UNIVERSITY OF CALIFORNIA
SANTA CRUZ

QCD INSTANTONS AND INFLATION

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PHYSICS

by

Lawrence Pack

June 2012

The Dissertation of Lawrence Pack
is approved:

Michael Dine, Chair

Tom Banks

Howard Haber

Dean Tyrus Miller
Vice Provost and Dean of Graduate Studies

Table of Contents

Abstract

QCD Instantons and Inflation

by

Lawrence Pack

In the first half of this dissertation, after giving a pedagogical introduction to quantum chromodynamics, we revisit the question of whether or not one can perform reliable semiclassical QCD computations at zero temperature. We study correlation functions with no perturbative contributions, and organize the problem by means of the operator product expansion, establishing a precise criterion for the validity of a semiclassical calculation. For $N_f > N$, a systematic computation is possible; for $N_f < N$, it is not. $N_f = N$ is a borderline case. In our analysis, we see explicitly the exponential suppression of instanton effects at large N. As an application, we describe a test of QCD lattice gauge theory computations in the chiral limit.

For the second half, we turn our attention to inflation. Once again, a pedagogical overview of inflation is given, after which we explore some issues in slow roll inflation in situations where field excursions are small compared to M_p. We argue that for small field inflation, minimizing fine tuning requires low energy supersymmetry and a tightly constrained structure. Hybrid inflation is *almost* an inevitable outcome. The resulting theory can be described in terms of a supersymmetric low energy effective action and inflation completely characterized in terms of a small number of parameters. Demanding slow roll inflation significantly constrains these parameters. In this context,

the generic level of fine tuning can be described as a function of the number of light fields, there is an upper bound on the scale of inflation, and an (almost) universal prediction for the spectral index. Models of this type need not suffer from a cosmological moduli problem.

To my wife, Kate.

Acknowledgments

I want to thank Michael Dine for his support and encouragement.

Part I

Instantons in Quantum Chromodynamics

Chapter 1

Introduction

Quantum chromodynamics, hereafter referred to simply as QCD, is the quantum field theory that governs the dynamics of the strong force. The strong force, in turn, is that force which describes the interactions of quarks and gluons. Quarks are the fundamental constituents of the hadrons (protons, neutrons, pions) and are fermionic in nature. That is, they carry half-integer spin. Just as photons are the mediators of the electromagnetic force, gluons are the mediators of the strong force, and are bosonic (carry integer spin).

Much like the electric charge of the electromagnetic force, the strong force carries what is known as "color charge", or simply color. However, unlike the electromagnetic force, the mediators of the strong force, the gluons, carry charge. Because gluons carry color, they interact amongst themselves. This sets the strong force in distinct contrast to electromagnetism where photons are neutral, and therefore don't self-interact. This self-interaction amongst gluons adds rich dynamics and leads to some exotic configurations, known as glueballs. A quantum field theory in which the force mediators carry charge is known as a non-abelian gauge theory, a topic which we address in a later section. QCD is an $SU(3)$ non-abelian gauge theory and is part of the Standard Model of particle physics.

Chapter 2

Some Preliminaries

2.1 Spontaneous Symmetry Breaking

2.1.1 Continuous Global Symmetries

Let us consider a Lagrangian with N continuous global symmetries. By Noether's theorem, there exist N conserved charges, Q^a, $1 \leq a \leq N$. The Q^a therefore commute with the Hamiltonian, $[Q^a, H] = 0$. Moreover, we can continue to construct new conserved charges via commutators among the Q^a's, until we have exhausted all the possible conserved charges. In other words, if $[Q^a, H] = 0$, then $[[Q^a, Q^b], H] = 0$ and so $[Q^a, Q^b]$ is also conserved. We continue in this manner until the new conserved charges become functions of the existing ones. At this point, we must now distinguish two cases:

1. The Q's generate internal symmetries only.

2. The Q's include spacetime symmetries.

Let's begin the discussion for the case of internal symmetries and consider an energy eigenfunction E_α, $H|\alpha\rangle = E_\alpha|\alpha\rangle$. Using the Q's, we can generate states degenerate with $|\alpha\rangle$, since

$$H(Q^a|\alpha\rangle) = Q^a(H|\alpha\rangle) = E_\alpha(Q^a|\alpha\rangle). \tag{2.1}$$

We continue this process and apply the Q's until further applications only generate linear combinations of states already obtained. We can then define a subspace $\mathcal{H}_\alpha \subset \mathcal{H}$ closed under the action of the Q's. By construction, this forms an irreducible representation since any state in $\mathcal{H}_\alpha$ can be connected to any other state by an appropriate application of Q's. Moreover, we can do this for each eigenstate, thereby classifying the states of the system into one of these irreducible representations.

We now allow the Q's to include spacetime symmetries. By including H in our symmetry algebra, states of the system still fall into irreducible representations. States within an irreducible representation are degenerate with respect to some invariant operators. These representations will be infinite dimensional since spacetime symmetries are non-compact.

There is, however, one caveat to the above arguments. Consider the state $Q^a|\alpha\rangle$. In order for the above arguments to go through, it is necessary that this state be part of the Hilbert space $\mathcal{H}$. Is it possible for $Q^a|\alpha\rangle$ to not be an element of $\mathcal{H}$? The answer is yes. If $Q^a|\alpha\rangle$ is not normalizable, then it is not in $\mathcal{H}$ and the states of the system do not get partitioned into irreducible representations. This phenomenon is called *spontaneous symmetry breaking.*

2.1.2 Spontaneous Symmetry Breaking

Let's examine spontaneous symmetry breaking a little further. In quantum field theory, other states are built upon the vacuum. The state $|\alpha\rangle$ can be expressed as $X_\alpha|0\rangle$ for some operator X_α. Now consider an operator $Y^a_\alpha \equiv [Q^a, X_\alpha]$. The normalization of the states $X_\alpha|0\rangle$ and $Y_\alpha|0\rangle$ must be defined with respect to some regularization. With this in mind, we can take them to be normalizable. Then,

$$Q^a|\alpha\rangle = Y^a_\alpha|0\rangle + X_\alpha Q^a|0\rangle. \tag{2.2}$$

From equation (2.2) we see that if the vacuum respects the symmetry, i.e. $Q^a|0\rangle = 0$, then $Q^a|\alpha\rangle$ is normalizable and the argument of the previous section goes through. The Hilbert space breaks up into irreducible representations of the Q-algebra. However, if $Q^a|0\rangle$ is not normalizable, then the vacuum must necessarily not be invariant under the action of Q^a. So we see that a consequence of spontaneous symmetry breaking is that the vacuum does not share the same symmetries as the Lagrangian.

We have seen, therefore, that a continuous global symmetry of the Lagrangian falls into one of two categories:

1. The vacuum respects the symmetry, and the Hilbert space breaks up into irreducible representations of the symmetry algebra.

2. The vacuum does not respect the symmetry and we get spontaneous symmetry breaking.

As a concrete example, consider Poincare symmetry. As far as we know, the vacuum is Lorentz invariant, and so particles fall into irreducible representations of the Poincare group.

Now, let's see what effect spontaneous symmetry breaking has on vacuum expectation values. Suppose the lagrangian is invariant under the continuous symmetry group G, and consider a set of operators Φ_A which transform non-trivially under G. The transformation law for the Φ_A is

$$e^{iQ^a\theta^a}\Phi_A e^{-iQ^a\theta^a} = M_{AB}\Phi_B \tag{2.3}$$

for some matrix M. Taking the expectation value of both sides gives

$$\langle 0|e^{iQ^a\theta^a}\Phi_A e^{-iQ^a\theta^a}|0\rangle = M_{AB}\langle 0|\Phi_B|0\rangle. \tag{2.4}$$

Let's suppose further that the vacuum is invariant under G. That is, $Q^a|0\rangle = 0$. Then we are left with

$$\langle 0|\Phi_A|0\rangle = M_{AB}\langle 0|\Phi_B|0\rangle. \tag{2.5}$$

However, by supposition, the matrix M is not the identity, and therefore we must have

$$\langle 0|\Phi_A|0\rangle = 0. \tag{2.6}$$

This tells us that anytime the vacuum is invariant under a symmetry, the vacuum expectation value of any operator that is not a singlet must necessarily be zero. In other words, the existence of non-zero vacuum expectation values for operators transforming non-trivially signals the existence of spontaneous symmetry breaking.

Suppose we want to determine if our theory is invariant under a group G. Since all states are built upon the vacuum, it suffices to check if the vacuum respects

G. Let us further suppose there is a particle field, call it Φ, which has a non-zero vacuum expectation value. If the vacuum is to preserve the symmetry of G then Φ must transform as a singlet under G. For if Φ transforms non-trivially, the vacuum expectation value of Φ will change under a suitable transformation thereby rendering the vacuum asymmetric. This is symmetry breaking. If, for example, our symmetry group is $G_1 \times G_2 \times G_3 \times G_4$ and we want to break it down to $G_1 \times G_4$ then we require a field which is a singlet under G_1 and G_4 but not a singlet of G_2 or G_3. Once this field acquires a non-zero vacuum expectation value, the symmetry will be broken. The vacuum of our world appears to be Lorentz invariant, so our field Φ, let's call it the Higgs, must be a singlet under Lorentz transformations. That is, it must be a scalar field.

To summarize, given a lagrangian invariant under a continuous Lie group G, we have learned the following statements are equivalent:

1. There exists an operator transforming non-trivially under G which has a non-zero vacuum expectation value.

2. The vacuum does not respect the symmetry. That is, $Q|0\rangle \neq 0$, where Q is a Noether charge.

3. There exists a state $|\alpha\rangle$ such that $Q|\alpha\rangle$ is not normalizable.

4. The system exhibits spontaneous symmetry breaking.

2.2 Gauge Theory

2.2.1 Classical Electrodynamics

Let's begin by considering the laws of classical electrodynamics. In appropriate units, Maxwell's equations in vacuum are:

$$\nabla \cdot \mathbf{E} = 0 \tag{2.7}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{2.8}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{2.9}$$

$$\nabla \times \mathbf{B} = \mathbf{j} + \frac{\partial \mathbf{E}}{\partial t} \tag{2.10}$$

Taking the divergence of equation (2.10) gives us the continuity equation,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0. \tag{2.11}$$

Consider an arbitrary volume Ω. The continuity equation states that no charge can be created or destroyed inside Ω. In other words, the rate of decrease of charge inside Ω is due precisely to the flux of current flowing out its surface. The power of this argument is that we can take Ω arbitrarily small, which means that electric charge is *locally* conserved. As we will see, it is the local nature of this conservation law that will be important for gauge theory.

