GAUGE TORSION GRAVITY, STRING THEORY,

AND ANTISYMMETRIC TENSOR INTERACTIONS

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By
Terry Glenn Pilling

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Major Department:
Physics

April 2002

Fargo, North Dakota

UMI Number: 3103621

UMI Microform 3103621

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

North Dakota State University

Graduate School

Title

GAUGE TORSION GRAVITY, STRING THEORY,

AND ANTISYMMETRIC TENSOR INTERACTIONS

By

TERRY GLENN PILLING

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Chair

Approved by Department Chair:

16 May 02

Date

Signature

ABSTRACT

Pilling, Terry Glenn, Ph.D., Department of Physics, College of Science and Mathematics, North Dakota State University, April 2002. Gauge Torsion Gravity, String Theory, and Antisymmetric Tensor Interactions. Major Professor: Dr. Patrick F. Kelly.

The antisymmetric tensor field is derived in the context of general relativity with torsion as well as the context of string theory. The interaction between antisymmetric tensor fields and fermion fields is examined. The tree level scattering amplitude and the differential and total cross section for massless fermions are derived. The one-loop contribution of torsion exchange to the fermion anomalous magnetic moment is shown to present a solution to a recent problem with the standard model of particle physics. The experimental discrepancy is used to place an upper bound on the torsion coupling to matter.

ACKNOWLEDGMENTS

I would like to thank my adviser, Dr. Patrick Kelly, who invited me to North Dakota State University and has been an inspiration and a friend. I am very grateful to the faculty and fellow graduate students. In particular, I am indebted to Dr. Richard Hammond, Patty Hartsoch, Scott Atkins, Feng Hong, Nathan Schoenack, Bin Lu, Tim Storsved, Liess Vantine, and Darren Evans for their assistance and friendship. Special thanks to Bonnie Cooper for meticulously reading the manuscript. Her detailed comments and suggestions have improved this dissertation immensely. Thanks to Dave Hornidge, Trevor Fulton, Darren White, and Derek Harnett for their valued friendship, support, humor, and daily political debates via email. I thank Glenn and Linda Pilling, Rick Pilling, Tammy and Wes Schock, as well as Tyler Schock and Danielle Schock for their love and interest throughout this long trek. Finally, I thank Melanie for her love, support, and amazing patience. Without her, the completion of this dissertation would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

To verify a theory of gravity or particle physics experimentally, one must make physical observations. These observations are made using photons of light or some other intermediate particle quanta. We detect photons or other particle quanta from stars, galaxies, supernovae, and particle physics experiments. From these observations, we are able to test the predictions of candidate theories.

In order for a theory of gravity be tested, one actually needs a theory of gravity coupled to electromagnetism [1]. The coupled Einstein-Maxwell system correctly describes a wealth of experiments, such as the gravitational bending of light, the gravitational red shift, the time delay of radar pulses in the gravitational field of the sun, and the lensing and microlensing of starlight in the gravitational field of galaxies. In all of these experiments, we are studying the propagation of light along null-geodesics in a gravitational field which is the solution of Einstein's *vacuum* field equations, *not* the electro-vacuum equations. In other words, we are treating gravity as a background metric with the motion of a photon described by a null-geodesic on this background. A truly novel effect of the Einstein-Maxwell theory would be, for example, the generation of electromagnetic waves by gravitational waves. Because of their smallness, such effects have never been observed. We would need very high energy particles or gravitational fields before this aspect of the Einstein-Maxwell theory could be tested directly.

When we test theories of gravity, we are testing theories of particles traveling in a fixed gravitational background, thus any theory of gravity which reduces to the same background theory at low energy must also be considered as a viable candidate theory of nature. In particular, Einstein-Cartan (EC) gravity [2, 3, 4], metric affine (MA) gravity [5], superstring theory [6] and 11-dimensional supergravity and M-theory [7] have this property. In order to decide whether one of these theories is the correct theory of nature, we must find predictions of the theory which differ from predictions of competing theories. If these predictions turn out to be verified experimentally, we would be able to discard the non-complying theories.

In this dissertation, we discuss the existence and interactions of a particle called by various sources the Kalb-Ramond antisymmetric tensor field, the axion, or the torsion tensor. Theories such as EC-gravity, MA-gravity, superstring theory, and supergravity predict the existence of such a particle and propose interactions between it and other particles of nature. It is interesting that the conventional Einstein gravity theory, general relativity, does not naturally allow for this type of particle interaction, thus it is important to test the predictions of this possibility.

In Chapter 2, we introduce the formalism of Einstein-Cartan gravity which is a generalization of Einstein gravity to allow a connection which is not symmetric. We will see that Einstein-Cartan gravity contains a gravitational "torsion" interaction between particles with spin.

In Chapter 3, we derive the interaction of torsion with spinning particles from the gauge principle in analogy to the derivation of the gauge fields mediating the strong and electroweak theories of particle physics.

Chapter 4 details how the antisymmetric tensor field and its interactions arise in superstring theory and supergravity theories. Many theoretical physicists around the world hope that a theory of this type turns out to be the correct theory of nature, hence it is important to find predictions of the theory that can be tested in the laboratory.

We examine the interactions between antisymmetric tensor fields and fermion fields in Chapter 5. We calculate scattering amplitudes and use a current high-precision problem with the standard model as an example of where this type of interaction may be evident. We use the experimental and theoretical discrepancy for the anomalous magnetic moment of the muon to set a bound on the antisymmetric tensor coupling to fermions.

Finally, we summarize our findings in Chapter 6 and propose future work. In order to keep the main text of this dissertation brief and concise, we have moved many of the details and much of the background material to Appendices. This material is included for completeness and to set our notation conventions, and may be necessary to the reader who finds any of the material in the main body of the dissertation unfamiliar.

CHAPTER 2

EINSTEIN-CARTAN GRAVITY

We begin this chapter with a discussion of differential geometry [5, 8, 9] and the general theory of relativity, and show how non-zero spacetime torsion in Einstein-Cartan gravity can give rise to an antisymmetric tensor field.

2.1. Riemannian manifolds and Cartan's equations

Suppose we are given a 4-manifold, M, and a metric, $g_{\mu\nu}(x)$, on M in local coordinates x^μ. The distance, ds, between two infinitesimally near points, x^μ and $x^\mu + dx^\mu$, is given by

$$ds^2 = g_{\mu\nu}(x)dx^\mu dx^\nu. \tag{2.1}$$

We decompose the metric into vierbeins or tetrads,[1] ${e_\mu}^a(x)$, as

$$\begin{aligned} g_{\mu\nu} &= \eta_{ab}{e_\mu}^a{e_\nu}^b \\ \eta^{ab} &= g^{\mu\nu}{e_\mu}^a{e_\nu}^b, \end{aligned} \tag{2.2}$$

where η_{ab} is a flat metric (such as δ_{ab} in Euclidean space). In this fashion, we can isolate the information about the curvature of space into the tetrad.

We raise and lower Greek indices with $g_{\mu\nu}$ or its inverse, $g^{\mu\nu}$, and Latin indices with

[1] We will use the terms "vierbein" and "tetrad" interchangeably throughout this dissertation.

η_{ab} or η^{ab}. We define the inverse of $e_\mu{}^a$ by

$$e^\mu{}_a = \eta_{ab} g^{\mu\nu} e_\nu{}^b \tag{2.3}$$

which obeys

$$\begin{aligned} e^\mu{}_a e_\mu{}^b &= \delta^b_a \\ \eta^{ab} e^\mu{}_a e^\nu{}_b &= g^{\mu\nu} \quad \text{etc.} \end{aligned} \tag{2.4}$$

In this fashion, we can use $e_\mu{}^a$ and $e^\mu{}_a$ to convert between Greek and Latin indices on tensorial quantities.

The tetrad, $e^\mu{}_a$, is a transformation from the basis, $\partial/\partial x^\mu$, of the tangent space, $T_x M$, of a manifold M to an orthonormal basis of $T_x M$,

$$e_a = e^\mu{}_a \frac{\partial}{\partial x^\mu}.$$

Similarly, $e_\mu{}^a$ is the matrix which transforms the coordinate basis, dx^μ, of the cotangent space $T^*_x M$ to an orthonormal basis of $T^*_x M$,

$$e^a = e_\mu{}^a dx^\mu.$$

While the coordinate basis, dx^μ, is always an exact differential, e^a is not necessarily an exact 1-form, and while $\partial/\partial x^\mu$ and $\partial/\partial x^\nu$ commute, e_a and e_b do not necessarily commute

$$[e_a, e_b] = e^{\mu}{}_a \left(\partial_\mu e_b\right) - e^{\nu}{}_b \left(\partial_\nu e_a\right). \tag{2.5}$$

The object e_a, with lowered index, is called a *frame*. The object e^a, with raised index, is called a *coframe*. Latin indices, $a, b, ...$, are called *anholonomic* or frame indices while the Greek indices, $\mu, \nu, ...$, are called *holonomic* or coordinate indices.

Define the torsion 2-form T^a and the curvature 2-form $R_b{}^a$ of the manifold in terms of the spin connection 1-form $\omega_b{}^a$ as follows:

$$T^a = De^a = de^a + \omega_b{}^a \wedge e^b = e^b \wedge K_b{}^a = \frac{1}{2} T_{bc}{}^a e^b \wedge e^c \tag{2.6}$$

and

$$R_a{}^b = D\omega_a{}^b = d\omega_a{}^b - \omega_a{}^c \wedge \omega_c{}^b = \frac{1}{2} R_{cda}{}^b e^c \wedge e^d, \tag{2.7}$$

where de^a is sometimes called the *anholonomity* 2-form. We have implicitly defined the *contortion* 1-form $K_b{}^a = -K^a{}_b$ and the covariant derivative of a 1-form De^a. Equations (2.6) and (2.7) are called Cartan's structure equations. Taking the exterior derivative of each of these equations gives the consistency conditions and the Bianchi identities, respectively, as

$$\begin{aligned}
dT^a &= d\left(de^a + \omega_b{}^a \wedge e^b\right) = d\omega_b{}^a \wedge e^b - \omega_b{}^a \wedge de^b \\
&= d\omega_b{}^a \wedge e^b - \omega_b{}^a \wedge \left(T^b - \omega_c{}^b \wedge e^c\right) \\
&= d\omega_b{}^a \wedge e^b - \omega_b{}^a \wedge T^b + \omega_b{}^a \wedge \omega_c{}^b \wedge e^c
\end{aligned} \tag{2.8}$$

$$dT^a + \omega_b{}^a \wedge T^b = d\omega_b{}^a \wedge e^b + \omega_c{}^a \wedge \omega_b{}^c \wedge e^b$$

$$\boxed{DT^a = D^2 e^a = dT^a + \omega_b{}^a \wedge T^b = R_b{}^a \wedge e^b}$$

and

$$\begin{aligned}
dR_b{}^a &= d\left(d\omega_b{}^a - \omega_b{}^c \wedge \omega_c{}^a\right) \\
&= \omega_b{}^c \wedge d\omega_c{}^a - d\omega_b{}^c \wedge \omega_c{}^a \\
&= \left(\omega_b{}^c \wedge R_c{}^a + \omega_b{}^c \wedge \omega_c{}^d \wedge \omega_d{}^a\right) - \left(R_b{}^c \wedge \omega_c{}^a + \omega_b{}^d \wedge \omega_d{}^c \wedge \omega_c{}^a\right)
\end{aligned} \tag{2.9}$$

$$dR_b{}^a = \omega_b{}^c \wedge R_c{}^a - R_b{}^c \wedge \omega_c{}^a$$

$$\boxed{dR_b{}^a + R_b{}^c \wedge \omega_c{}^a - \omega_b{}^c \wedge R_c{}^a = 0}$$

Define the covariant derivative of a p-form $V_b{}^a$ as

$$DV_b{}^a = dV_b{}^a + \omega_c{}^a \wedge V_b{}^c - (-1)^p V_c{}^a \wedge \omega_b{}^c \tag{2.10}$$

which allows us to write our Bianchi identity (2.9) as

$$DR_b{}^a = 0. \tag{2.11}$$

In other words, the Bianchi identity says that the covariant derivative of the curvature 2-form vanishes.

Consider on orthogonal rotation of the orthonormal frame (or gauge transformation)

$$e^a \to e'^a = \Phi_b{}^a e^b,$$

where

$$\eta_{ab}\Phi_c{}^a\Phi_d{}^b = \eta_{cd}.$$

We can use the fact that $(d\Phi)_b{}^a(\Phi^{-1})_c{}^b = -\Phi_b{}^a(d\Phi^{-1})_c{}^b$ to find the transformation law for the torsion

$$\begin{aligned}
T'^a = \Phi_b{}^a T^b &= \Phi_b{}^a(de^b) + \Phi_b{}^a(\omega_c{}^b \wedge e^c),\\
&= \Phi_b{}^a d\left((\Phi^{-1})_c{}^b e'^c\right) + \Phi_b{}^a\left(\omega_c{}^b \wedge (\Phi^{-1})_d^c e'^d\right),\\
&= \Phi_b{}^a\left((d\Phi^{-1})_c{}^b e'^c + (\Phi^{-1})_c{}^b de'^c\right) + \Phi_b{}^a\omega_c{}^b \wedge (\Phi^{-1})_d{}^c e'^d,\\
&= \Phi_b{}^a(d\Phi^{-1})_c{}^b e'^c + \Phi_b{}^a(\Phi^{-1})_c{}^b de'^c + \Phi_b{}^a\omega_c{}^b \wedge (\Phi^{-1})_d{}^c e'^d, \qquad (2.12)\\
&= \Phi_b{}^a(d\Phi^{-1})_c{}^b e'^c + \delta_c^a de'^c + \Phi_b{}^a\omega_c{}^b(\Phi^{-1})_d{}^c \wedge e'^d,\\
&= de'^a + \left(\Phi_b{}^a\omega_c{}^b(\Phi^{-1})_d{}^c + \Phi_b{}^a(d\Phi^{-1})_d{}^b\right) \wedge e'^d,\\
T'^a &= de'^a + \omega'_d{}^a \wedge e'^d,
\end{aligned}$$

where we have written

$$\omega'_d{}^a = \Phi_b{}^a\omega_c{}^b(\Phi^{-1})_d{}^c + \Phi_b{}^a(d\Phi^{-1})_d{}^b \qquad (2.13)$$

as our new connection. Notice that, given two different connections, $\omega_d{}^a$ and $\widetilde{\omega}_d{}^a$, their difference transforms as

$$\begin{aligned} \omega_d'^{\,a} - \widetilde{\omega}_d'^{\,a} &= \Phi_b{}^a \omega_c{}^b (\Phi^{-1})_d{}^c - \Phi_b{}^a \widetilde{\omega}_c{}^b (\Phi^{-1})_d{}^c + \Phi_b{}^a (d\Phi^{-1})_d{}^b - \Phi_b{}^a (d\Phi^{-1})_d{}^b \\ &= \Phi_b{}^a \left(\omega_c{}^b - \widetilde{\omega}_c{}^b\right) (\Phi^{-1})_d{}^c \end{aligned} \tag{2.14}$$

which is the transformation law for a tensor in the case that the transformation is simply a coordinate transformation. The difference between two connections is, therefore, a tensor. This fact will be useful when we derive Einstein's equations in Section 2.2. The transformation law for the curvature 2-form is

$$R_b'^{\,a} = d\omega_b'^{\,a} - \omega_b'^{\,c} \wedge \omega_c'^{\,a} = \Phi_c{}^a R_d{}^c (\Phi^{-1})_b{}^d. \tag{2.15}$$

The covariant derivative transforms covariantly

$$(DV)_b'^{\,a} = \Phi_c{}^a (DV)_d{}^c (\Phi^{-1})_b{}^d. \tag{2.16}$$

2.1.1. The Yang-Mills equations

It is instructive at this point to show the utility of differential geometric formalism in gauge theories to prepare us for the next chapter where we will formulate gravity as a gauge theory. (See Section E.5. for more a detailed treatment.)

Maxwell's theory of electromagnetism is described by the $U(1)$ gauge group. $U(1)$ is one dimensional and abelian. (The structure constants are zero.) We can use the language of fibre bundles (Appendix E) to view the gauge group as a principal bundle over M.

Suppose the base space M is a four-dimensional Minkowski spacetime. The $U(1)$ bundle is then trivial, $P = \mathbb{R}^4 \times U(1)$. A single local trivialization over M is all that is required to cover the entire manifold with a coordinate chart. The gauge potential is a 1-form connection on this principal bundle

$$A = A_\mu dx^\mu, \tag{2.17}$$

where A_μ is the usual four-vector potential. Exponentiation of A yields a map to the Lie group $U(1)$. The field strength, or curvature, in this abelian case, is given by

$$F = dA, \tag{2.18}$$

where d is the exterior derivative. Equation (2.18) is of the same form as equation (2.7) where $A \wedge A = 0$ in this abelian case. In components, we have

$$\begin{aligned} \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu = d(A_\mu dx^\mu) &= \frac{1}{2} \left(\partial_\mu A_\nu dx^\mu \wedge dx^\nu + \partial_\nu A_\mu dx^\nu \wedge dx^\mu \right) \\ &= \frac{1}{2} \left(\partial_\mu A_\nu - \partial_\nu A_\mu \right) dx^\nu \wedge dx^\mu. \end{aligned} \tag{2.19}$$

F satisfies the Bianchi identity,

$$dF = F \wedge A - A \wedge F = 0, \tag{2.20}$$

which is merely geometrical since F is exact, $F = dA$ and $d^2 = 0$. In components,

$$\partial_\lambda F_{\mu\nu} + \partial_\nu F_{\lambda\mu} + \partial_\mu F_{\nu\lambda} = 0. \tag{2.21}$$

If we identify components $F_{\mu\nu}$ with the electric and magnetic fields as

$$E_i = F_{i0}, \quad B_i = \frac{1}{2}\epsilon_{ijk}F_{jk}, \tag{2.22}$$

the Bianchi identity reduces to two of Maxwell's equations

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \quad \text{and} \quad \nabla \cdot \mathbf{B} = 0. \tag{2.23}$$

These two equations are geometrical rather than dynamical since they arise from the properties of the bundle, not the equations of motion. The free photon action is

$$S = \int_M L_{\mathrm{free}} d^4x = \frac{1}{4}\int_{\mathbf{R}^4} F_{\mu\nu}F^{\mu\nu} d^4x = -\frac{1}{4}\int_{\mathbf{R}^4} F_{\mu\nu}F^{\mu\nu} d^4x, \tag{2.24}$$

and, if we denote the Hodge dual of F by $*F = \frac{1}{2}F^{\alpha\beta}\epsilon_{\alpha\beta\mu\nu}$, we can write the action as

$$S = -\frac{1}{4}\int_{\mathbf{R}^4} F \wedge *F = -\frac{1}{4}\int_{\mathbf{R}^4} F_{\mu\nu}F^{\mu\nu} d^4x. \tag{2.25}$$

We have

$$-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} = \frac{1}{2}\left(\mathbf{E}^2 - \mathbf{B}^2\right) \quad \text{and } F_{\mu\nu} * F^{\mu\nu} = \mathbf{B}\cdot\mathbf{E}. \tag{2.26}$$

Variation of the action with respect to A_μ gives

$$d * F = \partial_\mu F^{\mu\nu} = 0, \tag{2.27}$$

which is the second set of vacuum Maxwell's equations (this time coming from dynamics)

$$\nabla \cdot \mathbf{E} = 0 \quad \text{and} \quad \nabla \times \mathbf{B} - \frac{\partial \mathbf{E}}{\partial t} = 0. \tag{2.28}$$

We see that Maxwell's equations in vacuum follow from the Bianchi identity, $dF = 0$, and the Yang-Mills equations, $d * F = 0$.

The generalization of this process to non-abelian gauge groups is straightforward. The gauge potential then takes values in a non-abelian Lie algebra. In the non-abelian case, the field strength generalizes, as in (2.7), to

$$F = dA + A \wedge A \tag{2.29}$$

and is non-linear in the connection. This non-linearity is typical of non-abelian theories and gives rise to self interactions of the gauge field as can be seen by computing the free lagrangian

$$L_{\mathsf{free}} = -\frac{1}{2} \int_M \mathrm{Tr}\, (F \wedge *F)\,, \tag{2.30}$$

and noticing the interaction terms. The equations of motion are still given by the Bianchi identity and the Yang-Mills equations.

2.1.2. Tensor formulation

There is a tensor formulation of differential geometry which is equivalent to the differential form version that we have used. Define the covariant derivative by

$$\nabla_X f = X^\mu \partial_\mu f \tag{2.31}$$

so that it becomes a directional derivative when acting on scalar functions, f. The covariant derivative acts on basis vectors, $e_\mu = \partial_\mu$, and basis one-forms, $e^\mu = dx^\mu$, as

$$\begin{aligned} \nabla_X e_\nu &= X^\mu \nabla_\mu e_\nu = X^\mu e_\lambda \Gamma_{\nu\mu}{}^\lambda \\ \nabla_X e^\nu &= X^\mu \nabla_\mu e^\nu = -X^\mu e^\lambda \Gamma_{\mu\lambda}{}^\nu, \end{aligned} \tag{2.32}$$

respectively. We can use this process to calculate the covariant derivative of an arbitrary tensor. For example,

$$\begin{aligned} \nabla_X A^\alpha_\beta \, e_\alpha \otimes e^\beta &= X^\mu \nabla_\mu \left(A^\alpha_\beta e_\alpha \otimes e^\beta \right) \\ &= X^\mu \left[\left(\nabla_\mu A^\alpha_\beta \right) e_\alpha \otimes e^\beta + A^\alpha_\beta \left(\nabla_\mu e_\alpha \right) \otimes e^\beta + A^\alpha_\beta e_\alpha \otimes \left(\nabla_\mu e^\beta \right) \right] \\ &= X^\mu \left[\left(\partial_\mu A^\alpha_\beta \right) + A^\rho_\beta \Gamma_{\rho\mu}{}^\alpha - A^\alpha_\rho \Gamma_{\mu\beta}{}^\rho \right] e_\alpha \otimes e^\beta . \end{aligned} \tag{2.33}$$

We will henceforth use the standard notation wherein we denote a tensor by its components and write the covariant derivative as

$$\nabla_\mu A^\alpha_\beta = \partial_\mu A^\alpha_\beta + \Gamma_{\mu\rho}{}^\alpha A^\rho_\beta - \Gamma_{\mu\beta}{}^\rho A^\alpha_\rho . \tag{2.34}$$

The various tensors with flat indices are related to the ones with curved indices by factors of the vierbeins, $e^a_{\ \mu}$ and $e_a^{\ \mu}$. We can rewrite the curvature 2-form and the torsion 2-form in curved indices as

$$R^a_{\ b} = \frac{1}{2} R^a_{\ bcd} e^c \wedge e^d = \frac{1}{2} R^a_{\ b\mu\nu} dx^\mu \wedge dx^\nu \tag{2.35}$$

and

$$T^a = \frac{1}{2} T_{bc}^{\ \ a} e^b \wedge e^c = \frac{1}{2} T_{\mu\nu}^{\ \ a} dx^\mu \wedge dx^\nu. \tag{2.36}$$

The Riemann tensor and the torsion tensor are then

$$R^\alpha_{\ \beta\mu\nu} = e_a^{\ \alpha} e^b_{\ \beta} R^a_{\ b\mu\nu} \tag{2.37}$$

and

$$T_{\mu\nu}^{\ \ \alpha} = e_a^{\ \alpha} T_{\mu\nu}^{\ \ a}. \tag{2.38}$$

We define the non-metricity

$$Q_{\mu\nu\alpha} = \nabla_\alpha g_{\mu\nu} \equiv g_{\mu\nu;\alpha}. \tag{2.39}$$

Requiring that the metric be covariantly constant and that there be no torsion yields the conditions

$$g_{\mu\nu;\alpha} = Q_{\alpha\mu\nu} = \partial_\alpha g_{\mu\nu} - \Gamma_{\alpha\mu}^{\ \ \lambda} g_{\lambda\nu} - \Gamma_{\alpha\nu}^{\ \ \lambda} g_{\mu\lambda} = 0 \tag{2.40}$$

and

$$T_{\alpha\beta}{}^{\mu} = \left(\Gamma_{\alpha\beta}{}^{\mu} - \Gamma_{\beta\alpha}{}^{\mu}\right) = 0. \tag{2.41}$$

We can solve these equations for the Levi-Civita connection (Christoffel symbol)

$$\Gamma_{\alpha\beta}{}^{\mu} = \frac{1}{2} g^{\mu\nu} \left(\partial_\alpha g_{\beta\nu} + \partial_\beta g_{\nu\alpha} - \partial_\nu g_{\alpha\beta}\right). \tag{2.42}$$

The analogous expressions in terms of the spin connection, ω_{ab}, are given by

$$\text{metricity: } \omega_{ab} = -\omega_{ba} \tag{2.43}$$

and

$$\text{no torsion: } T^a = de^a + \omega_b{}^a \wedge e^b = 0. \tag{2.44}$$

The covariant derivative in holonomic coordinates defined by equation (2.34) leads to an expression for the non-holonomic connection, $\omega_b{}^a$, by looking at the covariant derivative in a non-holonomic basis. Since the covariant derivative is a tensor, we have

$$\begin{aligned}
A^a_{;b} &= e_\mu{}^a e^\nu{}_b A^\mu_{;\nu} \\
A^a_{,b} + \omega_{bc}{}^a A^c &= e_\mu{}^a e^\nu{}_b \left(A^\mu_{,\nu} + \Gamma_{\nu\rho}{}^\mu A^\rho\right) \\
e^\alpha{}_b \partial_\alpha \left(e_\beta{}^a A^\beta\right) + \omega_{bc}{}^a A^c &= e_\mu{}^a e^\nu{}_b A^\mu_{,\nu} + e_\mu{}^a e^\nu{}_b \Gamma_{\nu\rho}{}^\mu A^\rho \\
e^\alpha{}_b \left(\partial_\alpha e_\rho{}^a\right) A^\rho + e_\rho{}^c \omega_{bc}{}^a A^\rho &= e_\mu{}^a e^\nu{}_b \Gamma_{\nu\rho}{}^\mu A^\rho \\
\omega_{bc}{}^a = -e^\rho{}_c e^\alpha{}_b e_{\rho,\alpha}^{\ a} + e^\rho{}_c e_\mu{}^a e^\nu{}_b \Gamma_{\nu\rho}{}^\mu &= -e^\rho{}_c e^\alpha{}_b e_{\rho;\alpha}^{\ a}.
\end{aligned} \tag{2.45}$$

We can, therefore, write the torsion as

$$\begin{aligned} T^a &= \frac{1}{2} T_{\mu\nu}{}^a dx^\mu \wedge dx^\nu = de^a + \omega_b{}^a \wedge e^b \\ &= \frac{1}{2}\left[\partial_\mu e_\nu{}^a - \partial_\nu e_\mu{}^a + \omega_{\mu b}{}^a e_\nu{}^b - \omega_{\nu b}{}^a e_\mu{}^b\right] dx^\mu \wedge dx^\nu . \end{aligned} \tag{2.46}$$

Thus,

$$\begin{aligned} T_{\mu\nu}{}^a &= e_{\nu,\mu}^{\ a} - e_{\mu,\nu}^{\ a} + \left(-e^\alpha{}_b e_{\alpha;\mu}^{\ a}\right) e_\nu{}^b - \left(-e^\alpha{}_b e_{\alpha;\nu}^{\ a}\right) e_\mu{}^b \\ &= e_{\nu,\mu}^{\ a} - e_{\mu,\nu}^{\ a} - e^\alpha{}_b \left(e_{\alpha,\mu}^{\ a} - \Gamma_{\mu\alpha}{}^\lambda e_\lambda{}^a\right) e_\nu{}^b + e^\alpha{}_b \left(e_{\alpha,\nu}^{\ a} - \Gamma_{\nu\alpha}{}^\lambda e_\lambda{}^a\right) e_\mu{}^b \\ &= e_{\nu,\mu}^{\ a} - e_{\mu,\nu}^{\ a} - g_\nu^\alpha \left(e_{\alpha,\mu}^{\ a} - \Gamma_{\mu\alpha}{}^\lambda e_\lambda{}^a\right) + g_\mu^\alpha \left(e_{\alpha,\nu}^{\ a} - \Gamma_{\nu\alpha}{}^\lambda e_\lambda{}^a\right) \\ &= e_{\nu,\mu}^{\ a} - e_{\mu,\nu}^{\ a} - e_{\nu,\mu}^{\ a} + e_{\mu,\nu}^{\ a} + \Gamma_{\mu\nu}{}^\lambda e_\lambda{}^a - \Gamma_{\nu\mu}{}^\lambda e_\lambda{}^a \\ &= \left(\Gamma_{\mu\nu}{}^\lambda - \Gamma_{\nu\mu}{}^\lambda\right) e_\lambda{}^a . \end{aligned} \tag{2.47}$$

If the torsion is zero, we have

$$\Gamma_{\mu\nu}{}^\lambda = \Gamma_{\nu\mu}{}^\lambda \tag{2.48}$$

so that the symmetry of the Christoffel connection comes from the torsion-free condition.

Cartan's structure equation for the curvature (2.7) gives

$$\begin{aligned} R_b{}^a = \frac{1}{2} R_{cdb}{}^a e^c \wedge e^d &= d\omega_b{}^a - \omega_b{}^c \wedge \omega_c{}^a \\ \frac{1}{2} R_{cdb}{}^a e_\mu{}^c dx^\mu \wedge e_\nu{}^d dx^\nu &= \frac{1}{2}\left[\partial_\mu \omega_{\nu b}{}^a - \partial_\nu \omega_{\mu b}{}^a\right] dx^\mu \wedge dx^\nu - \omega_b{}^c \wedge \omega_c{}^a \\ \frac{1}{2} R_{\mu\nu b}{}^a dx^\mu \wedge dx^\nu &= \frac{1}{2}\left[\partial_\mu \omega_{\nu b}{}^a - \partial_\nu \omega_{\mu b}{}^a - \omega_{\mu b}{}^c \omega_{\nu c}{}^a + \omega_{\nu b}{}^c \omega_{\mu c}{}^a\right] dx^\mu \wedge dx^\nu \end{aligned} \tag{2.49}$$

so that

$$R_{\mu\nu b}{}^{a} = \partial_\mu \omega_{\nu b}{}^{a} - \partial_\nu \omega_{\mu b}{}^{a} - \omega_{\mu b}{}^{c} \omega_{\nu c}{}^{a} + \omega_{\nu b}{}^{c} \omega_{\mu c}{}^{a}. \tag{2.50}$$

Using equation (2.45), (2.50) becomes

$$R_{\mu\nu\beta}{}^{\alpha} = \Gamma_{\nu\beta,\mu}^{\alpha} - \Gamma_{\mu\beta,\nu}^{\alpha} + \Gamma_{\mu\rho}{}^{\alpha}\Gamma_{\nu\beta}{}^{\rho} - \Gamma_{\nu\rho}{}^{\alpha}\Gamma_{\mu\beta}{}^{\rho}. \tag{2.51}$$

The Ricci tensor and the scalar curvature are defined by $R_{\mu\nu} = R_{\mu\alpha\beta}{}^{\mu}$ and $R = R_{\mu\nu}g^{\mu\nu}$, respectively.

Example: The 2-sphere

It is instructive to see how the tetrads can be used to find the spin connection 1-form, the curvature 2-form, and the Gaussian curvature (the scalar curvature) in a physical example. Consider the following metric on S^2:

$$ds^2 = r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2 = (e^1)^2 + (e^2)^2. \tag{2.52}$$

We choose

$$e^1 = r d\theta, \quad e^2 = r\sin\theta d\phi \tag{2.53}$$

and use Cartan's structure equation, $0 = de^a + {\omega_b}^a \wedge e^b$, to get

$$
\begin{aligned}
0 = de^1 + {\omega_2}^1 \wedge e^2 &= \frac{\partial {e_1}^1}{\partial\theta} d\theta \wedge d\theta + \frac{\partial {e_2}^1}{\partial\phi} d\phi \wedge d\theta + {\omega_2}^1 \wedge (r \sin\theta d\phi) \\
&= 0 + {\omega_2}^1 \wedge (r \sin\theta d\phi) \\
0 &= {\omega_2}^1 \wedge d\phi
\end{aligned}
\tag{2.54}
$$

$$
\begin{aligned}
0 = de^2 + {\omega_1}^2 \wedge e^1 &= \frac{\partial {e_1}^2}{\partial\theta} d\theta \wedge d\phi + {\omega_1}^2 \wedge (r d\theta) \\
-\frac{\partial(r\sin\theta)}{\partial\theta} d\theta \wedge d\phi &= r{\omega_1}^2 \wedge d\theta \\
-r\cos\theta d\theta \wedge d\phi &= r{\omega_1}^2 \wedge d\theta \\
\cos\theta d\phi \wedge d\theta &= {\omega_1}^2 \wedge d\theta \\
\cos\theta d\phi = {\omega_1}^2 \text{and} &\; -\cos\theta d\phi = {\omega_2}^1.
\end{aligned}
\tag{2.55}
$$

The curvature is given by the other structure equation: ${R_a}^b = d{\omega_a}^b - {\omega_a}^c \wedge {\omega_c}^b$.

$$
\begin{aligned}
{R_a}^b &= d{\omega_a}^b + {\omega_1}^b \wedge {\omega_a}^1 + {\omega_2}^b \wedge {\omega_a}^2 = d{\omega_a}^b + 0 \\
{R_2}^1 &= d{\omega_2}^1 = -{R_1}^2 = d{\omega_2}^1 = d(\cos\theta d\phi) = \sin\theta \, d\theta \wedge d\phi
\end{aligned}
\tag{2.56}
$$

so that

$$
\begin{aligned}
{R_{122}}^1 e^1 \wedge e^2 &= \sin\theta d\theta \wedge d\phi \\
{R_{122}}^1 r^2 \sin\theta d\theta \wedge d\phi &= \sin\theta d\theta \wedge d\phi \\
{R_{122}}^1 &= \frac{1}{r^2} = {R_{211}}^2.
\end{aligned}
\tag{2.57}
$$

The scalar curvature is then

$$\begin{aligned}\eta^{bd} R_{abd}{}^{a} &= R_{ab}{}^{ba} = R_{12}{}^{21} + R_{21}{}^{12} \\ &= \frac{1}{r^2} + \frac{1}{r^2} = \frac{2}{r^2},\end{aligned} \tag{2.58}$$

which is constant on S^2 and positive definite.

2.2. General relativity

The Einstein-Hilbert action with cosmological constant and matter lagrangian is given by

$$S_{EH} = \frac{1}{16\pi G} \int d^4x \left(\sqrt{-g}R - \sqrt{-g}\Lambda + 16\pi G\, \mathcal{L}\right), \tag{2.59}$$

where G is the gravitational constant, R is the scalar curvature, Λ is a cosmological constant, and $\mathcal{L}$ is a possible matter lagrangian.[2] In general relativity, it is assumed that the non-metricity and the torsion are both zero.

We vary the action with respect to the metric as follows:

$$\begin{aligned}\delta\mathcal{L}_{EH} &= \left(\delta\sqrt{-g}\right)[R - \Lambda] + \sqrt{-g}\left(\delta R - \delta\Lambda\right) + 16\pi G\, \delta\mathcal{L} \\ &= \left(\delta\sqrt{-g}\right)[R - \Lambda] + \sqrt{-g}\left(\delta g^{\mu\nu} R_{\mu\nu} + g^{\mu\nu}\delta R_{\mu\nu}\right) + 16\pi G\, \delta\mathcal{L}.\end{aligned} \tag{2.60}$$

We now need the results of the following derivations:

$$\delta\left(\ln\det g_{\mu\nu}\right) = \delta\left(\mathrm{Tr}\ln g_{\mu\nu}\right) \Rightarrow \frac{1}{g}\delta g = \mathrm{Tr}\left(\delta\ln g_{\mu\nu}\right) = g^{\nu\mu}\delta g_{\mu\nu} \Rightarrow \delta g = g g^{\nu\mu}\delta g_{\mu\nu} \tag{2.61}$$

[2]The matter lagrangian in curved space will contain factors of the metric and the overall $\sqrt{-g}$ factor.

and

$$0 = \delta\left(g_{\mu\nu} g^{\nu\beta}\right) = \left(\delta g_{\mu\nu}\right) g^{\nu\beta} + g_{\mu\nu}\left(\delta g^{\nu\beta}\right) \Rightarrow -g^{\mu\alpha} g^{\beta\nu} \delta g_{\mu\nu} = \delta g^{\alpha\beta}, \tag{2.62}$$

The variation of the Ricci tensor is a bit more tricky.

$$\delta R_{\nu\beta} = \delta \Gamma_{\nu\beta,\alpha}^{\ \ \alpha} - \delta \Gamma_{\alpha\beta,\nu}^{\ \ \alpha} + \delta \Gamma_{\alpha\rho}{}^{\alpha} \Gamma_{\nu\beta}{}^{\rho} + \Gamma_{\alpha\rho}{}^{\alpha} \delta \Gamma_{\nu\beta}{}^{\rho} - \delta \Gamma_{\nu\rho}{}^{\alpha} \Gamma_{\alpha\beta}{}^{\rho} - \Gamma_{\nu\rho}{}^{\alpha} \delta \Gamma_{\alpha\beta}{}^{\rho}. \tag{2.63}$$

We choose a coordinate system where $\Gamma = 0$ so that

$$\delta R_{\nu\beta} = \delta\left(\nabla_{\alpha} \Gamma_{\nu\beta}{}^{\alpha} - \nabla_{\nu} \Gamma_{\alpha\beta}{}^{\alpha}\right). \tag{2.64}$$

The covariant derivative of a connection is another connection. We have shown in equation (2.14) that the difference of two connections is a tensor, so the expression for $\delta R_{\nu\beta}$ is valid in any coordinate system.

We have

$$\delta \mathcal{L}_{EH} = \left(\frac{1}{2}\sqrt{-g} g^{\nu\mu} \delta g_{\mu\nu}\right)[R - \Lambda] + \sqrt{-g}\left(-g^{\mu\alpha} g^{\beta\nu} \delta g_{\alpha\beta} R_{\mu\nu}\right) + 16\pi G\, \delta \mathcal{L}. \tag{2.65}$$

We have dropped the $\delta R_{\mu\nu}$ term since, in Einstein gravity, the metric is covariantly constant and, thus, this term is a total divergence and drops out of the variation. We have

$$\delta \mathcal{L}_{EH} = \sqrt{-g}\left(\frac{1}{2} g^{\alpha\beta}[R - \Lambda] - g^{\mu\alpha} g^{\beta\nu} R_{\mu\nu} + \frac{16\pi G}{\sqrt{-g}} \frac{\partial \mathcal{L}}{\partial g_{\alpha\beta}}\right) \delta g_{\alpha\beta} \tag{2.66}$$

and the Einstein-Hilbert action becomes

$$S_{EH} = \frac{1}{16\pi G} \int d^4x \sqrt{-g} \left(\frac{1}{2} g^{\alpha\beta} R - R^{\alpha\beta} - \frac{1}{2} g^{\alpha\beta} \Lambda + \frac{16\pi G}{\sqrt{-g}} \frac{\partial \mathcal{L}}{\partial g_{\alpha\beta}} \right) \delta g_{\alpha\beta}. \tag{2.67}$$

Under an arbitrary variation $\delta g_{\alpha\beta}$ we have

$$R^{\alpha\beta} - \frac{1}{2} g^{\alpha\beta} R = \frac{1}{2} g^{\alpha\beta} \Lambda + 8\pi G \left[-\frac{2}{\sqrt{-g}} \frac{\partial \mathcal{L}}{\partial g_{\alpha\beta}} \right]. \tag{2.68}$$

Defining

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g^{\mu\nu} R \tag{2.69}$$

along with

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}} \frac{\partial \mathcal{L}}{\partial g^{\mu\nu}}, \tag{2.70}$$

we arrive at

$$G_{\mu\nu} = \frac{1}{2} g_{\mu\nu} \Lambda + 8\pi G T_{\mu\nu}, \tag{2.71}$$

which is Einstein's equation with a source term and cosmological constant.

2.3. Torsion

In this section, we want to describe the well-known extension of Einstein's theory of gravity which includes non-zero spacetime torsion [4, 10, 11, 12]. This extension is called Einstein-Cartan gravity.

Elementary particles are classified by irreducible unitary representations of the Poincaré group. They can be labeled by mass and spin. Mass arises from the translational part of the Poincaré group and spin from the rotational part. Distributing mass-energy and spin over spacetime leads to the energy-momentum tensor as well as the spin angular momentum tensor of matter. Energy-momentum adds up in the classical regime due to its monopole character, whereas spin usually averages out due to its multipole character. The dynamical characterization of a continuous distribution of macroscopic matter can usually be achieved by energy-momentum alone. In Einstein's theory of gravity, we see from equation (2.71) that energy-momentum is the source of the gravitational field since the energy-momentum tensor is coupled to the metric tensor of spacetime.

When we descend to the microscopic regime, spin angular momentum is needed to characterize matter dynamically. In other words, spin angular momentum may also be a source of a "gravitational" field which is directly coupled to the geometry of spacetime. This field is called the spacetime torsion. Spin angular momentum coupling to torsion is the rotational analogue to energy-momentum coupling to the metric.

To describe spacetime with non-zero torsion, we need to generalize the 4-dimensional spacetime of general relativity to the 4-dimensional spacetime known as Riemann-Cartan spacetime, sometimes called U_4 or Einstein-Cartan-Sciama-Kibble spacetime [4]. Let us go back to equation (2.47) and define

$$T_{\alpha\beta}{}^{\mu} = \widetilde{\Gamma}_{\alpha\beta}{}^{\mu} - \widetilde{\Gamma}_{\beta\alpha}{}^{\mu} \equiv 2S_{\alpha\beta}{}^{\mu}, \tag{2.72}$$

where we are now following the notation of [12] in writing

$$S_{\alpha\beta}{}^{\mu} = \widetilde{\Gamma}_{[\alpha\beta]}{}^{\mu}. \tag{2.73}$$

The tilde above quantities indicates that they are defined in a space with non-zero torsion so that writing $\Gamma_{\beta\alpha}{}^{\mu}$ without the tilde indicates that we refer to the torsion-free Christoffel connection.

In terms of the *contortion tensor*,[3] where $K_{\mu\nu}{}^{\alpha}$, and the *non-metricity*, $Q_{\mu\nu}{}^{\alpha}$, the affine connection can be written

$$\widetilde{\Gamma}_{\mu\nu}{}^{\alpha} = \Gamma_{\mu\nu}{}^{\alpha} + K_{\mu\nu}{}^{\alpha} + \widetilde{Q}_{\mu\nu}{}^{\alpha}, \tag{2.74}$$

$$K_{\mu\nu}{}^{\alpha} = \frac{1}{2}\left(T_{\mu\nu}{}^{\alpha} + T^{\alpha}{}_{\mu\nu} + T^{\alpha}{}_{\nu\mu}\right) = \left(S_{\mu\nu}{}^{\alpha} + S^{\alpha}{}_{\mu\nu} + S^{\alpha}{}_{\nu\mu}\right) \tag{2.75}$$

and

$$\widetilde{Q}_{\mu\nu}{}^{\alpha} = \frac{1}{2}\left(Q_{\mu\nu}{}^{\alpha} + Q_{\nu\mu}{}^{\alpha} - Q^{\alpha}{}_{\mu\nu}\right) \equiv \left(N_{\mu\nu}{}^{\alpha} + N_{\nu\mu}{}^{\alpha} - N^{\alpha}{}_{\mu\nu}\right). \tag{2.76}$$

The metricity condition (2.40) results in $Q_{\mu\nu}{}^{\alpha} = 0$. Since the torsion is antisymmetric in the first two indices, we see that the contortion tensor is antisymmetric in the last two indices. When the contortion tensor and the non-metricity are zero, equation (2.74) reduces to the usual Christoffel connection given by expression (2.42).

We can now find the Riemann curvature tensor from (2.51) with the general affine

[3]Note that our convention differs by a minus sign from that of many authors.

connection replacing the Christoffel connection

$$\widetilde{R}_{\mu\nu\beta}{}^{\alpha} = \widetilde{\Gamma}_{\nu\beta,\mu}^{\ \alpha} - \widetilde{\Gamma}_{\mu\beta,\nu}^{\ \alpha} + \widetilde{\Gamma}_{\mu\rho}{}^{\alpha}\widetilde{\Gamma}_{\nu\beta}{}^{\rho} - \widetilde{\Gamma}_{\nu\rho}{}^{\alpha}\widetilde{\Gamma}_{\mu\beta}{}^{\rho}. \tag{2.77}$$

Using equation (2.74) in the above expression and

$$K_{\mu\nu|\sigma}^{\ \alpha} = K_{\mu\nu,\sigma}^{\ \alpha} + \Gamma_{\sigma\rho}{}^{\alpha}K_{\mu\nu}{}^{\rho} - \Gamma_{\sigma\mu}{}^{\rho}K_{\rho\nu}{}^{\alpha} - \Gamma_{\sigma\nu}{}^{\rho}K_{\mu\rho}{}^{\alpha}, \tag{2.78}$$

where we have denoted with a vertical bar, $_|$, the covariant derivative with respect to the Christoffel connection, we have

$$\begin{aligned}\widetilde{R}_{\mu\nu\sigma}{}^{\alpha} = R_{\mu\nu\sigma}{}^{\alpha} &+ K_{\mu\nu|\sigma}^{\ \alpha} - K_{\sigma\nu|\mu}^{\ \alpha} + K_{\sigma\rho}{}^{\alpha}K_{\mu\nu}{}^{\rho} - K_{\mu\rho}{}^{\alpha}K_{\sigma\nu}{}^{\rho}\\ &+ \widetilde{Q}_{\mu\nu|\sigma}^{\ \alpha} - \widetilde{Q}_{\sigma\nu|\mu}^{\ \alpha} + \widetilde{Q}_{\sigma\rho}{}^{\alpha}\widetilde{Q}_{\mu\nu}{}^{\rho} - \widetilde{Q}_{\mu\rho}{}^{\alpha}\widetilde{Q}_{\sigma\nu}{}^{\rho}\\ &+ K_{\sigma\rho}{}^{\alpha}\widetilde{Q}_{\mu\nu}{}^{\rho} - K_{\mu\rho}{}^{\alpha}\widetilde{Q}_{\sigma\nu}{}^{\rho} + \widetilde{Q}_{\sigma\rho}{}^{\alpha}K_{\mu\nu}{}^{\rho} - \widetilde{Q}_{\mu\rho}{}^{\alpha}K_{\sigma\nu}{}^{\rho}.\end{aligned} \tag{2.79}$$

Let us keep in mind the symmetry $\widetilde{Q}_{\mu\nu}{}^{\rho} = \widetilde{Q}_{\nu\mu}{}^{\rho}$ following from the property $Q_{\mu\nu}{}^{\rho} = Q_{\mu}{}^{\rho}{}_{\nu}$ of the non-metricity, and the symmetry $K_{\mu\nu}{}^{\rho} = -K_{\mu}{}^{\rho}{}_{\nu}$ following from the property $S_{\mu\nu}{}^{\rho} = -S_{\nu\mu}{}^{\rho}$ of the torsion. These symmetries are useful in the calculation of the contracted curvature

$$\begin{aligned}\widetilde{R}_{\mu\nu} \equiv \widetilde{R}_{\alpha\mu\nu}{}^{\alpha} = R_{\mu\nu} &+ K_{\mu\nu|\alpha}^{\ \alpha} - K_{\alpha\nu|\mu}^{\ \alpha} + K_{\alpha\rho}{}^{\alpha}K_{\mu\nu}{}^{\rho} - K_{\mu\rho}{}^{\alpha}K_{\alpha\nu}{}^{\rho}\\ &+ \widetilde{Q}_{\mu\nu|\alpha}^{\ \alpha} - \widetilde{Q}_{\alpha\nu|\mu}^{\ \alpha} + \widetilde{Q}_{\alpha\rho}{}^{\alpha}\widetilde{Q}_{\mu\nu}{}^{\rho} - \widetilde{Q}_{\mu\rho}{}^{\alpha}\widetilde{Q}_{\alpha\nu}{}^{\rho}\\ &+ K_{\alpha\rho}{}^{\alpha}\widetilde{Q}_{\mu\nu}{}^{\rho} - K_{\mu\rho}{}^{\alpha}\widetilde{Q}_{\alpha\nu}{}^{\rho} + \widetilde{Q}_{\alpha\rho}{}^{\alpha}K_{\mu\nu}{}^{\rho} - \widetilde{Q}_{\mu\rho}{}^{\alpha}K_{\alpha\nu}{}^{\rho}\end{aligned} \tag{2.80}$$

which allows us to form the scalar curvature

$$\begin{aligned}\widetilde{R} \equiv \widetilde{R}^{\alpha}{}_{\alpha} = R + 2K^{\alpha}{}_{|\alpha} - K_{\rho}K^{\rho} - K_{\mu\rho\alpha}K^{\alpha\mu\rho} \\ + \widetilde{Q}_{\mu}{}^{\mu\alpha}{}_{|\alpha} - \widetilde{Q}_{\alpha}{}^{\mu\alpha}{}_{|\mu} + \widetilde{Q}_{\alpha\rho}{}^{\alpha}\widetilde{Q}_{\mu}{}^{\mu\rho} - \widetilde{Q}_{\mu\rho\alpha}\widetilde{Q}^{\alpha\rho\mu} \\ - K_{\rho}\widetilde{Q}^{\mu}{}_{\mu}{}^{\rho} - K_{\mu\rho\alpha}\widetilde{Q}^{\alpha\mu\rho} + \widetilde{Q}_{\alpha\rho}{}^{\alpha}K^{\rho} - \widetilde{Q}_{\mu\rho\alpha}K^{\alpha\mu\rho}.\end{aligned} \tag{2.81}$$

We have defined the torsion vector, $K_{\nu} = K_{\alpha}{}^{\alpha}{}_{\nu} = 2S_{\nu\alpha}{}^{\alpha} = 2S_{\nu}$. The last term in equation (2.81) is zero by symmetry. Using

$$\begin{aligned}\widetilde{Q}_{\mu}{}^{\mu\rho} = Q_{\mu}{}^{\mu\alpha} - \frac{1}{2}Q^{\alpha}{}_{\mu}{}^{\mu} \equiv \overrightarrow{Q}^{\alpha} - \frac{1}{2}\overleftarrow{Q}^{\alpha} = 2\overrightarrow{N}^{\alpha} - \overleftarrow{N}^{\alpha} \\ \widetilde{Q}^{\mu}{}_{\alpha}{}^{\alpha} = \frac{1}{2}Q^{\mu}{}_{\alpha}{}^{\alpha} \equiv \frac{1}{2}\overleftarrow{Q}^{\mu} = \overleftarrow{N}^{\mu},\end{aligned} \tag{2.82}$$

we can write equation (2.81) as

$$\begin{aligned}\widetilde{R} = R + 2K^{\alpha}{}_{|\alpha} - K_{\rho}K^{\rho} + K_{\mu\rho\alpha}K^{\alpha\rho\mu} - \widetilde{Q}_{\mu\rho\alpha}\widetilde{Q}^{\alpha\rho\mu} \\ + \overleftarrow{Q}_{\rho}K^{\rho} - \overrightarrow{Q}_{\rho}K^{\rho} + \overrightarrow{Q}^{\alpha}{}_{|\alpha} - \overleftarrow{Q}^{\alpha}{}_{|\alpha} + \frac{1}{2}\overleftarrow{Q}_{\rho}\overrightarrow{Q}^{\rho} - \frac{1}{4}\overleftarrow{Q}_{\rho}\overleftarrow{Q}^{\rho}.\end{aligned} \tag{2.83}$$

Writing this equation in terms of the torsion and non-metricity gives

$$\begin{aligned}\widetilde{R} = R + 4S^{\alpha}{}_{|\alpha} - 4S_{\rho}S^{\rho} + S_{\alpha\beta\gamma}S^{\alpha\beta\gamma} + 2S_{\alpha\beta\gamma}S^{\gamma\beta\alpha} + N_{\alpha\beta\gamma}N^{\alpha\beta\gamma} \\ - 2N_{\alpha\beta\gamma}N^{\gamma\beta\alpha} + 4\overleftarrow{N}_{\rho}S^{\rho} - 4\overrightarrow{N}_{\rho}S^{\rho} + 2\overrightarrow{N}^{\alpha}{}_{|\alpha} - 2\overleftarrow{N}^{\alpha}{}_{|\alpha} + 2\overleftarrow{N}_{\rho}\overrightarrow{N}^{\rho} - \overleftarrow{N}_{\rho}\overleftarrow{N}^{\rho}.\end{aligned} \tag{2.84}$$

2.3.1. Non-propagating torsion

The Einstein-Hilbert action can now be generalized to spacetimes with non-zero torsion and non-metricity:

$$S = \frac{1}{16\pi G}\int d^4x\sqrt{-g}\widetilde{R} = S_{EH} + S_T \tag{2.85}$$

which is the usual Einstein-Hilbert action, S_{EH}, plus a new piece depending on the torsion and non-metricity. Now arises the possibility of observing torsion and non-metricity as new particle fields.

If the non-metricity is zero, this action reduces to the Einstein-Cartan action

$$\widetilde{R} = R + 4S^{\alpha}{}_{|\alpha} - 4S_\rho S^\rho + S_{\alpha\beta\gamma}S^{\alpha\beta\gamma} + 2S_{\alpha\beta\gamma}S^{\gamma\beta\alpha}. \tag{2.86}$$

In the case that the torsion is completely antisymmetric, the torsion trace[4] will vanish, leaving

$$\widetilde{R} = R - H_{\alpha\beta\gamma}H^{\alpha\beta\gamma} \tag{2.87}$$

which we will find of interest later when comparing to similar actions arising in string theory.

The first thing that one notices about the action (2.85) with equation (2.84) is that there are no second derivative terms in the torsion or the non-metricity. Hence, these fields do not propagate in the theory as we have constructed it. This fact has caused

[4]It is interesting that the torsion trace has a linear derivative term in the kinetic part (2.86) of the action. This term may result in some topological instanton-like effects arising from this field.

many authors to propose modified theories which do allow propagating torsion. (See [4] and references therein.)

Even with non-propagation torsion, we can have torsion effects propagate through some other method. If the region of space had a non-zero spin density, then the effects of torsion would still propagate. For example, we could have the torsion coupling to photons via vacuum polarization [13].

2.3.2. Propagating torsion

There are many ways to extend Einstein-Cartan gravity to give a theory with propagating torsion. One way which we will mention at the end of the next chapter is to give up the notion of the metric being the fundamental field of gravity. We must then look beyond the scalar curvature as the lagrangian governing the gravitational force in order to arrive at a theory with propagating torsion.

For the time being, we take the much less drastic step proposed by several authors [4, 12] which is to assume that the torsion can be derived from a potential. The argument is that, since the metric is a potential and the torsion is a force, one should not regard the metric and the torsion as independent quantities and vary the action with respect to each of them independently. Instead, we need to put them on equal footing by finding a potential for the torsion. We, therefore, assume that the torsion vector can be written as

$$S_\alpha = \partial_\alpha \Theta, \tag{2.88}$$

where Θ is called the *torsion-dilaton* field [14]. Then, we can see by the form of equation

(2.86) that we have quadratic terms in our action leading to a propagating torsion-dilaton field. Notice that we could have just as easily kept the torsion as a fundamental field and discarded the metric as one. We could then consider the symmetric and the antisymmetric parts of the connection as the fundamental entities, the antisymmetric part being the torsion and the symmetric part being a new field. This approach is more closely related to the methods used in quantum field theory, so we will look at it again in the next chapter.

Another option is to treat the vierbein field as a physical field and assume that it can be written as [15]

$$e^a_\mu = \delta^a_\mu \phi. \tag{2.89}$$

Putting this *Ansatz* into the scalar curvature, one gets the lagrangian for a gravitational field coupled to a massless scalar field, ϕ. If we write $\phi' = \ln \phi$, this interaction term becomes a normal kinetic energy term for ϕ'. Another interesting facet of this model is seen when one rescales the metric so that $\hat{g}_{\mu\nu} = \phi^{-2} g_{\mu\nu}$. In that case, the lagrangian becomes exactly the one for dilaton gravity

$$\mathcal{L} = -\frac{1}{2\kappa}\phi^2\sqrt{-g}R. \tag{2.90}$$

We will look at a particular example from [12] in which the trace and symmetric parts of the torsion are zero and the totally antisymmetric part is derivable from a tensor potential

$$H_{\mu\nu\sigma} = B_{\mu\nu,\sigma} + B_{\sigma\mu,\nu} + B_{\nu\sigma,\mu}, \tag{2.91}$$

where $B_{\mu\nu}$ is the antisymmetric potential.[5]

Inserting the expression (2.91) into the action (2.85) we have

$$S = -\frac{1}{\kappa^2} \int d^4x \sqrt{-g} \left(R - H_{\alpha\beta\gamma} H^{\alpha\beta\gamma} \right) \tag{2.92}$$

We see that the kinetic energy term for the torsion field will now contain second-order derivatives of the torsion *potential* $B_{\sigma\mu}$ which means that we now have a theory where torsion propagates. The antisymmetric tensor field arising here, when the potential given in equation (2.91) is assumed, is identical to the so-called *Kalb-Ramond* antisymmetric tensor field found in the spectrum of string theory and M-theory (Chapter 4).

Notice that, in postulating a potential from which the torsion can be derived, we are saying that the antisymmetric part of the torsion is an exact 3-form, $H = dB$. This 3-form is invariant under $B \to B + d\Lambda$ for some function Λ, thus the action (2.85) is also invariant under such a gauge transformation.

At this point, we have propagating torsion without matter interactions. In the next chapter, we will derive an interaction between the torsion field and spin-$\frac{1}{2}$ particles. The interaction so derived will then be used in Chapter 5 to compute some relevant scattering amplitudes.

[5]Our notation differs from [12] in using the symbol $B_{\mu\nu}$ for the tensor potential rather than $\psi_{\mu\nu}$. We do this to more easily facilitate comparison to field theory in the next chapter.

CHAPTER 3

GAUGE TORSION GRAVITY

Our goal in this chapter is to derive a theory of propagating torsion interacting with matter as a gauge field [16].

We first show how invariances of the Dirac action under various compact gauge groups give rise to the usual interacting theories of particle physics. We next follow the lead of Sciama [2, 17] and Kibble [3] who, in the early 1960s, investigated gravity as resulting from gauge invariance of the Dirac action under local Poincaré transformations.

3.1. Gauge theories in particle physics

The techniques of variational calculus that we will use for the gravitational field with local Poincaré invariance are the same as those used in quantum field theory. To see the analogy to quantum field theory more easily, we will look at first at how the process works in quantum electrodynamics with local $U(1)$ invariance and in low-energy nuclear physics with local $SU(2)$ invariance. The same technique also works for the strong interactions with local $SU(3)$ invariance, but in the interest of brevity, we will simply refer the interested reader to modern field theory textbooks [18, 19] for more information.

3.1.1. Electrodynamic gauge theory

If transformations leaving a physical system invariant form a group, then the consequences of the symmetry can be deduced through a group theoretical analysis. For example, if a quantum mechanical system has no preferred direction in space, then the Hamiltonian which governs the system will be invariant under the rotation group, i.e.,

$$R(\theta)H(r)R^{-1}(\theta) = H(r) \tag{3.1}$$

In terms of the generators of the rotation, $R(\theta) = e^{i\theta \cdot J}$, equation (3.1) gives

$$[H, J_i] = 0. \tag{3.2}$$

The consequence of this symmetry is that

$$H\left(J_i \left|n\right\rangle\right) = E_n \left(J_i \left|n\right\rangle\right) \tag{3.3}$$

if $H\left|n\right\rangle = E_n \left|n\right\rangle$. Thus, all states that are connected by a rotation transformation are degenerate in energy. These states form the basis vectors for irreducible representations (j) of the group. Since the rotation group, $SO(3)$, has $(2j+1)$ dimensional irreducible representations, we have that the energy levels of the system have a $(2j+1)$-fold degeneracy. In other words, the hamiltonian is only sensitive to the equivalence classes of particles under rotations.

When we look at internal symmetries, the states are identified with various particles.

Such symmetry transformations change the particle labels, not the coordinate system, and irreducible representations of the group manifest themselves as degenerate particle multiplets.

Consider the electron, for example. The theory of the electron is governed by the Dirac lagrangian:

$$\mathcal{L} = \overline{\psi}\left(\frac{i}{2}\overleftrightarrow{\not\partial} - m\right)\psi.$$

The wave function, ψ, of the free electron is represented as a plane wave proportional to $\exp(ip \cdot x)$. We see from the form of the lagrangian that the theory is invariant under a global (constant) phase transformation, $\psi \to e^{i\theta}\psi$. What would happen if we allowed the phase change to depend on the point, x^μ, of spacetime? In that case, the phase is transformed by a different amount depending on the position in spacetime. The phase angle, θ, becomes $\theta(x)$. The transformation becomes an element of the Lie group $U(1)$. We are giving the particle an extra degree of freedom. Simply specifying the global wave function is not enough. To describe the particle fully, we must also know the value of the phase $\theta(x)$ at each spacetime point. We are constructing a circle, S^1, at each point in spacetime and specifying not only the wave as a function on spacetime, but also its position on this circle. Placing a copy of S^1 at each point in spacetime means that we are constructing a $U(1)$ bundle over spacetime. The wave function is a section of the $U(1)$ bundle.

In Appendix E, we have given a brief explanation of the mathematical theory of fibre bundles. We encourage readers to refer to this appendix as well as references

[20, 21, 22] if they are unfamiliar with this material.[1]

If the phase change is constant throughout space time (as in the case of a global gauge transformation), then the section of the $U(1)$ bundle is equivalent to the zero section. The $U(1)$ bundle is then unnecessary since the zero section is just a curve in spacetime itself.

Let us examine the abelian gauge transformations defined by the generators of the Lie group U(1),

$$\begin{aligned} \psi'(x) &= e^{-iq\Lambda(x)}\psi(x) \\ \overline{\psi}'(x) &= \overline{\psi}(x)e^{iq\Lambda(x)}. \end{aligned} \tag{3.4}$$

The infinitesimal parameter, Λ, is real and depends on x^μ. It is, therefore, a local gauge transformation. The infinitesimal (and therefore linear) form of equation (3.4) is given by

$$\psi'(x) \approx (1 - iq\Lambda(x))\psi(x) \tag{3.5}$$

and

$$\overline{\psi}'(x) \approx \overline{\psi}(x)(1 + iq\Lambda(x)). \tag{3.6}$$

[1]Later, we will be looking at frame bundles over spacetime where one must not blur the distinction between the bundle coordinates and the spacetime coordinates. A three index object may not be a tensor if all three indices are viewed as spacetime indices, but when two of them are frame bundle indices and the remaining index is a spacetime index, the object may be tensorial in the spacetime index; e.g., it may transform as a vector. It is, therefore, necessary to remember that the bundle over spacetime is a distinct space attached to each point in spacetime.

We have $\frac{\partial\psi'}{\partial\Lambda} = -iq\psi$, so the Noether current (3.22) is given by

$$J^\mu = -iq\left(\frac{\partial\mathcal{L}}{\partial\psi_{,\mu}}\,\psi - \overline{\psi}\frac{\partial\mathcal{L}}{\partial\overline{\psi}_{,\mu}}\right). \tag{3.7}$$

For a free spinor theory (Dirac lagrangian) where

$$\mathcal{L}_0 = \overline{\psi}\left(\frac{i}{2}\gamma^\mu\overleftrightarrow{\partial}_\mu - m\right)\psi, \tag{3.8}$$

we have $J^\mu = q\overline{\psi}\gamma^\mu\psi$.

What is the effect of this $U(1)$ degree of freedom on the particle? We know that the lagrangian of the particle is no longer invariant under this transformation.

$$\begin{aligned}
\mathcal{L}_0' &= \overline{\psi}'\left(\frac{i}{2}\gamma^\mu\overleftrightarrow{\partial}_\mu - m\right)\psi' \\
&= \overline{\psi}e^{iq\Lambda(x)}\left(\frac{i}{2}\gamma^\mu\overleftrightarrow{\partial}_\mu - m\right)e^{-iq\Lambda(x)}\psi \\
&= \overline{\psi}q\gamma^\mu\left(\partial_\mu\Lambda\right)\psi + \mathcal{L}_0 \\
&= J^\mu\,\partial_\mu\Lambda + \mathcal{L}_0.
\end{aligned}$$

Since $\mathcal{L}_0' \neq \mathcal{L}_0$, we define a new lagrangian, $\mathcal{L}$, containing an interaction term with the hope that we will then have enough freedom to make $\mathcal{L}$ gauge invariant. We add an interaction term to the original lagrangian containing a new field which couples to the current with strength, e, responsible for the loss or gain of energy. In this way, the entire lagrangian can be made invariant.

$$\mathcal{L} = \mathcal{L}_0 - \frac{e}{q}J^\mu A_\mu \tag{3.9}$$

so that (using the fact that J^μ is invariant)

$$\begin{aligned}\mathcal{L}' &= \mathcal{L}'_0 - \frac{e}{q} J^\mu A'_\mu \\ &= \mathcal{L}_0 + J^\mu \partial_\mu \Lambda - \frac{e}{q} J^\mu A'_\mu \\ &= \mathcal{L} + J^\mu \left(\frac{e}{q} A_\mu + \partial_\mu \Lambda - \frac{e}{q} A'_\mu \right).\end{aligned} \tag{3.10}$$

Hence, if $\mathcal{L}' = \mathcal{L}$, we must have $\frac{e}{q} A_\mu + \partial_\mu \Lambda - \frac{e}{q} A'_\mu = 0$, or if, for simplicity, we set $q = e$, we have finally

$$A'_\mu = A_\mu + \partial_\mu \Lambda. \tag{3.11}$$

Our general lagrangian, including a free part for the A_μ field, is now

$$\begin{aligned}\mathcal{L} &= \overline{\psi} \left(\frac{i}{2} \gamma^\mu \overleftrightarrow{\partial}_\mu - m \right) \psi - J^\mu A_\mu + \mathcal{L}_{\mathsf{free}} \\ &= \overline{\psi} \left(i\gamma^\mu (\partial_\mu + ieA_\mu) - m \right) \psi + \mathcal{L}_{\mathsf{free}},\end{aligned} \tag{3.12}$$

where we have simplified matters on the second line by having the derivative operate to the right only and inserting a factor of 2. We have discarded the total derivative term which would not contribute to the action.[2] Defining $D_\mu = \partial_\mu + ieA_\mu$ as the covariant derivative,

$$\mathcal{L} = \overline{\psi} \left(i\gamma^\mu D_\mu - m \right) \psi + \mathcal{L}_{\mathsf{free}}. \tag{3.13}$$

We would next like to find the form of the free lagrangian for the fields A^μ. We

[2]Here we are using the fact that the fields are localized and vanish asymptotically.

begin with the most general allowed form of the lagrangian given by

$$\mathcal{L}_{\text{free}} = AF_{\mu\nu}F^{\mu\nu} + BG_{\mu\nu}G^{\mu\nu} + M^2A^{\mu}A_{\mu}, \tag{3.14}$$

where M^2 is a possible mass term, and $F_{\mu\nu}$ and $G_{\mu\nu}$ are the general antisymmetric and symmetric combinations of the gauge fields and their derivatives

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{3.15}$$

and

$$G_{\mu\nu} = \partial_\mu A_\nu + \partial_\nu A_\mu. \tag{3.16}$$

We see from equation (3.11) that a gauge transformation leaves $F_{\mu\nu}$ invariant, but

$$G_{\mu\nu} \longrightarrow G_{\mu\nu} + 2\partial_\mu\partial_\nu\Lambda \tag{3.17}$$

and the mass term

$$A^\mu A_\mu \longrightarrow A^\mu A_\mu + 2A^\mu\partial_\mu\Lambda + \partial_\mu\Lambda\partial^\mu\Lambda. \tag{3.18}$$

Since neither of these terms is gauge invariant and our lagrangian is required to be, the simplest choice is $B = M^2 = 0$. Therefore, the gauge field, A^μ, is a massless vector field with a free lagrangian given by $AF_{\mu\nu}F^{\mu\nu}$. We will choose the standard normalization and write $A = -\frac{1}{4}$ so that

$$\mathcal{L} = \overline{\psi}\left(i\gamma^\mu D_\mu - m\right)\psi \; - \; \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{3.19}$$

is our full QED lagrangian. Notice that we can write $F_{\mu\nu}$ as an exact two form, $F = dA$, which means that F is closed and thus obeys the Bianchi identity, $dF = 0$, and also that F is only sensitive to the d'Rham cohomology equivalence classes on the gauge group manifold $[G] = \{A/d\Lambda\}$. The interested reader is referred to Appendix D for a more detailed discussion of the homology and cohomology of manifolds.

We now state several interesting results. $F_{\mu\nu}$, given in equation (3.15), is antisymmetric. Writing the Euler-Lagrange equations in the region of a source,

$$\partial_\mu F^{\mu\nu} = J^\nu, \tag{3.20}$$

we can see that

$$\partial_\nu J^\nu = \partial_\nu \partial_\mu F^{\mu\nu} = 0, \tag{3.21}$$

thus J^ν is conserved. Recognizing that equation (3.20) comprises two of Maxwell's equations (the other two following from $dF = 0$), we have electromagnetism coming from local abelian gauge invariance of a Dirac lagrangian.

According to Noether's theorem, symmetries of the action lead to conserved currents. Let us find the conserved current associated with *global* phase invariance. Given the global infinitesimal gauge transformation,

$$\psi' = e^{-i\theta}\psi \approx (1 - i\theta)\psi = \psi + \delta\psi, \tag{3.22}$$

we can vary the Dirac action as follows

$$
\begin{aligned}
S = \int_M L d^4x &= \int_M \left\{ \frac{\delta L}{\delta \psi}\delta\psi + \frac{\delta L}{\delta \partial_\mu \psi}\delta(\partial_\mu \psi) + \frac{\delta L}{\delta x^\mu}\delta x^\mu \right\} d^4x \\
&= \int_M \left\{ \frac{\delta L}{\delta \psi} - \partial_\mu \frac{\delta L}{\delta \partial_\mu \psi} \right\} \delta\psi d^4x + \int_M \partial_\mu \left[\frac{\delta L}{\delta \partial_\mu \psi}\delta\psi \right] d^4x,
\end{aligned}
\tag{3.23}
$$

where we have dropped the term that depends on x^μ since there is no explicit spacetime dependence in L. The variation of the action must vanish. The first term vanishes, producing the Euler-Lagrange equations. The surface term is

$$
\int_M \partial_\mu \left[\frac{\delta L}{\delta \partial_\mu \psi}\delta\psi \right] d^4x = \int_M \partial_\mu \left[\frac{\delta L}{\delta \partial_\mu \psi}(-i\theta\psi) \right] d^4x \equiv \int_M \partial_\mu J^\mu \theta d^4x. \tag{3.24}
$$

For this integral to vanish, we must have

$$
\partial_\mu J^\mu = 0 \tag{3.25}
$$

as our conserved Noether current where

$$
J^\mu = -i \frac{\delta L}{\delta \partial_\mu \psi}\psi = \overline{\psi}\gamma^\mu \psi.
$$

We can find the conserved charge associated with this current by integrating the charge density over all space. The charge density is defined as the time component of J^μ. We get

$$
\begin{aligned}
N &= \int_M J^0 d^3x = \int_M \overline{\psi}\gamma^0\psi d^3x \\
&= \int_M \psi^\dagger\gamma^0\gamma^0\psi d^3x = \int_M \left(\psi^\dagger\psi\right) d^3x.
\end{aligned}
\tag{3.26}
$$

We recognize the quantity in the integral as the quantum mechanical "number operator," thus we see that the net number of particles in space, N, is the conserved charge.

3.1.2. Yang-Mills gauge theory

In 1932, Heisenberg introduced the idea of isotopic spin (or isospin) as a way to distinguish the two charge states of the nucleon. It was found around that time, and confirmed experimentally many times since, that the neutron and the proton have nearly identical mass[3] and interact in the same way via the strong force. This observation indicated that the strong force treats the proton and the neutron like the same particle. Put another way, the strong force is only sensitive to isospin equivalence classes. The electromagnetic force, however, breaks this symmetry since the proton is electrically charged and the neutron is not, causing an isospin splitting. Thus, the proton and neutron can be distinguished from one another. However, at high enough energies where the strong force dominates the electromagnetic force, the approximation of treating them as two states of a single particle remains a good one.

Yang and Mills [24, 25] proposed, in analogy with quantum electrodynamics (QED),

[3] According to [23], the proton mass is $m_p = 938.27200 \pm 0.00004$ MeV $\simeq 1.672 \times 10^{-27}$ kg, and the neutron mass is $m_n = 939.56533 \pm 0.00004$ MeV $\simeq 1.675 \times 10^{-27}$ kg.

that perhaps there is a group of local transformations, a rotation in "isospin space," under which the QED lagrangian should be invariant. If we require our lagrangian to be invariant under independent rotations of the isospin vector (protons into neutrons and vice versa) at each point in spacetime, we are forcing the theory to describe interactions that are insensitive to isospin. Any isospin changes that occur in the lagrangian will be compensated by terms that absorb the change, leaving the lagrangian invariant. What we are essentially doing is asking, "What fields are necessary to compensate for any possible isospin change at any point in space time?" In another way, "What type of fields are required to interact with our particles to change their isospin?"

Such a group would have to rotate protons into neutrons and vice versa. Since we are thinking of protons and neutrons as vectors in two-dimensional isospin space, the transformations must be 2×2 matrices. A convenient set of matrices, used by Heisenberg in the quantum mechanics of isospin, is the set of Pauli matrices. Raising and lowering operators can be constructed which can then be used to increase and decrease the isospin quantum number of the nucleon.

The Pauli matrices form a Lie algebra. By exponentiation we can form a Lie group and use group elements as local transformations on quantum fields. A Lie group is just the type of transformation that will work in the technique that we introduced for $U(1)$ gauge invariance. Yang and Mills [24, 25] postulated that the fields representing the nucleons should take values in an $SU(2)$ bundle over spacetime.

Let $\psi = \begin{pmatrix} \psi_p \\ \psi_n \end{pmatrix}$ be a two-component wave function describing a field with isospin

$\frac{1}{2}$ like the nucleon. We transform the nucleon, ψ, as

$$\psi' = e^{-ig\vec{\tau}\cdot\vec{\theta}(x)}\psi = e^{-ig\tau_a\theta^a(x)}\psi, \tag{3.27}$$

where $\tau_a = \sigma_a$ is a Pauli matrix, a takes values in $\{1, 2, 3\}$, and g is a coupling strength parameter (not to be confused with the determinant of the metric). The Dirac lagrangian now becomes

$$\begin{aligned} L_0' &= \overline{\psi}'\left(i\,\partial\!\!\!/ - m\right)\psi' = \overline{\psi}e^{ig\tau_a\theta^a}\left(i\,\partial\!\!\!/ - m\right)e^{-ig\tau_a\theta^a}\psi \\ &= \overline{\psi}\left(i\,\partial\!\!\!/ + g\tau_a\,\partial\!\!\!/\theta^a(x) - m\right)\psi = \overline{\psi}\left(i\,\partial\!\!\!/ - m\right)\psi + g\overline{\psi}\tau_a\,\partial\!\!\!/\theta^a(x)\psi \\ &= L_0 + g\overline{\psi}\tau_a\gamma^\mu\partial_\mu\theta^a\psi \equiv L_0 + J_a^\mu\partial_\mu\theta^a, \end{aligned} \tag{3.28}$$

where the current, $J_a^\mu = g\tau_a\overline{\psi}\gamma^\mu\psi$, is now an Lorentz vector and an isovector carrying a representation of the su(2) algebra.

As we did in the $U(1)$ case, we introduce a new field, B_μ, which will make this lagrangian invariant. This time, however, we can see by the form of the current that B_μ will have to be a Lorentz vector and it must also carry a representation of the Lie algebra. Anticipating this difference, we define

$$B_\mu = i\tau_a\pi_\mu^a, \tag{3.29}$$

and we will insert the covariant derivative, $D_\mu = \partial_\mu + gB_\mu$, into our lagrangian. Invariance is achieved if our covariant derivative is actually covariant (since the mass term is already invariant). Denoting $e^{-ig\tau_b\theta^b} \equiv U$, we have

$$(D_\mu)' = UD_\mu U^{-1}. \tag{3.30}$$

Expanding both sides gives

$$\partial_\mu + gB'_\mu = \partial_\mu + U\left(\partial_\mu U^{-1}\right) + gUB_\mu U^{-1}. \tag{3.31}$$

Therefore, we must have

$$\begin{aligned} B'_\mu &= \frac{1}{g}U\partial_\mu U^{-1} + UB_\mu U^{-1} = i\tau_b\partial_\mu\theta^b + iU\tau_a\pi^a_\mu U^{-1} \\ &= i\tau_b\partial_\mu\theta^b + i\tau_a\pi^a_\mu + ig\left[\tau_a, \tau_b\right]\theta^b\pi^a_\mu \end{aligned} \tag{3.32}$$

and

$$\begin{aligned} i\tau_b\pi^{b\prime}_\mu &= i\tau_b\left(\partial_\mu\theta^b + \pi^b_\mu\right) - 2g\epsilon^{acb}\tau_b\theta_c\pi_{\mu a} \\ \pi^{b\prime}_\mu &= \pi^b_\mu + \partial_\mu\theta^b + 2ig\left(\vec{\pi}_\mu \times \vec{\theta}\right)^b. \end{aligned} \tag{3.33}$$

The current transforms, for small g, as

$$\begin{aligned} (J^a_\mu)' &= g\overline{\psi}'\tau^a\gamma^\mu\psi' = g\overline{\psi}e^{ig\tau_b\theta^b}\tau^a\gamma^\mu e^{-ig\tau_c\theta^c}\psi \\ &\approx g\overline{\psi}\left(1 + ig\tau^b\theta_b\right)\tau^a\gamma^\mu\left(1 - ig\tau^c\theta_c\right)\psi \\ &= J^\mu_a - 2g^2\overline{\psi}\gamma^\mu(\vec{\tau} \times \vec{\theta})_a\psi. \end{aligned} \tag{3.34}$$

Our total lagrangian for this theory can then be written as

$$\begin{aligned} L &= \overline{\psi}\left(i\,\partial\!\!\!/ - m\right)\psi - g\overline{\psi}\tau_a\gamma^\mu\psi\pi^a_\mu - \frac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a \\ &= \overline{\psi}\left(i\,D\!\!\!\!/ - m\right)\psi - \frac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a. \end{aligned} \tag{3.35}$$

The field strength can be derived as follows:

$$F = ig\tau_a F^a_{\mu\nu}dx^\mu \wedge dx^\nu = \left[D_\mu, D_\nu\right]dx^\mu \wedge dx^\nu \tag{3.36}$$

which gives

$$\begin{aligned} \tau_a F^a_{\mu\nu} &= -i\left(\partial_\mu B_\nu - \partial_\nu B_\mu\right) - ig\left[B_\mu, B_\nu\right] \\ &= \tau_a\left(\partial_\mu\pi^a_\nu - \partial_\nu\pi^a_\mu\right) + ig\left[\tau_b, \tau_c\right]\pi^b_\mu\pi^c_\nu \end{aligned} \tag{3.37}$$

so that

$$F^a_{\mu\nu} = \left(\partial_\mu\pi^a_\nu - \partial_\nu\pi^a_\mu\right) - 2g\left(\vec{\pi}_\mu \times \vec{\pi}_\nu\right)^a \equiv f^a_{\mu\nu} - 2g\left(\vec{\pi}_\mu \times \vec{\pi}_\nu\right)^a. \tag{3.38}$$

Finally, we can write our lagrangian,

$$\begin{aligned} L = &\overline{\psi}\left(i\,\partial\!\!\!/ - m\right)\psi - \frac{1}{4}f^a_{\mu\nu}f^{\mu\nu}_a \\ &- g\overline{\psi}\tau_a\gamma^\mu\psi\pi^a_\mu + gf^a_{\mu\nu}\left(\vec{\pi}_\mu \times \vec{\pi}_\nu\right)_a - g^2\left(\vec{\pi}^\mu \times \vec{\pi}^\nu\right)^a\left(\vec{\pi}_\mu \times \vec{\pi}_\nu\right)_a. \end{aligned} \tag{3.39}$$

We immediately see that the first two terms are the propagation of the free nucleon and the free π field, the third term is an interaction vertex with two nucleons and a π interacting at a point, and the last two terms represent three and four π fields, respectively, interacting with each other at a point.[4]

The equations of motion are

$$
\begin{aligned}
(i\,\partial\!\!\!/ - m)\,\psi - g\tau_a\gamma^\mu\pi^a_\mu\psi = (i\,D\!\!\!\!/ - m)\,\psi = 0 \\
\partial_\alpha f^{\alpha\beta b} + 2g\left(\vec{\pi}^\mu \times \vec{f_\mu}^{\beta}\right)^b - g\bar{\psi}\tau^b\gamma^\alpha\psi = 0 \\
\Rightarrow \partial_\alpha f^{\alpha\beta b} - 2g\left(\vec{\pi}_\mu \times \vec{f}^{\beta\mu}\right)^b - J^{\alpha b} = 0
\end{aligned}
\tag{3.40}
$$

The last equation is called the Yang-Mills equation and can be written

$$
\partial_\alpha \vec{f}^{\alpha\beta} = \vec{\mathcal{J}}^\alpha, \tag{3.41}
$$

where the conserved current is

$$
\vec{\mathcal{J}}^\alpha = \vec{J}^\alpha + 2g\left(\vec{\pi}_\mu \times \vec{f}^{\beta\mu}\right). \tag{3.42}
$$

We notice that isospin gives rise to the π. Integrating the zeroth component of the conserved current shows us that the π field also contributes to the total isospin, resulting in non-linear field equations for the π field. Non-linear equations are common in non-abelian theories, resulting in self-interactions among the gauge fields.

Let us pause for a moment and discuss this theory in its historical context. It is a theory describing the interactions of massless vector particles. In the particle spectrum of the 1940s, there were none other than the photon. The photon has no isotopic spin and is already described nicely by the formalism of QED. This massless vector would,

[4]Technically, we should also include ghost interactions coming from the Fadeev-Popov determinant which is included in the path integral during quantization to fix the gauge.

therefore, have to represent a new particle. The fact that the theory predicted a new massless particle was considered a problem[5] by Yang and Mills [24, 25] since they could not verify the theory experimentally. Another problem mentioned in the original paper by Yang and Mills [24, 25] was that the theory is beset by divergences like the ones that, at that time, plagued all field theories, including QED.

We could perhaps add a mass term to the lagrangian (3.39) of the form $m_\pi^2 \pi_\mu^a \pi^{\mu a}$ which would then imply that π is an isospin-1, spin-1 field. We could identify it with a vector meson. Unfortunately, such a mass term would no longer be invariant to our transformations. We would have to add more terms containing yet another new field to cancel the new changes.

Yang and Mills [24, 25] had a theory describing vector mesons interacting with nucleons. The theory could possibly be modified to describe scalar or pseudo-scalar mesons like the pion. It seemed unfortunate to have $U(1)$ gauge invariance giving QED immediately and fitting so well with experiment, but having to make major modifications to $SU(2)$ gauge invariance before it describes anything in reality. Therefore, the problem of pions and nucleons which forms the observational basis of low energy nuclear physics was not immediately solved by their work.

The $SU(2)$ gauge theory of the nucleon is obsolete now since it is known that the nucleons and mesons are composites containing quarks. This formulation of nuclear physics is fruitful and has led to effective theories that are still used extensively in nuclear physics research. At energies which are low enough so that the quark substructure

[5]This prediction would certainly not constitute a problem today. The prediction of new particles is a welcome, testable part of any modern particle theory.

of these particles remains hidden (i.e., A probe with wavelength much larger than the dimensions of the quarks is used.) The "effective" theory given here is useful.

Goldstone, Weinberg, and Salam [26] showed that if the lagrangian of a theory is invariant to a symmetry but the vacuum state is not invariant, some of the aforementioned problems are solved. In fact, forgetting about nucleons altogether, it has been shown that the theory governing the weak force, the Glashow-Weinberg-Salam model, is based, in part, on $SU(2)$ gauge invariance.

3.2. Gauge theory of gravity

We have seen in the previous section that an effective theory of nuclear physics incorporating $SU(2)$ invariance of the Dirac lagrangian results in a physical spectrum and interactions that look identical to those found in the laboratory at low energies (mesons). It is then interesting to wonder whether other theories that we believe to be true based on observation may also be merely a low energy limit of some more fundamental theory.

String theorists believe that gravitation and particle physics have, as their fundamental entities, strings rather than particles. The spectrum of the standard model is thought to be merely the low energy manifestation of the interactions among open and closed strings. Thus, it is possible that general relativity is merely a limit of a more fundamental quantum theory of gravity, such as string theory.

We know from experience that the theories of conventional particle physics in flat spacetime are derived in accordance with special relativity and are, therefore, Poincaré invariant. When we formulate these theories in curved spacetime, this Poincaré invariance must be a local, rather than a global, symmetry. Since gravity is thought to be a manifestation of a curved spacetime, we should learn something about gravitational interactions by insisting that the Dirac lagrangian be locally Poincaré invariant and finding the gauge fields necessary to preserve this invariance.

In this section, we will apply the same analysis to Poincaré invariance that we have used above with $U(1)$ and $SU(2)$ gauge invariance [3]. In order to facilitate our analysis, we need to construct an orthonormal frame bundle over spacetime. A frame bundle assigns a tetrad basis to each point in spacetime just as we assigned a Lie group to each point in spacetime in our analysis of $U(1)$ and $SU(2)$ gauge invariance. One difference is that the transformations in the fibre for the case of the Lie group were given by the actual Lie group elements themselves, whereas in the case of a frame bundle, our transformations in the fibre will be the Poincaré group acting on the tetrads. Another difference is that the gauge groups used before formed a compact fibre space, whereas the gauge group used here is non-compact.

Let us explain these points more clearly. Recall that, in the case of $U(1)$ theory, each particle wave function had a phase associated with it. At each point of spacetime, the wave function described a point on the circle S^1 which is the fibre of the $U(1)$ bundle. Gauge transformations (phases) which moved this point in the fibre were themselves elements of the group, i.e., points in the group manifold. In the case of a local frame bundle, the transformations which translate or rotate the frame at each point (move points within the fibre) are not themselves elements of the fibre. They are elements of the Poincaré group. In this case, our connection will not be "so(3, 1)-valued" in the same way that the connection was "u(1)-valued" or "su(2)-valued" in previous cases. Instead, it will take values in the Poincaré group (in a spinor representation). On the other hand, the particle wave functions themselves will be tetrad-valued in that their components will be given with respect to the local tetrad basis. Gauge transformations will rotate this basis and gauge fields must be introduced to compensate for the changes to the action.

Postulating that spacetime is curved, with the fields defined locally as functions in the frame bundle, we must eventually re-investigate gauge invariance for electrodynamics to see that it still holds. If it no longer holds, we may have to find further compensating fields (interactions with gauge fields) to fix the situation [27].

3.2.1. Poincaré transformations

Consider a Dirac spinor, ψ, in four-dimensional curved spacetime.[6] The Dirac matrices in flat spacetime satisfy the usual Clifford algebra relation

$$\{\gamma^a, \gamma^b\} = 2\eta^{ab}. \tag{3.43}$$

Under a local Lorentz transformation, ${\Lambda_b}^a(x)$, the spinor transforms as

$$\psi(x) \to \rho(\Lambda)\psi(x), \quad \overline{\psi}(x) \to \overline{\psi}(x)\rho(\Lambda)^{-1}, \tag{3.44}$$

[6]See [28] to review the techniques of quantum field theory in curved spacetime.

where $\rho(\Lambda)$ is the spinor representation of Λ. The massless Dirac lagrangian is

$$\mathcal{L} = \det\left(e^{\mu}{}_{a}\right)\frac{i}{2}\left(\overline{\psi}\gamma^{a}e^{\mu}{}_{a}\partial_{\mu}\psi - e^{\mu}{}_{a}\partial_{\mu}\overline{\psi}\gamma^{a}\psi\right), \tag{3.45}$$

where $\det\left(e^{\mu}{}_{a}\right) = \sqrt{-g}$.

We want a derivative which is covariant under local Lorentz transformations,

$$\nabla_{a}\psi \to \rho(\Lambda)\left(\Lambda^{-1}\right)_{a}{}^{b}\nabla_{b}\psi, \tag{3.46}$$

so we postulate one of the form

$$\nabla_{a}\psi = e^{\mu}{}_{a}\left[\partial_{\mu} + \Omega_{\mu}\right]\psi. \tag{3.47}$$

The gauge field Ω_{μ} must act like a derivation and obey the Leibnitz rule when acting on products so that the entire covariant derivative is both covariant and a derivative. Therefore, since Ω_{μ} is a matrix valued operator, its action must be the commutator. Under a local Lorentz transformation, the mass term of the Dirac lagrangian is invariant while the kinetic term transforms as

$$\begin{aligned}
\overline{\psi}\gamma^{a}\nabla_{a}\psi \to \overline{\psi}'\gamma^{a}\nabla'_{a}\psi' &= \left(\overline{\psi}\rho^{-1}\right)\gamma^{a}\left(\Lambda^{-1}\right)_{a}{}^{b}e^{\mu}{}_{b}\left(\partial_{\mu} + \Omega'_{\mu}\right)(\rho\psi) \\
&= \overline{\psi}\left(\rho^{-1}\gamma^{a}\rho\right)\rho^{-1}\left(\Lambda^{-1}\right)_{a}{}^{b}e^{\mu}{}_{b}\left(\partial_{\mu} + \Omega'_{\mu}\right)(\rho\psi) \\
&= \overline{\psi}\gamma^{c}\Lambda_{c}{}^{a}\left(\Lambda^{-1}\right)_{a}{}^{b}e^{\mu}{}_{b}\rho^{-1}\left(\partial_{\mu} + \Omega'_{\mu}\right)(\rho\psi) \\
&= \overline{\psi}\gamma^{a}e^{\mu}{}_{a}\left(\rho^{-1}\left(\partial_{\mu}\rho\right) + \partial_{\mu} + \rho^{-1}\left[\Omega'_{\mu}, \rho\right] + \Omega'_{\mu}\right)\psi.
\end{aligned} \tag{3.48}$$

If

$$\Omega' = \rho\Omega_\mu\rho^{-1} - (\partial_\mu\rho)\,\rho^{-1}, \tag{3.49}$$

then

$$i\overline{\psi}\gamma^a\nabla_a\psi \to i\overline{\psi}\gamma^a e_a^\mu\,(\partial_\mu + \Omega_\mu)\,\psi \tag{3.50}$$

is invariant.

Ω_μ is the compensating gauge field to keep the Dirac lagrangian invariant under a local Lorentz transformation in curved spacetime. To find an explicit form for Ω, we perform an infinitesimal transformation. Under an infinitesimal Poincaré transformation,

$$x^\mu \to x^\mu + \epsilon^\mu{}_\nu x^\nu + \epsilon^\mu, \tag{3.51}$$

a Dirac spinor, transforms as

$$\begin{aligned}\psi \to \exp\left\{\frac{i}{2}\epsilon^{ab}\sigma_{ab} - i\epsilon^\mu\partial_\mu\right\}\psi &\approx \left[1 - i\epsilon^\mu\partial_\mu + \frac{i}{2}\epsilon^{ab}\sigma_{ab}\right]\psi \\ &= \exp\left\{\frac{i}{2}\epsilon^{ab}\sigma_{ab}\right\}\exp\left\{-i\epsilon^\mu\partial_\mu\right\}\psi, \\ &= \exp\left\{\frac{i}{2}\epsilon^{ab}\sigma_{ab}\right\}\exp\left\{-\epsilon^\mu p_\mu\right\}\psi.\end{aligned} \tag{3.52}$$

Here, $\sigma_{ab} = \frac{i}{4}\left[\gamma_a, \gamma_b\right]$ is the spinor representation of the Lorentz generators satisfying the Lie algebra

$$\left[\sigma_{ab}, \sigma_{cd}\right] = i\left(\eta_{bc}\sigma_{ad} - \eta_{ac}\sigma_{bd} - \eta_{bd}\sigma_{ac} + \eta_{ad}\sigma_{bc}\right). \tag{3.53}$$

A word or two about the form of the transformation is now in order. We would like a covariant derivative to make the lagrangian invariant under Poincaré transformations.

The traditional rule in curved space physics has been "comma goes to semicolon" in which we replace all derivatives by covariant derivatives when we go from flat space to curved space. According to this rule, we should, in fact, replace the derivative which appears in the Poincaré infinitesimal transformation by a covariant derivative.

There are two reasons why we are not going to make this replacement, although it is standard practice in the gravitational literature. The first reason is to preserve the analogy with gauge theory in which the connection is placed in the action in order to compensate for the effect of the gauge transformation. It is only combined into a covariant derivative after the fact. If we include it in the gauge transformation, then we are, in fact, introducing a non-linearity in that we will now have to compensate for the compensating field itself. The second reason is because the "comma goes to semicolon" rule is applied in the lagrangian; therefore, if one writes down the transformed lagrangian, there will be terms proportional to the second derivative of the fields. One of the derivatives coming from the lagrangian itself and the other from the infinitesimal translation transformation. Now if we were to follow the rule and replace derivatives by covariant ones, we would be replacing second-order derivatives with covariant derivatives. The rule does not apply to second-order derivatives. It only applies to first-order differential equations. (See [1] for a discussion of the applicability of the "comma goes to semicolon" rule.) For these reasons, we shall keep the ordinary derivative in the transformation and find the gauge compensating field accordingly.

As in equation (3.49) the gauge field, Ω_μ, transforms as

$$\Omega_\mu \to \left[1 - \epsilon^\alpha p'_\alpha + \frac{i}{2}\epsilon^{ab}\sigma_{ab}\right]\Omega_\mu\left[1 + \epsilon^\beta p_\beta - \frac{i}{2}\epsilon^{cd}\sigma_{cd}\right]$$
$$- \left[-\left(\partial_\mu\epsilon^\alpha\right)p'_\alpha + \frac{i}{2}\partial_\mu\epsilon^{ab}\sigma_{ab} + \frac{i}{2}\epsilon^{ab}\sigma_{ab}\partial_\mu\right]\left[1 + i\epsilon^\beta p_\beta - \frac{i}{2}\epsilon^{cd}\sigma_{cd}\right], \quad (3.54)$$

where p'_α is the final fermion momentum and p_β is the initial fermion momentum. Thus,

$$\Omega'_\mu = \Omega_\mu + \frac{i}{2}\epsilon^{ab}\left[\sigma_{ab}, \Omega_\mu\right] - \frac{i}{2}\partial_\mu\epsilon^{ab}\sigma_{ab} - \epsilon^\alpha\left(p' - p\right)_\alpha\Omega_\mu \quad (3.55)$$

We see that the gauge field transforms under translations like a field of momentum, $q = p' - p$. Using equation (3.51) with equation (2.13), we find that the field transforms like a connection 1-form under the Lorentz part of the transformation. Keeping linear terms in ϵ, we have

$$\omega_b{}^a \to \omega_b{}^a + \epsilon_c{}^a\omega_b{}^c - \omega_c{}^a\epsilon_b{}^c - d\epsilon_b{}^a \quad (3.56)$$

which, in components, becomes

$$\Gamma_{\mu b}{}^a \to \Gamma_{\mu b}{}^a + \epsilon_c{}^a\Gamma_{\mu b}{}^c - \Gamma_{\mu c}{}^a\epsilon_b{}^c - \partial_\mu\epsilon_b{}^a. \quad (3.57)$$

The combination

$$\Omega_\mu = \frac{i}{2}\Gamma_\mu{}^{ab}\sigma_{ab} \quad (3.58)$$

transforms under Lorentz transformations as

$$\begin{aligned}
\Omega_\mu &\to \frac{i}{2}\left(\Gamma_\mu^{\ ab} + \epsilon_c^{\ b}\Gamma_\mu^{\ ac} - \Gamma_{\mu c}^{\ \ b}\epsilon^{ac} - \partial_\mu \epsilon^{ab}\right)\sigma_{ab} \\
&= \Omega_\mu + \frac{i}{2}\left(\epsilon_c^{\ b}\Gamma_\mu^{\ ac} - \Gamma_{\mu c}^{\ \ b}\epsilon^{ac}\right)\sigma_{ab} - \frac{i}{2}\partial_\mu \epsilon^{ab}\sigma_{ab} \\
&= \Omega_\mu + \frac{i}{2}\left(\Gamma_\mu^{\ bc} - \Gamma_\mu^{\ cb}\right)\epsilon^a_{\ c}\sigma_{ab} - \frac{i}{2}\partial_\mu \epsilon^{ab}\sigma_{ab},
\end{aligned} \tag{3.59}$$

where we have used the antisymmetry of σ_{ab}. Notice that

$$\begin{aligned}
\frac{i}{2}\epsilon^{ab}\left[\sigma_{ab}, \Omega_\mu\right] &= -\frac{1}{4}\epsilon^{ab}\Gamma_\mu^{cd}\left[\sigma_{ab}, \sigma_{cd}\right] \\
&= \frac{i}{4}\epsilon^{ab}\Gamma_\mu^{cd}\left(\eta_{bc}\sigma_{ad} - \eta_{ac}\sigma_{bd} - \eta_{bd}\sigma_{ac} + \eta_{ad}\sigma_{bc}\right) \\
&= -\frac{i}{2}\left(\epsilon^a_{\ c}\Gamma_\mu^{\ cb} - \epsilon^a_{\ c}\Gamma_\mu^{\ bc}\right)\sigma_{ab}.
\end{aligned} \tag{3.60}$$

Thus equation (3.59) becomes

$$\Omega_\mu \to \Omega_\mu + \frac{i}{2}\epsilon^{ab}\left[\sigma_{ab}, \Omega_\mu\right] - \frac{i}{2}\partial_\mu \epsilon^{ab}\sigma_{ab} \tag{3.61}$$

which, according to equation (3.55), is the proper transformation under Lorentz transformations. Ω_μ must also transform under translations as

$$\Omega_\mu \to \Omega_\mu - \epsilon^\alpha q_\alpha \Omega_\mu, \tag{3.62}$$

where q_α is the momentum transfer.

Now, we can write down the kinetic term which is a scalar under both coordinate changes and Lorentz rotations,

$$\mathcal{L} = \det\left(e^{\mu}{}_{a}\right)\frac{i}{2}\left[\overline{\psi}\gamma^{a}e^{\mu}{}_{a}\left(\partial_{\mu}+\Omega_{\mu}\right)\psi - \left(\partial_{\mu}\overline{\psi}-\overline{\psi}\Omega_{\mu}\right)e^{\mu}{}_{a}\gamma^{a}\psi\right]$$
$$= \det\left(e^{\mu}{}_{a}\right)\frac{i}{2}\left[\overline{\psi}\gamma^{a}e^{\mu}{}_{a}\partial_{\mu}\psi - \partial_{\mu}\overline{\psi}e^{\mu}{}_{a}\gamma^{a}\psi + \frac{i}{2}\overline{\psi}e^{\mu}{}_{a}\Gamma_{\mu}{}^{cd}\left(\gamma^{a}\sigma_{cd}+\sigma_{cd}\gamma^{a}\right)\psi\right]. \tag{3.63}$$

Using the identity

$$\gamma^{a}\gamma^{b}\gamma^{c} = \gamma^{a}\eta^{bc} - \gamma^{b}\eta^{ac} + \gamma^{c}\eta^{ab} + i\epsilon^{abcd}\gamma^{5}\gamma_{d}, \tag{3.64}$$

equation (3.63) becomes

$$\mathcal{L} = \det\left(e^{\mu}{}_{a}\right)\frac{i}{2}\left[\overline{\psi}\gamma^{a}e^{\mu}{}_{a}\partial_{\mu}\psi - \partial_{\mu}\overline{\psi}e^{\mu}{}_{a}\gamma^{a}\psi - \frac{i}{2}\overline{\psi}e^{\mu}{}_{a}\epsilon^{a}{}_{cdb}\gamma^{5}\gamma^{b}\Gamma_{\mu}{}^{cd}\psi\right]. \tag{3.65}$$

Our invariant action is then

$$S = \int_{M} d^{4}x\sqrt{-g}\,\frac{i}{2}\left[\overline{\psi}\gamma^{a}e^{\mu}{}_{a}\partial_{\mu}\psi - \partial_{\mu}\overline{\psi}e^{\mu}{}_{a}\gamma^{a}\psi - \frac{1}{2}\overline{\psi}\sigma_{abc}\widetilde{\Gamma}^{abc}\psi\right], \tag{3.66}$$

where we have placed a tilde on the connection to remind us that we have assumed no symmetry properties other than antisymmetry in the ab indices due to contraction with σ_{ab}. We have also defined

$$\sigma^{abc} = i\epsilon^{abcd}\gamma^{5}\gamma_{d}. \tag{3.67}$$

Separating the connection into parts according to equation (2.74) gives

$$S = \int_M d^4x\sqrt{-g}\,\frac{i}{2}\left[\overline{\psi}\gamma^a e^\mu{}_a\partial_\mu\psi - \partial_\mu\overline{\psi}e^\mu{}_a\gamma^a\psi - \frac{1}{2}\overline{\psi}\sigma^a{}_{bc}e^\mu{}_a\left(\Gamma_\mu{}^{bc} + K_\mu{}^{bc} + \widetilde{Q}_\mu{}^{bc}\right)\psi\right] \tag{3.68}$$

The term containing the non-metricity is symmetric in its last two indices, which are contracted with an antisymmetric object. Therefore, the term vanishes, leaving only

$$S = \int_M d^4x\sqrt{-g}\left\{\frac{i}{2}\overline{\psi}\gamma^a e^\mu{}_a\partial_\mu\psi - \frac{i}{2}\partial_\mu\overline{\psi}e^\mu{}_a\gamma^a\psi - \frac{i}{4}\overline{\psi}\sigma^a{}_{bc}e^\mu{}_a\left(\Gamma_\mu{}^{bc} + K_\mu{}^{bc}\right)\psi\right\}. \tag{3.69}$$

The symmetry of the Christoffel connection reduces our action to

$$S = \int_M d^4x\sqrt{-g}\left\{\frac{i}{2}\overline{\psi}\gamma^a e^\mu{}_a\partial_\mu\psi - \frac{i}{2}\partial_\mu\overline{\psi}e^\mu{}_a\gamma^a\psi - \frac{i}{4}\overline{\psi}\sigma_{abc}H^{abc}\psi\right\}, \tag{3.70}$$

where we have written

$$\sigma_{abc}K^{abc} = \sigma_{abc}H^{abc} \tag{3.71}$$

since only the totally antisymmetric part of K^{abc} will survive the contraction with σ_{abc}. We see that the fermions interact only with the totally antisymmetric part of the torsion.

The antisymmetric tensor field as it is written has no kinetic term in the Einstein-Cartan action. Hence, the field cannot propagate. In order to get a kinetic term and thus a physical field, we will have to augment the theory in some way. One way is to assume that the torsion is derivable from a potential. Then, the usual Einstein-Cartan action would contain a kinetic energy term. We will explore the implications of this theory in Chapter 5 by calculating some important scattering diagrams.

For the time being, the most natural course from the point of view of the particle physicist is to continue in the same fashion as we did in the $U(1)$ and $SU(2)$ cases. We form a kinetic energy lagrangian with the curvature. We form the curvature from the commutator of covariant derivatives, and we form a scalar from the curvature to constitute our action. One possibility will give us the scalar curvature of the usual Einstein-Cartan gravity. In that case, the metric has to be considered the fundamental field (the potential from which the other fields are derived) since the other fields do not propagate. Another option is to use a quadratic action in the fashion done in the previous section. We would then have fourth-order derivatives of the metric field so that it would not be a kinetic energy term. Hence, the metric could not be a fundamental field in the theory. The fundamental gauge field for local Lorentz invariance of the Dirac action is the connection. It should be this field which propagates. We are, thus, postulating a direct analogy:

$$A^\mu(\mathrm{QED}) \to \lambda^a A^\mu_a(\mathrm{QCD}) \to \sigma^{ab} A^\mu_{ab}(\mathrm{gravity}), \tag{3.72}$$

where we have written $\frac{1}{2}\Gamma^\mu_{ab} \equiv A^\mu_{ab}$.

Let us proceed and form the field strength as the commutator of covariant derivatives,

$$e^\mu{}_c e^\nu{}_d F_{\mu\nu} = \frac{i}{2} e^\mu{}_c e^\nu{}_d \left(\sigma_{ab} F^{ab}_{\mu\nu} + i F^\beta_{\mu\nu} D_\beta\right) = \left[D_c, D_d\right], \tag{3.73}$$

where

$$[D_c, D_d] = e^\mu{}_c (\partial_\mu + \Omega_\mu\cdot)\, e^\nu{}_d (\partial_\nu + \Omega_\nu\cdot) - e^\nu{}_d (\partial_\nu + \Omega_\nu\cdot)\, e^\mu{}_c (\partial_\mu + \Omega_\mu\cdot)$$
$$= e^\mu{}_c e^\nu{}_d (\partial_\mu\Omega_\nu - \partial_\nu\Omega_\mu + [\Omega_\mu, \Omega_\nu]) + (e^\mu{}_c D_\mu e^\nu{}_d - e^\mu{}_d D_\mu e^\nu{}_c)\, D_\nu, \tag{3.74}$$

and we have used the fact that Ω_μ is a derivation with its product being the commutator as is usual for Lie algebra-valued fields.[7] For example,

$$\Omega_\mu \cdot (e^\nu{}_d \Omega_\nu) = [\Omega_\mu, e^\nu{}_d]\, \Omega_\nu + e^\nu{}_d\, [\Omega_\mu, \Omega_\nu] + e^\nu{}_d \Omega_\nu \Omega_\mu. \tag{3.75}$$

Putting in expression (3.58) for Ω_μ, we get

$$\frac{i}{2}\left(\sigma_{ab} F^{ab}_{\mu\nu} + i F^\beta_{\mu\nu} D_\beta\right) = \frac{i}{2}\left(\partial_\mu \Gamma_\nu{}^{ab}\sigma_{ab} - \partial_\nu \Gamma_\mu{}^{ab}\sigma_{ab}\right) - \frac{1}{4}\left[\sigma_{rs}, \sigma_{tu}\right] \Gamma_\mu{}^{rs}\Gamma_\nu{}^{tu}$$
$$+\, e_\mu{}^c e_\nu{}^d \left(e^\alpha{}_c D_\alpha e^\beta{}_d - e^\alpha{}_d D_\alpha e^\beta{}_c\right) D_\beta. \tag{3.76}$$

Using the defining properties of the Lie algebra (3.53), equation (3.76) reduces to

$$\sigma_{ab} F^{ab}_{\mu\nu} + i F^\beta_{\mu\nu} D_\beta = \sigma_{ab}\left(\partial_\mu \Gamma_\nu{}^{ab} - \partial_\nu \Gamma_\mu{}^{ab} + 2\Gamma_\mu{}^{la}\Gamma_{\nu l}{}^{b}\right) + 2\left(e_\mu{}^c D_\nu e^\beta{}_c - e_\nu{}^c D_\mu e^\beta{}_c\right) i D_\beta, \tag{3.77}$$

and our field strength is, therefore, given by the two quantities:

$$F^{ab}_{\mu\nu} = 2\left(\partial_{[\mu}\Gamma_{\nu]}{}^{ab} + \Gamma_{[\mu|l}{}^{a}\Gamma_{|\nu]}{}^{lb}\right) \tag{3.78}$$

[7] If the connection were not a derivation, then equation (3.79) would give the anholonomity rather than the torsion. Hehl et. al. [4] define the object of anholonomity as $\Omega_{\mu\nu}{}^a = \partial_{[\mu} e_{\nu]}{}^a$. We have used the fact that $\partial_\mu\left(e_\nu{}^a e^\rho{}_a\right) = 0$ to find $e^\beta{}_a \Omega_{\mu\nu}{}^a = 2 e_{[\mu}{}^a \partial_{\nu]} e^\beta{}_a$.

and

$$F^{\beta}_{\mu\nu} = 4e_{[\mu}{}^{c}D_{\nu]}e^{\beta}{}_{c} = 2e^{\beta}{}_{c}S_{\mu\nu}{}^{c}, \tag{3.79}$$

where we have used the antisymmetry of the Γ's in the Latin indices and explicitly antisymmetrized over $\mu\nu$. The field strength separates into two pieces. The first is a curvature-type piece, and the second is proportional to the torsion. Notice that, if the Latin indices were changed to Greek, then the first term would be exactly the Riemann curvature tensor.

