this procedure for θ and ϕ to get the covariant derivatives for the (2-)sphere, but this has already been done earlier. Besides, we don't need those expressions explicitly, since spherical symmetry means they vanish on anything, and we already know the curvature of a sphere. (So, we can also avoid choosing a coordinate system for the sphere.) Thus we can immediately take the direct product of the sphere with the above 2D space, and make the final scale transformation

$$\Phi = \tfrac{1}{r}$$

The result for the final covariant derivatives is

$$\nabla_r = B\partial_r, \quad \nabla_t = A\partial_t + \frac{B}{A} A' M_{tr}, \quad \nabla_i = \tfrac{1}{r} \overset{\circ}{\nabla}_i - \tfrac{1}{r} B M_{ir}$$

where $\overset{\circ}{\nabla}_i$ are the covariant derivatives for the sphere, in agreement with the previous method. The curvatures are also easy to find: Besides the $\nabla_r \Phi$ we needed for the covariant derivatives, the only second-order derivatives we need are $\nabla_r^2 \Phi$ and $\nabla_t^2 \Phi$. ($\nabla_t \Phi$ vanishes, but $\nabla_t^2 \Phi$ is nonvanishing because of the M_{tr} connection term in ∇_t converting $\nabla_t \Phi$ into $\nabla_r \Phi$.) Thus we need to evaluate

$$\begin{aligned} R'_{ij}{}^{kl} &= \delta_{[i}^{k} \delta_{j]}^{l} [\tfrac{1}{r^2} \cdot 1 - (\nabla_r \tfrac{1}{r})^2] \\ R'_{ii'}{}^{jj'} &= \delta_i^j [\tfrac{1}{r} \nabla_{i'} \nabla^{j'} \tfrac{1}{r} - \delta_{i'}^{j'} (\nabla_r \tfrac{1}{r})^2] \\ R'_{i'j'}{}^{k'l'} &= \delta_{[i'}^{k'} \delta_{j']}^{l'} \tfrac{1}{r^2} [\tfrac{1}{2} R + (\nabla_r^2 - \nabla_t^2) ln \; \tfrac{1}{r}] \end{aligned}$$

where $i' = (t, r)$, and the ∇'s and R on the right refer to the 2D t-r space just before or after the direct product. The result also reproduces the previous. The final result for R_{trtr} can be obtained even more simply by noting that it agrees with what we would have obtained by a single scaling for the 2D space $-ds^2 = -A^{-2}dt^2 + B^{-2}dr^2$, because of the triviality of the θ and ϕ derivatives.

Having all the curvatures, we can now calculate the Ricci tensor, which appears in the field equations. The nonvanishing components are:

$$R_{tt} = R_{trtr} + 2R_{t\theta t\theta}, \quad R_{rr} = -R_{trtr} + 2R_{r\theta r\theta}, \quad R_{\theta\theta} = R_{\phi\phi} = -R_{t\theta t\theta} + R_{r\theta r\theta} + R_{\theta\phi\theta\phi}$$

Vanishing of $R_{ab} - \frac{1}{2}\eta_{ab}R$ is equivalent to vanishing of R_{ab}. In terms of these curvatures, we see it also implies

$$-R_{trtr} = 2R_{t\theta t\theta} = -2R_{r\theta r\theta} = R_{\theta\phi\theta\phi}$$

These are easy to solve: First,

$$R_{t\theta t\theta} = -R_{r\theta r\theta} \quad \Rightarrow \quad (ln\ A)' = -(ln\ B)' \quad \Rightarrow \quad A = B^{-1}$$

where we have fixed the proportionality constant by requiring $A, B \to 1$ as $r \to \infty$ (redefining t by a constant scale transformation). Also,

$$-2R_{r\theta r\theta} = R_{\theta\phi\theta\phi} \quad \Rightarrow \quad (1 - B^2)' = -\frac{1}{r}(1 - B^2) \quad \Rightarrow \quad 1 - B^2 = \frac{k}{r}$$

$$\Rightarrow \quad B = \sqrt{1 - \frac{k}{r}}$$

for some constant k. The last field equation is then redundant. (As usual, the field equations are related by the Bianchi identity.) The constant k can be related to the nonrelativistic result by comparing at large distances. (See exercise IXB1.1.) We then find $k = 2GM$, so the final result is:

$$\boxed{-ds^2 = -\left(1 - \frac{2GM}{r}\right) dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 + r^2(d\theta^2 + sin^2\theta\ d\phi^2)}$$

Exercise IXC5.2

Repeat this Weyl scale derivation of covariant derivatives and curvatures for the Schwarzschild metric in dimensions D>4. (Hint: Do not use explicit expressions for the covariant derivatives of the higher-dimensional sphere.) Solve for A and B.

More generally, if we have some spherically symmetric, static matter distribution, then the only nonvanishing components of the energy-momentum tensor will be

T_{tt}, T_{rr}, and $T_{\theta\theta} = T_{\phi\phi}$ (representing energy density, radial pressure, and isotropic pressure), all functions of just r. Repeating the above procedure, we integrate

$$[r(1-B^2)]' = 2r^2 T_{tt}, \qquad [ln(AB)]' = -\frac{r}{B^2}(T_{tt} + T_{rr})$$

while the remaining equation is redundant.

For example, for a spherically symmetric, static electromagnetic field the only nonvanishing components of the field strength are F_{tr} and $F_{\theta\phi}$, corresponding to electric and magnetic charges, respectively. Then the invariance of T under a duality transformation (see subsections IIA7, IIIA4) implies

$$T_{\theta\theta} = T_{\phi\phi} \quad \Rightarrow \quad T_{tt} = -T_{rr} \quad \Rightarrow \quad A = B^{-1}$$

again, since on this F_{ab} duality effectively replaces $(\theta, \phi) \leftrightarrow (it, r)$, with the i from Wick rotation. Local scale invariance (see subsection IXA7) then tells us

$$T_a{}^a = 0 \quad \Rightarrow \quad T_{tt} = -T_{rr} = T_{\theta\theta} = T_{\phi\phi}$$

Exercise IXC5.3

Let's rederive these results by brute force:

a Derive T_{ab} for a general electromagnetic field by varying its action with respect to $e_a{}^m$ or g^{mn}.

b Find each of T's components explicitly in terms of F_{tr} and $F_{\theta\phi}$ in the case where those are the only nonvanishing components, and show they appear only in the combination $F_{tr}^2 + F_{\theta\phi}^2$.

As usual, these field strengths can be found easily from the integral form of Gauss' law by integrating over a sphere: For example, for the magnetic field

$$magnetic\ charge \sim \tfrac{1}{2} \int dx^m dx^n \ F_{mn} = 4\pi r^2 F_{\theta\phi}$$

for the $F_{\theta\phi}$ component of F_{ab} (integrating over θ and ϕ), since the metric (and vierbein) for θ and ϕ is the same as for flat space. By duality, the solution for F_{tr} in terms of the electric charge is the same. The result is

$$T_{tt} = \frac{Q}{r^4}, \qquad Q = \pi^2(e^2 + g^2)$$

for electric charge e and magnetic charge g. The $1/r^4$ dependence also follows from scale invariance, since the charges are dimensionless (and the matter field equations decouple from A and B). (Again, since the solution does not extend to $r = 0$, we normalize by comparing F_{ab} or T_{ab} at $r = \infty$ to the flat-space solution.) The net effect on the Schwarzschild metric is

$$1 - \frac{2GM}{r} \to 1 - \frac{2GM}{r} + \frac{2Q}{r^2}$$

Our solution relates to the usual mechanics normalization of the charges (see subsection VIIA3), restoring G, as

$$2Q = G2\pi(e^2 + g^2) = G(e_m^2 + g_m^2)$$

Exercise IXC5.4

Let's also apply brute force to solving Maxwell's equations $\nabla_a F^{ab} = \nabla_{[a} F_{bc]} = 0$ (outside the matter).

a As a warm-up, using directly the above covariant derivatives, show that *in flat space*

$$V^a = \delta_r^a V_r \quad \Rightarrow \quad \nabla_a J^a = r^{-2}\partial_r r^2 J_r$$

Note that the covariant derivative of a vanishing component doesn't necessarily vanish (just as the ordinary derivative of a function that vanishes at some point doesn't necessarily vanish at that point): Components of ∇ other than ∇_r contain Lorentz generators that rotate other components of V to V_r.

b Solve Maxwell's equations in differential form for F_{ab} in the above case. Use the empty-space solution to define the normalization at infinity. (Actually, in this case, the charge is well-defined in terms of the flux of the fields, as described above, but gives the same result here because the space is asymptotically flat.)

Exercise IXC5.5

Spherically symmetric solutions can also be written in Eddington-Finkelstein coordinates as

$$-ds^2 = -dt^2 + dr^2 + r^2(d\theta^2 + sin^2\theta \ d\phi^2) + H(r)(dt + dr)^2$$

where $H = 1 - B^2 = G[2M/r - 2\pi(e^2 + g^2)/r^2]$ in terms of the above results, and is linear in G. (In these coordinates, gravity looks Abelian for this solution.) Note that this form (and its Abelian nature) closely resembles the general wave solutions of exercise IXC1.2b (but the "Cartesian" coordinate x^1 is now replaced with r).

a Obtain this form from the above forms (where $A = B^{-1}$) by a coordinate transformation. (Hint: The angular term didn't change.)

b Find ∇ directly from this form of the metric. (Note: It might differ from the previous by not only coordinate but also local Lorentz transformations.)

Exercise IXC5.6

Consider the plane wave in the coordinates

$$-ds^2 = -2dx^+dx^- + L^2(x^-)\left(e^{2\beta(x^-)}dy^2 + e^{-2\beta(x^-)}dz^2\right)$$

Calculate the covariant derivatives and curvature tensor by the first method of this subsection (double-counting and subtracting, not the Weyl scale method). Show that the field equations reduce to

$$L'' + (\beta'^2)L = 0$$

Exercise IXC5.7

Use the first method of this subsection to calculate the covariant derivative and curvature tensor for the metric

$$-ds^2 = -dt^2 + 2e^x dt\ dy - \tfrac{1}{2}e^{2x}dy^2 + dx^2 + dz^2$$

Show that this metric satisfies the field equations with a cosmological term for a dust at rest with respect to this time coordinate; i.e.

$$R^{mn} - \tfrac{1}{2}g^{mn}(R - 4\Lambda) = \rho\delta_0^m\delta_0^n$$

where Λ and ρ are both constants.

Exercise IXC5.8

Use this method to calculate the covariant derivative and curvature tensor for the cylindrically symmetric metric

$$-ds^2 = -A^{-2}(r)dt^2 + B^{-2}(r)dr^2 + r^2d\theta^2 + dz^2$$

Assume the matter in this problem is a “perfect fluid”,

$$T^{ab} = \rho u^a u^b + P(\eta^{ab} + u^a u^b) \qquad (u^2 = -1)$$

Solve the equations of motion for the gravitational field to find A and B, as well as the pressure P and particle density ρ. What is the implied relation between P and ρ?

Exercise IXC5.9

Use this method to calculate the covariant derivative and curvature tensor for the following metric, corresponding to that outside a planar mass distribution:

$$-ds^2 = -A^{-2}(z)dt^2 + B^{-2}(z)(dx^2 + dy^2) + dz^2$$

Solve Einstein's equations in empty space to find A and B (up to some constants of integration).

6. Experiments

When comparing to the real world, it is useful to know some astrophysical radii:

(1) Earth's orbit (1 AU): 1.5 $\times 10^8$ km

(2) Solar radius: 7 $\times 10^5$ km

(3) Earth radius: 6000 km

(4) Solar gravitational (Schwarzschild) radius ($2GM$): 3 km

(5) Earth gravitational radius: 0.9 cm (1 shoe size).

Exercise IXC6.1

Consider the following very crude approximations to a star: Assume a star has the density of a hydrogen atom (i.e., of a sphere with the mass of the hydrogen atom and radius equal to the Bohr radius). Assume also that the radius of this (spherical) star is equal to its gravitational radius. (This is roughly a "white dwarf".) Find the mass and radius, in terms of both physical constants and conventional units. Do the same assuming the density of a neutron (a sphere with the mass of a neutron and radius equal to $\hbar/m$.) (This is roughly a "neutron star".) Compare to the mass and radius of the Sun.

All experiments (excluding cosmology) are based on the Schwarzschild metric. The first type of experiment involves gravitational redshift, but unlike the cosmological case, the relevant reference frames of observation are not local inertial frames but the static reference frame in which the Schwarzschild metric is defined. (There are also measurements of redshift from airplanes, whose reference frame is defined with respect to the Schwarzschild one.) In this reference frame the relevant Killing vector is the one which expresses the fact that the space is static, $K^m\partial_m = \partial/\partial t$. The momentum which is measured by the observer is p^a, not p^m or p_m, since the observer still uses a reference frame for which the metric at his position is flat (but not its first derivative, since he is not in free fall). (In fact, this is one of the purposes for using a vierbein, as a frame of reference.) The conserved quantity is then

$$E = -K_a p^a = \sqrt{1 - \frac{2GM}{r}}\widehat{E}$$

where the energy of a particle $\widehat{E}$ is the time component of p^a as measured in this frame. Thus, conservation of E for a photon gives the r-dependence of the observed energy $\widehat{E}$ (and thus the frequency, which in turn determines the wavelength, since $p^2 = 0$).

To compare with nonrelativistic mechanics, we instead evaluate E for a massive particle in the Newtonian limit:

$$E \approx \left(1 - \frac{GM}{r}\right)(m + K) \approx m + K - \frac{GMm}{r}$$

giving the "conserved energy" E in terms of the "particle energy" $\widehat{E}$ (rest mass m + kinetic K), including the potential energy.

The other type of experiment involves properties of geodesics, so we need to solve the geodesic equations of motion. Without loss of generality, we can choose the angular coordinates such that the initial position and direction of the particle is in the equatorial plane $\theta = \pi/2$, where it remains because of the symmetry $\theta \leftrightarrow \pi - \theta$, as in the nonrelativistic case. Also as in the nonrelativistic case, we can find constants of the motion corresponding to the energy E and (z-component of) angular momentum L by using the Killing vectors $K^m\partial_m = \partial/\partial t$ and $\partial/\partial\phi$ to find the conserved quantities $K^a p_a = K^m g_{mn}\dot{x}^n$ (in the parametrization $v = 1$):

$$E \equiv -g_{tm}\dot{x}^m = \left(1 - \frac{2GM}{r}\right)\dot{t}, \quad L \equiv g_{\phi m}\dot{x}^m = r^2\dot{\phi}$$

In the case where the particles come from infinity, these are the initial kinetic energy and angular momentum. We have chosen an affine parametrization, which requires

$$-m^2 = g_{mn}\dot{x}^m\dot{x}^n = -\left(1 - \frac{2GM}{r}\right)\dot{t}^2 + \left(1 - \frac{2GM}{r}\right)^{-1}\dot{r}^2 + r^2\dot{\phi}^2$$

Solving the previous equations for $\dot{t}$ and $\dot{\phi}$, this reduces to the radial equation

$$0 = -E^2 + \dot{r}^2 + \left(1 - \frac{2GM}{r}\right)\left(\frac{L^2}{r^2} + m^2\right)$$

$$\Rightarrow \quad \tfrac{1}{2}\dot{r}^2 + \left(-\frac{GMm^2}{r} + \frac{L^2}{2r^2} - \frac{GML^2}{r^3}\right) = \tfrac{1}{2}(E^2 - m^2)$$

This looks like a typical nonrelativistic Hamiltonian for "energy" $\frac{1}{2}(E^2 - m^2)$ with the same terms as in the Newtonian case but with an extra r^{-3} term. (To take the nonrelativistic limit for the massive case, first scale the affine parameter $\tau \to s/m$.) Since there are good coordinate systems for a "black hole" using r as a coordinate (e.g., see the following subsection: r and $r'' + t''$, as seen from the figure for Kruskal-Szkeres), this equation can even be used to descibe a fall into a black hole. (For example, for $L = 0$ we get the same cycloid solution as in cosmology and in Newtonian gravity, reaching the singularity at $r = 0$ in finite proper time.)

Because of the r^{-3} term in the potential, noncircular orbits are no longer closed. In particular, let's consider orbits which are close to circular. Circular orbits are found by minimizing the potential for the r-equation:

$$0 = \frac{dV}{dr} = \frac{GMm^2}{r^2} - \frac{L^2}{r^3} + \frac{3GML^2}{r^4}$$
$$0 < \frac{d^2V}{dr^2} = -\frac{2GMm^2}{r^3} + \frac{3L^2}{r^4} - \frac{12GML^2}{r^5}$$

The near-circular orbits are described by small (harmonic) oscillations about this minimum, with angular frequency given by

$$\omega_r^2 = \frac{d^2V}{dr^2} = \frac{GMm^2(r-6GM)}{r^3(r-3GM)}$$

from solving for $L^2 = GMm^2r^2/(r-3GM)$. On the other hand, the frequency of the circular orbit itself in terms of its angular dependence is just $\dot{\phi} = L/r^2$, giving

$$\omega_\phi^2 = \frac{GMm^2}{r^2(r-3GM)}$$

This means that the perihelion (closest approach to the Sun) of an orbit, which occurs every period $2\pi/\omega_r$ of the radial motion, results in the change of angle

$$2\pi + \delta\phi = \int_0^{2\pi/\omega_r} d\tau \frac{d\phi}{d\tau} = \frac{2\pi}{\omega_r}\omega_\phi = 2\pi\left(1 - 6\frac{GM}{r}\right)^{-1/2}$$
$$\Rightarrow \quad \delta\phi \approx 6\pi\frac{GM}{r}$$

in the weak-field approximation. This effect contributes to the measurement of the precession of the perihelion of the (elliptical) orbit of Mercury, but so do the precession of Earth's axis, the oblateness of the Sun, and gravitational interaction with other planets. As a result, this relativistic effect contributes less than 1% to the observed precession. In particular, the solar oblateness is difficult to measure.

The effects on geodesics of photons are much easier to measure, since there are no Newtonian effects. As a result, the weak field approximation is sufficient. We first consider bending of light by the Sun: A photon comes in from infinity and goes back out to infinity (actually to the Earth, which we assume is much farther from the Sun than the photon's closest approach to it), and we measure what angle its trajectory was bent by. (For example, we look at the apprarent change of position in stars when the Sun passes in their direction during an eclipse.) Starting with the exact solution for a photon's geodesic (case $m^2 = 0$ above), we use the equations for $\dot{r}$ and $\dot{\phi}$ to find

$$\frac{dr}{d\phi} = \sqrt{\frac{E^2}{L^2}r^4 - r^2 + 2GMr}$$

Changing variables,

$$u \equiv \frac{b}{r}, \quad b \equiv \frac{L}{E}, \quad a \equiv \frac{GM}{b} \quad \Rightarrow \quad d\phi = \frac{du}{\sqrt{1-u^2+2au^3}}$$

The *impact parameter* $b \equiv L/E$ would be the closest approach to the Sun neglecting gravitational effects ($L = rp = bE$). We now make the weak field approximation: For a small,

$$\begin{aligned} d\phi &\approx \frac{du}{\sqrt{1-u^2}}\left(1 - a\frac{u^3}{1-u^2}\right) \\ &= d\chi\left(1 - a\frac{sin^3\chi}{cos^2\chi}\right) \qquad (u \equiv sin\ \chi) \\ &= d\left[\chi - a\left(cos\ \chi + \frac{1}{cos\ \chi}\right)\right] \end{aligned}$$

Defining $\phi = 0$ at $r = \infty$, the integral is

$$\phi \approx \chi - a\frac{(1-cos\ \chi)^2}{cos\ \chi} \quad \Rightarrow \quad \chi \approx \phi + a\frac{(1-cos\ \phi)^2}{cos\ \phi}$$

$$\Rightarrow \quad u = sin\ \chi \approx sin\ \phi + a(1-cos\ \phi)^2$$

The change in ϕ from incoming photon to outgoing photon follows from the two solutions for $r = \infty$:

$$u = 0 \quad \Rightarrow \quad \phi = 0,\ \pi + 4a$$

Therefore the deviation of ϕ from a straight line is $4GME/L$. (Mathematical note: All variable changes were those suggested by the flat space case $a = 0$: E.g., $b/r = sin\ \chi$, where χ is what ϕ would be in flat space.)

A similar experiment involves measuring the round-trip travel time for radio waves from Earth to some reflector (on another planet or an artificial solar satellite), with and without the Sun near the path of the waves. Now, instead of $dr/d\phi$ we want, in units $b = 1$

$$\frac{dr}{dt} = \left(1 - \frac{2a}{r}\right)\sqrt{1 - \frac{1}{r^2} + \frac{2a}{r^3}}$$

$$\begin{aligned} \Rightarrow \quad dt &\approx \frac{rdr}{\sqrt{r^2-1}} + 2a\frac{dr}{\sqrt{r^2-1}} \\ &= d\left[\sqrt{r^2-1} + 2a\ cosh^{-1}r\right] \\ &= d\left[\sqrt{r^2-b^2} + 2GM\ cosh^{-1}\frac{r}{b}\right] \end{aligned}$$

putting the b's back. (We have neglected the gravitational effect on $\dot{r}$, which is negligible compared to that on $\dot{t}$, since for most of the path $r \gg b$.) We then integrate from $r = r_{min} \approx b$ to $r = r_{Earth}$, add the integral from $r = r_{min}$ to $r = r_{reflector}$, multiply by 2 for the round trip, and throw in a factor to convert to the proper time s of the observer (which turns out to have a negligible effect to this order in a). This result is then compared to the same measurement when both observer and reflector have revolved further about the Sun, so b changes significantly (but not r_{Earth} nor $r_{reflector}$). The biggest contribution (as seen from $\Delta t/\Delta b$), for $b \ll r_{Earth}$ and $r_{reflector}$, comes from the *cosh* term: For $x \gg 1$, $cosh^{-1}x \approx ln(2x)$, so

$$\Delta s \approx -8GM\Delta(ln\ b)$$

7. Black holes

For physical massive bodies the Schwarzschild solution applies only outside the body, where $T_{ab} = 0$. The form of the solution inside the body depends on the distribution of matter, which is determined by its dynamics. Generally the surface of the body is at $r \gg GM$, but we can try to find a solution corresponding to a point mass by extending the coordinates as far as possible, till the curvature components $R_{ab}{}^{cd}$ blow up. The Schwarzschild metric is singular at $r = 2GM$. In fact, r and t switch their roles as space and time coordinates there. There is no corresponding singularity there in the curvatures, which are $\sim r^{-3}$. This unphysical singularity can be eliminated by first making the coordinate transformation, for $r > 2GM$,

$$r' = \int dr \left(1 - \frac{2GM}{r}\right)^{-1} = r + 2GM\ ln\left(\frac{r}{2GM} - 1\right)$$

and then making a second coordinate transformation by rescaling the "lightcone" coordinates as

$$r'' \pm t'' = 4GMe^{(r'\pm t)/4GM} = 4GM\sqrt{\frac{r}{2GM} - 1}\ e^{(r\pm t)/4GM}$$

The result is the "Kruskal-Szekeres coordinates"

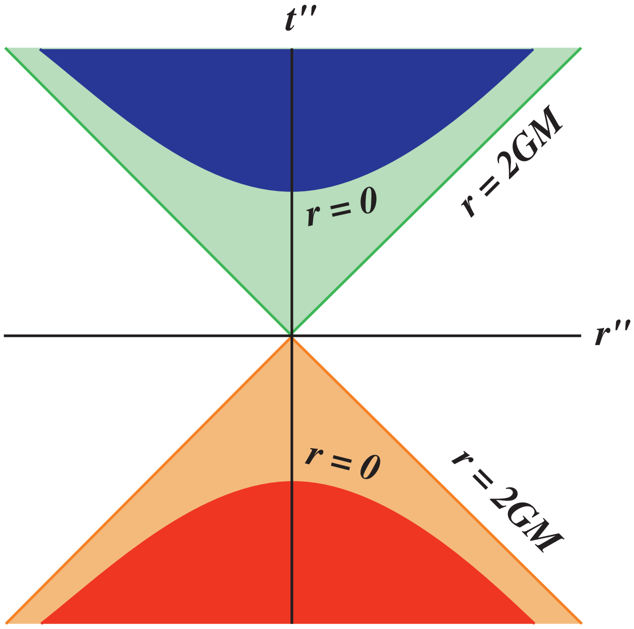
$$-ds^2 = \frac{2GM}{r}e^{-r/2GM}(-dt''^2 + dr''^2) + r^2(d\theta^2 + sin^2\theta\ d\phi^2)$$

where $r(r'', t'')$ is defined by

$$r''^2 - t''^2 = (4GM)^2\left(\frac{r}{2GM} - 1\right)e^{r/2GM}$$

This can now be extended past $r = 2GM$ down to the physical singularity at $r = 0$.

The complete space now looks like (plotting just r'' and t''):



In this diagram lines at 45° to the axes represent radial lightlike geodesics. Since nothing travels faster than light, this indicates the allowed paths of physical objects. Curves of fixed r are hyperbolas: In particular, the physical singularity is the curve $t''^2 - r''^2 = (4GM)^2$ ($r = 0$), while $t''^2 - r''^2 = 0$ ($r = 2GM$) is the "event horizon" which allows things to go only one way (out from the bottom half or into the top half), and $r = \infty$ is both $r'' = \pm\infty$. Nothing can communicate between the 2 "outside worlds" of the left and right 90° wedges. In particular, a star which collapses ("gravitational collapse") inside its "gravitational radius" $2GM$ is crushed to a singularity, and the spherically symmetric approximation to this collapse must be represented by part of the Kruskal-Szekeres solution (outside the star) by Birkhoff's theorem, patched to another solution inside the star representing the contribution of the matter (energy) there to the field equations. This means using just the top and right 90° wedges, with parts near the left edge of this modified appropriately. The top wedge is called a "black hole". (If a situation should exist described by just the bottom and right wedges, the bottom wedge would be called a "white hole".) Similarly, stable stars are described by just the right wedge, patched to some interior solution. This right wedge represents the original Schwarzschild solution in the region $r > 2GM$ where its coordinates are nonsingular. In that region lines of constant t are just "straight" radial lines in the Kruskal-Szekeres coordinate system ($r'' \sim t''$).

Besides the fact that nothing can get out, another interesting feature of the black hole is that an outside observer never sees something falling in actually reach the event horizon: Consider an observer at fixed $r > 2GM$ using Schwarzschild coordinates, so his proper time $s \sim t$. Then light radiating radially from an in-falling object is received later and later, up till $t = \infty$, by the observer as the object approaches the event horizon, although it takes the object a finite amount of proper time to reach the event horizon and the physical singularity.

Exercise IXC7.1

Apply the methods of subsection IXC3 to the equations of motion in a Schwarzschild metric of subsection IXC6 for a massive object falling straight into a black hole (angular momentum $L = 0$): Solve for r, t, τ in an appropriate parametrization to show that it takes a finite proper time to reach the event horizon from any finite r outside it.

There are also more complicated black-hole solutions with spin and electric charge.

Another interesting effect of the event horizon is the eventual decay of the black hole ("Hawking radiation"): Pair creation can result in a similar way to that in an electrostatic potential of sufficient strength (see exercise IIIB5.1). Particles are emitted near the event horizon (the edge of the gravitational barrier), carrying energy off to infinity, while their antiparticles fall into the singularity.

There are two features of the black hole that are less than desirable: the existence of singularities indicates a breakdown in the field equations, and the existence of event horizons results in an "information loss". Both these properties might be avoidable quantum mechanically: For example, quantum effects can generate curvature-squared terms in the effective action, which modify the short-distance behavior of the theory. One might think that such short-distance effects would have an effect only at short distances away from regions of high curvature such as the singularity, and thus remove the singularities but not the event horizons. However, it is possible (and examples of such solutions have been given) that the prevention of the creation of the singularity in stellar collapse would eventually result in a reversal of the collapse ("gravitational bounce"): The would-be black hole solution is patched to a would-have-been white hole by short-distance modifications, resulting in an exploding star that initially resembled a black hole but has no true event horizon.

REFERENCES

1 H.W. Brinkmann, *Math. Annalen* **94** (1925) 119;
I. Robinson, unpublished lectures (1956);
J. Hély, *Compt. Rend. Acad. Sci.* **249** (1959) 1867;
A. Peres, hep-th/0205040, *Phys. Rev. Lett.* **3** (1959) 571;
J. Ehlers and W. Kundt, Exact solutions of gravitational field equations, in *Gravitation: An introduction to current research*, ed. L. Witten (Wiley, 1962) p. 49:
nonlinear gravity waves.

2 J.F. Plebański, *J. Math. Phys.* **16** (1975) 2395:
reduction of self-dual metric to single component.
3 Siegel, *loc. cit.* (IVC, 2nd ref. 17):
lightcone gauge for self-dual supergravity.
4 W. de Sitter, *Proc. Kon. Ned. Akad. Wet.* **19** (1917) 1217, **20** (1917) 229.
5 A. Friedmann, *Z. Phys.* **10** (1922) 377;
H.P. Robertson, *Astrophys. J.* **82** (1935) 284, **83** (1936) 187, 257;
A.G. Walker, *Proc. Lond. Math. Soc.* **42** (1936) 90:
cosmology in general relativity.
6 Particle Data Group, *loc. cit.* (IC):
astrophysical "constants".
7 K. Schwarzschild, *Sitz. Preuss. Akad. Wiss. Berlin, Math.-phys. Kl.* (1916) 189.
8 A.S. Eddington, *Nature* **113** (1924) 192;
D. Finkelstein, *Phys. Rev.* **110** (1958) 965.
9 M.D. Kruskal, *Phys. Rev.* **119** (1960) 1743;
G. Szekeres, *Publ. Mat. Debrecen* **7** (1960) 285.
10 S.W. Hawking, *Comm. Math. Phys.* **43** (1975) 199.
11 V.P. Frolov and G.A. Vilkoviskii, *Phys. Lett.* **106B** (1981) 307:
possibility that quantum corrections to gravity eliminate black holes.

X. SUPERGRAVITY

In the previous chapter we studied the symmetry principles behind general relativity; now we add supersymmetry to the picture. Supergravity is a fundamental part of many of the applications of supersymmetry.

A. SUPERSPACE

We first need to understand the "geometry" associated with local supersymmetry.

1. Covariant derivatives

In subsection IVC3 we discussed superspace covariant derivatives for super Yang-Mills. Similar methods can be applied to supergravity, the theory of the graviton (spin 2) and gravitino (spin 3/2). In that case we want to gauge the complete (unbroken) global symmetry of the theory: Besides the obvious Poincaré and supersymmetry, there is also the axial U(1) ("R") symmetry that transforms the spin-3/2 field. (The best we might have expected is superconformal symmetry, which also has conformal boosts and scale, but which are broken by the vacuum just as in ordinary gravity, and S-supersymmetry, which is also broken because it's the square root of conformal boosts.) It is introduced in the same way as local Lorentz invariance in ordinary gravity, and acts on flat spinor indices (but cancels on vector indices). We therefore want to gauge the translations ∂_M (which have been generalized naturally to superspace from ∂_m appearing in ordinary gravity to include supersymmetry), the Lorentz generators $M_{\alpha\beta}$, $M_{\dot\alpha\dot\beta}$ of ordinary gravity, and the (second-quantized) *hermitian* U(1) generator Y, defined to act on the covariant derivatives as

$$[Y, \nabla_\alpha] = -\tfrac{1}{2}\nabla_\alpha, \qquad [Y, \overline{\nabla}_{\dot\alpha}] = \tfrac{1}{2}\overline{\nabla}_{\dot\alpha}, \qquad [Y, \nabla_a] = 0$$

We now use the "$\overline{}$" to refer to hermitian conjugation without reordering, i.e., keeping the partial derivatives and other generators on the right. (As in ordinary gravity, transformations are not truly unitary, and covariant derivatives truly hermitian, because of ordering.)

Then the gauge parameter, covariant derivative, and field strengths are expanded over these generators, as in ordinary gravity:

$$K = K^M\partial_M + \tfrac{1}{2}K^{\alpha\beta}M_{\beta\alpha} + \tfrac{1}{2}K^{\dot\alpha\dot\beta}M_{\dot\beta\dot\alpha} + iK_{-1}Y$$

$$\nabla_A = E_A{}^M\partial_M + \tfrac{1}{2}\Omega_A{}^{\beta\gamma}M_{\gamma\beta} + \tfrac{1}{2}\Omega_A{}^{\dot\beta\dot\gamma}M_{\dot\gamma\dot\beta} + iA_AY$$

$$[\nabla_A, \nabla_B\} = T_{AB}{}^C \nabla_C + \tfrac{1}{2} R_{AB}{}^{\gamma\delta} M_{\delta\gamma} + \tfrac{1}{2} R_{AB}{}^{\dot\gamma\dot\delta} M_{\dot\delta\dot\gamma} + i F_{AB} Y$$

($E_A{}^M$ is known as the "supervierbein" or "vielbein".) Alternatively, we can write the M and Y terms collectively as $\frac{1}{2} K^{AB} M_{BA}$ (and similarly for the covariant derivative and field strengths), where M_{AB} are the generators of OSp(3,1|4), by algebraically constraining K^{AB} to contain just the appropriate pieces (and relating K^{ab} to $K^{\alpha\beta}$ in the usual way). Also, the shorthand $K^I M_I$ now includes Lorentz and U(1) terms.

Exercise XA1.1

Use the definition in the above commutation relations to express the torsion $T_{AB}{}^C$ directly in terms of the structure functions $C_{AB}{}^C$, Lorentz connection $\Omega_A{}^{\beta\gamma}$, and U(1) connection A_A.

The constraints in supergravity are a combination of the kinds used in ordinary gravity and super Yang-Mills: those that (1) define the vector derivative in terms of the spinor ones

$$-i\nabla_{\alpha\dot\beta} = \{\nabla_\alpha, \overline{\nabla}_{\dot\beta}\}$$

(2) define the spinor (Lorentz and R) connections

$$T_{\alpha\beta}{}^\gamma = T_{\alpha,\beta(\dot\beta}{}^{\beta\dot\gamma)} = T_{\alpha b}{}^b = 0$$

and (3) allow the existence of chiral (scalar) superfields

$$\nabla_\alpha \bar\Phi = 0 \quad \Rightarrow \quad \{\nabla_\alpha, \nabla_\beta\}\bar\Phi = 0 \qquad (Y\Phi = 0)$$

(The first two constraints imply the generalization of this chirality condition to $Y \neq 0$ and chiral superfields with undotted indices, like the Yang-Mills field strength.)

Exercise XA1.2

Rewrite the first and last set of constraints directly in terms of field strengths.

The explicit solution of all these constraints is a bit messy, but we will need only a certain subset of them to find the prepotentials and supergravity action. The form of the solution is a generalization of super Yang-Mills in a way similar to how general relativity generalizes ordinary Yang-Mills. In particular, just as the vierbein $e_a = e_a{}^m \partial_m$ is a generalization of the Yang-Mills vector A_a to describe gauging of the translations, the generalization of the super Yang-Mills prepotential V to supergravity is $H = H^m(-i)\partial_m$, which appears in an exponential e^H just as V appears as e^V: The chirality-preserving constraints, expressed explicitly in terms of the vielbein, is

$$\{E_\alpha, E_\beta\} = C_{\alpha\beta}{}^\gamma E_\gamma$$

If the commutator vanished like the Yang-Mills case, E_α would be partial derivatives on some complex two-dimensional subspace, the usual ∂_α up to some complex (super)coordinate transformation, as for d_α in flat superspace. However, the fact that their algebra still closes means they still generate translations within such a subspace, and are thus linear combinations of such partial derivatives:

$$E_\alpha = \psi N_\alpha{}^\mu \hat{E}_\mu, \qquad \hat{E}_\mu = e^{-\Omega}\partial_\mu e^{\Omega}, \qquad \Omega = \Omega^M(-i)\partial_M$$

where we have separated the matrix coefficient into a local complex scale (scale $\oplus$ U(1)) ψ and a local Lorentz transformation $N_\alpha{}^\mu$. For most purposes we will find it convenient to fix all these invariances by choosing the gauge

$$\psi N_\alpha{}^\mu = \delta_\alpha^\mu \quad \Rightarrow \quad E_\alpha = \hat{E}_\alpha$$

As for Yang-Mills, solution of the chirality condition introduces a new, chiral gauge invariance:

$$e^{\Omega'} = e^{i\bar{\Lambda}}e^{\Omega}e^{-iK}; \qquad \Lambda = \Lambda^M(-i)\partial_M, \quad K = K^M(-i)\partial_M$$

$$[\bar{\partial}_{\dot{\mu}}, \Lambda] \sim \bar{\partial}_{\dot{\nu}} \quad \Rightarrow \quad \bar{\partial}_{\dot{\mu}}\Lambda^m = \bar{\partial}_{\dot{\mu}}\Lambda^\mu = 0$$

where $\Lambda^{\dot{\mu}}$ is not chiral, since it generates terms in the transformation law of $E_{\dot{\mu}}$ that can be canceled by including $\bar{\partial}_{\dot{\mu}}\Lambda^{\dot{\nu}}$ terms in the transformation law of $\bar{\psi}\bar{N}_{\dot{\alpha}}{}^{\dot{\mu}}$. This means we can use $\Lambda^{\dot{\mu}}$ and K^μ to gauge

$$\Omega^\mu = \Omega^{\dot{\mu}} = 0 \quad \Rightarrow \quad \Omega = \Omega^m(-i)\partial_m \quad \Rightarrow \quad \hat{E}_\mu = \partial_\mu + \hat{E}_\mu{}^m\partial_m$$

where $\hat{E}_\mu{}^m = -i\partial_\mu\Omega^m + ...$ from expanding the exponentials as multiple commutators. We can again transform to a chiral representation, and work in terms of

$$e^U = e^{\Omega}e^{\bar{\Omega}}, \qquad e^{U'} = e^{i\bar{\Lambda}}e^U e^{-i\Lambda}; \qquad \hat{E}_{\dot{\mu}} = \partial_{\dot{\mu}}, \qquad \hat{E}_\mu = e^{-U}\partial_\mu e^U$$

where U now generalizes the constant $\langle U\rangle = \theta^\mu\bar{\theta}^{\dot{\mu}}(-i)\partial_{\mu\dot{\mu}}$ used in flat superspace. Also as for Yang-Mills, the usual local component transformations (now for coordinate, supersymmetry, scale, U(1), and S-supersymmetry) reappear in the chiral parameters Λ^m and Λ^μ.

For a component analysis, we look at the linearized transformation (see exercise IVC4.3)

$$\delta U^m \approx i(\bar{\Lambda} - \Lambda)^m - i\tfrac{1}{2}[\langle U\rangle, \bar{\Lambda} + \Lambda]^m$$

$$= i(\bar{\Lambda} - \Lambda)^m - \tfrac{1}{2}\theta^\nu\bar{\theta}^{\dot{\nu}}\partial_{\nu\dot{\nu}}(\bar{\Lambda} + \Lambda)^m + (\theta^\mu\bar{\Lambda}^{\dot{\mu}} - \bar{\theta}^{\dot{\mu}}\Lambda^\mu) + i\tfrac{1}{2}(\bar{\theta}^2\theta^\nu\partial_\nu{}^{\dot{\mu}}\Lambda^\mu - \theta^2\bar{\theta}^{\dot{\nu}}\partial^\mu{}_{\dot{\nu}}\bar{\Lambda}^{\dot{\mu}})$$

where for Λ we use as independent just the chiral parameters Λ^m and Λ^μ; the nonchiral $\Lambda^{\dot\mu}$, having already been used to gauge away U^μ, is now fixed in terms of the others as

$$\Lambda^{\dot\mu} = e^{-U}\bar\Lambda^{\dot\mu}e^U$$

to maintain $U^\mu = 0$. The first term in the transformation tells us that the surviving component fields are the same as for super Yang-Mills, with "m" as the group index:

$$U^m = e_a{}^m(\theta\bar\theta), \psi_\alpha{}^m(\bar\theta^2\theta), \bar\psi_{\dot\alpha}{}^m(\theta^2\bar\theta), A^m(\theta^2\bar\theta^2)$$

The second term in the transformation law gives $\delta e_a{}^m \approx -\partial_a\lambda^m$ from $\bar\Lambda^m| = \Lambda^m| = \lambda^m$. Then Λ^μ contains the rest of the gauge parameters:

$$\begin{aligned}\Lambda^\mu &= \epsilon^\mu, a+ib(\theta), \lambda_\nu{}^\mu(\theta), \zeta^\mu(\theta^2)\\ &= \text{supersymmetry, scale +iU(1), Lorentz, S-supersymmetry}\end{aligned}$$

The third term in the transformation law then shows scale and Lorentz gauge away pieces of the vierbein, as usual, while S-supersymmetry gauges away the trace of $\psi_\alpha{}^m$. It also forces Λ^m to include $\bar\epsilon^{\dot\mu}$ at order θ to maintain the gauge; we then see that $\psi_\alpha{}^m$ is the gauge field for supersymmetry, with contributions from the second and fourth terms. Finally, the fourth term also shows that A^m is the gauge field for U(1). The resulting component content is that of "conformal supergravity", which will be transformed later to ordinary supergravity through a compensator superfield.

For perturbation theory, or comparison with global supersymmetry, we should expand about "flat" superspace (which is nontrivial because of nonvanishing torsion $T_{\alpha\dot\beta}{}^c$ in empty superspace). We then modify the chiral representation:

$$e^{\Omega}e^{\bar\Omega} = e^U = e^{\langle U\rangle/2}e^H e^{\langle U\rangle/2} \quad \Rightarrow \quad \hat E_{\dot\mu} = d_{\dot\mu}, \qquad \hat E_\mu = e^{-H}d_\mu e^H$$

We now expand all derivatives over the covariant derivatives d_M of global supersymmetry (constructed from $\langle U\rangle$ as before)

$$H = H^M(-i)d_M, \qquad E_A = E_A{}^M d_M$$

instead of over partial derivatives ∂_M, which is just a change of basis. This also modifies the description of the Λ gauge parameters:

$$e^{H'} = e^{i\bar\Lambda}e^U e^{-i\Lambda}; \qquad \Lambda = \Lambda^M(-i)d_M, \quad K = K^M(-i)d_M$$

$$\begin{aligned}[\bar d_{\dot\mu}, \Lambda] \sim \bar d_{\dot\nu} \quad &\Rightarrow \quad \Lambda^\mu d_\mu + \Lambda^m\partial_m = \tfrac{1}{2}\{\bar d^{\dot\nu}, [\bar d_{\dot\nu}, L^\mu d_\mu]\}\\ &\Rightarrow \quad \Lambda^\mu = \bar d^2 L^\mu, \qquad \Lambda^{\mu\dot\mu} = i\bar d^{\dot\mu}L^\mu\end{aligned}$$

in terms of a new parameter L^μ. From this we find the linearized transformation law

$$\delta H^{\mu\dot{\mu}} \approx d^\mu \bar{L}^{\dot{\mu}} - \bar{d}^{\dot{\mu}} L^\mu$$

Exercise XA1.3

Expand this transformation law in components, and compare with the previous analysis.

Besides chirality, we also need a certain combination of the other constraints:

$$0 = T_{\alpha b}{}^b - T_{\alpha\dot{\beta}}{}^{\dot{\beta}} = (-1)^B C_{\alpha B}{}^B + C_{\alpha\beta}{}^\beta - iA_\alpha$$

where the A term comes from the contribution $\frac{1}{2}iA_\alpha\delta_{\dot{\beta}}^{\dot{\gamma}}$ to $T_{\alpha\dot{\beta}}{}^{\dot{\gamma}}$. (See exercise XA1.1.) Using the gauge $E_\alpha = \hat{E}_\alpha$ without loss of generality, we then find (comparing similar manipulations in subsection IXA2)

$$-iA_\alpha = E\partial_M E^{-1} E_\alpha{}^M = E^{-1}\overleftarrow{E}_\alpha E$$

where the backwards arrow on $\overleftarrow{E}_\alpha = E_\alpha{}^M \overleftarrow{\partial}_M$ means all derivatives act on everything to the left (see subsection IA2), and

$$E \equiv sdet\ E_A{}^M$$

(The superdeterminant was defined in subsection IIC3.)

We now need the general identity, for any function A and first-order differential operator B,

$$\begin{aligned} Ae^{\overleftarrow{B}} &= (1 \cdot e^{\overleftarrow{B}} e^{-\overleftarrow{B}})Ae^{\overleftarrow{B}} = (1 \cdot e^{\overleftarrow{B}})(e^{-\overleftarrow{B}}Ae^{\overleftarrow{B}}) = (1 \cdot e^{\overleftarrow{B}})(e^B A e^{-B}) \\ &= (1 \cdot e^{\overleftarrow{B}})(e^B A) \end{aligned}$$

$$\Rightarrow \quad 1 = (1 \cdot e^{-\overleftarrow{B}})e^{\overleftarrow{B}} = (1 \cdot e^{\overleftarrow{B}})[e^B(1 \cdot e^{-\overleftarrow{B}})]$$

The final result in the gauge $\psi = 1$ ($N_\alpha{}^\mu$ is trivially restored) is then

$$iA_\alpha = E_\alpha T, \qquad e^T = E(1 \cdot e^{-\overleftarrow{\Omega}})$$

This can be used to solve chirality conditions on matter fields: In this gauge, we have

$$\begin{aligned} Y\Phi = y\Phi \quad &\Rightarrow \quad 0 = \overline{\nabla}_{\dot{\alpha}}\Phi = (\bar{E}_{\dot{\alpha}} + iy\bar{A}_{\dot{\alpha}})\Phi \\ &\Rightarrow \quad \Phi = e^{y\bar{T}} e^{\bar{\Omega}}\phi, \qquad \bar{\partial}_{\dot{\mu}}\phi = 0 \end{aligned}$$

Again as for Yang-Mills, the chiral-representation field ϕ transforms under only the Λ transformations:

$$\phi' = (1 \cdot e^{i\overleftarrow{\Lambda}})^y e^{i\Lambda}\phi, \qquad \Lambda = -i(\Lambda^m \partial_m + \Lambda^\mu \partial_\mu)$$

Thus, scalars in the real representation become densities in the chiral representation (except for $y = 0$). In particular, we have for the special case

$$y = 1 \quad \Rightarrow \quad \phi' = \phi e^{i\overleftarrow{\Lambda}} \quad \Rightarrow \quad \delta \int dx\ d^2\theta\ \phi = 0$$

which will prove useful later for chiral integration. For now, we note that such a chiral scalar, with $y \neq 0$, can be seen to compensate from the $\partial_\mu \Lambda^\mu$ term: This term allows U^m to eat the complex "physical" scalar and spinor, fixing scale, U(1), and S-supersymmetry, while the complex auxiliary scalar survives, along with coordinate, Lorentz, and supersymmetry invariance. We also note that for perturbation about flat superspace we have

$$i(1 \cdot \overleftarrow{\Lambda}) = \partial_m \Lambda^m - d_\mu \Lambda^\mu = -\bar{d}^2 d_\mu L^\mu$$

Exercise XA1.4

Show that preservation of the chirality of ϕ implies the previous chirality conditions on Λ. Thus, as for nonsupersymmetric or nongravitational theories, the gauge group follows more simply from starting with matter representations.

We also note that in the gauge $\psi = 1$, and also $\Omega^\mu = \Omega^{\dot{\mu}} = 0$, the superdeterminant is simply (see subsection IIC3)

$$E^{-1} = det(E_m{}^a)$$

where $E_m{}^a$ is a component of $E_M{}^A$ (not the inverse of $E_a{}^m$).

2. Field strengths

These constraints can be completely solved for all the field strengths. Alternatively, we can impose them, together with the Bianchi identities (Jacobi identities of the covariant derivatives), to find a smaller set of algebraically independent field strengths, and the differential equations that relate them. The method is analogous to the case of super Yang-Mills treated in subsection IVC3. We begin with the constraints analogous to the Yang-Mills ones:

$$\{\nabla_\alpha, \overline{\nabla}_{\dot{\beta}}\} = -i\nabla_{\alpha\dot{\beta}}, \qquad \{\nabla_\alpha, \nabla_\beta\} = R_{\alpha\beta}{}^I M_I$$

(where the latter will simplify from later results). From just the latter, we find

$$[\nabla_{(\alpha},\{\nabla_\beta,\nabla_{\gamma)}\}]=0 \quad\Rightarrow\quad R_{(\alpha\beta\gamma)}{}^\delta - i\tfrac{1}{2}\delta_{(\alpha}{}^\delta F_{\beta\gamma)} = \nabla_{(\alpha}R_{\beta\gamma)}{}^I = 0$$

Using both constraints, we also have

$$[\nabla_{(\alpha},\{\nabla_{\beta)},\overline{\nabla}_{\dot\gamma}\}]+[\overline{\nabla}_{\dot\gamma},\{\nabla_\alpha,\nabla_\beta\}]=0$$

$$\Rightarrow\quad [\nabla_{(\alpha},\nabla_{\beta)\dot\gamma}] = iR_{\alpha\beta\dot\gamma}{}^{\dot\delta}\overline{\nabla}_{\dot\delta} - \tfrac{1}{2}F_{\alpha\beta}\overline{\nabla}_{\dot\gamma} - i(\overline{\nabla}_{\dot\gamma}R_{\alpha\beta}{}^I)M_I$$

$$\Rightarrow\quad [\nabla_\alpha,\nabla_{\beta\dot\gamma}] = -iC_{\alpha\beta}\overline{\mathcal{W}}_{\dot\gamma} + \tfrac{1}{2}iR_{\alpha\beta\dot\gamma}{}^{\dot\delta}\overline{\nabla}_{\dot\delta} - \tfrac{1}{4}F_{\alpha\beta}\overline{\nabla}_{\dot\gamma} - \tfrac{1}{2}i(\overline{\nabla}_{\dot\gamma}R_{\alpha\beta}{}^I)M_I$$

for some operator $\overline{\mathcal{W}}_{\dot\gamma}=\overline{\mathcal{W}}_{\dot\gamma}{}^A\nabla_A+\overline{\mathcal{W}}_{\dot\gamma}{}^I M_I$. So far the exercise has been analogous to the super Yang-Mills case (where the extra "i" in the definition of $\overline{\mathcal{W}}$ is due to our use of antihermitian generators, except for Y). Now we impose the remaining constraints, which can be combined conveniently as

$$0 = T_{\alpha,\beta\dot\gamma}{}^{\beta\dot\delta} = -iC_{\alpha\beta}\overline{\mathcal{W}}_{\dot\gamma}{}^{\beta\dot\delta} \quad\Rightarrow\quad \overline{\mathcal{W}}_{\dot\alpha} = \overline{\mathcal{W}}_{\dot\alpha}{}^\beta\nabla_\beta + \overline{\mathcal{W}}_{\dot\alpha}{}^{\dot\beta}\overline{\nabla}_{\dot\beta} + \overline{\mathcal{W}}_{\dot\alpha}{}^I M_I$$

Following again the steps for Yang-Mills, we analyze the next-higher-dimension Jacobis, beginning with

$$0 = \{\nabla_{(\alpha},[\nabla_{\beta)},\nabla_{\gamma\dot\delta}]\} + [\nabla_{\gamma\dot\delta},\{\nabla_\alpha,\nabla_\beta\}] = iC_{\gamma(\alpha}\{\nabla_{\beta)},\overline{\mathcal{W}}_{\dot\delta}\} + \mathit{\Delta}_{\gamma\alpha\beta\dot\delta}$$

$$\begin{aligned}\mathit{\Delta}_{\gamma\alpha\beta\dot\delta} =& \tfrac{1}{2}i(\nabla_{(\alpha}R_{\beta)\gamma\dot\delta}{}^{\dot\epsilon})\overline{\nabla}_{\dot\epsilon} - \tfrac{1}{4}(\nabla_{(\alpha}F_{\beta)\gamma})\overline{\nabla}_{\dot\delta} - \tfrac{1}{2}i(\nabla_{(\alpha}\overline{\nabla}_{\dot\delta}R_{\beta)\gamma}{}^I)M_I\\ &+\tfrac{1}{2}R_{\gamma(\alpha\dot\delta}{}^{\dot\epsilon}\nabla_{\beta)\dot\epsilon} + \tfrac{1}{4}iF_{\gamma(\alpha}\nabla_{\beta)\dot\delta} - \tfrac{1}{2}i(\overline{\nabla}_{\dot\delta}R_{\gamma(\alpha\beta)}{}^\delta)\nabla_\delta - \tfrac{1}{4}(\overline{\nabla}_{\dot\delta}F_{\gamma(\alpha})\nabla_{\beta)}\\ &+(\nabla_{\gamma\dot\delta}R_{\alpha\beta}{}^I)M_I - R_{\alpha\beta\gamma}{}^\delta\nabla_{\delta\dot\gamma} - R_{\alpha\beta\dot\delta}{}^{\dot\epsilon}\nabla_{\gamma\dot\epsilon}\end{aligned}$$

By inspection, or applying the previous Jacobis, we see

$$\mathit{\Delta}_{(\gamma\alpha\beta)\dot\delta}=0 \quad\Rightarrow\quad \mathit{\Delta}_{\gamma\alpha\beta\dot\delta} = C_{\gamma(\alpha}\mathit{\Delta}_{\beta)\dot\delta}, \qquad \mathit{\Delta}_{\alpha\dot\beta} = -\tfrac{1}{3}\mathit{\Delta}^\gamma{}_{\gamma\alpha\dot\beta}$$

automatically, so the only new information comes from the trace of this Jacobi,

$$\{\nabla_\alpha,\overline{\mathcal{W}}_{\dot\beta}\} = i\mathit{\Delta}_{\alpha\dot\beta}$$

Evaluating $\{\nabla,\overline{\mathcal{W}}\}$ in terms of its pieces, we find

$$R_{\alpha\beta}{}^{\dot\gamma\dot\delta} = F_{\alpha\beta} = 0, \qquad R_{\alpha\beta}{}^{\gamma\delta} = \delta_{(\alpha}^\gamma\delta_{\beta)}^\delta\bar B, \qquad \overline{\mathcal{W}}_{\dot\alpha}{}^{\dot\beta} = -\bar B\delta_{\dot\alpha}^{\dot\beta}$$

$$\overline{W}_{\dot\alpha} = \overline{\nabla}_{\dot\alpha}\bar B + \nabla^\beta\overline{\mathcal{W}}_{\dot\alpha\beta}, \qquad \overline{\mathcal{W}}_{\dot\alpha}{}^{\beta\gamma} = -\tfrac{1}{2}\nabla^{(\beta}\overline{\mathcal{W}}_{\dot\alpha}{}^{\gamma)}$$

$$\nabla_\alpha\overline{W}_{\dot\beta} = \nabla_\alpha\overline{\mathcal{W}}_{\dot\beta}{}^{\dot\gamma\dot\delta} = 0, \qquad \nabla_\alpha\overline{\mathcal{W}}_{\dot\beta}{}^{\gamma\delta} = -\delta_\alpha^{(\gamma}(\overline{\mathcal{W}}_{\dot\beta}{}^{\delta)} + \tfrac{1}{2}i\nabla^{\delta)}{}_{\dot\beta})\bar B$$

where W_α is the Y part of $\mathcal{W}_\alpha$ $(=W_\alpha iY+...)$.

Exercise XA2.1

Show that

$$\{\nabla_\alpha, \nabla_\beta\} = \bar{B} M_{\alpha\beta}, \qquad \nabla_{[\alpha}\nabla_\beta\nabla_{\gamma]} = 0 \quad \Rightarrow \quad \nabla_\alpha(\nabla^2 + \bar{B}) = -\tfrac{1}{2}\bar{B}\nabla^\beta M_{\beta\alpha}$$

and thus $\overline{\nabla}^2 + B$ gives a chiral superfield when acting on any superfield without dotted indices.

For the other Jacobi of this dimension, we have

$$0 = [\nabla_{\alpha\dot{\alpha}}, \{\nabla_\beta, \overline{\nabla}_{\dot{\gamma}}\}] + \{\nabla_\beta, [\overline{\nabla}_{\dot{\gamma}}, \nabla_{\alpha\dot{\alpha}}]\} + \{\overline{\nabla}_{\dot{\gamma}}, [\nabla_\beta, \nabla_{\alpha\dot{\alpha}}]\}$$

$$= -i[\nabla_{\alpha\dot{\alpha}}, \nabla_{\beta\dot{\gamma}}] - i\{\nabla_\beta, C_{\dot{\gamma}\dot{\alpha}}\mathcal{W}_\alpha + \tfrac{1}{2}(\nabla_\alpha B)\overline{M}_{\dot{\gamma}\dot{\alpha}}\} - i\{\overline{\nabla}_{\dot{\gamma}}, C_{\beta\alpha}\overline{\mathcal{W}}_{\dot{\alpha}} + \tfrac{1}{2}(\overline{\nabla}_{\dot{\alpha}}\bar{B})M_{\beta\alpha}\}$$

$$= -iC_{\dot{\alpha}\dot{\gamma}}[f_{\alpha\beta} - \{\nabla_\beta, \mathcal{W}_\alpha\} + \tfrac{1}{2}(\overline{\nabla}^2\bar{B})M_{\alpha\beta}] - h.c.$$

$$\Rightarrow \quad f_{\alpha\beta} = \tfrac{1}{2}\{\nabla_{(\alpha}, \mathcal{W}_{\beta)}\} - \tfrac{1}{2}(\overline{\nabla}^2\bar{B})M_{\alpha\beta}, \qquad \{\nabla^\alpha, \mathcal{W}_\alpha\} + \{\overline{\nabla}^{\dot{\alpha}}, \overline{\mathcal{W}}_{\dot{\alpha}}\} = 0$$

(Here "$h.c.$" means "hermitian conjugate" without the reordering, which would generate non-operator terms.)

Evaluating $\{\nabla, \mathcal{W}\}$ in terms of its pieces, and combining with the results of the previous Jacobi, we obtain the final result:

$$\boxed{\begin{gathered}
\{\overline{\nabla}_{\dot{\alpha}}, \overline{\nabla}_{\dot{\beta}}\} = B\overline{M}_{\dot{\alpha}\dot{\beta}}, \qquad \{\nabla_\alpha, \overline{\nabla}_{\dot{\beta}}\} = -i\nabla_{\alpha\dot{\beta}} \\
[\overline{\nabla}_{\dot{\alpha}}, -i\nabla_{\beta\dot{\beta}}] = C_{\dot{\beta}\dot{\alpha}}\mathcal{W}_\beta - \tfrac{1}{2}(\nabla_\beta B)\overline{M}_{\dot{\alpha}\dot{\beta}}, \qquad [-i\nabla_{\alpha\dot{\alpha}}, -i\nabla_{\beta\dot{\beta}}] = C_{\dot{\beta}\dot{\alpha}}f_{\alpha\beta} - h.c. \\[1em]
\begin{aligned}
\mathcal{W}_\alpha &= -B\nabla_\alpha - G_\alpha{}^{\dot{\beta}}\overline{\nabla}_{\dot{\beta}} + \tfrac{1}{2}(\overline{\nabla}^{\dot{\beta}}G_\alpha{}^{\dot{\gamma}})\overline{M}_{\dot{\beta}\dot{\gamma}} + \tfrac{1}{2}W_\alpha{}^{\beta\gamma}M_{\gamma\beta} + iW_\alpha Y + i\tfrac{1}{6}W^\beta M_{\beta\alpha} \\
f_{\alpha\beta} &= i\tfrac{1}{2}G_{(\alpha}{}^{\dot{\gamma}}\nabla_{\beta)\dot{\gamma}} - \tfrac{1}{2}(\nabla_{(\alpha}B + i\tfrac{1}{3}W_{(\alpha})\nabla_{\beta)} + W_{\alpha\beta}{}^\gamma\nabla_\gamma - \tfrac{1}{2}(\nabla_{(\alpha}G_{\beta)}{}^{\dot{\gamma}})\overline{\nabla}_{\dot{\gamma}} \\
&\quad - (\tfrac{1}{2}\overline{\nabla}^2\bar{B} + B\bar{B} + \tfrac{1}{12}i\nabla^\gamma W_\gamma)M_{\alpha\beta} - i\tfrac{1}{16}[(\nabla_{(\alpha}{}^{\dot{\delta}}G_{\gamma)\dot{\delta}})M_\beta{}^\gamma + \alpha \leftrightarrow \beta] \\
&\quad + \tfrac{1}{2}W_{\alpha\beta}{}^{\gamma\delta}M_{\delta\gamma} + \tfrac{1}{4}(\nabla_{(\alpha}\overline{\nabla}^{\dot{\gamma}}G_{\beta)}{}^{\dot{\delta}})\overline{M}_{\dot{\delta}\dot{\gamma}} + i\tfrac{1}{2}(\nabla_{(\alpha}W_{\beta)})Y \\
W_{\alpha\beta\gamma\delta} &= \tfrac{1}{4!}\nabla_{(\alpha}W_{\beta\gamma\delta)}
\end{aligned}
\end{gathered}}$$

The "reduced tensors" $B, G_a, W_\alpha, W_{\alpha\beta\gamma}$ satisfy the "reduced Bianchi identities"

$$\boxed{\begin{gathered}
G_a = \bar{G}_a, \qquad \overline{\nabla}_{\dot{\alpha}}B = \overline{\nabla}_{\dot{\alpha}}W_\alpha = \overline{\nabla}_{\dot{\alpha}}W_{\alpha\beta\gamma} = 0, \qquad \overline{\nabla}^{\dot{\alpha}}G_{\alpha\dot{\alpha}} = \nabla_\alpha B - iW_\alpha \\
\nabla^\alpha W_{\alpha\beta\gamma} - i\tfrac{1}{3}\nabla_{(\beta}W_{\gamma)} = -i\tfrac{1}{2}\nabla_{(\beta}{}^{\dot{\alpha}}G_{\gamma)\dot{\alpha}}, \qquad \nabla^\alpha W_\alpha + \overline{\nabla}^{\dot{\alpha}}\overline{W}_{\dot{\alpha}} = 0
\end{gathered}}$$

Note that B or W_α may vanish in certain gauges, for reasons to be explained in subsection XA4.

Exercise XA2.2

In IXA4 we saw that integrals of total covariant derivatives vanished in curved space by virtue of the identity $T_{ab}{}^b = 0$. Show that these torsions satisfy the superpace generalization

$$(-1)^B T_{AB}{}^B = 0$$

Exercise XA2.3

Using the expression for ϵ_{abcd} in terms of spinor indices from subsection IIA5, show

$$T_{bcd} = G^a \epsilon_{abcd}$$

Thus G_a is an axial vector.

Exercise XA2.4

By hermitian conjugation, find the commutators not written explicitly above, and show the result is essentially the same as switching dotted and undotted indices (and similarly for bars), except that G_a, Y, and W^α (and $\overline{W}^{\dot{\alpha}}$) get extra minus signs. This illustrates CP invariance, and the fact that G_a is an axial vector, while Y is a pseudoscalar (and similarly for W^α).

Exercise XA2.5

Derive the Bianchi identities in the absence of constraints, in terms of the torsions and curvatures (as follow from the Jacobi identity):

$$\nabla_{[A} T_{BC)}{}^D - T_{[AB|}{}^E T_{E|C)}{}^D = R_{[ABC)}{}^D$$

$$\nabla_{[A} R_{BC)}{}^I - T_{[AB|}{}^E R_{E|C)}{}^I = 0$$

Exercise XA2.6

Show that in *4-component* notation we can write

$$T_{\alpha\beta}{}^c = -i\gamma^c_{\alpha\beta}; \qquad T_{a\beta}{}^\gamma = \gamma_{a\beta\delta} G^{\delta\gamma}, \qquad G^{\alpha\beta} = -G^{\beta\alpha}, \qquad \nabla_\beta G^{\beta\alpha} = W^\alpha$$

This gives another way to see the result of exercise XA2.2. Show that this expression for $G^{\alpha\beta} = (G^a, B, \bar{B})$ gives it an interpretation as an SO(3,3) 6-vector in SL(4) notation (see subsection IC5).

3. Compensators

Just as in ordinary gravity, compensators for scale transformations can be introduced, but for supergravity the compensator should be a supersymmetric multiplet. The simplest choice is the chiral scalar superfield Φ considered earlier: Its complex "physical" scalar (scalar $+i$ pseudoscalar) compensates local scale (the real part) and

U(1) (the imaginary part), its spinor compensates local S-supersymmetry, and its auxiliary complex scalar appears as one of the auxiliary fields of supergravity.