Let us now introduce the vector potential $\mathbf{A}$ and the scalar potential V, defined via,

$$\mathbf{E} = -\nabla V - \frac{\partial \mathbf{A}}{\partial t} \tag{2.12}$$

$$\mathbf{B} = \nabla \times \mathbf{A}. \tag{2.13}$$

An interesting observation, is that under certain transformations of V and $\mathbf{A}$, $\mathbf{E}$ and $\mathbf{B}$ don't change. If we let

$$\mathbf{A} \mapsto \mathbf{A} + \nabla\lambda \tag{2.14}$$

$$V \mapsto V - \frac{\partial\lambda}{\partial t} \tag{2.15}$$

then neither $\mathbf{E}$ nor $\mathbf{B}$ change. This is the origin of gauge invariance.

In preparation for the quantum case, let's write everything in an overtly Lorentz invariant fashion. We put

$$A^\mu \equiv (V, \mathbf{A}) \tag{2.16}$$

$$j^\mu \equiv (\rho, \mathbf{j}) \tag{2.17}$$

$$F^{\mu\nu} \equiv \partial^\mu A^\nu - \partial^\nu A^\mu. \tag{2.18}$$

Maxwell's equations then simplify to

$$\partial_\mu F^{\mu\nu} = j^\nu. \tag{2.19}$$

The continuity equation becomes

$$\partial_\mu j^\mu = 0 \tag{2.20}$$

and the statement of gauge invariance is now

$$A^\mu \mapsto A^\mu - \partial^\mu\lambda. \tag{2.21}$$

The equation of motion for the vector potential A^μ is

$$\Box A^\mu - \partial^\mu(\partial_\nu A^\nu) = j^\mu. \tag{2.22}$$

Using the principle of gauge invariance, we can make this equation more tractable. By gauge symmetry, we can choose the function λ such that the gauge transformed A^μ satisfies $\partial_\mu A^\mu = 0$. The equation of motion then simplifies to

$$\Box A^\mu = j^\mu. \tag{2.23}$$

It was mentioned earlier that it is the *local* nature of charge conservation that is important for gauge symmetry. Let's delve into this a little more. Consider the scalar potential V and let the gauge transformation function λ be constant. What effect does this have on the magnetic potential A? From equation (2.14) we see that it has absolutely no effect. The reason for this is because we are considering a *global* transformation. Now suppose we promote the global transformation to a local one. In other words, the function λ is no longer constant, but instead is a function of spacetime. This induces a change in the magnetic vector potential. So, by including magnetic effects and a compensating field, we can promote a global invariance to a local one [32]. This fact holds true in general. When promoting a global invariance to a local one, the existence of a new compensating field is required.

2.2.2 Quantum Electrodynamics

We now turn our attention from classical electrodynamics to quantum electrodynamics (QED). Let the electron field be denoted by $\psi(x)$. Let us first recall the free Dirac lagrangian:

$$\mathcal{L}_{Dirac} = \overline{\psi}(i\partial_\mu\gamma^\mu - m)\psi. \tag{2.24}$$

Consider a gauge transformation

$$\psi(x) \mapsto e^{i\alpha}\psi(x). \tag{2.25}$$

This is called a $U(1)$ transformation because the factor $e^{i\alpha}$ can be considered to be an element of the group $U(1)$. Let's now consider two cases:

1. The symmetry is global.

2. The symmetry is local.

First, we consider the case of a global symmetry. Then, the parameter α is constant. Moreover, the terms $\overline{\psi}(x)\psi(x)$ and $\overline{\psi}(x)\partial_\mu\gamma^\mu\psi(x)$ are invariant, and therefore so is $\mathcal{L}_{Dirac}$.

Now, let's promote this to a local symmetry, in which case α becomes a function on spacetime, $\alpha(x)$. Like before, the term $\overline{\psi}(x)\psi(x)$ is invariant, however, the $\overline{\psi}(x)\partial_\mu\gamma^\mu\psi(x)$ term is not invariant because of the derivative. Therefore, if we want a lagrangian invariant under this local symmetry, we cannot include this term. Evidently, the free Dirac lagrangian is not gauge invariant. Just as the magnetic vector potential played the role of the compensating field in classical electrodynamics in order to allow V to be invariant under local symmetries $V \mapsto V - \frac{\partial\lambda}{\partial t}$, the photon field A^μ must be included to allow invariance of the electron field under the local symmetry $\psi(x) \mapsto e^{i\alpha(x)}\psi(x)$. Moreover, the field A^μ must transform in a specified way, which we will see shortly. So, instead of the Dirac lagrangian, let's consider a slightly modified version:

$$\mathcal{L} = \mathcal{L}_{Dirac} + \mathcal{L}_{int} \tag{2.26}$$

where

$$\mathcal{L}_{int} = -e\overline{\psi}A_\mu\gamma^\mu\psi. \tag{2.27}$$

If we then require A to transform in such a way so as to cancel off the offending term in the transformation of $\overline{\psi}(x)\partial_\mu\gamma^\mu\psi(x)$, our lagrangian (2.26) will be invariant. As can be easily checked, the necessary transformation law for A is

$$A_\mu \mapsto A_\mu - \frac{1}{e}\partial_\mu\alpha(x). \tag{2.28}$$

Now that we have the photon field A in our lagrangian, we must add a kinetic term

$$\mathcal{L}_{kin} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{2.29}$$

The full QED lagrangian is then

$$\mathcal{L}_{QED} = \mathcal{L}_{Dirac} + \mathcal{L}_{int} + \mathcal{L}_{kin}, \tag{2.30}$$

and $\mathcal{L}_{QED}$ is invariant under the following local symmetry

$$\psi(x) \mapsto e^{i\alpha(x)}\psi(x) \tag{2.31}$$

$$\overline{\psi}(x) \mapsto \overline{\psi}(x)e^{-i\alpha(x)} \tag{2.32}$$

$$A_\mu \mapsto A_\mu - \frac{1}{e}\partial_\mu\alpha(x). \tag{2.33}$$

We can rewrite the QED lagrangian in a more enlightening way:

$$\mathcal{L}_{QED} = \overline{\psi}(x)(i\not{D} - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.34}$$

where D is called the *covariant derivative* and is given by

$$D_\mu \equiv \partial_\mu + ieA_\mu. \tag{2.35}$$

We have also introduced the slash notation, which denotes contraction with the gamma matrices: $\not{A} \equiv A_\mu \gamma^\mu$.

To summarize, we started out with a lagrangian, $\mathcal{L}_{Dirac}$ invariant under a global symmetry. In order to promote this symmetry to a local one, we had to add an additional em gauge field A_μ with a specified transformation law. We then amended the lagrangian to include a gauge kinetic term and interactions with the gauge field.

This local symmetry is what is called a *gauge symmetry*. What we have described is how to obtain a gauge invariant lagrangian from a global symmetry. Because the transformation of $\psi(x)$ was a $U(1)$ transformation, and $U(1)$ is an abelian group, quantum electrodynamics is known as an *abelian gauge theory*. In the next section we show how to extend this to non-abelian groups.

2.2.3 Non-Abelian Gauge Theory

In their seminal paper [34], Yang and Mills showed how to extend the construction of local invariance to non-abelian gauge groups. Before getting to the non-abelian case, let's summarize what we did for the abelian case. We started out with a lagrangian

$$\mathcal{L}_{Dirac} = \overline{\psi}(i\not{\partial} - m)\psi \tag{2.36}$$

invariant under the global transformation

$$\psi(x) \mapsto \Omega\psi(x) \tag{2.37}$$

$$\overline{\psi}(x) \mapsto \overline{\psi}(x)\Omega^{-1} \tag{2.38}$$

where $\Omega = e^{i\theta} \in U(1)$ and θ is constant. To localize this invariance, we then allow $\theta = \theta(x)$ to be a function on spacetime and introduce a gauge field A_μ transforming via

$$A_\mu \mapsto \Omega(x) A_\mu \Omega^{-1}(x) - \frac{i}{g} \left(\partial_\mu \Omega(x)\right) \Omega^{-1}(x). \tag{2.39}$$

We then defined a covariant derivative

$$D_\mu \equiv \partial_\mu - igA_\mu \tag{2.40}$$

and make the replacement

$$\partial_\mu \mapsto D_\mu \tag{2.41}$$

in the original lagrangian. Next, we add a gauge kinetic term to the lagrangian

$$\mathcal{L}_{kin} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}. \tag{2.42}$$

The locally invariant QED lagrangian then reads

$$\mathcal{L}_{QED} = \overline{\psi}(i\not{D} - m)\psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}. \tag{2.43}$$

We will mimic this procedure for the non-abelian case. Consider a lagrangian

$$\mathcal{L}_{global} = \mathcal{L}(\phi_i, \partial_\mu \phi_i) \tag{2.44}$$

invariant under the global transformation

$$\phi_i \mapsto \Omega_{ij} \phi_j \tag{2.45}$$

where $\Omega \in G$, is constant and is an element of the non-abelian lie group G. To localize this symmetry, we perform the following steps:

1. We allow $\Omega \in G$ to be a function on spacetime, $\Omega = \Omega(x)$.

2. Introduce a new field $A_\mu \in \mathfrak{g}$, where $\mathfrak{g}$ represents the lie algebra of G, transforming via

$$A_\mu \mapsto \Omega(x) A_\mu \Omega^{-1}(x) - \frac{i}{g} \left(\partial_\mu \Omega(x) \right) \Omega^{-1}(x). \quad (2.46)$$

   Here, g denotes the coupling constant.