The most natural step here would be to form the quadratic action, Tr $F_{\mu\nu}F^{\mu\nu}$, in exact analogy with previous sections. Instead, most authors [3, 4, 12] form a scalar curvature action since it is the simplest scalar action and gives the metric tensor as the propagating graviton

$$\begin{aligned} F^{ab}_{\mu\nu} &= \partial_{\mu}\left(e_{\alpha}{}^{a}e_{\beta}{}^{b}\right)\Gamma_{\nu}{}^{\alpha\beta} - \partial_{\nu}\left(e_{\alpha}{}^{a}e_{\beta}{}^{b}\right)\Gamma_{\mu}{}^{\alpha\beta}e_{\alpha}{}^{a}e_{\alpha}{}^{a} \\ &\quad + \left(\partial_{\mu}\Gamma_{\nu}{}^{\alpha\beta} - \partial_{\nu}\Gamma_{\mu}{}^{\alpha\beta} + \Gamma_{\mu\lambda}{}^{\beta}\Gamma_{\nu}{}^{\alpha\lambda} - \Gamma_{\nu\lambda}{}^{\beta}\Gamma_{\mu}{}^{\alpha\lambda}\right) \\ &= \partial_{\mu}\left(e_{\alpha}{}^{a}e_{\beta}{}^{b}\right)\Gamma_{\nu}{}^{\alpha\beta} - \partial_{\nu}\left(e_{\alpha}{}^{a}e_{\beta}{}^{b}\right)\Gamma_{\mu}{}^{\alpha\beta} + e_{\alpha}{}^{a}e_{\beta}{}^{b}R_{\mu\nu}{}^{\alpha\beta} \\ &= \partial_{\mu}\left(e_{\alpha}{}^{a}e_{\beta}{}^{b}\right)\Gamma_{\nu}{}^{\alpha\beta} - \partial_{\nu}\left(e_{\alpha}{}^{a}e_{\beta}{}^{b}\right)\Gamma_{\mu}{}^{\alpha\beta} + e_{\alpha}{}^{a}e_{\beta}{}^{b}R_{\mu\nu}{}^{\alpha\beta}. \end{aligned} \tag{3.80}$$

Writing equation (3.80) as

$$F^{\sigma\rho}_{\mu\nu} = e^{\sigma}{}_{a}e^{\rho}{}_{b}\partial_{\mu}\left(e_{\alpha}{}^{a}e_{\beta}{}^{b}\right)\Gamma_{\nu}{}^{\alpha\beta} - e^{\sigma}{}_{a}e^{\rho}{}_{b}\partial_{\nu}\left(e_{\alpha}{}^{a}e_{\beta}{}^{b}\right)\Gamma_{\mu}{}^{\alpha\beta} + R_{\mu\nu}{}^{\sigma\rho} \tag{3.81}$$

and comparing with equation (2.77), we have identified the curvature.

We have found that the field strength of our gauge field contains the curvature tensor with antisymmetric connection. Hence, if the curvature scalar is the gauge invariant lagrangian density for our theory, it contains exactly the lagrangian for Einstein-Cartan gravity with totally antisymmetric torsion as in equation (2.87). It is remarkable that local Lorentz invariance of the Dirac lagrangian leads to interaction with a Kalb-Ramond antisymmetric tensor field and the Einstein-Cartan action.

CHAPTER 4

STRING THEORY

Antisymmetric tensor fields arise naturally as fundamental fields in 11-dimensional supergravity [29], M-theory [30, 31, 32], and string theory. It is well known that the field strength of the antisymmetric tensor field in string theory can be identified with spacetime torsion. It has been argued [27] that such an identification is indeed necessary to preserve $U(1)$ gauge symmetries in spacetime with torsion.

String theory and supergravity theories frequently invoke mathematical techniques which may be unfamiliar to many physicists. We have, therefore, included extensive appendices introducing the mathematical background. Appendix B is an introduction to supersymmetry, Appendix C is an introduction to the classical non-relativistic string and superstring theory, and Appendices E and D introduce mathematical subjects which have become integral to much of modern high energy physics including string theory.

In this chapter, we take a brief look at the origin of the antisymmetric tensor field in string theory [6, 33, 34, 35]. We generally follow the paper by Kalb and Ramond [33] where the antisymmetric "Kalb-Ramond" field was first shown to arise via interacting string world sheets.

4.1. Introduction

The standard model is the current theory of the elementary particles of nature. These particles constitute the fundamental building blocks of everything that we see around us. The many particles included in the standard model are all assumed to be fundamental and indivisible. String theory suggests that these particles are not, in fact, truly fundamental and are actually distinct states of fundamental strings. We can group the particles of the standard model into two general classes, fermions and bosons. The two classes differ by a quantum number called *spin*. Fermions have half integral spin while bosons have integral spin. Physically, the difference is that fermions must obey a Pauli exclusion principle, causing them to build up the matter that we see around us, while the bosons make up the force fields acting on the matter.

In order to classify the particles of the standard model even further, we group them into three theories based on the force carrying bosons by which the fermion matter fields interact with each other. The three theories are QED, the theory of weak interactions, and the theory of strong interactions. The weak interaction, responsible for radioactive beta decay, is a force between quarks and leptons mediated by bosons called the $W^{\pm}$ and the Z^0. The strong interaction is the part of the standard model called Quantum Chromodynamics (QCD), describing the interactions between quarks mediated by gluons.

The standard model is a theory of the dynamics and interactions of these basic particles. The reader is invited to refer to [18] for the theoretical details of the standard model. Details on the classification of the particles, the current values of the parameters, and the experiments that are underway which test the standard model can be found in the most recent review of particle physics [23].

Before it was realized that protons, neutrons, and other particles were made up of quarks, physicists struggled to understand why so many high-mass resonances show up in accelerator experiments. The quark model explains these resonances as excited states of the proton and neutron due to the state of the quarks inside. Before this explanation was accepted, many people thought that these resonances could be explained if the particles were actually little loops of string and the resonances were like the normal modes of the string. This model was successful initially and was taken quite seriously until quarks were discovered. The theory of fundamental strings was then dropped as an explanation for these resonances.

The standard model of particle physics does not include a description of the force of gravity. Thus, high energy physics theorists have speculated on possible extensions of the standard model to include the gravitational force. There have been many attempts to extend the standard model. They are grouped under the subject heading *quantum gravity*. String theory is now considered to be one such attempt.

One of the reasons why it has been so difficult to achieve this unification of gravity with the rest of particle physics is because the classical theory that accurately describes gravity is based on the curvature of spacetime rather than a particle mediating the force. This change in viewpoint causes problems since, to quantize this force, one would be quantizing spacetime itself. There are many pitfalls associated with this idea.

There have been two distinct, but "dual," viewpoints that theorists have taken when developing unified theories. On the one hand, they appreciate the calculational methods and experimental successes of quantum field theory. They would like to express gravity as a gauge theory on par with QED and QCD. Gravitational forces would be caused by the exchange of field quanta called gravitons. On the other hand, some theorists see the beauty of the geometrical viewpoint taken by Einstein and would like the geometry of the universe to be of fundamental significance. In this paradigm, the other apparent "forces" (electroweak and strong) between particles are due merely to the motion of particles along geodesics in a curved geometry. Rather than postulating force-mediating particles, in this point of view, we postulate extra dimensions in our universe such that the other known forces of nature also arise from geometry. Gravity is a manifestation of curvature in the usual four spacetime dimensions while the other forces are a manifestation of curvature in the extra, otherwise unseen, dimensions. Theories based on this point of view are now generally called Kaluza-Klein theories [36, 37]. A review of Kaluza-Klein gravity can be found in [38].

It is very difficult to choose between the various unification schemes due to the scarcity of predictions which lend themselves to unequivocal experimental verification. Instead, theorists have used mathematical consistency as the major requirement. It is for these reasons that any possible experimental signatures of these theories must be found and examined.

In the remainder of this chapter, we would like to show how superstring theory includes an antisymmetric tensor field identical to the one that we found in Einstein-Cartan gravity in the previous chapters. In this way, we can view the interactions of the antisymmetric field as testable predictions of these theories.

4.2. Kalb-Ramond field

In 1974, M. Kalb and P. Ramond [33] formulated a theory of interactions among open and closed strings in the course of generalizing the interactions between fields representing point particles. In this chapter, we will describe these interactions and show that they give rise to the antisymmetric tensor field we have studied in the previous chapters.

4.2.1. Free point particles and strings

A point particle in a flat spacetime with metric $\eta_{\mu\nu}$ traces out a world line with squared length

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu. \tag{4.1}$$

If we introduce an intrinsic "time" parameter, τ, along the world line, equation (4.1) can be written

$$ds^2 = \eta_{\mu\nu} \frac{dx^\mu}{d\tau} d\tau \frac{dx^\nu}{d\tau} d\tau \tag{4.2}$$

so that, with $\dot{x}^\mu = \frac{dx^\mu}{d\tau}$, we have

$$ds = \sqrt{\eta_{\mu\nu} \dot{x}^\mu \dot{x}^\nu} d\tau \tag{4.3}$$

as the world line length element. Similarly, a string traces out a world sheet, $x^\mu(\tau, \sigma)$, parametrized by τ and an intrinsic length parameter, σ, with infinitesimal area,

$$\begin{aligned} dS^{\mu\nu} = dx^\mu \wedge dx^\nu &= \left(\frac{dx^\mu}{d\tau}d\tau + \frac{dx^\mu}{d\sigma}d\sigma\right) \wedge \left(\frac{dx^\nu}{d\tau}d\tau + \frac{dx^\nu}{d\sigma}d\sigma\right) \\ &= \left(\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\sigma} - \frac{dx^\mu}{d\sigma}\frac{dx^\nu}{d\tau}\right) d\tau_a \wedge d\sigma_a \equiv S^{\mu\nu} d\tau \wedge d\sigma, \end{aligned} \tag{4.4}$$

where

$$S^{\mu\nu} = \dot{x}^\mu x'^\nu - x'^\mu \dot{x}^\nu \tag{4.5}$$

and $x'^\mu = \frac{dx^\mu}{d\sigma}$. Note that the integration of the differential form, $dS^{\mu\nu}$, is defined by

$$\int dS^{\mu\nu} = \int S^{\mu\nu} d\tau d\sigma \tag{4.6}$$

so that the wedge product becomes the ordinary tensor product inside the integrand.

Let us introduce a convenient operator,

$$D^\mu = \frac{\partial x^\mu}{\partial \tau}\frac{\partial}{\partial \sigma} - \frac{\partial x^\mu}{\partial \sigma}\frac{\partial}{\partial \tau}, \tag{4.7}$$

so that, when acting on a function $f(x)$, it gives

$$\begin{aligned} D^\mu f(x) &= \frac{\partial x^\mu}{\partial \tau}\frac{\partial f}{\partial \sigma} - \frac{\partial x^\mu}{\partial \sigma}\frac{\partial f}{\partial \tau} = \frac{\partial x^\mu}{\partial \tau}\frac{\partial x^\nu}{\partial \sigma}\frac{\partial f}{\partial x^\nu} - \frac{\partial x^\mu}{\partial \sigma}\frac{\partial x^\nu}{\partial \tau}\frac{\partial f}{\partial x^\nu} \\ &= S^{\mu\nu}\partial_\nu f. \end{aligned} \tag{4.8}$$

In particular, we have

$$D^\mu x^\nu = S^{\mu\nu}. \tag{4.9}$$

The action for the free point particle and free string are given by the length of the world line and the area of the world sheet, respectively:

$$S_{\text{free particle}} = m \int ds = m \int_{\tau_i}^{\tau_f} \sqrt{dx^\mu dx_\mu} \tag{4.10}$$

and

$$S_{\text{free string}} = -\mu^2 \int dA = -\mu^2 \int_{\tau_i}^{\tau_f} \int_{\sigma=0}^{\sigma=l} \sqrt{-dS^{\mu\nu} dS_{\mu\nu}}. \tag{4.11}$$

We have set the length of the string to l, we have inserted constants m and μ to make the action dimensionless in natural units, and the minus sign is in the square root of the string action to give a positive argument since we are only considering spacelike strings. In each case, the intrinsic time is integrated from some initial time, τ_i, to some final time, τ_f.

To find the free particle and free string equations of motion, we require the action to be stationary under $x^\mu \to x^\mu + \delta x^\mu$. We assume that the variation δx^μ vanishes at the endpoints of the particle world line and at the temporal boundary of the string world sheet. (i.e., We assume fixed initial and final string positions. We do not require that the spatial endpoints of the string be fixed.) Variation of the free particle action gives

$$\begin{aligned} \delta S_{\text{free particle}} &= 0 = m \frac{d}{d\tau} \frac{\dot{x}^\mu}{\sqrt{\dot{x} \cdot \dot{x}}} \\ &\Rightarrow \frac{d\dot{x}^\mu}{d\tau} = 0, \end{aligned} \tag{4.12}$$

where we have defined $\dot{x} \cdot \dot{x} = \dot{x}_\mu \dot{x}^\mu$. The equations of motion of the free point particle generate straight lines in spacetime or, in other words, constant velocity motion in space as intuitively expected. Varying the string action, we get

$$\delta S_{\text{free string}} = \mu^2 \int_{\tau_i}^{\tau_f} \int_{\sigma=0}^{\sigma=l} d\tau \, d\sigma \left(-S^{\alpha\beta} S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu} \delta S^{\mu\nu}. \tag{4.13}$$

Notice that

$$\begin{aligned} S_{\mu\nu}\delta S^{\mu\nu} &= S_{\mu\nu}\left[(\delta \dot{x}^{\mu}) x'^{\nu} + \dot{x}^{\mu}(\delta x'^{\nu}) - (\delta \dot{x}^{\nu}) x'^{\mu} - \dot{x}^{\nu}(\delta x'^{\mu})\right] \\ &= S_{\mu\nu}\left[x'^{\nu}\frac{d}{d\tau}\delta x^{\mu} + \dot{x}^{\mu}\frac{d}{d\sigma}\delta x^{\nu} - x'^{\mu}\frac{d}{d\tau}\delta x^{\nu} - \dot{x}^{\nu}\frac{d}{d\sigma}\delta x^{\mu}\right] \\ &= 2S_{\mu\nu}\left[\dot{x}^{\mu}\frac{d}{d\sigma}\delta x^{\nu} - x'^{\mu}\frac{d}{d\tau}\delta x^{\nu}\right] = 2S_{\mu\nu}D^{\mu}\delta x^{\nu}, \end{aligned} \tag{4.14}$$

where we have used the antisymmetry of $S_{\mu\nu}$ in deriving the last step. We now use equation (4.14) in the integrand of equation (4.13). Integrating by parts, we obtain the following expression for the variation of the integrand:

$$\begin{aligned} &\left(-S^{\alpha\beta}S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu}\delta S^{\mu\nu} = \\ &\left\{2\frac{d}{d\sigma}\left[\left(-S^{\alpha\beta}S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu}\dot{x}^{\mu}\delta x^{\nu}\right] - 2\frac{d}{d\tau}\left[\left(-S^{\alpha\beta}S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu}x'^{\mu}\delta x^{\nu}\right]\right. \\ &\left. - 2\frac{d}{d\sigma}\left[\left(-S^{\alpha\beta}S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu}\dot{x}^{\mu}\right]\delta x^{\nu} + 2\frac{d}{d\tau}\left[\left(-S^{\alpha\beta}S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu}x'^{\mu}\right]\delta x^{\nu}\right\} \end{aligned} \tag{4.15}$$

so that

$$
\begin{aligned}
\delta S_{\text{free string}} = 2\mu^2 \Bigg\{ & \int_{\tau_i}^{\tau_f} d\tau \left[\left(-S^{\alpha\beta} S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu} \dot{x}^\mu \delta x^\nu \right]_{\sigma=0}^{\sigma=l} \\
& - \int_{\sigma=0}^{\sigma=l} d\sigma \left[\left(-S^{\alpha\beta} S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu} x'^\mu \delta x^\nu \right]_{\tau_i}^{\tau_f} \\
& - \int_{\tau_i}^{\tau_f} \int_{\sigma=0}^{\sigma=l} d\tau \, d\sigma \, \dot{x}^\mu \frac{d}{d\sigma} \left[\left(-S^{\alpha\beta} S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu} \right] \delta x^\nu \\
& + \int_{\tau_i}^{\tau_f} \int_{\sigma=0}^{\sigma=l} d\tau \, d\sigma \, x'^\mu \frac{d}{d\tau} \left[\left(-S^{\alpha\beta} S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu} \right] \delta x^\nu \\
& - \int_{\tau_i}^{\tau_f} \int_{\sigma=0}^{\sigma=l} d\tau \, d\sigma \, \left[\left(-S^{\alpha\beta} S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu} \right] \left(\frac{d}{d\sigma} \dot{x}^\mu \right) \delta x^\nu \\
& + \int_{\tau_i}^{\tau_f} \int_{\sigma=0}^{\sigma=l} d\tau \, d\sigma \, \left[\left(-S^{\alpha\beta} S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu} \right] \left(\frac{d}{d\tau} x'^\mu \right) \delta x^\nu \Bigg\}.
\end{aligned}
\tag{4.16}
$$

The first two terms are boundary terms which we will discuss in a moment. The second two terms can be combined using equation (4.8). The last two terms cancel each other since partials derivatives commute here. We are left with

$$
\begin{aligned}
\delta S_{\text{free string}} = 2\mu^2 \Bigg\{ & \int_{\tau_i}^{\tau_f} d\tau \left[\left(-S^{\alpha\beta} S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu} \dot{x}^\mu \delta x^\nu \right]_{\sigma=0}^{\sigma=l} \\
& - \int_{\sigma=0}^{\sigma=l} d\sigma \left[\left(-S^{\alpha\beta} S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu} x'^\mu \delta x^\nu \right]_{\tau_i}^{\tau_f} \\
& - \int_{\tau_i}^{\tau_f} \int_{\sigma=0}^{\sigma=l} d\tau \, d\sigma \, D^\mu \left[\left(-S^{\alpha\beta} S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu} \right] \delta x^\nu \Bigg\}.
\end{aligned}
\tag{4.17}
$$

The second term in equation (4.17) is zero since the initial and final positions of the string are assumed fixed. Let us examine the first term in more detail. The integrand is

$$
2\mu^2 \left(-S^{\alpha\beta} S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu} \dot{x}^\mu \delta x^\nu \Big|_{\sigma=0}^{\sigma=l}. \tag{4.18}
$$

For closed strings, there is no string boundary since $x^\mu(\tau, 0) = x^\mu(\tau, l)$. Therefore,

$\delta x^\mu(\tau,0) = \delta x^\mu(\tau,l)$, and the integrand vanishes automatically. On the other hand, if we are dealing with open strings, we have the boundary term vanishing only if

$$\left(-S^{\alpha\beta}S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu}\dot{x}^\mu \delta x^\nu\Big|_{\sigma=l} = \left(-S^{\alpha\beta}S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu}\dot{x}^\mu \delta x^\nu\Big|_{\sigma=0} = 0, \tag{4.19}$$

and for arbitrary independent variations at the endpoints, we must have

$$\left(-S^{\alpha\beta}S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu}\dot{x}^\mu\Big|_{\sigma=0} = \left(-S^{\alpha\beta}S_{\alpha\beta}\right)^{-\frac{1}{2}} S_{\mu\nu}\dot{x}^\mu\Big|_{\sigma=l} = 0. \tag{4.20}$$

Hence,

$$\dot{x}_\lambda S^{\lambda\nu} S_{\mu\nu}\dot{x}^\mu\Big|_{\sigma=0} = \dot{x}_\lambda S^{\lambda\nu} S_{\mu\nu}\dot{x}^\mu\Big|_{\sigma=l} = 0 \tag{4.21}$$

or, explicitly,

$$\begin{aligned} &\left(\dot{x}\cdot\dot{x}x'^\nu - \dot{x}\cdot x'\dot{x}^\nu\right)\left(\dot{x}\cdot\dot{x}x'_\nu - x'\cdot\dot{x}\dot{x}_\nu\right) \\ &= \left(\left[\dot{x}\cdot\dot{x}\right]^2 x'\cdot x' + \dot{x}\cdot\dot{x}\left[x'\cdot\dot{x}\right]^2\right)\Big|_{\sigma=0 \text{ and } \sigma=l} = 0. \end{aligned} \tag{4.22}$$

One of the following two conditions must be true

$$\left(\dot{x}\cdot\dot{x}\right)\left(x'\cdot x'\right) = -\left[x'\cdot\dot{x}\right]^2\Big|_{\sigma=0 \text{ and } \sigma=l} \tag{4.23}$$

or

$$\dot{x}\cdot\dot{x}\Big|_{\sigma=0 \text{ and } \sigma=l} = 0. \tag{4.24}$$

The first condition is never true, so we must have $\dot{x}\cdot\dot{x} = 0$ at the endpoints which means that the endpoints of the string are tracing out null geodesics. i.e., They are moving at the speed of light.

In either the closed or the open string case, equation (4.17) gives the equations of motion

$$2\mu^2 D^\mu \left[\left(-S^{\alpha\beta} S_{\alpha\beta} \right)^{-\frac{1}{2}} S_{\mu\nu} \right] = 0. \tag{4.25}$$

Notice that we did not require the variation on the spatial boundary for open strings to vanish as we did for the temporal boundary. We could have required the spatial variation on the boundary to vanish by assuming the string endpoints are fixed. Let us return briefly to the condition for a stationary action

$$\int_{\sigma=0}^{\sigma=l} d\sigma\, D^\mu \left[\left(-S^{\alpha\beta} S_{\alpha\beta} \right)^{-\frac{1}{2}} S_{\mu\nu} \right] \delta x^\nu = \left[\left(-S^{\alpha\beta} S_{\alpha\beta} \right)^{-\frac{1}{2}} S_{\mu\nu} \dot{x}^\mu \delta x^\nu \right]_{\sigma=0}^{\sigma=l}. \tag{4.26}$$

Now suppose $\delta x^a = 0$ at $\sigma = 0$ for a $(n-p)$-dimensional subset of the coordinates and $\delta x^a = 0$ at $\sigma = l$ for a $(n-q)$-dimensional subset where n is the dimension of the spacetime. Then, if $\dot{x}^\mu = 0$ at $\sigma = 0$ for a p-dimensional subset, U_p, of the coordinates and $\dot{x}^\mu = 0$ at $\sigma = l$ for a q-dimensional subset, U_q, of the coordinates, we would then still have the equations of motion (4.25) for the string coordinates, but the position, $x^\mu(\tau, 0)$ and $x^\mu(\tau, l)$, of the endpoints would be fixed on a p-dimensional and a q-dimensional subspace of spacetime, respectively. We say that the string has one endpoint on a p-brane and the other endpoint on a q-brane. Further discussion of D-branes is beyond the scope of this dissertation. We refer the interested reader to a sample of the vast literature on the subject [39, 40, 41, 42, 43] for more details.

4.2.2. Interacting point particles and strings

We would now like to extend our discussion from the free point particle and the free string to the corresponding interacting theories. For point particles, we postulate the following general form of the multiparticle action:

$$S_{\text{particle}} = \sum_a m_a \int ds_a + \sum_{a<b} \int d\tau_a \, d\tau_b \, \mathcal{L}_{ab} \left(x_a, x_b, \dot{x}_a, \dot{x}_b \right), \tag{4.27}$$

where we have now added indices a, b for particle labels, repeated Latin indices are not summed over, and $\mathcal{L}_{ab}$ denotes our interaction lagrangian. We have limited the dependence of the interaction lagrangian to first derivatives, and we require that it satisfy the symmetry $\mathcal{L}_{ab} = \mathcal{L}_{ba}$. The variation of this action gives

$$m_a \frac{d}{d\tau_a} \frac{\dot{x}_a^\mu}{\sqrt{\dot{x}_a \cdot \dot{x}_a}} = \sum_{a \neq b} \int d\tau_b \left(\frac{\partial \mathcal{L}_{ab}}{\partial x_{a\mu}} - \frac{d}{d\tau_a} \frac{\partial \mathcal{L}_{ab}}{\partial \dot{x}_{a\mu}} \right), \quad a = 1, 2, ... \tag{4.28}$$

which is

$$m_a \left(\dot{x} \cdot \dot{x} \right)^{-\frac{1}{2}} \left[\frac{d\dot{x}_a^\mu}{d\tau_a} - \frac{\dot{x}_a^\mu \dot{x}_{a\alpha}}{\left(\dot{x} \cdot \dot{x} \right)} \frac{d\dot{x}_a^\alpha}{d\tau_a} \right] = \sum_{a \neq b} \int d\tau_b \left(\frac{\partial \mathcal{L}_{ab}}{\partial x_{a\mu}} - \frac{d}{d\tau_a} \frac{\partial \mathcal{L}_{ab}}{\partial \dot{x}_{a\mu}} \right). \tag{4.29}$$

Contracting both sides of the equation with $\dot{x}_{a\mu}$ gives

$$\begin{aligned} 0 &= \dot{x}_{a\mu} \sum_{a \neq b} \int d\tau_b \left(\frac{\partial \mathcal{L}_{ab}}{\partial x_{a\mu}} - \frac{d}{d\tau_a} \frac{\partial \mathcal{L}_{ab}}{\partial \dot{x}_{a\mu}} \right) \\ &= \frac{d}{d\tau_a} \left[\left(1 - \dot{x}_a^\mu \frac{\partial}{\partial \dot{x}_a^\mu} \right) \sum_{a \neq b} \int d\tau_b \, \mathcal{L}_{ab} \right] \end{aligned} \tag{4.30}$$

which can be interpreted as a constraint on the possible lagrangians, $\mathcal{L}_{ab}$, which give consistent equations of motion. The constraint also arises from the fact that the parametrization of the particle world line is arbitrary. It is like having a gauge invariance. Requiring the action to be stationary under re-parametrization of the world line, $\tau_a \to \tau_a + \delta\tau_a$, we get

$$\begin{aligned} 0 = \delta S = &\sum_a m_a \int d\tau_a \left(\dot{x}_a \cdot \dot{x}_a\right)^{-\frac{1}{2}} \dot{x}_a \cdot \frac{d\dot{x}_a}{d\tau_a}\delta\tau_a \\ &+ \sum_{b<a} \int d\tau_a \int d\tau_b \left[\frac{\partial \mathcal{L}_{ab}}{\partial x_a^\mu}\dot{x}^\mu \delta\tau_a + \frac{\partial \mathcal{L}_{ab}}{\partial \dot{x}_a^\mu}\frac{d\dot{x}_a^\mu}{d\tau_a}\delta\tau_a\right]. \end{aligned} \tag{4.31}$$

Integrating each term by parts gives

$$m_a \frac{d}{d\tau}\left[\left(\dot{x}_a \cdot \dot{x}_a\right)^{-\frac{1}{2}} \dot{x}_{a\mu}\right]\dot{x}_a^\mu = \sum_{b\neq a} \int d\tau_b \left[\dot{x}_a^\mu \frac{\partial \mathcal{L}_{ab}}{\partial x_a^\mu} - \dot{x}_a^\mu \frac{d}{d\tau_a}\frac{\partial \mathcal{L}_{ab}}{\partial \dot{x}_a^\mu}\right]. \tag{4.32}$$

The left-hand side is zero, so we are left with exactly equation (4.30). We see that equation (4.30) indeed arises from the re-parametrization invariance. If we choose only from the $\mathcal{L}_{ab}$ which satisfy this condition, then our action will be re-parametrization invariant (as is physically necessary since our choice of parametrization was arbitrary). We can, therefore, choose whatever parametrization is convenient for our purposes without affecting the physics.

For the interacting string, the equation analogous to (4.27) is

$$S_{\text{string}} = \sum_a S_{\text{free string}} + \sum_{a<b} \int d\tau_a\, d\sigma_a \int d\tau_b\, d\sigma_b\, \mathcal{L}_{ab}\left(x_a, x_b, \dot{x}_a, x_a', \dot{x}_b, x_b'\right), \tag{4.33}$$

depending on up to first-order derivatives and with $\mathcal{L}_{ab} = \mathcal{L}_{ba}$. Requiring the variation of this action to vanish gives the equation of motion

$$
\begin{aligned}
0 = \delta S = & \, 2\mu_a^2 \left[\left(-S_a^{\alpha\beta} S_{a\alpha\beta} \right)^{-\frac{1}{2}} S_{a\mu\nu} \dot{x}_a^\mu \delta x_a^\nu \right]_{\sigma=0}^{\sigma=l} \\
& - 2\mu_a^2 \int d\sigma_a \, D_a^\mu \left[\left(-S_a^{\alpha\beta} S_{a\alpha\beta} \right)^{-\frac{1}{2}} S_{a\mu\nu} \right] \delta x_a^\nu \\
& + \sum_{b \neq a} \int d\sigma_a \int d\tau_b \, d\sigma_b \left[\frac{\partial \mathcal{L}_{ab}}{\partial x_a^\mu} \delta x_a^\mu - \frac{d}{d\tau_a} \left(\frac{\partial \mathcal{L}_{ab}}{\partial \dot{x}_a^\mu} \right) \delta x_a^\mu - \frac{d}{d\sigma_a} \left(\frac{\partial \mathcal{L}_{ab}}{\partial x_a'^\mu} \right) \delta x_a^\mu \right] \\
& + \sum_{b \neq a} \int d\tau_b \, d\sigma_b \left(\frac{\partial \mathcal{L}_{ab}}{\partial x_a'^\mu} \delta x_a^\mu \right)_{\sigma=0}^{\sigma=l} ,
\end{aligned}
\tag{4.34}
$$

which reduces to

$$
2\mu_a^2 D_a^\mu \left[\frac{S_{a\mu\nu}}{\sqrt{-S_a^{\alpha\beta} S_{a\alpha\beta}}} \right] = \sum_{b \neq a} \int d\tau_b \, d\sigma_b \left[\frac{\partial \mathcal{L}_{ab}}{\partial x_a^\nu} - \frac{d}{d\tau_a} \left(\frac{\partial \mathcal{L}_{ab}}{\partial \dot{x}_a^\nu} \right) - \frac{d}{d\sigma_a} \left(\frac{\partial \mathcal{L}_{ab}}{\partial x_a'^\nu} \right) \right] . \tag{4.35}
$$

At the endpoints, we have

$$
2\mu_a^2 \left(-S_a^{\alpha\beta} S_{a\alpha\beta} \right)^{-\frac{1}{2}} S_{a\mu\nu} \dot{x}_a^\nu = \sum_{b \neq a} \int d\tau_b \, d\sigma_b \, \frac{\partial \mathcal{L}_{ab}}{\partial x_a'^\mu} . \tag{4.36}
$$

In this case, the endpoints of the string have interactions and no longer move at the speed of light as they did in the case of the free string. There are also cases of equation (4.34) where a subset of the coordinates are fixed, thus the endpoints are confined to branes as in the free case. These cases give the conditions

$$
\sum_{b \neq a} \int d\tau_b \, d\sigma_b \, \frac{\partial \mathcal{L}_{ab}}{\partial x_a'^\lambda} = 0 \quad \text{and} \quad \dot{x}_a^\lambda = 0 \tag{4.37}
$$

on the branes.

We now multiply equation (4.35) by $\dot{x}^{\nu}_{a}$ and x'^{ν}_{a}, generalizing what we did in the case of the point particle in equation (4.30). The left-hand side is (dropping subscript a for notational simplicity)

$$
\begin{aligned}
&2\mu^2\left(\dot{x}^{\mu}\frac{\partial}{\partial\sigma}-x'^{\mu}\frac{\partial}{\partial\tau}\right)\left[S_{\mu\nu}\left(-S\cdot S\right)^{-\frac{1}{2}}\right]\\
&=\frac{2\mu^2}{\left(-S\cdot S\right)^{-\frac{1}{2}}}\left[\dot{x}^{\mu}S'_{\mu\nu}-\frac{\dot{x}^{\mu}S_{\mu\nu}\left(S\cdot S'\right)}{\left(S\cdot S\right)}-x'^{\mu}\dot{S}_{\mu\nu}+\frac{x'^{\mu}S_{\mu\nu}\left(S\cdot\dot{S}\right)}{\left(S\cdot S\right)}\right].
\end{aligned}
\tag{4.38}
$$

Now

$$
\begin{aligned}
&S'_{\mu\nu}=-S'_{\nu\mu},\quad \dot{S}_{\mu\nu}=-\dot{S}_{\nu\mu}\\
&S\cdot S=2\dot{x}^2x'^2-2\left(\dot{x}\cdot x'\right)^2=2x'^{\mu}S_{\mu\nu}\dot{x}^{\nu}.
\end{aligned}
\tag{4.39}
$$

Dotting $\dot{x}^{\nu}$ into the expression for the left-hand side, equation (4.38) becomes

$$
\begin{aligned}
&\frac{2\mu^2}{\left(-S\cdot S\right)^{-\frac{1}{2}}}\left[0-0-x'^{\mu}\dot{S}_{\mu\nu}\dot{x}^{\nu}+\frac{x'^{\mu}S_{\mu\nu}\dot{x}^{\nu}\left(S\cdot\dot{S}\right)}{\left(S\cdot S\right)}\right]\\
&=\frac{2\mu^2}{\left(-S\cdot S\right)^{-\frac{1}{2}}}\left[-x'^{\mu}\dot{S}_{\mu\nu}\dot{x}^{\nu}+\frac{1}{2}S\cdot\dot{S}\right]=0
\end{aligned}
\tag{4.40}
$$

since the terms in the brackets exactly cancel. Similarly, dotting x'^{ν} into expression (4.38) also gives zero.

Multiplying the right-hand side of equation (4.35) by $\dot{x}^{\nu}_{a}$ gives the constraint

$$
\left[\frac{\partial}{\partial\tau_a}\left(1-\dot{x}^{\nu}_{a}\frac{\partial}{\partial\dot{x}^{\nu}_{a}}\right)-\frac{\partial}{\partial\sigma_a}\dot{x}^{\nu}_{a}\frac{\partial}{\partial x'^{\nu}_{a}}\right]\sum_{b\neq a}\int d\tau_b\, d\sigma_b\,\mathcal{L}_{ab}=0.
\tag{4.41}
$$

In a similar fashion, when we multiply by $x_a'^{\nu}$, we get

$$\left[\frac{\partial}{\partial\sigma_a}\left(1-\dot{x}_a'^{\nu}\frac{\partial}{\partial x_a'^{\nu}}\right)-\frac{\partial}{\partial\tau_a}x_a'^{\nu}\frac{\partial}{\partial\dot{x}_a^{\nu}}\right]\sum_{b\neq a}\int d\tau_b\,d\sigma_b\,\mathcal{L}_{ab}=0. \tag{4.42}$$

Performing the same operation on the boundary conditions for the open string (4.36) yields

$$\begin{aligned} 0 &= \dot{x}_a^{\mu}\frac{\partial}{\partial x'^{\mu}}\sum_{b\neq a}\int d\tau_b\,d\sigma_b\,\mathcal{L}_{ab} \quad \text{at} \quad \sigma_a=0,l. \\ \mu_a^2\left(-S_a\cdot S_a\right)^{\frac{1}{2}} &= x_a'^{\mu}\frac{\partial}{\partial x'^{\mu}}\sum_{b\neq a}\int d\tau_b\,d\sigma_b\,\mathcal{L}_{ab} \quad \text{at} \quad \sigma_a=0,l. \end{aligned} \tag{4.43}$$

As with the point particle case, constraints (4.41)-(4.43) can be obtained by insisting that the action be invariant under re-parametrizations $\tau_a \to \tau_a+\delta\tau_a$ and $\sigma_a \to \sigma_a+\delta\sigma_a$. If we use a lagrangian which satisfies these conditions, we are then free to choose whatever parameters, τ_a and σ_a, are most convenient.

4.2.3. The antisymmetric tensor field

We will now look at specific forms of the interaction lagrangian for point particles and strings, and show how the antisymmetric tensor arises quite naturally. First, we couple the world lines of two point particles by writing

$$S_{\text{int}}=\int d\tau_a d\tau_b\mathcal{L}_{ab}=\sum_{a<b}\int e_a e_b\, g_{\mu\nu}dx_a^{\mu}dx_b^{\nu}\,\delta\left((x_a-x_b)^2\right), \tag{4.44}$$

where e_a and e_b are coupling constants (which, for concreteness, may be thought of as the electric charges of particles a and b, respectively). We now write

$$g_{\mu\nu}\, dx_a^\mu dx_b^\nu = g_{\mu\nu}\, \frac{\partial x_a^\mu}{\partial \tau_a}\, d\tau_a \frac{\partial x_b^\nu}{\partial \tau_b}\, d\tau_b = \dot{x}_a \cdot \dot{x}_b\, d\tau_a d\tau_b, \tag{4.45}$$

from which we identify

$$\mathcal{L}_{ab} = e_a e_b\, \dot{x}_a \cdot \dot{x}_b\, \delta\left((x_a - x_b)^2\right). \tag{4.46}$$

Setting the magnitude of four velocities $\dot{x}_a \cdot \dot{x}_a = 1$ and using equation (4.28), we find the equations of motion

$$\begin{aligned}
m_a \ddot{x}_a^\mu &= \sum_{b\neq a} \int d\tau_b \left(\frac{\partial \mathcal{L}_{ab}}{\partial x_{a\mu}} - \frac{d}{d\tau_a} \frac{\partial \mathcal{L}_{ab}}{\partial \dot{x}_{a\mu}} \right) \\
&= \sum_{b\neq a} e_a e_b \int d\tau_b \left\{ \dot{x}_a \cdot \dot{x}_b \frac{\partial}{\partial x_{a\mu}} \delta\left((x_a - x_b)^2\right) - \dot{x}_b^\mu \dot{x}_a^\alpha \frac{d}{dx_a^\alpha} \delta\left((x_a - x_b)^2\right) \right\} \\
&= \sum_{b\neq a} e_a \dot{x}_a^\alpha \left\{ \frac{\partial}{\partial x_{a\mu}} \left(e_b \int d\tau_b \dot{x}_{b\alpha} \delta\left((x_a - x_b)^2\right) \right) - \frac{d}{dx_a^\alpha} \left(\int d\tau_b \dot{x}_b^\mu \delta\left((x_a - x_b)^2\right) \right) \right\} \\
&= \sum_{b\neq a} e_a \dot{x}_{a\nu} \left(\partial_a^\mu A_b^\nu - \partial_a^\nu A_b^\mu \right) \\
&= \sum_{b\neq a} e_a F_b^{\mu\nu}(x_a) \dot{x}_{a\nu},
\end{aligned} \tag{4.47}$$

where we have defined the field strength, $F_b^{\mu\nu}$, in terms of the potential, A_b^α, given by

$$A_b^\alpha = e_b \int d\tau_b\, \dot{x}_b^\alpha\, \delta\left((x_a - x_b)^2\right). \tag{4.48}$$

It is easy to show that the vector potential, A_b^μ, satisfies

$$\Box A_b^\mu(x) = -4\pi e_b \int dx_b^\mu \delta^{(4)}\left(x - x_b\right) = -4\pi j_b^\mu(x), \tag{4.49}$$

where j_b^μ is the current generated by the charge carried by particle b. Also,

$$\partial_\mu A_b^\mu(x) = 0. \tag{4.50}$$

A_b^μ can indeed be viewed as the electromagnetic vector potential.

Notice that we can now write the interaction (4.44) as

$$S_{\text{int}} = e_a \int A_{b\mu}(x_a) dx_a^\mu = e_a \int A_b \tag{4.51}$$

which is the coupling of a 1-form A_b to the particle world line.

For the case of strings, we would like to generalize this interaction in the most natural way by coupling string world sheets in much the same way that we coupled particle world lines. Equation (4.33) is then

$$S_{\text{string}} = -\sum_a \mu_a^2 \int \sqrt{-dS^{\mu\nu} dS_{\mu\nu}} + \sum_{a<b} g_a g_b \int \left[dx_a^\mu \wedge dx_a^\nu\right]\left[dx_{b\mu} \wedge dx_{b\nu}\right] G\left((x_a - x_b)^2\right), \tag{4.52}$$

where we have contracted the world sheet area elements of the two strings in a fashion analogous to the coupling of world lines in equation (4.44) for the point particle case. We have defined coupling constants, g_a and g_b, and Green's function, $G\left((x_a - x_b)^2\right)$.

We now use the decomposition derived in equation (4.4) to write the string action as

$$S_{\text{string}} = -\sum_a \mu_a^2 \int \sqrt{-dS^{\mu\nu}dS_{\mu\nu}} + \sum_{a<b} g_a g_b \int d\tau_a d\sigma_a d\tau_b d\sigma_b \, S_a^{\mu\nu} S_{b\mu\nu} \, G\left((x_a - x_b)^2\right). \tag{4.53}$$

Define

$$\begin{aligned} B_b^{\mu\nu} &= g_b \int d\tau_b d\sigma_b \, S_b^{\mu\nu} \, G\left((x_a - x_b)^2\right) \\ &= g_b \int dx_b^\mu \wedge dx_b^\nu \, G\left((x_a - x_b)^2\right) \end{aligned} \tag{4.54}$$

so that

$$S_{\text{int}} = \sum_{a<b} g_a \int B_{b\mu\nu} dx_a^\mu \wedge dx_a^\nu = \sum_{a<b} g_a \int B_b. \tag{4.55}$$

We see that the field $B_b^{\mu\nu}$, as defined in equation (4.54), will be antisymmetric due to the antisymmetry of $S_b^{\mu\nu}$. We now have the coupling of a 2-form B_b to the world sheet of string a.

We now have an antisymmetric tensor field arising in a natural way by generalizing the "electrodynamical" coupling of a 1-form to the world line in the point particle case to a 2-form coupling to the world sheet in the case of strings. We now find the field strength associated with this field from the equations of motion (4.35) as follows:

$$2\mu_a^2 D_{a\nu} \left[\frac{S_a^{\mu\nu}}{(-S_a \cdot S_a)^{\frac{1}{2}}} \right] = g_a \sum_{b \neq a} H_b^{\mu\rho\lambda}(x_a) S_{a\rho\lambda}, \tag{4.56}$$

where we have written the field strength as $H_b^{\mu\rho\lambda}(x_a)$. Our string interaction lagrangian is

$$\mathcal{L}_{ab} = g_a g_b \, S_a^{\mu\nu} S_{b\mu\nu} \, G\left((x_a - x_b)^2\right) \tag{4.57}$$

from which we find

$$\begin{aligned}
g_a \sum_{b\neq a} H_b^{\mu\alpha\beta}(x_a) S_{a\alpha\beta} &= \sum_{b\neq a} \int d\tau_b \, d\sigma_b \left[\frac{\partial \mathcal{L}_{ab}}{\partial x_a^\nu} - \frac{d}{d\tau_a}\left(\frac{\partial \mathcal{L}_{ab}}{\partial \dot{x}_a^\nu}\right) - \frac{d}{d\sigma_a}\left(\frac{\partial \mathcal{L}_{ab}}{\partial x_a'^\nu}\right)\right] \\
&= \sum_{b\neq a} \int d\tau_b \, d\sigma_b \, g_a g_b \Big[S_a^{\alpha\beta} S_{b\alpha\beta} \frac{\partial G}{\partial x_a^\mu} + 2 S_{b\beta\mu} \left(\dot{x}_a'^\beta + \dot{x}_a^\rho x_a'^\beta \frac{\partial}{\partial x_a^\rho}\right) G \\
&\qquad + 2 S_{b\mu\alpha} \left(\dot{x}_a'^\alpha + \dot{x}_a^\alpha x_a'^\rho \frac{\partial}{\partial x_a^\rho}\right) G \Big] \\
&= \sum_{b\neq a} \int d\tau_b \, d\sigma_b \, g_a g_b \left[S_a^{\alpha\beta} S_{b\alpha\beta} \frac{\partial G}{\partial x_a^\mu} + 2 S_{b\beta\mu} S_a^{\rho\beta} \frac{\partial G}{\partial x_a^\rho}\right].
\end{aligned} \tag{4.58}$$

The antisymmetry of $S_a^{\rho\beta}$ reduces this to

$$\begin{aligned}
g_a \sum_{b\neq a} H_b^{\mu\alpha\beta}(x_a) S_{a\alpha\beta} &= \sum_{b\neq a} \int d\tau_b \, d\sigma_b \, g_a g_b S_a^{\alpha\beta} \left[S_{b\alpha 3} \frac{\partial G}{\partial x_a^\mu} S_{b\beta\mu} \frac{\partial G}{\partial x_a^\alpha} S_{b\mu\alpha} \frac{\partial G}{\partial x_a^\beta}\right] \\
&= \sum_{b\neq a} g_a S_a^{\alpha\beta} \Big[\frac{\partial}{\partial x_a^\mu} \left(\int d\tau_b \, d\sigma_b \, g_b S_{b\alpha\beta} G\right) \\
&\qquad + \frac{\partial}{\partial x_a^\alpha} \left(\int d\tau_b \, d\sigma_b \, g_b S_{b\beta\mu} G\right) \\
&\qquad + \frac{\partial}{\partial x_a^\beta} \left(\int d\tau_b \, d\sigma_b \, g_b S_{b\mu\alpha} G\right)\Big] \\
&= g_a \sum_{b\neq a} S_{a\alpha\beta} \left[\partial_a^\alpha B_b^{\beta\mu} + \partial_a^\beta B_b^{\mu\alpha} + \partial_a^\mu B_b^{\alpha\beta}\right],
\end{aligned} \tag{4.59}$$

where we have used definition (4.54). We can now extract the field strength,

$$H_b^{\mu\alpha\beta}(x_a) = \partial_a^\alpha B_b^{\beta\mu} + \partial_a^\beta B_b^{\mu\alpha} + \partial_a^\mu B_b^{\alpha 3}, \tag{4.60}$$

which may be compared with equation (2.91) for the torsion field.

4.2.4. Properties of the antisymmetric tensor field

Now that we have shown how the antisymmetric tensor field arises from interacting string world sheets, we would like to derive some more of its properties. We first note that

$$\begin{aligned}\frac{\partial}{\partial x^\alpha} B_b^{\alpha\beta}(x) &= g_b \frac{\partial}{\partial x^\alpha} \int d\tau_b \, d\sigma_b \, S_b^{\alpha\beta} G\left((x-x_b)^2\right) \\ &= -g_b \int d\tau_b \, d\sigma_b \, S_b^{\alpha\beta} \frac{\partial}{\partial x_b^\alpha} G\left((x-x_b)^2\right).\end{aligned} \tag{4.61}$$

Equation (4.8) reduces equation (4.61) to

$$\begin{aligned}\frac{\partial}{\partial x^\alpha} B_b^{\alpha\beta}(x) &= g_b \int d\tau_b \, d\sigma_b \, D_b^\beta G\left((x-x_b)^2\right) \\ &= g_b \int d\tau_b \, d\sigma_b \left[\frac{\partial}{\partial \sigma_b}\left(\dot{x}_b G\right) - \dot{x}_b' G + \dot{x}_b' G - \frac{\partial}{\partial \tau_b}\left(x_b' G\right)\right] \\ &= g_b \int d\tau_b \, \dot{x}_b G\Big|_{\sigma=0}^{\sigma=l} \qquad (=0 \quad \text{for closed strings}).\end{aligned} \tag{4.62}$$

This property of closed strings is analogous to the Lorentz gauge condition for the photon field in electrodynamics.

From the expression for the field strength (4.60), we see that $H^{\mu\alpha\beta}$ is invariant under the transformation

$$B^{\alpha\beta} \to B^{\alpha\beta} + \partial^\alpha \Lambda^\beta - \partial^\beta \Lambda^\alpha, \tag{4.63}$$

where, by equation (4.62), the Λ's must, for closed strings, satisfy

$$\partial_\mu \left(\partial^\nu \Lambda^\mu - \partial^\mu \Lambda^\nu\right) = 0. \tag{4.64}$$

In differential form notation, the field strength, $H = dB$, is invariant under $B \to B + d\Lambda$ since $d^2 = 0$. The lagrangian density for $B^{\alpha\beta}$ in the case of closed strings is then

$$\mathcal{L} = -\frac{1}{12} H_{\alpha\beta\lambda} H^{\alpha\beta\lambda} - \frac{1}{4\alpha} \left(\partial_\alpha B^{\alpha\beta} \partial^\lambda B_{\lambda\beta} + \partial_\alpha B^{\beta\alpha} \partial^\lambda B_{\beta\lambda} \right), \tag{4.65}$$

where the gauge parameter, α, accounts for our choice of Λ in equation (4.63). Choosing $\alpha = 1$ leads to the equations of motion for $B^{\alpha\beta}$

$$\Box B^{\alpha\beta} = 0. \tag{4.66}$$

In the case of open strings, we have differences due to the non-zero boundary term in equation (4.62). The reader interested in this case should refer to [33].

We concentrate on the closed string sector since closed strings give rise to *massless* antisymmetric tensor fields. In string theory, the closed strings give rise to gravitation, and it is, therefore, appropriate to use them if we want to identify the Kalb-Ramond antisymmetric tensor field from string theory with the torsion field from gravity. In this way, we hypothesize that the formulation of gravity in interacting closed string theory contains non-zero torsion in a natural way through the interaction of string world sheets. It would be interesting to derive the fermion-antisymmetric tensor interaction given in equation (3.70) using the formalism for supersymmetric strings given in Appendix C. In the following chapter, we will look at some of the physical predictions given by the interactions of the antisymmetric tensor field with the fermions of the standard model.

CHAPTER 5

THE ANTISYMMETRIC TENSOR INTERACTION

There are many possible experimental signatures [44] of non-zero torsion. For example, torsion has been shown to give gravitational parity violating interactions [27, 45, 46]. In theories with large extra dimensions [47, 48], the parity violations which occur in four dimensions disappear on the physical 3-brane when the torsion originates in the bulk [49]. There are also many effects that arise from the torsion interactions between particles of spin [16]. It has even been proposed that torsion in large extra dimensions is responsible for neutrino mass [50]. In this chapter, we derive the propagator and vertex rules for the fermion-antisymmetric tensor interaction [51] and then proceed to calculate several interesting diagrams resulting from this interaction.

5.1. Feynman rules for the Kalb-Ramond field

We begin with the following lagrangian which was derived as equation (3.70) in Chapter 3 (also see equation (4.65) in Chapter 4):

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{12}H_{\mu\nu\lambda}H^{\mu\nu\lambda} + \overline{\psi}\left(i\,\partial\!\!\!/ - e\,A\!\!\!/ - \frac{g}{M}\sigma_{\mu\nu\lambda}H^{\mu\nu\lambda} - m\right)\psi. \tag{5.1}$$

Notice that we have added a coupling constant to be consistent with the literature [13] and to simplify the calculation. The factor of M in the denominator represents a mass scale to make the coupling constant dimensionless. We assume that the antisymmetric torsion, or Kalb-Ramond antisymmetric tensor field $H^{\mu\nu\lambda}$, can be derived from a potential $B_{\mu\nu}$ as follows (i.e., $H = dB$ is exact):

$$H_{\mu\nu\lambda} = \partial_\mu B_{\nu\lambda} + \partial_\nu B_{\lambda\mu} + \partial_\lambda B_{\mu\nu} \tag{5.2}$$

and

$$\sigma_{\alpha\beta\gamma} = i\epsilon_{\alpha\beta\gamma\mu}\gamma_5\gamma^\mu . \tag{5.3}$$

We would like to calculate some amplitudes for physical processes from this interaction lagrangian.[1]

The first step is to find the propagator for the antisymmetric tensor particle. Doing so requires adding a gauge fixing term to the free lagrangian density,

$$\begin{aligned}\mathcal{L}_0 + \mathcal{L}_{\text{gf}} = & -\frac{1}{12}\left(\partial_\mu B_{\nu\lambda} + \partial_\nu B_{\lambda\mu} + \partial_\lambda B_{\mu\nu}\right)\left(\partial^\mu B^{\nu\lambda} + \partial^\nu B^{\lambda\mu} + \partial^\lambda B^{\mu\nu}\right) \\ & - \frac{1}{4\alpha}\left(\partial_\mu B^{\mu\nu}\partial^\sigma B_{\sigma\nu} + \partial_\mu B^{\nu\mu}\partial^\sigma B_{\nu\sigma}\right),\end{aligned} \tag{5.4}$$

where α is a gauge fixing parameter. The first term expands into nine terms which are seen to reduce to three terms by relabelling dummy indices. The result is

$$\begin{aligned}\mathcal{L}_0 + \mathcal{L}_{\text{gf}} = -\frac{1}{4}\Big[& \partial_\mu B_{\nu\lambda}\partial^\mu B^{\nu\lambda} + \partial_\mu B_{\nu\lambda}\partial^\nu B^{\lambda\mu} + \partial_\mu B_{\nu\lambda}\partial^\lambda B^{\mu\nu} \\ & + \frac{1}{\alpha}\left(\partial_\mu B^{\mu\nu}\partial^\sigma B_{\sigma\nu} + \partial_\mu B^{\nu\mu}\partial^\sigma B_{\nu\sigma}\right)\Big].\end{aligned} \tag{5.5}$$

[1]Note that, in curved spacetime, we would need to include a factor of $\sqrt{-g}$ in the action which would change the expression for $\Box$ to $\Box = (-g)^{-\frac{1}{2}}\partial_\mu\left(\sqrt{-g}g^{\mu\nu}\partial_\nu\right)$.

Now, we integrate each term by parts, discarding the surface terms,

$$\mathcal{L}_0 + \mathcal{L}_{\text{gf}} = \frac{1}{4}\Big[B_{\nu\lambda}\Box B^{\nu\lambda} + B_{\nu\lambda}\partial_\mu\partial^\nu B^{\lambda\mu} + B_{\nu\lambda}\partial_\mu\partial^\lambda B^{\mu\nu} + \frac{1}{\alpha}\left(\partial_\mu B^{\mu\nu}\partial^\sigma B_{\sigma\nu} + \partial_\mu B^{\nu\mu}\partial^\sigma B_{\nu\sigma}\right)\Big], \tag{5.6}$$

which can be written as

$$\mathcal{L}_0 + \mathcal{L}_{\text{gf}} = \frac{1}{4}B^{\alpha\beta}\Big[g_{\alpha\nu}g_{\beta\lambda}\Box + g_{\beta\nu}\partial_\lambda\partial_\alpha + g_{\alpha\lambda}\partial_\nu\partial_\beta + \frac{1}{\alpha}\left(g_{\beta\lambda}\partial_\alpha\partial_\nu + g_{\alpha\nu}\partial_\beta\partial_\lambda\right)\Big]B^{\nu\lambda}. \tag{5.7}$$

We use the antisymmetry property, $B_{\mu\nu} = -B_{\nu\mu}$, to further reduce equation (5.7) to

$$\mathcal{L}_0 + \mathcal{L}_{\text{gf}} = \frac{1}{4}B^{\alpha\beta}\Big[g_{\alpha\nu}g_{\beta\lambda}\Box + \left(\frac{1}{\alpha} - 1\right)\left(g_{\beta\lambda}\partial_\alpha\partial_\nu + g_{\alpha\nu}\partial_\beta\partial_\lambda\right)\Big]B^{\nu\lambda}. \tag{5.8}$$

We substitute $i\partial_\mu \to k_\mu$ to arrive at the following result in momentum space

$$\mathcal{L}_0 + \mathcal{L}_{\text{gf}} = -\frac{1}{4}B^{\alpha\beta}\Big[g_{\alpha\nu}g_{\beta\lambda}k^2 + \left(\frac{1}{\alpha} - 1\right)\left(g_{\beta\lambda}k_\alpha k_\nu + g_{\alpha\nu}k_\beta k_\lambda\right)\Big]B^{\nu\lambda}. \tag{5.9}$$

The propagator is the inverse of the quadratic operator in the lagrangian:

$$\mathcal{L} = \frac{1}{2}B^{\alpha\beta}\Delta_{\alpha\beta\nu\lambda}B^{\nu\lambda}. \tag{5.10}$$

From equation (5.9) we can extract the form of the operator

$$\Delta_{\alpha\beta\nu\lambda} = -\frac{1}{2}\Big[g_{\alpha\nu}g_{\beta\lambda}k^2 + \left(\frac{1}{\alpha} - 1\right)\left(g_{\beta\lambda}k_\alpha k_\nu + g_{\alpha\nu}k_\beta k_\lambda\right)\Big]. \tag{5.11}$$

The propagator, being the inverse of this operator, is defined by

$$\Delta_{\alpha\beta\nu\lambda} G^{\nu a\lambda b} = \frac{1}{2}\left(\delta^{a}_{\alpha}\delta^{b}_{\beta} - \delta^{b}_{\alpha}\delta^{a}_{\beta}\right). \tag{5.12}$$

Note that we make no distinction between Latin and Greek indices in this section since all of the calculations are performed in a flat background spacetime. We make an *Ansatz* for $G^{\nu a\lambda b}$ of the form

$$G^{\nu a\lambda b} = \frac{A}{k^2}\left[g^{\nu b}g^{\lambda a} + \frac{(1-\alpha)}{k^2}\left(g^{\nu a}k^{\lambda}k^{b} + g^{\lambda b}k^{\nu}k^{a}\right)\right] - (a \leftrightarrow b) \tag{5.13}$$

and use condition (5.12) to fix the constant A. Substituting equation (5.13) into equation (5.12) and using the fact that terms symmetric under $a \leftrightarrow b$ cancel, we find that $A = 1$.

Our gauge invariant propagator for the antisymmetric tensor field is, therefore,

$$\begin{aligned} G^{\nu\alpha\lambda\beta} = \frac{1}{k^2}\Big[& g^{\nu\beta}g^{\lambda\alpha} - g^{\nu\alpha}g^{\lambda\beta} \\ & + \frac{(1-\alpha)}{k^2}\left(g^{\nu\alpha}k^{\lambda}k^{\beta} + g^{\lambda\beta}k^{\nu}k^{\alpha} - g^{\nu\beta}k^{\lambda}k^{\alpha} - g^{\lambda\alpha}k^{\nu}k^{\beta}\right)\Big]. \end{aligned} \tag{5.14}$$

Breaking gauge invariance with the choice $\alpha = 1$ reduces equation (5.14) to

$$G^{\nu\alpha\lambda\beta} = \frac{1}{k^2}\left(g^{\nu\beta}g^{\lambda\alpha} - g^{\nu\alpha}g^{\lambda\beta}\right). \tag{5.15}$$

The second step is to derive the vertex rule for coupling the antisymmetric tensor field to a spin-$\frac{1}{2}$ Dirac field. We can extract this rule from the interaction lagrangian

$$\mathcal{L}_{\text{int}} = -\frac{g}{M}\overline{\psi}\,\sigma_{\mu\nu\lambda}H^{\mu\nu\lambda}\,\psi. \tag{5.16}$$

Substituting definition (5.2) into equation (5.16) this gives

$$\begin{aligned}\mathcal{L}_{\text{int}} &= -\frac{g}{M}\overline{\psi}\left[\left(i\epsilon_{\mu\nu\lambda\sigma}\gamma_5\gamma^\sigma\right)\left(\partial^\mu B^{\nu\lambda} + \partial^\nu B^{\lambda\mu} + \partial^\lambda B^{\mu\nu}\right)\right]\psi \\ &\equiv -\frac{g}{M}\overline{\psi}\,\Lambda^{\alpha\beta}\,\psi\,B_{\alpha\beta}.\end{aligned} \tag{5.17}$$

Our vertex rule in momentum space is then seen to be

$$-\frac{g}{M}\Lambda^{\alpha\beta} = -\frac{g}{M}\epsilon_{\mu\nu\lambda\sigma}\gamma_5\gamma^\sigma\left(g^{\nu\alpha}g^{\lambda\beta}k^\mu + g^{\lambda\alpha}g^{\mu\beta}k^\nu + g^{\mu\alpha}g^{\nu\beta}k^\lambda\right). \tag{5.18}$$

We summarize the Feynman rules in Figure 5.1.. The Feynman rules for the antisymmetric tensor field, along with those of electrodynamics, are all that is required to calculate the interactions among fermions, photons, and the Kalb-Ramond field.

5.2. Tree-level torsion exchange

The first scattering amplitude that we would like to examine is the amplitude for the tree level exchange of an antisymmetric tensor field as shown in Figure 5.2..

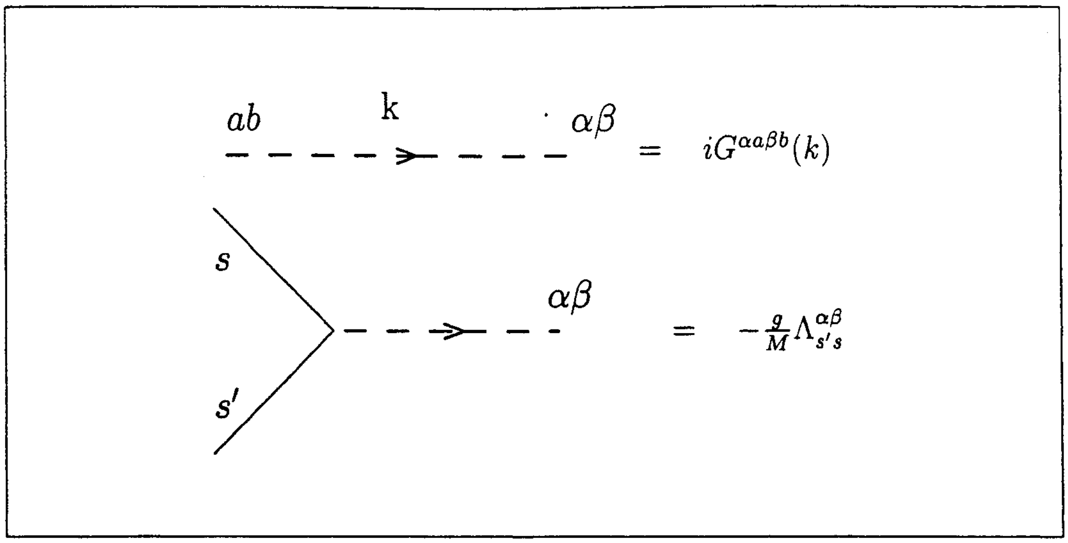


Figure 5.1. Feynman rules for fermion-antisymmetric tensor field interactions.

5.2.1. Scattering amplitude and cross section

Using the Feynman rules given in Figure 5.1., we can form the $\mathcal{M}$-matrix element as follows:

$$i\mathcal{M} = i\overline{U}^{s'}(p')\Lambda^{\alpha\beta}_{s's}U^{s}(p)G_{\alpha\mu\beta\nu}(q)\overline{U}^{r'}(k')\Lambda^{\mu\nu}_{r'r}(q)U^{r}(k) - \{k' \leftrightarrow p'\}\,, \tag{5.19}$$

where the relative minus sign between the two terms comes from the Fermi statistics of the final state. We define

$$\begin{aligned} -\frac{g}{M}\Lambda^{\alpha\beta}_{s's}(q) &= -\frac{g}{M}\epsilon_{abc\delta}\left(g^{b\alpha}g^{c\beta}q^{a} + g^{c\alpha}g^{a\beta}q^{b} + g^{a\alpha}g^{b\beta}q^{c}\right)\left(\gamma_5\gamma^{\delta}\right)_{s's} \\ &\equiv \Lambda^{\alpha\beta}_{\delta}\left(\gamma_5\gamma^{\delta}\right)_{s's} \end{aligned} \tag{5.20}$$

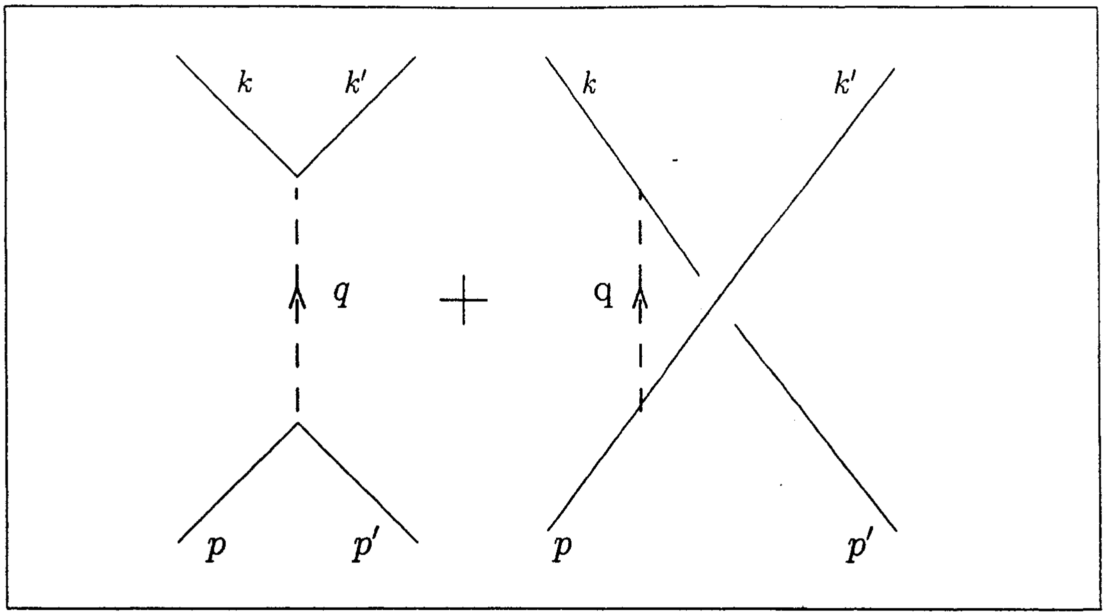


Figure 5.2. The Feynman diagram for the scattering of two identical fermions via the exchange of an antisymmetric tensor field.

so that we can work with the tensor part and the spinor part separately. Thus,

$$i\mathcal{M} = i\Lambda_{\delta}^{\alpha\beta}(q)G_{\alpha\mu\beta\nu}(q)\Lambda_{\lambda}^{\mu\nu}(q)\overline{U}^{s'}(p')\left(\gamma_5\gamma^{\delta}\right)U^{s}(p)\overline{U}^{r'}(k')\left(\gamma_5\gamma^{\lambda}\right)U^{r}(k) - \{k' \leftrightarrow p'\}. \tag{5.21}$$

Working out the pure tensor part gives

$$i\Lambda_{\delta}^{\alpha\beta}(q)G_{\alpha\mu\beta\nu}(q)\Lambda_{\lambda}^{\mu\nu}(q) = -i\frac{36g^2}{M^2}\left(\frac{q_\delta q_\lambda}{q^2} - g_{\delta\lambda}\right) \equiv iN_{\delta\lambda}. \tag{5.22}$$

Our $\mathcal{M}$-matrix is finally

$$i\mathcal{M} = iN_{\delta\lambda}(p'-p)\overline{U}^{s'}(p')\left(\gamma_5\gamma^{\delta}\right)_{s's}U^{s}(p)\overline{U}^{r'}(k')\left(\gamma_5\gamma^{\lambda}\right)_{r'r}U^{r}(k) - \{k' \leftrightarrow p'\}. \tag{5.23}$$

We now use expression (5.22) along with the Dirac equation to write our $\mathcal{M}$-matrix as

$$
\begin{aligned}
i\mathcal{M} = -\frac{i36g^2}{M^2}\Bigg\{ & \frac{4m^2}{(p'-p)^2}\overline{U}^{s'}(p')\left(\gamma_5\right)_{s's}U^s(p)\overline{U}^{r'}(k')\left(\gamma_5\right)_{r'r}U^r(k) \\
& - \overline{U}^{s'}(p')\left(\gamma_5\gamma^\lambda\right)_{s's}U^s(p)\overline{U}^{r'}(k')\left(\gamma_5\gamma_\lambda\right)_{r'r}U^r(k)\Bigg\} \\
+ \frac{i36g^2}{M^2}\Big\{ & p' \leftrightarrow k'\Big\}.
\end{aligned}
\tag{5.24}
$$

To form the unpolarized cross section, we need to square this amplitude, sum over the final state spins, and average over the initial state spins. For simplicity, we look at the limit of massless fermions which eliminates many terms, leaving

$$
\begin{aligned}
\frac{1}{4}\sum_{\text{spins}}|\mathcal{M}|^2 = \frac{4\cdot 9^2 g^4}{M^4}\Big\{ & \mathrm{Tr}\left[\not{k}\gamma_5\gamma_\delta\,\not{p}'\gamma_5\gamma_\lambda\right]\mathrm{Tr}\left[\not{p}\gamma_5\gamma^\delta\,\not{k}'\gamma_5\gamma^\lambda\right] \\
& + \mathrm{Tr}\left[\not{k}\gamma_5\gamma_\delta\,\not{k}'\gamma_5\gamma_\lambda\right]\mathrm{Tr}\left[\not{p}\gamma_5\gamma^\delta\,\not{p}'\gamma_5\gamma^\lambda\right] \\
& - \mathrm{Tr}\left[\not{k}\gamma_5\gamma_\delta\,\not{p}'\gamma_5\gamma_\lambda\,\not{p}\gamma_5\gamma^\delta\,\not{k}'\gamma_5\gamma^\lambda\right] \\
& - \mathrm{Tr}\left[\not{k}\gamma_5\gamma_\delta\,\not{k}'\gamma_5\gamma_\lambda\,\not{p}\gamma_5\gamma^\delta\,\not{p}'\gamma_5\gamma^\lambda\right]\Big\}.
\end{aligned}
\tag{5.25}
$$

Using the standard trace relations (A.17), we finally obtain

$$
\frac{1}{4}\sum_{\text{spins}}|\mathcal{M}|^2 = \frac{2\cdot 8^2\cdot 9^2 g^4}{M^4}\left\{4p\cdot kp'\cdot k' + p\cdot p'k\cdot k' + p\cdot k'k\cdot p'\right\}. \tag{5.26}
$$

We would now like to work in the center of momentum frame, with the incident fermions directed along the z-axis. We have the following expressions for kinematical quantities:

$$\begin{aligned} p &= (E, |p|\hat{z}) \quad p' = (E, -\vec{k}) \\ k &= (E, -|p|\hat{z}) \quad k' = (E, \vec{k}) \\ p \cdot k' &= E^2 - |p||\vec{k}| \cos\theta. \end{aligned} \tag{5.27}$$

The angle between the incoming and outgoing fermions is characterized by θ. In the limit of small fermion mass, relation (5.27) reduces to

$$\begin{aligned} p &= (E, E\hat{z}) \quad p' = (E, -E\hat{k}) \\ k &= (E, -E\hat{z}) \quad k' = (E, E\hat{k}) \\ p \cdot k' &= E^2(1 - \cos\theta). \end{aligned} \tag{5.28}$$

We have

$$\frac{1}{4} \sum_{\text{spins}} |\mathcal{M}|^2 = \frac{16^2 \cdot 9^2 g^4 E^4}{M^4} \left[9 + \cos^2\theta\right]. \tag{5.29}$$

Finally, since each incident particle contributes one-half of the center of mass energy, $E = E_{\text{cm}}/2$, our differential cross section is given by

$$\frac{d\sigma}{d\Omega} = \frac{1}{2E_{\text{cm}}^2} \frac{|\vec{k}|}{16\pi^2 E_{\text{cm}}} \frac{1}{4} \sum_{\text{spins}} |\mathcal{M}|^2 = \left(\frac{3g^2 E_{\text{cm}}}{2\pi M^2}\right)^2 \left[1 + \frac{1}{9}\cos^2\theta\right]. \tag{5.30}$$

Integrating over the solid angle gives the total cross section for this scattering process

$$\sigma = \left(\frac{g}{M}\right)^4 \frac{28 E_{\text{cm}}^2}{3\pi}. \tag{5.31}$$

5.3. Fermion anomalous magnetic moment

The measured magnetic moments of particles have provided a valuable test of QED. In the case of the magnetic moments of the electron and the muon, we have the important situation that both the experimental measurements and the standard model predictions are extremely precise. The experimental results for the tau lepton are much less precise [23]. The importance of the muon result over that of the electron stems from the fact that the larger rest mass of the muon makes it more sensitive to massive virtual particles and to any new physics that has not yet been included in the standard model. See [52] for a review of the electron and the muon anomalous magnetic moments.

It was recently found that the complete standard model prediction for the muon magnetic moment differs from the experimental value [53, 54]. This difference is outside the experimental and theoretical error bars.[2] The experimental value is $a = 11659203 \pm 15 \times 10^{-10}$, where $a = (g-2)/2$, if one assumes CPT invariance. The standard model prediction is $a = 11659176.7 \pm 6.7 \times 10^{-10}$.

There are many theories which exist as possible extensions of the standard model, for example, the minimal supersymmetric extension of the standard model (MSSM), superstring theories, quantum gravity theories, higher dimensional Kaluza-Klein theories, and theories based on heretofore undiscovered particles. These theories are very interesting in their own right as mathematical problems, but it is important to physi-

[2] The discrepancy originally stated in [53] and discussed in [55, 56] turned out to be incorrect due to a sign error found in the part of the standard model calculation dealing with the pion pole contribution. We refer the interested reader to [54, 57, 58, 59, 60] For a nice summary of the experimental and theoretical values, as well as history and prospects, see [55]. and references therein for the details.

cists trying to understand the nature of our universe that we be able to test them experimentally so that we can discard unviable candidates.

Testing theories beyond the standard model has been very difficult due to the remarkable accuracy of quantum electrodynamics and the difficulty of performing precise tests of the gauge theory of quarks which are confined inside hadrons. It is, therefore, fortunate that a measurement has been found which apparently cannot be explained using the standard model.

We will now give a brief pedagogical explanation of the experiment and elements involved in the standard model calculation. Brookhaven National Laboratory (BNL) as well as the european particle physics laboratory, CERN, have recently completed experiments with the positive muon. They have measured the anomalous magnetic moment to unprecedented accuracy. Both labs have independently confirmed a 1.6 standard deviation discrepancy with the standard model prediction. In the next section, we will briefly outline the techniques involved in measuring the magnetic moment at an accelerator laboratory. Our discussion is based on the review article [56].

The magnetic moment of a particle is related to its spin

$$\mu = g\left(\frac{e}{2mc}\right)\mathbf{S}. \tag{5.32}$$

For point-like particles, the gyromagnetic ratio, g, is equal to 2, whereas for composite particles, such as baryons, g may differ substantially from 2. For electrons, it was found that g is slightly larger than 2 even though they are assumed to be point-like in the standard model. The deviation is called the magnetic moment anomaly and arises through the exchange of virtual particles as allowed by the uncertainty principle. The magnetic moment anomaly, a, is defined through the relation

$$\mu = 2(1+a)\frac{e\hbar}{2m} \tag{5.33}$$

Since the 1940s, there have been increasingly refined experiments and more accurate calculations, making the anomalous magnetic moment of the electron one of the best understood phenomena in physics. The muon is much heavier than the electron, with $(m_\mu/m_e)^2 \simeq 40000$, making the muon much more susceptible to radiative corrections from heavier virtual particles and an excellent source of information about the standard model.

5.3.1. The $g-2$ experimental result

The goal of experiment E821 at BNL was to reduce the uncertainty in muon magnetic moment anomaly, $a(\mu)$, to 0.35 ppm, which would yield the direct observation of the weak corrections due to virtual $W^\pm$ and Z^0 bosons. In order to determine experimentally the anomalous magnetic moment of the muon, one measures the relative spin precession frequency, ω_a. This frequency is the rate at which the muon spin precesses relative to the orbital frequency in a magnetic field. The formula for ω_a is

$$\omega_a = -\frac{e}{m}\left[a\mathbf{B} - \left(a - \frac{1}{\gamma^2-1}\right)\beta \times \mathbf{E}\right], \tag{5.34}$$

where β is the velocity, v/c, and γ is the usual relativistic factor. If the muons in the experiment have a momentum of $p = 3.094$ GeV/c, then $\gamma \rightarrow \sqrt{1 + 1/a}$, and ω_a becomes nearly independent of the electric field. In this way, one needs only to have precise measurements of **B** and ω_a in order to accuratly determine $a(\mu)$.

The experimental setup is as follows: A synchrotron is used to accelerate protons to an energy of 24.3 GeV. The protons are then extracted and directed onto a target to create pions. The pions with momentum 1.7% above the magic value of 3.094 Gev/c are then transported through a beam line until they decay via the process $\pi^{\pm} \rightarrow \mu^{\pm} + \nu_{\mu}$. Since the pion has spin zero and the neutrino is left-handed, the muon must be right-handed. In other words, the muons so produced will have their spins aligned with their momenta. These polarized muons are then sent through dipoles and collimators to select those which have the magic value of momentum. The polarized muons at the correct energy are then placed into a storage ring with magnetic field, the **B** in equation (5.34), measured using NMR. The stored muons will eventually decay through the weak interaction with an intrinsic lifetime of about 2.2 μs, $\mu^{+} \rightarrow e^{+}\nu_{e}\overline{\nu}_{\mu}$. The angular distribution of decay electrons in the muon center of mass system is spin dependent,

$$\frac{dN}{dEd\Omega} = \mathcal{N}(E)\left[1 + \mathcal{A}(E)\cos\theta\right],$$

where N is the electron flux with energy in the range E to $E + dE$ through the solid angle, $d\Omega$. The angle θ is between the muon spin and the electron momentum. $\mathcal{N}(E)$ is the energy-dependent phase space factor. $\mathcal{A}(E)$ is the parity violating asymmetry. There is a correlation in the lab frame between the lab energy and the center of mass emission angle. The highest energy electrons are emitted in the direction of the muon momentum. The observed total electron flux thus depends on the orientation of the spin with respect to the muon momentum as well as energy and time. It is given by

$$N(E,t) = \mathcal{N}(E)e^{-t/\gamma\tau}\left[1 + \mathcal{A}(E)\cos\left(\omega_a t + \phi(E)\right)\right], \tag{5.35}$$

where $\mathcal{N}$ and $\mathcal{A}$ are computed in the lab frame, $\gamma\tau$ is the dilated lifetime of 64.4 μs, and ϕ is a phase depending on the flight path lengths of the decay electrons. The energy and arrival time of the decay electrons are measured. The observed rate then gives the precession rate, ω_a, by a fit to function (5.35). Using this ω_a and the measured magnetic field $\mathbf{B}$, equation (5.34) can then be solved for the anomalous part, a.

5.3.2. The standard model prediction

The coupling which gives the muon magnetic moment is shown in Figure 5.3.2.. The "anomalous" corrections to the magnetic moment are due to exchange of virtual particles which affect the measured value of the photon-muon coupling.

The electrodynamics prediction is (to 5 loops) [53, 54, 61]

$$a(\text{QED}) = 116584705.7(2.9) \times 10^{-11},$$

which differs in the fifth significant figure from the experimental value. The Z^0 and Higgs bosons, along with $W^{\pm}$ bosons and neutrino exchange, make up the electroweak corrections to the vertex. The contribution from the above electroweak effects is (to 2

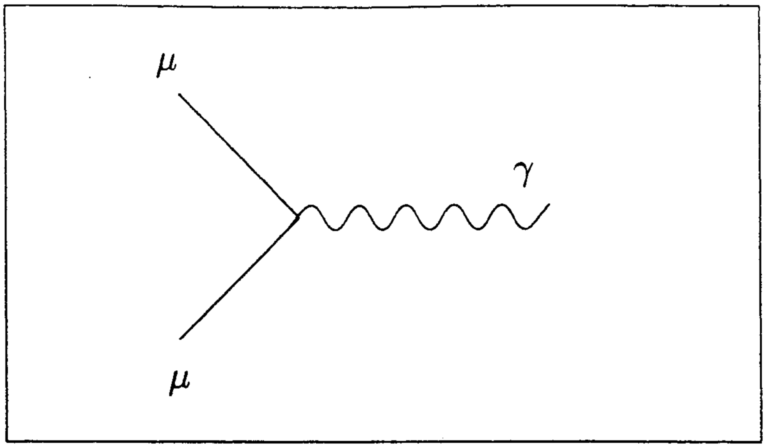


Figure 5.3. The photon-muon vertex which gives rise to the magnetic moment.

loops) [53, 62, 63]

$$a(\mathrm{EW}) = 152(4) \times 10^{-11}.$$

There is one more standard model correction given by hadron exchange where the first muon emits a photon which is energetic enough to produce hadron loops in the diagram; in other words, quark anti-quark pairs are produced, which then annihilate into another photon absorbed by the final state muon. The hadron loop correction to the muon moment (to 3 loops) [53, 64, 65] is

$$a(\mathrm{hadron1}) = 6739(67) \times 10^{-11}$$

or

$$a(\mathrm{hadron2}) = 6803(114) \times 10^{-11}.$$

We must note that this correction is very difficult to calculate, thus it contributes the highest amount of uncertainty in the total standard model prediction.

The complete standard model prediction is then

$$
\begin{aligned}
a(\mathrm{SM}) &= a(\mathrm{QED}) + a(\mathrm{EW}) + a(\mathrm{hadron1}) \\
&= 116591597(67) \times 10^{-11} \\
&\text{or (with a(hadron2))} \\
&= 116591660(114) \times 10^{-11}
\end{aligned}
\tag{5.36}
$$

We now see that the standard model prediction differs from experiment in the sixth digit, which is much better than the QED calculation.

The current standard model calculation differs from the latest experimental results and is outside of the experimental error bars. This discrepancy motivates a search for contributing sources beyond the standard model.

5.4. Torsion contribution to the magnetic moment

Using the Feynman rules calculated in Section 5.1., we can find the contribution of the antisymmetric tensor to fermion anomalous magnetic moments. The experimental discrepancy may then place a useful bound on the fermion-antisymmetric tensor coupling.

The Feynman diagram for the muon magnetic moment with an antisymmetric tensor interaction is shown in Figure 5.4.. We use the Feynman rules for QED as well as the

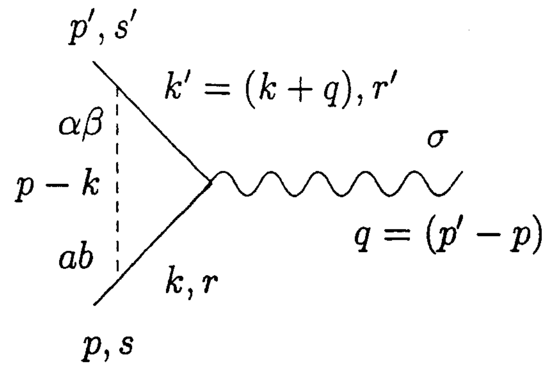


Figure 5.4. The Feynman diagram for the antisymmetric tensor contribution to the muon anomalous magnetic moment. We have written the photon polarization index, σ; the antisymmetric tensor indices, a, b, α, and β; and the spinor indices, s, s', r, and r'. The initial and final muon momenta are p and p'. The momentum of the antisymmetric tensor field is $p - k$. The momentum transfer to the photon is q.

ones given in Figure 5.1. to give the vertex correction:

$$\Gamma^{\mu} = \gamma^{\mu} + \delta\Gamma^{\mu}, \tag{5.37}$$

where

$$\begin{aligned} \delta\Gamma^{\sigma}(p', p)_{s's} = \int \frac{d^4k}{(2\pi)^4} \Bigg\{ & \left[-\frac{g}{m} \Lambda^{\alpha\beta}_{s'r'}(p-k) \right] \left[iG_{\alpha\beta ab}(p-k) \right] \left[-\frac{g}{m} \Lambda^{ab}_{rs}(p-k) \right] \\ & \times \left[iS_F(k)_{lr} \right] \left(-\gamma^{\sigma}_{l'l} \right) \left[iS_F(k')_{r'l'} \right] \Bigg\}. \end{aligned} \tag{5.38}$$

Using the propagator (5.15) and vertex rule (5.18), we have

$$\delta\Gamma^{\sigma}(p',p)_{ss'} = \frac{ig^2}{M^2}\int\frac{d^4k}{(2\pi)^4}\Big\{\epsilon_{\mu\nu\lambda\rho}\left(g^{\nu\alpha}g^{\lambda\beta}(p-k)^{\mu}+g^{\lambda\alpha}g^{\mu\beta}(p-k)^{\nu}+g^{\mu\alpha}g^{\nu\beta}(p-k)^{\lambda}\right)$$
$$\epsilon_{mnlp}\left(g^{na}g^{lb}(p-k)^{m}+g^{la}g^{mb}(p-k)^{n}+g^{ma}g^{nb}(p-k)^{l}\right)$$
$$\left(g_{a\beta}g_{\alpha b}-g_{\alpha a}g_{\beta b}\right)\left[\frac{\gamma_5\gamma^{\rho}\left(\not{k}'+m\right)\gamma^{\sigma}\left(\not{k}+m\right)\gamma_5\gamma^{p}}{\left[k^2-m^2\right]\left[(k')^2-m^2\right](k-p)^2}\right]\Big\}. \tag{5.39}$$

Working out the tensor part first using the identity

$$\epsilon_{\mu\nu\alpha\beta}\epsilon^{\alpha\beta mn} = -4\delta^{m}_{[\mu}\delta^{n}_{\nu]} \tag{5.40}$$

reduces the tensor part to 36 terms of the form $((k-p)^2 g_{p\rho}-(k-p)_{\rho}(k-p)_{p})$. Our vertex function reduces to

$$\delta\Gamma^{\sigma}(p',p)_{ss'} = \frac{i36g^2}{M^2}\int\frac{d^4k}{(2\pi)^4}\Big\{$$
$$\frac{\gamma^{\rho}\left[\not{k}'+m\right]\gamma^{\sigma}\left[\not{k}+m\right]\gamma^{p}\Big([k-p]^2 g_{p\rho}-(k-p)_{p}(k-p)_{\rho}\Big)}{[k-p]^2\left[(k')^2-m^2\right]\left[k^2-m^2\right]}\Big\}. \tag{5.41}$$

We can see that there are eight powers of momentum in the numerator and six in the denominator, which gives a quadratic divergence that we will need to regularize. We now concentrate on reducing the integrand

$$\frac{\gamma^{\rho}\left[\not{k}'+m\right]\gamma^{\sigma}\left[\not{k}+m\right]\gamma_{\rho}}{\left[(k')^2-m^2\right]\left[k^2-m^2\right]} - \frac{\left[\not{k}-\not{p}\right]\left[\not{k}'+m\right]\gamma^{\sigma}}{[k-p]^2\left[(k')^2-m^2\right]} \tag{5.42}$$

using the following identities (notational details in Appendix A)

$$\begin{aligned}\gamma^{\rho}\,\not{k}'\gamma^{\sigma}\,\not{k}\gamma_{\rho} &= -2\,\not{k}\gamma^{\sigma}\,\not{k}' + \epsilon\,\not{k}'\gamma^{\sigma}\,\not{k}\\ \gamma^{\rho}\gamma^{\sigma}\,\not{k}\gamma_{\rho} &= 4k^{\sigma} - \epsilon\gamma^{\sigma}\,\not{k}\\ \gamma^{\rho}\gamma^{\sigma}\gamma_{\rho} &= (\epsilon-2)\,\gamma^{\sigma},\end{aligned} \tag{5.43}$$

and the fact that $k' = k + q$ to give

$$\delta\Gamma^{\sigma}(p',p) = \frac{i36g^2}{M^2}\int\frac{d^4k}{(2\pi)^4}\left\{\frac{A^{\sigma}}{[(k')^2-m^2]\,[k^2-m^2]} - \frac{B^{\sigma}}{[k-p]^2\,[(k')^2-m^2]}\right\}, \tag{5.44}$$

where

$$\begin{aligned}A^{\sigma} &= \left(8mk^{\sigma} + 4mq^{\sigma} - 2\,\not{k}\gamma^{\sigma}\,\not{k} - 2\,\not{k}\gamma^{\sigma}\,\not{q} - 2m^2\gamma^{\sigma}\right)\\ &\quad + \epsilon\left(\not{k}\gamma^{\sigma}\,\not{k} + \not{q}\gamma^{\sigma}\,\not{k} - 2mk^{\sigma} - m\,\not{q}\gamma^{\sigma} + m^2\gamma^{\sigma}\right),\\ B^{\sigma} &= 2p\cdot k\gamma^{\sigma} + 2p\cdot q\gamma^{\sigma} - 2p^{\sigma}\,\not{k} - k^2\gamma^{\sigma} - \not{k}\,\not{q}\gamma^{\sigma} + 2mp^{\sigma} - m^2\gamma^{\sigma}.\end{aligned} \tag{5.45}$$

We now use the following Feynman parametrization

$$\begin{aligned}\frac{1}{[(k')^2-m^2]\,[k^2-m^2]} &= \int_0^1 dxdy\;\delta(x+y-1)\frac{1}{[(k')^2x - m^2x + k^2y - m^2y]^2}\\ &= \int_0^1 dxdy\;\delta(x+y-1)\frac{1}{[k^2 + xq^2 + 2xk\cdot q - m^2]^2},\end{aligned} \tag{5.46}$$

and shifting the integration variable, $k \to k - xq$, leaves

$$\frac{1}{[(k')^2-m^2]\,[k^2-m^2]} = \int_0^1 dxdy\;\delta(x+y-1)\frac{1}{[k^2-\Delta_1]^2}, \tag{5.47}$$

where $\Delta_1 = m^2 - xyq^2$. Similarly, with the second term

$$\frac{1}{[k-p]^2\left[(k')^2 - m^2\right]} = \int_0^1 dxdy\ \delta(x+y-1)\frac{1}{[k^2-\Delta_2]^2}, \tag{5.48}$$

where we have shifted $k \to k - (xp + yq)$ and defined $\Delta_2 = m^2y - (q-p)^2xy$. Now, making these variable shifts in the numerators, we arrive at

$$\delta\Gamma^\sigma = \frac{i36g^2}{M^2}\int\frac{d^4k}{(2\pi)^4}\int_0^1 dxdy\ \delta(x+y-1)\left\{\frac{N_1^\sigma + \epsilon N_2^\sigma}{[k^2-\Delta_1]^2} - \frac{N_3^\sigma}{[k^2-\Delta_2]^2}\right\}, \tag{5.49}$$

where, with $Q \equiv (p+p')$,

$$\begin{aligned} N_1^\sigma &= 4m(1-2x)q^\sigma - 2\ \not{k}\gamma^\sigma\ \not{k} - 2xy\left(2m^2+q^2\right)\gamma^\sigma - 2m^2\gamma^\sigma \\ N_2^\sigma &= \not{k}\gamma^\sigma\ \not{k} - x^2q^2\gamma^\sigma - 2m^2x^2\gamma^\sigma - m(1-2x)q^\sigma - m^2\gamma^\sigma + mQ^\sigma \\ N_3^\sigma &= \left(k^2 + m^2(1+x^2) + xq^2(1+x)\right)\gamma^\sigma \end{aligned} \tag{5.50}$$

and

$$\begin{aligned} \Delta_1 &= m^2 - xyq^2 \\ \Delta_2 &= m^2y^2 - 2xyq^2. \end{aligned} \tag{5.51}$$

We have used the fact that $Q\cdot q = (p+p')\cdot q = 0$. The terms containing q^σ will vanish after performing the x, y integrations as it should be according to the Ward identity [19].

We can now see that the only term that will contribute to the anomalous magnetic moment is the term in N_2^σ involving Q^σ. Extracting only this term, we have

$$\delta\Gamma^\sigma_{\text{anom.}} = \frac{i36g^2}{M^2}\int\frac{d^4k}{(2\pi)^4}\int_0^1 dxdy\,\delta(x+y-1)\left\{\frac{\epsilon m Q^\sigma}{[k^2-\Delta_1]^2}\right\}. \tag{5.52}$$

We now need to use dimensional regularization to perform the k integration. Using formula [19] (See Section A.4. in Appendix A.)

$$\int\frac{d^4k}{(2\pi)^4}\frac{1}{(k^2-\Delta)^n} = \frac{(-1)^n i}{(4\pi)^{\frac{d}{2}}}\frac{\Gamma(n-\frac{d}{2})}{\Gamma(n)}\frac{1}{\Delta^{n-\frac{d}{2}}}, \tag{5.53}$$

we have

$$\begin{aligned}\delta\Gamma^\sigma_{\text{anom.}} &= \frac{-36g^2}{M^2}\int_0^1 dxdy\,\delta(x+y-1)\left\{\frac{\epsilon m Q^\sigma(4\pi)^{\frac{\epsilon}{2}}\mu^\epsilon\Gamma(\frac{\epsilon}{2})}{\Delta_1^{\frac{\epsilon}{2}}}\right\}\\ &= \frac{-36g^2}{(4\pi)^2M^2}\int_0^1 dxdy\,\delta(x+y-1)\left\{\epsilon m Q^\sigma\left[\frac{2}{\epsilon}-\gamma+\ln\left(\frac{4\pi\mu^2}{\Delta}\right)\right]\right\},\end{aligned} \tag{5.54}$$

and keeping only the zeroth order term in ϵ, we obtain

$$\delta\Gamma^\sigma_{\text{anom.}} = \frac{-72mg^2}{(4\pi)^2M^2}Q^\sigma. \tag{5.55}$$

The Gordon identity [19] is

$$Q^\sigma = 2m\gamma^\sigma - i\sigma^{\sigma\nu}q_\nu, \tag{5.56}$$

and substituting into (5.55) and isolating the magnetic moment part gives

$$\delta\Gamma^\sigma_{\text{anom.}} = \frac{72img^2}{(4\pi)^2M^2}\sigma^{\sigma\nu}q_\nu \equiv \frac{i\sigma^{\sigma\nu}q_\nu}{2m}F_2(q^2), \tag{5.57}$$

which isolates the structure function, $F_2(q^2)$, as

$$F_2(q^2) = \frac{144m^2}{M^2}\frac{g^2}{(4\pi)^2}, \tag{5.58}$$

where we notice that there is no momentum dependence to this loop order. The g-factor is defined by $\frac{g-2}{2} = F_2(0)$ and, hence,

$$\frac{g_\mu - 2}{2} = \frac{9m_\mu^2 g^2}{M^2\pi^2}. \tag{5.59}$$

We can now bound the antisymmetric tensor coupling [51] by fitting to the experimental discrepancy for the muon g-factor.

The standard model prediction differs from experiment [54] by $\delta a = 25(16) \times 10^{-10}$. Using this difference as an upper bound to the torsion contribution, we have

$$a_\mu(\text{torsion}) \leq 25 \times 10^{-10}. \tag{5.60}$$

The muon mass is given in [23] as $m_\mu = 105.658$ MeV which we insert into equation (5.59) to give

$$\frac{g^2}{M^2} \leq 2.456 \times 10^{-7}\ \text{GeV}^{-2} \simeq 2.5 \times 10^{-7}\ \text{GeV}^{-2} \tag{5.61}$$

as an upper bound on the antisymmetric tensor coupling.[3] Notice that, for g to be of order 1, we must have a mass scale of $M \sim 2$ TeV which is much smaller than the

[3]Other authors [12, 66] have used $\mathcal{L}_{\text{int}} = -i\sqrt{\frac{\pi G}{12}}\ \kappa\bar{\psi}\sigma_{\mu\nu\lambda}H^{\mu\nu\lambda}\psi$ as their interaction lagrangian rather than the one that we have used in (3.70). In order to facilitate comparison with their results, it should be noted that our bound on g^2/M^2 given in equation (5.61) translates into a bound of $\kappa \leq 1.19 \times 10^{16}$.

Planck scale, $M_p = \sqrt{\frac{\hbar c}{G}} = 1.221 \times 10^{19}$ GeV, and may be near the supersymmetry scale.

Our coupling is valid for all fermions, thus it must also have an effect on the anomalous magnetic moment of the electron. The standard model prediction for the electron g-factor is verified by experiment to a very high accuracy and precision, thus we need to check that our result for the antisymmetric tensor contribution to the electron anomalous magnetic moment does not destroy the agreement but remains within the experimental error bars. The electron anomalous magnetic moment is given in [23] as

$$a_e = \frac{g_e - 2}{2} = 1159652187 \pm 4 \times 10^{-12}. \tag{5.62}$$

With our upper bound for the coupling constant, we get

$$a_e(\text{torsion}) = 0.05858 \times 10^{-12} \tag{5.63}$$

for the electron, which is well within the present day experimental error.

We also have a predicted torsion contribution to the tau lepton. The mass of the tau is 1.777 GeV, hence the 1-loop torsion exchange gives a upper bound contribution of

$$a_\tau(\text{torsion}) = 7.199 \times 10^{-7}$$

to the tau anomalous magnetic moment.

We have found that the interaction between a fermion and an antisymmetric tensor field such as the one arising in string theory and in Einstein-Cartan gravity can solve the problem with the muon anomalous magnetic moment without having a significant effect on the electron anomalous magnetic moment if the torsion coupling constant satisfies the bound given by equation (5.61). There are many other possible contributions to the muon anomalous magnetic moment which are outside the standard model. The most likely candidates are supersymmetric partners to the standard model spectrum [67, 68, 69, 70, 71]. If these particles are found to exist, our bound on the coupling would then change correspondingly.

CHAPTER 6

SUMMARY, CONCLUSIONS, AND FUTURE PROSPECTS

6.1. Summary

We have studied the antisymmetric tensor interaction from several points of view. First, we have shown how it arises in Einstein-Cartan gravity with non-zero spacetime torsion. We then showed how a gauge theory of gravity gives rise to interacting torsion which can be made to propagate by either postulating a potential from which the torsion is derived or by postulating a departure from conventional Einstein gravity.

Next, we discussed how torsion arises in string theory as the Kalb-Ramond antisymmetric tensor field. It comes from the geometrical structure of the theory and is also present in the generalization of string theory called M-theory.

Having motivated the existence, propagation, and interaction of completely antisymmetric torsion, we then proceeded to calculate two examples of possible physical effects of non-zero torsion. The first example was the tree-level scattering of two fermions by the exchange of a virtual torsion field. The second example was the torsion contribution to the fermion anomalous magnetic moment showing that such an interaction could resolve the muon anomalous magnetic moment problem. We used recent experimental results to place a bound on the torsion coupling constant.

6.2. Future ideas

There are many possibilities for future study motivated by this dissertation. We will now briefly discuss some ideas that could be developed as a continuation of this work.

6.2.1. Quadratic actions

In Chapter 5, we used the conventional Einstein-Cartan action and calculated torsion scattering based on the torsion being derivable from a potential in order to get propagation. The results of that analysis are valid in the case that the scalar curvature is the proper action of the theory. Instead, we could have used the alternate theory with a quadratic action, resulting in fully propagating torsion without postulating a potential. We could examine the possibility that the gauge field is the fundamental propagating field of gravity in analogy with the gauge fields of particle physics. The metric is forced into a role of non-propagating background geometry. To get a theory of propagating torsion, we must introduce an action which is quadratic in the field strength which also constitutes a drastic deviation from the action of general relativity which is the scalar curvature. We hold to the viewpoint that these possibilities must be checked since they need merely to reduce to the verified prediction of Einstein-Maxwell theory at low energy to be consistent with experiments.

If a new theory of this form reduces to Newton's gravitation in the appropriate limits, then it would constitute an exciting possibility for a quantum theory of gravity. The difference between this theory and most theories of quantum gravity is that we are suggesting that it is the connection field which constitutes the fundamental degrees of freedom and is responsible for gravitational interactions rather than the spacetime metric. The metric is demoted to a background field with no dynamics.

In analogy with the gauge theories constructed in Sections 3.1.1. and 3.1.2., the natural action for the gauge theory of torsion is

$$
\begin{aligned}
-\frac{1}{4}F^{ab}_{\mu\nu}F^{\mu\nu}_{ab} &= -\left(\partial_{[\mu}\Gamma_{\nu]}{}^{ab} + \Gamma_{[\mu l}{}^{b}\Gamma_{\nu]}{}^{al}\right)\left(\partial^{[\mu}\Gamma^{\nu]}{}_{ab} + \Gamma^{[\mu l}{}_{b}\Gamma^{\nu]}{}_{al}\right) \\
&= -\left(\partial_{[\mu}\Gamma_{\nu]}{}^{ab}\partial^{[\mu}\Gamma^{\nu]}{}_{ab} + \Gamma_{[\mu l}{}^{b}\Gamma_{\nu]}{}^{al}\Gamma^{[\mu l}{}_{b}\Gamma^{\nu]}{}_{al} + 2\partial_{[\mu}\Gamma_{\nu]}{}^{ab}\Gamma^{[\mu l}{}_{b}\Gamma^{\nu]}{}_{al}\right).
\end{aligned}
\tag{6.1}
$$

We have a quadratic term for the propagation of $\Gamma_{\mu}{}^{ab}$, and we have 3-Γ and 4-Γ self-interaction terms just like one gets in other non-abelian gauge theories like QCD. The propagation of the connection implies that, when it is separated into symmetric and antisymmetric parts, we will have propagating torsion also. We use equation (3.70) to write the full action in this new theory as

$$
\begin{aligned}
S = \int_M d^4x\sqrt{-g}\Big\{ &\frac{i}{2}\bar{\psi}\gamma^a e^{\mu}{}_a\partial_\mu\psi - \frac{i}{2}\partial_\mu\bar{\psi}e^{\mu}{}_a\gamma^a\psi - \frac{1}{2}\bar{\psi}\sigma_{abc}H^{abc}\psi \\
&- \partial_{[\mu}\tilde{\Gamma}_{\nu]}{}^{ab}\partial^{[\mu}\tilde{\Gamma}^{\nu]}{}_{ab} - \tilde{\Gamma}_{[\mu l}{}^{b}\tilde{\Gamma}_{\nu]}{}^{al}\tilde{\Gamma}^{[\mu l}{}_{b}\tilde{\Gamma}^{\nu]}{}_{al} - 2\partial_{[\mu}\tilde{\Gamma}_{\nu]}{}^{ab}\tilde{\Gamma}^{[\mu l}{}_{b}\tilde{\Gamma}^{\nu]}{}_{al}\Big\}.
\end{aligned}
\tag{6.2}
$$

6.2.2. Dual variables

It may be useful to separate the connection into "electric" and "magnetic" parts, and use these new fields as our fundamental variables. Let us examine the field strength (3.78) a bit more closely.

$$
F^{ab}_{\mu\nu} = 2\partial_{[\mu}\Gamma_{\nu]}{}^{ab} + 2\eta_{mn}\Gamma_{[\mu}{}^{ma}\Gamma_{\nu]}{}^{nb} \tag{6.3}
$$

Due to the antisymmetry in the Latin indices, there are only six different ab combinations. We can think of $\Gamma_\mu{}^{ab} = \Gamma_\mu{}^A$ as a collection of six vectors. Let us introduce

$$E^i_\mu = \Gamma_\mu{}^{0i} \tag{6.4}$$

and

$$\epsilon^{ijk} B^k_\mu = \Gamma_\mu{}^{ij} \tag{6.5}$$

so that equation (6.3) becomes

$$F^{0i}_{\mu\nu} = 2\partial_{[\mu} E^i_{\nu]} - 2\left(B_{[\mu} \times E_{\nu]}\right)^i \tag{6.6}$$

and

$$F^{ij}_{\mu\nu} = 2\epsilon^{ijk}\partial_{[\mu} B^k_{\nu]} + 2E^i_{[\mu}E^j_{\nu]} - 2B^i_{[\mu}B^j_{\nu]} \tag{6.7}$$

The quadratic action (6.1) now becomes

$$\begin{aligned} -\frac{1}{4}F^{ab}_{\mu\nu}F^{\mu\nu}_{ab} &= 2\partial_{[\mu}E^i_{\nu]}\partial^{[\mu}E^{\nu]}_i + 2\partial_{[\mu}B^i_{\nu]}\partial^{[\mu}B^{\nu]}_i - 4\epsilon_{ijk}\partial^{[\mu}E^{\nu]i}B^j_{[\mu}E^k_{\nu]} \\ &+ 2B^i_{[\mu}E^j_{\nu]}B^{[\mu}_i E^{\nu]}_j - 2B^i_{[\mu}E^j_{\nu]}B^{[\mu}_j E^{\nu]}_i + 2E^i_{[\mu}E^j_{\nu]}B^{[\mu}_i B^{\nu]}_j \\ &- E^i_{[\mu}E^j_{\nu]}E^{[\mu}_i E^{\nu]}_j - B^i_{[\mu}B^j_{\nu]}B^{[\mu}_i B^{\nu]}_j, \end{aligned} \tag{6.8}$$

and the Feynman diagrams for these gauge field self interactions are shown in Figure 6.2.2..

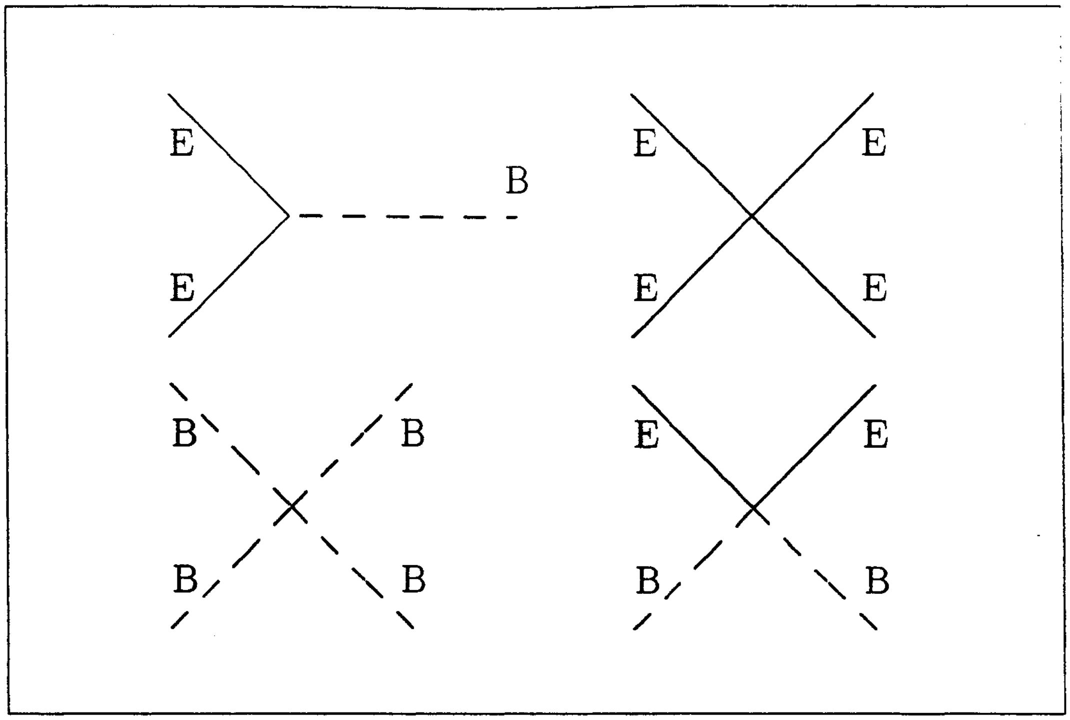


Figure 6.1. Feynman diagrams for the self-interactions of electric and magnetic gauge fields of gravity.

6.2.3. Torsion as an effective theory

The torsion coupling (3.70) which we used in the previous chapters had dimensions of GeV^{-2}. This fact hints that, perhaps, our theory really is an effective theory since inverse mass dimensional couplings are indicative that some heavy particle has been integrated out. Let us look at how this works.