Compensators are much more important in supergravity than in ordinary gravity: Almost any flat space action can be coupled to gravity by the minimal coupling prescription — replacing derivatives with covariant ones, and throwing a factor of $\mathbf{e}^{-1}$ in for the measure. In supergravity this is not the case: As we'll see in the next section, we have both integrals over all superspace, which use E^{-1}, but also integrals over chiral superspace (for integrating chiral superfields), which instead use Φ for the measure. The minimal coupling procedure is then:

(1) Use Φ (and $\bar{\Phi}$) to make a flat-superspace action superconformally invariant,

(2) replace the flat derivatives d_A with the curved ones ∇_A, and

(3) throw in the measure factors appropriate for the integrals.

(The last two steps couple conformal supergravity to a globally conformally invariant theory.)

Another compensator that is commonly used is the "tensor multiplet". (This is sometimes confused in the literature with the "(complex) linear multiplet", another version of the scalar multiplet with no gauge fields whatsoever.) Treated as a matter multiplet, it has the same physical content as the scalar multiplet, but the pseudoscalar is replaced with a second-rank antisymmetric tensor gauge field,

$$\delta B_{mn} = \partial_{[m}\lambda_{n]}$$

To make things simpler, let's look at flat space. We first note that this tensor is "dual" to a pseudoscalar in the sense of switching field equations and constraints of the field strength (see exercises IIB2.1 and VIIIA8.2): For the free fields,

$$F_a = \partial_a \varphi \quad \Rightarrow \quad \partial_{[a}F_{b]} = 0, \qquad G_a = \tfrac{1}{2}\epsilon_{abcd}\partial^b B^{cd} \quad \Rightarrow \quad \partial^a G_a = 0$$

with the field equations following from "self-duality" under $F \leftrightarrow G$:

$$F_a = G_a \quad \Rightarrow \quad \partial^a F_a = \partial_{[a}G_{b]} = 0$$

Since the theory of B_{ab} must be described in terms of G_a alone (because of gauge invariance), no renormalizable self-interactions are allowed; thus, this field is of little interest in quantum field theory outside of supergravity. In terms of the scalar, the fact that only the field strength F_a appears in the field equations means there is the global symmetry

$$\delta\varphi = \zeta$$

for constant parameter ζ. This generalizes to the nonabelian symmetries of nonlinear σ models, resulting in derivative interactions (again nonrenormalizable) but no potentials.

Exercise XA3.1

Consider coupling the tensor field to Yang-Mills: To preserve the tensor's own gauge symmetry, this coupling must be nonminimal. To produce such a coupling, we start with the scalar and duality transform. The coupling we choose is another 4D analog to the 2D model we considered in exercise VIIIA8.2, replacing the pseudoscalar and total derivative $\frac{1}{2}\epsilon^{ab}F_{ab}$ with $tr(\frac{1}{8}\epsilon^{abcd}F_{ab}F_{cd})$. (In general dimensions, the dual to a scalar is a rank-D$-$2 antisymmetric tensor.) We start with the Lagrangian

$$L = -\tfrac{1}{4}\phi\Box\phi + \lambda\phi\tfrac{1}{16}tr(\epsilon^{abcd}F_{ab}F_{cd})$$

for some coupling constant λ. Making use of the Chern-Simons form B_{abc} of subsection IIIC6 to write ϕ in this action only as $\partial_a\phi$, write a first-order form of this action and perform a duality transformation to obtain

$$L' = \tfrac{1}{24}\widetilde{H}^2, \qquad \widetilde{H}_{abc} = \tfrac{1}{2}\partial_{[a}B_{bc]} + \lambda B_{abc}$$

Find the Yang-Mills gauge transformation of B_{ab}. (Hint: $\widetilde{H}$ is gauge invariant.)

The tensor multiplet is described by a chiral spinor gauge field

$$\delta\phi_\alpha = i\bar{d}^2 d_\alpha K \qquad (K = \bar{K})$$

Duality is then described in terms of the real scalar superfield strength (in the free case)

$$F = \phi + \bar{\phi} \quad\Rightarrow\quad \bar{d}^2 d_\alpha F = 0, \qquad G = \tfrac{1}{2}(d_\alpha\phi^\alpha + \bar{d}_{\dot\alpha}\bar{\phi}^{\dot\alpha}) \quad\Rightarrow\quad \bar{d}^2 G = 0$$

with the field equation

$$F = G$$

(F_a appears at order $\theta\bar{\theta}$ in G, and B_{ab} at order θ in ϕ_α.) Again the pseudoscalar has a global symmetry: In terms of the superfield,

$$\delta\phi = i\zeta$$

Now we return to curved space, covariantizing the above with respect to conformal supergravity. We now identify the above global symmetry with the local axial

U(1) (R-)symmetry of supergravity. Thus, the superfield G does not compensate for this symmetry; it remains as a symmetry in actions that use this compensator. In particular, there are no $\int d^2\theta$ terms in such theories, except those that are locally superscale invariant (so the compensator decouples). The matter tensor multiplet also differs from the scalar multiplet in that it has no auxiliary fields (except for the auxiliary components of the gauge field).

4. Scale gauges

Since all the covariant derivatives are built up from the spinor part of the vielbein, we define the local superscale transformations for the covariant derivatives by first defining

$$E'_\alpha = L E_\alpha$$

where L is a real, unconstrained superfield. The constraints then imply

$$\nabla'_\alpha = L\nabla_\alpha + 2(\nabla^\beta L)M_{\beta\alpha} + 6(\nabla_\alpha L)Y, \qquad \overline{\nabla}'_{\dot\alpha} = L\overline{\nabla}_{\dot\alpha} + 2(\overline{\nabla}^{\dot\beta} L)\overline{M}_{\dot\beta\dot\alpha} - 6(\overline{\nabla}_{\dot\alpha} L)Y$$

From the anticommutator we find

$$\begin{aligned} -i\nabla'_{\alpha\dot\alpha} =& L^2(-i)\nabla_{\alpha\dot\alpha} + 4L(\nabla_\alpha L)\overline{\nabla}_{\dot\alpha} + 4L(\overline{\nabla}_{\dot\alpha} L)\nabla_\alpha \\ &+ \tfrac{1}{2}L^{-2}(\nabla_\alpha \overline{\nabla}^{\dot\beta} L^4)\overline{M}_{\dot\beta\dot\alpha} + \tfrac{1}{2}L^{-2}(\overline{\nabla}_{\dot\alpha}\nabla^\beta L^4)M_{\beta\alpha} - \tfrac{3}{2}L^{-2}([\nabla_\alpha, \overline{\nabla}_{\dot\alpha}]L^4)Y \end{aligned}$$

Using the commutation relations, we then can show

$$B' = L^6(\overline{\nabla}^2 + B)L^{-4}, \qquad W'_\alpha = L^3[W_\alpha - 12i(\overline{\nabla}^2 + B)\nabla_\alpha\ ln\ L]$$

$$G'_{\alpha\dot\alpha} = (2[\nabla_\alpha, \overline{\nabla}_{\dot\alpha}] + G_{\alpha\dot\alpha})L^2, \qquad W'_{\alpha\beta\gamma} = L^3 W_{\alpha\beta\gamma}$$

From the way they appear in the commutators we also have that

$$YG_a = 0, \qquad YW_\alpha = \tfrac{1}{2}W_\alpha, \qquad YW_{\alpha\beta\gamma} = \tfrac{1}{2}W_{\alpha\beta\gamma}, \qquad YB = B$$

From linearization, we see that B and W^α pick out exactly the two irreducible halves of the real scalar superfield L: The "vector multiplet" in W'_α and the "scalar multiplet" in B'. (Compare the vector multiplet field strength and chiral scalar gauge fixing for the prepotential V as described in subsections IVC4 and VIB9.) This means we can completely fix the superscale gauge by the choice

$$B = W_\alpha = 0$$

as the generalization of the scale gauge in ordinary gravity that fixes the Ricci scalar to vanish.

Exercise XA4.1

Derive the superscale transformations by use of the Bianchi identities:

a Use the commutation relations of the covariant derivatives (and the solution to the Jacobi identities) to find all the transformations above. Show they imply $E' = L^4 E$.

b An easier way is to use the *reduced* Bianchi identities: Determine the transformations of the reduced field strengths, up to constants, using chirality, dimensional analysis, etc., and then solve for the constants by plugging into the reduced identities.

We then define the scale (and U(1)) transformations of the compensators:

$$\Phi' = L^2 \Phi, \qquad Y\Phi = \tfrac{1}{3}\Phi; \qquad \overline{\nabla}_{\dot\alpha}\Phi = 0$$

$$G' = L^4 G, \qquad YG = 0; \qquad (\overline{\nabla}^2 + B)G = 0, \qquad G = \bar{G}$$

where the scale weights follow from the U(1) weights (vanishing for G by reality) by consistency with the constraints they satisfy.

Exercise XA4.2

Show that a superfield can be chiral only if it has no dotted indices. Then show the relation that any such superfield has between scale and U(1) weights.

All these transformations can be derived either by consistency with the constraints, or by using the solution of the constraints: In terms of the unconstrained superfields that solve the constraints, the superscale transformation is trivial:

$$\psi' = L\psi, \qquad N'_\alpha{}^\mu = N_\alpha{}^\mu, \qquad \Omega' = \Omega$$

The net result, as for super Yang-Mills, is that all the superficial transformations of the constrained covariant derivatives are completely replaced with the new invariances that appear upon solving the constraints: K_{-1} and L eliminate ψ, $K^{\alpha\beta}$ kills $N_\alpha{}^\mu$, and K^M reduces Ω^M to its real part U^M, which transforms only under Λ^M.

Exercise XA4.3

Rederive A_α as in subsection XA1, but in a general gauge, to find

$$iA_\alpha = E_\alpha T, \qquad e^T = \psi^2 E(1 \cdot e^{-\overleftarrow{\Omega}})$$

Show this result gives a superscale transformation for A_α that agrees with the result above. Show the explicit solution for Φ in terms of ϕ and T also gives it a superscale transformation that agrees with the above.

Exercise XA4.4

Often it is easier to use the solution to the constraints than the Jacobi identities:

a Solve for F_{AB} in terms of A_α, and use the solution for A_α from subsection XA1, to derive

$$W_\alpha = -i(\overline{\nabla}^2 + B)\nabla_\alpha(T + \bar{T})$$

and use this to rederive the superscale transformation above. (Hint: Define and use the chiral representation.)

b Find an explicit expression for B, and use it to rederive its superscale transformation. (Hint: You will need to find $\Omega_\alpha{}^{\beta\gamma}$ first. Since B is a scalar, you can choose the Lorentz gauge $N_\alpha{}^\mu = \delta_\alpha^\mu$.)

In subsection XA1 we found that a convenient way to simultaneously fix Lorentz, U(1), and scale gauges was to choose $E_\alpha = \hat{E}_\alpha$. (However, the corresponding component invariances reappeared in the chiral gauge invariances.) Compensators allow more freedom for gauge fixing: For example, we can fix the gauge $B = W_\alpha = 0$ as described above, or we can fix to 1 the compensator or a *physical* matter multiplet (string gauge, as for gravity in subsection IXB5): The possibility of gauges such as $\Phi = 1$ or $G = 1$ depends on the existence in the action of such fields, and not on the details of how they appear (as long as the gauge choice is consistent with the allowed vacuum values). In particular, it does not depend on the signs of their kinetic terms, which is the only thing that determines what is physical and what is a compensator.

Note that either $\Phi = 1$ or $G = 1$ completely fixes the superscale gauge, in spite of the constraints on these superfields. (E.g, $\Phi = \Phi' = 1 \Rightarrow L = 1$.) This is due to the appearance of the U(1) connection: For example, before fixing the scale and U(1) gauges the chirality condition on Φ, rather than constraining Φ, actually determines the spinor U(1) connection A_α:

$$\nabla_\alpha \bar{\Phi} = (E_\alpha - i\tfrac{1}{3}A_\alpha)\bar{\Phi} = 0 \quad \Rightarrow \quad A_\alpha = -3iE_\alpha \, ln \, \bar{\Phi}$$

(But the chirality of the ratio of two chiral superfields with the same weights really fixes it to be chiral; in other words, chirality of scalars makes all but one truly chiral, since the U(1) connection can be determined only once.) As a result, the scale ($\Phi\bar{\Phi} = 1$) and U(1) ($\Phi/\bar{\Phi} = 1$) gauge choice $\Phi = 1$ determines W_α:

$$\Phi = 1 \quad \Rightarrow \quad A_\alpha = 0 \quad \Rightarrow \quad W_\alpha = 0$$

Similarly,

$$G = 1 \quad \Rightarrow \quad 0 = (\overline{\nabla}^2 + B)G = B$$

Conversely, we see that whenever one of the two field strengths B and W^α is eliminated by a superscale(/U(1)) gauge choice in terms of one of the two compensators Φ and G, the other field strength can be made superscale invariant: If we introduce the compensator by a superscale transformation (as for gravity in subsection IXA7), substituting either

$$L^{-4} \to \bar{\Phi}\Phi \ or \ G$$

in the above transformation laws, we find

$$\tilde{B} = (\bar{\Phi}\Phi)^{-3/2}(\overline{\nabla}^2 + B)\bar{\Phi}\Phi = \Phi^{-1/2}\bar{\Phi}^{-3/2}(\overline{\nabla}^2 + B)\bar{\Phi}$$

$$\widetilde{W}_\alpha = G^{-3/4}[W_\alpha + 3i(\overline{\nabla}^2 + B)\nabla_\alpha ln\ G]$$

as locally superscale invariant, where using $\bar{\Phi}\Phi$ for $\widetilde{W}_\alpha$ or G for $\tilde{B}$ yields zero. We can therefore interpret gauging away the compensators as gauging them into the field strengths: We have a choice of either

$$\Phi = 1 \quad \Rightarrow \quad W_\alpha = 0, \qquad \tilde{B} = B$$

$$G = 1 \quad \Rightarrow \quad B = 0, \qquad \widetilde{W}_\alpha = W_\alpha$$

This is analogous to Stückelberg gauges (and their nonlinear generalizations): One of these two tensors (gauge fields with respect to superscale) "eats" the compensator. However, it differs from Stückelberg in that a second "gauge field" is completely gauged away.

In fact, we'll see in the next section that the pure supergravity actions constructed using either of these compensators gives the corresponding field strength as its field equation:

$$\frac{\delta}{\delta\Phi} \quad \Rightarrow \quad \tilde{B} = 0$$

$$\frac{\delta}{\delta\Phi_\alpha} \quad \Rightarrow \quad \widetilde{W}_\alpha = 0$$

Thus, either compensator can be used to eliminate both B and W_α, one as a field equation and the other as a gauge choice. This result is already clear at this point from dimensional analysis and chirality; similarly, we must have

$$\frac{\delta}{\delta U^m} \quad \Rightarrow \quad \tilde{G}_a = 0$$

where $\tilde{G}_a$ is the result of applying a superscale transformation to G_a with whichever of the two compensators is being used in the action. This leaves $W_{\alpha\beta\gamma}$ as the on-shell field strength. The analogy to ordinary gravity is

$$(R, R_{ab} - \tfrac{1}{2}\eta_{ab}R, W_{abcd}) \leftrightarrow (B/W_\alpha, G_a, W_{\alpha\beta\gamma})$$

Although the Ricci tensor must appear in G_a, superscale invariance allows the choice of gauges where the $\theta = 0$ component is arbitrary: From the above we find the linearized transformations

$$\delta G_{\alpha\dot{\alpha}} \approx 4[\nabla_\alpha, \overline{\nabla}_{\dot{\alpha}}]L, \qquad \delta A_{\alpha\dot{\alpha}} \approx -6[\nabla_\alpha, \overline{\nabla}_{\dot{\alpha}}]L$$

(For purposes of evaluating at $\theta = 0$ we can neglect $A_\alpha|$ in δA_a.) Thus, this axial vector component field can be moved around as convenient for component expansions. In $\tilde{G}_a$, they appear only in their invariant combination, $G_a + \frac{2}{3}A_a$ in this approximation. In *nonsupersymmetric gauges*, we can even gauge $G_a| = 0$.

Exercise XA4.5

Use the Bianchi identities instead of explicit superscale to track down the axial vector:

a Use the relation of $W_{\alpha\beta\gamma}$ (which is scale covariant) to W_α (the field strength for A_A) and G_a to show that it is just this combination $G_a + \frac{2}{3}A_a$ that appears in $W_{\alpha\beta\gamma}$ (as its curl, for U(1) invariance).

b Show that

$$G = 1 \quad \Rightarrow \quad B = 0 \quad \Rightarrow \quad \nabla^a G_a = 0$$

Thus, in this gauge the axial vector gauge field A_a has been gauged out of $G_a|$ (although its field strength may appear at higher order: the gauge $G = 1$ doesn't fix U(1)). What replaces it? (Hint: What's in G?)

REFERENCES

1 D.Z. Freedman, P. van Nieuwenhuizen, and S. Ferrara, *Phys. Rev.* **D13** (1976) 3214: supergravity.

2 V.P. Akulov, D.V. Volkov, and V.A. Soroka, *JETP Lett.* **22** (1975) 187: covariant derivatives for supergravity.

3 V. Ogievetsky and E. Sokatchev, *Nucl. Phys.* **B124** (1977) 309;
V.P. Akulov, D.V. Volkov, and V.A. Soroka, *Theor. Math. Phys.* **31** (1977) 285;
S. Ferrara and B. Zumino, *Nucl. Phys.* **B134** (1978) 301:
linearized off-shell supergravity (including auxiliary fields).

4 P. Breitenlohner, *Phys. Lett.* **67B** (1977) 49; *Nucl. Phys.* **B124** (1977) 500: off-shell supergravity in components.

5 Wess and Zumino; Wess; *loc. cit.* (IVC, ref. 5): superspace field equations.

6 W. Siegel, Supergravity superfields without a supermetric, Harvard preprint HUTP-77/A068 (November 1977);
The superfield supergravity action, Harvard preprint HUTP-77/A080 (December 1977);
A polynomial action for a massive, self-interacting chiral superfield coupled to supergravity, Harvard preprint HUTP-77/A077 (December 1977);

A derivation of the supercurrent superfield, Harvard preprint HUTP-77/A089 (December 1977):
off-shell supergravity in superspace, compensator superfield.
7 M. Kaku, P.K. Townsend, and P. van Nieuwenhuizen, *Phys. Rev.* **D17** (1978) 3179:
conformal supergravity in components.
8 M.F. Sohnius and P.C. West, *Phys. Lett.* **105B** (1981) 353:
R-symmetry generator in superspace covariant derivative.
9 V.I. Ogievetsky and I.V. Polubarinov, *Sov. J. Nucl. Phys.* **4** (1967) 156;
K. Hayashi, *Phys. Lett.* **44B** (1973) 497;
M. Kalb and P. Ramond, *Phys. Rev.* **D9** (1974) 2273;
E. Cremmer and J. Scherk, *Nucl. Phys.* **B72** (1974) 117:
antisymmetric tensor gauge field.
10 J. Wess, *Acta Phys. Austriaca* **41** (1975) 409:
N=2 supersymmetric tensor multiplet.
11 W. Siegel, *Phys. Lett.* **85B** (1979) 333:
N=1 supersymmetric tensor multiplet.
12 H. Nicolai and P.K. Townsend, *Phys. Lett.* **98B** (1981) 257;
E. Bergshoeff, M. de Roo, B. de Wit, and P. van Nieuwenhuizen, *Nucl. Phys.* **B195** (1982) 97;
G.F. Chapline and N.S. Manton, *Phys. Lett.* **120B** (1983) 105:
Yang-Mills appearing as Chern-Simons contribution to tensor field strength.
13 P.S. Howe and R.W. Tucker, *Phys. Lett.* **80B** (1978) 138:
superscale transformations.
14 Gates, Grisaru, Roček, and Siegel, *loc. cit.*:
complete superspace treatment of N=1 supergravity.

B. ACTIONS

Now that we understand the structure of superfields in curved superspace, we analyze various supergravity theories through their actions. We will use several methods for finding and evaluating supergravity actions. These actions are significantly more complicated than those we have encountered previously, and it is difficult to see all their features simultaneously, so for any particular application we use the method which best simplifies the property we most need:

(1) Superspace methods are the best for finding general actions and their symmetry properties, manifesting supersymmetry, using globally supersymmetric gauges, and performing quantum calculations.

(2) Component methods are useful for comparing actions and other properties to nonsupersymmetric theories. Such approaches sometimes make some use of superspace, but not superspace integration.

(3) Compensators are useful in conjunction with either of these methods, and can extract many important features and terms in the action with little more than the results of global supersymmetry. They reveal useful broken symmetries, and are the simplest way to analyze the "superhiggs effect" (Higgs for local supersymmetry).

1. Integration

The action for supergravity follows from dimensional analysis: Since the usual Einstein-Hilbert Lagrangian has dimension +2, as does $\int d^4\theta$, the superspace Lagrangian must be dimensionless. The only covariant possibility in terms of the potentials ($E_A{}^M$ and $\Omega_A{}^I$) is thus

$$S_{SG} = 3\int dx\ d^4\theta\ E^{-1}$$

including a normalization factor that will prove convenient later. Introducing the compensator and local scale invariance must also make the usual action for supergravity look like the kinetic term for the compensator multiplet, i.e.,

$$S_{SG,c} = 3\int dx\ d^4\theta\ E^{-1}\bar{\Phi}\Phi$$

(For simplicity we will restrict ourselves for the most part to the simplest compensator, the chiral scalar.) The previous form then corresponds to the scale gauge $\bar{\Phi}\Phi = 1$; often the scale + U(1) gauge $\psi = 1$ is more convenient.

There should also be a supersymmetrization of the cosmological term. This might seem difficult, requiring explicit prepotentials. However, we know from our study of de Sitter space that the cosmological term is basically a statement about the conformal compensator. Therefore, the cosmological term for supergravity, in terms of the superconformal compensator Φ, should be the supersymmetrization of the corresponding term in ordinary gravity, a dimensionless self-interaction for a scalar. The solution of the chirality condition can be written as

$$Y\Phi^3 = \Phi^3 \quad \Rightarrow \quad \Phi^3 = \phi^3 e^{\overleftarrow{\bar{\Omega}}} E$$

in the gauge $\psi = 1$. The result for the cosmological term is then

$$S_{scosmo} = \int dx\ d^2\theta\ E^{-1}\Phi^3 + h.c. = \int dx\ d^2\theta\ \phi^3 + h.c.$$

independent of scale or U(1) gauge: As we saw in subsection XA1, this expression is invariant under Λ transformations, and the integrand itself is invariant under K and L transformations.

Exercise XB1.1

Although this final result for the cosmological term in terms of ϕ is locally superscale invariant, the derivation started in the gauge $\psi = 1$. Generalize the derivation, and the result in terms of Φ, to arbitrary gauges (see exercise XA4.3).

A chiral expression for S_{SG} can be found by similar methods:

$$S_{SG} = 3\int dx\ d^2\theta\ E^{-1}B = 3\int dx\ d^2\bar{\theta}\ E^{-1}\bar{B}$$

Thus, as for super Yang-Mills, the action can be expressed as a real, chiral, or antichiral integral. With the compensator,

$$S_{SG} = 3\int dx\ d^2\theta\ E^{-1}\Phi(\overline{\nabla}^2 + B)\bar{\Phi} = 3\int dx\ d^2\bar{\theta}\ E^{-1}\bar{\Phi}(\nabla^2 + \bar{B})\Phi$$

or, more generally,

$$\int dx\ d^4\theta\ E^{-1}\mathbf{L} = \int dx\ d^2\theta\ E^{-1}(\overline{\nabla}^2 + B)\mathbf{L} = \int dx\ d^2\bar{\theta}\ E^{-1}(\nabla^2 + \bar{B})\mathbf{L}$$

which is just the naive covariantization of the flat-space result (at least in the gauge $\psi = 1$: see exercise XB1.1). Clearly this method generalizes to coupling to other multiplets, and allows both $\int d^2\theta$ and $\int d^4\theta$ integrals to be generalized to curved superspace. In fact, the analysis of the compensator is much simpler than that of the conformal supergravity that couples to it to produce ordinary supergravity: Just as in ordinary gravity, where use of just the compensator allowed us to study certain interesting solutions in gravity, namely de Sitter space and cosomology, some properties of supergravity can be analyzed in terms of just the compensator.

A simple expression (as simple as the super Yang-Mills case) can be written for the supergravity action in terms of unconstrained superfields. We first give a first-order action, analogous to the one for Yang-Mills (subsection IVC5): In that case the action was in terms of V and A_a; here it is in terms of U^m and $E_m{}^a$. Using just the constraints solved in subsection XA1, we write the action as

$$S_{SG,1} = 3 \int dx\ d^4\theta\ E^{-1}\bar{\Phi}\Phi(1 - \tfrac{1}{4}T_{\alpha,\dot{\alpha}}{}^{\alpha\dot{\alpha}})$$

in terms of the torsion $T_{\alpha\dot{\beta}}{}^c$. (Note the similarity to the Yang-Mills case, replacing the Chern-Simons form with the same component of the torsion.) We already evaluated everything except this torsion, which is also easily found in the gauge $\psi = 1$, $N_\alpha{}^\mu = \delta_\alpha^\mu$, $\Omega^\mu = \Omega^{\dot{\mu}} = 0$:

$$S_{SG,1} = 3 \int dx\ d^4\theta\ det(E_m{}^a)\bar{\Phi}\Phi(1 - \tfrac{1}{4}\hat{E}_a{}^m E_m{}^a)$$

$$\Phi = [det(E_m{}^a)]^{-1/3}(1 \cdot e^{\overleftarrow{\bar{\Omega}}})^{1/3} e^{\bar{\Omega}}\phi, \qquad \{\hat{E}_\alpha, \hat{\bar{E}}_{\dot{\alpha}}\} = -i\hat{E}_{\alpha\dot{\alpha}}{}^m \partial_m$$

In the chiral representation this simplifies to

$$S_{SG,1} = 3 \int dx\ d^4\theta\ [det(E_m{}^a)]^{1/3}(1 \cdot e^{-\overleftarrow{U}})^{1/3}\phi(e^{-U}\bar{\phi})(1 - \tfrac{1}{4}E_m{}^{\alpha\dot{\alpha}} i\bar{\partial}_{\dot{\alpha}}\hat{E}_\alpha{}^m)$$

Exercise XB1.2

Find the algebraic field equation for $E_m{}^a$. Use this to eliminate it from the action, yielding expressions for $E_m{}^a$, E, and the (second-order) action in terms of U^m only.

The expansion of the superspace action in terms of unconstrained superfields is needed for supergraphs, the most efficient way to do quantum calculations. We will not consider quantization here; the methods are similar to those described in subsections VIB5, 9-10, and C5 for super Yang-Mills. In particular, one uses background field methods: For example, as for super Yang-Mills,

$$e^\Omega \to e^{\Omega_B} e^{\Omega_Q} \quad \Rightarrow \quad \hat{E}_\alpha \to e^{-H}\hat{\mathcal{E}}_\alpha e^H, \qquad e^H = e^{\Omega_Q} e^{\bar{\Omega}_Q}$$

The end result is that the expansion is about background-covariant derivatives $\mathcal{D}_A$, e.g.,

$$\hat{\nabla}_\alpha = e^{-H}\mathcal{D}_\alpha e^H, \qquad \nabla_\alpha = e^{-H}(\psi\mathcal{D}_\alpha + \Omega_\alpha{}^I M_I)e^H$$

etc. However, $\mathcal{D}_A$ satisfy the same constraints as the full covariant derivatives: For example, they have nonvanishing torsion $T_{\alpha\dot{\beta}}{}^c = -i\delta_\alpha^\gamma\delta_{\dot{\beta}}^{\dot{\gamma}}$. This differs from the expansion implied above in U^m about partial derivatives, which anticommute without torsion: For a perturbation expansion useful for quantum calculations, one must expand in h_{ab} about $\langle e_a{}^m\rangle$, rather than in $e_a{}^m$ itself; thus (at least) the vacuum value $\langle U\rangle$ must be separated from U.

2. Ectoplasm

Although all supersymmetric theories can be analyzed directly in superspace (including classical solutions, effective potentials, Feynman graphs, etc.), for comparison to nonsupersymmetric theories it is necessary to expand superfields in components. Since all fundamental theories are described by actions, it is sufficient to give a prescription for evaluating any action in terms of component fields, as in subsection IVC2 for global supersymmetry. In locally supersymmetric theories, the vielbein needs to be expanded in terms of the prepotentials for supergraphs. We can also find component actions by a straightforward Taylor expansion in θ of the prepotentials in the superspace action. However, for component expansions of classical actions, one can get by more simply by applying Bianchi identities to the covariant derivatives and differential forms (antisymmetric tensors). It is unnecessary to know even the explicit form of the measure in terms of the vielbein or prepotentials.

The fundamental idea is to think of the Lagrangian not as a scalar times a measure, but more "geometrically" as an antisymmetric tensor. Although this approach does not work for the usual superspace Lagrangians because of the peculiarities of fermionic integration, it can be applied to component Lagrangians integrated over 4D spacetime, treated as the bosonic subspace of superspace. We thus write the component action as

$$S = \tfrac{1}{4!}\int dx^m dx^n dx^p dx^q \; L_{mnpq}(x,\theta)$$

where L_{MNPQ} is a graded antisymmetric superfield. Of course, the action should be independent of θ, even though we have integrated over only x. This is equivalent to requiring that the integral should be independent of the choice of 4D hypersurface in superspace. We are familiar with a similar requirement for conserved charges, which are defined as integrals over 3D hypersurfaces: Treating the conserved current in terms of the 3-form dual to the vector,

$$J_{mnp} = \sqrt{-g}\epsilon_{mnpq}J^q \quad \Rightarrow \quad Q = \tfrac{1}{3!}\int dx^m dx^n dx^p J_{mnp}$$

the dual to the usual conservation law as vanishing (covariant) divergence of the vector is vanishing curl of the 3-form:

$$\frac{d}{dt}Q = 0 \quad \Rightarrow \quad \partial_{[m}J_{npq]} = 0$$

An important point is that neither the definition of the charge nor the conservation law requires a metric, since integration in general does not. We thus require for our supersymmetric action

$$\frac{\partial}{\partial\theta}S = 0 \quad \Rightarrow \quad \partial_{[M}L_{NPQR)} = 0, \quad \delta L_{MNPQ} = \tfrac{1}{3!}\partial_{[M}\lambda_{NPQ)}$$

where the gauge invariance allows us to drop terms in the Lagrangian that are total derivatives (surface terms). Note that both L and λ are assumed to be local functions of the fields and their (finite-order) derivatives. (As a result, this is not the usual "cohomology", where both would be allowed to be arbitrary functions of the coordinates.)

Converting the curl-free condition to flat indices (see subsection IVC5 for the Chern-Simons superform),

$$\tfrac{1}{4!}\nabla_{[A}L_{BCDE)} - \tfrac{1}{2!3!}T_{[AB|}{}^{F}L_{F|CDE)} = 0, \qquad \delta L_{ABCD} = \tfrac{1}{3!}\nabla_{[A}\lambda_{BCD)} - \tfrac{1}{(2!)^2}T_{[AB|}{}^{E}\lambda_{E|CD)}$$

The plan is then to find L_{ABCD}, in terms of which the action can be written as

$$S = \int dx\ (-\tfrac{1}{4!})\epsilon^{mnpq}E_q{}^D E_p{}^C E_n{}^B E_m{}^A L_{ABCD}$$

where $E_m{}^A$ is exactly the nontrivial part of the inverse vielbein $E_M{}^A$:

$$E_m{}^A = (e_m{}^a, \psi_m{}^\alpha)$$

namely the inverse vierbein and the gravitino. (If we also write $\psi_m{}^\alpha = e_m{}^a\psi_a{}^\alpha$, we can collect all $e_m{}^a$ factors into a factor of $\mathbf{e}^{-1}$ using the ϵ tensor.)

The next step is to explicitly solve the curl-free condition on the 4-form in terms of the usual scalar superspace Lagrangian. (An alternative is to solve the Bianchis for the field strength of the 3-form gauge field, which is also a 4-form.) The procedure is the same as that used to solve the Bianchi identities for covariant derivatives (in subsection XA2): We start with the lowest-dimension equations and work up. The equations that include the constant (vacuum/flat-space) part of the torsion can be solved algebraically, the rest give differential constraints. Of course, we will need to use the results of subsection XA2 for the torsions and their constraints. The result is

$$L_{\alpha\beta cd} = \epsilon_\alpha{}^{\dot\alpha}{}_{,\beta\dot\alpha,cd}\bar{\mathcal{L}}, \qquad L_{\alpha bcd} = i\epsilon_{\alpha\dot\alpha,bcd}\overline{\nabla}^{\dot\alpha}\bar{\mathcal{L}}, \qquad L_{abcd} = \epsilon_{abcd}[(\overline{\nabla}^2 + 3B)\bar{\mathcal{L}} + h.c.]$$

and their complex conjugates, the rest vanishing, where $\mathcal{L}$ is the usual chiral superspace Lagrangian (superpotential):

$$\overline{\nabla}_{\dot{\alpha}}\mathcal{L} = 0$$

(We use the shorthand notation $a = (\alpha\dot{\alpha})$, etc.)

The result is thus the component expansion of the usual curved superspace action

$$S = \int dx\ d^2\theta\ E^{-1}\mathcal{L} + h.c.$$

Using the Bianchi identities of the covariant derivatives one can also covariantize the usual solution to the chirality condition:

$$\mathcal{L} = (\overline{\nabla}^2 + B)\mathbf{L}$$

which allows us to identify the action as

$$S = \int dx\ d^4\theta\ E^{-1}(\mathbf{L} + \bar{\mathbf{L}})$$

so $\mathbf{L}$ can be taken real without loss of generality for general $d^4\theta$ integrals. On the other hand, for supergravity we can take

$$\mathbf{L}_{SG} = 3, \qquad \bar{\mathbf{L}}_{SG} = 0 \quad \Rightarrow \quad \mathcal{L}_{SG} = 3B, \qquad \bar{\mathcal{L}}_{SG} = 0$$

or vice versa, and the curvature appears in terms of $R_{\alpha\beta}{}^{\alpha\beta}$ and not $\bar{R}_{\dot{\alpha}\dot{\beta}}{}^{\dot{\alpha}\dot{\beta}}$ (like the corresponding $f_{\alpha\beta}$ without $\bar{f}_{\dot{\alpha}\dot{\beta}}$ for Yang-Mills), with half as many terms to collect (for the same final result). In general, we thus have an expression for S in terms of $E_m{}^A$ and the components of L_{ABCD}, and for the latter in terms of curvatures and covariant derivatives of $\mathbf{L}$, which can be evaluated by the same methods as in flat space (except that the commutation relations of the covariant derivatives are more complicated).

Components of a superfield are again defined by evaluating its covariant derivatives at $\theta = 0$. However, as in the case of global supersymmetry, the value of θ is arbitrary, since the result for the action is θ independent: We therefore will generally drop the "$|$" in component expansions of actions; any superfield then implicitly refers to the corresponding component. This θ independence also means that it is not necessary to make any gauge choices: These methods automatically express the action in terms of just the component fields that cannot be completely gauged away. (For example, $E_\alpha{}^m|$ never appears.)

Exercise XB2.1

Collect all the above results:

a Find the complete component expression for the most general chiral ($\int d^2\theta$) action in terms of covariant derivatives of the superpotential, and the supergravity fields.

b Do the same for a real ($\int d^4\theta$) action.

Exercise XB2.2

Evaluate the above actions for massive ϕ^3 theory (for a chiral matter field ϕ) in terms of the components of ϕ.

Exercise XB2.3

Do the same for super Yang-Mills:

a Solve the Bianchi identities for super Yang-Mills in curved superspace.

b Use this result to evaluate the component expansion of its action.

3. Component transformations

We saw in the previous subsection that in component expansions the supergravity gauge fields naturally appear as $E_m{}^A$, since by definition we restrict to the bosonic submanifold. Similar remarks apply to the component form of their supercoordinate transformations (i.e., local supersymmetry), and the related component expansion of their field strengths: In subsection IXB4, we saw that coordinate transformations (as applied to solving the radial gauge condition) were simpler for $E_M{}^A$ because the derivative term on the parameter was just $\partial_M K^A$.

We therefore begin by rewriting the gauge (coordinate) transformations in terms of $E_m{}^A$. As for gravity (see subsections IXA2 and IXB4), we can choose to make the transformation laws more manifestly covariant by writing the generators in terms of covariant derivatives. Then, as for gravity,

$$K = K^A\nabla_A + K^I M_I, \qquad \delta\nabla_A = [K, \nabla_A] \quad \Rightarrow$$

$$\delta E_M{}^A = \nabla_M K^A - E_M{}^B K^C T_{CB}{}^A + K^I M_I E_M{}^A, \qquad \delta \Omega_M{}^I = -\nabla_M K^I + E_M{}^B K^C R_{CB}{}^I$$

$$\delta(T_{AB}{}^C, R_{AB}{}^I) = K(T_{AB}{}^C, R_{AB}{}^I)$$

using $(\delta E_A{}^M)E_M{}^B = -E_A{}^M \delta E_M{}^B$. Here $\nabla_M = E_M{}^A\nabla_A = \partial_M + \Omega_M{}^I M_I$, so these transformations on $E_m{}^A$ and $\Omega_m{}^A$ contain only bosonic derivatives ∂_m, other than those implicit in the torsions and curvatures. (Similar remarks apply to only $E_m{}^A$ and $\Omega_m{}^I$ appearing on the right.) Since the component expansion is an expansion in θ, it is really only supersymmetry $K^\alpha = \epsilon^\alpha$ which is no longer manifest; specializing to that, the transformations on the physical gauge fields become, using the results of subsection XA2 for the torsions,

$$\delta e_m{}^{\alpha\dot{\alpha}} = -i(\epsilon^\alpha \bar{\psi}_m{}^{\dot{\alpha}} + \bar{\epsilon}^{\dot{\alpha}} \psi_m{}^\alpha), \qquad \delta \psi_m{}^\alpha = \nabla_m \epsilon^\alpha + i e_m{}^{\beta\dot{\beta}} (\epsilon_\beta G^\alpha{}_{\dot{\beta}} - \bar{\epsilon}_{\dot{\beta}} \delta_\beta^\alpha B)$$

We have not yet determined the solution for the Lorentz connection $\omega_m{}^{ab}$ and the transformation laws for the auxiliary fields. We will also need the relation between the usual curvature and the components of the superfield strengths. All of these can be found by use of the identities (see subsection IXA2):

$$\begin{aligned} -E_n{}^C E_m{}^B T_{BC}{}^A &= -T_{mn}{}^A = \partial_{[m} E_{n]}{}^A + E_{[m}{}^B \Omega_{n]B}{}^A \\ E_n{}^D E_m{}^C R_{CD}{}^{ab} &= R_{mn}{}^{ab} = \partial_{[m} \Omega_{n]}{}^{ab} + \Omega_{[m}{}^{ac} \Omega_{n]c}{}^b \end{aligned}$$

The application of these identities is similar to that for expansion of the action in the previous section: The separation of the factors of $E_m{}^A$ into bosonic and fermionic parts yields an expansion in powers of the gravitino field. For the torsion case where the index $A = a$ we solve for ω_{mab} in terms of the torsions (auxiliary fields and constants) and vielbein; for the torsion case $A = \alpha$ we solve for $T_{ab}{}^\gamma$, used for the transformation law of the auxiliary fields, and for the curvature case we solve for $R_{ab}{}^{cd}$, used in the component expansion of the action, in terms of these auxiliaries, the vielbein, and the just-determined connection. The U(1) connection A_m needs no solution: It is pure (superscale) gauge, and will cancel in actions (after perhaps an appropriate redefinition of G_a). For these manipulations we use the relations that $T_{AB}{}^C$ and $R_{AB}{}^{cd}$ have to B, G_a, W_α, $W_{\alpha\beta\gamma}$, and their derivatives (as expressed by the solution to the Bianchi identities given in subsection XA2). The solution is

$$\omega_{mbc} = e_m{}^a [\overset{\circ}{\omega}_{abc} - \tfrac{1}{2}(\hat{T}_{bca} - \hat{T}_{a[bc]})], \qquad \hat{T}_{ab}{}^c \equiv e_a{}^m e_b{}^n T_{mn}{}^c = \epsilon_{ab}{}^{cd} G_d + i \psi_{[a}{}^\gamma \bar{\psi}_{b]}{}^{\dot{\gamma}}$$

$$T_{ab}{}^\gamma = -e_a{}^m e_b{}^n \nabla_{[m} \psi_{n]}{}^\gamma + i(\psi_{a\beta} G^\gamma{}_{\dot{\beta}} - \bar{\psi}_{a\dot{\beta}} \delta_\beta^\gamma B - a \leftrightarrow b) \equiv C_{\alpha\beta} t_{\dot{\alpha}\dot{\beta}}{}^\gamma + C_{\dot{\alpha}\dot{\beta}} t_{\alpha\beta}{}^\gamma$$

$$\delta B = -\tfrac{2}{3} \epsilon^\alpha t_{\alpha\beta}{}^\beta, \qquad \delta G_{\alpha\dot{\alpha}} = -\epsilon^\beta (t_{\beta\alpha\dot{\alpha}} + \tfrac{1}{3} C_{\beta\alpha} t_{\dot{\gamma}\dot{\beta}}{}^{\dot{\beta}}) + h.c.$$

$$\begin{aligned} R_{ab}{}^{\gamma\delta} &= e_a{}^m e_b{}^n R_{mn}{}^{\gamma\delta} + \bar{B} \psi_a{}^{(\gamma} \psi_b{}^{\delta)} \\ &\quad - 2i[\psi_a{}^\epsilon (C_{\beta\epsilon} t^{\gamma\delta}{}_{\dot{\beta}} + \tfrac{1}{3} \delta_{(\beta}^\gamma \delta_{\epsilon)}^\beta t_{\dot{\beta}\dot{\epsilon}}{}^{\dot{\epsilon}}) - \bar{\psi}_a{}^{\dot{\beta}} \tfrac{1}{6} t_{(\beta}{}^{\gamma\delta)} - a \leftrightarrow b] \end{aligned}$$

where "$^\circ$" refers to the usual expression for pure gravity, $\psi_a{}^\beta \equiv e_a{}^m \psi_m{}^\beta$, and we have chosen the superscale gauge $W^\alpha = 0$ for simplicity.

Exercise XB3.1

Find the extra terms for $W^\alpha \neq 0$.

4. Component approach

We can now take the superspace action of subsection XB1, as expanded in components by the ectoplasm method of subsection XB2, and substitute the component expansions of the field strengths found in subsection XB3, to find the component action

$$L_{SG} = L_G + L_\psi + \mathbf{e}^{-1} L_a$$

$$L_G = -\tfrac{1}{4}\mathbf{e}^{-1}R, \qquad L_\psi = \epsilon^{mnpq}\bar{\psi}_{m\dot{\alpha}}\tfrac{1}{2}\{e_n{}^{\alpha\dot{\alpha}}, \nabla_p\}\psi_{q\alpha}, \qquad L_a = -\tfrac{3}{8}(G_a)^2 + 3\bar{B}B$$

(Note the signs are again consistent with G_a and $B, \bar{B}$ forming a 6-vector of SO(3,3), though not in the same way as in exercise XA2.6.) Here ∇ and L_G are the usual covariant derivative and Einstein-Hilbert action of general relativity in terms of e and ω, but ω is slightly different from any of the connections used previously (see exercise XB4.1 below). It also differs from the ω given above in that we have explicitly extracted the G_a piece (which is the sole source of ω in the ectoplasm approach). An alternative to ectoplasm to determine ω is to use a first-order formalism: Rather than imposing the usual torsion constraint, we can leave the Lorentz connection as an independent field in R and in the ∇ in L_ψ. Eliminating the Lorentz connection by its field equation yields a modified torsion constraint, and produces ψ^4 terms in the action. We have written L_ψ in a form manifestly symmetric with respect to integration by parts. (Alternatively, we can write $\bar{\psi}e\nabla\psi - \psi e\nabla\bar{\psi}$.)

As an alternative to deriving the component action from the simpler superspace expression, we can postulate the component action directly. In the component approach writing the action in components is more direct than the superspace approach by definition, but proving supersymmetry invariance is less so. This is not so true when coupling to matter, where writing component actions can also be as complicated as deriving them from superspace, so here we consider the simplest case, pure supergravity. We thus begin by postulating $L_{SG} = L_G + L_\psi$; the first term is obvious, while the second follows from minimal coupling for the free gravitino action, which can be derived easily by many methods (see, e.g., subsection XIIA5 below). We ignore the auxiliary fields, which are necessary for off-shell closure of the supersymmetry algebra, but not for supersymmetry invariance of the action.

We write the action for gravity in a form that more resembles the gravitino action (see exercise IXA5.5):

$$L_G = -\tfrac{1}{4}\mathbf{e}^{-1}R = \tfrac{1}{16}\epsilon^{mnpq}\epsilon_{abcd}e_m{}^a e_n{}^b R_{pq}{}^{cd} = \tfrac{1}{8}\epsilon^{mnpq}e_m{}^a e_n{}^b \tilde{R}_{pqab}$$

$$= i\tfrac{1}{8}\epsilon^{mnpq}e_m{}^{\alpha\dot{\alpha}}(\bar{R}_{np\dot{\alpha}}{}^{\dot{\beta}}e_{q\alpha\dot{\beta}} - R_{np\alpha}{}^{\beta}e_{q\beta\dot{\alpha}})$$

where we have switched to spinor notation for the curvature (see subsection IXA1) and used duality in both vector and spinor notation (see subsection IIA7). (Alternatively, we can regard this as the definition of the gravity action.) We then have for the variation of this part of the action (after some integration by parts)

$$\delta L_G = i\tfrac{1}{4}\epsilon^{mnpq} e_m{}^{\alpha\dot{\alpha}}\{[\bar{R}_{np\dot{\alpha}}{}^{\dot{\beta}}\delta e_{q\alpha\dot{\beta}} - R_{np\alpha}{}^{\beta}\delta e_{q\beta\dot{\alpha}}] - [(\delta\bar{\omega}_{n\dot{\alpha}}{}^{\dot{\beta}})T_{pq\alpha\dot{\beta}} - (\delta\omega_{n\alpha}{}^{\beta})T_{pq\beta\dot{\alpha}}]\}$$

where we have used (see exercise IIIC1.2)

$$\delta R_{mn}{}^I = \nabla_{[m}\delta\omega_{n]}{}^I, \qquad \int dx\ \nabla_m V^m = \int dx\ \partial_m V^m = 0$$

Next, we pick the obvious transformation law for the gravitino field as the gauge field of supersymmetry:

$$\delta\psi_m{}^\alpha = \nabla_m\epsilon^\alpha$$

The transformation laws for e and ω will be derived as a by-product of the invariance proof, as will the explicit expression for ω in terms of e and ψ. Substituting this expression for $\delta\psi$ into L_ψ,

$$\begin{aligned}\delta L_\psi = \epsilon^{mnpq}[&\epsilon_\alpha\nabla_m\tfrac{1}{2}\{e_n{}^{\alpha\dot{\alpha}},\nabla_p\}\bar{\psi}_{q\dot{\alpha}} - \bar{\epsilon}_{\dot{\alpha}}\nabla_m\tfrac{1}{2}\{e_n{}^{\alpha\dot{\alpha}},\nabla_p\}\psi_{q\alpha}\\ &+\tfrac{1}{2}\bar{\psi}_{m\dot{\alpha}}(\delta e_n{}^{\alpha\dot{\alpha}})\nabla_p\psi_{q\alpha} - \tfrac{1}{2}\psi_{m\alpha}(\delta e_n{}^{\alpha\dot{\alpha}})\nabla_p\bar{\psi}_{q\dot{\alpha}}\\ &+\tfrac{1}{2}\bar{\psi}_{m\dot{\alpha}}e_n{}^{\alpha\dot{\alpha}}(\delta\omega_{p\alpha}{}^\beta)\psi_{q\beta} - \tfrac{1}{2}\psi_{m\alpha}e_n{}^{\alpha\dot{\alpha}}(\delta\bar{\omega}_{p\dot{\alpha}}{}^{\dot{\beta}})\bar{\psi}_{q\dot{\beta}}]\end{aligned}$$

where we have integrated by parts to free the supersymmetry parameters of derivatives. We then use the antisymmetrization on all curved indices to collect the resulting terms into torsion and curvature as

$$\nabla\{e,\nabla\} = \{e,\nabla\nabla\} + (\nabla e)\nabla = \tfrac{1}{2}\{e,R\} - \tfrac{1}{2}T\nabla$$

The curvature terms then cancel those from δL_G, if we choose for δe in δL_G the transformation law

$$\delta e_m{}^{\alpha\dot{\alpha}} = -i(\epsilon^\alpha\bar{\psi}_m{}^{\dot{\alpha}} + \bar{\epsilon}^{\dot{\alpha}}\psi_m{}^\alpha)$$

Then we also substitute this expression for δe in δL_ψ, and note that half those terms immediately drop out, since

$$\psi_{[m}{}^\alpha\psi_{n]\alpha} = 0$$

by antisymmetry. The remaining terms from both L_G and L_ψ then can be collected as

$$\delta L_{SG} = -\tfrac{1}{8}\epsilon^{mnpq}\widetilde{T}_{mn}{}^{\alpha\dot{\alpha}}\Delta_{pq\alpha\dot{\alpha}}$$

$$\widetilde{T}_{mn}{}^{\alpha\dot{\alpha}} \equiv T_{mn}{}^{\alpha\dot{\alpha}} - i\bar{\psi}_{[m}{}^{\dot{\alpha}}\psi_{n]}{}^\alpha$$

$$\Delta_{pq\alpha\dot{\alpha}} \equiv i(\delta\bar{\omega}_{[p\dot{\alpha}}{}^{\dot{\beta}})e_{q]\dot{\beta}\alpha} - i(\delta\omega_{[p\alpha}{}^{\beta})e_{q]\beta\dot{\alpha}} + \epsilon_\alpha\nabla_{[p}\bar{\psi}_{q]\dot{\alpha}} - \bar{\epsilon}_{\dot{\alpha}}\nabla_{[p}\psi_{q]\alpha}$$

We now note that the former factor in δL_{SG} vanishes by virtue of the equation of motion from varying the connection: Rather than vanishing, the torsion now satisfies

$$second - order: \quad \widetilde{T}_{mn}{}^{\alpha\dot{\alpha}} = 0$$

We can regard $\widetilde{T}$ as the "supersymmetrized torsion"; this is equivalent on shell to the result we found in the previous subsection from superspace. We can therefore quit now, since in a second-order formalism the torsion (and thus the Lorentz connection) would satisfy this equation even off shell. (This approach, using the second-order formalism but not bothering to substitute the supersymmetry variation of the connection, is called the "1.5-order formalism".) On the other hand, we can just as easily recognize that in the first-order formalism cancelation of δL_{SG} is also guaranteed by allowing vanishing of the *latter* factor to define the supersymmetry variation of the (independent) connection:

$$first - order: \quad \Delta_{pq\alpha\dot{\alpha}} = 0$$

Thus, use of the first-order formalism requires no more work than 1.5-order (contrary to remarks in the literature), which is really the same as second-order, and provides the bonus of yielding the transformation law for ω. However, it is useful to note that not all quantities should have their longer forms substituted at the beginning of a calculation (just as we learned in high-school algebra not to plug in numbers till the end).

Exercise XB4.1

Let's complete this calculation to the bitter end, finding all the properties of the connection:

- **a** Solve the torsion constraint for ω (see subsection IXA3).
- **b** Find the transformation law for ω that follows from cancelation of the above terms off shell (i.e., without imposing the torsion constraint).
- **c** Show the above two results are consistent (modulo terms with ψ field equations, which can be canceled by contributions from auxiliary fields) by plugging the expressions for δe and $\delta\psi$ into the variation of the result for part **a** and comparing with the result for part **b**.
- **d** Compare these results with the connection found in the previous subsection. How does the appearance of G_a affect the transformation law?

5. Duality

Although antisymmetric tensor gauge fields can be avoided in general, they tend to turn up in string theory, so we now look at them a little more generally, examining their actions and how they relate to those for scalars. In particular, we note that a sensible action for such a tensor alone cannot be constructed that is conformally invariant: From the same analysis as for electromagnetism or Yang-Mills (subsection IXA7), we see that $(F_a)^2$ does not give a scale-invariant action in four dimensions. Thus, such a field is not suitable as a compensator for pure gravity. However, in supergravity the tensor multiplet (see subsection XA3) also has an ordinary scalar, and an appropriate power of it can make the tensor's action conformal. Therefore, we now examine general duality transformations for the supersymmetric case, which is more relevant for understanding its use in gravity. We will consider explicitly a flat superspace background for simplicity, but generalization to curved superspace by covariantization is straightforward, replacing flat superspace derivatives with (superconformal) covariant derivatives, introducing supergravity field strengths where necessary (in this case, just $\bar{d}^2 \to \overline{\nabla}^2 + B$ for chirality), and using the covariant integration measures.

Duality transformations can be performed directly in the action by use of first-order formulations. Starting with the general tensor multiplet action

$$S_{tm} = \int dx\ d^4\theta\ K(G)$$

where K is some function and $G = d^\alpha \phi_\alpha + h.c.$, we write this in first-order form as

$$S'_{tm} = \int dx\ d^4\theta\ [\widetilde{K}(V) + VG]$$

where V is an unconstrained real superfield and $\widetilde{K}$ is the Legendre transform of K: For this action to reduce to the previous upon applying the algebraic field equation of V, we must have

$$[\widetilde{K}(V) + VG]|_{\frac{\partial \tilde{K}(V)}{\partial V} = -G} = K(G)$$

The duality transformation is then performed by varying ϕ_α instead of V in S'_G: Remembering that ϕ_α is chiral, so

$$\delta \int d^4x\ d^4\theta\ VG = \tfrac{1}{2} \int d^4x\ d^2\theta\ (\delta\phi^\alpha) \bar{d}^2 d_\alpha V + h.c.$$

we solve the condition on V as

$$\bar{d}^2 d_\alpha V = 0 \quad \Rightarrow \quad V = \phi + \bar{\phi}$$

since thinking of V as the prepotential for a vector multiplet says that it is pure gauge. The dualized action is then

$$S_\phi = \int dx\ d^4\theta\ \widetilde{K}(\phi+\bar{\phi})$$

We can also reverse the procedure through another first-order action

$$S'_\phi = \int dx\ d^4\theta\ [K(V) - V(\phi+\bar{\phi})]$$

where in this case varying with respect to ϕ implies

$$\bar{d}^2 V = 0 \quad \Rightarrow \quad V = G$$

while varying with respect to V gives the inverse Legendre transform

$$[K(V) - V(\phi+\bar{\phi})]|_{\frac{\partial K(V)}{\partial V}=\phi+\bar{\phi}} = \widetilde{K}(\phi+\bar{\phi})$$

The simplest case is the Lagrangian $\frac{1}{2}G^2$: We then find

$$K(V) = \tfrac{1}{2}V^2 \quad \Leftrightarrow \quad \widetilde{K}(V) = -\tfrac{1}{2}V^2$$

so the duality is

$$L_{tm} = \tfrac{1}{2}G^2 \quad \Leftrightarrow \quad L_\phi = -\tfrac{1}{2}(\phi+\bar{\phi})^2$$

In flat space, this gives the usual free result $L_\phi = -\bar{\phi}\phi$, but in curved space the $-\frac{1}{2}\phi^2 + h.c.$ part does not vanish because E^{-1} is not chiral. Consequently, this action is not the conformal one (as we already knew from the component argument above). However, the conformal one is easy to find by starting with $-\bar{\phi}\phi$: Making the field redefinition $\phi \to e^\phi$ expresses the action in terms of $\phi+\bar{\phi}$. Legendre transforming,

$$\widetilde{K}(V) = -e^V \quad \Leftrightarrow \quad K(V) = V(ln\ V - 1)$$

so the duality is

$$L_\phi = -e^{\phi+\bar{\phi}} \quad \Leftrightarrow \quad L_{tm} = G(ln\ G - 1)$$

These two conformal actions for matter, when coupled to conformal supergravity, become the two "minimal" actions for supergravity, when the overall sign is changed to make the matter fields into compensators: The version with ϕ as the compensator is called "old minimal", while that with ϕ_α is called "new minimal". They differ only off-shell, in their choice of auxiliary fields. Note that the field equations for the two conformal multiplets,

$$\bar{d}^2\bar{\Phi} = 0 \qquad (\Phi \equiv e^\phi), \qquad \bar{d}^2 d_\alpha ln\ G = 0$$

reproduce the compensator part of the supergravity field equations $\tilde{B} = 0$ and $\widetilde{W}_\alpha = 0$ described in subsection XA4. Again, the full expressions follow from the usual supergravitational and superscale invariances, which were used to find $\tilde{B}$ and $\widetilde{W}_\alpha$; the compensator dependence is enough to identify them as the appropriate covariantizations.

Exercise XB5.1

We saw in subsection IVC5 for the Chern-Simons form, or XB2 for ectoplasmic integrals, that differential forms can be defined in superspace. Do the same for the tensor multiplet:

a By generalizing the bosonic case to superspace with curved indices, and then "flattening" the indices (as for the Chern-Simons superform), show that the super 2-form B_{AB} with field strength H_{ABC} and gauge parameter λ_A is described by

$$\delta B_{AB} = \nabla_{[A}\lambda_{B)} - T_{AB}{}^C\lambda_C, \qquad H_{ABC} = \tfrac{1}{2}\nabla_{[A}B_{BC)} - \tfrac{1}{2}T_{[AB|}{}^D B_{D|C)}$$

(Hint: Show that replacing $E_A \to \nabla_A$ and $C_{AB}{}^C \to T_{AB}{}^C$ yields only canceling connection terms.)

b Show that the torsions given in subsection XA2 satisfy

$$\int dx\ d^4\theta\ E^{-1} H_{\alpha,\dot{\alpha},}{}^{\alpha\dot{\alpha}} = 0$$

Note that this also implies the gauge invariance of the Chern-Simons form of the super Yang-Mills action in curved superspace.

c Show that the constraints

$$H_{\alpha\beta\gamma} = H_{\alpha\beta\dot{\gamma}} = H_{\alpha\beta c} = 0, \qquad H_{\alpha,\dot{\beta},\gamma\dot{\gamma}} = -iC_{\alpha\gamma}C_{\dot{\beta}\dot{\gamma}}G$$

(and complex conjugates) can be solved by

$$B_{\alpha\beta} = B_{\alpha\dot{\beta}} = 0, \qquad B_{\dot{\alpha},\beta\dot{\beta}} = -iC_{\dot{\alpha}\dot{\beta}}\phi_\beta; \qquad \overline{\nabla}_{\dot{\alpha}}\phi_\beta = 0$$

$$B_{ab} = C_{\dot{\alpha}\dot{\beta}}b_{\alpha\beta} + C_{\alpha\beta}\bar{b}_{\dot{\alpha}\dot{\beta}}, \qquad b_{\alpha\beta} = \tfrac{1}{2}\nabla_{(\alpha}\phi_{\beta)}; \qquad G = \tfrac{1}{2}(\nabla_\alpha\phi^\alpha + \overline{\nabla}_{\dot{\alpha}}\bar{\phi}^{\dot{\alpha}})$$

Relate the results for the *gauge fields* B_{AB} to those for the Yang-Mills *field strengths* F_{AB} (subsection IVC3).

d Supersymmetrize the construction of exercise XA3.1: Show one can define a field strength

$$\widetilde{H}_{ABC} = H_{ABC} + B_{ABC}$$

using the Chern-Simons superform B_{ABC}.

The supergravity component action with the tensor multiplet compensator differs from the one of the previous subsection in that B and the longitudinal part of G have been replaced by the gauge field B_{mn}:

$$L_a \to \tfrac{1}{2}\epsilon^{mnpq} G_m \partial_n B_{pq}$$

which is the only possibility that preserves the gauge invariances of both G and B while leaving them both auxiliary. (Their field equations are that their field strengths vanish.)

6. Superhiggs

Supergravity affects spontaneous supersymmetry breaking in a simple way: From the discussion of the immediately preceding subsections, we know that supergravity can be described more simply as conformal supergravity coupled to a compensator. Simple (N=1) conformal supergravity contains no scalars: It consists of only conformal gravity (the traceless part of the metric), the conformal (traceless) part of the gravitino field, and an auxiliary gauge vector. Since symmetry breaking involves giving vacuum values to only scalars, we can replace supergravity by just its compensator for these purposes.

For a general analysis, consider a kinetic term

$$S_K = \int dx\ d^4\theta\ 3\bar{\phi}\phi e^{-K(\chi^i,\bar{\chi}_i)/3}$$

(The exponential form will prove convenient for later component analysis.) This is the most general kinetic term with the usual number of spacetime derivatives: Any term of the form $f(\phi,\bar{\phi},\chi^i,\bar{\chi}_i)$ can be rewritten in this form after appropriate field redefinitions. In particular, if we start with fields with arbitrary Weyl scale weight, then this form follows after rescaling fields so only ϕ carries scale weight, since all terms in the Lagrangian must have the same scale weight, fixed by (super)conformal invariance. ϕ is then the only field to carry U(1) weight, which is proportional to scale weight by superconformal invariance. Then ϕ appears only as $\bar{\phi}\phi$, while K is an arbitrary function of χ^i and $\bar{\chi}_i$. The first step in evaluating this action in components is to simply ignore conformal supergravity altogether, and evaluate this action as is, in terms of matter and compensator multiplets, by the methods we have considered previously for evaluating θ integration. The next step is to add back in some parts of conformal supergravity:

(1) the conformal graviton, which can be put back in easily and uniquely using coordinate and local scale invariance;

(2) the U(1) axial gauge vector, whose coupling is minimal, and thus follows directly from U(1) covariantizing the spacetime derivatives; and

(3) the conformal gravitino, whose quartic couplings can be quite complicated, but as a practical matter we are interested in only the mass term, which is determined from the mass of the Goldstone fermion it eats, which appears in the compensator (and the kinetic term, which is the usual one).

Before considering the general case, we look at the pure supergravity case under this analysis: Looking at just the bosons, we find

$$S_{SG,b} = \int dx\ \mathbf{e}^{-1}\tfrac{1}{2}\{-3[(\nabla - i\tfrac{1}{3}A)\bar{\phi}]\cdot[(\nabla + i\tfrac{1}{3}A)\phi] - \tfrac{1}{2}\bar{\phi}\phi R + 6\bar{B}B\}$$

where ∇ is the usual covariant derivative of general relativity, A is the U(1) gauge vector, and the relative coefficient of the R term was fixed by local scale invariance (see subsection IXA7). Note that here B is the usual auxiliary field from ϕ, and is not associated with conformal supergravity. Choosing the component U(1) and scale gauges $\phi| = 1$, this reduces to

$$S_{SG,b} \to \int dx\ \mathbf{e}^{-1}(-\tfrac{1}{4}R - \tfrac{1}{6}A^2 + 3\bar{B}B)$$

Relating to G_a, we recall that if we had included it from conformal supergravity, for this compensator $\tilde{G}_a = G_a + \frac{2}{3}A_a$, so we can identify A_a with $\frac{3}{2}\tilde{G}_a$. Thus, the compensator method immediately yields the bosonic action, including auxiliary fields.