3. Define gauge fields A_μ^a via

$$A_\mu = t^a A_\mu^a, \quad (2.47)$$

   where t^a are generators of the lie algebra $\mathfrak{g}$ and satisfy

$$[t^a, t^b] = i f^{abc} t^c. \quad (2.48)$$

4. Define the covariant derivative

$$D_\mu \equiv \partial_\mu - igA_\mu \quad (2.49)$$

$$= \partial_\mu - igt^a A_\mu^a. \quad (2.50)$$

   If we include color indices, $D\mu$ becomes

$$(D_\mu)_{ij} = \partial_\mu \delta_{ij} - igA_\mu^a t_{ij}^a. \quad (2.51)$$

   It should be noted that, like A_μ, D_μ is an element of the lie algebra $\mathfrak{g}$.

5. Introduce the following generalization of the field strength tensor

$$F_{\mu\nu} \equiv \frac{i}{g}[D_\mu, D_\nu] \quad (2.52)$$

$$= \partial_\mu A_\nu - \partial\nu A_\mu - ig[A_\mu, A_\nu]. \quad (2.53)$$

Like A_μ and D_μ, $F_{\mu\nu} \in \mathfrak{g}$. Therefore we can define $F^a_{\mu\nu} \equiv 2Tr(F_{\mu\nu}t^a)$. Written in a more convenient form,

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu. \tag{2.54}$$

Moreover, note that $Tr[F^2]$ is manifestly gauge invariant.

6. And finally, the desired locally invariant lagrangian is given by

$$\begin{aligned} \mathcal{L}_{local} &= \mathcal{L}_{global}(\phi_i, D_\mu\phi_i) - \frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} && (2.55) \\ &= \mathcal{L}_{global}(\phi_i, D_\mu\phi_i) - \frac{1}{2}Tr(F_{\mu\nu}F^{\mu\nu}) && (2.56) \end{aligned}$$

where the trace in (2.56) is over color indices.

Of course, it must be verified that $F^a_{\mu\nu}F^{a\mu\nu}$ and $Tr(F_{\mu\nu}F^{\mu\nu})$ are gauge invariant, but we leave this as an exercise for the reader.

Chapter 3

The Lagrangian of Quantum Chromodynamics

In QCD, the quarks, ψ_i, are in the fundamental representation of the gauge group $SU(N)$. The free-quark lagrangian is

$$\mathcal{L}_{Dirac} = \overline{\psi}_i(\partial_\mu - m)\psi_i. \tag{3.1}$$

In the previous section, we learned, by introducing A_μ, D_μ, and $F_{\mu\nu}$ how to convert this lagrangian into a locally invariant one, thereby giving us the requisite gauge theory. At this point, we'd like to add one additional complication. Namely, the existence of N_f flavors of quarks. The QCD lagrangian then becomes, with indices explicitly written

$$\mathcal{L}_{QCD} = \overline{\psi}_{iI}\left(i\not{D}_{ij}\delta_{IJ} - M_{IJ}\delta_{ij}\right)\psi_{jJ} - \frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu}, \tag{3.2}$$

where i, j are color indices and range from 1 to N, I is the flavor index and ranges from 1 to N_f, a is the gauge index and runs from 1 to $N^2 - 1$, and M_{IJ} is the diagonal flavor mass matrix. In the future, we will normally suppress flavor and color indices, and simply write

$$\mathcal{L}_{QCD} = \overline{\psi}(i\not{D} - M)\psi - \frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu}. \tag{3.3}$$

3.1 Symmetries of the QCD Lagrangian

The quark lagrangian without the masses is given by

$$\begin{aligned} \mathcal{L} &= i\overline{\psi}_{iI}\not{D}_{ij}\psi_{jI} && (3.4) \\ &= i\overline{\psi}^L_{iI}\not{D}_{ij}\psi^L_{jI} + i\overline{\psi}^R_{iI}\not{D}_{ij}\psi^R_{jI} && (3.5) \end{aligned}$$

where the superscripts L and R refer to the left and right-handed components of ψ, respectively. That is,

$$\psi^L \equiv \frac{1}{2}(1+\gamma^5)\psi \tag{3.6}$$

$$\psi^R \equiv \frac{1}{2}(1-\gamma^5)\psi \tag{3.7}$$

This lagrangian exhibits a $U(N)_L \times U(N)_R$ symmetry where

$$\psi^L_{iI} \mapsto U_{IJ}\psi^L_{iJ} \tag{3.8}$$

$$\psi^R_{iI} \mapsto V_{IJ}\psi^R_{iJ}, \tag{3.9}$$

and $U \in U(N)_L$ and $V \in U(N)_R$. This is known as *chiral symmetry*. However, when including the mass terms, this symmetry is no longer present, so it would appear that chiral symmetry plays no role in QCD. In the real world, there are 6 flavors of quark: u, d, s, c, b, and t with masses $1.7MeV$, $3.9MeV$, $76MeV$, $1.3GeV$, $4.3GeV$, and

$174 GeV$, respectively. These 6 quarks can be broken up into two categories: the light quarks u,d,s, and the heavy quarks c, b, t. If we restrict our lagrangian to contain only the light quarks, then the chiral symmetry, while not exact, can be considered to be an approximate symmetry.

Let us examine the symmetries of the $N = 3$ QCD lagrangian, keeping only the light quarks. Then, the lagrangian is

$$\mathcal{L}_{QCD} = i\overline{\psi}^L \not{D} \psi^L + i\overline{\psi}^R \not{D} \psi^R - \overline{\psi}^L M \psi^R - \overline{\psi}^R M \psi^L \tag{3.10}$$

where the mass matrix M is given by

$$M = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_d & 0 \\ 0 & 0 & m_s \end{pmatrix}. \tag{3.11}$$

Since the quark masses are small, we will approximate them as massless and examine the symmetries of the massless lagrangian.

1. **Gauge symmetry.** As we have already shown, $\mathcal{L}_{QCD}$ exhibits an $SU(3)$ gauge symmetry whereby

$$\psi_{iI} \mapsto \Omega_{ij}(x)\psi_{jI} \tag{3.12}$$

$$A_\mu \mapsto \Omega(x) A_\mu \Omega^{-1}(x) - \frac{i}{g}\left(\partial_\mu \Omega(x)\right)\Omega^{-1}(x) \tag{3.13}$$

   for $\Omega(x) \in SU(3)$.

2. **Discrete symmetries**. Ignoring topological considerations and the so-called θ-term, which we will discuss in subsequent sections, this lagrangian has the standard discrete symmetries of charge conjugation, parity, and time-reversal.

3. **Chiral symmetry**. As mentioned already, this system possesses a $U(3)_L \times U(3)_R$ chiral symmetry. From the $U(N) \cong SU(N) \times U(1)$ homomorphism, we can rewrite the chiral symmetry as $SU(3)_L \times SU(3)_R \times U(1) \times U(1)$, where

$$\psi^L_{iI} \mapsto U_{IJ}\psi^L_{iJ} \tag{3.14}$$

$$\psi^R_{iI} \mapsto V_{IJ}\psi^R_{iJ}, \tag{3.15}$$

and $U \in SU(3)_L$ and $V \in SU(3)_R$. We will touch upon the various $U(1)$ symmetries separately. There is an important subgroup of $SU(3)_L \times SU(3)_R$ that should be mentioned. This is the $SU(3)_V$ symmetry and is given by equations (3.14) and (3.15) with $U = V$. We will touch upon this symmetry more when discussing chiral symmetry breaking in the next section.

4. **U(1) symmetries**. There exist two $U(1) \times U(1)$ symmetries of note: $U(1)_L \times U(1)_R$ and $U(1)_V \times U(1)_A$. The $U(1)_L$ and $U(1)_R$ symmetries are the $U(1)$'s arising from the $U(N)_L \cong SU(N)_L \times U(1)_L$ and $U(N)_R \cong SU(N)_R \times U(1)_R$ homomorphisms. They are given by

$$\psi_L \overset{U(1)_L}{\mapsto} e^{i\theta}\psi_L, \quad \psi_R \overset{U(1)_L}{\mapsto} \psi_R \tag{3.16}$$

$$\psi_L \overset{U(1)_R}{\mapsto} \psi_L, \quad \psi_R \overset{U(1)_R}{\mapsto} e^{i\alpha}\psi_R \tag{3.17}$$

$U(1)_V$ and $U(1)_A$ are linear combinations of $U(1)_L$ and $U(1)_R$ and are given by

$$\psi_L \overset{U(1)_V}{\mapsto} e^{i\beta/3}\psi_L, \quad \psi_R \overset{U(1)_V}{\mapsto} e^{i\beta/3}\psi_R \tag{3.18}$$

$$\psi_L \overset{U(1)_A}{\mapsto} e^{i\phi}\psi_L, \quad \psi_R \overset{U(1)_A}{\mapsto} e^{-i\phi}\psi_R \tag{3.19}$$

We'll discuss the $U(1)_V$ and $U(1)_A$ symmetries in more detail in the next section on chiral symmetry breaking.

3.2 Condensates and Chiral Symmetry Breaking in QCD

It was mentioned that the QCD lagrangian contains a $U(3) \times U(3) \cong SU(3)_L \times SU(3)_R \times U(1)_V \times U(1)_A$ chiral symmetry. However, due to quantum corrections, the $U(1)_A$ is anomalous and is broken down to $\mathbf{Z}_6$. Therefore, the non-anomalous flavor symmetry is $SU(3)_L \times SU(3)_R \times U(1)_V$.

The Noether charge corresponding to the conserved $U(1)_V$ is quark number, which is 1/3 of the baryon number. Hadrons are therefore classified by baryon number.

The $SU(3)_V$ symmetry is known as isospin, and hadrons do in fact come in representations of isospin. Of course, it should be noted that $SU(3)_V$ is not an exact symmetry, but is only approximate. The small $SU(3)_V$ breaking is due to differences in quark masses.

However, hadrons do not appear to lie in representations of another $SU(3)$ group, which on the surface seems contradictory to the statement of $SU(3)_L \times SU(3)_R$ symmetry. The answer lies in the phenomenon of spontaneous symmetry breaking. Let us define

$$SU(3)_A \equiv \frac{SU(3)_L \times SU(3)_R}{SU(3)_V}. \qquad (3.20)$$

Since hadrons lie in representations of $SU(3)_V$, at least to a good approximation, we know that $SU(3)_V$ is not broken. Therefore, evidently $SU(3)_A$ is spontaneously broken.