The generating functional is defined as the integral of all possible field configurations in spacetime weighted by the exponential of the action.

$$Z[\eta^a, \bar{\eta}^b, j^\mu] = \int \mathcal{D}\psi \, \mathcal{D}\bar{\psi} \, \mathcal{D}A \; \exp\Big(i \int d^4x \, \big[\mathcal{L}_0 + \bar{\psi}_a \xi^a + \bar{\eta}^b \psi_b + j^\mu A_\mu + \mathcal{L}_{\text{int}}\big]\Big), \quad (6.9)$$

where we use lowercase Latin indices, a, b, and c, as possible isospin indices for the fields, ψ and $\overline{\psi}$. $\mathcal{L}_0$ is the sum of all free particle lagrangians, and $\mathcal{L}_{\text{int}}$ is the interaction lagrangian. The fermion source currents, η and $\overline{\eta}$, as well as the fermion fields are now Grassmann numbers that obey anticommutation relations, whereas the bosonic fields and currents are commuting.

We now define the n-point function (or n-point Green's function) to be the vacuum expectation value of the time-ordered product of fields at n spacetime points $x_1, ..., x_n$ as follows:

$$< 0|T\left(\phi(x_1)\cdots\phi(x_n)\right)|0 >\equiv \left(\frac{1}{i}\frac{\delta}{\delta J(x_1)}\right)\cdots\left(\frac{1}{i}\frac{\delta}{\delta J(x_n)}\right)Z[J]\Big|_{J=0}, \qquad (6.10)$$

where $Z[J]$ is the generating functional, (6.9), and the fields $\phi(x_i)$ can be any of the $\psi, \overline{\psi}$ or A^μ fields.

We see by the form of the generating functional, (6.9), that an application of a particular functional derivative will bring down a factor of the corresponding field. In this way, we can construct any polynomial in the fields by merely acting on the generating functional with functional derivatives. In particular, we can expand a given interaction lagrangian as a polynomial in the fields and then re-write it in terms of the functional derivatives. We can, therefore, re-write $Z[J]$ as

$$Z = \frac{1}{N}\exp\left[i\int d^4x\mathcal{L}_{\text{int}}\left(\frac{\delta}{\delta\eta_a(x)}, \frac{\delta}{\delta\overline{\eta}_b(x)}, \frac{\delta}{\delta j_\mu(x)}\right)\right]Z_0, \qquad (6.11)$$

where Z_0 is the remaining terms in equation (6.9) after removing the interaction part.

We have divided by $N = Z|_{j=\eta=\overline{\eta}=0}$, which has the effect of canceling all of the so-called vacuum bubble diagrams which have no external lines and are, hence, unobservable. To use this expression, one applies well-known methods [19] to reduce the free field generating functional Z_0 (which is simply a product of Gaussians) into the following form:

$$Z_0 = \exp\left[-i\int d^4x\, d^4y\, \left(\overline{\eta}_a(x)S_F^{ab}(x-y)\eta_b(y) + \frac{1}{2}j_\mu(x)D_F^{\mu\nu}(x-y)j_\nu(y)\right)\right], \quad (6.12)$$

where the Feynman propagators are defined in Appendix A. Finally, we form the generator of connected graphs by writing $Z \to -i\ \ln(Z)$, which removes all of the diagrams that are disconnected. In the path integral method, one uses this generating functional to find the propagator; the vertex functions; and, as a result, the Feynman rules of the theory.

Let us now look at a generic spin-$\frac{1}{2}$ theory and show how one forms an effective theory from it in this manner. If we view the fermions as very heavy static sources, we can integrate them out of the theory, resulting in mass corrections to the vertices.[1]

The lagrangian for a free fermion[2] χ is

$$\mathcal{L} = \overline{\chi}\,(i\ \partial\!\!\!/ - M)\,\chi + \overline{\eta}\chi + \overline{\chi}\xi, \quad (6.13)$$

[1]For an example of this procedure using heavy boson fields, see reference [18].

[2]The results derived here are also valid for any spin-$\frac{1}{2}$ particle or resonance since they depend only on the form of the free particle propagator. See Chapter 6 in [72] (which is also available as [73]) for the complete derivation in the case of spin-$\frac{1}{2}$ and spin-$\frac{3}{2}$ nucleon resonances.

where ξ and $\overline{\eta}$ are sources. The action is given by

$$
\begin{aligned}
S &= \int d^4x \; \mathcal{L}(\chi, \overline{\chi}, \xi, \overline{\eta}) + \overline{\eta}\chi + \overline{\chi}\xi \\
&= \int d^4x \; [\overline{\chi}\mathcal{D}\chi + \overline{\eta}\chi + \overline{\chi}\xi] \\
&= \int d^4x \; \left[(\overline{\chi} - \overline{\eta}\mathcal{D}^{-1})\mathcal{D}(\chi - \mathcal{D}^{-1}\xi) + \overline{\eta}\mathcal{D}^{-1}\xi\right],
\end{aligned}
\tag{6.14}
$$

where we have "completed the square" in the last line and made the following convenient definitions,

$$
\begin{aligned}
\mathcal{D} &= (i\,\partial\!\!\!/ - M) \\
\mathcal{D}^{-1}\mathcal{D} &= -\delta^4(x-y) \\
\mathcal{D}^{-1}\xi &= -\int d^4y \; S_F(x-y)\xi(y) \\
\overline{\eta}\mathcal{D}^{-1} &= -\int d^4y \; \overline{\eta}(y)S_F(x-y),
\end{aligned}
\tag{6.15}
$$

which allow us to work without explicitly showing the integrations which are taking place. We now make the change of variables

$$
\chi' = \chi + \int d^4y \; S_F(x-y)\xi(y) = \chi - \mathcal{D}^{-1}\xi \tag{6.16}
$$

and

$$
\overline{\chi}' = \overline{\chi} + \int d^4y \; \overline{\eta}(y)S_F(x-y) = \overline{\chi} - \overline{\eta}\mathcal{D}^{-1}, \tag{6.17}
$$

which gives our action as

$$S = \int d^4x \,\left[\overline{\chi'}\mathcal{D}\chi' + \overline{\eta}\mathcal{D}^{-1}\xi\right]. \tag{6.18}$$

If we recall the expression for the path integral, (6.9), we see that the integration ranges over all possible fields, χ and $\overline{\chi}$, and hence $\mathcal{D}\chi = \mathcal{D}\chi'$ and $\mathcal{D}\overline{\chi} = \mathcal{D}\overline{\chi}'$. Our generating functional is then

$$\begin{aligned} Z &= \frac{\int D\overline{\chi'}\, D\chi' e^{i\int d^4x\, \mathcal{L}(\overline{\chi}',\chi',\overline{\eta},\xi)}}{\int D\overline{\chi}\, D\chi\, e^{i\int d^4x\, \mathcal{L}(\overline{\chi},\chi,0,0)}} = \frac{\int D\overline{\chi'}\, D\chi' e^{i\int d^4x\, [\overline{\chi}'\mathcal{D}\chi']} e^{i\int d^4x [\overline{\eta}\mathcal{D}^{-1}\xi]}}{\int D\overline{\chi}\, D\chi\, e^{i\int d^4x\, [\overline{\chi}\mathcal{D}\chi]}} \\ &= e^{i\int d^4x [\overline{\eta}\mathcal{D}^{-1}\xi]}. \end{aligned} \tag{6.19}$$

Since $Z = e^{iS}$, we have our effective action as

$$S_{\text{eff}} = \int d^4x\, \overline{\eta}\mathcal{D}^{-1}\xi = -\int d^4x\, d^4y\, \overline{\eta} S_F(x-y)\xi. \tag{6.20}$$

The lagrangian implicit in equation (6.20) can be made local by noticing that the heavy field propagator, $S_F(x-y)$, is peaked at small distances. We can, therefore, Taylor expand $\xi(y)$ as

$$\xi(y) = \xi(x) + (y-x)^\mu \left[\partial_\mu \xi(y)\right]_{y=x} + \dots \tag{6.21}$$

and keep the leading term.

Using the fact[3] that

$$\int d^4x\, S_F(x-y) = \int \frac{d^4p}{(2\pi)^4}\, d^4x\, e^{-ip\cdot(x-y)} \frac{\not{p}+M}{p^2-M^2}$$
$$= \int \frac{d^4p}{(2\pi)^4}\, e^{ip\cdot y}(2\pi)^4\delta^4(p)\frac{\not{p}+M}{p^2-M^2} \tag{6.22}$$
$$= -\frac{1}{M},$$

we have from the action (6.20)

$$S_{\text{eff}} = -\int d^4x\, d^4y\, \overline{\eta}(x) S_F(x-y)\xi(x) + \dots$$
$$= -\int d^4x\, \frac{-1}{M}\overline{\eta}(x)\xi(x) + \dots \tag{6.23}$$
$$\approx \int d^4x\, \frac{1}{M}\overline{\eta}(x)\xi(x)$$

which gives an expression for our effective lagrangian with the heavy field integrated out

$$\mathcal{L}_{\text{eff}} = \frac{1}{M}\overline{\eta}(x)\xi(x). \tag{6.24}$$

We now look at the lagrangian (5.1)

$$\mathcal{L} = \det\left(e^{\mu}{}_{a}\right)\overline{\psi}\left(i\,\not{\partial} - m\right)\psi + \text{H.c.} - \det\left(e^{\mu}{}_{a}\right)\overline{\psi}\frac{g}{9M}\sigma_{\mu\nu\lambda}H^{\mu\nu\lambda}\psi \tag{6.25}$$

and interpret the dimensionful coupling as an indication that we are dealing with an

[3]This equation is where the difference occurs between various heavy fields which may have different propagators.

effective interaction with some heavy fermion integrated out. We postulate

$$\mathcal{L}_{\text{int}} = \frac{1}{M}\overline{\eta}(x)\xi(x) = -\frac{1}{M}\overline{\psi}\det\left(e^{\mu}{}_{a}\right)\frac{g}{9}\sigma_{\mu\nu\lambda}H^{\mu\nu\lambda}\,\psi. \tag{6.26}$$

The underlying theory is of the form

$$\mathcal{L} = \overline{\chi}\left(i\,\partial\!\!\!/ - M\right)\chi + \overline{\eta}\chi + \overline{\chi}\xi. \tag{6.27}$$

If we write

$$\overline{\eta}(x) = \overline{\psi}A \quad\text{and}\quad \xi(x) = B\psi, \tag{6.28}$$

then we have

$$\mathcal{L} = \overline{\chi}\left(i\,\partial\!\!\!/ - M\right)\chi + \overline{\psi}\left(i\,\partial\!\!\!/ - m\right)\psi + \overline{\psi}A\chi + \overline{\chi}B\psi \tag{6.29}$$

as our fundamental theory. The effective theory is

$$\mathcal{L} = \overline{\psi}\left(i\,\partial\!\!\!/ - m\right)\psi + \frac{1}{M}\overline{\psi}AB\psi. \tag{6.30}$$

We, therefore, need to identify

$$\begin{aligned} AB &= -\det\left(e^{\mu}{}_{a}\right)\frac{g}{9}\sigma_{\mu\nu\lambda}H^{\mu\nu\lambda} \\ &= -\det\left(e^{\mu}{}_{a}\right)\frac{ig}{9}\epsilon_{\mu\nu\lambda\sigma}\gamma_5\gamma^{\sigma}\left(\partial^{\mu}B^{\nu\lambda} + \partial^{\nu}B^{\lambda\mu} + \partial^{\lambda}B^{\mu\nu}\right). \end{aligned} \tag{6.31}$$

One possible way to implement this condition is to let $A = \det\left(e^{\mu}{}_{a}\right) = \sqrt{g}$ and $B = -\frac{g}{9}\sigma_{\mu\nu\lambda}H^{\mu\nu\lambda}$, giving a theory of a heavy and a light fermion interacting via derivative coupling with torsion along with a gravitational interaction with the determinant of the vierbein. The heavy particle is only very short lived. The effective theory given by integrating it out is a curved space formulation of theory of interacting torsion that we have used in the main body of this dissertation. It would be interesting to work out some of the details of this possibility in future work.

6.2.4. Topological effects

Another interesting possibility would be to examine the topological effects caused in various non-trivial spaces by the fact that the torsion field strength is an exact 3-form in the model that we have used. It is, therefore, invariant under the addition of an exact 2-form to B. In other words, the field strength is invariant under the gauge transformation (4.63)

$$B \to B + d\Lambda, \quad \text{or} \quad B_{\mu\nu} \to B_{\mu\nu} + \partial_\mu \Lambda_\nu - \partial_\nu \Lambda_\mu \tag{6.32}$$

in components. Thus, if the space has a non-trivial second cohomology group so that the second Betti number is non-zero (See Appendix D.), we could see effects such as torsion solitons and instantons due to the fact that there will be closed non-contractible 2-surfaces, K, and

$$I = \int_K B \tag{6.33}$$

will be a gauge invariant quantity. This idea is pursued further by Rohm and Witten [74].

REFERENCES

[1] Friedrich W. Hehl and Yuri N. Obukhov. How does the electromagnetic field couple to gravity, in particular to metric, nonmetricity, torsion and curvature? *gr-qc/0001010.*

[2] D. W. Sciama. The physical structure of general relativity. *Rev. Mod. Phys.*, 36(1):463, 1964.

[3] T. W. B. Kibble. Lorentz invariance and the gravitational field. *Jour. Math. Phys.*, 2(2):212, 1961.

[4] Friedrich W. Hehl, Paul von Heyde, and G. David Kerlick. General relativity with spin and torsion: Foundations and prospects. *Rev. Mod. Phys.*, 48(3):393, 1976.

[5] Friedrich W. Hehl, J. Dermott McCrea, Eckehard W. Mielde, and Yuval Ne'eman. Metric-affine gauge theory of gravity: Field equations, noether identities, world spinors and breaking of dilation invariance. *Phys. Rep.*, 258:1–171, 1995.

[6] M. B. Green, J. H. Schwartz, and E. Witten. *Superstring Theory, Vols 1 & 2.* Cambridge University Press, Cambridge, UK, 1987.

[7] M. J. Duff. *The World in Eleven Dimensions, Supergravity, Supermembranes and M-Theory.* IOP Publishing, London, 1999.

[8] Peter Peterson. Global riemannian geometry. *Bulletin of the AMS*, 36(3):297, 1999.

[9] Tohru Eguchi, Peter B. Gilkey, and Andrew J. Hanson. Gravitation, gauge theories and differential geometry. *Physics Reports*, 66(6):214, 1980.

[10] Richard T. Hammond. New fields in general relativity. *Contemporary Phys.*, 36(2):103, 1995.

[11] I. L. Shapiro. Physical aspects of the space-time torsion. *hep-th/0103093 and Phys. Rep.*, 357(2):213, 2002.

[12] Richard T. Hammond. Torsion gravity. *Rep. Prog. Phys.*, 65:559, 2002.

[13] N. D. Hari Dass and K. V. Shajesh. Vacuum polarization induced coupling between maxwell and Kalb-Ramond fields. *hep-th/0107006.*

[14] P. P. Fiziev. Graviation theory with propagating torsion. *gr-qc/9808006*, 1998.

[15] Hong-Jun Xie and Takeshi Shirafuji. Dynamical torsion and torsion potential. *gr-qc/9603006.*

[16] V. de Sabbata and C. Sivaram. Strong spin-torsion interaction between spinning protons. *Nuovo Cimento*, 101A(2):273. 1989.

[17] D. W. Sciama. *On The Analogy Between Charge and Spin in General Relativity.* McMillan, 1962.

[18] John F. Donoghue, Eugene Golowich, and Barry R. Holstein. *Dynamics of the Standard Model.* Cambridge University Press, Cambridge Massachusetts, 1992.

[19] M. E. Peskin and D. V. Shroeder. *An Introduction to Quantum Field Theory.* Addison-Wesley Publishing Company, Cambridge, Massachusetts, 1995.

[20] Michael Spivak. *A Comprehensive Introduction to Differential Geometry, Vols. 1 & 2.* Publish or Perish, Inc., New York, 1979.

[21] Michael Atiyah. *The Geometry and Physics of Knots.* Cambridge University Press, New York, 1990.

[22] S.W. Hawking and G.F.R. Ellis. *The Large Scale Structure of Space-Time.* Cambridge University Press, Cambridge, UK, 1973.

[23] Particle Data Group. Review of particle physics. *European Physical Journal*, C15(1-4):81–82, 2000.

[24] C. N. Yang and R. L. Mills. Conservation of isotopic spin and isotopic gauge invariance. *Phys. Rev.*, 96(1):191, 1954.

[25] C. N. Yang and R. L. Mills. Isotopic spin conservation and generalized gauge invariance. *Phys. Rev.*, 95(1):631, 1954.

[26] J. Goldstone, A. Salam, and S. Weinberg. Broken symmetries. *Phys. Rev.*, 127:965, 1962.

[27] Parthasarathi Majumdar and Soumitra SenGupta. Parity-violating gravitational coupling of electromagnetic fields. *Class. Quan. Grav.*, 16:L89, 1999.

[28] N. D. Birrell and P. C. W. Davies. *Quantum Fields in Curved Space.* Cambridge University Press, Cambridge, UK, 1982.

[29] B. Julia E. Cremmer and J. Scherk. Supergravity theory in 11 dimensions. *Phys. Lett.*, B76:409, 1978.

[30] N.A. Obers and B. Pioline. U-duality and M-theory. *Physics Reports*, 318:113, 1999.

[31] Miao Li. Introduction to M theory. *hep-th/9811019.*

[32] Gary T. Horowitz and Leonard Susskind. Bosonic M theory. *hep-th/0012037.*

[33] M. Kalb and P. Ramond. Classical direct interstring action. *Phys. Rev.*, D9:2273, 1974.

[34] Pierre Deligne, Pavel Etingof, Daniel S. Freed, Lisa C. Jeffrey, David Kazhdan, John W. Morgan, David R. Morrison, and Edward Witten. *Quantum Fields and Strings: A Course for Mathematicians, Vols 1 & 2.* American Mathematical Society, 1999.

[35] David Balin and Alexander Love. *Supersymmetric Gauge Field Theory & String Theory.* Inst. of Phys. Graduate Student Series, Philadelphia, 1994.

[36] T. Kaluza. Zum unitätsproblem der physik. *Sitz. Preuss. Akad. Wiss. Phys. Math.*, K1:966, 1921.

[37] O. Klein. Quantentheorie und fünfdimensionale relativitätstheorie. *Zeits. Phys.*, 37:895, 1926.

[38] J. M. Overduin and P. S. Wesson. Kaluza-klein gravity. *Phys. Rep.*, 283:303, 1997.

[39] J. Polchinski. *String Theory, Vols. 1 & 2.* Cambridge University Press, Cambridge, UK, 1998.

[40] J. Polchinski. Dirichlet-branes and Ramond-Ramond charges. *hep-th/9510017.*

[41] Micheal R. Douglas. Branes within branes. *hep-th/9512077.*

[42] Kasper Olsen and Richard J. Szabo. Constructing D-branes from K-theory. *hep-th/9907140.*

[43] Edward Witten. D-branes and K-theory. *hep-th/9810188.*

[44] Venzo de Sabbata and C. Sivaram. *Spin and Torsion in Gravitation.* World Scientific, Singapore, 1994.

[45] Biswarup Mukhopadhyaya and Soumitra SenGupta. A geometrical interpretation of parity violation in gravity with torsion. *Phys. Lett. B*, 458:8, 1999.

[46] Biswarup Mukhopadhyaya, Soumitra SenGupta, and Saurabh Sur. Spacetime torsion and parity violation: a gauge invariant formulation. *hep-th/0106236.*

[47] L. Randall and R. Sundrum. A large mass hierarchy from a small extra dimension. *Phys Rev. Lett.*, 83:3370, 1999.

[48] L. Randall and R. Sundrum. An alternative to compactification. *Phys Rev. Lett.*, 83:4690, 1999.

[49] Biswarup Mukhopadhyaya and Somasri Sen. Bulk torsion fields in theories with large extra dimensions. *hep-ph/0110308.*

[50] Soumitra SenGupta and Aninda Sinha. Neutrino mass from torsion coupled to fermion in large extra dimension. *hep-th/0110190.*

[51] Donald E. Neville. Experimental bounds on the coupling strength of torsion potentials. *Phys. Rev.*, D21(8):2075, 1980.

[52] V. W. Hughes and T. Kinoshita. Anomalous g values of the electron and muon. *Rev. Mod. Phys.*, 71:S133, 1999.

[53] H. N. Brown et. al. Precise measurement of the positive muon anomalous magnetic moment. *hep-ex/0102017 and Phys. Rev. Lett.*, 86:2227, 2001.

[54] David W. Hertzog. The BNL muon anomalous magnetic moment measurement. *hep-ex/0202024.*

[55] J. P. Miller. The muon anomaly: Experiment and theory. *hep-ex/0111036.*

[56] Gerco Onderwater. A new precise measurement of the muon anomalous magnetic moment. *hep-ex/0105036.*

[57] Masashi Hayakawa and Toichiro Kinoshita. Comment on the sign of the pseudoscalar pole contribution to the muon $g-2$. *hep-ph/0112102.*

[58] Andrzej Czarnecki Ian Blokland and Kirill Melnikov. Pion pole contribution to hadronic light-by-light scattering and muon anomalous magnetic moment. *hep-ph/0112117.*

[59] Marc Knecht and Andreas Nyffeler. Hadronic light-by-light corrections to the muon $g-2$: the pion-pole contribution. *hep-ph/0111058.*

[60] Elisabetta Pallante Johan Bijnens and Joaquim Prades. Comment on the pion pole part of the light-by-light contribution ot the muon $g-2$. *hep-ph/0112255.*

[61] T. Kinoshita, B. Nizic, and Y. Okamoto. Eighth-order QED contribution to the anomalous magnetic moment of the muon. *Phys. Rev.*, D41(2):593, 1990.

[62] Andrzej Czarnecki, Bernd Krause, and William J. Marciano. Electroweak fermion-loop contributions to the muon anomalous magnetic moment. *Phys. Rev.*, D52(5):R2619, 1995.

[63] Andrzej Czarnecki, Bernd Krause, and William J. Marciano. Electroweak corrections to the muon anomalous magnetic moment. *Phys. Rev. Lett.*, 76(18):3267, 1996.

[64] T. Kinoshita, B. Nizic, and Y. Okamoto. Hadronic contributions to the anomalous magnetic moment of the muon. *Phys. Rev.*, D31(8):2108, 1985.

[65] Jens Erler and Mingxing Luo. Hadronic loop corrections to the muon anomalous magnetic moment. *Phys. Rev. Lett.*, 87(7):071804–1, 2001.

[66] Richard T. Hammond. Upper limit on the torsion coupling constant. *Phys. Rev.*, D52(12):6918, 1995.

[67] J.A. Grifols and A. Méndez. Constraints on supersymmetric-particle masses from $(g-2)_\mu$. *Phys. Rev.*, D26(7):1809, 1982.

[68] Lisa Everett, Gordon L. Kane, Stefano Rigolin, and Lian-Tao Wang. Implications of muon $g-2$ for supersymmetry and for discovering superpartners directly. *hep-ph/0102145.*

[69] Jonathan L. Feng and Konstantin T. Matchev. Supersymmetry and the anomalous magnetic moment of the muon. *Phys. Rev.*, D64:013014, 2001.

[70] F. Ferrer and J. A. Grifols. Constraints on the mass of the superlight gravitino from the anomalous magnetic moment of the muon. *Phys. Rev.*, D56(11):7466, 1997.

[71] D.V. Nanopoulos John Ellis and Keith A. Olive. Combining the anomalous magnetic moment with other constraints on the cmssm. *hep-ph/0102331.*

[72] Terry Pilling. Threshold pion photoproduction. *Saskatchewan Accelerator Laboratory, SAL Report*, 58:168, 1998.

[73] Terry Pilling. *Threshold Pion Photoproduction.* Master's Thesis - University of Saskatchewan, Saskatoon, Saskatchewan, Canada, 1998.

[74] R. Rohm and E. Witten. The antisymmetric tensor field in superstring theory. *Annals of Physics*, 170:454–489, 1986.

[75] Martin F. Sohnius. Introducing supersymmetry. *Physics Reports*, 128(2&3):39, 1985.

[76] Joseph D. Lykken. Introduction to supersymmetry. *Fermilab-pub-96/445-T*, page 67, 1996.

[77] Keith A. Olive. Introduction to supersymmetry: Astrophysical and phenomenological constraints. *hep-ph/9911307.*

[78] S. Coleman and J. Mandula. All possible symmetries of the S matrix. *Phys. Rev.*, 159:1251, 1967.

[79] Sergei Ketov. Solitons, monopoles and duality. *hep-th/9611209.*

[80] Eric D'Hoker and D. H. Phong. The geometry of string perturbation theory. *Rev. Mod. Phys.*, 60(4):917, 1988.

[81] Edward Witten. Quantum field theory and the Jones polynomial. *IAMP Congress Lecture*, page 361, 1988.

[82] F. Gross. *Relativistic Quantum Mechanics and Field Theory.* John Wiley and Sons, Toronto, 1993.

[83] Kasper Olsen. Aspects of duality. *hep-th/9908199.*

[84] John G. Hocking and Gail S. Young. *Topology.* Addison-Wesley, Reading, Massachusetts, 1961.

[85] William Fulton. *Algebraic Topology, a first course.* Springer, New York, 1995.

[86] Morton Hammermesh. *Group Theory and Its Application to Physical Problems.* Addison-Wesley, Reading, Massachusetts, 1962.

[87] Howard Georgi. *Lie Algebras in Particle Physics.* Perseus Books, Reading, Massachusetts, 1999.

[88] Seth Warner. *Modern Algebra.* Prentice Hall, Princeton, New Jersey, 1965.

[89] Louis Shapiro. *Introduction to Abstract Algebra.* McGraw-Hill, 1975.

[90] Michael Atiyah. *K-Theory.* W. A. Benjamin, Inc., New York, 1967.

[91] Raoul Bott. *Lectures on $K(X)$.* W. A. Benjamin, Inc., New York, 1969.

[92] Norman Steenrod. *The Topology of Fibre Bundles.* Princeton University Press, Princeton, New Jersey, 1951.

[93] Phillip Griffiths and Joseph Harris. *Principles of Algebraic Geometry.* Wiley and Sons, New York, 1978.

[94] M. Nakahara. *Geometry, Topology and Physics.* Institute of Physics Publishing, London, 1990.

APPENDIX A
FIELD THEORY CONVENTIONS

A.1. General definitions

We work in the standard units, setting $\hbar = c = 1$ in which we have the relations

$$[\text{length}] = [\text{time}] = [\text{energy}]^{-1} = [\text{mass}]^{-1}.$$

The metric tensor is given in the usual Bjorken and Drell convention

$$g_{\mu\nu} = g^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \tag{A.1}$$

where we let Greek indices range from 0 to 4 and Latin indices from 1 to 3. Four-vectors are given in contravariant and covariant forms as

$$x^\mu = \left(x^0, \mathbf{x}\right) \;, \; x_\mu = g_{\mu\nu} x^\nu = \left(x^0, -\mathbf{x}\right). \tag{A.2}$$

The inner product is then naturally

$$p \cdot x = p^\mu x_\mu = g_{\mu\nu} p^\mu x^\nu = p^0 x^0 - \mathbf{p} \cdot \mathbf{x}. \tag{A.3}$$

A massive particle has

$$p^2 = p^\mu p_\mu = E^2 - |\mathbf{p}|^2 = m^2. \tag{A.4}$$

A.1.1. Pauli matrices

The Pauli spin matrices that generate $SU(2)$ can be represented by

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{A.5}$$

The Pauli matrices satisfy the algebra

$$[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k, \tag{A.6}$$

where ϵ_{ijk} is the usual Levi-Civita totally antisymmetric tensor which reverses sign under the interchange of any two indices and is equal to ± 1 for cyclic/anti-cyclic permutations of the indices. A useful relation involving this symbol is

$$\epsilon_{ijk}\epsilon_{abk} = \delta_{ia}\delta_{jb} - \delta_{ib}\delta_{ja},$$

where δ_{ij} is the Kronecker delta symbol. The Pauli matrices are traceless, have determinant equal to -1, and have anticommutation relations

$$\{\sigma_i, \sigma_j\} = 2I\delta_{ij} = Tr(\sigma_i\sigma_j), \quad \text{(A.7)}$$

where I is the unit 2×2 matrix. The completeness relation for the Pauli matrices is given by

$$\sum_i (\sigma_i)_{ab}(\sigma_i)_{cd} = 2(\delta_{bc}\delta_{ad} - \frac{1}{2}\delta_{ab}\delta_{cd}) \quad \text{(A.8)}$$

The algebra of the Lorentz group is normally written in terms of six Hermitian generators, arranged into an antisymmetric second-rank tensor, $M_{\mu\nu}$. The algebra can also be written in terms of six non-Hermitian generators, $\mathbf{J}_\pm$, with

$$J^1_\pm = \frac{1}{2}\left(M^{23} \pm iM^{01}\right) \quad \text{and cyclic } (1 \to 2 \to 3 \to 1). \quad \text{(A.9)}$$

The algebra then becomes

$$\begin{aligned} \left[J^1_\pm, J^2_\pm\right] &= iJ^3_\pm \quad \text{and cyclic} \\ [\mathbf{J}_\pm, \mathbf{J}_\pm] &= 0. \end{aligned} \quad \text{(A.10)}$$

Finite dimensional, irreducible representations are classified by pairs of half-integer numbers (j_+, j_-) which are taken from the eigenvalues $j_\pm(j_\pm + 1)$ of the two Casimir operators $(\mathbf{J}_\pm)^2$. The algebra is not that of $SU(2) \times SU(2)$ because the J's are not Hermitian. The J's can be represented by finite dimensional Hermitian matrices, but the $M_{\mu\nu}$'s cannot. For example, the representation, $(\frac{1}{2}, 0)$, has

$$r(\mathbf{J}_+) = \frac{1}{2}\boldsymbol{\sigma}; \quad r(\mathbf{J}_-) = 0.$$

Hence

$$r(M^{0i}) = \frac{1}{2}\sigma^{0i} = -\frac{i}{2}\sigma^i; \quad r(M^{12}) = \frac{1}{2}\sigma^{12} = \frac{1}{2}\sigma^3 \text{ and cyclic.}$$

The group generated by these matrices is $SL(2, \mathbb{C})$. For the representation, $(0, \frac{1}{2})$, we have

$$r(\mathbf{J}_-) = \frac{1}{2}\boldsymbol{\sigma}; \quad r(\mathbf{J}_+) = 0.$$

Hence

$$r(M^{0i}) = \frac{1}{2}\overline{\sigma}^{0i} = \frac{i}{2}\sigma^i; \quad r(M^{12}) = \frac{1}{2}\overline{\sigma}^{12} = \frac{1}{2}\sigma^3 \text{ and cyclic.}$$

We see that, in either case, the generators, $\mathbf{L}$, of the proper rotations are represented by $\frac{1}{2}\boldsymbol{\sigma}$.

Since $\overline{\sigma}^{\mu\nu} = (\sigma^{\mu\nu})^\dagger$, it is appropriate to assign index types, $(\sigma^{\mu\nu})_\alpha^{\ \beta}$ and $(\overline{\sigma}^{\mu\nu})_{\dot{\alpha}}^{\ \dot{\beta}}$, to these representations. We also have the self-duality relations,

$$\sigma_{\mu\nu} = \frac{i}{2}\epsilon_{\mu\nu\rho\lambda}\sigma^{\rho\lambda}; \quad \overline{\sigma}_{\mu\nu} = -\frac{i}{2}\epsilon_{\mu\nu\rho\lambda}\overline{\sigma}^{\rho\lambda}, \quad \text{(A.11)}$$

which reflect the fact that there are only three, rather than six, linearly independent traceless 2×2 matrices.

A.2. Spinor definitions and formulae

A.2.1. Gamma matrices

Dirac representation

The gamma matrices that act on the Dirac spinors to couple spin angular momenta are given in the Dirac representation as the following 4×4 matrices

$$\gamma^0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \gamma^i = \begin{pmatrix} 0 & \boldsymbol{\sigma}^i \\ -\boldsymbol{\sigma}^i & 0 \end{pmatrix} \tag{A.12}$$

and the axial matrix

$$\gamma^5 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} = i\gamma^0\gamma^1\gamma^2\gamma^3. \tag{A.13}$$

Weyl representation

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \overline{\sigma}^\mu & 0 \end{pmatrix}, \tag{A.14}$$

where $\sigma^\mu = (I, \boldsymbol{\sigma})$ and $\overline{\sigma}^\mu = (I, -\boldsymbol{\sigma}) = \sigma_\mu$, and then

$$\gamma^5 = \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix} = i\gamma^0\gamma^1\gamma^2\gamma^3. \tag{A.15}$$

In the Weyl representation, we see that the upper two components of a Dirac spinor have left chirality while the lower two components have right chirality.

$$\begin{aligned} \psi_D &= \psi_L + \psi_R \\ \psi_L &= \frac{1}{2}(1 - \gamma_5)\psi_D \\ \psi_R &= \frac{1}{2}(1 + \gamma_5)\psi_D \end{aligned} \tag{A.16}$$

The following relations satisfied by the γ-matrices will be found useful

$$\begin{aligned}
\gamma^5\gamma^\mu &= -\gamma^\mu\gamma^5 \\
\{\gamma^\mu, \gamma^\nu\} &= \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu} \\
\left(\gamma^5\right)^2 &= 1 \\
\gamma^\mu\gamma_\mu &= 4 \\
\gamma^\mu\gamma^\nu\gamma_\mu &= -2\gamma^\nu \\
\gamma^\mu\gamma^\nu\gamma^\lambda\gamma_\mu &= 4g^{\mu\nu} \\
Tr\{\not a \not b\} &= 4a\cdot b \\
Tr\{\not a \not b \not c \not d\} &= 4\left(a\cdot b\, c\cdot d - a\cdot c\, b\cdot d + a\cdot d\, b\cdot c\right) \\
Tr\{\gamma^5 \not a \not b\} &= 0 \\
Tr\{\gamma^5 \not a \not b \not c \not d\} &= 4i\epsilon_{\mu\nu\lambda\rho}a^\mu b^\nu c^\lambda d^\rho \\
\not a \not b &= a\cdot b - 2i\sigma_{\mu\nu}a^\mu b^\nu \\
Tr\{\gamma^{\mu_1\cdots\mu_n}\} &= 0; \quad n = \text{odd},
\end{aligned} \tag{A.17}$$

where the slash indicates contraction with the γ-matrix (i.e., $\not a = \gamma\cdot a$) and

$$\sigma_{\mu\nu} = \frac{i}{4}\left[\gamma_\mu, \gamma_\nu\right]. \tag{A.18}$$

The Dirac spinors are the following 4×2 matrices, $u(\mathbf{p}, s)$ for the fermions and $v(\mathbf{p}, s)$ for the anti-fermions.

$$\begin{aligned}
u(\mathbf{p}, s) &= \sqrt{p^0+m}\begin{pmatrix} I \\ \boldsymbol{\sigma}\cdot\frac{\mathbf{p}}{p^0+m} \end{pmatrix}\chi^{(s)} \\
v(\mathbf{p}, s) &= \sqrt{p^0+m}\begin{pmatrix} \boldsymbol{\sigma}\cdot\frac{\mathbf{p}}{p^0+m} \\ I \end{pmatrix}\eta^{(s)},
\end{aligned} \tag{A.19}$$

where $\chi^{(s)}$ and the $\eta^{(s)}$ are vectors in spin space given by

$$\begin{aligned}
&\chi^{(\frac{1}{2})} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \; \chi^{(-\frac{1}{2})} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
&\eta^{(\frac{1}{2})} = -i\sigma_2\chi^{(-\frac{1}{2})} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \; \eta^{(-\frac{1}{2})} = -i\sigma_2\chi^{(\frac{1}{2})} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}.
\end{aligned} \tag{A.20}$$

We also have the conjugate spinors

$$\begin{aligned}
\overline{u}(\mathbf{p}, s) &= u^\dagger(\mathbf{p}, s)\gamma^0 \\
\overline{v}(\mathbf{p}, s) &= v^\dagger(\mathbf{p}, s)\gamma^0
\end{aligned} \tag{A.21}$$

and the spin sum relations

$$\sum_s u(\mathbf{p},s)\overline{u}(\mathbf{p},s) = \frac{\not{p}+M}{2M} = \Lambda_+(p)$$
$$\sum_s v(\mathbf{p},s)\overline{v}(\mathbf{p},s) = \frac{\not{p}-M}{2M} = \Lambda_-(p). \tag{A.22}$$

The on-shell spinors satisfy the Dirac equation

$$\begin{aligned} (\not{p}-m)\,u(\mathbf{p}) = 0, \quad (\not{p}+m)\,v(\mathbf{p}) = 0 \\ \overline{u}(\mathbf{p})\,(\not{p}-m) = 0, \quad \overline{v}(\mathbf{p})\,(\not{p}+m) = 0. \end{aligned} \tag{A.23}$$

Note that the matrix elements formed out of diagrams with initial and final nucleon spinors will always be of the form

$$\overline{u}(P_f,s_f)\Gamma u(P_i,s_i) = \chi^\dagger_{s_f} M(P_f,P_i)\chi_s \tag{A.24}$$

with the Γ containing one of $1, \gamma_5, \gamma^\mu, \gamma^\mu\gamma_5$, or $\sigma^{\mu\nu}$. One can then write down the $M(P_f,P_i)$ by substituting the following two-component forms

$$\begin{aligned}
1 &\to N'N\left[1 - \frac{(\boldsymbol{\sigma}\cdot\mathbf{P}_f)(\boldsymbol{\sigma}\cdot\mathbf{P}_i)}{(E_f+M)(E_i+M)}\right], \\
\gamma_5 &\to N'N\left[\frac{(\boldsymbol{\sigma}\cdot\mathbf{P}_i)}{(E_i+M)} - \frac{(\boldsymbol{\sigma}\cdot\mathbf{P}_f)}{(E_f+M)}\right], \\
\gamma^0 &\to N'N\left[1 + \frac{(\boldsymbol{\sigma}\cdot\mathbf{P}_f)(\boldsymbol{\sigma}\cdot\mathbf{P}_i)}{(E_f+M)(E_i+M)}\right], \\
\boldsymbol{\gamma} &\to N'N\left[\boldsymbol{\sigma}\frac{(\boldsymbol{\sigma}\cdot\mathbf{P}_i)}{(E_i+M)} + \frac{(\boldsymbol{\sigma}\cdot\mathbf{P}_f)}{(E_f+M)}\boldsymbol{\sigma}\right], \\
\gamma^0\gamma_5 &\to N'N\left[\frac{(\boldsymbol{\sigma}\cdot\mathbf{P}_i)}{(E_i+M)} + \frac{(\boldsymbol{\sigma}\cdot\mathbf{P}_f)}{(E_f+M)}\right], \\
\boldsymbol{\gamma}\gamma_5 &\to N'N\left[\frac{(\boldsymbol{\sigma}\cdot\mathbf{P}_i)}{(E_i+M)} + \frac{(\boldsymbol{\sigma}\cdot\mathbf{P}_f)}{(E_f+M)}\right], \\
\sigma^{0j} &\to N'Ni\left[\sigma^j\frac{(\boldsymbol{\sigma}\cdot\mathbf{P}_i)}{(E_i+M)} - \frac{(\boldsymbol{\sigma}\cdot\mathbf{P}_f)}{(E_f+M)}\sigma^j\right], \\
\sigma^{ij} &\to N'N\left[\sigma^k - \frac{(\boldsymbol{\sigma}\cdot\mathbf{P}_f)}{(E_f+M)}\sigma^k\frac{(\boldsymbol{\sigma}\cdot\mathbf{P}_i)}{(E_i+M)}\right]\epsilon_{ijk}.
\end{aligned} \tag{A.25}$$

The photon polarization vectors, ϵ^μ and $\epsilon^{\mu*}$, give the following polarization sum

$$\sum_\lambda \epsilon^{\lambda*}\cdot\epsilon^\lambda = -2. \tag{A.26}$$

Note also the product $\epsilon^*_\mu\epsilon^\mu = 1$.

A.2.2. Weyl, Dirac, and Majorana spinors

Two component Weyl spinors, ψ_α and $\overline{\chi}^{\dot{\alpha}}$, obey the following relations

$$\overline{\psi}_{\dot{\alpha}} = (\psi^{\alpha})^* \quad \text{and} \quad \chi^{\alpha} = (\overline{\psi}^{\dot{\alpha}})^*. \tag{A.27}$$

Indices are raised and lowered as

$$\begin{aligned} \chi_\alpha &= \epsilon_{\alpha\beta}\chi^{\beta} \quad \chi^{\alpha} = \epsilon^{\alpha\beta}\chi_{\beta} \\ \overline{\psi}^{\dot{\alpha}} &= \epsilon^{\dot{\alpha}\dot{\beta}}\overline{\psi}_{\dot{\beta}} \quad \overline{\psi}_{\dot{\alpha}} = \epsilon_{\dot{\alpha}\dot{\beta}}\overline{\psi}^{\dot{\beta}}, \end{aligned} \tag{A.28}$$

where

$$\begin{aligned} \epsilon^{\alpha\beta} &= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \epsilon^{\dot{\alpha}\dot{\beta}} \\ \epsilon_{\alpha\beta} &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \epsilon_{\dot{\alpha}\dot{\beta}}. \end{aligned} \tag{A.29}$$

A Dirac spinor and its charge conjugate are defined in terms of Weyl spinors as

$$\psi_D = \begin{pmatrix} \psi_\alpha \\ \overline{\chi}^{\dot{\alpha}} \end{pmatrix} \quad \psi_D^c = \begin{pmatrix} \chi_\alpha \\ \overline{\psi}^{\dot{\alpha}} \end{pmatrix}, \tag{A.30}$$

which has four independant components, two for each Weyl spinor.

A Majorana spinor is defined as one which is equal to its charge conjugate,

$$\psi_M = \begin{pmatrix} \psi_\alpha \\ \overline{\psi}^{\dot{\alpha}} \end{pmatrix} = \psi_M^c, \tag{A.31}$$

so that it has only two independant components corresponding to a single Weyl spinor.

We can always construct a Majorana spinor from a Weyl spinor in the obvious way, and we can construct a Dirac spinor from two Majorana spinors as

$$\psi_D = \psi_{M1} + i\psi_{M2}, \tag{A.32}$$

where

$$\begin{aligned} \psi_{M1} &= \frac{1}{2}\left(\psi_D + \psi_D^c\right) \\ \psi_{M2} &= \frac{1}{2}\left(\psi_D - \psi_D^c\right) \end{aligned} \tag{A.33}$$

A.2.3. Spinor indices

We define the contraction of indices on combinations of Weyl spinors as follows

$$\begin{aligned} \chi^{\alpha}\psi_{\alpha} &= -\chi_{\alpha}\psi^{\alpha} \equiv \chi\psi = \psi\chi \\ \overline{\chi}_{\dot{\alpha}}\overline{\psi}^{\dot{\alpha}} &= -\overline{\chi}^{\dot{\alpha}}\overline{\psi}_{\dot{\alpha}} \equiv \overline{\chi}\overline{\psi} = \overline{\psi}\overline{\chi}. \end{aligned} \tag{A.34}$$

These products are Lorentz invariant quantities. Covariant four-vectors are constructed as

$$\overline{\chi}_{\dot{\alpha}} \left(\overline{\sigma}^{\mu}\right)^{\dot{\alpha}\alpha} \psi_{\alpha} = \overline{\chi}^{\dot{\alpha}} \left(\sigma^{\mu}\right)_{\alpha\dot{\alpha}} \psi^{\alpha} \tag{A.35}$$

and

$$\chi^{\alpha} \left(\sigma^{\mu}\right)_{\alpha\dot{\alpha}} \overline{\psi}^{\dot{\alpha}} = \chi_{\alpha} \left(\overline{\sigma}^{\mu}\right)^{\dot{\alpha}\alpha} \overline{\psi}_{\dot{\alpha}}. \tag{A.36}$$

Tensors are constructed from

$$\chi^{\alpha} \left(\sigma^{\mu\nu}\right)_{\alpha}^{\beta} \psi_{\beta} = \chi_{\alpha} \left(\sigma^{\mu\nu}\right)_{\beta}^{\alpha} \psi^{\beta} \tag{A.37}$$

and

$$\overline{\chi}_{\dot{\alpha}} \left(\overline{\sigma}^{\mu\nu}\right)_{\dot{\beta}}^{\dot{\alpha}} \overline{\psi}^{\dot{\beta}} = \overline{\chi}^{\dot{\alpha}} \left(\overline{\sigma}^{\mu\nu}\right)_{\dot{\alpha}}^{\dot{\beta}} \overline{\psi}_{\dot{\beta}}. \tag{A.38}$$

Notice the way that the indices are contracted on the right hand sides of these equations. We can put them into the natural order by interchanging the two spinors. The spinors are anti-commuting Grassmann variables so that interchanging gives the following identities (remembering our index contraction conventions):

$$\begin{aligned} \chi^{\alpha}\psi_{\alpha} &= \psi^{\alpha}\chi_{\alpha} \\ \overline{\chi}_{\dot{\alpha}}\overline{\psi}^{\dot{\alpha}} &= \overline{\psi}_{\dot{\alpha}}\overline{\chi}^{\dot{\alpha}} \\ \overline{\chi}_{\dot{\alpha}} \left(\overline{\sigma}^{\mu}\right)^{\alpha\dot{\alpha}} \psi_{\alpha} &= -\psi^{\alpha} \left(\sigma^{\mu}\right)_{\alpha\dot{\alpha}} \overline{\chi}^{\dot{\alpha}} \\ \chi^{\alpha} \left(\sigma^{\mu\nu}\right)_{\alpha}^{\beta} \psi^{\beta} &= -\psi^{\alpha} \left(\sigma^{\mu\nu}\right)_{\alpha}^{\beta} \chi_{\beta} \\ \overline{\chi}^{\alpha} \left(\overline{\sigma}^{\mu\nu}\right)_{\dot{\beta}}^{\dot{\alpha}} \overline{\psi}^{\dot{\beta}} &= -\overline{\psi}_{\dot{\alpha}} \left(\overline{\sigma}^{\mu\nu}\right)_{\dot{\beta}}^{\dot{\alpha}} \overline{\chi}^{\dot{\beta}}. \end{aligned} \tag{A.39}$$

The Hermitian conjugate of a product is defined as

$$\left(\chi\psi\right)^{\dagger} = \psi^{\dagger}\chi^{\dagger} = \overline{\psi}\overline{\chi} = \overline{\chi}\overline{\psi} \tag{A.40}$$

since

$$\left(\psi_{\alpha}\right)^{\dagger} = \left(\psi_{\alpha}\right)^{*} = \overline{\psi}_{\dot{\alpha}} \tag{A.41}$$

and

$$\left(\chi^{\alpha}\right)^{\dagger} = \left(\chi^{\alpha}\right)^{*} = \overline{\chi}^{\dot{\alpha}}. \tag{A.42}$$

We define

$$\begin{aligned} \chi\sigma^{\mu}\overline{\psi} &= \chi^{\alpha} \left(\sigma^{\mu}\right)_{\alpha\dot{\beta}} \overline{\psi}^{\dot{\beta}} \\ \chi\sigma^{\mu\nu}\overline{\psi} &= \chi^{\alpha} \left(\sigma^{\mu\nu}\right)_{\alpha}^{\beta} \psi_{\beta} \end{aligned} \tag{A.43}$$

so that

$$\begin{aligned}\left(\chi\sigma^{\mu}\overline{\psi}\right)^{\dagger} &= \psi\sigma^{\mu}\overline{\chi} = -\overline{\chi}\overline{\sigma}^{\mu}\psi = -\left(\overline{\psi}\overline{\sigma}^{\mu}\chi\right)^{\dagger} \\ \left(\chi\sigma^{\mu\nu}\psi\right)^{\dagger} &= -\overline{\psi}\overline{\sigma}^{\mu\nu}\overline{\chi} = \overline{\chi}\overline{\sigma}^{\mu\nu}\overline{\psi} = -\left(\psi\sigma^{\mu\nu}\chi\right)^{\dagger},\end{aligned} \tag{A.44}$$

allowing us to re-write the usual Dirac covariant bilinears in terms of Weyl spinors which appear in the Dirac spinors. We write

$$\Psi = \begin{pmatrix} \psi_{\alpha} \\ \overline{\chi}^{\dot{\alpha}} \end{pmatrix} \quad \Phi = \begin{pmatrix} \phi_{\alpha} \\ \overline{\eta}^{\dot{\alpha}} \end{pmatrix} \tag{A.45}$$

in the Weyl representation. We then have

$$\begin{aligned}\overline{\Psi}\Phi &= \overline{\psi}\overline{\eta} + \chi\phi = \left(\overline{\Phi}\Psi\right)^{\dagger} \\ \overline{\Psi}\gamma_5\Phi &= \overline{\psi}\overline{\eta} - \chi\phi = -\left(\overline{\Phi}\gamma_5\Psi\right)^{\dagger} \\ \overline{\Psi}\gamma^{\mu}\Phi &= \chi\sigma^{\mu}\overline{\eta} + \overline{\psi}\overline{\sigma}^{\mu}\phi = \left(\overline{\Phi}\gamma^{\mu}\Psi\right)^{\dagger} \\ \overline{\Psi}\gamma^{\mu}\gamma_5\Phi &= \chi\sigma^{\mu}\overline{\eta} - \overline{\psi}\overline{\sigma}^{\mu}\phi = \left(\overline{\Phi}\gamma^{\mu}\gamma_5\Psi\right)^{\dagger} \\ \overline{\Psi}\Sigma^{\mu\nu}\Phi &= i\chi\sigma^{\mu\nu}\phi + i\overline{\psi}\overline{\sigma}^{\mu\nu}\overline{\eta} = \left(\overline{\Phi}\Sigma^{\mu\nu}\Psi\right)^{\dagger}.\end{aligned} \tag{A.46}$$

It is straightforward to derive the similar identities for Majorana spinors from these. Notice, for example,

$$\begin{aligned}\overline{\Psi}_M\gamma^{\mu}\Phi_M &= \psi\sigma^{\mu}\overline{\phi} + \overline{\psi}\overline{\sigma}^{\mu}\phi \\ &= \psi^{\alpha}\left(\sigma^{\mu}\right)_{\alpha\dot{\beta}}\overline{\phi}^{\dot{\beta}} + \overline{\psi}_{\dot{\alpha}}\left(\overline{\sigma}^{\mu}\right)^{\dot{\alpha}\beta}\phi_{\beta} \\ &= -\overline{\phi}^{\dot{\beta}}\left(\sigma^{\mu}\right)_{\alpha\dot{\beta}}\psi^{\alpha} - \phi_{\beta}\left(\overline{\sigma}^{\mu}\right)^{\dot{\alpha}\beta}\overline{\psi}_{\dot{\alpha}} \\ &= -\overline{\phi}_{\dot{\beta}}\left(\overline{\sigma}^{\mu}\right)^{\dot{\beta}\alpha}\psi_{\alpha} - \phi^{\beta}\left(\sigma^{\mu}\right)_{\beta\dot{\alpha}}\overline{\psi}^{\dot{\alpha}} \\ &= -\phi^{\beta}\left(\sigma^{\mu}\right)_{\beta\dot{\alpha}}\overline{\psi}^{\dot{\alpha}} - \overline{\phi}_{\dot{\beta}}\left(\overline{\sigma}^{\mu}\right)^{\dot{\beta}\alpha}\psi_{\alpha} \\ &= -\overline{\Phi}_M\gamma^{\mu}\Psi_M = -\left(\overline{\Psi}_M\gamma^{\mu}\Phi_M\right)^{\dagger}.\end{aligned} \tag{A.47}$$

A.2.4. Fierz identities

The Weyl spinors also satisfy various Fierz identities. All of these identities may be derived from the basic identity

$$\delta_{\alpha\beta}\delta_{\gamma\delta} = \frac{1}{2}\left[\delta_{\alpha\delta}\delta_{\gamma\beta} + \sigma^{i}_{\alpha\delta}\sigma^{i}_{\gamma\beta}\right] \tag{A.48}$$

which expresses the completeness of (I, σ^i) as a set of 2×2 matrices. For example,

$$\begin{aligned}\frac{1}{2}(\sigma^\mu)_{\alpha\dot{\delta}}(\overline{\sigma}_\mu)^{\dot{\gamma}\beta} &= \frac{1}{2}\left[(I,\sigma^i)_{\alpha\dot{\delta}} \cdot (I, -\sigma^i)^{\dot{\gamma}\beta}\right] \\ &= \frac{1}{2}\left[\delta^{\dot{\gamma}}_{\alpha}\delta^{\beta}_{\dot{\delta}} + (\sigma^i)^{\dot{\gamma}}_{\alpha}(\sigma^i)^{\beta}_{\dot{\delta}}\right] \\ &= \delta^{\beta}_{\alpha}\delta^{\dot{\gamma}}_{\dot{\delta}} \quad \text{from (A.48).}\end{aligned} \tag{A.49}$$

This relation leads to the following identity,

$$(\theta\phi)(\overline{\chi}\overline{\eta}) = -\frac{1}{2}(\theta\sigma^\mu\overline{\eta})(\overline{\chi}\overline{\sigma}_\mu\phi), \tag{A.50}$$

through equation (A.49) with the minus sign coming from anti-commutation. Simiarly, we can write

$$\delta^{\beta}_{\alpha}\delta^{\delta}_{\gamma} = \frac{1}{2}\left[\delta^{\delta}_{\alpha}\delta^{\beta}_{\gamma} - (\sigma^{\mu\nu})^{\delta}_{\alpha}(\sigma_{\mu\nu})^{\beta}_{\gamma}\right] \tag{A.51}$$

which allows us to deduce

$$(\theta\phi)(\chi\eta) = -\frac{1}{2}\left[(\theta\eta)(\chi\phi) - (\theta\sigma^{\mu\nu}\eta)(\chi\sigma_{\mu\nu}\phi)\right]. \tag{A.52}$$

In the case where $\theta = \eta$, we find

$$(\theta\phi)(\chi\theta) = -\frac{1}{2}(\theta\theta)(\chi\phi) = (\theta\phi)(\theta\chi) \tag{A.53}$$

since $\theta\sigma_{\mu\nu}\theta = 0$. Also from equation (A.51), contracting both sides with $\theta^\alpha\theta_\delta$, we have

$$\theta^\beta\theta_\gamma = \frac{1}{2}(\theta\theta)\delta^{\beta}_{\gamma} \tag{A.54}$$

which is one of the most useful identities for doing calculations. Other useful, related identities are

$$\begin{aligned}\theta^\alpha\theta^\beta &= -\frac{1}{2}(\theta\theta)\epsilon^{\alpha\beta} \\ \overline{\theta}^{\dot{\alpha}}\overline{\theta}^{\dot{\beta}} &= \frac{1}{2}(\overline{\theta}\overline{\theta})\epsilon^{\dot{\alpha}\dot{\beta}}\end{aligned} \tag{A.55}$$

and

$$\begin{aligned}(\theta)^2 &= \theta^\alpha\theta_\alpha = -2\theta^1\theta^2 \\ (\overline{\theta})^2 &= \overline{\theta}_{\dot{\alpha}}\overline{\theta}^{\dot{\alpha}} = 2\overline{\theta}^{\dot{1}}\overline{\theta}^{\dot{2}}\end{aligned} \tag{A.56}$$

Note that we can re-write equations (A.49) and (A.51) as

$$1 \times 1 = \frac{1}{2}\sigma^\mu \otimes \overline{\sigma}_\mu \tag{A.57}$$

and

$$1 \times 1 = \frac{1}{2}\left[1 \otimes 1 - \sigma^{\mu\nu} \otimes \overline{\sigma}_{\mu\nu}\right]. \tag{A.58}$$

We have the following useful identities involving the σ's:

$$\begin{aligned}
\sigma^{\mu\nu}\sigma^{\lambda} &= -\frac{1}{2}\left[\eta^{\mu\lambda}\sigma^{\nu} - \eta^{\nu\lambda}\sigma^{\mu} + i\epsilon^{\mu\nu\lambda\rho}\sigma_{\rho}\right] \\
\overline{\sigma}^{\mu}\sigma^{\nu\lambda} &= \frac{1}{2}\left[\eta^{\mu\nu}\overline{\sigma}^{\lambda} - \eta^{\mu\lambda}\overline{\sigma}^{\nu} + i\epsilon^{\mu\nu\lambda\rho}\overline{\sigma}_{\rho}\right] \\
\sigma^{\mu\nu}\sigma^{\kappa\lambda} &= -\frac{1}{4}\Big[\eta^{\mu\kappa}\eta^{\nu\lambda} - \eta^{\mu\lambda}\eta^{\nu\kappa} + i\epsilon^{\mu\nu\kappa\lambda} \\
&\qquad + 2\left(\eta^{\mu\kappa}\sigma^{\nu\lambda} + \eta^{\nu\lambda}\sigma^{\mu\kappa} - \eta^{\mu\lambda}\sigma^{\nu\kappa} - \eta^{\nu\kappa}\sigma^{\mu\lambda}\right)\Big].
\end{aligned} \tag{A.59}$$

Using these identities, we can derive the following complete set of Fierz identities:

$$(\theta\phi)(\overline{\chi}\overline{\eta}) = -\frac{1}{2}(\theta\sigma^{\mu}\overline{\eta})(\overline{\chi}\overline{\sigma}_{\mu}\phi) \tag{A.60}$$

$$(\theta\phi)(\chi\eta) = -\frac{1}{2}\left[(\theta\eta)(\chi\phi) - (\theta\sigma^{\mu\nu}\eta)(\chi\sigma_{\mu\nu}\phi)\right] \tag{A.61}$$

$$(\theta\phi)(\chi\sigma^{\mu}\overline{\eta}) = -\frac{1}{2}\left[(\theta\sigma^{\mu}\overline{\eta})(\chi\phi) + 2(\theta\sigma_{\nu}\overline{\eta})(\chi\sigma^{\mu\nu}\phi)\right] \tag{A.62}$$

$$(\theta\phi)(\overline{\chi}\overline{\sigma}^{\mu}\eta) = -\frac{1}{2}\left[(\theta\eta)(\overline{\chi}\overline{\sigma}^{\mu}\phi) - 2(\theta\sigma^{\mu\nu}\eta)(\overline{\chi}\overline{\sigma}_{\nu}\phi)\right] \tag{A.63}$$

$$\begin{aligned}
(\theta\sigma^{\mu}\overline{\phi})(\chi\sigma^{\nu}\overline{\eta}) &= -\frac{1}{2}\left[(\theta\sigma^{\mu}\overline{\eta})(\chi\sigma^{\nu}\overline{\phi}) + (\theta\sigma^{\nu}\overline{\eta})(\chi\sigma_{\mu}\overline{\phi})\right. \\
&\quad \left. - \eta^{\mu\nu}(\theta\sigma^{\lambda}\overline{\eta})(\chi\sigma_{\lambda}\overline{\phi}) - i\epsilon^{\mu\nu\kappa\lambda}(\theta\sigma_{\kappa}\overline{\eta})(\chi\sigma_{\lambda}\overline{\phi})\right]
\end{aligned} \tag{A.64}$$

$$\begin{aligned}
(\theta\sigma^{\mu}\overline{\phi})(\overline{\chi}\overline{\sigma}^{\nu}\eta) &= -\frac{1}{2}\left[\eta^{\mu\nu}(\theta\eta)(\overline{\chi}\overline{\phi}) + 2(\theta\sigma^{\mu\nu}\eta)(\overline{\chi}\overline{\phi})\right. \\
&\quad \left. - 2(\theta\eta)(\overline{\chi}\overline{\sigma}^{\mu\nu}\overline{\phi}) - 4(\theta\sigma^{\nu\lambda}\eta)(\overline{\chi}\overline{\sigma}^{\mu}_{\lambda}\overline{\phi})\right]
\end{aligned} \tag{A.65}$$

$$\begin{aligned}
(\theta\phi)(\chi\sigma^{\mu\nu}\eta) &= -\frac{1}{2}\left[(\theta\eta)(\chi\sigma^{\mu\nu}\phi) + (\theta\sigma^{\mu\nu}\eta)(\chi\phi)\right. \\
&\quad \left. - (\theta\sigma^{\mu\lambda}\eta)(\chi\sigma^{\nu}_{\lambda}\phi) + (\theta\sigma^{\nu\lambda}\eta)(\chi\sigma^{\mu}_{\lambda}\phi)\right]
\end{aligned} \tag{A.66}$$

$$\begin{aligned}
(\theta\phi)(\overline{\chi}\overline{\sigma}^{\mu\nu}\overline{\eta}) &= -\frac{1}{4}\left[(\theta\sigma^{\nu}\overline{\eta})(\overline{\chi}\overline{\sigma}^{\mu}\phi) - (\theta\sigma^{\mu}\overline{\eta})(\overline{\chi}\overline{\sigma}^{\nu}\phi)\right. \\
&\quad \left. + i\epsilon^{\mu\nu\kappa\lambda}(\theta\sigma_{\kappa}\overline{\eta})(\overline{\chi}\overline{\sigma}_{\lambda}\phi)\right]
\end{aligned} \tag{A.67}$$

$$(\theta\sigma^{\mu\nu}\phi)(\chi\sigma^{\lambda}\overline{\eta}) = \frac{1}{4}\left[\eta^{\mu\lambda}(\theta\sigma^{\nu}\overline{\eta})(\chi\phi) - \eta^{\nu\lambda}(\theta\sigma^{\mu}\overline{\eta})(\chi\phi) + i\epsilon^{\mu\nu\lambda\rho}(\theta\sigma_{\rho}\overline{\eta})(\chi\phi)\right] + \frac{1}{2}\left[(\theta\sigma^{\nu}\overline{\eta})(\chi\sigma^{\lambda\mu}\phi) - (\theta\sigma^{\mu}\overline{\eta})(\chi\sigma^{\lambda\nu}\phi) + i\epsilon^{\mu\nu\kappa\rho}(\theta\sigma_{\kappa}\overline{\eta})(\chi\sigma^{\lambda}_{\rho}\phi)\right] \tag{A.68}$$

$$(\theta\sigma^{\mu\nu}\phi)(\overline{\chi}\overline{\sigma}^{\lambda}\eta) = -\frac{1}{4}\left[\eta^{\mu\lambda}(\theta\eta)(\overline{\chi}\overline{\sigma}^{\nu}\phi) - \eta^{\nu\lambda}(\theta\eta)(\overline{\chi}\overline{\sigma}^{\mu}\phi) + i\epsilon^{\mu\nu\lambda\rho}(\theta\eta)(\overline{\chi}\overline{\sigma}_{\rho}\phi)\right] - \frac{1}{2}\left[(\theta\sigma^{\mu\lambda}\eta)(\overline{\chi}\overline{\sigma}^{\nu}\phi) - (\theta\sigma^{\nu\lambda}\eta)(\overline{\chi}\overline{\sigma}^{\mu}\phi) + i\epsilon^{\mu\nu\kappa\rho}(\theta\sigma^{\lambda}_{\kappa}\eta)(\overline{\chi}\overline{\sigma}_{\rho}\phi)\right] \tag{A.69}$$

$$\begin{aligned}(\theta\sigma^{\mu\nu}\phi)(\overline{\chi}\sigma^{\kappa\lambda}\overline{\eta}) = -\frac{1}{8}\big[&(\eta^{\mu\lambda}\eta^{\nu\kappa} - \eta^{\mu\kappa}\eta^{\nu\lambda})(\theta\sigma^{\rho}\overline{\eta})(\overline{\chi}\overline{\sigma}_{\rho}\phi)\\ &+\eta^{\mu\kappa}(\theta\sigma^{\nu}\eta)(\overline{\chi}\overline{\sigma}^{\lambda}\phi) + \eta^{\mu\kappa}(\theta\sigma^{\lambda}\eta)(\overline{\chi}\overline{\sigma}^{\nu}\phi)\\ &-\eta^{\nu\kappa}(\theta\sigma^{\mu}\eta)(\overline{\chi}\overline{\sigma}^{\lambda}\phi) + \eta^{\nu\kappa}(\theta\sigma^{\lambda}\eta)(\overline{\chi}\overline{\sigma}^{\mu}\phi)\\ &+\eta^{\nu\lambda}(\theta\sigma^{\mu}\eta)(\overline{\chi}\overline{\sigma}^{\kappa}\phi) + \eta^{\nu\lambda}(\theta\sigma^{\kappa}\eta)(\overline{\chi}\overline{\sigma}^{\mu}\phi)\\ &-\eta^{\mu\lambda}(\theta\sigma^{\nu}\eta)(\overline{\chi}\overline{\sigma}^{\kappa}\phi) + \eta^{\mu\lambda}(\theta\sigma^{\kappa}\eta)(\overline{\chi}\overline{\sigma}^{\nu}\phi)\\ &+i\epsilon^{\mu\nu\kappa\rho}(\theta\sigma_{\rho}\eta)(\overline{\chi}\overline{\sigma}^{\lambda}\phi) - i\epsilon^{\mu\nu\lambda\rho}(\theta\sigma_{\rho}\eta)(\overline{\chi}\overline{\sigma}^{\kappa}\phi)\\ &- i\epsilon^{\kappa\lambda\mu\rho}(\theta\sigma^{\nu}\eta)(\overline{\chi}\overline{\sigma}_{\rho}\phi) + i\epsilon^{\kappa\lambda\nu\rho}(\theta\sigma^{\mu}\eta)(\overline{\chi}\overline{\sigma}_{\rho}\phi)\big].\end{aligned} \tag{A.70}$$

A.3. Free field propagators

The mesons are described by the Klein-Gordon lagrangian

$$\mathcal{L} = \frac{1}{2}\left[\partial^{\mu}\vec{\phi}\cdot\partial_{\mu}\vec{\phi} - m^2\vec{\phi}\cdot\vec{\phi}\right]. \tag{A.71}$$

Application of the Euler-Lagrange equations gives the following equation of motion

$$\left(\partial^{\mu}\partial_{\mu} + m^2\right)\phi = 0 \tag{A.72}$$

which is the Klein-Gordon equation. The propagator, which propagates the particle between two points in spacetime is defined as the vacuum expectation value of the time-ordered product

$$\langle 0|\, T(\phi^i(x)\phi^j(y))\,|0\rangle = i\Delta_F^{ij}(x-y), \tag{A.73}$$

where

$$\Delta_F^{ij}(x-y) = \int \frac{d^4q}{(2\pi)^4}\frac{\delta^{ij}e^{-iq\cdot(x-y)}}{q^2 - m^2 + i\epsilon}. \tag{A.74}$$

The small complex piece, $i\epsilon$, added in the denominator moves it into the complex plane to avoid the singularity at $q^2 = m^2$. This addition is what is meant by "off-mass-shell." The expression is computed using the residue theorem in complex analysis, where one integrates around a closed curve in the complex plane surrounding the singularity and computes the residue.

Photons are descibed by the following lagrangian, which is the most general lagrangian consistent with $U(1)$ gauge invariance

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \tag{A.75}$$

where $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. The photon propagator is slightly different from the meson one due to the presence of the polarization vectors and the fact that the wavefunction is a vector field. It is given by

$$\langle 0 | T(A^\mu(x) A^\nu(y)) | 0 \rangle = i D_F^{\mu\nu}(x-y), \tag{A.76}$$

where

$$D^{\mu\nu}(x-y) = \int \frac{d^4k}{(2\pi)^4} \frac{e^{-ik\cdot(x-y)}}{-k^2 - i\epsilon} \left[g^{\mu\nu} - \frac{k^\mu k^\nu}{k^2}(1-\alpha) \right] \tag{A.77}$$

and α is the gauge parameter. In what follows, we will use the Feynman gauge $\alpha = 1$.

The nucleons are fermions and are described by the Dirac lagrangian

$$\mathcal{L} = \overline{\psi} \left[\frac{i}{2} \gamma^\mu \overleftrightarrow{\partial}_\mu - m \right] \psi \tag{A.78}$$

which gives the Dirac equation as the equation of motion describing the free particle

$$(i\gamma^\mu \partial_\mu - m)\psi = (\gamma^\mu p_\mu - m)\psi = 0. \tag{A.79}$$

The propagator, defined as the renormalized 2-point function, is

$$\langle 0 | T(\psi_\alpha(x) \overline{\psi}_\beta(y)) | 0 \rangle = i S_{\alpha\beta}(x-y), \tag{A.80}$$

where

$$S^{\alpha\beta}(x-y) = \int \frac{d^4p}{(2\pi)^4} e^{-ip\cdot(x-y)} \frac{(\not{p} + m)^{\alpha\beta}}{p^2 - m^2 + i\epsilon}. \tag{A.81}$$

The interaction lagrangian and the propagator for a spin-$\frac{3}{2}$ particle (We use the $\Delta(1232)$ resonance as an example.) are given by

$$\mathcal{L}_\Delta = \mathcal{L}_{\pi N \Delta} + \mathcal{L}_{\gamma N \Delta}^{g1} + \mathcal{L}_{\gamma N \Delta}^{g2}, \tag{A.82}$$

where the individual interaction lagrangians are given by

$$\begin{aligned} \mathcal{L}_{\pi N \Delta} &= \frac{f_{\pi N \Delta}}{m_\pi} \overline{\Delta}^\mu T_a \Theta_{\mu\nu}(Z) N \partial^\nu \pi_a + H.c. \\ \mathcal{L}_{\gamma N \Delta}^{g1} &= \frac{ieg_1}{2M} \overline{\Delta}^\rho \Theta_{\rho\alpha}(Y) \gamma_\nu \gamma_5 T_3 N F^{\alpha\nu} + H.c. \\ \mathcal{L}_{\gamma N \Delta}^{g2} &= \frac{eg_2}{4M^2} \overline{\Delta}^\rho \Theta_{\rho\nu}(X) \gamma_5 T_3 (\partial_\alpha N) F^{\nu\alpha} + H.c., \end{aligned} \tag{A.83}$$

where now Δ^μ is the Lorentz vector, Dirac spinor, isovector field describing the $\Delta(1232)$

resonance; T_a is the isospin $\frac{1}{2} \to \frac{3}{2}$ transition operator; and $\Theta_{\rho\nu}(X) = g_{\rho\nu} + [\frac{1}{2}(1 + 4X)A + X]\gamma_\rho\gamma_\nu$ for off-shell parameter X. We choose the arbitrary parameter $A = -1$ for algebraic convenience.[1]

We have given two separate electromagnetic lagrangians for photon-resonance interactions. The reason for this can be seen through multipole analysis as follows.

The Δ field is of spin-$\frac{3}{2}$ and even parity, meaning that it can only be excited by an $M1$ or an $E2$ photon[2]. The g_1 and g_2 couplings correspond to these two possible transitions to the Δ. The final state pion-nucleon system produced through the resonance excitation must, therefore, consist of M_{1+} and E_{1+} multipoles.

The propagator for the resonance field is given by

$$P^{\mu\nu}(\not{R}) = \frac{\not{R} + M_\Delta}{R^2 - M_\Delta^2}\left[g^{\mu\nu} - \frac{1}{3}\gamma^\mu\gamma^\nu - \frac{2R^\mu R^\nu}{3M_\Delta^2} + \frac{R^\mu\gamma^\nu - R^\nu\gamma^\mu}{3M_\Delta}\right] \tag{A.84}$$

for resonance 4-momentum R^μ. The spin-$\frac{3}{2}$ lagrangian contains $\theta_{\mu\nu}$ defined by

$$\theta_{\mu\nu} = g_{\mu\nu} + \left[\frac{1}{2}(1 + 4X)A + X\right]\gamma_\mu\gamma_\nu, \tag{A.85}$$

where A is an arbitrary parameter subject to the restriction that $A \neq -\frac{1}{2}$. This parameter drops out of the observables and is usually chosen as $A = -1$ for calculational ease.

T_a is the spin-$\frac{1}{2} \to$ spin-$\frac{3}{2}$ transition operator, which is a 2×4 matrix defined by its matrix element

$$\left\langle \frac{3}{2}\lambda_\Delta \middle| T_\lambda^+ \middle| \lambda \frac{1}{2}\lambda_N \right\rangle = C^{\frac{3}{2}\frac{1}{2}\frac{1}{2}}_{\lambda_\Delta \lambda \lambda_N}, \tag{A.86}$$

where $C^{I_1 I_2 I}_{m_1 m_2 m}$ is a Clebsch-Gordon coefficient. Another important relation is

$$\sum_{\lambda_\Delta} T_b \left| \frac{3}{2}\lambda_\Delta \right\rangle \left\langle \frac{3}{2}\lambda_\Delta \right| T_a^+ = \delta_{ab} - \frac{1}{2}\tau_b\tau_a, \tag{A.87}$$

relating the spin-$\frac{3}{2}$ to spin-$\frac{1}{2}$ transition operator, T_a, to the Pauli matrices.

[1]Notice that the matrices $\Theta_{\rho\nu}(c)$ where $c = \frac{1}{2}(1 + 4X)A + X = -\frac{1}{2}(1 + 2X)$ form a group with the multiplication given by $\Theta_{\alpha\beta}(a)\Theta^{\beta\delta}(b) = \Theta_\alpha^\delta(a + b + 4ab)$. The identity element is $\Theta(0)$ and the multiplicative inverse element $\Theta(a)^{-1}$ is given by $\Theta(-\frac{a}{1+4a})$ if we remove the element $\Theta(a = Z = -\frac{1}{4})$ which does not have an inverse due to $\left(\Theta(-\frac{1}{4})\Theta(b) = \Theta(-\frac{1}{4}) \ \forall\, b \in \Re\right)$. The non-invertable element can be removed since, if $\Theta(a)\Theta(b) = \Theta(-\frac{1}{4})$, either $a = -\frac{1}{4}$ or $b = -\frac{1}{4}$.

[2]For electroproduction, one also needs a "g_3" coupling due to the existence of $L0$ longitudinal photons.

A.4. Renormalization formulae

A.4.1. Dimensional regularization

In dimensional regularization, the general 1-loop integral has the form

$$I = \int \frac{d^dk}{(2\pi)^d} \frac{N}{A_1^{\alpha_1} A_2^{\alpha_2} \cdots A_n^{\alpha_n}}, \tag{A.88}$$

where d is the number of spacetime dimensions (not nessecarily an integer) and the numbers, α_i, are integers unless equation (A.88) is the result of a multi-loop calculation in which some loops have already been evaluated. This integral is solved in the following way:

(a) The different denominators are combined into a single denominator and then the combined denominator is reduced to a standard form by translating or shifting the internal momenta.

(b) The integral is then evaluated using an integral identity.

Step (a) makes use of Feynman parametrization identities of the form

$$\begin{aligned}
\frac{1}{A_1 A_2} &= \int_0^1 dx \frac{1}{[A_1 x + A_2(1-x)]^2} = \int_0^1 \frac{dx}{[D(x)]^2}, \\
\frac{1}{A_1 A_2 A_3} &= 2\int_0^1 dx_1 \int_0^{1-x_1} \frac{dx_2}{[A_1 x_1 + A_2 x_2 + A_3(1 - x_1 - x_2)]^3} \\
&= 2\int_0^1 dx_1 \int_0^{1-x_1} \frac{dx_2}{[D(x_1, x_2)]^3}, \\
&\ \ \vdots \\
\frac{1}{A_1^{\alpha_1} \cdots A_n^{\alpha_n}} &= \frac{\Gamma(\alpha_1 + \cdots + \alpha_n)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_n)} \int_0^1 dx_1 \cdots dx_n \delta(1 - \sum_i x_i) \\
&\quad \times \frac{x_1^{\alpha_1 - 1} \cdots x_n^{\alpha_n - 1}}{[A_1 x_1 + \cdots + A_n x_n]^{\alpha_1 + \cdots + \alpha_n}}.
\end{aligned} \tag{A.89}$$

Note that the combined denominator, D, always has the form

$$D = k^2 + 2k \cdot q + b^2, \tag{A.90}$$

where k is the internal loop momentum and q is a vector function of the external momenta and the Feynman parameters. The square of the denominator can always be completed by shifting $k \to k' - q$, leaving

$$D \to (k')^2 + b^2 - q^2. \tag{A.91}$$

This shift must also be carried out in the numerator, N, which assumes the general form

$$N = N_0 + k'_\mu N_1^\mu + k'_\mu k'_\nu N_2^{\mu\nu} + k'_\mu k'_\nu k'_\sigma N_3^{\mu\nu\sigma} + \cdots , \tag{A.92}$$

where the N_i are tensors which do not depend on k'. Since the denominator is even in k' (depends on $(k')^2$ only), all of the odd terms reduce to zero. The even terms can be simplified using

$$\begin{aligned} \int \frac{d^dk}{(2\pi)^d} \frac{k^\mu k^\nu}{D(k^2)} &= \frac{g^{\mu\nu}}{d} \int \frac{d^dk}{(2\pi)^d} \frac{k^2}{D(k^2)}, \\ \int \frac{d^dk}{(2\pi)^d} \frac{k^\mu k^\nu k^\sigma k^\rho}{D(k^2)} &= \frac{g^{\mu\nu}g^{\sigma\rho} + g^{\mu\sigma}g^{\nu\rho} + g^{\mu\rho}g^{\nu\sigma}}{d(d+2)} \int \frac{d^dk}{(2\pi)^d} \frac{k^4}{D(k^2)}. \end{aligned} \tag{A.93}$$

If the integrals are finite, we can set $d = 4$. The integrals will now have the standard form

$$I = \int \frac{d^dk}{(2\pi)^d} \frac{(k^2)^n}{[C^2 - k^2 - i\epsilon]^\alpha}, \tag{A.94}$$

where n is an integer. These integrals can always be reduced to a sum of integrals of the form

$$\int \frac{d^dk}{(2\pi)^d} \frac{1}{[C^2 - k^2 - i\epsilon]^\alpha} = \frac{i}{(4\pi)^{\frac{d}{2}}} \frac{\Gamma(\alpha - \frac{d}{2})}{\Gamma(\alpha)} \left(\frac{1}{C^2}\right)^{(\alpha - \frac{d}{2})}. \tag{A.95}$$

A convenient combination is

$$\int \frac{d^dk}{(2\pi)^d} \frac{k^\mu k^\nu}{[C^2 - k^2 - i\epsilon]^\alpha} = -\frac{i g^{\mu\nu}}{2(4\pi)^{\frac{d}{2}}} \frac{\Gamma(\alpha - 1 - \frac{d}{2})}{\Gamma(\alpha)} \left(\frac{1}{C^2}\right)^{(\alpha - 1 - \frac{d}{2})}. \tag{A.96}$$

The following formulae are now used to further reduce the result

$$\begin{aligned} \Gamma(-n+\epsilon) &= \frac{(-1)^n}{n!} \left[\frac{1}{\epsilon} + \psi(n+1) + \frac{\epsilon}{2}\left\{\frac{\pi^2}{3} + \psi^2(n+1) - \psi'(n+1)\right\} + \mathcal{O}(\epsilon^2)\right] \\ \psi(n+1) &= 1 + \frac{1}{2} + \cdots + \frac{1}{n} - \gamma \\ \psi(1) &= -\gamma = -0.577215664901532... = \text{ Euler's constant} \\ \psi'(n+1) &= \frac{\pi^2}{6} + \sum_{j=1}^{n} \frac{1}{j^2} \\ \psi'(1) &= \frac{\pi^2}{6} \\ X^\epsilon &= 1 + \epsilon \ln(X) + \frac{\epsilon^2}{2!} \ln^2(X) + \cdots . \end{aligned} \tag{A.97}$$

APPENDIX B
INTRODUCTION TO SUPERSYMMETRY

In this appendix, we introduce supersymmetry following the excellent report by Sohnius [75] and the lectures by Lykken [76] and Olive [77]. Supersymmetry is a symmetry between fermions and bosons. A supersymmetric field theory consists of a set of quantum fields and a lagrangian for these fields which exhibits the symmetry. A supersymmetric model which is covariant under general coordinate transformations, i.e., a model which possesses local (gauged) supersymmetry, is called a *supergravity* model.

Supersymmetry and supergravity aim at a unified description of fermions and bosons, and hence of matter fields and force fields. Forces are mediated by gauge potentials which are either vector fields with spin one as in the case of electromagnetism and the weak and strong nuclear forces, or tensor fields of spin two in the case of gravitation. Matter, on the other hand, is built from quarks and leptons which are spin-$\frac{1}{2}$ fermions. Supersymmetric theories unite the fermions and bosons into multiplets and lift the basic distinction between matter and interaction. The gluinos, which are the supersymmetric partner of the gluon (and are therefore fermions), are thought of as the carriers of the strong force as much as the gluons except that they are fermions and, hence, they must obey an exclusion principle. Thus, gluinos will never conspire to form a coherent, measurable potential. The distinction between forces and matter now becomes phenomenological only. Bosons manifest themselves as forces because they can build up coherent classical fields while fermions are seen as matter since no two identical ones can occupy the same point in space, an intuitive definition of material existence.

B.1. The supersymmetry algebra

The generator of a symmetry is an operator in a Hilbert space that replaces one incoming or outgoing multiparticle state with another, leaving the "physics," or the S-matrix, unchanged. For example, permuting identical incoming particles should be the generator of a symmetry. Such an operator can be written as an annihilation operator which removes the original state and a creation operator which replaces it with the new one which may have different properties and 3-momentum. The most general such product is the convolution

$$G = a^\dagger * K * a = \sum_{ij} \int d^3p\, d^3q\; a_i^\dagger(\mathbf{p}) K_{ij}(\mathbf{p},\mathbf{q}) a_j(\mathbf{q}). \tag{B.1}$$

This operator is completely defined by the integral kernel, $K_{ij}(\mathbf{p},\mathbf{q})$. The operator, G, will be called a generator of a symmetry if it leaves the physics unchanged. This statement is expressed mathematically by saying that the symmetry operator commutes with the S-matrix ($[S, G] = 0$). The result of a scattering experiment (the S-matrix) is unchanged when we first permute the initial state with the symmetry operator so that the two different initial states give the same observational results and are, therefore, physically indistinguishable.

The operator, G, can be decomposed into an even part, B, which changes fermions into fermions and bosons into bosons, and an odd part, F, which changes fermions into bosons and vice versa. B will change the total spin of a state by integer amounts while F will change it by half-integer amounts. $G = B + F$, where

$$\begin{aligned} B &= b^\dagger * K_{bb} * b + f^\dagger * K_{ff} * f \\ F &= f^\dagger * K_{fb} * b + b^\dagger * K_{bf} * f, \end{aligned} \tag{B.2}$$

and b and f are the annihilation operators for bosons and fermions, respectively. The canonical quantization relations for the particle operators are

$$\left[a, a^\dagger\right\} = \mathbb{I}; \quad \left[a, a\right\} = \left[a^\dagger, a^\dagger\right\} = 0, \tag{B.3}$$

where $\mathbb{I} = \delta_{ij}\delta^3(\mathbf{p} - \mathbf{q})$ is the unit element of the convolution product, $*$, and the graded commutator, $\left[\cdots\right\}$, is the anticommutator if both particle operators are fermionic, and is the commutator in all other cases.

B.1.1. Graded Lie algebras

Graded Lie algebras are defined by the following algebraic relations:

$$\begin{aligned} [B_i, B_j] &= i c_{ij}{}^k B_k \\ [F_\alpha, B_i] &= s_{\alpha i}{}^\beta F_\beta \\ \{F_\alpha, F_\beta\} &= \gamma_{\alpha\beta}{}^i B_i, \end{aligned} \tag{B.4}$$

where the structure constants have the symmetries

$$\begin{aligned} c_{ij}{}^k &= -c_{ji}{}^k \\ s_{ij}{}^k &= -s_{ji}{}^k \\ \gamma_{\alpha\beta}{}^i &= \gamma_{\beta\alpha}{}^i \end{aligned} \tag{B.5}$$

and the graded Jacobi identities

$$\left[\left[G^1, G^2\right\}, G^3\right\} + \text{graded cyclic} = 0. \tag{B.6}$$

Graded cyclic means that there is an extra minus sign if two fermionic operators are exchanged. For example,

$$F_\alpha F_\beta B_i + \text{graded cyclic} = F_\alpha F_\beta B_i + B_i F_\alpha F_\beta - F_\beta B_i F_\alpha.$$

The requirement that the graded Jacobi identities be satisfied is equivalent to demanding that the structure constants form the adjoint representation of the algebra. The representation is

$$r(B_i) = \begin{pmatrix} C_i & 0 \\ 0 & S_i \end{pmatrix} \quad \text{and} \quad r(F_\alpha) = \begin{pmatrix} 0 & \Sigma_\alpha \\ \Gamma_\alpha & 0 \end{pmatrix} \tag{B.7}$$

with matrix elements

$$\begin{aligned} (C_i)_j^k = ic_{ji}^k; \quad (S_i)_\alpha^\beta = s_{\alpha i}^\beta \\ (\Gamma_\alpha)_\beta^i = \gamma_{\beta\alpha}^i; \quad (\Sigma_\alpha)_i^\beta = s_{i\alpha}^\beta. \end{aligned} \tag{B.8}$$

The bosonic (even) generators

In 1967, Coleman and Mandula wrote a paper [78] which gave a proof that, under certain assumptions, the only possible symmetries of the scattering matrix, or S-matrix, are as follows:

- Poincaré invariance: The semi-direct product of rotations and Lorentz transformations, with generators P_μ and $M_{\mu\nu}$.
- Internal global symmetries: Symmetries related to conservation of quantum numbers such as electric charge, color, and isospin. The symmetry generators are Lorentz scalars and generate a Lie algebra,

  $$[B_i, B_j] = iC_{ij}^k B_k,$$

  where the C_{ij}^k are the structure constants.
- Discrete symmetries: The charge conjugation (C), parity reversal (P), and time reversal (T) symmetries.

It was, therefore, shown that any group of bosonic symmetries of the S-matrix in relativistic field theory is the direct product of the Poincaré group with an internal symmetry group, and the internal group must be the product of a compact semisimple group with $U(1)$ factors.

The bosonic generators are then the four momenta, P_μ, and the six Lorentz generators, $M_{\mu\nu}$, plus a certain number of internal hermitian symmetry generators, B_r. The algebra is that of the Poincaré group

$$\begin{aligned} [P_\mu, P_\nu] &= 0 \\ [P_\mu, M_{\rho,\sigma}] &= i\,(\eta_{\mu\rho}P_\sigma - \eta_{\mu\sigma}P_\rho) \\ [M_{\mu\nu}, M_{\rho,\sigma}] &= i\,(\eta_{\nu\rho}M_{\mu\sigma} - \eta_{\nu\sigma}M_{\mu\rho} - \eta_{\mu\rho}M_{\nu\sigma} + \eta_{\nu\sigma}M_{\nu\rho}) \end{aligned} \tag{B.9}$$

together with the internal symmetry group

$$[B_r, B_s] = ic_{rs}^t B_t. \tag{B.10}$$

The direct product structure means that the B_r must be translationally invariant Lorentz scalars with the vanishing commutators

$$[B_r, P_\mu] = [B_r, M_{\mu\nu}] = 0. \tag{B.11}$$

The Casimir operators of the Poincaré symmetry group are the mass-square operator, $P^2 = P_\mu P^\mu$, and the generalized spin operator, $W^2 = W_\mu W^\mu$, where W^μ is the Pauli-Lubanski vector

$$W^\mu = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma} P_\nu M_{\rho\sigma}. \tag{B.12}$$

In the rest frame of a massive particle, we have

$$P_\mu = (m, 0, 0, 0) \quad W^2 = -m^2\mathbf{L}^2$$

with $\mathbf{L} = (M_{23}, M_{31}, M_{12})$. These Casimir operators commute with the entire Poincaré group and also with the internal symmetry generators

$$\left[B_r, P^2\right] = 0; \quad \left[B_r, W^2\right] = 0.$$

The first of these commutators says that all members of an irreducible multiplet of the internal symmetry group must have the same mass, and the second says that they must also have the same spin.

For massless states with discrete helicities, we have $W_\mu = \lambda P_\mu$ with λ half integral and $P^2 = W^2 = 0$. No B_r can change the helicity. The Coleman-Mandula *no-go theorem* states that all generators of supersymmetries (spin-changing symmetries) must be fermionic in that they must change the spin by half-integral amounts and change the statistics of the state.

The fermionic (odd, supersymmetry) generators

The odd elements of the graded algebra are the fermionic generators. The Jacobi identities link the properties of the fermionic sector with those of the bosonic sector. The positive metric assumption of the Hilbert space says that the anticommutator of an operator with its adjoint is non-negative and it is zero only if the operator itself is zero. If there is no element of the bosonic subset of the generators which possesses all of the properties required of $\{Q, Q^\dagger\}$, we can conclude that $Q = 0$.

The supersymmetry generators, which we will now call Q rather than F, must carry a representation of the bosonic symmetry group. If Q sits in the (j, j') representation of the Lorentz group, then $\{Q, Q^\dagger\}$ will contain the representation $(j + j', j + j')$. Since P_μ is the only object in the bosonic sector which is in such a representation, namely in $(\frac{1}{2}, \frac{1}{2})$, we know that all Q's must be in one of the two 2-dimensional representations, $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$, of the algebra of the Lorentz group. Thus, we have

$$\begin{aligned} \left[Q_{\alpha i}, M_{\mu\nu}\right] &= \frac{1}{2}\left(\sigma_{\mu\nu}\right)_\alpha^\beta Q_{\beta i} \\ \left[\overline{Q}^i_{\dot\alpha}, M_{\mu\nu}\right] &= -\frac{1}{2}\overline{Q}^i_{\dot\beta}\left(\overline{\sigma}_{\mu\nu}\right)^{\dot\beta}_{\dot\alpha}, \end{aligned} \tag{B.13}$$

where the σ-matrices are defined in Section A.1.1.

The index, i, in $Q_{\alpha i}$ labels all of the different 2-spinors, Q_α, that are present and runs from 1 to some integer N. We assume that. if Q is the generator of a symmetry, then so is $Q^\dagger$. Since $(Q_{\alpha i})^\dagger$ must be in the complex conjugate representation (and, therefore, must be one of the $\overline{Q}$'s), we number them in such a way that $\overline{Q}^i_{\dot\alpha} = (Q_{\alpha i})^\dagger$ always. The anticommutator $\{Q, \overline{Q}\}$ transforms as $(\frac{1}{2}, \frac{1}{2})$ and must, therefore, be proportional to the energy-momentum operator

$$\left\{Q_{\alpha i}, \overline{Q}^j_{\dot\beta}\right\} = 2\delta^j_i \left(\sigma^\mu\right)_{\alpha\dot\beta} P_\mu. \tag{B.14}$$

The sign on the right-hand side of this equation is determined by the requirement that the energy, $E = P_0$, should be a positive definite operator. For each value of i, we get

$$\sum_{\alpha=1}^{2}\left\{Q_{\alpha i}, (Q_{\alpha i})^\dagger\right\} = 2\mathrm{Tr}\left(\sigma^\mu P_\mu\right) = 4P_0 \quad \text{(no sum over } i\text{)}, \tag{B.15}$$

and since the commutator is always positive,[1] the energy is also. We also have that the supersymmetry generators commute with the momenta

$$\left[Q_{\alpha i}, P_\mu\right] = \left[\overline{Q}^i_{\dot\alpha}, P_\mu\right] = 0. \tag{B.16}$$

The Q's generally carry some representation of the internal symmetry

$$\left[Q_{\alpha i}, B_r\right] = (b_r)^j_i\, Q_{\alpha j}. \tag{B.17}$$

Since the internal symmetry group is compact, the representation matrices can be chosen to be hermitian, $b_r = b_r^\dagger$. For $\overline{Q}$, we have

$$\left[\overline{Q}^i_{\dot\alpha}, B_r\right] = -\overline{Q}^j_{\dot\alpha}\,(b_r)^i_j\,. \tag{B.18}$$

The largest possible internal symmetry group that can act non-trivially on Q is $U(N)$. The indices, α and $\dot\alpha$, belong to the Poincaré group while i is an internal symmetry index. The anticommutator, $\{Q, Q\}$, is required from Lorentz invariance to be a linear combination of bosonic operators in the representations $(0,0)$ and $(1,0)$ of the Lorentz group. The only $(1,0)$ present in the bosonic sector of the algebra is the self-dual part of $M_{\mu\nu}$, but such a term would not commute with P_μ as required. Hence, the most general form of the anticommutator[2] is

$$\left\{Q_{\alpha i}, Q_{\beta j}\right\} = 2\epsilon_{\alpha\beta} Z_{ij}, \tag{B.19}$$

where the Z_{ij} are some linear combination of the internal symmetry generators, $Z_{ij} = a^r_{ij} B_r$. The Z_{ij} commute with anything; therefore, they are elements of the center of

[1] This is due to the positive metric assumption of the Hilbert space.

[2] The two-dimensional Levi-Civita tensors are defined by $\epsilon_{12} = \epsilon^{12} = -\epsilon_{\dot1\dot2} = -\epsilon^{\dot1\dot2} = 1$.

the group and are called central charges for this reason. The adjoint gives

$$\left\{\overline{Q}^i_{\dot\alpha}, \overline{Q}^j_{\dot\beta}\right\} = -2\epsilon_{\dot\alpha\dot\beta} Z^{ij} \tag{B.20}$$

with $Z^{ij} = (Z_{ij})^\dagger$. We also have $Z_{ij} = -Z_{ji}$ which means $a^r_{ij} = -a^r_{ji}$. For each non-vanishing central charge, there must be a different antisymmetric $N \times N$ matrix, a^r, which is a numerical invariant of the internal symmetry group:

$$(b^s)^k_i a^r_{kj} + (b^s)^k_j a^r_{ik} = 0. \tag{B.21}$$

Central charges in the algebra, therefore, impose a symplectic structure on the semi-simple part of the internal symmetry group. The largest internal symmetry group which can sustain a central charge is $USp(N)$, the compact version of the symplectic group $Sp(N)$.

If there is only one 2-spinor supercharge Q_α, i.e., if $N = 1$, we say that a theory exhibits simple, or unextended, supersymmetry. If $N > 1$, we speak of extended supersymmetry. For simple supersymmetry, the only internal symmetry acting non-trivially is a single $U(1)$ generated by a charge which has become known as R:

$$[Q, R] = Q; \quad [\overline{Q}, R] = -\overline{Q}. \tag{B.22}$$

Since, under parity, $Q \to \overline{Q}$ and $\overline{Q} \to Q$, we must have $R \to -R$, which means that the $U(1)$ symmetry group is chiral.

B.2. Supersymmetry representations

Recall that a representation of a group is a set of matrices which satisfy the algebra of the group. An n-dimensional representation will be a set of $n \times n$ matrices which act on an n-dimensional vector space of state vectors.

Supersymmetric theories deal with sets of fields which carry representations of one of the graded Lie algebras. Each state vector in the representation space contains equal numbers of fermionic and bosonic components. The even elements of the supersymmetry generators (P_μ, $M_{\mu\nu}$, and B_r) map these subspaces into themselves while the odd elements (Q's) map the fermionic components to bosonic ones and vice versa. The supersymmetry relation $\{Q, \overline{Q}\} = 2\sigma^\mu P_\mu$ means that, if we map the fermionic subspace to the bosonic one and then back again, the net effect is the same as operating with P_μ on the original subspace.

On a physical multiparticle state, the 4-momentum operator can be represented by $r(P_\mu) = (E, \mathbf{p})$ while, on quantum fields, $\phi(x)$, P_μ is the generator of translations and is represented as $r(P_\mu) = i\partial_\mu$.

B.2.1. Massless one-particle states

A massless one-particle state can always be rotated into a standard frame where its movement is in the z-direction so that $P_\mu = (E, 0, 0, E)$. The spacetime properties of the state are then determined by its energy, E, and its helicity, λ. The helicity is the projection of its spin onto the direction of motion, so it is the eigenvalue of $E^{-1}\mathbf{L}\cdot\mathbf{P}$. Using the definitions of the Pauli-Lubanski vector and the angular momentum operator, we find that $W_0 = \mathbf{L}\cdot\mathbf{P}$ so that, for a massless helicity eigenstate, $|E,\lambda\rangle$, we have $W_0 = \lambda E$. Lorentz covariance then allows us to write

$$W_\mu |E,\lambda\rangle = \lambda P_\mu |E,\lambda\rangle . \tag{B.23}$$

Let one of the supersymmetry operators, $Q_{\alpha i}$, operate on our state, $|E,\lambda\rangle$. The energy and the momentum of the state remain unchanged since $[Q, P_\mu] = 0$. Since application of W_0 returns the helicity of the state, we can find the helicity of $Q_\alpha |E,\lambda\rangle$ and, hence, the effect of Q on helicity by checking with W_0. We find

$$\begin{aligned} W_0 Q_\alpha |E,\lambda\rangle &= Q_\alpha W_0 |E,\lambda\rangle + [W_0, Q_\alpha] |E,\lambda\rangle \\ &= \lambda E Q_\alpha |E,\lambda\rangle + [\mathbf{L}\cdot\mathbf{P}, Q_\alpha] |E,\lambda\rangle , \end{aligned} \tag{B.24}$$

and, since $\mathbf{L} = (M_{23}, M_{31}, M_{12})$,

$$\begin{aligned} [\mathbf{L}\cdot\mathbf{P}, Q_\alpha] &= [M_{23}P_1 + M_{31}P_2 + M_{12}P_3, Q_\alpha] \\ &= ([M_{23}, Q_\alpha] P_1 + [M_{31}, Q_\alpha] P_2 + [M_{12}, Q_\alpha] P_3) \\ &= \left(-\frac{1}{2} (\sigma_{23})_\alpha^{\ \beta} Q_\beta P_1 - \frac{1}{2} (\sigma_{31})_\alpha^{\ \beta} Q_\beta P_2 - \frac{1}{2} (\sigma_{12})_\alpha^{\ \beta} Q_\beta P_3 \right) \\ &= -\frac{1}{2} \left((\sigma_1)_\alpha^{\ \beta} P_1 + (\sigma_2)_\alpha^{\ \beta} P_2 + (\sigma_3)_\alpha^{\ \beta} P_3 \right) Q_\beta \\ &= -\frac{1}{2} (\boldsymbol{\sigma}\cdot\mathbf{P})_\alpha^{\ \beta} Q_\beta , \end{aligned} \tag{B.25}$$

where we used $[M_{\mu\nu}, Q_\alpha] = -\frac{1}{2} (\sigma_{\mu\nu})_\alpha^{\ \beta} Q_\beta$ and the fact that $\sigma_{12} = \sigma_3$ and cyclic. Note also that $\sigma_{0i} = -i\sigma_i$. Using $\mathbf{P} = (0, 0, E)$, we now have

$$W_0 Q_\alpha |E,\lambda\rangle = E \left(\lambda \mathbb{I} - \frac{1}{2}\sigma^3 \right)_\alpha^{\ \beta} Q_\beta |E,\lambda\rangle , \tag{B.26}$$

where $\mathbb{I}$ is the unit matrix. Now substituting the explicit form for the Pauli matrix, $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, we have for each supersymmetry generator labeled by i

$$\begin{aligned} W_0 \begin{pmatrix} Q_{1i} \\ Q_{2i} \end{pmatrix} |E,\lambda\rangle &= E \begin{pmatrix} \lambda - \frac{1}{2} & 0 \\ 0 & \lambda + \frac{1}{2} \end{pmatrix} \begin{pmatrix} Q_{1i} \\ Q_{2i} \end{pmatrix} |E,\lambda\rangle \\ &= E \begin{pmatrix} \left(\lambda - \frac{1}{2}\right) Q_{1i} \\ \left(\lambda + \frac{1}{2}\right) Q_{2i} \end{pmatrix} |E,\lambda\rangle . \end{aligned} \tag{B.27}$$

Finally,

$$W_0 Q_{1i} |E,\lambda\rangle = E\left(\lambda - \frac{1}{2}\right) Q_{1i} |E,\lambda\rangle$$
$$W_0 Q_{2i} |E,\lambda\rangle = E\left(\lambda + \frac{1}{2}\right) Q_{2i} |E,\lambda\rangle \tag{B.28}$$

Since $W_0 |E,\lambda\rangle = E\lambda |E,\lambda\rangle$, we see that $Q_{1i} |E,\lambda\rangle$ is a state with helicity $\lambda - \frac{1}{2}$ and $Q_{2i} |E,\lambda\rangle$ is a state with helicity $\lambda + \frac{1}{2}$. The Q's raise and lower the helicity by $\frac{1}{2}$ (changing bosons into fermions and vice versa). Similarly, the $\overline{Q}$'s also raise and lower helicity but in the opposite sense since $\left[M_{\mu\nu}, \overline{Q}_\alpha\right] = \frac{1}{2}\left(\sigma_{\mu\nu}\right)_\alpha^\beta Q_\beta$, which differs in sign with that of the Q's. $\overline{Q}_{\dot{1}}$ raises the helicity by $\frac{1}{2}$ and $\overline{Q}_{\dot{2}}$ lowers it.

Using the fact that $\sigma^\mu = (\mathbb{I}, \boldsymbol{\sigma})$, we find that, in the absence of central charges, the supersymmetry algebra (B.14, B.19, and B.20) reduces to[3]

$$\{Q,Q\} = \{\overline{Q},\overline{Q}\} = 0$$
$$\left\{Q_{2i}, \overline{Q}^j_{\dot{2}}\right\} = 0 \tag{B.29}$$
$$\left\{Q_{1i}, \overline{Q}^j_{\dot{1}}\right\} = 4\delta_i^j E.$$

If we rescale $q_i \equiv (4E)^{-1/2} Q_{1i}$, we have

$$\left\{q_i, \overline{q}^j\right\} = \delta_i^j, \quad \{q_i, q_j\} = \left\{\overline{q}^i, \overline{q}^j\right\} = 0. \tag{B.30}$$

Our positivity requirement gives $Q_{2i} = 0$. The algebra of the q's is the Clifford algebra of N fermionic degrees of freedom, and any irreducible representation of this algebra is characterized by a Clifford ground state, $|E,\lambda_0\rangle$, with $q\,|E,\lambda_0\rangle = 0$. The other states are generated by successive application of the N operators, $\overline{q}^i$.

$$\overline{q}^i |E,\lambda_0\rangle = |E, \lambda + 1/2, i\rangle\,,$$
$$\overline{q}^i\overline{q}^j |E,\lambda_0\rangle = |E, \lambda + 1, ij\rangle\,, \text{ etc.} \tag{B.31}$$

We continue in this fashion until we have reached the top state for which any further application of $\overline{q}$ gives zero. In addition to whatever other labels the ground state may have, the other states will also contain the internal symmetry labels of the $\overline{q}$'s which generated them. These states must be totally antisymmetric in the internal labels, i.e., $|E,\lambda,ij\rangle = -\,|E,\lambda,ji\rangle$. The helicity has been raised by each application of a $\overline{q}^i$, so we have the following spectrum

$$\begin{array}{llllll} \text{helicity:} & \lambda_0 & \lambda_0 + 1/2 & \lambda_0 + 1 & \cdots & \lambda_0 + N/2 \\ \text{no. of states:} & \binom{N}{0} = 1 & \binom{N}{1} = N & \binom{N}{2} & \cdots & \binom{N}{N} = 1 \end{array} \tag{B.32}$$

[3] Actually, $\left\{Q_{2i}, \overline{Q}^j_{\dot{2}}\right\} = -4\delta_i^j \left(\sigma^3\right)_{22} P_3 = -4\delta_i^j E$. This implies that $Q_{2i} = 0$ from positivity.

We can see that the number of states, 2^N, and the fact that the number of fermions is equal to the number of bosons follows from

$$(1+1)^N = \sum_{k=0}^{N} \begin{pmatrix} N \\ k \end{pmatrix} = 2^N \tag{B.33}$$

and

$$(1-1)^N = \sum_{k=0}^{N/2} \begin{pmatrix} N \\ 2k \end{pmatrix} - \sum_{k=0}^{N/2} \begin{pmatrix} N \\ 2k+1 \end{pmatrix} = 0. \tag{B.34}$$

Two important examples of these spectra are the $N = 4$ Yang-Mills multiplet with $\lambda_0 = -1$ and the $N = 8$ supergravity multiplet with $\lambda_0 = -2$. These multiplets are

$$\begin{array}{llccccccccc} \underline{N=4} & \text{helicity:} & -1 & -\frac{1}{2} & 0 & \frac{1}{2} & 1 \\ & \text{states:} & 1 & 4 & 6 & 4 & 1 \\ \underline{N=8} & \text{helicity:} & -2 & -\frac{3}{2} & -1 & -\frac{1}{2} & 0 & \frac{1}{2} & 1 & \frac{3}{2} & 2 \\ & \text{states:} & 1 & 8 & 28 & 56 & 70 & 56 & 28 & 8 & 1. \end{array} \tag{B.35}$$

Since spin-$\frac{3}{2}$ does not allow a renormalizable coupling and spin-$\frac{5}{2}$ does not allow consistent coupling to gravity, we must have

$N \leq 4$ for renormalizable theories

$N \leq 8$ for consistent theories of supergravity.

Normally, any spectrum of states that is derived from a Lorentz covariant field theory will exhibit PCT-symmetry which implies that, for every state with helicity λ, there should be a parity reflected state with helicity $-\lambda$. Our spectra do not, in general, have this property. For example, $N = 1, \lambda_0 = 0$ only has helicities 0 and 1. In order to get a Lorentz covariant field theory for $N = 1$, we need to include the PCT-conjugate multiplet[4] with $\lambda_0 = -\frac{1}{2}$ to give

$$\begin{array}{llccc} \underline{N=1} & \text{helicity:} & -\frac{1}{2} & 0 & \frac{1}{2} \\ & \text{states:} & 1 & 2 & 1 \end{array} \tag{B.36}$$

which is the spectrum of the massless Wess-Zumino model which contains a scalar, a pseudoscalar, and a Majorana spin-$\frac{1}{2}$ field in interaction.

B.2.2. Massive one-particle states

First, we discuss the case where there are no central charges. The spacetime properties of a massive one-particle state are given by the mass, m, the total spin, s, and the spin projection onto the z-axis, s_z. We assume that the particles are in the rest frame, where $P_\mu = (m, 0, 0, 0)$. Q is a tensor operator of spin-$\frac{1}{2}$ as can be seen from

[4]It is a property of the $N = 4$ Yang-Mills and the $N = 8$ supergravity that their multiplets are already PCT-self conjugate.

$[Q, \mathrm{L}] = \frac{1}{2}\sigma Q$. Therefore, the action of Q on a state with spin s will be a linear combination of states with spins $s + \frac{1}{2}$ and $s - \frac{1}{2}$:

$$Q\left|mss_3\right\rangle = \sum_{s_3'} c^{(+)}_{s_3 s_3'} \left|ms + \frac{1}{2}s_3'\right\rangle + \sum_{s_3'} c^{(-)}_{s_3 s_3'} \left|ms - \frac{1}{2}s_3'\right\rangle \tag{B.37}$$

and similarly for $\overline{Q}$ with different coefficients.

The algebra in the rest frame without central charges is then

$$\left\{Q_{\alpha i}, \overline{Q}^j_{\dot\beta}\right\} = 2m\delta^j_i \delta_{\alpha\dot\beta} \quad \{Q,Q\} = \{\overline{Q},\overline{Q}\} = 0. \tag{B.38}$$

Notice that $Q_{2i} \neq 0$ in the massive case, giving $2N$ generators rather than just N as in the massless case.

$$\left\{Q_{\alpha i}, \overline{Q}^j_{\dot\beta}\right\} = 2\delta^j_i \left(\sigma^\mu\right)_{\alpha\dot\beta} P_\mu = 2\delta^j_i \left(\mathbb{I}\right)_{\alpha\dot\beta} P_0 = 2m\delta^j_i \delta_{\alpha\dot\beta}$$

is positive definite unlike $\left\{Q_{\alpha i}, \overline{Q}^j_{\dot\beta}\right\} = 2\delta^j_i \left(\sigma^3\right)_{\alpha\dot\beta} P_3$ in the massless case.