Returning to the general case, the part of the action for the "physical" scalars ($\phi|$ and $\chi|$) then starts out as

$$S_{K,ps} = \int dx\ \mathbf{e}^{-1}e^{-K/3}\tfrac{1}{2}\{-3[(\nabla - i\tfrac{1}{3}A)\bar{\phi}]\cdot[(\nabla + i\tfrac{1}{3}A)\phi] + \bar{\phi}[(\nabla + i\tfrac{1}{3}A)\phi]\cdot(\bar{\partial}^i K)\nabla\bar{\chi}_i$$
$$+\phi[(\nabla - i\tfrac{1}{3}A)\bar{\phi}]\cdot(\partial_i K)\nabla\chi^i + \bar{\phi}\phi[(\partial_i\bar{\partial}^j K) - \tfrac{1}{3}(\partial_i K)(\bar{\partial}^j K)](\nabla\bar{\chi}_j)\cdot(\nabla\chi^i) - \tfrac{1}{2}\bar{\phi}\phi R\}$$

ignoring until the following subsection the auxiliary scalars, which are irrelevant for the kinetic term. We use the notation $\partial_i = \partial/\partial\chi^i$, $\bar{\partial}^i = \partial/\partial\bar{\chi}_i$. We then choose the U(1) and scale gauges

$$\phi| = e^{K(\chi^i|,\bar{\chi}_i|)/6}$$

where we have explicitly written the $|$'s to emphasize that this is a nonsupersymmetric gauge choice for the component $\phi|$. Finally, we eliminate A by its algebraic field equation. We thus obtain

$$S_{K,ps} \to \int dx\ \mathbf{e}^{-1}\tfrac{1}{2}[(\partial_i\bar{\partial}^j K)(\nabla\bar{\chi}_j)\cdot(\nabla\chi^i) - \tfrac{1}{2}R]$$

Except for the R term and covariant derivatives, this is what would follow in flat superspace from the action $-\int dx\ d^4\theta\ K$. For supersymmetry breaking, we also need the super cosmological term

$$S_c = \int dx\ d^2\theta\ \lambda\phi^3 + h.c.$$

for some constant λ. We could consider more general potentials $\phi^3 e^{f(\chi^i)}$ (again the power of ϕ is fixed by scale and U(1)), but then the field redefinition $\phi \to \phi e^{-f/3}$ would remove it while replacing $K \to K + f + \bar{f}$. (This invariance, and the form of the "metric" on the space of fields χ and $\bar{\chi}$ appearing in the action, identify K as a "Kähler potential".)

The analysis for S_K can also be made by performing a duality transformation on the compensator. Following the same steps as described in the previous subsection for the case without matter (factoring the overall -3 out of the process for convenience), we find

$$S_K \to -\int dx\ d^4\theta\ [3G\ ln\ G + GK(\chi, \bar{\chi})]$$

Since in this form A decouples, the result is obvious from the flat-space result.

Exercise XB6.1

Repeat the above analysis using the compensator G: Evaluate explicitly all the contributions from the bosons in G, couple A, find the R term, show the result is the same.

Normally any kind of symmetry breaking will generate a cosmological term, since a scalar getting a vacuum value implies the potential itself getting one, giving a term $\int dx\ \mathbf{e}^{-1} constant$. This would require adding a cosmological term to the action by hand to cancel the generated one, since within observational limits no cosmological constant is observed in nature. (In any case, the constant generated would correspond to a subatomic length scale, whereas a realistic cosmological constant requires a cosmological length scale, which means a constant, going as 1/length^2, of the order of 10^{-80} in subatomic units.) An exception is when the potential is flat in some direction: In supersymmetry energy is always positive, and the supersymmetric vacuum has zero energy, but some potentials allow other, perhaps nonsupersymmetric, vacua that also have zero energy, and thus generate no cosmological constant. This avoids the ad hoc procedure of "fine tuning" the cosmological constant of an added term for exact cancelation (or at least to order 10^{-80}).

7. No-scale

A useful example of the superhiggs effect with a flat potential is "no-scale supergravity". This theory has an explicit super-cosmological term, but the kinetic term is such that this term does not generate a component cosmological term, but does spontaneously break supersymmetry. The simplest example describes supergravity coupled to a single chiral scalar multiplet. The kinetic term has an SU(1,1) symmetry, and also appears in N=4 supergravity (see subsection XC6 below). Written in terms of just the compensator part of supergravity, it is

$$S_K = \int dx\ d^4\theta\ 3(\bar{\phi}\chi + \bar{\chi}\phi)$$

where ϕ is the compensator and χ is the matter. We have written it in a manifestly U(1,1) covariant form, where the U(1,1) metric is off-diagonal ($\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ instead of the usual diagonalized $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$). For the above component analysis we redefine

$$\chi \to \phi\chi \quad \Rightarrow \quad S_K \to \int dx\ d^4\theta\ 3\bar{\phi}\phi(\bar{\chi} + \chi) \quad \Rightarrow \quad K = -3\ ln(\bar{\chi} + \chi)$$

(Many other superfield redefinitions are possible to put this in more conventional forms, such as $(3\bar{\phi}\phi - \bar{\chi}\chi)$, $\bar{\phi}\phi(3 - \bar{\chi}\chi)$, etc.) The kinetic term for the physical scalars follows from the same analysis we applied to the CP(1) model in subsection IVA2. The only differences here are: (1) the symmetry is U(1,1), not U(2), and (2) the constraint on the norm of the complex 2-vector follows not from a Lagrange multipler (or a low-energy limit), but as a local scale gauge chosen to give the Einstein-Hilbert curvature term the usual normalization. Alternatively, we can use the analysis given in the previous subsection for the general case to find

$$S_{K,ps} \to \int dx\ \mathbf{e}^{-1}\tfrac{1}{2}\left[3\frac{|\nabla\chi|^2}{(\bar{\chi} + \chi)^2} - \tfrac{1}{2}R\right]$$

However, to study just the supersymmetry breaking, we want to look at the "potential" terms: terms that involve the auxiliary scalars instead of spacetime derivatives. We thus now need to include the super cosmological term, which breaks the SU(1,1) invariance. Again evaluating at first without conformal supergravity, then putting some (all but the conformal gravitino) back in, we find the contributions from S_K and S_c

$$S_{aux} = \int dx\ \mathbf{e}^{-1}3[\bar{B}B(\bar{\chi} + \chi) + (\bar{B}\phi b + B\bar{\phi}\bar{b}) + \lambda(B\phi^2 + \bar{B}\bar{\phi}^2)]$$

where $B = d^2\phi$ and $b = d^2\chi$. We then see that eliminating the auxiliaries gives nothing, so there is no potential to generate a cosmological term. However, there is still a mass term for the gravitino: As always, S_c also contains the spinor term

$$6\lambda(\phi\zeta^2 + h.c.)$$

where $\zeta_\alpha = d_\alpha \phi$ is the trace of the gravitino. The gravitino in this model therefore has a mass proportional to $\lambda\langle\bar{\chi}+\chi\rangle^{-1/2}$.

Exercise XB6.1

Explicitly evaluate the spinor part of the kinetic term, and thus determine the exact value of the mass of the spinor, and thus the gravitino.

SU(1,1) invariant kinetic terms also appear in superstring theory, but unlike N=4 and no-scale supergravity, the kinetic term is $(\bar{\phi}\chi + \bar{\chi}\phi)^{1/3}$ instead of just $\bar{\phi}\chi + \bar{\chi}\phi$. (See subsection XIB6.) When applying no-scale supergravity to nature, more matter multiplets are added,

$$S = \int dx\ d^4\theta\ 3(\bar{\phi}\chi + \bar{\chi}\phi - \bar{\chi}_i\chi^i) + \left(\int dx\ d^2\theta\ \lambda\phi^3 e^{f(\chi^i/\phi)} + h.c.\right)$$

generalizing SU(1,1) to SU(n,1) in the first term. (N=5 supergravity has such an SU(5,1) symmetry; see below.) Then χ acts as the "hidden" matter sector that doesn't directly couple to the observed matter χ^i, but serves only to break supersymmetry.

REFERENCES

1 W. Rarita and J. Schwinger, *Phys. Rev.* **60** (1941) 61:
action for spin 3/2.

2 Freedman, van Nieuwenhuizen, and Ferrara, *loc. cit.* (XA):
complete component formalism for supergravity, except for auxiliary fields.

3 S. Deser and B. Zumino, *Phys. Lett.* **62B** (1976) 335:
first-order formalism for supergravity.

4 A.H. Chamseddine and P.C. West, *Nucl. Phys.* **B129** (1977) 39;
P.K. Townsend and P. van Nieuwenhuizen, *Phys. Lett.* **67B** (1977) 439:
1.5-order formalism.

5 P. van Nieuwenhuizen, *Phys. Rep.* **68** (1981) 189:
general review of supergravity.

6 S.J. Gates, Jr., hep-th/9709104, Ectoplasm has no topology: the prelude, in *Supersymmetries and quantum symmetries*, proc., July 22-26, 1997, Dubna (Lecture notes in physics, v. 524), eds. J. Wess and E.A. Ivanov (Springer, 1999) p. 46;
S.J. Gates, Jr., M.T. Grisaru, M.E. Knutt-Wehlau, and W. Siegel, hep-th/9711151, *Phys. Lett.* **421B** (1998) 203:
ectoplasm.

7 Wess and Zumino; Wess; *loc. cit.* (IVC, ref. 5):
component expansions for supergravity.

8 W. Siegel, *loc. cit.* (XA);
K.S. Stelle and P.C. West, *Phys. Lett.* **74B** (1978) 330;
S. Ferrara and P. van Nieuwenhuizen, *Phys. Lett.* **74B** (1978) 333:
old-minimal supergravity.

9 Sohnius and West, *loc. cit.* (XA):
new-minimal supergravity.

10 D.V. Volkov and V.A. Soroka, *JETP Lett.* **18** (1973) 312;
S. Deser and B. Zumino, *Phys. Rev. Lett.* **38** (1977) 1433:
superhiggs.

11 B. Zumino, *Phys. Lett.* **87B** (1979) 203:
Kähler in N=1 nonlinear σ models.

12 E. Cremmer, S. Ferrara, C. Kounnas, and D.V. Nanopoulos, *Phys. Lett.* **133B** (1983) 61:
no-scale supergravity.

C. HIGHER DIMENSIONS

A convenient method for describing extended supersymmetry in D=4 is to apply dimensional reduction to supersymmetry in D>4, since (1) spinors are bigger in D>4, so even simple supersymmetry reduces to extended supersymmetry, and (2) the Lorentz group is bigger in D>4, so some 4D scalars arise as parts of higher-D vectors, etc., meaning fewer Lorentz representations in the multiplet in D>4.

1. Dirac spinors

We saw in subsection IC1 that coordinate representations of orthogonal groups SO(D) could be defined in terms of self-conjugate fermions,

$$G_{ab} = \tfrac{1}{2}[\gamma_a, \gamma_b], \qquad \{\gamma_a, \gamma_b\} = \delta_{ab}$$

We now will construct explicit matrix representations of the Dirac matrices for arbitrary D, and examine their properties. This is useful for understanding:

(1) representations of internal symmetries, such as in Grand Unified Theories;

(2) theories in higher dimensions, which give simpler formulations of certain four-dimensional theories when the extra dimensions are eliminated, and appear in string theory; and

(3) properties of spinors that are independent of D, or their dependence on D, which is useful for comparison and for perturbation in quantum field theory.

An explicit solution can be found easily by first looking at even dimensions, and breaking up the problem into $D/2{=}n$ two-dimensional problems. Furthermore, we can look first at the Euclidean case (SO(D)), and solve for the other cases (SO(D_+,D_-)) by Wick rotation. The solution for SO(2) is just two of the Pauli σ matrices. The general solution then comes from the direct product of the two-dimensional cases, using the third σ matrix to introduce appropriate "Klein factors" (see exercise IA2.3) to insure that the γ matrices from one two-dimensional subspace anticommute with those from another. The resulting γ matrices are then:

$$\frac{1}{\sqrt{2}}(\sqrt{2}\sigma_3 \otimes \cdots \otimes \sqrt{2}\sigma_3 \otimes \sqrt{2}\sigma_i \otimes I \otimes \cdots \otimes I), \qquad \frac{1}{\sqrt{2}}(\sqrt{2}\sigma_3 \otimes \cdots \otimes \sqrt{2}\sigma_3)$$

where $i = 1, 2$, there are a total of n factors, and the number of $\sqrt{2}\sigma_3$ and I factors in the first expression ranges from 0 to $n{-}1$. The last matrix can always be included to extend SO($2n$) to SO($2n{+}1$); in fact, up to normalization, it's simply the product of all the other γ's. (In other words, the product of all the γ matrices is proportional to the identity.)

Exercise XC1.1

Apply exercise IC1.2 to this construction: Show how this representation relates simply to creation and annihilation operators. Show that these Klein factors are identical to those of exercise IA2.3.

The next step is to notice that this construction generally gives a reducible representation. Reducibility comes from two properties: (1) For SO($2n$) we really have SO($2n$+1); and (2) the representation may be real. In fact, most of the interesting cases involve SO($2n$) (in particular, SO(3,1) for Lorentz and SO(4,2) for conformal in four dimensions). In that case we can call the first (or any other) γ matrix $(\sigma_1 \otimes I \otimes \cdots \otimes I)$ for SO($2n$+1) "γ_{-1}", and take the rest as those for SO($2n$). Then the projection operators

$$\varPi_\pm = \tfrac{1}{2}(1 \pm \sqrt{2}\gamma_{-1}) \quad \Rightarrow \quad {\varPi_\pm}^2 = \varPi_\pm,\ \varPi_+\varPi_- = \varPi_-\varPi_+ = 0,\ \varPi_+ + \varPi_- = 1$$

commute with the SO($2n$) generators $G_{ab} \sim \gamma\gamma$, so they can be used to project the representation of the γ's into two representations of SO($2n$). These two halves of a Dirac spinor are known as "Weyl spinors". A convenient representation of the γ matrices for this purpose is the one given in subsection IIA6, with the representation of the Pauli matrices used in our SU(2)/SL(2,C) discussion of subsections IIA1 and 5,

$$\sigma_1 = \tfrac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma_2 = \tfrac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_3 = \tfrac{1}{\sqrt{2}}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

We then can write the spinor, which has 2^n components (since it represents the direct product of n representations of σ matrices, each of which has two components) as two 2^{n-1}-component spinors projected by

$$\varPi_\pm = \sigma_\pm \otimes I \otimes \ldots \otimes I, \qquad \sigma_+ = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \sigma_- = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

The γ matrices then take the block-diagonal form

$$\gamma_{-1} = \tfrac{1}{\sqrt{2}}\begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \qquad \textit{other}\ \gamma = \begin{pmatrix} 0 & \tilde{\sigma} \\ \sigma & 0 \end{pmatrix}$$

We will refer to these reduced matrices σ (and $\tilde{\sigma}$), and the γ matrices themselves for SO($2n$+1), as generalized Pauli (σ) matrices.

The reality properties of the representation depend on the existence of a metric $\eta_{\dot{A}}{}^B$ (or $\varOmega_{\dot{A}}{}^B$ for psuedoreality, which doesn't reduce the representation), as in our discussion of classical groups of subsection IB5. In fact, all the spinor representations of any orthogonal group are also defining representations of another group: For less than seven dimensions, this leads to the identification of covering groups discussed in subsection IC5; for more than six dimensions, it only identifies the orthogonal group as a subgroup of this new group. (An interesting exception is SO(8), where the spinor representations are also 8-dimensional, and are the two other defining representations of SO(8).) In matrix notation, we look for a matrix $C = \eta$ or Ω such that we can define the operation of charge conjugation as

$$\Psi \to C^{-1}\Psi^*,\ G\Psi \to C^{-1}(G\Psi)^* \quad \Rightarrow \quad G = C^{-1}G^*C$$

If we like, we can also choose

$$C = C^\dagger = C^{-1}$$

without loss of generality. For a representation to be invariant under charge conjugation (i.e., real)

$$\Psi = C^{-1}\Psi^* \quad \Rightarrow \quad C^* = C^{-1}$$

For our γ matrix representation, the matrix to look at is

$$C = ... \otimes C_2\sqrt{2}\sigma_3 \otimes C_2 \otimes C_2\sqrt{2}\sigma_3 \otimes C_2$$

where

$$C_2 = \left(\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right)$$

independent of the representation used for the Pauli matrices. (Our representation is simplest, since then $C_2\sqrt{2}\sigma_3 = I$, and $C = C^\dagger$. In other representations, C may also need an n-dependent factor of i if we want $C = C^\dagger$.) Using properties of σ matrices we found in our discussion of SO(3) in subsection IIA2, such as $C_2\sigma^* C_2 = -\sigma$, we find

$$C^{-1}\gamma^* C = (-1)^n\gamma, \quad C^* = (-1)^{n(n+1)/2}C^{-1}, \quad C^T = (-1)^{n(n+1)/2}C, \quad C^\dagger = C^{-1}$$

This distinguishes 8 cases, where the irreducible spinors are:

SO($8m$):	Weyl and real
SO($8m$+1):	real
SO($8m$+2):	Weyl
SO($8m$+3):	pseudoreal
SO($8m$+4):	Weyl and pseudoreal
SO($8m$+5):	pseudoreal
SO($8m$+6):	Weyl
SO($8m$+7):	real

For SO($4m$+2), charge conjugation does not preserve γ_{-1}, and thus $\Pi_\pm$. Therefore, in those cases there is no metric $\eta_{\dot{A}}{}^B$ or $\Omega_{\dot{A}}{}^B$ on the irreducible spinor: The Dirac spinor consists of two irreducible spinors that are complex conjugate representations of each other. In general, a Dirac spinor has 2^n complex components for SO($2n$) and SO($2n$+1); the Weyl condition reduces this a factor of two for SO($2n$), as does reality where applicable. (Pseudoreality does nothing.)

It is useful to know the other group metrics, if they exist. For unitarity properties we look for a metric $\Upsilon^{\dot{A}B}$ such that

$$G = -\Upsilon^{-1}G^\dagger\Upsilon \quad \Rightarrow \quad \Upsilon = \Upsilon^\dagger$$

(Thus $g = e^G$ satisfies $\Upsilon g^{-1} = g^\dagger\Upsilon$.) We can also choose

$$\Upsilon = \Upsilon^{-1}$$

without loss of generality. We therefore look for a metric satisfying

$$\Upsilon^{-1}\gamma^\dagger\Upsilon = \gamma$$

so $G \sim [\gamma,\gamma]$ is antihermitian with respect to Υ. For SO(D), we have simply

$$\Upsilon = I$$

since the hermiticity of the σ matrices implies that of the γ matrices. We then can also define a metric to raise and lower indices in terms of these two metrics, by contracting the dotted (or undotted) indices: In matrix notation, we then have

$$(C^T\Upsilon)^{-1}\gamma^T(C^T\Upsilon) = (-1)^n\gamma \quad \Rightarrow \quad G = -(C^T\Upsilon)^{-1}G^T(C^T\Upsilon)$$

For all cases except SO($4m$), this also defines the symmetry properties of the generalized σ matrices: They can be defined as $C^T\Upsilon\gamma$ for SO($2n$+1), and as its diagonal blocks with respect to $\Pi_\pm$ for SO($4m$+2); but for SO($4m$) it's off-diagonal, so the generalized σ matrices appearing there carry one each of the two different kinds of spinor indices, and thus have no symmetry. Then we rewrite the above result as

$$(C^T\Upsilon\gamma)^T = (-1)^{n(n-1)/2}(C^T\Upsilon\gamma)$$

2. Wick rotation

The indefinite-metric groups SO(D_+,D_-) ($D_+ \neq 0 \neq D_-$) can be treated by Wick rotation: giving i's to D_- of the γ_a's, so the corresponding components of

η_{ab} get minus signs. Since this affects γ^T in the same way as γ, the metric $C^T\Upsilon$ is unchanged. In other words,

$$C^T\Upsilon = C_E^T \qquad (\Upsilon_E = I)$$

in terms of the Euclidean C of the previous subsection. However, γ^* and $\gamma^\dagger$ are affected in the opposite way to γ and γ^T ($-i$'s instead of i's). Υ then becomes (up to normalization) the product of the timelike γ's,

$$\Upsilon \sim \prod_{\eta_{aa}<0} \sqrt{2}\gamma_a$$

which also determines the modification of C. In the above equations for $\gamma^\dagger$ and γ^T, we then find a factor of -1 for each rotated dimension, coming from anticommutation with each timelike γ in Υ, so we redefine all γ's by an overall factor of i^{D_-} to preserve their pseudohermiticity. This changes the normalization to

$$G_{ab} = (-1)^{D_-}\tfrac{1}{2}[\gamma_a, \gamma_b], \qquad \{\gamma_a, \gamma_b\} = (-1)^{D_-}\eta_{ab}$$

In odd dimensions the σ matrices are the γ matrices (up to multiplication by one of the metrics), while in even dimensions the γ matrices consist of two off-diagonal blocks of the σ matrices. To write actions we also need the "dual" Dirac spinor, in the sense of a Hilbert-space inner product,

$$\bar{\Psi} = \Psi^\dagger \Upsilon$$

with Υ as defined above. In particular, it is just $\sqrt{2}\gamma_0$ in $D_- = 1$.

We then find (e.g., using the explicit representation given above) that C has the same properties with regard to symmetry and γ_{-1} for SO(D_+,D_-) as for SO($D_+ - 1$, $D_- - 1$). Thus, the properties of these metrics on the *irreducible* (as opposed to Dirac) spinors follow easily from the Euclidean case by using

$$C^T\Upsilon: \qquad SO(D_+, D_-) \leftarrow SO(D_+ + D_-)$$

$$C: \qquad SO(D_+, D_-) \leftarrow SO(D_+ - D_-)$$

Then the properties of Υ follow from the above two in the cases where all 3 exist; the few cases where only Υ exists, which have D even, follow from the next higher D (increasing D_+ by 1).

Exercise XC2.1

Find the explicit γ matrices for $D = 2$ and 4 from the construction of the previous section, and apply this Wick rotation. Compare the results with the conventions of subsections VIIIA7 and IIA6.

We can use these properties to determine that the number of real components D' of an irreducible spinor is

$$D' = 2^{[D-2+f(D_+-D_-)]/2}, \qquad \begin{array}{c|cccccccc} x \bmod 8 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \hline f(x) & 0 & 1 & 2 & 3 & 2 & 3 & 2 & 1 \end{array}$$

The complete results can be summarized by the following table, showing for each case of SO($D-D_-$,D_-), for D mod 8 and D_- mod 4, the types of irreducible spinors ψ, the types of metrics η (symmetric) and Ω (antisymmetric) for these irreducible spinors, and the type of generalized σ matrices (and its symmetry, where relevant):

D \ D_-	0 Euclidean	1 Lorentz	2 conformal	3
0	$\psi_\alpha\ \psi_{\alpha'}$ $\eta^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{\alpha\beta'}$	$\psi_\alpha\ \psi_{\dot\alpha}$ $\eta^{\alpha\beta}$ $\sigma_{\alpha\dot\beta}$	$\psi_\alpha\ \psi_{\alpha'}$ $\eta^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{\alpha\beta'}$	$\psi_\alpha\ \psi_{\dot\alpha}$ $\eta^{\alpha\beta}$ $\sigma_{\alpha\dot\beta}$
1	ψ_α $\eta^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}$	ψ_α $\eta^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}$	ψ_α $\eta^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}$	ψ_α $\eta^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}$
2	$\psi_\alpha\ \psi^\alpha$ $\eta^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}\ \sigma^{(\alpha\beta)}$	$\psi_\alpha\ \psi^\alpha$ $\eta_{\dot\alpha}{}^\beta$ $\sigma_{(\alpha\beta)}\ \sigma^{(\alpha\beta)}$	$\psi_\alpha\ \psi^\alpha$ $\Omega^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}\ \sigma^{(\alpha\beta)}$	$\psi_\alpha\ \psi^\alpha$ $\Omega_{\dot\alpha}{}^\beta$ $\sigma_{(\alpha\beta)}\ \sigma^{(\alpha\beta)}$
3	ψ_α $\Omega^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}$	ψ_α $\Omega^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}$	ψ_α $\Omega^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}$	ψ_α $\Omega^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}$
4	$\psi_\alpha\ \psi_{\alpha'}$ $\Omega^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{\alpha\beta'}$	$\psi_\alpha\ \psi_{\dot\alpha}$ $\Omega^{\alpha\beta}$ $\sigma_{\alpha\dot\beta}$	$\psi_\alpha\ \psi_{\alpha'}$ $\Omega^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{\alpha\beta'}$	$\psi_\alpha\ \psi_{\dot\alpha}$ $\Omega^{\alpha\beta}$ $\sigma_{\alpha\dot\beta}$
5	ψ_α $\Omega^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}$	ψ_α $\Omega^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}$	ψ_α $\Omega^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}$	ψ_α $\Omega^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}$
6	$\psi_\alpha\ \psi^\alpha$ $\eta^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}\ \sigma^{[\alpha\beta]}$	$\psi_\alpha\ \psi^\alpha$ $\Omega_{\dot\alpha}{}^\beta$ $\sigma_{[\alpha\beta]}\ \sigma^{[\alpha\beta]}$	$\psi_\alpha\ \psi^\alpha$ $\Omega^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}\ \sigma^{[\alpha\beta]}$	$\psi_\alpha\ \psi^\alpha$ $\eta_{\dot\alpha}{}^\beta$ $\sigma_{[\alpha\beta]}\ \sigma^{[\alpha\beta]}$
7	ψ_α $\eta^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}$	ψ_α $\eta^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}$	ψ_α $\eta^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}$	ψ_α $\eta^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}$

Exercise XC2.2

The vectors of SO(D_+,D_-) for 3≤ $D_+ + D_-$ ≤6 were expressed as tensors with two spinor indices, with the appropriate symmetry and tracelessness conditions, in subsection IC5.

a Show that the spinor metrics found from the Dirac analysis are sufficient to identify each of these covering groups.

b Show the equivalence of each orthogonal group to its covering group by showing that (1) the Lie algebras have the same dimension, and (2) the determinant (or its square root) of this tensor gives the appropriate orthogonal metric. (Hint: Do the cases $D_+ + D_- = 4$ and 6 first, and specialize to 3 and 5.)

Exercise XC2.3

Consider the groups SO(n,n) and SO(n+1,n). Explicitly construct a real representation of the γ matrices by modifying the method of subsection XC1, demonstrating that all such spinors are real.

Besides the Dirac spinors and Dirac matrices γ, and the irreducible spinors and Pauli matrices σ, it is also useful to introduce irreducible real ("Majorana") spinors and corresponding matrices Γ. When the irreducible spinors are already real these are the same, but when the irreducible spinors are complex this real spinor is just the direct sum of the irreducible spinor and its complex conjugate, a spinor with twice as many components. In general, these generalized Majorana spinors and matrices have many properties that are independent of the number of dimensions, but depend on the number of time dimensions:

D_-	0 Euclidean	1 Lorentz	2 conformal	3
	$\psi_\alpha\ \psi_{\alpha'}$ $\eta^{\alpha\beta}$ $\Gamma_{\alpha\beta'}$	$\psi_\alpha\ \psi^\alpha$ $\Gamma_{(\alpha\beta)}\ \Gamma^{(\alpha\beta)}$	$\psi_\alpha\ \psi_{\alpha'}$ $\Omega^{\alpha\beta}$ $\Gamma_{\alpha\beta'}$	$\psi_\alpha\ \psi^\alpha$ $\Gamma_{[\alpha\beta]}\ \Gamma^{[\alpha\beta]}$

For D odd, there is only one irreducible spinor, so there is a metric $M^{\alpha\beta}$ or $M^{\alpha\beta'}$ to relate the two spinors listed. For $D - 2D_-$ twice odd (2 *mod* 4), the original irreducible spinor was complex, so there is a metric representing a U(1) generator that rotates the complex spinor and its complex conjugate oppositely. (I.e., it's the identity on the complex spinor and minus the identity for the complex conjugate.) For $D - 2D_- = 3, 4, 5$ *mod* 8, the original spinor was pseudoreal, and this U(1) can be extended to an SU(2): Since the complex spinor and its complex conjugate transform the same way under the orthogonal group, they can be paired as a doublet of SU(2). This doubled representation is a real representation of the orthogonal group ⊗ SU(2), since the direct product of the two antisymmetric charge-conjugation matrices is symmetric (two −'s under transposition).

3. Other spins

Before considering supersymmetry in higher dimensions, we first study representations of the Poincaré group there. From the general analysis of section IIB, we know that general on-shell representations follow from the massless ones, which can be classified by their representation of the lightcone little group SO(D−2). Specifically, the bosons can be described as traceless tensors of a certain symmetry (labeled by a Young tableau), while the fermions can be labeled as the direct product of such tensors with an irreducible spinor, with a tracelessness condition imposed between any vector index and the spinor index using a γ or σ matrix. Similar methods can be used to find the off-shell representations in terms of representations of SO(D−1,1), but without subtracting traces. (For full details, see chapter XII.) The gauge degrees of freedom can be subtracted from these Lorentz representations by dropping all lower vector indices with the value "−", by the usual lightcone gauge condition; this tells us the number of total physical + auxiliary degrees of freedom.

In practice, the only interesting massless fields in higher dimensions are:

(1) the metric (graviton),

(2) totally antisymmetric tensors (including scalars and vectors),

(3) spin-3/2 (gravitino), desribed by vector⊗spinor, and

(4) spinors.

By the methods described above, the counting of physical, auxiliary, and gauge degrees of freedom for these fields is (where D' is the number of components of an irreducible spinor of SO(D−1,1) — see the previous subsection):

field	physical	auxiliary	gauge
$h_{(ab)}$	$\frac{1}{2}D(D-3)$	D	D
$A_{[a_1 \ldots a_n]}$	$\binom{D-2}{n}$	$\binom{D-2}{n-1}$	$\binom{D-1}{n-1}$
$\psi_{a\alpha}$	$\frac{1}{2}D'(D-3)$	$\frac{1}{2}D'(D+1)$	D'
χ_α	$\frac{1}{2}D'$	$\frac{1}{2}D'$	0

Exercise XC3.1

Derive all the entries in the table. For each type of field, find the *minimum* D for which physical degrees exist.

We next consider exactly how many higher dimensions are relevant. From the previous subsection, we see that an irreducible spinor (which we use for the supersymmetry generators) has 1 component in D=2, 2 in D=3, 4 in D=4, 8 in D= 5 or 6, 16 in D=7, 8, 9, or 10, 32 in D=11, etc. Since the maximal Lorentz symmetry can be obtained by looking at the maximum D for which a certain size spinor exists, we see that the appropriate D for which an irreducible spinor reduces to N irreducible spinors (for N-extended supersymmetry) in D=4 is

N	D
1	4
2	6
4	10
8	11

etc. From the discussion of subsection IIC5, we know that supergravity exists only for N≤8, and super Yang-Mills only for N≤4. This means that simple supergravity (i.e., any supergravity) exists only for D≤11, and simple super Yang-Mills for D≤10. Since theories with massless states of spin>2 are not of physical interest (in fact, no interacting examples have been constructed), we can restrict ourselves to looking at just D=4, 6, 10, and 11. In general, an irreducible multiplet in some D can become reducible in lower D. However, since irreducible multiplets of supersymmetry are constructed as the direct product of the smallest representation of supersymmetry with an arbitrary representation of the Poincaré group, this reducibility corresponds directly to the reducibility of that Poincaré representation, which occurs simply because the Lorentz group gets smaller upon reduction. In particular, the smallest representation of supersymmetry is itself irreducible. For the case of simple supersymmetry, this is the scalar multiplet (scalars and spinors) in D=6, the vector multiplet (super Yang-Mills: vectors, spinors, and scalars) in D=10, and supergravity in D=11. The statement that it is the smallest multiplet in that number of dimensions is directly related to the fact that it does not exist in higher dimensions.

4. Supersymmetry

We first generalize to arbitrary dimensions some definitions used earlier: To discuss the properties of supersymmetry that are common to all dimensions (but one time), it's most convenient to use the Majorana form

$$\{q_\alpha, q_\beta\} = \Gamma^a_{\alpha\beta} p_a$$

which is consistent with the general symmetry of these matrices. The supersymmetry generators are then

$$q_\alpha = -i\frac{\partial}{\partial\theta^\alpha} + \tfrac{1}{2}\Gamma^a_{\alpha\beta}\theta^\beta \frac{\partial}{\partial x^a}$$

and $\epsilon^\alpha q_\alpha$ generates the infinitesimal transformations

$$\delta\theta^\alpha = \epsilon^\alpha, \qquad \delta x^a = i\tfrac{1}{2}\Gamma^a_{\alpha\beta}\epsilon^\alpha\theta^\beta$$

where $(q_\alpha)^\dagger = -q_\alpha$. The covariant derivatives are

$$d_\alpha = \frac{\partial}{\partial\theta^\alpha} + \tfrac{1}{2}\Gamma^a_{\alpha\beta}\theta^\beta p_a$$

and they satisfy the same algebra as supersymmetry

$$\{d_\alpha, d_\beta\} = \Gamma^a_{\alpha\beta} p_a$$

but with the opposite hermiticity condition $(d_\alpha)^\dagger = +d_\alpha$. The invariant infinitesimals are

$$d\theta^\alpha, \qquad dx^a + i\tfrac{1}{2}(d\theta^\alpha)\Gamma^a_{\alpha\beta}\theta^\beta$$

Superfields can be expanded as either

$$\Phi(x,\theta) = \phi(x) + \theta^\alpha\psi_\alpha(x) + ...$$

or

$$\psi_\alpha = d_\alpha\Phi, \qquad ...$$

giving the transformations

$$\delta\phi = \epsilon^\alpha\psi_\alpha, \qquad \delta\psi_\alpha = -i\epsilon^\beta\tfrac{1}{2}\Gamma^a_{\alpha\beta}\partial_a\phi + ..., \qquad ...$$

Representations can be found as for D=4; we don't have twistors in general, but we can always use a lightcone frame. We first need to define $\Gamma^{a\alpha\beta}$, which in general is independent of $\Gamma^a_{\alpha\beta}$ (only the latter was needed to define supersymmetry above): The analog of the Dirac anticommutation relations (which can be reconstructed if we combine the two Γ's, as generalized σ's, to form a generalized γ) is

$$\Gamma^{(a}_{\alpha\gamma}\Gamma^{b)\gamma\beta} = \eta^{ab}\delta^\beta_\alpha$$

In the lightcone frame the momentum is just $p^a = \delta^a_+ p^+$ with $p^+ = \pm 1$ being the sign of the (canonical) energy. In this frame we have the constraint $\Gamma^- q = 0$. This projects away half the q's, since $-\Gamma^{\pm}\Gamma^{\mp}$ are projection operators: Using the anticommutation relations of $\Gamma^{\pm}$,

$$\Pi_{\pm} = -\Gamma^{\pm}\Gamma^{\mp} \quad \Rightarrow \quad {\Pi_{\pm}}^2 = \Pi_{\pm},\ \Pi_{+}\Pi_{-} = \Pi_{-}\Pi_{+} = 0,\ \Pi_{+} + \Pi_{-} = 1$$

The equality of the sizes of the two subspaces follows from parity symmetry, $\Gamma^{\pm} \leftrightarrow \Gamma^{\mp}$. We thus need to consider only half of the q's, namely $\Gamma^{+}q$. We therefore switch to a notation where we consider the truncated spinor q_μ with just that half of the components. This "lightcone spinor" is an irreducible spinor of SO(D−2). In a Majorana basis it satisfies the same commutation relations as Dirac matrices,

$$\{q_\mu, q_\nu\} = \delta_{\mu\nu}$$

Since q_μ has an even number of components (2^n, $n > 0$) in $D > 3$, the states that represent this algebra form a Dirac spinor of SO(2^n) that is reducible to two Weyl spinors. (These spinors should not be confused with those of SO(D−2), such as q_μ, which is a vector of this SO(2^n).) Since supersymmetry takes each of these "spinors" into the other, one spinor contains all the bosons, while the other contains all the fermions. There are an equal number of physical boson and fermion states because the two Weyl spinors are equal in size. Since SO(D−2)⊂SO(2^n), each Weyl spinor of SO(2^n) is reducible with respect to SO(D−2). The only exceptions are (1) D=4, where SO(D−2)=SO(2^n)=SO(2), and there is one bosonic state and one fermionic one, and (2) D=10, where SO(D−2)=SO(2^n)=SO(8).

Exercise XC4.1

Let's look more closely at these exceptions:

a Show that SO(D−2)=SO(2^n) only in D=3,4,6,10.

b Show that in D=6 the bosons form a reducible representation of the little group SO(D−2). How is this possible, when the group SO(2^n) is the same?

c For D=10, what representations of the little group are the bosons and the fermions? Compare this to the representations of SO(2^n) formed by the bosons, fermions, and q itself, and apply this "symmetry" to the cases D=4,6.

This "Dirac spinor" of SO(2^n) is the smallest representation of supersymmetry. It can also be represented in terms of anticommuting coordinates, by dividing up q_μ into two halves, one of which is complex coordinates, the other half being both the complex and canonical conjugate (as for the fermionic harmonic oscillators of exercise IA2.3). The most general representation of supersymmetry is then the direct product of this one with an arbitrary representation of the Poincaré group.

All the results of this section can be extended to "extended supersymmetry", with supersymmetry generators $q_{i\alpha}$ for an N-valued "internal" index i, as expected from our discussion of supergroups in subsection IIC4: For example, in D=4 the supergroup describing extended conformal supersymmetry, SU(2,2|N), includes conformal symmetry SU(2,2), internal symmetry U(N), N supersymmetries, and N S-supersymmetries. In general, the supersymmetries then satisfy the algbera

$$\{q_{i\alpha}, q_{j\beta}\} = \delta_{ij} \mathit{\Gamma}^a_{\alpha\beta} p_a$$

The smallest representation of an extended supersymmetry follows as before, where now the complete lightcone q acts as Dirac matrices for SO(N2^n). Other representations are again found by direct product, now between this smallest supersymmetry representation and an arbitrary representation of both Poincaré and the internal symmetry. For the more interesting cases, where N itself is a power of 2, the smallest representation can also be derived by dimensional reduction from higher dimensions of N=1 ("simple") supersymmetry, changing the higher-dimensional algebra only by setting some components of the momentum to vanish, and noting that a spinor of higher dimensions reduces to many spinors, as clear from our explicit construction earlier. (Other representations tend to be reducible, since the Poincaré representation in the direct product is reducible upon dimensional reduction.) Dimensional reduction can also be defined for an action (for supersymmetric or nonsupersymmetric theories), by again setting the derivatives with respect to the "extra" coordinates to vanish, and also restricting the integration to the reduced set of coordinates. Another interpretation is that we expand the fields over all momentum modes in the extra coordinates, and then drop all but the zero (constant) modes.

We also recall from subsection XC2 the index structure of spinors in D=6, 10, and 11, which we need to write supersymmetry covariant derivatives. We thus have, for simple supersymmetry,

$$\begin{aligned} D=6: &\quad \{d_{i\alpha}, d_{j\beta}\} = -C_{ji} i\partial_{\alpha\beta} \\ D=10: &\quad \{d_\alpha, d_\beta\} = -\sigma^a_{\alpha\beta} i\partial_a \\ D=11: &\quad \{d_\alpha, d_\beta\} = -\sigma^a_{\alpha\beta} i\partial_a \end{aligned}$$

where in the case of D=6 we have taken advantage of the fact that SO(5,1)=SU*(4) to eliminate vector indices, and introduced the SU(2) index i for spinors to make them Majorana.

5. Theories

We first consider the scalar multiplet in D=6. The constraints and field equations are given by the statement, in terms of supersymmetry covariant derivatives, that there are only scalars and spinors on shell, and by supersymmetry their physical polarizations must be equal in number. Since a spinor has 4 polarizations in D=6, we must have 4 real scalars, and thus

$$d_{i\alpha}\phi_{jk'} = C_{ji}\psi_{k'\alpha}$$

The second SU(2) index k' is introduced again to make a spinor (this time the field) Majorana, and performs a similar service for the scalars. This one equation is sufficient to completely describe this multiplet on shell in the free case; interactions require derivatives, so we won't consider them here. This multiplet reduces to N=2 in D=4 in a very simple way: The SU(2) index on d labels the 2 supersymmetries, and the 4-component spinor index reduces in the obvious way to SL(2,C) indices, $\alpha \to (\alpha, \dot{\alpha})$, with appropriate 6D spinor conventions.

Exercise XC5.1

Show the equations given for the 6D scalar multiplet give the complete field equations for all the components, and that only the scalars and spinors shown explicitly in that equation survive on shell.

This six-dimensional theory gives a simple example of nontrivial dimensional reduction: Assume we have a 5-dimensional theory with a nontrivial U(1) symmetry. Then we can dimensionally reduce by choosing the fields to depend on the fifth coordinate in such a way that the fifth component of the momentum of each field is equal to a constant m (with dimensions of mass) times its U(1) charge Q:

$$p_4 = Z = mQ$$

This is consistent at the interacting level because each term in the action satsifies conservation of the U(1) charge as well as conservation of momentum. This is equivalent to how we introduced masses by dimensional reduction in subsection IIB4 for free fields, since any free field can be "complexified". This has an interesting effect on the supersymmetry algebra: It introduces a U(1) charge Z (called "central" because it commutes with the rest of the algebra). For example, if we start with the 6D supersymmetry algebra (like the above algebra for the supersymmetry covariant derivatives), introduce the central charge in reducing to 5, and then do an ordinary reduction to 4 (or vice versa), the supersymmetry algebra becomes (see subsection IVC7)

$$\{q_{i\alpha}, \bar{q}^j_{\dot\beta}\} = \delta_i^j p_{\alpha\dot\beta}, \quad \{q_{i\alpha}, q_{j\beta}\} = C_{\alpha\beta}C_{ij}Z, \quad \{\bar{q}^i_{\dot\alpha}, \bar{q}^j_{\dot\beta}\} = C_{\dot\alpha\dot\beta}C^{ij}Z$$

If the higher-dimensional theory was massless, then $p^2 + Z^2 = 0$ for the 4D theory. More generally, if the higher-dimensional theory already had masses before the central charge was introduced, then by supersymmetry it satisfied $p^2 + M_0^2 = 0$, $M_0^2 \geq 0$ (since supersymmetry always has positive potentials), while afterwards the 4D theory satisfies

$$p^2 + Z^2 + M_0^2 = 0 \quad \Rightarrow \quad M^2 = M_0^2 + Z^2 \geq Z^2$$

where M is the 4D mass, in terms of the higher-D mass M_0. However, in general, in the absence of central charges, massive representations of supersymmetry are bigger than massless ones (because there are twice as many independent supersymmetry generators on shell, since q is a spinor with 1 helicity for the massless case, but an SU(2) doublet for the massive). So, $M^2 = Z^2 > 0$ has the advantage of allowing smaller massive representations than when $M^2 > Z^2 = 0$ or when $M^2 > Z^2 > 0$. Note that when $M^2 = Z^2$, so all masses arise from the central charge, (total) mass is conserved, just as in nonrelativistic physics, although in the relativistic case the mass Z can be negative. (Of course, its square is always positive, as is physical energy. The relation between the relativistic and nonrelativistic cases can be understood through dimensional reduction: See exercise IA4.5. The mass is also a central charge for the Galilean group, but there the reduction is for a lightlike dimension.)

In the present case, we can choose our U(1) symmetry to be a subgroup of the extra SU(2) internal symmetry (k' index) of the 6D scalar multiplet. Note that the algebra of the d's is modified in the same way as that of the q's.

Super Yang-Mills is a bit more interesting, because interactions are easier to introduce. From the counting arguments given in subsection XC3, we see that a supersymmetric theory consisting of 1 vector and 1 spinor can exist in D=3, 4, 6, or 10. This corresponds directly with our analysis of the largest dimensions for simple supersymmetries: Dimensional reduction of a vector gives also scalars, so the condition of no scalars gives maximum dimensions. We now make an analysis similar to that of the previous subsection: By dimensional analysis for physical fields, and using single-Majorana-spinor-index notation,

$$\{\nabla_\alpha, \nabla_\beta\} = -\mathit{\Gamma}^a_{\alpha\beta} i\nabla_a$$

$$[\nabla_\alpha, \nabla_a] = \mathit{\Gamma}_{a\alpha\beta} W^\beta$$

$$[\nabla_a, \nabla_b] = iF_{ab}$$

Applying the Jacobi (Bianchi) identities, we find

$$\Gamma_{a(\alpha\beta}\Gamma^a_{\gamma)\delta} = 0$$

This identity can be satisfied only in D=3, 4, 6, or 10. The Bianchi identities imply the field equations for D=10.

Exercise XC5.2

Multiply the identity $\Gamma_{a(\alpha\beta}\Gamma^a_{\gamma)\delta} = 0$ by $\Gamma^{b\alpha\beta}$, and use the Γ matrix anticommutation relation $\Gamma^{(a}_{\alpha\beta}\Gamma^{b)\beta\gamma} = \eta^{ab}\delta^\gamma_\alpha$ to show that D=3, 4, 6, or 10.

Similar methods can be applied to D=11 supergravity. Our component counting for general dimensions, and our helicity analysis for general extended supersymmetric theories in D=4 (applied to the dimensionally reduced theory), can be satisfied by adding to the metric (44 physical components) and gravitino (128) a third-rank antisymmetric tensor gauge field (84) A_{mnp} (with field strength $F_{mnpq} = \frac{1}{6}\partial_{[m}A_{npq]}$). The action for the graviton and gravitino are like those in 4D N=1, while A has not only the obvious quadratic term but also a "Chern-Simons term":

$$L = \mathbf{e}^{-1}[-\tfrac{1}{4}R + \bar{\psi}_m\gamma^{mnp}\nabla_n\psi_p + \tfrac{1}{96}(F_{abcd})^2 + \psi^2 F + \psi^4]$$

$$+\tfrac{1}{4\cdot 3!(4!)^2}\epsilon^{mnpqrstuvwx}A_{mnp}F_{qrst}F_{uvwx}$$

(There are also more-complicated fermion interaction terms than in 4D N=1.) The necessity of the last term can be shown by finding the component form of the supersymmetry transformations, or by finding the field equations implied by the superspace formulation.

6. Reduction to D=4

We now look instead at the component formulation of higher-dimensional super Yang-Mills. This formulation is off shell except for the lack of auxiliary fields. Since the fields are just a vector and a spinor, the Lagrangian consists of just that of super Yang-Mills coupled to a spinor in the adjoint representation of the Yang-Mills group. Upon dimensional reduction, the vector produces some scalars. For example, the D=10 theory has an SO(9,1) symmetry, which reduces in D=4 to the SO(3,1)⊗SO(6) subgroup. The SO(6) symmetry of the 6 flattened dimensions is the SU(4) symmetry of the N=4 supersymmetries. Under this reduction, the vector becomes $10 \to (4,1) \oplus (1,6)$, namely a 4-vector and scalars that form a 6 of SU(4), while the spinor becomes $16 \to (4,4)$, a 4D spinor that is also a 4 of SU(4) (like the supersymmetry generators).

Although γ (or σ, or Γ) matrices are necessary in D=10, in D=4 we can convert to spinor notation for both SO(3,1) (=SL(2,C)) and SO(6) (=SU(4)). Thus vectors and the Minkowski metric reduce as

$$V^a \to (V^{\alpha\dot{\alpha}}, V^{ij}); \qquad (V^{\beta\dot{\alpha}})^* \equiv \bar{V}^{\alpha\dot{\beta}} = V^{\alpha\dot{\beta}}, \qquad (V^{ij})^* \equiv \bar{V}_{ij} = \tfrac{1}{2}\epsilon_{ijkl}V^{kl}$$

$$\eta_{ab} \to (C_{\alpha\beta}C_{\dot{\alpha}\dot{\beta}}, \epsilon_{ijkl}) : \qquad V\cdot W \to V^{\alpha\dot{\alpha}}W_{\alpha\dot{\alpha}} + \tfrac{1}{2}V^{ij}\overline{W}_{ij}$$

while spinors and Pauli matrices reduce as

$$\psi^\alpha \to \tfrac{1}{\sqrt{2}}\begin{pmatrix}\psi^{i\alpha}\\ \bar{\psi}_i{}^{\dot{\alpha}}\end{pmatrix}, \qquad \sigma^a_{\alpha\beta}V_a \to \begin{pmatrix} iC_{\beta\alpha}\bar{V}_{ij} & \delta_i^j V_{\alpha\dot{\beta}} \\ \delta_j^i V_{\beta\dot{\alpha}} & iC_{\dot{\beta}\dot{\alpha}}V^{ij}\end{pmatrix}$$

$$\psi \not{V} \chi \to V_{\alpha\dot{\alpha}}\tfrac{1}{2}(\psi^{i\alpha}\bar{\chi}_i{}^{\dot{\alpha}} - \chi^{i\alpha}\bar{\psi}_i{}^{\dot{\alpha}}) + \tfrac{1}{2}i(\bar{V}_{ij}\psi^{i\alpha}\chi^j{}_\alpha + V^{ij}\bar{\psi}_i{}^{\dot{\alpha}}\bar{\chi}_{j\dot{\alpha}})$$

The two terms in the 10D Lagrangian then reduce as

$$\begin{aligned} \tfrac{1}{8}F^2 \quad &\to \quad \tfrac{1}{8}F^2 + \tfrac{1}{8}[\nabla, \bar{\phi}_{ij}]\cdot[\nabla, \phi^{ij}] - \tfrac{1}{32}[\bar{\phi}_{ij}, \bar{\phi}_{kl}][\phi^{ij}, \phi^{kl}] \\ \psi^\alpha \sigma^a_{\alpha\beta}[-i\nabla_a, \psi^\beta] \quad &\to \quad \bar{\psi}_i{}^{\dot{\alpha}}[-i\nabla_{\alpha\dot{\alpha}}, \psi^{i\alpha}] + \tfrac{1}{2}i(\psi^{i\alpha}[\bar{\phi}_{ij}, \psi^j_\alpha] + \bar{\psi}_i{}^{\dot{\alpha}}[\phi^{ij}, \bar{\psi}_{j\dot{\alpha}}]) \end{aligned}$$

Exercise XC6.1

Looking at the SU(3) subgroup of SU(4), decompose the states of N=4 super Yang-Mills into those of N=3. (Use the analysis of subsection IIC5 to count states, in SU(N) representations.) Do the same to decompose N=4 into N=2 super Yang-Mills plus scalar multiplet, this time using the SU(2)⊗SU(2) subgroup for which $4 \to (\frac{1}{2},0)\oplus(0,\frac{1}{2})$ (i.e., $i \to (i,i')$). This is another way of understanding where the second SU(2) of the scalar multiplet comes from.

Exercise XC6.2

Derive the commutation relations of the N=4 Yang-Mills covariant derivatives of subsection IVC7 by dimensional reduction of those for 10D N=1 given in the previous subsection. (Don't forget the scalars come from the components of the vector covariant derivative in the extra dimensions.)

Dimensional reduction of (super)gravity is an example of the comparative simplicity of the vierbein (covariant derivative) formalism vs. the metric or even inverse vierbein (differential form) formalisms. The reason in this case is that gravity is treated like Yang-Mills theory, and gauge vectors result from reducing the graviton. This is seen most easily from comparison of the coordinate transformation laws:

$$\delta e_a{}^m = \lambda^n\partial_n e_a{}^m - e_a{}^n\partial_n\lambda^m$$

$$\delta e_m{}^a = \lambda^n\partial_n e_m{}^a + e_n{}^a\partial_m\lambda^n$$

$$\delta g_{mn} = \lambda^p \partial_p g_{mn} + g_{p(m}\partial_{n)}\lambda^p$$

Fixing the index $m = -1$ on λ^m to get the gauge transformations of an Abelian vector resulting from reduction from one extra dimension, and setting $\partial_{-1} = 0$ when acting on any field as the definition of reduction, we see the identification (in an appropriate gauge for the SO(D,1)/SO(D−1,1) generators M_{-1a})

$$e_a{}^m \quad \to \quad \begin{array}{cc} & \begin{array}{cc} m & -1 \end{array} \\ \begin{array}{c} a \\ -1 \end{array} & \begin{pmatrix} e_a{}^m & A_a \\ 0 & \psi \end{pmatrix} \end{array}$$

where A transforms in the usual way for a gauge vector, and ψ is an additional scalar. A more transparent way to write this is as

$$\lambda \equiv \lambda^m \partial_m, \qquad e_a \equiv e_a{}^m \partial_m; \qquad \delta e_a = [\lambda, e_a]$$

$$\lambda \to \lambda + \lambda^{-1}\partial_{-1}, \qquad e_a \to (e_a + A_a \partial_{-1}, \psi \partial_{-1})$$

which makes it clear that reduction has simply U(1)-covariantized the gauge parameter, transformation, and field, where ∂_{-1} is the U(1) generator. (Under reduction all fields are U(1) neutral.) On the other hand, the reduction of $e_m{}^a$, being the inverse of $e_a{}^m$, and g_{mn}, being the square of that, yields nonlinear reductions, and the U(1) covariantization is not manifest. (In particular, in the metric formalism the metric, and thus the U(1) vector, does not even appear in the covariant derivative, except in terms with its derivatives.)

Exercise XC6.3

Derive the result of exercise IXC1.1 by dimensional reduction.

Exercise XC6.4

Let's work out the details of this simple example, reduction of pure gravity from one extra dimension:

a Find the reduction of $c_{ab}{}^c$ by examining the commutators of the reduced e_a. (F_{ab} comes out directly.) Using the expression of the Lagrangian in terms of the c's from exercise IXA5.2, find the reduced action, including a cosmological term. (Drop the $\int dx^{-1}$. We can think of this as compactification on a circle, independence from x^{-1} yielding a constant factor upon integration, which can be absorbed.)

b The scalar appears in a funny way, seen previously in subsection IXB5. Rather than field redefinitions, it is more convenient to reintroduce local scale invariance (after the reduction), as in subsection IXB5, introducing the dilaton ϕ.

Then make a simple redefinition that replaces ψ and ϕ with the "canonical" fields $\phi_\pm$. (The F^2 and cosmological terms then appear with powers of $\phi_\pm$.)

Reduction from one extra dimension can give only a single (Abelian) gauge vector, but two or more dimensions can yield nonabelian gauge groups as the spacetime symmetries of the compactified dimensions. For example, compactifying n extra dimensions into an n-sphere gives SO(n+1). (However, compactifying to a box with periodic boundary conditions gives an Abelian group again.) The generalization is then

$$\lambda \quad \to \quad \lambda + \lambda^I G_I, \qquad e_a \quad \to \quad (e_a + A_a{}^I G_I, \psi_i{}^I G_I)$$

where the only dependence on the extra dimensions is implicit in the group generators G_I. If we add matter fields (before reduction), then the fields can be constrained to be independent of the extra dimensions (i.e., singlets of G_I) when their indices are flat.

Another possible modification is to make the action of the generators on matter fields nontrivial. If we already have an internal symmetry group, with generators $\hat{G}_I$, identical to that of the G_I, then we can impose on all matter fields ϕ

$$G_I \phi = \hat{G}_I \phi$$

to determine their dependence on the extra coordinates. The fact that the original higher-dimensional action was invariant under the $\hat{G}_I$ guarantees that the resultant dependence on the extra coordinates will cancel. The simplest example was applied to supersymmetry in the previous subsection: In the Abelian case we can set

$$-i\partial_{-1}\phi = m\hat{G}\phi$$

where we are free to scale the Abelian generator by a mass parameter m (unlike the nonabelian case, where it would change the algebra).

Similar results can be obtained for supergravity, but the results are more complicated, because the scalars (which appear for N>3) appear in nonlinear σ models. Furthermore, although these models can be constructed by the coset method discussed in subsection IVA3, the coset space G/H is noncompact, because the group G is noncompact, although the subgroup H is compact. This is a consequence of the fact that the "compensating" scalars of the group H=U(N) (or SU(8) for N=8) appear with the wrong-sign kinetic term (as the dilaton even in ordinary gravity). Thus, conformal supergravity is coupled to "matter" with scalars in the adjoint representation of the noncompact group G, while gauging away the compensating scalars leaves the physical scalars of the coset space G/H. A simpler analog is N=1 supergravity coupled to a scalar multiplet (see subsection XB7). This is the same as conformal supergravity coupled to the matter action $\bar{\phi}\phi - \bar{\chi}\chi$, which has a symmetry G=U(1,1), while N=1 supergravity has a gauge group U(1). Including Weyl scale invariance GL(1), the physical scalars then inhabit the coset space U(1,1)/U(1)⊗GL(1)=SU(1,1)/GL(1).

In the case of extended supergravity, the group G can be found by noting that the physical scalars parametrizing G/H form the representation ϕ^{ijkl} (totally antisymmetric, and complex conjugate) of the group H. We then look for the group G whose adjoint representation transforms under the H subgroup as these scalars + adjoint of H. We can also determine G by defining group generators for G as $M_i{}^j$ for H, and M_{ijkl} (and hermitian conjugate $\overline{M}^{ijkl}$) for G/H, and write commutation relations consistent with covariance under H. For N=8 we also have $\overline{M}^{ijkl} = \frac{1}{4!}\epsilon^{ijklmnpq}M_{mnpq}$ (and the same for the corresponding physical scalars). The result for the coset space G/H is

$$
\begin{aligned}
N = 4 &: SU(4) \otimes SU(1,1)/U(4) = SU(1,1)/U(1) \\
5 &: SU(5,1)/U(5) \\
6 &: SO^*(12)/U(6) \\
8 &: E_{7(+7)}/SU(8)
\end{aligned}
$$

where $E_{7(+7)}$ is a noncompact form (Wick rotation) of the exceptional group E_7.

An additional complication is that the vectors represent the full H symmetry only on shell. For example, for N=2 we have a single vector, as in electromagnetism. Maxwell's equations without sources have a U(1) symmetry, "S-duality", that transforms $f_{\alpha\beta}$ by a phase (and $\bar{f}_{\dot{\alpha}\dot{\beta}}$ by the opposite), that mixes the field equations with the Bianchi identities. With sources, it mixes electric and magnetic charge, since it mixes electric and magnetic fields. So, in general we must introduce both electric and magnetic potentials for each vector. Furthermore, for N=6 the vectors appear as both $f_{\alpha\beta}^{ij}$ and $f_{ijklmn\alpha\beta}$ (one extra vector). (For N=8, the two are related by an ϵ tensor, just as ϕ and $\bar{\phi}$.) In this version of extended supergravity, all the vectors are Abelian. There is also a version where they gauge SO(N), but that theory has a cosmological constant.

REFERENCES

1 Georgi, *loc. cit.* (IB):
Dirac and irreducible spinors for SO(n).
2 Kugo and Townsend, *loc. cit.* (IIC):
Dirac and irreducible spinors for SO(n_+,n_-).

3 Siegel, *loc. cit.*:
spinors and γ, σ, Γ matrices in SO(n_+,n_-).
4 L. Brink, J.H. Schwarz, and J. Scherk, *Nucl. Phys.* **B121** (1977) 77:
D=10 super Yang-Mills.
5 E. Cremmer, B. Julia, and J. Scherk, *Phys. Lett.* **76B** (1978) 409:
D=11 supergravity.
6 Haag, Łopuszański, Sohnius, *loc. cit.* (IIC):
central charges.
7 W. Nahm, *Nucl. Phys.* **B135** (1978) 149:
arbitrary free representations of supersymmetry in arbitrary dimensions, central charges from dimensional reduction.
8 W. Siegel, *Phys. Lett.* **80B** (1979) 220:
higher-dimensional superspace.
9 E. Cremmer and B. Julia, *Phys. Lett.* **80B** (1978) 48, *Nucl. Phys.* **B159** (1979) 141:
N=8 supergravity.
10 L. Brink and P. Howe, *Phys. Lett.* **88B** (1979) 268;
W. Siegel, *Nucl. Phys.* **B177** (1981) 325:
N=8 superspace formulation of supergravity.
11 E. Cremmer and S. Ferrara, *Phys. Lett.* **91B** (1980) 61;
L. Brink and P. Howe, *Phys. Lett.* **91B** (1980) 384:
superspace formulation of D=11 supergravity.
12 Th. Kaluza, *Sitz. Preuss. Akad. Wiss. Berlin, Math.-phys. Kl.* (1921) 966;
Klein, *loc. cit.* (IIB);
H. Mandel, *Z. Phys.* **39** (1926) 136:
dimensional reduction/compactification of general relativity, Abelian gauge field.
13 Klein, Pauli, *loc. cit.* (IIIC):
same with nonabelian gauge fields.
14 J. Scherk and J.H. Schwarz, *Phys. Lett.* **82B** (1979) 60:
identifying symmetries of compactified spaces with existing internal symmetries — masses from dimensional reduction in interacting theories.

XI. STRINGS

There are three areas of application of QCD, as defined by the region of momentum space they address: (1) One is perturbative QCD, which applies to large relative, "transverse" velocity of (some of) the constituents of the hadrons. In this approach such an amplitude is divided into a half consisting of high-energy, asymptotically-free partons, which is calculated perturbatively in the gauge coupling, and a half consisting of low-energy, confined partons, which is nonperturbative, and therefore not calculated.

(2) Another area deals with the low-energy behavior of QCD — with properties of the vacuum (e.g., broken chiral symmetry), or the lowest-mass hadrons, scattering at small relative velocities. This approach is nonperturbative with respect to the gauge coupling, and instead perturbs in derivatives, as in first-quantized JWKB. The methods used include instantons, lattice QCD, current algebra, dispersion relations, nonlinear σ models, and duality. This low-energy behavior really says nothing about confinement, just as the low-energy states of the hydrogen atom tell us nothing about ionization.

A closely related problem is that the nonperturbative information about QCD that comes from (electromagnetic-type) duality considerations, which relates "weak" coupling to "strong" coupling as $g \leftrightarrow 1/g$, is not really relating quark-gluon physics to hadronic physics, but is relating quark-gluon physics to monopole physics; i.e., it relates a description of weakly coupled "electric" color charges to a similar looking theory of weakly coupled "magnetic" color charges. Thus, the dual theory, being formally of the same type as the original, except for a relabeling of what is called "electric" and what is called "magnetic", does not give anything that looks any more like hadrons, or make it any easier to calculate.

(3) The one nonperturbative approach that does deal with high (hadron) energies is string theory: It incorporates hadrons of arbitrarily high mass, and studies their scattering at high energies. It also shows that stringy (hadron-like) behavior is a characteristic of QCD coupling $g \approx 1$, while $g \approx 0$ or ∞ have non-stringy (parton-like) behavior: String perturbation expands in $G = ln\ g$, not g nor $1/g$; duality is the symmetry $G \leftrightarrow -G$. Furthermore, this G is the coupling that defines the free string, i.e., how partons bind to form strings. The coupling that determines how hadrons couple to each other is $1/N_c$, as described topologically in subsection VC9. (However, the relation of duality to the $1/N_c$ expansion is unclear, since duality has been studied so far only in relation to theories where the group is spontaneously broken to U(1),

so effectively $N_c = 1$, or with respect to instantons, which are always defined for SU(2) subgroups, so effectively $N_c = 2$. Thus, it has not been possible to apply such duality arguments simultaneously with a $1/N_c$ analysis. Similarly, since both these duality approaches deal only with low-energy behavior, they are difficult to relate to confinement.)

Most of the research effort on string theory has been directed toward models with "critical dimension" D$\gg$4 (10, 11, or 26): Attempts to relate to the real world involve choosing a solution of minimal energy with all but 4 dimensions "compactified", and considering perturbations about that solution. (The extra dimensions cannot be completely eliminated without losing renormalizability.) Unfortunately, the ambiguity in choice of such a solution results in the loss of predictability, which is what unification and renormalizability are all about.

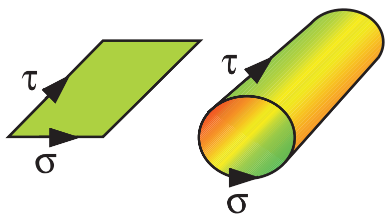
However, historically the true usefulness of such string theories has been for the concepts and features of field theory they have revealed: For example, supersymmetry (sections IIC and IVC, and chapter X), the Gervais-Neveu gauge (subsection VIB4), topological (1/N) expansion (subsection VIIC4), first-quantized BRST approach to gauge theory (chapter XII), and certain simplifications in one-loop amplitudes were all discovered through studies of 10- and 26-dimensional string theory, even though they all are now understood more simply through ordinary field theory. This is due to the fact that string theories are so complex and restrictive that they require the most powerful techniques available. Clearly such strings are useful toy models for learning about particle field theories, and about general properties of string theory that might lead to generalizations to include realistic 4-dimensional string theories. (In fact, the first paper on string theory was written in 1747 by d'Alembert, and was the first appearance of the wave equation and the d'Alembertian. Thus, field theory, quantum mechanics, and special relativity can trace their origins to string theory.)

In subsections IVB1 and VIIC4 we briefly discussed how hadrons are expected to arise as strings from QCD. In this chapter we analyze the dynamics of this mechanism. We begin by formulating the theory in terms of strings directly. Perturbative calculations are performed using first-quantized path integrals. The only experimental evidence for strings is as a description of hadrons; to some extent the way that QCD leads to strings can be understood with similar first-quantized methods, based on random lattices.