From what we learned in the section on spontaneous symmetry breaking, we are looking for an operator O which satisfies all of the following:

1. O has a non-zero vacuum expectation value: $\langle 0|O|0\rangle = v$.
2. O transforms non-trivially under $SU(3)_A$.
3. O must be a singlet under $SU(3)_V$ since $SU(3)_V$ is not broken.
4. We want to maintain Lorentz invariance, and therefore O must be a scalar.
5. In order to maintain the gauge symmetry, O must be a color singlet.

Is there such an operator that satisfies all these conditions? There are no fundamental scalar particles, so evidently O must be composite. But what could it be? Because of the strong attractive interaction of quarks and anti-quarks, if the quarks are nearly massless, it costs little energy to create a quark-antiquark pair. The QCD vacuum should therefore consist of quark-antiquark composites. We might guess that this is the required operator. To be precise,

$$O \equiv \overline{\psi}\psi \tag{3.21}$$

$$= \overline{\psi}_L\psi_R + \overline{\psi}_R\psi_L. \tag{3.22}$$

This is called a *quark condensate* and is responsible for the breaking of chiral symmetry in QCD. O is manifestly $SU(3)_V$ invariant thereby preserving the approximate isospin symmetry. The breaking of $SU(3)_A$ leads to 8 pseudogoldstone bosons. These are identified with the $\pi^{\pm}$, π^0, $K^{\pm}$, K^0, $\bar{K}^0$, and η mesons. The pseudo refers to the fact that because the symmetry is only approximate, the goldstone bosons aren't actually massless. In the limit of zero quark mass and vanishing weak and electromagnetic gauge couplings, the masses of these particles should go to zero. The strange quark has much higher mass than the up or down, and as such, the chiral symmetry becomes much better when considering only the u and d quarks. In this case, the broken $SU(2)$ leads to the pions as the pseudogoldstone bosons. Treating the $SU(3)$ symmetry as an approximate symmety, the K and η mesons are also pseudogoldstone bosons. It should also be noted that the non-zero vacuum expectation value of O is strictly non-perturbative. Indeed, $\langle 0|O|0|\rangle$ vanishes at tree level. Due to the chiral symmetry of the lagrangian, perturbative corrections therefore vanish as well.

We have discussed spontaneous breaking of chiral symmetry, but it should be emphasized that chirality is also explicitly broken by the quark masses. If all quarks have the same non-zero mass, the unbroken subgroup becomes $SU(3)_V \times U(1)_V$. However, if the quarks are given different masses, as is the case in the real world, $SU(3)_V$, or isospin, is broken further.

Chapter 4

QCD Dynamics

4.1 Asymptotic Freedom in QCD

4.1.1 The Running of the Coupling Constant

One of the major differences between QED and QCD, is the value of the coupling constant. In QED, each vertex of a Feynman diagram introduces a factor $\alpha = \frac{1}{137}$ so that diagrams with more and more vertices contribute less and less. This is the basis of perturbation theory. However, with there being a caveat that will explain shortly, $\alpha_s \approx 1$. This would seem to indicate that Feynman graphs with more and more vertices would contribute more and more, thereby rendering the breakdown of perturbation theory, and destroying the viability of QCD. Now, the caveat. One of the triumphs of quantum field theory was the discovery that the coupling constant α_s is, in fact, not constant. Instead, it is a function of energy, or equivalently, distance. As it turns out, α_s gets smaller at larger energies. This means, experimentally, as long as we probe at sufficiently high energies, or sufficiently short distances, α_s is small so we can use perturbation theory. This is called the *running of the coupling constant.*

4.1.2 Asymptotic Freedom

The function which describes the running of the coupling constant, is called the beta function, and is given by

$$\beta(g) = \frac{\partial g}{\partial \ln \mu}. \tag{4.1}$$

For a non-abelian gauge theory with gauge group G,

$$\beta(g) = -\frac{g^3}{(4\pi)^2}\left(\frac{11}{3}C_2(G) - \frac{4}{3}N_f C(r)\right) \tag{4.2}$$

where $C_2(G)$ is the quadratic casimir of the adjoint representation of G, and $C(r)$ is defined by

$$Tr[t_r^a t_r^b] = C(r)\delta^{ab}. \tag{4.3}$$

If we assume $G = SU(N)$ and that the fermions lie in the fundamental representation, the result becomes

$$\beta(g) = -\frac{g^3}{(4\pi)^2}\left(\frac{11}{3}N - \frac{2}{3}N_f\right). \tag{4.4}$$

The solution to (4.1) is [26]:

$$g^2(q) = \frac{g^2}{1 + \frac{g^2}{16\pi^2}(\frac{11}{3}N - \frac{2}{3}N_f)\ln(q^2/\mu^2)}. \tag{4.5}$$

From this equation, we see the crucial quantity is $11N - 2N_f$, or equivalently, the sign of the beta function. If $\beta < 0$, g gets small on small distance scales. Qualitatively, this means quarks and gluons are free at short distances and large energies. This is known as *asymptotic freedom*. It was shown by 't Hooft, Politzer, Gross, and Wilczek [14,27] that the only asymptotically free theories in 4-dimensions are non-abelian gauge theories. From equation (4.4), we see that for an $SU(N)$ gauge theory, $N_f < \frac{11}{2}N$ guarantees the theory is asymptotically free. If we take $N = 3$, as it is in the real world, any theory where the number of flavors is 16 or less will be asymptotically free. Because of the smallness of the coupling constant, the high energy regime in asymptotically free theories is accessible by perturbation theory.

4.2 Confinement

Let's continue to assume $\beta < 0$, but instead of examining the high energy, small distance regime, let's examine the low energy, long distance regime. From equation (4.5), g gets bigger as q gets smaller. Thus, the force between two color carrying particles does not diminish on increasing distance scales, which therefore invalidates the use of perturbation theory for large distances, or small energies. Moreover, lattice computations have shown that the coupling grows linearly with distance. This is the phenomenon known as *color confinement*, or simply *confinement*. Because of this force law, it would take an infinite amount of energy to separate quarks, and therefore quarks are bound together in hadrons. This is why searches for free quarks have not turned up any results. While widely believed to be true, confinement has not been analytically proven. The problem is that the regime needed to be probed to prove confinement is where the coupling constant is large, and therefore perturbation theory is not valid. An analytic proof of confinement would therefore require non-perturbative methods.

Chapter 5

Instanton Solutions

5.1 A Lightning Quick Introduction to Homotopy Theory

The basic idea of the homotopy groups is to encode information about the hole structure, and therefore the topology, of a given space. Let's begin by considering a space X and a point $p \in X$. Intuitively, we want to examine all continuous paths in X that both begin and end at the point p. We will then consider two paths to be equivalent if they can be deformed into each other. It will turn out that the set of equivalence classes forms a group, known as the *fundamental group*. Let's be more precise and consider the following set of continuous functions on X

$$\mathcal{F}(p) \equiv \{f : [0,1] \to X | f(0) = f(1) = p\}. \tag{5.1}$$

To capture the notion that two paths are considered to be the same if they can be continuously deformed into one another, we define an equivalence relation:

$$f \sim g \quad \text{iff there exists a} \quad \phi : [0,1] \to \mathcal{F}(p) \quad \text{such that} \quad \phi(0) = f \quad \text{and} \quad \phi(1) = g. \tag{5.2}$$

Now, we define the set

$$\pi_1(X,p) \equiv \mathcal{F}(p)/\sim . \tag{5.3}$$

We can also define multiplication on π_1 by defining

$$(f * g)(t) \equiv \begin{cases} f(2t) & : \quad 0 \le t \le 1/2 \\ g(2t-1) & : \quad 1/2 \le t \le 1 \end{cases} . \tag{5.4}$$

This turns π_1 into a group, known as the fundamental group. It can be shown that $\pi_1(X,p)$ is a topological invariant, and is independent of the point p. π_1 is a good way to characterize topological spaces.

As an example, let D denote a two dimensional disc. Since any closed loop in D can be deformed to a point, the fundamental group is trivial

$$\pi_1(D) = 0. \tag{5.5}$$

Let A be an annulus. A closed loop which winds around the hole, cannot be continuously unwound. In fact, a closed loop which wraps around m times cannot be continuously deformed to a path which winds around n times. Therefore, the fundamental group of the annulus is

$$\pi_1(A) = \mathbf{Z}. \tag{5.6}$$

The fundamental group involves maps $S^1 \to X$, but one can consider higher dimensional spheres $S^n \to X$. This leads to the higher homotopy groups, $\pi_n(X)$. We will be primarily concerned with maps $S^3 \to X$ and the group π_3. It turns out that if X is a simple, non-abelian, compact Lie group, then $\pi_3(X) = \mathbf{Z}$. In particular,

$$\pi_3(SU(3)) = \mathbf{Z}. \tag{5.7}$$

This result will play a critical role in the development of instantons.

5.2 Instantons in Quantum Mechanics

For our purposes, it will be convenient to work with the Euclidean path integral, S_{eucl}, obtained from the standard Minkowski path integral through a Wick rotation. For the approximate evaluation of S_{eucl}, one typically finds the global minimum of S (hereafter, we drop the eucl subscript) and then replace S near this minimum with its quadratic fluctuations. However, this very same procedure can be applied to local minima of S as well. Typically, the result obtained from expanding about a local minimum will be negligible compared to that obtained by expanding about a global minimum, but this is not always the case. The contribution from the functional integral about a local minimum is dampened by the factor $e^{-\delta S}$ where $\delta S = |S_{global} - S_{local}|$. However, for certain non-perturbative processes, the contribution from the global minimum is zero, and so the local minimum contribution becomes dominant. The local minima of S_{eucl} are known as *instantons.*

As an example, let's consider a quantum mechanical system whose Hamiltonian is given by [29]

$$H = \frac{1}{2}p^2 + \lambda(q^2 - a^2)^2. \tag{5.8}$$

The Euclidean action we need to consider is

$$S = \int \left(\frac{1}{2} \left(\frac{dq}{dt} \right)^2 + \lambda(q^2 - a^2)^2 \right) dt. \tag{5.9}$$

The global minima of (5.9) are given by

$$q(t) = \pm a. \tag{5.10}$$

However, there are also local minima, given by

$$q(t) = \pm a \tanh \left[a(t - c)\sqrt{2\lambda} \right]. \tag{5.11}$$

The two local minima are going to be our instantons, and as we will shortly see, they are topologically non-trivial. By topologically non-trivial, we mean that the solution cannot be continuously deformed to the global minimum through finite action configurations. In this simple system, this is equivalent to having different limits at $\pm\infty$. With this in mind, our local minima instantons become topologically non-trivial.