With $q_{\alpha i} = (2m)^{-1/2} Q_{\alpha i}$ (B.38) becomes

$$\left\{q_{\alpha i}, \overline{q}^j_{\dot\beta}\right\} = \delta^j_i \delta_{\alpha\dot\beta}; \quad \{q,q\} = \{\overline{q},\overline{q}\} = 0, \tag{B.39}$$

which is the Clifford algebra for $2N$ fermionic degrees of freedom. A representation is characterized by a spin multiplet of ground states, $\left|ms_0s_3\right\rangle, s_3 = -s_0, ..., +s_0$, which are annihilated by $q_{\alpha i}$. The other states are again generated by successive application of the $\overline{q}$'s. The states are totally antisymmetric under interchange of index pairs $(\dot\alpha\ i) \leftrightarrow (\dot\beta\ j)$. The maximal spin ($s_{\mathsf{max}} = s_0 + N/2$) is carried by states like

$$\overline{q}^1_1 \overline{q}^2_1 \cdots \overline{q}^N_1 \left|ms_0\right\rangle.$$

The minimal spin is zero if $N/2 \geq s_0$ and is $s_0 - N/2$ otherwise. The dimension of a representation with $s_0 = 0$ is 2^{2N} because there are $2N$ independent raising operators.

Since renormalizability requires massive matter to have spin $\leq \frac{1}{2}$, we deduce from the expression for s_{max} that we must have $N = 1$ for the renormalizable coupling of massive matter. The only important massive multiplet in the absence of central charges (apart from supergravity where renormalization is not an issue) is the massive Wess-Zumino model with $N = 1$ and $s_0 = 0$ which has a scalar, a pseudoscalar, and the two spin states of a massive Majorana spinor:

<u>$N = 1$</u>	s^P	0^+	$\frac{1}{2}$	0^-
	states:	1	2	1.

(B.40)

Now, we want to discuss one-particle representations with central charges included. Since the central charges, Z_{ij}, commute with everything, we can choose a basis in which they are diagonal and represented by complex numbers, z_{ij}. These numbers form an

$N \times N$ matrix which can be brought into standard form by using a unitary matrix, $\overline{z}_{ij} = U_i^k U_j^l z_{kl}$. The standard form for even N is

$$\overline{z} = \begin{pmatrix} 0 & D \\ -D & 0 \end{pmatrix}, \tag{B.41}$$

where D is a real, diagonal matrix with non-negative eigenvalues. We label the rth eigenvalue by $z(r)$ where $r = 1, ..., N/2$. If N is odd, there is an extra row and an extra column of zeros, i.e.,

$$\overline{z} = \begin{pmatrix} 0 & D & 0 \\ -D & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{B.42}$$

We now use the unitary matrix, U, to redefine our Q's so that $U_i^j Q_{\alpha j} \to Q_{\alpha i}$ and $\overline{Q}^j_{\dot{\alpha}} (U^{-1})^i_j \to \overline{Q}^i_{\dot{\alpha}}$. We introduce double indices, $i = (a, r)$, where $a = 1, 2$ and $r = 1, ..., N/2$ which arise from the form of equation (B.41). For odd N, the last charge, $Q_{\alpha N}$, is not touched by this change.

The algebra of the Q's is now

$$\begin{aligned} \left\{ Q_{\alpha ar}, \overline{Q}^{bs}_{\dot{\beta}} \right\} &= 2\delta^b_a \delta^s_r (\sigma^\mu)_{\alpha\dot{\beta}} P_\mu \\ \{ Q_{\alpha ar}, Q_{\beta bs} \} &= 2\epsilon_{\alpha\beta}\epsilon_{ab}\delta_{rs} z_{(r)} \\ \left\{ \overline{Q}^{ar}_{\dot{\alpha}}, \overline{Q}^{bs}_{\dot{\beta}} \right\} &= -2\epsilon_{\dot{\alpha}\dot{\beta}} \epsilon^{ab} \delta^{rs} z_{(r)}. \end{aligned} \tag{B.43}$$

For odd N, we also have

$$\{ Q_{\alpha N}, Q_{\beta i} \} = 0; \quad \left\{ Q_{\alpha N}, \overline{Q}^i_{\dot{\beta}} \right\} = 2\delta^i_N (\sigma^\mu)_{\alpha\dot{\beta}} P_\mu. \tag{B.44}$$

In the massless case, $Q_{2i} = 0$, implying that $z_{(r)} = 0$ and the central charges are trivial in the massless particle representations. For the massive case, we introduce the linear combinations

$$A^{\pm}_{\alpha r} \equiv \frac{1}{2} \left(Q_{\alpha 1 r} \pm \overline{Q}^{\dot{\alpha} 2 r} \right) \tag{B.45}$$

and their hermitian adjoints. As a reminder, recall

- α and $\dot{\alpha}$ label the component of the Q or $\overline{Q}$ (i.e., the component which is a spin raising or spin lowering operator)
- Q_{1ar} lowers spin and Q_{2ar} raises spin
- $\overline{Q}^{1ar}$ raises spin and $\overline{Q}^{2ar}$ lowers spin
- $(1r)$ and $(2r)$ label the rth diagonal element of the 1st and 2nd $N/2 \times N/2$ matrices in the off-diagonals of Q_1 or Q_2. (See equation (B.41).)

In terms of the $A^{\pm}$, the rest frame algebra is now

$$\begin{aligned}\{A^{\pm}, A^{\pm}\} = \{A^{\pm}, A^{\mp}\} = \left\{A^{\pm}, (A^{\mp})^{\dagger}\right\} = 0 \\ \left\{A^{\pm}_{\alpha r}, (A^{\pm}_{\beta s})^{\dagger}\right\} = \delta_{\alpha\beta}\delta_{rs}\left(m \pm z_{(r)}\right).\end{aligned} \tag{B.46}$$

The positivity of the last equation gives $z_{(r)} \leq m$. Let us assume that, for n_0 of the eigenvalues $z_{(r)}$ of the central charges, we have $m = z_{(r)}$. Then, the corresponding A^- are represented trivially, and after rescaling the remaining generators,

$$\begin{aligned} q^{\pm}_{\alpha r} &= \left(m \pm z_{(r)}\right)^{-1/2} A^{\pm}_{\alpha r} \\ q_{\alpha N} &= m^{-1/2} Q_{\alpha N} \quad \text{(if } N \text{ is odd.)}\end{aligned} \tag{B.47}$$

These equations form the Clifford algebra of $2(N - n_0)$ fermionic degrees of freedom. The spectrum is the same as without central charges, except that N is effectively reduced by n_0, which is the number of central charges that satisfy $m = z$. The simplest representation with central charge is the $N = 2$ hypermultiplet, which has one central charge with $z = m$, and the spectrum is a doubled version of the massive Wess-Zumino model.

B.2.3. Representations on fields

We now construct the field representation of the simplest supersymmetry algebra called the $N = 1$ chiral multiplet. Elements of the Hilbert space of a quantum field theory can be generated by the action of field-valued operators, $\phi(x)$, on a translationally invariant vacuum

$$|x\rangle = \phi^{\dagger}(x)\,|0\rangle\,, \quad |x, x'\rangle = \phi^{\dagger}(x)\phi^{\dagger}(x')\,|0\rangle\,, \text{ etc.} \tag{B.48}$$

Translations of a state are generated by the energy-momentum operator

$$|x + y\rangle = e^{iy\cdot P}\,|x\rangle\,, \quad |x + y, x' + y\rangle = e^{iy\cdot P}\,|x, x'\rangle\,, \text{ etc.}, \tag{B.49}$$

where $y \cdot P = y^{\mu} P_{\mu}$. A displacement of the field itself is, therefore[5],

$$\phi^{\dagger}(x + y) = e^{-iy\cdot P}\phi^{\dagger}(x)e^{iy\cdot P}. \tag{B.50}$$

We can write this displacement in differential form as $\left[P_{\mu}, \phi^{\dagger}\right] = i\partial_{\mu}\phi^{\dagger}$. Now, since we know the action of P_{μ} on fields, the structure of the supersymmetry algebra of the Q's is very similar to that of a Clifford algebra

$$\{Q, Q\} = \{\overline{Q}, \overline{Q}\} = 0 \quad \text{and} \quad \{Q, \overline{Q}\} = \text{ something known},$$

and we can construct a representation explicitly using the following steps:

[5]This is from $\langle x + y|x + y\rangle = \langle 0|\, e^{-iy\cdot P}\phi(x)e^{iy\cdot P}e^{-iy\cdot P}\phi^{\dagger}(x)e^{iy\cdot P}\,|0\rangle$, which is $\langle 0|\,\phi(x + y)\phi^{\dagger}(x + y)\,|0\rangle$.

Steps to construct the chiral multiplet

Step 1: Choose some complex scalar field, $A(x)$, as the ground state of our representation.

Step 2: Impose the constraint

$$[A, \overline{Q}_{\dot{\alpha}}] = 0.$$

This constraint defines what is known as the chiral multiplet. The graded Jacobi identity, the algebra, and $[P_\mu, \phi^\dagger] = i\partial_\mu \phi^\dagger$ give

$$\{[A, Q], \overline{Q}\} + \{[A, \overline{Q}], Q\} = [A, \{Q, \overline{Q}\}] = 2i\sigma^\mu \partial_\mu A, \qquad \text{(B.51)}$$

which implies that A must be complex since, otherwise, $[A, Q] = [A, \overline{Q}] = 0$ and that would mean that A is constant, $\partial_\mu A = 0$.

Step 3: Define fields $\psi_\alpha(x), F_{\alpha\beta}(x)$, and $X_{\alpha\dot{\beta}}$ by

$$[A, Q_\alpha] = 2i\psi_\alpha, \quad \{\psi_\alpha, Q_\beta\} = iF_{\alpha\beta}, \quad \left\{\psi_\alpha, \overline{Q}_{\dot{\beta}}\right\} = X_{\alpha\dot{\beta}}.$$

Step 4: Enforce the algebra on A. We have $\left[A, \overline{Q}_{\dot{\beta}}\right] = 0$ substituted into equation (B.51), leaving

$$\begin{aligned} [A, Q_\alpha] \overline{Q}_{\dot{\beta}} - \overline{Q}_{\dot{\beta}} [A, Q_\alpha] &= 2i \left(\sigma^\mu\right)_{\alpha\dot{\beta}} \partial_\mu A \\ 2i\psi_\alpha \overline{Q}_{\dot{\beta}} - \overline{Q}_{\dot{\beta}} 2i\psi_\alpha = 2i \left\{\psi_\alpha, \overline{Q}_{\dot{\beta}}\right\} &= 2i \left(\sigma^\mu\right)_{\alpha\dot{\beta}} \partial_\mu A \\ 2iX_{\alpha\dot{\beta}} &= 2i \left(\sigma^\mu\right)_{\alpha\dot{\beta}} \partial_\mu A. \end{aligned} \qquad \text{(B.52)}$$

Since $\{Q, Q\} = 0$, we have

$$[A, \{Q, Q\}] = 2i \{\psi_\alpha, Q_\beta\} + 2i \{\psi_\beta, Q_\alpha\} = 2 \left(F_{\alpha\beta} + F_{\beta\alpha}\right) = 0,$$

which means that

$$F_{\alpha\beta} = \epsilon_{\alpha\beta} F \qquad \text{(B.53)}$$

with $F(x)$ a complex scalar field.

Step 5: Define fields λ_α and $\overline{\chi}_{\dot{\alpha}}$ by

$$[F, Q_\alpha] = \lambda_\alpha, \quad [F, \overline{Q}_{\dot{\alpha}}] = \overline{\chi}_{\dot{\alpha}}. \qquad \text{(B.54)}$$

Step 6: Enforce the algebra on ψ starting with the graded Jacobi identities

$$\begin{aligned}
\left[\{\psi_\alpha, Q_\beta\}, \overline{Q}_{\dot\beta}\right] + \left[\left\{\psi_\alpha, \overline{Q}_{\dot\beta}\right\}, Q_\beta\right] &= \left[\psi_\alpha, \left\{Q_\beta, \overline{Q}_{\dot\beta}\right\}\right] \\
\left[-iF_{\alpha\beta}, \overline{Q}_{\dot\beta}\right] + \left[X_{\alpha\dot\beta}, Q_\beta\right] &= 2i\left(\sigma^\mu\right)_{\beta\dot\beta}\partial_\mu\psi_\alpha \\
-i\epsilon_{\alpha\beta}\overline{\chi}_{\dot\beta} + \left[\left(\sigma^\mu\right)_{\alpha\dot\beta}\partial_\mu A, Q_\beta\right] &= 2i\left(\sigma^\mu\right)_{\beta\dot\beta}\partial_\mu\psi_\alpha \\
-i\epsilon_{\alpha\beta}\overline{\chi}_{\dot\beta} + \left(\sigma^\mu\right)_{\alpha\dot\beta}\partial_\mu\left[A, Q_\beta\right] &= 2i\left(\sigma^\mu\right)_{\beta\dot\beta}\partial_\mu\psi_\alpha \\
-i\epsilon_{\alpha\beta}\overline{\chi}_{\dot\beta} + 2i\left(\sigma^\mu\right)_{\alpha\dot\beta}\partial_\mu\psi_\beta &= 2i\left(\sigma^\mu\right)_{\beta\dot\beta}\partial_\mu\psi_\alpha
\end{aligned} \tag{B.55}$$

and

$$\begin{aligned}
\left[\{\psi_\alpha, Q_\beta\}, Q_\gamma\right] + \left[\{\psi_\alpha, Q_\gamma\}, Q_\beta\right] &= \left[\psi_\alpha, \{Q_\beta, Q_\gamma\}\right] = 0 \\
-i\epsilon_{\alpha\beta}\left[F, Q_\gamma\right] - i\epsilon_{\alpha\gamma}\left[F, Q_\beta\right] &= 0 \\
\epsilon_{\alpha\beta}\lambda_\gamma + \epsilon_{\alpha\gamma}\lambda_\beta &= 0.
\end{aligned} \tag{B.56}$$

The only solution to these two equations is found by

$$\begin{aligned}
-i\epsilon^{\alpha\beta}\epsilon_{\alpha\beta}\overline{\chi}_{\dot\beta} &= -2i\epsilon^{\alpha\beta}\left(\sigma^\mu\right)_{\alpha\dot\beta}\partial_\mu\psi_\beta + 2i\epsilon^{\alpha\beta}\left(\sigma^\mu\right)_{\beta\dot\beta}\partial_\mu\psi_\alpha \\
-i2\overline{\chi}_{\dot\beta} &= -2i\left(\epsilon^{\alpha\beta}\left(\sigma^\mu\right)_{\alpha\dot\beta}\partial_\mu\psi_\beta - \epsilon^{\alpha\beta}\left(\sigma^\mu\right)_{\beta\dot\beta}\partial_\mu\psi_\alpha\right) \\
\overline{\chi}_{\dot\beta} &= \epsilon^{\alpha\beta}\left(\sigma^\mu\right)_{\alpha\dot\beta}\partial_\mu\psi_\beta + \epsilon^{\alpha\beta}\left(\sigma^\mu\right)_{\alpha\dot\beta}\partial_\mu\psi_\beta \\
&= 2\epsilon^{\alpha\beta}\partial_\mu\psi_\beta\left(\sigma^\mu\right)_{\alpha\dot\beta} = 2\partial_\mu\psi^\beta\left(\sigma^\mu\right)_{\beta\dot\beta} \\
\overline{\chi}_{\dot\alpha} &= 2\partial_\mu\psi^\beta\left(\sigma\right)_{\beta\dot\alpha}
\end{aligned} \tag{B.57}$$

and

$$\begin{aligned}
0 &= \epsilon^{\alpha\beta}\epsilon_{\alpha\beta}\lambda_\gamma + \epsilon^{\alpha\beta}\epsilon_{\alpha\gamma}\lambda_\beta = 2\lambda_\gamma + \epsilon_{\alpha\gamma}\lambda^\alpha \\
&= 2\lambda_\gamma - \lambda_\gamma = \lambda_\gamma.
\end{aligned} \tag{B.58}$$

Step 7: Check the remaining conditions

$$\begin{aligned}
&\left[\psi, \left\{\overline{Q}, \overline{Q}\right\}\right] = \left[F, \{Q, Q\}\right] = \left[F, \left\{\overline{Q}, \overline{Q}\right\}\right] = 0 \\
&\left[F, \left\{Q, \overline{Q}\right\}\right] = 2i\sigma^\mu\partial_\mu F.
\end{aligned} \tag{B.59}$$

We have, therefore, constructed a representation of the $N = 1$ supersymmetry algebra on a multiplet, ϕ, of fields, $\phi = (A; \psi; F)$. The representation is in terms of the commutators and anticommutators

$$\begin{aligned}
&\left[A, Q_\alpha\right] = 2i\psi_\alpha, \quad \left[A, \overline{Q}_{\dot\alpha}\right] = 0 \\
&\{\psi_\alpha, Q_\beta\} = -i\epsilon_{\alpha\beta}F, \quad \left\{\psi_\alpha, \overline{Q}_{\dot\beta}\right\} = \left(\sigma^\mu\right)_{\alpha\dot\beta}\partial_\mu A \\
&\left[F, Q_\alpha\right] = 0, \quad \left[F, \overline{Q}_{\dot\alpha}\right] = 2\partial_\mu\psi^\beta\left(\sigma^\mu\right)_{\beta\dot\alpha}.
\end{aligned} \tag{B.60}$$

Now, one introduces anticommuting spinor parameters (Grassmann numbers), ξ^α and $\overline{\xi}^{\dot{\alpha}} = (\xi^\alpha)^*$, which anticommute with everything fermionic and commute with everything bosonic.

With these spinor parameters, we can define infinitesimal variations of ϕ by

$$\delta\phi = -i\left[\phi, \xi Q + \overline{Q\xi}\right]. \tag{B.61}$$

For our multiplet, the transformation laws are

$$\begin{aligned} \delta A &= 2\xi\psi \\ \delta\psi &= -\xi F - i\partial_\mu A\sigma^\mu\overline{\xi} \\ \delta F &= -2i\partial_\mu\psi\sigma^\mu\overline{\xi}. \end{aligned} \tag{B.62}$$

The entire algebra can be written as

$$[\delta_1, \delta_2]\,\phi = 2i\left(\xi_1\sigma^\mu\overline{\xi}_2 - \xi_2\sigma^\mu\overline{\xi}_1\right)\partial_\mu\phi. \tag{B.63}$$

The degrees of freedom of multiplet ϕ, which is an unconstrained single real field, are four bosonic ones, Re A, Im A, Re F, and Im F, as well as four fermionic ones, Re ψ_1, Im ψ_1, Re ψ_2, and Im ψ_2. The multiplet has $4+4$ degrees of freedom and is the smallest possible number in four spacetime dimensions since any multiplet must contain a spinor, and spinors have at least two complex (Weyl) or four real (Majorana) components. The multiplet is, therefore, irreducible.

The algebra prevents $\delta\phi = 0$ since we would have ϕ =constant. The only trivial representation of supersymmetry is constant fields. If we had decided to use $[A, Q] = 0$ instead of $\left[A, \overline{Q}\right] = 0$ in step 2, we would have gotten an anti-chiral multiplet, $\overline{\phi}$. Such a multiplet can be constructed from ϕ by hermitian conjugation because we have $\left[A^\dagger, Q\right] = 0$. The components of this multiplet are $\overline{\phi} = \left(A^\dagger; \overline{\psi}; F^\dagger\right)$, and the transformation laws are

$$\begin{aligned} \delta A^\dagger &= 2\overline{\xi\psi} \\ \delta\overline{\psi} &= -F^\dagger\overline{\xi} + i\xi\sigma^\mu\partial_\mu A^\dagger \\ \delta F^\dagger &= 2i\xi\sigma^\mu\partial_\mu\overline{\psi}. \end{aligned} \tag{B.64}$$

Further chiral multiplets can be derived by giving each field, ϕ or $\overline{\phi}$, an additional Lorentz index. These multiplets are not irreducible. The seven steps can also be used in much more complicated situations such as supergravity.

Supergravity arises when we consider supersymmetry as a local symmetry and allow the group parameters, ξ and a_μ, in the super Poincare transformations to be functions of spacetime. Since the supersymmetry transformations contain the generator of translations, P^μ, we must consider translations that depend on the point of spacetime which means that we are considering a theory of general coordinate transformations of spacetime and, therefore, a theory of gravity. Hence, we refer to a theory of local supersymmetry as supergravity.

Manipulation of 2-spinors

Raising and lowering of spinor indices is accomplished by

$$\psi^{\alpha} = \epsilon^{\alpha\beta}\psi_{\beta}; \quad \overline{\psi}^{\dot{\alpha}} = \overline{\psi}_{\dot{\beta}}\epsilon^{\dot{\beta}\dot{\alpha}} \tag{B.65}$$

and

$$\psi_{\alpha} = \psi^{\beta}\epsilon_{\alpha\beta}; \quad \overline{\psi}_{\dot{\alpha}} = \epsilon_{\dot{\alpha}\dot{\beta}}\overline{\psi}^{\dot{\beta}}. \tag{B.66}$$

We also have

$$(\psi^{\alpha})^{*} = \overline{\psi}^{\dot{\alpha}}; \quad (\psi_{\alpha})^{*} = \overline{\psi}_{\dot{\alpha}}. \tag{B.67}$$

The mixed tensors are antisymmetric:

$$\epsilon_{\alpha}^{\beta} = -\epsilon_{\alpha}^{\beta} = \delta_{\alpha}^{\beta}; \quad \epsilon_{\dot{\alpha}}^{\dot{\beta}} = -\epsilon_{\dot{\alpha}}^{\dot{\beta}} = \delta_{\dot{\alpha}}^{\dot{\beta}} \tag{B.68}$$

and $(\epsilon_{\alpha\beta})^{*} = -\epsilon_{\dot{\alpha}\dot{\beta}}$. The contraction of unwritten spinor indices is defined as

$$\xi\psi = \xi^{\alpha}\psi_{\alpha} = -\xi_{\alpha}\psi^{\alpha}; \quad \overline{\psi\xi} = \overline{\xi}_{\dot{\alpha}}\overline{\psi}^{\dot{\alpha}} = -\overline{\psi}^{\dot{\alpha}}\overline{\xi}_{\dot{\alpha}}. \tag{B.69}$$

Complex conjugation also reverses the order of the spinors, i.e.,

$$(\xi\psi)^{*} = (\xi^{\alpha}\psi_{\alpha})^{*} = (\psi_{\alpha})^{*}(\xi^{\alpha})^{*} = \overline{\psi}_{\dot{\alpha}}\overline{\xi}^{\dot{\alpha}} = \overline{\psi\xi}. \tag{B.70}$$

We now drop the chiral notation with the dotted and undotted indices, and adopt the usual four-component spinor notation defined as follows.

Manipulation of 4-spinors

Dirac matrices can be constructed from the 2×2 matrices, σ^{μ} and $\overline{\sigma}^{\mu}$, as follows

$$\gamma^{\mu} = \begin{pmatrix} 0 & \sigma^{\mu} \\ \overline{\sigma}^{\mu} & 0 \end{pmatrix}, \quad \text{and} \quad \sigma^{\mu\nu} = \begin{pmatrix} \sigma^{\mu\nu} & 0 \\ 0 & \overline{\sigma}^{\mu\nu} \end{pmatrix}, \tag{B.71}$$

which satisfy

$$\{\gamma^{\mu}, \gamma^{\nu}\} = 2\eta^{\mu\nu}, \text{ and } \frac{i}{2}[\gamma^{\mu}, \gamma^{\nu}] = \sigma^{\mu\nu}. \tag{B.72}$$

Four-spinors are composed from chiral and anti-chiral two-spinors as

$$\psi = \begin{pmatrix} \chi_{\alpha} \\ \overline{\lambda}^{\dot{\alpha}} \end{pmatrix} \quad \overline{\psi} = \begin{pmatrix} \lambda^{\alpha} & \overline{\chi}_{\dot{\alpha}} \end{pmatrix}, \tag{B.73}$$

and the charge conjugate spinors are

$$\psi^{c} = \begin{pmatrix} \lambda_{\alpha} \\ \overline{\chi}^{\dot{\alpha}} \end{pmatrix} \quad \overline{\psi}^{c} = (\psi^{c})^{\dagger} A = \begin{pmatrix} \chi^{\alpha} & \overline{\lambda}_{\dot{\alpha}} \end{pmatrix}. \tag{B.74}$$

These spinors are related through 4×4 matrices

$$A = \begin{pmatrix} 0 & \mathbb{I} \\ \mathbb{I} & 0 \end{pmatrix}; \quad C = \begin{pmatrix} -\epsilon_{\alpha\beta} & 0 \\ 0 & -\epsilon^{\dot{\alpha}\dot{\beta}} \end{pmatrix}; \quad C^{-1} = \begin{pmatrix} \epsilon^{\alpha\beta} & 0 \\ 0 & \epsilon_{\dot{\alpha}\dot{\beta}} \end{pmatrix} \tag{B.75}$$

as

$$\overline{\psi} = \psi^{\dagger} A; \quad \psi^c = C\overline{\psi}^T \tag{B.76}$$

and

$$A\gamma_{\mu}A^{-1} = \gamma_{\mu}^{\dagger}; \quad C^{-1}\gamma_{\mu}C = -\gamma_{\mu}^{T}. \tag{B.77}$$

The matrix, γ_5, is defined by

$$\gamma_5 = \gamma_0\gamma_1\gamma_2\gamma_3; \quad (\gamma_5)^2 = -1; \quad \gamma_5 = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}, \tag{B.78}$$

and the projection operators, $\frac{1}{2}(\mathbb{I} \pm i\gamma_5)$, project out the chiral components, χ_{α} and $\overline{\lambda}^{\dot{\alpha}}$, of ψ. The 16 linearly independent matrices, $A, A\gamma_{\mu}, A\sigma_{\mu\nu}, iA\gamma_{\mu}\gamma_5$, and $A\gamma_5$, are hermitian. The 6 matrices, $C, \gamma_{\mu}\gamma_5 C$, and $\gamma_5 C$ are antisymmetric. The 10 matrices, $\gamma_{\mu}C$, and $\sigma_{\mu\nu}C$, are symmetric.

For any N-extended superalgebra, we can define Majorana spinors from our chiral $Q_{\alpha i}$

$$Q_i = \begin{pmatrix} Q_{\alpha i} \\ \overline{Q}^{\dot{\alpha} i} \end{pmatrix}; \quad \overline{Q}_i = \begin{pmatrix} Q_i^{\alpha} & \overline{Q}^i_{\dot{\alpha}} \end{pmatrix}. \tag{B.79}$$

In this four-spinor notation, the supersymmetry algebra of the Q's becomes

$$\begin{aligned} \left\{Q_i, \overline{Q}_j\right\} &= 2\left(\delta_{ij}\gamma^{\mu}P_{\mu} + i\mathrm{Im}Z_{ij} + i\gamma_5\mathrm{Re}Z_{ij}\right) \\ [Q, P_{\mu}] = 0; \quad [Q, M_{\mu\nu}] &= \frac{1}{2}\sigma_{\mu\nu}Q; \quad [Q, R] = i\gamma_5 Q. \end{aligned} \tag{B.80}$$

We also have a Majorana spinor, ξ^i, defined from $\xi^{\alpha i}$

$$\begin{aligned} \xi^{\alpha i}Q_{\alpha i} + \overline{Q}^i_{\dot{\alpha}}\overline{\xi}^{\dot{\alpha}}_i &= \overline{\xi}Q; \quad \delta\phi = -i\left[\phi, \overline{\xi}Q\right] \\ [\delta_1, \delta_2]\,\phi &= 2i\overline{\xi}_1\gamma^{\mu}\xi_2\partial_{\mu}\phi + 2i\overline{\xi}^i_1\left[\phi, \mathrm{Im}Z_{ij} + \gamma_5\mathrm{Re}Z_{ij}\right]\xi_2^j. \end{aligned} \tag{B.81}$$

A real (Majorana) form of the $N = 1$ chiral multiplet is found by calling the real and imaginary parts of the complex fields, A and $F^{\dagger}$, in the chiral multiplet by the names A, B and F, G and then constructing a Majorana spinor, ψ, from the chiral spinor, ψ_{α}, and its conjugate, $\overline{\psi}_{\dot{\alpha}}$. The variation transformation laws for the resulting multiplet, $\phi = (A, B; \psi; F, G)$, is found from equation (B.62) to be

$$\begin{aligned} \delta A &= \overline{\xi}\psi; \quad \delta B = \overline{\xi}\gamma_5\psi \\ \delta\psi &= -(F + \gamma_5 G)\xi - i\,\partial\!\!\!/(A + \gamma_5 B)\xi \\ \delta F &= -i\overline{\xi}\,\partial\!\!\!/\psi; \quad \delta G = i\overline{\xi}\gamma_5\,\partial\!\!\!/\psi. \end{aligned} \tag{B.82}$$

The presence of the γ_5 gives clear parity assignments: A and F are scalars, and B and G are pseudoscalars.

B.2.4. The general multiplet

If we had constructed the chiral multiplet without the chirality constraint, $[A, \overline{Q}_{\dot\alpha}] = 0$, then we could still use the idea behind the seven steps (which would now be more than seven) which is that the anticommutator, $\{Q, Q\}$, is known. We can derive a multiplet from a general ground state field, $C(x)$. The result is a larger multiplet

$$V = (C; \chi; M, N; A_\mu; \lambda; D) \tag{B.83}$$

with transformations given in four-component notation by

$$\begin{aligned}
\delta C &= \bar\xi\gamma_5\chi \\
\delta\chi &= (M + \gamma_5 N)\xi - i\gamma^\mu(A_\mu + \gamma_5\partial_\mu C)\xi \\
\delta M &= \bar\xi\,(\lambda - i\,\partial\!\!\!/\chi) \\
\delta N &= \bar\xi\gamma_5\,(\lambda - i\,\partial\!\!\!/\chi) \\
\delta A_\mu &= i\bar\xi\gamma_\mu\lambda + \bar\xi\partial_\mu\chi \\
\delta\lambda &= -i\sigma^{\mu\nu}\xi\partial_\mu A_\nu - \gamma_5\xi D \\
\delta D &= -i\bar\xi\,\partial\!\!\!/\gamma_5\lambda.
\end{aligned} \tag{B.84}$$

The scalar, M; the pseudoscalars, C, N, and D; and the vector, A_μ are a priori complex. The spinors, χ and λ, are Dirac spinors. In contrast to the chiral multiplet, we could impose a reality condition, $V = V^\dagger$, which is defined to mean that all components are real or Majorana (called the real general multiplet, general because there are no chiral constraints).

Irreducible submultiplets

The general multiplet has $8+8$ components and is not irreducible. We can see that the fields, λ, D, and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, transform among themselves and form an irreducible submultiplet called the curl multiplet, dV.

$$dV = (\lambda; F_{\mu\nu}; D) \tag{B.85}$$

with

$$\begin{aligned}
\delta\lambda &= -\frac{i}{2}\sigma^{\mu\nu}\xi F_{\mu\nu} - \gamma_5\xi D \\
\delta F_{\mu\nu} &= -i\bar\xi\,(\gamma_\mu\partial_\nu - \gamma_\nu\partial_\mu)\,\lambda \\
\delta D &= -i\bar\xi\,\partial\!\!\!/\gamma_5\lambda
\end{aligned} \tag{B.86}$$

which represents the supersymmetry algebra as long as $\partial_{[\kappa}F_{\mu\nu]} = 0$ holds, which means $F_{\mu\nu}$ is a curl.

Another submultiplet of the general multiplet is the chiral multiplet

$$\partial V = (M, N; \lambda - i \not{\partial}\chi; \partial^\mu A_\mu, D + \Box C) \tag{B.87}$$

Since dV and ∂V form submultiplets, we can constrain V not to contain one or the other. For $dV = 0$, the surviving components of V can be arranged into a chiral multiplet, $\phi = (A, C; \chi; -M, -N)$, where the scalar A is defined by the solution, $A_\mu = \partial_\mu A$, of $F_{\mu\nu} = 0$. For $\partial V = 0$, the surviving components form an irreducible multiplet called the linear multiplet

$$L = (C; \chi; A_\mu)\,, \tag{B.88}$$

where A_μ is divergence free, $\partial^\mu A_\mu = 0$. The variations given by equation (B.84) with $\partial V = 0$ are

$$\begin{aligned}
\delta C &= \bar{\xi}\gamma_5\chi \\
\delta\chi &= -i\gamma^\mu\left(A_\mu + \gamma_5\partial_\mu C\right)\xi \\
\delta A_\mu &= i\bar{\xi}\gamma_\mu\lambda + \bar{\xi}\partial_\mu\chi = i\bar{\xi}\gamma_\mu(i\not{\partial}\chi) + \bar{\xi}\partial_\mu\chi \\
&= -\bar{\xi}\gamma_\mu\gamma_\nu\partial^\nu\chi + \bar{\xi}\eta_{\mu\nu}\partial^\nu\chi = -\bar{\xi}\left(\gamma_\mu\gamma_\nu - \eta_{\mu\nu}\right)\partial^\nu\chi \\
&= i\bar{\xi}\sigma_{\mu\nu}\partial^\nu\chi,
\end{aligned} \tag{B.89}$$

where we have used

$$\sigma_{\mu\nu} = \frac{i}{2}\left[\gamma_\mu, \gamma_\nu\right] \quad \text{and} \quad \eta_{\mu\nu} = \frac{1}{2}\left\{\gamma_\mu, \gamma_\nu\right\}. \tag{B.90}$$

Products of multiplets

One can build a chiral multiplet, ϕ_3, from the product of two others, ϕ_1 and ϕ_2, with $\left[\mathcal{A}_1, \overline{Q}\right] = 0$ and $\left[\mathcal{A}_2, \overline{Q}\right] = 0$. We also have $\left[\mathcal{A}_1\mathcal{A}_2, \overline{Q}\right] = 0$ for all complex $\mathcal{A}$'s. In real components, the spinor in $\delta(A_1B_2 + B_1A_2)$ is then γ_5 times that in $\delta(A_1A_2 - B_1B_2)$. We now find the components of the product of two chiral multiplets, $\phi_3 = \phi_1 \cdot \phi_2$. First, to find $\mathcal{A}_3$, we use the fact that, with $\mathcal{A}_1 = A_1 + iB_1$, we have

$$\mathcal{A}_3 = \mathcal{A}_1\mathcal{A}_2 = (A_1 + iB_1)(A_2 + iB_2) = (A_1A_2 - B_1B_2) + i(A_1B_2 + B_1A_2), \tag{B.91}$$

so, since $\mathcal{A}_3 = A_3 + iB_3$, we have

$$\begin{aligned}
A_3 &= A_1A_2 - B_1B_2 \\
B_3 &= A_1B_2 + B_1A_2.
\end{aligned} \tag{B.92}$$

We can now find ψ_3 using A_3 and the variations given in equation (B.82)

$$\begin{aligned}\delta A_3 &= \bar{\xi}\psi_3 \\ \delta(A_1A_2 - B_1B_2) &= \delta(A_1A_2) - \delta(B_1B_2) \\ &= \delta(A_1)A_2 + A_1\delta(A_2) - \delta(B_1)B_2 - B_1\delta(B_2) \\ &= \bar{\xi}\psi_1A_2 + A_1\bar{\xi}\psi_2 - \bar{\xi}\gamma_5\psi_1B_2 - B_1\bar{\xi}\gamma_5\psi_2 \\ \Rightarrow \bar{\xi}\psi_3 &= \bar{\xi}\left(\psi_1A_2 + A_1\psi_2 - \gamma_5\psi_1B_2 - B_1\gamma_5\psi_2\right) \\ \psi_3 &= (A_1 - \gamma_5B_1)\psi_2 + (A_2 - \gamma_5B_2)\psi_1.\end{aligned} \tag{B.93}$$

Next we use variations, $\delta\psi_3$ and $\delta\overline{\psi}_3$ along with our previous results. After some algebra, we get F_3 and G_3, and the components of the inner product are summarized as

$$\begin{aligned}A_3 &= A_1A_2 - B_1B_2 \\ B_3 &= B_1A_2 + A_1B_2 \\ \psi_3 &= (A_1 - \gamma_5B_1)\,\psi_2 + (A_2 - \gamma_5B_2)\,\psi_1 \\ F_3 &= F_1A_2 + A_1F_2 + B_1G_2 + G_1B_2 + \overline{\psi}_1\psi_2 \\ G_3 &= G_1A_2 + A_1G_2 - B_1F_2 - F_1B_2 - \overline{\psi}_1\gamma_5\psi_2.\end{aligned} \tag{B.94}$$

The product of three multiplets is associative. Therefore, the product of any number of them is well defined.

Example:

As an example, we prove the formulae for the inner product components, F_3 and G_3, in equation (B.94). We use the identities

$$\begin{aligned}&\bar{\xi}\psi = \overline{\psi^c}\xi^c, \quad \bar{\xi}\gamma_\mu\gamma_5\psi = \overline{\psi^c}\gamma_\mu\gamma_5\xi^c, \\ &\bar{\xi}\gamma_5\psi = \overline{\psi^c}\gamma_5\xi^c, \quad \bar{\xi}\gamma_\mu\psi = -\overline{\psi^c}\gamma_\mu\xi^c, \quad \text{and} \quad \bar{\xi}\sigma_{\mu\nu}\psi = -\overline{\psi^c}\sigma_{\mu\nu}\xi^c\end{aligned} \tag{B.95}$$

along with the Majorana condition, $\psi = \psi^c$. We have

$$\begin{aligned}\delta\psi_3 &= -(F_3 + \gamma_5G_3)\xi - i\,\partial\!\!\!/(A_3 + \gamma_5B_3)\xi \\ &= -(F_3 + i\,\partial\!\!\!/A_3)\xi - \gamma_5(G_3 - i\,\partial\!\!\!/B_3)\xi.\end{aligned}$$

Now we use

$$
\begin{aligned}
\delta\psi_3 &= \delta\left[(A_1 - \gamma_5 B_1)\psi_2 + (1 \leftrightarrow 2)\right] \\
&= \left(\delta(A_1) - \gamma_5\delta(B_1)\right)\psi_2 + (A_1 - \gamma_5 B_1)\,\delta(\psi_2) + (1 \leftrightarrow 2) \\
&= \left(\bar{\xi}\psi_1 - \gamma_5\bar{\xi}\gamma_5\psi_1\right)\psi_2 \\
&\qquad - (A_1 - \gamma_5 B_1)\left[(F_2 + \gamma_5 G_2)\xi + i\,\partial\!\!\!/(A_2 + \gamma_5 B_2)\xi\right] + (1 \leftrightarrow 2) \\
&= -\overline{\psi}_1\psi_2\xi + \gamma_5\overline{\psi}_1\gamma_5\psi_2\xi \\
&\qquad - \left\{(A_1 - \gamma_5 B_1)\left[(F_2 + \gamma_5 G_2)\xi + i\,\partial\!\!\!/(A_2 + \gamma_5 B_2)\xi\right] + (1 \leftrightarrow 2)\right\} \\
&= -\left\{A_1F_2 + F_1A_2 + B_1G_2 + G_1B_2 + i\,\partial\!\!\!/(A_1A_2 - B_1B_2)\right\}\xi \\
&\qquad - \gamma_5\left\{A_1G_2 + A_2G_1 - B_1F_2 - B_2F_1 - i\,\partial\!\!\!/(A_1B_2 + B_1A_2)\right\}\xi \\
&\qquad\quad - \overline{\psi}_1\psi_2\xi + \gamma_5\overline{\psi}_1\gamma_5\psi_2\xi.
\end{aligned}
$$

Equating the terms

$$
\begin{aligned}
\delta\psi_3 &= -(F_3 + i\,\partial\!\!\!/A_3)\xi - \gamma_5(G_3 - i\,\partial\!\!\!/B_3)\xi \\
&= -\left\{A_1F_2 + F_1A_2 + B_1G_2 + G_1B_2 + i\,\partial\!\!\!/A_3\right\}\xi \\
&\qquad - \gamma_5\left\{A_1G_2 + A_2G_1 - B_1F_2 - B_2F_1 - i\,\partial\!\!\!/B_3\right\}\xi \\
&\qquad\quad - \overline{\psi}_1\psi_2\xi + \gamma_5\overline{\psi}_1\gamma_5\psi_2\xi
\end{aligned}
$$

gives

$$
\begin{aligned}
F_3 + i\,\partial\!\!\!/A_3 &= A_1F_2 + F_1A_2 + B_1G_2 + G_1B_2 + i\,\partial\!\!\!/A_3 + \overline{\psi}_1\psi_2 \\
F_3 &= A_1F_2 + F_1A_2 + B_1G_2 + G_1B_2 + \overline{\psi}_1\psi_2 \\
G_3 - i\,\partial\!\!\!/B_3 &= A_1G_2 + A_2G_1 - B_1F_2 - B_2F_1 - i\,\partial\!\!\!/B_3 - \overline{\psi}_1\gamma_5\psi_2 \\
G_3 &= A_1G_2 + A_2G_1 - B_1F_2 - B_2F_1 - \overline{\psi}_1\gamma_5\psi_2
\end{aligned}
$$

which proves our result.

□

Two other products of chiral multiplets can be defined. The first is denoted by $V = \phi_1 \times \phi_2$ and is a general, real multiplet with components

$$
\begin{aligned}
C &= A_1A_2 + B_1B_2 \\
\chi &= (B_1 - \gamma_5 A_1)\,\psi_2 + (B_2 - \gamma_5 A_2)\,\psi_1 \\
M &= -G_1A_2 - F_1B_2 - A_1G_2 - B_1F_2 \\
N &= F_1A_2 - G_1B_2 + A_1F_2 - B_1G_2 \\
A_\mu &= B_1 \overleftrightarrow{\partial}_\mu A_2 + B_2 \overleftrightarrow{\partial}_\mu A_1 + i\overline{\psi}_1\gamma_\mu\gamma_5\psi_2 \\
\lambda &= -\left(G_1 + \gamma_5 F_1 - i\,\partial\!\!\!/ B_1 - i\gamma_5\,\partial\!\!\!/ A_1\right)\psi_2 + (1 \leftrightarrow 2) \\
D &= -2F_1F_2 - 2G_1G_2 - 2\partial_\mu A_1\partial^\mu A_2 - 2\partial_\mu B_1\partial^\mu B_2 - i\overline{\psi}_1\,\overleftrightarrow{\partial\!\!\!/}\,\psi_2.
\end{aligned}
\tag{B.96}
$$

This product is symmetric under the interchange of ϕ_1 and ϕ_2 as is $\phi_1 \cdot \phi_2$ above. The last type of product is antisymmetric under this interchange, denoted $\phi_1 \wedge \phi_2$, and is a real multiplet with components

$$
\begin{aligned}
C &= A_1A_2 - B_1B_2 \\
\chi &= (A_1 + \gamma_5 B_1)\,\psi_2 + (A_2 + \gamma_5 B_2)\,\psi_1 \\
M &= F_1A_2 - G_1B_2 - A_1F_2 + B_1G_2 \\
N &= F_1B_2 + G_1A_2 - B_1F_2 - A_1G_2 \\
A_\mu &= A_1 \overleftrightarrow{\partial}_\mu A_2 + B_1 \overleftrightarrow{\partial}_\mu B_2 - i\overline{\psi}_1\gamma_\mu\psi_2 \\
\lambda &= -\left(F_1 - \gamma_5 G_1 - i\,\partial\!\!\!/ A_1 - i\,\partial\!\!\!/\gamma_5 B_1\right)\psi_2 - (1 \leftrightarrow 2) \\
D &= -2G_1F_2 + 2F_1G_2 + 2\partial_\mu B_1\partial^\mu A_2 - 2\partial_\mu A_1\partial^\mu B_2 + i\overline{\psi}_1\gamma_5\,\overleftrightarrow{\partial\!\!\!/}\,\psi_2.
\end{aligned}
\tag{B.97}
$$

Finally, we can also define a product for two general real multiplets, $V_3 = V_1 \cdot V_2$, which is again a general, real multiplet; is symmetric in V_1 and V_2; and has components

$$
\begin{aligned}
C_3 &= C_1C_2 \\
\chi_3 &= C_1\chi_2 + \chi_1C_2 \\
M_3 &= C_1M_2 + M_1C_2 - \frac{1}{2}\overline{\chi}_1\gamma_5\chi_2 \\
N_3 &= C_1N_2 + N_1C_2 - \frac{1}{2}\overline{\chi}_1\chi_2 \\
A_3^\mu &= C_1A_2^\mu + A_1^\mu C_2 + \frac{i}{2}\overline{\chi}_1\gamma^\mu\gamma_5\chi_2 \\
\lambda_3 &= C_1\lambda_2 + \frac{1}{2}\left(N_1 + \gamma_5 M_1 + i\,\partial\!\!\!/ C_1 - i\,A\!\!\!/_1\gamma_5\right)\chi_2 + (1 \leftrightarrow 2) \\
D_3 &= C_1D_2 + D_1C_2 - M_1M_2 - N_1N_2 - \partial_\mu C_1\partial^\mu C_2 - A_{1\mu}A_2^\mu \\
&\qquad + \overline{\lambda}_1\chi_2 + \overline{\chi}_1\lambda_2 - \frac{i}{2}\overline{\chi}_1\,\overleftrightarrow{\partial\!\!\!/}\,\chi_2.
\end{aligned}
\tag{B.98}
$$

Dimensions of the fields

We define the mass dimension, Δ, of a chiral multiplet as that of its A-component. This definition gives

$$\begin{aligned}\dim A &= \dim B = \Delta \equiv \dim \phi \\ \dim \psi &= \Delta + \frac{1}{2} \\ \dim F &= \dim G = \Delta + 1.\end{aligned} \tag{B.99}$$

The parameter, ξ, must have dimension $-\frac{1}{2}$ so that $i\bar{\xi}_1\gamma^\mu\xi_2$ is a length, the dimensions follow from equation (B.62). We also have

$$\begin{aligned}\dim(\phi \cdot \phi) &= \Delta_1 + \Delta_2 \\ \dim \mathbf{T}\phi &= \Delta + 1 \\ \dim (\phi \cdot \mathbf{T}\phi)_F &= 2\Delta + 2 \\ \dim (\phi^n)_F &= n\Delta + 1,\end{aligned} \tag{B.100}$$

where $\mathbf{T}$ is a generalization of $i\,\partial\!\!\!/$ and is defined by its action on a multiplet.

$$\mathbf{T}\phi = (F, G; i\,\partial\!\!\!/\psi; -\Box A, -\Box B)\,, \tag{B.101}$$

and $\mathbf{T}\mathbf{T}\phi = -\Box\phi$.

B.3. Quaternions and the scalar multiplet

We now introduce a method of constructing supersymmetry by combining complex and Grassmann valued functions into a quaternion and identifying it with a supersymmetric scalar (chiral) multiplet.

A supermultiplet consists of component fields that are made up of complex-valued functions. For example, a scalar (chiral) multiplet,

$$\Phi = (A, \psi, F)\,, \tag{B.102}$$

has one Weyl spinor, ψ, which has two complex components; one complex scalar, A; and one complex pseudoscalar, F, for a total of $4+4$ real components. We can think of a product of scalar multiplets which returns a scalar multiplet as a product from $\mathbb{R}^8 \times \mathbb{R}^8 \to \mathbb{R}^8$ defined at a specific point, x^μ, in spacetime (the interaction point).

Supersymmetric theories require a multiplet product. This product takes two multiplets and returns a new multiplet

$$\phi_1 \cdot \phi_2 = \phi_3. \tag{B.103}$$

The number of fermion components, n, is determined by the dimension of the underlying spacetime. We can make this statement clear by noticing that a supersymmetric theory must have equal numbers of fermion and boson degrees of freedom. Since, upon quantization, the number of fermion degrees of freedom is given by $n/2$. where n is the number of spinor components, and the gauge fields have $d-2$ physical degrees of freedom, we, therefore, have a relationship between the number of spinor components, n, and the dimension of spacetime, d. Explicitly, we have

$$d-2=n/2 \Rightarrow d=2+n/2. \tag{B.104}$$

Now from the theory of Clifford algebras, we know that the number of spinor components must be equal to the rank of the Dirac matrices which are $n = 2,4,8,16$ (for minimal supersymmetry). We conclude that the dimension of spacetime can be $d = 3,4,6,10$. Hence, the supersymmetry rule that there are equal numbers of bosons and fermions implies the dimension of the space.

An interacting field theory cannot have two superfields annihilate at a point leaving vacuum. This fact implies that the product of superfields can have no zero divisors and and leads us to a division algebra formulation.

B.3.1. Complex division algebra

We begin by writing a pair of multiplets as complex fields

$$\begin{aligned}\phi_1 &= A_1 e_0 + \psi e_1 \\ \phi_2 &= A_2 e_0 + \chi e_1,\end{aligned} \tag{B.105}$$

where A_i is a real or complex scalar field, and ψ and χ are real or compex Grassmann fields.

Now, we simply multiply the two multiplets together to form the product multiplet.

$$\phi_1 \cdot \phi_2 = \phi_3 = A_3 e_0 + \lambda e_1. \tag{B.106}$$

$$\begin{aligned}A_3 &= A_1 A_2 - \psi\chi \\ \lambda &= A_1\chi + \psi A_2\end{aligned} \tag{B.107}$$

If these components are real numbers, then our lagrangian density will be constructed from the A term of the product. The product appearing in a mass term of a lagrangian will be $(\phi_1 \cdot \phi_1)_A$, which is the A component of

$$\begin{aligned}A_3 &= A^2 \\ \lambda &= 2A\psi.\end{aligned} \tag{B.108}$$

Therefore, the boson field has a mass and the fermion field must be massless. A (3-point) interaction term would be of the form $(\phi \cdot \phi \cdot \phi)_A$, which is the A component of

$$\begin{aligned}A_3 &= A^3 \\ \lambda &= 3A^2\chi.\end{aligned} \tag{B.109}$$

Similarly for an n-point interaction. In these cases, the fermion is not interacting.

Now, write fields as complex, $A = A + iB$ and $\psi = \psi_0 + i\psi_1$, with the analogous definitions for χ. The product is then

$$\begin{aligned} A_3 &= A_1A_2 - B_1B_2 - \psi_0\chi_0 + \psi_1\chi_1 \\ B_3 &= A_1B_2 + B_1A_2 - \psi_0\chi_1 - \psi_1\chi_0 \\ \lambda_0 &= A_1\chi_0 - B_1\chi_1 + \psi_0A_2 - \psi_1B_2 \\ \lambda_1 &= B_1\chi_0 + A_1\chi_1 + \psi_0B_2 + \psi_1A_2. \end{aligned} \tag{B.110}$$

We can write this result in the form

$$\begin{aligned} A_3 &= A_1A_2 - B_1B_2 - \psi_0\chi_0 + \psi_1\chi_1 \\ B_3 &= A_1B_2 + B_1A_2 - \psi_0\chi_1 - \psi_1\chi_0 \\ \lambda_a &= (A_1\delta_{ab} + B_1\epsilon_{ab})\chi_b + (A_2\delta_{ab} + B_2\epsilon_{ab})\psi_b. \end{aligned} \tag{B.111}$$

The lagrangian would have mass given by the A term of the product of two identical mulitplets

$$\begin{aligned} A_3 &= A^2 - B^2 \\ B_3 &= 2AB \\ \lambda_a &= 2(A\delta_{ab} + B\epsilon_{ab})\psi_b. \end{aligned} \tag{B.112}$$

The A term contains no fermion. Hence, the fermion remains massless. The fields A and B have masses (one is a negative mass tachyon). For a 3-point interaction,

$$\begin{aligned} A_3 &= A^3 - 3B^2A + 2B\psi_0\psi_1 \\ B_3 &= 3A^2B - B^3 - 2A\psi_1\psi_0 \\ \lambda_0 &= 3(A^2 - B^2)\psi_0 - 6AB\psi_1 \\ \lambda_1 &= 3(A^2 - B^2)\psi_1 + 6AB\psi_0. \end{aligned} \tag{B.113}$$

If we let $\psi_1 = \gamma^5\psi_0$ and B is a pseudoscalar, then our A_3 term would represent a cubic scalar interaction, A^3; a scalar-pseudoscalar interaction, B^2A; and an interaction between the pseudoscalar and the massless fermion. Instead, we would like to take ψ_0 and ψ_1 to be pure Grassmann without any γ^5 involved. It seems that keeping the fermion components as mere Grassmann numbers does not give a Lorentz invariant interaction.

B.3.2. Quaternion division algebra

We now write a pair of scalar multiplets as quaternion fields

$$\begin{aligned} \phi_1 &= A_1e_0 + F_1e_1 + \psi_1e_2 + \psi_2e_3 \\ \phi_1 &= A_2e_0 + F_2e_1 + \chi_1e_2 + \chi_2e_3, \end{aligned} \tag{B.114}$$

where A_i is a complex scalar field; F_i is a complex pseudoscalar field; and ψ_i and χ_i, where $i \in \{1, 2\}$, are compex Grassmann fields.

Now, we simply multiply the two quaternions together to form the product multiplet

$$\phi_1 \cdot \phi_2 = \phi_3 = A_3 e_0 + F_3 e_1 + \lambda_1 e_2 + \lambda_2 e_3. \quad \text{(B.115)}$$

The result is given in Table B.3.2.

Table B.1. The product of two scalar multiplets.

$\cdot$	A_2e_0	F_2e_1	χ_1e_2	χ_2e_3
A_1e_0	$A_1A_2e_0$	$A_1F_2e_1$	$A_1\chi_1e_2$	$A_1\chi_2e_3$
F_1e_1	$F_1A_2e_1$	$-F_1F_2e_0$	$F_1\chi_1e_3$	$-F_1\chi_2e_2$
ψ_1e_2	$\psi_1A_2e_2$	$-\psi_1F_2e_3$	$-\psi_1\chi_1e_0$	$\psi_1\chi_2e_1$
ψ_2e_3	$\psi_2A_2e_3$	$\psi_2F_2e_2$	$-\psi_2\chi_1e_1$	$-\psi_2\chi_2e_0$

The diagonal 2×2 quadrants of Table B.3.2. give the bosonic piece of the product (by examining the basis units) while the off-diagonal quadrants give the fermionic piece. We see the following pattern:

$$\begin{aligned} \text{Boson} \cdot \text{Boson} &= \text{Boson} \\ \text{Boson} \cdot \text{Fermion} &= \text{Fermion} \\ \text{Fermion} \cdot \text{Boson} &= \text{Fermion} \\ \text{Fermion} \cdot \text{Fermion} &= \text{Boson} \end{aligned} \quad \text{(B.116)}$$

which is the familiar supersymmetric relationship. The components of the product multiplet are read from the table

$$\begin{aligned} A_3 &= A_1A_2 - F_1F_2 - \psi_1\chi_1 - \psi_2\chi_2 \\ F_3 &= A_1F_2 + F_1A_2 + \psi_1\chi_2 - \psi_2\chi_1 \\ \lambda_1 &= A_1\chi_1 + \psi_1A_2 - F_1\chi_2 + \psi_2F_2 \\ \lambda_2 &= A_1\chi_2 + F_1\chi_1 + \psi_2A_2 - \psi_1F_2. \end{aligned} \quad \text{(B.117)}$$

Now, we write the explicit real components of the complex fields, $A = A + iB$, and $F = iF - G$, where we are using the notation of Sohnius [75], and B and G are pseudoscalar fields. We also write the Grassmann components of the spinors as $\psi_1 = \psi_0 + i\psi_1$ and

$\psi_2 = \psi_2 - i\psi_3$ with the analogous definitions for χ_i. Inserting the new definitions gives

$$\begin{aligned}
A_3 &= A_1A_2 - B_1B_2 + F_1F_2 - G_1G_2 - \psi_0\chi_0 + \psi_1\chi_1 - \psi_2\chi_2 + \psi_3\chi_3 \\
B_3 &= B_1A_2 + A_1B_2 - G_1F_2 - F_1G_2 - \psi_0\chi_1 - \psi_1\chi_0 + \psi_2\chi_3 + \psi_3\chi_2 \\
F_3 &= F_1A_2 + A_1F_2 + B_1G_2 + G_1B_2 + \psi_0\chi_3 - \psi_1\chi_2 + \psi_2\chi_1 - \psi_3\chi_0 \\
G_3 &= G_1A_2 + A_1G_2 - B_1F_2 - F_1B_2 - \psi_0\chi_2 - \psi_1\chi_3 + \psi_2\chi_0 + \psi_3\chi_1 \\
\lambda_0 &= A_1\chi_0 - B_1\chi_1 + G_1\chi_2 + F_1\chi_3 + \psi_0A_2 - \psi_1B_2 - \psi_2G_2 - \psi_3F_2 \\
\lambda_1 &= B_1\chi_0 + A_1\chi_1 + F_1\chi_2 - G_1\chi_3 + \psi_0B_2 + \psi_1A_2 - \psi_2F_2 + \psi_3G_2 \\
\lambda_2 &= -G_1\chi_0 + F_1\chi_1 + A_1\chi_2 + B_1\chi_3 + \psi_0G_2 - \psi_1F_2 + \psi_2A_2 + \psi_3B_2 \\
\lambda_3 &= F_1\chi_0 + G_1\chi_1 - B_1\chi_2 + A_1\chi_3 - \psi_0F_2 - \psi_1G_2 - \psi_2B_2 + \psi_3A_2.
\end{aligned} \tag{B.118}$$

Define $\psi^T = (\psi_0, \psi_1, \psi_2, \psi_3)$, $\overline{\psi} = \psi^T\tilde{\gamma}^0$, and the following Majorana representations of the gamma matrices

$$\begin{aligned}
\gamma^0 &= \begin{pmatrix} 0 & -\sigma_2 \\ -\sigma_2 & 0 \end{pmatrix} = i\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \\
\gamma^1 &= \begin{pmatrix} i\sigma_3 & 0 \\ 0 & i\sigma_3 \end{pmatrix} = i\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \\
\gamma^2 &= \begin{pmatrix} 0 & -\sigma^2 \\ \sigma^2 & 0 \end{pmatrix} = i\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \\
\gamma^3 &= \begin{pmatrix} -i\sigma_1 & 0 \\ 0 & -i\sigma_1 \end{pmatrix} = i\begin{pmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \\
\gamma^5 &= \begin{pmatrix} i\sigma_2 & 0 \\ 0 & -i\sigma_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \\
H = \tilde{\gamma}^0\tilde{\gamma}^1 &= \begin{pmatrix} 0 & -I \\ -I & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}.
\end{aligned} \tag{B.119}$$

The $\tilde{\gamma}$ matrices are defined by $\gamma^\mu = i\tilde{\gamma}^\mu$ for $\mu \in \{0, ..., 3\}$. We can write our product as

$$\begin{aligned} A_3 &= A_1A_2 - B_1B_2 + F_1F_2 - G_1G_2 + \overline{\psi}H\chi \\ B_3 &= B_1A_2 + A_1B_2 - G_1F_2 - F_1G_2 - \overline{\psi}H\gamma^5\chi \\ F_3 &= F_1A_2 + A_1F_2 + B_1G_2 + G_1B_2 + \overline{\psi}\chi \\ G_3 &= G_1A_2 + A_1G_2 - B_1F_2 - F_1B_2 - \overline{\psi}\gamma^5\chi \\ \lambda &= (A_1 - \gamma^5B_1)\chi + (A_2 - \gamma^5B_2)\psi - H(F_1 + \gamma^5G_1)\chi + H(F_2 + \gamma^5G_2)\psi \end{aligned} \tag{B.120}$$

which is similar in form to that of Sohnius [75]

$$\begin{aligned} A_3 &= A_1A_2 - B_1B_2 \\ B_3 &= B_1A_2 + A_1B_2 \\ F_3 &= F_1A_2 + A_1F_2 + B_1G_2 + G_1B_2 + \overline{\psi}\chi \\ G_3 &= G_1A_2 + A_1G_2 - B_1F_2 - F_1B_2 - \overline{\psi}\gamma_5\chi \\ \lambda &= (A_1 - \gamma_5B_1)\,\chi + (A_2 - \gamma_5B_2)\,\psi. \end{aligned} \tag{B.121}$$

Notice that our expression differs from the usual one (B.94) only in terms involving the auxiliary fields, F and G. Further, our F-term, which is the one used in the lagrangian density, is the same. When one forms the product of supermultiplets which are the same like, for example, in the lagrangian (B.124):

$$L = \left(\frac{1}{2}\phi \cdot T\phi - \frac{m}{2}\phi \cdot \phi - \frac{g}{3}\phi \cdot \phi \cdot \phi\right)_F, \tag{B.122}$$

then the terms with H in them actually vanish due to the Grassmann nature of the fermion components and cancellation. For example, for $\phi \cdot \phi$, we have

$$\begin{aligned} A_3 &= A^2 - B^2 + F^2 - G^2 \\ B_3 &= 2AB - 2FG \\ F_3 &= 2AF + 2BG + \overline{\psi}\psi \\ G_3 &= 2AG - 2BF - \overline{\psi}\gamma^5\psi \\ \lambda &= 2(A - \gamma^5B)\psi \end{aligned} \tag{B.123}$$

which is precisely what equation (B.121) reduces to also, modulo auxiliary fields in A_3 and B_3. Mass and interaction terms will be the same. Another possibility is to take the limit of F and G going to zero, in which case the two scalar multiplets are identical. In any case, we find this relationship between product multiplets and the division algebras quite interesting. It would worthwhile to examine the octonion division algebra in this context. We now turn to a specific example of a supersymmetric field theory.

B.4. The Wess-Zumino model

The Wess-Zumino model consists of a single chiral multiplet, ϕ, in renormalizable self-interaction. We begin by constructing the lagrangian which describes the motion of fields in time. A lagrangian must be a scalar density, which means its variation is a divergence, i.e., $\delta L = \partial^\mu K_\mu$, so that the action, $\int d^4x\ L$, is an invariant constant. We can use the fact that

(a) Any F-component of a chiral multiplet is a scalar density.

(b) Any D-component of a general multiplet is a scalar density.

We can construct our lagrangian from the chiral multiplet out of terms which comprise the F-component of a product of multiplets. The lagrangian must contain derivatives. If the maximal number of derivatives is to be two (for the boson fields), we must include exactly one of the T-operations. Demanding that $\dim L = 4$, determines the dimension of the multiplet as $\Delta = 1$. The most general renormalizable lagrangian for a single chiral multiplet (i.e., one with no coupling constants of negative dimension) is the Wess-Zumino lagrangian given by

$$L = \left(\frac{1}{2}\phi \cdot \mathbf{T}\phi - \frac{m}{2}\phi \cdot \phi - \frac{g}{3}\phi \cdot \phi \cdot \phi\right)_F = L_0 + L_m + L_g. \tag{B.124}$$

In components (using the definitions of the inner product given by equation (B.94)), we have

$$\begin{aligned} L_0 &= \frac{1}{2}\left[\phi \cdot \mathbf{T}\phi\right]_F = \frac{1}{2}\left[(A, B; \psi; F, G) \cdot (F, G; i\,\partial\!\!\!/\psi; -\Box A, -\Box B)\right]_F \\ &= \frac{1}{2}\Big[\big(AF - BG, FB + AG; (A - \gamma_5 B) i\,\partial\!\!\!/\psi + (F - \gamma_5 G)\psi; F^2 - A\Box A \\ &\qquad -B\Box B + G^2 + \overline{\psi} i\,\partial\!\!\!/\psi, GF - A\Box B + B\Box A - FG - \overline{\psi}\gamma_5 i\,\partial\!\!\!/\psi\big)\Big]_F \\ &= \frac{1}{2}\Big[\big(AF - BG, FB + AG; i(A - \gamma_5 B)\,\partial\!\!\!/\psi + (F - \gamma_5 G)\psi; F^2 + \partial_\mu A \partial^\mu A \\ &\qquad +\partial_\mu B \partial^\mu B + G^2 + i\overline{\psi}\,\partial\!\!\!/\psi + \text{4-div.}, -i\overline{\psi}\gamma_5 i\,\partial\!\!\!/\psi + \text{4-div.}\big)\Big]_F \\ &= \frac{1}{2}\left[F^2 + \partial_\mu A \partial^\mu A + \partial_\mu B \partial^\mu B + G^2 + i\overline{\psi}\,\partial\!\!\!/\psi\right] + \text{4-div.} \\ L_0 &= \frac{1}{2}\left[\partial_\mu A \partial^\mu A + \partial_\mu B \partial^\mu B + i\overline{\psi}\,\partial\!\!\!/\psi + F^2 + G^2\right] + \text{4-div.} \end{aligned} \tag{B.125}$$

$$
\begin{aligned}
L_m &= -\frac{m}{2}\left[\phi\cdot\phi\right]_F = -\frac{m}{2}\left[(A,B;\psi;F,G)\cdot(A,B;\psi;F,G)\right]_F \\
&= -\frac{m}{2}\left[A^2-B^2, 2AB; 2(A-\gamma_5 B)\psi; 2AF+2BG+\overline{\psi}\psi, 2AG-2BF-\overline{\psi}\gamma_5\psi\right]_F \\
&= -\frac{m}{2}\left(2AF+2BG+\overline{\psi}\psi\right) \\
L_m &= -m\left(AF+BG\right) - \frac{m}{2}\overline{\psi}\psi
\end{aligned}
\tag{B.126}
$$

$$
\begin{aligned}
L_g &= -\frac{g}{3}\left[\phi\cdot\phi\cdot\phi\right]_F = -\frac{g}{3}\left[(\phi\cdot\phi)\cdot\phi\right]_F \\
&= -\frac{g}{3}\left[\left(A^2-B^2, 2AB; 2(A-\gamma_5 B)\psi; 2AF+2BG+\overline{\psi}\psi, 2AG-2BF-\overline{\psi}\gamma_5\psi\right)\cdot\phi\right]_F \\
&= -\frac{g}{3}\Big[\left(2AF+2BG+\overline{\psi}\psi\right)A + (A^2-B^2)F + 2ABG \\
&\qquad + (2AG-2BF)B - \left(\overline{\psi}\gamma_5\psi\right)B + \left[2\overline{(A-\gamma_5 B)\psi}\right]\psi\Big] \\
&= -\frac{g}{3}\Big[2A^2F + 2ABG + \overline{\psi}A\psi + A^2F - B^2F + 2ABG \\
&\qquad + 2ABG - 2B^2F - \overline{\psi}\gamma_5 B\psi + 2\left((A-\gamma_5 B)\psi\right)^\dagger\gamma^0\psi\Big] \\
&= -\frac{g}{3}\left[3A^2F + 6ABG + \overline{\psi}(A-\gamma_5 B)\psi - 3B^2F + 2\psi^\dagger(A+\gamma_5 B)\gamma^0\psi\right] \\
&= -\frac{g}{3}\left[3A^2F + 6ABG + \overline{\psi}\left(A-\gamma_5 B\right)\psi - 3B^2F + 2\overline{\psi}(A-\gamma_5 B)\psi\right] \\
L_g &= -g\left[(A^2-B^2)F + 2ABG + \overline{\psi}\left(A-\gamma_5 B\right)\psi\right].
\end{aligned}
\tag{B.127}
$$

To summarize,

$$
\begin{aligned}
L &= L_0 + L_m + L_g \\
L_0 &= \frac{1}{2}\left[\partial_\mu A\partial^\mu A + \partial_\mu B\partial^\mu B + i\overline{\psi}\,\partial\!\!\!/\psi + F^2 + G^2\right] + \text{4-div.} \\
L_m &= -m\left(AF+BG\right) - \frac{m}{2}\overline{\psi}\psi \\
L_g &= -g\left[(A^2-B^2)F + 2ABG + \overline{\psi}\left(A-\gamma_5 B\right)\psi\right].
\end{aligned}
\tag{B.128}
$$

We can now derive the Euler-Lagrange equations of motion for each of the component fields.

$$
\begin{aligned}
0 &= \frac{\delta L}{\delta \psi} - \partial_\mu \frac{\delta L}{\delta(\partial_\mu \psi)} = -\frac{m}{2}\overline{\psi} - g\overline{\psi}(A - \gamma_5 B) - \partial_\mu \left(\frac{i}{2}\gamma^\mu \overline{\psi}\right) \\
&= -\overline{\psi}\left(\frac{m}{2} + g(A - \gamma_5 B)\right) - \frac{i}{2}\, \partial\!\!\!/ \overline{\psi} \\
&\Rightarrow i\, \partial\!\!\!/ \overline{\psi} = -m\overline{\psi} - 2g\overline{\psi}(A - \gamma_5 B) \\
0 &= \frac{\delta L}{\delta \overline{\psi}} - \partial_\mu \frac{\delta L}{\delta(\partial_\mu \overline{\psi})} = \frac{i}{2}\, \partial\!\!\!/ \psi - \frac{m}{2}\psi - g(A - \gamma_5 B)\psi \\
&\Rightarrow i\, \partial\!\!\!/ \psi = m\psi + 2g(A - \gamma_5 B)\psi
\end{aligned}
$$

$$
\begin{aligned}
0 &= \frac{\delta L}{\delta F} = F - mA - g(A^2 - B^2) \\
&\Rightarrow F = mA + g(A^2 - B^2) \\
0 &= \frac{\delta L}{\delta G} = G - mB - 2gAB \\
&\Rightarrow G = mB + 2gAB
\end{aligned}
$$

$$
\begin{aligned}
0 &= \frac{\delta L}{\delta A} - \partial_\mu \frac{\delta L}{\delta(\partial_\mu A)} = -mF - 2gAF - 2BG - g\overline{\psi}\psi - \partial_\mu \partial^\mu A \\
&\Rightarrow -\Box A = mF + 2g\left(AF + BG + \frac{1}{2}\overline{\psi}\psi\right) \\
0 &= \frac{\delta L}{\delta B} - \partial_\mu \frac{\delta L}{\delta(\partial_\mu B)} = -mG + 2gBF - 2AG + g\overline{\psi}\gamma_5\psi - \partial_\mu \partial^\mu B \\
&\Rightarrow -\Box B = mG - 2g\left(BF - AG + \frac{1}{2}\overline{\psi}\gamma_5\psi\right).
\end{aligned}
$$

To summarize,

$$
\begin{aligned}
F &= mA + g(A^2 - B^2) \\
G &= mB + 2gAB \\
i\, \partial\!\!\!/ \psi &= m\psi + 2g(A - \gamma_5 B)\psi \\
-\Box A &= mF + 2g\left(AF + BG + \frac{1}{2}\overline{\psi}\psi\right) \\
-\Box B &= mG - 2g\left(BF - AG + \frac{1}{2}\overline{\psi}\gamma_5\psi\right).
\end{aligned}
\tag{B.129}
$$

From the form of $\mathbf{T}\phi$ and that of $\phi \cdot \phi$, we see that we can write these equations simply as

$$
\mathbf{T}\phi = m\phi + g\phi \cdot \phi. \tag{B.130}
$$

Notice that the equations for F and G are purely algebraic and contain no derivatives of fields. These fields are called auxiliary fields since their equations of motion do not describe propagation in space and time. We can, therefore, use these equations of motion to eliminate these fields in both the lagrangian and the equations of motion for the other fields. The result is the "on-shell" lagrangian

$$\begin{aligned} L = \frac{1}{2}\left(\partial_\mu A \partial^\mu A - m^2 A^2\right) + \frac{1}{2}\left(\partial_\mu B \partial^\mu B - m^2 B^2\right) + \frac{1}{2}\overline{\psi}\left(i\not\partial - m\right)\psi \\ - mgA\left(A^2 + B^2\right) - g\overline{\psi}\left(A - \gamma_5 B\right)\psi - \frac{g^2}{2}\left(A^2 + B^2\right)^2 \end{aligned} \tag{B.131}$$

and the "on-shell" equations of motion

$$\begin{aligned} \left(\Box + m^2\right) A &= -mg\left(3A^2 + B^2\right) - 2g^2 A\left(A^2 + B^2\right) - g\overline{\psi}\psi \\ \left(\Box + m^2\right) B &= -2mgAB - 2g^2 B\left(A^2 + B^2\right) + g\overline{\psi}\gamma_5\psi \\ \left(i\not\partial - m\right)\psi &= 2g(A - \gamma_5 B)\psi. \end{aligned} \tag{B.132}$$

This lagrangian contains all seven possible parity-invariant interaction terms (or diagrams), $A^3, AB^2, A^4, B^4, A^2B^2, A\overline{\psi}\psi$, and $B\overline{\psi}\gamma_5\psi$. It contains the three possible mass terms. The masses, however, are all the same. The seven couplings are fixed in terms of the mass and a single coupling g due to the fact that the on-shell lagrangian still transforms as a density under the on-shell transformations (derived by eliminating F and G from equation (B.82))

$$\begin{aligned} \delta A &= \overline{\xi}\psi \\ \delta B &= \overline{\xi}\gamma_5\psi \\ \delta\psi &= -\left(i\not\partial + m + g(A + \gamma_5 B)\right)(A + \gamma_5 B)\xi. \end{aligned} \tag{B.133}$$

We can see that it is actually a particular relationship between the coupling constants of the lagrangian that constitutes the supersymmetry of the model. Also, independent renormalizations of m and g are not needed since all of the contributions to the linearly and quadratically divergent terms cancel due to fermion and boson loops contributing with opposite signs. This relationship between fermions and bosons is a supersymmetry. There is only a single, logarithmically divergent infinity which can be absorbed into a wavefunction renormalization. This fact is common for all the fields in the multiplet. This absence of infinities is understood in terms of non-renormalization theorems derived in superspace which will be discussed later.

B.4.1. On-shell and off-shell representations

One of the important problems in supersymmetry is to find off-shell formulations for the higher-N extended models or to prove that they do not exist. We use the Wess-Zumino model to explain what is meant by on-shell and off-shell representations.

The transformation laws (B.82) represent the supersymmetry algebra on the $4 + 4$ field components of the chiral multiplet. These laws are independent of either the lagrangian or the dynamics. Each of the terms in the Wess-Zumino lagrangian separately transforms as a density under those transformations. This separation is called "off-shell supersymmetry."

The complete set of field equations (B.129) is again a multiplet (a chiral multiplet in this case), but the structure of these field equations separates them into two classes, algebraic equations for some fields (auxiliary) and wave equations for the others (dynamical degrees of freedom). The elimination of the auxiliary fields is accomplished by enforcing their equations of motion on the lagrangian and transformation rules. This procedure is not supersymmetric since all of the field equations together form an irreducible multiplet. We can retain supersymmetry only by taking all fields on-shell by enforcing their equations of motion.

The on-shell lagrangian (B.131) is still a density under the on-shell transformations (B.133), but the transformations have become dependent on the model (m and g). There is no part of the lagrangian that separately transforms as a density under them.

The most serious drawback of the on-shell formulation appears when we calculate the commutator of two of the transformations. The result is $2i\bar{\xi}_1 \not{\partial}\xi_2$ on A and B as it should be, but

$$[\delta_1, \delta_2]\,\psi = 2i\bar{\xi}_1 \not{\partial}\xi_2\psi - \gamma^\mu \left(i \not{\partial} - m - 2g(A - \gamma_5 B)\right)\psi\bar{\xi}_1\gamma_\mu\xi_2 \tag{B.134}$$

on ψ. Thus, the on-shell algebra closes only if the equations of motion hold, which sets the last term in equation (B.134) to zero. This situation could, in principle, be disastrous for quantum corrections where the fields must be taken off-shell, away from their classical paths through configuration space. What this problem means for the on-shell supersymmetric theories is unclear. If there exists some unknown off-shell version, then there is no problem. If the theory is only intrinsically on-shell supersymmetric or if there are several competing off-shell versions, then things are also unclear.

As for any continuous symmetry, there should be a Noether current associated with the transformations and the lagrangian. The current is a Majorana spinor-vector current given by

$$J_\mu = \frac{\partial}{\partial\bar{\xi}}\left(\delta A\frac{\partial L}{\partial\partial^\mu A} + \delta B\frac{\partial L}{\partial\partial^\mu B} + \delta\psi\frac{\partial L}{\partial\partial^\mu\psi} - k_\mu\right) \tag{B.135}$$

if $\delta L = \partial^\mu k_\mu$. The term k_μ is determined only up to terms of the form $\partial^\nu a_{\mu\nu}$ with $a_{\mu\nu} = -a_{\nu\mu}$. One form of the supercurrent is

$$J_\mu = \not{\partial}(A - \gamma_5 B)\gamma_\mu\psi + im\gamma_\mu(A - \gamma_5 B)\psi + ig\gamma_\mu(A - \gamma_5 B)^2\psi. \tag{B.136}$$

It is derived by beginning with the off-shell lagrangian

$$\begin{aligned}\delta L &= i\bar{\xi}\,\not{\partial}\left[\frac{1}{2}\phi\cdot\mathbf{T}\phi - \frac{m}{2}\phi\cdot\phi - \frac{g}{3}\circ\cdot\phi\cdot\phi\right]_\psi + \frac{1}{4}\Box\delta(A^2 + B^2)\\ &= \partial^\mu k_\mu\end{aligned}$$

with

$$-k_\mu = \frac{1}{4}\bar{\xi}(A+\gamma_5 B)\left[\gamma_\mu, \partial\!\!\!/\right]\psi - \frac{1}{2}\bar{\xi}\partial_\mu(A+\gamma_5 B)\psi - \frac{i}{2}\bar{\xi}(F+\gamma_5 G)\gamma_\mu\psi$$
$$+ im\bar{\xi}(A+\gamma_5 B)\gamma_\mu\psi + ig\bar{\xi}\gamma_\mu(A+\gamma_5 B)^2\psi.$$

The result for the on-shell lagrangian is the same with F and G expressed in terms of A and B. The other terms in J_μ are

$$\delta A\frac{\partial L}{\partial\partial^\mu A} + \delta B\frac{\partial L}{\partial\partial^\mu B} + \delta\psi\frac{\partial L}{\partial\partial^\mu\psi}$$
$$= \bar{\xi}\partial_\mu(A+\gamma_5 B)\psi + \frac{i}{2}\bar{\xi}(F+\gamma_5 G)\gamma_\mu\psi + \frac{1}{2}\bar{\xi}\ \partial\!\!\!/\gamma_\mu(A+\gamma_5 B)\psi,$$

and with the improvement term, $\partial^\nu a_{\mu\nu} = -\frac{i}{2}\partial^\nu\left[\sigma_{\mu\nu}(A+\gamma_5 B)\psi\right]$, we get the current (B.136).

Using the equations of motion (B.132) and the fact that $\psi(\overline{\psi}\psi) = \gamma_5\psi(\overline{\psi}\gamma_5\psi)$, one can prove that the supercurrent is conserved. It is important that the supercurrent exist for interacting models. For a toy model given by the lagrangian

$$L_{\text{toy}} = \frac{1}{2}\partial^\mu A\partial_\mu A + \frac{i}{2}\overline{\psi}\ \partial\!\!\!/\psi,$$

we have the supersymmetry

$$\delta A = \bar{\xi}\psi; \quad \delta\psi = -i\ \partial\!\!\!/ A\xi$$

and the current

$$J_\mu = \partial\!\!\!/ A\gamma_\mu\psi$$

which is conserved due to the free equations of motion, $\Box A = \partial\!\!\!/\psi = 0$. There is no way to make this toy model interact and preserve a conserved supercurrent though since it is not supersymmetric in a non-trivial sense. It cannot be since it violates the fermions = bosons rule.

B.5. Spontaneous symmetry breaking

Using self-interactions of chiral multiplets, we can spontaneously break supersymmetry. We break up the Wess-Zumino lagrangian into superkinetic and superpotential parts

$$L = \frac{1}{2}\left(\phi\cdot\mathbf{T}\phi\right)_F - \left[V(\phi)\right]_F, \text{ with } V(\phi) = \lambda\phi + \frac{m}{2}\phi\cdot\phi + \frac{g}{3}\phi\cdot\phi\cdot\phi, \tag{B.137}$$

where we have included a possible $\lambda\phi$ term in the lagrangian which we can use to see exactly how and why it can be eliminated. In this notation, the equations of motion are

$$\mathbf{T}\phi = V'(\phi), \tag{B.138}$$

where V' is the derivative of the function V with respect to its argument.

Since all of the terms in $[V(\phi)]_F$ are either linear in F or G, or quadratic in ψ, we have a scaling equation

$$[V(\phi)]_F = \left(F\frac{\partial}{\partial F} + G\frac{\partial}{\partial G} + \frac{1}{2}\overline{\psi}\frac{\partial}{\partial \overline{\psi}} \right) [V(\phi)]_F . \tag{B.139}$$

To evaluate this equation in a general way, we use an adapted chain rule and the inner product of chiral multiplets

$$\frac{\partial}{\partial F}V(\phi) = \frac{\partial \phi}{\partial F} \cdot \frac{d}{d\phi}V(\phi) = (0,0;0;1,0) \cdot V'(\phi) = (0,0;0;[V'(\phi)]_A, -[V'(\phi)]_B) . \tag{B.140}$$

Using this rule and the equation of motion (B.138), we get

$$\frac{\partial}{\partial F}[V(\phi)]_F = [V'(\phi)]_A = F, \tag{B.141}$$

and a similar procedure gives

$$\begin{aligned} \frac{\partial}{\partial G}[V(\phi)]_F &= [V'(\phi)]_B = G \\ \frac{\partial}{\partial \overline{\psi}}[V(\phi)]_F &= [V'(\phi)]_\psi . \end{aligned} \tag{B.142}$$

Now, we can use F and G to denote $[V'(\phi)]_A$ and $[V'(\phi)]_B$ rather than independent fields. The superpotential becomes

$$[V(\phi)]_F = F^2 + G^2 + \frac{1}{2}\overline{\psi}[V'(\phi)]_\psi . \tag{B.143}$$

The "true" potential U must also contain the non-kinetic terms in $\frac{1}{2}(\phi \cdot \mathbf{T}\phi)_F$ so that

$$U = -\frac{1}{2}\left(F^2 + G^2\right) + [V(\phi)]_F = \frac{1}{2}\left(F^2 + G^2\right) + \frac{1}{2}\overline{\psi}[V'(\phi)]_\psi . \tag{B.144}$$

We see that, for the vacuum, where $\langle\psi\rangle = 0$ (otherwise Lorentz invariance would be spontaneously broken), U is never negative. This positivity is a consequence of the positivity of the energy in any supersymmetric theory, spontaneously broken or not.

Supersymmetry will be spontaneously broken if and only if the lowest energy (vacuum energy) is larger than zero,

$$\langle 0| E |0\rangle \neq 0 \Leftrightarrow Q_\alpha |0\rangle \neq 0, \tag{B.145}$$

due to

$$E_{\text{min}} = \langle 0| E |0\rangle = \frac{1}{4}\sum_{\alpha=1}^{4} |Q_\alpha |0\rangle|^2 . \tag{B.146}$$

Since, on the other hand,

$$E_{\text{min}} = \langle 0| U |0\rangle , \tag{B.147}$$

we find that supersymmetry is spontaneously broken if and only if the minimum of the potential is positive.

For the Wess-Zumino model, the potential is $\langle U\rangle = \frac{1}{2}\langle F\rangle^2 + \frac{1}{2}\langle G\rangle^2$ with

$$F = \lambda + mA + g(A^2 - B^2), \text{ and } G = mB + 2gAB.$$

The minimum of the potential is at $\langle F\rangle = \langle G\rangle = 0$, which means it is at

$$\begin{aligned}\langle A\rangle &= -\frac{1}{2g}(m \pm \sqrt{m^2 - 4g\lambda}) \\ \langle B\rangle &= 0\end{aligned} \tag{B.148}$$

for $m^2 \geq 4g\lambda$ and

$$\begin{aligned}\langle A\rangle &= -\frac{m}{2g} \\ \langle B\rangle &= \pm\frac{1}{2g}\sqrt{4g\lambda - m^2}\end{aligned} \tag{B.149}$$

for $m^2 \leq 4g\lambda$. The field redefinitions[6],

$$A^{new} = A - \langle A\rangle , \quad B^{new} = B - \langle B\rangle ,$$

will then eliminate $\lambda\phi$ from the lagrangian and shift one of the minima to $A = B = 0$.

B.6. Superspace

First, consider ordinary quantum fields, $\phi(x)$, which depend only on the four coordinates x^μ of Minkowski space. Translations of these coordinates are generated by the operators, P_μ, so we can consider $\phi(x)$ to have been translated from $x^\mu = 0$

$$\phi(x) = e^{ix\cdot P}\phi(0)\phi(x) = e^{-ix\cdot P}. \tag{B.150}$$

This transformation law is compatible with the multiplication law,

$$e^{iy\cdot P}e^{-ix\cdot P} = e^{i(x+y)\cdot P}, \tag{B.151}$$

which holds because the operators, P_μ, commute with each other.

[6]A further redefinition, $\psi^{new} = \exp(-1/4\pi\gamma_5)\psi$, is necessary to recover manifest parity invariance if $\langle B\rangle = 0$.

We can use the formula

$$e^{-B}Ae^{B}=\sum_{n=0}^{\infty}\frac{1}{n!}[A,B]_{(n)} \tag{B.152}$$

with

$$[A,B]_{(0)}=A;[A,B]_{(n+1)}=\left[[A,B]_{(n)},B\right] \tag{B.153}$$

so that $\phi(x)$ is completely determined be the properties at any given point, x_0, by the multiple commutators of $\phi(x_0)$ with P_μ to arbitrary order.

For a Lorentz rotation, we have

$$e^{\frac{i}{2}\lambda\cdot M}\phi(x)e^{-\frac{i}{2}\lambda\cdot M}=e^{ix'\cdot P}e^{\frac{i}{2}\lambda\cdot M}\phi(x)e^{-\frac{i}{2}\lambda\cdot M}e^{-ix'\cdot P}, \tag{B.154}$$

where $\lambda\cdot M=\lambda^{\mu\nu}M_{\mu\nu}$ and x' is given implicitly by

$$e^{\frac{i}{2}\lambda\cdot M}e^{ix\cdot P}=e^{ix'\cdot P}e^{\frac{i}{2}\lambda\cdot M}. \tag{B.155}$$

For infinitesimal λ, we can calculate x'

$$\begin{aligned}\left(1+\frac{i}{2}\lambda\cdot M\right)e^{ix\cdot P}&=e^{ix\cdot P}e^{-ix\cdot P}\left(1+\frac{i}{2}\lambda\cdot M\right)e^{ix\cdot P}\\&=e^{ix\cdot P}\left(1+\frac{i}{2}\lambda\cdot M+\left[\frac{i}{2}\lambda\cdot M,ix\cdot P\right]\right)\\&=e^{ix\cdot P}\left(1+\frac{i}{2}\lambda\cdot M+i\lambda^{\mu\nu}x_\mu P_\nu\right)\\&=e^{i(x^\mu+x^\nu\lambda^\mu_\nu)P_\mu}\left(1+\frac{i}{2}\lambda\cdot M\right)+\mathcal{O}(\lambda^2).\end{aligned} \tag{B.156}$$

For finite λ, we would get

$$\begin{aligned}e^{ix'\cdot P}e^{\frac{i}{2}\lambda\cdot M}&=e^{ix^\alpha(\eta^\mu_\alpha+\lambda^\mu_\alpha+\cdots)P_\mu}e^{\frac{i}{2}\lambda\cdot M}\\e^{ix'\cdot P}&=e^{ix^\alpha(e^\lambda)^\mu_\alpha P_\mu}\\\Rightarrow x'^\mu&=x^\alpha(e^\lambda)^\mu_\alpha=(e^{-\lambda})^\mu_\alpha x^\alpha\end{aligned} \tag{B.157}$$

which are the expected transformation laws for the coordinates under Lorentz transformations. Finally, the action of the Lorentz transformation on $\phi(0)$ can at most be some linear transformation which acts on the unwritten indices which $\phi(0)$, and hence $\phi(x)$, may have

$$e^{\frac{i}{2}\lambda\cdot M}\phi(0)e^{-\frac{i}{2}\lambda\cdot M}=e^{-\frac{i}{2}\lambda\cdot\Sigma}\phi(0), \tag{B.158}$$

where $\Sigma_{\mu\nu}$ are some matrix representations of the algebra of the $M_{\mu\nu}$ and a further Lorentz transformation does not see the Σ and acts directly on $\phi(0)$, so we have

$$e^{\frac{i}{2}\lambda\cdot M}\phi(x)e^{-\frac{i}{2}\lambda\cdot M}=e^{-\frac{i}{2}\lambda\cdot\Sigma}\phi(e^{-\lambda}x). \tag{B.159}$$

The differential version of this equation is

$$[\phi(x), M_{\mu\nu}] = i\left(x_\mu \partial_\nu - x_\nu \partial_\nu\right)\phi(x) + \Sigma_{\mu\nu}\phi(x). \tag{B.160}$$

To understand these ideas further, let g be an arbitrary element of a group G which contains a subgroup H.(This group could be the Poincare' group with the Lorentz subgroup.) We now define equivalence classes of elements of G by the right and left cosets of g with respect to H. g is equivalent to g' if there exists an element, $h \in H$, such that $g' = g \circ h$ or, equivalently, $g^{-1} \circ g \in H$. The set of all cosets is a manifold denoted G/H.

A set of group elements, $L(x)$, labeled by as many parameters as necessary, parameterizes the manifold if each coset contains exactly one of the L's. Once we have chosen a parameterization, $L(x)$, each group element, g, can be uniquely decomposed into a product, $g = L(x) \circ h$, where L is the representative member of the coset to which g belongs and h connects L to g within the coset. A product of g with an arbitrary group element, and in particular with some $L(x)$, will thus define another L and an h according to

$$g \circ L(x) = L(x') \circ h. \tag{B.161}$$

Here x' and h are in general functions of both x and g: $x' = x'(x, g)$ and $h = h(x, g)$. The Minkowski space considered at the beginning of this section was the coset manifold Poincare'/Lorentz. We parametrized it by

$$L(x) = e^{ix\cdot P}$$

and determined x' and h explicitly as

$$\left.\begin{array}{l} x' = x + y \\ h = \mathbb{I} \end{array}\right\} \text{ for } g = L(y)$$
$$\left.\begin{array}{l} x' = e^{-\lambda}x \\ h = g \end{array}\right\} \text{ for } g = e^{\frac{i}{2}\lambda\cdot M} \in H. \tag{B.162}$$

That case was particularly simple because $[P, P] = 0$ and $[P, M] = P$. The quantum fields were written as

$$\phi(x) = L(x)\phi(0)L^{-1}(x), \tag{B.163}$$

and the action of a group element on them was completely determined,

$$g\phi(x)g^{-1} = L(x')h\phi(0)h^{-1}L^{-1}(x'), \tag{B.164}$$

once we knew how h acted on $\phi(0)$,

$$h\phi(0)h^{-1} = e^{-\frac{i}{2}\lambda\cdot\Sigma}\phi(0). \tag{B.165}$$

This case is not the most general since the translations are an abelian invariant subgroup, which will not always be the case.

The supersymmetry algebra is not the Lie algebra of a group since it involves anticommutators. We must still be able to exponentiate the algebra to get group elements

$$e^{i\theta Q}e^{i\overline{Q\theta}} = e^{\text{something}}. \tag{B.166}$$

If we assume that the θ^{α} and $\overline{\theta}^{\dot{\alpha}}$ are anticommuting spinorial quantities, then we can express commutators like $\left[\theta Q, \overline{Q\theta}\right]$ in terms of only anticommutators. In this way, the the Campbell-Baker-Hausdorff formula,

$$e^{A}e^{B} = e^{\sum_{n=1}^{\infty}\frac{1}{n!}C_n(A,B)}, \tag{B.167}$$

where the $C_n(A, B)$ are defined in terms of the commutators of A and B, can be changed into anticommutators and used to define a multiplication.

Using these anticommuting parameters, we can form a group from the supersymmetry algebra called the *Super-Poincaré group* with typical elements

$$g = e^{ix\cdot P + i\theta Q + i\overline{Q\theta} + \frac{i}{2}\lambda\cdot M}. \tag{B.168}$$

Superspace is the coset space Super-Poincaré/Lorentz. There are infinitely many different ways to parameterize this manifold. The most commonly used way gives the "real," or symmetric, superspace parametrized by

$$L(x, \theta, \overline{\theta}) = e^{ix\cdot P + i\theta^{\alpha}Q_{\alpha} + iQ_{\dot{\alpha}}\overline{\theta}^{\dot{\alpha}}}. \tag{B.169}$$

The space has eight coordinates, four bosonic ones, x^{μ}, and four fermionic ones, θ^{α}, if we assume that $\overline{\theta}^{\dot{\alpha}} = (\theta^{\alpha})^{*}$.

Using equation (B.167), multiplication law (B.161) gives

$$L(y, \xi, \overline{\xi})L(x, \theta, \overline{\theta}) = L(x', \theta', \overline{\theta}') \tag{B.170}$$

with transformed coordinates

$$\begin{aligned} x'^{\mu} &= x^{\mu} + y^{\mu} + i\xi\sigma^{\mu}\overline{\theta} - i\theta\sigma^{\mu}\overline{\xi} \\ \theta' &= \theta + \xi \\ \overline{\theta}' &= \overline{\theta} + \overline{\xi}. \end{aligned} \tag{B.171}$$

Multiplication with an element of the Lorentz group gives

$$e^{\frac{i}{2}\lambda\cdot M}L(x, \theta, \overline{\theta}) = L(x', \theta', \overline{\theta}')e^{\frac{i}{2}\lambda\cdot M} \tag{B.172}$$

with transformed coordinates

$$\begin{aligned} x'^{\mu} &= \left(e^{-\lambda}\right)^{\mu}_{\ \nu}x^{\nu} \\ \theta' &= \theta e^{-\frac{i}{4}\lambda^{\mu\nu}\sigma_{\mu\nu}} \\ \overline{\theta}' &= e^{\frac{i}{4}\lambda^{\mu\nu}\overline{\sigma}_{\mu\nu}}\overline{\theta}. \end{aligned} \tag{B.173}$$

B.6.1. Representations on superfields

Let $r : G \to M$ be a matrix representation of a group. G, such that $r(g_1)r(g_2) = r(g_1 \circ g_2)$. $r(g)$, $r^*(g)$, $r^{-1T}(g)$, and $r^{-1\dagger}(g)$ are all representations, and u_α, $u_{\dot\alpha}$, u^α, and $u^{\dot\alpha}$ are the vectors on which these representations act, respectively. The group transformations on the vectors are then

$$\begin{aligned} u'_\alpha &= r_\alpha^\beta u_\beta \\ u'_{\dot\alpha} &= (r^*)_{\dot\alpha}^{\dot\beta} u_{\dot\beta} = u_{\dot\beta} \left(r^\dagger\right)_{\dot\alpha}^{\dot\beta} \\ u'^\alpha &= \left(r^{-1T}\right)_\beta^\alpha u^\beta = u^\beta \left(r^{-1}\right)_\beta^\alpha \\ u'^{\dot\alpha} &= \left(r^{-1\dagger}\right)_{\dot\beta}^{\dot\alpha} u^{\dot\beta} = u^{\dot\beta} \left(r^{-1*}\right)_{\dot\beta}^{\dot\alpha}. \end{aligned} \tag{B.174}$$

If the group is a Lie group, we have infinitesimal group elements for which the representations take the form

$$\begin{aligned} r_\alpha^\beta &\approx (\mathbb{I} + i\epsilon\Sigma)_\alpha^\beta; \quad \text{generator: } \Sigma \\ (r^*)_{\dot\alpha}^{\dot\beta} &\approx (\mathbb{I} - i\epsilon\Sigma^*)_{\dot\alpha}^{\dot\beta}; \quad \text{generator: } -\Sigma^* \\ \left(r^{-1T}\right)_\beta^\alpha &\approx \left(\mathbb{I} - i\epsilon\Sigma^T\right)_\beta^\alpha; \quad \text{generator: } -\Sigma^T \\ \left(r^{-1\dagger}\right)_{\dot\beta}^{\dot\alpha} &\approx \left(\mathbb{I} + i\epsilon\Sigma^\dagger\right)_{\dot\beta}^{\dot\alpha}; \quad \text{generator: } \Sigma^\dagger, \end{aligned} \tag{B.175}$$

and the infinitesimal variation of a vector becomes

$$\delta u_\alpha = u'_\alpha - u_\alpha = i\epsilon\Sigma_\alpha^\beta u_\beta. \tag{B.176}$$

Similar equations hold for $u_{\dot\alpha}$, u^α, and $u^{\dot\alpha}$. The variations of tensors are defined accordingly:

$$\begin{aligned} \delta t_{\alpha\beta} &= i\epsilon\Sigma_\alpha^\gamma t_{\gamma\beta} + i\epsilon\Sigma_\beta^\gamma t_{\alpha\gamma} \\ \delta t_{\alpha\dot\beta} &= i\epsilon\Sigma_\alpha^\gamma t_{\gamma\dot\beta} - i\epsilon\left(\Sigma^*\right)_{\dot\beta}^{\dot\gamma} t_{\alpha\dot\gamma} = i\epsilon\Sigma_\alpha^\gamma t_{\gamma\dot\beta} - i\epsilon t_{\alpha\dot\gamma}\left(\Sigma^\dagger\right)_{\dot\beta}^{\dot\gamma} \\ \delta t_\alpha^\beta &= i\epsilon\Sigma_\alpha^\gamma t_\gamma^\beta - i\epsilon\left(\Sigma^T\right)_\gamma^\beta t_\alpha^\gamma = i\epsilon\Sigma_\alpha^\gamma t_\gamma^\beta - i\epsilon t_\alpha^\gamma\left(\Sigma\right)_\gamma^\beta. \end{aligned} \tag{B.177}$$

Also

$$\delta t_\alpha^\alpha = \delta t_{\dot\alpha}^{\dot\alpha} = 0; \quad \delta\delta_\alpha^\beta = \delta\delta_{\dot\alpha}^{\dot\beta} = 0. \tag{B.178}$$

We define a superfield in the same way as equation (B.163)

$$\phi(x,\theta,\overline{\theta}) = L(x,\theta,\overline{\theta})\phi(0,0,0)L^{-1}(x,\theta,\overline{\theta}). \tag{B.179}$$

Notice that this also allows us to conclude that, if two superfields have the same lowest dimensional component, they are equivalent everywhere. For any group element, the action on ϕ is given by the coordinate transformations (B.171 and B.173) in conjunction with

$$e^{\frac{i}{2}\lambda\cdot M}\phi(0,0,0)e^{-\frac{i}{2}\lambda\cdot M} = e^{-\frac{i}{2}\lambda\cdot\Sigma}\phi(0,0,0). \tag{B.180}$$

Infinitesimally, we can write the group action on a superfield $\phi(x,\theta,\overline{\theta})$ as

$$\begin{aligned}
\delta_\xi\phi &= -i\left[\phi,\xi Q\right] = -i\xi^\alpha r(Q_\alpha)\phi \\
\delta_{\overline{\xi}}\phi &= -i\left[\phi,\overline{Q\xi}\right] = i\overline{\xi}^{\dot\alpha} r(\overline{Q}_{\dot\alpha})\phi \\
\delta_y\phi &= -i\left[\phi,y\cdot P\right] = -iy\cdot r(P)\phi \\
\delta_\lambda\phi &= -\frac{i}{2}\left[\phi,\lambda\cdot M\right] = -\frac{i}{2}\lambda\cdot r(M)\phi
\end{aligned} \tag{B.181}$$

with the differential operator representation, r, of the algebra given by

$$\begin{aligned}
r(Q_\alpha) &= i\frac{\partial}{\partial\theta^\alpha} - \left(\sigma^\mu\overline{\theta}\right)_\alpha\partial_\mu \\
r(\overline{Q}_{\dot\alpha}) &= -i\frac{\partial}{\partial\overline{\theta}^{\dot\alpha}} + (\theta\sigma^\mu)_{\dot\alpha}\partial_\mu \\
r(P_\mu) &= i\partial_\mu \\
r(M_{\mu\nu}) &= i\left(x_\mu\partial_\nu - x_\nu\partial_\mu\right) + \frac{1}{2}\theta^\alpha\sigma_{\mu\nu}\frac{\partial}{\partial\theta^\alpha} - \frac{1}{2}\left(\overline{\theta}\overline{\sigma}_{\mu\nu}\right)^{\dot\alpha}\frac{\partial}{\partial\overline{\theta}^{\dot\alpha}} + \Sigma_{\mu\nu}.
\end{aligned} \tag{B.182}$$

The optional chiral transformations which are generated by R act as

$$\delta_\alpha\phi = -i\left[\phi,\alpha R\right] = -i\alpha r(R)\phi \tag{B.183}$$

with

$$r(R) = \theta^\alpha\frac{\partial}{\partial\theta^\alpha} - \overline{\theta}^{\dot\alpha}\frac{\partial}{\partial\overline{\theta}^{\dot\alpha}} + n, \tag{B.184}$$

where n is the chiral weight of the superfield and must be real in order to render the R-transformations compact.

If ϕ has no spinor indices, the first two equations in equation (B.181) can be written

$$[\phi,Q] = r(Q)\phi, \quad \left[\phi,\overline{Q}\right] = r(\overline{Q})\phi.$$

The differentiations are defined by

$$\frac{\partial}{\partial\theta^\alpha}\theta^\beta = \delta^\beta_\alpha; \quad \frac{\partial}{\partial\overline{\theta}^{\dot\alpha}}\overline{\theta}^{\dot\beta} = \delta^{\dot\beta}_{\dot\alpha} \tag{B.185}$$

The Campbell-Baker-Hausdorff formula, along with the definition of the superfield (B.179), mean that a superfield is actually defined as a Taylor expansion in θ and $\overline{\theta}$ with coefficients which are themselves local fields over Minkowski space. Since the third powers of θQ and $\overline{Q\theta}$ are zero due to the vanishing of the square of each component, $(\theta)^2\theta = \theta^\alpha\epsilon_{\alpha\beta}\theta^\beta\theta = \left(\theta^1\theta^2 - \theta^2\theta^1\right)\theta = 0$, the series terminates, and the most general superfield is

$$V(x,\theta,\overline{\theta}) = C - i\theta\chi + i\overline{\chi}'\overline{\theta} - \frac{i}{2}\theta^2(M - iN) + \frac{i}{2}\overline{\theta}^2(M + iN) - \theta\sigma^\mu\overline{\theta}A_\mu$$
$$+ i\overline{\theta}^2\theta\left(\lambda - \frac{i}{2}\,\partial\!\!\!/\overline{\chi}'\right) - i\theta^2\overline{\theta}\left(\overline{\lambda}' - \frac{i}{2}\overline{\partial\!\!\!/}\chi\right) - \frac{1}{2}\theta^2\overline{\theta}^2\left(D + \frac{1}{2}\Box C\right). \tag{B.186}$$

Assuming that it has no overall Lorentz indices, the superfield contains as Taylor coefficients four complex scalar fields $C(x)$, $M(x)$, $N(x)$, and $D(x)$; one complex vector $A_\mu(x)$; two spinors $\chi(x)$ and $\lambda(x)$ in the $(\frac{1}{2}, 0)$ representation; and two unrelated spinors $\overline{\chi}'(x)$ and $\overline{\lambda}'(x)$ in the $(0, \frac{1}{2})$ representation of the Lorentz group. Altogether, there are 16 bosonic + 16 fermionic field components.

If we define $(\theta^\alpha)^* = \overline{\theta}^{\dot{\alpha}}$ and impose the rule that fermionic quantities reverse their order under complex conjugation, we have

$$\left(L(x,\theta,\overline{\theta})\right)^\dagger = L^{-1}(x,\theta,\overline{\theta}), \tag{B.187}$$

and a reality condition on a superfield

$$\left(V(x,\theta,\overline{\theta})\right)^\dagger = V(x,\theta,\overline{\theta}) \tag{B.188}$$

is covariant since $V^\dagger$ is a superfield just as V was. We can impose this condition on the superfield (B.186) and find

$$\begin{aligned} &C = C^\dagger; \quad M = M^\dagger; \quad N = N^\dagger; \quad D = D^\dagger \\ &\overline{\lambda}' = \lambda^\dagger; \quad \overline{\chi}' = \chi^\dagger \end{aligned} \tag{B.189}$$

so that the total number of component fields is now only 8 + 8. This superfield is called the real, or general, superfield. In order to maintain this reality condition, the supersymmetry transformations must be restricted by $(\xi^\alpha)^\dagger = \overline{\xi}^{\dot{\alpha}}$ and the chiral weight of a real V must be zero.

The transformation laws under the supersymmetry transformations for the components of V are calculated by comparing coefficients in the expressions $\delta V = \delta C - i\theta\delta\chi + \cdots$ and $\delta_\xi V + \delta_{\overline{\xi}} V$ given by equation (B.181) and operating with equation (B.182) on V. The result is the transformation laws (B.84) for the real general multiplet. We see that the component fields of a general superfield form a general multiplet. This representation is not, however, an irreducible representation. We need to impose constraints on the superfields to obtain an irreducible representation.

We need to impose supersymmetric conditions on the superfields to reduce the number of degrees of freedom that they describe. To impose these constraints, we need to develop a particular kind of covariant derivative.

Covariant spinor derivatives

The associativity of group multiplication can be interpreted to mean that the right action and the left action commute. Correspondingly, the left action of one L on another,

$$L(y,\xi,\overline{\xi})L(x,\theta,\overline{\theta}) \approx \left[1 - iy^\mu r(P_\mu) - i\xi^\alpha r(Q_\alpha) + i\overline{\xi}^{\dot{\alpha}} r(\overline{Q}_{\dot{\alpha}})\right] L(x,\theta,\overline{\theta}), \tag{B.190}$$

which we interpreted as a transformation of the superfield, commutes with the right action, which we can write as

$$L(x,\theta,\overline{\theta})L(y,\xi,\overline{\xi}) \approx \left[1 - y^\mu D_\mu + \xi^\alpha D_\alpha - \overline{\xi}^{\dot{\alpha}} \overline{D}_{\dot{\alpha}}\right] L(x,\theta,\overline{\theta}). \tag{B.191}$$

We can explicitly calculate the D's and find

$$\begin{aligned} D_\mu &= \partial_\mu \\ D_\alpha &= \frac{\partial}{\partial\theta^\alpha} - i\left(\sigma^\mu\overline{\theta}\right)_\alpha \partial_\mu \\ \overline{D}_{\dot{\alpha}} &= -\frac{\partial}{\partial\overline{\theta}^{\dot{\alpha}}} + i\left(\theta\sigma^\mu\right)_{\dot{\alpha}} \partial_\mu. \end{aligned} \tag{B.192}$$

Whereas the left action induces a mapping of the manifold G/H on itself, which is a realization of the group, the right action induces an anti-realization. (The order is inverted.) The associative law for group multiplication, $g_1 \circ (g_2 \circ g_3) = (g_1 \circ g_2) \circ g_3$, implies

$$\begin{aligned} \{D, r(Q)\} = \{\overline{D}, r(Q)\} &= 0 \\ \{D, r(\overline{Q})\} = \{\overline{D}, r(\overline{Q})\} &= 0. \end{aligned} \tag{B.193}$$

The commutators of the D's with each other

$$\begin{aligned} \{D, D\} = \{\overline{D}, \overline{D}\} &= 0 \\ \left\{D_\alpha, \overline{D}_{\dot{\beta}}\right\} &= 2i\left(\sigma^\mu\right)_{\alpha\dot{\beta}} \partial_\mu \\ [D, \partial_\mu] = [\overline{D}, \partial_\mu] = [\partial_\mu, \partial_\nu] &= 0 \end{aligned} \tag{B.194}$$

resemble the fact that the right action induces an anti-realization on superspace. Having defined the D's off by a factor of i from $r(Q)$, this algebra is formally the same as that for the $r(Q)$. The conventions for D and $\overline{D}$ were chosen such that $(D_\alpha\phi)^\dagger = \overline{D}_{\dot{\alpha}}\overline{\phi}$ with ϕ bosonic.

The anticommutation relations (B.193) have as a consequence that, if ϕ is a superfield, then so are $D\phi$, $\overline{D}\phi$ and $\partial_\mu\phi$. The only change is in the representation matrix $\Sigma_{\mu\nu}$ of the Lorentz generators since the derivatives carry Lorentz indices, and in the chiral weight, n, which is raised by one for $D\phi$ and lowered by one for $\overline{D}\phi$. These derivatives are, therefore, covariant under super-Poincaré transformations of the superspace coordinates.