A. SCATTERING

String theories are the only known theories that exhibit (S-matrix) duality. Unlike other S-matrix approaches, they provide an explicit perturbative calculational scheme, like field theory, and string theory can be formulated as a field theory. Also like field theory, string theory has consistency conditions at the classical and quantum levels, related to gauge invariance and renormalizability.

The defining concept of the string is that it is a two-dimensional object: Just as the particle is defined as a point object whose trajectory through spacetime is one-dimensional (a worldline), the string has as its trajectory a two-dimensional surface, the "worldsheet". This leads to a much simpler picture of interactions: For interacting particles, the geometric picture of a worldline becomes a graph, whose geometry is singular at the interaction points. For interacting strings, we have instead a worldsheet with nontrivial topology: sphere, disk, torus (doughnut), etc. For example, a tree graph now looks more like a real tree, in that the branches now have thickness, and they join smoothly to the rest of the tree. There are two types of free strings: open (two ends) and closed (no boundary). Their worldsheets are a rectangle and a tube (cylinder).



1. Regge theory

In principle there is no difference between a fundamental state and a bound state: We can always write an action with every state represented by an independent field. Of course, such an action might not be renormalizable, but that seems more of a formal distinction. A more physical one is based on the qualitative property that bound states have radial and other excitations with related properties, while fundamental states are more unique.

Regge theory is an approach to bound states that treats them as fundamental. A family of states that are different excitations of the same ground state is treated as a single entity. Although basically an approach based on fundamental properties of the S-matrix, when combined with perturbation theory it leads directly to string theory.

A quantitative definition of this concept follows from a generalization of a concept seen in perturbative field theory. In amplitudes following from Feynman diagrams the nature of intermediate states can be seen from the momentum-space behavior: Single-particle states appear as poles (in the sense of complex analysis) in some momentum invariants, $1/(p^2 + m^2)$, where this p is the sum of some of the external momenta, representing the momentum of the internal state. (Any tree graph is a simple example.) Two-particle states appear as cuts in these invariants, where the branch point represents the state where the two particles are at rest with respect to one another, and the rest of the cut corresponds to arbitrary relative velocities. (For example, a one-loop propagator correction has a branch point at $-p^2 = (m_1 + m_2)^2$ for intermediate particles of masses m_1 and m_2.) Similar remarks apply to other multi-particle states. "Analytic S-matrix theory" was an attempt to formulate particle physics in terms of the S-matrix by replacing the property of locality of the action with "maximal" analyticity of the S-matrix in momentum space. (Of course, unitarity and Poincaré invariance can be described easily in terms of the S-matrix; even analogs of renormalizability can be formulated in terms of certain properties of the high-energy behavior.) Unfortunately, the most general form of such nonanalytic behavior (poles, cuts, etc.), as discovered from analyzing Feynman diagrams, proved to be too complicated to provide a practical method for defining a theory.

Since Poincaré invariance means that not only momentum is conserved but also angular momentum, a natural next step was to consider the analytic behavior in that variable as well. This behavior is seen already in nonrelativistic theories; here we will approach the concept in a language most relevant to relativistic physics. The simplest example of "Regge behavior" is the 4-point S-matrix; this is the relativistic analog of a nonrelativistic particle in a potential. (We can think of an infinitely massive second particle as producing the potential, or separate center-of-mass and relative coordinates for two finite-mass particles.) Also, the Feynman diagrams that appear in the nonrelativistic problem are "ladder diagrams": The sides of the ladder represent the two scattering particles, while the rungs represent a perturbation expansion for the potential. It can be shown that such diagrams give the leading behavior of this amplitude at high energies. Here the appropriate high energy limit is defined in terms of the Mandelstam variables (see subsection IA4); by high energy we mean, e.g., $s \to -\infty$ for fixed t.

The high-energy behavior of ladder diagrams can be shown to be of the form (in units of an appropriate mass)

$$\mathcal{A}(s,t) = kg^2 \Gamma[-\alpha(t)](-s)^{\alpha(t)}, \qquad \alpha(t) = a + g^2 b(t)$$

where a is a constant that describes the behavior of the tree graph, and $b(t)$ is determined by the one-loop graph. (Both, and the constant k, are independent of g.) This amplitude takes a simple form under a modified type of "Sommerfeld-Watson transform":

$$\mathcal{A}(s,t) = \oint \frac{dJ}{2\pi i}\ \Gamma(-J)(-s)^J f(J,t) = \sum_{J=0}^{\infty} \tfrac{1}{J!} s^J f(J,t)$$

The contour integral is taken as counterclockwise about the positive real axis to obtain the last form, where it picks up the poles of $\Gamma(-J)$ (see exercise VIIA2.3b), but can be deformed to surround the singularities of f. In this case, a pole at $J = \alpha(t)$ in f will reproduce our original ladder amplitude, while integrating it around the positive real axis gives

$$f(J,t) = -k\frac{g^2}{J-\alpha(t)} \qquad \Rightarrow \qquad \mathcal{A}(s,t) = k\sum_{J=0}^{\infty} \tfrac{1}{J!} s^J \frac{g^2}{J-\alpha(t)}$$

which shows that particles of spin J contribute simple poles in t to the amplitude at $\alpha(t) = J$ when $\alpha(t)$ can be approximated as linear near that value. The spin of the intermediate particle follows from the s^J factor. (This is clear from examining a 4-point tree graph where the external lines are scalars and the internal line carries J indices, and must contract the momenta of its two ends. There are also contributions of lower spins from traces.) Thus the "Regge trajectory" $\alpha(t)$ determines not only the high-energy behavior of the amplitude (for negative t), but also the spins and masses of the bound states (for positive t): Looking at the graph for $J = \alpha(t)$, there is a bound state of spin J and mass $\sqrt{t}$ whenever the curve crosses an integer value of J. The contribution at n loops in perturbation theory to $f(J,t)$ is a multiple pole $(J-a)^{-(n+1)}$, which contributes to $\mathcal{A}(s,t)$ a term proportional to $(-s)^a[ln(-s)]^n$.

Exercise XIA1.1

Consider the amplitude

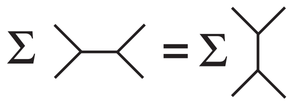
$$\mathcal{A} = \int_0^\infty d\tau\ e^{\tau s}[f(\tau)]^{-\alpha(t)-1}, \qquad f(0) = 0, \quad f'(0) \neq 0$$

where f is Taylor expandable. By expanding f, show that a sum of Regge amplitudes is obtained, where the "leading trajectory" is $\alpha(t)$, and there are "daughter trajectories" $\alpha(t) - n$ for positive integer n.

Unfortunately, for field theories with a finite number of fundamental particles, the trajectories are rather boring, containing only a finite number of bound states. In certain cases a trajectory may include one of the fundamental particles itself ("Reggeization"). Because of the usual infrared divergences, such calculations can be applied directly to S-matrix elements only for fundamental massive particles; for fundamental massless particles, as in confining theories (like QCD), these results require external lines to be off-shell, and some knowledge of the parton wave functions is needed. Regge behavior thus gives a measurable definition of confinement: If the scattering amplitudes of color-singlet states (or color-singlet channels of off-shell amplitudes of color-nonsinglet states) have linear trajectories, the constitutent color-nonsinglet particles can be said to be "confined". On the other hand, if the Regge trajectory rises only to finite spin and then falls, as with the Higgs effect, then there is only "color screening"; color-singlet states might not be observable, but we do not see the infinite number of radial excitations characteristic of confinement. Another possibility is that arbitrarily high spin is reached at finite energy: This is characteristic of Coloumb binding, and indicates that a new, "ionized" phase is reached above that energy.

Experimentally, hadrons are observed to have Regge behavior with respect to both high-energy behavior and spectrum. However, those Regge trajectories are approximately linear, thus indicating an (near) infinite number of bound states. The linearity of the trajectories can be shown to be related to the relative stability of these unstable particles (as compared to what is found in ladder approximations). This suggests a formulation of the theory of hadrons where the whole Regge trajectory is treated as fundamental. It can be shown that in any such "Regge theory" based on a perturbation expansion where the "tree" graphs have only poles in the angular momentum J (whose accuracy is implied by the linearity of the observed trajectories), that the theory has a further property called (s-t) "duality": This property states that the amplitude can be expressed as a sum of poles in either the s or t "channel", rather than as a sum over both:

$$\mathcal{A}(s,t) = \sum_n \frac{C_n(t)}{s - s_n} = \sum_n \frac{\tilde{C}_n(s)}{t - t_n}$$



This holds even when the sets of particles exchanged in the two channels are different, due to quantum numbers of the external states. This relation has also been experimentally verified (approximately).

Explicit realizations of such "dual models" of the S-matrix in terms of first-quantized systems are called "string theories". They explain the linearity of the Regge trajectories by the harmonic-oscillator structure of the string Hamiltonian, and the duality of the amplitudes by the conformal invariance ("stretchiness") of the string worldsheet.

2. Classical mechanics

We now consider string theory as derived by first quantization. As for particles, the first step is to study the classical mechanics, which determines the appropriate set of variables, the kinetic term of the field theoretic action, some properties of the interactions, and some techniques useful for perturbation. Just as the simplest such action for the particle produces only the relatively uninteresting case of the scalar, the most obvious action for the string yields a model that is not only too simple, but quantum mechanically consistent only in 26 dimensions. However, this toy model exhibits many relevant qualitative features, such as Regge behavior and duality. Later we'll consider the source of its problems by relating to four-dimensional particle theories.

The simplest classical mechanics action for the string is a direct generalization of that for the massless scalar particle: For the Lagrangian form of this action we write

$$S_L = \frac{1}{\alpha'} \int \frac{d^2\sigma}{2\pi} \sqrt{-g} g^{mn} \tfrac{1}{2} (\partial_m X^a) \cdot (\partial_n X^b) \eta_{ab}$$

where $X^a(\sigma^m)$ is the position in spacetime of a point at worldsheet coordinates $\sigma^m = (\sigma^0, \sigma^1) = (\tau, \sigma)$, $g^{mn}(\sigma^m)$ is the (inverse) worldsheet metric, and α' is a normalization constant, the string tension. It can also be associated with the flat-space spacetime metric η_{ab}; if we couple a spacetime metric, then its vacuum value can be taken as η_{ab}/α', where α' is the gravitational coupling, as discussed in subsection IXA5.

Exercise XIA2.1

Analyze the classical mechanics of the string by approximating σ by a set of discrete points, so $X'(\sigma) \to X_{n+1} - X_n$, etc. Choose the gauge $X^0 = \tau$. Show that the string then acts as a bunch of particles connected by springs, and find all the usual spring properties: tension, speed of wave propagation, etc.

A new feature of this action (compared to the particle's) is that it is (2D) Weyl scale invariant. This gauge invariance can be used to gauge away one component of the metric, in addition to the two that can be gauged away using 2D general coordinate invariance. The net result is that the worldsheet metric can be completely gauged away (except for some bits at boundaries), just as for the particle. However, this same invariance prevents the addition of a worldsheet cosmological term: In the particle case, such a term was needed to introduce mass. Here, mass is introduced through the coefficient $1/\alpha'$ of the $(\partial X)^2$ term: The same scale invariance that prevents use of a cosmological term also prevents this coefficient from being absorbed into the definition of the worldsheet metric.

Just as for the particle, the metric can be eliminated by its equation of motion, resulting in a more geometrical, but less useful, form of the action: In this case the equation of motion ("Virasoro constraints")

$$(\partial_m X) \cdot (\partial_n X) = \tfrac{1}{2} g_{mn} g^{pq} (\partial_p X) \cdot (\partial_q X)$$

after taking the determinant of both sides, gives

$$S = \frac{1}{\alpha'} \int \frac{d^2\sigma}{2\pi} \sqrt{-\tilde{g}}, \qquad \tilde{g}_{mn} = (\partial_m X) \cdot (\partial_n X)$$

This is the area of the string in terms of the "induced" metric $\tilde{g}_{mn}$, analogously to the particle case. The induced metric measures length as usually measured in spacetime:

$$d\sigma^m d\sigma^n \tilde{g}_{mn} = (d\sigma^m \partial_m X) \cdot (d\sigma^n \partial_n X) = (dX)^2$$

Equivalently, this action can be written in terms of the area element $dX^a \wedge dX^b$:

$$S = \frac{1}{2\pi\alpha'} \int \sqrt{-\tfrac{1}{2}(dX^a \wedge dX^b)^2}, \qquad dX^a \wedge dX^b = (d\sigma^0 \partial_0 X^{[a})(d\sigma^1 \partial_1 X^{b]})$$

For purposes of quantization, it's also useful to have the Hamiltonian form of the action. This also allows us to see how the Virasoro constraints generalize the Klein-Gordon equation, and then find the BRST operator. By the usual methods of converting from Lagrangian to Hamiltonian, we find

$$S_H = \int \frac{d^2\sigma}{2\pi} (-\dot{X} \cdot P + \mathcal{H}), \qquad \mathcal{H} = \frac{\sqrt{-g}}{g_{11}} \tfrac{1}{2} (\alpha' P^2 + \alpha'^{-1} X'^2) + \frac{g_{01}}{g_{11}} X' \cdot P$$

where $\dot{} = \partial_0$ and $' = \partial_1$. Various combinations of components of the worldsheet metric now appear explicitly as Lagrange multipliers. If we define

$$\hat{P}_{(\pm)} = \tfrac{1}{\sqrt{2}} (\alpha'^{1/2} P \pm \alpha'^{-1/2} X') \quad \Rightarrow \quad [\hat{P}_{(+)}, \hat{P}_{(-)}] = 0$$

the constraints can be written as two independent sets $\hat{P}^2_{(\pm)}$.

Exercise XIA2.2

Show that if we call $g_\pm$ the Lagrange multipliers for $\hat{P}^2_{(\pm)}$, then in convenient local Lorentz and Weyl scale (but not coordinate) gauges we can write in a lightcone basis

$$e_\pm = e_\pm{}^m \partial_m = \tfrac{1}{\sqrt{2}} (\partial_0 \pm g_\mp \partial_1)$$

while in another scale gauge we can write

$$dx^m e_m{}^\pm = \tfrac{1}{\sqrt{2}} (dx^0 g_\pm \pm dx^1)$$

Exercise XIA2.3

Find the (equal-τ) commutation relations $[\hat{P}_{(\pm)}, \hat{P}_{(\pm)}]$. Show that the (semi-classical) commutation relations of the constraints $\hat{P}^2_{(\pm)}$ close. (Hint: Use the identity $f(a)\delta'(a-b) = f(b)\delta'(a-b) - f'(b)\delta(a-b)$.)

Since 2D general coordinate (and even just Lorentz) invariance is no longer manifest, for some purposes we need to generalize this to a form that is first-order with respect to both τ and σ derivatives:

$$S_1 = -\frac{1}{\alpha'}\int \frac{d^2\sigma}{2\pi}[(\partial_m X)\cdot P^m + (-g)^{-1/2} g_{mn}\tfrac{1}{2}P^m \cdot P^n]$$

obviously reproduces S_L after eliminating P^m. Eliminating just P^1 gives a simpler way of deriving S_H (with $P^0 = \alpha' P$).

Since open strings have boundaries, the action implies boundary conditions, originating from integration by parts when deriving the field equations. In the last form of the action variation of the first term gives, in addition to the $\int d^2\sigma$ terms $(\delta P)\cdot\partial X$ and $-(\delta X)\cdot\partial P$ for the field equations, a boundary term $\int d\sigma^m \epsilon_{mn}(\delta X)\cdot P^n$, where $d\sigma^m$ is a line integral along the boundary, and the ϵ_{mn} picks the component of P^m normal to the boundary. We thus have

$$n_m P^m = 0 \textit{ at boundaries}$$

where n_m is a vector normal to the boundary. From the constraint imposed by varying g_{mn}, it then follows that

$$(t_m P^m)^2 = 0 \textit{ at boundaries}$$

where t_m is a vector tangent to the boundary (or any vector, for that matter). Since by the field equations $P^m \sim g^{mn}\partial_n X$, this means that the boundary is lightlike in spacetime: The ends of the string travel at the speed of light.

3. Gauges

In direct analogy to the particle (subsection IIIB2), the two most useful gauges are the "conformal gauge", defined by completely fixing the worldsheet metric, and the lightcone gauge, which is not manifestly globally covariant but is a complete fixing of the residual gauge invariance of the conformal gauge. In the conformal gauge we set

$$g_{mn} = \eta_{mn}$$

by using the 2 coordinate invariances and the 1 scale invariance to fix the 3 components of the symmetric tensor g_{mn}. The coordinate part of this gauge is essentially the temporal gauge $g_{0m} = \eta_{0m}$, just as for the particle ($-g_{00} = v^2 = 1$). Also as for the particle, this gauge can't be fixed everywhere (see also subsections IIIA5 and IIIC2), but the equation of motion from the metric is implied everywhere by imposing it at the just the boundaries in τ. In this gauge the equations of motion for X are just the 2D Klein-Gordon equation, which is easy to solve in 2D lightcone coordinates:

$$\partial_+\partial_- X = 0 \quad \Rightarrow \quad X = X_{(+)}(\tau+\sigma) + X_{(-)}(\tau-\sigma)$$

(We have used $\tau \pm \sigma$ in place of $\sigma^{\pm}$ for later convenience.) The constraints are then $\hat{P}^2_{(\pm)} \sim (X'_{(\pm)})^2 = 0$. This directly relates to the form of 2D conformal transformations, which are infinite-dimensional in D=2:

$$ds^2 = 2d\sigma^+ d\sigma^- \quad \Rightarrow \quad \sigma'^+ = f_{(+)}(\sigma^+), \quad \sigma'^- = f_{(-)}(\sigma^-)$$

The constraints are the generators of these conformal transformations. (As described in subsection IIIA5, the constraints generate the gauge transformations; the global transformations are those that preserve the temporal gauge.)

For the lightcone gauge, we again fix the (spacetime) +-components of the variables, and solve the +-components of the equations of motion (found by varying the −-components). Looking at the equations of motion first, using the first-order form of the action,

$$0 = \frac{\delta S}{P^{-m}} \sim \partial_m X^+ + (-g)^{-1/2} g_{mn} P^{+n}$$

$$\Rightarrow \quad (-g)^{-1/2} g_{mn} = (A\cdot B)^{-1}(\epsilon_{mp}A^p \epsilon_{nq}A^q - B_m B_n); \qquad A^m = P^{+m}, \quad B_m = \partial_m X^+$$

(as seen, e.g., by using $\epsilon_{mn}A^n, B_m$ as a basis), and

$$0 = \frac{\delta S}{X^-} \sim \partial_m P^{+m} \quad \Rightarrow \quad \frac{d}{d\tau}\int d\sigma\ P^{+0} = 0$$

which identifies $\int d\sigma\ P^{+0}$ as the conserved momentum p^+, up to a factor of $2\pi\alpha'$ (since p is really the coefficient of $\dot{x}$ in the action, where $X(\sigma) = x + ...$). Similarly, $\delta S/\delta g^{mn}$ determines P^{-m}, and thus X^-.

We then choose as our main set of gauge conditions

$$X^+ = k\tau, \qquad P^{+0} = k$$

for some constant k, which explicitly determines τ, and determines σ up to a function of τ: An equivalent way to define the lightcone σ in terms of an arbitrary spacelike coordinate σ' is

$$\sigma = k^{-1}\int_0^{\sigma} d\sigma'\ P^{+0}(\sigma')$$

which identifies σ as the amount of momentum p^+ between that value of σ and $\sigma = 0$ (at fixed τ). We thus have that the length of the string (the range of σ, not the physical length) is

$$l = k^{-1} \int d\sigma \ P^{+0} = 2\pi\alpha' p^+ k^{-1}$$

We then need to fix the location of $\sigma = 0$ as some function $\sigma'(\tau)$: Since in this gauge

$$\partial_1 P^{+1} = 0$$

so P^{+1} is also a function of just τ, we further fix the gauge for σ by choosing

$$P^{+1} = 0 \quad \Rightarrow \quad (-g)^{-1/2} g_{mn} = \eta_{mn}$$

Thus the lightcone gauge is a special case of the conformal gauge, after also fixing scale gauge $g = -1$. For the open string, this almost fixes $\sigma'(\tau)$ at $\sigma = 0$, which we can take as one boundary: The boundary condition for X^+ is now

$$0 = n \cdot \partial X^+ \sim n_0$$

since in this (and any conformal) gauge $\partial_m X \sim \eta_{mn} P^n$. Thus the normal to the boundary must be in the σ direction, so the boundary is at constant σ. This means we have one constant left to fix:

$$\sigma = 0 \ \mathit{at\ one\ boundary\ (open\ string)}$$

This invariance was left because all our previous gauge conditions preserved global σ translation. Unfortunately, there is no corresponding convenient gauge choice for the closed string, so there we leave just this one invariance. In summary, our complete set of ligtcone gauge conditions is now:

$$gauge: \qquad X^+ = k\tau, \qquad P^{+m} = k\delta_0^m, \qquad \sigma = 0 \ \mathit{at\ one\ boundary\ (open\ string)}$$

The lightcone action is now, in Hamiltonian form,

$$S_{lc} = -\int d\tau \left\{ \dot{x}^- p^+ + \int \frac{d\sigma}{2\pi} \left[\dot{X}_i P_i - \tfrac{1}{2}(\alpha' P_i^2 + \alpha'^{-1} X_i'^2) \right] \right\}$$

The only distinction between open and closed strings is the boundary condition (since closed strings by definition have no boundary). For closed strings we have only periodicity in σ (by definition of "closed"), while for open strings we have

$$X'(\tau, 0) = X'(\tau, l) = 0$$

One consequence, as we just saw, is that closed strings have one residual gauge invariance in the lightcone gauge. These two strings can be made to resemble each other more closely by extending the open string to twice its length, defining X for negative σ by

$$X(\tau,-\sigma) = X(\tau,\sigma)$$

This is the known as the "method of images": $X(\tau,-\sigma)$ is identified with its mirror image in the τ axis, $X(\tau,-\sigma)$. Then the two strings both satisfy periodic boundary conditions, while the open string has this one additional condition. We also choose

$$k = 2\kappa\alpha' p^+, \qquad \kappa = \begin{cases} 1 & (open) \\ \frac{1}{2} & (closed) \end{cases} \quad \Rightarrow \quad l = \frac{\pi}{\kappa}$$

so the length of the closed string is 2π, while the open string has original length π that has now been doubled to match the closed string. Our choice of "phase" in relating X for positive and negative σ for the open string automatically enforces the boundary condition $X'(\tau,0) = 0$ at one end of the string, while the condition $X'(\tau,\pi) = 0$ at the other end is implied in the same way by the closed string "boundary condition" of periodicity, which can be written as $X(\tau,\pi) = X(\tau,-\pi)$. The picture is then that the open string is a closed string that has collapsed on itself, so that for half of the range of σ X doubles back over the path it covered for the other half.

Because σ has a finite range, X can always be expanded in Fourier modes in that variable; the boundary conditions slightly restrict the form of this expansion. We saw that the equations of motion, being second-order in τ-derivatives, gave two modes for each initial state: a left-handed one and a right-handed one. We need to be a bit more precise about the zero-modes (independent of σ): We can separate them out as

$$X(\tau,\sigma) = x + \frac{2\pi\alpha'}{l} p\tau + \sqrt{\tfrac{\alpha'}{2}}[Y_{(+)}(\tau+\sigma) + Y_{(-)}(\tau-\sigma)], \qquad \int d\sigma\ Y_{(\pm)} = 0$$

where Y contains only nonzero-modes. (The normalization of p, conjugate to x, comes from the $-\dot{x}\cdot p$ term in the Lagrangian.) Then x represents the "center of mass" of the string, and p its total momentum. Note that this implies $X_{(\pm)}$ aren't quite periodic:

$$X_{(\pm)}(\sigma + 2\pi) = X_{(\pm)}(\sigma) + 2\pi\kappa\alpha' p$$

Now the periodicity boundary conditions shared by open and closed strings imply

$$Y_{(\pm)}(\sigma + 2\pi) = Y_{(\pm)}(\sigma)$$

while the extra boundary condition for the open string implies

$$Y_{(+)}(\sigma) = Y_{(-)}(\sigma) = Y(\sigma)$$

allowing us to drop the subscript in that case. Thus, the closed string has twice as many modes as the open, except for the nonperiodic part, corresponding to the total momentum and average position. This is related to the interpretation that the open string is a closed string with its two halves occupying the same path. This doubling also shows up in the constraints: For the closed string we have $\hat{P}^2_{(\pm)}$, while for the open string we can consider just $\hat{P}^2_{(+)}$, since $\hat{P}^2_{(-)}(\sigma) = \hat{P}^2_{(+)}(-\sigma)$. In the lightcone gauge we solve these constraints for X^-, by integrating

$$0 = \hat{P}^2_{(\pm)} \sim \dot{X}^2_{(\pm)} = \left(\dot{X}^i_{(\pm)}\right)^2 - k\dot{X}^-_{(\pm)} \sim (\dot{Y}_{(\pm)} + \kappa\sqrt{2\alpha'}p)^2$$

Exercise XIA3.1

Rederive the solution to the boundary conditions for the open string without using $X(\tau,-\sigma) = X(\tau,\sigma)$ (and periodicity): The string, as originally, extends between boundaries at 0 and π.

This separation of zero-modes from nonzero-modes also allows us to find the spin and mass of the string: In any conformal gauge,

$$0 = p^2 + M^2 = \frac{1}{\kappa\alpha'^2}\int_0^{\pi/\kappa}\frac{d\sigma}{2\pi}\ \tfrac{1}{2}(\dot{X}^2 + X'^2) \quad \Rightarrow \quad M^2 = \frac{1}{2\kappa\alpha'}\sum_{\pm}\int\frac{d\sigma}{2\pi}\ \dot{Y}^2_{(\pm)}$$

$$J^{ab} = x^{[a}p^{b]} + S^{ab} = \frac{1}{\alpha'}\int_0^{\pi/\kappa}\frac{d\sigma}{2\pi}\ X^{[a}\dot{X}^{b]} \quad \Rightarrow \quad S^{ab} = \sum_{\pm}\int\frac{d\sigma}{2\pi}\ Y^a_{(\pm)}\dot{Y}^b_{(\pm)}$$

(using the Hermitian form of the Lorentz generators, for classical purposes), where for the open string we can replace

$$\sum_{\pm}\int_0^{\pi}\frac{d\sigma}{2\pi} \to \int_0^{2\pi}\frac{d\sigma}{2\pi}, \qquad Y_{(\pm)} \to Y$$

For the lightcone gauge we then have the gauge condition to determine X^+ and Virasoro constraints to determine X^-:

$$Y^+_{(\pm)} = 0, \qquad \kappa\alpha' p^- + \sqrt{\tfrac{1}{2}\alpha'}\,\dot{Y}^-_{(\pm)} = \frac{1}{2\kappa\alpha' p^+}\left(\kappa\alpha' p^i + \sqrt{\tfrac{1}{2}\alpha'}\,\dot{Y}^i_{(\pm)}\right)^2$$

Exercise XIA3.2

Consider gauge fixing in the temporal gauge, replacing X^+ with X^0. The classical interpretation is now simpler, since τ and X^0 can now be identified with the usual time. Everything is similar except that the Virasoro constraints can't be solved (e.g., for X^1) in general without square roots.

a Show that some 3D solutions (2 space, 1 time) for the open string are given by

$$\tfrac{1}{\sqrt{2}}(\dot{Y}^1 - i\dot{Y}^2)(\tau) = ce^{-in\tau},$$

for nonzero integer n. (Without loss of generality, we can choose c real and positive.) Find the mass (energy) and spin as

$$M = \frac{c}{\sqrt{\alpha'}}, \qquad S^{12} = \frac{c^2}{n} = \frac{\alpha'}{n} M^2$$

Find X explicitly, and show it describes an "n-fold spinning rod".

b Show that the above solution can be generalized to closed strings by using two such Y's, and fixing the relative magnitude of the two c's. Consider the special cases where $n_- = \pm n_+$. Find the explicit masses, spins, and X's, and show that one describes another n-fold spinning rod, while the other is an "n-fold oscillating ring".

4. Quantum mechanics

The more interesting features of the string don't appear until quantization. In particular, we can already see at the free level the discrete mass spectrum characteristic of Regge theory, or of bound states in general.

Canonical quantization is simplest in the lightcone gauge. As for particles, canonical quantization is convenient only in mechanics (first quantization), not field theory (second quantization). As can be seen from the lightcone action, the Hamiltonian is part of the constraints: For the spinless particle, we had only the constraint $p^2 + m^2 = 0$, which became $E = H$ in the lightcone gauge $X^+ = p^+\tau$ after identifying the lightcone "energy" $E = p^+p^-$ and its Hamiltonian $H = \frac{1}{2}(p_i^2 + m^2)$. (See subsection IIIB2.) The string Hamiltonian can be rewritten conveniently in terms of $\hat{P}$. Since the closed string is effectively just a doubling of the open string, we treat the open string first. The Hamiltonian is simply

$$H = \int_{-\pi}^{\pi} \frac{d\sigma}{2\pi} \tfrac{1}{2} \hat{P}_i^2$$

where $\hat{P} = \hat{P}_{(+)}$. Since we have chosen $X^+ = 2\alpha' p^+ \tau$, we have $E = 2\alpha' p^+ p^-$.

To identify the individual particle states, we Fourier expand the worldsheet variables in σ. As for the particle, we can work at $\tau = 0$, since all the dynamics is contained in the constraints. Equivalently, from the nonrelativistic view of the lightcone formalism, we can work in the Schrödinger picture where the τ dependence is in the wave function instead of the operators. We expand as

$$\hat{P}(\sigma) = \sum_{n=-\infty}^{\infty} \tilde{a}_n e^{-in\sigma}, \qquad \tilde{a}_0 = \sqrt{2\alpha'} p, \qquad \tilde{a}_{-n} = \tilde{a}_n{}^\dagger$$

From the canonical commutation relations for P and X,

$$[\hat{P}_i(\sigma_1), \hat{P}_j(\sigma_2)] = -2\pi i \delta'(\sigma_2 - \sigma_1)\delta_{ij}$$

we have

$$[\tilde{a}_{im}, \tilde{a}_{jn}] = m\delta_{m+n,0}\delta_{ij}$$

as well as the usual $[p_i, x_j] = -i\delta_{ij}$, and thus can relate the modes to the usual harmonic oscillator creation and annihilation operators:

$$\tilde{a}_n = \sqrt{n} a_n, \qquad \tilde{a}_{-n} = \sqrt{n} a_n{}^\dagger \quad \Rightarrow \quad [a_m, a_n{}^\dagger] = \delta_{mn}$$

for positive n. After normal ordering, we find for the Hamiltonian

$$H = \alpha' p_i^2 + N - \alpha_0, \qquad N = \sum_{n=1}^{\infty} n a_{in}{}^\dagger a_{in}$$

$$\Rightarrow \quad H - E = \alpha'(p_a^2 + M^2), \qquad M^2 = \alpha'^{-1}(N - \alpha_0)$$

for some constant α_0.

From the expression for the mass in terms of the number operator N, we see that the nth oscillator $a_{in}{}^\dagger$ raises the mass-squared of the ground state $|0\rangle$ by n (and similarly for multiple applications of these oscillators). For any given mass, the highest-spin state is the symmetric, traceless tensor part of multiple $a_{i1}{}^\dagger$'s acting on $|0\rangle$: This describes the leading Regge trajectory, with spins

$$j = \alpha' M^2 + \alpha_0$$

Let's look first at the first excited level, obtained by acting on the scalar ground state $|0\rangle$ with the lowest-mass oscillators $a_{i1}{}^\dagger$. Clearly this describes a (lightcone) transverse vector, with no Stückelberg scalar for describing a massive vector. (I.e., it has only D−2 components, not the D−1 necessary for a massive vector.) Thus this state describes a massless vector, so

$$\alpha_0 = 1$$

The ground state is then a scalar tachyon with $M^2 = -\alpha'^{-1}$. For any given level past the first excited level, one can check explicitly that the states coming from the various oscillators include the necessary Stückelberg fields. For example, at the second excited level, $a_{i1}{}^\dagger a_{j1}{}^\dagger$ contains a traceless, symmetric tensor and a scalar (coming from the trace), while $a_{i2}{}^\dagger$ is a vector; they combine to describe a massive tensor. The proof that this works to all mass levels is closure of the Poincaré algebra quantum mechanically: The only nontrivial commutator is $[J^{-i}, J^{-j}] = 0$, since only J^{-i} is higher than quadratic (cubic) in oscillators (from the form of X^- and P^- after solving constraints), so normal-ordering ambiguities lead to more than just constant terms. For reasons to be explained in chapter XII, the algebra(ic computations) in calculating this commutator is the same in any first-quantized theory as the first-quantized BRST algebra. Of course, the proof of closure is already the same in principle because both algebras are a consequence of the constraints, the conformal algebra. Thus, any anomaly must show up in the conformal algebra itself, which will be considered in the following subsection.

Exercise XIA4.1

Check the third excited level massive representations.

The closed string works similarly to the open, but with two sets of harmonic oscillators, but with

$$p_{(+)} = p_{(-)} = \tfrac{1}{2}p$$

In that case we find

$$M^2 = 2\alpha'^{-1}(N_{(+)} + N_{(-)} - 2)$$

where $N_{(+)}$ and $N_{(-)}$ are the number operators for the two independent sets of oscillators. In the lightcone gauge the closed string has the residual gauge invariance generated by $\int d\sigma\ X' \cdot \delta/\delta X$; this gives the residual constraint

$$N_{(+)} = N_{(-)}$$

The closed-string states are thus the direct product of two open-string states of the same mass: For example, the ground state is a scalar tachyon with $M^2 = -2\alpha'^{-1}$, while the first excited states are massless ones from the product of two vectors — a scalar, an antisymmetric tensor, and a symmetric, traceless tensor. The leading Regge trajectory consists of states created with equal numbers of ${a_{i1(+)}}^\dagger$'s and ${a_{i1(-)}}^\dagger$'s, with

$$j = \tfrac{1}{2}\alpha' M^2 + 2$$

In summary, the leading trajectory for open or closed string is given by

$$\boxed{j = \kappa\alpha' M^2 + \frac{1}{\kappa}}$$

Covariant quantization of the string can be performed in several ways: One is to use the OSp methods of chapter XII, as applied to the Lorentz generators derived from the lightcone analysis. Another is to use the usual BRST of subsection VIA, as applied to gravity in subsection IXB1, treating the mechanics of the string as a 2D field theory. For the case of the conformal gauge, introducing ghosts C_m corresponding to the gauge parameters, and antighosts B^{mn} paired with the Lagrange multipliers of the gauge conditions (Nakanishi-Lautrup fields), we find the ghost action

$$\sqrt{-g}g_{mn} = \eta_{mn}, \qquad \delta\sqrt{-g}g_{mn} = \nabla_{(m}\lambda_{n)} - g_{mn}g^{pq}\nabla_p\lambda_q$$

$$\Rightarrow \quad L_g = B_{++}\nabla_- C_- + B_{--}\nabla_+ C_+$$

where in the last step we have introduced a background "zweibein" for applications such as the background field gauge, or geometries that do not admit the conformal gauge globally, and flattened the indices on the ghosts so the tracelessness of B (which follows from that of $\delta\sqrt{-g}g_{mn}$) can be solved explicitly. (The conformal-gauge Weyl and Lorentz gauge conditions do not involve derivatives in the transformation laws of the gauge-fixing conditions, so the ghosts are just algebraic.)

5. Anomaly

Closure of the BRST algebra, or lightcone Poincaré algebra, is nontrivial because of the infinite summations over oscillators, or equivalently because of integration over the two-dimensional "momentum" (which is quantized as mode number in the σ direction because of the finite extent of σ). As usual, BRST invariance is equivalent to gauge invariance, and we can check for anomalies in the usual way, now applied to the 2D "field theory" corresponding to the mechanics of the string. Because of conformal invariance the boundaries of the worldsheet are irrelevant; we can conformally transform them to infinity; then the anomaly calculations are similar to those applied to the Schwinger model in subsection VIIIA8. (In particular, see exercise VIIIA8.1.) The gauge invariances in question are coordinate invariance and local scale invariance, whose preservation is the vanishing of the divergence and trace of the energy-momentum tensor. As seen from our analysis for the Schwinger model, this implies that the quantum corrections to the energy-momentum tensor must themselves vanish when the external fields are restricted to gravity only. The calculation again involves one-loop propagator corrections; performed in position space, we get the product of two propagators between the same two points, with various numbers of derivatives acting on either end of either propagator.

The classical energy-momentum tensors can be found by various methods: A straightforward way is to use the methods of chapter IX. A simpler method is to note that the local Lorentz group in D=2 is Abelian (SO(1,1)=GL(1)), and the coupling of the Lorentz connection is proportional to the "spin" (number of + minus − indices, as used in the 2D spinor notation of subsections VIIIA7-8). This immediately tells us the coupling of the "zweibein" from the action for terms linear in derivatives:

$$iL = \psi_w \nabla_- \psi_{1-w} \quad \Rightarrow \quad -iT_{++} = \tfrac{1}{2}\psi_w \overleftrightarrow{\partial}_+ \psi_{1-w} + (w - \tfrac{1}{2})\partial_+(\psi_w \psi_{1-w})$$

where ψ_w is any field with w the number of + vector indices minus − (half the number of spinor indices). We have written the most general terms with one derivative. (It is ∂_+ to get the right Lorentz weight for T_{++}, but also because ∂_- vanishes by the free field equation.) The first term is the "zweibein" term, the second is the Lorentz connection term; the coefficient of the second term is fixed by considering the cases $w = 0$ (scalar, so all derivatives act on it) and $1 - w = 0$. (As a check, the second term vanishes by antisymmetry for $w = 1 - w = 1/2$.) For convenience we rewrite this result as

$$-iT_{++} = w\psi_w \partial_+ \psi_{1-w} + (w - 1)(\partial_+ \psi_w)\psi_{1-w}$$

The same result holds for T_{--} from a ∇_+ action. This includes the classical string action

$$P_+ \cdot P_- + P_+ \cdot \nabla_- X + P_- \cdot \nabla_+ X$$

as well as the conformal-gauge ghost action

$$B_{++}\nabla_- C_- + B_{--}\nabla_+ C_+$$

(The factor of $\mathbf{e}^{-1}$ can be absorbed into the fields by a local scale transformation; it's irrelevant for defining $T_{\pm\pm}$.)

The calculation for the matrix element for two T_{++}'s (i.e., for two h_{--} background fields) is now simple: The propagators go as $1/(x - x')$, and the various directions of derivatives from the vertices give either both hitting the same propagator or one hitting each, giving either $(\partial x^{-1})^2 = x^{-4}$ or $x^{-1}\partial^2 x^{-1} = 2x^{-4}$. The matrix element is then proportional to

$$[w^2 + (w - 1)^2] + 2[2w(w - 1)] = 6(w - \tfrac{1}{2})^2 - \tfrac{1}{2}$$

with the usual extra minus sign in the fermionic case. (As a check, this result is consistent with bosonization: A real, bosonic scalar gives the same result as a complex, fermionic spinor.) Taking (spacetime dimension) D $w = 0$ X's and −2 (fermionic) $w = 2$ ghosts (actually one with $w = 2$ and the conjugate with $w = -1$), this adds up to

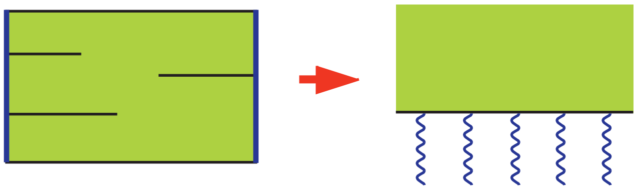
$$D - 26 = 0 \quad \Rightarrow \quad D = 26$$

for anomaly cancelation. Thus, we have two conditions on this string that make it unsuitable for describing mesons: unphysical intercept α_0 for the leading Regge trajectory (massless particles) and unphysical spacetime dimension D.

Exercise XIA5.1

In the derivation of the anomaly for X used for anomaly cancelation, the effect of the P^2 term was neglected. Using the result for the $h_{--}T_{++}$ term from $P_+ \cdot \nabla_- X$, eliminate $P_\pm$ from the action, and redo the anomaly calculation in terms of this form of S_L. (Hint: Show each step in the above calculation is still correct, even though P and X aren't independent.)

6. Tree amplitudes



Unlike particles, Feynman diagrams for strings can be treated by first-quantized methods for arbitrary loops. The basic idea is that interacting strings are just strings with nontrivial geometries: For example, while an open-string propagator can be described by a rectangle, an open-string tree graph can be described by a rectangle that has parallel slits cut from two opposite ends of the rectangle part-way into the interior; this describes initial strings that join and split at their ends (interactions). This is the lightcone picture of interactions, where conservation of total p^+ means conservation of the sum of the lengths of the strings. This corresponds to the choice $k = 1$ in the language of subsection XIA3, since the worldsheet coordinates must be chosen consistently over the whole worldsheet:

$$X^+ = \tau \quad \Rightarrow \quad l = 2\pi\alpha' p^+$$

More general conformal gauges are defined by conformal transformations of this configuration: For example, the boundary of this slit rectangle can be transformed to a single straight line by the usual methods of complex analysis, so the worldsheet becomes simply a half-plane. (For the infinite rectangle, relevant for asymptotic states, the transformation is $\rho = \sum p_r^+ ln(z - Z_r)$.) Then even the geometry is irrelevant; all that matters is the topology, which tells how many loops the diagram has. (See the discussion of subsection VIIC4.)

First-quantized path integrals are then the easiest way to calculate arbitrary S-matrix elements in string theory. However, the calculations still can be quite complicated (as expected from a theory with an infinite number of one-particle states), so here we will consider just the tree-level scattering of ground states, which is sufficient to illustrate the qualitative features. We can start from a gauge where the string is an infinite strip (a rectangle of infinite length but finite width), with all but two of the external states associated with points on one side of the strip, the remaining two states being at the ends at infinity. This is equivalent to a picture of a propagator in an external field, with all but two of the external states associated with the external field. Similar calculations are possible for particles, but give only a single graph; for strings this gives the only graph, since different cyclic orderings are related by conformal transformations. (We are restricted to cyclic orderings by group theory, as for the 1/N expansion for particles.) This method can be used in either the lightcone gauge or Lorentz covariant conformal gauge. In the latter case, we integrate over the relative position of the vertices on the end of the string; this is equivalent to integrating $\int_0^\infty d\tau\ e^{-\tau H} = 1/H$ for the propagators between vertices. The result is the usual particle type expression $VH^{-1}V...H^{-1}V$. The ground-state vertex of external momentum k is simply $e^{ik\cdot X(T,0)}$, where the vertex is positioned at the end of the string at $\sigma = 0$.

Since the functional integral is Gaussian, all we need to evaluate it is the two-dimensional propagator, subject to the boundary condition $X' = 0$. As for 2D electrostatic problems, we simplify the problem by a conformal transformation from the infinite strip to the half-plane:

$$z = e^{-\rho}$$

where $\rho = \tau - i\sigma$ is the complex (Wick-rotated worldsheet) coordinate for the strip. The bottom of the strip ($\sigma = 0$, the real axis) is mapped to the positive real axis, while the top of the strip ($\sigma = \pi$) is mapped to the negative real axis; the interior is mapped to the upper-half complex plane. The 2D propagator (in worldsheet coordinate space, needed for the explicit $\tau = T_r$ dependence of the vertices) now follows from that for the whole plane by the method of images:

$$\Box G(z,z') = \pi\delta^2(z-z') \quad \Rightarrow \quad G(z,z') = \tfrac{1}{2}(ln|z-z'| + ln|z-z'^*|)$$

(Note that we have changed normalization from our earlier discussions of general dimensions.)

Thus, the N-point amplitude is given by the path integral

$$\mathcal{A}_N = g^{N-2} \lim_{\substack{z_1\to\infty \\ z_N\to 0}} \prod_{r=3}^{N-1} \int_{z_{r+1}}^{z_{r-1}} dz_r z_1^2 \int DX\ e^{-S}$$

$$S = -i\sum_{r=1}^{N} k_r \cdot X(z_r) + \int \frac{d^2\sigma}{\pi}\ \tfrac{1}{2}(\partial X)^2$$

where we have chosen units $2\alpha' = 1$. (So $H = \frac{1}{2}p^2 + N - 1$.) Since the z_r are the values of z for the vertices, we integrate over z_r from 0 to 1, but

$$z_r > z_{r+1}, \quad z_1 = \infty, \quad z_2 = 1, \quad z_N = 0$$

While only $z_3, ..., z_{N-1}$ are integrated over, and vertices are inserted originally at only $z_2, ..., z_{N-1}$, we have inserted two extra vertices at the ends of the string at z_1 and z_N (i.e., $\tau = \pm\infty$), to excite the true ground states from $k = 0$ to nonzero momentum (as in ordinary quantum mechanics, where $e^{ikx}|0\rangle = |k\rangle$ for $P|k\rangle = k|k\rangle$.) The extra factor of z_1^2 was inserted to amputate the external lines of these two vertices:

$$|k\rangle = -ze^{ik\cdot X(z)}z^{-(\frac{1}{2}p^2+N-1)}|0\rangle = -e^{ik\cdot X(z)}z^2|0\rangle$$

$$\langle k| = -\langle 0|z^{\frac{1}{2}p^2+N-1}e^{ik\cdot X(z)}z = -\langle 0|e^{ik\cdot X(z)}$$

The extra factors of $-z$ come from the conformal transformation of this "vertex operator": Conformal invariance requires that $d\rho\ V(\rho) = dz\ V(z)$, and it is actually $V(\rho) = -zV(z)$ that creates an appropriately normalized state acting on $|0\rangle$.

As usual, we perform the functional integral by completing the square,

$$-S = \int \tfrac{1}{2}X\Box X + iJX \to \int \tfrac{1}{2}J\Box^{-1}J = \tfrac{1}{2}\int\int JGJ$$

Here J is a density in terms of k_r and a δ function; we can thus use its integration to define it (rather then bothering with its explicit form in terms of δ functions in ρ or z):

$$\int J\cdot X = \sum_r k_r \cdot X(z_r) \quad \Rightarrow \quad \int\int J\cdot GJ = \sum_{r,s} k_r\cdot k_s G(z_r, z_s)$$

(using also the fact that X and G are scalars). We drop the infinite terms coming from connecting a vertex to itself with a propagator: This is a "renormalization" of the vertex, called "normal ordering" in the operator approach. (It can also be treated in a more careful way by taking the vertices to correspond to finite-width strings, as they would in the lightcone approach, and taking the limit where their widths vanish.) The result of the functional integral is then

$$\mathcal{A}_N = g^{N-2} \lim_{\substack{z_1\to\infty \\ z_N\to 0}} \prod_{r=3}^{N-1}\int_{z_{r+1}}^{z_{r-1}} dz_r z_1^2 \prod_{1\le r<s\le N}(z_r - z_s)^{k_r\cdot k_s}$$

$$= g^{N-2}\prod_{r=3}^{N-1}\int_{z_{r+1}}^{z_{r-1}} dz_r \prod_{2\le r<s\le N}(z_r - z_s)^{k_r\cdot k_s}$$

where we have used the ground-state mass-shell condition $k^2 = 2$ for k_1.

The simplest case is the four-point function,

$$\mathcal{A}_4 = g^2 \int_0^1 dz\ z^{-\alpha(s)-1}(1-z)^{-\alpha(t)-1}$$

$$\alpha(s) = \tfrac{1}{2}s + 1, \quad s = -(k_3 + k_4)^2, \quad t = -(k_2 + k_3)^2$$

which we recognize as the Beta function (see subsection VIIA2)

$$\mathcal{A}_4 = g^2 B[-\alpha(s), -\alpha(t)] = g^2 \frac{\Gamma[-\alpha(s)]\Gamma[-\alpha(t)]}{\Gamma[-\alpha(s)-\alpha(t)]}$$

Duality of this amplitude follows from the fact that it can be written as a sum of poles in either the s or t channel (see exercise XIA6.1 below):

$$\mathcal{A}_4 = \sum_{j=0}^{\infty} \left\{ \frac{[\alpha(s)+j][\alpha(s)+j-1]\cdots[\alpha(s)+1]}{j!} \right\} \frac{1}{j-\alpha(t)}$$

where the pole at $j = \alpha(t)$ has maximum spin j, as indicated by the residue of order s^j. Regge behavior is also found in the appropriate limit (see exercise XIA6.2 below):

$$\lim_{\substack{s\to-\infty \\ t\ fixed}} \mathcal{A}_4 = g^2 \Gamma[-\alpha(t)][-\alpha(s)]^{\alpha(t)}$$

Exercise XIA6.1

Derive the pole structure of the 4-point string amplitude by Taylor expanding the integrand of the Beta function in z.

Exercise XIA6.2

Use the Stirling approximation (see exercise VIIC2.1) to derive the Regge limit of the 4-point string amplitude. Show that the same result can be obtained directly from the integral (Beta function) representation of the amplitude, where the main contribution comes from z near 1.

Exercise XIA6.3

Show the Regge limit can be obtained from the Sommerfeld-Watson transform of subsection XIA1. (Hint: The Beta function is a sum of Regge trajectories.)

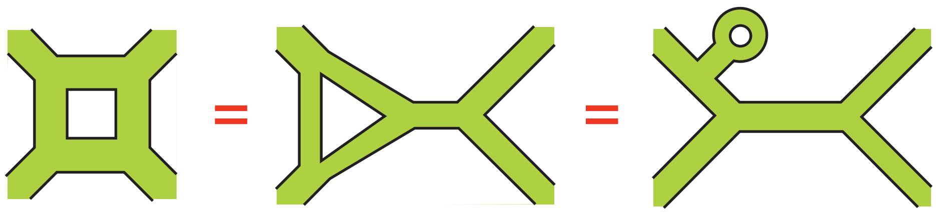
On the other hand, in the corresponding limit with fixed angle (i.e., fixed t/s; again using the Stirling approximation),

$$\lim_{\substack{s\to-\infty \\ \theta\ fixed}} \mathcal{A}_4 \sim e^{-f(cos\ \theta)\alpha(s)}$$

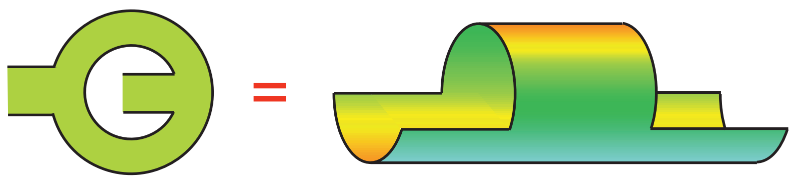
$$cos\ \theta \approx 1 + 2\frac{t}{s}, \qquad f \approx \frac{t}{-s} ln\left(\frac{-s}{t}\right) + \frac{u}{-s} ln\left(\frac{-s}{u}\right)$$

This limit corresponds to that treated in perturbative QCD, with large transverse energies. However, this Gaussian behavior in the momentum differs from the power-law behavior (with logarithmic corrections) found in QCD (see subsection XIC3 below).

Similar methods can be used for calculating closed string diagrams: The vertices are then located anywhere on the worldshseet instead of just on the boundary, so there are integrals over both z and $\bar{z}$. Since the closed-string Hilbert space is the direct product of two open-string Hilbert spaces (except for momentum), for left- and right-handed modes, the z-$\bar{z}$ integrands are products of two open-string integrands. There are also generalizations to strings with worldsheet fermions; the main differences are supersymmetry and D=10 (instead of 26).



Because of the local scale invariance of string theory, different string graphs are distinguished by their topology rather than their geometry: Any two graphs that can be "stretched" into one another are equivalent. In particular, in any loop graph any hole or window can be pulled out so that it appears as a tadpole. The result is that any graph is equivalent to a tree graph with insertions of some one-loop open- or closed-string tadpoles. However, this does not mean that any graph constructed with only open-string propagators and interactions can be expressed as an open-string tree graph with tadpole insertions: The one-loop open-string graph with two "half-twists" on the open-string propagators in the loop is equivalent to a tree graph with a closed-string intermediate state, as can be seen by stretching the surface, or by tracing the routes of the boundaries. (For example, drawing this graph in a psuedo-planar way, as a flat ring with external states connected to both the inner and outer edges, pulling the inner edge out of the plane reveals a closed string connecting the two edges.) This phenomenon is similar to 2D bosonization: A closed string can be represented as the "bound state" of two *free* open strings just as a massless scalar in D=2 can be represented as the bound state of two free massless spinors.



REFERENCES

1 T. Regge, *Nuo. Cim.* **14** (1959) 951, **18** (1960) 947.

2 R. Dolen, D. Horn, and C. Schmid, *Phys. Rev. Lett.* **19** (1967) 402:
duality.

3 G. Veneziano, *Nuo. Cim.* **57A** (1968) 190;
M. Suzuki, unpublished:
birth of string theory (as dual models).

4 K. Bardakçi and H. Ruegg, *Phys. Rev.* **181** (1969) 1884;
C.J. Goebel and B. Sakita, *Phys. Rev. Lett.* **22** (1969) 257;
Chan H.-M. and T.S. Tsun, *Phys. Lett.* **28B** (1969) 485;
Z. Koba and H.B. Nielsen, *Nucl. Phys.* **B10** (1969) 633, **12** (1969) 517:
generalization of 4-point amplitude.

5 M. Virasoro, *Phys. Rev.* **D1** (1970) 2933;
I.M. Gelfand and D.B. Fuchs, *Functs. Anal. Prilozhen* **2** (1968) 92:
Virasoro constraints.

6 D.J. Gross, A. Neveu, J. Scherk, and J.H. Schwarz, *Phys. Lett.* **31B** (1970) 592;
C. Lovelace, *Phys. Lett.* **34B** (1971) 500;
E. Cremmer and J. Scherk, *Nucl. Phys.* **B50** (1972) 222:
string loops as 1-loop tadpole insertions.

7 Y. Nambu, Quark model and the factorization of the Veneziano amplitude, in *Proc. of International conference on Symmetries and quark model*, ed. R. Chaud, Wayne State U., June, 1969 (Gordon and Breach, 1970) p. 269;
L. Susskind, *Phys. Rev.* **D1** (1970) 1182, *Nuo. Cim.* **69A** (1970) 457;
H.B. Nielsen, An almost physical interpretation of the integrand of the n-point Veneziano model, 15th international conference on high energy physics, Kiev, 1970:
dual model as string, through mechanics.

8 Y. Nambu, lectures at Copenhagen Symposium, 1970, unpublished;
O. Hara, *Prog. Theo. Phys.* **46** (1971) 1549;
T. Goto, *Prog. Theo. Phys.* **46** (1971) 1560;
H. Noskowitz, unpublished:
area action for string.

9 P. Goddard, J. Goldstone, C. Rebbi, and C.B. Thorn, *Nucl. Phys.* **B56** (1973) 109:
lightcone quantization for string.

10 P.A. Collins and R.W. Tucker, *Phys. Lett.* **64B** (1976) 207;
L. Brink, P. Di Vecchia, and P. Howe, *Phys. Lett.* **65B** (1976) 471;
S. Deser and B. Zumino, *Phys. Lett.* **65B** (1976) 369:
mechanics of string with worldsheet metric.

11 A.M. Polyakov, *Phys. Lett.* **103B** (1981) 207:
worldsheet ghosts.

12 C. Lovelace, *Phys. Lett.* **34B** (1971) 500:
discovery of D=26 requirement.

13 D. Friedan, Introduction to Polyakov's string theory, in *Recent advances in field theory and statistical mechanics*, eds. J.B. Zuber and R. Stora, proc. 1982 Les Houches summer school (Elsevier, 1984) p. 839:
covariant vertex operators with ghosts.

14 S. Mandelstam, *Nucl. Phys.* **B64** (1973) 205:
lightcone path integrals for interacting strings.

15 O. Alvarez, *Nucl. Phys.* **B216** (1983) 125:
covariant path integrals for interacting strings.

16 M.B. Green, J.H. Schwarz, and E. Witten, *Superstring theory*, 2 v. (Cambridge University, 1987);
J. Polchinski, *String theory*, 2 v. (Cambridge University, 1998):
comprehensive texts on strings.

17 P.H. Frampton, *Dual resonance models and string theories* (World Scientific, 1986);
Siegel, *loc. cit.*;
M. Kaku, *Introduction to superstrings and M-theory*, 2nd ed. (Springer-Verlag, 1999):
other string texts.

18 S. Fubini; G. Veneziano; V. Alessandrini, D. Amati, M. Le Bellac, and D. Olive; J.H. Schwarz; C. Rebbi; S. Mandelstam; *Dual theory*, ed. M. Jacob (North-Holland, 1974);
J. Scherk, *Rev. Mod. Phys.* **47** (1975) 123:
old string reviews.

B. SYMMETRIES

In the previous section we looked in detail at the simplest of the string models. In this section we examine some of the general properties of string theory, shared by all known models, but expected to apply also to more realistic strings. These features can be used for phenomenological applications of string theory, but also may help point to new generalizations.

When string theory is used as a unified theory of gravity and other forces, the most interesting predictions are those for the "low-energy" (with respect to the Planck mass) part of the theory. Although the possible low-energy limits of known string theories have not all been explored, the indications are that there are only a few restrictions beyond the usual field theoretic ones:

(1) The dilaton counts the number of loops.

(2) The spectrum of the closed string is given by the direct product of two open strings.

(3) String theory has noncompact symmetries, "S-duality" and "T-duality", resulting from the amplitudes also having this direct-product structure.

All these properties survive the low-energy limit. In the supersymmetric case, the last property follows from the first two. (However, the D=10 superstring is actually a D=11 supermembrane nonperturbatively. Since the observed features of hadrons, as well as qualitative arguments from QCD, indicate stringy but not membrane-like behavior, we will abstract only the perturbative features of higher-dimensional strings.)

1. Massless spectrum

The known string models all have massless particles. A string model with massless particles can be applied to hadrons only if masses are given to all these states through the Higgs mechanism or some other change in the vacuum. An alternative is to use such a model to describe fundamental massless particles (graviton, photon, gluons, neutrinos), although this would also require the usual Higgs of the Standard Model for generating masses for some particles (W, Z, quarks, charged leptons, Higgs). In particular, all known string models have a graviton, and there is no known method whereby this graviton would gain mass, so these models seem suited only for unified theories of gravity plus matter. For this purpose, the massive fields have little phenomenological interest. They might improve high-energy behavior, but only near the

Planck scale, which is effectively unobservable. Therefore, it is necessary to analyze the massless subsector of such string theories to find signs of fundamental strings in nature.

The low-energy limits of the known string models follow simply from gauge invariance once the spectrum is known, and the spectrum itself follows from simple considerations. In particular, we saw in subsection XIA4 for the simplest model that the massless spectrum of the closed string was simply the direct product of that of two open strings. This is true for all known models, and is expected to be a general feature: It is a consequence of the fact that all free massless fields in D=2 propagate to either the left or right; the left-handed modes are associated with one open-string factor, the right-handed with the other. In general, the two factors may be from different open-string theories.

The massless sector of the open string includes spin 1 and no higher. This is true for the known string models, and also is expected to be a general result, since otherwise the closed string would include massless states with spin higher than 2, for which no consistent interacting theory is known. Spin 1/2 leads to supersymmetry, as described in subsection XIB3 below; for now we limit ourselves to bosonic strings.

For the spin-1 particles to obtain nontrivial self-interactions (Yang-Mills), we need to introduce internal symmetry. This is naturally introduced by associating indices with the ends of the open string (as if they were quarks), as applied to particle field theory in subsection VC9 (but which originated in string theory). These "Chan-Paton factors" can also be associated with worldsheet variables that live only on string boundaries. As in the field theory case, these indices are associated with orientation of the boundaries (arrows) only for U groups, not for SO or USp. (Exceptional groups cannot be described by Chan-Paton factors.)

The bosonic string contains at least the graviton, a scalar (usually going by the misnomer of "dilaton"), and a pseudoscalar (the "axion", described by an antisymmetric-tensor gauge field), and can contain additional vectors and scalars (if the open strings had scalars in addition to the vector). This analysis can be performed covariantly, but it is simpler to use a helicity or lightcone analysis. Then the helicities of the closed-string states are just the sums of those of the open strings: For the product of two vectors (the minimal case),

$$(+1 \oplus -1) \otimes (+1 \oplus -1) = +2 \oplus 0 \oplus 0 \oplus -2$$

giving the graviton, scalar, and axion. Similarly, additional scalars for one open string give additional vectors for the closed, while additional scalars for both open strings give also additional scalars.

Exercise XIB1.1

Make the same analysis in terms of covariant fields, both for the fields themselves *and* their gauge transformations. Note that the trace of the gravitational field h_{ab} (determinant of the metric g_{mn}) is missing. (It's unphysical, and can be found from the ghost sector, as explained in chapter XII.)

2. Reality and orientation

In the case that the two open strings are the same, it is possible to restrict this direct product to its symmetric part. This eliminates the axion, but not the scalar. Thus, a massless scalar appears even in the simplest case.

To understand this restriction better, we examine the discrete symmetries of the worldsheet. As in D=4, local, unitary, Poincaré invariant 2D field theories are always CPT invariant. (In particular, CPT doesn't switch left- and right-handed modes, which differ in some string theories.) Thus, we can always impose invariance of the wave function/field under the worldsheet CPT transformation

$$worldsheet \quad CPT: \qquad \Phi[X(\sigma)] = \Phi^\dagger[X(\pi-\sigma)]$$

where parity transforms $\sigma \to \pi - \sigma$ to preserve $\sigma\epsilon[0,\pi]$ for the open string; for the closed string the π is irrelevant because of periodicity and invariance under σ translation. (As for the particle, we drop τ dependence. We have written only the X coordinate explicitly for simplicity; similar remarks apply to other coordinates, such as ghosts, with possible extra signs due to 2D Lorentz indices.) Hermitian conjugation for the open string (for the closed string the field is not a matrix), instead of just complex conjugation (for C), simply switches the internal symmetry (Chan-Paton) factors associated with the left and right ends of the open string (matrix transposition), as also required by parity. In particular, this implies that the matrices associated with the Yang-Mills fields are hermitian, so the Yang-Mills group is unitary.

In addition to this reality condition, if the 2D theory is also invariant under CP and T, it is also possible, though not necessarily required, to impose such a quantum mechanical invariance under CP, and thus T:

$$worldsheet \quad T: \qquad \Phi[X(\sigma)] = M\Phi^*[X(\sigma)]M^{-1} \quad \Leftrightarrow$$

$$worldsheet \quad CP: \qquad \Phi[X(\sigma)] = M\Phi^T[X(\pi-\sigma)]M^{-1}$$

where the matrix "M" is the Yang-Mills group metric for the open string (we drop the M and the ^{T} for the closed string, which is not a matrix), either symmetric or antisymmetric depending on whether the group is orthogonal or symplectic; without imposing T and CP the group is just unitary. (The T condition is the usual reality condition for the particle, as discussed in subsection IA5.) Thus, all the classical groups are allowed (at least in the classical field theory). Since imposing invariance of the states (not just the action) under CP and T makes it impossible to observe the left/right handedness of the worldsheet, such strings are “unoriented”, as opposed to the “oriented” strings that satisfy just the CPT condition. Thus, orientability of the surface is directly related to orientability of its boundaries (oriented for U, unoriented for SO or USp). Also, as in the particle field theory, unorientability allows “twisted” worldsheets that are prohibited in the oriented case (because we can distinguish the “front” of the worldsheet from the “back”): This allows such exotic geometries as Möbius strips and Klein bottles. Open strings produce closed ones as bound states (open and closed strings are parts of the same worldsheet with different boundaries); in theories of open and closed strings, they must be both oriented or both unoriented. Since the worldsheet-CP and -T switch lefty and righty modes, this invariance on the closed string results in the restriction introduced earlier, keeping only the symmetric part of the direct product.

Exercise XIB2.1

What is the difference between an unoriented closed string (satisfying this worldsheet CP condition) and the interpretation of subsection XIB3 of the open string as a closed string folded over on itself?

3. Supergravity

Again focusing on just the massless spectrum, we now look at the restrictions imposed by supersymmetry. The open string can also contain massless spin 1/2, but only if it is related by supersymmetry to its massless spin 1, since it leads to spin 3/2 in the closed string, and massless spin 3/2 is known to be inconsistent in an interacting theory unless related by supersymmetry to the graviton. (Spin 3/2 gauges supersymmetry. But spin 1 can’t couple minimally to spin 3/2: see exercise XIIB7.2b below. So, spin 3/2 needs spin 2 as its supersymmetric partner.) Thus there are two possibilities for the massless sector of each open string: (1) vectors and scalars for an open bosonic string, or (2) vector multiplets (vectors, spinors, and scalars, all related by some number of supersymmetries) for an open superstring. From our analysis of subsection IIC5, there are furthermore 3 types of vector multiplets in D=4, corresponding to N=1,2, or 4 supersymmetries.