Classically, the two global minima of our system give two degenerate ground states. However, due to quantum mechanical tunneling, these two states are only quasi stationary, and hence are in fact non-degenerate. If we now consider the transition matrix from one classical vacuum to the other, it is evident that we must consider instantons, because topologically trivial solutions do not contribute to this process. The physical meaning of instantons is now apparent: they are solutions of the euclidean equations of motion and mediate tunneling processes between classical vacuums.

5.3 Instantons in Gauge Theories

5.3.1 Definition of an Instanton

Consider a gauge theory with gauge group G and let's study the space of gauge fields for which the Euclidean action is finite. Finite Euclidean action implies $F_{\mu\nu}$ is $O(1/r^3)$. Naively, one would say this implies that A_μ is $O(1/r^2)$. But in fact, this only implies A_μ is a gauge transformation of zero:

$$A_\mu = ih\partial_\mu h^{-1} + O(1/r^2) \tag{5.12}$$

where h is a function of angular variables only and takes values in the gauge group G. That is, h is a mapping $S^3 \to G$ so every finite-action field configuration is associated with a mapping $S^3 \to G$. Since $\pi_3(G) = \mathbf{Z}$ the space of gauge fields splits into components. The gauge field which minimizes the action in a given component is an instanton.

Definition 5.1 (Instanton) *An instanton is a finite action solution of the Euclidean equations of motion.*

5.3.2 Instantons in Gauge Theories

We now turn our attention to instantons in gauge theories. Once again, we will be considering the Euclidean path integral. Since we are looking for finite-action solutions, $F_{\mu\nu}$ must fall off faster than $1/r^2$, because the action contains a term $F_{\mu\nu}^2$. Naively, this would then imply that A_μ goes as $O(1/r^2)$. However, this is incorrect.

Rather than implying that A_μ is $O(1/r^2)$, this implies that A_μ is a gauge transformation of zero. That is,

$$A_\mu = ih\partial_\mu h^{-1} + O(1/r^2) \tag{5.13}$$

where h is a function of angular variables only and takes values in the gauge group G. This means h is a mapping $S^3 \to G$, and therefore every finite action field configuration is associated with a mapping $S^3 \to G$. Moreover, we know that $\pi_3(G) = \mathbf{Z}$ and so the space of gauge fields splits into components. The gauge field which minimizes the action in a given component is an instanton.

5.3.3 A Two-Dimensional Example

Let's consider a two-dimensional example. Consider a $U(1)$ theory on two-dimensional spacetime. $U(1)$ is topologically equivalent to S^1 and therefore h is a mapping $S^1 \to S^1$. Since the fundamental group of S^1 is $\mathbf{Z}$, every finite action field is associated with an integer. This integer goes by several names:

- Winding number
- Topological charge
- Pontryagin index.

If we now parameterize S^1 by ϕ, then the canonical mapping of $S^1 \to S^1$ of winding number n is given by

$$h_n(\phi) = e^{in\phi} \in U(1). \tag{5.14}$$

Every mapping $S^1 \to S^1$ is homotopic to h_n, for some n.

5.3.4 SU(2) on Four-Dimensional Spacetime

Let's now return to four-dimensions and consider an $SU(2)$ gauge theory. $SU(2)$ is topologically equivalent to S^3 so we consider maps $h : S^3 \to S^3$. Since $\pi_3(SU(2)) = \mathbf{Z}$, every h, and therefore A_μ, has a winding number. The mapping of winding number n is given by

$$h_n(x) = [h_1(x)]^n \tag{5.15}$$

where

$$h_1(x) = (x_4 \cdot \mathbf{1} + i\vec{x} \cdot \vec{\sigma})/r. \tag{5.16}$$

Every mapping $S^3 \to S^3$ is homotopic to one of these.

5.3.5 The Structure of the Vacuum

Let's now analyze the structure of the vacuum in gauge theories. Consider a theory of gauge fields only. Classically, the ground state is given by $F_{\mu\nu} = 0$, which means that A_μ must be a gauge transformation of zero. Now let A_μ and $\widetilde{A}_\mu$ be two gauge transformations of zero with different winding numbers. If we try to deform one into the other, we must pass through vector potentials which are not gauge transformations of zero, and therefore we must pass through field strengths $F_{\mu\nu}$ which do not vanish. These non vanishing field strengths imply non vanishing energy. In other words, there is an energy barrier between A_μ and $\widetilde{A}_\mu$. Therefore, A_μ and $\widetilde{A}_\mu$ represent different vacuum states. Thus, for each integer n, there is a quantum vacuum state $|n\rangle$. These are known as *n vacua*.

5.3.6 The Action of an Instanton

It can be shown that the winding number n of a gauge field A_μ is given by

$$n = \frac{g^2}{16\pi^2} \int d^4x Tr[\widetilde{F}^{\mu\nu} F_{\mu\nu}] \tag{5.17}$$

where $\widetilde{F}^{\mu\nu}$ is the *dual field strength* and is given by

$$\widetilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\sigma\tau} F_{\sigma\tau}. \tag{5.18}$$

We have

$$\frac{1}{2} Tr[\widetilde{F}_{\mu\nu} \pm F_{\mu\nu}]^2 = Tr[F^{\mu\nu} F_{\mu\nu}] \pm Tr[\widetilde{F}^{\mu\nu} F_{\mu\nu}]. \tag{5.19}$$

Therefore,

$$\int d^4x Tr[F^{\mu\nu} F_{\mu\nu}] \geq \left| \int d^4x Tr[\widetilde{F}^{\mu\nu} F_{\mu\nu}] \right|. \tag{5.20}$$

This, in turn implies

$$S \geq 8\pi^2 |n|/g^2. \tag{5.21}$$

Configurations with $F = \pm\tilde{F}$ saturate the bound, so such configurations are instantons since instantons minimize the action of a given winding number. It therefore follows that the action of an instanton of winding number n is

$$S_n = 8\pi^2 |n|/g^2. \tag{5.22}$$

5.3.7 A Summary

Here is a summary of what we have learned so far:

- Every gauge field has a winding number.

- The space of gauge fields separates into components, one for each winding number.
- An instanton of winding number n minimizes the Euclidean action in the nth component.
- An instanton is a classical solution of the Euclidean equations of motion.
- Gauge theory yields a countably infinite set of vacua.
- The action of an instanton of winding number n is given by $S_n = 8\pi^2|n|/g^2$.

5.4 The Strong CP Problem

5.4.1 Θ Vacua

The action of an n-instanton is $S_n = \frac{8\pi^2}{g^2}|n|$. Between vacua $|m\rangle$ and $|n\rangle$ there is a tunneling amplitude of the form $\langle m|H|n\rangle \sim \mathrm{e}^{-S}$ where $S = \frac{8\pi^2}{g^2}|n-m|$ is the action of the instanton of winding number $|n-m|$. This means the n-vacua are not eigenstates of the Hamiltonian. We therefore define the Θ-vacua, which are eigenstates of the Hamiltonian:

$$|\theta\rangle = \sum_{n=-\infty}^{\infty} \mathrm{e}^{-in\theta}|n\rangle. \tag{5.23}$$

With the inclusion of the θ-vacuum, the Euclidean path integral becomes

$$Z_\theta[J] = \int [dA] \exp \int \mathrm{Tr}\left[-\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{ig^2\theta}{32\pi^2}\widetilde{F}^{\mu\nu}F_{\mu\nu} + J^\mu A_\mu\right] d^4x. \tag{5.24}$$

Therefore the Lagrangian of QCD can contain a term

$$\frac{ig^2\theta}{32\pi^2}\widetilde{F}^{\mu\nu}F_{\mu\nu} \tag{5.25}$$

where $\theta \in [0, 2\pi)$. However, this term violates CP and experimental constraints place $\theta < 10^{-10}$ [4]. This begs the question, why is Θ so small? This conundrum is known as the *Strong CP Problem.*

5.4.2 A Possible Solution

Consider QCD with one flavor of massless quark u. The path integral is

$$\int [dA][du][d\bar{u}] \exp i \int \left[i\bar{u}\not{D}u - \frac{1}{4}F^{a\mu\nu}F^{a}_{\mu\nu} - \frac{g^2\theta}{32\pi^2}\widetilde{F}^{a\mu\nu}F^{a}_{\mu\nu} \right] d^4x. \tag{5.26}$$

The quark action is invariant under a $U(1)_A$ symmetry under which

$$u \mapsto \mathrm{e}^{-i\alpha\gamma_5}u \tag{5.27}$$

$$\bar{u} \mapsto \bar{u}\mathrm{e}^{-i\alpha\gamma_5}. \tag{5.28}$$

But the $U(1)_A$ is anomalous and the integration measure picks up a factor

$$[du][d\bar{u}] \mapsto \exp\left[-i\frac{g^2\alpha}{16\pi^2} \int \widetilde{F}^{a\mu\nu}F^{a}_{\mu\nu}d^4x \right] [du][d\bar{u}] \tag{5.29}$$

The effect of a $U(1)_A$ transformation is that $\theta \mapsto \theta + 2\alpha$. Since the value of θ can change by a $U(1)_A$ transformation, the path integral cannot depend on θ. In other words, adding a massless quark turns the theta angle into a physically irrelevant, unobservable parameter. To see this another way, integrate over the quark field in the path integral to get

$$Z = \int [dA] \det(i\not{D})\mathrm{e}^{iS}\mathrm{e}^{in\theta}. \tag{5.30}$$

If Z is to be independent of θ, gauge fields with non-zero winding number must not contribute. Evidently, it must be the case that $i\not{D}$ has a zero eigenvalue whenever the gauge field has a non-zero winding number.