Like all covariant derivatives, D and $\overline{D}$ can be used to impose covariant conditions on superfields. The simplest conditions are those for a chiral superfield ϕ with

$$\overline{D}_{\dot{\alpha}}\phi = 0 \tag{B.195}$$

and an anti-chiral superfield $\overline{\phi}$ with

$$D_\alpha \overline{\phi} = 0. \tag{B.196}$$

Note that, if ϕ is chiral, then $\phi^\dagger$ is anti-chiral and that a superfield cannot be both chiral and anti-chiral unless it is a constant. The covariant conditions are first-order differential equations and can be solved to yield

$$\phi(x,\theta,\overline{\theta}) = \exp\left(-i\theta\ \partial\!\!\!/\overline{\theta}\right)\phi(x,\theta); \quad \overline{\phi}(x,\theta,\overline{\theta}) = \exp\left(i\theta\ \partial\!\!\!/\overline{\theta}\right)\overline{\phi}(x,\overline{\theta}), \tag{B.197}$$

with $\phi(x,\theta)$ and $\overline{\phi}(x,\overline{\theta})$ not depending on $\overline{\theta}$ and θ, respectively. Their Taylor expansion in terms of ordinary fields is

$$\phi(x,\theta,\overline{\theta}) = A + 2\theta\psi - \theta^2 F \quad \text{and} \quad \overline{\phi}(x,\theta,\overline{\theta}) = A^\dagger + 2\overline{\psi}\overline{\theta} - \overline{\theta}^2 F^\dagger, \tag{B.198}$$

and gives rise to irreducible multiplets which turn out to be exactly the chiral and anti-chiral multiplets, (B.62) and (B.64).

B.6.2. Superfield tensor calculus

In the superfield formulation, the critical property which makes tensor calculus possible is that the product of two superfields is again a superfield

$$\phi_1(x,\theta,\overline{\theta})\phi_2(x,\theta,\overline{\theta}) = \phi_3(x,\theta,\overline{\theta}). \tag{B.199}$$

This fact can be seen as a consequence of definition (B.179) or, alternatively, of the fact that the representation, (B.182), of the super-algebra is in terms of first-order differential operators.[7]

The superfield expressions for the four products are

$$\begin{aligned}
\phi_3 = \phi_1 \cdot \phi_2 \quad &\Leftrightarrow \quad \phi_3 = \phi_1\phi_2 \\
V = \phi_1 \times \phi_2 \quad &\Leftrightarrow \quad V = \frac{1}{2}\left(\phi_1\overline{\phi}_2 + \overline{\phi}_1\phi_2\right) \\
V = \phi_1 \wedge \phi_2 \quad &\Leftrightarrow \quad V = \frac{i}{2}\left(\phi_1\overline{\phi}_2 - \overline{\phi}_1\phi_2\right) \\
V_3 = V_1 \cdot V_2 \quad &\Leftrightarrow \quad V_3 = V_1V_2,
\end{aligned} \tag{B.200}$$

and the fact that ϕ_3 is again chiral is due to the fact that $\overline{D}$ is a first-order differential

[7]This multiplication law has deep consequences for supersymmetric field theories. In particular, we have shown in Section (B.3.), using properties of division algebras, that the allowed dimension of the spacetime is restricted.

operator which obeys the Leibniz rule

$$\overline{D}(\phi_1\phi_2) = \left(\overline{D}\phi_1\right)\phi_2 + \phi_1\left(\overline{D}\phi_2\right) = 0. \qquad \text{(B.201)}$$

Note that second order equations like $D^2\phi = 0$ are not preserved for a product. The kinetic superfield is given by

$$\mathbf{T}\phi = \frac{1}{4}\overline{D}^2\overline{\phi}. \qquad \text{(B.202)}$$

We see that $\mathbf{T}\phi$ is chiral; $\overline{D}(\mathbf{T}\phi) = 0$. The property $\mathbf{TT} = -\Box$ follows from $\overline{\mathbf{T}\phi} = \frac{1}{4}D^2\phi$ and

$$\left[\overline{D}^2, D^2\right] = -16\Box - 8iD\,\partial\!\!\!/\overline{D}. \qquad \text{(B.203)}$$

B.6.3. Superspace invariants

The general method by which a translation-invariant action is derived from fields is to integrate a lagrangian density, $L(x)$, over d^4x. The result is translationally invariant if surface terms vanish. A similar procedure can be used to construct supersymmetry invariant actions in superspace. Once we can define an integral, $\int d\theta$, in such a way that it is invariant under $\theta \to \theta + \xi$, the integral of any superfield over the entire superspace (θ, $\overline{\theta}$, and x^μ) will be invariant.

Such an integral is known and is called the Berezin integral, defined by

$$0 = \int d\theta; \quad 1 = \int d\theta\,\theta \qquad \text{(B.204)}$$

for each different θ. Since a function of any one anticommuting θ is always of the form

$$f(\theta) = f_0 + \theta f_1, \qquad \text{(B.205)}$$

these definitions are sufficient to define a general $\int d\theta\, f(\theta)$. Assuming that θ is not a multiple of ξ, the translational invariance of the integral follows:

$$\int d\theta\, f(\theta+\xi) = \int d\theta\,(f_0 + \theta f_1 + \xi f_1) = f_1 = \int d\theta\, f(\theta). \qquad \text{(B.206)}$$

Formally, differentiation and integration are the same,

$$\int d\theta\, f(\theta) = \frac{\partial}{\partial\theta} f(\theta), \qquad \text{(B.207)}$$

which we can understand by visualizing power series in θ as modulo 2 so that raising the power (integration) and lowering the power (differentiation) are the same thing. This property results in equations like that of the δ-function

$$\delta(\theta) = \theta; \quad \delta(-\theta) = -\delta(\theta). \qquad \text{(B.208)}$$

We define

$$\int d^2\theta = \int d\theta^1\, d\theta^2; \quad \int d^2\overline{\theta} = \int d\overline{\theta}^{\dot{2}}\, d\overline{\theta}^{\dot{1}} \tag{B.209}$$

so that

$$\int d^2\theta\, \theta^2 = \int d^2\overline{\theta}\, \overline{\theta}^2 = -2. \tag{B.210}$$

The integral of any superfield over the entire superspace is an invariant

$$\delta \int d^4x\, d^2\theta\, d^2\overline{\theta}\, \phi(x,\theta,\overline{\theta}) = 0, \tag{B.211}$$

provided there is no jacobian. Also, for chiral and anti-chiral superfields, we have the following invariant integrals:

$$\delta \int d^4x\, d^2\theta\, \phi(x,\theta,\overline{\theta}) = 0 \quad \text{if } \overline{D}\phi = 0 \tag{B.212}$$

and

$$\delta \int d^4x\, d^2\overline{\theta}\, \phi(x,\theta,\overline{\theta}) = 0 \quad \text{if } D\phi = 0. \tag{B.213}$$

We can see that the jacobian is not needed for the transformations (B.171) since

$$\frac{\partial(x',\theta,\overline{\theta}')}{\partial(x,\theta,\overline{\theta})} = \begin{bmatrix} \delta_\mu^\nu & -i(\sigma^\nu\overline{\xi})_\alpha & -i(\xi\sigma^\nu)_{\dot{\alpha}} \\ 0 & \delta_\alpha^\beta & 0 \\ 0 & 0 & \delta_{\dot{\alpha}}^{\dot{\beta}} \end{bmatrix} = e^X \tag{B.214}$$

with

$$X = \begin{bmatrix} 0 & -i(\sigma^\nu\overline{\xi})_\alpha & -i(\xi\sigma^\nu)_{\dot{\alpha}} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{B.215}$$

has unit determinant.

The superspace transformations (B.171) which represent supersymmetry transformations in flat-space field theory have the functional matrix (B.214) with unit jacobian determinant. Therefore, we had no need to include a determinant in the formula for an invariant. In the context of supergravity, one encounters general coordinate transformations of superspace, and there is, therefore, a need for a consistent prescription for a jacobian. Since the matrix has anticommuting parameters in it, we need a prescription for a super-determinant which obeys the law

$$\text{Sdet}(M_1M_2) = (\text{Sdet}M_1)(\text{Sdet}M_2). \tag{B.216}$$

We start by defining the super-trace. For a matrix of the form

$$M = \begin{bmatrix} m_1 & \mu_2 \\ \mu_1 & m_2 \end{bmatrix}, \tag{B.217}$$

where m_i are commuting elements and μ_i are anticommuting elements. We define the trace so that the trace of a commutator is $\mathrm{STr}[M,N]=0$. This condition gives the definition as

$$\mathrm{STr}M = \mathrm{Tr}m_1 - \mathrm{Tr}m_2. \tag{B.218}$$

We define our super-determinant as

$$\mathrm{Sdet}M = \exp\{\mathrm{STr}(\ln M)\} \tag{B.219}$$

so that

$$\begin{aligned}\mathrm{Sdet}(M_1M_2) &= \mathrm{Sdet}\exp\left\{\ln M_1 + \ln M_2 + \frac{1}{2}[\ln M_1, \ln M_2] + \cdots\right\}\\ &= \exp\{\mathrm{STr}(\ln M_1 + \ln M_2)\}\\ &= \exp\{\mathrm{STr}(\ln M_2)\}\exp\{\mathrm{STr}(\ln M_1)\}\\ &= \mathrm{Sdet}(M_1)\mathrm{Sdet}(M_2)\end{aligned} \tag{B.220}$$

as required (where the first step is from Campbell-Baker-Hausdorff). This definition works for a determinant as long as the anti-commuting elements are off diagonal.

B.6.4. Wess-Zumino action in superspace

We want to write down an invariant action for a chiral superfield, ϕ, which has field components A, B, ψ, F, and G. First, we perform some dimensional analysis. Each θ carries a mass dimension of $-\frac{1}{2}$ so that the exponents in equation (B.197) are dimensionless. If we want the spinor ψ to have dimension $\frac{3}{2}$, we find from equation (B.198) that the entire field has $\dim\phi = 1$. The integrations, $d^2\theta$ and $d^2\overline{\theta}$, have dimension $+1$ each, thus the chiral measure, $d^4x\, d^2\theta$, has dimension -3. In order to construct a dimensionless action, the most general term involving only ϕ's and no negative dimension coupling constants is

$$\lambda\phi + \frac{m}{2}\phi^2 + \frac{g}{3}\phi^3. \tag{B.221}$$

We need a kinetic term to describe the dynamics, and it must be bilinear in the superfield with either one derivative, ∂_μ, (dimension 1) which would not be Lorentz covariant, or two derivatives, $\overline{D}$ or D, (dimension $\frac{1}{2}$). The kinetic superfield, $\mathbf{T}\phi = \frac{1}{4}\overline{D}^2\overline{\phi}$ can, therefore, be used

$$\begin{aligned}S &= \frac{1}{4}\int d^4x\, d^2\theta\left(\frac{1}{2}\phi\mathbf{T}\phi - \frac{m}{2}\phi^2 - \frac{g}{3}\phi^3\right) + \text{h.c.}\\ &= \frac{1}{4}\int d^4x\, d^2\theta\left(\frac{1}{8}\phi\overline{D}^2\overline{\phi} - \frac{m}{2}\phi^2 - \frac{g}{3}\phi^3\right) + \text{h.c.}\end{aligned} \tag{B.222}$$

which gives the same lagrangian density that we had before if we use the fact that

$$\frac{1}{4}\int d^2\theta\phi + \text{h.c.} = [\phi]_F + \text{4-div.} \tag{B.223}$$

for any chiral ϕ. For the integral over the entire superspace of a general field, V, we have

$$-\frac{1}{2}\int d^2\theta d^2\overline{\theta}\,V = [V]_D + \text{4-div.} \tag{B.224}$$

The action of the Wess-Zumino model can be written in a different form when we note that

$$\ldots + \text{h.c.} = \ldots + \frac{1}{4}\int d^4x\, d^2\overline{\theta}\left(\frac{1}{8}\overline{\phi}D^2\phi - \frac{m}{2}\overline{\phi}^2 - \frac{g}{3}\overline{\phi}^3\right) \tag{B.225}$$

and we concentrate on the first term, which we can rewrite as an integral over the entire superspace

$$\begin{aligned}\frac{1}{32}\int d^4x\, d^2\overline{\theta}\,\overline{\phi}D^2\phi &= \frac{1}{32}\int d^4x\, d^2\overline{\theta}\, D^2(\overline{\phi}\phi) = \frac{1}{16}\int d^4x\, d^2\theta\, d^2\overline{\theta}\,\overline{\phi}\phi \\ &= \frac{1}{32}\int d^4x\, d^2\theta\,\phi\overline{D}^2\overline{\phi}.\end{aligned} \tag{B.226}$$

We have

$$S = \frac{1}{4}\int d^4x\, d^2\theta\left(\frac{1}{4}\phi\overline{D}^2\overline{\phi} - \frac{m}{2}\phi^2 - \frac{g}{3}\phi^3\right) - \frac{1}{4}\int d^4x\, d^2\overline{\theta}\left(\frac{m}{2}\overline{\phi}^2 + \frac{g}{3}\overline{\phi}^3\right), \tag{B.227}$$

and the equation of motion for ϕ is easily calculated to be

$$\frac{1}{4}\overline{D}^2\overline{\phi} - m\phi - g\phi^2 = 0, \tag{B.228}$$

which is $\mathbf{T}\phi = m\phi + g\phi^2$. Note that, if we would have chosen a different kinetic term, we would have gotten the correct result only by imposing the chiral constraint via a Lagrange multiplier, i.e., Say we had chosen $S_{\text{kinetic}} = \frac{1}{8}\int d^4x\, d^2\theta\,\overline{\phi}\phi$; then, we insert the constraints before finding the equations of motion

$$S_{\text{kinetic}} = \frac{1}{8}\int d^4x\, d^2\theta\left(\overline{\phi}\phi + \overline{D}_{\dot\alpha}\phi\Lambda^{\dot\alpha} + \Lambda^{\alpha}D_{\alpha}\overline{\phi}\right). \tag{B.229}$$

Varying with respect to $\overline{\Lambda}$, we get the constraint, $\overline{D}\phi = 0$, and when varied with respect to ϕ, we get

$$0 = \overline{D}_{\dot\alpha}\frac{\partial L}{\partial\overline{D}_{\dot\alpha}\phi} - \frac{\partial L}{\partial\phi}, \tag{B.230}$$

which gives

$$\overline{\phi} = \overline{D}_{\dot\alpha}\overline{\Lambda}^{\dot\alpha}, \tag{B.231}$$

which is the general solution of the free equation of motion, $\frac{1}{4}\overline{D}^2\overline{\phi} = 0$.

B.6.5. Non-renormalization theorem

The fact that the kinetic part of the Wess-Zumino action can be written as an integral over the entire superspace, but the mass and interaction terms cannot has important consequences. There is a theorem which states that those parts of a lagrangian which can, in principle, only be written as chiral integrals will not receive quantum corrections.

The observed renormalization behavior of the Wess-Zumino model is a direct and predictable consequence of this theorem. The kinetic term must be renormalized, resulting in a logarithmically divergent wavefunction renormalization, but there are no independent quadratically and linearly divergent mass and coupling constant renormalizations. Furthermore, since the vacuum energy is strictly zero in a supersymmetric theory, we know that there will also be no contributions to the vacuum energy if supersymmetry is unbroken by renormalization as it is in the Wess-Zumino model. We now turn to a discussion of the theories which result when we apply the techniques of supersymmetry in this chapter to quantum field theories in physics.

B.7. Supersymmetric quantum field theory

B.7.1. Chiral feynman rules

The first aims of this section are somewhat a review of the previous sections using slightly different notation, where we derive the chiral and vector multiplets again, form supersymmetric lagrangians from them, and find the corresponding Feynman rules. The main purpose of this repetition is to familiarize ourselves with the manipulation of the spinor quantities and become accustomed to writing Weyl 2-spinor equations in terms of Majorana 4-spinors instead, which allows the use of Dirac gamma matrices in calculations rather than the more tedious Pauli matrices.

Recall that the supersymmetry algebra is generated by the momentum operators, P_μ, and the Weyl spinor supersymmetry generators, Q_α and $\overline{Q}_{\dot{\alpha}}$. The algebra is

$$\begin{aligned} [Q_\alpha, P_\mu] &= [\overline{Q}_{\dot{\alpha}}, P_\mu] = [P_\mu, P_\nu] = 0 \\ \{Q_\alpha, Q_\beta\} &= \left\{\overline{Q}_{\dot{\alpha}}, \overline{Q}_{\dot{\beta}}\right\} = 0 \\ \left\{Q_\alpha, \overline{Q}_{\dot{\beta}}\right\} &= 2\sigma^\mu_{\alpha\dot{\beta}} P_\mu. \end{aligned} \tag{B.232}$$

A finite element of the group corresponding to this algebra is

$$G(x^\mu, \theta, \overline{\theta}) = e^{i\left(\theta Q + \overline{\theta}\overline{Q} - x^\mu P_\mu\right)}. \tag{B.233}$$

We wish to construct linear representations of this group (derivatives acting on superfields) by considering the action of the group on the parameter space $(x^\mu, \theta, \overline{\theta})$. Using the Hausdorff formula (which terminates at the first commutator due to the supersymmetry algebra), we have

$$G(x^\mu,\theta,\overline{\theta})G(a^\mu,\xi,\overline{\xi}) = G(x^\mu + a^\mu - i\xi\sigma^\mu\overline{\theta} + i\theta\sigma^\mu\overline{\xi}, \theta+\xi, \overline{\theta}+\overline{\xi}). \tag{B.234}$$

Thus, the supersymmetry generators induce the following motion in group parameter space:

$$e^{i\left(\theta Q+\overline{\theta Q}-a^\mu P_\mu\right)} : (x^\mu,\theta,\overline{\theta}) \to (x^\mu + a^\mu - i\xi\sigma^\mu\overline{\theta} + i\theta\sigma^\mu\overline{\xi}, \theta+\xi, \overline{\theta}+\overline{\xi}). \tag{B.235}$$

For a superfield, $S(x^\mu,\theta,\overline{\theta})$, we have the expansion

$$\begin{aligned} &S(x^\mu + a^\mu - i\xi\sigma^\mu\overline{\theta} + i\theta\sigma^\mu\overline{\xi}, \theta+\xi, \overline{\theta}+\overline{\xi}) \\ &\quad = S(x^\mu,\theta,\overline{\theta}) + \left(a^\mu - i\xi\sigma^\mu\overline{\theta} + i\theta\sigma^\mu\overline{\xi}\right)\frac{\partial S}{\partial x^\mu} + \xi^\alpha\frac{\partial S}{\partial\theta^\alpha} + \overline{\xi}_{\dot\alpha}\frac{\partial S}{\partial\overline{\theta}_{\dot\alpha}} + \cdots \end{aligned} \tag{B.236}$$

Therefore, the action of the supersymmetry algebra on superfields,

$$S(x^\mu,\theta,\overline{\theta}) \to e^{i\left(\theta Q+\overline{\theta Q}-a^\mu P_\mu\right)}S(x^\mu,\theta,\overline{\theta}), \tag{B.237}$$

is generated by the linear representation

$$\begin{aligned} P_\mu &= i\partial_\mu \\ iQ_\alpha &= \frac{\partial}{\partial\theta^\alpha} - i\sigma^\mu_{\alpha\dot\alpha}\overline{\theta}^{\dot\alpha}\partial_\mu \\ i\overline{Q}_{\dot\alpha} &= -\frac{\partial}{\partial\overline{\theta}^{\dot\alpha}} + i\theta^\alpha\sigma^\mu_{\alpha\dot\alpha}\partial_\mu. \end{aligned} \tag{B.238}$$

The general superfield, $S(x^\mu,\theta,\overline{\theta})$, may be expanded as a power series in θ and $\overline{\theta}$ with the series terminating as the second power of each as we have shown in Section B.6.1. equation (B.186). Such a superfield provides a representation of the supersymmetry algebra which is reducible, and we obtain irreducible representations by imposing covariant constraints such as $S = S^\dagger$ or $\overline{D}_{\dot\alpha}S = 0$.

The constraint, $\overline{D}_{\dot\alpha}S = 0$, gives a chiral superfield as its irreducible representation. Any function of θ and $y^\mu = x^\mu + i\theta\sigma^\mu\overline{\theta}$ will satisfy this constraint as can be seen by the definition (B.192) due to $\overline{D}_{\dot\alpha}\theta = 0$ and $\overline{D}_{\dot\alpha}y^\mu = 0$.

We can expand a chiral superfield, $\Phi(y^\mu,\theta)$, in powers of θ as

$$\Phi(y^\mu,\theta) = \phi(y) + 2\theta\psi(y) + \theta\theta F(y), \tag{B.239}$$

where ϕ and F are complex scalar fields, and ψ is a left-handed Weyl spinor field. (It is contracted with θ^α, so it must be ψ_α, which means left-handed, i.e., the upper two components of a Dirac spinor.) Thus, the general expansion of a chiral superfield in component fields is

$$\Phi(x^\mu,\theta,\overline{\theta}) = \phi(x) + 2\theta\psi(x) + \theta\theta F(x) + i\partial_\mu\phi\theta\sigma^\mu\overline{\theta} + 2i\theta\partial_\mu\psi\theta\sigma^\mu\overline{\theta} - \frac{1}{2}\partial_\mu\partial_\nu\phi\theta\sigma^\mu\overline{\theta}\theta\sigma^\nu\overline{\theta}. \tag{B.240}$$

The last two terms of this expansion can be written

$$\begin{aligned}
2i(\theta\partial_\mu\psi)(\theta\sigma^\mu\overline{\theta}) &= 2i\theta^\gamma\partial_\mu\psi_\gamma\theta^\alpha(\sigma^\mu)_{\alpha\dot{\alpha}}\overline{\theta}^{\dot{\alpha}} \\
&= -2i\theta^\gamma\theta^\alpha\partial_\mu\psi_\gamma(\sigma^\mu)_{\alpha\dot{\alpha}}\overline{\theta}^{\dot{\alpha}} \\
&= -i\left((\theta\theta)\delta^{\gamma\alpha}\right)\partial_\mu\psi_\gamma(\sigma^\mu)_{\alpha\dot{\alpha}}\overline{\theta}^{\dot{\alpha}} \\
&= -i(\theta\theta)\partial_\mu\psi^\alpha(\sigma^\mu)_{\alpha\dot{\alpha}}\overline{\theta}^{\dot{\alpha}} = -i\theta\theta\partial_\mu\psi\sigma^\mu\overline{\theta},
\end{aligned} \tag{B.241}$$

where we used equation (A.54), and

$$\begin{aligned}
&-\frac{1}{2}\partial_\mu\partial_\nu\phi\theta\sigma^\mu\overline{\theta}\theta\sigma^\nu\overline{\theta} = -\frac{1}{2}\partial_\mu\partial_\nu\phi\theta^\alpha(\sigma^\mu)_{\alpha\dot{\alpha}}\overline{\theta}^{\dot{\alpha}}\theta^\beta(\sigma^\nu)_{\beta\dot{\beta}}\overline{\theta}^{\dot{\beta}} \\
&= \frac{1}{2}\partial_\mu\partial_\nu\phi\theta^\alpha(\sigma^\mu)_{\alpha\dot{\alpha}}\overline{\theta}^{\dot{\alpha}}\overline{\theta}_{\dot{\beta}}(\overline{\sigma}^\nu)^{\dot{\beta}\beta}\theta_\beta = \frac{1}{2}\partial_\mu\partial_\nu\phi\theta^\alpha(\sigma^\mu)_{\alpha\dot{\alpha}}\left(-\frac{1}{2}\overline{\theta}\overline{\theta}\delta^{\dot{\alpha}}_{\dot{\beta}}\right)(\overline{\sigma}^\nu)^{\dot{\beta}\beta}\theta_\beta \\
&= -\frac{1}{4}\partial_\mu\partial_\nu\phi\theta^\alpha(\sigma^\mu)_{\alpha\dot{\alpha}}\overline{\theta}\overline{\theta}(\overline{\sigma}^\nu)^{\dot{\alpha}\beta}\theta_\beta = -\frac{1}{4}\partial_\mu\partial_\nu\phi\left(\frac{1}{2}\theta\theta\delta^\alpha_\beta\right)(\sigma^\mu)_{\alpha\dot{\alpha}}\overline{\theta}\overline{\theta}(\overline{\sigma}^\nu)^{\dot{\alpha}\beta} \\
&= -\frac{1}{8}\partial_\mu\partial_\nu\phi\theta\theta(\sigma^\mu)_{\beta\dot{\alpha}}\overline{\theta}\overline{\theta}(\overline{\sigma}^\nu)^{\dot{\alpha}\beta} = -\frac{1}{8}\partial_\mu\partial_\nu\phi\theta\theta\mathrm{Tr}\left(\sigma^\mu\overline{\sigma}^\nu\right)\overline{\theta}\overline{\theta} \\
&= -\frac{1}{8}\partial_\mu\partial_\nu\phi\theta\theta\left(2\eta^{\mu\nu}\right)\overline{\theta}\overline{\theta} = -\frac{1}{4}\partial_\mu\partial^\mu\phi\theta\theta\overline{\theta}\overline{\theta},
\end{aligned} \tag{B.242}$$

which allows us to write our left and right chiral superfields[8] as

$$\begin{aligned}
\Phi(x^\mu,\theta,\overline{\theta}) &= \phi(x) + 2\theta\psi(x) + \theta\theta F(x) + i\partial_\mu\phi\theta\sigma^\mu\overline{\theta} \\
&\quad - i\theta\theta\partial_\mu\psi\sigma^\mu\overline{\theta} - \frac{1}{4}\partial_\mu\partial^\mu\phi\theta\theta\overline{\theta}\overline{\theta} \\
\Phi^\dagger(x^\mu,\theta,\overline{\theta}) &= \phi^\dagger(x) + 2\overline{\theta}\overline{\psi}(x) + \overline{\theta}\overline{\theta}F^\dagger(x) - i\partial_\mu\phi^\dagger\theta\sigma^\mu\overline{\theta} \\
&\quad + i\overline{\theta}\overline{\theta}\theta\sigma^\mu\partial_\mu\overline{\psi} - \frac{1}{4}\partial_\mu\partial^\mu\phi^\dagger\theta\theta\overline{\theta}\overline{\theta}.
\end{aligned} \tag{B.243}$$

We can find the behavior of Φ under infinitesimal supersymmetry transformations by

[8]We call it a left chiral superfield since it is chiral and it is written in terms of the left-handed Weyl spinor, ψ. The right-handed chiral superfield is found by hermitian conjugating the entire equation.

using equations (B.237) and (B.238) as follows. First, we note that $\Phi \to \Phi + \delta\Phi$, where

$$\begin{aligned}
\delta\Phi &= i\left(\xi Q + \overline{\xi Q}\right)\Phi \\
&= i\left(\xi^\alpha Q_\alpha - \overline{\xi}^{\dot\alpha}\overline{Q}_{\dot\alpha}\right)\Phi \text{ minus sign from reversing contraction order} \\
&= \xi^\alpha\left(\partial_\alpha - i\sigma^\mu_{\alpha\dot\alpha}\overline{\theta}^{\dot\alpha}\partial_\mu\right) - \overline{\xi}^{\dot\alpha}\left(-\partial_{\dot\alpha} + i\theta^\alpha\sigma^\mu_{\alpha\dot\alpha}\partial_\mu\right)\Phi \\
&= \xi^\alpha\partial_\alpha\Phi - i\xi^\alpha\sigma^\mu_{\alpha\dot\alpha}\overline{\theta}^{\dot\alpha}\partial_\mu\Phi + \overline{\xi}^{\dot\alpha}\partial_{\dot\alpha}\Phi + i\theta^\alpha\sigma^\mu_{\alpha\dot\alpha}\overline{\xi}^{\dot\alpha}\partial_\mu\Phi.
\end{aligned} \tag{B.244}$$

Working out the individual terms,

$$\begin{aligned}
\delta\Phi = &\left(2\xi\psi + 2\xi\theta F + i\partial_\mu\phi\xi\sigma^\mu\overline{\theta} - 2i\xi\theta\partial_\mu\psi\sigma^\mu\overline{\theta} - \frac{1}{2}\partial_\mu\partial^\mu\phi\xi\theta\overline{\theta\theta}\right) \\
&- i\xi\sigma^\nu\overline{\theta}\left(\partial_\nu\phi(x) + 2\theta\partial_\nu\psi(x) + \theta\theta\partial_\nu F(x) + i\partial_\nu\partial_\mu\phi\theta\sigma^\mu\overline{\theta} - i\theta\theta\partial_\nu\partial_\mu\psi\sigma^\mu\overline{\theta}\right) \\
&+ \left(i\partial_\mu\phi\theta\sigma^\mu\overline{\xi} - i\theta\theta\partial_\mu\psi\sigma^\mu\overline{\xi} - \frac{1}{2}\partial_\mu\partial^\mu\phi\theta\theta\overline{\xi\theta}\right) \\
&+ i\theta\sigma^\nu\overline{\xi}\left(\partial_\nu\phi(x) + 2\theta\partial_\nu\psi(x) + i\partial_\nu\partial_\mu\phi\theta\sigma^\mu\overline{\theta}\right),
\end{aligned} \tag{B.245}$$

where we used the fact that terms with 3 θ's or 3 $\overline{\theta}$'s are zero. Collecting powers of θ gives

$$\begin{aligned}
\delta\Phi = &\,2\xi\psi + 2\xi\theta F + 2i\partial_\mu\phi\theta\sigma^\mu\overline{\xi} \\
&- 2i\xi\theta\partial_\mu\psi\sigma^\mu\overline{\theta} - 2i\xi\sigma^\nu\overline{\theta}\theta\partial_\nu\psi(x) - i\theta\theta\partial_\mu\psi\sigma^\mu\overline{\xi} + 2i\theta\sigma^\nu\overline{\xi}\theta\partial_\nu\psi(x) \\
&- \frac{1}{2}\partial_\mu\partial^\mu\phi\xi\theta\overline{\theta\theta} - \frac{1}{2}\partial_\mu\partial^\mu\phi\theta\theta\overline{\xi\theta} - \partial_\nu\partial_\mu\phi\theta\sigma^\nu\overline{\xi}\theta\sigma^\mu\overline{\theta} - i\partial_\nu F(x)\xi\sigma^\nu\overline{\theta}\theta\theta \\
&+ \partial_\nu\partial_\mu\phi\xi\sigma^\nu\overline{\theta}\theta\sigma^\mu\overline{\theta} - \xi\sigma^\nu\overline{\theta}\theta\theta\partial_\nu\partial_\mu\psi\sigma^\mu\overline{\theta}.
\end{aligned} \tag{B.246}$$

Since we can also write the variation as

$$\delta\Phi = \delta\phi(x) + 2\theta\delta\psi(x) + \theta\theta\delta F(x) + i\partial_\mu\delta\phi\theta\sigma^\mu\overline{\theta} - i\theta\theta\partial_\mu\delta\psi\sigma^\mu\overline{\theta} - \frac{1}{4}\partial_\mu\partial^\mu\delta\phi\theta\theta\overline{\theta\theta}, \tag{B.247}$$

we can compare powers of θ and pick out the variations:

$$\delta\phi = \boxed{2\xi\psi} \tag{B.248}$$

$$\begin{aligned}
\delta\psi^\alpha &= -\xi^\alpha F - i\partial_\mu\phi(\sigma^\mu)^\alpha_{\dot\alpha}\overline{\xi}^{\dot\alpha} \\
&= \boxed{-\xi^\alpha F - i\partial_\mu\phi(\sigma^\mu)^{\alpha\dot\alpha}\overline{\xi}_{\dot\alpha}}
\end{aligned} \tag{B.249}$$

$$\begin{aligned}
\theta\theta\delta F(x) &= -i\theta\theta\partial_\mu\psi\sigma^\mu\bar{\xi} + 2i\theta\sigma^\nu\bar{\xi}\theta\partial_\nu\psi \\
&= -i\theta\theta\partial_\mu\psi\sigma^\mu\bar{\xi} + 2i\theta^\alpha\theta_\gamma(\sigma^\nu)_{\alpha\dot{\alpha}}\bar{\xi}^{\dot{\alpha}}\partial_\nu\psi^\gamma \\
&= -i\theta\theta\partial_\mu\psi\sigma^\mu\bar{\xi} + 2i\left(\frac{1}{2}\theta\theta\delta^\alpha_\gamma\right)(\sigma^\nu)_{\alpha\dot{\alpha}}\bar{\xi}^{\dot{\alpha}}\partial_\nu\psi^\gamma \\
&= -i\theta\theta\partial_\mu\psi\sigma^\mu\bar{\xi} - i\theta\theta\partial_\nu\psi^\alpha(\sigma^\nu)_{\alpha\dot{\alpha}}\bar{\xi}^{\dot{\alpha}} \\
\delta F(x) &= \boxed{-2i\partial_\mu\psi\sigma^\mu\bar{\xi}}.
\end{aligned} \tag{B.250}$$

We see that the boxed formulae are exactly the transformation laws (B.62) as they should be. It is important to note that the change in F under a supersymmetry transformation is a total derivative so that, if we form an action from the F term, the action will also change by a total derivative and will, therefore, be invariant.

The product of chiral superfields is given as one would expect by simply multiplying the expansions of each field and collecting powers of θ. The anti-chiral superfield is given by the hermitian conjugate of the chiral superfield. We can form products of chiral and anti-chiral superfields also, but in that case, it is the D term which is a total divergence rather than the F term. The F term is the coefficient of $\theta\theta$, and the D term is the coefficient of $\bar{\theta}\bar{\theta}\theta\theta$. We can, therefore, construct a lagrangian wherein we have the D-term of a product of a chiral and an anti-chiral superfield, and we can add a "potential" term which is the F component of a product of only chiral or only anti-chiral superfields. A general lagrangian has the form

$$\mathcal{L} = \sum_i \left[\Phi_i^\dagger\Phi_i\right]_D + \left([W(\Phi)]_F + \text{h.c.}\right), \tag{B.251}$$

where, for example,

$$W(\Phi) = \frac{1}{2}m_{ij}\Phi_i\Phi_j + \frac{1}{3}\lambda_{ijk}\Phi_i\Phi_j\Phi_k \tag{B.252}$$

for real m_{ij} and λ_{ijk} which are symmetric in their indices.

Let us now write out the lagrangian in component fields using the potential term.

$$\mathcal{L} = \sum_i \left[\Phi_i^\dagger\Phi_i\right]_D + \left(\left[\frac{1}{2}m_{ij}\Phi_i\Phi_j + \frac{1}{3}\lambda_{ijk}\Phi_i\Phi_j\Phi_k\right]_F + \text{h.c.}\right) \tag{B.253}$$

We work out the terms as

$$\begin{aligned}
\left[\Phi_i^\dagger\Phi_j\right]_D &= \left[\left(\phi_i^\dagger(y) + 2\bar{\theta}\bar{\psi}_i(y) + F_i^\dagger(y)\bar{\theta}\bar{\theta}\right)\left(\phi_j(y) + 2\theta\psi_j(y) + \theta\theta F_j(y)\right)\right]_D \\
&= \left\{\left[\phi_i^\dagger + 2\bar{\theta}\bar{\psi}_i + \bar{\theta}\bar{\theta}F_i^\dagger - i\partial_\mu\phi_i^\dagger\theta\sigma^\mu\bar{\theta} + i\bar{\theta}\bar{\theta}\theta\sigma^\mu\partial_\mu\bar{\psi}_i - \frac{1}{4}\partial_\mu\partial^\mu\phi_i^\dagger\theta\theta\bar{\theta}\bar{\theta}\right]\right. \\
&\times \left.\left[\phi_j + 2\theta\psi_j + \theta\theta F_j + i\partial_\nu\phi_j\theta\sigma^\nu\bar{\theta} - i\theta\theta\partial_\nu\psi_j\sigma^\nu\bar{\theta} - \frac{1}{4}\partial_\nu\partial^\nu\phi_j\theta\theta\bar{\theta}\bar{\theta}\right]\right\}_D \\
&= \left\{F_i^\dagger F_j - \frac{1}{4}\phi_i^\dagger\partial_\mu\partial^\mu\phi_j - \frac{1}{4}\partial_\mu\partial^\mu\phi_i^\dagger\phi_j + \frac{1}{2}\partial_\mu\phi_i^\dagger\partial^\mu\phi_j + i\psi_j\sigma^\mu\partial_\mu\bar{\psi}_i + i\bar{\psi}_i\bar{\sigma}^\mu\partial_\mu\psi_j\right\}
\end{aligned} \tag{B.254}$$

$$
\begin{aligned}
[\Phi_i\Phi_j]_F &= \{[\phi_i + 2\theta\psi_i + \theta\theta F_i] \times [\phi_j + 2\theta\psi_j + \theta\theta F_j]\}_F \\
&= \left[\phi_i F_j - 4\theta^\alpha(\psi_i)_\alpha\theta_\beta(\psi_j)^\beta + \theta\theta F_i\phi_j\right] \\
&= \left[\phi_i F_j + 4\left(\frac{1}{2}\delta^\alpha_\beta\right)(\psi_i)_\alpha(\psi_j)^\beta\theta\theta + \theta\theta F_i\phi_j\right] \\
&= [\phi_i F_j + F_i\phi_j - 2\psi_i\psi_j]
\end{aligned}
\tag{B.255}
$$

$$
\left[\Phi_j^\dagger\Phi_i^\dagger\right]_F = \left[\phi_j^\dagger F_i^\dagger + F_j^\dagger\phi_i^\dagger - 2\overline{\psi}_i\overline{\psi}_j\right] \tag{B.256}
$$

$$
[\Phi_i\Phi_j\Phi_k]_F = [\phi_i\phi_j F_k + \phi_i F_j\phi_k + F_i\phi_j\phi_k - 2\psi_i\psi_j\phi_k - 2\psi_i\psi_k\phi_j - 2\psi_j\psi_k\phi_i] \tag{B.257}
$$

and

$$
\left[\Phi_k^\dagger\Phi_j^\dagger\Phi_i^\dagger\right]_F = \left[F_k^\dagger\phi_j\phi_i + \phi_k^\dagger F_j^\dagger\phi_i^\dagger + \phi_k^\dagger\phi_j^\dagger F_i^\dagger - 2\overline{\psi}_i\overline{\psi}_j\phi_k^\dagger - 2\overline{\psi}_i\overline{\psi}_k\phi_j^\dagger - 2\overline{\psi}_j\overline{\psi}_k\phi_i^\dagger\right]. \tag{B.258}
$$

We can now put these equations together to form our complete supersymmetric lagrangian for our left- and right-handed chiral multiplets

$$
\begin{aligned}
\mathcal{L} = &\left[F_i^\dagger F_i - \frac{1}{4}\phi_i^\dagger\partial_\mu\partial^\mu\phi_i - \frac{1}{4}\partial_\mu\partial^\mu\phi_i^\dagger\phi_i + \frac{1}{2}\partial_\mu\phi_i^\dagger\partial^\mu\phi_i + i\psi_i\sigma^\mu\partial_\mu\overline{\psi}_i + i\overline{\psi}_i\overline{\sigma}^\mu\partial_\mu\psi_j\right] \\
&+ \left\{\frac{1}{2}m_{ij}\left[\phi_i F_j + F_i\phi_j - 2\psi_i\psi_j\right]\right. \\
&\quad \left. + \frac{1}{3}\lambda_{ijk}\left[\phi_i\phi_j F_k + \phi_i F_j\phi_k + F_i\phi_j\phi_k - 2\psi_i\psi_j\phi_k - 2\psi_i\psi_k\phi_j - 2\psi_j\psi_k\phi_i\right]\right\} \\
&+ \left\{\frac{1}{2}m_{ij}^\dagger\left[\phi_j^\dagger F_i^\dagger + F_j^\dagger\phi_i^\dagger - 2\overline{\psi}_i\overline{\psi}_j\right]\right. \\
&\quad \left. + \frac{1}{3}\lambda_{ijk}^\dagger\left[F_k^\dagger\phi_j\phi_i + \phi_k^\dagger F_j^\dagger\phi_i^\dagger + \phi_k^\dagger\phi_j^\dagger F_i^\dagger - 2\overline{\psi}_i\overline{\psi}_j\phi_k^\dagger - 2\overline{\psi}_i\overline{\psi}_k\phi_j^\dagger - 2\overline{\psi}_j\overline{\psi}_k\phi_i^\dagger\right]\right\}.
\end{aligned}
\tag{B.259}
$$

We can translate the lagrangian into Majorana 4-spinors using the following identities (denoting Weyl spinors by ψ and Majorana spinors by Ψ)

$$
\begin{aligned}
\overline{\Psi}_i\Psi_j &= \overline{\psi}_i\overline{\psi}_j + \psi_i\psi_j \\
\overline{\Psi}_i\gamma_5\Psi_j &= \overline{\psi}_i\overline{\psi}_j - \psi_i\psi_j \\
\overline{\Psi}\gamma^\mu\partial_\mu\Psi &= \overline{\psi}\overline{\sigma}^\mu\partial_\mu\psi + \psi\sigma^\mu\partial_\mu\overline{\psi} = 2\overline{\psi}\overline{\sigma}^\mu\partial_\mu\psi,
\end{aligned}
\tag{B.260}
$$

where we dropped a total divergence. We also note that the constants, m_{ij} and λ_{ijk}, are real.

$$\mathcal{L} = \left[F_i^\dagger F_i - \frac{1}{4}\phi_i^\dagger\partial_\mu\partial^\mu\phi_i - \frac{1}{4}\phi_i\partial_\mu\partial^\mu\phi_i^\dagger + \frac{1}{2}\partial_\mu\phi_i^\dagger\partial^\mu\phi_i - i\overline{\Psi}\gamma^\mu\partial_\mu\Psi\right]$$
$$+ m_{ij}\left\{F_i\phi_j + F_i^\dagger\phi_j^\dagger - \left(\psi_i\psi_j + \overline{\psi}_i\overline{\psi}_j\right)\right\} \tag{B.261}$$
$$+ \lambda_{ijk}\left\{\phi_i\phi_j F_k + \phi_i^\dagger\phi_j^\dagger F_k^\dagger - 2\overline{\psi}_i\overline{\psi}_j\phi_k^\dagger - 2\psi_i\psi_j\phi_k\right\},$$

where we have used the symmetry of λ and m to combine terms. Also

$$-\frac{1}{4}\phi_i^\dagger\partial_\mu\partial^\mu\phi_i - \frac{1}{4}\phi_i\partial_\mu\partial^\mu\phi_i^\dagger = -\frac{1}{4}\partial_\mu\partial^\mu\left(\phi_i^\dagger\phi_i\right) + \frac{1}{2}\partial_\mu\phi_i^\dagger\partial^\mu\phi_i. \tag{B.262}$$

Substituting this result and the relations from equation (B.260) gives

$$\mathcal{L} = \left[F_i^\dagger F_i - \frac{1}{4}\partial_\mu\partial^\mu\left(\phi_i^\dagger\phi_i\right) + \partial_\mu\phi_i^\dagger\partial^\mu\phi_i + i\overline{\Psi}_i\gamma^\mu\partial_\mu\Psi_i\right]$$
$$+ m_{ij}\left\{F_i\phi_j + F_i^\dagger\phi_j^\dagger - \overline{\Psi}_i\Psi_j\right\} \tag{B.263}$$
$$+ \lambda_{ijk}\left\{\phi_i\phi_j F_k + \phi_i^\dagger\phi_j^\dagger F_k^\dagger - \overline{\Psi}_i\Psi_j\left(\phi_k^\dagger + \phi_k\right) - \overline{\Psi}_i\gamma_5\Psi_j\left(\phi^\dagger - \phi_k\right)\right\}.$$

We dropped the total divergence. This is written in a more familiar way as

$$\mathcal{L} = \partial_\mu\phi_i^\dagger\partial^\mu\phi_i + m_{ij}\left(F_i\phi_j + F_i^\dagger\phi_j^\dagger\right) + i\overline{\Psi}_i\,\partial\!\!\!/\Psi_i - m_{ij}\overline{\Psi}_i\Psi_j + F_i^\dagger F_i$$
$$+ \lambda_{ijk}\left\{\phi_i\phi_j F_k + \phi_i^\dagger\phi_j^\dagger F_k^\dagger - \overline{\Psi}_i(1+\gamma_5)\Psi_j\phi_k^\dagger - \overline{\Psi}_i(1-\gamma_5)\Psi_j\phi_k\right\}. \tag{B.264}$$

Since the F field has no kinetic term, it is auxiliary, and we can use the equations of motion to eliminate it. The equations of motion for the F are

$$\frac{\delta\mathcal{L}}{\delta F_i} = m_{ij}\phi_j + \lambda_{ijk}\phi_j\phi_k + F_i^\dagger = 0$$
$$\Rightarrow F_i^\dagger = -m_{ij}\phi_j - \lambda_{ijk}\phi_j\phi_k. \tag{B.265}$$

The equations of motion can be written

$$F_i^\dagger = -\frac{\partial W(\phi)}{\partial\phi_i}, \tag{B.266}$$

where $W(\phi)$ is the potential function with $\Phi_i \to \phi_i$. Sometimes, we call $V = |F|^2$ the *effective potential* at tree-level since it is the potential term of the lagrangian when you only consider tree-level processes. Putting the equations of motion of the F_i into our lagrangian gives

$$\begin{aligned}\mathcal{L} = &\partial_\mu\phi_i^\dagger\partial^\mu\phi_i + i\overline{\Psi}_i \not\partial\Psi_i - m_{ij}\overline{\Psi}_i\Psi_j + |m_{ij}\phi_j + \lambda_{ijk}\phi_j\phi_k|^2 \\ &- 2m_{kl}\lambda_{ijk}\left(\phi_i\phi_j\phi_l^\dagger + \phi_l\phi_i^\dagger\phi_j^\dagger\right) - 2\lambda_{ijk}\lambda_{klm}\phi_i\phi_j\phi_l^\dagger\phi_m^\dagger - 2m_{ij}m_{il}\phi_l^\dagger\phi_j \\ &- \lambda_{ijk}\overline{\Psi}_i(1+\gamma_5)\Psi_j\phi_k^\dagger - \lambda_{ijk}\overline{\Psi}_i(1-\gamma_5)\Psi_j\phi_k.\end{aligned} \tag{B.267}$$

We can condense this expression by using

$$\begin{aligned}|m_{ij}\phi_j + \lambda_{ijk}\phi_j\phi_k|^2 &= (m_{ij}\phi_j + \lambda_{ijk}\phi_j\phi_k)\left(m_{im}\phi_m^\dagger + \lambda_{imn}\phi_m^\dagger\phi_n^\dagger\right) \\ &= \left(m_{ij}m_{im}\phi_j\phi_m^\dagger + \lambda_{ijk}m_{im}\phi_j\phi_k\phi_m^\dagger + m_{ij}\lambda_{imn}\phi_j\phi_m^\dagger\phi_n^\dagger + \lambda_{ijk}\lambda_{imn}\phi_j\phi_k\phi_m^\dagger\phi_n^\dagger\right) \\ &= \left(m_{ij}m_{im}\phi_j\phi_m^\dagger + \lambda_{ijk}m_{im}\left(\phi_j\phi_k\phi_m^\dagger + \phi_m\phi_j^\dagger\phi_k^\dagger\right) + \lambda_{ijk}\lambda_{imn}\phi_j\phi_k\phi_m^\dagger\phi_n^\dagger\right).\end{aligned}$$

The lagrangian becomes

$$\begin{aligned}\mathcal{L} = &\partial_\mu\phi_i^\dagger\partial^\mu\phi_i + i\overline{\Psi}_i \not\partial\Psi_i - m_{ij}\overline{\Psi}_i\Psi_j \\ &- \lambda_{ijk}\overline{\Psi}_i(1+\gamma_5)\Psi_j\phi_k^\dagger - \lambda_{ijk}\overline{\Psi}_i(1-\gamma_5)\Psi_j\phi_k \\ &+ m_{ij}m_{im}\phi_m\phi_j^\dagger + \lambda_{ijk}m_{im}\left(\phi_j\phi_k\phi_m^\dagger + \phi_m\phi_j^\dagger\phi_k^\dagger\right) + \lambda_{ijk}\lambda_{imn}\phi_j\phi_k\phi_m^\dagger\phi_n^\dagger \\ &- 2m_{im}\lambda_{ijk}\left(\phi_k\phi_j\phi_m^\dagger + \phi_m\phi_k^\dagger\phi_j^\dagger\right) - 2\lambda_{ijk}\lambda_{ilm}\phi_k\phi_j\phi_l^\dagger\phi_m^\dagger - 2m_{ij}m_{im}\phi_m^\dagger\phi_j\end{aligned}$$

which finally reduces to

$$\begin{aligned}\mathcal{L} = &\partial_\mu\phi_i^\dagger\partial^\mu\phi_i + i\overline{\Psi}_i \not\partial\Psi_i - m_{ij}\overline{\Psi}_i\Psi_j \\ &- \lambda_{ijk}\overline{\Psi}_i(1+\gamma_5)\Psi_j\phi_k^\dagger - \lambda_{ijk}\overline{\Psi}_i(1-\gamma_5)\Psi_j\phi_k \\ &- m_{im}\lambda_{ijk}\left(\phi_k\phi_j\phi_m^\dagger + \phi_m\phi_k^\dagger\phi_j^\dagger\right) - \lambda_{ijk}\lambda_{ilm}\phi_k\phi_j\phi_l^\dagger\phi_m^\dagger - m_{ij}m_{im}\phi_m^\dagger\phi_j.\end{aligned} \tag{B.268}$$

We can now read the Feynman rules from this lagrangian, noting that indices i, j, and k label different particles. First, let us write the m_{ij} as a diagonal matrix, $m_{ij} = m_i\delta_{ij}$, so that each particle multiplet has an independent mass

$$\begin{aligned}\mathcal{L} = &\partial_\mu\phi_i^\dagger\partial^\mu\phi_i - m_i^2\phi_i^\dagger\phi_i + \overline{\Psi}_i(i\not\partial - m_i)\Psi_i - \lambda_{ijk}\overline{\Psi}_i(1+\gamma_5)\Psi_j\phi_k^\dagger \\ &- \lambda_{ijk}\overline{\Psi}_i(1-\gamma_5)\Psi_j\phi_k - m_i\lambda_{ijk}\left(\phi_k\phi_j\phi_i^\dagger + \phi_i\phi_k^\dagger\phi_j^\dagger\right) - \lambda_{ijk}\lambda_{ilm}\phi_k\phi_j\phi_l^\dagger\phi_m^\dagger.\end{aligned} \tag{B.269}$$

The lagrangian describes multiplets of scalar particles and Majorana spinors interacting. Indices i, j, and k label the multiplets. We can see that, for each spinor of mass m_i, there is a corresponding scalar with the same mass. The Feynman rules for this theory are shown in Figures B.7.1. and B.7.1..

B.7.2. Renormalization and non-renormalization

The one-loop diagrams can be worked out using the rules from Figures B.7.1. and B.7.1. where we write the dimension of the loop integrations as $d = 2\mu$ as usual. We

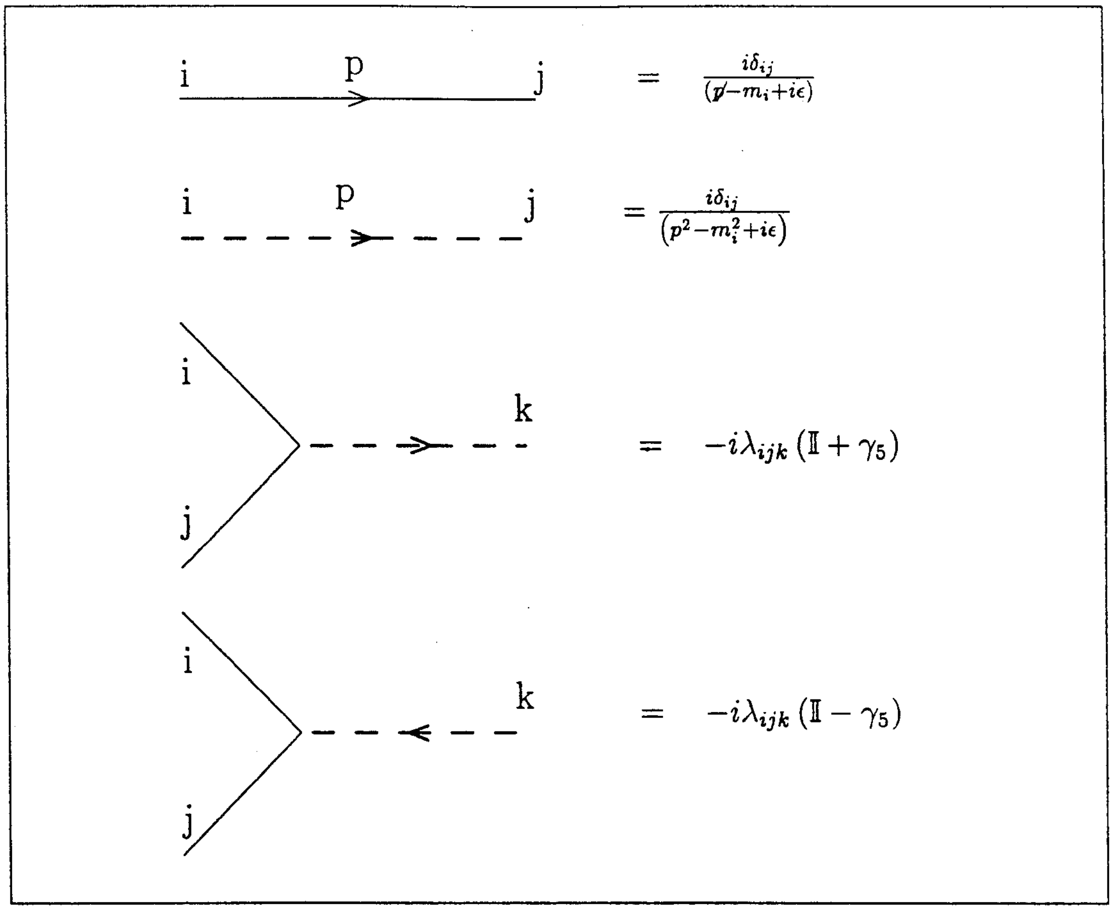


Figure B.1. The first set of Feynman rules for chiral supermultiplets. Each of the indices, $i, j, k, l,$ and m (labeling different multiplets), are assumed to be different here, and appropriate symmetry factors must be used whenever two or more are the same.

show the results to one loop in Figure B.7.2. where we are assuming that the initial and final particle are from the same multiplet, and the internal particle lines can be from any of the N different multiplets. If we add the 1-loop diagrams from Figure B.7.2., we have

$$\begin{aligned}
\text{Total} &= \int \frac{d^{2\mu}q}{(2\pi)^{2\mu}} \sum_{j,k=1}^{N} \lambda_{ijk}^2 \frac{2q^2 - 2m_k^2 - 4q^2 + m_i^2 + 2m_j^2}{(q^2 - m_k^2)(q^2 - m_j^2)} \\
\text{Total} &= \int \frac{d^{2\mu}q}{(2\pi)^{2\mu}} \sum_{j,k=1}^{N} \lambda_{ijk}^2 \frac{-2(q^2 - m_j^2) + m_i^2 - 2m_k^2}{(q^2 - m_k^2)(q^2 - m_j^2)} \qquad \text{(B.270)} \\
\text{Total} &= \int \frac{d^{2\mu}q}{(2\pi)^{2\mu}} \sum_{j,k=1}^{N} \lambda_{ijk}^2 \left\{ \frac{-2}{(q^2 - m_k^2)} + \frac{m_i^2 - 2m_k^2}{(q^2 - m_k^2)(q^2 - m_j^2)} \right\}.
\end{aligned}$$

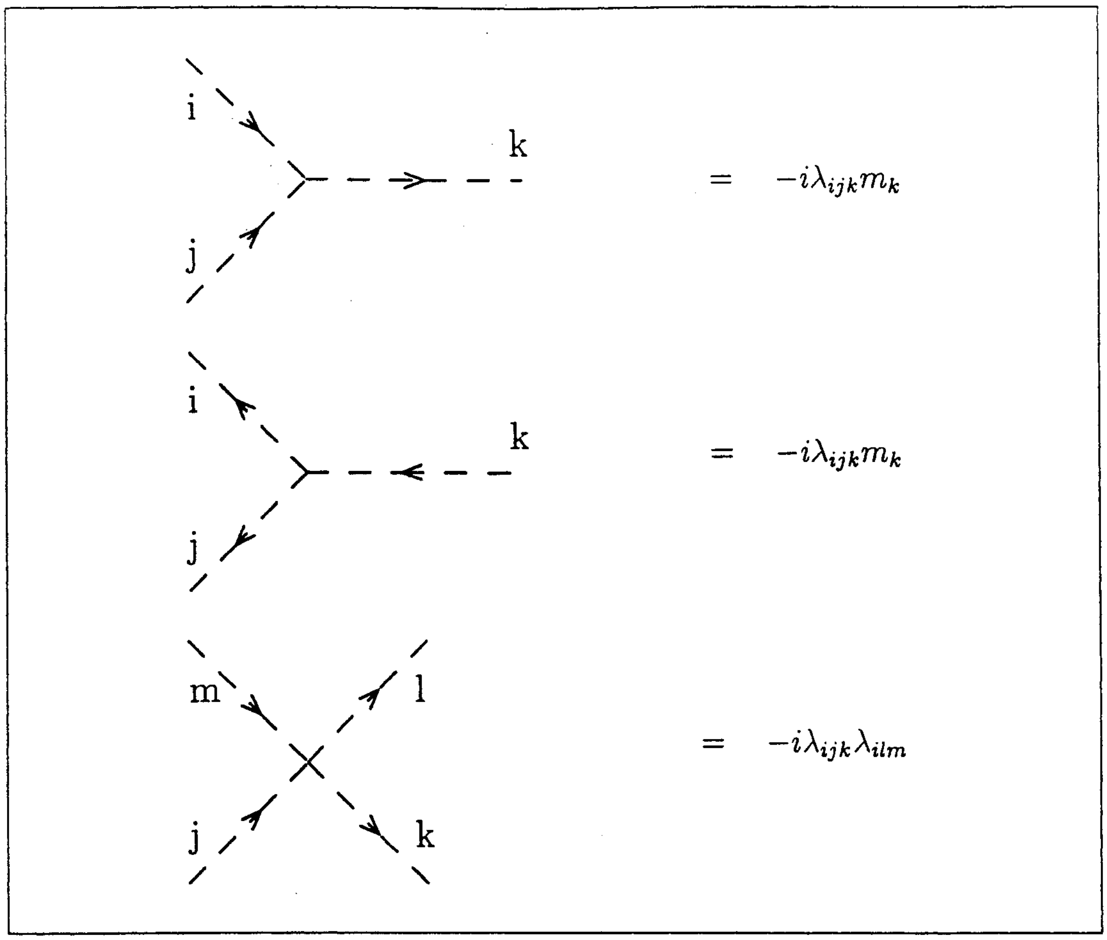
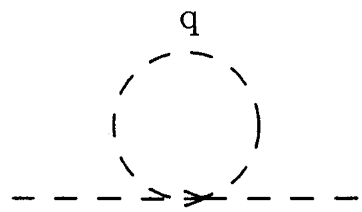
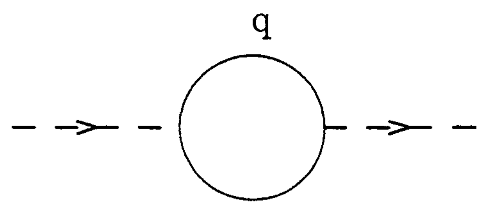


Figure B.2. The second set of Feynman rules for chiral supermultiplets. Each of the indices, i, j, k, l, and m (labeling different multiplets) are assumed to be different here, and appropriate symmetry factors must be used whenever two or more are the same.

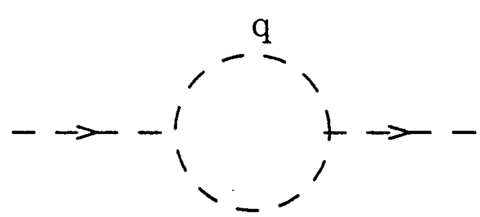
The first term is a quadratically divergent integral that is equal to zero in dimensional regularization, and the second term is a logarithmic divergence which leads to a wave function renormalization. There is no mass renormalization necessary since the contribution of the internal fermion loop cancels the contribution of the internal boson loops. We also notice that the cancellation occurs regardless of the masses, m_i and m_j, which can be different. Even if there is a large mass particle in the theory, it will not induce large masses in the other particles due to radiative correction counterterms. A mass hierarchy is, therefore, preserved.



$$= 2\int \frac{d^{2\mu}q}{(2\pi)^{2\mu}} \sum_{j,k=1}^{N} \frac{\lambda^2_{kij}}{\left(q^2-m_j^2\right)}$$



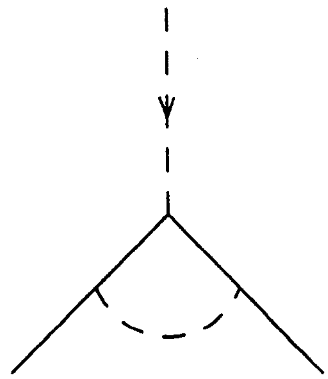
$$\begin{aligned} &= -\tfrac{1}{2}\int \frac{d^{2\mu}q}{(2\pi)^{2\mu}} \sum_{j,k=1}^{N} \lambda^2_{ijk} \mathrm{Tr}\left((\mathbb{I}+\gamma_5)\frac{1}{\not{q}-m_k}(\mathbb{I}-\gamma_5)\frac{1}{\not{q}-m_j}\right) \\ &= -\tfrac{1}{2}\int \frac{d^{2\mu}q}{(2\pi)^{2\mu}} \sum_{j,k=1}^{N} \lambda^2_{ijk} \frac{2\mathrm{Tr}[\not{q}\not{q}]}{\left(q^2-m_k^2\right)\left(q^2-m_j^2\right)} \\ &= -4\int \frac{d^{2\mu}q}{(2\pi)^{2\mu}} \sum_{j,k=1}^{N} \lambda^2_{ijk} \frac{q^2}{\left(q^2-m_k^2\right)\left(q^2-m_j^2\right)} \end{aligned}$$



$$= \int \frac{d^{2\mu}q}{(2\pi)^{2\mu}} \sum_{j,k=1}^{N} \frac{(m_i^2+2m_j^2)\lambda^2_{ijk}}{\left(q^2-m_k^2\right)\left(q^2-m_j^2\right)}$$

Figure B.3. 1-loop propagator corrections for the scalar particle in a chiral supermultiplet. The supermultiplet is labeled by i.

Notice also that there is no infinite renormalization of the coupling constants due to the vanishing of diagrams like Figure B.7.2.



$$= \frac{1}{6}\lambda^3 \int \frac{d^{2\mu}q}{(2\pi)^{2\mu}} \frac{1}{(q^2-m_i^2)} \mathrm{Tr}\left\{(1-\gamma_5)\frac{\not{r}+m_j}{(r^2-m_j^2)}(1-\gamma_5)\frac{\not{s}+m_k}{(s^2-m_k^2)}(1+\gamma_5)\right\}$$

$= 0$ since the trace is cyclic and $(1-\gamma_5)(1+\gamma_5)$ is zero.

Figure B.4. The 1-loop vertex correction for chiral supermultiplets.

Since $d^4\theta = d^2\theta d^2\overline{\theta}$, we have

$$\int d^2\theta\, \theta^2 = \int d^2\overline{\theta}\, \overline{\theta} = 1 \tag{B.271}$$

derived as follows

$$\begin{aligned}
\int d^2\theta\, \theta^2 &= -\frac{1}{4}\int [d\theta^\alpha d\theta_\alpha]\left[\theta^\beta\theta_\beta\right] = -\frac{1}{4}\int [d\theta^\alpha \epsilon_{\alpha\gamma} d\theta^\gamma]\left[\theta^\beta\epsilon_{\beta\delta}\theta^\delta\right] \\
&= -\frac{1}{4}\int \left[d\theta^2 d\theta^1 - d\theta^1 d\theta^2\right]\left[\theta^2\theta^1 - \theta^1\theta^2\right] \\
&= -\frac{1}{4}\int \left[d\theta^2 d\theta^1\theta^2\theta^1 - d\theta^2 d\theta^1\theta^1\theta^2 - d\theta^1 d\theta^2\theta^2\theta^1 + d\theta^1 d\theta^2\theta^1\theta^2\right] \\
&= -\frac{1}{4}\int [-1-1-1-1] = 1.
\end{aligned} \tag{B.272}$$

We can write our lagrangian in a form that automatically picks out the F and the D

terms where appropriate. This form is

$$\mathcal{L} = \int d^4\theta \sum_i \Phi_i^\dagger \Phi_i + \left(\int d^2\theta \, W(\Phi) + \text{h.c.} \right). \tag{B.273}$$

The non-renormalization theorem comes from the fact that any radiative correction to the effective action can be written as a single superspace integration, $\int d^4\theta$, over a product of factors that are local in θ and $\overline{\theta}$ with no superspace δ-functions. Since the lagrangian only has one term of this form, the kinetic term, the only terms that are renormalized are the $\Phi_i^\dagger \Phi_i$ terms, which results in a wavefunction renormalization.

B.7.3. Supersymmetry breaking

We do not see fermions accompanied by bosons degenerate in mass with them and vice versa; therefore, supersymmetry must be broken. Spontaneous breakdown of supersymmetry occurs when the vacuum state is not invariant to one or more of the supersymmetry generators. Either $Q_\alpha |0\rangle \neq 0$ or $\overline{Q}_{\dot\alpha} |0\rangle \neq 0$ for some α. When we calculate the energy of the ground state, we have

$$\begin{aligned} \{Q_\alpha, \overline{Q}_{\dot\alpha}\} |0\rangle &= 2\sigma^\mu_{\alpha\dot\alpha} P_\mu |0\rangle \\ \Rightarrow (\overline{\sigma}^\nu)^{\dot\alpha\alpha} \{Q_\alpha, \overline{Q}_{\dot\alpha}\} |0\rangle &= 2(\overline{\sigma}^\nu)^{\dot\alpha\alpha} \sigma^\mu_{\alpha\dot\alpha} P_\mu |0\rangle \\ &= 4\eta^{\nu\mu} P_\mu |0\rangle = 4P^\nu |0\rangle \end{aligned} \tag{B.274}$$

so that

$$\begin{aligned} P^\nu |0\rangle &= \frac{1}{4} (\overline{\sigma}^\nu)^{\dot\alpha\alpha} \{Q_\alpha, \overline{Q}_{\dot\alpha}\} |0\rangle \\ \Rightarrow P^0 |0\rangle &= \frac{1}{4} \left(Q_1 \overline{Q}_{\dot 1} + \overline{Q}_{\dot 1} Q_1 + Q_2 \overline{Q}_{\dot 2} + \overline{Q}_{\dot 2} Q_2 \right) |0\rangle \end{aligned} \tag{B.275}$$

which is positive semi-definite. If all of the Q's and $\overline{Q}$'s annihilate the vacuum $|0\rangle$, then the energy of the vacuum, $p^0 |0\rangle$, is zero. Suppose there is spontaneous symmetry breaking and $Q_1 |0\rangle = \sqrt{4E} \, |p, \lambda\rangle \neq 0$, then

$$\begin{aligned} P^0 |0\rangle &= \frac{1}{4} \left(\overline{Q}_{\dot 1} Q_1 \right) |0\rangle = \frac{1}{4} \sqrt{4E} \, \overline{Q}_{\dot 1} \, |p, \lambda\rangle \\ &= \frac{1}{4} \sqrt{4E} \sqrt{4E} \, |0\rangle = E \, |0\rangle \, . \end{aligned} \tag{B.276}$$

When one of the supersymmetry generators fails to annihilate the vacuum, the vacuum state has positive energy. As a result, whenever a supersymmetric vacuum state exists as a local minimum of the effective potential, it is a global minimum. If there are more than one supersymmetric vacuum states, they will all be global minima, each with zero energy. If the effective potential has a supersymmetric minimum, then it will be the vacuum state and will have zero energy. If the potential has no supersymmetric minimum, then the symmetry is spontaneously broken and the global minimum will be a positive energy vacuum state.

Another way of looking at spontaneous supersymmetry breaking is to notice that it must arise from one of the fields having a vacuum expectation value that is not invariant under supersymmetry transformations. For example, for the left-handed chiral multiplet, one of the following variations must have a non-zero vacuum expectation value

$$\begin{aligned}\langle 0|\,\delta\phi\,|0\rangle &= 2\xi\,\langle 0|\,\psi\,|0\rangle \\ \langle 0|\,\delta\psi^{\alpha}\,|0\rangle &= -\xi^{\alpha}\,\langle 0|\,F\,|0\rangle - i\partial_{\mu}\,\langle 0|\,\phi\,|0\rangle\,(\sigma^{\mu})^{\alpha\dot{\alpha}}\overline{\xi}_{\dot{\alpha}} \\ \langle 0|\,\delta F(x)\,|0\rangle &= -2i\partial_{\mu}\,\langle 0|\,\psi\,|0\rangle\,\sigma^{\mu}\overline{\xi}\end{aligned} \tag{B.277}$$

and the only one that can be non-zero without breaking Lorentz invariance is the auxiliary field

$$\langle 0|\,\delta\psi_i^{\alpha}\,|0\rangle = -\xi^{\alpha}\,\langle 0|\,F_i\,|0\rangle \neq 0 \tag{B.278}$$

for some chiral multiplet, Φ_i. Since we can write $F_i^{\dagger} = \frac{\partial W(\phi)}{\partial\phi_i}$, we see that, for chiral superfields, the auxiliary field, F_i, is the order parameter for spontaneous supersymmetry breaking.

When spontaneous symmetry breaking occurs, we expect to see a massless Goldstone fermion in the spectrum corresponding to the broken fermionic generator. This Goldstone fermion is analogous to the usual symmetry breaking where we see a massless Goldstone boson. The Goldstone fermion will be the ψ field in the multiplet where the F field has acquired a vacuum expectation value. When the symmetry becomes a local theory, such as in supergravity, the Goldstone fermion will be eaten by one of the gauge fields (the gravitino), and the other field will, thereby, acquire a mass.

Let us now briefly discuss the form that the effective potential must have in order for spontaneous symmetry breaking to occur. We know that the potential must have no supersymmetric minimum. A supersymmetric minimum of our effective potential will be a solution of

$$\begin{aligned}F_i^{\dagger} &= -\frac{\partial W(\phi)}{\partial\phi_i} = 0 \\ &= -m_{ij}\phi_j - \lambda_{ijk}\phi_j\phi_k \\ &\Rightarrow m_{ij}\phi_j + \lambda_{ijk}\phi_j\phi_k = 0\end{aligned} \tag{B.279}$$

so that $\phi = 0$ and $\phi = -\frac{m}{\lambda}$ are both zero energy supersymmetric minima. The only way we can get supersymmetry breaking is when there is no solution so that there is no zero energy minimum. An example is the O'Raifeartaigh model which has the effective potential

$$W(\Phi_1, \Phi_2, \Phi_3) = \lambda_1\Phi_1\left(\Phi_3^2 - M^2\right) + \mu\Phi_2\Phi_3. \tag{B.280}$$

The auxiliary fields in this model are

$$
\begin{aligned}
-F_1^\dagger &= \frac{\partial W}{\partial \phi_1} = \lambda_1 \left(\phi_3^2 - M^2\right) \\
-F_2^\dagger &= \frac{\partial W}{\partial \phi_2} = \mu \phi_3 \\
-F_3^\dagger &= \frac{\partial W}{\partial \phi_3} = 2\lambda_1 \phi_1 \phi_3 + \mu \phi_2 .
\end{aligned}
\tag{B.281}
$$

There is no solution which has all three of the F's equal to zero. Therefore, in this model, there is spontaneous supersymmetry breaking. The effective potential is given by

$$
V = \sum_{i=1}^{3} |F_i|^2 = \lambda_1^2 \left|\phi_3^2 - M^2\right|^2 + \mu^2 |\phi_3|^2 + |\mu\phi_2 + 2\lambda_1\phi_1\phi_3|^2 . \tag{B.282}
$$

The minimum of this effective potential is found by

$$
\begin{aligned}
\frac{\partial V}{\partial \phi_1} &= 2\lambda_1 \phi_3 \left(\mu\phi_2^\dagger + 2\lambda_1 \phi_1^\dagger \phi_3^\dagger\right) = 0 \\
\frac{\partial V}{\partial \phi_2} &= \mu \left(\mu\phi_2^\dagger + 2\lambda_1 \phi_1^\dagger \phi_3^\dagger\right) = 0 \\
\frac{\partial V}{\partial \phi_3} &= 2\lambda_1^2 \phi_3 \left(\phi_3^\dagger \phi_3^\dagger - M^2\right) + \mu^2 \phi_3^\dagger + 2\lambda_1 \phi_1 \left(\mu\phi_2^\dagger + 2\lambda_1 \phi_1^\dagger \phi_3^\dagger\right) = 0.
\end{aligned}
\tag{B.283}
$$

The solution is $\langle\phi_2\rangle = \langle\phi_3\rangle = 0$ with $\langle\phi_1\rangle$ left undetermined. Thus, the potential has a flat direction. This minimum has

$$
F_1^\dagger = \lambda_1 M^2 \quad F_2^\dagger = F_3^\dagger = 0 \tag{B.284}
$$

so that

$$
V = |F_1|^2 = \lambda_1^2 M^4 \tag{B.285}
$$

is positive definite as expected.

Since F_1 is non-zero, we expect that ψ_1 will be a Goldstone fermion. We look at the terms in equation (B.261) which are quadratic in the fermion fields

$$
\begin{aligned}
\mathcal{L}_m^F &= -m_{ij}\left(\psi_i\psi_j + \overline{\psi}_i\overline{\psi}_j\right) - 2\lambda_{ijk}\left(\overline{\psi}_i\overline{\psi}_j\phi_k^\dagger + \psi_i\psi_j\phi_k\right) \\
&= -\left(m_{ij} - 2\lambda_{ijk}\phi_k\right)\left(\psi_i\psi_j + \overline{\psi}_i\overline{\psi}_j\right),
\end{aligned}
\tag{B.286}
$$

which has a minimum value of

$$
\begin{aligned}
\mathcal{L}_m^F &= -\left(m_{ij} - 2\lambda_{ij1}\langle\phi_1\rangle\right)\left(\psi_i\psi_j + \overline{\psi}_i\overline{\psi}_j\right) \\
&= -\left(m_{ij} - 2\lambda_{ij1}\langle\phi_1\rangle\right)\overline{\Psi}_i\Psi_j \\
&= -M_{ij}\overline{\Psi}_i\Psi_j .
\end{aligned}
\tag{B.287}
$$

By the form of the potential in the present model (B.280), we see that

$$\begin{aligned} \lambda_{ijk} &= \lambda_{133} = \lambda_{313} = \lambda_{331} = \lambda_1 \\ m_{ij} &= m_{23} = m_{32} = \mu \end{aligned} \tag{B.288}$$

and all the other components are zero. We have

$$\begin{aligned} \mathcal{L}_m^F &= -\mu\overline{\Psi}_2\Psi_3 - \mu\overline{\Psi}_3\Psi_2 + 2\lambda_1 \langle\phi_1\rangle \overline{\Psi}_3\Psi_3 \\ &= -\left(\overline{\Psi}_1\overline{\Psi}_2\overline{\Psi}_3\right) \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \mu \\ 0 & \mu & -2\lambda_1\langle\phi_1\rangle \end{pmatrix} \begin{pmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \end{pmatrix} \\ &= -M_{ij}\overline{\Psi}_i\Psi_j. \end{aligned} \tag{B.289}$$

We see that Ψ_1 is massless as supposed.

Now, we look at the bosonic mass terms

$$\begin{aligned} \mathcal{L}_m^B &= \left[F_i^\dagger F_i\right]_m + \lambda_{ijk}\left\{\phi_i\phi_j F_k + \phi_i^\dagger\phi_j^\dagger F_k^\dagger\right\} \\ &= -\mu^2\phi_3^\dagger\phi_3 - \mu^2\phi_2^\dagger\phi_2 + \lambda_1^2 M^2\phi_3\phi_3 + \lambda_1^2 M^2\phi_3^\dagger\phi_3^\dagger. \end{aligned} \tag{B.290}$$

Let $\phi_3 = (a_3 + ib_3)$, giving

$$\begin{aligned} \mathcal{L}_m^B &= -\mu^2 a_3^2 - \mu^2 b_3^2 + \mu^2\phi_2^\dagger\phi_2 + 2\lambda_1^2 M^2 a_3^2 - 2\lambda_1^2 M^2 b_3^2 \\ &= -\left(\mu^2 - 2\lambda_1^2 M^2\right) a_3^2 - \left(\mu^2 + 2\lambda_1^2 M^2\right) b_3^2 - \mu^2\phi_2^\dagger\phi_2. \end{aligned} \tag{B.291}$$

The complex scalar field, ϕ_2, has mass μ; the field, ϕ_1, is massless, just like its fermionic superpartner, ψ_1; and the two real scalar fields, a_3 and b_3, have masses

$$m_{a_3}^2 = \mu^2 - 2\lambda_1^2 M^2 \quad m_{b_3}^2 = \mu^2 + 2\lambda_1^2 M^2. \tag{B.292}$$

The masses of a_3 and b_3 now differ from those of their superpartner ψ_3, indicating that, with supersymmetry breaking, we have the possibility of mass splitting between superpartners and we can hope to reproduce the physical spectrum of particles.

B.8. Superspace methods

A convenient fact about superspace is that we can perform many of our calculations of Feynman diagrams and Green's functions directly in superspace by manipulating entire sets of component field Feynman diagrams at the same time. The component field Green's functions can be simply derived from the superfield ones by taking the appropriate coefficients of the θ and $\overline{\theta}$'s. We will see that the non-renormalization theorems that were arrived at so tediously come about much more simply and are seen to vanish more explicitly since cancellation of component diagrams inside a superfield diagram shows up as the superfield diagram vanishing.

Let us begin by looking, once again, at our lagrangian,

$$\mathcal{L} = \int d^4\theta \left\{ \Phi_i^\dagger \Phi_i + W(\Phi)\delta(\overline{\theta}) + \text{h.c.} \right\}, \tag{B.293}$$

where it is now written as a single integral over superspace with a δ-function factor in the potential piece. Let us insert our particular form for the potential piece and write

$$\begin{aligned} \mathcal{L} = \int d^4\theta \Big\{ & \Phi_i^\dagger \Phi_i + \frac{1}{2} m_{ij} \Phi_i \Phi_j \delta(\overline{\theta}) + \frac{1}{2} m_{ij}^* \Phi_i^\dagger \Phi_j^\dagger \delta(\theta) \\ & + \frac{1}{3} \lambda_{ijk} \Phi_i \Phi_j \Phi_k \delta(\overline{\theta}) + \frac{1}{3} \lambda_{ijk}^* \Phi_i^\dagger \Phi_j^\dagger \Phi_k^\dagger \delta(\theta) \Big\}. \end{aligned} \tag{B.294}$$

To derive expressions for the propagators, we look at the free field terms in the lagrangian; these terms are

$$\begin{aligned} \mathcal{L}_0 &= \int d^4\theta \left\{ \Phi^\dagger \Phi + \frac{1}{2} m \Phi \Phi \delta(\overline{\theta}) + \frac{1}{2} m^* \Phi^\dagger \Phi^\dagger \delta(\theta) \right\} \\ &= F^\dagger F - \phi^\dagger \partial_\mu \partial^\mu \phi + i \overline{\psi} \sigma^\mu \partial_\mu \psi + m \left\{ F\phi + F^\dagger \phi^\dagger - \left(\psi\psi + \overline{\psi\psi} \right) \right\}. \end{aligned} \tag{B.295}$$

The propagators are

$$\begin{aligned} \langle 0 | T \left(\phi(y) \phi^\dagger(y') \right) | 0 \rangle &= i \Delta_F(y - y') \\ \langle 0 | T \left(\phi(y) F(y') \right) | 0 \rangle = \langle 0 | T(\phi(y) F(y') | 0 \rangle &= -im \Delta_F(y - y') \\ \langle 0 | T \left(F(y) F^\dagger(y') \right) | 0 \rangle &= i \Box \Delta_F(y - y') \\ \langle 0 | T \left(\psi_\alpha(y) \psi^\beta(y') \right) | 0 \rangle &= \frac{i}{2} \delta_\alpha^\beta m \Delta_F(y - y') \\ \langle 0 | T \left(\overline{\psi}^{\dot{\alpha}}(y) \overline{\psi}_{\dot{\beta}}(y') \right) | 0 \rangle &= \frac{i}{2} \delta_{\dot{\beta}}^{\dot{\alpha}} m \Delta_F(y - y') \\ \langle 0 | T \left(\psi_\alpha(y) \overline{\psi}_{\dot{\beta}}(y') \right) | 0 \rangle &= \frac{1}{2} \sigma^\mu_{\alpha\dot{\beta}} \partial_\mu \Delta_F(y - y'), \end{aligned} \tag{B.296}$$

where

$$\Delta_F(y - y') = \frac{1}{\Box - m^2}. \tag{B.297}$$

We then use these component propagators to construct the superfield propagators. The superfield propagators are

$$\begin{aligned} & \langle 0 | T \left(\Phi(y, \theta) \Phi^\dagger(y', \theta') \right) | 0 \rangle \\ &\quad = \langle 0 | T \left([\phi(y) + 2\theta\psi(y) + \theta\theta F(y)] \times [\phi(y') + 2\theta'\psi(y') + \theta'\theta' F(y')] \right) | 0 \rangle \\ &\quad = \theta'\theta' \langle 0 | T \left(\phi(y) F(y') \right) | 0 \rangle + \theta\theta \langle 0 | T \left(F(y) \phi(y') \right) | 0 \rangle \\ &\qquad\quad + 2\theta'^\alpha \theta^\beta \langle 0 | T \left(\psi_\alpha(y) \psi_\beta(y') \right) | 0 \rangle \\ &\quad = -im(\theta - \theta')^2 \Delta_F(y - y'). \end{aligned} \tag{B.298}$$

From the definition $y = x + i\theta\sigma\overline{\theta}$, we have

$$\langle 0| T\left(\Phi(x,\theta,\overline{\theta})\Phi^\dagger(x',\theta',\overline{\theta}')\right)|0\rangle = ie^{i(\theta\sigma^\mu\overline{\theta}+\theta'\sigma^\mu\overline{\theta}'-2\theta\sigma^\mu\overline{\theta}')\partial^x_\mu}\Delta_F(x-x')$$
$$\langle 0| T\left(\Phi(x,\theta,\overline{\theta})\Phi(x',\theta',\overline{\theta}')\right)|0\rangle = -im\delta(\theta-\theta')e^{i(\theta\sigma^\mu\overline{\theta}-\theta'\sigma^\mu\overline{\theta}')\partial^x_\mu}\Delta_F(x-x') \quad \text{(B.299)}$$
$$\langle 0| T\left(\Phi^\dagger(x,\theta,\overline{\theta})\Phi^\dagger(x',\theta',\overline{\theta}')\right)|0\rangle = im\delta(\overline{\theta}-\overline{\theta}')e^{-i(\theta\sigma^\mu\overline{\theta}-\theta'\sigma^\mu\overline{\theta}')\partial^x_\mu}\Delta_F(x-x').$$

We can use these propagators to find the superfield Green's functions to any order in perturbation theory. Using the expressions for the propagators, we can now easily show that the mass renormalization vanishes as we found in the previous section. We now calculate the one-loop corrections to the three propagators shown in Figure B.5..

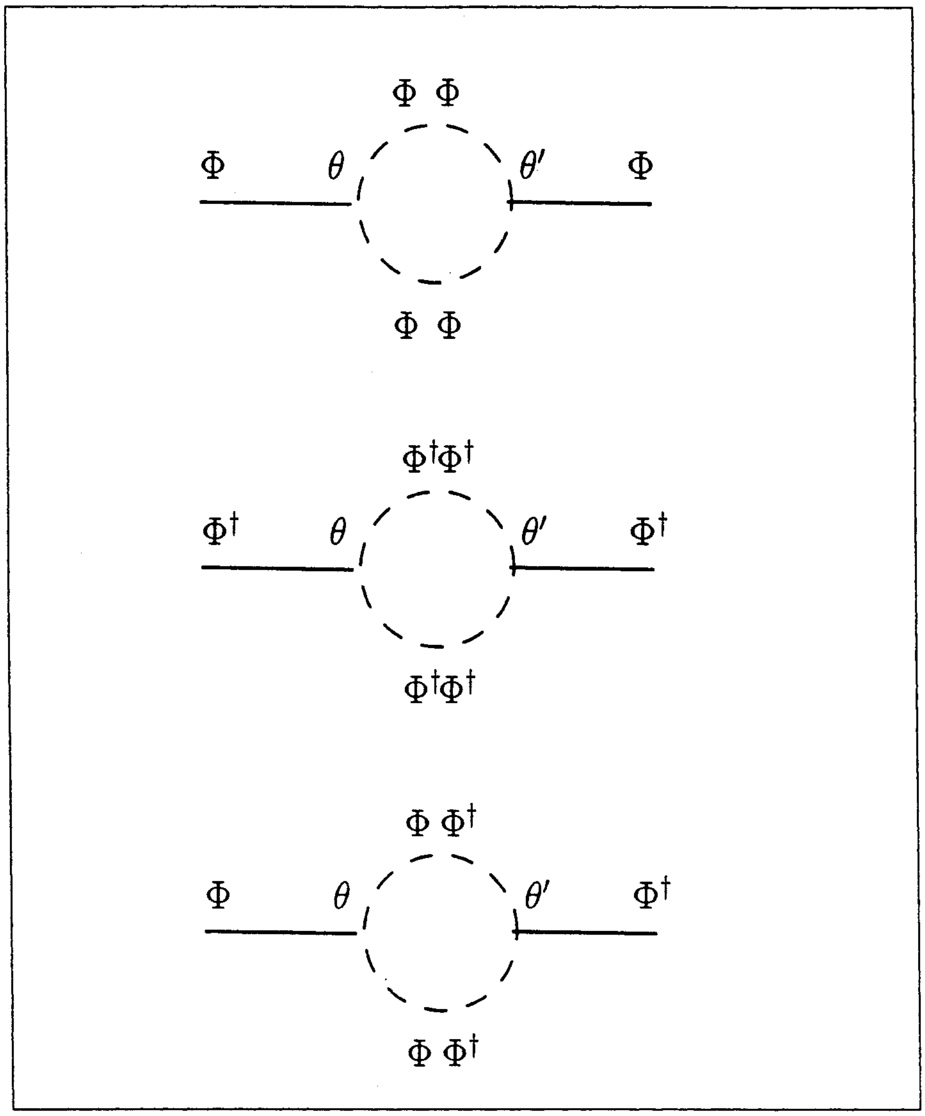


Figure B.5. 1-loop propagator corrections for the superfield Φ.

The first diagram in Figure B.5. is proportional to

$$\begin{aligned}
&\int d^4x\, d^4x'\, d^2\theta\, d^2\overline{\theta}\, d^2\theta'\, d^2\overline{\theta}'\, \delta(\theta)\delta(\theta')\Phi(x,\theta,\overline{\theta}) \\
&\times \delta(\theta-\theta')\exp\left\{i(\theta\sigma^\mu\overline{\theta}-\theta'\sigma^\mu\overline{\theta}')\partial^x_\mu\right\}\Delta_F(x-x') \\
&\times \delta(\theta-\theta')\exp\left\{i(\theta\sigma^\mu\overline{\theta}-\theta'\sigma^\mu\overline{\theta}')\partial^x_\mu\right\}\Delta_F(x-x')\Phi(x',\theta',\overline{\theta}')
\end{aligned} \tag{B.300}$$

which is zero by the following arguments. The delta functions are defined by

$$\int f(\eta)\delta(\eta)d\eta = f(0) \tag{B.301}$$

and since we have

$$\begin{aligned}
\int \frac{\partial}{\partial\eta} f(\eta)d\eta &= 0 \\
\int (A+B\eta)\, d\eta &= B \\
\int (A+B\eta)\, \eta d\eta &= A,
\end{aligned} \tag{B.302}$$

we have $\delta(\eta) = \eta$ and, therefore, $\delta(\eta)\delta(\eta) = 0$, so our contribution from the first loop is identically zero. Similarly, the second loop in Figure B.5. also has two factors of the same delta function and, therefore, also vanishes. We have shown that all of the component diagrams which these superfield diagrams represent cancel out. Hence, there are no mass renormalizations required. Notice also that $\delta(0) = 0$ so that any tadpole diagrams are also zero since they have a closed loop attached at a single point and, therefore, necessarily have a factor of $\delta(\theta-\theta)$ in them. The third diagram in Figure B.5. is proportional to

$$\begin{aligned}
&\int d^4xd^4x'd^2\theta d^2\overline{\theta}d^2\theta'd^2\overline{\theta}'\delta(\theta)\delta(\theta') \\
&\quad\times \Phi(x,\theta,\overline{\theta})\exp\left\{i(\theta\sigma^\mu\overline{\theta}+\theta'\sigma^\mu\overline{\theta}'-2\theta\sigma^\mu\overline{\theta}')\partial^x_\mu\right\}\Delta_F(x-x') \\
&\quad\times \delta(\theta-\theta')\exp\left\{i(\theta\sigma^\mu\overline{\theta}+\theta'\sigma^\mu\overline{\theta}'-2\theta\sigma^\mu\overline{\theta}')\partial^x_\mu\right\}\Delta_F(x-x')\Phi^\dagger(x',\theta',\overline{\theta}') \\
&= \int d^4xd^4x'd^2\theta d^2\overline{\theta}\Delta_F^2(x-x')\Phi(x,\theta,0)\exp\left(-2i\theta\sigma^\mu\overline{\theta}\partial^{x'}_\mu\right)\Phi^\dagger(x',0,\overline{\theta}).
\end{aligned} \tag{B.303}$$

The Δ_F^2 in the above expression will lead to a logarithmic divergence which will require a wavefunction renormalization. Notice how much simpler it was to arrive at this result using the superfield Feynman diagrams rather than the component fields as we did in Section B.7.2..

B.8.1. Superspace gauge theory

For each of the superspace derivatives. ∂_μ, D_α, and $\overline{D}_{\dot{\alpha}}$, we introduce a corresponding gauge potential superfield, $A_\mu(x,\theta,\overline{\theta})$, $A_\alpha(x,\theta,\overline{\theta})$, and $A_{\dot{\alpha}}(x,\theta,\overline{\theta})$. As a gauge parameter, we take a general superfield, $X(x,\theta,\overline{\theta})$. We do not make reality assumptions about any of these fields.

We now simplify notation by using capital Latin characters to denote an entire set of superspace indices,

$$\begin{aligned} D_A &= \left(\partial_\mu, D_\alpha, \overline{D}_{\dot{\alpha}}\right) \\ A_A &= \left(A_\mu, A_\alpha, A_{\dot{\alpha}}\right). \end{aligned} \tag{B.304}$$

The gauge connections transform as

$$A_A \to e^{-iX}\left(A_A - iD_A\right)e^{iX}. \tag{B.305}$$

We define the covariant derivatives

$$\nabla_A = D_A + iA_A. \tag{B.306}$$

In contrast to ordinary space, where $[\partial_\mu, \partial_\nu] = 0$, we now have

$$[D_A, D_B\} = iT^C_{AB} D_C, \tag{B.307}$$

which is the algebra, (B.194), of the spinor derivatives, with

$$T^\mu_{\alpha\dot{\beta}} = T^\mu_{\dot{\beta}\alpha} = 2\sigma^\mu_{\alpha\dot{\beta}} \tag{B.308}$$

and all other T's zero. The non-abelian structure gives us the Ricci identity

$$[\nabla_A, \nabla_B\} = iT^C_{AB}\nabla_C + iF_{AB}, \tag{B.309}$$

where

$$F_{AB} = D_A A_B \mp D_B A_A + i\left[A_A, A_B\right\} - iT^C_{AB} A_C. \tag{B.310}$$

With this definition, including the last term, F is covariant

$$F_{AB} \to e^{-iX} F_{AB} e^{iX} \tag{B.311}$$

and fulfills Bianchi identities which are modified by the presence of the "torsion" term

$$\sum_{[ABC\}} \left[\nabla_A F_{BC} - iT^D_{AB} F_{DC}\right] = 0. \tag{B.312}$$

In extending gauge theory to superspace, we had to introduce the complex connection superfields, A_A, which, between them, have 16 times as many components as we expect for the minimal version with just one real V. In order to get to the minimal scheme, we need super- and gauge-covariant constraints on the theory.

Constraints

Constraints must set covariant superfields to zero. The constraints must act on covariant object so that they do not break gauge invariance. The constraints must also act on superfields since they should not break supersymmetry. Therefore, the F_{AB} are suitable candidates for being constrained.[9]

F_{AB} decomposes into Lorentz representations as

$$F_{AB} : 2 \cdot (1,0) + 2 \cdot (0,1) + \left(1, \frac{1}{2}\right) + \left(\frac{1}{2}, 1\right) + \left(\frac{1}{2}, 0\right) + \left(0, \frac{1}{2}\right) + \left(\frac{1}{2}, \frac{1}{2}\right) \quad \text{(B.313)}$$

which leaves plenty of choice for constraints. We need to decide which constraints to impose. First, there are conventional constraints which come from the fact that the gauge potential remains unchanged if we add a covariant field in the adjoint representation, $A_A \to A_A + F_A$, to it. For example, we could add a superfield, $F_\mu = -\frac{i}{4}\bar{\sigma}_\mu^{\beta\alpha} F_{\alpha\dot{\beta}}$, to A_μ and get a new A_μ. With this redefinition, we get $F^{\text{new}}_{\alpha\dot{\beta}} = 0$, so $F_{\alpha\dot{\beta}} = 0$ is a safe constraint that we can always impose.

The second type of constraint comes from requiring that derivatives be gauge covariant. We expect the following constraints on "gauge-chiral" and "gauge-anti-chiral" superfields

$$\begin{aligned} \nabla_{\dot{\alpha}} \phi &= 0 \\ \nabla_\alpha \bar{\phi} &= 0. \end{aligned} \quad \text{(B.314)}$$

Such differential equations have integrability conditions. The equations are only consistent for all ϕ if

$$\{\nabla_\alpha, \nabla_\beta\} = iF_{\alpha\beta} = 0, \quad \left\{\nabla_{\dot{\alpha}}, \nabla_{\dot{\beta}}\right\} = iF_{\dot{\alpha}\dot{\beta}} = 0. \quad \text{(B.315)}$$

These constriants are called representation preserving constraints.

Another type of constraint which is necessary to get a minimal scheme is the soft reality constraint

$$A_\mu = (A_\mu)^\dagger + \text{gauge transformation.} \quad \text{(B.316)}$$

Summarizing our constraints,

$$\text{Im } A_\mu = \text{pure gauge}, \quad F_{\alpha\beta} = F_{\alpha\dot{\beta}} = F_{\dot{\alpha}\dot{\beta}} = F_{\dot{\alpha}\beta} = 0. \quad \text{(B.317)}$$

To solve these constraints means to express a field which is subject to some differential condition in terms of derivatives of other fields in such a way that the properties of the derivatives ensure that the original condition holds. For example, in electrodynamics the curl of the vector potential is a solution to the constraint which the homogeneous

[9]For an ordinary gauge theory, $F_{\mu\nu}$ contains the Lorentz representations $(1,0) + (0,1)$ which are linked to each other via the Bianchi identity. Setting one of them to zero, say $0 = F_{\mu\nu} + \frac{i}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}$, gives the equation of motion, $\nabla^\mu F_{\mu\nu} = 0$. We end up with on-shell conditions for the fields or even a trivial theory, $F_{\mu\nu} = 0$.

Maxwell's equations impose on the field strength.

To solve constraint $F_{\alpha\dot{\beta}} = 0$,

$$\begin{aligned} 0 = F_{\alpha\dot{\beta}} &= D_\alpha A_{\dot{\beta}} + \overline{D}_{\dot{\beta}} A_\alpha + i\left\{A_\alpha, A_{\dot{\beta}}\right\} - iT^{C}_{\alpha\dot{\beta}} A_C \\ &= D_\alpha A_{\dot{\beta}} + \overline{D}_{\dot{\beta}} A_\alpha + i\left\{A_\alpha, A_{\dot{\beta}}\right\} - i2\sigma^{\mu}_{\alpha\dot{\beta}} A_\mu \\ 2\sigma^{\mu}_{\alpha\dot{\beta}} A_\mu &= i\left(D_\alpha A_{\dot{\beta}} + \overline{D}_{\dot{\beta}} A_\alpha + i\left\{A_\alpha, A_{\dot{\beta}}\right\}\right) \end{aligned} \tag{B.318}$$

and multiplying each side by $\overline{\sigma}^{\dot{\beta}\alpha}_{\mu}$ gives

$$\begin{aligned} 2\overline{\sigma}^{\dot{\beta}\alpha}_{\mu}\sigma^{\mu}_{\alpha\dot{\beta}} A_\mu &= i\overline{\sigma}^{\dot{\beta}\alpha}_{\mu}\left(D_\alpha A_{\dot{\beta}} + \overline{D}_{\dot{\beta}} A_\alpha + i\left\{A_\alpha, A_{\dot{\beta}}\right\}\right) \\ 2\left(\overline{\sigma}_\mu \sigma^\mu\right) A_\mu &= i\overline{\sigma}^{\dot{\beta}\alpha}_{\mu}\left(D_\alpha A_{\dot{\beta}} + \overline{D}_{\dot{\beta}} A_\alpha + i\left\{A_\alpha, A_{\dot{\beta}}\right\}\right) \\ 2\left(-2\mathbb{I}\right) A_\mu &= i\overline{\sigma}^{\dot{\beta}\alpha}_{\mu}\left(D_\alpha A_{\dot{\beta}} + \overline{D}_{\dot{\beta}} A_\alpha + i\left\{A_\alpha, A_{\dot{\beta}}\right\}\right) \\ -4A_\mu &= i\overline{\sigma}^{\dot{\beta}\alpha}_{\mu}\left(D_\alpha A_{\dot{\beta}} + \overline{D}_{\dot{\beta}} A_\alpha + i\left\{A_\alpha, A_{\dot{\beta}}\right\}\right) \\ A_\mu &= -\frac{i}{4}\overline{\sigma}^{\dot{\beta}\alpha}_{\mu}\left(D_\alpha A_{\dot{\beta}} + \overline{D}_{\dot{\beta}} A_\alpha + i\left\{A_\alpha, A_{\dot{\beta}}\right\}\right). \end{aligned} \tag{B.319}$$

To solve the constraint, we end up with the connection on spacetime, A_μ, expressed in terms of the connections on the spinorial directions, A_α and $A_{\dot{\alpha}}$.

The representation preserving constraints are a bit more difficult to solve. First, we observe that the space is spinorially flat in that the curvatures that have both indices in a spinorial direction are zero. The solution to these constraints will be connections which are spinorially pure gauges

$$\begin{aligned} A_\alpha &= -ie^{2V} D_\alpha e^{-2V} = 2iD_\alpha V \\ A_{\dot{\alpha}} &= -ie^{2U}\overline{D}_{\dot{\alpha}} e^{-2U} = 2i\overline{D}_{\dot{\alpha}} U, \end{aligned} \tag{B.320}$$

which are a solution of equation (B.315) for general complex superfields V and U. Now, A_α and $A_{\dot{\alpha}}$ are pure gauges, and A_μ is expressed in terms of them, but since U and V are different and we have only one gauge parameter superfield, X, we can, at most, gauge one of the connection superfields away. In order to transform properly, the connections must satisfy equation (B.305) so that

$$\begin{aligned} A'_\alpha = e^{-iX}(A_\alpha - iD_\alpha)e^{iX} &= e^{-iX}(2iD_\alpha V - iD_\alpha)e^{iX} \\ -ie^{2V'}\left(D_\alpha e^{-2V'}\right) &= -i\left(-2D_\alpha V + D_\alpha(iX)\right) \\ D_\alpha(-2V') &= -2D_\alpha V + D_\alpha(iX) \\ 2V' &= 2V - iX \end{aligned} \tag{B.321}$$

and similar conditions hold for $A_{\dot{\alpha}}$ in equation (B.320). We have

$$e^{-2V'} = e^{-2V+iX}$$
$$e^{-2U'} = e^{-2U+iX}. \tag{B.322}$$

Notice that equations (B.320, B.321, and B.324) are still invariant under

$$e^{-2V} \to e^{-2V-i\overline{\Lambda}'}$$
$$e^{-2U} \to e^{-2U-i\Lambda}, \tag{B.323}$$

where Λ is chiral ($\overline{D}_{\dot{\alpha}}\Lambda = 0$) and $\overline{\Lambda}'$ is anti-chiral ($D_\alpha \overline{\Lambda}' = 0$).

Now we make a *partial* gauge choice (partial since we still have the invariance under transformations (B.323)); we choose X so that $iX = 2U$; therefore,

$$A'_{\dot{\alpha}} = e^{-2U}\left(2i\overline{D}_{\dot{\alpha}}U - i\overline{D}_{\dot{\alpha}}\right)e^{2U} = \left(2i\overline{D}_{\dot{\alpha}}U - 2i\overline{D}_{\dot{\alpha}}U\right) = 0, \tag{B.324}$$

so $A_{\dot{\alpha}} = 0$. Our covariant derivative and spacetime connection are now

$$\nabla_{\dot{\alpha}} = \overline{D}_{\dot{\alpha}} \tag{B.325}$$

and

$$A_\mu = -\frac{1}{4}\overline{\sigma}_\mu^{\dot{\beta}\alpha}\overline{D}_{\dot{\beta}}\left(e^{2V}D_\alpha e^{-2V}\right) = \frac{1}{2}\overline{\sigma}_\mu^{\dot{\beta}\alpha}\overline{D}_{\dot{\beta}}D_\alpha V. \tag{B.326}$$

The remaining gauge freedom on V is now

$$e^{-2V} \to e^{-i\overline{\Lambda}'}e^{-2V}e^{i\Lambda} \tag{B.327}$$

which leaves everything we have found so far invariant. The gauge choice, $A_{\dot{\alpha}} = 0$, is supersymmetric since a superfield has been gauged away, a superfield's worth of gauge parameters have been fixed, and the remaining gauge freedom is expressed in terms of superfields.

Our final constraint that we can use to fix the remaining gauge freedom is the soft reality constraint, $A_\mu = (A_\mu)^\dagger +$ (gauge transformation). If we choose $\Lambda' = \Lambda$, then we have

$$-2V^\dagger \to \left(-i\overline{\Lambda} - 2V + i\Lambda\right)^\dagger = i\Lambda - 2V^\dagger - i\overline{\Lambda} = -i\overline{\Lambda} - 2V^\dagger + i\Lambda, \tag{B.328}$$

which means that $V^\dagger = V$ and that A_μ is gauge equivalent to $(A_\mu)^\dagger$.

We have succeeded in imposing the constraints and two gauge choices, reducing everything to the desired minimal theory containing one prepotential superfield, V (which gives the connection superfield), and one chiral gauge parameter Λ. We still have the gauge transformations

$$e^{-2V} \to e^{-i\overline{\Lambda}}e^{-2V}e^{i\Lambda}. \tag{B.329}$$

There are three field strengths left; let us look at the lowest dimensional ones, $F_{\mu\alpha}$

and $F_{\mu\dot{\alpha}}$. Using our solutions for the connections,

$$\begin{aligned} F_{\mu\dot{\alpha}} &= D_\mu A_{\dot{\alpha}} - \overline{D}_{\dot{\alpha}} A_\mu + i\left[A_\mu, A_{\dot{\alpha}}\right] - iT^C_{\mu\dot{\alpha}} A_C \\ &= -\overline{D}_{\dot{\alpha}} A_\mu = \frac{1}{4}\overline{\sigma}^{\dot{\beta}\alpha}_\mu \overline{D}_{\dot{\alpha}} \overline{D}_{\dot{\beta}} \left(e^{2V} D_\alpha e^{-2V}\right) \\ &= -\frac{i}{4}\overline{\sigma}^{\dot{\beta}\alpha}_\mu \overline{D}_{\dot{\alpha}} \overline{D}_{\dot{\beta}} A_\alpha = -\frac{i}{8}\overline{\sigma}^{\alpha}_{\mu\dot{\beta}} \overline{D}^2 A_\alpha \equiv -i\overline{\sigma}^{\alpha}_{\mu\dot{\beta}} W_\alpha, \end{aligned} \tag{B.330}$$

where $W_\alpha = \frac{1}{8}\overline{D}^2 A_\alpha$ is obviously chiral (since three derivatives gives zero) and gauge covariant. Similarly,

$$F_{\mu\alpha} = -i\overline{\sigma}_{\mu\alpha\dot{\beta}} \overline{W}^{\dot{\beta}} \tag{B.331}$$

with $\overline{W}_{\dot{\alpha}} = (W_\alpha)^\dagger$. From the Bianchi identities (and $D^\alpha \overline{D}^2 D_\alpha = \overline{D}^{\dot{\alpha}} D^2 \overline{D}_{\dot{\alpha}}$), we have

$$\nabla^\alpha W_\alpha + \nabla_{\dot{\alpha}} \overline{W}^{\dot{\alpha}} = 0. \tag{B.332}$$

The superfield, W_α, is

$$W_\alpha = \frac{1}{8}\overline{D}^2 A_\alpha = \frac{i}{4}\overline{D}^2 D_\alpha V. \tag{B.333}$$

Since our general superfield, V, is real, $V = V^\dagger$, we can use equation (B.186) restricted to real fields to write

$$\begin{aligned} V = C &- i\theta\chi + i\overline{\chi}\overline{\theta} - \frac{i}{2}\theta^2 (M - iN) + \frac{i}{2}\overline{\theta}^2 (M + iN) - \theta\sigma^\mu\overline{\theta} A_\mu \\ &+ i\overline{\theta}^2\theta\left(\lambda - \frac{i}{2}\,\partial\!\!\!/\overline{\chi}\right) - i\theta^2\overline{\theta}\left(\overline{\lambda} - \frac{i}{2}\overline{\partial\!\!\!/}\chi\right) - \frac{1}{2}\theta^2\overline{\theta}^2\left(D + \frac{1}{2}\Box C\right), \end{aligned} \tag{B.334}$$

and now, we calculate W_α as

$$\begin{aligned} W_\alpha = \frac{i}{4}\overline{D}^2 D_\alpha \Big[C &- i\theta\chi + i\overline{\chi}\overline{\theta} - \frac{i}{2}\theta^2 (M - iN) + \frac{i}{2}\overline{\theta}^2 (M + iN) - \theta\sigma^\mu\overline{\theta} A_\mu \\ &+ i\overline{\theta}^2\theta\left(\lambda - \frac{i}{2}\,\partial\!\!\!/\overline{\chi}\right) - i\theta^2\overline{\theta}\left(\overline{\lambda} - \frac{i}{2}\overline{\partial\!\!\!/}\chi\right) - \frac{1}{2}\theta^2\overline{\theta}^2\left(D + \frac{1}{2}\Box C\right)\Big]. \end{aligned} \tag{B.335}$$

In the Wess-Zumino gauge, where $C = \chi = M = N = 0$,

$$W_\alpha = \frac{i}{4}\epsilon^{\dot{\beta}\dot{\alpha}} \overline{D}_{\dot{\alpha}} \overline{D}_{\dot{\beta}} D_\alpha \left[-\theta\sigma^\mu\overline{\theta} A_\mu + i\overline{\theta}^2\theta\lambda - i\theta^2\overline{\theta}\overline{\lambda} - \frac{1}{2}\theta^2\overline{\theta}^2 D \right]. \tag{B.336}$$

After much algebra, along with switching to 4-component notation, we arrive at

$$W = \begin{pmatrix} W_\alpha \\ \overline{W}^{\dot{\alpha}} \end{pmatrix} = e^{-i\theta\partial\!\!\!/\overline{\theta}} \left[\lambda - \frac{i}{2}\sigma^{\mu\nu}\theta F_{\mu\nu} + i\theta D - i\theta^2\sigma^\mu \nabla_\mu \overline{\lambda} \right], \tag{B.337}$$

where $\nabla_\mu \lambda = \partial_\mu \lambda + i\left[A_\mu, \lambda\right]$ is the covariant derivative of the "gaugino" field.