This leads to four types of closed strings:

(1) The bosonic string, from bosonic ⊗ bosonic, was discussed in subsection XIB1.

(2) The "heterotic" string, comes from bosonic ⊗ super. It thus can have N=1,2, or 4 supersymmetries.

(3) The "(Type II) superstring" comes from super ⊗ super. The total number of its supersymmetries is the sum of those from the open strings: Depending on the type of supersymmetric open strings used, the superstring can have N=2,3,4,5,6, or 8 (in other words, anything greater than 1, since N=7 supersymmetry is equivalent to N=8, and 8 is the maximum for supergravity).

(4) In the super case, if the left and right open strings are the same, we can impose symmetry as in the bosonic case ("Type I"). Then we may also include the open strings in the spectrum: This symmetrization also identifies the left and right supersymmetries, so then N=1,2 or 4, the same for open and closed states (so they can be consistently coupled).

The spectrum again can be analyzed by helicity: For example, for the N=1 heterotic string, we have

$$(1 \oplus \tfrac{1}{2} \oplus -\tfrac{1}{2} \oplus -1) \otimes (1 \oplus -1) = (2 \oplus \tfrac{3}{2} \oplus -\tfrac{3}{2} \oplus -2) \oplus (\tfrac{1}{2} \oplus 0 \oplus 0 \oplus -\tfrac{1}{2})$$

which is supergravity plus a scalar multiplet. As for the bosonic string, all supersymmetric closed strings include the scalar, again coming from vector ⊗ vector.

Exercise XIB3.1

Make the same analysis for the superstring.

4. T-duality

Another symmetry of all known string models is "T-duality". It is closely related to the open ⊗ open structure of closed string states, and thus expected to be a general property of string theory. We consider the simple bosonic model as an example. Including constant background fields, working in the conformal gauge for convenience, the Lagrangian is

$$L = -(\partial_+ X^m)(\partial_- X^n) M_{mn}, \qquad M_{mn} = G_{mn} + B_{mn}$$

where the curved indices now refer to spacetime, G_{mn} is the spacetime metric, and B_{mn} is an antisymmetric tensor gauge field ("axion"). Writing the action in first-order form

$$L' = -P_{+m}\partial_- X^m - P_{-m}\partial_+ X^m + P_{+m}P_{-n}M^{mn}$$

where M^{mn} is the inverse of M_{mn}, we vary X instead of P to solve the field equation

$$\partial_+ P_{-m} + \partial_- P_{+m} = 0 \quad \Rightarrow \quad P_{+m} = \partial_+ \tilde{X}_m , \quad P_{-m} = -\partial_- \tilde{X}_m$$

and substitute to find the "dual" Lagrangian

$$L'' = -(\partial_+ \tilde{X}_m)(\partial_- \tilde{X}_n) M^{mn}$$

Thus the "duality transformation" from X to $\tilde{X}$ is an invariance of the theory, as long as we also transform the background:

$$X^m \to \tilde{X}_m , \qquad M_{mn} \to M^{mn}$$

Note that in flat space ($M^{mn} = \eta^{mn}$), using the P equation of motion in L', we have

$$P_{+m} = \eta_{mn} \partial_+ X^n , \qquad P_{-m} = \eta_{mn} \partial_- X^n$$

so duality just changes the sign of the right-handed modes ($\partial_- X = -\partial_- \tilde{X}$) while leaving invariant the left-handed ones ($\partial_+ X = \partial_+ \tilde{X}$). (The treatment of the zero-modes is more tricky: We have ignored them by taking the background constant.) We can see this to lowest order in the background, since $M \to M^{-1}$, to lowest order in perturbation about $\langle M \rangle = \eta$, changes the sign of the field, corresponding to the fact that their vertex operators are linear in both left- and right-handed modes ($(\partial_+ X)(\partial_- X)$). However, in full nonlinearity, duality mixes the spacetime metric $G_{mn}(X)$ and axion $B_{mn}(X)$.

This invariance can be generalized to a continuous O(D,D) symmetry by combining it with (global) Lorentz transformations. The above discrete symmetry is a kind of "parity" for this larger group: There are also "reflections" from performing the duality on just one component of X^m. The easiest way to see the full symmetry is in the Hamiltonian formalism, where it can be made manifest: We first combine X'^m and the canonical momentum P_m into an O(D,D) vector:

$$Z_M = (P_m , X'^m)$$

$$\Rightarrow \quad [Z_M(1), Z_N(2)] = -i\delta'(2-1)\eta_{MN} , \qquad \eta_{MN} = \begin{pmatrix} 0 & \delta_m^n \\ \delta_n^m & 0 \end{pmatrix}$$

where the O(D,D) metric η_{MN} is constant even in curved space. (We have abbreviated "1" for "σ_1", etc.) The Virasoro constraints are then

$$\tfrac{1}{2}\eta^{MN} Z_M Z_N = \tfrac{1}{2} M^{MN} Z_M Z_N = 0$$

where M is not only symmetric but also an element of the O(D,D) group:

$$M^{MN} = \begin{pmatrix} G^{mn} & G^{mp}B_{pn} \\ -B_{mp}G^{pn} & G_{mn} - B_{mp}G^{pq}B_{qn} \end{pmatrix} = M^{NM} = \eta^{MP}(M^{-1})_{PQ}\eta^{QN}$$

If the fields are constant in only d of the D dimensions, than the symmetry is reduced to O(d,d); thus O(d,d) is a symmetry of the dimensionally reduced theory with arbitrary fields.

Exercise XIB4.1

Show that the conditions on M can be solved in a manifestly O(D,D) covariant way by use of a "vielbein" $E_A{}^M$:

$$M = M^T, \quad M\eta M = \eta \quad (\eta = \eta^T = \eta^{-1})$$

$$\Rightarrow \quad M = E^T\tilde{\eta}E, \quad E\eta E^T = \eta, \quad \tilde{\eta} = \begin{pmatrix} \eta^{ab} & 0 \\ 0 & \eta^{a'b'} \end{pmatrix}$$

Show that M is invariant under a local O(D−1,1)⊗O(D−1,1) transformation on E, so E is an element of the coset space O(D,D)/O(D−1,1)⊗O(D−1,1) (see subsection IVA3).

5. Dilaton

We can extend the spectrum analysis off shell: The procedure (to be justified in chapter XII) includes the ghost and antighost (multiplets) for the vector (multiplet) as a doublet of the ghostly Sp(2) symmetry. The direct product of vector ⊗ vector now clearly gives a traceless symmetric tensor (graviton), the corresponding trace (physical scalar), and an antisymmetric tensor (axion). In the direct product of the ghosts, the Sp(2) singlet gives the trace part of the metric tensor, which is the true dilaton. This dilaton (the determinant of the metric tensor in the nonlinear case) is required in gravity for constructing local actions (see subsection IXA7), but does not contain a physical degree of freedom. The physical polarizations of the graviton are contained in the traceless (actually $det = -1$) part of the metric, which describes the conformal part of gravity. The direct products involving ghosts also give Sp(2) nonsinglets, which are the ghosts of the massless sector of the closed string. BRST transformations (and thus gauge transformations) can also be obtained by this direct-product procedure.

The natural coupling of background fields in the classical mechanics of the string reflects this direct-product structure, as seen in the previous subsection. This means that the background metric as we have defined it has as its determinant not the usual one, but that times a power of the physical scalar: It is a physical degree of freedom. T-duality mixes physical degrees of freedom with each other.

If we try to construct a low-energy action for the massless fields of the bosonic string, it is not too difficult to find a scalar invariant under T-duality to act as the Lagrangian. However, it is impossible to use the usual measure $\int dx\ \sqrt{-g}$ because g is not invariant under T-duality. This problem is solved by including the spacetime dilaton field $\Phi(X)$: It couples to the string as

$$S_{dil} = -\int \frac{d^2\sigma}{2\pi}\sqrt{-g}\ \tfrac{1}{2}r\ ln\ \Phi(X)$$

where we denote the worldsheet curvature by $r(\sigma)$ (only in this subsection) to distinguish it from the spacetime curvature $R(x)$. (There are also boundary contributions: see exercise IXA7.3.) This term can also be expressed as a coupling to the worldsheet ghosts (according to the above arguments), allowing the worldsheet metric to be completely fixed by gauge transformations, as usual.

Since there is no X dependence of S_{dil} for constant dilaton field (no ∂X factors, unlike G and B), the constant dilaton is invariant under T-duality. Furthermore, since it couples to the worldsheet curvature, which counts the number of loops, the dilaton must appear homogeneously in the classical action. The dilaton that appears as above in the string action transforms as a density under general coordinate transformations, allowing the construction of actions invariant under both T-duality and coordinate transformations. The resulting spacetime action is

$$S_{massless} = \int dx\ \Phi(\Box - \tfrac{1}{4}R + \tfrac{1}{24}H^{abc}H_{abc} + \Lambda)\Phi$$

where $H_{abc} = \frac{1}{2}\nabla_{[a}B_{bc]}$ is the field strength for the axion. T-duality determines the only arbitrary coefficient, the relative weight of the $\Box$ and R terms. Note the absence of the factor $\mathbf{e}^{-1}$, which has been absorbed into the definition of Φ: The covariant derivative acting on Φ, since it is a density that transforms as $\mathbf{e}^{-1/2}$, acts as $\nabla_a\Phi = \mathbf{e}^{-1/2}e_a\mathbf{e}^{1/2}\Phi$.

Exercise XIB5.1

Find the field equations following from this action. Then make the field redefinition $\Phi = \mathbf{e}^{-1/2}e^{\phi}$, to find the result:

$$\begin{aligned}
\frac{\delta}{\delta\Phi} &\quad\Rightarrow\quad (\nabla\phi)^2 + \Box\phi - \tfrac{1}{4}R + \tfrac{1}{24}H^2 + \Lambda = 0 \\
\frac{\delta}{\delta B_{ab}} &\quad\Rightarrow\quad \nabla^a H_{abc} + 2H_{abc}\nabla^a\phi = 0 \\
\frac{\delta}{\delta e_a{}^m} &\quad\Rightarrow\quad R_{ab} = 2\nabla_a\nabla_b\phi
\end{aligned}$$

Both coupling constants in string theory can be associated with vacuum values: (1) The string coupling appears as the vacuum value of the dilaton, since it counts loops. (2) α' comes from the vacuum value of the (spacetime) metric, as can be seen from the worldsheet action. This is the string-gauge equivalent of the fact that the gravitational constant naturally arises as the vacuum (or asymptotic) value of the metric in ordinary gravity (see subsection IXA5). In string theory, the fact that the gravitational constant is a combination of α' and the string coupling is equivalent to the field redefinition from the string gauge to the particular Weyl gauge where the Einstein term in the action appears in the usual way.

Exercise XIB5.2

This action is in the string gauge (see subsection IXB5).

a Make the physical scalar explicit in the action by the field redefinition (Weyl scaling: see subsection IXA7)

$$e_a{}^m \to \chi e_a{}^m$$

leaving Φ and B_{mn} unchanged.

b The resulting scalar action can be (off-)diagonalized by further redefinitions: Noting that the known string theories are defined for $\sqrt{D-1}$ an (odd) integer (5 or 3), write the dimension in general as (for any $D > 1$, n not necessarily integer)

$$D = n^2 + 1$$

Restoring the $\mathbf{e}^{-1}$ to the action, redefine

$$\Phi = \mathbf{e}^{-1/2}\phi_+^{(n-1)/2(n+1)}\phi_-^{(n+1)/2(n-1)}, \qquad \chi = \phi_+^{1/(n+1)}\phi_-^{-1/(n-1)}$$

which also gives the scalars $\phi_\pm$ the canonical Weyl scale weights, to obtain the final result for the Lagrangian L (where $S = \int dx\ \mathbf{e}^{-1}L$)

$$L = \phi_+(\tfrac{n^2}{n^2-1}\Box - \tfrac{1}{4}R)\phi_- + \tfrac{1}{24}\phi_+^{(n+5)/(n+1)}\phi_-^{(n-5)/(n-1)}H^2$$
$$+\Lambda\phi_+^{(n-1)/(n+1)}\phi_-^{(n+1)/(n-1)}$$

c This redefinition is singular for $D = 2$ ($n = 1$). Fix this by making the additional redefinition

$$\phi_\pm \to \phi_\pm^{(n\pm1)/2}$$

and then taking the limit $n \to 1$. (We can also use redefinitions equivalent in the limit, such as $\phi_- \to \phi_-^{(D-2)/4}$.) Show the result is then

$$L \to \tfrac{1}{4}\phi_+(\Box\ ln\ \phi_- - R) + \tfrac{1}{24}\phi_+^3\phi_-^{-2}H^2 + \Lambda\phi_-$$

We can no longer choose the gauge $\phi_+ = 1$, since it is now scale invariant, but we can still choose $\phi_- = 1$.

6. Superdilaton

When applied to the supersymmetric cases (superstring or heterotic string), the inclusion of ghosts in the direct-product procedure also gives the auxiliary fields. (The dilaton itself is an auxiliary field.) For example, in the heterotic case, the direct product of the physical parts of the vector and vector multiplet give conformal supergravity (the supersymmetrization of the traceless part of the metric) and a physical tensor multiplet (the supersymmetrization of the axion and scalar). On the other hand, the ghosts of the vector multiplet form a chiral scalar superfield; its procuct with the scalar ghost of the vector gives another chiral scalar superfield, the compensator, containing the dilaton. (See chapter X.)

The two conditions of supersymmetry and that the dilaton must appear homogeneously (quadratically after an appropriate field redefinition) are now enough to fix the form of the action (except for the nonminimal heterotic cases, where the open string's scalars introduce extra vector multiplets). For convenience we redefine the chiral scalar compensator as $\phi \to \phi^{2/3}$ so that it appears quadratically in the cosmological term $\int d^4x\ d^2\theta\ \phi^2$. Thus, by dimensional analysis ϕ now has scale weight $\frac{3}{2}$. The axial-vector field strength of the axion appears as $[\nabla_\alpha, \overline{\nabla}_{\dot\alpha}]G$, so G has scale weight 2. (Gauge fields are Weyl scale invariant with curved indices for consistency with gauge transformations; thus H_{mnp} has weight 0 while H_{abc} has weight 3.) Of course, these weights also follow from local superscale transformations, the global part of which transforms fields as L^{2w} (see subsection XA4). The only action quadratic in the dilaton consistent with global scale and U(1) (R) invariance is then (with implicit covariantization with respect to conformal supergravity, which makes these invariances local)

$$S = \int dx\ d^4\theta\ \bar\phi\phi G^{-1/2} + \left(\varLambda \int dx\ d^2\theta\ \phi^2 + h.c.\right)$$

Exercise XIB6.1

Use the methods of subsection XB6 to find all of the terms in this action involving only bosonic fields. Compare to the bosonic string action of the previous subsection.

Besides T-duality, string theories also have "S-duality" symmetries that are realized only on the field equations, or after performing electromagnetic-type duality transformations on the fields: If we convert G into a second, physical chiral multiplet χ by such a duality as described in subsection XB5, the above action is converted to

$$S = \int dx\ d^4\theta\ (\bar{\phi}\phi)^{2/3}(\chi+\bar{\chi})^{1/3} + \left(\Lambda \int dx\ d^2\theta\ \phi^2 + h.c.\right)$$

After the redefinitions

$$\phi^2 \to \phi, \qquad \phi^2\chi \to \chi$$

the first term becomes manifestly SU(1,1) invariant (see subsection XB7):

$$S = \int dx\ d^4\theta\ (\bar{\phi}\chi+\bar{\chi}\phi)^{1/3} + \left(\Lambda \int dx\ d^2\theta\ \phi + h.c.\right)$$

It is now the original T-duality that can be realized only on shell. Also, in this form the condition that the dilaton should appear homogeneously is obscured. (Such S-dualities were first seen in extended supergravity theories, especially when obtained by reduction from higher dimensions, where antisymmetric tensors are often required.)

Exercise XIB6.2

Apply the results of exercise XB5.1 to include vector multiplets in the above actions by replacing $G \to \tilde{G}$ in the first action and performing duality transformations. (The super Yang-Mills appears in the spectrum from the product (vector $\oplus$ scalars) $\otimes$ vector multiplet in the heterotic string.) This substitution is dictated by homogeneity in the dilaton, which prevents the usual conformal $\int d^2\theta\ W^2$ term. Such terms occur naturally in higher-dimensional couplings of supergravity to super Yang-Mills.

Classical and quantum symmetries of mechanics formulations of particle and string theories in background fields are often used to derive equations for those backgrounds. These features are not peculiar to these theories or their formulations: They are a general feature of describing a particle/field of some (super)spin in a gauge background. These equations fall into two distinct types: (1) A *supersymmetric* system in a gauge background of *higher* superspin generates *constraints* on the background, necessary for consistently defining the coupling (see subsections IVC4 and XA1). (2) Any gauge system in a background of the *same* gauge field generates *field equations* for the background (see exercise VIB8.2).

For example, the classical symmetries of the superparticle always generate constraints on its background, but give field equations for it only if the number of supersymmetries is enough to insure its superspin is as high as that of its background (e.g., 10D N=1 in background super Yang-Mills or 11D N=1 in background supergravity). Similarly, the bosonic string generates field equations for background gravity at the quantum mechanical level because quantization is required to reveal the massless graviton excited state contained in the string itself. On the other hand, the 10D superstring already generates field equations for background supergravity classically, since the ground state of the superstring (closed if boundary conditions are ignored), the only part that is evident (semi)classically, already contains supergravity.

7. Conformal field theory

We saw in subsections VIIIA7-8 some unusual features of massless theories in D=2. Since the mechanics of the string is mathematically equivalent to 2D field theory (as the mechanics of the particle is to 1D field theory), we now examine such field theories in a little more detail. In particular, since the string we studied in section XIA possessed local Weyl scale invariance on the worldsheet, we are directed to 2D conformal field theories coupled to 2D gravity.

Since for the most part we will be interested in free fields, quantization will be described most easily by the path-integral method. Although 2D field theory already looks quite different from the 1D field theory of particle mechanics, free 2D massless fields depend on only one of the two lightcone coordinates $\sigma^{\pm}$ (or are the sum of two such terms), and hence 2D conformal field theory is similar to 1D massive field theory. Consequently some of the features of particle mechanics or nonrelativistic field theory, such as the commutator, can still be useful and 2D Lorentz covariant. In particle mechanics, the (equal-time) commutator is evaluated by path-integral methods as

$$\langle [\mathcal{A},\mathcal{B}](t)\rangle \equiv \lim_{\epsilon\to 0}\langle \mathcal{A}(t+\epsilon)\mathcal{B}(t) - \mathcal{B}(t+\epsilon)\mathcal{A}(t)\rangle$$

(and similarly for the anticommutator), since $\mathcal{A}$ and $\mathcal{B}$ are treated as classical functions when evaluating the path integral

$$\langle f\rangle \equiv \int D\phi\ f e^{-iS}$$

where now "$\langle\ \rangle$" refers not to just the vacuum expectation value, but incorporates arbitrary initial and final states through the boundary conditions, or explicit wave functions in the path integral (see subsections VA1, XIA6). In general, this definition of $\langle\ \rangle$ actually gives the time-ordered expectation value, as follows from the derivation of subsection VA1: The ϵ's were introduced to enforce the appropriate ordering. For the rest of this subsection time ordering will be implicit in expectation values.

For simplicity, we assume the worldsheet boundary is at infinity; this can be achieved by conformal (coordinate) transformation ($z = e^{-\rho}$ on the Euclidean worldsheet: see subsection XIA6), or we can look at just short-distance effects. The upper-half plane, for the open string, can be completed into the full plane, as describes the closed string, by extending the open string to have closed-string boundary conditions, but with half as many fields, as described in subsection XIA3: In the following we consider only fields that are left- or right-handed (functions on-shell of only σ^+ or σ^-, as $X_{(\pm)}$ or $\hat{P}_{(\pm)}$), which are always defined on the whole worldsheet (or, before conformal transformation, periodic on the cylinder).

The simplest example is the 2D "spinor". From subsection VIIIA7, we have

$$\langle\bar{\psi}_\alpha(\sigma^m)\psi_\alpha(\sigma'^m)\rangle = \frac{-i}{(\sigma-\sigma')^\pm - i\epsilon\epsilon(\tau-\tau')}\langle 1\rangle + ...$$

(not summed over α) or in operator notation simply

$$\bar{\psi}_\alpha(\sigma^m)\psi_\alpha(\sigma'^m) = \frac{-i}{(\sigma-\sigma')^\pm - i\epsilon\epsilon(\tau-\tau')} + ...$$

where "..." means terms that are finite in the limit $\tau \to \tau'$, resulting from propagators that extend not between these two fermions, but between either one and the initial or final wave function. Using the identity (see exercise VA3.1)

$$\frac{i}{x+i\epsilon} - \frac{i}{x-i\epsilon} = 2\pi\delta(x)$$

we then have the usual

$$\{\bar{\psi}_\alpha(\sigma^\pm), \psi_\alpha(\sigma'^\pm)\} = 2\pi\delta(\sigma^\pm - \sigma'^\pm)$$

Normally, this would taken at equal τ, but in 2D conformal field theory, since the fields' time-dependence is given by their depending on just σ^+ or just σ^-, and since even interacting string calculations in terms of these "free" (with respect to first-quantization) 2D fields factorizes into two separate calculations for the left-handed fields and for the right-handed fields, we generally treat just σ^+ or just σ^- as the only argument. (After Wick rotation, this becomes just the complex variable z or just $\bar{z}$.)

This limiting procedure can be avoided when evaluating commutators of σ-integrated quantities, such as conserved charges: Since such "surface" integrals in D=2 are basically contour integrals (see exercise IIA1.2c), a commutator of the form

$$\left[\int \frac{d\sigma}{2\pi}\ \lambda(\sigma)\bar{\psi}(\sigma), \psi(\sigma')\right] = \lambda(\sigma')$$

(now dropping the α index) for some anticommuting function λ, which by the above definition is the integral over the difference of two contours, becomes the integral over a closed contour surrounding σ', directly picking up the contribution from the pole in $\sigma - \sigma'$ (see exercise VA3.1b).

One unusual consequence of 2D massless field theory is bosonization. In subsection VIIIA7 we saw from propagators that the usual massless scalar $\phi = \phi_{(+)} + \phi_{(-)}$ can be used to construct fermions as $\psi_\alpha = e^{-i\phi_{(\pm)}}$, $\bar{\psi}_\alpha = e^{i\phi_{(\pm)}}$. From the above arguments this implies the usual anticommutation relations. (Note that such commutators are completely quantum mechanical and not merely semiclassical: The calculation of subsection VIIIA7 involved multiloop diagrams with arbitrary numbers of propagators of scalars, whereas Poisson brackets effectively use only a single propagator of a fundamental field.)

As a further check of this equivalence we now examine their conformal transformations. The commutator of the scalar conformal generators

$$T_{\pm\pm} = \tfrac{1}{2}(\partial_\pm \phi)^2 = \tfrac{1}{2}(\partial_\pm \phi_{(\pm)})^2$$

with that scalar follows from a single propagator (see subsection VIIIA7):

$$\phi_{(\pm)}(\sigma^m)\phi_{(\pm)}(\sigma'^m) = -ln[i(\sigma^\pm - \sigma'^\pm)] + ...$$

$$\Rightarrow \quad [\partial_\pm \phi_{(\pm)}(\sigma^m)]\phi_{(\pm)}(\sigma'^m) = -\frac{1}{\sigma^\pm - \sigma'^\pm} + ...$$

$$\Rightarrow \quad [\partial_\pm \phi_{(\pm)}(1), \phi_{(\pm)}(2)] = -i2\pi\delta(1-2)$$

(again dropping the $\pm$ on σ, using "1" for "σ_1", etc.), confirming that $\partial\phi$ is canonically conjugate to ϕ. We then find

$$[T_{\pm\pm}(1), \phi_{(\pm)}(2)] = -i2\pi\delta(1-2)\phi'_{(\pm)}(2), \qquad [T_{\mp\mp}, \phi_{(\pm)}] = 0$$

Similarly, for the composite fermion we find

$$[\partial_\pm \phi_{(\pm)}(\sigma^m)]e^{ia\phi_{(\pm)}(\sigma'^m)} = -\frac{ia}{\sigma^\pm - \sigma'^\pm}e^{ia\phi_{(\pm)}(\sigma'^m)} + ...$$

$$\Rightarrow \quad [\partial_\pm \phi_{(\pm)}(1), \bar{\psi}_\alpha(2)] = 2\pi\delta(1-2)\bar{\psi}_\alpha(2),$$

$$[\partial_\pm \phi_{(\pm)}(1), \psi_\alpha(2)] = -2\pi\delta(1-2)\psi_\alpha(2)$$

in agreement with the identification $\partial_\pm \phi_{(\pm)} = \bar{\psi}_\alpha \psi_\alpha$. Then the product with T gives two factors of $1/(\sigma - \sigma')$ for the most singular term, as well as a less-singular term from connecting a propagator to either $\partial\phi$ (but not both):

$$\tfrac{1}{2}(\partial\phi)^2(\sigma^m)e^{ia\phi(\sigma'^m)} = -\tfrac{1}{2}\left(\frac{a}{\sigma^\pm - \sigma'^\pm}\right)^2 e^{ia\phi(\sigma'^m)} - \frac{1}{\sigma^\pm - \sigma'^\pm}\left(e^{ia\phi(\sigma'^m)}\right)' + ...$$

where we have used

$$[ia\partial\phi(\sigma^m)]e^{ia\phi(\sigma'^m)} = \left(e^{ia\phi(\sigma'^m)}\right)' + \mathcal{O}(\sigma^\pm - \sigma'^\pm)$$

The final result is thus:

$$[T_{\pm\pm}(1), \psi_\alpha(2)] = -i2\pi\delta(2-1)\psi'_\alpha(2) - i\pi\delta'(2-1)\psi_\alpha(2), \qquad [T_{\mp\mp}, \psi_\alpha] = 0$$

These results generalize easily to general linear exponentials for multiple scalars: Including an indefinite metric

$$\langle 0|\phi_{i(\pm)}(\sigma^m)\phi_{j(\pm)}(\sigma'^m)|0\rangle = -\eta_{ij}[i\ ln(\sigma - \sigma')^\pm]$$

we have

$$e^{ia\cdot\phi_{(\pm)}(\sigma^m)}e^{ib\cdot\phi_{(\pm)}(\sigma'^m)} = [i(\sigma^\pm - \sigma'^\pm)]^{a\cdot b}e^{ia\cdot\phi_{(\pm)}(\sigma^m)+ib\cdot\phi_{(\pm)}(\sigma'^m)} + ...$$

as the most singular contribution to the path integral, using the same method as in subsection VIIIA7, where indices are raised and lowered with η_{ij}, as usual. Furthermore, we can generalize the conformal generators with a linear term, as results from a $\frac{1}{2}R\mu^i\phi_i$ term in the action:

$$L = \tfrac{1}{4}\eta^{ij}(\nabla\phi_i)\cdot(\nabla\phi_j) + \tfrac{1}{2}R\mu^i\phi_i \quad \Rightarrow \quad T_{\pm\pm} = \tfrac{1}{2}\eta^{ij}(\partial_\pm\phi_{i(\pm)})(\partial_\pm\phi_{j(\pm)}) - \mu^i\partial^2_\pm\phi_{i(\pm)}$$

Its commutators then generalize, for any "covariant" fields χ, as

$$\left[i\int\frac{d\sigma'}{2\pi}\lambda(\sigma')T_{\pm\pm}(\sigma'), \chi(\sigma)\right] = \lambda(\sigma)\chi'(\sigma) + w_{(\pm)}\lambda'(\sigma)\chi(\sigma)$$

$$i\frac{1}{2\pi}[T_{\pm\pm}(1), \chi(2)] = \delta(2-1)\chi'(2) + w_{(\pm)}\delta'(2-1)\chi(2)$$

where $w_{(\pm)}$ are the conformal weights: $w_{(+)} + w_{(-)}$ is the usual conformal weight, while $w_{(+)} - w_{(-)}$ is the "spin". For the fields $e^{ia\cdot\phi_{(\pm)}}$ we then find

$$w_{(\pm)}(a) = \tfrac{1}{2}a^2 + i\mu\cdot a, \qquad w_{(\mp)} = 0$$

In particular, for $\mu = 0$ we see that ψ_α has $w_{(\pm)} = \frac{1}{2}$, while $\phi_{(\pm)}$ is covariant with weight $w_{(\pm)} = 0$.

Exercise XIB7.1

Use the result for $e^{ia\phi}e^{ib\phi}$ to derive $\bar{\psi}_\alpha\psi_\alpha = \partial_\pm\phi_{(\pm)}$ (up to an infinite constant), defining the equal-time product as a limit, as above.

Exercise XIB7.2

The $i\mu\cdot a$ term in w is classical, since it comes from a single propagator:

a Derive the above Lagrangian for a single scalar by starting with the nonlocal term $R(1/\square)R$ and applying a local Weyl scale transformation, introducing the scalar as the compensator.

b Find the classical scale weight of $e^{ia\phi}$ from its local scale transformation. (Hint: In deriving the local scale transformation in part **a**, an exponential will be needed, so that e^{ϕ} transforms homogeneously.)

The most important use of such exponentials, outside of bosonization, is in the use of external fields (or "vertex operators"): The conformal generators (energy-momentum tensor) have conformal weight 2; requiring that background fields preserve conformal invariance implies that such vertex operators must have conformal weight 1 and be local on the worldsheet:

$$i\frac{1}{2\pi}[\hat{T}_{\pm\pm}(1),\hat{T}_{\pm\pm}(2)] = \delta'(2-1)[\hat{T}_{\pm\pm}(1)+\hat{T}_{\pm\pm}(2)]$$

$$\hat{T}_{\pm\pm}(\sigma) = T_{\pm\pm}(\sigma) + \widetilde{W}(\sigma) \quad \Rightarrow \quad \widetilde{W} \ \mathit{has}\ w = 1, \qquad [\widetilde{W}(\sigma),\widetilde{W}(\sigma')] = 0$$

$$\Rightarrow \quad \widetilde{W}(\sigma) = 2\pi\delta(\sigma-\sigma_0)W(\sigma_0)$$

where we have assumed the conformal anomaly cancels (or ignored its contribution), and solved for closure of the algebra perturbatively in the background.

If we write a background spacetime field $\Phi(X(\sigma^m))$ as a Fourier transform, then we see that its conformal weight is proportional to k^2, the square of the external momentum. Hence a vertex operator consisting of just a scalar field produces the tachyonic ground state. Excited states are created by products of derivatives of X times fields (with spacetime Lorentz indices contracted); the derivatives add to the conformal weight, forcing k^2 to decrease in compensation, resulting in massless and massive ($m^2 > 0$) states.

8. Triality

Bosonization can also be applied to representations of groups (Lorentz or internal). In particular, to obtain the correct anticommutation relations for a fermion $e^{ia\cdot\phi}$ and its conjugate $e^{-ia\cdot\phi}$ we require $a^2 = 1$. Their weights are then $\frac{1}{2} \pm i\mu\cdot a$. The simplest choice to obtain $2n$ fermions from n scalars is, for $\eta_{ij} = \delta_{ij}$, in a (complex) null basis for $SO(2n)$,

$$SO(2n)\ \mathit{vector}: \quad a^i_V = (\pm1,0,0,...),(0,\pm1,0,...),...$$

Klein factors should be included to make fermions using different scalars anticommute (see subsection IA2). This construction follows that for the Dirac matrices in subsection XC1: Each scalar corresponds to a 2D subspace of $SO(2n)$, and each component of a D-vector a^i is the corresponding eigenvalue of the 2D spin. From that construction we see that spinors should be

$$SO(2n)\ spinors: \quad a^i_{S,S'} = (\pm\tfrac{1}{2}, \pm\tfrac{1}{2}, ...)$$

with independent $\pm$'s, with the product of all of them equal to 1 for one Weyl spinor and -1 for the other. (The conventions are slightly different from subsection XC1: Now we use a representation where σ_3 is diagonal, and γ_{-1} is chosen as the *last* γ.)

However, the resulting operators have the correct anticommutation relations and conformal weights only for SO(8) ($n = 4$), corresponding to the lightcone symmetry for the 10D superstring. This follows from the "triality" symmetry between the vector and two spinors: By simply changing the scalar-field basis so that the a's for one of the spinors are $\pm$ the basis elements, the a's for the vector and the other spinor take the same form as above for the two spinors (i.e., the above expressions for the vector and spinor a's are permuted).

Similar triality constructions apply to lower dimensions by taking into account supersymmetry, which also relates a vector to a spinor. In D=6, simple supersymmetry has an internal SU(2) (R) symmetry: Thus, there is a triality relating this SU(2) to the two SU(2)'s of the lightcone's SO(4). In terms of these, the vector is the $(\frac{1}{2}, \frac{1}{2}, 0)$ representation, while the spinors are $(\frac{1}{2}, 0, \frac{1}{2})$ and $(0, \frac{1}{2}, \frac{1}{2})$. The resulting operators are given by

$$a_V = (\pm\tfrac{1}{\sqrt{2}}, \pm\tfrac{1}{\sqrt{2}}, 0), \qquad a_S = (\pm\tfrac{1}{\sqrt{2}}, 0, \pm\tfrac{1}{\sqrt{2}}), \qquad a_{S'} = (0, \pm\tfrac{1}{\sqrt{2}}, \pm\tfrac{1}{\sqrt{2}})$$

To relate to the SO(8) results we use the vector to define the basis, yielding

$$a_V = (\pm 1, 0, 0), (0, \pm 1, 0)$$

$$\Rightarrow \quad a_S = (\tfrac{1}{2}, \tfrac{1}{2}, \pm\tfrac{1}{\sqrt{2}}), (-\tfrac{1}{2}, -\tfrac{1}{2}, \pm\tfrac{1}{\sqrt{2}}); \quad a_{S'} = (\tfrac{1}{2}, -\tfrac{1}{2}, \pm\tfrac{1}{\sqrt{2}}), (-\tfrac{1}{2}, \tfrac{1}{2}, \pm\tfrac{1}{\sqrt{2}})$$

This also follows directly from the SO(8) result by dropping the third and fourth scalars for the vector, and using only $(1/\sqrt{2}\times)$ their sum for the spinors. (I.e., it represents only internal symmetry.) For D=4 the construction is even simpler: Besides the SO(2)=U(1) of the lightcone, there is a second U(1) for R symmetry. In terms of the complex plane defined by these two quantum numbers, there is an obvious triality for the three cube roots of 1; thus

$$a_V = \pm(1, 0), \qquad a_S = \pm(-\tfrac{1}{2}, \tfrac{\sqrt{3}}{2}), \qquad a_{S'} = \pm(-\tfrac{1}{2}, -\tfrac{\sqrt{3}}{2})$$

which again also follows from SO(8), now combining its last 3 scalars.

Exercise XIB8.1

We now extend the analogy to the construction of subsection XC1:

a Show for general SO(2n), in analogy to the Dirac γ's, that the (integral of) products of two vector fermions, antisymmetrized in the vector indices, act in the same way as the group generators, by examining their commutators with each other and with the vector and spinor operators.

b Show for the triality cases that the (anti)commutator of two representations yields the third (supersymmetry).

These constructions can be generalized from the lightcone to manifest Lorentz covariance by adding equal numbers of scalars of positive and negative metric (at least one of each): Their contributions to the spinors' operator product (power of σ) then cancel, preserving the anticommutation relations. One of the extra scalars of positive metric yields the two "longitudinal" spacetime directions to complete the SO(D−2) vector and spinor representations to SO(D−1,1). The rest of the scalars come from "ghosts". Note that the spacetime metric is unrelated to the metric for the scalars: The Minkowski spacetime metric comes from Wick rotation of the scalars, as applied in subsection XC2 to the construction of subsection XC1 for Dirac spinors. For either Euclidean or Minkowski spacetime the basis is null; the only difference is in reality.

REFERENCES

1 Ramond; Neveu and Schwarz; *loc. cit.* (IIC):
almost superstrings (fermions included).

2 M.B. Green and J.H. Schwarz, *Phys. Lett.* **109B** (1982) 444, **149B** (1984) 117:
superstrings.

3 D.J. Gross, J.A. Harvey, E. Martinec, and R. Rohm, *Phys. Rev. Lett.* **54** (1985) 502, *Nucl. Phys.* **B256** (1985) 253, **267** (1986) 75:
heterotic string.

4 W. Siegel, *Phys. Lett.* **134B** (1984) 318;
T.H. Buscher, *Phys. Lett.* **194B** (1987) 59, **201B** (1988) 466:
T-duality as transformation on worldsheet.

5 K. Kikkawa and M. Yamasaki, *Phys. Lett.* **149B** (1984) 357;
N. Sakai and I. Senda, *Prog. Theor. Phys.* **75** (1986) 692;
V.P. Nair, A. Shapere, A. Strominger, and F. Wilczek, *Nucl. Phys.* **B287** (1987) 402;
B. Sathiapalan, *Phys. Rev. Lett.* **58** (1987) 1597;
R. Dijkgraaf, E. Verlinde, and H. Verlinde, *Comm. Math. Phys.* **115** (1988) 649;
K.S. Narain, M.H. Sarmadi, and E. Witten, *Nucl. Phys.* **B279** (1987) 369;
P. Ginsparg, *Phys. Rev.* **D35** (1987) 648;
P. Ginsparg and C. Vafa, *Nucl. Phys.* **B289** (1987) 414;
S. Cecotti, S. Ferrara, and L. Girardello, *Nucl. Phys.* **B308** (1988) 436;
R. Brandenberger and C. Vafa, *Nucl. Phys.* **B316** (1988) 391;

A. Giveon, E. Rabinovici, and G. Veneziano, *Nucl. Phys.* **B322** (1989) 167;
A. Shapere and F. Wilczek, *Nucl. Phys.* **B320** (1989) 669;
M. Dine, P. Huet, and N. Seiberg, *Nucl. Phys.* **B322** (1989) 301;
J. Molera and B. Ovrut, *Phys. Rev.* **D40** (1989) 1146;
K.A. Meissner and G. Veneziano, *Phys. Lett.* **267B** (1991) 33;
A.A. Tseytlin and C. Vafa, hep-th/9109048, *Nucl. Phys.* **B372** (1992) 443;
M. Roček and E. Verlinde, hep-th/9110053, *Nucl. Phys.* **B373** (1992) 630;
J.H. Horne, G.T. Horowitz, and A.R. Steif, hep-th/9110065, *Phys. Rev. Lett.* **68** (1992) 568;
A. Sen, *Phys. Lett.* **271B** (1991) 295;
A. Giveon and M. Roček, hep-th/9112070, *Nucl. Phys.* **B380** (1992) 128:
T-duality as symmetry of spacetime fields.

6 M.J. Duff, *Nucl. Phys.* **B335** (1990) 610;
J. Maharana and J.H. Schwarz, hep-th/9207016, *Nucl. Phys.* **B390** (1993) 3;
W. Siegel, hep-th/9302036, *Phys. Rev.* **D47** (1993) 5453, hep-th/9305073, *Phys. Rev.* **D48** (1993) 2826, hep-th/9308133, Manifest duality in low-energy superstrings, in *Proc. of the Conference Strings '93*, Berkeley, CA, May 24-29, eds. M.B. Halpern, G. Rivlis, and A. Sevrin (World Scientific, 1995) p. 353:
Metric and axion as coset space for T-duality.

7 Siegel, *loc. cit.* (IXB, 1st and 2nd papers of ref. 5):
coupling of (unphysical) dilaton to string (through ghosts).

8 E.S. Fradkin and A.A. Tseytlin, *Phys. Lett.* **158B** (1985) 316, *Nucl. Phys.* **B261** (1985) 1:
coupling of dilaton through worldsheet curvature.

9 T. Banks, D. Nemeschansky, and A. Sen, *Nucl. Phys.* **B277** (1986) 67:
relation between above two couplings.

10 Fradkin and Tseytlin, *loc. cit.*;
C.G. Callan, D. Friedan, E.J. Martinec, and M.J. Perry, *Nucl. Phys.* **B262** (1985) 593:
low-energy closed-string actions for massless fields.

11 S. Cecotti, S. Ferrara, and M. Villasante, *Int. J. Mod. Phys. A* **2** (1987) 1839:
superspace action for 4D massless part of heterotic string.

12 W. Siegel, *Phys. Lett.* **211B** (1988) 55:
massless part of heterotic string is old-minimal supergravity coupled to tensor multiplet, as follows from direct product of open strings.

13 W. Siegel, hep-th/9510150, *Phys. Rev.* **D53** (1996) 3324:
massless part of string actions from direct product and dilaton homogeneity.

14 Virasoro; Gelfand and Fuchs; *loc. cit.* (XIA);
S. Fubini and G. Veneziano, *Nuo. Com.* **67A** (1970) 29, *Ann. Phys.* **63** (1971) 12;
A. Galli, *Nuo. Cim.* **69A** (1970) 275;
J.L. Gervais, *Nucl. Phys.* **B21** (1970) 192;
J.-L. Gervais and B. Sakita, *Nucl. Phys.* **B34** (1971) 477;
A. Chodos and C.B. Thorn, *Nucl. Phys.* **B72** (1974) 509:
theory of free conformal fields.

15 R. Marnelius, *Nucl. Phys.* **B211** (1983) 14:
effect of $R\phi$ term on Virasoro operators.

16 E. Witten, D=10 superstring theory, in *Fourth workshop on grand unification*, University of Pennsylvania, Philadelphia, April 21-23, 1983, eds. H.A. Weldon, P. Langacker,

and P.J. Steinhardt (Birkhäuser, 1983) p. 395:
vectors and spinors of SO(8) as different exponentials of bosons.

17 W. Siegel and B. Zwiebach, *Nucl. Phys.* **B263** (1986) 105:
bosonization of ghosts (bosonic string).

18 D. Friedan, E. Martinec, and S. Shenker, *Phys. Lett.* **160B** (1985) 55, *Nucl. Phys.* **B271** (1986) 93;
V.G. Knizhnik, *Phys. Lett.* **160B** (1985) 403:
vectors and spinors of SO(9,1) as different exponentials of bosons.

19 M. Kaku, *Strings, conformal fields, and topology: an introduction* (Springer-Verlag, 1991);
P. Ginsparg and G. Moore, hep-th/9304011, Lectures on 2D gravity and 2D string theory, in *Recent directions in particle theory: from superstrings and black holes to the standard model*, Proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 1992), Boulder, CO, June 1-26, 1992, eds. J. Harvey and J. Polchinski (World Scientific, 1993) p. 277;
S.V. Ketov, *Conformal field theory* (World Scientific, 1995);
P. Di Francesco, P. Mathieu, and D. Sénéchal, *Conformal field theory* (Springer, 1997).

C. LATTICES

Integrals are defined as limits of sums. For some cases it can be convenient to define quantum theories on discrete spacetimes ("lattices"), perform all calculations there, and then take the limit of continuous spacetime. Two types of such lattices will be considered here: (1) Physical four-dimensional spacetime can be treated as a regular hypercubic lattice. Then the existence and uniqueness of a continuum limit where Lorentz invariance is restored must be proven. (2) In first-quantization of particles or strings, the worldline or worldsheet can be approximated as a random lattice. Integration over the metric of the worldline or worldsheet is then replaced with summation over lattices with different geometries. The continuum limit is not required by physical criteria, but only for purposes of comparison to the theory as defined in the continuum.

1. Spacetime lattice

The use of a regular 4D lattice for quantizing QCD has three main advantages:

(1) The lattice acts as a gauge invariant regulator for UV divergences (and, if the lattice is finite, also IR ones).

(2) Gauge fixing is no longer necessary, since the path integral can be performed without it.

(3) Nonperturbative calculations are possible, some analytically and some numerically (if the lattice is small enough).

Gauge fields are associated with translations through the covariant derivative. However, on a lattice, even a regular one, infinitesimal translations are no longer possible: For example, scalar fields are defined only at vertices of the lattice. We therefore consider covariantizing finite translations, as in subsections IIIA5 and IIIC2,

$$e^{-k^m \nabla_m} = \mathcal{P}\left[exp\left(-i\int_{x-k}^{x} dx' \cdot A\right)\right] e^{-k\cdot\partial} = U_{x,x-k}e^{-k\cdot\partial}$$

Without loss of generality, we can restrict ourselves to translations along links, from one vertex straight to an adjacent one (keeping all coordinates but one constant), and successive combinations of these. Then the gauge field is replaced with the group element $U_{x,x-k}$ associated with each link, where k is now any of the 4 orthonormal basis vectors (in Euclidean space). The gauge transformation of this representation of the gauge field follows from either the path-ordered definition or the covariant-translation definition:

$$e^{-k\cdot\nabla(x)'} = \mathbf{g}(x)e^{-k\cdot\nabla}\mathbf{g}^{-1}(x) \quad \Rightarrow \quad U'_{x,x-k} = \mathbf{g}(x)U_{x,x-k}\mathbf{g}^{-1}(x-k)$$

Note that, while the gauge field is a group element associated with a link, the gauge transformation is a group element associated with a vertex. Furthermore, the field strength can be associated with the product of these group elements of the links bounding a "plaquet":

$$U_{x,x-k}U_{x-k,x-k-k'}U_{x-k-k',x-k'}U_{x-k',x} = \mathcal{P}\left(e^{-i\oint dx\cdot A}\right) = e^{-k\cdot\nabla}e^{-k'\cdot\nabla}e^{k\cdot\nabla}e^{k'\cdot\nabla}$$

$$\approx e^{[k\cdot\nabla,k'\cdot\nabla]} \approx 1 + ik^a k'^b F_{ab}$$

where we have used

$$e^B e^C = e^{B+C+\frac{1}{2}[B,C]+\ldots}$$

(In general, there is a geometric prescription associating a scalar with a point, a vector with a line, a second-rank antisymmetric tensor with a surface, etc.)

We now define a gauge-invariant action by looking for an expression in terms of these group elements that approximates the usual Yang-Mills action to lowest order in the lattice spacing, while involving the least number of factors of the group elements. The result is:

$$S = -\tfrac{1}{g^2}tr\sum_{plaquets}(U_{x,x-k}U_{x-k,x-k-k'}U_{x-k-k',x-k'}U_{x-k',x} - 1)$$

$$\approx -\tfrac{1}{g^2}tr\sum_{plaquets}\tfrac{1}{2}(ik^a k'^b F_{ab})^2 \sim \tfrac{1}{g^2}tr\sum_x F^2(x)$$

(expanding the exponential as above to quadratic order, and noting that total commutators vanish when traced). Since our fields are now represented by group elements, we no longer need to fix the gauge to make the functional path integral well defined: In contrast to the continuum case, where integrating a gauge-invariant action over gauge transformations would produce an infinite factor, here such an integral at any one point is just an integral over the group space, which is finite (for compact groups, which have finite volume). The functional integration is now integration over U for each link, where the range of U is the group space (which is finite, since the group is compact).

Matter can also be introduced: Scalars are naturally associated with vertices, just as vectors are with links, and second-rank antisymmetric tensors with plaquets. However, fermions do not have such a natural geometric interpretation. In particular, it has been proven (the "Nielsen-Ninomiya theorem") that massless fermions can't be defined in a useful way on the lattice without "fermion doubling": There must be a multiple of 2^D massless fermion fields for D lattice dimensions. This is closely related to the existence of axial anomalies: The absence of an anomaly is implied by the existence of a regularization that manifestly preserves a symmetry (in this case, chiral symmetry as a consequence of the existence of lattice-regularized fermions). However, massless fermions can be defined as limits of massive ones (so chiral invariance is not manifest). Alternatively, *nonlocal* spinor kinetic operators can be found that preserve masslessness and chirality without doubling. (The nonlocality can be controlled, but at the cost of a significantly more complicated action.)

Exercise XIC1.1

For the lattice action for a spinor in D=1, use

$$S = \tfrac{1}{2} i \sum_n \psi_n \psi_{n+1}$$

where ψ is a real one-component fermion.

a Show this has the correct continuum limit.

b Find the equations of motion.

c Solve the equations of motion for both the lattice and continuum cases, and show the lattice has twice as many solutions.

d Repeat all the above for the single-component complex (Dirac) fermion,

$$S = \tfrac{1}{2} i \sum_n \bar{\psi}_n \psi_{n+1}$$

e Make the same analysis for the D=1 scalar, and show it has no such problems. (Hint: Try an expansion in plane waves.)

Exercise XIC1.2

In the book by Feynman and Hibbs, exercise 2-6 states rules for the path integral for a Dirac spinor in D=2. These rules are equivalent to the use of a lightlike lattice, where the lightcone coordinates are discretized. The rules are to consider all paths that are piecewise lightlike forward in time, with a factor of $im\epsilon$ for each right-angle "kink" (where m is the mass and ϵ the lattice spacing). Show these rules follow from the 2D action for a Dirac spinor (see subsection VIIIA7, and include a mass term), using a term as in the previous exercise for the derivative term for each of the two component fields (each of which has a derivative in only one of the two lightlike coordinates).

In general, fermions are more difficult to integrate over, particularly when using "numerical methods" (computers), since fermions are not numbers. In principle one can integrate out the fermions analytically to produce functional determinants in terms of bosonic fields, but nonlocality makes them hard to evaluate by iterative schemes. In practice fermion loops are usually ignored ("quenched approximation"), which corresponds to leading order in an expansion in the inverse of the number of flavors, or the approximation of heavy quarks. The resulting accuracy of QCD calculations for low-energy parameters (masses of light hadrons, decay constants, etc.) is of the order of 5-10%. (Getting good numbers in nonperturbative calculations is significantly harder than in perturbative ones. The situation is expected to improve somewhat with the advent of faster computers.) Finding scattering amplitudes, or other properties that involve high-mass hadrons, is presently beyond the scope of lattice methods. However, lattice QCD is one of the few methods so far to obtain numbers for comparison with experiment from nonperturbative calculations with the QCD action. (Other nonperturbative methods have also been restricted to low-mass hadrons, and basically study effects of chiral symmetry breaking, not confinement.)

The spacetime lattice allows a direct nonperturbative analysis of confinement. For example, consider the potential between a heavy quark-antiquark pair. The heaviness again allows us to ignore pair creation, and to treat the quarks as static. For simplicity, consider scalar quarks, as described by first-quantization. Since we approximate the quarks as static, the only relevant term in the quark mechanics action is the interaction term $\int d\tau\ \dot{x} \cdot A = \int dx \cdot A$. Taking into account the nonabelian nature of the group, and ignoring the first-quantized path integration Dx (since x is assumed fixed), the factor e^{-S} for the quark becomes just the path-ordered expression $\mathcal{P}(e^{-i\int dx\cdot A})$ we have been considering, while for the antiquark we get the inverse expression. To get a gauge-invariant expression, we connect the paths at top and bottom, since the fields will be fixed at the boundaries at $t = \pm\infty$. (Functional integration over any gauge-field link picks out just the singlet part of the integrand, since the integral is over the group, and nonsinglet representations can be rotated to minus themselves by an appropriate group element, canceling the contribution.) The result is a "Wilson loop"

$$tr\ \mathcal{P}\left(e^{-i\oint dx\cdot A}\right)$$

The strong-coupling expansion is applied by expanding the functional integrand e^{-S} in powers of S, which is an expansion in powers of $1/g^2$, and which is also an expansion in the number of plaquets. Clearly the dominant term in this expansion is the one with the fewest factors of S. To be nonvanishing, each link variable must appear in a singlet combination: The function of that link, when expanded in irreducible group representations, must include a term that is proportional to the identity. For example, for any unitary group, this is true for the product $U_i{}^j(U^{-1})_k{}^l$, where the two U's are for the same link, and the indices are the group indices; this has the constant piece $U_i{}^j(U^{-1})_j{}^i$. For the case of the Wilson loop, if we assume the simplest case where the path is a rectangle, then we need at least a factor of S for each plaquet enclosed by the loop, so there will a UU^{-1} for each link on the boundary (one factor from the loop, one from S), as well as for each link enclosed by it (both factors from the contribution to S from either side). The result for the path integral is then

$$\mathcal{A} = \left\langle tr\ \mathcal{P}\left(e^{-i\oint dx\cdot A}\right)\right\rangle \sim e^{-Vt} \sim \left(\frac{1}{g^2}\right)^{rt}$$

where V is the (potential) energy ($S = \int dt(V+T)$ in Euclidean space), t is the time separation between the top and bottom of the rectangle, and r is the spatial separation between the two sides. We thus have a linear quark-antiquark potential

$$V(r) \sim (ln\ g)r$$

so the quark-antiquark pair is confined.

Unfortunately, we can get a similar result from QED, by defining the U(1) group in terms of a phase factor (so effectively the range of group integration is 2π, defining a "compact" group). The reason is that for this U(1) theory this strong coupling expansion is not accurate. The approximation is better for nonabelian theories, but the persistence of confinement has not been proven in the continuum limit (small coupling). In fact, while the transition to deconfinement in Abelian theories has been found at finite coupling, it has been proven that such a phenomenon can occur in the nonabelian theory only near zero coupling. However, the perturbative properties of the continuum theory show that this is exactly where one expects the appearance of ambiguities in the theory (known in lattice terminology as "nonuniversality"); this is directly related to the ambiguities in the Borel transform discussed in subsection VIIC3.

2. Worldsheet lattice

In string theory there are two spaces, the two-dimensional space of the worldsheet, and physical spacetime. In the previous subsection we considered approximating spacetime by a lattice; in this subsection we instead approximate the worldsheet by lattices. For the spacetime of QCD we used a regular lattice, representing the fixed geometry of flat spacetime. In string theory we considered worldsheets of arbitrary geometry, described by a worldsheet metric, so our lattices should be more arbitrary; in fact, functional integration over the worldsheet metric must be replaced by summation over different lattices. We saw that the topological expansion of QCD in 1/N generated polyhedra analogous to the worldsheet, with 1/N acting as the string coupling. We therefore identify the Feynman diagrams themselves, with faces chosen by the 1/N expansion, as these lattices, to give a more precise correlation between the second-quantized path integral of QCD (and other field theories) and the first-quantized path integral of string theory.

Presently the relation between such field theories and string theory is not well understood, and has been described only for the bosonic string. Since the bosonic string has only the worldsheet metric and spacetime coordinates as degrees of freedom, it corresponds to a (N×N-matrix) scalar field theory. Since a lattice requires a scale, while conformal invariance includes scale invariance, we must break the conformal invariance of the worldsheet. The simplest coordinate-invariant yet scale-variant property of a space is its volume, so we add a volume (area) term to the string action. Furthermore, to describe interactions we need to include a term containing the string coupling constant. In string theory the coupling is topological, in the sense that the power of the coupling constant is counted by the "genus" of the worldsheet (the number of "windows" plus twice the number of "handles"). This is the "Euler number" χ, counted by the integral of the worldsheet curvature (see exercise IXA7.3). Our worldsheet action thus consists of the three terms

$$S = \int \frac{d^2\sigma}{2\pi}\sqrt{-g}\left[\frac{1}{\alpha'}g^{mn}\tfrac{1}{2}(\partial_m X)\cdot(\partial_n X) + \mu + (ln\ \kappa)\tfrac{1}{2}R\right]$$

A lattice version of this action is (with χ given in subsection VIIC4)

$$S_1 = \frac{1}{\alpha'}\sum_{\langle jk\rangle}\tfrac{1}{2}(X_j - X_k)^2 + \mu\sum_j 1 + (ln\ \kappa)\left(\sum_j 1 - \sum_{\langle jk\rangle} 1 + \sum_J 1\right)$$

where j are vertices of the lattice, $\langle jk\rangle$ are the links, and J are the plaquets.

Exercise XIC2.1

Put the particle on a random lattice "Minkowski" worldline. (See exercise VB1.2.) Show the propagator for a massless particle, written in momentum space, before taking the limit lattice spacing $\epsilon \to 0$, is

$$\Delta = \frac{2\epsilon}{1 - e^{-i\epsilon p^2}}$$

Show this has unphysical poles at $p^2 = 2\pi n/\epsilon$ for arbitrary integer n. How do these results differ if the propagator is defined for Wick-rotated τ?

The corresponding field theory is easily found, according to our earlier discussions, by (1) identifying the worldsheet lattice with a position-space Feynman diagram (the vertices of the lattice being those of the diagram, the links of the lattice being the Feynman propagators; see subsection VC8), and (2) using the 1/N expansion to associate the faces of the worldsheet polyhedra with the U(N) indices of the scalar field (see subsection VIIC4).

We then can immediately identify the three terms in the string action with their counterparts in the scalar field theory:

(1) The X term gives the propagators,

(2) the area term (which counts the vertices) gives the vertex factor (coupling constant), and

(3) the curvature term (which classifies the topology) gives the 1/N factors of the topological expansion.

Thus, the three constants in the string action can be identified with the mass, coupling, and number of colors of the scalar field theory. Explicitly, the field theory action is

$$S_2 = N\ tr \int \frac{d^D x}{(2\pi\alpha')^{D/2}} (\tfrac{1}{2}\phi e^{-\alpha'\Box/2}\phi - G\tfrac{1}{n}\phi^n)$$

where we have identified

$$G = e^{-\mu}, \quad \tfrac{1}{N} = \kappa, \quad m^2 = \tfrac{2}{\alpha'}$$

and we have put an overall factor of N (associated with $1/\hbar$) so that G (and m^2) is fixed (rather than G times some power of N) when the $1/N$ expansion is performed. (The reverse can be made true by rescaling ϕ.) The unusual kinetic operator

$$e^{-\alpha'\Box/2} = \tfrac{1}{m^2}(m^2 - \Box + ...)$$

comes from identifying the second-quantized particle propagator as it appears in the first-quantized path integral for the string:

$$\mathcal{A} = \int \prod_j \frac{d^D X_j}{(2\pi\alpha')^{D/2}}\ e^{-S_1} \quad \Rightarrow \quad \Delta(x,y) = e^{-(x-y)^2/2\alpha'}$$

Unlike the spacetime lattice, the worldsheet lattice preserves spacetime Poincaré symmetry, so it's not necessary to take any limits to define a physical theory (or at least taking limits won't improve the physical relevance of this model). This model thus describes a stiff or lumpy string. The usual continuum-worldsheet string then can be identified with a particular limit of this more general string. Explicit calculations have demonstrated that this lattice regularization of the worldsheet reproduces the results of the continuum approach. These results have been limited to spacetime dimension ≤ 1 because of the inconsistencies introduced by the tachyon, which is the ground state in higher dimensions. Unfortunately, this prevents study of the more interesting properties, such as scattering amplitudes and the precise form of the potential (we have left n arbitrary in S_2), since it's superrenormalizable in D$\leq$2 regardless of its form. However, these limitations probably would not appear in a corresponding formulation of the superstring, which has no tachyons.

An interesting feature of this model is the use of Gaussian propagators to get rid of the usual perturbative divergences of momentum integration. Naively, one might suspect that such field theories were completely finite. However, we know in this case that the bosonic string does have divergences perturbatively in the string coupling, and that there are further problems unless D=26. This demonstrates that modifying a theory to fix problems seen in perturbation theory does not preclude the reappearance of such difficulties nonperturbatively.

3. QCD strings

The main problem with known string theories is that they describe spacetimes of dimension D=10. To describe physics in D=4, it is usually assumed that 6 of the dimensions choose to "compactify" to submicroscopic dimensions, corresponding to length scales well below the range of present experiments. Such a solution to the classical field equations with minimal energy is chosen as the "vacuum", about which perturbations are performed, but nothing is known to preclude contributions to the functional integral from other vacuua, whether 4-dimensional, 10-dimensional, or elsewhere. Although strings have been chosen to describe quantum gravity because of their renormalizability (finiteness), this advantage is lost after compactification, since the arbitrariness in choice of compactification is tantamount to the loss of predictability in nonrenormalizable theories. However, there are some general features of string theories, such as the appearance of a dilaton-like physical scalar, that have suggested certain 4D models (such as no-scale supergravity, which is also suggested by extended supergravity: see subsection XB7).

Furthermore, D=10 string theories have recently been discovered to be (compactifications of) D=11 membrane theories in disguise, where the eleventh dimension shows up only nonperturbatively. Not only is using a formalism where not all of the dimensions are manifest a technical obstacle, but the quantum mechanics of membranes suffers from several problems, including nonrenormalizability, which is what string theories were chosen to avoid in the first place. This suggests that D=10 strings are nonrenormalizable at the nonperturbative level. Furthermore, even the few properties of 4D theories suggested by strings are no longer required, since compactification from D=11 need not proceed by way of D=10.

On the other hand, renormalizabilty of theories with a finite number of fields predicts D=4, since theories in higher dimensions are all nonrenormalizable (or have unbounded potentials: ϕ^3 theory). Furthermore, both experiments with hadrons and theoretical arguments in QCD suggest the existence of an inherently 4D string theory. (For example, the existence of a continuum limit for confining spacetime-lattice theories requires asymptotic freedom.)

The reason 10D strings avoid the D=4 limit is clear from the worldsheet lattice approach of the previous subsection: The partons that make up the string have Gaussian propagators, and Gaussian integrals always converge. This leads to Gaussian behavior of fixed-angle scattering (subsection XIA6), in conflict with hadronic physics, where power-law behavior is observed for partons with large transverse momenta, and is a theoretical consequence of asymptotic freedom with the usual propagators. (In fact, it is the main empirical verification of QCD.)

Since nonrelativistic first-quantization gives Gaussian propagators $e^{-x^2/t}$, it is not surprising that the simplest strings should result in partons with Gaussian propagators e^{-x^2}. However, the fact that first-quantization for particles leads instead to, e.g., $1/x^2$ propagators for massless particles in 4D position space suggests that an analogous treatment for strings should be possible. We thus attempt to follow the derivation of the previous subsection from parton to string, but starting with realistic parton propagators. The first step is to exponentiate the propagator so that the exponent can be identified with a first-quantized action. The easiest way, and that most analogous to the nonrelativistic case, is to use the Schwinger parametrization of the propagator, which follows from the appearance of the worldline metric in the action:

$$\frac{1}{\frac{1}{2}p^2} = \int_0^\infty d\tau\ e^{-\tau p^2/2}$$

As we saw in subsection VC8, a Feynman diagram in a scalar field theory with nonderivative self-interactions is then written as

$$\int dx_i' dp_{ij} d\tau_{ij}\ e^{-\sum_{\langle ij\rangle}[\tau_{ij}p_{ij}^2/2 - i(x_i - x_j)\cdot p_{ij}]}$$

In the (worldsheet) continuum limit of this expression, p becomes a worldsheet vector, so τ must become a symmetric worldsheet tensor. Since on a regular square lattice ("flat" worldsheet) there are two propagators per vertex (for the two independent directions), τ must be a traceless tensor. (This also explains why τ can't be just a scalar.) Imposing this tracelessness through a Lagrange multiplier λ, we can write the (Wicked rotated) continuum action as

$$S = \int \frac{d^2\sigma}{2\pi}\{-iP^m \cdot \partial_m X + \tfrac{1}{2}\tau_{mn}(P^m \cdot P^n + \lambda g g^{mn}) + \sqrt{-g}[\mu + (ln\ \kappa)\tfrac{1}{2}R]\}$$

Thus τ acts as a kind of second worldsheet metric. However, since Schwinger parameters are positive, τ_{mn} must be positive definite, and thus a Euclidean metric. This also implies that g_{mn} must be Minkowskian, to be consistent with the tracelessness condition. Note that if we set λ equal to a constant, and ignore the positivity condition on τ, then eliminating τ by the equation of motion from varying g_{mn} reproduces the usual string action, where we can identify $\alpha' = \mu/\langle\lambda\rangle$. This indicates a possible approximation scheme.