The possibility of a massless quark is the main motivation for the following section. Indeed, lattice gauge calculations appear to demonstrate that $m_u \neq 0$. How reliable are these results? The goal of the next chapter is to find a non-perturbative, calculable quantity, that, in theory, can be used as a benchmark for lattice QCD. As we will see shortly, the quantity we will study is the Green's function

$$\langle \overline{u}(x)u(x)\overline{d}(x)d(x)\overline{s}(0)s(0)\rangle. \tag{5.31}$$

Chapter 6

Reliable Semiclassical Computations in QCD

6.1 Introduction and Motivation

Over the last few decades, semiclassical computations have proven valuable in understanding a variety of effects in quantum field theory and string theory. In four dimensional gauge theories, 't Hooft showed in his original work that systematic computations are possible in a Higgs phase [31]. It was soon realized that in QCD at high temperatures, instantons provide the leading contribution to θ-dependent quantities, as well as to certain chirality violating amplitudes [1,13]. These methods found applications to the understanding of baryon number violation in the Standard Model, the axion potential at high temperatures and the dynamics of supersymmetric gauge theories. But the question of the utility of instantons at zero temperature in real QCD has remained elusive. Even for short distance quantities, such computations are plagued by infrared divergences. These divergences arise precisely in the region of strong coupling, suggesting that instantons are, at best, just one of many important non-perturbative effects. Witten, indeed, argued compellingly that at large N, instanton effects are negligible [33].

A number of authors in the past have considered possible contributions to physical processes in QCD by studying momentum space Green's functions [2, 3, 25]. Such computations, remarkably, are infrared finite, and arguably might represent distinguishable effects. But there are reasons to be skeptical. Most strikingly, in the cases which have been studied, these instanton contributions fall very rapidly with momentum. It is not clear that, e.g. in large N, the leading contributions to the corresponding quantities won't swamp the semiclassical contributions.

In this section, we resolve this question by considering a set of correlation functions which receive no perturbative contribution but do receive contributions from instantons. They receive other non-perturbative contributions as well, as readily seen in the large N limit. By considering the structure of the Operator Product Expansion (OPE), we can show that, for particular numbers of flavors and colors, instantons provide the most singular contribution to these Green's functions at short distances, while for others they are sub dominant. In the former cases, as we demonstrate self-consistently, a complete semiclassical computation is possible. There is a systematic expansion of the most singular part of the coefficient function. The leading instanton contribution to the singularity is infrared finite, and corrections can be calculated systematically in powers of α_s.

6.2 The Vacuum to Vacuum Amplitude

6.2.1 Setup

In order to calculate the Green's function of equation (5.31), we first must know how to compute the vacuum to vacuum amplitude

$$W^{(1)} = \int [dA][d\psi][d\phi]\mathrm{e}^{-S}. \tag{6.1}$$

In the expression above, A represents the gauge field, ψ the quarks, and ϕ the ghosts. In this section, we outline the basic procedure of how to calculate $W^{(1)}$, and in subsequent sections we go into more detail.

To calculate $W^{(1)}$, we will expand about the $n = 1$ instanton solution and write

$$A_\mu = A_\mu^{inst} + A_\mu^{qu}. \tag{6.2}$$

The action becomes

$$S = S^{inst} - \frac{1}{2} A^{qu} \mathcal{M}_A A^{qu} + \overline{\psi} \mathcal{M}_\psi \psi - \phi^* \mathcal{M}_{gh} \phi. \tag{6.3}$$

Naively, one might say that

$$W^{(1)} = e^{-S^{inst}} \frac{(\det \mathcal{M}_A)^{-1/2} \det \mathcal{M}_\psi \det \mathcal{M}_{gh}}{\left((\det \mathcal{M}_A)^{-1/2} \det \mathcal{M}_\psi \det \mathcal{M}_{gh}\right)_{A=0}}. \tag{6.4}$$

However, this is wrong because $\mathcal{M}$ has zero eigenvalues. Normally, the minimum of the action occurs at a point. In this case, however, the minimum occurs over an entire manifold. There is a zero eigenvalue of $\mathcal{M}$ for each symmetry of the theory under which the instanton solution is not invariant. It is then necessary to integrate over this manifold of solutions. The correct expression is

$$W^{(1)} = e^{-S^{inst}} \int Q(\gamma) d\gamma, \tag{6.5}$$

where:

- $d\gamma$ is the measure on the instanton manifold.
- γ describes the *collective coordinates.*
- $Q(\gamma)$ is the determinantal expression above, but taken only over non-zero eigenvalues.

For $SU(2)$,

$$A_\mu^{a\ inst} = \frac{2}{g} \frac{\eta_{a\mu\nu}(x-z)^\nu}{(x-z)^2 + \rho^2}. \tag{6.6}$$

ρ is called the "size" of the instanton, z is called the "center" of the instanton, and $d\gamma = \frac{d\rho d^4 z}{\rho^5}$. 't Hooft [31] found $Q(\gamma)$ by explicitly finding and multiplying the non-zero eigenvalues. The regularized result for $SU(2)$ and three massless flavors in the $\overline{MS}$ scheme is:

$$W^{(1)} = \frac{2^{10}\pi^6 \mu^{16/3}}{g^8} \exp\left[-\alpha(1) + 6\alpha(1/2) - 5/36\right] \int \rho^{1/3} d\rho d^4 z. \tag{6.7}$$

The function $\alpha(t)$ in the above expression is defined by

$$\alpha(t) \equiv C(t) \left[2R - \frac{1}{6}\ln 2 + \frac{1}{2}\sum_{s=1}^{2t+1} s(2t+1-s)(s-t-\frac{1}{2})\ln s - \frac{1}{6}t(t+1) - \frac{1}{9} \right] \tag{6.8}$$

where

$$C(t) \equiv \frac{2}{3}t(t+1)(2t+1) \tag{6.9}$$

and

$$R \equiv \frac{1}{12}(\ln 2\pi + \gamma) + \frac{1}{2\pi^2}\sum_{2}^{\infty}\frac{\ln s}{s^2} \approx 0.248754. \tag{6.10}$$

6.2.2 Details of the $SU(2)$ Calculation

Now that we have the basic ingredients and have defined the important quantities, we are ready to tackle the problem of calculating $W^{(1)}$ in some detail. This calculation was first done by 't Hooft [31].

Consider an $SU(2)$ gauge theory with N_f flavors and a Lagrangian given by

$$\mathcal{L} = -\frac{1}{4}F^a_{\mu\nu}F^a_{\mu\nu} - \bar{\psi}\not{D}\psi. \tag{6.11}$$

We then write $A^a_\mu = A^{a\ inst}_\mu + A^{a\ qu}_\mu$ where $A^{a\ inst}_\mu$ is the classical instanton solution

$$A^{a\ inst}_\mu = \frac{2}{g}\frac{\eta_{a\mu\nu}(x-x_0)^\nu}{(x-x_0)^2+\rho^2}. \tag{6.12}$$

Using background field gauge $G^a = D_\mu A^{a\ qu}_\mu$ the action now becomes

$$S = S^{inst} - \frac{1}{2}A^{qu}\mathcal{M}_A A^{qu} + \bar{\psi}\mathcal{M}_\psi\psi - \phi^*\mathcal{M}_{gh}\phi \tag{6.13}$$

where the fields ϕ are the ghosts and

$$\mathcal{M}_A A^{a\ qu}_\mu = -D^2 A^{a\ qu}_\mu - 2g\epsilon_{abc}F^{b\ inst}_{\mu\nu}A^{c\ qu}_\nu \tag{6.14}$$

$$-\mathcal{M}^2_\psi\psi = -D^2\psi + \frac{i}{4}\tau^a F^{a\ inst}_{\mu\nu}\gamma_\mu\gamma_\nu\psi \tag{6.15}$$

$$\mathcal{M}_{gh}\phi = -D^2\phi. \tag{6.16}$$

$S^{inst} = 8\pi^2/g^2$ and the covariant derivative D is to be taken with the background field A^{inst}. We wish to calculate the normalized one-loop vacuum to vacuum amplitude $W^{(1)}$, expanded about the instanton solution. Let γ_i denote the collective coordinates of the instanton (to be discussed later). Then

$$\begin{aligned} W^{(1)} &= \int [dA][d\psi][d\phi] e^{S} && (6.17) \\ &= \int \prod_i d\gamma_i Q(\gamma) J(\gamma) F(\gamma) e^{-8\pi^2/g^2}, && (6.18) \end{aligned}$$

where

$$Q(\gamma) = \frac{(\det \mathcal{M}_A)^{-1/2} \det \mathcal{M}_\psi \det \mathcal{M}_{gh}}{\left((\det \mathcal{M}_A)^{-1/2} \det \mathcal{M}_\psi \det \mathcal{M}_{gh}\right)_{A=0}}, \tag{6.19}$$

$J(\gamma)$ is the Jacobian from passing to collective coordinates, and $F(\gamma)$ is a fermionic factor to be calculated later.

We will calculate the determinants by finding the eigenvalues and then multiplying them together. We consider here the eigenvalue equation $\mathcal{M}\Psi = E\Psi$ for one complex scalar of isospin t. We need to compute $\prod_n \frac{E(n)}{E_0(n)}$, however, the solutions to $\mathcal{M}\Psi = E\Psi$ cannot be expressed in elementary functions. The equation $\mathcal{M}\Psi = \frac{\lambda}{(1+r^2)^2}\Psi$, however is a hypergeometric equation with known solutions. We therefore define the operators [31]

$$\begin{aligned} \mathcal{V} &= \frac{1}{4}(1+r^2)\mathcal{M}(1+r^2) && (6.20) \\ \mathcal{V}_0 &= \frac{1}{4}(1+r^2)\mathcal{M}_0(1+r^2). && (6.21) \end{aligned}$$

The determinants $\det(\mathcal{M}/\mathcal{M}_0)$ and $\det(\mathcal{V}/\mathcal{V}_0)$ are identical so we choose to solve the equation $\mathcal{V}\Psi = \lambda\Psi$. These eigenvalues are

$$\lambda_n = (n + l + j_1 + 1 - t)(n + l + j_1 + 2 + t) \tag{6.22}$$

The eigenvalues of $\mathcal{V}_0$ are given by the same expression but with $j_1 = l$ and $t = 0$.