The action for the gauge fields is

$$S_{YM} = \frac{1}{16} \int d^4x\, d^2\theta\, \mathrm{Tr} W^2 + \mathrm{h.c.} \tag{B.338}$$

which is supersymmetric and gauge invariant. In components,

$$L_{YM} = \mathrm{Tr}\left(-\frac{1}{4} F_{\mu\nu}F^{\mu\nu} + \frac{i}{2}\overline{\lambda}\gamma^\mu \nabla_\mu \lambda + \frac{1}{2} D^2 \right). \tag{B.339}$$

The equations of motion are

$$\begin{aligned} \nabla^\nu F_{\mu\nu} &= \frac{1}{2}\left\{\overline{\lambda}, \gamma_\mu \lambda\right\} \\ i\gamma^\mu \nabla_\nu \lambda &= 0 \\ D &= 0. \end{aligned} \tag{B.340}$$

In superfield language,

$$\nabla^\alpha W_\alpha - \nabla_{\dot{\alpha}} \overline{W}^{\dot{\alpha}} = 0, \tag{B.341}$$

which differs from the Bianchi identity (B.332) by the minus sign in the middle. The action can be written as an integral over all of superspace, so we expect a wavefunction renormalization.

The kinetic part of the action is given by interacting the V superfield with ϕ as

$$S_{kin} = \frac{1}{8} \int d^4x\, d^2\theta\, d^2\overline{\theta}\, \overline{\phi} e^{-2V} \phi, \tag{B.342}$$

which is

$$\begin{aligned} L_{kin} = \mathrm{Tr}\Big[&\frac{1}{2}\nabla_\mu A \nabla^\mu A + \frac{1}{2}\nabla_\mu B \nabla^\mu B + \frac{i}{2}\overline{\psi}\gamma^\mu \nabla_\mu \psi + \frac{1}{2}F^2 + \frac{1}{2}G^2 \\ &- iA\,[B, D] - i\overline{\psi}\,[\lambda, A] - i\overline{\psi}\gamma_5\,[\lambda, B] \Big]. \end{aligned} \tag{B.343}$$

The most general renormalizable $N = 1$ supersymmetric lagrangian is the sum of L_{YM} and L_{kin}, along with gauge invariant self interactions. With indices denoting the gauge group representation, we have

$$\begin{aligned} L = \mathrm{Tr}\left[-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\lambda\sigma^\mu \nabla_\mu \overline{\lambda} + \frac{1}{2}D^2 \right] &+ \frac{1}{2}\nabla_\mu A_a \nabla^\mu A^{\dagger a} + i\psi_a \sigma^\mu \nabla_\mu \overline{\psi}^a + \frac{1}{2}F_a F^{\dagger a} \\ &- iA^{\dagger a}\lambda_a^b \psi_b - i\overline{\psi}^a \overline{\lambda}_a^b A_b + \frac{1}{2}A^{\dagger a} D_a^b A_b \\ &- \frac{1}{2}\left[b^a F_a + m^{ab}(F_a A_b + \psi_a \psi_b) + g^{abc}(A_a A_b F_c + 2\psi_a \psi_b A_c) + \mathrm{h.c.} \right]. \end{aligned} \tag{B.344}$$

In four component notation,

$$
\begin{aligned}
L = \mathrm{Tr}\Bigg\{ & -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{i}{2}\overline{\lambda}\gamma^\mu\nabla_\mu\lambda + \frac{1}{2}D^2 \\
& + \sum_{i=1}^{3}\Bigg(\frac{1}{2}\nabla_\mu A_i\nabla^\mu A_i + \frac{1}{2}\nabla_\mu B_i\nabla^\mu B_i + \frac{i}{2}\overline{\psi}_i\gamma^\mu\nabla_\mu\psi_i + \frac{1}{2}F_i^2 + \frac{1}{2}G_i^2 \\
& \qquad - i\left[A_i, B_i\right]D - i\overline{\psi}_i\left[\lambda, A_i\right] - i\overline{\psi}_i\gamma_5\left[\lambda, B_i\right]\Bigg) \\
& - \frac{i}{2}\sum_{ijk}\epsilon^{ijk}\left(\overline{\psi}_i\left[\psi_j, A_k\right] - \overline{\psi}_i\gamma_5\left[\psi_j, B_k\right] + \left[A_i, A_j\right]F_k - \left[B_i, B_j\right]F_k + 2\left[A_i, B_j\right]G_k\right)\Bigg\}.
\end{aligned}
\tag{B.345}
$$

B.9. Supersymmetric electroweak theory

In order to discuss the supersymmetric version of the electroweak theory, we will have to construct a supersymmetric version of the non-abelian Higgs mechanism which is required to break the gauge symmetry and generate masses. Therefore, we will have to assign chiral superfields for each of the chiral components of the known fermions. We also need Higgs scalars in our chiral superfields. The chiral supermultiplet, $\Phi(E_L)$, which contains the electron and neutrino (ν_{eL}, e_L) will have a scalar doublet component with the same electric charges $(\tilde{\nu}_{eL}, \tilde{e}_L)$ which could perhaps be used to generate masses for the down-like quarks and charged leptons.

We denote the three doublet superfields that contain the left chiral lepton doublets by $L^{(l)}$, where $l = e, \mu, \tau$, and the quark doublets are contained in the superfields, $Q^{(f)}$, where $f = 1, 2, 3$ and there is an undisplayed color index running over the three color labels of the 3 representation of $SU(3)$. The singlet superfields are denoted by l^c, $U^{c(f)}$, and $D^{c(f)}$, where the three family labels indicate the flavors u, c, and t in $U^{(f)}$ as well as d, s, and b in $D^{(f)}$. The superscript c indicates charge conjugate, and for the quark fields, there is an undisplayed color index running over the three components of the $\overline{3}$ representation. We denote the two Higgs doublet superfields by H_1 and H_2. Then, the Yukawa couplings necessary to generate masses for the charged leptons and quarks arise from the F-part of a superpotential of the form

$$
\begin{aligned}
W = \sum_l m^{(l)}\left(L^{(l)T}i\tau_2 H_2\right)l^c + \sum_{f,g} m^{(d)}_{fg}\left(Q^{(f)T}i\tau_2 H_2\right)D^{c(g)} \\
+ \sum_{f,g} m^{(u)}_{fg}\left(Q^{(f)T}i\tau_2 H_1\right)U^{c(g)},
\end{aligned}
\tag{B.346}
$$

where $m^{(u)}$ and $m^{(d)}$ are the up-like and down-like quark mass matrices, and there is an implicit sum over the undisplayed color labels in the two terms involving quark fields. Note that the factor $i\tau_2$ is just the matrix $\epsilon_{\alpha\beta}$ used to construct an $SU(2)$ singlet from two doublets of the internal symmetry group just like we did with the spacetime spinors, $\chi^\alpha\psi_\alpha$. To write down the remaining parts of the lagrangian, we denote the field strength superfields of the $SU(2)$ gauge theory by W_α^i, where $i = 1, 2, 3$, and the field strength of the $U(1)$ gauge theory by B_α. We have

$$\begin{aligned} W_\alpha^i &= \overline{D}^2 D_\alpha W^i + ig\epsilon^{ijk}\overline{D}^2 \left(D_\alpha W^i\right) W^k \\ B_\alpha &= \overline{D}^2 D_\alpha B, \end{aligned} \tag{B.347}$$

where W^i and B are the vector superfields.

The supersymmetric pure gauge lagrangian is then

$$L_V = \frac{1}{64}\left[W^{i\alpha}W_\alpha^i + W_\alpha^{i\dagger}W^{i\alpha\dagger} + 2B^\alpha B_\alpha\right]_F . \tag{B.348}$$

The interaction of the gauge supermultiplets with the chiral matter and the Higgs supermultiplets is fixed by their weak isospin and hypercharge quantum numbers, and is given by

$$\begin{aligned} L_\Phi = \Bigg[&\sum_l L^{(l)\dagger} \exp\left(ig\boldsymbol{\tau}\cdot\mathbf{W} - ig'B\right) L^{(l)} + \sum_f Q^{(f)\dagger}\exp\left(ig\boldsymbol{\tau}\cdot\mathbf{W} + \frac{i}{3}g'B\right) Q^{(f)} \\ &+ \sum_f U^{c(f)\dagger}\exp\left(-\frac{4}{3}ig'B\right) U^{c(f)} + \sum_f D^{c(f)\dagger}\exp\left(\frac{2}{3}ig'B\right) D^{c(f)} \\ &+ \sum_l l^{c\dagger}\exp\left(2ig'B\right) l^c + H_1^\dagger \exp\left(ig\boldsymbol{\tau}\cdot\mathbf{W} + ig'B\right) H_1 \\ &+ H_2^\dagger \exp\left(ig\boldsymbol{\tau}\cdot\mathbf{W} - ig'B\right) H_2 \Bigg]_D . \end{aligned} \tag{B.349}$$

B.9.1. Renormalization group equations

Recall that the renormalized coupling constants necessarily depend on the scale, M, used in their definition. Since the physics described by the bare lagrangian is independent of M, it must be that the coupling constants run with M, so physical quantities calculated with different values of M have the same values provided that they are calculated to sufficiently high order. The renormalization group equations specify precisely how the renormalized coupling constants vary. For a general gauge group, G, with coupling constant, g, the fine structure constant, $\alpha = g^2/4\pi$, satisfies

$$M\frac{d\alpha}{dM} = -\frac{b}{2\pi}\alpha^2 + \mathcal{O}(\alpha^3), \tag{B.350}$$

where

$$b = \frac{11}{3}C_1(G) - \frac{2}{3}\sum_R C_2(R) - \frac{1}{3}\sum_S C_2(S) \tag{B.351}$$

with $C_1(G)\delta^{ab} = f^{acd}f^{bcd}$ and f^{abc} as the structure constants of G. The sum over R is for Weyl fermions in representation, T_R, of G, and $C_2(R)\delta^{ab} = \mathrm{Tr}(T_R^a T_R^b)$. The sum over S is for scalars in representation, T_S, of G, and $C_2(S)\delta^{ab} = \mathrm{Tr}(T_S^a T_S^b)$.

In a supersymmetric theory, the gauge bosons are accompanied by gauginos in the same (adjoint) representation of G. Thus, the vector supermultiplet contributes

$$b(V) = \frac{11}{3}C_1(G) - \frac{2}{3}C_1(G) = 3C_1(G). \tag{B.352}$$

Similarly, in a chiral supermultiplet, each Weyl fermion is accompanied by a scalar in the same representation of G, so the contribution is

$$b(\Phi) = -\frac{2}{3}C_2(R) - \frac{1}{3}C_2(R) = -C_2(R). \tag{B.353}$$

In all, we have

$$b = 3C_1(G) - \sum_R C_2(R) \tag{B.354}$$

with the sum over the representations, R, of all chiral supermultiplets. For the supersymmetric standard model, we have for the QCD group $SU(3)$

$$b_3 = 9 - 2n_G, \tag{B.355}$$

where n_G is the number of fermion generations. Similarly, for the $SU(2)$, group we have

$$b_2 = 6 - 2n_G - \frac{1}{2}n_H, \tag{B.356}$$

where n_H is the number of Higgs doublets. For the $U(1)_Y$ group,

$$b_1 = -\frac{10}{3}n_G - \frac{1}{2}n_H. \tag{B.357}$$

Integrating the renormalization group equation between the mass of the Z boson, m_Z, and the unification scale, m_X, gives

$$\begin{aligned}\int_{\alpha_i(m_Z)}^{\alpha_i(m_X)} \frac{1}{\alpha_i^2}d\alpha_i &= -\int_{m_Z}^{m_X} \frac{b}{2\pi M}dM \\ \alpha_i^{-1}(M_Z) - \alpha_i^{-1}(M_X) &= \frac{b}{2\pi}\ln\left(\frac{M_Z}{M_X}\right).\end{aligned} \tag{B.358}$$

The unification scale is defined as the point at which all three coupling constants are equal.

$$\alpha_3(m_X) = \alpha_2(m_X) = \frac{5}{3}\alpha_1(m_X) \equiv \alpha_{GUT}(m_X), \tag{B.359}$$

where the factor of $\frac{5}{3}$ comes from the requirement that $U(1)_Y$ is associated with a diagonal generator of $SU(5)$ with normalization is determined by $C_2(R)\delta^{ab} = \mathrm{Tr}\left(T_R^a T_R^b\right)$. From the electromagnetic fine structure constant, the the weak mixing angle, $\tan^2\theta_W =$

α_1/α_2. We also have the relation

$$[8b_3 - 3(b_1 + b_2)]\sin^2\theta_W(m_Z) = 3(b_3 - b_2) + (5b_2 - 3b_1)\frac{\alpha_{em}(m_Z)}{\alpha_3(m_Z)}. \quad \text{(B.360)}$$

Using our supersymmetric values of b_i gives

$$(54 + 3n_H)\sin^2\theta_W(m_Z) = 9 + \frac{3}{2}n_H + (30 - n_H)\frac{\alpha_{em}(m_Z)}{\alpha_3(m_Z)} \quad \text{(B.361)}$$

which is independent of n_G. In the minimal supersymmetric model, we have $n_H = 2$. With the fine structure constants, $\alpha_{em}^{-1}(m_Z) = 128.8$ and $\alpha_3^{-1}(m_Z) = 0.108 \pm 0.005$, we have

$$\sin^2\theta_W(m_Z) = 0.234 \quad \text{(B.362)}$$

which is compared with the experimental value of $\sin^2\theta_W(m_Z) = 0.2336 \pm 0.0018$. This result is to be contrasted with the non-supersymmetric prediction of 0.21. We can also find the unification scale through

$$(18 + n_H)\ln\frac{m_X}{m_Z} = 2\pi\left[\alpha_{em}^{-1}(M_Z) - \frac{8}{3}\alpha_3^{-1}(M_Z)\right]. \quad \text{(B.363)}$$

With $m_Z = 91.176 \pm 0.023 GeV$, we have $m_X = 1.46 \times 10^{16} GeV$, which is only three orders of magnitude from the Plank scale and consistent with the lower value of proton decay. Non-supersymmetric theory gives $m_X = 5 \times 10^{14} GeV$. The common value of the three coupling constants is given by

$$(18 + n_H)\alpha_G^{-1}(m_X) = \alpha_3^{-1}(M_Z)\left(-6 + \frac{16}{3}n_G + n_H\right) + (9 - 2n_H)\alpha_{em}^{-1}(M_Z). \quad \text{(B.364)}$$

Taking $n_G = 3$ gives

$$\alpha_G^{-1}(m_X) = 25.8 \quad \text{(B.365)}$$

as the grand unified coupling constant.

APPENDIX C
INTRODUCTION TO STRING THEORY

In this appendix, we give a brief introduction to string theory. We begin with the non-relativistic classical string and then continue to with the superstring. We refer frequently to [6, 21, 39, 79, 80, 81].

C.1. The non-relativistic string

Let us begin with a discussion of the classical string [82]. We will model the classical string as a string of length L which consists of a collection of N point particles of mass m connected to one another by springs of spring constant k. We join the ends of this string together so that it forms a closed loop. The equilibrium separation, l, of the point masses will be considered small as compared with the radius of the loop so that a small length along the circumference can be approximated as a linear system.

The loop of string contains N particles. We will denote the displacement of the ith particle from its equilibrium position as $\overline{\phi}_i$. Since the first particle and the Nth particle are actually the same, we have the following periodic boundary conditions.

$$\begin{aligned} \overline{\phi}_0 &= \overline{\phi}_N \\ \frac{d\overline{\phi}_0}{dt} &= \frac{d\overline{\phi}_N}{dt}. \end{aligned} \tag{C.1}$$

The kinetic and potential energy of the string will be the sum of the kinetic and potential energies of the individual masses.

$$\begin{aligned} \mathrm{KE} &= \frac{1}{2}m\sum_{i=0}^{N-1}\left(\frac{d\overline{\phi}_i}{dt}\right)^2 \\ \mathrm{PE} &= \frac{1}{2}k\sum_{i=0}^{N-1}\left(\overline{\phi}_{i+1}-\overline{\phi}_i\right)^2 \end{aligned} \tag{C.2}$$

The total length of the loop of string is given by $L = Nl$ since there is a distance, l, between each of the N masses. The mass density, or mass per unit length of the string, is $\mu = m/l$, and the tension in the string is $T = kl$. If we parametrize the string with coordinate z, we can write the position of the ith particle as a function of z and time t so that $\overline{\phi}_i = \overline{\phi}_i(z_i, t)$.

Now we want to let the distance between the masses $l \to 0$ and let the number of masses $N \to \infty$ while keeping the total length, mass density, and tension of the string constant. In this way, the displacement and energy of the string can be defined in terms of a continuous field $\overline{\phi}(z, t)$. Our kinetic and potential energy equations now become

$$\begin{aligned}
\mathrm{KE} &= \frac{1}{2}m\sum_{i=0}^{N-1}\left(\frac{d\overline{\phi}_i}{dt}\right)^2 = \frac{1}{2}\frac{m}{l}\sum_{i=0}^{N-1} l\left(\frac{d\overline{\phi}_i}{dt}\right)^2 \\
&\rightarrow \frac{1}{2}\mu\int_0^L dz\left(\frac{d\overline{\phi}(z,t)}{dt}\right)^2 \\
\mathrm{PE} &= \frac{1}{2}k\sum_{i=0}^{N-1}\left(\overline{\phi}_{i+1}-\overline{\phi}_i\right)^2 = \frac{1}{2}kl\sum_{i=0}^{N-1} l\left(\frac{\overline{\phi}_{i+1}-\overline{\phi}_i}{l}\right)^2 \\
&\rightarrow \frac{1}{2}T\int_0^L dz\left(\frac{d\overline{\phi}(z,t)}{dz}\right)^2.
\end{aligned} \tag{C.3}$$

We can now write down the hamiltonian, which is the total energy, and the lagrangian for the classical string as

$$\begin{aligned}
H = \mathrm{KE}+\mathrm{PE} &= \int_0^L dz\left\{\frac{1}{2}\mu\left(\frac{d\overline{\phi}(z,t)}{dt}\right)^2 + \frac{1}{2}T\left(\frac{d\overline{\phi}(z,t)}{dz}\right)^2\right\} = \int_0^L dz\mathcal{H}(z,t), \\
L = \mathrm{KE}-\mathrm{PE} &= \int_0^L dz\left\{\frac{1}{2}\mu\left(\frac{d\overline{\phi}(z,t)}{dt}\right)^2 - \frac{1}{2}T\left(\frac{d\overline{\phi}(z,t)}{dz}\right)^2\right\} = \int_0^L dz\mathcal{L}(z,t),
\end{aligned} \tag{C.4}$$

where we have defined the hamiltonian and lagrangian densities, $\mathcal{H}$ and $\mathcal{L}$, respectively. Let us now redefine $\overline{\phi}(z,t)$ to absorb the tension by letting $\overline{\phi} \rightarrow \sqrt{T}\,\overline{\phi} \equiv \phi$. Since the wave velocity is defined by $v^2 = T/\mu$, we can write our hamiltonian and lagrangian densities as

$$\mathcal{H} = \frac{1}{2}\left[\frac{1}{v^2}\left(\frac{d\phi}{dt}\right)^2 + \left(\frac{d\phi}{dz}\right)^2\right] \tag{C.5}$$

and

$$\mathcal{L} = \frac{1}{2}\left[\frac{1}{v^2}\left(\frac{d\phi}{dt}\right)^2 - \left(\frac{d\phi}{dz}\right)^2\right]. \tag{C.6}$$

The action is $S = S(\dot{\phi}, \phi', \phi)$ with variation δS given by

$$\begin{aligned}
\delta S &= \frac{\partial S}{\partial\dot{\phi}}\delta\dot{\phi} + \frac{\partial S}{\partial\phi'}\delta\phi' + \frac{\partial S}{\partial\phi}\delta\phi \\
&= \int dtdz\left(\frac{\partial\mathcal{L}}{\partial\dot{\phi}}\delta\dot{\phi} + \frac{\partial\mathcal{L}}{\partial\phi'}\delta\phi' + \frac{\partial\mathcal{L}}{\partial\phi}\delta\phi\right) = 0.
\end{aligned} \tag{C.7}$$

Since $\dot{\phi} = \frac{\partial \phi}{\partial t}$ and $\phi' = \frac{\partial \phi}{\partial z}$, we have

$$\delta S = \int dtdz \left(\frac{\partial \mathcal{L}}{\partial \dot{\phi}} \delta \frac{\partial \phi}{\partial t} + \frac{\partial \mathcal{L}}{\partial \phi'} \delta \frac{\partial \phi}{\partial z} + \frac{\partial \mathcal{L}}{\partial \phi} \delta \phi \right),$$
$$= \int dtdz \left(\frac{\partial}{\partial t} \left[\frac{\partial \mathcal{L}}{\partial \dot{\phi}} \delta \phi \right] + \frac{\partial}{\partial z} \left[\frac{\partial \mathcal{L}}{\partial \phi'} \delta \phi \right] - \frac{\partial}{\partial t} \frac{\partial \mathcal{L}}{\partial \dot{\phi}} \delta \phi - \frac{\partial}{\partial z} \frac{\partial \mathcal{L}}{\partial \phi'} \delta \phi + \frac{\partial \mathcal{L}}{\partial \phi} \delta \phi \right) = 0. \tag{C.8}$$

We see that the first two terms in the integral are total derivatives and should be evaluated at the endpoints of the integration. In the first case, the endpoints are at the infinite past and the infinite future. We will assume that the variation, $\delta\phi$, is localized in time and vanishes at infinity. We see that, since we are dealing with a closed string with periodic boundary conditions, the second term, $\frac{\partial \mathcal{L}}{\partial \phi'} \delta \phi(L) - \frac{\partial \mathcal{L}}{\partial \phi'} \delta \phi(0)$, will also vanish. We are left with

$$\delta S = \int dtdz \left(-\frac{\partial}{\partial t} \frac{\partial \mathcal{L}}{\partial \dot{\phi}} - \frac{\partial}{\partial z} \frac{\partial \mathcal{L}}{\partial \phi'} + \frac{\partial \mathcal{L}}{\partial \phi} \right) \delta \phi = 0, \tag{C.9}$$

and, since the variation, $\delta\phi$, is arbitrary, the term inside the brackets must vanish. We have

$$\frac{\partial \mathcal{L}}{\partial \phi} - \frac{\partial}{\partial t} \frac{\partial \mathcal{L}}{\partial \dot{\phi}} - \frac{\partial}{\partial z} \frac{\partial \mathcal{L}}{\partial \phi'} = 0. \tag{C.10}$$

Plugging in expression (C.6) for $\mathcal{L}$, we get our equation of motion

$$\frac{1}{v^2} \frac{\partial^2 \phi}{\partial t^2} - \frac{\partial^2 \phi}{\partial z^2} = 0. \tag{C.11}$$

We would like to solve the equation of motion for the function, $\phi(z,t)$, which will give us the shape of the string at each point of the string over all time. We use separation of variables and write $\phi(z,t) = F(t)G(z)$, giving

$$\frac{1}{v^2} \frac{1}{F} \frac{\partial^2 F}{\partial t^2} = \frac{1}{G} \frac{\partial^2 G}{\partial z^2}, \tag{C.12}$$

and since the left-hand side is a function of t only and the right-hand side is a function of z only, they must be equal to a constant, C.

$$\begin{aligned} &\frac{1}{v^2} \frac{1}{F} \frac{\partial^2 F}{\partial t^2} = C \\ &\qquad \Rightarrow F = F(0) e^{\pm i\omega t} \quad \text{where } \omega = -v^2 C \\ &\frac{1}{G} \frac{\partial^2 G}{\partial z^2} = C \\ &\qquad \Rightarrow G = G(0) e^{\pm ikz} \quad \text{where } k^2 = -C. \end{aligned} \tag{C.13}$$

We have

$$\phi(z,t) = \phi(0)e^{\pm i(kz\pm\omega t)} \text{ where } v = \pm\frac{\omega}{k}. \tag{C.14}$$

Our periodic boundary conditions (C.1) tell us that

$$k = \frac{2\pi n}{L} \equiv k_n \text{ where } n \in \mathbb{Z}. \tag{C.15}$$

Thus, $\omega_n = vk_n$. Our solution is

$$\phi_n(z,t) = \phi(0)e^{i(k_n z\pm\omega_n t)}. \tag{C.16}$$

These quantities are the normal modes of vibration of the string (since the periodic boundary conditions require the solutions to be standing waves). We have both positive and negative frequency solutions depending on the sign in the exponential. k_n takes on positive and negative values depending on n, but we will choose the convention that ω_n is always positive so that $-\omega_n$ will denote the negative frequency waves.

To fix the constant factor, $\phi(0)$, we write our solution as

$$\phi_n(z,t) = \phi(0)e^{ik_n(z\pm vt)} = \phi_n^L(z+vt) + \phi_n^R(z-vt), \tag{C.17}$$

meaning we have left-moving and right-moving waves for the plus and minus signs in the exponential, respectively. Mathematically, we have $e^{-i\theta}e^{i\theta} = 1$ and $e^{-i\theta} = e^{i(\theta-\frac{\pi}{2})}$ so that $\phi_n^*(z,t)$ is a standing wave which is phase shifted by $\frac{\pi}{2}$. If we integrate $\phi_n^*\phi_n$ over the entire string, we should simply get the length of the string since the waves will have canceled each other out, leaving an unstretched string. Mathematically,

$$\int_0^L dz\phi_n^*(z,t)\phi_n(z,t) = L = L\phi(0)^2, \tag{C.18}$$

which gives $\phi(0)^2 = 1$. Instead, it is more convenient for us to redefine our wavefunctions so that $\phi(0) = \frac{1}{\sqrt{L}}$ which allows us to rewrite our normalization condition as

$$\int_0^L dz\phi_n^*(z,t)\phi_n(z,t) = 1. \tag{C.19}$$

If we have two different modes, we have the condition

$$\begin{aligned}\int_0^L dz\phi_n^*(z,t)\phi_m(z,t) &= \int_0^L dz\frac{1}{L}e^{-ik_n(z\pm vt)}e^{ik_m(z\pm vt)} = \int_0^L dz\frac{1}{L}e^{i(k_m-k_n)(z\pm vt)} \\ &= \frac{1}{L}e^{\pm i(k_m-k_n)vt}\int_0^L dze^{i(k_m-k_n)z} = \frac{1}{L}e^{\pm i(k_m-k_n)vt}\int_0^L dze^{i\frac{2\pi(m-n)}{L}z} = \delta_{n,m},\end{aligned} \tag{C.20}$$

which allows us to use the convenient normalization

$$\int_0^L dz\phi_n^*(z,t)\phi_m(z,t) = \delta_{n,m}. \tag{C.21}$$

Now, we have found the normal modes of the string, the standing wave solutions, but we want to express a general *real* field, $\phi(z,t)$, which is not necessarily stationary. We write the field as a linear combination of the complete set of normalized modes.

$$\phi(z,t) = \sum_{n=-\infty}^{+\infty} \frac{c_n}{\sqrt{L}} \left[a_n(t)e^{ik_n z} + a_n^*(t)e^{-ik_n z}\right], \tag{C.22}$$

where we have absorbed the time dependence into the $a_n(t)$ and $a_n^*(t)$ coefficients of the expansion. The coefficients must also be complex conjugates of each other in order for ϕ to be real (complex conjugates added together give real). We will fix the real normalization factor, c_n, later.

Notice now that the coefficients, $a_n(t) = a_n(0)e^{-i\omega_n t}$, satisfy the equation of a simple harmonic oscillator

$$\ddot{a}_n(t) + \omega_n^2 a_n(t) = 0,$$

meaning that each normal mode in the expansion behaves as an independent, simple harmonic oscillator.

In plugging our expressions back into the equation for kinetic energy, we set the constant, c_n, so that the kinetic energy assumes the simple form

$$H = \sum \omega_n a_n^*(t) a_n(t). \tag{C.23}$$

Now, we want to quantize this system. The a_n and a_n^* will become annihilation and creation operators of the states of the string. These states are called *phonons*. The $a_n^* a_n$ becomes a number operator. The total energy of the string is given by adding up the number of phonons in each state, n, multiplied by the energy of the state, ω_n (in units where $\hbar = 1$).

To derive the energy, we first want to form the generalized position and momentum, $q_n(t)$ and $p_n(t)$, from the $a_n(t)$ and $a_n^*(t)$. Since the position and momentum must be real, we need to form them with $a + a^*$. We choose

$$\begin{aligned} q_n(t) &= \frac{1}{\sqrt{2\omega_n}} \left[a_n(t) + a_n^*(t)\right] \\ p_n(t) &= \frac{dq_n}{dt} \\ &= -\frac{i\omega_n}{\sqrt{2\omega_n}} \left[a_n(t) - a_n^*(t)\right] \end{aligned} \tag{C.24}$$

which means that

$$\begin{aligned} a_n(t) &= \frac{ip_n + \omega_n q_n}{\sqrt{2\omega_n}} \\ a_n^*(t) &= \frac{-ip_n + \omega_n q_n}{\sqrt{2\omega_n}}. \end{aligned} \tag{C.25}$$

The hamiltonian becomes

$$\begin{aligned} H &= \sum \omega_n a_n^*(t) a_n(t) = \sum \omega_n \frac{(-ip_n + \omega_n q_n)(ip_n + \omega_n q_n)}{2\omega_n} \\ &= \sum \frac{1}{2}\left(p_n^2 + \omega_n^2 q_n^2\right) \end{aligned} \tag{C.26}$$

which is the sum of independent harmonic oscillator hamiltonians as can be seen by Hamilton's equations of motion

$$\begin{aligned} \dot{q}_n &= \frac{\partial H}{\partial p_n} = p_n \\ \dot{p}_n &= -\frac{\partial H}{\partial q_n} = -\omega_n^2 q_n \quad \Rightarrow \quad \ddot{q}_n + \omega_n^2 q_n = 0. \end{aligned} \tag{C.27}$$

The Poisson bracket is defined as

$$\{A, B\}_{PB} = \sum_n \left(\frac{\partial A}{\partial p_n}\frac{\partial B}{\partial q_n} - \frac{\partial A}{\partial q_n}\frac{\partial B}{\partial p_n} \right), \tag{C.28}$$

so we have

$$\begin{aligned} \{q_n, p_m\}_{PB} &= \sum_k \left(\frac{\partial q_n}{\partial p_k}\frac{\partial p_m}{\partial q_k} - \frac{\partial q_n}{\partial q_k}\frac{\partial p_m}{\partial p_k} \right) \\ &= -\sum_k \frac{\partial q_n}{\partial q_k}\frac{\partial p_m}{\partial p_k} \\ &= -\sum_k \delta_{nk}\delta_{mk} = -\delta_{mn} \\ \{q_n, q_m\}_{PB} &= \{p_n, p_m\}_{PB} = 0. \end{aligned} \tag{C.29}$$

To quantize, the Poisson bracket becomes the commutator as follows

$$\{A, B\}_{PB} \longrightarrow \frac{i}{\hbar}[A, B], \tag{C.30}$$

and we have the following commutation relations

$$\begin{aligned} [q_n, p_m] &= i\hbar\delta_{mn} \\ [q_n, q_m] &= [p_n, q_m] = 0. \end{aligned} \tag{C.31}$$

Plugging in the expressions from equation (C.24), we have

$$\begin{aligned} \left[a_n, a_m^\dagger\right] &= \delta_{mn} \\ [a_n, a_m] &= \left[a_n^\dagger, a_m^\dagger\right] = 0, \end{aligned} \tag{C.32}$$

where we have generalized the complex conjugation to hermitian conjugation, and we have set $\hbar = 1$. Our fields now become

$$\phi(z,t) = \sum_{n=-\infty}^{+\infty} \frac{1}{\sqrt{2\omega_n L}} \left[a_n e^{i(k_n z - \omega_n t)} + a_n^\dagger e^{-i(k_n z - \omega_n t)}\right] = \phi^{(+)}(z,t) + \phi^{(-)}(z,t), \quad \text{(C.33)}$$

and we can interpret $\phi^{(+)}(z,t)$ as an annihilation operator and $\phi^{(-)}(z,t)$ as a creation operator.

Now that we have seen how to get a quantized string from a classical, non-relativistic string, we want to generalize some of the aspects of our construction. What if we would have started with a rubber membrane rather than a string? In that case, we would have had to add one dimension to the space. Each mass would then be connected to the other masses by springs in a 2-d lattice. Instead, add two dimensions of space to our problem so that we have point masses connected by springs in each direction of 3 dimensional space. The way that these extra dimensions change our equations is as follows. The hamiltonian and lagrangian densities (C.5 and C.6) become

$$\begin{aligned} \mathcal{H} &= \frac{1}{2}\left[\frac{1}{v^2}\left(\frac{d\phi}{dt}\right)^2 + \left(\frac{d\phi}{dx}\right)^2 + \left(\frac{d\phi}{dy}\right)^2 + \left(\frac{d\phi}{dz}\right)^2\right] = \frac{1}{2}\left[\frac{1}{v^2}\left(\frac{d\phi}{dt}\right)^2 + \nabla^2\phi\right] \\ \mathcal{L} &= \frac{1}{2}\left[\frac{1}{v^2}\left(\frac{d\phi}{dt}\right)^2 - \left(\frac{d\phi}{dx}\right)^2 - \left(\frac{d\phi}{dy}\right)^2 - \left(\frac{d\phi}{dz}\right)^2\right] = \frac{1}{2}\left[\frac{1}{v^2}\left(\frac{d\phi}{dt}\right)^2 - \nabla^2\phi\right] \end{aligned} \quad \text{(C.34)}$$

or

$$\mathcal{L} = \frac{1}{2}\eta^{\mu\nu}\partial_\mu\phi\;\partial_\nu\phi \equiv \frac{1}{2}\partial^\mu\phi\;\partial_\mu\phi \quad \text{(C.35)}$$

which we recognize as the Klein-Gordon lagrangian.

Let us now return to strings of only one spatial dimension, allowing the displacement of the string, ϕ, to be in any of the three spatial dimensions. Let σ^0 and σ^1 be our new string coordinates replacing the ones that we had previously called vt and z. Then, we have $d\sigma^0 = vdt$ and $d\sigma^1 = dz$. We write our displacement as a four-dimensional vector, X^μ, so that our lagrangian density becomes

$$\mathcal{L} = \frac{1}{2}\eta^{ab}\eta_{\mu\nu}\partial_a X^\mu \partial_a X^\nu \;\equiv \frac{1}{2}\eta^{ab}\partial_a X^\mu \partial_b X_\mu, \quad \text{(C.36)}$$

where we have defined

$$\eta^{ab} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad \text{(C.37)}$$

The action action for this system is then

$$S = \int_M d^2\sigma\;\eta^{ab}\partial_a X^\mu \partial_b X^\nu G_{\mu\nu}(X), \quad \text{(C.38)}$$

where we have absorbed the factor of $\frac{1}{2}$. The integration is over the world sheet of the string, and $G_{\mu\nu}(X)$ is a general metric on spacetime which may be a function of the position of the points on the string, X^μ. In fact, our action is the same as the free string action that we have alreadv seen in Chapter 4, equation (4.11).

If we consider a flat spacetime so that the metric, $G_{\mu\nu}$, becomes the flat metric, $\eta_{\mu\nu}$, we can take the variation and find the wave equation analogous to equation (C.11)

$$\eta^{ab}\partial_a\partial_b X^\mu = (\partial_0^2 - \partial_1^2)X^\mu = 0. \tag{C.39}$$

The solution is a sum of independent left- and right-moving degrees of freedom exactly as before

$$X^\mu(\sigma_0, \sigma_1) = X_L^\mu(\sigma_0 + \sigma_1) + X_R^\mu(\sigma_0 - \sigma_1), \tag{C.40}$$

and in the quantized case, these terms become annihilation and creation operators for left and right moving modes of the the string. We are now in a position to consider a generalization of the string theory that we have discussed in this dissertation by using supersymmetry as outlined in Appendix B.

C.2. Superstring theory

In the conformal gauge where the world sheet metric is $h_{\alpha\beta} = \eta_{\alpha\beta}e^\phi$, the free superstring action is given by

$$S = -\frac{1}{2\pi}\int d^2\sigma \left[\partial_\alpha X^\mu \partial^\alpha X_\mu - i\overline{\psi}^\mu \rho^\alpha \partial_\alpha \psi_\mu\right]. \tag{C.41}$$

The conformal factor, e^ϕ, drops out of the integration measure. The $X^\mu(\sigma,\tau), \mu = 0, 1, ..., D-1$ are coordinates for a bosonic string propagating in D spacetime dimensions, and (σ, τ) are the usual world sheet coordinates. $\psi_A^\mu(\sigma,\tau)$ is a D-plet of Majorana fermions, where A is a two component chirality index, and $\overline{\psi} = \psi^T\rho^0$, as is usual for real Majorana spinors. ψ_A^μ transforms in the vector representation of the Lorentz group, $SO(D-1,1)$ and satisfies the equal time anti-commutation relations:

$$\{\psi_A^\mu(\sigma), \psi_B^\nu(\sigma')\} = \pi\eta^{\mu\nu}\delta_{AB}\delta(\sigma - \sigma'). \tag{C.42}$$

The ρ^α are the two dimensional Dirac matrices

$$\rho^0 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \rho^1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \tag{C.43}$$

in an imaginary basis. They satisfy the Clifford algebra

$$\{\rho^\alpha, \rho^\beta\} = -2\eta^{\alpha\beta}\cdot 1. \tag{C.44}$$

Notice that $\rho^0\rho^{\alpha T}\rho^0 = -\rho^\alpha$ and $\partial\!\!\!/^2 = -\Box$.

C.2.1. World sheet supersymmetry

For an introduction to supersymmetry, the reader is encouraged to refer to Appendix B. The free superstring action (C.41) is invariant under the infinitesimal supersymmetry transformations

$$\delta X^\mu = \bar{\epsilon}\psi^\mu, \quad \delta\psi^\mu = -i\rho^\alpha\partial_\alpha X^\mu\epsilon \quad \delta\overline{\psi}^\mu = i\bar{\epsilon}\rho^\alpha\partial_\alpha X^\mu. \tag{C.45}$$

where ϵ is a constant infinitesimal anti-commuting spinor.

The variation of the action (C.41) is given by

$$\delta S = -\frac{1}{2\pi}\int d^2\sigma\Big(\delta\left(\partial_\alpha X^\mu \partial^\alpha X_\mu\right) + \delta\left(-i\overline{\psi}^\mu\ \partial\!\!\!/\psi_\mu\right)\Big). \tag{C.46}$$

We will compute each of the terms above separately, beginning with the bosonic piece.

$$\begin{aligned}\delta\left(\partial_\alpha X^\mu \partial^\alpha X_\mu\right) &= 2\partial_\alpha(\delta X^\mu)\partial^\alpha X_\mu = 2\partial_\alpha(\bar{\epsilon}\psi^\mu)\partial^\alpha X_\mu \\ &= 2(\partial_\alpha\bar{\epsilon})\psi^\mu\partial^\alpha X_\mu + 2\bar{\epsilon}\partial_\alpha\psi^\mu\partial^\alpha X_\mu.\end{aligned} \tag{C.47}$$

Integrating the second term by parts leaves

$$\begin{aligned}\delta\left(\partial_\alpha X^\mu \partial^\alpha X_\mu\right) &= 2(\partial_\alpha\bar{\epsilon})\psi^\mu\partial^\alpha X_\mu - 2(\partial_\alpha\bar{\epsilon})\psi^\mu\partial^\alpha X_\mu - 2\bar{\epsilon}\psi^\mu\partial_\alpha\partial^\alpha X_\mu \\ &= -2\bar{\epsilon}\psi^\mu\Box X_\mu.\end{aligned} \tag{C.48}$$

The variation of the fermion piece is

$$\begin{aligned}\delta\left(-i\overline{\psi}^\mu\ \partial\!\!\!/\psi_\mu\right) &= -i(\delta\overline{\psi}^\mu)\ \partial\!\!\!/\psi_\mu - i\overline{\psi}^\mu\ \partial\!\!\!/(\delta\psi_\mu) \\ &= -i(i\bar{\epsilon}\ \partial\!\!\!/X^\mu)\ \partial\!\!\!/\psi_\mu - i\overline{\psi}^\mu\ \partial\!\!\!/(-i\ \partial\!\!\!/X_\mu\epsilon) \\ &= \bar{\epsilon}\ \partial\!\!\!/X^\mu\ \partial\!\!\!/\psi_\mu + \overline{\psi}^\mu(\Box X_\mu)\epsilon - \overline{\psi}^\mu\rho^\alpha\rho^\beta\partial_\alpha\epsilon\partial_\beta X_\mu.\end{aligned} \tag{C.49}$$

Integrating the first term by parts leaves

$$\delta\left(-i\overline{\psi}^\mu\ \partial\!\!\!/\psi_\mu\right) = -(\partial_\alpha\bar{\epsilon})\ \partial\!\!\!/X^\mu\rho^\alpha\psi_\mu + 2\overline{\psi}^\mu(\Box X_\mu)\epsilon - \overline{\psi}^\mu\rho^\alpha\rho^\beta\partial_\alpha\epsilon\partial_\beta X_\mu. \tag{C.50}$$

Now, we use the fact that $\overline{\psi}^\mu\epsilon = \bar{\epsilon}\psi^\mu$ and also a related formula given by

$$\begin{aligned}\partial_\alpha\bar{\epsilon}\rho^\beta\rho^\alpha\psi_\mu\partial_\beta X^\mu &= \overline{\partial_\alpha\epsilon}(\rho^\beta\rho^\alpha\psi_\mu)\partial_\beta X^\mu \\ &= \overline{\rho^\beta\rho^\alpha\psi_\mu}\partial_\alpha\epsilon\partial_\beta X^\mu = (\rho^\beta\rho^\alpha\psi_\mu)^T\rho^0\partial_\alpha\epsilon\partial_\beta X^\mu \\ &= \psi_\mu^T\rho^{\alpha T}\rho^{\beta T}\rho^0\partial_\alpha\epsilon\partial_\beta X^\mu = \psi_\mu^T\rho^0\rho^0\rho^{\alpha T}\rho^0\rho^0\rho^{\beta T}\rho^0\rho^0\rho^0\partial_\alpha\epsilon\partial_\beta X^\mu \\ &= \overline{\psi_\mu}(-\rho^\alpha)(-\rho^\beta)\partial_\alpha\epsilon\partial_\beta X^\mu = \overline{\psi_\mu}\rho^\alpha\rho^\beta\partial_\alpha\epsilon\partial_\beta X^\mu\end{aligned} \tag{C.51}$$

which tells us that the first and last terms in the variation (C.50) are identical. We are left with

$$\delta\left(-i\overline{\psi}^\mu\ \partial\!\!\!/\psi_\mu\right) = -2(\partial_\alpha\bar{\epsilon})\ \partial\!\!\!/X^\mu\rho^\alpha\psi_\mu + 2\bar{\epsilon}\psi^\mu\Box X_\mu. \tag{C.52}$$

We can put the fermion and boson pieces back into the variation of the action

$$\begin{aligned}\delta S &= -\frac{1}{2\pi}\int d^2\sigma\Big(-2\bar{\epsilon}\psi^\mu\Box X_\mu - 2(\partial_\alpha\bar{\epsilon})\ \partial\!\!\!/ X^\mu\rho^\alpha\psi_\mu + 2\bar{\epsilon}\psi^\mu\Box X_\mu\Big)\\ &= \frac{1}{\pi}\int d^2\sigma\Big((\partial_\alpha\bar{\epsilon})\ \partial\!\!\!/ X^\mu\rho^\alpha\psi_\mu\Big).\end{aligned} \tag{C.53}$$

Finally,

$$\delta S = \frac{1}{\pi}\int d^2\sigma\Big((\partial_\alpha\bar{\epsilon})\rho^\beta\rho^\alpha\psi_\mu\partial_\beta X^\mu\Big) \equiv \frac{2}{\pi}\int d^2\sigma(\partial_\alpha\bar{\epsilon})J^\alpha, \tag{C.54}$$

where we have defined the supercurrent as $J^\alpha = \frac{1}{2}\rho^\beta\rho^\alpha\psi_\mu\partial_\beta X^\mu$.

We can easily see that, if ϵ is constant, then $\delta S = 0$, and otherwise, it is proportional to the supercurrent. Notice that supersymmetry transformations do not commute, and in fact, the commutator of two supersymmetry transformations is a translation of the string world sheet. Let us prove this fact for the commutator acting on first a bosonic and then a fermionic field:

$$\begin{aligned}[\delta_1,\delta_2]\,X^\mu &= \delta_1(\delta_2 X^\mu) - \delta_2(\delta_1 X^\mu) = \delta_1(\bar{\epsilon}_2\psi^\mu) - \delta_2(\bar{\epsilon}_1\psi^\mu) = \bar{\epsilon}_2(\delta_1\psi^\mu) - \bar{\epsilon}_1(\delta_2\psi^\mu)\\ &= \bar{\epsilon}_2(-i\rho^\alpha\partial_\alpha X^\mu\epsilon_1) - \bar{\epsilon}_1(-i\rho^\alpha\partial_\alpha X^\mu\epsilon_2) = i\bar{\epsilon}_1\rho^\alpha\partial_\alpha X^\mu\epsilon_2 + i\bar{\epsilon}_1\rho^\alpha\partial_\alpha X^\mu\epsilon_2\\ &= 2i\bar{\epsilon}_1\rho^\alpha\epsilon_2\partial_\alpha X^\mu = a^\alpha\partial_\alpha X^\mu\end{aligned} \tag{C.55}$$

and for a fermion

$$\begin{aligned}[\delta_1,\delta_2]\,\psi^\mu_A &= \delta_1(\delta_2\psi^\mu_A) - \delta_2(\delta_1\psi^\mu_A)\\ &= \delta_1(-i\rho^\alpha_{AB}\partial_\alpha X^\mu\epsilon_{2B}) - \delta_2(-i\rho^\alpha_{AB}\partial_\alpha X^\mu\epsilon_{1B})\\ &= -i\rho^\alpha_{AB}\partial_\alpha(\delta_1 X^\mu)\epsilon_{2B} + i\rho^\alpha_{AB}\partial_\alpha(\delta_2 X^\mu)\epsilon_{1B}\\ &= -i\rho^\alpha_{AB}\partial_\alpha(\bar{\epsilon}_{1C}\psi^\mu_C)\epsilon_{2B} + i\rho^\alpha_{AB}\partial_\alpha(\bar{\epsilon}_{2C}\psi^\mu_C)\epsilon_{1B}\\ &= i\rho^\alpha_{AB}\left[\epsilon_{1B}\bar{\epsilon}_{2C} - \epsilon_{2B}\bar{\epsilon}_{1C}\right]\partial_\alpha\psi^\mu_C\\ &\equiv i\rho^\alpha_{AB}\Lambda_{BC}\partial_\alpha\psi^\mu_C.\end{aligned} \tag{C.56}$$

Expanding Λ_{BC} in explicit components gives

$$\Lambda_{BC} = \begin{pmatrix} i[\epsilon_{10}\epsilon_{21} - \epsilon_{20}\epsilon_{11}] & -2i\epsilon_{10}\epsilon_{20}\\ 2i\epsilon_{11}\epsilon_{21} & i[\epsilon_{10}\epsilon_{21} - \epsilon_{20}\epsilon_{11}]\end{pmatrix} \equiv \begin{pmatrix} \alpha & b_1\\ b_2 & -\alpha\end{pmatrix}, \tag{C.57}$$

so we have (after a bit of work)

$$[\delta_1,\delta_2]\,\psi = i\begin{pmatrix} ib_2(\partial_1 - \partial_0)\psi_0 - i\alpha(\partial_1 - \partial_0)\psi_1\\ i\alpha(\partial_1 + \partial_0)\psi_0 + ib_1(\partial_1 + \partial_0)\psi_1\end{pmatrix}. \tag{C.58}$$

A bit more work gives

$$2i\bar{\epsilon}_1\ \partial\!\!\!/\epsilon_2\psi = 2i\begin{pmatrix} -\epsilon_{11}\epsilon_{21}(\partial_1 - \partial_0)\psi_0 + \epsilon_{10}\epsilon_{20}(\partial_1 + \partial_0)\psi_0\\ -\epsilon_{11}\epsilon_{21}(\partial_1 - \partial_0)\psi_1 + \epsilon_{10}\epsilon_{20}(\partial_1 + \partial_0)\psi_1\end{pmatrix}. \tag{C.59}$$

Substituting back in for b_1 and b_2 in equation (C.58) and using the Dirac equation,

$$(\partial_1 + \partial_0)\psi_0 = (\partial_1 - \partial_0)\psi_1 = 0. \tag{C.60}$$

We can easily see that equations (C.58) and (C.59) are the same but notice that they are *not* the same without using the equations of motion.

This problem is made simpler by an identity derived as follows.

$$\begin{aligned}(\rho^\alpha\epsilon_1)_A(\bar{\epsilon}_2\partial_\alpha\psi) &= \begin{bmatrix} \epsilon_{11}\epsilon_{21}(\partial_0-\partial_1)\psi_0 - \epsilon_{11}\epsilon_{20}(\partial_0-\partial_1)\psi_1 \\ \epsilon_{10}\epsilon_{20}(\partial_0+\partial_1)\psi_1 - \epsilon_{10}\epsilon_{21}(\partial_0+\partial_1)\psi_0 \end{bmatrix} \\ (\rho^\alpha\epsilon_2)_A(\bar{\epsilon}_1\partial_\alpha\psi) &= \left[\ \epsilon_1 \leftrightarrow \epsilon_2\ \right],\end{aligned} \tag{C.61}$$

and if we define the chirality matrix, $\rho^2 = \rho^0\rho^1$, with property $\rho^\alpha\rho^2 = -\rho^2\rho^\alpha$, we have

$$\begin{aligned}(\rho^\alpha\rho^2\partial_\alpha\psi) &= i\begin{pmatrix}(\partial_0-\partial_1)\psi_1 \\ (\partial_0+\partial_1)\psi_0\end{pmatrix} \\ \bar{\epsilon}_1\rho^2\epsilon_2 &= i(\epsilon_{10}\epsilon_{21}+\epsilon_{11}\epsilon_{20})\end{aligned} \tag{C.62}$$

so that

$$i\bar{\epsilon}_1\rho^2\epsilon_2\left(\rho^\alpha\rho^2\partial_\alpha\psi\right) = -i\begin{pmatrix}\epsilon_{10}\epsilon_{21}+\epsilon_{11}\epsilon_{20}(\partial_0-\partial_1)\psi_1 \\ \epsilon_{10}\epsilon_{21}+\epsilon_{11}\epsilon_{20}(\partial_0+\partial_1)\psi_0\end{pmatrix}. \tag{C.63}$$

We also have

$$\left(\rho^\alpha\rho^0\partial_\alpha\psi\right) = i\begin{pmatrix}(\partial_0-\partial_1)\psi_0 \\ (\partial_0+\partial_1)\psi_1\end{pmatrix} \text{ and } \left(\rho^\alpha\rho^1\partial_\alpha\psi\right) = i\begin{pmatrix}(\partial_0-\partial_1)\psi_0 \\ -(\partial_0+\partial_1)\psi_1\end{pmatrix}, \tag{C.64}$$

which gives

$$\left(\bar{\epsilon}_1\rho_\beta\epsilon_2\right)\left(\rho^\alpha\rho^\beta\partial_\alpha\psi\right) = -2\begin{pmatrix}\epsilon_{11}\epsilon_{21}(\partial_0-\partial_1)\psi_0 \\ \epsilon_{10}\epsilon_{20}(\partial_0+\partial_1)\psi_1\end{pmatrix}, \tag{C.65}$$

thus we see that equations (C.63) and (C.65) combine to give

$$\begin{aligned}[\delta_1,\delta_2]\,\psi^\mu_A &= i\left[(\rho^\alpha\epsilon_1)(\bar{\epsilon}_2\partial_\alpha\psi^\mu) - (\rho^\alpha\epsilon_2)(\bar{\epsilon}_1\partial_\alpha\psi^\mu)\right] \\ &= i(\bar{\epsilon}_1\rho^2\epsilon_2)(\rho^\alpha\rho^2\partial_\alpha\psi^\mu) - i(\bar{\epsilon}_1\rho_\beta\epsilon_2)(\rho^\alpha\rho^\beta\partial_\alpha\psi^\mu).\end{aligned} \tag{C.66}$$

Using the Clifford identity (C.44), we get

$$\begin{aligned}[\delta_1,\delta_2]\,\psi^\mu_A &= -i(\bar{\epsilon}_1\rho^2\epsilon_2)(\rho^2\ \partial\!\!\!/\psi^\mu) + i(\bar{\epsilon}_1\rho_\beta\epsilon_2)(\rho^\beta\ \partial\!\!\!/\psi^\mu) + 2i(\bar{\epsilon}_1\rho^\beta\epsilon_2)(\partial_\beta\psi^\mu) \\ &= 2i(\bar{\epsilon}_1\rho^\beta\epsilon_2)(\partial_\beta\psi^\mu) = 2ia^\beta\partial_\beta\psi^\mu\end{aligned} \tag{C.67}$$

by the equations of motion, $\partial\!\!\!/\psi^\mu = 0$.

We have, therefore, shown that, for both fermions and bosons, the commutator of two supersymmetry transformations is a translation of the world sheet. Hence, the supersymmetry algebra closes on-shell.

C.2.2. Superspace

In superspace, the supersymmetry algebra actually closes without the use of the Dirac equation. We supplement the world sheet coordinate, σ^α, by Grassman coordinates, θ^A, which form a 2-component Majorana spinor. A general function, Y^μ, in superspace is then

$$Y^\mu = X^\mu + \bar{\theta}\psi^\mu + \frac{1}{2}\bar{\theta}\theta B^\mu, \tag{C.68}$$

where B^μ is an auxilliary field. We call this generalized function a superfield.

Supersymmetry is represented on superspace by the generator

$$Q_A = \frac{\partial}{\partial\bar{\theta}^A} + i\left(\rho^\alpha\theta\right)_A \partial_\alpha. \tag{C.69}$$

If we introduce an arbitrary anticommuting parameter, ϵ_A, then we have the transformation of the superspace coordinates as

$$\begin{aligned} \delta\theta^A &= \left[\bar{\epsilon}Q, \theta^A\right] = \epsilon^A \\ \delta\sigma^\alpha &= \left[\bar{\epsilon}Q, \sigma^\alpha\right] = i\bar{\epsilon}\rho^\alpha\theta, \end{aligned} \tag{C.70}$$

and the transformation of the superfield is given by

$$\delta Y^\mu = \left[\bar{\epsilon}Q, Y^\mu\right] = \bar{\epsilon}QY^\mu. \tag{C.71}$$

Since

$$\left[\bar{\epsilon}_1 Q, \bar{\epsilon}_2 Q\right] = -2i\bar{\epsilon}_1\rho^\alpha\epsilon_2\partial_\alpha,$$

we have

$$\left[\delta_1, \delta_2\right] Y^\mu = -a^\alpha\partial_\alpha Y^\mu \tag{C.72}$$

without using the equations of motion.

We want to write δY^μ in components so that we can pick out the variations of the component fields and show that they are equivalent to our previous results.

$$\begin{aligned} \delta Y^\mu &= \bar{\epsilon}QY^\mu = \bar{\epsilon}\left(\frac{\partial}{\partial\bar{\theta}} + i\rho^\alpha\theta\partial_\alpha\right) Y^\mu \\ &= \bar{\epsilon}\frac{\partial Y^\mu}{\partial\bar{\theta}} + i\bar{\epsilon}\rho^\alpha\theta\partial_\alpha Y^\mu = \bar{\epsilon}_A\left[\psi^\mu_A + \theta_A B^\mu\right] + i\bar{\epsilon}_A\rho^\alpha_{AB}\theta_B\partial_\alpha\left[X^\mu + \bar{\theta}\psi^\mu\right] \\ &= \bar{\epsilon}\psi^\mu + \bar{\epsilon}_A\theta_A B^\mu + i\bar{\epsilon}_A\rho^\alpha_{AB}\theta_B\partial_\alpha X^\mu + i\bar{\epsilon}_A\rho^\alpha_{AB}\theta_B\partial_\alpha\bar{\theta}_C\psi^\mu_C \\ &= \bar{\epsilon}\psi^\mu + \bar{\theta}_A\epsilon_A B^\mu + i\overline{\rho^\alpha_{AB}\theta_B}\epsilon_A\partial_\alpha X^\mu + i\bar{\epsilon}_A\rho^\alpha_{AB}\theta_B\bar{\theta}_C\partial_\alpha\psi^\mu_C \\ &= \bar{\epsilon}\psi^\mu + \bar{\theta}\epsilon B^\mu - i\bar{\theta}_A\rho^\alpha_{AB}\epsilon_B\partial_\alpha X^\mu + i\bar{\epsilon}_A\rho^\alpha_{AB}\left(-\frac{1}{2}\delta_{BC}\bar{\theta}_D\theta_D\right)\partial_\alpha\psi^\mu_C, \end{aligned} \tag{C.73}$$

where we used the Fierz relation, $\theta_A \overline{\theta}_B = -\frac{1}{2}\delta_{AB}\overline{\theta}_C\theta_C$, in the last term.

$$\begin{aligned}\delta Y^\mu &= \overline{\epsilon}\psi^\mu + \overline{\theta}\epsilon B^\mu - i\overline{\theta}_A\rho^\alpha_{AB}\epsilon_B\partial_\alpha X^\mu - \overline{\theta}_D\theta_D\frac{i}{2}\overline{\epsilon}_A\rho^\alpha_{AC}\partial_\alpha\psi^\mu_C \\ &= \overline{\epsilon}\psi^\mu + \overline{\theta}\left(\epsilon B^\mu - i\rho^\alpha\epsilon\partial_\alpha X^\mu\right) + \frac{1}{2}\overline{\theta}\theta\left(-i\overline{\epsilon}\rho^\alpha\partial_\alpha\psi^\mu\right)\end{aligned} \tag{C.74}$$

Identifying each of the coefficients as

$$\delta Y^\mu = \delta X^\mu + \overline{\theta}\delta\psi^\mu + \frac{1}{2}\overline{\theta}\theta\delta B^\mu$$

gives the following variations

$$\begin{aligned}\delta X^\mu &= \overline{\epsilon}\psi^\mu \\ \delta\psi^\mu &= \epsilon B^\mu - i\rho^\alpha\epsilon\partial_\alpha X^\mu \\ \delta B^\mu &= -i\overline{\epsilon}\rho^\alpha\partial_\alpha\psi^\mu.\end{aligned} \tag{C.75}$$

We recognize that these variations reduce to the supersymmetry transformations defined earlier (C.45) if we set the auxilliary field, B^μ, to zero (which, as we will soon see, is essentially putting B^μ on-shell since the equations of motion are $B^\mu = 0$).

The product of two superfields is again a superfield since $\delta = \overline{\epsilon}Q$ is a first-order differential operator, thus

$$\delta(Y_1Y_2) = (\delta Y_1)Y_2 + Y_1(\delta Y_2).$$

Now, introduce the superspace covariant derivative,

$$D = \frac{\partial}{\partial\overline{\theta}} - i\rho^\alpha\theta\partial_\alpha, \tag{C.76}$$

having the properties

$$\begin{aligned}\{D_A, Q_B\} &= 0 \\ \{D_A, \overline{D}_B\} &= 2i(\rho^\alpha)_A B\partial_\alpha \\ \{D_A, D_B\} &= 2i(\rho^\alpha\rho^\beta)_A B\partial_\alpha.\end{aligned} \tag{C.77}$$

The covariant derivative of a superfield is again a superfield.

We define the superspace integration measure, $\int d^2\sigma d^2\theta$, using the Berezin integral as follows:

$$\int d^2\theta(a + \theta^1 b_1 + \theta^2 b_2 + \theta^1\theta^2 c) = c. \tag{C.78}$$

Since $\overline{\theta}\theta = \theta\rho^0\theta = -2i\theta^1\theta^2$, we have

$$\int d^2\theta\overline{\theta}\theta = -2i.$$

Integration by parts is valid due to $\int d^2\theta \partial V/\partial\theta^A = 0$ since $\partial V/\partial\theta^A$ has no $\theta^1\theta^2$ term.

An action can be constructed out of an arbitrary product of superfields and their covariant derivatives. Such an action will automatically be invariant under supersymmetry transformations, $\delta S = 0$. This fact can be easily be seen by noting that the arbitrary product will again be a superfield due to the properties of the product and of the covariant derivative. Hence, we need only prove it for a single superfield, Y, with $S = \int d^2\sigma d^2\theta Y$:

$$\begin{aligned}
\delta S &= \int d^2\sigma d^2\theta (\delta Y) = \int d^2\sigma d^2\theta (\bar{\epsilon} Q Y) \\
&= \int d^2\sigma d^2\theta \bar{\epsilon} \left(\psi + [B + i \not{\partial} X]\theta - \frac{i}{2}\bar{\theta}\theta \not{\partial}\psi \right) \\
&= \int d^2\sigma \bar{\epsilon} \frac{i}{2}(2i) \not{\partial}\psi = -\int d^2\sigma \bar{\epsilon} \not{\partial}\psi \\
&= \int d^2\sigma (\partial_\alpha \bar{\epsilon})\rho^\alpha \psi = 0
\end{aligned} \tag{C.79}$$

for constant ϵ.

Now, construct the following action:

$$S = \frac{i}{4\pi} \int d^2\sigma d^2\theta \overline{D} Y^\mu D Y_\mu . \tag{C.80}$$

Working out the derivatives gives

$$\begin{aligned}
DY_\mu &= \left(\frac{\partial}{\partial\bar{\theta}} - i\rho^\alpha \theta \partial_\alpha \right) \left(X_\mu + \bar{\theta}\psi_\mu + \frac{1}{2}\bar{\theta}\theta B_\mu \right) \\
&= \psi_\mu + \theta B_\mu - i\rho^\alpha \theta \partial_\alpha X_\mu - i\rho^\alpha \theta \bar{\theta} \partial_\alpha \psi_\mu \\
&= \psi_\mu + \theta B_\mu - i\theta \not{\partial} X_\mu + \frac{i}{2}\bar{\theta}\theta \not{\partial}\psi_\mu ,
\end{aligned} \tag{C.81}$$

where we used the Fierz identity in the last step. The expression for $\overline{D}Y$ is easily computed by conjugation

$$\overline{D} Y^\mu = \bar{\psi}^\mu + \bar{\theta} B^\mu + i\bar{\theta} \not{\partial} X^\mu - \frac{i}{2}\bar{\theta}\theta \partial_\alpha \bar{\psi}^\mu \rho^\alpha . \tag{C.82}$$

We write down the action, noting that only the terms in $\bar{\theta}\theta$ will survive.

$$\begin{aligned}
S &= \frac{i}{4\pi} \int d^2\sigma d^2\theta \left(\frac{i}{2}(\bar{\psi}^\mu \not{\partial}\psi_\mu - \partial_\alpha \bar{\psi}^\mu \rho^\alpha \psi_\mu) + B^\mu B_\mu \right) \bar{\theta}\theta + \left(\frac{1}{2}\bar{\theta}\{\rho^\alpha, \rho^\beta\}\partial_\alpha X^\mu \partial_\beta X_\mu \theta \right) \\
&= \frac{i}{4\pi} \int d^2\sigma d^2\theta \left(i\bar{\psi}^\mu \not{\partial}\psi_\mu + B^\mu B_\mu - \partial_\alpha X^\mu \partial^\alpha X_\mu \right) \bar{\theta}\theta
\end{aligned} \tag{C.83}$$

which is

$$S = \frac{1}{2\pi} \int d^2\sigma \left(i\overline{\psi}^{\mu} \not{\partial} \psi_{\mu} + B^{\mu} B_{\mu} - \partial_{\alpha} X^{\mu} \partial^{\alpha} X_{\mu} \right)$$
$$= -\frac{1}{2\pi} \int d^2\sigma \left(\partial_{\alpha} X^{\mu} \partial^{\alpha} X_{\mu} - i\overline{\psi}^{\mu} \not{\partial} \psi_{\mu} - B^{\mu} B_{\mu} \right). \quad \text{(C.84)}$$

The field equations say $B^{\mu} = 0$, thus putting this auxilliary field on-shell gives us the free action (C.41). Since this action is given by the product of two superfields, which is again a superfield, it is automatically invariant under supersymmetry transformations. This time, we do not even need to use the Dirac equation as we did previously. Hence, a fictitious auxilliary field and superspace allow us to close the supersymmetry algebra without the field equations.

C.3. Bott periodicity and the M-theory spectra

In this section, we want to outline the periodicity in the spectrum of M-theory. Readers may find it useful to consult Appendix D if they find some of the topological material unfamiliar.

The spectrum follows from supersymmetry and the algebra of spinors.

$$32_s \otimes 32_s = [1] \oplus [2] \oplus [5],$$

where $[n]$ is the n-fold antisymmetric tensor. The latter two terms are exactly the M2 and M5 brane charge.

The string theory solitons can be seen as the physical manifestation of $SO(32)$ periodicity. Since massless Ramond-Ramond fields are differential forms, we think of the nontrivial homotopy classes as representing the Ramond-Ramond charges. Table C.3. gives string theory interpretations for the non-trivial groups.

The table stops at $q = 9$ for ten dimensions. We only look at the $U(N)$ and $SO(N)$ columns in determining the spectrum. The $Sp(N)$ column is given for completeness and has to do with stacked type I branes. The $U(N)$ comes from a stack of N type II branes, and the $SO(N)$ is actually $SO(32)$ for the type I case.

The corresponding branes in the table are found by noticing that $\pi_q(SO(32))$ classifies $SO(32)$ bundles on S^{q+1} which is equivalent to $\mathbb{R}^{q+1}$. For example, $\pi_7 = Z$ classifies a $SO(32)$ bundle on $\mathbb{R}^8$ in ten spacetime dimensions. The codimension is then 2, which is the dimension of a string world sheet, so we write '1-brane' in the corresponding type I spot. The type IIA/B spots are found by using the fact that the non-trivial groups correspond to the q-form charge on branes.

Notice that, since the brane world volumes of type IIA have odd codimension, we actually use the eleven-dimensional spacetime, $S^1 \times M$, to compute the spectrum rather than just the ten-dimensional M as with type IIB. Consequently, we see the effect of the $U(N)$ charge shifted by 1 from that of type IIB. The 9-brane does not appear. I suppose it is because it is not a dynamical object. It is used, however, in the KO-theory construction of this table in the form of stacked annihilating 9-anti-9 brane pairs.

Table C.1. Periodic table of the branes. $\pi_q(x)$ for $N+1 > (q+2)/d$ The right-hand side shows the manifestation of the nontrivial $U(N)$ and $SO(N)$ homotopy groups in the string theory brane spectrum.

q	$U(1)$ ($d=2$)	SO(N) ($d=1$)	Sp(N) ($d=4$)	Type IIA	Type IIB	Type I
0	0	$\mathbb{Z}_2$	0			8-brane
1	$\mathbb{Z}$	$\mathbb{Z}_2$	0	8-brane	7-brane	7-brane
2	0	0	0			
3	$\mathbb{Z}$	$\mathbb{Z}$	$\mathbb{Z}$	6-brane	5-brane	5-brane
4	0	0	$\mathbb{Z}_2$			
5	$\mathbb{Z}$	0	$\mathbb{Z}_2$	4-brane	3-brane	
6	0	0	0			
7	$\mathbb{Z}$	$\mathbb{Z}$	$\mathbb{Z}$	2-brane	1-brane	1-brane
8	0	$\mathbb{Z}_2$	0			0-brane
9	$\mathbb{Z}$	$\mathbb{Z}_2$	0	0-brane	(-1)-brane	(-1)-brane
10	0	0	0			
11	$\mathbb{Z}$	$\mathbb{Z}$	$\mathbb{Z}$	F-theory?	F-theory?	F-theory?

Perhaps it is not surprising that the 9-brane does not "predict its own existence."

The Bott periodicity looks like a mapping between branes of differing dimension. It is tempting to hypothesize that branes in the same periodic equivalence class have properties in common. In type II, one would never know since there is only one class (or cycle), periodic with period 2. In type I, however, there is a period of 8, so, for example, the D8-brane and the D-particle, or the D7-brane and the D-instanton may have properties in common. Perhaps "Bott-dual" pairs -1, 7 and 0, 8 will pairwise cancel each other in some way due to this periodicity, leaving only the 1-brane and 5-brane of the type I theory. Then the type I spectrum above, modulo the periodicity, gives the actual type I brane spectrum. Indeed, it turns out that the D7 and D8 in type I are still debated, and may be unstable. It may be that the octonions may be relevant to the type IIA 2-brane, and the type IIB and type I D-strings.

The comment about F-theory is motivated by the following argument. We arrived at the type IIA spectrum by looking at eleven dimensional spacetime compactified on a circle. Perhaps twelve extended dimensions may give F-theory. The result is the branes in the table shifting down two notches and a 9-brane appearing at the top. Perhaps F-theory, compactified on a torus, gives type IIB the same way that M-theory, compactified on a circle, gives type IIA. Perhaps M-theory, compactified on a circle, with a T-duality [83] is the same as F-theory, compactified on a torus. The modular parameter of the torus would somehow correspond to scalar fields in type IIB. In fact, F-theory may possibly be a meta-theory which reduces to M-theory if one compactifies on a circle and type IIB if one compactifies on a torus. This situation would be nice, but we had better not speculate any further.

APPENDIX D
INTRODUCTION TO HOMOTOPY THEORY

The fundamental group is the group of equivalence classes of loops starting and ending at a fixed point in space. The equivalence relation that used to define these classes is called *homotopy.* We begin this appendix by showing how a large number of mathematical and physical concepts are unified by homology, and cohomology. We then continue with an exposition of the fundamental group and covering spaces. We have made extensive use of references on topology [84], algebraic topology [85], group theory [86, 87], and algebra [88, 89].

D.1. Introduction

It is exciting to see well-known physical laws re-cast with the precision and elegance of algebraic topology. Often, one sees that seemingly disparate concepts (or at least only vaguely related) turn out to be merely specific cases of an all-encompassing mathematical principle. The language of algebraic topology has become very common in high energy physics theory, so it is important to be familiar with it. The concepts we will discuss first are homotopy, homology and cohomology.

The first homology group of a space, U, is the the set of closed 1-chains (curves), Z, modulo the closed 1-chains, $B \subset Z$, which are also boundaries. A closed 1-chain is a boundary if it is homotopic (can be continuously deformed) to a point so that removing a boundary will separate U into disconnected components. This group is denoted $H_1(U) = Z_1(U)/B_1(U)$, where Z_1 are the cycles, or closed 1-chains, and B_1 are the 1-boundaries.

For example, on a torus, there are three homology classes of curves: ones that are boundaries, ones that go around the outside of the hole, and ones that go through the hole. The first homology group of the torus is, therefore, isomorphic to the free abelian group with two generators. This group consists of two-dimensional vectors with integer entries. The integer could be the number of times you go around each curve or merely a weight attached to each curve. The curve which does not encircle a hole is a boundary and is, therefore, equivalent to the zero element.

The boundary of a boundary is zero. For example, the boundary of a disk is a circle. The circle has no boundary. A line, on the other hand, has a boundary consisting of two points.

Not all curves without a boundary are themselves boundaries. If ∂ is the boundary operator, there are curves, γ, for which $\partial\gamma = 0$ but which are not boundaries themselves; i.e., there is no σ such that $\partial\sigma = \gamma$. To see an example of this, we examine the three curves mentioned for the torus. The curve that is contractible to a point is a boundary. The boundary of this curve is zero as required. The other two curves also have no boundary but are *not* themselves bounding any region of the torus. They do not split the torus into distinct connected components. If you removed either of these curves from the torus, you would be left with a surface which is topologically equivalent to a cylinder. These two curves are exactly the ones that make the first homology group of the torus non-trivial.

There is a duality between 1-forms and 1-chains, and all of the results above have analogues for 1-forms (1-forms are integrated over 1-chains to give real numbers, i.e., an inner product). The connectedness of the space we are integrating over defines whether the 1-chains themselves are boundaries.

A 1-form ω, (e.g., $f(x)dx$) is called *closed* if $d\omega = 0$, where d is the exterior derivative familiar from differential geometry. This definition is analogous to the fact that a 1-chain is closed (a cycle) if its boundary is zero ($\partial\gamma = 0$).

A 1-form ,ω, is called *exact* if $\omega = df$ for some 0-form f. This definition is analogous to the case of a 1-chain being a boundary. A 0-form is a continuous function; a 1-form is $f(x)dx$; a 2-form is $f(x,y)dx \wedge dy$; etc. The "coboundary" operator, d, maps from 0-forms $\to$ 1-forms $\to$ 2-forms..., whereas the boundary operator, ∂, maps the other direction in chains $n \to n-1 \to \cdots$.

Applying the exterior derivative twice is zero, $d^2F = 0$. This fact is equivalent to the curl of a gradient being zero or that the boundary of a boundary is zero.

The first deRham cohomology group, analogous to the first homology group, is the set of closed 1-forms modulo the set of exact 1-forms. The group is denoted

$$H^1(U) = Z^1(U)/B^1(U).$$

The group is trivial (consisting of only the zero element) if all closed 1-forms are exact. If, as we mentioned before, the space that we are dealing with is multiply connected, then there are closed 1-chains that are not themselves boundaries. Likewise, there are closed 1-forms that are not themselves exact.

What does this mean physically? It means that just because $\nabla \times f = 0$ does not mean that $f = \nabla g$ for some g. Just because $\nabla B = 0$ does not mean that there is a potential function A such that $B = \nabla \times A$. Whenever we have a region in our space that the field cannot penetrate (e.g., an infinite solenoid in the case of the Aharonov-Bohm effect), the space is multiply connected. We cannot, therefore, expect a global potential function to exist.

Green's Theorem, Stokes Theorem, and the fundamental theorem of calculus are all special cases of the generalized Stokes theorem, which can be stated as

$$\int_\gamma d\omega = \int_{\partial\gamma} \omega.$$

If any form, A, can be written as $d\omega$ for some ω (i.e., it is exact), then $\int_{\partial\gamma} A = 0$. If $f = dA$ for some A, then the integral only depends on the values of A on the boundary. In particular, if f is a 1-form, then A is a 0-form (i.e., a function), and γ is a 1-chain. $\partial\gamma$ is a set of two points. The integral then depends only on the value of F at those two points (the endpoints). This statement is the fundamental theorem of calculus. The other theorems arise simply by considering higher-dimensional n-forms and n-chains.

Let us look at an application of these ideas to electromagnetism. We can add an exact 1-form to the vector potential without changing the electric and magnetic fields, E and B So E and B depend only on the deRham cohomology equivalence class of A and not the actual A. The curvature is unaffected by an exact addition to the connection. Also, if $F = dA$ (i.e., The curvature is exact.), the homogeneous Maxwell's equations, $dF = 0$, become trivial. The inhomogeneous Maxwell's equations come from $d * F = J$, where the $*$ denotes Hodge dual.

If $A = f dx$ is a 1-form in $\mathbb{R}^3$, then

$$
\begin{aligned}
dA &= \partial_x A dx \wedge dx + \partial_y A dy \wedge dx = -\partial_y A dx \wedge dy \\
d * A &= d\left(A \frac{1}{3-1!}\epsilon_{123} dy \wedge dz + \epsilon_{132} dz \wedge dy\right) \\
&= dA dy \wedge dz, = \partial_x A dx \wedge dy \wedge dz.
\end{aligned}
$$

The adjoint of d is $\delta = - * d *$ and $\delta\delta\omega = 0$, and the laplacian on a manifold is $(d+\delta)^2$.

Definition: The *winding number*, W, of a curve, γ, about a point, p, is defined as

$$W(\gamma, p) = \frac{1}{2\pi}\int_\gamma \omega_{p,\theta}, \tag{D.1}$$

where

$$\omega_{p,\theta} = \frac{-(y-y_0)dx + (x-x_0)dy}{(x-x_0)^2 + (y-y_0)^2},$$

and $\gamma : [a,b] \to \mathbb{R}^2 \setminus p$ is a map from the closed interval $[a,b]$ to the complement of a point in the plane. Basically, $\gamma(t)$ is a curve in the plane minus p (doubly connected) with endpoints at $\gamma(a)$ and $\gamma(b)$.

The winding number counts the angle that a ray drawn from point p to the curve traces out as one traverses the curve. If the curve is closed, the winding number is an integer, being the number of times one has gone around the point. Notice that our curve could also be the orbits or trajectories of a vector field on the plane, and in this case, p would simply be a singular point (a zero) of that vector field.

Now, if we write the "angle function," $\omega_{p,\theta}$, in the complex plane instead of $\mathbb{R}^2$ by the identification $(x,y) \to (x+iy)$, we see that the angle function becomes $\omega \to \frac{1}{z}dz$. The winding number is, of course, still an integer. By the definition of the winding number, what do we get when we integrate this angle function around a curve surrounding p?

$$\int_\gamma \omega = 2\pi i W$$

The i comes from $dz = dx + idy$ which, for a curve of radius 1 in polar coordinates, is $dz = d\left(re^{i\theta}\right) = ie^{i\theta}d\theta$. If you only go around once, then the winding number is exactly the residue of the function $\frac{1}{z}$ since the coefficient of $\frac{1}{z}$ in the Laurent expansion of $\frac{1}{z}$ is just 1 here.

Cauchy Integral Theorem: If γ is a closed chain in U with winding number around any point not in U being zero and if f is an analytic function in U, then for any a in U that is not in the support of γ (closure of the region where γ is not zero).

$$W(\gamma, a)f(a) = \frac{1}{2\pi i}\int_{\gamma}\frac{f(z)}{z-a}dz.$$

The physicist will notice that the Dirac delta function can be defined in the complex plane as $\delta(z-a) = \frac{1}{2\pi i}\frac{1}{z-a}$.

Residue Theorem: If f is analytic in $U - a_1, ..., a_n$ and γ is a closed 1-chain such that $W(\gamma, P) = 0$ for all P not in U, then

$$\frac{1}{2\pi i}\int_{\gamma} f(z)dz = \sum_{i=1}^{r} W(\gamma, a_i)Res_{a_i}(f)$$

with $f(z)dz$ a 1-form. Winding number, which is deeply related to the fact that closed curves cannot be contracted to a point, leads to the residue theorem.

The Euler characteristic of a surface tells us (among other things) the number of sources, sinks, and saddle points of a vector field on a surface (the "index"). It is basically an alternating sum, $\sum_{i=o}^{N}(-1)^i C_i$, given by a triangulation (closed graph) drawn on the (N-dimensional) surface. C_i is the number of i-chains on the graph (the ith Betti numbers). For example, on a disk or plane, $\mathbb{R}^2$, we have any graph containing vertices, edges, and faces. The Euler characteristic is then $\#V - \#E + \#F$ since vertices are 0-chains, edges are 1-chains, and the faces are 2-chains. An important point is that any graph in the space will have exactly the same Euler characteristic! We need only compute it once. Why is this point useful? Let us list a few examples:

- There is a direct relationship between graphs and vector fields:

  Vertices = sources, zeros

  Edges = saddles

  Faces = sinks

  Any vector field on a disk will have index 1 since we can find a single vector field that does, and the index, or Euler characteristic, is an invariant. The field we are thinking of is the one that consists of increasing circles about the origin with one singular point (zero).

- On a sphere with g handles $v - e + f = 2 - 2g$, so a torus being a sphere with 1 handle has $v - e + f = 0$. You can "comb the hair" on a torus; i.e., there are vector fields on the torus with no zeros. Any planet (sphere) will have # Mountains − # passes + # valleys = 2 since any vector field on a sphere has two zeros (can't comb the hair on a sphere!).

- In classical mechanics, the phase space of a system of particles is the cotangent bundle, and when the energy (or any function of the cotangent vectors) of the system has periodic boundary conditions, the cotangent space can then be considered as the universal covering space of the space of observables. For example, in a system consisting of a particle in a box, with coordinates (q, p) in the cotangent bundle, the energy of the system is a function of the coordinates (q', p') given by the projection of the cotangent bundle onto its base space which is the torus. Since it is a function on the torus, it can be expanded in terms of the basis of 1-forms on the torus, which are the two equivalence classes of the deRham cohomology group on the torus. These classes can be represented by $e^{2\pi imp}dp$ and $e^{2\pi inq}dq$ for coordinates $0 \leq p < 1$ and $0 \leq q < 1$ on the torus (Fourier modes). These modes are closed 1-forms which are not exact. They are not global coordinates since they are only defined modulo 1 and they form a basis for the first deRham cohomology group.

- In Dynamical Systems (chaos theory), attractors are sinks, etc...

The list of applications continues.

D.2. Groups

A *path* in a topological space, X, is defined as a continuous map, $\gamma : [0,1] \to X$, which runs from a point, $x = \gamma(0)$, to another point, $x' = \gamma(1)$. Saying two paths are homotopic is the same as saying there is a continuous deformation, called a homotopy, deforming one into the other. A homotopy is a function of two variables, $H(t, s)$, where the first variable t denotes the position along path $\gamma_s(t)$ and the second variable denotes which of a continuous selection of paths we are following. The value $s = 0$ denotes the original path while $s = 1$ denotes the final path homotopic to it.

The product of two paths is given by first traversing the first path and then the second. For example, if $\sigma(t)$ is a path from $x = \sigma(0)$ to $x' = \sigma(1)$ and $\tau(t)$ is a path from $x' = \tau(0)$ to $x'' = \tau(1)$, then the product path, $(\sigma \cdot \tau)(t)$, is a new path, parametrized by $t \in [0,1]$ from $x = (\sigma \cdot \tau)(0)$ to $x'' = (\sigma \cdot \tau)(1)$ wherein $(\sigma \cdot \tau)(\frac{1}{2}) = x'$.

The inverse of a path formed by traversing the path in the opposite direction, which amounts to reparametrizing the original path by $t \to (1 - t)$ so that $\sigma^{-1}(t) = \sigma(1 - t)$.

The identity is a constant path: we stay at one point without moving. We denote this path by $\epsilon_x(t) = x, \quad 0 \leq t \leq 1$.

Finally, there is a notion of equivalence given by homotopy. We say that two paths are equal if there is a homotopy between them. We denote a path as an equivalence class, $[\sigma]$, rather than σ since any path that is homotopic to σ is equivalent.

We have a product, an inverse, an identity, and an equivalence relation. We can, therefore, form a group. To form a group, one begins with an *algebraic structure*. An algebraic structure with one composition[1] is a set (in our case a set of *loops*) with a composition, allowing us to combine two elements of the set to get a third. In our present case of interest, the composition is the product of paths as defined previously.

We will define our set as, not the set of paths between two points, but the set of *closed* paths, or loops, starting and ending at a point, x, in our space, X. This set, along with our composition, gives an algebraic structure. If our composition is associative, we have a *semigroup*, and if the composition is also commutative, we have

[1] An algebraic structure with more than one composition leads to a *ring* rather than a group.

an abelian semigroup. To prove that our product is associative, we need only show that, for three loops (or more precisely, the equivalence classes of three loops via homotopy equivalence), $[\sigma]$, $[\tau]$, and $[\mu]$, we have $([\sigma]\cdot[\tau])\cdot[\mu]=[\sigma]\cdot([\tau]\cdot[\mu])$. This equivalence is proven by finding a homotopy between the left-hand and right-hand sides of the equal sign. One such homotopy is seen to be

$$H(t,s)=\begin{cases}\sigma(\frac{4t}{1+s}) & \text{if } 0\leq t\leq \frac{1}{4}(1+s)\\ \tau(4t-s-1) & \text{if } \frac{1}{4}(1+s)\leq t\leq \frac{1}{4}(2+s)\\ \mu(\frac{4t-2-s}{2-s}) & \text{if } \frac{1}{4}(2+s)\leq t\leq 1\end{cases}.$$

If we set $s=0$ in the above homotopy, we get the left-hand side of our equation, and when $s=1$, we get the right-hand side. Since we have a valid homotopy we see that our product operation, or composition, is associative. Thus, we have a semigroup, and we can now omit the brackets and just write the product of the three paths as $\sigma\cdot\tau\cdot\mu=[\sigma\cdot\tau\cdot\mu]$.

If our composition satisfies two further requirements, we will have a group rather than merely a semigroup. First, our composition must have a neutral element defined by the property that the product of any loop in our set with the neutral element leaves the loop unaffected. In fact, we have already defined a neutral element called the constant path, $\epsilon_x(t)$, $0\leq t\leq 1$. (We will drop the label, x, attached to ϵ and just keep in mind that our entire group is defined with respect to a single point, x, in our space X.) To show that the element is indeed neutral with respect to our product, we need to show that, for a loop σ, we have that $\epsilon\cdot\sigma=\sigma\cdot\epsilon=[\sigma]\cdot[\epsilon]=[\sigma\cdot\epsilon]$ is homotopic to σ. It is easy to see that the following homotopy works fine:

$$H(t,s)=\begin{cases}\epsilon(2t) & \text{if } 0\leq t\leq \frac{1}{2}(1-s)\\ \sigma(\frac{2t-1+s}{1+s}) & \text{if } \frac{1}{2}(1-s)\leq t\leq 1\end{cases}.$$

Second, we need every element in our set of loops to be invertible with respect to our composition; i.e., for a given loop σ, we need an element, σ^{-1}, such that the composition of the two is homotopic to the neutral element, $\sigma\cdot\sigma^{-1}=\epsilon$. A homotopy which shows that $\sigma(1-t)=\sigma^{-1}$ is

$$H(t,s)=\begin{cases}\sigma^{-1}(\frac{2t}{1-s}) & \text{if } 0\leq t\leq \frac{1}{2}(1-s)\\ \epsilon(\frac{2t+s-1}{2s}) & \text{if } \frac{1}{2}(1-s)\leq t\leq \frac{1}{2}(1+s)\\ \sigma(\frac{2t-1-s}{1-s}) & \text{if } \frac{1}{2}(1+s)\leq \frac{1}{2}(1+s)\leq t\leq 1\end{cases}$$

which is equal to $\sigma^{-1}\cdot\sigma$ for $s=0$ and equal to ϵ for $s=1$.

To summarize, we have an algebraic structure with an associative composition (loops, with "traversing them in order" being the composition), making it a semigroup. We also showed that the composition had a neutral element (identity, constant loop) and each element had an inverse (traverse in opposite direction) with respect to the composition. Thus, we have defined a group at each point $x\in X$.

This group is called *The fundamental group of the space, X, with basepoint, x.* It is denoted by $\pi_1(X,x)$ and is the set of equivalence classes of loops at x under homotopy equivalence.

D.2.1. Homomorphisms

The fundamental group is defined at a specific point. How do we relate the fundamental groups at two different points? How do we relate, or find functions between, the fundamental groups defined in different base spaces? The answers to these questions involve finding group homomorphisms.

A *homomorphism* is a mapping, ϕ, from one group, G, to another group, G', which preserves the group product. If g_1, g_2, and g_3 are elements of G such that $g_1 \cdot g_2 = g_3$, then $\phi(g_1) \cdot \phi(g_2) = \phi(g_3)$ in the group G'. In another way, $\phi(g_1 \cdot g_2) = \phi(g_1) \cdot \phi(g_2)$. Homomorphisms are not, in general, one-to-one (*injective*), so several elements of G may be mapped through ϕ to the same element in G'.

If the identity element in G is e, then $\phi(e) = e'$ will be the identity element in G' since $g_1 \cdot e = g_1$ implies $\phi(g_1) \cdot \phi(e) = \phi(g_1)$ due to the definition of group homomorphisms. Let us assume, since a homomorphism may be many-to-one, that elements $g_1, \cdots, g_n$ of G all map to the identity, e', of G'. Then choosing some other element, g_k, of G, we find n distinct elements of G given by $g_k \cdot g_1, \cdots, g_k \cdot g_n$ which are all mapped to the same element, $\phi(g_k)$, in G'. Although our homomorphism is many-to-one, it is of a specific type in that the same number of elements of G will map to each of the elements of G'. This number is uniquely specified when we find how many elements are mapped to the identity, e'. The equation, $g \cdot g^{-1} = e$, implies that $\phi(g) \cdot \phi(g^{-1}) = e'$. Thus, if g is mapped to the identity, then g^{-1} also is. Hence, the set of all elements of G that are mapped to the identity in G' form an *invariant subgroup* of G, which is defined as a subset of G which is also a group and is invariant under multiplication by elements of G'. This subgroup is called the *kernel* of the homomorphism, ϕ. Each of the elements in this kernel is mapped to the identity in G' so that if a different element, g_k, of G is multiplied by any of the elements in the kernel, it will not change the element in G' to which g_k is mapped. The moral of the story is that we can easily create a one-to-one map out of this many-to-one homomorphism by dividing out the kernel. Create a new homomorphism, $\overline{\phi} : G/Ker(\phi) \to G'$, where two elements in $G/Ker(\phi)$ are equivalent if their difference is an element in the kernel of ϕ. If ϕ is *surjective*, then every element in G' has an element in G that is mapped to it, and $\overline{\phi}$ is an isomorphism. The groups $G/Ker(\phi)$ and G' are identical. We note that homomorphisms are not invertible, in general, since many-to-one maps project out structure.

There may be many different homomorphisms between two groups. The set of all possible homomorphisms from G to G' is denoted $Hom(G, G')$. If N is an invariant subgroup of G, then $Hom(G/N, G')$ is equal to $\{\phi \in Hom(G, G') : \phi(N) = e'\}$.

D.2.2. Normal subgroups

A subgroup, N, of a group, G, is normal in G if $g^{-1}Ng \subseteq N$ for all $g \in G$. This relationship is denoted $N \lhd G$ where we are now suppressing the symbol $\cdot$ since the group composition is understood.

An interesting normal subgroup of G is called the *commutator subgroup*, denoted $[G, G]$, consisting of all products of the form

$$(g_i g_j g_i^{-1} g_j^{-1}) \cdots (g_k g_l g_k^{-1} g_l^{-1}).$$

If G' is abelian, then any homomorphism from $G \to G'$ sends all of these commutators to the identity in G'. Hence, $Hom(G/[G,G],G') = Hom(G,G')$. Often, when groups are abelian, the composition is written as $+$ rather than $\cdot$; the identity is written as 0 rather than e or 1; and the inverse of an element, g, is written $-g$ rather than g^{-1}.

D.3. Fundamental group homomorphisms

We have developed group theory and defined the notion of a homomorphism between groups. We are now prepared to find homomorphisms which relate the fundamental groups at two different points in a space and between two different spaces.

Let $f : X \to Y$ be a continuous function from the space X on which we have a fundamental group to a space Y on which we also have a fundamental group. If $f(x) = y$, then f determines a homomorphism of groups,

$$f_* : \pi_1(X,x) \to \pi_1(Y,y),$$

that takes the class $[\sigma]$ in X to the class $[f \circ \sigma]$ in Y.

We need to verify that this is well defined and that it is indeed a group homomorphism. Since f is continuous, each x in X will map to a single y in Y, and the fundamental group, $\pi_1(X,x)$, will map into a well-defined fundamental group, $\pi_1(Y,y)$. To show that it is a group homomorphism, we need to show that it preserves the group composition. If $\sigma \cdot \tau = \mu$ in $\pi_1(X,x)$, then the map gives $[f \circ \sigma] \cdot [f \circ \tau]$. If $\sigma(t)$ goes from $x = \sigma(0)$ to $x' = \sigma(1)$, then $f \circ \sigma = f(\sigma(t))$ goes from $y = f(x)$ to $y' = f(x')$. Similarly, $f \circ \tau$ goes from $y' = f(x')$ to $y'' = f(x'')$. Now, $\mu(t)$ is defined as

$$(\sigma \cdot \tau)(t) = \begin{cases} \sigma(2t) & \text{if } 0 \leq t \leq \frac{1}{2} \\ \tau(2t-1) & \text{if } \frac{1}{2} \leq t \leq 1 \end{cases} .$$

Hence, $f(\sigma \cdot \tau(t))$ then goes from $y = f(x)$ to $y'' = f(x'')$ as required. The product is preserved, and f is a homomorphism. Our equivalence relation is given by homotopy so that, if two loops in X are homotopic, they will still be homotopic in Y. On the other hand, if two loops are not homotopic in X, it is *not* necessarily true that they will not be homotopic in Y. Homomorphisms can be many-to-one, and two distinct (non-homotopic) elements in X may be mapped to the same element (homotopy class) in Y (homomorphisms may project out structure), as may happen, for example, if X has more than one connected component. For example, if the order of $\pi_1(X,x)$ is greater than that of $\pi_1(Y,y)$, then there are no homomorphisms, f_* and g_*, such that the composite $\pi_1(X,x) \xrightarrow{f_*} \pi_1(Y,y) \xrightarrow{g_*} \pi_1(X,x)$ is the identity. If $\pi_1(X,x)$ is a non-trivial group such as the integers in the case of the circle and $\pi_1(Y,y)$ is trivial as in the case of the disk, then X cannot be embedded in Y as a retract; one cannot continuously retract a disk to its boundary circle.

We have shown how to relate the fundamental groups of two different spaces, X and Y, at basepoints x and y. We will now show how to relate the fundamental groups at two different basepoints, x and x', in the same space, X. The idea here is similar to that of finding a *connection* in differential geometry or the theory of fibre bundles relating the tangent spaces or fibres on a manifold at two different basepoints.

If X is path connected, the fundamental groups at two different base points will be the same up to isomorphism. Suppose τ is a path from x to x', and define a map $\tau_{\#} : \pi_1(X, x) \to \pi_1(X, x')$ by $[\gamma] \mapsto [\tau^{-1} \cdot (\gamma \cdot \tau)] = [(\tau^{-1} \cdot \gamma) \cdot \tau]$. This map is a homomorphism and an isomorphism since $(\tau^{-1})_{\#}$ gives the inverse homomorphism. What this map essentially does is take a class of loops at x and leave them exactly as they were, changing only the basepoint; i.e., start at the point x', travel the inverse of path τ to the basepoint x; traverse the loop, and finally travel back along τ to end up at x'. Since we travel both directions along τ, this line is homotopic to the constant path x', or the identity. In essence, we have merely substituted basepoints. If the space was not path connected, this method would not work since we would not be able to find a suitable τ connecting x to each possible x'.

This isomorphism shows us that, for path connected spaces at least, the basepoint need not be stated. The same fundamental group would result had we chosen any other point in the space X as our basepoint. It is clear that any choice of τ works as long as they are all homotopic.

Proposition: The fundamental group of a Cartesian product is the product of the fundamental groups of the spaces:

$$\pi_1(X \times Y, x \times y) \cong \pi_1(X, x) \times \pi_1(Y, y).$$

Proof: Let $\pi_1(X, x) = \{g_1, ..., g_n\}$ and $\pi_1(Y, y) = \{h_1, ..., h_m\}$, where each element represents a different equivalence class of loops at the basepoint in the respective spaces. Let $I : \pi_1(X \times Y, x \times y) \to \pi_1(X, x) \times \pi_1(Y, y)$ be the map that takes the product, $g_i \cdot h_k \mapsto (g_i, h_k)$. Then, I is an isomorphism.

□

Proposition: If $\mathfrak{G}$ is a topological group, then $\pi_1(\mathfrak{G}, e)$ is commutative.

Proof: A *topological group* is a topological space which is also a group such that the multiplication and inverse maps are continuous. A topological space is a space with a topology. The topology consists of open sets, and these open sets are the group elements. For example, we say "a neighborhood of the identity" to mean that open set which represents the identity element e of the group. To say that a group is *commutative* means that its composition is abelian as we defined previously for abelian semigroups.

Our space is topological group $\mathfrak{G}$, and our basepoint is the neighborhood of the identity e. A loop is, therefore, a sequence of open sets which begins and ends at the identity open set. Let $\pi_1(\mathfrak{G}, e) = \{e, g_2, ..., g_n\}$ be the equivalence classes of loops in $\mathfrak{G}$. We want to show that our composition, $\cdot$, is such that, for loops g_i and g_j, we have $g_i \cdot g_j = g_j \cdot g_i$.

Since $\mathfrak{G}$ is a topological space, the union of elements in any subcollection of $\mathfrak{G}$ is an open set which is itself an element of $\mathfrak{G}$. Thus, if $a_1 \cup ... \cup a_n$ form a loop in $\mathfrak{G}$, then the union is also an element of $\mathfrak{G}$, say a_k. Hence, the group elements forming equivalence classes of loops in $\mathfrak{G}$ are merely a subset of the open sets of $\mathfrak{G}$. The product, $g_i \cdot g_j$, is a union of open sets: a_i forming the loop g_i and b_j forming the loop g_j. The resulting loop (equivalence class) formed by the product. call it g_k, is given by the union

$$g_i \cdot g_j = a_i \cup b_j = b_j \cup a_i = g_j \cdot g_i,$$

and our product is, therefore, commutative.

□

D.4. Homotopy

We have shown how two paths can be homotopic with homotopy $H(t,s)$, where t labels the position on the curve and s labels the particular curve we are traversing. We now want to show that two functions can also be homotopic. The point will be to give a simple relation between the $\tau_\#$ map defined earlier relating groups at basepoints and the f_* map relating groups in different spaces. We say that two continuous maps, $f_0 : X \to Y$ and $f_1 : X \to Y$, are homotopic if there exists a continuous mapping, $H : X \times [0,1] \to Y$, such that $H(z,0) = f_0(z)$ and $H(z,1) = f_1(z)$ *for all points*, z, in X.

Let x be a basepoint in X, and let $y_0 = f_0(x)$ and $y_1 = f_1(x)$. Then, f_0 and f_1 map the same basepoint in X to two different points in Y. $H(x,t)$ is a homotopy between the two functions (sets of points) in Y and is, therefore, a path in Y for each x. We will call it $\tau(t)$. We have a path between the two points, thus we also have a map between the fundamental groups at these points. This map (isomorphism) is called $\tau_\#$ and takes an element of the group $\pi_1(Y, y_0)$ to an element of the group $\pi_1(Y, y_1)$. Since f_0 and f_1 are maps between basepoints in two different spaces we need another map (homomorphism), $(f_0)_*$ and $(f_1)_*$, mapping the fundamental group in X at basepoint x to the fundamental group in Y at basepoints $f_0(x)$ and $f_1(x)$, respectively.

The point of this discussion is that $\tau_\# \circ (f_0)_* = (f_1)_*$. In other words, we map the fundamental group in X at x to the fundamental group in Y at y_0 and then follow this map by mapping the group in Y at y_0 to the group in Y at y_1. This process is identical to simply mapping the group in X at x straight to the group in Y at y_1 directly via $(f_1)_*$.

To prove this equality, we need a homotopy between the left- and right-hand sides, i.e., between loops $\tau^{-1} \cdot ((f_0 \circ \gamma) \cdot \tau)$ and $f_1 \circ \gamma$. Our homotopy is given by $H(\gamma(t), s) = \tau(s) = \tau^{-1} \cdot \tau \cdot \tau$. When $s = 0$, we have $\tau^{-1} \cdot \tau(0) \cdot \tau = \tau^{-1} \cdot f_0(\gamma(t)) \cdot \tau$, and when $s = 1$, we have $\tau(1) = f_1(\gamma(t)) = f_1 \circ \gamma$ as required. If the point $f_0(x) = f_1(x)$ and if $H(x,s) = y$ for all s, then the map $(f_0)_* = (f_1)_*$.

One last point worth mentioning is this: If $f : X \to Y$ and $g : Y \to X$ are two maps such that $g \circ f$ is homotopic to the identity in X and $g \circ f$ is homotopic to the identity in Y, then the two spaces are the same in that they have the same homotopy type. If such a g exists, then f is called a homotopy equivalence. The spaces will have the same fundamental groups, and the map, f_* or g_*, will provide an isomorphism between the fundamental groups.

Proposition:

(a) Homotopy type is an equivalence relation.

(b) Homotopy equivalence $f : X \to Y$ determines an isomorphism $f_* : \pi_1(X, x) \to \pi_1(Y, f(x))$ of fundamental groups. In particular, if $i : X \to Y$ embeds X as a deformation retract of Y, then $i_* : \pi_1(X, x) \to \pi_1(Y, i(x))$ is an isomorphism.

Proof:

(a) An equivalence relation is an identification of elements of sets which is reflexive, symmetric, and transitive. In our case, the set is that of spaces which have fundamental groups. Having the same homotopy type is reflexive, $X \cong X$. This reflexivity is clear since a space always has the same fundamental group as itself. It is also symmetric in that, if $X \cong Y$, then $Y \cong X$. The fundamental groups are the same. The fact that it is transitive is only slightly more difficult. Being transitive means that, if $X \cong Y$ and $Y \cong Z$, then $X \cong Z$. Let f, g, h, and p be maps such that (as in the definition of having the same homotopy type) $f \circ g = id_Y$, $g \circ f = id_X$, $h \circ p = id_Y$, and $p \circ h = id_Z$. Then, $id_Z = p \circ h = p \circ id_Y \circ h = (p \circ f) \circ (g \circ h)$ and $id_X = g \circ f = g \circ id_Y \circ f = (g \circ h) \circ (p \circ f)$ which shows that it is transitive. Therefore, having the same homotopy type is an equivalence relation.

(b) An isomorphism is a function, f_*, which is injective and surjective (a bijection) such that $f_*(g_1 \cdot g_2) = f_*(g_1) \cdot f_*(g_2)$ for fundamental group elements, g_1 and g_2, of $\pi_1(X, x)$. (This composition rule actually comes from the fact that f_* is a homomorphism.) The fact that f is a homotopy equivalence means that it has an inverse g and is, therefore, a bijection. If x_1 and x_2 are two elements of X such that $f(x_1) = f(x_2)$ in Y, then $g \circ f(x_1) = g \circ f(x_2)$, meaning that $id_X(x_1) = id_X(x_2) \Rightarrow x_1 = x_2$ so it is injective. If $y \in Y$, then $g(y) \in X$ and $f \circ g(y) = id_Y(y) = y$, so f is surjective. We have then that f_* is also a bijection between fundamental groups since $g_* : \pi_1(Y, f(x)) \to \pi_1(X, g \circ f(x)) = \pi_1(X, x)$ is its inverse, and our previous method follows the same way with replacements $x \to X$, $y \to Y$, $id_X \to e_X$, etc. Therefore, f_* is an isomorphism.

□

We can use the proposition to prove that, if the identity on X embeds X in Y as a deformation retract, then i_* is an isomorphism. A deformation retract is a subspace, X, of a space, Y, which has a continuous mapping, called a retract $r : Y \to X$, such that the identity map in Y is homotopic to the map $i \circ r$, where i is the inclusion of X in Y. so i and r are the f and g as in the above proposition since $i \circ r$ is homotopic to the identity in Y. Therefore, by our previous argument, we have an isomorphism $i_* : \pi_1(X, x) \to \pi_1(Y, i(x))$.

We can see that, if X is a deformation retract, the retraction itself forms a map between all loops in Y. Therefore, the fundamental groups must be the same.

□

Proposition: The mapping from S^2 to itself that takes (x, y, z) to $(-x, -y, z)$ is homotopic to the identity map. The mapping that takes (x, y, z) to $(x, y, -z)$ is homotopic to the antipodal map.

Proof: The map, $(x, y, z) \mapsto (\cos\theta \sin\phi, \sin\theta \sin\phi, \cos\phi)$ is the identity map on S^2. The map $(x, y, z) \mapsto (-\cos\theta \sin\phi, -\sin\theta \sin\phi, \cos\phi)$ is homotopic to the identity through the homotopy, $H(\vec{x}, s) = (\cos(\theta + s\pi)\sin\phi, \sin(\theta + s\pi)\sin\phi, \cos\phi)$, so that $H(\vec{x}, 0) = id_{S^2}$ and $H(\vec{x}, 1)$ is that map given.

The map, $(x, y, z) \mapsto (\cos\theta \sin\phi, \sin\theta \sin\phi, -\cos\phi)$, is homotopic to the antipodal map $(x, y, z) \mapsto (-x, -y, -z)$, through the homotopy,

$$H(\vec{x}, s) = (\cos(\theta + s\pi)\sin\phi, \sin(\theta + s\pi)\sin\phi, -\cos\phi),$$

so that $H(\vec{x}, 0)$ is the map given and $H(\vec{x}, 1)$ is the antipodal map.

□

Proposition: If n is odd, the identity mapping on S^n is homotopic to the antipodal mapping.

Proof: *At first glance one might expect the following method to work*: An expression for the identity map in S^n is

$$(x_1, ..., x_{n+1}) \mapsto (\cos(\theta_1), \sin(\theta_1)\cos(\theta_2), \sin(\theta_1)\sin(\theta_2)\cos(\theta_3), \\ ..., \sin(\theta_1)\cdots\sin(\theta_n)).$$

The antipodal map is

$$(x_1, ..., x_{n+1}) \mapsto (-\cos(\theta_1), -\sin(\theta_1)\cos(\theta_2), -\sin(\theta_1)\sin(\theta_2)\cos(\theta_3), \\ ..., -\sin(\theta_1)\cdots\sin(\theta_n)).$$

The homotopy between the two is given by

$$H(\vec{x}, s) = (\cos(\theta_1 + s\pi), \sin(\theta_1 + s\pi)\cos(\theta_2), \sin(\theta_1 + s\pi)\sin(\theta_2)\cos(\theta_3), \\ ..., \sin(\theta_1 + s\pi)\cdots\sin(\theta_{n-1})\cos(\theta_n), \sin(\theta_1 + s\pi)\cdots\sin(\theta_n)).$$

This expression is the identity for $s = 0$ and is the antipodal for $s = 1$.

The problem with this homotopy is firstly that it seems to work for n both, even or odd, which is in contradiction to the stated problem. The serious flaw is that it is not continuous.

The homotopy map has to be one that changes the entire coordinate system in a continuous fashion. In S^1, there is no problem since the antipodal map and rotation of all of S^1 by 180° are the same thing. However, in S^2, there is no way to change all of the vectors into their antipodal counterparts by a continuous change of the entire sphere. This reason is the same reason why you cannot comb the hair on a sphere. There are going to be fixed points, and in fact, only the points in a subset, equivalent to S^1, will be mapped to their antipodal points by any homotopy. If we were looking at S^2 in a 4-dimensional space, we could turn the entire thing inside out using the 4th dimensionl It would be a perfectly reasonable homotopy of the identity to its antipodal map.

Four dimensions notwithstanding, the homotopy that we need is as follows:

$$H(\vec{x}, s) = (\cos(\theta_1 + s\pi), \sin(\theta_1 + s\pi)\cos(\theta_2(1-2s)),$$
$$\sin(\theta_1 + s\pi)\sin(\theta_2(1-2s))\cos(\theta_3 + s\pi), ..., \sin(\theta_1 + s\pi)$$
$$\cdots \sin(\theta_{n-1}(1-2s))\cos(\theta_n + s\pi), \sin(\theta_1 + s\pi)\cdots\sin(\theta_n + s\pi)).$$

if n is odd.

Notice that all we have done is set $\theta_i \to \theta_i + s\pi$ if i is odd and $\theta_i \to \theta_i(1-2s)$ if i is even. We see that, for $s = 0$, we get the identity. If $s = 1$, we get a net minus sign in each term which gives the antipodal map. If n is even by this method, we would get

$$H(\vec{x}, s) = (\cos(\theta_1 + s\pi), \sin(\theta_1 + s\pi)\cos(\theta_2(1-2s)),$$
$$\sin(\theta_1 + s\pi)\sin(\theta_2(1-2s))\cos(\theta_3 + s\pi), ..., \sin(\theta_1 + s\pi)$$
$$\cdots \sin(\theta_{n-1} + s\pi)\cos(\theta_n(1-2s)), \sin(\theta_1 + s\pi)\cdots\sin(\theta_n(1-2s))),$$

which is still the identity if $s = 0$, but if $s = 1$, we have

$$(x_1, ..., x_{n+1}) \to (-x_1, ..., -x_n, +x_{n+1}),$$

where the last term (which contains all sin functions) remains positive. The reason is that, in our mapping, all sin functions give a minus sign and all odd indexed cos's give a minus sign. Each component gets an odd number of minus signs in this fashion except the last which has an even number of sin factors.