The two components of τ that survive this tracelessness condition correspond to the two lightlike directions defined by g_{mn}: If we use a "zweibein", defined as usual by $g_{mn} = -e_{(m}{}^{+}e_{n)}{}^{-}$, to flatten the indices on τ, then the Lagrange multiplier constraint can be solved by simply setting $\tau_{+-} = 0$. The action is then

$$S = \int \frac{d^2\sigma}{2\pi}\sqrt{-g}[iP_{\pm} \cdot e_{\mp}{}^{m}\partial_m X + \tfrac{1}{2}\tau_{\pm\pm}P_{\mp} \cdot P_{\mp} + \mu + (ln\ \kappa)\tfrac{1}{2}R]$$

Back on the lattice, this implies that the directions chosen by the propagators (links) on which P is defined are lightlike. Thus, the matrix model defined by this theory should have only 4-point vertices, with the four propagators coming from any vertex forming the worldsheet lightcone at that point on the worldsheet. The field theory action is thus

$$S_2 = N\ tr \int \frac{d^D x}{(2\pi)^{D/2}}(-\tfrac{1}{4}\phi\Box\phi - G\tfrac{1}{4}\phi^4)$$

For $D = 4$, this action describes an asymptotically free theory, "wrong-sign" ϕ^4 theory.

Unlike conventional strings, the QCD string has critical dimension D=4 for renormalizability. (In conventional strings all momentum integrals are Gaussian and thus converge.) Another reason for D=4 is T-duality: T-duality interchanges the positions of the vertices with the momenta of the loops. This is clear from our discussion of the classical mechanics of Feynman diagrams in subsection VC8, if we note that the procedure we used there to translate from coordinates to loop momenta is exactly the random lattice version of the T-duality transformation performed in subsection XIB4 (with $\tilde{X}$ as the loop momenta). Thus, invariance of a string theory under T-duality must include invariance of the propagators of the underlying field theory under Fourier transformation. This is trivial for conventional strings, since the Fourier transform of a Gaussian is a Gaussian. However, by dimensional analysis (or explicit evaluation: see exercise VIIB4.2), we see that the Fourier transform of $1/p^2$ is $1/x^2$ *only* in D=4: T-duality implies both D=4 and masslessness. Furthermore, we can look at interactions by considering the simplest case: The flat worldsheet is represented by a regular, flat lattice. For ϕ^4 theory we have the usual square lattice, which is self-dual under switching vertices with loops (T-duality). On the other hand, triangular and hexagonal lattices, corresponding to ϕ^6 and ϕ^3 theory, are dual to each other (i.e., ϕ^n is dual to $\phi^{2n/(n-2)}$, as follows from geometry). Thus T-duality also implies the ϕ^4 interaction.

Exercise XIC3.1

Let's examine T-duality for the random lattice more carefully:

a Repeat the T-duality transformation of subsection XIB4, but for the QCD string (see subsection VC8), without a background ($M^{mn} = \eta^{mn}$). Show that invariance under $X^m \to \tilde{X}_m$ requires that the matrix τ also be replaced by its inverse, with some factors of the 2D ϵ tensor. (λ also transforms; you can avoid this complication by using the zweibein form of the action.)

b Write the massless scalar propagator in momentum space of arbitrary dimension D as an exponential using a Schwinger parameter τ. Show that after T-duality — Fourier transformation combined with $\tau \to 1/\tau$ (which leaves the exponent invariant) — a τ-dependent "measure" factor is introduced, *except* for D=4.

REFERENCES

1 K.G. Wilson, *Phys. Rev.* **D10** (1974) 2445:
QCD lattice.

2 H.B. Nielsen and M. Ninomiya, *Nucl. Phys.* **B185** (1981) 20, **195** (1982) 541, **193** (1981) 173, *Phys. Lett.* **105B** (1981) 219.

3 P.H. Ginsparg and K.G. Wilson, *Phys. Rev.* **D25** (1982) 2649;
P. Hasenfratz, hep-lat/9709110, *Nucl. Phys. Proc. Suppl.* **63A-C** (1998) 53, hep-lat/9802007, *Nucl. Phys.* **B525** (1998) 401, P. Hasenfratz, V. Laliena, and F. Niedermayer, hep-lat/9801021, *Phys. Lett.* **427B** (1998) 125;
H. Neuberger, hep-lat/9707022, *Phys. Lett.* **417B** (1998) 141, hep-lat/9801031, *Phys. Lett.* **427B** (1998) 353;
M. Lüscher, hep-lat/9802011, *Phys. Lett.* **428B** (1998) 342:
avoiding Nielsen-Ninomiya, while preserving chiral invariance.

4 Feynman and Hibbs, *loc. cit.* (VA).

5 C. Destri and H.J. de Vega, *Nucl. Phys.* **B290** (1987) 363, *Phys. Lett.* **201B** (1988) 261, *J. Phys. A* **22** (1987) 1329:
regular (square) 2D lightcone lattice.

6 M. Creutz, *Phys. Rev.* **D21** (1980) 2308;
E. Marinari, G. Parisi, and C. Rebbi, *Phys. Rev. Lett.* **47** (1981) 1795:
computer calculations with lattice QCD.

7 E.T. Tomboulis, *Phys. Rev.* **D25** (1982) 606, *Phys. Rev. Lett.* **50** (1983) 885:
deconfinement can occur only at zero coupling.

8 't Hooft, *loc. cit.* (VIIC, ref. 3):
convergence of singularities in the complex coupling plane at zero coupling.

9 H.J. Rothe, *Lattice gauge theories, an introduction* (World Scientific, 1992).

10 H.B. Nielsen and P. Olesen, *Phys. Lett.* **32B** (1970) 203;
D.B. Fairlie and H.B. Nielsen, *Nucl. Phys.* **B20** (1970) 637;
B. Sakita and M.A. Virasoro, *Phys. Rev. Lett.* **24** (1970) 1146:
worldsheet lattice as Feynman diagrams.

11 F. David, *Nucl. Phys.* **B257** [FS14] (1985) 543;
V.A. Kazakov, I.K. Kostov, and A.A. Migdal, *Phys. Lett.* **157B** (1985) 295;
J. Ambjørn, B. Durhuus, and J. Fröhlich, *Nucl. Phys.* **B257** (1985) 433:
integration over worldsheet metric as sum over Feynman diagrams.

12 M.R. Douglas and S.H. Shenker, *Nucl. Phys.* **B335** (1990) 635;
D.J. Gross and A.A. Migdal, *Phys. Rev. Lett.* **64** (1990) 127;
E. Brézin and V.A. Kazakov, *Phys. Lett.* **236B** (1990) 144:
continuum limit for worldsheet lattice with metric and 1/N expansion.

13 Veneziano, *loc. cit.* (XIA);
V. Alessandrini, D. Amati, and B. Morel, *Nuo. Cim.* **7A** (1971) 797;
D.J. Gross and P.F. Mende, *Phys. Lett.* **197B** (1987) 129, *Nucl. Phys.* **B303** (1988) 407;
D.J. Gross and J.L. Mañes, *Nucl. Phys.* **B326** (1989) 73:
Gaussian behavior of string amplitudes.

14 P.A. Collins and R.W. Tucker, *Nucl. Phys.* **B112** (1976) 150;
B. de Wit, M. Lüscher, and H. Nicolai, *Nucl. Phys.* **B305** (1988) 545:
problems quantizing membranes.

15 C.M. Hull and P.K. Townsend, hep-th/9410167, *Nucl. Phys.* **B438** (1995) 109:
the 11th dimension arises nonperturbatively in string theory.

16 Siegel, *loc. cit.* (VC):
QCD string with Schwinger parameters as second worldsheet metric.

17 A.M. Polyakov, *Nucl. Phys.* **B268** (1986) 406:
two-metric formulation of usual string.

XII. MECHANICS

String theories describe particles of arbitrarily large spins: So far in this text we have concentrated on lower spins, but we can describe (at least) free gauge-invariant actions for arbitrary spins based on quantum mechanical BRST.

Gauge invariance is required in field theory to manifest Lorentz invariance. The basic problem is that a four-vector wave function cannot have the obvious Minkowski inner product, since the time component would have a minus sign in its normalization, resulting in negative probability. In the classical action there is a gauge invariance that allows the time component to be dropped from the action. However, such gauges destroy manifest Lorentz invariance, since a three-vector cannot represent Lorentz transformations in a local way. More useful gauges keep all components of the four-vector, while also introducing scalar fermionic "ghosts" to cancel the effects of the bad part of the four-vector. A certain symmetry between the bosonic and fermionic unphysical degrees of freedom is needed to enforce this cancelation: It is the field theoretic version of the BRST symmetry discussed in section VIA.

Another complication is that gauge transformations do not allow the elimination of traces in a simple way: Although it is Lorentz covariant to constrain a tensor to vanish when a pair of its vector indices is contracted, this interferes with gauge invariance in interacting theories, such as gravity. A related complication is massive theories, which can't always be described simply by adding mass terms to massless theories.

There is a simple solution to all these problems, which determines the free part of the action for any theory. (Interactions are a separate problem.) This method automatically introduces all the correct fields, including ghosts, for any massless or massive theory. It also gives a simple universal expression for the BRST symmetry that cancels unphysical modes, as well as providing a simple proof that these modes disappear in the lightcone gauge. The method is based on the idea of introducing extra fermionic dimensions to spacetime that are unphysical (unlike superspace for supersymmetry), which cancel unphysical degrees of freedom associated with the time dimension.

Although for most purposes the only spins of fundamental particles relevant in field theory for are 0, 1/2, 1, 3/2 (maybe), and 2, and these few cases can be studied separately, in this chapter we'll analyze all free theories because:

(1) The ultimate theory of particles may require them;

(2) some of the theories presently under most active investigation (such as strings and membranes) require them;

(3) many observed, though perhaps not fundamental, particles have higher spin; and

(4) a better understanding of field theory can be obtained by determining exactly which properties all fields have in common as well as how they differ.

A. OSp(1,1|2)

This construction involves the introduction of spacetime symmetries that are not manifest on the physical coordinates. An important analog is the conformal group in D dimensions, which acts nonlinearly on the usual D spacetime coordinates, but can be represented linearly on D+2 coordinates, since the group is SO(D,2). As described in subsection IA6 for spin 0, 1 space and 1 time coordinate can be eliminated, so that SO(D,2) is still represented, but SO(D−1,1) is the largest orthogonal group that is still manifest. We have also seen that in the lightcone gauge this manifest symmetry is reduced again in the same way, leaving SO(D−2). In our case the relevant group is OSp(D,2|2), the natural generalization of the orthogonal group to D space, 2 time, and 2 anticommuting dimensions. This allows rotations between timelike and fermionic directions, eventually resulting in their cancelation.

1. Lightcone

We saw in subsection IIB3 how the single equation of motion $S_a{}^b \partial_b + w\partial_a = 0$, applied to field strengths, universally described all spins in all dimensions, for free, massless particles. (A possible exception is the spinless case, where we need $\Box = 0$, which is redundant otherwise. However, we can use the universal field equation even in that case if we use the vector field strength formulation of spin 0.) One way we solved this equation was to perform a unitary transformation. We can use the same unitary transformation, plus the constraints, to simplify the Lorentz generators. To further simplify matters, we can use the constraint $\Box = 0$, solved for ∂^-, to choose the gauge $x^+ = 0$, which is equivalent to working in the Schrödinger picture (no time dependence for operators). The procedure is thus: (1) Start with the manifest (antihermitian) representation of the Lorentz generators,

$$J^{ab} = x^{[a}\partial^{b]} + S^{ab}$$

(2) Apply the transformation

$$J \to UJU^{-1}, \qquad ln\ U = S^{+i}\frac{\partial^i}{\partial^+}$$

which eliminates the only S^{+i} term, in J^{+i} (while complicating J^{-i}). (3) Finally, apply the constraints, which have already been transformed by this same transformation,

$$\square = 0 \quad (\Rightarrow \ \textit{gauge} \ x^+ = 0), \qquad S^{ab}\partial_b + w\partial^a = 0 \to S^{+-} - w = S^{i-} = 0$$

Our Lorentz generators are then

$$J^{+i} = -x^i\partial^+, \qquad J^{+-} = -x^-\partial^+ + w, \qquad J^{ij} = x^{[i}\partial^{j]} + S^{ij}$$

$$J^{i-} = -x^-\partial^i + \frac{1}{\partial^+}[\tfrac{1}{2}x^i(\partial^j)^2 + S^{ij}\partial_j + w\partial^i]$$

These generators satisfy the pseudo(anti)hermiticity condition

$$J^{ab\dagger}(\partial^+)^{1-2w} = -(\partial^+)^{1-2w}J^{ab}$$

This means that the Hilbert-space metric needs a factor of $(\partial^+)^{1-2w}$. This is related to the fact that the w terms can be eliminated by a *non*unitary transformation with the appropriate power of ∂^+: As part of step 2, we could have applied a second transformation

$$U_2 = (\partial^+)^{S^{-+}}$$

with the result of eliminating all S^{+-} terms, so the redundant constraint $S^{+-} = w$ would not have been needed, so w would not appear. U_2 is in fact just the transformation that takes the surviving independent part of the field strength $F^{+\cdots+i\cdots j}$ to the lightcone gauge field $A^{i\cdots j}$, taking us from the original constrained field strengths to unconstrained gauge fields. In any case, we generally choose $w = 0$ for bosons.

We previously applied dimensional reduction to the field equations for the field strengths, to obtain the equations for the massive, free theories from the massless ones. The same methods can be applied to the Lorentz (or Poincaré) generators. (The Lorentz generators will be used later to find the BRST operator, to obtain the field equations in terms of the gauge fields, and the action. For that purpose the dimensional reduction can be performed at any stage in the derivation.) We thus find the general result

$$J^{+i} = -x^i\partial^+, \qquad J^{+-} = -x^-\partial^+ + w, \qquad J^{ij} = x^{[i}\partial^{j]} + S^{ij}$$

$$J^{i-} = -x^-\partial^i + \frac{1}{\partial^+}\{\tfrac{1}{2}x^i[(\partial^j)^2 - m^2] + S^{ij}\partial_j + S^{i-1}im + w\partial^i\}$$

(The -1 is still an index, and should not be confused with an inverse.)

Note that S^{-i} and S^{-+} were eliminated (after the unitary transformation) by the constraints, and that S^{+i} just dropped out. (In other words, $S^{+i} = 0$ was the gauge choice for the constraint $S^{-i} = 0$.) This leaves only S^{ij} (and S^{i-1} in the massive case), whose representation is that of the highest-weight part of the original field strength. However, we can more simply choose the representation of S^{ij} as our starting point, since it is just the transverse part of the gauge field; it defines the representation of the Poincaré group. We therefore have an explicit construction of the generators of the Poincaré group, for arbitrary representations, defined on just the physical degrees of freedom, given directly by the little group SO(D−2) spin generators S^{ij} (or SO(D−1) generators S^{ij} and S^{i-1} in the massive case) that identify the representation. For example, in D=4, SO(2) has just one generator, the helicity, so for any state of a given helicity we know the action of the Poincaré generators.

Exercise XIIA1.1

Check that this lightcone representation of the Lorentz generators satisfies the correct commutation relations.

Classical free field theory is easy to define in the lightcone, since solving the constraints in the lightcone formalism has picked out just the physical components, so the only remaining constraint is the Klein-Gordon equation. Thus, the kinetic term for *any* massless bosonic field is simply $-\frac{1}{2}\phi(\frac{1}{2}\Box)\phi$, where $\frac{1}{2}\Box = -\partial^+\partial^- + \frac{1}{2}(\partial^i)^2$, and ∂^- is considered the time derivative. (In general, the kinetic operator for a massless boson is some second-order differential operator, which reduces to $\Box$ on the physical components.) For fermions we have instead $\Box/\partial^+$, since we must then have an odd number of derivatives to avoid getting a trivial result after integration by parts. (For a boson, $\phi\partial\phi = \partial(\frac{1}{2}\phi^2)$, for a fermion $\psi\partial\partial\psi = \partial(\psi\partial\psi) - (\partial\psi)^2 = \partial(\psi\partial\psi)$ by anticommutativity. In general, the kinetic operator for a massless fermion is some first-order differential operator, which reduces to $\Box/\partial^+$ after eliminating auxiliary fields.)

This quantum mechanical representation of the Lorentz generators has a simple translation into classical field theory, in terms of field theory Poisson brackets. The definition of Poisson brackets in lightcone quantum field theory follows directly from the action: Defining as usual the canonical momentum π as (minus, in our conventions) the variation of the Lagrangian with respect to the time derivative ∂_+ $(= -\partial^-)$ of the variable ϕ, we find the fundamental bracket for bosons

$$\pi = -\partial^+\phi \quad \Rightarrow \quad [\phi(x^-, x^i), \phi(x'^-, x'^i)] = i\frac{1}{\partial^+}(2\pi)^{D/2}\delta(x^- - x'^-)\delta^{D-2}(x^i - x'^i)$$

(Note that the ∂^+ was essential for the antisymmetry of the bracket. We can also evaluate its inverse as an integral, so $\frac{1}{\partial^+}\delta(x^- - x'^-) = \frac{1}{2}\epsilon(x^- - x'^-)$. For fermions we have instead an anticommutator and no $1/\partial^+$.) We then find that any quantum mechanical group generator J (including internal symmetries) can be represented in field theoretic form as

$$\mathcal{J} = \int \frac{dx^- d^{D-2}x^i}{(2\pi)^{D/2}}\ \tfrac{1}{2}\phi\partial^+ J\phi = i\tfrac{1}{2}\langle\phi|J\phi\rangle$$

where we have used the relativistic inner product of subsection VB2, but for a lightlike hypersurface: For positive-energy solutions

$$\langle 1|2\rangle = \int \frac{dx^- d^{D-2}x^i}{(2\pi)^{D/2}}\ \psi_1{}^*\tfrac{1}{2}(-i)\overleftrightarrow{\partial}^+\psi_2$$

Note that the free Poincaré generators are local in this form, from cancelation of ∂^+'s. In interacting theories, the generator $\mathcal{J}^{-i}$, as well as the translation generator $\mathcal{P}^-$, which is also the Hamiltonian, gets additional terms higher-order in the fields. In this manner, relativistic quantum field theory can be quantized in a way that more resembles nonrelativistic field theory than in non-lightcone methods, since $\Box$ is quadratic in the usual time derivative ∂_0. We won't consider lightcone quantum field theory further; however, in the following sections we'll use this construction to derive free gauge theory and its covariant quantization, in a way that we'll generalize straightforwardly to interactions. Thus, the same construction directly gives the formulation of free representations of the Poincaré group, from field strengths to transverse fields to covariant gauge fields.

2. Algebra

From the definition of the graded determinant in terms of Gaussian integrals (see subsection IIC3), we see that anticommuting coordinates act like negative dimensions: For example, $sdet(kI) = k^{a-b}$ for a commuting and b anticommuting dimensions. Thus, if we add equal numbers of commuting and anticommuting dimensions, they effectively cancel. Here we'll do the same for theories with spin, which allows the restoration of manifest Lorentz covariance to lightcone theories: Adding 2 commuting and 2 anticommuting dimensions to SO(D−2) gives OSp(D−1,1|2) (see also subsection IIC3), which has an SO(D−1,1) subgroup.

We have seen that quantum field theory requires unphysical anticommuting fields to cancel the commuting unphysical fields introduced by using gauges that do not eliminate longitudinal polarizations. For example, the gauge field for electromagnetism has only D−2 components in the lightcone gauge, but needs to keep all D components to maintain manifest Lorentz covariance; this requires 2 "ghosts" to cancel the 2 extra components of the gauge field. The general result, at least for bosonic gauge fields, is to produce fields that form representations of OSp(D−1,1|2), including gauge fields and ghosts. Furthermore, by adding 2 anticommuting dimensions the BRST transformations that relate the ghosts to the longitudinal degrees of freedom can be introduced in a natural way, as *translations* in the new coordinates. The result is that OSp(D−1,1|2) multiplets are automatic, and gauge fixing receives a geometric interpretation. In this section we'll see that an even more natural interpretation of these BRST transformations is as *rotations* of the anticommuting coordinates, and that they not only make gauge fixing to the simplest Lorentz covariant gauge trivial, but also give a simple derivation of the gauge *invariant* action itself. (The two points of view are related in that translations can be considered as part of "conformal rotations", as we saw in subsection IA6.)

The basic idea is very simple: Take the lightcone representation of the Poincaré generators, found in the previous subsection, and extend the SO(D−2) indices and representations to OSp(D−1,1|2) ones (including appropriate signs for the grading). Conversely, we can begin the construction with the "conformal" group OSp(D+1,3|2), find the equations of motion for the "Poincaré" group OSp(D,2|2), and solve them for the "lightcone group" OSp(D−1,1|2). So we can use the same expressions for the generators, but now the "transverse" OSp(D−1,1|2) index is

$$i = (a, \alpha)$$

where a is a D-component index of SO(D−1,1) and α is a 2-component index of Sp(2). The OSp(D−1,1|2) metric is

$$\eta^{ij} = (\eta^{ab}, C^{\alpha\beta})$$

Furthermore, we divide up the full OSp(D,2|2) index as

$$(\pm, i) = (A, a), \qquad A = (\pm, \alpha)$$

We now interpret the SO(D−1,1) subgroup that acts on the a index as the usual physical one, since the generators take the usual covariant form (because all the transverse generators are linear). The orthogonal subgroup OSp(1,1|2) that acts on the A index, and leaves the a index alone, is then interpreted as a symmetry group of the unphysical degrees of freedom, an extension of BRST. However, the generators with − indices are nonlinear, since the ± indices are no longer independent from the rest. (The + were gauged away, the − were fixed by equations of motion.) As a result, they act on transverse indices in a nontrivial way.

$$\begin{array}{rl} \text{longitudinal, nonlinear SO(1,1):} & \left.\begin{array}{l} \pm \\ \begin{cases} \alpha \\ a \end{cases} \end{array}\right. \end{array}$$

$$\begin{array}{rll} \text{longitudinal, nonlinear SO(1,1):} & \pm & \multirow{2}{*}{$\Big\} = A : \text{ghost OSp(1,1|2)}$} \\ \text{transverse, manifest OSp(D−1,1|2):}\ i = \Big\{ & \alpha & \\ & a & : \text{Lorentz SO(D−1,1)} \end{array}$$

Since the OSp(1,1|2) generators act only in the unphysical directions, all physical states should be singlets (with respect to the cohomology) under this symmetry. This is clear from the original construction: We started with linear generators for OSp(D,2|2) (and translations and dilatations for (D,2|2) dimensions), applied the equations of motion in terms of them, and now we apply the OSp(1,1|2) singlet condition last. If we had instead applied the OSp(1,1|2) singlet condition first to the (D,2|2) dimensional space, we would have gotten the usual (D−1,1|0) dimensional space, and finally applying the equations of motion would have given us the lightcone results of subsection IIB3.

manifest symmetry:

$\searrow$ field equations (fix $\pm$)
$\swarrow$ BRST singlets (fix A)
$\nearrow$ add 2+2 (extend $i \to (a,\alpha)$)

$$\begin{array}{ccccc} & & \text{OSp(D,2|2)} & & \\ & \swarrow\!\!\!\!\nearrow & & \searrow & \\ \text{SO(D}-1,1) & & \Longrightarrow & & \text{OSp(D}-1,1|2) \\ \text{field strengths} & & & & \text{gauge fields/BRST} \\ & \searrow & & \swarrow\!\!\!\!\nearrow & \\ & & \text{SO(D}-2) & & \\ & & \text{lightcone} & & \end{array}$$

Explicitly, the OSp(1,1|2) generators are (choosing $w = 0$)

$$J^{+\alpha} = -x^\alpha \partial^+, \qquad J^{-+} = x^- \partial^+, \qquad J^{\alpha\beta} = x^{(\alpha}\partial^{\beta)} + S^{\alpha\beta}$$

$$J^{\alpha-} = -x^- \partial^\alpha + \frac{1}{\partial^+}[\tfrac{1}{2}x^\alpha(\Box - m^2 + \partial^\beta \partial_\beta) + S^{\alpha b}\partial_b + S^{\alpha-1}im + S^{\alpha\beta}\partial_\beta]$$

while the SO(D−1,1) generators take their usual manifest form

$$J^{ab} = x^{[a}\partial^{b]} + S^{ab}$$

Exercise XIIA2.1

Write the general commutation relations of OSp(D−1,1|2). Specialize to the case OSp(1,1|2), in lightcone notation. Show that this representation satisfies them, paying special attention to signs. (Use the OSp(D−1,1|2) commutators for the S's.)

We can also add a "nonminimal" part to the general "minimal" part of the OSp(1,1|2) algebra we have already derived, in the sense that the two parts commute and separately satisfy the commutation relations:

$$J^{AB} \to J^{AB} + \tilde{S}^{AB}$$

(This is similar to adding spin pieces to orbital in the absence of constraints relating them.) The simplest choices are to choose this new part to be just quadratic in new, "nonminimal" coordinates and momenta. It will prove convenient to perform some transformations $J \to UJU^{-1}$ that make the OSp(1,1|2) generators more similar to what they were before adding the spin parts. We therefore make two consecutive transformations:

$$J \to U_2 U_1 J U_1^{-1} U_2^{-1} : \qquad U_1 = e^{\tilde{S}^{+\alpha}\partial_\alpha/\partial^+}, \qquad U_2 = (\partial_+)^{\tilde{S}^{-+}}$$

to return J^{-+} and $J^{+\alpha}$ to their previous forms. In fact, these are just the OSp(1,1|2) version of the same transformations we used in the previous subsection to remove S^{+i} and S^{+-}. The result is

$$J^{+\alpha} = -x^\alpha \partial^+, \qquad J^{-+} = x^- \partial^+, \qquad J^{\alpha\beta} = x^{(\alpha}\partial^{\beta)} + \hat{S}^{\alpha\beta}$$

$$J^{\alpha -} = -x^- \partial^\alpha + \frac{1}{\partial^+}[x^\alpha(-K + \tfrac{1}{2}\partial^\beta\partial_\beta) + \mathcal{Q}^\alpha + \hat{S}^{\alpha\beta}\partial_\beta]$$

Here we have

$$\boxed{K = -\tfrac{1}{2}(\Box - m^2), \quad \hat{S}^{\alpha\beta} = S^{\alpha\beta} + \tilde{S}^{\alpha\beta}, \quad \mathcal{Q}^\alpha = \tilde{S}^{\alpha -} + S^{\alpha b}\partial_b + S^{\alpha -1} im - \tilde{S}^{+\alpha}K}$$

but more generally we can satisfy the commutation relations by requiring only that K, $\mathcal{Q}_\alpha$, and $\hat{S}_{\alpha\beta}$ are independent of the unphysical coordinates x^- and x^α and their momenta, and satisfy that their only nontrivial commutators are

$$\{\mathcal{Q}^\alpha, \mathcal{Q}^\beta\} = 2K\hat{S}^{\alpha\beta}, \qquad [\hat{S}^{\alpha\beta}, \mathcal{Q}^\gamma] = \mathcal{Q}^{(\alpha}C^{\beta)\gamma}, \qquad [\hat{S}_{\alpha\beta}, \hat{S}^{\gamma\delta}] = \delta_{(\alpha}^{(\gamma}\hat{S}_{\beta)}{}^{\delta)}$$

U_2 is nonunitary, which makes $\tilde{S}^{\pm\alpha}$ hermitian (rather than antihermitian) after the transformation, requiring a modification of the usual representation for $\tilde{S}$. The usual representation can also be used by introducing an i into the transformation, which gives $\tilde{S}^{\pm\alpha}$ a factor of $\pm i$ in $\mathcal{Q}^\alpha$. However, this i can be removed by the same method used in subsection IIB4 to remove i's associated with the index -1, only now it is applied to both the $+$ and $-$ indices.

3. Action

We saw in the previous subsection that physical states are singlets under the OSp(1,1|2) BRST symmetry. It was introduced in a trivial way, but became nontrivial after solving the equations of motion; on the other hand, applying the singlet condition first reproduced the usual lightcone analysis. Reversing the order of applying the two conditions has the advantage of allowing the physical state condition to be expressed as a single equation, which can be derived from an action.

There are two ways of doing this: One is to use this algebra to generalize to gauge fields the first-quantized BRST as applied to field theory in subsection VIA3. Because of their quantum mechanical origin, the gauge-invariant $\Phi Q \Phi$ actions directly give a form suitable for choosing the Fermi-Feynman gauge, where the kinetic operator is simply $\square - m^2$. However, this is somewhat unusual for fermions, whose simplest field equation is first-order. (But it is useful for supersymmetry, where bosons and fermions are treated symmetrically.) As a result, this approach gives actions for fermions with an infinite number of auxiliary and ghost fields. The most convenient way to discover the usual finite-component gauge-invariant first-order actions hidden there is by performing an appropriate unitary transformation, after which this action (for bosons or fermions) appears as the sum of three terms: the usual gauge-invariant action, a term giving the usual second-quantized BRST transformations, and a non-derivative term that would be considered nonminimal under second-quantized BRST. This approach will be described in detail in the following sections.

The other way is to define a δ function in the generators of the group OSp(1,1|2), and use it as the kinetic operator for the action:

$$S = -\int dx \ dx^- d^2x^\alpha \ \tfrac{1}{4}\Phi \partial^+ \delta(J_{AB})\Phi$$

where the integration is over all the coordinates appearing in the OSp(1,1|2) generators. (The dx part is the usual $d^D x/(2\pi)^{D/2}$.) ∂^+ comes from the usual relativistic inner product; it is also a "measure" factor, which is a consequence of our using generators satisfying the pseudohermiticity condition

$$J^\dagger{}_{AB}\partial^+ = -\partial^+ J_{AB}$$

Equivalently, we could redefine $\Phi \to (\partial^+)^{-1/2}\Phi$, $J \to (\partial^+)^{-1/2} J (\partial^+)^{1/2}$ (assuming $\partial^+ \neq 0$, as usual in lightcone formalisms) to eliminate it and restore hermiticity. (This would only affect $J^{-+} \to J^{-+} - \frac{1}{2}$, $J^{\alpha -} \to J^{\alpha -} + \partial^\alpha / 2\partial^+$, making hermitian the terms $\frac{1}{2}\{x^-, \partial^+\}$ and $\frac{1}{4}\{x^\alpha, \partial^\beta \partial_\beta\}/\partial^+$.) Because of the δ function, this action has the gauge invariance

$$\delta\Phi = \tfrac{1}{2} J^{BA} \Lambda_{AB}$$

Thus, the field equations and gauge invariance reduce Φ to states in the OSp(1,1|2) cohomology.

More explicitly, the δ function can be written as

$$\begin{aligned}\partial^+ \delta(J_{AB}) &= \partial^+ \delta(J_{\alpha\beta}{}^2)\delta(J^{-+})\delta^2(J^{+\alpha})\delta^2(J^{\alpha -}) \\ &= \delta(x^-)\delta^2(x^\alpha)\delta(\hat{S}_{\alpha\beta}{}^2)\partial^{+2} J^{-\alpha 2}\end{aligned}$$

where we have used

$$J^{-+}\delta(J^{-+}) = \delta(J^{-+})J^{-+} = 0 \quad \Rightarrow \quad \delta(J^{-+}) = \frac{1}{\partial^+}\delta(x^-)$$

(There is freedom in ordering of the original δ functions: Reordering of any two δ's produces terms that are killed by the other δ's.) The $\delta(\hat{S}_{\alpha\beta}{}^2)$ can be interpreted as a Kronecker δ_{s0} in the Sp(2) "spin" s:

$$-\tfrac{1}{2}\hat{S}^{\alpha\beta}\hat{S}_{\alpha\beta} = 4s(s+1)$$

(remember $\hat{S}^{\alpha\beta}$ is antihermitian, and $i\hat{S}^{\oplus\ominus}$ is always integer while s can be half-integer). The rest of the explicit δ's are Dirac δ's in the unphysical coordinates, which can therefore be trivially integrated out, leaving:

$$\boxed{S = \int dx\ L_{gi}, \qquad L_{gi} = \tfrac{1}{2}\phi K_{gi}\phi, \qquad K_{gi} = \tfrac{1}{2}(-\Box + m^2 + \tfrac{1}{2}\mathcal{Q}^\alpha\mathcal{Q}_\alpha)}$$

where ϕ is Φ evaluated at $x^\alpha = x^- = s = 0$. Furthermore, the remaining gauge invariance is

$$\boxed{\delta\phi = \delta_{s0}\tfrac{1}{2}\mathcal{Q}^\alpha\Lambda_\alpha}$$

from $J^{\alpha-}$, since $J^{\alpha\beta}$, $J^{+\alpha}$, and J^{-+} have been used to gauge to $s = x^\alpha = x^- = 0$, respectively.

Exercise XIIA3.1

Show explicitly that this action is invariant under the OSp(1,1|2) gauge transformations. (Hint: Use the same method as exercise XA2.1.)

4. Spinors

As we saw in subsection VIA3, the BRST algebra for the (Dirac) spinor requires nonminimal terms. For the general case of fermions we add these terms in the general way described in subsection XIIA2, choosing them in terms of a (second) set of OSp(1,1|2) γ matrices:

$$\tilde{S}^{AB} = -\tfrac{1}{2}[\tilde{\gamma}^A,\tilde{\gamma}^B\}, \qquad \{\tilde{\gamma}^A,\tilde{\gamma}^B] = -\eta^{AB}$$

where $\tilde{\gamma}^A = (-\kappa, -\mu; \tilde{\xi}, i\tilde{\zeta})$ in the notation of subsection VIA3. In particular, we find

$$\hat{S}^{\alpha\beta} = S^{\alpha\beta} - \tfrac{1}{2}\tilde{\gamma}^{(\alpha}\tilde{\gamma}^{\beta)}$$

$$\mathcal{Q}^\alpha = S^{\alpha b}\partial_b + S^{\alpha -1}im + \tilde{\gamma}^\alpha[\tilde{\gamma}^- + \tilde{\gamma}^+\tfrac{1}{2}(\Box - m^2)]$$

The next step for general massless fermions (and similarly for the massive case) is to apply

$$\mathcal{Q}^2 = \tfrac{1}{2}S^{\alpha a}S_\alpha{}^b\partial_a\partial_b - (\tilde{\gamma}^- - \tilde{\gamma}^+K)\tilde{\gamma}^\alpha S_\alpha{}^a\partial_a - \tfrac{1}{2}K$$

where we have used

$$\tilde{\gamma}^\alpha\tilde{\gamma}_\alpha = C^{\alpha\beta}\tfrac{1}{2}[\tilde{\gamma}_\beta, \tilde{\gamma}_\alpha] = 1$$

At this point we note that the gauge invariance generated by $\mathcal{Q}^\alpha$, for gauge parameter $\Lambda_\alpha = \tilde{\gamma}_\alpha\Lambda$, includes a term $\tilde{\gamma}^-\Lambda$ that allows us to choose the gauge

$$\tilde{\gamma}^+\phi = 0$$

One way to think of this is to treat $\tilde{\gamma}^+$ as an anticommuting coordinate and $\tilde{\gamma}^-$ as its derivative; another way is to treat them as 2×2 matrices. Alternatively, we can unitarily transform the action to contain just the $\tilde{\gamma}^-$ term of the operator: From the discussion of 2D γ matrices of subsection VIIIA7 we find, including that part of the spinor metric,

$$\tilde{\Upsilon}\tilde{\gamma}^- = -\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$$

then acts as a projection operator. Either way, the net result is to reduce the action to, now restoring the mass,

$$\boxed{S_f = \int dx\ L_{gi,f}, \qquad L_{gi,f} = \tfrac{1}{2}\hat{\phi}K_{gi,f}\hat{\phi}, \qquad K_{gi,f} = \tfrac{1}{2}\tilde{\gamma}^\alpha(S_\alpha{}^a\partial_a + S_{\alpha-1}im)}$$

where $\hat{\phi}$ is ϕ with the $\tilde{\gamma}^\pm$-dependence eliminated (the top component in the above matrix representation). Thus, $\hat{\phi}$ differs from the bosonic case in that it not only depends on x^a and is a representation of S^{ij}, but is also a representation of $\tilde{\gamma}^\alpha$, which appears in $\hat{S}^{\alpha\beta}$ to define $s = 0$.

The only type of representation we have missed in this analysis is self-dual antisymmetric tensors. In terms of field strengths, these satisfy

$$F_{a_1 \ldots a_{D/2}} = \pm\tfrac{1}{(D/2)!}\epsilon_{a_1 \ldots a_{D/2} b_1 \ldots b_{D/2}} F^{b_1 \ldots b_{D/2}}$$

which is consistent, with Lorentz metric, if $D/2$ is odd (as seen from applying the ϵ tensor twice). A similar condition holds for the gauge field in the lightcone gauge (with a (D−2)-dimensional ϵ tensor). Because of the ϵ tensor, this condition can't be described by adding extra dimensions to the lightcone. However, the direct product of two spinors contains all antisymmetric tensors, and the rank $D/2$ one can be picked out by an appropriate OSp invariant constraint. The self-dual part of this tensor comes from the direct product of chiral spinors.

Exercise XIIA4.1

We now consider this construction in more detail:

a Derive the generalization of γ_{-1} to OSp γ matrices, anticommuting with both the fermionic and bosonic γ's, $\{\gamma_{-1}, \gamma_A\} = 0$. We can use the usual product for the fermionic γ's, but obviously the bosonic ones will need something different. (Hint: For each pair of fermionic or bosonic γ's there is a Klein factor, as in exercise IA2.3e; for the fermionic γ's the exponential is equal to the usual product.)

b In twice-odd dimensions, consider the direct product of two spinors by representing the OSp spin operators as a sum in terms of the two different sets of OSp γ matrices acting on the two different spinor indices. Define the U(1) (O(2)) symmetry that mixes the two γ matrices by taking linear complex combinations of the γ matrices to form fermionic creation and annihilation operators, so the OSp-invariant U(1) generator is $a^{\dagger A} a_A$. Show that the eigenvalues of this generator pick out the different Lorentz representations. These can be made irreducible by including γ_{-1} projections. Using explicit U(1) and (both) γ_{-1} projectors in the action, show that self-dual tensors can be described. (Note: This description contains an infinite number of auxiliary fields.)

5. Examples

The OSp(1,1|2) method is thus an efficient method for finding gauge-invariant actions (though not so useful for gauge fixing). We begin with examples of massless bosons, for which the gauge-invariant kinetic operator is

$$K_{gi} = \tfrac{1}{2}(-\Box + \tfrac{1}{2}\mathcal{Q}^\alpha \mathcal{Q}_\alpha), \qquad \mathcal{Q}^\alpha = S^{\alpha a}\partial_a$$

The scalar is a trivial example; the simplest nontrivial example is the massless vector: In terms of the basis $|^i\rangle$ for an OSp(D$-$1,1|2) vector (D-vector plus 2 ghosts), normalized to

$$\langle^i|^j\rangle = \eta^{ij} \quad \Rightarrow \quad \langle^a|^b\rangle = \eta^{ab}, \qquad \langle^\alpha|^\beta\rangle = C^{\alpha\beta}$$

we can write the OSp(D$-$1,1|2) generators as

$$S_{(1)}^{ij} = |^{[i}\rangle\langle^{j)}|$$

The Sp(2)-singlet field is then (dropping the $|^\alpha\rangle$ term)

$$\phi = |^a\rangle A_a(x)$$

We then have

$$\mathcal{Q}^\alpha = (|^\alpha\rangle\langle^a| - |^a\rangle\langle^\alpha|)\partial_a \quad\Rightarrow\quad \mathcal{Q}^2 \equiv \tfrac{1}{2}\mathcal{Q}^\alpha\mathcal{Q}_\alpha = |^a\rangle\langle^b|\partial_a\partial_b - \tfrac{1}{2}|^\alpha\rangle\langle_\alpha|\Box$$

$$\Rightarrow\quad L_{gi(1)} = \tfrac{1}{8}(F_{ab})^2$$

and for the gauge invariance

$$\delta\phi = \delta_{s0}\tfrac{1}{2}\mathcal{Q}^\alpha\Lambda_\alpha \quad\Rightarrow\quad \Lambda_\alpha = |_\alpha\rangle\lambda(x) \quad\Rightarrow\quad \delta A_a = \partial_a\lambda$$

A more complicated example is the graviton (massless spin 2): We write the field, a graded symmetric, traceless OSp(D−1,1|2) tensor, in terms of the direct product of two vectors, with basis $|^i\rangle|^j\rangle$. The spin operators are thus

$$S^{ij} = S^{ij}_{(1)} \otimes I_{(1)} + I_{(1)} \otimes S^{ij}_{(1)}$$

where the first factor in each term acts on the first factor in $|^i\rangle|^j\rangle$, etc.; $I_{(1)}$ is the spin-1 identity. The $s = 0$ part of the field is then

$$\phi = |^i\rangle|^j\rangle h_{ji}, \qquad h^i{}_i = h^a{}_a + h^\alpha{}_\alpha = 0 \quad\Rightarrow\quad \phi = (|^a\rangle|^b\rangle + \tfrac{1}{2}|^\alpha\rangle|_\alpha\rangle\eta^{ab})h_{ab}$$

where h_{ab} includes its trace. The rest is straightforward algebra; we use identities such as:

$$\mathcal{Q}^2 = \mathcal{Q}^2_{(1)} \otimes I_{(1)} + I_{(1)} \otimes \mathcal{Q}^2_{(1)} + \mathcal{Q}^\alpha_{(1)} \otimes \mathcal{Q}_{(1)\alpha}$$

$$\mathcal{Q}^\alpha_{(1)} \otimes \mathcal{Q}_{(1)\alpha} = (|^\alpha\rangle|_\alpha\rangle\langle^a|\langle^b| + |^a\rangle|^b\rangle\langle^\alpha|\langle_\alpha|)\partial_a\partial_b$$

$$(\langle^\alpha|\langle_\alpha|)(|^\beta\rangle|_\beta\rangle) = -\langle^\alpha|^\beta\rangle\langle_\alpha|_\beta\rangle = -2$$

where $\mathcal{Q}^2_{(1)}$ was evaluated above, and in the last identity we used the fact that $|^\alpha\rangle$ is anticommuting. The final result is then

$$L_{gi(2)} = -\tfrac{1}{4}[h^{ab}\Box h_{ab} + 2(\partial^b h_{ab})^2 - h^a{}_a\Box h^b{}_b + 2h^a{}_a\partial^b\partial^c h_{bc}]$$

in agreement with subsection IXB1. The original OSp(1,1|2) gauge invariance reduces to

$$\Lambda_\alpha = (|_\alpha\rangle|^a\rangle + |^a\rangle|_\alpha\rangle)\lambda_a \quad\Rightarrow\quad \delta h_{ab} = \partial_{(a}\lambda_{b)}$$

Exercise XIIA5.1

Consider a (D−2)-rank antisymmetric tensor (i.e., totally antisymmetric in D−2 indices in D dimensions; see exercises IIB2.1 and VIIIA8.2, and subsection XA3).

a Show from a lightcone analysis that it is equivalent to a scalar. Derive the gauge-invariant action using OSp methods. Find the gauge transformations and field strength.

b Find a first-order form for the action, (auxiliary field)^2 + (auxiliary field) $\times$ (field strength). Show that eliminating the gauge field as a Lagrange multipler results in the action for a scalar. Show that switching between scalar and antisymmetric tensor is equivalent to switching field equation and constraint for the field strength.

c Find the description for the massive case by dimensional reduction.

Exercise XIIA5.2

Consider a tensor totally symmetric in its vector indices. In the lightcone gauge, the irreducible tensor is traceless. Show that, upon covariantization, the field appearing in the gauge-invariant action satisfies a double-tracelessness condition (or equivalently the fields appearing there are the totally traceless tensor and another totally tracelsss tensor with two less indices).

For massless fermions we saw

$$K_{gi,f} = \tfrac{1}{2}\tilde{\gamma}^\alpha S_\alpha{}^a \partial_a$$

The next step is to use the fact that arbitrary fermionic representations are constructed by taking the γ-traceless piece of the direct product of a (Dirac) spinor with an irreducible bosonic representation. (Just as an irreducible bosonic representation of an orthogonal group is found by taking the direct product of vectors, choosing an appropriate symmetry, as described by the Young tableau, and requiring the trace in any two vector indices to vanish; here we also require that using a γ matrix to contract the spinor index with any vector index also vanishes. Of course, simpler methods can be used for SO(3,1), but we need methods that apply to all dimensions, so they can be applied to orthosymplectic groups.) We then can write

$$S^{ij} = \check{S}^{ij} - \tfrac{1}{2}[\gamma^i, \gamma^j\} \quad \Rightarrow \quad \hat{S}^{\alpha\beta} = \check{S}^{\alpha\beta} - a^{\dagger(\alpha}a^{\beta)}$$

where $\check{S}^{ij}$ is the part of the spin acting on just the vector indices, and we have combined γ^α and $\tilde{\gamma}^\alpha$ into creation and annihilation operators, as in subsection VIA3:

$$a^\alpha = \tfrac{1}{\sqrt{2}}(\gamma^\alpha + i\tilde{\gamma}^\alpha), \quad a^{\dagger\alpha} = \tfrac{1}{\sqrt{2}}(\gamma^\alpha - i\tilde{\gamma}^\alpha); \qquad [a_\alpha, a^{\dagger\beta}] = -\delta_\alpha^\beta$$

For spin 1/2 $\check{S} = 0$, δ_{s0} projects to the ground-state of the oscillators, and we immediately find

$$S_\alpha{}^a = -\gamma_\alpha\gamma^a \quad \Rightarrow \quad L_{gi(1/2)} = \tfrac{1}{4}\hat{\phi}\gamma^a i\partial_a\hat{\phi}$$

where we have used

$$\tilde{\gamma}^\alpha \gamma_\alpha = i\tfrac{1}{2}(a^{\dagger\alpha}a_\alpha - a^\alpha a^\dagger{}_\alpha) = i(a^{\dagger\alpha}a_\alpha - 1) = -i(N+1)$$

and this "N" counts the $a^{\dagger\alpha}$ excitation level. (Note that the Hilbert-space inner product between the spinors includes the usual factor of γ_0.) A less trivial case is spin 3/2: Now

$$\phi = |^i\rangle\phi_i, \qquad S^{\alpha a} = -\gamma^\alpha\gamma^a + |^{[\alpha}\rangle\langle^{a]}|$$

where ϕ_i has an explicit vector index, and an implicit spinor index. Then from γ-tracelessness (for irreducibility) we have for the Sp(2) singlets

$$\gamma^i\phi_i = 0 \quad \Rightarrow \quad \phi_\alpha = -\gamma_\alpha\gamma^a\phi_a \quad \Rightarrow \quad a_\alpha\phi_a = 0, \quad \phi_\alpha = -\tfrac{1}{\sqrt{2}}a^\dagger{}_\alpha\gamma^a\phi_a$$

After a little algebra, using identities such as

$$\tfrac{1}{6}\gamma^{[a}\gamma^b\gamma^{c]} = \gamma^a\gamma^b\gamma^c + \tfrac{1}{2}(\eta^{b(a}\gamma^{c)} - \eta^{ac}\gamma^b)$$

we find

$$L_{gi(3/2)} = -\tfrac{1}{12}\hat{\phi}_a\gamma^{[a}\gamma^b\gamma^{c]}i\partial_b\hat{\phi}_c$$

From inspection, or from $\delta\phi = \delta_{s0}\frac{1}{2}\mathcal{Q}^\alpha\Lambda_\alpha$, we find the gauge invariance

$$\delta\hat{\phi}_a = \partial_a\lambda$$

Exercise XIIA5.3

Let's now examine some massive examples:

a Find the gauge-invariant actions for massive spin 2 and spin 3/2 by dimensional reduction of the massless cases.

b Note that for the spin-2 case the part of the mass term quadratic in h is proportional to $-h_{[a}{}^a h_{b]}{}^b$. More generally, we might have expected $(h_{ab})^2 + k(h_a{}^a)^2$ for arbitrary constant k, since the first part gives mass to the physical (transverse, traceless) part of h, while the second term affects only the unphysical pieces. Find the Stückelberg terms generated from this generalized mass term by the linearized gauge invariance. Looking at just the terms quadratic in the Stückelberg vector, what is special about $k = -1$, and why do other values of k give ghosts? (Hint: Compare gauge-fixed electromagnetism.)

REFERENCES

1 K. Bardakçi and M.B. Halpern, *Phys. Rev.* **176** (1968) 1686:
lightcone representations of Poincaré algebra.

2 M. Fierz and W. Pauli, *Proc. Roy. Soc.* **A173** (1939) 211;
S.J. Chang, *Phys. Rev.* **161** (1967) 1308;
L.P.S. Singh and C.R. Hagen, *Phys. Rev.* **D9** (1974) 898;
C. Fronsdal, *Phys. Rev.* **D18** (1978) 3624;
T. Curtright, *Phys. Lett.* **85B** (1979) 219;
B. deWit and D.Z. Freedman, *Phys. Rev.* **D21** (1980) 358;
T. Curtright and P.G.O. Freund, *Nucl. Phys.* **B172** (1980) 413;
T. Curtright, *Phys. Lett.* **165B** (1985) 304:
covariant actions for totally symmetric representations of all (4D) spin.
3 W. Siegel and B. Zwiebach, *Nucl. Phys.* **B282** (1987) 125:
lightcone representations of Poincaré algebra, and covariant actions, for arbitrary spin and dimension.
4 G. Parisi and N. Sourlas, *Phys. Rev. Lett.* **43** (1979) 744:
cancelation of extra dimensions.
5 R. Delbourgo and P.D. Jarvis, *J. Phys.* **A15** (1982) 611;
J. Thierry-Mieg, *Nucl. Phys.* **B261** (1985) 55;
J.A. Henderson and P.D. Jarvis, *Class. and Quant. Grav.* **3** (1986) L61:
ghosts from extra dimensions for indices.
6 W. Siegel, *Nucl. Phys.* **B284** (1987) 632:
1st-quantized BRST for fermions.
7 N. Berkovits, hep-th/9607070, *Phys. Lett.* **388B** (1996) 743:
1st-quantized BRST for self-dual tensors.
8 Siegel, *loc. cit.*
9 Fierz and Pauli, *loc. cit.* (ref. 2 above):
mass term for spin 2.

B. IGL(1)

Although the OSp(1,1|2) method is the simplest way to derive general free gauge-invariant actions, it does not yield a simple method for gauge fixing, even though the Hilbert space contains exactly the right set of ghosts. We now describe a related method that is slightly less useful for finding gauge-invariant actions (it includes redundant auxiliary fields), but allows gauges to be fixed easily.

1. Algebra

For this method we use a subset of the OSp(1,1|2) constraints, and show they are sufficient. A simple analog is SU(2): To find SU(2) singlets, it's sufficient to look for states that are killed by both T_3 and the raising operator $T_1 + iT_2$. This approach gives a formalism that turns out to be easier to generalize to interacting theories, as well as allowing a simple gauge-fixing procedure. We first divide up the Sp(2) indices as $\alpha = (\oplus, \ominus)$ (not to be confused with $\pm$). We then make a similarity transformation that simplifies some of the generators (while making others more complicated):

$$J \to UJU^{-1}: \qquad U = (\partial^+)^{iJ^{\oplus\ominus}}$$

which changes the Hilbert-space metric (and corresponding hermiticity conditions) to

$$\Upsilon = U^\dagger U = (-1)^{iJ^{\oplus\ominus}}, \qquad \langle \Psi | \Xi \rangle = \int \Psi^\dagger \Upsilon \Xi$$

This simplifies (looking at the massless case without $\tilde{S}^{AB}$ for simplicity)

$$J^{+\ominus} \to \frac{1}{\partial^+} J^{+\ominus} = -x^\ominus$$

$$J^{-+} \to J^{-+} - iJ^{\oplus\ominus}, \qquad J^{\oplus\ominus} \to J^{\oplus\ominus}$$

$$J^{\oplus -} \to J^{\oplus -}\partial^+ + iJ^{\oplus\ominus}\partial^\oplus = (-x^-\partial^+)\partial^\oplus + \tfrac{1}{2}x^\oplus \Box + S^{\oplus a}\partial_a + S^{\oplus\oplus}\partial_\oplus$$

(We use the same conventions for raising and lowering Sp(2) indices as for SU(2) in subsection IIA4 and SL(2,C) in subsection IIA5.) These four generators form the subgroup GL(1|1) of OSp(1,1|2) (=SL(1|2)): We can write the generators as $J_I{}^J$, where $I = (+, \oplus)$. In subsection XIIB4 we'll see that the singlets of this GL(1|1) are the same as those of OSp(1,1|2).

On the other hand, because of the simplified form of these generators, it's easy to see how to reduce the group even further: Applying some of the constraints on wave functions/fields to the right,

$$J^{+\ominus} = -x^\ominus = 0, \qquad J^{-+} + iJ^{\oplus\ominus} = x^-\partial^+ = 0$$

$$\Rightarrow \quad J^{\oplus\ominus} = -ix^{\oplus}\partial_{\oplus} + i + S^{\oplus\ominus}, \qquad J^{\oplus -} = \tfrac{1}{2}x^{\oplus}\Box + S^{\oplus a}\partial_a + S^{\oplus\oplus}\partial_{\oplus}$$

(Of course, this further reduction could have been performed even without the transformation.) We are now left with the group IGL(1), with just $J^{\oplus\ominus}$ and $J^{\oplus -}$ as generators. ($J^{\oplus -}$ acts as translations for the GL(1) generator $J^{\oplus\ominus}$.) Also, we have reduced the unphysical coordinates to just $x^{\oplus}$. We now simplify notation by relabeling

$$c = x^{\oplus}, \qquad b = \partial_{\oplus}, \qquad S^3 = iS^{\oplus\ominus}, \qquad J = iJ^{\oplus\ominus} + 1, \qquad Q = J^{\oplus -}$$

$$\Rightarrow \quad \boxed{J = cb + S^3, \qquad Q = \tfrac{1}{2}c\Box + S^{\oplus a}\partial_a + S^{\oplus\oplus}b}$$

J and Q are versions of the ghost-number and BRST operators introduced in subsection VIA1. The net result for obtaining these IGL(1) generators from the original OSp(1,1|2) generators can also be stated as

$$J = iJ^{\oplus\ominus}|_{\partial^{\oplus}=0,\partial^{+}=1}, \qquad Q = J^{\oplus -}|_{\partial^{\oplus}=0,\partial^{+}=1}$$

where $\partial^{\oplus} = 0$ can be regarded as a gauge condition for the constraint $x^{\ominus} = 0$, and $\partial^{+} = 1$ for $x^{-}\partial^{+}$ ($\sim \partial/\partial(ln\ \partial^{+})) = 0$. The IGL(1) algebra is

$$[J, Q] = Q, \qquad Q^2 = \tfrac{1}{2}\{Q, Q\} = 0$$

Exercise XIIB1.1

Show that any IGL(1) subgroup of OSp(1,1|2) ($J = iJ^{\oplus\ominus}$, $Q = J^{\oplus -}$) satisfies these commutation relations. Check that this final representation of the IGL(1) algebra satisfies them.

Exercise XIIB1.2

Using the results of subsection XIIA2,

a Give expressions for Q and J in terms of (c, b,) K, $\mathcal{Q}^{\alpha}$, and $\hat{S}^{\alpha\beta}$.

b Derive $\{\mathcal{Q}^{\alpha}, \mathcal{Q}^{\beta}\}$, assuming only $Q^2 = 0$ and the previous results for $[\hat{S}, \hat{S}]$ and $[\hat{S}, \mathcal{Q}]$.

2. Inner product

The new inner product can be derived by the same steps: Starting with the lightcone inner product of subsection XIIA1, we add extra dimensions to get the OSp(D,2|2) inner product. We next drop dependence on $x^{\ominus}$, which will be eliminated in the IGL(1) formalism. Then we perform the transformation with $(\partial^{+})^{iJ^{\oplus\ominus}} = (\partial^{+})^{J-1}$ used to simplify the BRST operator. This acts on both fields in the inner product; applying integration by parts turns one such factor into $(-\partial^{+})^{-J}$. The net effect is that it cancels the ∂^+ in the Hilbert-space metric, which allows us to drop the x^- integration, and it introduces a factor of $(-1)^J$.

Rather than defining a Hilbert-space inner product, which is sesquilinear, it is slightly more convenient to define a symplectic inner product, replacing the Hermitian conjugate of the wave function/state on the left with the transpose, in analogy to an ordinary vector inner product. The inner product is then

$$\boxed{\langle \Psi | \Xi \rangle = -i(-1)^{\Psi} \int dx\ dc\ \Psi^T(x,c)(-1)^J \Xi(x,c)}$$

A Hilbert-space inner product can then be defined simply as $\langle \Psi^* | \Xi \rangle$ (where $(\Psi^*)^T = \Psi^\dagger$). We have included a sign factor corresponding to what would be obtained if the dc integration were moved to the symmetric position between the two wave functions: By $(-1)^\Psi$ we mean take $\Psi = 0$ in the exponent if Ψ is bosonic and 1 if Ψ is fermionic. We can make this manifest by defining

$$\Psi(x,c) = \langle x, c | \Psi \rangle \quad \Rightarrow \quad I = -i \int dx\ dc\ |x,c\rangle (-1)^J \langle x, c|$$

which allows the inner product to be evaluated between $\langle \Psi |$ and $| \Xi \rangle$ by inserting this form for the "identity" I.

Exercise XIIB2.1

Work out the inner product for the vector field in terms of all of the components (both expanding over c and separating physical and ghost parts of the OSp(D−1,1|2) vector).

As a result, any commuting or anticommuting constant factor "a" can be moved out of the inner product from the left or right in the usual way:

$$\langle \Psi | \Xi a \rangle = \langle \Psi | \Xi \rangle a, \qquad \langle a\Psi | \Xi \rangle = a \langle \Psi | \Xi \rangle$$

As a consequence of the anticommutativity of the integration measure, we have

$$(-1)^{\langle \Psi | \Xi \rangle} = (-1)^{\Psi + \Xi + 1}$$

meaning that the statistics of the inner product is the opposite of an ordinary product; we can think of "$|$" as a fermion.

Because of the change in metric from the lightcone, the IGL(1) generators now satisfy

$$J^T = 1 - J, \qquad Q^T = Q$$

where the constant comes from dropping the extra coordinates, and the transpose "^{T}" indicates integration by parts (the usual transpose in the infinite-matrix representation of operators):

$$\int \varPsi \mathcal{O} \varXi = \int (-1)^{\varPsi \mathcal{O}} (\mathcal{O}^T \varPsi) \varXi$$

Before our transformations the generators all satisfied $G^T = -G$; now they are pseudoantisymmetric with respect to the metric $(-1)^J$, up to the constant:

$$\langle \varPsi | \mathcal{O} \varXi \rangle = \langle \tilde{\mathcal{O}} \varPsi | \varXi \rangle \quad \Rightarrow \quad \tilde{\mathcal{O}} = (-1)^J \mathcal{O}^T (-1)^J$$

$$\tilde{J} = 1 - J, \qquad \tilde{Q} = -Q$$

From $J^T = 1 - J$ also follows the symmetry property of the inner product:

$$\langle \varPsi | \varXi \rangle = (-1)^{(\varPsi+1)(\varXi+1)} \langle \varXi | \varPsi \rangle$$

This can be interpreted as antisymmetry once the anticommutativity of the "$|$" (metric) is taken into account.

The hermiticity conditions that follow from the change from the lightcone are

$$J^\dagger = 1 - J, \qquad Q^\dagger = Q$$

Before all generators were antihermitian; now they are pseudoantihermitian, up to a constant:

$$\hat{\mathcal{O}} = (-1)^J \mathcal{O}^\dagger (-1)^J$$

$$\Rightarrow \quad \hat{J} = 1 - J, \qquad \hat{Q} = -Q$$

The factor of i in the inner product compensates for the funny hermiticity of $(-1)^J$. We then find the usual hermiticity condition for a vector inner product,

$$\langle \varPsi | \varXi \rangle^* = \langle \varXi^\dagger | \varPsi^\dagger \rangle$$

3. Action

As explained in subsection VIA1, we are interested in states in the cohomology of the BRST operator Q, which means states satisfying

$$Q\varPhi = 0, \qquad \delta\varPhi = iQ\varLambda$$

In particular, the physical states are states in the cohomology of Q at ghost number $J = 0$. However, now $Q\varPhi = 0$ is the wave equation (as in subsection VIA3), and $Q\varLambda$

contains the usual gauge transformations (as in the OSp(1,1|2) action of the previous section).

The free gauge-invariant action for an arbitrary field theory is then

$$S_0 = -(-1)^{\Phi} \int dx\ dc\ \tfrac{1}{2}\Phi^T \delta_{J1} Q\Phi = -(-1)^{\Phi} \int dx\ dc\ \tfrac{1}{2}\Phi^T Q\delta_{J0}\Phi$$

for a real column-vector field Φ. (Complex fields can be decomposed into their real and imaginary parts. For relating to quantum mechanics, we will usually consider the column-vector in our OSp Hilbert-space notation.) This action gives $Q\Phi = 0$ as the equations of motion, and has $\delta\Phi = Q\Lambda$ as a gauge invariance, so the solutions are the cohomology of Q. The projector δ_{J0} is a Kronecker δ restricting Φ to vanishing ghost-number (and thus $Q\Phi$ to ghost number 1, since $[J, Q] = Q$). As we'll see in the next section, this projector is redundant: The states in the cohomology with nonvanishing momentum automatically have vanishing ghost-number, and the states with nonvanishing ghost-number are needed for gauge fixing. The projection is useful only for eliminating states that are redundant for discussing gauge invariance; we'll drop it for the remainder of this section. The "complete" free action is then

$$\boxed{S_0 = -(-1)^{\Phi} \int dx\ dc\ \tfrac{1}{2}\Phi^T (-1)^{J-1} Q\Phi = \tfrac{1}{2}\langle\Phi|iQ\Phi\rangle = S_0{}^{\dagger}}$$

where we have included the inner-product metric. This is just the translation of the free BRST operator from first- to second-quantized form, as for the lightcone in subsection XIIA1.

We now consider some simple examples, to see how this method reproduces the usual results. The simplest example is the scalar: As shown in subsection VIA3,

$$Q = c\tfrac{1}{2}(\Box - m^2) \quad \Rightarrow \quad -\int dc\ \tfrac{1}{2}\Phi(-1)^{J-1}Q\Phi = -\tfrac{1}{2}\phi\tfrac{1}{2}(\Box - m^2)\phi$$

without restrictions to vanishing ghost number, unitary transformations, gauge fixing, etc. Thus the scalar is in no way a gauge field: The kinetic operator follows from simple kinematic considerations.

The fundamental example of a gauge theory is a vector: It is the defining representation of the Lorentz group, and of the extended Lorentz group we used to define the BRST operator. In the rest of this chapter we will see from its action most of the general properties of gauge theories: ghosts, gauge invariance, BRST transformations of the fields, the gauge-invariant action, gauge fixing, backgrounds, mass, etc. In addition to the equations given for this case in subsection XIIA5, we will use

$$\langle{}^{\oplus}|{}^{\ominus}\rangle = -\langle{}^{\ominus}|{}^{\oplus}\rangle = i$$

The BRST and ghost-number operators are

$$Q = \tfrac{1}{2}c\Box + (|^{\oplus}\rangle\langle^{a}| - |^{a}\rangle\langle^{\oplus}|)\partial_a + 2|^{\oplus}\rangle\langle^{\oplus}|b, \qquad J = cb + i(|^{\oplus}\rangle\langle^{\ominus}| + |^{\ominus}\rangle\langle^{\oplus}|)$$

The field is real; it can't be called hermitian, since it is a column vector, but each component of that vector is hermitian. (This is the same as reality, but for anti-commuting objects reality includes extra signs that are defined to be exactly those coming from hermitian conjugation.) Thus,

$$\Phi = \Phi^* = (|^{a}\rangle A_a - i|^{\ominus}\rangle C + i|^{\oplus}\rangle\tilde{C}) - ic(|^{a}\rangle\check{A}_a - |^{\ominus}\rangle\check{\tilde{C}} - |^{\oplus}\rangle\check{C})$$

$$\Phi^T = \Phi^\dagger = (A_a\langle^{a}| + iC\langle^{\ominus}| - i\tilde{C}\langle^{\oplus}|) + (\check{A}_a\langle^{a}| - \check{\tilde{C}}\langle^{\ominus}| - \check{C}\langle^{\oplus}|)ic$$

where we denote the "antifields" (those at order c) by a " ˇ ".