The product $\prod \lambda(n)/\lambda_0(n)$ diverges so we must regularize. We therefore introduce a set of Pauli-Villars regulators of mass μ_i and alternating metric $e_i = \pm 1$. With this regularization we are now interested in the product

$$\det \mathcal{M} \prod_i (\det \mathcal{M}_i)^{e_i}. \tag{6.23}$$

Choosing a mass for the Pauli-Villars regulators independent of spacetime renders the eigenvalue equation for M_i difficult to solve. If, however, we choose

$$\mu_i = \frac{M_i}{(1 + r^2)^2}, \tag{6.24}$$

for M_i independent of spacetime, the eigenvalues are easily seen to be

$$\lambda_n^{M_i} = \lambda_n + M_i^2. \tag{6.25}$$

We now have all the pieces to calculate the regularized product $\Pi(t)/\Pi_0(t)$. After some algebraic manipulations, we find [31]

$$\begin{aligned} \ln[\Pi(t)/\Pi_0(t)] = \frac{t(t+1)(2t+1)}{3} \Bigg[& \frac{1}{3} \ln M + 4R \\ & + \frac{1}{t(t+1)} \sum_{s=1}^{2t+1} s(2t + 1 - s)(s - t - \frac{1}{2}) \ln s - \frac{1}{3} t(t+1) - \frac{1}{2} \Bigg] \end{aligned} \tag{6.26}$$

where $R = 0.248754477$ and M is defined by $\sum e_i \ln M_i^2 = -\ln M$. Since we would like to compare the result with standard QCD computations, it is desirable to switch to a new regulator that is independent of spacetime. In doing so, we must introduce a counter term Lagrangian $\Delta\mathcal{L}$. If we were switching from a spacetime independent regulator μ to another spacetime independent regulator μ_0, the counterterm lagrangian would be

$$\Delta\mathcal{L} = -\frac{g^2}{32\pi^2} F^a_{\mu\nu} F^a_{\mu\nu} \times \frac{1}{9} t(t+1)(2t+1)\ln(\mu/\mu_0). \tag{6.27}$$

In our case, the regulator μ is not independent of spacetime, but instead is given by

$$\mu^2 = \frac{M^2}{(1+r^2)^2}. \tag{6.28}$$

Since the regulator is spacetime dependent, the counter term lagrangian must be spacetime dependent as well, and in fact, by locality, must also be given by equation (6.27). Inputing our expressions for μ and $F^{a\ inst}_{\mu\nu}$, the change in the classical action becomes,

$$\begin{aligned} \Delta S^{cl} &= \int \Delta\mathcal{L} d^4x & (6.29) \\ &= \frac{2}{3} t(t+1)(2t+1)\left(\frac{1}{6}\ln\frac{\mu_0}{2M} + \frac{5}{36}\right). & (6.30) \end{aligned}$$

We add this term to the regularized product to get,

$$\begin{aligned} \ln[\Pi(t)/\Pi_0(t)] = \frac{t(t+1)(2t+1)}{3} \Bigg[\frac{1}{3}\ln\frac{\mu_0}{2} + 4R \\ + \frac{1}{t(t+1)} \sum_{s=1}^{2t+1} s(2t+1-s)(s-t-\frac{1}{2})\ln s - \frac{1}{3}t(t+1) - \frac{2}{9} \Bigg] \end{aligned} \tag{6.31}$$

This expression represents the regularized product of nonzero eigenvalues for one complex scalar of isospin t. The vector field has four *real* components and isospin one. Moreover, for vector fields we need $\Pi_0(t)/\Pi(t)$ and so for the vector field contribution we take $t = 1$ and multiply by -2. This gives,

$$\left(\frac{\det \mathcal{M}_A}{(\det \mathcal{M}_A)_{A=0}}\right)^{-1/2} = \mu_0^{-4/3} \exp\left(-16R - \frac{2}{3}\ln 2 + \frac{32}{9}\right). \tag{6.32}$$

For the ghost field we take $t = 1$ to obtain

$$\frac{\det \mathcal{M}_{gh}}{(\det \mathcal{M}_{gh})_{A=0}} = \mu_0^{-2/3} \exp\left(8R + \frac{1}{3}\ln 2 - \frac{16}{9}\right). \tag{6.33}$$

Each fermion flavor carries four complex Dirac components, but the above product applies to $\mathcal{M}_\psi^2$. So for the total fermionic contribution we take $t = 1/2$ and multiply by $2N_f$. This gives,

$$\frac{\det \mathcal{M}_\psi}{(\det \mathcal{M}_\psi)_{A=0}} = \mu_0^{N_f/3} \exp\left[N_f\left(4R - \frac{1}{3}\ln 2 - \frac{17}{36}\right)\right]. \tag{6.34}$$

Putting it all together, and anticipating the collective coordinate ρ, we get the determinantal factor,

$$Q(\rho) = (\mu_0\rho)^{\frac{N_f-2}{3}} \exp\left[4(N_f-2)R - \frac{N_f+1}{3}\ln 2 + \frac{64-17N_f}{36}\right]. \tag{6.35}$$

Each massive flavor has a term $m\bar{\psi}\psi$ in the action. If m is sufficiently small, we can do an expansion in m. This gives a factor of

$$m\int \bar{\psi}\psi d^4x \;\rightarrow\; m\int d\bar{c}_0\bar{c}_0 dc_0 c_0 \int d^4x \bar{\psi}_0\psi_0 \tag{6.36}$$

$$= m\langle\psi^{(0)}|\psi^{(0)}\rangle \tag{6.37}$$

$$= \rho m \tag{6.38}$$

for each massive flavor. Moreover, each flavor contributes a factor $(\mu_0\rho)^{-1}$ due to its zero mode, and therefore,

$$F(\rho) = \frac{(\rho m)^{N_m}}{(\mu_0\rho)^{N_f}}, \tag{6.39}$$

where N_m is the number of massive flavors (all assumed to have the same mass m).

Combining $Q(\rho)$, $J(\rho, z)$, and $F(\rho)$ we obtain $W^{(1)}$:

$$W^{(1)} = \frac{2^{10}\pi^6 m^{N_m} \mu_0^{\frac{2}{3}(11-N_f)}}{g^8} \exp\left[-\frac{8\pi^2}{g^2} + (N_f - 2)R - \frac{N_f + 1}{3} \ln 2 + \frac{64 - 17N_f}{36}\right] \int \rho^{\frac{7-2N_f+3N_m}{3}} d^4 z d\rho. \tag{6.40}$$

6.2.3 Generalizing to N Colors and N_f Flavors

We consider an $SU(N)$ gauge theory with N_f flavors of massless quarks. The one loop amplitude in the presence of an instanton is given by

$$W^{(1)}_{\overline{MS}}(N, N_f) = \int d\gamma J(\gamma) Q(\gamma) F(\gamma) V M \exp\frac{-8\pi^2}{g^2} \tag{6.41}$$

where:

- $J(\gamma)$ is the collective coordinate Jacobian.
- $Q(\gamma)$ is the contribution to the determinant from the gauge bosons and ghosts.
- $F(\gamma)$ is the total fermionic contribution coming from the zero modes and the determinant.
- V is the integral over the collective coordinates which describe the orientation of an instanton in the group $SU(N)$. Equivalently, it is the volume of $SU(N)/T_N$ where T_N is the stability subgroup of the instanton.
- M is a factor coming from changing to the $\overline{MS}$ scheme.

Bernard [7] calculated these quantities for the general $SU(N)$ case. They are given as follows:

- $d\gamma = d\rho d^4 z$.

- $J(\gamma) = 2^{4N+2}\pi^{2N}\rho^{4N-5}g^{-4N}$.

- $Q(\gamma) = \mu^{11N/3}\rho^{-N/3}\exp\left[-\alpha(1) - 2(N-2)\alpha(1/2)\right]$.

- $F(\gamma) = \mu^{-2N_f/3}\rho^{-2N_f/3}\exp\left[2N_f\alpha(1/2)\right]$.

- $V = \frac{\pi^{2N-2}}{(N-1)!(N-2)!}$.

- $M = \exp\left[\frac{5(N-N_f)}{36}\right]$.

Combining all these gives:

$$\begin{aligned} W^{(1)}_{\overline{MS}}(N, N_f) &= \frac{2^{4N+2}\pi^{4N-2}\mu^{b_0}}{g^{4N}(N-1)!(N-2)!} \\ &\times \exp\left[-\alpha(1) + 2(N_f - N + 2)\alpha(1/2) + \frac{5(N-N_f)}{36} - \frac{8\pi^2}{g^2(\mu)}\right] \\ &\times \int \rho^{b_0-5} d\rho d^4 z. \end{aligned} \tag{6.42}$$

6.2.4 Green's Function Calculation

The fermionic zero modes are given by

$$q_i^\alpha(x-z) = \frac{\sqrt{2}}{\pi}\frac{\rho^{3/2}}{((x-z)^2+\rho^2)^{3/2}}\delta_i^\alpha. \tag{6.43}$$

Here, i is the color index and α the spin index. These are normalized to

$$\langle q_i^\alpha | q_\beta^j \rangle = \rho \delta_i^\alpha \delta_\beta^j. \tag{6.44}$$