We can see that implicit in the definition of S^{2n}, we need to specify that it is embedded in a $2n + 1$ dimensional space and no higher, or there would be no difference between even and odd dimensions. It is interesting that the presence of another component, even if it is zero, can make a difference. For example, (x, y, z) has much different properties than $(x, y, z, 0)$ in homotopy theory. These ideas lead to stable homotopy and K-theory [90, 91]

□

Proposition: An element of O_{n+1} determines a mapping of S^n to itself. Two mappings are homotopic if and only if they have the same determinant. The determinant of a map homotopic to the identity is $+1$, and the determinant of a map homotopic to the antipodal map is -1.

Proof: An element A of O_{n+1} is an orthogonal $(n+1) \times (n+1)$ matrix with one condition on each of its rows given by $O^T O = 1$. It is, thus, a map between unit vectors in $\mathbb{R}^{n+1}$, i.e., a map from S^n to S^n. Notice:

$$\begin{aligned}\det(A^T A) &= \det(A^T)\det(A),\\ \det(1) &= \det(A)\det(A),\\ 1 &= (\det(A))^2,\\ \Rightarrow \det(A) &= \pm 1.\end{aligned} \tag{D.2}$$

Let A and B be two such maps *which are homotopic* and such that $\det(A) > 0$ and $\det(B) < 0$. Then, there exists a homotopy, $H(x,t)$, such that $H(x,0) = A(x)$ and $H(x,1) = B(x)$. Also, $\det(A) = 0$ is a condition that represents a surface in $\mathbb{R}^{(n+1)(N+1)}$. Removing this surface by the requirement that the determinant be either 1 or -1 separates the space into two connected components. Hence, we *cannot* have a path connecting a point on one side with any point on the other. (They are in different connected components.) Since our homotopy, $H(x,t)$, for a fixed x is defined as a path from the point $H(x,0) = A(x)$ on one side to the point $H(x,1)$ on the other side, we have a contradiction. Therefore, A and B must have the same determinant in order that a homotopy exists.

Now, we want to show that, if a matrix has determinant 1, it is homotopic to the identity map and, if it has determinant -1, it is homotopic to the antipodal map. The identity map is a matrix with 1's on the diagonal and, hence, has determinant 1. The antipodal map is a matrix with -1's on the diagonal. Therefore, if n is even (i.e., $n+1$ is odd), then the determinant will be -1. It is clear that each connected component of the space is, in itself, path connected (trivial homology group). Therefore, all matrices of determinant 1 are homotopic to each other (We can easily construct a homotopy in a path connected space.), and in particular, since the identity has determinant 1, they are all homotopic to the identity. A similar argument applies for the matrices of determinant -1.

It is important to notice that, if n is odd, $n+1$ is even and the antipodal map itself now has determinant +1. The antipodal map is then homotopic to the identity. In particular, in $\mathbb{R}^2$, the orthogonal 2×2 matrices form a map from $S^1 \to S^1$ such that the identity is homotopic to the antipodal map. We saw this homotopy in the previous proposition.

□

As an example we will compute the fundamental group of the real projective line. The antipodal map, a, is homotopic to the identity on S^1. Therefore, we have

$$\pi(S^1, x) \xrightarrow{a_*} \pi(S^1, a(x)) = \pi(\mathbb{RP}, x) \cong \pi(S^1, x) = \mathbb{F}_1 = \mathbb{Z}.$$

The fundamental group of the real projective line is $\mathbb{Z}$, exactly like S^1.

D.5. Homology

We want to relate homotopy and the fundamental group to homology and the homology groups (closed chains modulo boundaries). For any topological space, X, with basepoint x, there is a homomorphism, $AB : \pi_1(X, x) \to H_1(X)$, from the fundamental group, $\pi_1(X, x)$, to the homology group, $H_1(X)$, taking the class $[\gamma]$ of a loop γ at x to the homology class of γ, regarded as a 1-chain. It takes the constant path, ϵ_x, to 0. Homotopic paths define the same homology class.

The homology class of a product, $\sigma \cdot \tau$, of loops is the sum of the homology classes of the individual loops. $AB(\sigma\dot{\tau}) = AB(\sigma) + AB(\tau)$, where $+$ is the composition in the group $H_1(X)$. By the definition of homomorphisms, AB is a well-defined group homomorphism.

Proposition: If $f : X \to Y$ is a continuous mapping and $f(x) = y$, the following diagram commutes:

$$\begin{array}{ccc} \pi_1(X,x) & \xrightarrow{f_*} & \pi_1(Y,y) \\ AB\big\downarrow & & \big\downarrow AB \\ H_1(X) & \xrightarrow[f_*]{} & H_1(Y). \end{array}$$

Proof: We simply need to show that $AB \circ f_* = f_* \circ AB$, where the equivalence is that they are *homologous.* The difference between the 1-chain on the left and the one on the right must be a boundary.

We are finished if we show that, for two 1-chains, σ and τ, in X, $\sigma - \tau$ is a boundary if σ and τ are homotopic. We essentially want to prove that the notion of being homotopic implies that of being homologous.

If two loops, σ and τ, are homotopic, there is a continuous homotopy between them. The set of all points, $H(t, s)$ $0 \le t \le 1$, $0 < s < 1$, is a region in X which is bounded by the curves $H(t, 0) = \sigma$ and $H(t, 1) = \tau$. Hence, $\sigma - \tau$ is a boundary. Therefore, homotopic implies homologous.

□

Observe that homologous does not imply homotopic, since the abelianization homomorphism, AB, may project out structure (many-to-one). In fact, if X has more than one connected component, the fundamental group will be different in each component and, hence, the fundamental group at one point, x, will not determine the homology group.

Since $H_1(X)$ is an abelian group and $\pi_1(X, x)$ is not in general, the homomorphism must vanish on the commutator subgroup, $[\pi_1(X, x), \pi_1(X, x)]$. This subgroup is a normal subgroup. We define the abelianized fundamental group as (See our previous discussion of commutator subgroups.)

$$\pi_1(X, x)_{\mathsf{abel}} = \pi_1(X, x)/[\pi_1(X, x), \pi_1(X, x)].$$

If X is path connected, the abelianization homomorphism is an isomorphism from $\pi_1(X, x)_{\mathsf{abel}}$ to $H_1(X)$. To show this fact, we find a homomorphism from the group of 1-cycles to $\pi_1(X, x)_{\mathsf{abel}}$ and show that the 1-boundaries map to zero (constant paths). This homomorphism will give a map back from $H_1(X)$ which is the inverse that we need.

Proposition: If X is a path-connected space, then the canonical homomorphism from $\pi_1(X, x)_{\mathsf{abel}}$ to $H_1(X)$ is an isomorphism.

Proof: We must define a homomorphism from the abelian group, $Z_1(X)$, of 1-cycles to $\pi_1(X,x)_{\mathsf{abel}}$ and show that the 1-boundaries, $B_1(X)$, map to zero. This homomorphism will be a map from $H_1(X)$ back to $\pi_1(X,x)_{\mathsf{abel}}$. Then, we will show that the two homomorphisms are inverses of each other which tells us that the canonical homomorphism is an isomorphism.

Let $\gamma = \sum_i n_i \gamma_i$ be a 1-cycle so that each of the γ_i are a side of some little rectangle. Define a weight factor, n_i, such that the endpoint of the last path is equal to the starting point of the first one and the sum of all the coefficient weight factors is zero. Now, let the paths, γ_i, have endpoints, $a(i)$ and $b(i)$, in X. For each point c that occurs as an endpoint of any γ_i, choose a path, τ_c, from x (which is our basepoint) to c. Let γ_i' be the loop at x defined by

$$\gamma_i' = \tau_{a(i)} \cdot \gamma_i \cdot \tau_{b(i)}^{-1},$$

where γ_i' starts at the basepoint x, travels via $\tau_{a(i)}$ to $a(i)$ (the starting point of γ_i), then travels along the small path γ_i, and finally $\tau_{b(i)}^{-1}$ leads from the final point of γ_i back to the basepoint x. Hence, we have a small loop.

Define the map from $Z_1(X)$ to $\pi_1(X,x)$ by sending $\gamma = \sum_i n_i \gamma_i$ to the class of $\Pi_i [\gamma_i']^{n_i}$. The order in the products is unimportant since the group $\pi_1(X,x)_{\mathsf{abel}}$ is abelian. We see that sums in $Z_1(X)$ are sent to products in $\pi_1(X,x)_{\mathsf{abel}}$ (similar to exponentiation). Composition in $\pi_1(X,x)_{\mathsf{abel}}$ is defined by first traversing one and then the other path (with the reparametrization of t as you will recall) and similarly with loops. Since an element of $\pi_1(X,x)$ (no abelianization) depends on the basepoint, it generally matters which order you take. This dependence can be seen by imagining two paths from say a to b. Together, they form a loop. The basepoint of the loop depends on which order you take the paths. The composition in $Z_1(X)$ which we call summation is defined by traveling around one loop with a weight factor in direct sum with traveling around the other with a weight factor (taking values in some field). For example, you can write any group element as a vector, say $(n_1, .., n_k)$, where each component represents the weight factor associated with an equivalence class of homologous loops and can, therefore, be re-written using the loops as basis vectors as $(n_1, ..., n_k) = n_1\gamma_1 + \cdots + n_k\gamma_k$. We see that a subset of the elements of $\pi_1(X,x)$ should be equivalent to the elements of $H_1(X)$ with the difference being the basepoint. Abelianization is, therefore, making every point on a loop an equivalent basepoint. Taking $\pi_1(X,x)$, modulo the points on each loop, gives equivalence classes of loops which do not depend on the basepoint. These equivalence classes are the elements of $Z_1(X)$ as required.

To continue, we want to verify that this map is independent of the choice of paths, τ_c. Suppose $\tilde{\tau}_c$ is another path from x to c for each c. Let

$$\tilde{\gamma}_i' = \tilde{\tau}_{a(i)} \cdot \gamma_i \cdot \tilde{\tau}_{b(i)}^{-1}.$$

Let $\mathcal{O}_c$ be the loop $\tilde{\tau}_c \cdot \tau_c^{-1}$. Then, $[\tilde{\gamma}_i'] = [\mathcal{O}_{a(i)}] \cdot [\gamma_i'] \cdot [\mathcal{O}_{b(i)}^{-1}]$ as can be seen by canceling all paths which are traversed in both forward and backward directions (leaving $\tilde{\gamma}_i'$). We have written (using $\mathcal{O}_c$) the new loop, $\tilde{\gamma}_i'$, in terms of the old loop, γ_i'. We take this loop to the $\pi_1(X,x)_{\mathsf{abel}}$ space via our map (changing sums of the path segments i to products) to get

$$\Pi_i[\tilde{\gamma}_i']^{n_i} = \Pi_i[\gamma_i']^{n_i} \cdot \left(\Pi_i[\mathcal{O}_{a(i)}]^{n_i} \cdot \Pi_i[\mathcal{O}_{b(i)}]^{-n_i}\right) \Pi_i[\tilde{\gamma}_i']^{n_i} \\ = \Pi_i[\gamma_i']^{n_i}, \tag{D.3}$$

where we have used the fact that the product in $\pi_1(X, x)_{\mathsf{abel}}$ is commutative and n_i is the weight factor associated with each segment. The last equality comes from the fact that the entire loop, γ, is a 1-cycle so that each point c (vertex) occurs as many times as an endpoint, $b(i)$, of a segment as it does a starting point, $a(i)$. Thus the term in the bracket will be the identity since each different subscript, $b(i)$, in $[\mathcal{O}_{b(i)}]^{-n_i}$ will have a corresponding subscript, $a(i)$, in $[\mathcal{O}_{a(i)}]^{n_i}$. These points are, in fact, the same point, hence they are inverses.

A further comment on this. Recall that we had a subgroup, G', of a group G as a normal subgroup if, for all $g \in G$ and $g' \in G'$, $g \cdot g' \cdot g^{-1} \in G'$. We have used this closure under conjugation to rewrite our $\tilde{\gamma}_i'$, which is an element of the normal subgroup of the 1-chains, $C_1(X)$, called the 1-cycles, $Z_1(X)$, in terms of another element in the normal subgroup called γ_i'. We used the elements, $\mathcal{O}$, of the group of chains, $C_1(X)$. This is a useful trick since we know that in $\pi(X, x)_{\mathsf{abel}}$ to which we are mapping, the product commutes, allowing us to get rid of the $\mathcal{O}$'s.

We have now shown that the choice of τ_c is irrelevant to the map, so the map is now a homomorphism. We separate our loop into tiny segments, making the segments into tiny loops, sending the loops to $\pi_1(X, x)_{\mathsf{abel}}$, and then reassembling them. In this way, a path surrounding a hole can be mapped in a well defined way.

Now that we have shown how to perform the map, we need to show that it maps boundary cycles to zero. It suffices to show that γ maps to zero when

$$\gamma = \partial\Gamma = \gamma_1 + \gamma_2 - \gamma_3 - \gamma_4,$$

where $\Gamma : [0,1] \times [0,1] \to X$ is a continuous mapping and the boundary paths, γ_i, are the straight line segments around the four sides of a small square (given by Γ). Recall that, by "maps to zero," we mean that the resulting path in $\pi_1(X, x)_{\mathsf{abel}}$ is homotopic to the constant path, ϵ_x.

Let τ_1 and τ_2 be paths from x to the starting and ending points of γ_1, and let τ_3 and τ_4 be paths from x to the starting and ending points of γ_3. Then, γ maps to the class of

$$[\tau_1 \cdot \gamma_1 \cdot \tau_2^{-1}] \cdot [\tau_2 \cdot \gamma_2 \cdot \tau_4^{-1}] \cdot [\tau_4 \cdot \gamma_3^{-1} \cdot \tau_3^{-1}] \cdot [\tau_3 \cdot \gamma_4^{-1} \cdot \tau_1^{-1}] \\ = [\tau_1 \cdot \gamma_1 \cdot \gamma_2 \cdot \gamma_3^{-1} \cdot \gamma_4^{-1} \cdot \tau_1^{-1}] \\ = [(\tau_1 \cdot \gamma_1 \cdot \gamma_2) \cdot (\tau_1 \cdot \gamma_4 \cdot \gamma_3)^{-1}]. \tag{D.4}$$

To show that this last equality is, in fact, trivial, we need a homotopy which identifies $(\tau_1 \cdot \gamma_1 \cdot \gamma_2)$ with $(\tau_1 \cdot \gamma_4 \cdot \gamma_3)$ with fixed endpoints. In fact, it is enough to show that

$(\gamma_1 \cdot \gamma_2)$ and $(\gamma_4 \cdot \gamma_3)$ are homotopic. The following homotopy works

$$H(t,s) = \begin{cases} \Gamma(2t(1+s), 2ts) & 0 \le t \le \frac{1}{2} \\ \Gamma((2t-1)s+1-s, (2t-1)(1-s)+s) & \frac{1}{2} \le t \le 1 \end{cases}.$$

We can see that

$$\begin{aligned} H(t,0) &= \gamma_1 \cdot \gamma_2 \\ H(t,1) &= \gamma_4 \cdot \gamma_3 \end{aligned} \tag{D.5}$$

Now that we have shown that it is indeed a homomorphism and that boundaries do map to zero, we need only show that this map is, in fact, the inverse of the map $AB : \pi_1(X,x) \to H_1(X)$ defined previously. Both maps will be isomorphisms, and our proposition will be proven.

It is immediate from the definitions that the composite

$$\pi_1(X,x)_{\mathsf{abel}} \to H_1(X) \to \pi_1(X,x)_{\mathsf{abel}}$$

is the identity. For the other composite,

$$H_1(X) \to \pi_1(X,x)_{\mathsf{abel}} \to H_1(X),$$

we have that, for a 1-cycle $\gamma = \sum_i n_i\gamma_i$, and γ_i' defined as before, the element $\Pi_i[\gamma_i']^{n_i}$ in $\pi_1(X,x)_{\mathsf{abel}}$ maps to

$$\sum_i n_i\gamma_i' = \sum_i n_i(\tau_{a(i)} + \gamma - \tau_{b(i)}) = \sum_i n_i\gamma_i = \gamma,$$

where we used the fact that the boundary of γ is zero. (i.e., γ is a 1-cycle so that $a(i)$ and $b(i)$ are the same point.)

□

D.6. Covering spaces

Let X and Y be topological spaces. A *covering* is a continuous map, $p : Y \to X$, with the property that each point of X has an open neighborhood, N, such that $p^{-1}(N)$ is a disjoint union of open sets, each of which is mapped homeomorphically onto N. If N is connected, these sets must be the components of $p^{-1}(N)$. A covering, p, is called a trivial covering if p is isomorphic to the projection of $X \times T \to X$, where T is any set with the discrete topology.[2]

An isomorphism between coverings $p : Y \to X$ and $p' : Y' \to X$ is a homeomorphism, $\phi : Y \to Y'$, such that $p' \circ \phi = p$.

An *n-sheeted covering* is a covering p such that $p^{-1}(x)$ has cardinality n for each x. Usually, the covering space will not have n distinct sheets unless it is trivial (but

[2]The discrete topology is defined such that every point is an open set and every subset is open.

locally it will). If $p : Y \to X$ is a covering and $f : Z \to X$ is a continuous mapping, define a continuous mapping called a *lifting* of f by $\tilde{f} : Z \to Y$ such that $p \circ \tilde{f} = f$. If Z is connected, a lifting is determined by which sheet it maps one point. For example, if a lifting is the image of a curve in the base space on one of the sheets, then $Z = [0, 1]$ is connected, $x \in X = f(t \in Z)$ and $\tilde{f}(t) = y \in Y$, such that $p(y) = x$. There are two useful propositions that we will state without proof.

Path Lifting Let $p : Y \to X$ be a covering, and let $\gamma : [a, b] \to X$ be a continuous path in X. Let y be a point of Y with $p(y) = \gamma(a)$. There is a unique continuous path $\tilde{\gamma} : [a, b] \to Y$, such that $\tilde{\gamma}(a) = y$ and $p \circ \tilde{\gamma}(t) = \gamma(t)$ for all t in the interval $[a, b]$.

The uniqueness is due to the fact that we specified the sheet. We can use the Lebesgue lemma (defined below) on the open sets, $\gamma^{-1}(N)$, where N varies over open sets in X which are evenly covered by p. We have now subdivided the t interval such that each $\gamma([t_{i-1}, t_i])$ is contained in some open set of X evenly covered by p. The lifting of each of the path segments combines to give a lifting of the entire curve.

Lebesgue Lemma: Given any covering of a compact metric space, K, by open sets, there is an $\epsilon > 0$ such that any subset of K of diameter less than ϵ is contained in some open set in the covering.

Homotopy Lifting Let $p : Y \to X$ be a covering, and let H be a homotopy of paths in X. i.e., $H : [a, b] \times [0, 1] \to X$ is a continuous mapping. Let $\gamma_0(t) = H(t, 0)$, $a \leq t \leq b$, be the initial path. Suppose $\tilde{\gamma}_0$ is a lifting of γ_0. Then, there is a unique lifting, $\tilde{H}$, of H with initial path $\tilde{\gamma}_0$. Hence, $H : [a, b] \times [0, 1] \to Y$ is continuous, with $p \circ \tilde{H} = H$ and $\tilde{H}(t, 0) = \tilde{\gamma}_0(t)$, $a \leq t \leq b$.

Proposition: The following maps are covering maps, where $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$ is the group of non-zero complex numbers:

(a) the nth power mapping $\mathbb{C}^* \to \mathbb{C}^*, z \to z^n$ and

(b) the exponential mapping exp:$\mathbb{C} \to \mathbb{C}^*$.

Proof: To show that these maps are covering maps, we must show that they are continuous, the inverse image of open sets are disjoint unions of open sets, and they map each of the disjoint open sets in the covering sets homeomorphically to the open set in the base space.

(a) The nth power map is obviously continuous. Since

$$\begin{aligned} z &\to z^n, \\ re^{i\theta} &\to r^n e^{in\theta}, \\ r\left(\cos(\theta) + i\sin(\theta)\right) &\to r^n\left(\cos(n\theta) + i\sin(n\theta)\right), \end{aligned} \tag{D.6}$$

we have that all points of the form $\sqrt[n]{a}e^{i\frac{\theta+2k\pi}{n}}$, where k is an integer, will map to the same point, $ae^{i\theta}$, in the base space. For each k, we have an open set in the covering space surrounding the point which is disjoint from the open set surrounding a point given by a different k. Each open set in the covering space is homeomorphic to the open set in the base space which it is mapped to since the map and its inverse are both continuous bijections.

(b) Due to the identity $e^{x+iy} = e^x(\cos y + i \sin y)$, the exponential mapping here is similar to part (a). The argument is the same as part (a) with $r = e^x$ and $y = n\theta$.

□

D.6.1. G-coverings

A left *action* of a group G on Y is a mapping $G \times Y \to Y$ with $(g, y) \mapsto g \cdot y$ such that

(a) $g \cdot (h \cdot y) = (g \cdot h) \cdot y \ \forall g, h \in G$ and $y \in Y$.

(b) $e \cdot y = y \ \forall y \in Y$, where e is the identity in G.

(c) The mapping, $y \mapsto g \cdot y$, is a homeomorphism of $Y \ \forall g \in G$.

G defines a group of homeomorphisms of the base space Y so that $G \times Y$ is a covering of Y and each sheet is homeomorphic to Y through a homeomorphism given by one of the group elements. Two points, y and y', are in the same *orbit* if there is an element of g in G that maps one to the other, $y' = g \cdot y$. Since G is a group, the property of being in the same orbit is an equivalence relation. The sets of orbits $X = Y/G$ are the equivalence classes. There is a *projection* map, $p : Y \to X$, which maps each point to the orbit containing it. The *orbit space* is the set of equivalence classes in Y/G. These classes consist of points that are not related by an application of a group element. A group G acts *evenly* if each different g in G will map a neighborhood in the base space to a different sheet in the covering space, $G \times Y$. A covering $p : Y \to X$ is called a *G-covering* if it arises from an even action of G on Y with the base space being the space of orbits. An isomorphism of G-coverings is an isomorphism of coverings that commutes with the action of G. To be precise, an isomorphism of $p : Y \to X$ with $p' : Y' \to X$ is a homeomorphism, $\phi : Y \to Y'$, such that $p' \circ \phi = p$ and $\phi(g \cdot y) = g \cdot \phi(y)$ for $g \in G$ and $y \in Y$. The *trivial* G-covering of X is the product $X \times G \to X$, where G acts by (left) multiplication on the second factor. A *section* of a covering $p : Y \to X$ is a continuous mapping, $s : X \to Y$, such that $p \circ s$ is the identity mapping of X. If a G-covering has a section, then it is a trivial G-covering.

For example, let the group, $\mathbb{Z}$, act on $\mathbb{R}$ by translation, $n \cdot r = r + n$. The quotient, $\mathbb{R} \to \mathbb{R}/\mathbb{Z}$, can be identified with the covering map from $\mathbb{R}$ to S^1 given by $p(r) = (\cos(2\pi r), \sin(2\pi r))$. Our covering is simply $r \to p(r)$ since $n \cdot r \to r + n = r \in \mathbb{R}/\mathbb{Z}$ and $p(n \cdot r) = (\cos(2\pi r + 2\pi n), \sin(2\pi r + 2\pi n)) = (\cos(2\pi r), \sin(2\pi r)) = p(r) \in S^1$.

We have $\mathbb{R}$ as the covering space over S^1. This covering space can be visualized as an infinite helix where, above every point in S^1, we have a finite disconnected point set. The covering map takes each of these points down to the base space, and the group elements move along the fibre in the covering space.

For any covering $p : Y \to X$, there is a group, $Aut(Y/X)$, of deck transformations such that

$$Aut(Y/X) = \{\phi : Y \to Y \text{ such that } \phi \text{ is a homeomorphism and } p \circ \phi = p\}.$$

$Aut(Y/X)$ is a group by composition of mappings, and it acts on Y in the sense of the group action above. We can use this group to map along orbits, mapping points in the covering space to other points in the covering space so that their projections to the base space do not change.

For a G-covering, G acts *transitively* on each fibre $p^{-1}(x)$ of p. For y and y' in a fibre, there is a g in G with $g \cdot y = y'$. In addition, this action is *faithful* in that the element, g, taking y to y' is unique.

We can visualize what G is doing: Any element, g, will move a point in the covering space to another point in the same fibre, and each point in the cover will be moved to a unique point by a g.

D.7. The fundamental group and coverings

We will now show how the fundamental group controls the possible coverings a space can have. Coverings correspond to subgroups of the fundamental group. There is a correspondence between G-coverings and homomorphisms from the fundamental group to G. We have seen that a a G-covering will have a group, G, which moves points transitively along the fibres and that each sheet corresponds to a different group element. The fundamental group is such a group. For example, consider the infinite helix over S^1. The element corresponding to looping once around S^1 will lift to a path (not a loop) in the covering space that moves along the fibre between two points above the same x. Hence, there is a relationship between the group, G, in a G-covering and the fundamental group of the base space. It turns out that the group G in a G-covering will be a subgroup of the fundamental group of the base space. The reason for this relationship is as follows. If you have a homomorphism from $\pi_1(X,x)$ to a group, H, it may be many-to-one, but H will always be a subgroup of $\pi_1(X,x)$. This subgroup will define a G-covering with $H = G$. Hence, each different homomorphism from $\pi_1(X,x)$ to G will define a covering and will correspond to a subgroup of $\pi_1(X,x)$.

Proposition:

(a) If σ is a loop at x and $\tilde{\sigma}$ is the unique lifting of σ to a path starting at y, $\tilde{\sigma}$ ends at y if and only if its class is in the image of p_*. i.e., $[\sigma] \in p_*(\pi_1(Y,y))$.

(b) If σ and σ' are two paths in X from x to x' and $\tilde{\sigma}$ and $\tilde{\sigma}'$ are the lifts to paths in Y starting at y, $\tilde{\sigma}$ and $\tilde{\sigma}'$ have the same endpoint if and only if $[\sigma' \cdot \sigma^{-1}]$ is in $p_*(\pi_1(Y,y))$.

$$y * \sigma = y * \sigma' \Leftrightarrow [\sigma' \cdot \sigma^{-1}] \in p_*(\pi_1(Y,y)).$$

Recall that $y * \sigma = \tilde{\sigma}(1)$ is the endpoint of the curve in the covering space starting at y.

Proof: Note first that in the double cover of S^1, $\pi_1(S^1,x) = \mathbb{Z}$, the group elements given by the number of times you go around the circle. The loop in the base space that goes once around the circle lifts to a path that goes from $y \in p^{-1}(x)$ to $y' \in p^{-1}(x)$. which are two different points. The endpoint is not the same as the starting point, and the path is not in $\pi_1(Y,y)$. Thus, the loop in the base space is not in $p_*(\pi_1(Y,y))$. On the other hand, the element which loops twice around the circle maps up to a path that goes from y up through y' and then back down to y (since it is a double cover $= \mathbb{Z}/\mathbb{Z}_2$). This path is a loop in Y at y and is, therefore, an element of $\pi_1(Y, y)$. The original loop in S^1 is then an element of $p_*(\pi_1(Y, y))$.

(a) Since p_* is an injection, if $[\sigma] \in p_*(\pi_1(Y, y))$, then $[\sigma] \in \pi_1(X, x)$ has a unique pre-image, $[\tilde{\sigma}]$, such that $p_*[\tilde{\sigma}] = [\sigma]$. We have $\tilde{\sigma}(0) = y$ and $p \circ \tilde{\sigma}(t) = \sigma(t)$ by path lifting. Since $[\sigma] \in \pi_1(X, x)$,

$$\begin{aligned} p \circ \tilde{\sigma}(1) &= \sigma(1) = x = \sigma(0) = p(y) \\ \Rightarrow p(\tilde{\sigma}(1)) &= p(y) \\ \Rightarrow \tilde{\sigma}(1) &= y. \end{aligned}$$

Hence, if $[\sigma] \in p_*(\pi_1(Y, y))$, then $\tilde{\sigma}$ ends at y.

Now, we must show the converse. If $\tilde{\sigma}$ ends at y, then $[\sigma] \in p_*(\pi_1(Y, y))$. If $\tilde{\sigma}$ ends at y, then, since $p(y) = x$, we have that $p(\tilde{\sigma}(0)) = p(\tilde{\sigma}(1)) = x$ so that $\sigma \in p_*(\pi_1(Y, y))$ by path lifting. If σ and σ' are homotopic paths in X, then, by homotopy lifting, $\tilde{\sigma}$ and $\tilde{\sigma}'$ are homotopic paths in Y. Thus, if $\tilde{\sigma}$ ends at y, then $[\sigma] \in p_*(\pi_1(Y, y))$.

(b)

$$\begin{aligned} y * \sigma &= y * \sigma' \\ &\Rightarrow (\tilde{\sigma}' \cdot \tilde{\sigma}^{-1})(1) = y \\ &\Rightarrow [\tilde{\sigma}' \cdot \tilde{\sigma}^{-1}] \in \pi_1(Y, y) \\ &\Rightarrow p_*[(\tilde{\sigma}' \cdot \tilde{\sigma}^{-1})] = [p(\tilde{\sigma}' \cdot \tilde{\sigma}^{-1})] \\ &= [\sigma' \cdot \sigma^{-1}] \\ &\Rightarrow [\sigma' \cdot \sigma^{-1}] \in p_*(\pi_1(Y, y)) \end{aligned}$$

since p_* is an injection.

$$\begin{aligned} [\sigma' \cdot \sigma^{-1}] &\in p_*(\pi_1(Y, y)) \\ \Rightarrow p(\tilde{\sigma}(0)) &= x = p(\tilde{\sigma}'(0)) \\ \Rightarrow p(\tilde{\sigma}(t)) &= \sigma(t), \text{by path lifting} \\ \text{and } p(\tilde{\sigma}'(t)) &= \sigma'(t), \text{by path lifting} \\ \text{since } (\sigma' \cdot \sigma^{-1}) &\in \pi_1(X, x) \\ \Rightarrow \sigma'(1) &= \sigma(1) \\ \Rightarrow p(\tilde{\sigma}'(1)) &= p(\tilde{\sigma}(1)) \\ \Rightarrow \tilde{\sigma}'(1) &= \tilde{\sigma}(1) \\ y * \sigma &= y * \sigma' \end{aligned}$$

□

The preceeding proof has shown that all elements of the fundamental group in the covering space will have images in the fundamental group in the base space at $x = p(y)$ due to the fact that p_* is an injection. The second part tells us that two paths will form a loop in the covering space only if they also form a loop in the base space. Essentially, this comes from the fact that, if two paths have the same starting and ending points, they will be in $\pi_1(Y, y)$ (since they will be a loop), and they will be injected into $\pi_1(X, x)$ via p_*.

Notice that a path starting at y and going around the covering space to end at y' could project down to a loop in X if $p(y) = p(y')$. Thus, a path that was not a loop in Y is now a loop in X. This path in Y would then not be a loop in Y and would not be an element of $\pi_1(Y, y)$. In fact, p_* is not an injection unless the domain is restricted to the loops in Y as our proof shows.

Proposition: If y' is another point in Y with $p(y') = x$, and y and y' can be connected by a path in Y, the image of $\pi_1(Y, y')$ in $\pi_1(X, x)$ is a subgroup conjugate to the image of $\pi_1(Y, y)$. In fact, if σ is a path from y' to y and $\gamma = p \circ \sigma$ is its image,

$$p_*(\pi_1(Y, y')) = [\gamma] \cdot p_*(\pi_1(Y, y)) \cdot [\gamma]^{-1}.$$

Proof: Let H be a subgroup of G. Choose an element, $g \in G$, and form the set $g \cdot h \cdot g^{-1}$. If we let h run through all of the elements in H, we have another subgroup conjugate to H. Choosing various elements of G gives different conjugate subgroups.

The first part of our proof is to show that the image of the group $\pi_1(Y, y')$ at our new point, y', in $\pi_1(X, x)$ is a conjugate subgroup of the image of the group $\pi_1(Y, y)$ at our old point. This is not too difficult since the two points are connected by a path. The group of all *paths* in Y that project down to loops in X is analogous to G, whereas the group $\pi_1(Y, y)$ of loops is analogous to the subgroup H. Let the path from y' to y be called $\tilde{\tau}$. We have that $y' = \tilde{\tau}(0)$ and $y = \tilde{\tau}(1)$. $\tilde{\tau}^{-1}$ is a path back to y'. Also, $p(y) = p(y') = x$. Now suppose $[\tilde{\sigma}]$ is a class of loops at y such that $p_*[\tilde{\sigma}] \in \pi_1(X, x)$ (which will be the case since $p(y) = x$ and p_* is an injection of $\pi_1(Y, y)$ into $\pi_1(X, x)$). Form the loop $\tilde{\tau} \cdot \tilde{\sigma} \cdot \tilde{\tau}^{-1}$ This loop starts and ends at y' and is, therefore, an element of $\pi_1(Y, y')$. Since y and y' are path connected, the fundamental groups at each point will be the same (recall there is an isomorphism between them given by $\tau_{\#}$.), thus any element of $\pi_1(Y, y')$ can be written in this way. An arbitrary element of $\pi_1(Y, y')$ is given by $\tilde{\tau} \cdot \tilde{\sigma} \cdot \tilde{\tau}^{-1}$. Hence,

$$\pi_1(Y, y') = \tilde{\tau} \cdot \pi_1(Y, y) \cdot \tilde{\tau}^{-1} = \tilde{\tau}_{\#}(\pi_1(Y, y')).$$

We have shown that $\tilde{\tau} \cdot \pi_1(Y, y) \cdot \tilde{\tau}^{-1}$ is a conjugate subgroup of $\pi_1(Y, y')$. Since p_* is an injection, the images of these groups will also form c onjugate subgroups.

We want to show further that, if $\tilde{\tau}$ is a path from y' to y, and $\tau = p \circ \tilde{\tau}$, the image of the group at y' will, in fact, be an invariant subgroup of the image of the group at

y. We can project down to X by

$$\begin{aligned} p_*([\tilde{\tau}] \cdot [\tilde{\sigma}] \cdot [\tilde{\tau}^{-1}]) &= p_*([\tilde{\sigma}]) \\ [\tau] \cdot [\sigma] \cdot [\tau^{-1}] &= [\sigma] \\ [\tau] \cdot p_*(\pi_1(Y,y)) \cdot [\tau^{-1}] &= p_*(\pi_1(Y,y'). \end{aligned} \tag{D.7}$$

Since p_* is an injection, $\tau \cdot \pi_1(X,x) \cdot \tau^{-1}$ is a conjugate subgroup of $\pi_1(X,x)$.

□

Consider what the proposition means. It says that the paths between any two points in the fibre above x will project down to loops in X via the covering map and will be elements of the fundamental group of X. Those paths in Y which form loops are elements of $\pi_1(Y,y)$ and will project down to loops in X forming a subgroup of $\pi_1(X,x)$. Each sheet of the covering space corresponds to a different conjugate subgroup of $\pi_1(X,x)$. The number of possible conjugate subgroups determines the number of sheets in the cover since each element will lift from one sheet to the other. Further, if we take one of these group elements that move you along the fibre and apply it to the entire $\pi_1(X,x)$, it will give us back $\pi_1(X,x)$, but the sheets will be permuted. Thus, we can think of these elements as the group of automorphisms of the covering space.

We will now define a group action that will move us between points in a fibre in the covering space above a point in the base space.

Given a covering, $p : Y \to X$, for any point y with $p(y) = x$, and any loop σ at x, we defined $y * \sigma$ to be the end point of the lift of σ ($\tilde{\sigma}$) that starts at y. This point, $y * \sigma$, is also in $p^{-1}(x)$. (Recall that $p^{-1}(x)$ is a set of points, one on each sheet in the cover.) If σ' is a loop that is homotopic to σ then $y * \sigma' = y * \sigma$. For any homotopy class, $[\sigma]$, in $\pi_1(X,x)$, we can, therefore, define $y * [\sigma]$ to be $y * \sigma$. This definition gives a right action of the fundamental group, $\pi_1(X,x)$, on the fibre $p^{-1}(x)$:

$$p^{-1}(x) \times \pi_1(X,x) \to p^{-1}(x), \quad y \times [\sigma] \mapsto y * [\sigma],$$

taking $y \times [\sigma]$ to the endpoint of the lift of σ that starts at y.

The fundamental group in the base space moves you along the fibre in the covering space. To visualize this, consider the covering of S^1, $p : \mathbb{R} \to S^1$ given by $p(r) \mapsto (\cos(2\pi r), \sin(2\pi r))$. This is an infinite helix of $\mathbb{R}$ coiling upwards over S^1. The fundamental group of the covering space, $\mathbb{R}$, is trivial, whereas the fundamental group of S^1 is not. If one takes a path from a point r to a point $r + 2\pi$ in the covering space, both endpoints map down to the same point in S^1. This path defines the element, $\{1\}$, in $\pi_1(S^1, x)$. We call this element "1" since it loops once around S^1. In this way, we would call the identity element "0". Similarly, the path from r to $r + 4\pi$ would project down to $\{2\}$ in the base space. We see that $r \times [1] = r * [1] = r + 2\pi$ since it is the end point of the path given by the lift of $[1]$ starting at r, which is $r + 2\pi$. We then have $r \times [2] = r + 4\pi$ since the lift of the group element, $[2]$, is a path from r to $r + 2\pi$. We see how this map really does move between fibres. Given a point in the fibre, $p^{-1}(x)$, we can apply all possible group elements and move to a unique new point in the fibre for each one.

In this example, the covering space is trivial in that its fundamental group is the identity. Therefore, $p_*(\pi_1(Y, y) = \{\epsilon_x\}$, and the only conjugate group is the entire group $\pi_1(X, x)$. Every element of $\pi_1(X, x)$ will move you along a fibre and will also act as an automorphism of the covering space if applied to the rest of $\pi_1(X, x)$.

The fundamental group is useful for finding maps between spaces. If there is no map between the fundamental groups on the spaces, then there is no map between the spaces. We will now show that, if there is a map between the fundamental groups, then there is also a map between the spaces. If two coverings of a space X give conjugate subgroups of the fundamental group of X, then the coverings are isomorphic. They will differ only by a permutation of the sheets corresponding to the group element which makes them conjugate. We can decide whether two spaces are isomorphic by testing whether the images of their fundamental groups are conjugate subgroups. If the images of the fundamental groups are the same, then there is an isomorphism between coverings that also preserves the basepoints.

Proposition: Suppose $p : Y \to X$ is a covering and $f : Z \to X$ is a continuous mapping with Z a connected and locally path-connected space. Let $x \in X$, $y \in Y$, and $z \in Z$ be points with $p(y) = f(z) = x$. For there to be a continuous mapping, $\tilde{f} : Z \to Y$, with $p \circ \tilde{f} = f$ and $\tilde{f}(z) = y$, it is necessary and sufficient that $f_*(\pi_1(Z, z))$ be contained in $p_*(\pi_1(Y, y))$. Such a lifting, $\tilde{f}$, when it exists, is unique.

Visualization: Let $f_*(\pi_1(Z, z))$ be a bigger subgroup of $\pi_1(X, x)$ than $p_*(\pi_1(Y, y))$. Then, we cannot map to Y and then down to X since the most we can get is the elements of $p_*(\pi_1(Y, y))$. When we map a space with a bigger fundamental group to one with a smaller fundamental group, there is nowhere to put the extra group elements. They would be mapped to elements in a many-to-one fashion. Mapping from the second space down to the base space would not distinguish those elements that got projected out in the first mapping. This proposition says that *if we want to map Z to Y so that the projection to the base space is preserved, Z cannot have a bigger fundamental group than Y*. For example, the real line is a cover of S^1 which has one element in its fundamental group: the identity. (It is a trivial or *universal* cover.) On the other hand, the double cover of S^1 has $\mathbb{Z}$ elements in it. Mapping the double cover of S^1 to the real line will project out elements of the fundamental group. (They will all map to the identity.) If we then mapped to the base space, S^1, the net result would be the entire fundamental group of the double cover mapping to the identity in S^1. Mapping straight to S^1 from the double cover has $\pi_1(Z, z) = \mathbb{Z}$ mapping to the $\mathbb{Z}/\mathbb{Z}_2$ subgroup of $\pi_1(S^1, z) = \mathbb{Z}$. We could map from the real line to the double cover and then down to S^1 since the fundamental group of the real line maps to the identity and nothing is lost.

Proof: The necessity is clear from our visualization (which is a result of the fundamental group being a "functor"). The uniqueness is given by the following lemma.

Lemma: Let $p : Y \to X$ be a covering, and let Z be a connected topological space. Suppose $\tilde{f}_1$ and $\tilde{f}_2$ are continuous mappings from Z to Y such that $p \circ \tilde{f}_1 = p \circ \tilde{f}_2$. If $\tilde{f}_1(z) = \tilde{f}_2(z)$ for one point, z, in Z, then $\tilde{f}_1 = \tilde{f}_2$.

Proof: Since we have the condition $p \circ \tilde{f} = f$ on our function, $\tilde{f}$, we are fixing a point in Y for it to map to and, hence, makes it unique. Now to show that $f_*(\pi_1(Z,z))$ being contained in $p_*(\pi_1(Y,y))$ is enough to guarantee a mapping, $\tilde{f}$ (the sufficient condition), we construct $\tilde{f}$.

Given a point w in Z, choose a path, γ, in Z from z to w, and let $\sigma = f \circ \gamma$. σ is a path starting at x in X. Define $\tilde{f}(w)$ to be $y * \sigma$ so that $\tilde{f}(w)$ is the endpoint of the lifted σ which starts at y. $\tilde{f}(w)$ defines our continuous map from Z to Y which has the required property that

$$p \circ \tilde{f}(w) = p \circ (y * \sigma) = \sigma(1) = (f \circ \gamma)(1) = f(\gamma(1)) = f(w).$$

To make sure that we have a valid map, we must show that it is independent of our choice of γ and that it is continuous at w. Let γ' be another path from z to w; then, $f \circ (\gamma' \cdot \gamma^{-1}) = \sigma' \cdot \sigma^{-1}$, but since $[\sigma' \cdot \sigma^{-1}]$ is in the image of p_* (because we are saying that $f_*(\pi_1(Z,z))$ is contained in $p_*(\pi_1(Y,y))$ and $\gamma' \cdot \gamma^{-1} \in \pi_1(Z,z)$, then it comes from a group element in $\pi_1(Y,y)$, a loop, thus $\tilde{\sigma}'$ and $\tilde{\sigma}$ end at the same point in Y. To show that $\tilde{f}$ is continuous at w, we have to show that, if V is an open neighborhood of $\tilde{f}(w)$, we have $\tilde{f}^{-1}(V)$ open. Let N be any neighborhood of $f(w)$ that is evenly covered by p. Let V be the open set in $p^{-1}(N)$ that maps homeomorphically onto N (by covering mappings) that contains $\tilde{f}(w)$ (by $p \circ \tilde{f} = f$). Choose a path-connected neighborhood, U, of w so that $f(U) \subset N$ (using the fact that f is continuous to X). We need to show that $\tilde{f}$ maps U into V. For all points, w', in U, we may find a path, α, from w to w' in U. We can use $\gamma \cdot \alpha$ as the path from z to w'. The lifting of $f \circ (\gamma \cdot \alpha) = (f \circ \gamma) \cdot (f \circ \alpha)$ is obtained by first lifting $f \circ \gamma$ to $\tilde{\sigma}$ and then lifting $f \circ \alpha$. Since the latter lifting stays in V, we have $\tilde{f}(U) \subset V$.

□

Corollary: Let X be a connected and locally path-connected space. Let $p : Y \to X$ and $p' : Y' \to X$ be two covering maps with Y and Y' connected. Let $p(y) = x$ and $p'(y') = x$. In order for there to be an isomorphism between the coverings preserving the basepoints, it is necessary and sufficient that

$$p_*(\pi_1(Y,y)) = p'_*(\pi_1(Y',y')).$$

Proof: The fact that it is necessary is clear since the spaces cannot be isomorphic if they have different fundamental groups. (They would have a different number of holes.) The sufficient part comes by assuming that these two subgroups agree. The proposition then works both ways, with $p_*(\pi_1(Y,y)) \subset p'_*(\pi_1(Y',y'))$ and $p'_*(\pi_1(Y',y')) \subset p_*(\pi_1(Y,y))$ giving two maps, $\phi : Y \to Y'$ and $\psi : Y' \to Y$. Apply the lemma with $\tilde{f}_1 = \psi \circ \phi$ and $\tilde{f}_2 = \phi \circ \psi$. Since they agree, Y and Y' are isomorphic.

The only time that the basepoints are not preserved is when the fundamental groups have images that are conjugate subgroups rather than the same subgroup.

□

Corollary: A simply connected and locally path-connected space has only trivial coverings.

Proof: The proof is an obvious result from the proposition. For it to be a proper covering, it would have its fundamental group map to a subgroup of the fundamental group on the simply connected base space (trivial fundamental group); therefore, $p_*(\pi_1(Y,y)) = p_*(\pi_1(E,x))$, where E is the trivial covering space. Hence, by the last corollary, there is an isomorphism between them.

□

The fundamental group of a base space determines an automorphism (an isomorphism of a group onto itself) group of the covering space. We have shown how each different fundamental group element in $\pi_1(X,x)/p_*(\pi_1(Y,y))$ will map to a different sheet in the covering space. We can build another group from this one which will map between the sheets in the covering given by projecting down to the base space and then up to the new sheet.

We state two theorems and refer the reader to [85] for proof.

Theorem: Let $p: Y \to X$ be a covering, with Y connected and X locally path-connected, and let $p(y) = x$. If $p_*(\pi_1(Y,y))$ is a normal subgroup of $\pi_1(X,x)$, then there is a canonical isomorphism

$$\pi_1(X,x)/p_*(\pi_1(Y,y)) \xrightarrow{\cong} Aut(Y/X).$$

The covering is a G-covering, with G being the quotient group, $\pi_1(X,x)/p_*(\pi_1(Y,y))$.

Theorem: The following are equivalent:

(a) The covering is regular

(b) The action of $Aut(Y/X)$ on p^{-1} is transitive

(c) For every loop, σ, at x, if one lifting of σ is closed, then all liftings are closed.

These last statements result from the fact that, if the covering is regular, we have removed all the elements of the group which do not lift us to a new sheet. The other elements are unique (if $[\sigma] \cdot y = [\sigma'] \cdot y \Rightarrow [\sigma] = [\sigma']$), so the action is transitive on the fibre. If a lifting is closed, then it is in $p_*(\pi_1(Y,y))$ and cannot lift to an open path in another sheet. Hence, all liftings will be closed.

Notice that if Y is a simply connected space and if X is a locally path-connected space, we have $\pi_1(X,x) \cong Aut(Y/X)$. This is because, if Y is simply connected, it has a trivial fundamental group and all of $\pi_1(X,x)$ will be in the quotient group.

If a group G acts evenly on a simply connected and locally path-connected space Y, and $X = Y/G$ is the orbit space, then the fundamental group of X is isomorphic to G. This is because G must then be isomorphic to $Aut(Y/X)$, and if we choose a point y in Y over a point x in X, then $\pi_1(X,x)$ is also isomorphic to $Aut(Y/X) = G$.

The *universal cover* of a space is a simply connected cover and the automorphism group is isomorphic to the fundamental group of the base space. If we find an abelian automorphism group of the universal cover we have found the fundamental group of the base space (recall that an automorphism maps disconnected open sets to disconnected open sets).

Proposition:

(a) If the Klein bottle is constructed by identifying the sides of a rectangle in the usual fashion (See Section D.9.4.), the fundamental group has two generators, a and b, with one relation, $abab^{-1} = e$. The fundamental group is not abelian.

(b) The torus is a two-sheeted covering of the Klein bottle given by $\mathbb{R}^2/H$, where H is the two-element group, $(x, y) \mapsto (x+1, y)$ and $(x, y) \mapsto (x, y+1)$. (This is an obvious generalization of the covering of S^1.) The fundamental group of the torus is a normal subgroup of the fundamental group of the Klein bottle.

Proof:

(a) Given any point, x, in the Klein bottle, there are two different closed loops that can be made as can be seen by the usual diagram of the Klein bottle as a square with two oposite edges identified and the other two identified with a twist. The closed loops are vertical or horizontal lines on the diagram. The path traversing the entire edge is homotopic to the identity and gives our relation. The relation is $ab = ba^{-1}$, showing that the group is non-abelian.

(b) The typical diagram of the torus made from a square shows that it has two elements in its fundamental group, with $ghg^{-1}h^{-1} = e$, so that it is abelian. Our covering map is mapping both of these group elements to the generators, a and b^2, of the Klein bottle (since twice around b and once around a correspond to g and h, respectively). The remaining generator for the Klein bottle corresponds to going halfway around one of the circles in the torus; i.e., one loop around the Klein bottle maps only halfway around a loop on the torus. Another application of this element maps the rest of the way around the torus. Hence, $\pi_1(X, x)/p_*(\pi_1(Y, y)) = \{e, b\}$ such that $b^2 = e$ on the torus. We have $\mathbb{Z}/\mathbb{Z}_2$. Odd multiples of b map to a new sheet. Even multiples are the identity. To show that this group is a normal subgroup, we show that gHg^{-1} is in the subgroup, where g is in the original group $\{e, a, a^2, ..., b, b^2, ...\}$, and H are $\{e, b^2, b^4, ...\}$ which form $p_*(\pi_1(Y, y))$ in this case. Now, $abab^{-1} = e$ on the Klein bottle, so $b^n b^{2k} b^{-n} = b^n b^{2k-n} = b^{2k}$,

$$
\begin{aligned}
a^n b^{2k} a^{-n} &= aa \cdots abb \cdots ba^{-1}a^{-1} \cdots a^{-1} \\
&= aa \cdots (ba^{-1})b \cdots ba^{-1}a^{-1} \cdots a^{-1} \\
&= aa \cdots b(a^{-1}b)b \cdots ba^{-1}a^{-1} \cdots a^{-1} \\
&= aa \cdots b(ba)b \cdots ba^{-1}a^{-1} \cdots a^{-1} \\
&\vdots \text{ permuting } 2k \text{ times gives } (ab) \to (ba) \\
&= aa \cdots bbb \cdots (ba)a^{-1}a^{-1} \cdots a^{-1} \\
&= aa \cdots bbb \cdots b(e)a^{-1} \cdots a^{-1} \\
&= bbb \cdots b = b^{2k}
\end{aligned}
$$

It is indeed a normal subgroup. Therefore, S^1 is a double cover of the Möbius band by the same reasoning. Hence, **R** is also a cover of the Möbius band (called the universal cover).

□

D.7.1. The universal covering

A covering is called a *universal cover* if Y is simply connected. The fundamental group of the base space is then itself isomorphic to the automorphism group. The universal covering is automatically a G-covering with G being the fundamental group of the base space. The universal covering is unique up to canonical isomorphism since any group which acts evenly on a simply connected and locally path-connected space, with $X = Y/G$ being the orbit space. The fundamental group of X is isomorphic to G. Therefore, any other universal covering will also have its automorphism group isomorphic to G.

Now, we want to find the condition that guarantees the existence of a universal covering. Suppose we have a universal covering, $p : Y \to X$. Any point in X has an evenly covered path-connected neighborhood, N. Any loop, σ, in N (note that it must be in N) lifts to a loop $\tilde{\sigma}$ in Y, and since Y is simply connected, this loop is homotopic in Y to a constant path. It follows that the loop $\sigma = p \circ \tilde{\sigma}$ is homotopic to a constant path in X. A space, X, is called *semilocally simply connected* if every point has a neighborhood such that every loop in the neighborhood is homotopic to a constant path in X. Being semilocally simply connected is a necessary condition for the existence of a universal covering. Note that, if X is locally simply connected, then it is semilocally simply connected. (Locally simply connected means that every neighborhood of every point contains a simply connected neighborhood.)

For example the clam shell (the union of all circles, C_n, of radius $1/2^n$ centered at $(1/2^n, 0)$ in the plane) is connected, locally path connected, but not semilocally simply connected. In fact, let X be the cone over C, meaning every circle becomes the base of a different cone which goes up to the same point $(0,0,1)$. Then, X is now semilocally simply connected (and, therefore, has a universal covering) but not locally simply connected. The point, $(0,0,0)$, has a simply connected neighborhood, but not every neighborhood contains a simply connected neighborhood. (Any neighborhood in the plane does not have one.)

Suppose we have a universal covering with $p(y) = x$. For any point z in Y, there is a path, γ, from y to z which is unique up to homotopy since Y is simply connected. The image, $\alpha = p \circ \gamma$, is a path from x to $p(z)$ which is unique up to homotopy (with fixed endpoints). Conversely, a path, α, in X starting at x determines a point, $z = y * \alpha$, in Y. This observation identifies Y, as a set, with the set of homotopy classes of paths in X that start at a given point, x. We can use these homotopy classes to construct the universal covering. First note that the universal covering must be simply connected by definition. The projection of its fundamental group on the fundamental group of the base space must form the identity element. We know that the normal subgroups of the fundamental group of the base space define all possible coverings that the space may have. Then, the entire fundamental group is the automorphism group of the universal covering space. At any point in the fibre over x, we can apply any of the elements of the fundamental group to get another point in the fibre over x. Hence, constructing the universal covering is as simple as drawing all of these. For example, to find the universal cover of S^1, we simply start with a point $y = p^{-1}(x)$, and since there is one generator of the group, we apply it and its inverse. From each of these new points, we do the same thing again. In this fashion, we get the helical covering $\mathbb{R}$. Similarly, for a base space consisting of two circles joined at a point x, we get four paths leaving the point, $y = p^{-1}(x)$. These paths correspond to each of the generators and their inverses. At the new point, we draw four more lines leaving it (including the one we just came from, corresponding to the inverse of the group element or some multiple of it).

There is a correspondence between subgroups of the fundamental group and the coverings that a space has. Assume that X is connected, locally path-connected, and semilocally simply connected so that X has a universal covering.

Proposition:

(a) For every subgroup H of $\pi_1(X, x)$, there is a connected covering, $p_H : Y_H \to X$, with a basepoint, $y_H \in p_H^{-1}(x)$, so that the image of $\pi_1(Y_H, y_H)$ in $\pi_1(X, x)$ is H. Any other such covering (with the same basepoint) is canonically isomorphic to this one.

(b) If K is another subgroup of $\pi_1(X, x)$ containing H, there is a unique continuous mapping, $p_{H,K} : Y_H \to Y_K$, that maps y_H to y_K and is compatible with the projections to X. This mapping is a covering mapping, and if H is a normal subgroup of K, it is a G-covering with $G = K/H$.

Proof: Let $u : \tilde{X} \to X$ be a universal covering with $u(\tilde{x}) = x$ and identify $\pi_1(X, x)$ with $Aut(\tilde{X}/X)$. Any subgroup H of $\pi_1(X, x)$ acts evenly on $\tilde{X}$. The quotient, $\tilde{X} \to \tilde{X}/H = Y_H$, makes $\tilde{X}$ the universal covering of Y_H. For example, since $\mathbb{R}^2$ is the universal covering space of the torus and the torus is a double cover of the Klein bottle, then $\mathbb{R}^2$ is also the universal cover of the Klein bottle. The fundamental group, Y_H (with basepoint y_H as the image of $\tilde{x}$), is canonically isomorphic to H. The projection from Y_H to $\tilde{X}/\pi_1(X, x) = X$ is a covering. The image of fundamental group of this covering in $\pi_1(X, x)$ is H. If H is contained in K, there is a canonical map on the orbit spaces, $\tilde{X}/H \to \tilde{X}/K$.

□

The covering, $Y_H \to X$, corresponding to H can be identified with $\tilde{X}/H \to X$, where H acts on $\tilde{X}$ as a subgroup of $Aut(\tilde{X}/X) = \pi_1(X, x)$. We can form Y_H by simply starting with the universal covering and identifying all points that differ from each other by an application of an element of H. The regular coverings correspond to normal subgroups, H. Every connected G-covering, $p : Y \to X$, has the form $\tilde{X}/H \to X$, with

$$\pi_1(X, x)/H \cong Aut(Y/X) \cong G.$$

For an example, if we take the universal covering of the Klein bottle, ($\mathbb{R}^2$), and identify all points differing by an application of the subgroup of the generators given by a and b^2 (where a is the twisted one), we get the torus. Also notice that the subgroup, a^2, b^2, of the generators of the Klein bottle also gives the torus as a quotient space of the universal cover. It is a quadruple cover since the torus "squares" in the universal covering space now cover four of the original Klein bottle "squares." We have found the covering space, Y_H, which is the double cover of the Klein bottle given by the torus corresponding to the subgroup generated by $\{a, b^2\}$.

This correspondence is similar to that seen in Galois theory where subgroups correspond to field extensions and smaller subgroups corresponding to larger extensions. Here, smaller subgroups of the fundamental group correspond to larger coverings of the space:

$$\begin{array}{cc} \tilde{X}, \tilde{x} & \{e\} \\ \downarrow & \cap \\ Y_H, y_H & H \\ \downarrow & \cap \\ Y_K, y_K & K \\ \downarrow & \cap \\ X, x & G \end{array} \tag{D.8}$$

If H is a normal subgroup of K, then the covering, $Y_H \to Y_K$, is a K/H-covering.

For example, if $X = S^1$ is a circle, the coverings correspond to subgroups of $\pi_1(S^1, (1,0)) = \mathbb{Z}$. The trivial group corresponds to the universal covering, $\mathbb{R} \to S^1$. The subgroups, $\mathbb{Z}_n \subset \mathbb{Z}$, correspond to the n-sheeted covering, $p_n : S^1 \to S^1$. Up to isomorphism, these coverings are the only connected coverings of a circle.

The covering corresponding to the commutator subgroup of the fundamental group (The commutators of group elements can easily be shown to be a subgroup of the fundamental group, hence there is a covering corresponding to them.) of X is a covering we may denote by $\tilde{X}_{\text{abel}}$:

$$\tilde{X}_{\text{abel}} = \tilde{X}/[\pi_1(X,x), \pi_1(X,x)] \to X.$$

$\tilde{X}_{\text{abel}}$ is a G-covering with $G = \tilde{X}/[\pi_1(X,x), \pi_1(X,x)]$ and is the first homology group, $H_1(X)$. It is sometimes called the *universal abelian covering* of X.

The universal covering of the complement of the origin in the plane can be realized as the right half plane, via the polar coordinate mapping $(r, \theta) \mapsto (r\cos(\theta), r\sin(\theta))$, and as the entire complex plane, $\mathbb{C}$, via the mapping $z \mapsto \exp(z)$. We will find an isomorphism between these coverings.

This map indeed covers the entire plane (minus the origin) since the negative y-axis gets wrapped around counterclockwise to the negative x-axis and the positive y-axis goes there, too. The other map does precisely the same thing to the right half of $\mathbb{C}$ also as can be seen by $-i\pi \mapsto \exp(-i\pi) = -1$ and $i\pi \mapsto -1$. An isomorphism is given by $(r, \theta) \mapsto \log(r) + i\theta$ since $\exp(\log(r) + i\theta) = r\ \exp(i\theta)$.

D.8. General homology and homotopy

In this section, we will discuss the Siefert-Van Kampen theorem and how it can be used to construct the fundamental group for a union of two spaces by a knowledge of the fundamental groups on the two spaces. This theorem is quite important since it allows us to easily say things like: The fundamental group of n circles joined at a point is the free group[3] on n generators. It allows us to find the fundamental group for a sphere with g handles and other such surfaces by merely knowing the fundamental group of the sphere with one handle (the torus).

Suppose the base space X is a union of two opens sets, U and V. A covering of X restricts to coverings of U and of V which are isomorphic over $U \cap V$. The isomorphism is given by first projecting down to the intersection using one covering map and then projecting back up using the other.

We will show the generalization of this: Suppose we have a collection, X_α, of open sets, $\alpha \in \mathcal{A}$, the union of which is the base space X. For each α, we have a covering map, $p_\alpha : Y_\alpha \to X_\alpha$. Then, for each α and β, we have an isomorphism

$$\theta_{\alpha\beta} : p_\alpha^{-1}(X_\alpha \cap X_\beta) \to p_\beta^{-1}(X_\alpha \cap X_\beta)$$

of coverings of $X_\alpha \cap X_\beta$. Assume these coverings are compatible, i.e.,

(1) $\theta_{\alpha\alpha}$ is the identity on Y_α; and

(2) $\theta_{\gamma\alpha} = \theta_{\gamma\beta} \circ \theta_{\beta\alpha}$ on $p_\alpha^{-1}(X_\alpha \cap X_\beta \cap X_\gamma)$ for all α, β, γ in $\mathcal{A}$.

Then, we can patch these coverings together to obtain a covering $p : Y \to X$. We have isomorphisms, $\phi_\alpha : Y_\alpha \to p^{-1}(X_\alpha)$, of coverings of X_α such that $\theta_{\beta\alpha} = \phi_\beta^{-1} \circ \phi_\alpha$ on $p_\alpha^{-1}(X_\alpha \cap X_\beta)$. In addition, the space Y is the union of the open sets, $\phi_\alpha(Y_\alpha)$.

If each $p_\alpha : Y_\alpha \to X_\alpha$ is a G-covering with fixed G and each $\theta_{\beta\alpha}$ is an isomorphism of G-coverings, then there is a unique action of G on Y so that each ϕ_α commutes with the action of G, i.e., $\phi_\alpha(g \cdot y_\alpha) = g \cdot \phi_\alpha(y_\alpha)$ for $g \in G$ and $y_\alpha \in Y_\alpha$. The patched covering, $p : Y \to X$, then has the structure of a G-covering so that each ϕ_α is an isomorphism of G-coverings.

Let X be a space that is a union of two open subspaces, U and V. Assume that each of the spaces, U, V, and their intersection, $U \cap V$, is path-connected, and let x be a point in the intersection. Assume also that all these spaces X, U, V, and $U \cap V$ have universal covering spaces. We have a commutative diagram of homomorphisms of fundamental groups:

$$\begin{array}{ccc} \pi_1(U \cap V, x) & \xrightarrow{i_1} & \pi_1(U, x) \\ {\scriptstyle i_2}\downarrow & & \downarrow{\scriptstyle j_1} \\ \pi_1(V, x) & \underset{j_2}{\longrightarrow} & \pi_1(X, x). \end{array} \tag{D.9}$$

The maps are induced by the inclusions of subspaces. Commutativity means that $j_1 \circ i_1 = j_2 \circ i_2$.

We will describe how $\pi_1(X, x)$ is determined by the other groups and the maps between them. The description is not direct but through a universal property. Any homomorphism, h, from $\pi_1(X, x)$ to a group G determines a pair of homomorphisms.

[3]The free group on two generators, a and b, is given by the set of all words of the form $a^{n_0} \cdot b^{n_1} \cdot a^{n_2} \cdots b^{n_r}$, where the n_i are integer powers which are all non-zero except perhaps the first and last. The identity element is $e = a^0 b^0$. The product is formed by simply placing two words of this form next to each other and canceling factors as necessary. The other free groups are constructed similarly.

$h_1 = h \circ j_1$ from $\pi_1(U,x)$ to G and $h_2 = h \circ j_2$ from $\pi_1(V,x)$ to G; the two homomorphisms, $h_1 \circ i_1$ and $h_2 \circ i_2$ from $\pi_1(U \cap V, x)$ to G, determined determined in this way are the same. The Van Kampen theorem says that $\pi_1(X,x)$ is the universal group with this property.

Siefert-Van Kampen Theorem: For any homomorphisms $h_1 : \pi_1(U,x) \to G$ and $h_2 : \pi_1(V,x) \to G$, such that $h_1 \circ i_1 = h_2 \circ i_2$ (They agree on the intersection.), there is a unique homomorphism, $h : \pi_1(X,x) \to G$, such that $h \circ j_1 = h_1$ and $h \circ j_2 = h_2$. (It agrees on the intersection.)

Proof: The homomorphisms, h_1 and h_2, determine G-coverings, $Y_1 \to U$ and $Y_2 \to V$, together with base points y_1 and y_2 over x. The fact that $h_1 \circ i_1$ is equal to $h_2 \circ i_2$ means that the restrictions of these coverings to the intersection are isomorphic G-coverings. There is then a unique isomorphism between these G-coverings that maps y_1 to y_2. Using our patching method, we can patch these two coverings together and use this patched isomorphism on the intersection to give a G-covering, $Y \to X$, that restricts to the two given G-coverings (and has the same basepoint). This G-covering corresponds to a homomorphism, h, from $\pi_1(X,x)$ to G, and the fact that the restricted coverings agree means precisely that $h \circ j_1 = h_1$ and $h \circ j_2 = h_2$.

□

The group $\pi_1(X,x)$ is generated by the images of $\pi_1(U,x)$ and $\pi_1(V,x)$ in G. The implication is that if, say, a hole is shared between U and V, then the group elements will overlap when homomorphically mapped to G, and they will together count as only a single element of $\pi_1(X,x)$.

Corollary: If U and V are simply connected, then X is simply connected. Notice that one of the hypotheses of the Van Kampen theorem is that the intersection is connected! There is no forming an annulus from two elongated disks.

Corollary: If $U \cap V$ is simply connected, then, for any G,

$$Hom(\pi_1(X,x),G) = Hom(\pi_1(U,x),G) \times Hom(\pi_1(V,x),G).$$

$\pi_1(X,x)$ is then the free product of $\pi_1(U,x)$ and $\pi_1(V,x)$.

An example would be the case where we have a torus with $\pi_1(U,x)$ being the free abelian group on two generators $\{a,b\}$ with relation $aba^{-1}b^{-1} = e$ and another torus with $\pi_1(V,x)$ being another free abelian group $\{c,d\}$ with a similar abelian relation. The space, X, made from the union of the two with simply connected intersection has $\pi_1(X,x)$ being the free product of the two. Thus, it has generators, $\{a,b,c,d\}$, with the relation $aba^{-1}b^{-1}cdc^{-1}d^{-1} = e$ and is, therefore, non-abelian.

The following generalization can be used to compute the fundamental group of an increasing union of spaces, each of which having a known fundamental group.

Theorem: Suppose a space X is a union of a family of open subspaces, X_α, with the property that the intersection of any two of these subspaces is in the family. Assume that X and each X_α are path-connected and have a universal covering, and that the intersection of all the X_α contains a point x. When X_β is contained in X_α, let $i_{\alpha\beta}$ be the map from $\pi_1(X_\beta,x)$ to $\pi_1(X_\alpha,x)$ determined by the inclusion, and let j_α be the map from $\pi_1(X_\alpha,x)$ to $\pi_1(X,x)$ determined by inclusion.

With these hypotheses, $\pi_1(X,x)$ is the direct limit of the groups, $\pi_1(X_\alpha,x)$. That is, for any group G and any collection of homomorphisms, h_α, from $\pi_1(X_\alpha,x)$ to G such that $h_\beta = h_\alpha \circ i_{\alpha\beta}$ whenever $X_\beta \subset X_\alpha$, there is a unique homomorphism, h, from $\pi_1(X,x)$ to G such that $h_\alpha = h \circ j_\alpha$ for all α.

□

To understand this theorem, we will describe the fundamental group of $\mathbb{R}^2 \setminus \mathbb{Z}$. This space is the plane with a point removed at each of the integers on the x-axis. We want to use the theorem to find the fundamental group, so we construct open sets given by the strips

$$(x,y) = \begin{cases} |z| - \frac{1}{2} < x < \frac{1}{2} & \text{if } z < 0 \\ -\frac{1}{2} < x < \frac{1}{2} & \text{if } z = 0 \\ -\frac{1}{2} < x < |z| + \frac{1}{2} & \text{if } z > 0 \end{cases},$$

where z is an integer. The union of these strips covers $\mathbb{R}^2$ and the intersection of any two strips is in this family of strips since we have made the intersection of any two of them equal to either the smaller of the two or the strip about zero. These strips correspond to the subspaces, X_α, from the theorem. The zeroth strip is given by $(-\frac{1}{2}, +\frac{1}{2}) \times (-\infty, \infty)$ and has one point removed from it which is the origin, $z = 0$. Now, the map, $j_0 : \pi_1(X_0,x) \to \pi_1(X,x)$, is just the inclusion of the fundamental group in the zeroth strip in the total. This fundamental group is the free group on one generator. The fundamental group of the entire space is the direct limit of the fundamental groups of the subspaces which is $\mathbb{F}_{\aleph_0}$, the free group on countably infinite generators.

Example 1:

The n-sphere is simply connected if $n \geq 2$. To see this, write it as the union of two hemispheres homeomorphic to disks. The intersection is then connected (it is S^{n-1}). Therefore, the corollary, tells us that the n-sphere is simply connected. (We needed $n \geq 2$ so that the intersection would be connected.)

Example 2:

Consider a figure eight as the union, X, of two circles, U and V, meeting at a point x. Let γ_1 and γ_2 be loops, one around each circle. The fundamental group of each circle is infinite cyclic, generated by the classes of loops. It follows that to give a homomorphism from $\pi_1(X,x)$ to a group G is the same as specifying arbitrary elements g_1 and g_2 in G. There is a unique homomorphism from $\pi_1(X,x)$ to G mapping $[\gamma_1]$ to g_1 and $[\gamma_2]$ to g_2. $\pi_1(X,x)$ is the free group on the generators, $[\gamma_1]$ and $[\gamma_2]$.

Let $a = [\gamma_1]$ and $b = [\gamma_2]$. We will show that every element in $\pi_1(X,x)$ has a unique expression in the form

$$a^{m_0} \cdot b^{m_1} \cdot a^{m_2} \cdots b^{m_r},$$

as per the definition of the free group on two generators.

Every element in $\pi_1(X,x)$ will map the point x to a unique point y in the universal covering space of X via a universal covering, $p : \tilde{X} \to X$. All of the equivalence classes of paths in X will map onto $\tilde{X}$ Since the universal covering space is simply connected, the group acts evenly on it. (It is the space with two choices of path at each element of p^x.) Therefore, if $a \cdot x = b \cdot x$, then $a = b$. That the expression is of the proper form follows from the fact that any point in $p^{-1}(x)$ will consist of products of loops around a or b.

Example 3:

If G is a free group on n generators and H is a subgroup of G that has finite index, d, in G, then H is a free group with $dn - d + 1$ generators. This result can be seen as follows. Take G to be the fundamental group of a connected graph X that has v vertices and e edges, with $n = e - v + 1$ (number of faces). For simplicity, we assume each edge of X connects two distinct vertices. The subgroup, H, corresponds to a connected covering, $p : Y \to X$, with d sheets and with the fundamental group of Y isomorphic to H. Y is a connected graph since the d points over each of the vertices of X can be taken as the vertices of Y and the inverse images of the edges are the edges of Y. Since Y is a graph, the fundamental group is free (same as n circles joined at a point on each of the d sheets) with

$$de - dv + 1 = d(e - v + 1) - d + 1 = dn - d + 1$$

generators.

Example 4:

We will construct the universal covering of n circles joined at a point and compute the fundamental group. The universal covering will look very similar to the universal covering of the lemniscate, or figure eight. It will be an infinite tree or connected graph, but instead of having four choices at each vertex, we have $2n$ choices at each vertex. Starting at the center, we have $2n$ paths from origin, one forward path and one backward path for each circle. After traversing any path, we come to another vertex (which is mapped to the first one via the covering map), and we again have $2n$ choices, including the path backward.

A corollary to the Siefert-Van Kampen theorem is as follows: *If $X = U \cup V$ is a path connected space and $U \cap V$ is simply connected, then $\pi_1(X, x) = \pi_1(U, x) * \pi_1(V, x)$.*

Therefore, the fundamental group of n circles joined at a point is simply the direct product (all words formed from the elements) of the fundamental groups each circle individually. Note that the union of the circles is a single point and is simply connected since there exists a neighborhood of it which is simply connected. Therefore, $\pi_1(\text{n circles}, x) = \mathbb{F}_n$. We are using an induction argument with $X = U_1 \cup \cdots \cup U_n$, writing $\pi_1(X, x) = \pi_1([U_1 \cup \cdots \cup U_{n-1}], x) * \pi_1(U_n, x)$ and then applying the corollary again, etc.

Example 5:

We will use the Van Kampen theorem to compute the fundamental group of

(a) The complement of n points (or small disks) in a 2-sphere.

(b) The complement of n points in a Torus.

(c) The sphere with g handles.

(a) The complement of one point in a 2-sphere is the open disk. Therefore, the fundamental group is trivial. The complement of two points is the same as the complement of one point in the disk, making the fundamental group equal to the free abelian group on one generator, a, with $aa^{-1} = e$ and so it is then $\mathbb{Z}$. We can use the Van Kampen theorem to find the group of the complement of three points in the sphere from two copies of the complement of two points in the sphere. We merely think about each of them as the complements of one point in the disk. So now we have two disks. We join them so that the overlap is simply connected. The result is a disk with two holes removed or a sphere with three holes removed. The Van Kampen theorem tells us that the fundamental group will be the free product of the fundamental groups of the two original disks. Namely, the fundamental group is the free group on two generators. We can extend this result to n points by continuing the process so that, if we have the complement of k points, we generate the complement of $k+1$ points by forming the union of the disk with $k-1$ points removed and the disk with one point removed with the intersection being simply connected. We get a disk with k points removed or, equivalently, a sphere with $k+1$ points removed. We see that the fundamental group of the complement of n points is the free group on $n-1$ generators.

(b) The complement of n points in a torus can be derived in the same way as the sphere. First, we notice that the torus has the same fundamental group as the complement of two points in the open disk, the free abelian group on two generators. If we remove a point from the torus, we are left with a surface that has a figure eight as a deformation retract. This is identical to a disk with two points removed. Now, we form the union with disks having one point removed in the exact same fashion as we did for the sphere. The result is that the complement of two points in the torus is the same as the complement of three points in a disk, the free group on three generators. As before, we continue in this fashion until we find that the complement of n points in the torus is the free group on $n+1$ generators.

(c) A sphere with one handle is a torus. A sphere with g handles is the same as g tori joined (say each joined to another at a point so that the intersection is simply connected). The group is, therefore, the free product of all of them, $\mathbb{Z}^{2g}$.

Example 6:

The free group on two generators is the fundamental group of the complement of a point in the torus. The complement of a point in a torus has a figure eight as a retract. Since a covering of the figure eight is a grid covering, $\mathbb{R}^2$, with lines vertical and horizontal, intersecting at each of the integer coordinates the same covering will cover the torus minus a point if you "thicken" the lines and make them 2-dimensional strips.

This covering is the same as the plane minus a countably infinite number of points corresponding to the points at the center of the squares. The fundamental group of this covering space is the free group on a countably infinite number of generators. Now, since this covering is a covering of the torus minus a point, there is a homomorphism given by p_* which maps its fundamental group onto a normal subgroup of the fundamental group of the torus minus a point. This map is an injection if one point y is given so that $p(y) = x$. Hence, the free group on two generators has as a subgroup which is not finitely generated!

D.9. Simplicial homology theory

Simplicial homology is a standard technique used in computing the homology groups of various surfaces.[4] An *n-simplex* is the n-dimensional analogue of a triangle. It is built by drawing n orthogonal lines and then closing the ends. For a 3-simplex, you would draw three orthogonal lines and then connect the ends of these lines to each other via three more lines to give a tetrahedron. A 0-simplex is simply a point; a 1-simplex is a line; a 2-simplex is a triangle; etc. We can use simplexes to triangulate a surface and compute the Euler characteristic and other topological properties in this fashion.

We write an arbitrary simplex algebraically as

$$\sigma_n = \langle p_0 p_1 \cdots p_n \rangle , \tag{D.10}$$

where each p_i labels a vertex point. One could think of this notation as a formula for drawing a simplex in that you connect the adjacent points with straight lines

$$p_0 \to p_1 \to \cdots \to p_n \to p_0.$$

Since the p_i are all geometrically independent, we can think of the simplex as a vector in $\mathbb{R}^m$ where $m \geq n$. We, therefore, express the simplex as

$$\sigma_n = \left\{ x \in \mathbb{R}^m \text{ such that } x = \sum_{i=0}^{n} c_i p_i, \; c_i \geq 0, \text{ and } \sum_{i=0}^{n} c_i = 1 \right\}. \tag{D.11}$$

The coordinate $(c_0, ..., c_n)$ is called the barycentric coordinate of x.

If we choose $q+1$ points out of our n-simplex, we are left with a q-simplex called a q-face of our simplex. σ_q is called a proper face if it is a face of σ_n. To clarify this, consider a tetrahedron thought of as a 3-simplex. We can write it as $< p_0 p_1 p_2 p_3 >$. Any subset of these points is a proper face of the simplex since it will always form a lower-dimension simplex. The number of q-faces in an n-simplex is given by

$$\frac{(n+1)!}{(q+1)!(n-q-1)!} = \begin{pmatrix} n+1 \\ q+1 \end{pmatrix}$$

[4]There is another common method that is more convenient in higher-dimensional homotopy theory and K-theory which uses cells and CW-complexes. These cells are lines, squares, cubes, and their generalizations rather than triangles.

since they are formed by choosing $q+1$ points from $n+1$ points without regard to the ordering of the points. There are $(n+1)!/(n-q-1)!$ ways to pick a set of $q+1$ points from $n+1$, and there are $(q+1)!$ ways to reorder each of these sets that results in the same simplex.

We can fit simplexes together to form simplicial complexes. Simplicial complexes are formed by attaching simplexes together at points; along edges; on faces; or, in general, by identifying a proper q-face of one with a proper q-face of the other. The rules for putting two simplexes, σ and σ', together to form a simplicial complex, K, can be summarized as

(a) An arbitrary face of a simplex of K belongs to K. That is, if $\sigma \in K$ and $\sigma' \leq \sigma$, then $\sigma' \in K$.

(b) If σ and σ' are two simplexes of K, the intersection, $\sigma \cap \sigma'$, is either empty or a face of σ and σ'. That is, if $\sigma, \sigma' \in K$, then either $\sigma \cap \sigma' = \emptyset$, or $\sigma \cap \sigma' \leq \sigma$ and $\sigma \cap \sigma' \leq \sigma'$.

The first rule means that an arbitrary simplex is a subset of the points in the longest simplex. The second rule can be exemplified by our 3-simplex. The possible ways to combine two 3-simplexes would be so that

$$\sigma \cap \sigma' = \emptyset, < p_0 >, < p_0 p_1 >, < p_0 p_1 p_2 > .$$

The union of all of the simplexes in the simplicial complex is called the polyhedron, $|K|$, of the simplicial complex K.

Definition: (Triangulation) A topological space X is said to be triangulable if there exists a simplicial complex, K, and a homeomorphism, $f : |K| \to X$. The pair (K, f) is called the triangulation of X.

When forming triangulations, it is important to remember that the intersection of any two simplexes in the polygon must either be empty or a simplex. They cannot intersect in a disjoint fashion. For example, Figure D.1.(a) is a valid triangulation while Figure D.1.(b) is not.

We assign an orientation to each of our n-simplexes by assigning an orientation to each of its edges. Write an oriented simplex, $\sigma_n = (p_0 p_1 \cdots p_n)$, so that cyclic and anti-cyclic permutations, P, of the points give

$$(p_{i_0} p_{i_1} \cdots p_{i_n}) = sgn(P)(p_0 p_1 \cdots p_n).$$

We define the positive orientation as the one which is given by the right-hand rule on each of the faces.

We can now give a new definition of the group of n-chains defined earlier. Let $K = \{\sigma_\alpha\}$ be an n-dimensional simplicial complex where the simplexes, σ_α, in K are oriented. The r-chain group, $C_r(K)$, of K is a free abelian group generated by the oriented r-simplexes of K. If $r > \dim K$, $C_r(K)$ is defined to be 0. An element of $C_r(K)$ is then an r-chain.

Thus, r-chains are defined using simplexes rather than curves as we did earlier, allowing a generalization to higher dimensions in a more convenient way. Let there be

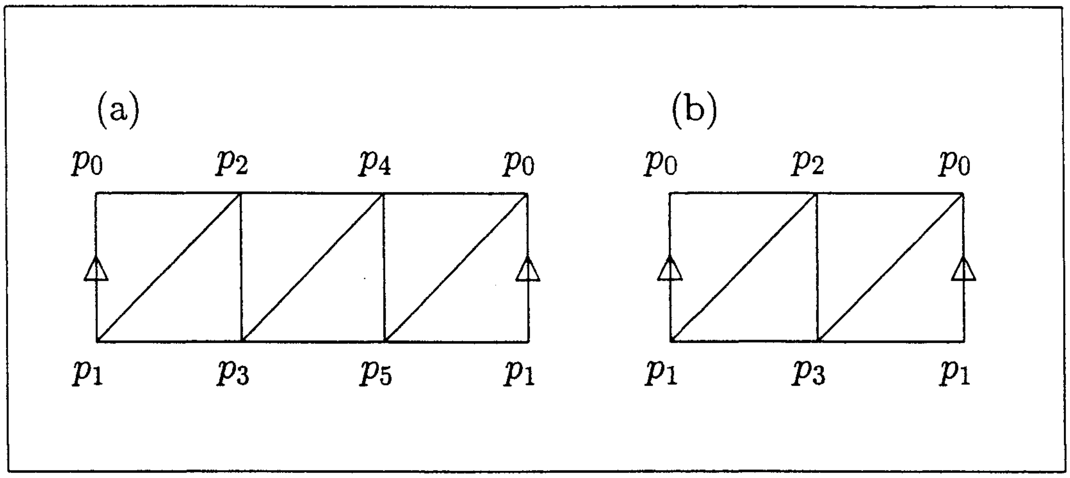


Figure D.1. (a) is a triangulation of the cylinder while (b) is not.

I_r r-simplexes in K, denoted by $\sigma_{r,i}(1 \leq i \leq I_r)$. Then, $c \in C_r(K)$ is expressed as

$$c = \sum_{i=1}^{I_r} c_i \sigma_{r,i}, \quad c_i \in \mathbb{Z}.$$

The group product is given by addition in the obvious fashion. The identity is $c = 0$. The inverse of a simplex is the simplex with the opposite orientation. Thus, $C_r(K)$ is a free abelian group of rank I_r,

$$C_r(K) \cong \underbrace{\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}}_{I_r}.$$

We can define the groups of boundaries and cycles in a general fashion. Let $\partial_r : C_r(K) \to C_{r-1}(K)$ be the boundary operator on r-chains. It acts on a 0-simplex and a 1-simplex as before:

$$\partial_0 \sigma_0 = 0, \quad \partial_1 \sigma_1 = p_1 - p_0.$$

It is easy to show that the boundary operator commutes with the group product operation which is simply addition. For a general r-simplex, we have

$$\partial_r \sigma_r = \sum_{i=0}^{r} (-1)^i (p_0 p_1 \cdots \hat{p}_i \cdots p_r), \tag{D.12}$$

where the hat indicates that we have omitted that vertex in the sum. For example,

$$\partial_2(p_0p_1p_2) = (p_1p_2) - (p_0p_2) + (p_0p_1),$$
$$\partial_3(p_0p_1p_2p_3) = (p_1p_2p_3) - (p_0p_2p_3) + (p_0p_1p_3) - (p_0p_1p_2),$$

and acting on a general r-chain, we have

$$\partial_r c = \sum_i c_i \partial_r \sigma_{r,i}.$$

Let K be an n-dimensional simplicial complex (and, therefore, homeomorphic to an n-dimensional space, as an n-simplex is homeomorphic to an n-sphere). There exists a sequence of free abelian groups and homomorphisms called the chain complex which is an exact sequence. (The image of one map is the kernel of the next map.) The chain complex is given by

$$0 \xrightarrow{i} C_n(K) \xrightarrow{\partial_n} C_{n-1}(K) \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_2} C_1(K) \xrightarrow{\partial_1} C_0(K) \xrightarrow{\partial_0} 0,$$

where i is the inclusion map. We can define an r-boundary, c, as an r-chain such that there exists and $r+1$-chain with the property that $c = \partial_{r+1} d$. The set of r-boundaries forms a subgroup of $C_r(K)$ denoted by $B_r(K)$. An r-cycle is an element of the subgroup $Z_r(K)$ of $C_r(K)$ with the property that $\partial_r c = 0$ for all $c \in Z_r(K)$. $B_r t(K) \subset Z_r(K)$ since all boundaries are closed chains (cycles).

The group of 1-chains of a triangle and that of a square are, respectively,

$$C_1(K_1) = \{i(p_0p_1) + j(p_1p_2) + k(p_2p_0) | i, j, k \in \mathbb{Z}\} \cong \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z} \tag{D.13}$$

and

$$C_1(K_2) \cong \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}. \tag{D.14}$$

The triangle and the square are homeomorphic spaces and are, therefore, topologically equivalent. Hence, the group of 1-chains is not a topological invariant.

Let K be an n-dimensional simplicial complex. The rth homology group, $H_r(K)$, where $0 \leq r \leq n$, associated with K is defined by

$$H_r(K) = Z_r(K)/B_r(K)$$

and is the set of equivalence classes of r-cycles. Each equivalence class, $[z]$, is called a homology class. Two r-cycles are homologous if their difference is a boundary (and hence the identity in $H_r(K)$).

If a topological space, X, is homeomorphic to a topological space, Y, and (K, f) and (L, g) are triangulations of X and Y, respectively, then $H_r(K) \cong H_r(L)$. This result is clearly independent of the triangulation so that $H_r(X) \cong H_r(Y)$. The rth homology group is a topological invariant. We have generalized what we have established earlier for the case $r = 1$.

D.9.1. Homology of the Möbius band

Figure D.2. shows a triangulation of the Möbius band with orientations chosen for the 2-simplexes. We see that, since there are no 3-chains, $B_2(K) = 0$. A 2-cycle, $z \in Z_2(K)$, will be a linear combination of oriented 2-simplexes shown in Figure D.2..

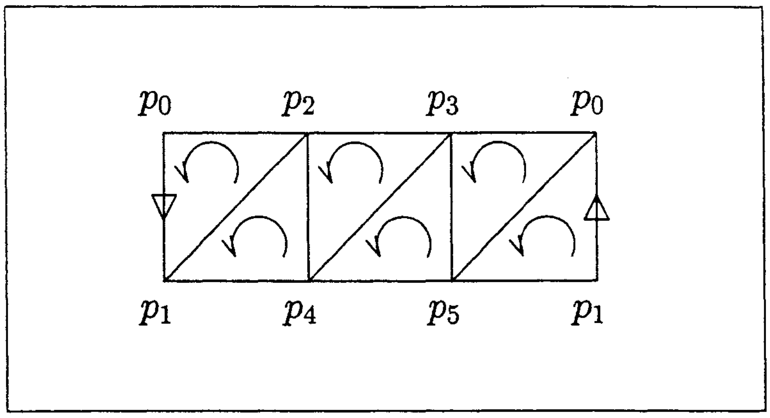


Figure D.2. A triangulation of the Möbius band.

We write it as

$$z = a_1(p_0p_1p_2) + a_2(p_2p_1p_4) + a_3(p_2p_4p_3) + a_4(p_3p_4p_5) + a_5(p_3p_5p_1) + a_6(p_1p_5p_0).$$

Since z is an element of $Z_2(K)$, it is a 2-cycle, or a closed 2-chain. Being closed gives the following requirement (using (D.12)):

$$\begin{aligned}\partial_2 z = {} & a_1\left[(p_1p_2) - (p_0p_2) + (p_0p_1)\right] \\ & + a_2\left[(p_1p_4) - (p_2p_4) + (p_2p_1)\right] \\ & + a_3\left[(p_4p_3) - (p_2p_3) + (p_2p_4)\right] \\ & + a_4\left[(p_4p_5) - (p_3p_5) + (p_3p_4)\right] \\ & + a_5\left[(p_5p_1) - (p_3p_1) + (p_3p_5)\right] \\ & + a_6\left[(p_5p_0) - (p_1p_0) + (p_1p_5)\right] = 0.\end{aligned} \tag{D.15}$$

Examining this equation, we see that the 1-chains that lie along the edge of the Möbius strip only appear once. The others are traversed twice each, once in each direction and, hence, with a relative minus sign. Since a 1-chain from the edge, i.e., one of $(p_0p_2), (p_2p_3), (p_3p_1), (p_1p_4), (p_4p_5), (p_5p_0)$, appears in each term, there is no way they can cancel each other unless $a_1 = \cdots = a_6 = 0$. Therefore, we have $Z_2(K) = 0$. (There are no closed 2-cycles.) Thus, we have

$$H_2(K) = Z_2(K)/B_2(K) \cong \{0\}. \tag{D.16}$$

The second homology group of the Möbius band is trivial. This statement is intuitive since the Möbius band encloses no volume. If the second homology group was non-trivial, it would indicate that the complex, if embedded in $\mathbb{R}^3$, actually separates the region into two connected components: the inside and the outside of the complex. The second homology group is similar to the zeroth homology group (when the complex is embedded in $\mathbb{R}^3$ at least) in that the zeroth homology group tells the number of connected components of the complex, whereas the second homology group (in $\mathbb{R}^3$) gives the number of connected components ($\dim H_2(K) - 1$) in the *complement* of the complex.

Since the Möbius band is connected, we have that the zeroth homology group has one generator. The generator has an integer coefficient defining the elements. (The closed zero chains, or points, are all boundaries of one chains.[5]) The group is, therefore, $H_0(K) \cong \mathbb{Z}$.

The first homology group, $H_1(K)$, is given by the group of 1-cycles modulo the 1-boundaries. By examining the diagram in Figure D.2., we see that the closed circuits along the edges of the 2-simplexes are 1-cycles. They are also 1-boundaries since they bound the 2-simplex which they circuit. The 1-cycles that do not bound a 2-simplex are the ones that traverse the Möbius band from one side to the other and connect when we identify the edges. For example, $(p_1p_4) + (p_4p_5) - (p_1p_5)$. All such 1-cycles are homologous in that their difference is a boundary. For example, take the difference between $(p_0p_1) + (p_1p_4) + (p_4p_5) + (p_5p_0)$ and our previous example, giving

$$\begin{aligned}(p_1p_4) + (p_4p_5) - (p_1p_5) - (p_0p_1) - (p_1p_4) - (p_4p_5) - (p_5p_0) \\ = -\left[(p_1p_5) + (p_5p_0) + (p_0p_1)\right]\end{aligned} \tag{D.17}$$

which is a 1-boundary. We have that this class of 1-cycles generates $H_1(K)$. Thus, $H_1(K) = \{n[z] | n \in \mathbb{Z}\} \cong \mathbb{Z}$. In Section E.1., we will show that the Möbius band also forms a generalization of a covering space called a fibre bundle.

D.9.2. Homology of the projective plane

The projective plane, $\mathbb{R}P^2$, is the sphere with antipodal points identified. Mapping the upper hemisphere of the sphere to the lower hemisphere, we are left with a disk. The projective plane is this disk with the S^1 boundary (which was the equator of the sphere) having its antipodal points identified. We now have a convenient method to triangulate the projective plane. Using the rules defined earlier for building a simplicial complex, we triangulate the projective plane as shown in Figure D.3..

Since there are no 3-chains in the projective plane, we have $B_2(K) = 0$. Taking all of the 2-simplexes and adding them together with coefficients, a_i, and then setting the boundary to zero (It is a cycle.), we find that all of the coefficients have to be zero. Hence, $Z_2(K) = 0$ and $H_2(K) = \{0\}$.

Examination of the diagram reveals that any 1-cycle is homologous to the 1-cycle given by $z_1 = (p_3p_5) + (p_5p_4) + (p_4p_3)$. Adding all of the 2-simplexes with the same coefficient, m, and taking the boundary, we arrive at $\partial_2 z_2 = 2m\left((p_3p_5) + (p_5p_4) + (p_4p_3)\right)$. This boundary is an even multiple of our 1-cycle, z_1. Any even multiple of our 1-cycle is a boundary. The first homology group is given by

$$H_1(K) = Z_1(K)/B_1(K) = \{[z] | [z] + [z] \sim [0]\} \cong \mathbb{Z}/2\mathbb{Z} = \mathbb{Z}_2. \tag{D.18}$$

[5] If there were more than one connected component, the 0-chains in one component would not be boundaries of the 1-chains in the other. We would then have more generators, one generator for each component.

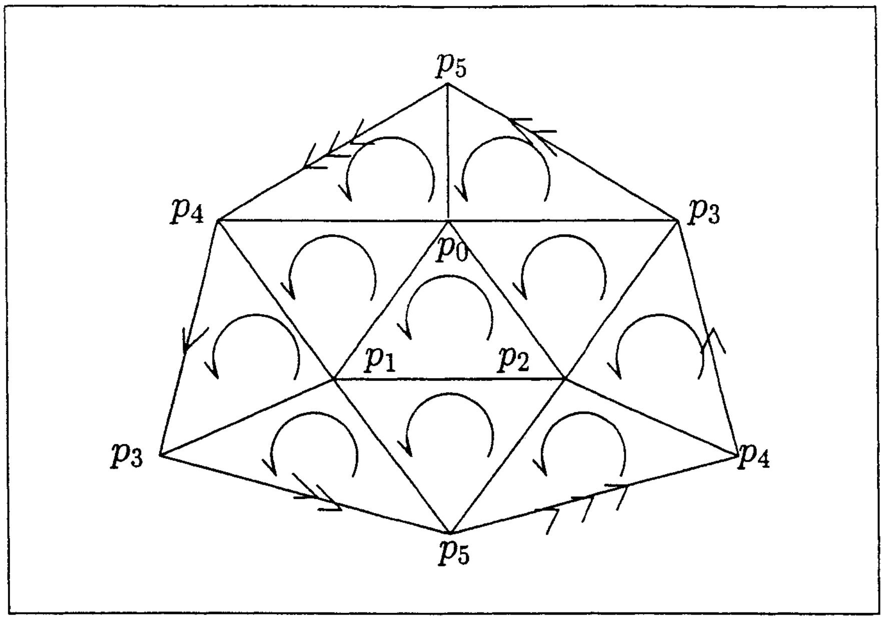


Figure D.3. A triangulation of the real projective plane $\mathbb{R}P^2$.

The first homology group is not free abelian but is a finitely generated abelian group (cyclic group of order 2). The zeroth homology group is $H_0(K) \cong \mathbb{Z}$, as expected, since the projective plane has a single connected component.

D.9.3. Homology of the genus-g torus

A sphere with g handles is homeomorphic to a torus with g holes in it. Since there are no 3-chains, $B_2(T_g) = 0$. Since the surface has no 1-chain as a boundary (It is closed.), the surface itself freely generates $H_2(T_g) \cong \mathbb{Z}$. Hence, the complement of the g-torus in $\mathbb{R}^3$ has two connected components.

$H_1(T_g)$ is generated by the 1-cycles which are not themselves 1-boundaries. These cycles are simply the 1-cycles surrounding each of the g holes. We have

$$H_1(T_g) \cong \underbrace{\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}}_{2g}. \tag{D.19}$$

There is only 1 connected component, so $H_0(T_g) \cong \mathbb{Z}$.

D.9.4. Homology of the Klein bottle

For the Klein bottle, we have $B_2(K) = 0$ since it is a 2-dimensional surface and there are no 3-chains. From the triangulation in Figure D.4., we form a 2-cycle by adding together all of the 2-simplexes with the same coefficient. The result is $z_2 = \sum_i n\sigma_{2,i}$. Taking the boundary, we see that the inner 1-simplexes cancel, leaving only the boundary

$$\partial_2 z_2 = -2na = -2n\left[(p_0p_1) + (p_1p_2) + (p_2p_0)\right]. \tag{D.20}$$

For this boundary to be zero, we must have $n = 0$, so $H_2(K) = Z_2(K) \cong \{0\}$.

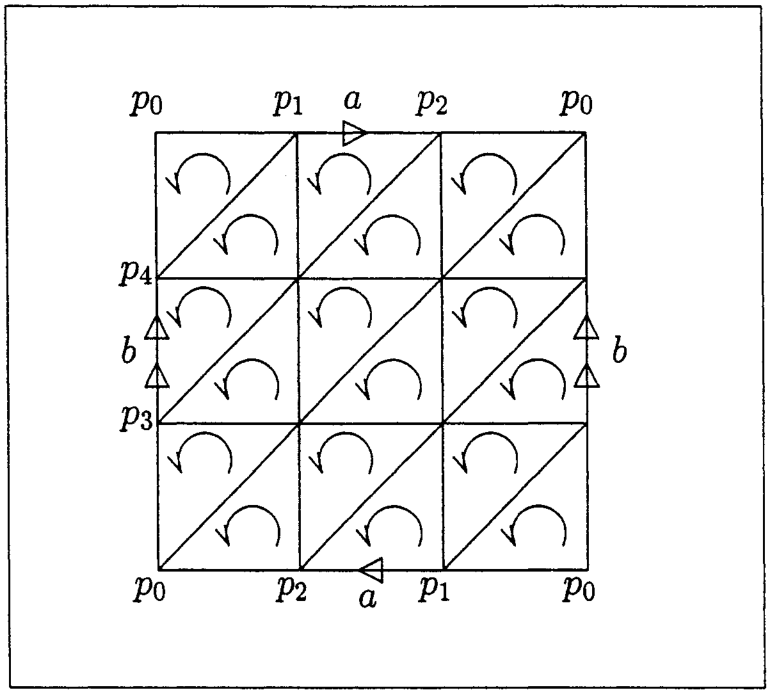


Figure D.4. A triangulation of the Klein bottle.

Every 1-cycle is homologous to $na + mb$ for $n, m \in \mathbb{Z}$. Any 1-cycle that is itself the boundary of a 2-chain (i.e., $z_1 = \partial_2 z_2$) must have the form $z_1 = 2na$ since 1-boundaries must have this form. Hence, $2na \sim 0$ (is homologous to the identity) and thus

$$H_1(K) = \{n[a] + m[b] | n, m \in \mathbb{Z} \text{ and } a + a = 0\} \cong \mathbb{Z}_2 \oplus \mathbb{Z}. \tag{D.21}$$

Again, $H_0(K) \cong \mathbb{Z}$ since it is connected.

The subgroup of the rth homology group that contains finite cyclic groups is called the torsion subgroup. It is interesting to notice that, since the dimension of a finite group is zero, the Euler characteristic (since we are taking coefficients in $\mathbb{R}$ rather than $\mathbb{Z}$) does not distinguish torsion subgroups

$$\chi(K) = \sum_{r=0}^{n}(-1)^r \mathrm{dim} H_r(K, \mathbb{R}) = \sum_{r=0}^{n}(-1)^r b_r(K) = \sum_{r=0}^{n}(-1)^r I_r(K), \tag{D.22}$$

where $b_r(K)$ is the rth Betti number of the surface and I_r is the number of r simplexes in the simplicial complex triangulating the surface.

Draw a graph on a surface with edges (1-simplexes/chains) corresponding to the saddle points (points of zero local curvature) of a vector field, vertices (0-simplexes/chains) corresponding to sources, and faces (2-simplexes/chains) corresponding to sinks. Any two of these numbers will give us the third by the Euler characteristic since the graph is nothing but a triangulation of the surface. (This was mentioned briefly in the introduction.) Hence, knowing the global topology of a surface gives information about the possible vector fields that can live on the surface.

Example 7:

We prove that a (triangulable) compact, non-orientable surface has a normal form

$$\alpha_1\alpha_1 \cdots \alpha_h\alpha_h.$$

In a polygon describing an arbitrary compact, non-orientable surface, the operation shown in Figure D.9.4. replaces two separated edges by two edges that are adjacent.

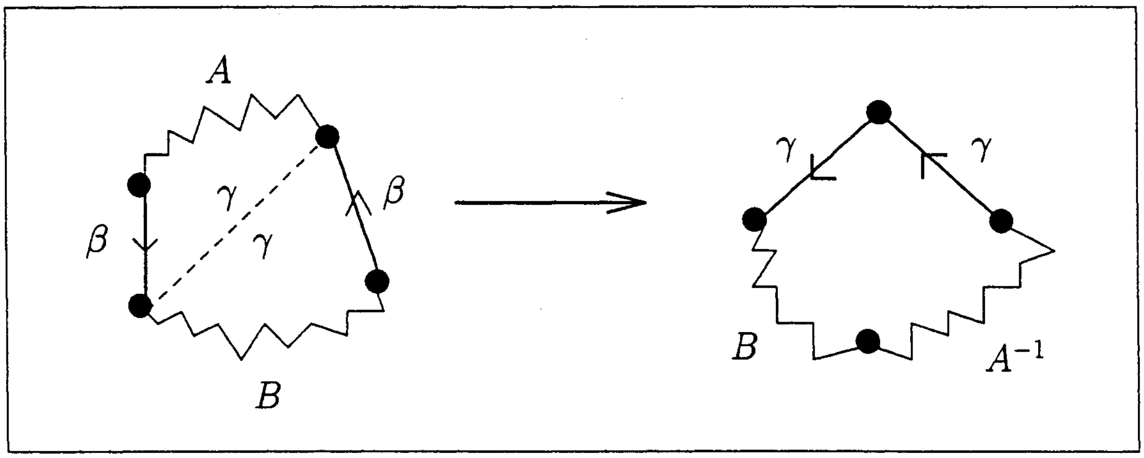


Figure D.5. Operation moving the two edges together.

The operation does not affect the relationships between edges residing in regions A and B as far as adjacency. It changes the orientation of the edges in A so that $A \to A^{-1}$. Also, $\alpha^{-1}\alpha^{-1} = \alpha\alpha$ since the edges are identified in the exact same way in each case. We can, therefore, write $\alpha^{-1}\alpha^{-1} \to \alpha\alpha$ when it occurs.

Let A, B, and C be arbitrary sequences of letters, and let γ and α_i represent edges. We have the following algebraic rules.

(a) $\alpha_1 \cdots \alpha_n\alpha_1 \cdots \alpha_n = \alpha_1^2$

(b) $\gamma_1^{-1}\gamma_1^{-1} = \gamma_1\gamma_1$

(c) $ABC = BCA = CAB$

(d) $\gamma A\gamma B = AB^{-1}\gamma^2 = BA^{-1}\gamma^2$

Our problem is reduced to algebraic manipulations rather than trying to deal with all of the cutting and pasting of polygon diagrams.[6] Suppose we have a sequence, $A\gamma B\gamma$. We can use (c) and (d) to write it as $AB^{-1}\gamma^2 = \gamma^2 AB^{-1}$. Now, suppose there is another pair somewhere in AB^{-1}. We can then continue the same procedure and pull them out until we are left with $\gamma_1^2 \cdots \gamma_n^2 C \equiv GC$, where G is our γs and C contains only α_i and α_i^{-1} in some sequence. It can be shown that we can leave the G untouched and put C into the form $\alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1} \cdots \alpha_k\beta_k\alpha_k^{-1}\beta_k^{-1}$.

So far, we have that an arbitrary, non-orientable surface can be put into the form

$$\alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1} \cdots \alpha_k\beta_k\alpha_k^{-1}\beta_k^{-1}\gamma_1^2 \cdots \gamma_n^2.$$

Using our identities

$$\begin{aligned}
C\alpha_k\beta_k\alpha_k^{-1}\beta_k^{-1}\gamma_1^2 G &= GC\alpha_k\beta_k\gamma_1\beta_k\alpha_k\gamma_1, \text{ from } (d) \\
&= \gamma_1 GC\alpha_k\beta_k\gamma_1\beta_k\alpha_k, \text{ from } (c) \\
&= \gamma_1 GC\beta_k^{-1}\gamma^{-1}\beta_k^{-1}\alpha_k^2 = \alpha_k^2\gamma_1 GC\beta_k^{-1}\gamma^{-1}\beta_k^{-1}, \\
&= \alpha_k^2\gamma_1 GC\gamma_1(\beta_k^{-1})^2 = (\beta_k^{-1})^2\alpha_k^2\gamma_1 GC\gamma_1, \\
&= GC\gamma_1(\beta_k^{-1})^2\alpha_k^2\gamma_1 = GC(\alpha_k^{-1})^2\beta_k^2\gamma_1^2 = CG', \\
\text{Where } C &= \alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1} \cdots \alpha_{k-1}\beta_{k-1}\alpha_{k-1}^{-1}\beta_{k-1}^{-1}, \\
\text{and } G' &= (\alpha_k^{-1})^2\beta_k^2\gamma_1^2 \cdots \gamma_n^2.
\end{aligned} \tag{D.23}$$

Clearly, this process can be continued until the code is of the form

$$\gamma_1\gamma_1 \cdots \gamma_h\gamma_h$$

which is what we wanted to show.

Example 8:

For a non-orientable, compact surface, X, with normal form,

$$\alpha_1\alpha_1 \cdots \alpha_h\alpha_h,$$

we will compute the fundamental group and the first homology group of the surface. We begin with the polygon representing an arbitrary, non-orientable surface and designate regions U, V, and $U \cap V$ similar to those shown in Figure D.9.4..

We call the entire boundary of the polygon K. K consists of generators, $\alpha_1, ..., \alpha_h$, and its fundamental group is, therefore, the free group, F_h, on h generators. Since K is a deformation retract of U, we have $\pi_1(U, x) = \pi_1(K, x)$. The region, V, is homeomorphic to a disk and, therefore, has a trivial fundamental group, $\pi_1(V, x) = \{e_x\}$. The intersection, $U \cap V$, has circle S^1 as a deformation retract and, therefore,

[6]Notice that (a) can be trivially proven from (d), or shown by beginning with the projective plane and adding factors of the identity. (b) and (c) are tautological.

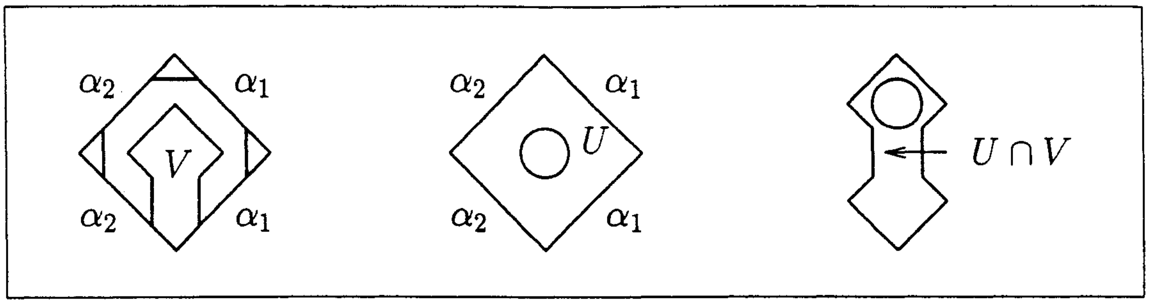
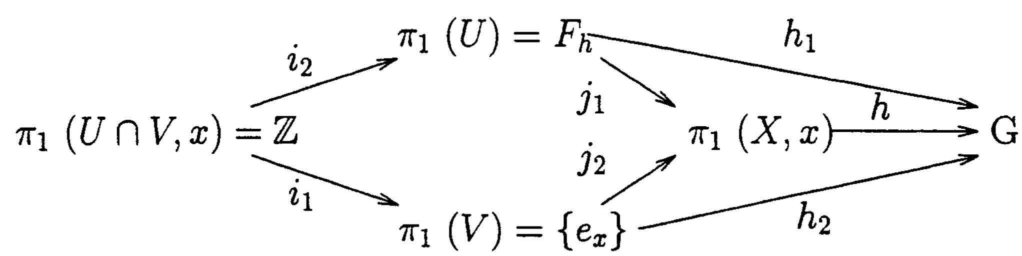


Figure D.6. Regions U, V, and $U \cap V$ of an arbitrary non-orientable surface.

has $\pi_1(U \cap V, x) = \mathbb{Z}$. The Siefert Van-Kampen theorem tells us that the following diagram commutes.