The BRST transformations of the fields can be found by comparing terms in Φ and $Q\Phi$: If we define a second-quantized BRST operator $\hat{Q}$ such that $Q\Phi = \hat{Q}\Phi$, but $\hat{Q}$ acts only on the fields while Q (as usual) acts on $|^{i}\rangle$ and the coordinates (x and c), then

$$\Phi = |^{i}\rangle(\phi_i - ic\psi_i) \quad \Rightarrow \quad Q\Phi = \hat{Q}\Phi = |^{i}\rangle[(-1)^i\hat{Q}\phi_i - ic(-1)^{i+1}\hat{Q}\psi_i]$$

In other words, we compare terms in Φ and $Q\Phi$, and throw in a minus sign for transformations of fermions. (So, e.g., "$\hat{Q}A_a$" is the coefficient of $|^{a}\rangle$ in $Q\Phi$.) Dropping the " ˆ " on Q, the result is

$$QA_a = -\partial_a C, \quad QC = 0, \quad Q\tilde{C} = -2i(\check{\tilde{C}} - \tfrac{1}{2}\partial\cdot A)$$

$$Q\check{A}_a = -i(\tfrac{1}{2}\Box A_a - \partial_a\check{\tilde{C}}), \quad Q\check{C} = \tfrac{1}{2}\Box\tilde{C} + \partial\cdot\check{A}, \quad Q\check{\tilde{C}} = -\tfrac{1}{2}\Box C$$

Note that although Q is hermitian, it is antihermitian with respect to the inner-product metric $(-1)^J$, as expected from our convention of using antihermitian generators for spacetime symmetries. (The same extra sign for hermiticity vs. pseudohermiticity, also because of ghosts introduced by relativistic quantum mechanics, occurs for the spatial Dirac matrices γ^i: see subsection XC2.) As a result, our transformations agree with those of subsection VIA4. However, while the first-quantized Abelian transformations also agree with those of subsections VIA1-3, the second-quantized nonabelian transformations will have the extra i demonstrated in subsection VIA4, following from the i introduced in the inner-product metric in the previous subsection. (This minor yet annoying factor will be further discussed in section XIIC when we relate first- and second-quantized BRST.)

The Lagrangian then can be expanded as (after some integration by parts)

$$L_0 = -\int dc\ \tfrac{1}{2}\Phi^T(-1)^{J-1}Q\Phi = -i\tfrac{1}{2}(A\cdot Q\check{A} - CQ\check{C} - \tilde{C}Q\check{\tilde{C}} + \check{A}\cdot QA - \check{C}QC - \check{\tilde{C}}Q\tilde{C})$$

$$= \tfrac{1}{8}(F_{ab})^2 + (\check{\tilde{C}} - \tfrac{1}{2}\partial\cdot A)^2 - i\tilde{C}\tfrac{1}{2}\Box C + i\check{A}\cdot\partial C$$

where we have used the transpose of the field on the left of $\Phi Q\Phi$. To find the gauge-invariant action, we can evaluate it by keeping just the (anti)fields with vanishing ghost number (A_a and $\check{\tilde{C}}$), and then eliminate the remaining antifields by their equations of motion:

$$L \to \tfrac{1}{8}(F_{ab})^2 + (\check{\tilde{C}} - \tfrac{1}{2}\partial\cdot A)^2 \quad \Rightarrow \quad L_{gi} = \tfrac{1}{8}(F_{ab})^2$$

Exercise XIIB3.1

Consider the example of the second-rank antisymmetric tensor (see exercises IIB2.1, VIIIA8.2, and XIIA5.1, and subsection XA3):

a Construct the states by direct product of two vectors. Decompose into fields plus antifields, physical plus ghost: In particular, note the Sp(2) representation of each SO(D−1,1) representation.

b Find the BRST transformations for all the (anti)fields. In particular, note that the tensor transforms into vector ghosts (as expected from the gauge invariance), which themselves transform into scalar ghosts ("ghosts for ghosts").

c Graph all the states for s (of the Sp(2) $S^{\alpha\beta}$) vs. J, and indicate there how BRST relates them.

d Find the gauge-invariant action from $\Phi Q\Phi$.

e Generalize to arbitrary-rank antisymmetric tensors. Compare the results of exercise XIIA5.1a.

4. Solution

The identity of the cohomology and the physical states can be proven most easily by making a unitary ("gauge") transformation to the "lightcone gauge":

$$(Q,J) \to U(Q,J)U^{-1}: \qquad U = e^{(S^{+i}\partial_i + S^{+\oplus}b)/\partial^+}$$

which simplifies Q while leaving J unchanged:

$$Q \to \tfrac{1}{2}c\Box - S^{\oplus-}\partial^+, \qquad J \to cb + S^3$$

These are the usual lightcone indices $\pm$ of any D-vector, not to be confused with the $\pm$ used earlier when reducing from D+2 bosonic dimensions. Except for the extension to include the $\oplus$ index, this is the same transformation used in subsection IIB3 (and XIIA1).

This makes the generators separable, allowing us to treat the two terms in Q and J independently. Specifically, if we integrate the action over c,

$$\Phi = \phi - ic\psi \quad \Rightarrow \quad -\int dc\ \tfrac{1}{2}\Phi^T(-1)^{J-1}Q\Phi = -\tfrac{1}{4}\phi^T(-1)^{S^3}\Box\phi - i\psi^T(-1)^{S^3-1}S^{\oplus-}\partial^+\phi$$

Then ψ is just a Lagrange multiplier enforcing the algebraic constraint $S^{\oplus-}\phi = 0$ (ignoring ∂^+, which we always assume is invertible in the lightcone approach), leaving just the Klein-Gordon term for the part of ϕ that satisfies the constraint. We also have the gauge invariance

$$\Lambda = \lambda + c\chi, \quad \delta\Phi = iQ\Lambda \quad \Rightarrow \quad \delta\phi = -iS^{\oplus-}\partial^+\lambda, \quad \delta\psi = -S^{\oplus-}\partial^+\chi - \tfrac{1}{2}\Box\lambda$$

so we can shoose the gauge where ϕ is restricted (algebraically) to be in the cohomology of $S^{\oplus-}$.

To solve for the cohomology of $S^{\oplus-}$ it is sufficient to consider the reducible representations formed by direct products of vectors (for bosons), or the direct products of these with a single Dirac spinor (for fermions), since by definition the OSp(D$-$1,1|2) generators S^{ij} don't mix different irreducible OSp(D$-$1,1|2) representations. We'll show that this cohomology restricts any reducible OSp(D$-$1,1|2) representation to the corresponding reducible SO(D$-$2) lightcone representation, and therefore restricts any irreducible OSp(D$-$1,1|2) representation to the irreducible SO(D$-$2) representation from which it was derived. (Also, the irreducible representations in arbitrary dimensions are most conveniently found by such a construction, where reduction is performed by symmetrization and antisymmetrization and subtracting traces of vector indices, and in the fermionic case also subtracting gamma-matrix traces and using Majorana/Weyl projection.)

For bosons, we first consider the representation from which all the rest are constructed, the vector. Writing the basis for the vector states as $|^i\rangle$, where $S^{ij}|^k\rangle = |^{[i}\rangle\eta^{j)k}$, we find

$$S^{\oplus-} = 0 \quad \Rightarrow \quad \textit{not}\ |^+\rangle, |^\ominus\rangle$$
$$\delta = S^{\oplus-} \quad \Rightarrow \quad \textit{not}\ |^-\rangle, |^\oplus\rangle$$

This leaves only the transverse lightcone states, as advertised. For the direct product of an arbitrary number of vectors, we find the same result: The unphysical directions are eliminated from each vector in the product.

We might worry that extra states in the cohomology would arise from a cancelation of two terms, resulting from the action of the "$\oplus$" and "$-$" parts of $S^{\oplus-}$. Specifically, this could happen if we could separate out the supertraceless part of (the graded symmetric part of) the product of two OSp(1,1|2) vectors. (For example, for an SO(n) vector we can separate the traceless part of a symmetric tensor as $T_{ij} - \frac{1}{n}\delta_{ij}T_{kk}$.) However, this is not possible, since for OSp(1,1|2)

$$str(\delta_A^B) = 2 - 2 = 0$$

Explicitly, we can look at the two likely candidates for extra states in the cohomology,

$$|^+\rangle|^-\rangle \pm i|^\oplus\rangle|^\ominus\rangle$$

(and their transposes). But using

$$S^{\oplus-}|^+\rangle = -|^\oplus\rangle, \qquad S^{\oplus-}|^\ominus\rangle = -i|^-\rangle$$

for these states we find

$$S^{\oplus-}(|^+\rangle|^-\rangle + i|^\oplus\rangle|^\ominus\rangle) = -2|^\oplus\rangle|^-\rangle$$

$$S^{\oplus-}i|^+\rangle|^\ominus\rangle = |^+\rangle|^-\rangle - i|^\oplus\rangle|^\ominus\rangle$$

so neither state is in the cohomology. Note that we take $S^{\oplus-}$ to anticommute with $|^\alpha\rangle$; the states in the Hilbert space are assigned statistics. (This is the simplest way to allow a direct relation between wave functions and fields.)

Exercise XIIB4.1

Check this analysis for spin 3.

Note that in the Lagrangian $-\frac{1}{4}\phi^T\Box\phi - i\psi^T S^{\oplus-}\partial^+\phi$ the fields in ϕ that are nonzero when acted upon by $S^{\oplus-}$ are auxiliary, killed by the Lagrange multiplier ψ. On the other hand, the fields that are $S^{\oplus-}$ on something are pure gauge, and do not appear in the $\psi S^{\oplus-}\phi$ term because $S^{\oplus-}$ is nilpotent, while they drop out of the $\phi\Box\phi$ term because the fields multiplying them there are exactly the auxiliary ones that were killed by varying ψ. This follows from the fact that a field that is pure gauge with respect to $S^{\oplus-}$ has a nonvanishing inner product only with an auxiliary field, since $\phi_1 = S^{\oplus-}\lambda \Rightarrow \phi_2\phi_1 = \phi_2 S^{\oplus-}\lambda$. Equivalently, a field redefinition $\psi \to \psi + A\Box\phi$ can cancel any terms in $\phi\Box\phi$ where one ϕ is $S^{\oplus-}$ on something.

For the example of the vector, we have explicitly for the transformation to the lightcone

$$ln\ U = \frac{1}{\partial^+}\left[(|^+\rangle\langle^i| - |^i\rangle\langle^+|)\partial_i + (|^+\rangle\langle^\oplus| - |^\oplus\rangle\langle^+|)b\right]$$

under which the Lagrangian becomes

$$L \to L' = -\tfrac{1}{4}A \cdot \Box A - i\tilde{C}\tfrac{1}{2}\Box C + \check{\tilde{C}}\partial^+ A^- - i\check{A}^- \partial^+ C$$

The lightcone gauge transformations are

$$\delta A^+ = -\partial^+\lambda, \qquad \delta\check{\tilde{C}} = -\tfrac{1}{2}\Box\lambda$$

$$\delta\tilde{C} = -\partial^+\zeta, \qquad \delta\check{A}^- = -\tfrac{1}{2}\Box\zeta$$

$$\delta\check{C} = \tfrac{1}{2}\Box\xi^\ominus, \qquad \delta\check{A}^+ = \tfrac{1}{2}\Box\xi^+, \qquad \delta\check{A}^i = \tfrac{1}{2}\Box\xi^i$$

where "i" here refers to the transverse (D$-$2) components.

5. Spinors

In general we can add nonminimal terms $\tilde{S}^{AB}$ of subsection XIIA2: The easiest way is to add them as the last step, remembering that Q comes from $J^{\oplus-}$ and J from J^3; this yields

$$Q \to \tfrac{1}{2}c\Box + S^{\oplus a}\partial_a + S^{\oplus\oplus}b + \tilde{S}^{\oplus-}, \qquad J \to cb + S^3 + \tilde{S}^3$$

This result can also be seen from first-quantization of spin 1/2 (subsection VIA3). Alternatively, if we add $\tilde{S}$ at the beginning as in subsection XIIA2, performing the transformations given there, followed by the transformation

$$U = (\partial^+)^{iJ^{\oplus\ominus}}$$

of subsection XIIB1, where $J^{\oplus\ominus}$ itself now contains $\tilde{S}$ terms, we again find

$$J^{+\ominus} \to \frac{1}{\partial^+}J^{+\ominus} = -x^\ominus, \qquad J^{-+} \to J^{-+} - iJ^{\oplus\ominus}$$

$$J^{\oplus\ominus} \to J^{\oplus\ominus}, \qquad J^{\oplus-} \to J^{\oplus-}\partial^+ + iJ^{\oplus\ominus}\partial^\oplus$$

Adding a final transformation

$$U = e^{-\tilde{S}^{+\oplus}b}$$

(which actually undoes part of an earlier one), we again obtain the above result.

As in the previous subsection, we can also transform to the lightcone gauge, to find

$$Q \to \tfrac{1}{2}c\Box - S^{\oplus-}\partial^+ + \tilde{S}^{\oplus-}$$

When analyzing the BRST cohomology in the lightcone gauge, the effect of this nonminimal term is to replace (again ignoring the factor of $-\partial^+$)

$$S^{AB} \to \hat{S}^{AB} = S^{AB} + \tilde{S}^{AB}$$

(although actually only the $S^{\oplus-}$ and S^3 parts are used here).

As in subsection XIIA4, when treating fermions we choose the Dirac spinor representation of OSp(1,1|2) for $\tilde{S}^{AB}$. Thus, for the case of spin 1/2, where S^{AB} also is a Dirac spinor representation, $\hat{S}^{AB}$ is represented by the direct product of two OSp(1,1|2) spinors. We now use the harmonic oscillator interpretation of the ghost coordinates used in subsection XIIA5 (extending it trivially to the fermionic ones), which can be applied to arbitrary OSp groups:

$$S^{AB} = -\tfrac{1}{2}[\gamma^A, \gamma^B\}, \qquad \tilde{S}^{AB} = -\tfrac{1}{2}[\tilde{\gamma}^A, \tilde{\gamma}^B\}$$

$$a^A = \tfrac{1}{\sqrt{2}}(\gamma^A + i\tilde{\gamma}^A), \qquad a^{\dagger A} = \tfrac{1}{\sqrt{2}}(\gamma^A - i\tilde{\gamma}^A)$$

$$\Rightarrow \quad \{a^A, a^{\dagger B}] = -\eta^{AB}, \qquad \hat{S}^{AB} = -a^{\dagger [A} a^{B)}$$

By expanding about the oscillator vacuum, we find this representation of OSp(1,1|2) consists of the direct sum of totally (graded) antisymmetrized tensors: $|0\rangle, |^A\rangle = a^{\dagger A}|0\rangle, |^{[AB)}\rangle = a^{\dagger A} a^{\dagger B}|0\rangle, ...$. But we have already treated this case for the bosons, the result being that only the singlet (vacuum) survives. Of course, the complete spin representation is given by the direct product of the representation of the unphysical variables (γ^A, $\tilde{\gamma}^A$) and the physical ones, namely the transverse lightcone gamma matrices. Thus, the states in the cohomology are given by the direct product of all the lightcone states with the vacuum of the unphysical variables. To treat arbitrary fermions, generalization to direct products of the Dirac spinor with arbitrary numbers of vectors works the same way, since the spinor looks like the direct sum of parts of direct products of vectors as far as $\hat{S}^{AB}$ is concerned.

Exercise XIIB5.1

Work out the explicit Q and Hilbert space for spin 3/2.

6. Masses

As usual masses can be added by dimensional reduction: Our complete result for application to massless and massive, bosons and fermions is then

$$\boxed{Q = \tfrac{1}{2}c(\Box - m^2) + S^{\oplus a}\partial_a + S^{\oplus -1} im + S^{\oplus\oplus} b \quad (+\tilde{S}^{\oplus -}), \qquad J = cb + S^3 \quad (+\tilde{S}^3)}$$

with extra i's introduced implicitly by the procedure given in subsection IIB4.

For example, for the vector we have (the "Stückelberg formalism")

$$Q = \tfrac{1}{2}c(\Box - m^2) + (|^{\oplus}\rangle\langle^a| - |^a\rangle\langle^{\oplus}|)\partial_a + (|^{\oplus}\rangle\langle^{-1}| + |^{-1}\rangle\langle^{\oplus}|)m + 2|^{\oplus}\rangle\langle^{\oplus}|b$$

$$J = cb + i(|^{\oplus}\rangle\langle^{\ominus}| + |^{\ominus}\rangle\langle^{\oplus}|)$$

Compared to the massless case treated in subsections XIIB3-4, the corresponding field now has the extra terms

$$\Phi \to \Phi + |^{-1}\rangle\phi - ic|^{-1}\rangle\check{\phi}$$

giving the action

$$L_0 = L_{gi} + [\check{\tilde{C}} - \tfrac{1}{2}(\partial\cdot A + m\phi)]^2 - i\tilde{C}\tfrac{1}{2}(\Box - m^2)C + iC(\partial\cdot\check{A} + m\check{\phi})$$

$$L_{gi} = \tfrac{1}{8}(F_{ab})^2 + \tfrac{1}{4}(mA + \partial\phi)^2$$

For the spinor

$$Q = \tfrac{1}{2}c(\Box - m^2) - \gamma^{\oplus}(\not{\partial} - i\tfrac{m}{\sqrt{2}}) - (\gamma^{\oplus})^2 b - \tilde{\gamma}^{\oplus}\tilde{\gamma}^{-}$$

$$J = cb - i\gamma^{\oplus}\gamma^{\ominus} - i\tilde{\gamma}^{\oplus}\tilde{\gamma}^{\ominus}$$

where in the notation of subsection VIA3,

$$\gamma^{\oplus} = \xi, \qquad \gamma^{\ominus} = i\zeta, \qquad \tilde{\gamma}^{\oplus} = \tilde{\xi}, \qquad \tilde{\gamma}^{\ominus} = i\tilde{\zeta}, \qquad \tilde{\gamma}^{-} = -\mu$$

Exercise XIIB6.1

Use this method to work out the action and gauge transformations for massive spin 2.

7. Background fields

The coupling of external fields can be treated by suitable modification of the BRST operator. In terms of self-interacting field theories, this corresponds to writing the field as the sum of quantum and background fields, and keeping in the action only the terms quadratic in the quantum fields, as discussed for semiclassical expansions in subsection VA2 and for the background field method in subsection VIB8.

One interesting case is the coupling of an external vector gauge field. Clearly the spacetime derivatives in Q must be modified by the minimal coupling prescription $\partial \to \nabla = \partial + iA$, but dimensional analysis and Lorentz covariance also allow the addition of a nonminimal term proportional to $F^{ab}S_{ab}$ to $\Box$. With the appropriate coefficient, the general result is (see subsection VIIIA3)

$$Q_I = \tfrac{1}{2}c(\Box - iF^{ab}S_{ba}) + S^{\oplus a}\nabla_a + S^{\oplus\oplus}b$$

where $\Box$ is now the covariant ∇^2. (J is unchanged.)

In the case of spin 0, this modification is trivial. For spin 1/2, we substitute the graded generalization of the Dirac matrices, $S^{ij} = -\frac{1}{2}[\gamma^i, \gamma^j\}$, as discussed in subsection XIIA5. We then find $Q_I^2 = 0$ fixes the above coefficient of the nonminimal term, the same as from squaring $\gamma^a \nabla_a$. This follows from the simple factorization of S^{ij} in Q_I:

$$Q_I = -c(\gamma^a \nabla_a)^2 - \gamma^\oplus(\gamma^a \nabla_a) - (\gamma^\oplus)^2 b = -(\gamma^a \nabla_a + \gamma^\oplus b)c(\gamma^b \nabla_b + \gamma^\oplus b)$$

where we have neglected the $\tilde{S}^{\oplus -}$ term of subsection XIIB5, and used

$$[\gamma^a, c] = \{\gamma^\oplus, c\} = 0$$

In the spin-1 case, we find the interesting result that $Q_I^2 = 0$ requires not only the above coefficient for the nonminimal term (as expected from supersymmetry), but also that the background terms satisfy the field equation $\nabla_b F^{ab} = 0$. On the other hand, for spins >1, $Q_I^2 = 0$ implies $F_{ab} = 0$, so these spins can't couple minimally (at least in flat spaces). Similar remarks apply to coupling gravity (spin 2) to spins >2.

Exercise XIIB7.1

Check these statements for spin 1. Compare the analogous result for background fields in the field theoretic approach from exercise VIB8.2 for Yang-Mills for both the gauge transformations of the gauge-invariant action and the *field-theoretic* BRST transformations of the gauge-fixed action.

Exercise XIIB7.2

Show that electromagnetism can't couple minimally to (massless) higher spins (i.e., they can't have charge):

a Show this for the graviton by considering $Q_I^2 = 0$ for spin 2 (symmetric traceless OSp tensor) in an external vector field.

b Do the same for spin 3/2.

Another interesting feature of the spin-1 case is that we can define a "vacuum" state

$$|0\rangle = |^\ominus\rangle$$

which is in the free BRST cohomology of Q only at zero momentum (constant field), where Q simplifies to $S^{\oplus\oplus}b$ without background. However, this state has ghost number $J = -1$. In fact, it corresponds to the global part of the gauge invariance of the theory: Gauge parameters satisfying $Q\Lambda = 0$ have no effect in the free theory (where $\delta\Phi = iQ\Lambda$), but can act in the interacting theory: They do not contribute an inhomogeneous term to gauge transformations. However, gauge parameters of the form $\Lambda + Q\Xi$ have the same effect as Λ, up to trivial transformations proportional to the field equations. Thus, while the BRST cohomology at $J = 0$ gives the physical states, that at $J = -1$ gives the global invariances associated with the gauge field.

Now the physical states can be derived by operating on the vacuum with appropriate vertex operators: If we expand Q_I about the free BRST operator Q,

$$Q_I = Q + V$$

$$Q_I^2 = 0 \qquad \Rightarrow \qquad \{Q, V\} + V^2 = 0$$

$$\delta Q_I = i[Q_I, \lambda] \qquad \Rightarrow \qquad \delta V = i[Q + V, \lambda]$$

where λ is the gauge parameter for the *background* field. The usual operator cohomology, relevant for asymptotic states, follows from linearization:

$$V \to V_0 \quad \Rightarrow \quad \{Q, V_0\} = 0, \quad \delta V_0 = i[Q, \lambda]$$

in the weak-coupling limit, where V_0 is the part of V linear in the background fields. The asymptotic states in the cohomology of Q are then given by

$$\Phi = V_0|0\rangle, \quad \Lambda = \lambda|0\rangle \qquad \Rightarrow \qquad Q\Phi = 0, \quad \delta\Phi = iQ\Lambda$$

We can check this explicitly, as

$$V = i\tfrac{1}{2}c(\{A_a, \partial^a\} + iA^2 - F^{ab}S_{ba}) + iA_a S^{\oplus a} \quad \Rightarrow \quad V_0|0\rangle = A_a|^a\rangle + ic\tfrac{1}{2}(\partial^a A_a)|0\rangle$$

The second term gives $\check{\check{C}} = \frac{1}{2}\partial \cdot A$, in agreement with the free field equations.

8. Strings

Another interesting example is strings. Since first-quantization is essential in string S-matrix calculations, it's natural to associate string field theory with quantum mechanical BRST. As usual, for massive fields this formalism automatically includes the Stückelberg fields that would have been found by dimensional reduction, as well as all the ghosts. However, the explicit expression for the BRST operator does not explicitly correspond to that obtained by dimensional reduction: Although the spin operators $S^{\oplus a}$, $S^{\oplus\oplus}$, and S^3 are quadratic in oscillators, $S^{\oplus -1}$ is cubic (because $\hat{P}^-$, and thus X^-, is quadratic in the lightcone gauge). Nevertheless, the representation on any particular irreducible Poincaré representation contained among all the string states is the same as obtained by dimensional reduction, as follows from the generality of our analysis.

As for any Poincaré representation, reducible or not, all we need is the lightcone spin operators, given for the general case in subsection XIIA1. In subsection XIIA2 we saw that the OSp(1,1|2) generators followed immediately from just a change in notation. The IGL(1) generators were then found in subsection XIIB1 by a unitary transformation and solving half the constraints of GL(1|1); the net result was equivalent to applying "gauge conditions" to the original OSp(1,1|2) generators:

$$J = iJ^{\oplus\ominus}|_{\partial^\oplus=0,\partial^+=1}, \qquad Q = J^{\oplus-}|_{\partial^\oplus=0,\partial^+=1}$$

For the string, the gauge condition $\partial^+ = 1$ simply removes the last vestige of X^+, whose oscillator modes were already eliminated by the string lightcone gauge. On the other hand, the condition $\partial^\oplus = 0$ makes $X^\oplus = X^\oplus_{(+)} + X^\oplus_{(-)}$ the sum of two conformally covariant objects: With the elimination of the linear τ term in the expansion of $X^\oplus$, $X^\oplus_{(\pm)}$ are periodic in their arguments, and have the usual mode expansion in terms of exponentials only (no linear term).

As a result, the decomposition of Q as obtained from the lightcone becomes trivial (J was easy anyway, since it's quadratic): Relabeling the result for the string's lightcone Lorentz generators from subsection XIA3,

$$J^{\oplus-} = i\int \frac{d\sigma}{2\pi\alpha'}\,(X^\oplus \dot{X}^- - X^- \dot{X}^\oplus)$$

(where now the "i" comes from using the antihermitian form of the Lorentz spin), separating $\dot{X}^\oplus$ into its $(\pm)$ pieces, using their "chirality" $\dot{X}^\oplus_{(\pm)} = \pm X'^\oplus_{(\pm)}$ to convert the τ derivative into a σ derivative, integrating by parts, and applying the definition

$$\hat{P}_{(\pm)} = \tfrac{1}{\sqrt{2\alpha'}}(\dot{X} \pm X')$$

to X^-, we obtain

$$Q = i\sqrt{\frac{2}{\alpha'}}\int \frac{d\sigma}{2\pi}\sum_\pm X^\oplus_{(\pm)}\hat{P}^-_{(\pm)}$$

In this form we can easily apply the Virasoro constraint, as solved in the lightcone,

$$\hat{P}^2 = 0, \qquad \hat{P}^+ = \kappa\sqrt{2\alpha'}p^+ \quad \Rightarrow \quad \hat{P}^- = i\frac{1}{2\kappa\sqrt{2\alpha'}}(\hat{P}^i)^2$$

where we have applied $\partial^+ = 1$. Finally, we relabel

$$X^\oplus_{(\pm)} = C_{(\pm)}$$

for purposes of identification with the usual BRST procedure in terms of ghosts C and antighosts B (see subsections XIA4-5); comparison of the (equal-time) commutation relations then gives the further identification

$$[\hat{P}^i_{(\pm)}(1), \hat{P}^j_{(\pm)}(2)\} = \mp i\eta^{ij}2\pi\delta'(2-1)$$

$$\Rightarrow \quad [\hat{P}^i_{(\pm)}(1), X^j_{(\pm)}(2)\} = -i\sqrt{\tfrac{\alpha'}{2}}\eta^{ij}2\pi\delta(2-1)$$

$$\{B_{(\pm)}(1), C_{(\pm)}(2)\} = 2\pi\delta(2-1)$$

$$\Rightarrow \quad B_{(\pm)} = -\sqrt{\tfrac{2}{\alpha'}}\hat{P}^{\ominus}_{(\pm)}$$

The final result is then

$$Q = \frac{1}{\kappa}\sum_{\pm}\int \frac{d\sigma}{2\pi\alpha'}\ C_{(\pm)}(-\tfrac{1}{2}\hat{P}^2_{(\pm)} \pm iC'_{(\pm)}B_{(\pm)})$$

in agreement with direct first-quantization of this string in the conformal gauge, in terms of the constraints $\hat{P}^2_{(\pm)}$. The ghosts can be separated into zero- and nonzero-modes as

$$C_{(\pm)} = \tfrac{1}{2}c + \sqrt{\tfrac{\alpha'}{2}}Y^{\oplus}_{(\pm)}, \qquad B_{(\pm)} = 2\kappa b \mp \sqrt{\tfrac{2}{\alpha'}}Y'^{\ominus}_{(\pm)}$$

(For the closed string there is also an extra zero-mode in C and B, enforcing the constraint that the "+" contributions to M^2 equal the "−".)

Exercise XIIB8.1

By separating zero-modes in the string's Q, and comparing with its generic expression

$$Q = \tfrac{1}{2}c(\Box - M^2) + S^{\oplus a}\partial_a + iS^{\oplus -1}M + S^{\oplus\oplus}b$$

show that

$$S^{ij} = i\sum_{\pm}\pm\int \frac{d\sigma}{2\pi}\ Y^i_{(\pm)}Y'^j_{(\pm)}$$

$$S^{i-1}M = \frac{1}{2\kappa\sqrt{2\alpha'}}\,i\sum_{\pm}\int \frac{d\sigma}{2\pi}\ Y^i_{(\pm)}(Y'^j_{(\pm)})^2$$

$$M^2 = \frac{1}{2\kappa\alpha'}\sum_{\pm}\int \frac{d\sigma}{2\pi}\ (Y'^i_{(\pm)})^2$$

Of particular interest is the massless level: As mentioned in subsection XIB5, using the fact that the Hilbert space of the closed string is the direct product of the Hilbert spaces of open strings gives a simple analysis of the massless states of any closed string, since the massless states of any open string are given by a vector (multiplet) plus perhaps some scalars in the nonsupersymmetric case. To find the complete off-shell structure, including auxiliary fields and ghosts, we can either take the direct product of the two lightcone representations and then add 2+2 dimensions, or first add the 2+2 dimensions and then take the direct product of the two OSp(D−1,1|2) representations. In the supersymmetric case the latter is more convenient, since the procedure of adding dimensions to superspace is not yet understood, but quantization of the vector multiplet is (at least for N=1, and probably for N=2, in D=4).

For example, for the bosonic closed string we just multiply two OSp vectors, producing a tensor t_{ij}, which we can decompose into its symmetric traceless part h_{ij} (graviton plus ghosts), antisymmetric part B_{ij} (axion plus ghosts), and trace χ (physical scalar):

$$t^{ij}|_j\rangle \otimes |_i\rangle \quad \to \quad h_{ij} = t_{(ij]} - \tfrac{2}{D-2}\eta_{ij}t^k{}_k, \qquad B_{ij} = t_{[ij)}, \qquad \chi = t^i{}_i$$

(See subsection XIIA5.) However, we know that string theory prefers to treat fields in the string gauge for Weyl invariance, where the action (kinetic term) is not diagonalized in the fields (until coordinate invariance is fixed also). This is understood from this direct product structure: The "natural" string fields are

$$string\ gauge: \quad \begin{cases} t_{(ab)} & \text{graviton} \\ t_{[ab]} & \text{axion} \\ t^\alpha{}_\alpha & \text{dilaton (T-duality invariant)} \end{cases}$$

since duality affects only the X modes, while the diagonal fields (representations of OSp(D−1,1|2)) are

$$normal\ gauge: \quad \begin{cases} h_{ab} = t_{(ab)} - \frac{2}{D-2}\eta_{ab}(t^c{}_c + t^\gamma{}_\gamma) & \text{graviton} \\ B_{ab} = t_{[ab]} & \text{axion} \\ \chi = t^a{}_a + t^\alpha{}_\alpha & \text{physical scalar} \end{cases}$$

Exercise XIIB8.2

Show that the above OSp analysis is consistent with the diagonalizing field redefinitions of the low-energy string action found in exercise XIB5.2. Explain the result in terms of the redefinitions

$$\Phi^2 \frac{1}{\sqrt{-g}} \to e^{2\phi}, \qquad \Phi^2 g^{mn} \to \sqrt{-g}g^{mn}$$

In the heterotic case, as mentioned in subsection XIB6, we take the product of the real prepotential plus two chiral ghosts of super Yang-Mills with the usual vector plus two scalar ghosts of bosonic Yang-Mills:

$$(V \oplus \phi_\alpha) \otimes (A_a \oplus C_\alpha) = H_a \oplus (V_\alpha \oplus \phi_{a\alpha}) \oplus \phi \oplus \phi_{(\alpha\beta)}$$

The result is a vector prepotential H_a describing the physical supergravity and tensor multiplets (in a string gauge, $G = 1$), a chiral scalar compensator ("superdilaton") ϕ appropriate for "old minimal" supergravity, first-generation ghosts (Sp(2) doublets) V_α and $\phi_{a\alpha}$, and second-generation ghosts (an Sp(2) triplet, for the tensor multiplet) $\phi_{(\alpha\beta)}$. If the vector is accompanied by scalars, the closed string also has additional vector multiplets:

$$(V \oplus \phi_\alpha) \otimes \varphi_I = V_I \oplus \phi_{I\alpha}$$

Background fields in string theory can be treated similarly to the previous subsection: Since any open string includes Yang-Mills, we again choose the Yang-Mills ghost as the vacuum state. (In the bosonic case, even the tachyon can be obtained from it, by what would have been considered an annihilation operator with respect to the tachyonic vacuum.) Again vertex operators can be considered as additions to the BRST operator, are in the BRST operator cohomology, create states (in the state cohomology) from the vacuum, etc. For strings, such operators are local functions of the string variables: They are associated with coupling an external field at a particular value of σ (and τ). These local, fermionic vertex operators with ghost number 1 (so they can be added to Q) can be derived from related, integrated, bosonic operators with vanishing ghost number:

$$[Q, \mathcal{W}] = 0, \qquad \mathcal{W} = \int W \quad \Rightarrow \quad [Q, W] = V'$$
$$\Rightarrow \quad \{Q, V\}' = 0 \quad \Rightarrow \quad \{Q, V\} = 0$$

(Usually it is convenient to use conformal field theory methods for such methods, so we use the notation of subsection XIB7, which mainly means dropping $\pm$'s.) Thus $\int V$ is the only part of V in the cohomology, the rest being a BRST variation: Moving V to a different value of σ is a gauge transformation. Thus $V(\sigma)$ has explicit σ dependence, while Q does not:

$$Q_I = Q + V(\sigma)$$

corresponding to a vertex local in σ (as found in subsection XIB7).

The operators W are those that appear with the gauge-fixed Hamiltonian:

$$H_I = H + W(\sigma) = \{Q_I, \Lambda\} = \{Q, \Lambda\} + \{V, \Lambda\}$$

where the gauge-fixing operator is

$$\Lambda = \int (-B + f), \qquad \Lambda^2 = 0$$

the first term in Λ giving the usual $\frac{1}{2}(p^2 + M^2)$ in H, while f is left arbitrary (as long as $\Lambda^2 = 0$ is preserved) to allow more general gauges. Thus, either of V or W can be derived from the other:

$$V' = [Q, W], \qquad W = \{\Lambda, V\} \qquad (\Rightarrow \quad \{Q, V\} = [\Lambda, W] = 0)$$

In general we have

$$i\tfrac{1}{2\pi}[\hat{G}(1), \chi(2)] = \delta(2-1)\chi'(2) + w\delta'(2-1)\chi(2)$$

$$\Rightarrow \quad i[H,\chi] = \chi', \qquad H = \int \hat{G}$$

for any conformally covariant operator, so assuming V is one,

$$\hat{G}(\sigma) = \{Q, (-B+f)(\sigma)\}, \qquad \hat{G}_I(\sigma) = \{Q_I, (-B+f)(\sigma)\}$$

$$\Rightarrow \quad [Q,W] = [Q,\{\Lambda,V\}] = [\{Q,\Lambda\},V] = [H,V] = V'$$

in agreement with the previous. (Similarly, we can plug $V' = [Q,W]$ into $\{\Lambda, V'\}$ to show the consistency of $W = \{\Lambda, V\}$.) In the notation of subsection XIB7, $\hat{G}$ are the conformal generators $T_{\pm\pm}$, and $\hat{G}_I$ are $\hat{T}_{\pm\pm}$, while

$$\{V(\sigma_0), (-B+f)(\sigma)\} = \widetilde{W}(\sigma) = 2\pi\delta(\sigma-\sigma_0)W(\sigma_0)$$

The simplest examples come from adding a background directly to the Virasoro operators (as for gravity in subsection XIB4): Then for the bosonic string

$$V = CW(X,P), \qquad Q = \int C(-\tfrac{1}{2}\hat{P}^2 + iC'B), \qquad \Lambda = \int -B,$$

If we assume that W has conformal weight 1 with respect to the gauge-fixed conformal (Virasoro) generators, so that $\mathcal{W}$ is conformally invariant, then V has conformal weight 0.

9. Relation to OSp(1,1|2)

For comparison to OSp(1,1|2), we perform a unitarity (gauge) transformation on the IGL(1) action. (The most general OSp(1,1|2) expressions are given at the end of subsection XIIA2.) We first define an almost-inverse of $\hat{S}^{\oplus\oplus}$: Since $\hat{S}^{\oplus\oplus}$ annihilates states with $s^3 = s$, where we define

$$-\tfrac{1}{2}\hat{S}^{\alpha\beta}\hat{S}_{\alpha\beta} = 4s(s+1), \qquad \hat{S}^3|s\rangle = 2s^3|s\rangle$$

(so that s and s^3 take their usual integer or half-integer values), we can define $\hat{S}^{\oplus\oplus-1}$ such that

$$\hat{S}^{\oplus\oplus-1}\hat{S}^{\oplus\oplus} = 1 - \delta_{s^3,s}, \qquad \hat{S}^{\oplus\oplus}\hat{S}^{\oplus\oplus-1} = 1 - \delta_{s^3,-s}$$

$$\hat{S}^{\oplus\oplus}\hat{S}^{\oplus\oplus-1}\hat{S}^{\oplus\oplus} = \hat{S}^{\oplus\oplus}, \qquad \hat{S}^{\oplus\oplus-1}\hat{S}^{\oplus\oplus}\hat{S}^{\oplus\oplus-1} = \hat{S}^{\oplus\oplus-1}$$

We then apply the transformation

$$Q \to Q_{diag} = UQU^{-1}$$

$$Q = -cK + \mathcal{Q}^{\oplus} + \hat{S}^{\oplus\oplus}b, \qquad ln\ U = c[\{\hat{S}^{\oplus\oplus-1}, \mathcal{Q}^{\oplus}\} - \hat{S}^{\oplus\oplus-1}\hat{S}^{\oplus\oplus}\mathcal{Q}^{\oplus}\hat{S}^{\oplus\oplus-1}]$$

The exponent of U is nilpotent from the c, so it generates only a linear term, $U = 1 + ln\ U$. Using the commutation relations from subsection XIIA2

$$[\hat{S}^{\oplus\oplus}, \mathcal{Q}^{\oplus}] = 0, \qquad (\mathcal{Q}^{\oplus})^2 = K\hat{S}^{\oplus\oplus}$$

we find

$$Q_{diag} = c\delta_{s^3,s}(-K + \mathcal{Q}^{\oplus}\hat{S}^{\oplus\oplus-1}\mathcal{Q}^{\oplus})\delta_{s^3,-s} + (cb\mathcal{Q}^{\oplus}\delta_{s^3,s} + bc\delta_{s^3,-s}\mathcal{Q}^{\oplus}) + \hat{S}^{\oplus\oplus}b$$

Now we apply the identity

$$[\hat{S}^3, A] = 0 \quad \Rightarrow \quad \delta_{s^3,s}A\delta_{s^3,-s} = A\delta_{s0}$$

since any matrix element between $\langle s, s^3|...|s', s'^3\rangle$ gives $s = s^3 = s'^3 = -s' \to s = s' = 0$, as well as the facts

$$\hat{S}^{\oplus\oplus}\mathcal{Q}^{\ominus}\delta_{s0} = [\hat{S}^{\oplus\oplus}, \mathcal{Q}^{\ominus}]\delta_{s0} = 2i\mathcal{Q}^{\oplus}\delta_{s0} \quad \Rightarrow \quad \hat{S}^{\oplus\oplus-1}\mathcal{Q}^{\oplus}\delta_{s0} = -\tfrac{1}{2}i\mathcal{Q}^{\ominus}\delta_{s0}$$

$$\{\mathcal{Q}^{\alpha}, \mathcal{Q}^{\beta}\}\delta_{s0} = 0$$

This yields the final result

$$\boxed{Q_{diag} = -c(K + \tfrac{1}{4}\mathcal{Q}^{\alpha}\mathcal{Q}_{\alpha})\delta_{s0} + (cb\mathcal{Q}^{\oplus}\delta_{s^3,s} + bc\delta_{s^3,-s}\mathcal{Q}^{\oplus}) + \hat{S}^{\oplus\oplus}b}$$

Exercise XIIB9.1

Use the commutation relations of $\hat{S}^{\alpha\beta}$, as well as $-\frac{1}{2}\hat{S}^{\alpha\beta}\hat{S}_{\alpha\beta} = 4s(s+1)$, to derive

$$\hat{S}^{\ominus\ominus}\hat{S}^{\oplus\oplus} = 4[s(s+1) - s^3(s^3-1)]$$

from which follows the explicit expression

$$\hat{S}^{\oplus\oplus-1} = \frac{1}{\hat{S}^{\ominus\ominus}\hat{S}^{\oplus\oplus}}\hat{S}^{\ominus\ominus} = \frac{1-\delta_{s^3,s}}{4(s-s^3)(s+s^3+1)}\hat{S}^{\ominus\ominus}$$

Integrating over c in the action $S_{diag} = -\int \frac{1}{2}\Phi^T(-1)^{J-1}Q_{diag}\Phi$ as in subsections XIIB3-4, we find the Lagrangian

$$L_{diag} = \tfrac{1}{2}\phi^T(K + \tfrac{1}{4}\mathcal{Q}^{\alpha}\mathcal{Q}_{\alpha})\delta_{s0}\phi + i\psi^T(-1)^{\hat{S}^3-1}\delta_{s^3,-s}\mathcal{Q}^{\oplus}\phi - \tfrac{1}{2}\psi^T(-1)^{\hat{S}^3}\hat{S}^{\oplus\oplus}\psi$$

(Note that in a product of the form $\psi^T\chi$ a state of eigenvalue s^3 multiplies one of eigenvalue $-s^3$, since $(\hat{S}^3)^\dagger = (\hat{S}^3)^T = -\hat{S}^3$.) We now see that in this action only the $s = 0$ (physical) part of ϕ appears in the $\phi\phi$ term, while only the $s^3 = -s$ ("minimal") part (including physical) appears in the $\phi\psi$ term. The only part of ψ that appears in the $\phi\psi$ term is the $s^3 = s$ (minimal) part of ψ (the "antifields" to the corresponding ones in ϕ), while all, but only, the remaining ("nonminimal") part of ψ appears in the $\psi\psi$ term. In particular, the $\phi\phi$ term is recognized as the OSp(1,1|2) action of subsection XIIA3. The terms involving ψ can be eliminated by ψ's gauge invariance and field equation, and ψ contains no propagating degrees of freedom (fields with $\Box$ equations of motion), as can be seen by the methods used to analyze the cohomology in subsection XIIB4. However, the auxiliary fields ψ, and the ghosts in (the $s \neq 0$ part of) ϕ to which they couple, are useful in gauge fixing, as we'll see in the next section.

Again looking at the example of the vector:

$$\hat{S}^{\oplus\oplus} = 2|^{\oplus}\rangle\langle^{\oplus}| \quad \Rightarrow \quad \hat{S}^{\oplus\oplus-1} = \tfrac{1}{4}\hat{S}^{\ominus\ominus} = \tfrac{1}{2}|^{\ominus}\rangle\langle^{\ominus}|$$

$$\mathcal{Q}^{\oplus} = (|^{\oplus}\rangle\langle^{a}| - |^{a}\rangle\langle^{\oplus}|)\partial_a \quad \Rightarrow \quad ln\ U = -i\tfrac{1}{2}c(|^{\ominus}\rangle\langle^{a}| + |^{a}\rangle\langle^{\ominus}|)\partial_a$$

The result of the transformation is then (cf. subsection XIIB3)

$$Q_{diag} = \tfrac{1}{2}c(\Box - |^{a}\rangle\langle^{b}|\partial_a\partial_b)\delta_{s0} + cb|^{\oplus}\rangle\langle^{a}|\partial_a - bc|^{a}\rangle\langle^{\oplus}|\partial_a + 2|^{\oplus}\rangle\langle^{\oplus}|b$$

$$L_{diag} = \tfrac{1}{8}(F_{ab})^2 + i\check{A}\cdot\partial C + \check{\check{C}}^2$$

The BRST transformations now simplify to

$$QA_a = -\partial_a C, \qquad QC = 0, \qquad Q\tilde{C} = -2i\check{\check{C}}$$

$$Q\check{A}_a = -i\tfrac{1}{2}\partial^b F_{ba}, \qquad Q\check{C} = \partial\cdot\check{A}, \qquad Q\check{\check{C}} = 0$$

Exercise XIIB9.2

Find the ghosts and simplified BRST transformations for massless spin 2.

REFERENCES

1 M. Kato and K. Ogawa, *Nucl. Phys.* **B212** (1983) 443;
S. Hwang, *Phys. Rev.* **D28** (1983) 2614;
K. Fujikawa, *Phys. Rev.* **D25** (1982) 2584:
1st-quantized BRST (for strings).

2 W. Siegel, *Phys. Lett.* **149B** (1984) 157, **151B** (1985) 391:
gauge-invariant actions from 1st-quantized BRST.

3 E. Witten, *Nucl. Phys.* **B268** (1986) 253;
A. Neveu, H. Nicolai, and P.C. West, *Phys. Lett.* **167B** (1986) 307:
BRST operator as kinetic operator.

C. GAUGE FIXING

Although the quantum mechanical BRST operator is clearly useful for gauge fixing, its relation to the second-quantized BRST we applied in chapter VI is not obvious, since the latter BRST operator does not include the gauge-invariant action. Here we relate the two, and extend the former to interacting field theories. In particular, we show how the $\Phi Q \Phi$ action leads directly to the gauge-fixed kinetic term as simply as it led to the gauge-invariant one, without applying any transformations.

1. Antibracket

In the usual Hamiltonian formalism we work in a phase space (q,p) on which is defined a Poisson bracket, useful for studying symmetry properties and equations of motion in the classical theory, and for relating to the commutator of the quantum theory (see subsections IA1-2). We want to interpret the present case of interest as an analogous phase space, for which the fields ϕ (in $\Phi = \phi - ic\psi$) correspond to q and the antifields ψ to p. This automatically follows from the lightcone commutator of subsection XIIA1, by the same steps used to derive the IGL(1) algebra and inner product

$$\langle \Psi | \Xi \rangle = -i(-1)^{\Psi} \int dx\ dc\ \Psi^T(x,c)(-1)^{J(c)}\Xi(x,c)$$

in subsections XIIB1-2: We thus define this generalization of the Poisson bracket (see subsection IA2) in terms of the inner product as

$$(f[\Phi], g[\Phi]) = f \circ g, \qquad \circ = \eta^{IJ} \left\langle \frac{\overleftarrow{\delta}}{\delta \Phi^I} \middle| \frac{\delta}{\delta \Phi^J} \right\rangle$$

where we have expanded the column vector Φ over a basis in the usual way (see subsections IB1,5),

$$\Phi = |^I\rangle \Phi_I, \qquad \Phi^T = \Phi^I \langle_I|$$

$$\langle^I|^J\rangle = \eta^{IJ} = (-1)^I \eta^{JI}, \qquad (\eta^{IJ})^* = \eta^{JI}, \qquad \eta^{IK}\eta_{JK} = \delta_I^J$$

etc., and the indices I, J (not to be confused with the ghost-number operator $J(c)$ appearing in the definition of the inner product above) run over all indices on the field, which determine the statistics of the corresponding component as $(-1)^I$. (Thus Φ is always bosonic, the statistics of the fields coming always from expansion over $|^I\rangle$.) This generalized commutator "(,)" is called an "antibracket" because of the unusual statistics associated with it, following from the same unusual statistics of the inner product. (The ordering of indices on η^{IJ} in the antibracket is the opposite of usual to take into account the extra sign factor from the intervention of the anticommuting "$|$" between the two Φ's.) Plugging in the definition of the inner product, we have more explicitly

$$\circ = i \int dx\ dc\ \frac{\overleftarrow{\delta}}{\delta \Phi^I(x,c)} \eta^{JI} (-1)^{J(c)} \frac{\delta}{\delta \Phi^J(x,c)}$$

Note that while the inner product is defined between two functions of the coordinates, the antibracket is defined between two functionals of Φ, f and g, which don't depend explicitly on (x,c) (although we can specialize to cases where they depend on other values of the coordinates (x',c')). Thus the orbital term cb in $J(c)$ acts only on the argument c of $\Phi^J(x,c)$ (and not on g), while the spin term S^3 acts only on the index J of $\Phi^J(x,c)$.

We have used the fact that $\delta/\delta\Phi$ is antihermitian, as follows from the fact that the graded commutator between it and Φ is always a commutator, since they always have opposite statistics:

$$\frac{\delta}{\delta \Phi^I(x,c)} \Phi^J(x',c') = \left(\frac{\delta}{\delta \Phi^I(x,c)} \Phi^J(x',c') \right)^\dagger$$

$$= \left[\frac{\delta}{\delta \Phi^I(x,c)}, \Phi^J(x',c') \right] = \delta_I^J \delta(x'-x)\delta(c'-c)$$

$$\Rightarrow \quad \left(\frac{\delta}{\delta \Phi} \right)^\dagger = \left(\frac{\delta}{\delta \Phi} \right)^T = -\frac{\delta}{\delta \Phi}$$

Thus,

$$\left(\frac{\delta}{\delta \Phi} f^\dagger \right)^\dagger = \left[\frac{\delta}{\delta \Phi}, f^\dagger \right\}^\dagger = \left[f, -\frac{\delta}{\delta \Phi} \right\} = f \frac{\overleftarrow{\delta}}{\delta \Phi}$$

Other properties of the bracket follow directly from those of the inner product:

$$(-1)^{(f,g)} = (-1)^{f+g+1}$$

$$(f, ga) = (f,g)a, \qquad (af, g) = a(f,g)$$

$$(f,g) = -(-1)^{(f+1)(g+1)}(g,f)$$

$$(f, gh) = (f,g)h + (-1)^{(f+1)g} g(f,h)$$

$$(-1)^{(f+1)(h+1)}(f,(g,h)) + cyc. = 0$$

$$(f,g)^\dagger = (g^\dagger, f^\dagger)$$

(for some commuting or anticommuting constant a). Thus, the bracket has the exact opposite symmetry as the inner product (as is the case with the usual brackets): It would be symmetric in its two arguments if not for the $\int dc$ that sits effectively between the two arguments. Most of the properties follow from this fact, and that the signs obtained from pushing things around are determined by moving things naively while treating the "," in the middle as anticommuting. Furthermore, the existence of a bracket with these properties allows the definition of a Lie derivative,

$$\mathcal{L}_A B \equiv (A, B)$$

Exercise XIIC1.1

Find all the usual properties of this derivative (statistics, linearity, distributivity, hermiticity, algebra, etc.), and relate to the usual Lie derivative.

We also have functional identities such as

$$\frac{\delta}{\delta \Phi^I(x,c)} \langle \Phi | \Psi \rangle = i \delta_I^J (-1)^{J(c)} \Psi_I(x,c)$$

$$(\Phi^I(x,c), f[\Phi]) = -i(-1)^{J(c)} \eta^{IJ} \frac{\delta}{\delta \Phi^J(x,c)} f[\Phi]$$

from which follow

$$(\Phi^I(x,c), \Phi^J(x',c')) = -i(-1)^{J(c)} \eta^{IJ} \delta(c'-c)\delta(x'-x) = -i(-1)^{S^3} \eta^{IJ}(c+c')\delta(x-x')$$

as well as

$$(\Phi(x,c), \langle \Phi | \Psi \rangle) = \Psi(x,c), \qquad (\langle \Psi | \Phi \rangle, \langle \Phi | \Xi \rangle) = \langle \Psi | \Xi \rangle$$

Here Ψ and Ξ are wave functions in the same space as Φ, but need not be taken as bosonic (or real): We can even take them as functionals of Φ when applying the chain rule, using the above expressions for the terms where the $\delta/\delta\Phi$'s don't act on them.

Expressions quadratic in Φ will be used to perform the second-quantized (or just classical field theoretic) version of linear first-quantized transformations:

$$\mathcal{O}_A \equiv \tfrac{1}{2} \langle \Phi | A \Phi \rangle \quad \Rightarrow \quad (\mathcal{O}_A, \mathcal{O}_B) = \mathcal{O}_{[A,B\}}, \qquad A\Phi = (\Phi, \mathcal{O}_A)$$

$$df = -\int dx\ dc\ (-1)^I d\Phi^I \frac{\delta}{\delta \Phi^I} f \quad \Rightarrow \quad (\mathcal{O}_A, f) = \int dx\ dc\ (-1)^I (A\Phi^I) \frac{\delta}{\delta \Phi^I} f$$

where A and B must satisfy

$$[(-1)^{J(c)} A]^T = (-1)^{J(c)} A \quad \Rightarrow \quad A = -(-1)^{J(c)} A^T (-1)^{J(c)}$$

to give nontrivial contributions when appearing symmetrically between the two factors of Φ. Corresponding group elements come from exponentiating bosonic first-quantized generators, yielding fermionic second-quantized generators:

$$\delta f = (\mathcal{O}_A, f) = \mathcal{L}_{\mathcal{O}_A} f \quad \Rightarrow \quad f' = e^{\mathcal{L}_{\mathcal{O}_A}} f$$

$$(-1)^A = 1 \quad \Rightarrow \quad (-1)^{\mathcal{O}_A} = -1 \quad \Rightarrow \quad (\mathcal{O}_A, f) = -(f, \mathcal{O}_A)$$

Clearly, the latter relations must hold when replacing $\mathcal{O}_A$ with their nonlinear second-quantized generalizations. We then find (also for bosonic A)

$$A\Phi = (\Phi, \mathcal{O}_A) = -(\mathcal{O}_A, \Phi)$$

where the minus sign is the usual from translating first-quantized to second-quantized language (see subsection IC1).

Expanding (Φ, Φ) in c as

$$\Phi^I = \phi^I - ic\eta^{IJ}(-1)^{S^3}\check{\phi}_J$$

we find

$$(\check{\phi}_I(x), \phi^J(x')) = -(\phi^J(x), \check{\phi}_I(x')) = \delta_I^J \delta(x - x')$$

This allows us to reexpress the antibracket as

$$\circ = -\int dx\ (-1)^I \left(\frac{\overleftarrow{\delta}}{\delta\check{\phi}_I} \frac{\delta}{\delta\phi^I} + \frac{\overleftarrow{\delta}}{\delta\phi^I} \frac{\delta}{\delta\check{\phi}_I} \right)$$

For example, the antibrackets of the component fields for the vector are

$$\Phi = |^i\rangle\Phi_i = |^i\rangle[\eta_{ji}A^j - ic(-1)^{S^3}\check{A}_i] = (|^a\rangle A_a - i|^{\ominus}\rangle C + i|^{\oplus}\rangle\tilde{C}) - ic(|^a\rangle\check{A}_a - |^{\ominus}\rangle\check{\tilde{C}} - |^{\oplus}\rangle\check{C})$$

$$(\check{A}_i(x), A^j(x')) = -(A^j(x), \check{A}_i(x')) = \delta_i^j \delta(x - x')$$

$$\Rightarrow \quad (\check{A}_a, A_b) = \eta_{ab}\delta, \quad (\check{C}, C) = \delta, \quad (\check{\tilde{C}}, \tilde{C}) = \delta$$

where now "i" refers to the OSp(D$-$1,1|2) index (and we use $C = A^{\oplus}$, $\tilde{C} = A^{\ominus}$, $\check{C} = \check{A}_{\oplus}$, $\check{\tilde{C}} = \check{A}_{\ominus}$).

2. ZJBV

To prove gauge independence of the path integral, it's useful to draw an analogy of relativistic quantum mechanical BRST to second-quantized BRST. (We'll see below that this is not just an analogy, but an equivalence.) Translating the BRST quantization of subsection VIA2 into path integral language, the general Lagrangian path integral for BRST quantization in quantum physics is

$$\mathcal{A} = \int Dq\ e^{-iS'}, \qquad S' = S + \{Q, \Lambda\}$$

where q is all coordinates, including ghosts. While S and Λ depend only on q, the BRST operator is linear in the conjugate momenta p — It generates a coordinate transformation:

$$[Q, q^m\} = -i\delta_Q q^m \quad \Rightarrow \quad Q = (\delta_Q q^m) p_m$$

Here the index "m" includes all dependence of q, including time. (We saw a more explicit expression of this result in subsection VIA2, assuming the constraints G_i are themselves linear in the physical momenta.) We also have

$$[Q, S] = 0$$

where S can include not only the gauge-invariant action, but arbitrary additional gauge-invariant pieces. Such pieces can be used to construct states in the BRST cohomology from the vacuum. (This is the path-integral translation of the operator construction given in subsection VIA1.) It can also include pure BRST variations, $\{Q, \Lambda_0\}$. Thus, to prove gauge independence of $\mathcal{A}$, we need only prove the vanishing of its variation under infinitesimal change of Λ,

$$\int Dq\ e^{-iS}\{Q, \delta\Lambda\} = 0$$

But this is trivial, since Q acts as a total derivative, and $[Q, S] = 0$. More generally, we require only

$$0 = 1 \cdot \overleftarrow{Q} - i[Q, S] = -i(\delta_Q q^m)\overleftarrow{\partial}_m - (\delta_Q q^m)\partial_m S$$

where $\partial_m = \partial/\partial q^m$, and "$1 \cdot \overleftarrow{Q}$" means the derivatives in Q act backwards onto the 1 (and itself), as found by integration by parts. In cases we have considered (and almost always), $1 \cdot \overleftarrow{Q}$ and $[Q, S]$ separately vanish. More generally, since there is a $1/\hbar$ multiplying S implicitly, nonvanishing values would require a "quantum correction" to S. We can also write this condition as

$$e^{-iS}\overleftarrow{Q} = 0$$

Furthermore, we can write

$$1 \cdot \overleftarrow{Q} = -i\frac{\partial^2 Q}{\partial p_m \partial q^m}$$

These manipulations can be applied to the field theory expressions $\mathcal{J}$ for group generators, as found in subsection XIIA1 for the lightcone: The BRST operator as found from these generators is

$$S = \tfrac{1}{2}\langle\Phi|iQ\Phi\rangle, \quad Q^2 = 0 \quad \Leftrightarrow \quad (S, S) = 0$$

Then a unitary transformation can be implemented by exponentiating the infinitesimal transformation

$$\delta Q = [G, Q] \quad \Leftrightarrow \quad \delta S = \tfrac{1}{2}\langle\Phi|i(\delta Q)\Phi\rangle = (\mathcal{G}, S), \qquad \mathcal{G} = \tfrac{1}{2}\langle\Phi|G\Phi\rangle$$

$$Q' = e^{G} Q e^{-G} \quad \Leftrightarrow \quad S' = \tfrac{1}{2}\langle\Phi|iQ'\Phi\rangle = e^{\mathcal{L}_{\mathcal{G}}} S$$

We want to implement gauge fixing by performing a unitary transformation on Q and then evaluating S at the antifields $\psi = 0$:

$$\mathcal{A} = \int D\phi \; e^{-iS'}|_{\psi=0}, \qquad S' = e^{\mathcal{L}_{\Lambda}} S \quad \Rightarrow \quad S_{gf} = (e^{\mathcal{L}_{\Lambda}} S)|_{\psi=0}$$

By similar manipulations to the BRST case, we see that gauge independence means

$$0 = \int D\phi \; e^{-iS}(S, \delta\Lambda) = i \int D\phi \; (e^{-iS}, \delta\Lambda)$$

where we evaluate this expression at $\psi = 0$, and we have again included arbitrary gauge-invariant pieces in S. We thus obtain gauge independence from $(S, S) = 0$, or more generally (again using integration by parts)

$$0 = \int dx \; (-1)^I \frac{\delta^2}{\delta\check{\phi}_I \delta\phi^I} e^{-iS} = \int dx \; (-1)^I \frac{\delta^2 S}{\delta\check{\phi}_I \delta\phi^I} + i\tfrac{1}{2}(S, S)$$

This is the approach to BRST of Zinn-Justin, Batalin, and Vilkovisky (ZJBV).

Exercise XIIC2.1

Find the unitary transformation, in ZJBV language, that transforms the "untransformed" action for the massive vector (that which gives the gauge-invariant action upon dropping antifields) into the action that has only the vector field (and not the scalar) upon dropping antifields.

Writing the BRST transformations in this second-quantized ZJBV notation will allow us to gauge fix interacting theories (found by adding interaction terms to the free S) in a gauge-independent way. In particular, it proves the equivalence of the manifestly unitary lightcone gauge (which has no ghosts, only physical degres of freedom) to the manifestly Lorentz covariant Fermi-Feynman gauges (where the kinetic operator is simply $\Box - m^2$). Specifically, the $\Phi Q \Phi$ action is already unitarily transformed to the Fermi-Feynman gauge: Keeping just the ϕ terms, we have for a bosonic theory

$$S_{FF} = S|_{\psi=0} = -\int dx\, dc\; \tfrac{1}{2}\Phi^T(-1)^{J-1} c \tfrac{1}{2}(\Box - m^2)\Phi = -\int dx \; \tfrac{1}{2}\phi^T(-1)^{S^3}\tfrac{1}{2}(\Box - m^2)\phi$$

In other words, the Fermi-Feynman kinetic term is just the sum over all fields (but not antifields) of a $\Box - m^2$ term (using the OSp(D−1,1|2)-invariant inner product:

the $(-1)^{S^3}$ is just a sign, and can be absorbed by a field redefinition). This can also be seen from the result for the complete $\Phi Q \Phi$ action after dropping the antifield terms: For example, for a vector the gauge-fixed (free) action is simply (see subsection XIIB3)

$$L_{FF} = -\tfrac{1}{4} A^a \Box A_a - i\tilde{C}\tfrac{1}{2}\Box C$$

(The Fermi-Feynman gauge for fermions also gives a $\Box - m^2$ kinetic term, but with an infinite number of ghosts; this may be useful for supersymmetry.)

Exercise XIIC2.2

Let's again consider arbitrary-rank antisymmetric tensors (see exercises XIIA5.1 and XIIB3.1):

a Find the Fermi-Feynman actions.

b Do the same for the massive case. (Note: There are more fields.)

On the other hand, we saw in subsection XIIB9 that a unitary transformation, and evaluation at $\psi = 0$, gave the gauge-invariant OSp(1,1|2) action in terms of just the physical fields:

$$\Lambda_0 = \tfrac{1}{2}\langle \Phi | c[\{S^{\oplus\oplus -1}, \mathcal{Q}^{\oplus}\} - S^{\oplus\oplus -1} S^{\oplus\oplus} \mathcal{Q}^{\oplus} S^{\oplus\oplus -1}] \Phi \rangle \quad \Rightarrow \quad S' = S_{diag}$$

$$S_{diag} = \int dx\ [\tfrac{1}{2}\phi^T (K + \tfrac{1}{4}\mathcal{Q}^\alpha \mathcal{Q}_\alpha)\delta_{s0}\phi + i\psi^T (-1)^{S^3 - 1}\delta_{s^3,-s}\mathcal{Q}^{\oplus}\phi - \tfrac{1}{2}\psi^T (-1)^{S^3} S^{\oplus\oplus}\psi]$$

$$S_{gi} = S_{diag}|_{\psi=0} = \int dx\ \tfrac{1}{2}\phi^T (K + \tfrac{1}{4}\mathcal{Q}^\alpha \mathcal{Q}_\alpha)\delta_{s0}\phi$$

In the usual gauge-fixing approach, we would start with S_{diag} and do the inverse transformation to obtain the Fermi-Feynman gauge:

$$\Lambda = -\Lambda_0 \quad \Rightarrow \quad S_{FF} = (e^{-\mathcal{L}_{\Lambda_0}} S_{diag})|_{\psi=0}$$

The same Λ can be used in the interacting case, since the effect on the quadratic piece of the action will be the same. Thus, to apply the usual ZJBV procedure we can either start with S_{diag} and apply some $\Lambda \neq 0$ sufficient to fix the gauge, or we can start with S (the one we found from quantum mechanical BRST) and apply some equivalent Λ, or no Λ at all (for Fermi-Feynman gauge).

To compare with the lightcone gauge, we start with

$$S_{lc} = \int dx\ [-\tfrac{1}{4}\phi(-1)^{S^3}\Box\phi - i\psi(-1)^{S^3 - 1} S^{\oplus -}\partial^+ \phi]$$

which was itself obtained by unitary transformation from S (see subsection XIIB4), and make a further unitary transformation, of the form $\Lambda = \int dx\ \tfrac{1}{2}\phi A \Box \phi$, that has the effect $\psi \to \psi + A\Box\phi$ for A such that all terms in $\phi\Box\phi$ containing auxiliary fields, and thus also pure-gauge fields, are canceled. For this transformed S_{lc} we then have

$$S_{lc,diag}|_{\psi=0} = -\int dx\ \tfrac{1}{4}\phi\Box\delta(S^{AB})\phi$$

where the projection operator $\delta(S^{AB})$ picks out the singlets of S^{AB} $(A = (\pm,\alpha))$, i.e., the transverse (physical) degrees of the light cone. A similar procedure can be applied in the interacting case.

For the example of the vector:

$$ln\ U = -i\tfrac{1}{2}c(|^{\ominus}\rangle\langle^{+}| + |^{+}\rangle\langle^{\ominus}|)(\partial^{+})^{-1}\Box$$

$$Q_{lc} = \tfrac{1}{2}c\Box - S^{\oplus -}\partial^{+} \quad \to \quad Q_{lc,diag} = \tfrac{1}{2}c\Box|^{i}\rangle\langle_{i}| - S^{\oplus -}\partial^{+}$$

$$L_{lc} \to L_{lc,diag} = -\tfrac{1}{4}A_i\Box A_i + \check{\check{C}}\partial^{+}A^{-} - i\check{A}^{-}\partial^{+}C$$

which has just the transverse (lightcone) degrees of freedom when the antifields are dropped:

$$L_{lc,gf} = -\tfrac{1}{4}A_i\Box A_i$$

3. BRST

In subsection VIA2, we saw that BRST could be used to gauge fix by adding a BRST variation to the gauge-invariant Lagrangian. In that case, physical states are those that are not only in the BRST cohomology, but also satisfy the equations of motion.

On the other hand, for relativistic mechanics we saw that the equations of motion are rather redundant, since τ is unphysical, and so $p^2(+m^2) = 0$ is already included as a constraint, and contained in the quantum mechanical BRST operator. In fact, we have seen how in the most general case of a free field the correct spectrum is specified by just the cohomology of the quantum mechanical BRST operator.

We therefore want to identify the quantum mechanical BRST cohomology condition with the combination of the second-quantized BRST cohomology condition and the wave equation. This essentially has been accomplished in subsection XIIC2 by decomposing Q_{diag} with respect to c, as we'll now see by some further analysis.

From subsection XIIB9 we have

$$Q_{diag} = -c(K + \tfrac{1}{4}\mathcal{Q}^{\alpha}\mathcal{Q}_{\alpha})\delta_{s0} + (cb\mathcal{Q}^{\oplus}\delta_{s^3,s} + bc\delta_{s^3,-s}\mathcal{Q}^{\oplus}) + S^{\oplus\oplus}b$$

where the first term gives L_{gi} in terms of the physical part $(s = 0)$ of ϕ, the second term gives the minimal BRST transformations in terms of the minimal (anti)fields, and the last term adds the nonminimal stuff needed for fixing to general gauges. In particular, we see that the BRST transformation of the physical fields are

$$\delta(\delta_{s0}\phi) \sim \delta_{s0}\mathcal{Q}^{\oplus}\phi = \delta_{s0}\mathcal{Q}^{\oplus}(\delta_{s,1/2}\delta_{s^3,-1/2}\phi)$$

(Note that this differs somewhat from the expression found from Q, since the transformation to Q_{diag} is effectively a redefinition of ψ, adding to it a piece proportional to an operator on ϕ.)

Thus the only occurrences of the physical fields in Q_{diag} are in the term that gives the gauge-invariant action and the term that gives their BRST transformation; the remaining terms introduce nonminimal fields, as well as account for the BRST transformations of the ghosts. But this is the definition of BRST: Take the classical action in terms of physical fields, construct the BRST transformation from the gauge transformation that leaves the classical action invariant, add terms to the BRST operator that insure its nilpotency on these ghosts, and add nonminimal terms to allow gauge fixing. We have just seen that Q_{diag} is exactly of this structure, where gauge fixing gives the desired Fermi-Feynman gauge by unitary transformation to Q (which becomes a canonical transformation in second-quantized language, using the antibracket).

All that is left to see is how the ZJBV combination of the gauge-invariant action with the BRST operator is equivalent to ordinary BRST. Expanding in antifields, ZJBV gives the gauge fixed action as

$$S = S_0 + \psi_m(Q\phi_m) + \psi_{nm}\psi_{nm} \;\Rightarrow\; S_{gf} = S_0 + (\delta\Lambda/\delta\phi_m)(Q\phi_m) + (\delta\Lambda/\delta\phi_{nm})(\delta\Lambda/\delta\phi_{nm})$$

where we have used the fact that the three terms in Q_{diag} contain only the physical, minimal ("m"; including physical), and nonminimal ("nm") fields, respectively. In the usual ZJBV and BRST formalisms, derived from BRST without antifields, there is no ψ^2 term, since this generates a BRST transformation

$$iQ\phi_{nm} = (S, \phi_{nm}) \sim \psi_{nm}$$

One instead introduces further nonminimal fields, the "Nakanishi-Lautrup fields", such that

$$Q\phi_{nm} = \phi_{NL}$$

and use an extended gauge-fixing function

$$\hat{\Lambda} = \Lambda + \phi_{nm}\phi_{NL}$$

Then the gauge fixed action is

$$S_{gf} = S_0 + \{Q, \hat{\Lambda}\} = S_0 + (\delta\Lambda/\delta\phi_m)(Q\phi_m) + (\delta\Lambda/\delta\phi_{nm})\phi_{NL} + \phi_{NL}^2$$

After eliminating the NL fields by their (algebraic) equations of motion, we obtain the same result as found from ZJBV. (Of course, the NL fields can also be introduced directly into the ZJBV formalism, but are redundant for purposes of finding Fermi-Feynman gauges.)

Consider the special case of Yang-Mills: Generalizing our results for free Yang-Mills to the interacting case, making use of the BRST transformations of subsection VIA4, we have

$$L_{ZJBV} = \tfrac{1}{8}(F_{ab})^2 + \check{\tilde{C}}^2 + i\check{A}\cdot[\nabla, C] - \check{C}C^2$$

The basic antibrackets are

$$(\check{A}_a, A_b) = \eta_{ab}\delta, \qquad (\check{C}, C) = \delta, \qquad (\check{\tilde{C}}, \tilde{C}) = \delta$$

From the general relations we saw earlier, and the definition of S in terms of Q for the free case, we have

$$iQ\Phi = (\Phi, S) = (S, \Phi)$$

Since in the above we have pulled out factors to the left of the fields, as

$$\Phi = |^i\rangle\Phi_i = |^i\rangle[\eta_{ji}A^j - ic(-1)^{S^3}\check{A}_i]$$

we pull them out of the left of the antibracket, to obtain

$$i(Q\phi)^I = (S, \phi^I), \qquad i(Q\check{\phi})^I = (S, \check{\phi}^I)$$

where as before $(Q\phi)^I$, etc., means to evaluate the corresponding component of $Q\Phi$ and introduce the corresponding signs for effectively pulling those factors to the left. We then find the previous results for the BRST transformations of the fields (by construction), but also those of the antifields:

$$QA_a = -[\nabla_a, C], \qquad QC = iC^2, \qquad Q\tilde{C} = -2i\check{\tilde{C}}$$

$$Q\check{A}_a = -i\tfrac{1}{2}[\nabla^b, F_{ba}] + i\{C, \check{A}_a\}, \qquad Q\check{C} = [\nabla\cdot, \check{A}] + i[C, \check{C}], \qquad Q\check{\tilde{C}} = 0$$

(Remember that the funny signs of the antibracket come from its symmetry, plus treating the comma in "(,)" as anticommuting. Note the generic terms $i[C,\ \}$.)

Exercise XIIC3.1

Generalize the above results for the action and BRST transformations with antifields when Yang-Mills is coupled to matter.

BRST was described in a different way in subsection VIA4: Here we apply quantum mechanical BRST, and find it equivalent to applying the ZJBV form of BRST to second-quantization. The ZJBV action consists of the gauge-invariant action, plus the antifields times the BRST transformations of the fields, plus (antifield)^2 terms. The difference between the BRST transformations obtained by the general methods of subsection VIA1 as applied to second-quantization, and those found in this chapter by applying OSp methods to first-quantization of relativistic systems, is that the Nakanishi-Lautrup field is treated as a field in the former approach and as an antifield in the latter. The two give equivalent results: The latter uses fewer fields, but is slightly more restricted in choices of gauge; however, this restriction is avoided in practice. (More "nonminimal" fields can be added to allow more general gauge choices in either case.) For example, the ZJBV action for the former treatment of Yang-Mills can be obtained from that for the latter by the replacement

$$\check{\tilde{C}}^2 \to \check{\tilde{C}} B$$

The gauge-fixed action is the canonically transformed action (with respect to the antibracket) evaluated at vanishing antifields:

$$S_{gf} = e^{\mathcal{L}_\Lambda} S_{ZJBV}|$$

Consider Yang-Mills in the most common type of gauge, where some function of A is fixed. From the usual BRST approach (see subsection VIA4), or the ZJBV approach with B, we find

$$\Lambda = tr \int \tfrac{1}{2}\tilde{C}[f(A) + \tfrac{1}{2}\alpha B] \quad \Rightarrow \quad L_{gf} = L_{gi} - \tfrac{1}{2}B[f(A) + \tfrac{1}{2}\alpha B] - \tfrac{1}{2}i\tilde{C}\frac{\partial f}{\partial A} \cdot [\nabla, C]$$

while in the ZJBV approach without B we have

$$\Lambda = tr \int \tfrac{1}{2}\tilde{C} f(A) \quad \Rightarrow \quad L_{gf} = L_{gi} + \tfrac{1}{4}f(A)^2 - \tfrac{1}{2}i\tilde{C}\frac{\partial f}{\partial A} \cdot [\nabla, C]$$

which is equivalent to the previous for positive α (after elimination of B).

REFERENCES

1 Siegel and Zwiebach, *loc. cit.* (XIIA).
2 Siegel, *loc. cit.*
3 J. Zinn-Justin, in *Trends in elementary particle theory*, eds. H. Rollnik and K. Dietz (Springer-Verlag, 1975) p. 2:
introduced antifields, antibracket, etc.
4 I.A. Batalin and G.A. Vilkovisky, *Phys. Lett.* **102B** (1983) 27, **120B** (1983) 166; *Phys. Rev.* **D28** (1983) 2567, **D30** (1984) 508; *Nucl. Phys.* **B234** (1984) 106; *J. Math. Phys.* **26** (1985) 172:
generalized Zinn-Justin's approach.
5 W. Siegel and B. Zwiebach, *Nucl. Phys.* **B299** (1988) 206:
2nd-quantized ZJBV from 1st-quantization.

AfterMath

Conversions to other common conventions

$$\eta_{ab} \to -\eta_{ab}, \qquad \gamma_a \to \tfrac{1}{\sqrt{2}}\gamma_a, \qquad S \to -S$$

$$g^2 \to \frac{g^2}{8\pi^2} \quad [via \quad \tfrac{1}{2}(m^2 - \Box) \to m^2 - \Box, \quad \frac{d^D x}{(2\pi)^{D/2}} \to d^D x] \quad or\ nonabelian \quad \frac{g^2}{16\pi^2}$$

Natural (Planck) units

$$c = \hbar = k = \kappa = 1 \quad (G = \pi)$$

$$\frac{1}{e^2} = \frac{8\pi^2}{e_{ft}^2} = \frac{2\pi}{e_m^2} = \frac{2\pi}{\alpha} = 861.022516(39), \qquad H^{-1} = 1.5(2) \times 10^{61}$$

$$1\ kg = 2.59194(17) \times 10^7, \quad 1\ m = 1.096782(70) \times 10^{35}, \quad 1\ s = 3.28807(21) \times 10^{43}$$

$$1\ K = 3.98170(25) \times 10^{-33}, \qquad 1\ GeV = 4.62054(30) \times 10^{-20}$$

Indices

$a, b, c, ...$ — (flat) vector
$i, j, k, ...$ — transverse (D−1 or D−2) vector or internal
$m, n, p, ...$ — (curved) vector or large summation
$A, B, C, ...$ — (flat) super or conformal vector
$I, J, K, ...$ — internal
$M, N, P, ...$ — (curved) super
$\mathcal{A}, \mathcal{B}, \mathcal{C}, ...$ — conformal spinor
$\alpha, \beta, \gamma, ...; \mu, \nu, \pi, ...$ — spinor (usually 2-valued) or fermionic
ι, κ — internal
0 — time
-1 — mass (dimensional reduction)
$\pm; t, \bar{t}$ — lightcone (longitudinal; transverse)
$\oplus, \ominus$ — spinor or spacecone reference line

Integration

$$\int dx \equiv \int \frac{d^D x}{(2\pi)^{D/2}}, \qquad \int dp \equiv \int \frac{d^D p}{(2\pi)^{D/2}}$$

$$\delta(x - x') \equiv (2\pi)^{D/2}\delta^D(x - x'), \qquad \delta(p - p') \equiv (2\pi)^{D/2}\delta^D(p - p')$$

$$\langle x|x'\rangle = \delta(x - x'),\ \langle p|p'\rangle = \delta(p - p'); \quad \langle x|p\rangle = e^{ip\cdot x},\ \langle p|x\rangle = e^{-ip\cdot x}; \quad p_a = -i\partial_a$$

$$on-shell: \qquad \langle p|p'\rangle = \frac{\delta(p - p')}{2\pi\delta[\frac{1}{2}(p^2 + m^2)]}$$

GROUP THEORY (I,X): Covering groups

SO(2) = U(1), SO(1,1) = GL(1)
SO(3) = SU(2) = SU*(2) = USp(2), SO(2,1) = SU(1,1) = SL(2) = Sp(2)
SO(4) = SU(2)⊗SU(2), SO(3,1) = SL(2,C) = Sp(2,C), SO(2,2) = SL(2)⊗SL(2)
SO(5) = USp(4), SO(4,1) = USp(2,2), SO(3,2) = Sp(4)
SO(6) = SU(4), SO(5,1) = SU*(4), SO(4,2) = SU(2,2), SO(3,3) = SL(4)

SO*(2) = U(1), SO*(4) = SU(2)⊗SL(2), SO*(6) = SU(3,1), SO*(8) = SO(6,2)

Spinors

D_- / D	0 Euclidean	1 Lorentz	2 conformal	3
0	$\psi_\alpha\ \psi_{\alpha'}$ $\eta^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{\alpha\beta'}$	$\psi_\alpha\ \psi_{\dot\alpha}$ $\eta^{\alpha\beta}$ $\sigma_{\alpha\dot\beta}$	$\psi_\alpha\ \psi_{\alpha'}$ $\eta^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{\alpha\beta'}$	$\psi_\alpha\ \psi_{\dot\alpha}$ $\eta^{\alpha\beta}$ $\sigma_{\alpha\dot\beta}$
1	ψ_α $\eta^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}$	ψ_α $\eta^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}$	ψ_α $\eta^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}$	ψ_α $\eta^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}$
2	$\psi_\alpha\ \psi^\alpha$ $\eta^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}\ \sigma^{(\alpha\beta)}$	$\psi_\alpha\ \psi^\alpha$ $\eta_{\dot\alpha}{}^\beta$ $\sigma_{(\alpha\beta)}\ \sigma^{(\alpha\beta)}$	$\psi_\alpha\ \psi^\alpha$ $\Omega^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}\ \sigma^{(\alpha\beta)}$	$\psi_\alpha\ \psi^\alpha$ $\Omega_{\dot\alpha}{}^\beta$ $\sigma_{(\alpha\beta)}\ \sigma^{(\alpha\beta)}$
3	ψ_α $\Omega^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}$	ψ_α $\Omega^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}$	ψ_α $\Omega^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}$	ψ_α $\Omega^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{(\alpha\beta)}$
4	$\psi_\alpha\ \psi_{\alpha'}$ $\Omega^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{\alpha\beta'}$	$\psi_\alpha\ \psi_{\dot\alpha}$ $\Omega^{\alpha\beta}$ $\sigma_{\alpha\dot\beta}$	$\psi_\alpha\ \psi_{\alpha'}$ $\Omega^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{\alpha\beta'}$	$\psi_\alpha\ \psi_{\dot\alpha}$ $\Omega^{\alpha\beta}$ $\sigma_{\alpha\dot\beta}$
5	ψ_α $\Omega^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}$	ψ_α $\Omega^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}$	ψ_α $\Omega^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}$	ψ_α $\Omega^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}$
6	$\psi_\alpha\ \psi^\alpha$ $\eta^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}\ \sigma^{[\alpha\beta]}$	$\psi_\alpha\ \psi^\alpha$ $\Omega_{\dot\alpha}{}^\beta$ $\sigma_{[\alpha\beta]}\ \sigma^{[\alpha\beta]}$	$\psi_\alpha\ \psi^\alpha$ $\Omega^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}\ \sigma^{[\alpha\beta]}$	$\psi_\alpha\ \psi^\alpha$ $\eta_{\dot\alpha}{}^\beta$ $\sigma_{[\alpha\beta]}\ \sigma^{[\alpha\beta]}$
7	ψ_α $\eta^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}$	ψ_α $\eta^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}$	ψ_α $\eta^{\alpha\beta}\ \Omega_{\dot\alpha}{}^\beta\ \Omega^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}$	ψ_α $\eta^{\alpha\beta}\ \eta_{\dot\alpha}{}^\beta\ \eta^{\dot\alpha\beta}$ $\sigma_{[\alpha\beta]}$

LORENTZ (I,II)

$$-m^2 = p^2 = p^a p^b \eta_{ab} = -(p^0)^2 + (p^1)^2 + (p^2)^2 + (p^3)^2 = -2p^+ p^- + 2p^t \bar{p}^t$$

$$E = p^0 = -p_0; \qquad p^\pm = \tfrac{1}{\sqrt{2}}(p^0 \pm p^1), \quad p^t = \tfrac{1}{\sqrt{2}}(p^2 - ip^3)$$

$$-ds^2 = dx^2 = dx^a dx^b \eta_{ab}, \quad p^a ds = m\, dx^a, \quad p^a d\tau = dx^a$$

$$p^2 + m^2 = S_a{}^b p_b + S_{a,-1} m + w p_a = S_{-1}{}^a p_a + wm = 0$$

2-spinor

$$C_{\alpha\beta} = -C^{\alpha\beta} = C_{\dot\alpha\dot\beta} = \left(\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right); \quad p^{\alpha\dot\beta} = \begin{pmatrix} p^+ & \bar{p}^t \\ p^t & p^- \end{pmatrix}$$

$$\psi_\alpha = \psi^\beta C_{\beta\alpha}, \quad \psi_{\dot\alpha} = \psi^{\dot\beta} C_{\dot\beta\dot\alpha}; \qquad \psi^2 = \tfrac{1}{2}\psi^\alpha \psi_\alpha = i\psi^\oplus \psi^\ominus = -i\psi_\oplus \psi_\ominus$$

$$\bar\psi^{\dot\alpha} = (\psi^\alpha)^\dagger \quad \Rightarrow \quad (\psi_\alpha)^\dagger = -\bar\psi_{\dot\alpha}, \quad (\psi^2)^\dagger = \bar\psi^2 = \tfrac{1}{2}\bar\psi^{\dot\alpha}\bar\psi_{\dot\alpha} = i\bar\psi^{\dot\oplus}\bar\psi^{\dot\ominus}$$

$$A_{[\alpha\beta]} = A_{\alpha\beta} - A_{\beta\alpha} = C_{\alpha\beta} C^{\gamma\delta} A_{\gamma\delta}, \quad A_{[\alpha\beta\gamma]} = 0; \qquad V^2 = -2\ det\ V = V^{\alpha\dot\beta} V_{\alpha\dot\beta}$$

$$\eta_{\alpha\dot\alpha,\beta\dot\beta} = C_{\alpha\beta} C_{\dot\alpha\dot\beta}, \quad \epsilon_{\alpha\dot\alpha,\beta\dot\beta,\gamma\dot\gamma,\delta\dot\delta} = i(C_{\alpha\beta}C_{\gamma\delta}C_{\dot\alpha\dot\delta}C_{\dot\beta\dot\gamma} - C_{\alpha\delta}C_{\beta\gamma}C_{\dot\alpha\dot\beta}C_{\dot\gamma\dot\delta})$$

$$\langle\psi| = \psi^\alpha \langle_\alpha|, \quad |\psi\rangle = |^\alpha\rangle \psi_\alpha; \quad [\psi| = \psi^{\dot\alpha}[_{\dot\alpha}|, \quad |\psi] = |^{\dot\alpha}]\psi_{\dot\alpha}$$

$$V = |^\alpha\rangle V_\alpha{}^{\dot\beta}[_{\dot\beta}|, \quad V^* = -|^{\dot\alpha}]V^\beta{}_{\dot\alpha}\langle_\beta|; \quad f = |^\alpha\rangle f_\alpha{}^\beta \langle_\beta|, \quad f^* = |^{\dot\alpha}] f_{\dot\alpha}{}^{\dot\beta}[_{\dot\beta}|$$

$$\langle\psi\chi\rangle = \langle\chi\psi\rangle = \psi^\alpha \chi_\alpha, \quad [\psi\chi] = \psi^{\dot\alpha}\chi_{\dot\alpha}; \quad \langle\psi\chi\rangle^\dagger = [\psi\chi]$$

$$\langle\psi|V|\chi] = \psi^\alpha V_\alpha{}^{\dot\beta}\chi_{\dot\beta}, \quad \langle\psi|f|\chi\rangle = \psi^\alpha f_\alpha{}^\beta \chi_\beta; \quad VW^* + WV^* = (V\cdot W)I$$

$$\epsilon_{0123} = -\epsilon^{0123} = 1, \qquad \epsilon(V,W,X,Y) = i\ tr(VW^*XY^* - Y^*XW^*V)$$

4-spinor

$$\varPsi = \begin{pmatrix} \psi_\alpha \\ \bar\chi_{\dot\alpha} \end{pmatrix}, \qquad \bar\varPsi = \varPsi^\dagger \varUpsilon = (\chi^\alpha\ \bar\psi^{\dot\alpha}); \qquad -\gamma^a\gamma^b = \tfrac{1}{2}\eta^{ab} + S^{ab}$$

$$\nabla\!\!\!/ = \begin{pmatrix} 0 & \nabla_\alpha{}^{\dot\beta} \\ \nabla^\beta{}_{\dot\alpha} & 0 \end{pmatrix}, \quad \varUpsilon = \sqrt{2}\gamma_0 = \begin{pmatrix} 0 & \bar{C}^{\dot\alpha\dot\beta} \\ C^{\alpha\beta} & 0 \end{pmatrix}, \quad \gamma_{-1} = \tfrac{1}{\sqrt{2}}\begin{pmatrix} -i\delta_\alpha^\beta & 0 \\ 0 & i\delta_{\dot\alpha}^{\dot\beta} \end{pmatrix}$$

$$\gamma^a\gamma_a = -2, \qquad \gamma^a a\!\!\!/ \gamma_a = a\!\!\!/, \qquad \gamma^a a\!\!\!/ b\!\!\!/ \gamma_a = a\cdot b, \qquad \gamma^a a\!\!\!/ b\!\!\!/ c\!\!\!/ \gamma_a = c\!\!\!/ b\!\!\!/ a\!\!\!/$$

$$tr(I) = 4, \qquad tr(a\!\!\!/ b\!\!\!/) = -2a\cdot b, \qquad tr(a\!\!\!/ b\!\!\!/ c\!\!\!/ d\!\!\!/) = a\cdot b\ c\cdot d + a\cdot d\ b\cdot c - a\cdot c\ b\cdot d$$

$$\varPsi = |^\alpha\rangle\psi_\alpha + |^{\dot\alpha}]\bar\chi_{\dot\alpha}, \qquad \bar\varPsi = \varPsi^\dagger = \chi^\alpha\langle_\alpha| + \bar\psi^{\dot\alpha}[_{\dot\alpha}|$$

$$\gamma_{\alpha\dot\beta} = -|_\alpha\rangle[_{\dot\beta}| - |_{\dot\beta}]\langle_\alpha|; \qquad \varPi_+ = |^\alpha\rangle\langle_\alpha|, \qquad \varPi_- = |^{\dot\alpha}\rangle\langle_{\dot\alpha}|$$

ACTIONS (III-VI,XII)

$$L = -\tfrac{1}{2}\dot{q}^2 g(q) + \dot{q}A(q) + U(q), \qquad S = \int dt\ \ L$$

$$L_H = -\dot{q}p + H(q,p), \qquad S_H = \int dt\ \ H - dq\ \ p$$

$$\mathcal{A} = \int D\phi\ \ e^{-iS}\Psi; \qquad Wick: \qquad \mathcal{A} = \int D\phi\ \ e^{-S}\Psi, \quad S \geq 0$$

Mechanics

OSp(1,1|2):

$$K = -\tfrac{1}{2}(\Box - m^2), \qquad \mathcal{Q}^\alpha = S^{\alpha a}\partial_a + S^{\alpha -1}im \quad [+\tilde{\gamma}^\alpha(\tilde{\gamma}^- - \tilde{\gamma}^+ K)]$$

$$\hat{S}^{\alpha\beta} = S^{\alpha\beta} \quad (-\tfrac{1}{2}\tilde{\gamma}^{(\alpha}\tilde{\gamma}^{\beta)}), \qquad -\tfrac{1}{2}\hat{S}^{\alpha\beta}\hat{S}_{\alpha\beta} = 4s(s+1)$$

$$S = \int dx\ L_{gi}, \qquad L_{gi} = \tfrac{1}{2}\phi^T K_{gi}\phi, \qquad K_{gi} = \tfrac{1}{2}(-\Box + m^2 + \tfrac{1}{2}\mathcal{Q}^\alpha\mathcal{Q}_\alpha) \qquad (\hat{S}^{\alpha\beta}\phi = 0)$$

$$\delta\phi = \delta_{s0}\tfrac{1}{2}\mathcal{Q}^\alpha\Lambda_\alpha$$

$$S_f = \int dx\ L_{gi,f}, \qquad L_{gi,f} = \tfrac{1}{2}\hat{\phi}K_{gi,f}\hat{\phi}, \qquad K_{gi,f} = \tfrac{1}{2}\tilde{\gamma}^\alpha(S_\alpha{}^a\partial_a + S_{\alpha -1}im)$$

IGL(1):

$$Q = \tfrac{1}{2}c(\Box - m^2) + S^{\oplus a}\partial_a + S^{\oplus -1}im + S^{\oplus\oplus}b \quad (+\tilde{S}^{\oplus -}), \qquad J = cb + S^3 \quad (+\tilde{S}^3)$$

$$S = -(-1)^\Phi \int dx\ dc\ \tfrac{1}{2}\Phi^T(-1)^{J-1}Q\Phi = \tfrac{1}{2}\langle\Phi|iQ\Phi\rangle, \qquad S_{gi} = S|_{J\Phi=0}$$

$$S_{FF} = S|_{b\Phi=0} = -\int dx\ \tfrac{1}{2}\phi^T(-1)^{S^3}\tfrac{1}{2}(\Box - m^2)\phi$$

Quantum ChromoDynamics

$$G_i{}^\dagger = G_i, \qquad [G_i, G_j] = -if_{ij}{}^k G_k, \qquad (G_i\psi)_A = (G_i)_A{}^B\psi_B$$

$$\nabla_a = \partial_a + iA_a = \partial_a + iA_a{}^i G_i, \qquad -i[\nabla_a, \nabla_b] = F_{ab} = F_{ab}{}^i G_i = \partial_{[a}A_{b]} + i[A_a, A_b]$$

$$L = \tfrac{1}{8g^2}tr\ F^{ab}F_{ab} + L(\nabla, \psi), \qquad tr_D(G_iG_j) = \delta_{ij}, \quad tr_A(G_iG_j) = 2N\delta_{ij}$$

$$QA_a = -[\nabla_a, C],\ QC = iC^2,\ Q\tilde{C} = -iB,\ QB = 0,\ Q\phi = iC\phi\ (for\ \delta\phi = i\lambda\phi)$$

$$S_{gf} = S_{gi} - iQ\Lambda,\ \Lambda = tr\int \tfrac{1}{2}\tilde{C}(f + \tfrac{1}{2}\alpha B)\ \Rightarrow\ L_{gf} = L_{gi} - \tfrac{1}{2}B(f + \tfrac{1}{2}\alpha B) + \tfrac{1}{2}i\tilde{C}(\delta f)|_{\lambda = C}$$

$$L_{Majorana} = \psi^\alpha i\nabla_\alpha{}^{\dot{\alpha}}\bar{\psi}_{\dot{\alpha}} + \tfrac{m}{2\sqrt{2}}(\psi^\alpha\psi_\alpha + \bar{\psi}^{\dot{\alpha}}\bar{\psi}_{\dot{\alpha}}) \quad \to \quad -\tfrac{1}{4}\psi^\alpha(\Box - m^2)\psi_\alpha - \tfrac{1}{2}\psi^\alpha f_\alpha{}^\beta\psi_\beta$$

$$L_{Dirac} = \overline{\Psi}(i\nabla\!\!\!/ + \tfrac{m}{\sqrt{2}})\Psi = (\bar{\psi}^{\dot{\alpha}}i\nabla^\alpha{}_{\dot{\alpha}}\psi_\alpha + \bar{\chi}^{\dot{\alpha}}i\nabla^\alpha{}_{\dot{\alpha}}\chi_\alpha) + \tfrac{m}{\sqrt{2}}(\psi^\alpha\chi_\alpha + \bar{\psi}^{\dot{\alpha}}\bar{\chi}_{\dot{\alpha}})$$

FEYNMAN (V)

$$S = S_0 + S_I, \qquad S_0 = \int \tfrac{1}{2}\phi K \phi; \qquad \mathcal{A}_N = \prod_{i=1}^{N} \left(\int \psi_{Ni} \frac{\delta}{\delta\phi} \right) Z[\phi]$$

$$Z[\varphi] = e^{-W[\varphi]} = \int D\phi\ e^{-(S_0[\phi]+S_I[\phi+\varphi])} = exp\left(\int \tfrac{1}{2} \frac{\delta}{\delta\varphi} \frac{1}{K} \frac{\delta}{\delta\varphi} \right) e^{-S_I[\varphi]}$$

Effective action $\Gamma[\phi]$ (unrenormalized): (E.g., $L = -\frac{1}{4}\phi(\Box - m^2)\phi + \frac{1}{6}g\phi^3$.)
(A1) 1PI graphs only (plus S_0). (For $W[\phi]$, connected graphs only.)
(A2) Momenta: label consistently with conservation, with $\int dp$ for each loop.
(A3) Propagators: $1/K$ for each internal line. (E.g., $1/\frac{1}{2}(p^2 + m^2)$.)
(A4) Vertices: read off of $-S_I$. (E.g., $-g$)
(A5) External lines: attach the appropriate (off-shell) fields and $\int dp$, with $\delta(\sum p)$.
(A6) Statistics: 1/n! for n-fold symmetry of internal/external lines
(or keep just 1 of n! related graphs); -1 for fermionic loop; overall -1.

Vacuum: (Renormalize before for minimal subtraction/after for MOM.)
(B1) Find the minimum of the effective potential (for scalars).
(B2) Shift (scalar) fields to perturb about minimum; drop constant in potential.
(B3) Find resulting masses; find wave function normalizations.

T-matrix:
(C1) Connected trees of (shifted, renormalized) Γ: (A2-4) for L=0 with $S \to \Gamma$.
(C2) Amputate external Γ_0-propagators.
(C3) External lines: appropriate to Γ_0 wave equation $\tilde{K}\psi = 0$. (E.g., 1.)
(C4) External-line statistics: No symmetry factors; -1 for fermion permutation.

Probabilities

$$\mathcal{S}_{connected} = i\delta\left(\sum p\right) T$$

$$dP = |T_{fi}|^2 \delta^D \left(\sum p\right) \prod_{all} \frac{(2\pi)^{D/2}}{\omega} \prod_{out} \frac{d^{D-1}p}{(2\pi)^{D-1}}, \quad P = 2(Im\ T_{ii})(2\pi)^{-D/2} \prod_{in} \frac{(2\pi)^{D/2}}{\omega}$$

$$\frac{dP}{dt} = \frac{2\ Im\ T_{ii}}{\omega}, \quad \frac{dP}{ds} = \frac{2\ Im\ T_{ii}}{m} = -2\ Im\ M, \quad \frac{dP}{d\tau} = 2\ Im\ T_{ii} = -Im\ M^2$$

$$d\sigma = \frac{dP}{v_{12}} = |T_{fi}|^2 \delta^D \left(\sum p\right) \frac{(2\pi)^D}{\lambda_{12}} \prod_{out} \frac{d^{D-1}p}{(2\pi)^{D/2-1}\omega}, \qquad \sigma = 2(Im\ T_{ii}) \frac{(2\pi)^{D/2}}{\lambda_{12}}$$

$$\lambda_{12}^2 = (p_1 \cdot p_2)^2 - m_1^2 m_2^2 = \tfrac{1}{4}[s - (m_1 + m_2)^2][s - (m_1 - m_2)^2]$$

$$\frac{d\sigma}{dt} = \tfrac{1}{2}(2\pi)^3 |T_{fi}|^2 \frac{1}{\lambda_{12}^2} \quad (4D); \quad s = -(p_1 + p_2)^2,\ t = -(p_1 + p_3)^2,\ u = -(p_1 + p_4)^2$$

$$\frac{d\sigma}{d\Omega} = (2\pi)^2 |T_{fi}|^2 \frac{|\vec{p}_3|^{D-1}}{\lambda_{12}[\frac{1}{2}(s - m_3^2 - m_4^2)\omega_3 - m_3^2\omega_4]} = (2\pi)^2 |T_{fi}|^2 \frac{\lambda_{34}^{D-3}}{\lambda_{12} s^{D/2-1}} \quad (CoM)$$

GAUGES (II,VI): Gervais-Neveu

$$L_A = -\tfrac{1}{4}A\cdot\Box A - iA^a A^b \partial_b A_a - \tfrac{1}{4}A^a A^b A_a A_b$$

$$Yang-Mills: \qquad L = L_A + L_C, \quad L_C = -\tfrac{1}{2}i\tilde{C}(\partial + iA)^2 C - \tfrac{1}{2}\tilde{C}C(\partial\cdot A + iA^2)$$

$$Gervais-Neveu: \qquad L = L_A + \tfrac{1}{4}m^2 A^2$$

Twistors

$$\langle p| = p^\alpha\langle_\alpha|, \quad |p\rangle = |^\alpha\rangle p_\alpha; \quad [p| = p^{\dot\alpha}[_{\dot\alpha}|, \quad |p] = |^{\dot\alpha}]p_{\dot\alpha}$$

$$\langle pq\rangle^* = [qp] = -[pq], \qquad \langle pq\rangle\langle rs\rangle + \langle qr\rangle\langle ps\rangle + \langle rp\rangle\langle qs\rangle = 0$$

$$P = |p\rangle[p| = p^+|+\rangle[+| + p^-|-\rangle[-| + p^t|-\rangle[+| + \bar{p}^t|+\rangle[-|, \qquad -P^* = |p]\langle p|$$

$$p^+ = \langle p-\rangle[-p], \; p^- = \langle +p\rangle[p+], \; p^t = \langle +p\rangle[-p], \; \bar{p}^t = \langle p-\rangle[p+], \quad \langle +-\rangle = [-+] = 1$$

Spacecone

axial gauges	*non-null*	*null (+ auxiliary field eq.)*
(partly) temporal	*timelike* : $A^0 = 0$	*lightcone* : $A^+ = 0, \; \delta/\delta A^-$
spacelike	*Arnowitt-Fickler* : $A^1 = 0$	*spacecone* : $A^t = 0, \quad \delta/\delta\bar{A}^t$
scalar	*unitary* : $\phi = \phi^\dagger$	*Gervais-Neveu* : $\phi = \langle\phi\rangle, \; \delta/\delta\phi^\dagger$

$$n\cdot A = 0, \quad n = |+\rangle[-|; \qquad tree \;\sim\; \langle\;\rangle^{2-E_+}[\;]^{2-E_-}$$

$$L = L_2 + L_3 + L_4$$

$$L_2 = A^+(-\tfrac{1}{2}P^2)A^- + \psi^+\frac{-\frac{1}{2}P^2}{p}\psi^-$$

$$L_3 = \left(\frac{p^\mp}{p}A^\pm\right)([A^\pm, pA^\mp] + \{\psi^+,\psi^-\}) + \left(\frac{p^\mp}{p}\psi^\pm\right)[A^\pm,\psi^\mp]$$

$$L_4 = ([A^+, pA^-] + \{\psi^+,\psi^-\})\frac{1}{p^2}([A^-, pA^+] + \{\psi^+,\psi^-\}) - [A^+,\psi^-]\frac{1}{p}[A^-,\psi^+]$$

$$A^+ = \frac{[-p]}{\langle +p\rangle}, \qquad A^- = \frac{\langle +p\rangle}{[-p]}; \qquad \psi^+ = [-p], \qquad \psi^- = \langle +p\rangle$$

$$ref.\;lines: \quad \frac{p^-}{p}A^+ = \frac{p^+}{p}A^- = \frac{p^-}{p}\psi^+ = \frac{p^+}{p}\psi^- = 1$$

$$P_\oplus = |-\rangle[-|, \quad P_\ominus = |+\rangle[+|; \qquad P_\oplus^a = \delta_-^a, \quad P_\ominus^a = \delta_+^a$$

Background-field

$$\phi \to \varphi + \phi; \qquad \nabla \to \mathcal{D} + iA, \qquad F_{ab} \to \mathcal{F}_{ab} + \mathcal{D}_{[a}A_{b]} + i[A_a, A_b]$$

$$\partial\cdot A \to \mathcal{D}\cdot A, \qquad \tilde{C}\partial\cdot\nabla C \to \tilde{C}\mathcal{D}^2 C + \tilde{C}\mathcal{D}\cdot i[A,C]$$

$$1-loop: \qquad K = -\tfrac{1}{2}(\Box - i\mathcal{F}^{ab}S_{ba})$$

SUPERSYMMETRY (II,IV,VI): Superspace

$$q_\alpha = -i\left(\frac{\partial}{\partial\theta^\alpha} - \tfrac{1}{2}\bar\theta^{\dot\beta}p_{\alpha\dot\beta}\right), \qquad \bar q_{\dot\alpha} = -i\left(\frac{\partial}{\partial\bar\theta^{\dot\alpha}} - \tfrac{1}{2}\theta^\beta p_{\beta\dot\alpha}\right); \qquad (q^\alpha)^\dagger = \bar q^{\dot\alpha}$$

$$d_\alpha = \frac{\partial}{\partial\theta^\alpha} + \tfrac{1}{2}\bar\theta^{\dot\beta}p_{\alpha\dot\beta}, \qquad \bar d_{\dot\alpha} = \frac{\partial}{\partial\bar\theta^{\dot\alpha}} + \tfrac{1}{2}\theta^\beta p_{\beta\dot\alpha}; \qquad (d^\alpha)^\dagger = -\bar d^{\dot\alpha}$$

$$\{q^\alpha, \bar q^{\dot\beta}\} = \{d^\alpha, \bar d^{\dot\beta}\} = p^{\alpha\dot\beta}, \qquad \int d^2\theta = d^2 = \tfrac{1}{2}d^\alpha d_\alpha, \qquad \bar d^2 d^2 \bar d^2 = \tfrac{1}{2}\Box\bar d^2$$

$$\partial_M = (\partial_\mu, \partial_{\dot\mu}, \partial_m) = \partial/\partial z^M, \qquad z^M = (\theta^\mu, \bar\theta^{\dot\mu}, x^m)$$

$$d_A = (d_\alpha, \bar d_{\dot\alpha}, \partial_{\alpha\dot\alpha}) = E_A{}^M\partial_M; \qquad [d_A, d_B\} = T_{AB}{}^C d_C$$

$$T_{\alpha\dot\beta}{}^{\gamma\dot\gamma} = T_{\dot\beta\alpha}{}^{\gamma\dot\gamma} = -i\delta_\alpha^\gamma\delta_{\dot\beta}^{\dot\gamma}, \quad rest = 0$$

Super Yang-Mills

$$\nabla_A = d_A + iA_A, \qquad [\nabla_A, \nabla_B\} = T_{AB}{}^C\nabla_C + iF_{AB}$$

$$\overline\nabla_{\dot\alpha}\phi = 0 \quad\Rightarrow\quad \{\nabla_\alpha, \nabla_\beta\} = \{\overline\nabla_{\dot\alpha}, \overline\nabla_{\dot\beta}\} = 0; \qquad \{\nabla_\alpha, \overline\nabla_{\dot\beta}\} = -i\nabla_{\alpha\dot\beta}$$

$$[\nabla_\alpha, \nabla_{\beta\dot\gamma}] = C_{\alpha\beta}\overline W_{\dot\gamma}; \qquad \overline\nabla_{\dot\alpha}W_\beta = 0, \quad \nabla^\alpha W_\alpha + \nabla^{\dot\alpha}W_{\dot\alpha} = 0$$

$$[\nabla_{\alpha\dot\alpha}, \nabla_{\beta\dot\beta}] = i(C_{\alpha\beta}\bar f_{\dot\alpha\dot\beta} + C_{\dot\alpha\dot\beta}f_{\alpha\beta}), \qquad f_{\alpha\beta} = \tfrac{1}{2}\nabla_{(\alpha}W_{\beta)}$$

Actions

$$L_{N=1} = -\tfrac{1}{g^2}tr\int d^2\theta\ \tfrac{1}{2}W^\alpha W_\alpha + \zeta\int d^4\theta\ V - \int d^4\theta\ \bar\phi e^V\phi + \left[\int d^2\theta\ f(\phi) + h.c.\right]$$

$$L_{N=2} = -\tfrac{1}{g^2}tr\left(\int d^2\theta\ W^2 + \int d^4\theta\ e^{-V}\bar\phi e^V\phi\right) + \int d^4\theta\ \zeta_0 V + \left(\int d^2\theta\ \zeta_+\phi + h.c.\right)$$

$$- \int d^4\theta\ \bar\phi^{i'}(e^{V\tau})_{i'}{}^{j'}\phi_{j'} + \tfrac{1}{2}\left[\int d^2\theta\ \tau^{i'j'}\phi_{i'}(\phi + M)\phi_{j'} + h.c.\right]$$

$$L_{N=4} = \tfrac{1}{g^2}tr\left[-\int d^2\theta\ W^2 - \int d^4\theta\ e^{-V}\bar\phi^I e^V\phi_I + \left(\int d^2\theta\ \tfrac{1}{6}\epsilon^{IJK}\phi_I[\phi_J, \phi_K] + h.c.\right)\right]$$

Supergraphs

(A2$\tfrac{1}{2}$) θ's: one for each vertex, with an $\int d^4\theta$.

(A3$'$) Propagators:

$$(VV, \bar\phi\phi, \phi\phi, \bar\phi\bar\phi): \qquad \left(1, -1, \tfrac{m}{\sqrt2}\frac{d^2}{-\frac{1}{2}p^2}, \tfrac{m}{\sqrt2}\frac{\bar d^2}{-\frac{1}{2}p^2}\right)\frac{1}{\frac{1}{2}(p^2+m^2)}\delta^4(\theta - \theta')$$

(A4$\tfrac{1}{2}$) Chiral vertex factors: $\bar d^2$ on the ϕ end(s) of every chiral propagator, d^2 on the $\bar\phi$ end(s), but drop any one such factor at a superpotential vertex.

LOOPS (VII,VIII): Gamma function

$$\Gamma(z) = \int_0^\infty d\lambda\ \lambda^{z-1}e^{-\lambda} = \tfrac{1}{z}e^{-\gamma z}\prod_{n=1}^{\infty}\frac{1}{1+\frac{z}{n}}e^{z/n} = \tfrac{1}{z}exp\left[-\gamma z + \sum_{n=2}^{\infty}\tfrac{1}{n}\zeta(n)(-z)^n\right]$$

$$\gamma = \lim_{n\to\infty}\left(-ln\ n + \sum_{m=1}^{n}\tfrac{1}{m}\right) = 0.5772156649..., \qquad \zeta(z) = \sum_{n=1}^{\infty}\frac{1}{n^z}$$

$$\Gamma(z+1) = z\Gamma(z), \qquad \Gamma(z)\Gamma(1-z) = \pi\ csc(\pi z), \qquad \frac{\Gamma(z)}{\Gamma(2z)} = \frac{2^{1-2z}\sqrt{\pi}}{\Gamma(z+\frac{1}{2})}$$

$$\Gamma(n+1) = n!, \qquad \Gamma(n+\tfrac{1}{2}) = (n-\tfrac{1}{2})(n-\tfrac{3}{2})...\tfrac{1}{2}\sqrt{\pi} = \frac{(2n)!}{n!2^{2n}}\sqrt{\pi}$$

$$\lim_{z\to\infty}\Gamma(z) \approx \sqrt{\frac{2\pi}{z}}\left(\frac{z}{e}\right)^z, \qquad \lim_{z\to\infty}\Gamma(az+b) \approx \sqrt{2\pi}(az)^{az+b-1/2}e^{-az}$$

$$B(x,y) = \int_0^1 dz\ z^{x-1}(1-z)^{y-1} = \int_0^\infty d\tau\ \tau^{x-1}(1+\tau)^{-x-y} = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

Regularization

$$\frac{\Gamma(a)}{[\frac{1}{2}(p^2+m^2)]^a} = \int_0^\infty d\tau\ \tau^{a-1}e^{-\tau(p^2+m^2)/2}$$

$$1 = \int_0^\infty d\lambda\ \delta\left(\lambda - \sum\tau_i\right), \qquad \tau_i = \lambda\alpha_i$$

$$\int dk\ e^{-k^2/2} = 1, \qquad \int dk\ \frac{k_a...k_b}{(\frac{1}{2}k^2)^a} = 0$$

$$\int dk\ e^{ik\cdot x}\frac{1}{\frac{1}{2}(k+\frac{1}{2}p)^2\frac{1}{2}(k-\frac{1}{2}p)^2} =$$

$$(\tfrac{1}{2}p^2)^{D/2-2}\Gamma(\tfrac{D}{2}-1)\Gamma(2-\tfrac{D}{2})\sum_{n=0}^{\infty}\frac{\Gamma(n+\frac{D}{2}-1)}{n!\Gamma(2n+D-2)}\{\tfrac{1}{4}[p^2x^2-(p\cdot x)^2]\}^n$$

$$\int dk\ \frac{\Gamma(a)}{(\frac{1}{2}k^2)^a}\frac{\Gamma(b)}{[\frac{1}{2}(k+p)^2]^b} = \frac{\Gamma(a+b-\frac{D}{2})}{(\frac{1}{2}p^2)^{a+b-D/2}}B(\tfrac{D}{2}-a,\tfrac{D}{2}-b)$$

Schemes

$$\mathrm{MS}: \quad \hbar$$

$$\overline{\mathrm{MS}}: \quad \Gamma(\tfrac{D}{2})\hbar, \quad \Gamma(1-\epsilon)\hbar, \quad etc.$$

$$\mathrm{G}: \quad \frac{(-1)^{D_0/2}}{\epsilon\Gamma(2-\frac{D}{2})B(\frac{D}{2}-1,\frac{D}{2}-1)}\hbar, \quad \frac{\Gamma(1-2\epsilon)}{\Gamma(1+\epsilon)[\Gamma(1-\epsilon)]^2}\hbar, \quad etc.$$

Running coupling

$$\Gamma_{1,2g} \approx tr\int dx\ \tfrac{1}{8}F^{ab}\beta_1(ln\,\Box - \tfrac{1}{\epsilon})F_{ab}; \quad \beta_1 = \tfrac{1}{2}c_R(-1)^{2s}(4s^2-\tfrac{1}{3});\ \ c_{D+\bar{D}} = 2,\ c_A = 2N$$

$$\Gamma_M \approx tr\int dx\ \tfrac{1}{8}F\left[\beta_1 ln\frac{\Box}{M^2} + \frac{\beta_2}{\beta_1}ln\left(\beta_1 ln\frac{\Box}{M^2}\right)\right]F, \qquad \frac{\mu^2}{M^2} = e^{1/\beta_1 g^2}(g^2)^{\beta_2/\beta_1^2}$$

GRAVITY (IX)

$$[M_{ab}, V_c] = V_{[a}\eta_{b]c} \quad \Rightarrow \quad \tfrac{1}{2}\lambda^{ab}[M_{ba}, V_c] = \lambda_c{}^a V_a, \qquad [M_{ab}, M^{cd}] = -\delta_{[a}^{[c} M_{b]}{}^{d]}$$

$$\tfrac{1}{2}\lambda^{ab}M_{ba} = \tfrac{1}{2}\lambda^{\alpha\beta}M_{\beta\alpha} + \tfrac{1}{2}\lambda^{\dot{\alpha}\dot{\beta}}M_{\dot{\beta}\dot{\alpha}}$$

$$[M_{\alpha\beta}, \psi_\gamma] = \psi_{(\alpha}C_{\beta)\gamma} \quad \Rightarrow \quad \tfrac{1}{2}\lambda^{\alpha\beta}[M_{\beta\alpha}, \psi_\gamma] = \lambda_\gamma{}^\alpha \psi_\alpha, \qquad [M_{\alpha\beta}, M^{\gamma\delta}] = \delta_{(\alpha}^{(\gamma} M_{\beta)}{}^{\delta)}$$

$$\nabla_a = e_a + \tfrac{1}{2}\omega_a{}^{bc}M_{cb}, \quad e_a = e_a{}^m \partial_m; \qquad [\nabla_a, \nabla_b] = T_{ab}{}^c \nabla_c + \tfrac{1}{2}R_{ab}{}^{cd}M_{dc}$$

$$-ds^2 = dx^m dx^n g_{mn}, \qquad g_{mn} = e_m{}^a e_n{}^b \eta_{ab}$$

$$S = \int dx \ \mathbf{e}^{-1} L, \qquad \mathbf{e} = det \ e_a{}^m \qquad L = -\tfrac{1}{4}R + L(\nabla, \psi), \qquad R = R_{ab}{}^{ab}$$

$$R_{ab} - \tfrac{1}{2}\eta_{ab}R = 2T_{ab}, \qquad R_{ab} = R^c{}_{acb}, \qquad \delta S_M = \int dx \ \mathbf{e}^{-1}(e_m{}^a \delta e^{bm})T_{ab}$$

Methods

$$\begin{aligned}
(1): &\qquad D = 1 \quad \Rightarrow \quad \nabla = \partial \\
(2): &\qquad ds^2 = ds_1^2 + ds_2^2 \quad \Rightarrow \quad \nabla = (\nabla_1, \nabla_2) \\
(3): &\qquad \nabla_a(x) \to \nabla_a(x') \\
(4): &\qquad ds'^2 = \Phi^{-2}ds^2 \quad \Rightarrow \quad \nabla'_a = \Phi\nabla_a + (\nabla^b\Phi)M_{ab}, \\
&\qquad R'_{ab}{}^{cd} = \Phi^2 R_{ab}{}^{cd} + \Phi\delta_{[a}^{[c}\nabla_{b]}\nabla^{d]}\Phi - \delta_{[a}^c\delta_{b]}^d(\nabla\Phi)^2
\end{aligned}$$

$$[M_{12}, V_2] = \eta_{22}V_1, \qquad [M_{12}, M_{23}] = \eta_{22}M_{13}$$

$$[\nabla_1, \nabla_2] = [e_1 + \omega_1, e_2 + \omega_2]$$

$$= \{[e_1, e_2] + (e_1\omega_2)M_2 - (e_2\omega_1)M_1\} + \{\omega_1[M_1, \nabla_2] - \omega_2[M_2, \nabla_1] - \omega_1\omega_2[M_1, M_2]\}$$

Examples

$$L_G = -\tfrac{1}{4}\mathbf{e}^{-1}R = \tfrac{1}{16}\epsilon^{mnpq}\epsilon_{abcd}e_m{}^a e_n{}^b R_{pq}{}^{cd} = \tfrac{1}{8}\epsilon^{mnpq}e_m{}^a e_n{}^b \tilde{R}_{pqab}$$

$$= i\tfrac{1}{8}\epsilon^{mnpq}e_m{}^{\alpha\dot{\alpha}}(\bar{R}_{np\dot{\alpha}}{}^{\dot{\beta}}e_{q\alpha\dot{\beta}} - R_{np\alpha}{}^\beta e_{q\beta\dot{\alpha}})$$

$$S_G = 4\tfrac{D-1}{D-2}\int dx \ \mathbf{e}^{-1}\tfrac{1}{4}\phi(\Box - \tfrac{1}{4}\tfrac{D-2}{D-1}R)\phi$$

$$-ds^2 = -\left(1 - \frac{2GM}{r}\right)dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1}dr^2 + r^2(d\theta^2 + sin^2\theta \ d\phi^2)$$

SUPERGRAVITY (X)

$$\nabla_A = E_A{}^M \partial_M + \tfrac{1}{2}\Omega_A{}^{\beta\gamma} M_{\gamma\beta} + \tfrac{1}{2}\Omega_A{}^{\dot\beta\dot\gamma} M_{\dot\gamma\dot\beta} + iA_A Y, \qquad [Y, \nabla_\alpha] = -\tfrac{1}{2}\nabla_\alpha$$

$$[\nabla_A, \nabla_B\} = T_{AB}{}^C \nabla_C + \tfrac{1}{2}R_{AB}{}^{\gamma\delta} M_{\delta\gamma} + \tfrac{1}{2}R_{AB}{}^{\dot\gamma\dot\delta} M_{\dot\delta\dot\gamma} + iF_{AB} Y$$

$$\{\overline\nabla_{\dot\alpha}, \overline\nabla_{\dot\beta}\} = B\overline{M}_{\dot\alpha\dot\beta}, \qquad \{\nabla_\alpha, \overline\nabla_{\dot\beta}\} = -i\nabla_{\alpha\dot\beta}$$

$$[\overline\nabla_{\dot\alpha}, -i\nabla_{\beta\dot\beta}] = C_{\dot\beta\dot\alpha}\mathcal{W}_\beta - \tfrac{1}{2}(\nabla_\beta B)\overline{M}_{\dot\alpha\dot\beta}, \qquad [-i\nabla_{\alpha\dot\alpha}, -i\nabla_{\beta\dot\beta}] = C_{\dot\beta\dot\alpha} f_{\alpha\beta} - h.c.$$

$$\begin{aligned}
\mathcal{W}_\alpha = &\ - B\nabla_\alpha - G_\alpha{}^{\dot\beta}\overline\nabla_{\dot\beta} + \tfrac{1}{2}(\overline\nabla^{\dot\beta} G_\alpha{}^{\dot\gamma})\overline{M}_{\dot\beta\dot\gamma} + \tfrac{1}{2}W_\alpha{}^{\beta\gamma}M_{\gamma\beta} + iW_\alpha Y + i\tfrac{1}{6}W^\beta M_{\beta\alpha} \\
f_{\alpha\beta} = &\ i\tfrac{1}{2}G_{(\alpha}{}^{\dot\gamma}\nabla_{\beta)\dot\gamma} - \tfrac{1}{2}(\nabla_{(\alpha}B + i\tfrac{1}{3}W_{(\alpha)}\nabla_{\beta)} + W_{\alpha\beta}{}^\gamma\nabla_\gamma - \tfrac{1}{2}(\nabla_{(\alpha}G_{\beta)}{}^{\dot\gamma})\overline\nabla_{\dot\gamma} \\
&- (\tfrac{1}{2}\overline\nabla^2\bar B + B\bar B + \tfrac{1}{12}i\nabla^\gamma W_\gamma)M_{\alpha\beta} - i\tfrac{1}{16}[(\nabla_{(\alpha}{}^{\dot\delta}G_{\gamma)\dot\delta})M_\beta{}^\gamma + \alpha\leftrightarrow\beta] \\
&+ \tfrac{1}{2}W_{\alpha\beta}{}^{\gamma\delta}M_{\delta\gamma} + \tfrac{1}{4}(\nabla_{(\alpha}\overline\nabla^{\dot\gamma}G_{\beta)}{}^{\dot\delta})\overline{M}_{\dot\delta\dot\gamma} + i\tfrac{1}{2}(\nabla_{(\alpha}W_{\beta)})Y \\
W_{\alpha\beta\gamma\delta} = &\ \tfrac{1}{4!}\nabla_{(\alpha}W_{\beta\gamma\delta)}
\end{aligned}$$

$$G_a = \bar G_a, \qquad \overline\nabla_{\dot\alpha}B = \overline\nabla_{\dot\alpha}W_\alpha = \overline\nabla_{\dot\alpha}W_{\alpha\beta\gamma} = 0, \qquad \overline\nabla^{\dot\alpha}G_{\alpha\dot\alpha} = \nabla_\alpha B - iW_\alpha$$

$$\nabla^\alpha W_{\alpha\beta\gamma} - i\tfrac{1}{3}\nabla_{(\beta}W_{\gamma)} = -i\tfrac{1}{2}\nabla_{(\beta}{}^{\dot\alpha}G_{\gamma)\dot\alpha}, \qquad \nabla^\alpha W_\alpha + \overline\nabla^{\dot\alpha}\overline{W}_{\dot\alpha} = 0$$

Ectoplasm

$$S = \int dx\ (-\tfrac{1}{4!})\epsilon^{mnpq} e_q{}^D e_p{}^C e_n{}^B e_m{}^A L_{ABCD}$$

$$L_{\alpha\beta cd} = \epsilon_\alpha{}^{\dot\alpha}{}_{,\beta\dot\alpha,cd}\bar{\mathcal{L}}, \qquad L_{\alpha bcd} = i\epsilon_{\alpha\dot\alpha,bcd}\overline\nabla^{\dot\alpha}\bar{\mathcal{L}}, \qquad L_{abcd} = \epsilon_{abcd}[(\overline\nabla^2 + 3B)\bar{\mathcal{L}} + h.c.]$$

$$\overline\nabla_{\dot\alpha}\mathcal{L} = 0 \quad \Rightarrow \quad \mathcal{L} = (\overline\nabla^2 + B)\mathbf{L}$$

Action

$$S_{SG,c} = 3\int dx\ d^4\theta\ E^{-1}\bar\Phi\Phi$$

$$L_{SG} = L_G + L_\psi + \mathbf{e}^{-1}L_a$$

$$L_G = -\tfrac{1}{4}\mathbf{e}^{-1}R, \qquad L_\psi = \epsilon^{mnpq}\bar\psi_{m\dot\alpha}\tfrac{1}{2}\{e_n{}^{\alpha\dot\alpha}, \nabla_p\}\psi_{q\alpha}, \qquad L_a = -\tfrac{3}{8}(G_a)^2 + 3\bar BB$$

$$\omega_{mbc} = e_m{}^a[\mathring\omega_{abc} - \tfrac{1}{2}(\hat T_{bca} - \hat T_{a[bc]})], \qquad \hat T_{ab}{}^c = \epsilon_{ab}{}^{cd}G_d + i\psi_{[a}{}^\gamma\bar\psi_{b]}{}^{\dot\gamma}$$

$$\delta e_m{}^{\alpha\dot\alpha} = -i(\epsilon^\alpha\bar\psi_m{}^{\dot\alpha} + \bar\epsilon^{\dot\alpha}\psi_m{}^\alpha), \qquad \delta\psi_m{}^\alpha = \nabla_m\epsilon^\alpha + ie_m{}^{\beta\dot\beta}(\epsilon_\beta G^\alpha{}_{\dot\beta} - \bar\epsilon_{\dot\beta}\delta_\beta^\alpha B)$$

$$\delta B = -\tfrac{2}{3}\epsilon^\alpha t_{\alpha\beta}{}^\beta, \qquad \delta G_{\alpha\dot\alpha} = -\epsilon^\beta(t_{\beta\alpha\dot\alpha} + \tfrac{1}{3}C_{\beta\alpha}t_{\dot\gamma\dot\beta}{}^{\dot\beta}) + h.c.$$

$$T_{ab}{}^\gamma = -e_a{}^m e_b{}^n\nabla_{[m}\psi_{n]}{}^\gamma + i(\psi_{a\beta}G^\gamma{}_{\dot\beta} - \bar\psi_{a\dot\beta}\delta_\beta^\gamma B - a\leftrightarrow b) \equiv C_{\alpha\beta}t_{\dot\alpha\dot\beta}{}^\gamma + C_{\dot\alpha\dot\beta}t_{\alpha\beta}{}^\gamma$$

STRINGS (XI)

$$S_L = \frac{1}{\alpha'}\int \frac{d^2\sigma}{2\pi}\sqrt{-g}g^{mn}\tfrac{1}{2}(\partial_m X^a)\cdot(\partial_n X^b)\eta_{ab}$$

$$\alpha(s) = \kappa\alpha' s + \frac{1}{\kappa}, \qquad \kappa = \begin{cases} 1 & (open) \\ \frac{1}{2} & (closed) \end{cases}$$

$$\begin{aligned}
\mathcal{A}_4 &= g^2 \int_0^1 dz\ z^{-\alpha(s)-1}(1-z)^{-\alpha(t)-1} \\
&= g^2 B[-\alpha(s),-\alpha(t)] = g^2\frac{\Gamma[-\alpha(s)]\Gamma[-\alpha(t)]}{\Gamma[-\alpha(s)-\alpha(t)]} \\
&= \sum_{j=0}^{\infty}\left\{\frac{[\alpha(s)+j][\alpha(s)+j-1]\cdots[\alpha(s)+1]}{j!}\right\}\frac{1}{j-\alpha(t)} \\
&= \left(in \lim_{\substack{s\to-\infty \\ t\ fixed}}\right)\ g^2\Gamma[-\alpha(t)][-\alpha(s)]^{\alpha(t)} \\
&\sim \left(in \lim_{\substack{s\to-\infty \\ \theta\ fixed}}\right)\ e^{-f(cos\ \theta)\alpha(s)}; \quad cos\ \theta \approx 1 + 2\frac{t}{s}, \\
&\qquad f \approx \frac{t}{-s}ln\left(\frac{-s}{t}\right) + \frac{u}{-s}ln\left(\frac{-s}{u}\right)
\end{aligned}$$

Conformal field theory

$$\langle 0|\phi_{i(\pm)}(\sigma^m)\phi_{j(\pm)}(\sigma'^m)|0\rangle = -\eta_{ij}[i\ ln(\sigma-\sigma')^{\pm}]$$

$$e^{ia\cdot\phi_{(\pm)}(\sigma^m)}e^{ib\cdot\phi_{(\pm)}(\sigma'^m)} = \left[i(\sigma^{\pm}-\sigma'^{\pm})\right]^{a\cdot b}e^{ia\cdot\phi_{(\pm)}(\sigma^m)+ib\cdot\phi_{(\pm)}(\sigma'^m)} + \ldots$$

$$L = \tfrac{1}{4}\eta^{ij}(\nabla\phi_i)\cdot(\nabla\phi_j) + \tfrac{1}{2}R\mu^i\phi_i \quad \Rightarrow \quad T_{\pm\pm} = \tfrac{1}{2}\eta^{ij}(\partial_\pm\phi_{i(\pm)})(\partial_\pm\phi_{j(\pm)}) - \mu^i\partial_\pm^2\phi_{i(\pm)}$$

$$\left[i\int\frac{d\sigma'}{2\pi}\lambda(\sigma')T_{\pm\pm}(\sigma'),\chi(\sigma)\right] = \lambda(\sigma)\chi'(\sigma) + w_{(\pm)}\lambda'(\sigma)\chi(\sigma)$$

$$i\frac{1}{2\pi}[T_{\pm\pm}(1),\chi(2)] = \delta(2-1)\chi'(2) + w_{(\pm)}\delta'(2-1)\chi(2)$$

$$w_{(\pm)}\left(e^{ia\cdot\phi_{(\pm)}}\right) = \tfrac{1}{2}a^2 + i\mu\cdot a, \qquad w_{(\mp)} = 0$$

$$anomaly\ \sim\ 6(w-\tfrac{1}{2})^2 - \tfrac{1}{2}$$

INDEX

Topics are listed by subsection (or section, etc.).

Comments on Warren Siegel's **Fields:**

"The price is right."

"Oh, Warren, what have you done to your students now?"

"I can see you put a lot of work into it."

"That's nice, honey."

"You might want to add a reference to my paper..."

"It's different."

"Is this going to be on the exam?"

"I'll have a look at it when I get the time."

"Aren't there enough field theory books already?"

"So this is why you haven't written any papers lately."

"Where are the jokes?"