The inclusion, i_1, of $U \cap V$ in U takes the single generator of $\pi_1(U \cap V, x)$ to $\alpha_1^2 \cdots \alpha_h^2$ in $\pi_1(U, x) = F_h$. The commuting diagram tells us that

$$h_1 \circ i_1 = h_2 \circ i_2 = h_2(e_x) = id_G \tag{D.24}$$

$$\Rightarrow h_1(\alpha_1^2 \cdots \alpha_h^2) = id_G, \tag{D.25}$$

which means that h_1 is a homomorphism from $F_h \rightarrow G$ such that $\alpha_1^2 \cdots \alpha_h^2 \mapsto id_G$. Since it is a homomorphism,[7] the normal subgroup $N_g = \{\alpha_i(\alpha_1^2 \cdots \alpha_h^2)\alpha_i^{-1}\} \mapsto id_G$.

Siefert Van-Kampen implies

$$h \circ j_1 = h_1,$$

which means that

$$Hom(\pi_1(X, x), G) = Hom(F_h/N_g, G).$$

[7] h_1 is a homomorphism because we have

$$h(\alpha_i(\alpha_1^2 \cdots \alpha_h^2)\alpha_i^{-1}) = h(\alpha_i)h(\alpha_1^2 \cdots \alpha_h^2)h(\alpha_i^{-1}) = h(\alpha_1^2 \cdots \alpha_h^2)h(\alpha_i)h(\alpha_i^{-1})$$
$$= h(\alpha_1^2 \cdots \alpha_h^2\alpha_i\alpha_i^{-1}) = h(\alpha_1^2 \cdots \alpha_h^2) = id_G.$$

Thus $\pi_1(X,x) = F_h/N_g$ where N_g is the least normal subgroup containing $\alpha_1^2 \cdots \alpha_h^2$.

The first homology group $H_1(X)$ is given by the abelianization of the fundamental group. This group is the subgroup formed by sending all commutators to the identity. A commutator is of the form

$$\begin{aligned} [\alpha, \beta] &= \alpha\beta\alpha^{-1}\beta^{-1} \rightarrow e, \\ \beta^2 &= \alpha\beta\alpha^{-1}\beta, \\ \beta^2 &= \alpha^2\beta^2 \text{ by problem \# 1, rule } (d) \text{ above}, \\ e &= \alpha^2, \end{aligned} \tag{D.26}$$

and we have $\alpha_i^2 \rightarrow e$ which is $\cong 2\mathbb{Z}$. Our homology group is, therefore,

$$\begin{aligned} H_1(X) &= \overbrace{\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}}^{h}/2\mathbb{Z}, \\ &= \overbrace{\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}}^{h-1} \oplus \mathbb{Z}_2. \end{aligned} \tag{D.27}$$

Example 9:

Let X be a compact, oriented surface. We will show that there is a perfect pairing

$$H^2(X) \times H^0(X) \rightarrow \mathbb{R} \quad (\mu, f) \mapsto \int_X f\mu.$$

For a compact, orientable surface, dim $H^2(X) = 1$. If, for every 2-form, μ, and locally constant functions, f and f', in $H^0(X)$, we have

$$\int_X f\mu = \int_X f'\mu,$$

then

$$\int_X (f - f')\mu = 0 \Rightarrow f = f'.$$

If

$$\int_X f\mu = \int_X f\mu'$$

for every f, then

$$\int_X f(\mu - \mu') = 0.$$

If $\int_X \omega = 0$, then $\omega = d\nu$ is exact so that $f(\mu - \mu')$ is an exact 2-form and[8] that $\mu - \mu'$ is an exact 2-form. We have shown that there is a unique $f \in H^0(X)$ and a unique $\mu \in H^2(X)$ such that

$$(\mu, f) \mapsto \int_X f\mu.$$

[8]Set the arbitrary function, f, to be the constant function, 1, and the result follows.

Therefore, we have a perfect pairing.

Example 10:

If X is a non-orientable, compact surface, $H^2(X) = 0$.

We realize our non-orientable surface as a sphere with h disks removed and h Möbius put in their places. Let $X = U \cup V$, where U is the disk with holes removed and V is the disjoint union of the Möbius bands. Then, $H^1(U) = \mathbb{R}^h$, and $H^1(V) = \mathbb{R}^h$. The intersection $U \cap V$, is on copies of S^1 around the removed disks. Therefore, $H^1(U \cap V) = \mathbb{R}^h$. We also have $H^2(U) = H^2(V) = 0$ by the fact that there are no closed 2-chains in a Möbius band. (It has an unidentified edge.)

$$\mathbb{R}^h \oplus \mathbb{R}^h \xrightarrow{-} \mathbb{R}^h \xrightarrow{\delta} H^2(U \cup V) \xrightarrow{+} 0. \tag{D.28}$$

The image of $-$ is the entire $H^1(U \cap V)$ which is then the kernel of δ. The image of δ is, therefore, zero. Hence, the kernel of $+$ is zero, but since $H^2(U) \oplus H^2(V) = 0$, we have $Img(\delta) = 0$ and $H^2(U \cup V) = 0$.

Example 11:

As a final example, we will show that, for any triangulable surface (orientable or non-orientable), if triangulated with v vertices, e edges, and f faces, the equality

$$\sum_{i=0}^{2} (-1)^i \dim H^i(X) = v - e + f$$

holds. (See Figure D.9.4..)

We have the following results:

$$\begin{array}{ll} \overbrace{\qquad\text{orientable}\qquad} & \overbrace{\qquad\text{non-orientable}\qquad} \\ \dim H^0(X) = 1 & \dim H^0(X) = 1 \\ \dim H^1(X) = 2g & \dim H^1(X) = h - 1 \\ \dim H^2(X) = 1 & \dim H^2(X) = 0 \\ \text{-------} & \text{-------} \\ \sum = 2 - 2g & \sum = 2 - h \end{array} \tag{D.29}$$

The formula for the Euler characteristic, $\chi = v - e + f$, gives $\chi = 2 - 2g$ for orientable surfaces (sphere with g handles) and $\chi = 2 - h$ for non-orientable surfaces as can be seen from Figure D.9.4..

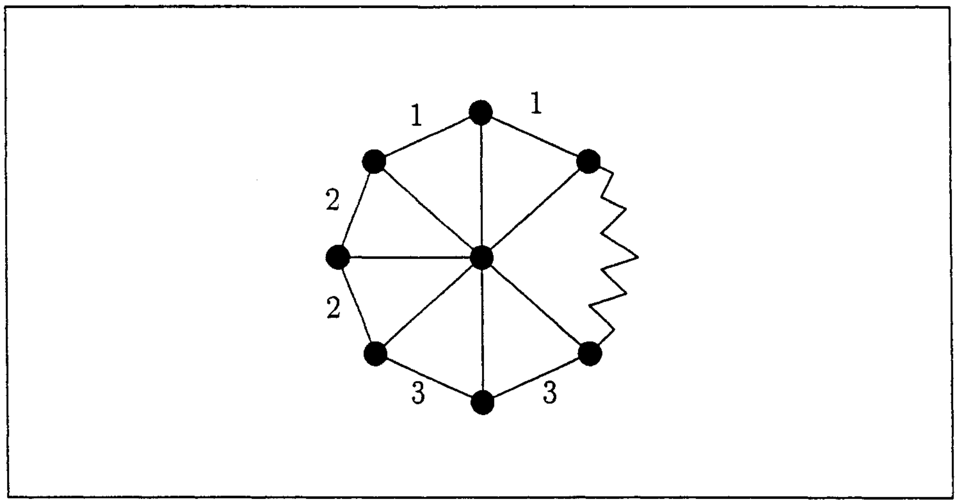


Figure D.7. Triangulation of a non-orientable surface to compute the Euler characteristic. There are 2 vertices, $3h$ edges, and $2h$ faces giving Euler characteristic $\chi = 2 - 3h + 2h = 2 - h$.

APPENDIX E
INTRODUCTION TO FIBRE BUNDLES

Fibre bundles are a generalization of the covering spaces discussed in Section D.6. Therefore, many of the results derived for covering spaces have corresponding results in the theory of fibre bundles.

E.1. Definition of a fibre bundle

A fibre bundle, $\mathcal{B}$, consists of at least the following (See [92].):

(a) A topological space, B, called the bundle space.

(b) A topological space, X, called the base space.

(c) A continuous projection map, $\pi : B \to X$, from the bundle to the base.

(d) A space, Y, called a typical fibre.

(e) For each $x \in X$, there is a neighborhood, V, of x and a homeomorphism, $\phi : V \times Y \to \pi^{-1}(V)$, such that $\pi\phi(x', y) = x'$ for $x' \in V$ and $y \in Y$.

(f) A group, H, of homeomorphisms of Y called the group of the bundle.

(g) A group, G, of homeomorphisms of Y called the group of the fibre.

We will give some examples of bundles and some additional terminology, but let us first review each of the mathematical concepts pertaining to the definition.

Topology: A topology on a set, S, is a collection, $\mathcal{T}$, of subsets of S having the following properties:

(a) $\emptyset$ and S are elements of $\mathcal{T}$.

(b) The union of the elements of any subcollection of $\mathcal{T}$ is an element of $\mathcal{T}$.

(c) The intersection of the elements of any finite subcollection of $\mathcal{T}$ is an element of $\mathcal{T}$.

A set for which a topology has been defined is called a **topological space**.

Topological subspace: Any subset Y of a topological space, B, with the induced topology is called a topological subspace. The induced topology consists of open sets of the form $U \cap Y$ for U open in B.

Product topology: If X and Y are topological spaces, the cartesian products, $U \times V$, of open sets, $U \subset X$ and $V \subset Y$, form a basis for the product topology of $X \times Y$.

Discrete topology: In the discrete topology, every subset of the space is open, and the points are all defined to be open.

Continuous map: A map, $p : Y \to X$, from one topological space to another is continuous if $p^{-1}(U)$ is open in Y for every U open in X. Hence, if a set in the image space, X, is open, it necessarily came from an open set in Y. A closed set in X, on the other hand, could have come from either an open or a closed set in Y.

Homeomorphism: A bijection, $f : Y \to Y'$, is a homeomorphism if f and f^{-1} are continuous.

Group Axioms: A group $(G, \cdot)$ is a set, G, together with any binary operation $\cdot$ such that

(G1) $\cdot$ is closed, i.e., for every ordered pair (g_1, g_2) of elements of G, there is a unique element, $g_1 \cdot g_2$, also in G.

(G2) $\cdot$ is associative.

(G3) There is an element, e (identity), in G such that $g \cdot e = e \cdot g = g$ for all g in G.

(G4) For every element g in G, there is an element, g^{-1} (inverse), in G such that $g \cdot g^{-1} = g^{-1} \cdot g = e$.

A **cross section** of a bundle is a continuous map, $f : X \to B$, from the base space into the bundle satisfying $\pi f(x) = x$ for each $x \in X$.

The set, Y_x, defined by $Y_x = \pi^{-1}(x)$, is called the fibre over the point x of X. This set can be seen as the points in the bundle that project down to the point x in the base.

Examples of bundles

Product bundle: The first example of a fibre bundle is the product bundle, $B = X \times Y$. It is just the direct product of the base space and the fibre. For example. $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ is a product bundle. The topology on the product bundle is just the product topology. In the case of $\mathbb{R}^2$, the open sets in the topology of the bundle are just the cartesian product of the open sets in two copies of $\mathbb{R}$; i.e., they are open squares or open disks. The projection map is simply $\pi(x, y) = x$, picking out the first component of any ordered pair denoting a point in the bundle space. The typical fibre is simply $\mathbb{R}$. Since the entire bundle is a direct product in this case, every point has a neighborhood which is a direct product. Axiom (e) is satisfied by taking $V = X$ and ϕ being the identity so that $\phi(x, y) = (x, y)$ and, therefore $\pi\phi(x, y) = \pi(x, y) = x$. Notice that ϕ is a function that takes a point from the fibre and a point from the base, and returns a point in the bundle space. In this case, the point in the bundle is simply the direct product of the two points. The cross sections of B are the graphs of maps $X \to Y$ since $\pi(x, f(x)) = x$. The fibres are all homeomorphic to $\mathbb{R}$ and to each other. The natural homeomorphism $Y_x \to Y$ from the fibre over x to the typical fibre Y given by $(x, y) \to y$. Thus, the group H of the bundle is the identity. The group G of the fibre is translations since the fibre is $\mathbb{R}$ and is unchanged by translation by any real number.

Möbius band: A more non-trivial example is the Möbius band. The base space is the circle S^1 obtained from a line segment L by identifying its ends and the typical fibre, Y, is a line segment. The bundle is obtained from the product, $L \times Y$, by matching the two ends with a twist when forming S^1 from L. In this fashion, the resulting bundle is not just the direct product, $S^1 \times Y$ (which would be a cylinder). The open sets are just direct products of the open neighboorhoods of the points in the base space with those of the points in the fibre space. The projection is the same as the product bundle, as is the homeomorphism $\phi(x, y)$, since the bundle is locally a direct product of the neighboorhoods, V_x, of any point, $x \in X$, and the fibre, Y_x, over x. The cross sections are any graph (curve) in the bundle which matchs up with its endpoints after the twist. Any two cross sections will have at least one point in common since each curve must cross the line $y = L/2$ in the bundle (where L is the length of Y) in order to meet itself again since the bundle is twisted. There are two classes of homeomorphisms of the fibre Y_x with the typical fibre Y. If we move Y_x around the bundle, we can see that it matches with some of the other fibres via the identity map, but we will eventually have to flip Y_x upside down to match. In other words, if the typical fibre is represented by an arrow of orientation, then some of the fibres in the bundle will have to be flipped over first before they can be identified with the typical fibre. A good way of thinking about this bundle is to choose a fibre and call it your typical fibre. Then, find all the ways that you can move it around the bundle and identify it again with itself. If you do not move it at all, you can identify it with the identity map, and if you move it all the way around the S^2 and come back to itself, you will need to use the reflection map to identify it to itself. It is important to notice that you cannot rotate the fibre Y_x in the bundle to achieve the identification with itself. (See the twisted torus example, given later, to see why this fact is important.) Therefore, we have the identity and the reflection maps as our two homeomorphisms. The group of the bundle is the cyclic group of order two, $\mathbb{Z}_2$ generated by the reflection element g ($g^2 = e$). For example, if we take the base space, X, to be the circle at the center of the bundle so that the points in B can be labeled (θ, y), where $\theta \in (0, 2\pi]$ and $y \in [-L/2, L/2]$, then the homeomorphisms are just multiplication by 1 and multiplication by -1, i.e., $(\theta, y) \to (\theta, \pm y)$. We can summarize all of this discussion by writing our generic bundle as

$$\mathcal{B} = \{\text{base } X, \text{fibre } Y, \text{projection } \pi, \text{local trivialization } \phi, \text{structure group } G\} \tag{E.1}$$

so that the Möbius band is

$$\mathcal{B} = \left\{X = S^1, Y = [0, 1], \pi(x, y) = x, \phi(x, y) = (x, y), G = \mathbb{Z}_2\right\}.$$

For a cylinder, we would have

$$\mathcal{B} = \{X = S^1, Y = [0,1], \pi(x,y) = x, \phi(x,y) = (x,y), G = 1\}.$$

We see that the only aspect of the bundle which distinguishes the two is the structure group. The group of homeomorphisms which maps an arbitrary fibre to the one we have chosen as our "typical" fibre is a measure of how the bundle twists.

Klein bottle: The next example is the Klein bottle

$$\mathcal{B} = \{X = S^1, Y = S^1, \pi(x,y) = x, \phi(x,y) = (x,y), G = \mathbb{Z}_2\}.$$

The structure group is generated by reflection in the diameter of $Y = S^1$ (not equivalent to a rotation about the center).

Torus: The torus is

$$\mathcal{B} = \{X = S^1, Y = S^1, \pi(x,y) = x, \phi(x,y) = (x,y), G = 1\},$$

and we see that the difference between the Klein bottle and the torus lies only in the group of the bundle.

Twisted torus: The Twisted torus is constructed by twisting the fibre 180° about its center before identifying it. The bundle is, therefore,

$$\mathcal{B} = \{X = S^1, Y = S^1, \pi(x,y) = x, \phi(x,y) = (x,y), G = \mathbb{Z}_2\}.$$

We now have a bit of a problem. This bundle looks identical to the Klein bottle! It is in fact homeomorphic to a direct product (i.e., the torus). This is because we can identify a fibre with itself using the identity or with the reflection map. The reflection in this case is *not* the same as that of the Klein bottle. In the case of the twisted torus, it is the same as rotation by π. The difference is that this rotation generates a subgroup of G. The rotation group is the full symmetry group, $SO(2)$, of the fibre, and our $G = \mathbb{Z}_2$. in this case is a subgroup of the symmetry group of the fibres $G \subset SO(2)$ whereas for the Klein bottle, G is not a subgroup of $SO(2)$. For the Klein bottle, it is a reflection in a diameter of Y rather than a reflection through the center of Y and is, therefore, a subgroup of $O(2)$.

The moral of the twisted torus story is that if G is equivalent to the identity up to a transformation in the group of the fibre, then the bundle is homeomorphic to the product space (trivial bundle). To see this equivalence clearly, we notice that, for one bundle to be G-equivalent to another, the coordinate transformations must satisfy (Lemma 2.10 in Steenrod [92]):

$$g'_{ji} = \lambda_j(x)^{-1} g_{ji}(x) \lambda_i(x),$$

where $\lambda_j : V_j \to G$.

Hence, if the transition functions for the two bundles are in the same conjugacy class in G, the bundles are G-equivalent. In particular, the trivial bundle has $g'_{ji} = 1$

so that, if a bundle is G-equivalent to the trivial bundle, it must have

$$g_{ji}(x) = \lambda_j(x)\lambda_i(x)^{-1}$$

which is given in Section 4.3 of Steenrod [92].

In the case of the circle, take the base space S^1; cover it with two open neighborhoods, V_1 and V_2, i.e., semi-circles. The intersection of the two neighborhoods is the union of two small arcs, U and W. Define $g_{12} = 1$ in U and $g_{12} = -1$ in W. *This "-1" is the same as rotation by* 180^o *in* $G = SO(2)$. Take $\lambda_2(V_2) = e$, and take $\lambda_1(x)$ be a function over V_1 which forms a continuous path in $SO(2)$ from 1 to -1. Therefore, the twisted torus is $SO(2)$-equivalent to the product bundle.

In the case of the Klein bottle we cannot use $SO(2)$ since the group of the bundle is not even in $SO(2)$ so it cannot be $SO(2)$-equivalent to the product. If we think for a second that it might be $O(2)$ equivalent to the product bundle, we immediately see that it is not. We attempt the same construction that we did for the twisted torus by defining $g_{12} = 1$ in U and $g_{12} = -1$ in W *This "-1" is the same as the reflection operation in* $O(2)$. In order for the Klein bottle to be $O(2)$ equivalent to the trivial bundle, we need to find a continuous path, $\lambda_1(x)$, leading from 1 to -1 in $O(2)$. Finding such a path is impossible since 1 and -1 live in different connected components of $O(2)$ and there can be no continuous path connecting them. Hence, the twisted torus and the Klein bottle are not equivalent bundles in any sense.

Covering spaces: We will now discuss some examples which are not so typical. A covering space, B, of a space, X, is an example of a fibre bundle where the fibres, Y_x, are discrete. The projection is the covering map. The local trivialization is just $\phi(x, y) = y$, or in other words, a discrete point. If X is arcwise connected, then motion of a point x along a continuous curve from x_0 to x_1 in X is covered by a continuous motion in B from Y_{x_0} to Y_{x_1}. The action of closed curves in X can put any Y_x in correspondence with a typical fibre, Y_{x_0}. The correspondence depends only on the homotopy class of the curve. The fundamental group, $\pi_1(X)$, is then a group of permutations on Y. Any two correspondences between Y_x and Y will differ only by a permutation of Y corresponding to an element of $\pi_1(X)$. The structure group of the bundle is, therefore, a factor group of the fundamental group of the base space. We can write our bundle as

$$\mathcal{B} = \{X, Y = \{y_1, y_2, ..., y_n\}, \pi(x, y) = x, \phi(x, y) = y, H = \pi_1(X)/n\mathbb{Z}, G = S_n\},$$

where $H = \pi_1(X)/n\mathbb{Z}$ means that the structure group of the bundle which maps fibres to the typical fibre is a factor group. You can visualize this bundle by thinking of the double cover of $X = S^1$ given by S^1 which loops twice around. Moving once around the base space will move from the first point in Y_x above x to the next point above x; then, moving around the loop again will bring you back again. Hence, twice around the loop in the base space is the identity in the structure group. The fundamental group of the base space is $\mathbb{Z}$. The structure group is $\mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$. Notice also that S_2, which is the permutation group on two symbols, is simply isomorphic to $\mathbb{Z}_2$. We refer the interested reader to Section D.6. for a very elaborate discussion of covering spaces.

Coset spaces: Coset spaces form another example of a fibre bundle. Let B be a Lie group operating as a transitive group of transformations on a manifold X. We first define the projection map by selecting a point, $x_0 \in X$, and defining $\pi(b) = b(x_0)$. Thus, we have defined the projection of all elements of the Lie group by their action on x_0, $b(x_0) = x$ for some $x \in X$. With the projection map, we can see that the fibre Y is the subgroup of B which leaves x_0 fixed. The other fibres are just the left cosets of Y in B since the elements of B which do not leave x_0 fixed will move it. For example, let $b_1 \in B$ not in Y and let $b_0 \in Y$. Then, $b_1(x_0) = x_1$ for some $x_1 \in X$. However, since $b_0(x_0) = x_0$, we have $b_1 b_0(x_0) = b_1(x_0) = x_1$. $b_1 b_0 = b_2$ for some $b_2 \in B$ not in Y and $b_2(x_0) = x_1$ is then in the fibre over x_1 (It projects down to x_1 according to the definition of the projection operator.), hence the fibre over x_1 is given by $b_1 Y$, etc. It is clear that all of the fibres are given by the left cosets of Y in B. The correspondences, $Y \to Y_x$, are given by any $b \in Y_x$ in the same way by letting $y \to b\dot{y}$. However, any two such correspondences, say $y \to b \cdot y$ and $y \to b' \cdot y$, differ by the left translation of Y corresponding to $b^{-1}b'$. Thus, the structure group of the bundle coincides with Y and acts on Y by left translations. Finding a cross section is the problem of constructing a simply transitive continuous family of transformations. The bundle is

$$\mathcal{B} = \Big\{X, Y = \{b \in B | b(x_0) = x_0\}, \pi(b) = b(x_0), \phi(x,y) = (x,y), \\ G(\text{bundle}) = Y, G(\text{fibre}) = B \times B\Big\}. \tag{E.2}$$

We will return to this type of bundle later.

Tangent bundle: Define the Tangent Bundle of a manifold as follows. Let X be an n-dimensional, differentiable manifold, and let B be the set of all tangent vectors at all points of X. Define the projection to assign to each tangent vector its initial point. Then, Y_x is the tangent plane at x. Choosing a typical fibre, Y, we can form linear correspondences, $Y_x \to Y$, and the group of the bundle will be the general linear group, $GL(n)$, acting on Y. A cross section is a vector field over X.

The problems that one deals with in the theory of fibre bundles are numerous. We will elaborate on these problems as we go along. For starters, the existence of a cross section is an important question. If we wanted to construct a tensor field over a certain manifold, then we would be asking whether a cross section of a tensor bundle over that manifold exists and how to construct it. There are also many problems relating to classification of space, homology and homotopy of bundles, and differential geometry on bundles.

E.2. Coordinate bundle

We will now modify our definition of a fibre bundle to allow different constructions. First, we modify the structure group so that it becomes a topological group. (In the last two examples in Section E.1., the structure group also had a topology.) We then modify the local trivialization so that it becomes a coordinate function, and we choose open neighboorhoods from the topology and index them. We will need to choose specific open sets from the topology on X.

A *topological group* is a set which has a group structure and a topology such that

(a) g^{-1} is continuous for $g \in G$.

(b) $g_1 g_2$ is continuous simultaneously in g_1 and g_2, meaning that the map, $G \times G \to G$, given by $(g_1, g_2) \to g_1 g_2$ is continuous when $G \times G$ has the usual topology of a product space.

If G is a topological group and Y is a topological space, G is called a topological transformation group of Y relative to a map $\cdot : G \times Y \to Y$. If $\cdot$ is a continous map, $e \cdot y = y$ for e, the identity of G, and $g_1 \cdot (g_2 \cdot y) = (g_1 \cdot g_2) \cdot y$. We assume $\cdot$ is effective in that, if $g \cdot y = y$ for all $y \in Y$, then $g = e$. G is then isomorphic to a group of homeomorphisms of Y.

We can now define a *coordinate bundle* by the following ten axioms

(1) A space B called the bundle space.

(2) A space X called the base space.

(3) A map $\pi : B \to X$ called the projection.

(4) A space Y called the typical fibre.

(5) An effective topological transformation group, G, of Y called the group of the bundle.

(6) A family, $\{V_j\}$, of open sets, covering X and indexed by a set J, called coordinate neighborhoods.

(7) For each $j \in J$, a homeomorphism

$$\phi_j : V_j \times Y \to \pi^{-1}(V_j)$$

called the coordinate function.

(8) $\pi\phi_j(x, y) = x$ for $x \in V_j$ and $y \in Y$, which means that the coordinate function preserves the base point.

(9) If the map $\phi_{j,x} : Y \to \pi^{-1}(x)$ is defined by setting

$$\phi_{j,x}(y) = \phi_j(x, y),$$

then, for each pair, $i, j \in J$, and each $x \in V_i \cap V_j$, the homeomorphism,

$$\phi_{j,x}^{-1}\phi_{i,x} : Y \to Y$$

coincides with the operation of a unique element of G.

(10) For each pair $i, j \in J$, the map

$$g_{ij} : V_i \cap V_j \to G$$

defined by

$$g_{ij}(x) = \phi_{i,x}^{-1}\phi_{j,x} \tag{E.3}$$

is continuous. These transformations are called the coordinate transformations of the bundle.

The coordinate functions are maps from the direct product space formed by $V_j \times Y$ back into the bundle. From axiom (9), mapping to the bundle and then back to the typical fibre is the same as acting on the typical fibre with a group element. In other words, two different coordinate functions will only differ from each other by an element of the group. The coordinate transformations defined in axiom (10) relate the open sets to group elements and, therefore, directly tie the topology of the bundle to the group.

We immediately see from the definition, $g_{ij}(x) = \phi_{i,x}^{-1}\phi_{j,x}$, that the following identities are satisfied by g_{ij}:

(a) $g_{ij}g_{jk} = g_{ik}$

(b) $g_{ii} = e$, the identity in G

(c) $g_{ij} = (g_{ji})^{-1}$.

We now introduce a map that projects a point in the fibre bundle in a coordinate neighborhood to a point in the typical fibre. Let

$$\pi_j : \pi^{-1}(V_j) \to Y \quad \text{and} \quad \pi_j(b) = \phi_{j,x}^{-1}(b), \quad \text{where } x = \pi(b),$$

then π_j satisfies

(a) $\pi_j\phi_j(x, y) = y$

(b) $\phi_j(\pi(b), \pi_j(b)) = b$

(c) $g_{ji}(\pi(b)) \cdot \pi_i(b) = \pi_j(b)$ for $\pi(b) \in V_i \cap V_j$.

We can think of the coordinates of a point b in the bundle as being $(\pi(b), \pi_j(b))$ with respect to the neighborhood V_j since $\phi_j(\pi(b), \pi_j(b)) = b$. If we have two different coordinate systems for the bundle (two sets of coordinate functions), the union of them will also be a set of coordinate functions for the bundle. Two bundles that differ only in coordinate system are equivalent. If $\{\phi_j\}$ is one coordinate system and $\{\phi'_k\}$ is another, then

$$\bar{g}_{kj}(x) = (\phi'_{k,x})^{-1}\phi_{j,x}, \quad x \in V_j \cap V'_k$$

coincides with an element of the group. One can define a fibre bundle as an equivalence class of coordinate bundles since a fibre bundle has all open sets in the topology and not the restricted set of coordinate neighborhoods.

E.3. Maps between bundles

Let $\mathcal{B}$ and $\mathcal{B}'$ be two coordinate bundles having the same fibre and the same group. Define a map $h : B \to B'$ having the following properties:

(a) h carries each fibre, Y_x, of B homeomorphically onto a fibre, $Y_{x'}$, of B', thus inducing a continuous map between the base spaces, $\overline{h} : X \to X'$, such that

$$\pi' h = \overline{h}\pi.$$

(b) If $x \in V_j \cap \overline{h}^{-1}(V_k')$ and $h_x : Y_x \to Y_{x'}$ is the map induced by h, then the map

$$\overline{g}_{kj}(x) = (\phi'_{k,x})^{-1} h_x \phi_{j,x} = \pi'_k h_x \phi_{j,x} \tag{E.4}$$

of Y into Y corresponds to a group element in G.

(c) The map

$$\overline{g}_{kj} : V_j \cap \overline{h}^{-1}(V_k') \to G$$

so obtained is continous.

h maps the bundle B to the bundle B', where $\overline{h}$ is the restriction of this map to the base space and h_x is the map from the fibre Y_x to $Y_{x'}$. Looking at the trivializations of each bundle, we can see that h_x induces a map between the typical fibre to itself. First map from Y to B with $\phi_{j,x}$; map to B' with h_x; and then map back to the trivialization with $(\phi'_{k,x'})^{-1}$. We have then mapped Y onto itself. This map must correspond to one of the group elements. (The two bundles have the same group.) We have called this element $\overline{g}_{kj}(x)$. It is important to become fluent in manipulating the objects that we have defined. We will, therefore, work out a typical problem.

Problem: Prove the following relations:

$$\begin{aligned} \overline{g}_{kj}(x) g_{ji}(x) &= \overline{g}_{ki}(x), \quad x \in V_i \cap V_j \cap \overline{h}^{-1}(V_k') \\ g'_{lk}\left(\overline{h}(x)\right) \overline{g}_{kj}(x) &= \overline{g}_{lj}(x), \quad x \in V_j \cap \overline{h}^{-1}(V_k' \cap V_l') \end{aligned} \tag{E.5}$$

Solution: We use the definitions (E.3) and (E.4) to write the relations (E.5) as

$$\begin{aligned} \phi'^{-1}_{k,x'} h_x \phi_{j,x} \phi^{-1}_{j,x} \phi_{i,x} &= \phi'^{-1}_{k,x'} h_x \phi_{i,x} = \overline{g}_{ki}(x) \\ \phi'^{-1}_{l,\overline{h}(x)} \phi_{k,\overline{h}(x)} \phi^{-1}_{k,x'} h_x \phi_{j,x} &= \phi'^{-1}_{l,x'} h_x \phi_{j,x} = \overline{g}_{lj}(x), \end{aligned}$$

where we used the fact that $\overline{h}(x) = x'$.

E.3.1. Construction of fibre bundles

Given M, $\{U_i\}$, $g_{ij}(p)$, F, and G, we can construct a unique fibre bundle. To perform this construction, we need to find a unique π, E, and ϕ_i. Define $X \equiv \bigcup_i U_i \times F$.

Introduce an equivalence relation $\sim$ between $(p, f) \in U_i \times F$ and $(q, f') \in U_j \times F$ by $(p, f) \sim (q, f')$ if and only if $p = q$ and $f' = g_{ij}(p)f$. A fibre bundle is then defined by

$$E = X/\sim .$$

Denote an element of E by $[(p, f)]$. The projection is given by

$$\pi : [(p, f)] \mapsto p,$$

and the local trivialisation, $\phi_i : U_i \times F \to \pi^{-1}(U_i)$, is given by

$$\phi_i : (p, f) \mapsto [(p, f)].$$

If $E \xrightarrow{\pi} M$ and $E' \xrightarrow{\pi'} M'$ are two fibre bundles, a smooth map, $\overline{f} : E' \to E$, is called a bundle map if it maps each fibre, F'_p, of E' onto a single fibre, F_q, of E. The map, $\overline{f}$, naturally induces a smooth map, $f : M' \to M$, such that $f(p) = q$. If $\overline{f}$ is a diffeomorphism and the induced map, f, is the identity map, then the two bundles are equivalent.

E.3.2. Pullbacks and homotopy

Let $E \xrightarrow{\pi} M$ be a fibre bundle with typical fibre F. If a map, $f : N \to M$, is given, the pair (E, f) defines a new fibre bundle over N with the same fibre, F. Let f^*E be a subspace of $N \times E$ consisting of points (p, u) such that $f(p) = \pi(u)$. $f^*E \equiv \{(p, u) \in N \times E | f(p) = \pi(u)\}$ is called the pullback of E by f. The fibre, F_p, of f^*E is just a copy of the fibre $F_{f(p)}$ of E. If we define $f^*E \xrightarrow{\pi_1} N$ by $\pi_1 : (p, u) \mapsto p$ and $\pi_2 : f^*E \to E$ by $(p, u) \mapsto u$, the pullback, f^*E, may be endowed with the structure of a fibre bundle. If $N = M$ and $f = id_M$, the two fibre bundles, f^*E and E, are equivalent.

For example, let M and N be differentiable manifolds with $\dim M = \dim N = m$, and let $f : N \to M$ be a smooth map. The map, f, induces a map $\pi_2 : TN \to TM$ such that the following diagram commutes

$$\begin{array}{ccc} TN & \xrightarrow{\pi_2} & TM \\ \pi_1 \downarrow & & \downarrow \pi \\ N & \xrightarrow[f]{} & M \end{array}$$

Let $W = W^\nu \frac{\partial}{\partial y^\nu}$ be a vector of T_pN and $V = V^\mu \frac{\partial}{\partial x^\mu}$ be the corresponding vector in $T_{f(p)}M$. If TN is a pullback bundle, $f^*(TM)$, π_2 maps T_pN to $T_{f(p)}M$ diffeomorphically. This diffeomorphism is possible if and only if π_2 has the maximal rank, m, at each point of N. Let $\phi(f(p)) = (f^1(y), ..., f^m(y))$ be the coordinates of $f(p)$ in a chart (U, ϕ) of M, where $y = \psi(p)$ are the coordinates of p in a chart (V, ψ) of N. The maximal rank condition is given by $\det(\partial f^\mu(y)/\partial y^\nu)_p \neq 0$ for $p \in N$. We state the following theorem without proof.

Theorem: Let $E \xrightarrow{\pi} M$ be a fibre bundle with fibre F, and let f and g be homotopic maps from N to M. Then, f^*E and g^*E are equivalent bundles over N.

E.3.3. Vector bundles and line bundles

A quasi-vector bundle is denoted by $\xi = (E, \pi, X)$, where E is the total space; X is the base space; and π is the continuous projection map, $\pi : E \to X$. ξ is a finite dimensional k-vector space, E_x, for every point, x, in the base space X which has a topology defined on the disjoint union of the fibres, E_x, over each x, and k is one of the fields, $\mathbb{R}, \mathbb{C}$, or $\mathbb{H}$. These are the reals, complexes, or quaternions, respectively.

Definition: A *general morphism* from $\xi = (E, \pi, X)$ to $\xi' = (E', \pi', X')$ is given by a pair (f, g) of continuous maps, $f : X \to X'$ and $g : E \to E'$, such that

- The map $g_x : E_x \to E'_{f(x)}$ from the fibre over x one bundle to the fibre over $f(x)$ in the other is k-linear.

- The following diagram commutes:

$$\begin{array}{ccc} E & \xrightarrow{g} & E' \\ \pi \downarrow & & \downarrow \pi' \\ X & \xrightarrow{f} & X' \end{array}$$

In this way, we can construct a category with objects being quasi-vector bundles and arrows being general morphisms. If the base spaces, X and X', of the two bundles are the same, then f is simply the identity map and can denote the morphism by the single map, g. These quasi-vector bundles form a subcategory.

For example, let E'' be the quotient of $E' = S^1 \times \mathbb{R}$ by the equivalence relation $(x, y) \sim (-x, -y)$. Then, E'' is the total space of a quasi-vector bundle over $P_1(\mathbb{R})$ called the infinite Möbius band. Indentify $P_1(\mathbb{R})$ with S^1 by the map $z \mapsto z^2$. Then E'' is also the quotient of $I \times \mathbb{R}$ by $(0, y) \sim (1, -y)$. Restricting y to have norm less than 1 gives the classical Möbius band.

Notice that the bundles, E' and E'', are not isomorphic since we would then have an isomorphism, $g : E' \to E''$, and $E' - X'$ would then be homeomorphic to $E'' - X''$, where X' and X'' are the base spaces and are both circles $\{(x, 0) \in E' \text{ or } (E'')|x \in S^1\}$. We see that $E'' - X''$ is connected, whereas $E' - X'$ is two cylinders and is not connected (their zeroth homotopy groups $\pi_0(E'' - X'') \neq \pi_0(E' - X')$ so that they cannot be isomorphic).[1].

Definition: A *vector bundle* is a quasi-vector bundle which is locally trivial. For every point x in the base space X, there exists a neighborhood which is isomorphic to a trivial bundle (i.e., a direct product).

For trivial vector bundles with the same base, $E = X \times V$ and $E' = X \times V'$, we have a morphism, $g : E \to E'$, for each point x in the base space X. Therefore, g induces a linear map, $g_x : V \to V'$, and we define a $\check{g} : X \to \mathcal{L}(V, V')$ so that $\check{g}(x) = g_x$.

[1]This example shows how it is not obvious as to whether a given quasi-vector bundle is isomorphic to a trivial bundle.

Theorem: The map, $\check{g} : X \to \mathcal{L}(V, V')$, is continuous relative to the natural topology of $\mathcal{L}(V, V')$. Conversely, let $h : X \to \mathcal{L}(V, V')$ be a continuous map, and let $\hat{h} : E \to E'$ be the map which induces $h(x)$ on each fibre. Then $\hat{h}$ is a morphism of quasi-vector bundles.

Proof: To prove this theorem, we choose a basis $e_1, ..., e_n$ of V and a basis $\epsilon_1, ..., \epsilon_p$ of V'. With respect to this basis, g_x may be regarded as the matrix, $\alpha_{ij}(x)$, where α_{ij} is the ith coordinate of the vector $g_x(e_j)$. Hence, the function, $x \mapsto \alpha_{ij}(x)$, is obtained from the composition of the following continuous maps

$$X \overset{\beta_j}{\to} X \times V \overset{g}{\to} X \times V' \overset{\gamma}{\to} V' \overset{p_i}{\to} k,$$

where $\beta_j(x) = (x, e_j)$, $\gamma(x, v') = v'$, and p_i is the ith projection of $V' \supseteq k^p$ on k. Since the functions, $\alpha_{ij}(x)$, are continuous the map, $\check{g}$, which they induce, is also continuous according to the definition of the topology of $\mathcal{L}(V, V')$.

Conversely, let $h : X \to \mathcal{L}(V, V')$ be a continuous map. Then, $\hat{h}$ is obtained from the composition of the continuous maps

$$X \times V \overset{\delta}{\to} X \times \mathcal{L}(V, V') \times V \overset{\epsilon}{\to} X \times V',$$

where $\delta(x, v) = (x, h(x), v)$ and $\epsilon(x, u, v) = (x, u(v))$. Hence, $\hat{h}$ is continuous and defines a morphism of quasi-vector bundles.

□

We see that any continuous map from a space X to the space of k-linear functions between vector spaces will induce a morphism of quasi-vector bundles over X. If $f : X' \to X$ is any continuous map where X' could be a subspace of X or even a different section of the total space, E, we can let $E'_{x'} = E_{f(x')}$ to make a new bundle called the fibre product, $X' \times_X E$. $X' \times_X E$ is then the subset of $X' \times E$ given by the pairs (x', e) such that $f(x') = \pi(e)$. The fibre product is essentially replacing X by X' in E. The new bundle has $\pi'(x', e) = x'$ and is a quasi-vector bundle over X'. For example, we could redefine the zero section of a bundle in this way, or we could rotate or transform the base space X and form a new bundle by leaving the fibres over each $x \in X$ fixed. Then, operating with f on X' is given by projecting the fibre over x' with the projection map, π, of ξ, giving x, i.e., $f(x') = \pi(e)$. Hence, given a map from X' to X and a bundle over X, we can construct a bundle over X'. If $X' \subset X$ and f is the inclusion map, then $f^*(\xi) = \xi|_{X'}$. The new bundle is written $\xi' = (E', \pi', X') = f^*(\xi) = f^*(E)$, where $f^*(E)$ is the inverse image of ξ by f. We also have $(f \cdot f')^*(\xi) = f'^*(f^*(\xi))$. Given $\alpha : E \to F$, a morphism between bundles E and F, we can define a morphism $\alpha' = f^*(\alpha)$, where $\alpha'_{x'} = \alpha_{f(x')}$ from bundles E' to F'.

The correspondence, $E \mapsto f^*(E)$ and $\alpha \mapsto f^*(\alpha)$, induces a functor between the category of quasi-vector bundles over X and the category of quasi-vector bundles over X'. We can devise maps to add to vectors in the same fibre and to multiply a vector by a constant (from k) by $s : E \times_X E \to E$ by $s(e, e') = e + e'$ and $p : k \times E \to E$ by $p(\lambda, e) = \lambda e$.

Define the *rank* of a vector bundle as a function, $r(x) : X \to \mathbb{N}$, which assigns to each $x \in X$ a natural number which is the dimension of the fibre space over x. If the base space is connected, then $r(x)$ is a constant.

Theorem (clutching of bundles): Let (U_i) be an open cover of a space X. Let $\xi_i = (E_i, \pi_i, U_i)$ be a vector bundle over each U_i. Let $g_{ji} : \xi_i|_{U_i \cap U_j} \to \xi_j|_{U_i \cap U_j}$ be isomorphisms which satisfy the compatibility condition, $g_{ki}|_{U_i \cap U_j \cap U_k} = g'_{kj} \cdot g'_{ji}$, where $g'_{kj} = g_{kj}|_{U_i \cap U_j \cap U_k}$ and $g'_{ji} = g_{ji}|_{U_i \cap U_j \cap U_k}$. Then, there exists a vector bundle, ξ, over X and isomorphisms, $g_i : \xi_i \to \xi|_{U_i}$, such that the following diagram commutes

$$\begin{array}{ccc} \xi_i|_{U_i \cap U_j} & \xrightarrow{g_{ij}} & \xi_j|_{U_i \cap U_j} \\ g_i|_{U_i \cap U_j} \searrow & & \swarrow g_j|_{U_i \cap U_j} \\ & \xi|_{U_i \cap U_j} & \end{array}$$

Let S^n be the sphere in $\mathbb{R}^{n+1}$ which consists of all vectors of length 1, i.e., the set of points, $x = (x_1, ..., x_{n+1})$, such that $||x||^2 = \sum_{i=1}^{n+1}(x_i)^2 = 1$. Let S^n_+ denote the upper hemisphere (the set of points in S^n with $x_{n+1} \geq 0$). Let S^n_- denote the lower hemisphere. The overlap of these two subsets is the equator which is $S^{n-1} = S^n_+ \cap S^n_-$ and has $x_{n+1} = 0$. Now, let $f : S^{n-1} \to GL_p(k)$ be a continuous map so that, for each point on the equator, we associate a $p \times p$ matrix with entries from k. Then, there is a bundle, E_f, over S^n which is naturally associated with f. It is obtained from the clutching of the trivial bundles, $E_1 = S^n_+ \times k^p$ and $E_2 = S^n_- \times k^p$, on the overlap by the "transition function" $g_{12} = \hat{f} : S^{n-1} \times k^p \to S^{n-1} \times k^p$. (Note that g_{11} and g_{22} are the identity map.) Hence, a bundle is defined over the entire space when we have defined a trivial bundle on each half and a collection of transition functions at the overlap (one transition function for each point in the overlap) which map the fibres of one bundle to the fibres of the other and glue them together along the equator.

Theorem: Let $\Phi^k_n(X)$ be the set of isomorphism classes of k-vector bundles of rank n over the topological space X. Then, $\Phi^k_n(X)$ is naturally isomorphic to the set $H^1(X; G)$, where $G = GL_n(k)$.

Proof: We define two maps

$$h : \Phi^k_n(X) \to H^1(X; G) \text{ and } h' : H^1(X; G) \to \Phi^k_n(X)$$

such that $h' = h^{-1}$.

Let $\xi = (E, \pi, X)$ be a vector bundle, and let (U_i) be a trivialisation cover of X. (The bundle is trivial over each set, U_i, in the cover.) Choose isomorphisms, $\phi_i : U_i \times k^n \to E_{U_i}$, and let g_{ij} be the map from $U_i \cap U_j$ to $G = GL_n(k)$ defined by $g_{ij}(x) = (\phi_j)_x^{-1} \cdot (\phi_i)_x$. In this way, we obtain a G-cocycle on X. Its class in the set $H^1(X; G)$ is independent of the choice of the trivialization cover and of the ϕ_i. In fact, if (V_r, h_{sr}) is another cocycle associated with trivializations, $\psi_r : V_r \times k^n \to E_{V_r}$, let

$g_i^r(x) = (\psi_r)_x^{-1} \cdot (\phi_i)_x$. Then, if $x \in U_i \cap U_j \cap V_r \cap V_s$, we have

$$\begin{aligned} g_j^s(x) \cdot g_{ji}(x) \cdot (g_i^r(x))^{-1} &= (\psi_s)_x^{-1} \cdot (\phi_j)_x \cdot (\phi_j)_x^{-1} \cdot (\phi_i)_x \cdot (\phi_i)_x^{-1} \cdot (\psi_r)_x \\ &= (\psi_s)_x^{-1} \cdot (\psi_r)_x = h_{sr}(x), \end{aligned}$$

showing that h is well defined.

Conversely, let (U_i, g_{ij}) be a G-cocycle, and let E by the vector bundle over X obtained by clutching the trivial bundles, $E_i = U_i \times k^n$, with the "transition functions," g_{ji}. Then, the class of E in $\Phi_n^k(X)$ depends only on the class of the cocycle in $H^1(X;G)$. In fact, consider a cocycle (V_r, h_{rs}) equivalent to (U_i, g_{ji}) and let F be the vector bundle obtained from this cocycle by clutching the trivial bundles, $F_r = V_r \times k^n$. Let $\alpha : E \to E'$ by the unique morphism which makes the following diagram commute for each pair (i, r):

$$\begin{array}{ccc} E_i|_{U_i\cap V_r} & \xrightarrow{\hat{g}_i^r} & F_r|_{U_i\cap V_r} \\ g_i|_{U_i\cap V_r} \downarrow & & \downarrow h_r|_{U_i\cap V_r} \\ E|_{U_i\cap V_r} & \xrightarrow{\alpha|_{U_i\cap V_r}} & F|_{U_i\cap V_r}, \end{array}$$

where the morphisms, g_i and h_r, are the morphisms defined by the clutching theorem. To see that α is well defined, we note the following identities for $x \in U_i \cap U_j \cap V_r \cap V_s$:

$$\begin{aligned} h_{sr}(x) &= g_j^s(x) \cdot g_{ij}(x) \cdot (g_i^r(x))^{-1} \\ g_j^s(x) &= h_{sr}(x) \cdot g_i^r(x) \cdot g_{ij}(x) \\ g_j^s(x) &= (h_s(x))^{-1} \cdot h^r(x) \cdot g_i^r(x) \cdot (g_i(x))^{-1} \cdot g_j(x) \\ h_r(x) \cdot g_i^r(x) \cdot (g_i(x))^{-1} &= h_s(x) \cdot g_j^s(x) \cdot (g_j(x))^{-1} \end{aligned}$$

showing that h' is also well defined. The fact that $h \cdot h'$ and $h' \cdot h$ are the identities of $H^1(X;G)$ and $\Phi_n^k(X)$, respectively, follows directly from their definitions.

□

Consider two continuous maps, f_0 and f_1, from S^{n-1} to $GL_p(k)$ which are homotopic (Appendix D). Let $\alpha : S^{n-1} \to GL_p(k)$ be the map defined by $\alpha(x) = (f_1(x))^{-1} \cdot f_0(x)$. Then, α is homotopic to 1 since there exists a continuous homotopy, $\beta : S^{n-1} \times I \to GL_p(k)$, such that $\beta(x,0) = \alpha(x)$ and $\beta(x,1) = 1$. We parametrize the upper half hemisphere, S_+^n of S^n, by writing each element, w, of S_+^n in the form $v\cos\theta + e_{n+1}\sin\theta$ with $v \in S^{n-1}$. We use β to define $\gamma : S_+^n \to GL_p(k)$ given by $\gamma(w) = \beta(v, \sin\theta)$. This map is well defined and continuous even for $\theta = \frac{\pi}{2}$ because $\beta(x,t)$ converges to 1 uniformly in x when t converges to 1. Hence, $E_{f_0} = E_{f_1\alpha}$ is isomorphic to E_{f_1}.

If we restrict our attention to maps, $f : S^{n-1} \to GL_p(k)$, such that $f(e) = 1$, where $e = (1, 0, ..., 0)$ is the base point of S^{n-1}, the previous discussion shows that the correspondence, $f \mapsto E_f$, defines a map from $\pi_{n-1}(GL_p(k))$ to $\Phi_p^k(S^n)$. On the other hand, $\pi_0(GL_p(k))$ acts on $\pi_{n-1}(GL_p(k))$ by the map defined on representatives by $(a, f) \to a \cdot f \cdot a^{-1}$. Since the vector bundles E_f and $E_{a\cdot f\cdot a^{-1}}$ are isomorphic, we can define a bijection from the quotient set, $\pi_n(GL_p(k))/\pi_0(GL_p(k))$, to $\Phi_p^k(S^n)$. Since $\Phi_p^k(X)$ isomorphic to $H^1(X;G)$, we have that $H^1(X;G) \approx \pi_n(G)/\pi_0(G)$.

Examples

If $k = \mathbb{C}$, the group $GL_p(k) = GL_p(\mathbb{C})$ may be regarded as the topological product of $U(p)$ by $\mathbb{R}^{p^2}$. Since $U(p)$ is arcwise connected, $\pi_0(U(p)) = \pi_0(GL_p(\mathbb{C})) = 0$. Hence, $\Phi_p^{\mathbb{C}}(S^n) \approx \pi_{n-1}(U(p))$. We have the locally trivial fibration

$$U(p) \to U(p+1) \to S^{2p+1},$$

hence the exact sequence of homotopy groups

$$\pi_{i+1}(S^{2p+1}) \to \pi_i(U(p)) \to \pi_i(U(p+1)) \to \pi_i(S^{2p+1}).$$

Since $\pi_j(S^r) = 0$ for $j > r$, it follows that, for $p > \frac{i}{2}$, we have $\pi_i(U(p)) \approx \pi_i(U(p+1))$.

If $k = \mathbb{R}$, the group $GL_p(k) = GL_p(\mathbb{R})$ may be regarded as the topological product, $O(p) \times \mathbb{R}^{p(p+1)/2}$. Hence, $\pi_i(GL_p(\mathbb{R})) \approx \pi_i(O(p))$ and $\pi_0(GL_p(\mathbb{R})) \approx \mathbb{Z}/2$. The homotopy exact sequence associated with the locally trivial fibration,

$$O(p) \to O(p+1) \to S^p,$$

is

$$\pi_{i+1}(S^p) \to \pi_i(O(p)) \to \pi_i(O(p+1)) \to \pi_i(S^p),$$

showing that $\pi_i(O(p)) \approx \pi_i(O(p+1))$ and that $\pi_i(O(p)) \approx \text{inj lim}\,\pi_i(O(m))$ when $p > i+1$. We will see later that the groups, $\pi_i = \text{inj lim}\ \pi_i(O(m))$, are respectively isomorphic to $\mathbb{Z}/2, \mathbb{Z}/2, 0, \mathbb{Z}, 0, 0, 0$ when $i = 0, 1, 2, 3, 4, 5, 6, 7 \bmod 8$. Hence, the homotopy groups, $\pi_i(O(p)) \approx \pi_i(GL_p(\mathbb{R}))$, are completely known for $p > i+1$. Moreover, in this case, the action of $\pi_0(GL_p(\mathbb{R})) = \mathbb{Z}/2$ is trivial. When p is odd, we have nothing to prove since $\det(-1) = -1$. When p is even, the isomorphism between $\pi_i(O(p))$ and $\pi_i(O(p+1))$ is compatible with the action. Therefore, $\pi_i(O(p))/(\mathbb{Z}/2) \approx \pi_i(O(p+1))/(\mathbb{Z}/2) \approx \pi_i(O(p+1)) \approx \pi_i(O(p))$.

The sections of a vector bundle, $s(x)$, can be point-wise added and multiplied in that, for local sections, $s(x)$ and $s'(x)$, we have new local sections given by $s(x) + s'(x)$ and $[fs](x) = s(x)f(x)$ for a smooth, continuous function, $f(x)$, on M.

There is a single global section of any vector bundle. It is called the zero section and is defined by $s(x) = 0$. It is possible that no other global sections may exist, as can be seen by the fact that the zero section can be identified with the base manifold which is globally defined.

Example: The Möbius band

Let E be a fibre bundle, $E \xrightarrow{\pi} S^1$, with typical fibre $F = [-1, 1]$; let $U_1 = (0, 2\pi)$ and $U_2 = (-\pi, \pi)$ be an open cover of S^1; and let $A = (0, \pi)$ and $B = (\pi, 2\pi)$ be the intersection $U_1 \cap U_2$. The local trivializations, ϕ_1 and ϕ_2, are given by

$$\phi_1^{-1}(u) = (\theta, t) \quad \text{and} \quad \phi_2^{-1}(u) = (\theta, t)$$

for $\theta \in A$ and $t \in F$. The transition function, $g_{12}(\theta)$, $\theta \in A$, is the identity map, $g_{12}(\theta) : t \mapsto t$. We have two choices on B:

$$\begin{array}{lll} (I) & \phi^{-1}(u) = (\theta, t) & \phi_2^{-1}(u) = (\theta, t) \\ (II) & \phi^{-1}(u) = (\theta, t) & \phi_2^{-1}(u) = (\theta, -t) \end{array}$$

In case (I), the two local pieces of the bundle are glued together to form a cylinder, and in case (II), we have $g_{12}(\theta) : t \to -t$, $\theta \in B$, and we obtain the Möbius strip. The cylinder, therefore, has the trivial structure group $G = \{e\}$, where e is the identity map from F to F. The Möbius strip has $G = \{e, g\}$, where $g : t \mapsto -t$. Since $g^2 = e$, we find that $G \cong \mathbb{Z}_2$. The cylinder is a trivial bundle, whereas the möbius strip is not. Notice that, in this case, we use a discrete group as the structure group rather than a Lie group.

Example: Tangent and cotangent bundles

The tangent and cotangent bundles are vector bundles with transition functions given by the change of basis functions in the overlapping neighborhoods. The complexified tangent and cotangent bundles, $TM \otimes \mathbb{C}$ and $T^*M \otimes \mathbb{C}$, of a real manifold, M, are defined by permitting the coefficients of the frames, $\{\partial/\partial x_i\}$ and $\{dx_i\}$, to be complex. If M is a complex manifold with local complex coordinates, z_j, we define the complex tangent bundle, T_cM, to be the sub-bundle of $TM \otimes \mathbb{C}$ which is spanned by the holomorphic tangent vectors, $\partial/\partial z_j$. The complex dimension of T_cM is half of the real dimension of TM. If we forget the complex structure on T_cM and consider it as a real bundle, then T_cM is isomorphic to TM.

If V is a vector space, we define the dual space, V^*, to be the set of linear functionals. If V and W are vector spaces, we can define the Whitney sum, $V \oplus W$ (as the set of all pairs (v, w)), and the tensor product, $V \otimes W$. These and other constructions can be carried over to the vector bundle case as follows.

If E and F are vector bundles over M, the fibre of the Whitney sum bundle, $E \oplus F$, is obtained by taking the Whitney direct sum of the fibres of E and F at each point, $x \in M$. If $\dim E = j$ and $\dim F = k$ and if the transition functions of E and F are the $j \times j$ matrices, Φ, and the $k \times k$ matrices, Ψ, then the transition matrices of $E \oplus F$ are just the $(j + k) \times (j + k)$ matrices, $\Phi \oplus \Psi$, given by

$$\begin{pmatrix} \Phi & 0 \\ 0 & \Psi \end{pmatrix} = \Phi \oplus \Psi.$$

If $\{\mathbf{e}_i\}$ and $\{\mathbf{f}_i\}$ are local frames for E and F, then $\{\mathbf{e}_1, ..., \mathbf{e}_j, \mathbf{f}_1, ..., \mathbf{f}_k\}$ is a local frame for $E \oplus F$ and $\dim(E \oplus F) = \dim E + \dim F$.

The tensor product bundle, $E \otimes F$, is obtained by taking the tensor product of the fibres at each point in M. The transition matrices for $E \otimes F$ are obtained by taking the tensor product of the transition functions of E and F. A local frame for $E \otimes F$ is given by $\{\mathbf{e}_i \otimes \mathbf{f}_j\}$ so $\dim(E \otimes F) = \dim E \dim F$. One can also construct the bundle of antisymmetric p-tensors, $\Lambda^p(E)$, and the bundle of symmetric p-tensors, $S^p(E)$, as sub-bundles of $\mathfrak{S}^p(E)$. One can also construct a complimentary bundle, $\overline{E}$, such that the Whitney sum $E \oplus \overline{E} \cong M \times \mathbb{C}^l$, is a trivial bundle with fibre $\mathbb{C}^l$.

A fibre metric is a pointwise inner product between two sections of a vector bundle which allows us to define the length of a section at a point x of the base. In local coordinates, a fibre metric is a positive definite, symmetric matrix, $h_{ij}(x)$. The inner product of two sections is then

$$(s, s') = h_{ij}(x) z^i(x) \overline{z}'^j(x), \tag{E.6}$$

where $\overline{z}$ denotes complex conjugation if the fibre is complex. Under a change of frame, we find

$$h \to (\Phi^t)^{-1} h \overline{\Phi}^{-1}.$$

Example: Tangent and cotangent bundles of S^2

Let $U_+ = S^2 - \{(0,0,-\frac{1}{2})\}$ and $U_- = S^2 - \{(0,0,+\frac{1}{2})\}$ be spheres of unit diameter minus the south and north poles, respectively. We stereographically project these two neighborhoods to the plane defining coordinates $x_+ = (x, y)$ and $x_- = (x', y')$. Let $r^2 = x^2 + y^2$. The standard metric is then

$$ds^2 = \frac{1}{(1+r^2)^2}(dx^2 + dy^2).$$

The U_- coordinates are related to the U coordinates by the inversion $x_- = \frac{1}{r^2} x_+$, which gives

$$dx_- = \frac{1}{r^4}\left(r^2 dx_+ - 2x_+(x_+ \cdot dx_+)\right).$$

The transition functions, $|\partial x_- / \partial x_+|$ for $T^*(S^2)$ are given on the intersection by

$$\Phi_{U_- U_+} = \frac{1}{r^4}(\delta^{ij} r^2 - 2x^i x^j)$$

We introduce polar coordinates on $\mathbb{R}^2 - (0,0)$ and restrict to $r = 1$ so that we are effectively working on the equator, S^1, of the sphere. Then, we find

$$\Phi_{U_- U_+}(\cos\theta, \sin\theta) = \begin{pmatrix} -\cos 2\theta & -\sin 2\theta \\ -\sin 2\theta & \cos 2\theta \end{pmatrix}.$$

The inverse, transposed matrix is the transition matrix for $T(S^2)$. $\Phi_{U_- U_+}$ represents a non-trivial map from $S^1 \to GL(2, \mathbb{R})$. This map is just twice the generator of $\pi_1(GL(2,\mathbb{R})) \cong \mathbb{Z}$.

The bundles, $T(S^2)$ and $T^*(S^2)$, are non-trivial and isomorphic. Let I denote the trivial bundle over S^2. We can identify I with the normal bundle of S^2 in $\mathbb{R}^3$ so $T(S^2) \oplus I = T(\mathbb{R}^3) = I^3$ is the trivial bundle given above as a map from S^1 to $GL(3, \mathbb{R})$. It is still twice the generator. Since $\pi_1(GL(3,\mathbb{R})) \cong \mathbb{Z}_2$, the map is null homotopic and $T^*(S^2) \oplus I$ is trivial.

Example: The natural line bundle over $P_n(\mathbb{C})$

Recall that $P_n(\mathbb{C})$ is the set of lines through the origin in $\mathbb{C}^{n+1}$. Let $I^{n+1} = P_n(\mathbb{C}) \times \mathbb{C}^{n+1}$ be the trivial bundle of dimension $n+1$ over $P_n(\mathbb{C})$. We denote a point of I^{n+1} (I is the bundle of normal vectors.) by the pair (p, z); scalar multiplication and addition are performed on the second factor while leaving the first factor unchanged. Let L be the sub-bundle of I^{n+1} defined by

$$L = \{(p, z) \in I^{n+1} = P_n(\mathbb{C}) \times \mathbb{C}^{n+1} \text{ such that } z \in p\}.$$

In other words, the fibre of L over a point p of $P_n(\mathbb{C})$ is just the set of points in $\mathbb{C}^{n+1}$ which belong to the line p. We define coordinates, $\zeta_i^{(j)} = z_j/z_i$, on neighborhoods, $U_j = \{p : z_j(p) \neq 0\}$. On U_j, we define the section s_j to L by

$$s_j(p) = (\zeta_0^{(j)}(p), ..., 1, ..., \zeta_n^{(j)}(p)).$$

The transition functions are 1×1 complex matrices (scalars).

$$s_k(p) = (\zeta_k^{(j)})^{-1} s_j(p)$$

Since the transition functions are holomorphic, L is a holomorphic line bundle.

The dual bundle L^* has sections,[2] s_j^*, so that $s_j^*(s_j) = 1$. Since we have a line bundle, a frame is given by a single section. The transition functions act as

$$s_k^* = s_j^* \zeta_k^{(j)}.$$

We now interpret the $\{s_j^*\}$ as homogeneous coordinates on $P_n(\mathbb{C})$ since it is clear that $s_j^*(p) = z_j$. Also note that $s_j^* = 0$ whenever $z_j = 0$, i.e., whenever p is not in the neighborhood U_j. The ratio of these global sections may be used to define the inhomogeneous coordinates, $\zeta_i^{(j)}$.

L^* has global holomorphic sections, s_j^*, the zeros of which lie in the complement of U_j, which is just a projective space of dimension $(n-1)$. The bundle L does not have any global holomorphic sections. Since $s_j s_j^* = 1$ and $s_j^* = 0$ on the complement of U_j, s_j must blow up like z_j^{-1} on the complement of U_j. The s_j are, therefore, *meromorphic* sections of L.

We define the line bundle, L^k, by

$$\begin{aligned} &L^* \otimes \cdots \otimes L^* \text{ if } k < 0 \\ &L^0 = I \text{ the trivial line bundle} \\ &L \otimes \cdots \otimes L \text{ if } k > 0. \end{aligned}$$

Because $L \otimes L^* = I, L^j \otimes L^k = L^{j+k}$ for all integers j, k. Any line bundle over $P_n(\mathbb{C})$ is isomorphic to L^k for some uniquely defined integer k. The integer k is related to the first Chern class of L^k.

[2]The subscripts here refer to different coordinate systems and not to elements of a frame.

Let $T_c(P_n(\mathbb{C}))$ and $T_c^*(P_n(\mathbb{C})) = \Lambda^{1,0}(p_n(\mathbb{C}))$ be the complex tangent and cotangent spaces. Then,

$$I \oplus T_c(P_n(\mathbb{C})) = \overbrace{L^* \oplus \cdots \oplus L^*}^{n+1}$$
$$I \oplus T_c^*(P_n(\mathbb{C})) = \underbrace{L \oplus \cdots \oplus L}_{n+1}$$

This identity does not preserve the holomorphic structures but is an isomorphism between complex vector bundles.

Example: Combining the last two examples

Using the relations $S^2 = P_1(\mathbb{C})$ and $T(S^2) = T_c(P_1(\mathbb{C}))$, we can combine the last two examples for $n = 1$ to show

$$T^*(S^2) = L \otimes L, \quad T(S^2) = L^* \otimes L^*.$$

We prove these relationships by recalling that we may choose complex coordinates on S^2 of the form $\zeta_0 = z_1/z_0$ on U_0 and $\zeta_1 = z_0/z_1$ on U_1. We choose the basis of $T^*(S^2)$ to be $d\zeta_0$ on U_0 and $-d\zeta_1$ on U_1. The transition functions are given by $(-d\zeta_1) = \zeta_0^{-2}(d\zeta_0)$. The local sections $s_0 = (1, \zeta_0)$ and $s_1 = (\zeta_0^{-1}, 1)$ of L give the transition function, $s_1 = \zeta_0^{-1} s_0$. The $L \otimes L$ transition functions are, thus,

$$s_1 \otimes s_1 = \zeta_0^{-2} s_0 \otimes s_0,$$

so $T^*(S^2)$ and $L \otimes L$ are isomorphic bundles. The isomorphism between $T(S^2)$ and $L^* \otimes L^*$ is obtained by dualizing the preceding argument.

E.4. The principal bundle

A vector bundle is a fibre bundle with fibre, F, is a linear vector space and with transition functions belonging to the general linear group of F. A principal bundle, P, is a fibre bundle with a Lie group, G (which is also a manifold) as its fibre. The transition functions of P belong to G and act on G by left multiplication. We can define a right action of G on P because left and right multiplication commute. This action is a map, $P \times G \to P$, which commutes with the projection π. $\pi(p \cdot g) = \pi(p)$ for any $p \in P$ and $g \in G$. Note that we can reverse the roles of left and right multiplication if desired.

We can construct a principal bundle known as a frame bundle or as the associated principal bundle from a given vector bundle, E. The fibre G_x of P at x is the set of all frames of the vector space, F_x, which is the fibre of E over the point x.

Consider the case of $F = \mathbb{C}^k$. The fibre G of the frame bundle P is the collection of the $k \times k$ non-singular matrices which form the group $G = GL(k, \mathbb{C})$. (i.e., G is the structure group of the vector bundle E.)

The associated principal bundle, P, has the same transition functions as the vector bundle, E (i.e., $GL(k,\mathbb{C})$ acting on the fibre G by left multiplication). The right action of the group $G = GL(k,\mathbb{C})$ on the principal G-bundle, P, takes a frame $e = \{e_1, ..., e_k\}$ to a new frame in the same fibre, $e \cdot g = \{e_i g_{i1}, ..., e_i g_{ik}\}$, for $g_{ij} \in GL(k,\mathbb{C})$.

If P is a principal G-bundle and if ρ is a representation of G on a finite-dimensional vector space, V, we can define the associated vector bundle, $P \times_\rho V$, by introducing an equivalence relation on $P \times V$

$$(p, \rho(g) \cdot v) \cong (p \cdot g, v) \quad \text{for all } p \in P.v \in V, g \in G. \tag{E.7}$$

The transition functions on $P \times_\rho V$ are given by the representation, $\rho(\Phi)$, applied to the transition functions, Φ, of P. If P is the frame bundle of E and if ρ is the identity representation of G on the fibre F, then $P \times_\rho F = E$. In this way, we may pass from a vector bundle E to its associated principal bundle, P, and back again by changing the space on which the transition functions act from the vector space to the general linear group and back.

If E is a vector bundle with an inner product, we can apply the Gram-Schmidt process to construct unitary frames. The bundle of unitary frames is a $U(k)$ principal bundle if E is complex and an $O(k)$ principal bundle if E is real. If E is an oriented real bundle, we may consider the set of oriented frames to define an $SO(k)$ principal bundle.

If E is a complex vector bundle with an inner product and if the transition functions are unitary with determinant 1, we can define an $SU(k)$ principal bundle associated with E. However, not every vector bundle admits $SU(k)$ transition functions; the first Chern class must vanish. If the curvature of the connection is zero, then the vector bundle will admit $SU(k)$ transition functions. This condition will be true if our connection is chosen from the moduli space of flat connections.

If $\gamma(x)$ is a local section to P over a neighborhood, U, in M we can use right multiplication to define a map $\Phi : U \times G \to \pi^{-1}(U)$, where $\Phi(x, g) = \gamma(x) \cdot g$. This map gives a local trivialisation of P. A principal bundle P is trivial if and only if it has a global section; non-trivial principal bundles do not admit global sections. (The identity element of G is not invariant. Thus, there is no analog of the zero section to a vector bundle.)

The Lie algebra, $\mathfrak{G}$, of G is the tangent space, $T_e(G)$, at the identity element, e, of G. By using left translation in the group, we may identify $\mathfrak{G}$ with the set of left-invariant vector fields on G. Let $\mathfrak{G}^*$ be the dual space. We can identify $\mathfrak{G}^*$ with the left-invariant 1-forms of G. Let $\{L_a\}$ be a basis for $\mathfrak{G}$ and ϕ_a a dual basis for $\mathfrak{G}^*$. The $\{L_a\}$ obey the Lie bracket algebra, $[L_a, L_b] = f_{abc} L_c$, where the f_{abc} are the structure constants for $\mathfrak{G}$. The Maurer-Cartan equation

$$d\phi_a = \frac{1}{2} f_{abc} \phi_b \wedge \phi_c$$

is the corresponding equation for $\mathfrak{G}^*$.

E.4.1. Magnetic monopole bundles

A magnetic monopole bundle is a $U(1)$-bundle over S^2. $U(1)$ is the gauge (Lie) group given by the elements

$$g = e^{i\psi},$$

where ψ is the coordinate in the fibre. Construct the G-bundle by taking the base manifold, $M = S^2$, with coordinates (θ, ϕ) where $0 \leq \theta < \pi$ and $0 \leq \phi < 2\pi$. Cover S^2 with two hemispherical neighborhoods $H_\pm$, the intersection $H_+ \cap H_-$ being a thin strip a the equator parametrized by the angle ϕ. Notice that the intersection of these two neighborhoods is *not* simply connected.

Locally, the bundle looks like

$$\begin{aligned} H_+ \times U(1) &\to \text{coordinates } (\theta, \phi, e^{i\psi_+}) \\ H_- \times U(1) &\to \text{coordinates } (\theta, \phi, e^{i\psi_-}) \end{aligned} \tag{E.8}$$

The transition functions must be functions of ϕ along $H_+ \cap H_-$ and must be elements of $U(1)$ to give a principal bundle. Let $g_0 = e^{ia\phi}$ be the transition function at each point in the intersection. In the intersection, the fibres in each neighborhood must be related by

$$e^{i\psi_-} = e^{ia\phi} e^{i\psi_+},$$

Since $\phi = 0$ and $\phi = 2\pi$ describe the same transition function (which must be uniquely defined at each point in $H_+ \cap H_-$), we have $a = n \in \mathbb{Z}$.

The integer n is what allows us to get non-trivial principal bundles. If $n = 0$ the bundle is trivial, meaning it is globally $S^2 \times S^1$. However, for $n \neq 0$, there is a twist in the fibres as we go around the equator and n indicated the number of twists corresponding to the first Chern class. n is the charge of the Dirac magnetic monopole.

E.4.2. Spin and Clifford bundles

Another important type of vector space which may appear as a fibre is a space of spinors. The structure group of a spinor space is the spin group, Spin(k). For example, the spin group corresponding to $SO(3)$ is just its double covering Spin$(3) = SU(2)$. The principal spin bundles associated with a bundle of spinors have fibres lying in Spin(k). We note that not all base manifolds admit well-defined spinor structures. Spinors arising from the tangent space can only be defined for manifolds where the second Stiefel-Whitney class vanishes.

Spinors must, in general, belong to an algebra of anticommuting variables. Such variables are a special case of the more general notion of a Clifford algebra which may also be used to define a type of fibre bundle. For example, if we start with a real vector bundle, E, of dimension k, we can construct the corresponding Clifford bundle, Cliff(E), as follows. The sections of Cliff(E) are constructed from section $s(x)$ and $s'(x)$ of E by introducing the Clifford multiplication

$$s \cdot s' + s' \cdot s = 2(s, s'),$$

where (s, s') is the vector bundle inner product. Cliff(E) is then a 2^k-dimensional bundle containing E as a sub-bundle. The Clifford algebra acts on itself by Clifford multiplication; relative to a matrix basis, this action admits a $2^k \times 2^k$ dimensional representation of the algebra. For $k = 1$, we find a 2×2 Pauli matrix representation while, for $k = 2$, we have the 4×4 Dirac matrices.

There is a natural isomorphism between the exterior algebra bundle, $\Lambda^*(E)$, and the Clifford bundle, Cliff(E). For example, the 16 independent Dirac matrix components can be matched with the elements of Λ^* as follows:

$$\begin{aligned} 1 &\longleftrightarrow 1 \\ \gamma_5 &\longleftrightarrow dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4 \\ \gamma_\mu &\longleftrightarrow dx^\mu \\ \gamma_\mu \gamma_5 &\longleftrightarrow \epsilon_{\mu\nu\lambda\sigma} dx^\nu \wedge dx^\lambda \wedge dx^\sigma \\ [\gamma_\mu, \gamma_\nu] &\longleftrightarrow dx^\mu \wedge dx^\nu \end{aligned} \tag{E.9}$$

E.5. Connections, curvature, and torsion on fibre bundles

Thus far, we have only considered fibre bundles as global geometric constructions. The notion of a connection plays an essential role in the local differential geometry of fibre bundles [93, 94]. A connection defines a covariant derivative which contains a gauge field and specifies the way in which a vector in the bundle E is to be parallel-transported along a curve lying in the base M. We shall first describe connections on vector bundles which form the foundation of the theory of general relativity and then proceed to treat connections on principal bundles which form the foundation for quantum field theory. Together, these concepts are of the most fundamental to modern physics as a whole.

E.5.1. Connections on vector bundles

Consider the unit sphere S^2 in $\mathbb{R}^3$. The following coordinates parameterize the sphere

$$x(\theta, \phi) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \quad 0 \leq \theta \leq \pi, 0 \leq \phi \leq 2\pi. \tag{E.10}$$

We observe that $x(\theta, \phi)$ is also the unit normal. The Riemannian metric induced by the chosen embedding is given by

$$g_{ij} = \begin{pmatrix} \partial_\theta x \cdot \partial_\theta x & \partial_\theta x \cdot \partial_\phi x \\ \partial_\phi x \cdot \partial_\theta x & \partial_\phi x \cdot \partial_\phi x \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{pmatrix} \tag{E.11}$$

so that

$$ds^2 = d\theta^2 + \sin^2\theta d\phi^2. \tag{E.12}$$

The two vector fields,

$$\mathbf{u}_1 = \partial_\theta x = (\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta)$$
$$\mathbf{u}_2 = \partial_\phi x = (-\sin\theta\sin o, \sin\theta\cos\phi, 0), \tag{E.13}$$

are tangent to the surface and span the tangent space provided that $0 < \theta < \pi$, i.e., away from the north and south poles. We can then decompose any derivative into tangential components proportional to $\mathbf{u}_1$ and $\mathbf{u}_2$, and a normal component $\hat{n}$ proportional to x. We identify $\mathbf{u}_1$ and $\mathbf{u}_2$ with the bases $\partial/\partial\theta$ and $\partial/\partial\phi$ for the tangent space because

$$\frac{\partial f(x)}{\partial\theta} = \mathbf{u}_1 \cdot \frac{\partial f}{\partial x}, \quad \frac{\partial f(x)}{\partial\phi} = \mathbf{u}_2 \cdot \frac{\partial f}{\partial x},$$

where $f(x)$ is a function on $\mathbb{R}^3$.

We want to differentiate tangential vector fields in a way which is intrinsic to the surface and not to the particular embedding involved. The ordinary partial derivatives are

$$\partial_\theta(u_1) = (-\sin\theta\cos\phi, -\sin\theta\sin\phi, -\cos\theta) = -x$$
$$\partial_\phi(u_1) = \partial_\theta(u_2) = (-\cos\theta\sin\phi, -\cos\theta\cos\phi, 0) = \frac{\cos\theta}{\sin\theta}\mathbf{u}_2 \tag{E.14}$$
$$\partial_\phi(u_2) = (-\sin\theta\cos\phi, -\sin\theta\sin\phi, 0) = -\sin^2\theta x - \cos\theta\sin\theta\mathbf{u}_1.$$

Define intrinsic covariant differentiation, ∇_X, with respect to a given tangent vector, X, by taking the ordinary derivative and projecting back the surface. ∇_X is the directional derivative obtained by throwing away the normal component of the ordinary partial derivatives.

$$\nabla_{u_1}(u_1) = 0$$
$$\nabla_{u_1}(u_2) = \nabla_{u_2}(u_1) = \cot\theta\mathbf{u}_2 \tag{E.15}$$
$$\nabla_{u_2}(u_2) = -\cos\theta\sin\theta\mathbf{u}_1$$

∇ is the Levi-Civita connection on S^2.

Using the identification of (u_1, u_2) with $(\partial/\partial\theta, \partial/\partial\phi)$, we write $\nabla_{\partial/\partial\theta} = \nabla_{u_1}$ and $\nabla_{\partial/\partial\phi} = \nabla_{u_2}$. The Christoffel symbol is defined by

$$\nabla_{u_i}(u_j) = u_k\Gamma^k_{ij}, \quad \nabla_{\partial_i}(\partial_j) = \Gamma^k_{ij}\partial_k.$$

Then, for our example, we find

$$\Gamma^2_{12} = \Gamma^2_{21} = \cot\theta; \quad \Gamma^1_{22} = -\cos\theta\sin\theta; \quad \Gamma^k_{ij} = 0, \quad \text{otherwise.}$$

Geodesic equation

Suppose $x(t)$ is a curve lying in S^2. This curve is a geodesic if there is no shear; in other words, the acceleration, $\ddot{x}$, has only components normal to the surface. This condition may be written $\nabla_{\dot{x}}(\dot{x}) = 0$. For example, if we consider a parallel to latitude

$x(t) = x(\theta = \theta_0, \phi = t)$, then $\dot{x} = u_2$ and $\nabla_{\dot{x}}(\dot{x}) = -\cos\theta_0 \sin\theta_0 \mathbf{u}_1$. The curve is a geodesic on the equator, $\theta_0 = \pi/2$. The curves, $x(t) = x(\theta = t, \phi = \phi_0)$, always satisfy the geodesic equations because $\dot{x} = \mathbf{u}_1$ and $\nabla_{\dot{x}}(\dot{x}) = 0$. These circles are great circles through the north and south poles.

Parallel transport

The Levi-Civita (Christoffel) connection provides a rule for the parallel transport of vectors on a surface. Let $x(t)$ be a curve in S^2 and $s(t)$ a vector field defined along the curve (i.e., a section of the vector bundle above the curve). We say that s is parallel transported along the curve if $\dot{s}$ is normal to the surface, meaning that it satisfies the equation

$$\nabla_{\dot{x}}(s) = 0.$$

Given an initial vector, $s(t_0)$, and the connection, $s(t)$ is uniquely determined by the parallel transport equation (very much like the lifted path is completely given by its initial point and the path lifting map in Section D.6.).

Holonomy

Holonomy is the process of assigning to each closed curve the linear transformation measuring the rotation which results when a vector is parallel transported around the given curve. The set of holonomy matrices forms a group called the holonomy group. The non-triviality of holonomy is related to the existence of curvature on the sphere: parallel transport around a closed curve in a place gives no rotation.

Example: Parallel transport and holonomy of S^2

Let x be the geodesic triangle in S^2 connecting the points $(1, 0, 0), (0, 1, 0)$, and $(0, 0, 1)$. x consists of three great circles:

$$x(t) = \begin{cases} (\cos(t), \sin(t), 0) & t \in [0, \pi/2] \\ (0, \sin(t), -\cos(t), 0) & t \in [\pi/2, \pi] \\ (-\sin(t), 0, -\cos(t)) & t \in [\pi, 3\pi/2] \end{cases}$$

Let $s(0)$ be the initial tangent vector, $s(0) = (0, \alpha, \beta)$, at $(1, 0, 0)$. When we parallel transport $s(0)$ along $x(t)$ using the Levi-Civita connection, we find

$$s(t) = \begin{cases} (-\alpha\sin(t), \alpha\cos(t), \beta) & t \in [0, \pi/2] \\ (-\alpha, \beta\cos(t), -\beta\sin(t), 0) & t \in [\pi/2, \pi] \\ (\alpha\cos(t), -\beta, -\alpha\sin(t)) & t \in [\pi, 3\pi/2] \end{cases}$$

One may verify that $s(t)$ is continuous at the corners, $\pi/2$, π, and satisfies $\nabla_{\dot{x}}(s) = 0$ since $\partial s/\partial t$ is normal to the surface. Parallel translation around the geodesic triangle changes s from $s(0) = (0, \alpha, \beta)$ to $s(3\pi/2) = (0, -\beta, \alpha)$, which represents a rotation through $\pi/2$. ($\pi/2$ is also the area of the spherical triangle.)

The holonomy matrix changing $s(0)$ to $s(3\pi/2)$ is

$$H_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}.$$

General definitions of the connection

Now, we will give the general formulation and axioms for connections on vector bundles. Let E be a general vector bundle. On each neighborhood, U, we choose a local frame $\{e_1, ..., e_k\}$ and express vectors in the fibre, $\pi^{-1}(U)$, in the form $Z = \sum_{i=1}^{k} e_i z^i$. This frame gives a local trivialisation of $\pi^{-1}(U) \cong U \times F$ and defines local coordinates (x, z). The vectors, e_i, themselves have the form $e_i = (0, ..., 0, 1, 0, ..., 0)$ in each local frame. The e_i may not be constant vectors on M since the local frames may be different in each neighborhood. The dependence of e_i on x due to the change of the local frame is dictated by the rule of covariant differentiation described later. A local section to the bundle is a smooth map from U to the fibre and can be regarded as a vector-valued function (a vector field),

$$s(x) = \sum_{i=1}^{k} e_i(x) z^i(x).$$

The tangent and cotangent spaces *of the bundle* $T(E)$ and $T^*(E)$ may be assigned local bases

$$T(E) : \left(\frac{\partial}{\partial x^\mu}, \frac{\partial}{\partial z^j} \right)$$
$$T^*(E) : \left(dx^\mu, dz^j \right)$$

The Levi-Civita connection lets us take the directional derivative of a tangent vector field and get another tangent vector field. We generalize this concept for vector bundles as follows.

Let X be a tangent vector, and let s be a section to E. A connection, ∇, is a rule, $\nabla_X(s)$, for taking the directional derivative of s in the direction X and getting another section to E. The assignment of a connection ∇ in a general vector bundle E provides a rule for the parallel transport of sections.

Let $x(t)$ be a curve in M. We say that $s(t)$ is parallel-transported along x if s satisfies the differential equation, $\nabla_{\dot{x}}(s) = 0$. There always exists a unique solution to this equation for given initial conditions. The generalized Christoffel symbols, ${\Gamma^j}_{\mu i}$, giving the action of a connection ∇ on a frame of the bundle E are defined by

$$\nabla_{\partial/\partial x^\mu}(e_i) = e_j {\Gamma^j}_{\mu i}.$$

We recall that we may associate the operator, d/dt, with $\dot{x}^\mu$ because $df/dt = \dot{x}^\mu \partial f / \partial x^\mu$.

In terms of the Christoffel symbols, the parallel transport equation takes the form

$$\begin{aligned}\nabla_{\dot{x}}(s) &= \nabla_{d/dt}(e_i z^i) = \nabla_{d/dt}(e_i) z^i + e_j \dot{z}^j \\ &= \dot{x}^\mu \left(\nabla_{\partial/\partial x^\mu}(e_i) z^i + e_j \partial_\mu z^j\right) \\ &= \dot{x}^\mu e_j \left(\Gamma^j{}_{\mu i} z^i + \partial_\mu z^j\right) = 0.\end{aligned} \tag{E.16}$$

Parallel transporting along a curve $x(t)$ lets us compare the fibres of the bundle E at different points of the curve. Thus, it becomes natural to think of lifting a curve $x(t)$ in M to a curve $c(t) = (x^\mu(t), z^i(t))$ in the bundle. Differentiation along $c(t)$ is defined by

$$\frac{d}{dt} = \dot{x}^\mu \frac{\partial}{\partial x^\mu} + \dot{z}^i \frac{\partial}{\partial z^i},$$

where $\dot{z}^i$ is given by solving the parallel transport equation:

$$\dot{z}^i + \Gamma^i{}_{\mu j} \dot{x}^\mu z^j = 0. \tag{E.17}$$

Thus, we may write

$$\frac{d}{dt} = \dot{x}^\mu \left(\frac{\partial}{\partial x^\mu} - \Gamma^i{}_{\mu j} z^j \frac{\partial}{\partial z^i}\right) = \dot{x}^\mu D_\mu, \tag{E.18}$$

where

$$D_\mu = \frac{\partial}{\partial x^\mu} - \Gamma^i{}_{\mu j} z^j \frac{\partial}{\partial z^i} \tag{E.19}$$

is the operator in $T(E)$ called the covariant derivative.

We can define a splitting of $T(E)$ at $x \in U$ into vertical $V(E)$ (lying strictly in the fibre) and horizontal $H(E)$ components with bases $\{\partial/\partial z^i\}$ and $\{D_\mu\}$, respectively, so that $T_x(E) = V_x(E) \oplus H_x(E)$.

In the cotangent space, one considers a vector-valued one-form

$$\omega^i = dz^i + \Gamma^i{}_{\mu j} dx^\mu z^j \tag{E.20}$$

in $T^*(E)$. We observe that ω^i is the unique, non-trivial solution to the conditions

$$\begin{aligned}\langle \omega^i, D_\mu \rangle &= 0 \\ \langle \omega^i, \partial/\partial z^j \rangle &= \delta^{ij}\end{aligned} \tag{E.21}$$

Conversely, these conditions determine D_μ if ω^i is given. The connection 1-form, ω^i, annihilates the horizontal subspace to $T(E)$ and is in some sense dual to it.

Now, introduce the matrix-valued connection 1-form Γ, where

$$\Gamma^i{}_j = \Gamma^i{}_{\mu j} dx^\mu.$$

The total covariant derivative, $\nabla(s)$, is defined by

$$\nabla(s) = e_i \otimes dz^i(x) + e_i \otimes \Gamma^i{}_j z^j(x) \tag{E.22}$$

which maps $C^\infty(E)$ to $C^\infty(E \otimes T^*(M))$. Note that this map is the pullback to M (using the section $z^i(x)$) of a covariant derivative in the bundle given by

$$\nabla(Z) = e_i \otimes \omega^i, \quad (Z = e_i z^i \in \pi^{-1}(U)),$$

where ω^i belongs to $T^*(E)$ rather than $T^*(M)$. The total covariant derivative contains all of the directional derivatives at the same time in the same way that $df = (\partial f/\partial x^\mu)dx^\mu$ contains all of the partial derivatives at the same time.

Properties of the connection

The properties of the connection are

(a) Linearity in s: $\nabla_X(s + s') = \nabla_X(s) + \nabla_X(s')$

(b) Linearity in X: $\nabla_{X+X'}(s) = \nabla_X(s) + \nabla_{X'}(s)$

(c) Differential operator: $\nabla_X(sf) = s \cdot X(f) + (\nabla_X(s))f$

(d) Tensorial in X: $\nabla_{fX}(s) = f\nabla_X(s)$,

where $s(x)$ is a section to E, X is a vector field on M, and $f(x)$ is a scalar function.
The properties of the total covariant derivative are

(a) Linearity in s: $\nabla(s + s') = \nabla(s) + \nabla(s')$

(b) Differential operator: $\nabla(sf) = s \otimes df + \nabla(s)f$

The relationship between these two differential operators is given by

(a) $\nabla(s) = \nabla_{\partial/\partial x^\mu}(s) \otimes dx^\mu$

(b) $\nabla_X(s) = \langle\nabla(s), X\rangle$,

where $X \in C^\infty(T(M))$ and $\nabla(s) \in C^\infty(E \otimes T^*(M))$.
One can extend total covariant differentiation to p-form-valued sections of E by the rule

$$\nabla(s \otimes \theta) = \nabla(s) \wedge \theta + s \otimes d\theta.$$

where $s \in C^\infty(E)$ and $\theta \in C^\infty(\Lambda^p(M))$. ∇, thus, extends to a differential operator with the following domain and range

$$\nabla : C^\infty(E \otimes \Lambda^p(M)) \to C^\infty(E \otimes \Lambda^{p-1}(M)).$$

Under a change of frame,

$$e'_j = e_i\Phi^{-1}_{ij}(x), \quad z'^i = \Phi_{ij}(x)z^j,$$

and sections are invariant

$$s(x) = e_i z^i = e'_i z'^i = s'(x).$$

We see that

$$\nabla(e_i') = \nabla(e_i) \otimes \Phi_{ij}^{-1} + e_i \otimes d\Phi_{ij}^{-1} = e_i'\Gamma'^{i}{}_{j},$$

where

$$\Gamma'^{i}_{j} = \Phi_{ik}\Gamma^{k}_{l}\Phi_{lj}^{-1} + \Phi_{ik}d\Phi_{kj}^{-1}, \tag{E.23}$$

so the connection 1-form $\Gamma^{i}{}_{j}$ transforms as a gauge field rather than as a tensor. We may *define* a connection as a collection of 1-forms obeying this transformation law (E.23). We need to check that ∇ is independent of the choice of frame and is well defined in the overlap region $U \cap U'$. We find

$$\nabla(s) = \nabla(e_i z^i) = e_j \otimes \Gamma^{j}{}_{i} z^i + e_j \otimes dz^j = e_j' \otimes \Gamma'^{j}{}_{i} z'^{i} + e_j' \otimes dz'^{j}.$$

Curvature

Curvature measures the extent to which parallel transport is path-dependent. If the curvature is zero and $x(t)$ is a path in a coordinate neighborhood of M, then the result of parallel transport is alway the identity transformation (unless the path encloses a hole in the space!). For curved manifolds, we get non-trivial results. For example, parallel transport around a geodesic triangle on S^2 gives a rotation equal to the area of the spherical triangle.

Let $(x^1, ..., x^k)$ be a local coordinate chart and take a square path $x(t)$ with vertices, say in the 1-2 plane. Let $H_{ij}(\tau)$ be the holonomy matrix obtained by traversing the path with vertices $(0, ..., 0), (0, \sqrt{\tau}, ..., 0), (\sqrt{\tau}, \sqrt{\tau}, ..., 0)(\sqrt{\tau}, ..., 0)$. Then, the curvature matrix in the 1-2 plane is

$$R_{ij}(1,2) = \frac{d}{d\tau} H_{ij}(\tau)|_{\tau=0} \tag{E.24}$$

The curvature is defined as the commutator of the components, D_μ, of the basis for the horizontal subspace of $T(E)$ (commutator of covariant derivatives).

$$[D_\mu, D_\nu] = -R^{i}{}_{j\mu\nu} z^j \frac{\partial}{\partial z^i}, \tag{E.25}$$

where $R^{i}{}_{j\mu\nu}$ can be expressed in terms of Christoffel symbols as in equation (2.51). Note that the right-hand side of the this equation has only vertical components. $R^{i}{}_{j\mu\nu}$ can be interpreted as the obstruction of integrability of the horizontal subspace.

In the cotangent space approach, the curvature appears as a matrix-valued 2-form

$$R^{i}{}_{j} = d\Gamma^{i}{}_{j} + \Gamma^{i}{}_{k} \wedge \Gamma^{k}{}_{j} = \frac{1}{2} R^{i}{}_{j\mu\nu} dx^\mu \wedge dx^\nu \tag{E.26}$$

which is equation (2.7). We observe that $R^{i}{}_{j} z^j$ is the covariant differential of the 1-form $\omega^i \in T^*(E)$:

$$R^{i}{}_{j} z^j = d\omega^i + \Gamma^{i}{}_{j} \wedge \omega^j. \tag{E.27}$$

Note that, although ω^i has dz^k components, they cancel out in $R^{i}{}_{j}$.

Properties of the Curvature:

We define the curvature operator as

$$R(X,Y)(s) = \nabla_X \nabla_Y(s) - \nabla_Y \nabla_X(s) - \nabla_{[X,Y]}(s), \tag{E.28}$$

where

$$R\left(\frac{\partial}{\partial x^\mu}, \frac{\partial}{\partial x^\nu}\right)(e_i) = e_j R^j_{\ i\mu\nu}.$$

The properties of the curvature are as follows:

(a) Multilinearity: $R(X+X',Y)(s) = R(X,Y)(s) + R(X',Y)(s)$

(b) Antisymmetry: $R(X,Y)(s) = -R(Y,X)(s)$

(c) Tensorial: $R(fX,Y)(s) = R(X,fY)(s) = R(X,Y)(fs) = fR(X,Y)(s)$,

where X and Y are vector fields, $s(x)$ is a section, and $f(x)$ is a scalar function.

The total curvature, R, is a matrix-valued 2-form given by

$$\begin{aligned} R(s) &= \nabla^2(s) = \nabla\left(e_j \otimes \Gamma^j_{\ i} z^i + e_j \otimes dz^j\right) \\ &= e_k \otimes \Gamma^k_{\ j} \wedge \Gamma^j_{\ i} z^i + e_k \otimes \left(d\Gamma^k_{\ i} z^i - \Gamma^k_{\ j} \wedge dz^j\right) + e_k \otimes \Gamma^k_{\ j} \wedge dz^j + 0 \\ &= e_k \otimes R^k_{\ i} z^i \end{aligned} \tag{E.29}$$

The matrix $R = ||R^i_{\ j}||$ is given by

$$R = \frac{1}{2} R\left(\frac{\partial}{\partial x^\mu}, \frac{\partial}{\partial x^\nu}\right) dx^\mu \wedge dx^\nu$$

acting on a section s. R is a 2-form-valued linear map from $E \to E$.

Under change of frame, $R^i_{\ j}$ transforms as

$$R'^i_{\ j} = \Phi^i_{\ k} R^k_{\ l} (\Phi^{-1})^l_{\ j},$$

and $R(s)$ is invariant under a change of frame.

The curvature can be regarded as an obstruction to finding locally flat frames (i.e., covariant constant frames). Given e_i, let us attempt to find a new frame $e'_i = e_j \Phi^{-1}_{ji}$ which is locally flat. If we set $\nabla(e'_i) = 0$, we find $\Phi\Gamma\Phi^{-1} + \Phi d\Phi^{-1} = 0$. This equation is solved if Γ is a pure gauge,

$$\Gamma^i_{\ j} = -\left(d\Phi^{-1}\right)^i_{\ k} \Phi^k_{\ j} = \left(\Phi^{-1}\right)^i_{\ k} d\Phi^k_{\ j}.$$

Conversely, if the curvature vanishes, Γ can be written as a pure gauge.

Torsion

An advantage to the cotangent space formulation of the vector bundle connection (E.22) is that it is independent of the coordinate system, $\{x^\mu\}$, on M. Multiple covariant differentiation of an invariant 1-form such as $p_\mu dx^\mu$ is independent of the connection chosen on the cotangent bundle $T^*(M)$. However, if we choose to take the covariant derivative of the individual components of the covariant derivative of a section $s(x) = e_i z^i(x)$, we must specify an additional connection on $T^*(M)$ to treat the μ index. Torsion is a property of the connection on the tangent bundle which must be introduced when we examine the double covariant derivative.

Let $\{\Gamma^i{}_{\mu j}\}$ be the Christoffel symbols on the vector bundle, E, and let $\{\Gamma^\nu{}_{\mu\lambda}\}$ be the Christoffel symbols on $T(M)$. We define the double covariant derivative of a section $s(x) = e_i z^i(x)$ as

$$z^i{}_{;\mu;\nu} = \partial_\nu\left(\partial_\mu z^i + \Gamma^i{}_{\mu j} z^j\right) + \Gamma^i{}_{\nu j}\left(\partial_\mu z^i + \Gamma^j{}_{\mu k} z^k\right) - \Gamma^\lambda{}_{\mu\nu}\left(\partial_\lambda z^i + \Gamma^i{}_{\mu j} z^j\right).$$

The commutator of double covariant differentiation on a section yields the following formula:

$$z^i{}_{;\mu;\nu} - z^i{}_{;\nu;\mu} = -R^i{}_{j\mu\nu} z^j - T^\lambda_{\mu\nu} z^i{}_{;\lambda}, \tag{E.30}$$

where we have written the torsion as

$$T^\lambda_{\mu\nu} = \Gamma^\lambda{}_{\mu\nu} - \Gamma^\lambda{}_{\nu\mu}.$$

Multiple covariant differentiation is a map

$$C^\infty(E) \xrightarrow{\nabla} C^\infty(E \otimes T^*(M)) \xrightarrow{\nabla} C^\infty(E \otimes T^*(M) \otimes T^*(M)).$$

Properties of the Torsion:

We define the torsion operator by

$$T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y].$$

This operator is a vector field with components

$$T\left(\frac{\partial}{\partial x^\mu}, \frac{\partial}{\partial x^\nu}\right) = \left(\Gamma^\lambda{}_{\mu\nu} - \Gamma^\lambda{}_{\nu\mu}\right)\frac{\partial}{\partial x^\lambda}.$$

Once a metric, $(X,Y) = g_{\mu\nu} x^\mu y^\nu$, has been chosen, the Levi-Civita connection is uniquely defined by the following properties:

(a) Torsion free: $T(X,Y) = 0$

(b) Covariant constancy of the metric: $d(X,Y) = (\nabla X, Y) + (X, \nabla Y)$.

The connection is said to be Riemannian if $\nabla^* = \nabla$ or $\Gamma^i{}_{\mu j} = -\Gamma^j{}_{\mu i}$ with the curvature of the connection then being $R^i{}_j = -R^j{}_i$.

Let $f : M \to M'$, and let ∇' be a connection on the vector bundle, E', over M'. There is a natural pullback connection, $\nabla = f^*\nabla'$,

$$\Gamma^i{}_{\mu j} = \Gamma'{}_{\alpha j}\frac{\partial x'^\alpha}{\partial x^\mu}.$$

The curvature of ∇ is the pullback of the curvature of ∇':

$$R^i_{\ j\mu\nu} = \frac{1}{2} R'^i_{\ j\alpha\beta} \left(\frac{\partial x'^\alpha}{\partial x^\mu} \frac{\partial x'^\beta}{\partial x^\nu} - \frac{\partial x'^\alpha}{\partial x^\nu} \frac{\partial x'^\beta}{\partial x^\mu} \right).$$

Example: Complex line bundle

Let L be the line bundle over $P_1(\mathbb{C})$ defined previously. This bundle is a natural sub-bundle of $P_1(\mathbb{C}) \times \mathbb{C}^2$. We denote a point of the bundle L by $(x; z_0, z_1)$, where (z_0, z_1) lie on the line in $\mathbb{C}^2$ corresponding to the point x in $P_1(\mathbb{C})$. The natural fibre metric on L is given by

$$((x; z_0, z_1), (x; w_0, w_1))\ z_0\overline{w}_0 + z_1\overline{w}_1.$$

Let $h(x; z_0, z_1) = |z_0|^2 + |z_1|^2$ be the length of a point in L and form a connection, ω, lying in $T^*(L)$ given by

$$\omega = h^{-1}\partial h = \frac{\overline{z}_0 dz_0 + \overline{z}_1 dz_1}{|z_0|^2 + |z_1|^2}.$$

The curvature is then

$$\Omega = d\omega + \omega \wedge \omega = (\partial + \overline{\partial})(h^{-1}\partial h) + 0 = -\partial\overline{\partial} \ln h.$$

In order to carry out practical computations we choose a gauge (that is a local section of L) with coordinates $(x; \zeta_0^{(1)}, 1)$, where $\zeta_0^{(1)} = z_0/z_1 = u + iv$ for $u, v \in \mathbb{R}$. Then, we compute $h = 1 + u^2 + v^2$ and

$$\Omega = -\partial\overline{\partial} \ln(1 + u^2 + v^2) = \frac{2i du \wedge dv}{(1 + u^2 + v^2)^2}$$

We recognize this expression as $(2i)$ times the Kähler form for $S^2 = P_1(\mathbb{C})$. We can thus read off the metric directly from Ω.

In some sense, $\omega = h^{-1}\partial h$ is a pure gauge with respect to a curvature involving only a ∂. We find non-trivial full curvature because Ω involves $d = \partial + \overline{\partial}$.

This same construction works for an arbitrary holomorphic line bundle over an arbitrary complex manifold once a fibre metric has been chosen.

E.5.2. Connections on principal bundles

Let G be a matrix group and $\mathfrak{G}$ be its Lie algebra. The Maurer-Cartan form, $g^{-1}dg$, is a matrix of 1-forms belonging to the Lie algebra, $\mathfrak{G}$. This form is invariant under the left action by a constant group element g_0,

$$(g_0 g)^{-1} d(g_0 g) = g^{-1} dg.$$

Let $\{\Phi_a\}$ be a basis for the left-invariant 1-forms. We then express the Maurer-Cartan form as

$$g^{-1}dg = \Phi_a \frac{\lambda_a}{2i}, \tag{E.31}$$

where $\lambda_a/2i$ is a constant matrix in $\mathfrak{G}$. Since $d(g^{-1}dg) + g^{-1}dg \wedge g^{-1}dg = 0$, we find that Φ_a obeys the Maurer-Cartan equations

$$d\Phi_a + \frac{1}{2} f_{abc} \Phi_b \wedge \Phi_c = 0, \tag{E.32}$$

where the f_{abc} are the structure constants of $\mathfrak{G}$.

The dual of Φ_a is the differential operator

$$L_a = Tr\left(g \frac{\lambda_a}{2i} \frac{\partial}{\partial g^T}\right) = \frac{1}{2i} g_{ik} [\lambda_a]_{kl} \frac{\partial}{\partial g_{jl}}$$

obeying

$$\langle \Phi_a, L_b \rangle = \delta_{ab}, \quad [L_a, L_b] = f_{abc} L_c. \tag{E.33}$$

$\{L_a\}$ is a left-invariant basis for the tangent space of G.

The corresponding right-invariant objects are defined by

$$dgg^{-1} = \frac{\lambda_a}{2i} \overline{\Phi}_a, \quad \overline{L}_a = Tr\left(\frac{\lambda_a}{2i} g \frac{\partial}{\partial g^T}\right), \tag{E.34}$$

where

$$\begin{aligned} \langle \overline{\Phi}_a, \overline{L}_b \rangle = \delta_{ab}, \quad [\overline{L}_a, \overline{L}_b] = -f_{abc} \overline{L}_c \\ d\overline{\Phi}_a - \frac{1}{2} f_{abc} \overline{\Phi}_b \wedge \overline{\Phi}_c = 0. \end{aligned} \tag{E.35}$$

That is, all of the structure equations have reversed sign. Note that L_a and $\overline{L}_a$ commute $[L_a, \overline{L}_a] = 0$. L_a and $\overline{L}_a$ generalize the distinction between fixed-space and fixed-body rotation generators.

Let P be a principal bundle. If we choose a local trivialisation, we have coordinates, (x, g), for P, where $g \in G$. A local section of P is a smooth map from a neighborhood U to G. The assignment of a connection on a principal bundle provides a rule for the parallel transport of sections. A connection A of a principal bundle is a Lie-algebra valued matrix of 1-forms in $T^*(M)$,

$$A(x) = A^\alpha_\mu(x) \frac{\lambda_a}{2i} dx^\mu. \tag{E.36}$$

If $x(t)$ is a curve in M, the section $g_{ij}(t)$ is defined to be parallel-transported along x if the following differential equation is satisfied:

$$\dot{g}_{ik} + A_{\mu ij}(x) \dot{x}^\mu g_{jk} = 0, \tag{E.37}$$

where A_μ is the connection on P. We may rewrite this expression as

$$g^{-1}\frac{dg}{dt} + g^{-1}\left(A^a_\mu(x)\frac{\lambda_a}{2i}\frac{dx^\mu}{dt}\right)g = 0. \tag{E.38}$$

Parallel transport along a curve $x(t)$ allows us to compare the fibres of P at different points of the curve. In analogy to the methods used for vector bundle connections, we may lift curves $x(t)$ in M to curves in P. We define differentiation along the lifted curve by

$$\frac{d}{dt} = \dot{x}^\mu\frac{\partial}{\partial x^\mu} + \dot{g}_{ij}\frac{\partial}{\partial g_{ij}} = \dot{x}^\mu\left(\frac{\partial}{\partial x^\mu} - A^a_{\ \mu}\frac{(\lambda_a)_{ik}}{2i}g_{kj}\frac{\partial}{\partial g_{ij}}\right) = \dot{x}^\mu\left(\frac{\partial}{\partial x^\mu} - A^a_{\ \mu}\overline{L}_a\right) \tag{E.39}$$

We are, thus, led to define a splitting of $T(P)$ into a horizontal component, $H(P)$, with basis D_μ and a vertical component, $V(P)$, lying in $T(G)$, $T(P) = H(P) \oplus V(P)$. This splitting is invariant under right multiplication by the group.

The curvature is defined by

$$[D_\mu, D_\nu] = -F^a_{\mu\nu}\overline{L}_a, \tag{E.40}$$

where

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f_{abc}A^b_\mu A^c_\nu. \tag{E.41}$$

As expected, the commutator of covariant derivatives only has vertical components (points in the fibre direction).

We may regard the connection on P as a $\mathfrak{G}$-valued 1-form, ω, in $T^*(P)$ with vertical component being the Maurer-Cartan form, $g^{-1}dg$. In local coordinates, we may write

$$\omega = g^{-1}Ag + g^{-1}dg,$$

where $A(x) = A^a_\mu(x)(\lambda_a/2i)dx^\mu$. We observe that, as in the vector bundle case, ω annihilates the horizontal basis of $T(P)$ and is constant on the vertical basis:

$$\langle\omega, D_\mu\rangle = 0, \quad \langle\omega, L_a\rangle = \lambda_a/2i. \tag{E.42}$$

Under the right action of the group, $g \to gg_0$, A remains invariant while ω transforms tensorially ($\omega \to g_0^{-1}\omega g_0$).

The curvature in this approach is a Lie-algebra valued matrix 2-form defined by

$$\Omega = d\omega + \omega\wedge\omega = g^{-1}Fg, \tag{E.43}$$

where

$$F = dA + A\wedge A = \frac{1}{2}F_{\mu\nu}^{\ \ a}\frac{\lambda_a}{2i}dx^\mu\wedge dx^\nu. \tag{E.44}$$

Ω obeys the Bianchi identity

$$d\Omega + \omega\wedge\Omega - \Omega\wedge\omega = 0. \tag{E.45}$$

Note that Ω has no vertical components and transforms tensorially under right action $\Omega \to g_0^{-1}\Omega g_0$.

Gauge transformations

The transition functions of a principal bundle act on fibres by left multiplication. Let us consider two overlapping neighborhoods, U and U', and a transition function, $\Phi_{UU'} = \Phi$. The local fibre coordinates, g and g', in U and U' are related by $g' = \Phi g$. Then, in order for the connection 1-form, ω, to be well defined in the overlapping region, $U \cap U'$, A must transform as

$$A' = \Phi A \Phi^{-1} + \Phi d\Phi^{-1}. \tag{E.46}$$

We verify that

$$\omega = g^{-1}Ag + g^{-1}dg = g'^{-1}A'g' + g'^{-1}dg', \tag{E.47}$$

so ω is indeed well defined in $T^*(P)$. The transformation is a gauge transformation of A. Using equation (E.41), we find the gauge transformation of F to be

$$F' = \Phi F \Phi^{-1}. \tag{E.48}$$

The curvature 2-form is also consistently defined over the manifold,

$$\Omega = g^{-1}Fg = g'^{-1}F'g'.$$

Choosing a section, $g = g(x)$, one can pull back ω and Ω to the base space. A and F are equivalent to the pullbacks, $g^*\omega$ and $g^*\Omega$, which are sometimes denoted simply as ω and Ω. Gauge transformations of A and F correspond to changes of the section.

In the theory of gauge fields, the structure group G is called the gauge group. The (pulled-back) connection A of a principal bundle is the gauge potential, and the (pulled-back) curvature F gives the field strength of the gauge field. When matter fields are present in the gauge theory, they are described by the associated vector bundles.