Put

$$O^{\alpha_1 j_1 \cdots \alpha_{N_f} j_{N_f}}_{i_1\beta_1\cdots i_{N_f}\beta_{N_f}}(x) = \left(\prod_{a=1}^{A} (\overline{q_a})^{\alpha_a}_{i_a}(x)(q_a)^{j_a}_{\beta_a}(x)\right)\left(\prod_{b=A+1}^{N_f} (\overline{q_b})^{\alpha_b}_{i_b}(0)(q_b)^{j_b}_{\beta_b}(0)\right) \tag{6.45}$$

where $A = \lfloor \frac{N_f}{2} \rfloor$, and a and b run over flavor. For ease of notation, we will suppress spin and color indices from here on. Define the green's function

$$G(x) = \langle O(x) \rangle. \tag{6.46}$$

This is then equal to

$$\begin{aligned} G(x) &= W^{(1)}_{\overline{MS}}(N, N_f) \left(\frac{\sqrt{2}}{\pi} \frac{\rho^{3/2}}{((x-z)^2 + \rho^2)^{3/2}} \right)^{2A} \left(\frac{\sqrt{2}}{\pi} \frac{\rho^{3/2}}{(z^2 + \rho^2)^{3/2}} \right)^{2(N_f - A)} \\ &= C(N, N_f) \frac{\Lambda_{QCD}^{b_0}}{\alpha_s^{2N}(\mu)} x^{b_0 - 3N_f} \end{aligned} \tag{6.47}$$

where

$$\begin{aligned} C(N, N_f) &= \frac{2^{N_f+2} \pi^{2(N - N_f - 1)}}{(N-1)!(N-2)!} \\ &\times \exp\left[-\alpha(1) + 2(N_f - N + 2)\alpha(1/2) + \frac{5(N - N_f)}{36} \right] x^{3N_f - b_0} \\ &\times \int \frac{\rho^{b_0 + 3N_f - 5}}{\left[(x-z)^2 + \rho^2\right]^{3A} \left[z^2 + \rho^2\right]^{3(N_f - A)}} d\rho d^4 z. \end{aligned} \tag{6.48}$$

The integral over ρ converges for

$$N < N_f \leq N + A. \tag{6.49}$$

We do the integral by introducing Feynman parameters and get

$$\begin{aligned} C(N, N_f) &= \frac{\Gamma(\frac{11}{6}N_f - \frac{11}{6}N)\Gamma(\frac{11}{6}N + \frac{7}{6}N_f - 2)}{\Gamma(3A)\Gamma(3N_f - 3A)\Gamma(\frac{11}{3}N - \frac{2}{3}N_f)} \\ &\times \frac{\Gamma(-\frac{11}{6}N_f + \frac{11}{6}N + 3A)\Gamma(\frac{7}{6}N_f + \frac{11}{6}N - 3A)2^{N_f+1}\pi^{-2(N_f - N)}}{(N-1)!(N-2)!} \\ &\times \exp\left[-\alpha(1) + 2(N_f - N + 2)\alpha(1/2) + \frac{5(N - N_f)}{36} \right]. \end{aligned} \tag{6.50}$$

In the case that $N = N_f$ the Green's function is log divergent and requires an infrared cutoff λ. We then get

$$G(x) = -C(N) \frac{\Lambda_{QCD}^{b_0}}{\alpha_s^{2N}(\mu)} \ln(\lambda x) \tag{6.51}$$

where

$$C(N) \approx \frac{2^{N+2}}{(N-1)!(N-2)!} \frac{\Gamma(3N-2)}{\Gamma(3N)} \exp\left[-\alpha(1) + 4\alpha(1/2)\right]. \tag{6.52}$$

Let us first take the case of two colors and three flavors and consider the Green's function

$$\langle \overline{u}(x)u(x)\overline{d}(0)d(0)\overline{s}(0)s(0)\rangle \tag{6.53}$$

where we have contracted color and spinor indices between neighboring quark pairs. This Green's function is equal to

$$\begin{aligned} \langle \overline{u}(x)u(x)\overline{d}(0)d(0)\overline{s}(0)s(0)\rangle &= 2^3 C(2,3)\frac{\Lambda_{QCD}^{16/3}}{\alpha_s^4(\mu)} x^{-11/3} && (6.54)\\ &= 0.361959 \cdot \frac{\Lambda_{QCD}^{16/3}}{\alpha_s^4(\mu)} x^{-11/3}. && (6.55)\end{aligned}$$

Next, we consider the case of three flavors and three colors.

$$\begin{aligned} \langle \overline{u}(x)u(x)\overline{d}(0)d(0)\overline{s}(0)s(0)\rangle &\approx -2^3 C(3)\frac{\Lambda_{QCD}^{9}}{\alpha_s^6(\mu)} \ln(\lambda x) && (6.56)\\ &\approx -2.62968 \cdot \frac{\Lambda_{QCD}^{9}}{\alpha_s^6(\mu)} \ln(\lambda x). && (6.57)\end{aligned}$$

6.3 The Operator Product Expansion

6.3.1 Instantons and Maximally Chirality Violating Green's Functions

Consider a theory with N colors and N_f *massless* flavors, with N_f small enough that the theory is asymptotically free. Label the flavors $q_f, \bar{q}_f$. In an asymptotically free theory, we might hope that the short distance behavior of correlation functions can be systematically analyzed. The (gauge-invariant) Green's function

$$G(x_i) = \langle \prod_{f=1}^{N_f} \bar{q}_f(x_f) q_f(x_f)\rangle \tag{6.58}$$

vanishes in perturbation theory, and receives a contribution from a single instanton. For simplicity, take one set of points, say $x_1, \ldots, x_A = x$, and the other, $x_{A+1}, \ldots x_{N_f} = 0$, where $N_f = 2A$ or $N_f = 2A + 1$, and call

$$\mathcal{O}_1(x) = \prod_{f=1}^{A} \bar{q}_f(x) q_f(x) \qquad \mathcal{O}_2(x) = \prod_{f=A+1}^{N_f} \bar{q}_f(x) q_f(x). \tag{6.59}$$

We study the two point function

$$\Delta(x) = \langle \mathcal{O}_1(x) \mathcal{O}_2(0) \rangle. \tag{6.60}$$

We are interested in the behavior of $\Delta(x)$ for small x. If the correlation function is singular at short distances, we should be able to understand the singularities in terms of the operator product expansion. As for any pair of operators, the operator product expansion, OPE, of $\mathcal{O}_1(x)$, $\mathcal{O}_2(x)$ takes the form:

$$\mathcal{O}_1(x) \mathcal{O}_2(0) \approx \sum C_{12}^n(x) \mathcal{O}_n(0) \tag{6.61}$$

The coefficient functions, in an asymptotically free theory like QCD, can be calculated in a short distance expansion. One expects that these include both perturbative and non-perturbative contributions. In the past, there has been much discussion of possible non-perturbative contributions to OPE's [2,3,25]. In general, this is a subtle problem. If there are perturbative contributions to a coefficient, any non-perturbative contribution which one might hope to isolate will be exponentially small. In the case of $\Delta(x)$, there are no perturbative contributions. One can ask: are any of the coefficient functions dominated by instantons?

In perturbation theory, the theory respects an axial $U(1)$ symmetry, as a result of which the lowest dimension operator allowed on the right hand side of eqn. 6.61

is simply $\mathcal{O}_{3N_f}(0) = \mathcal{O}_1(0)\mathcal{O}_2(0)$, with coefficient $C^{3N_f}(x)$. $C^{3N_f}(x)$ starts out as a constant, and receives corrections in powers of $\alpha_s(x)$ (we will discuss these in section 6.4). More precisely, in perturbation theory, various operators of the same dimension as $\mathcal{O}_{3N_f}$ appear on the right hand side, multiplied by powers of $\alpha_s(x)$; we will label these $\mathcal{O}^a_{3N_f}(0)$, and their coefficients $C^{3N_f}_a(x)$. These will be considered more explicitly in section 6.4. Non-perturbatively, and in particular at the level of a single instanton, the *unit operator* can appear on the right hand side of this expression. In other words,

$$\mathcal{O}_1(x)\mathcal{O}_2(0) \sim C^{(0)}(x) + C^{3N_f}_a(x)\mathcal{O}^a_{3N_f}(0) + \text{ higher dimension.} \tag{6.62}$$

Assuming $C^{(0)}(x)$ is dominated by a well-defined instanton computation, we can determine its behavior on dimensional grounds (up to logarithmic variation associated with the perturbative anomalous dimensions of $\mathcal{O}_1$ and $\mathcal{O}_2$):

$$C^{(0)} = \Lambda^{b_0} x^{-3N_f+b_0}, \tag{6.63}$$

where $b_0 - 3N_f = \frac{11}{3}(N - N_f)$. So $C^{(0)}(x)$ exhibits a power-law singularity at short distances if $N_f > N$, is non-singular if $N_f < N$, and possesses a logarithmic singularity for $N_f = N$. We will see that in the first case, the instanton computation is infrared finite, and there is a systematic expansion of the Green's function in powers of $e^{-\frac{8\pi^2}{g^2(x)}}$ and $g^2(x)$; in the second, the instanton computation is infrared divergent, and can be considered to correct the matrix elements of $\mathcal{O}_{3N_f}$. The $N_f = N$ case requires more careful analysis. Correspondingly, there are three behaviors for the correlation function. For $N_f > N$,

$$\langle \mathcal{O}_1(x)\mathcal{O}_2(0)\rangle \approx C(\Lambda^{b_0} x^{-3N_f+b_0} + \text{ non singular}) \tag{6.64}$$

where the coefficient C is calculable, but the non-singular terms are not. For $N_f < N$

$$\langle \mathcal{O}_1(x)\mathcal{O}_2(0)\rangle \approx D(\Lambda^{3N_f}\log(x\Lambda) + \text{ non singular}) \tag{6.65}$$

where the coefficient, D, is *not calculable.* For $N_f = N$, the effects of instantons and other non-perturbative effects are comparable, and a more detailed analysis is required. This case, of course, is particularly interesting in real QCD, and will be the subject of section 6.4.

Note, most importantly, that in what we refer to as the non-calculable cases, the contributions of the operator $\mathcal{O}_1(0)\mathcal{O}_2(0)$ fall off more slowly than the instanton contributions, so, for example, the contributions to this operator associated with dynamical breaking of the $SU(N_f)$ chiral symmetry are more important than those calculated in the semiclassical approximation.

Let's consider the instanton computation in more detail. In this section, we will examine the structure of the relevant integrals in order to determine their behavior with x; particular cases important for applications will be evaluated in more detail in section 6.5. The fermion zero modes have the form:

$$q^i_\alpha = \rho\frac{\sqrt{2}}{\pi}\frac{\delta^i_\alpha}{[(x-x_0)^2+\rho^2]^{3/2}}, \tag{6.66}$$

where α is a two component spinor index and i is the gauge index. So the instanton contribution to Δ is

$$\Delta(x) = C\int d^4x_0 d\rho \frac{(\Lambda\rho)^{\frac{11}{3}N-\frac{2}{3}N_f}\rho^{3N_f-5}}{[(x-x_0)^2+\rho^2]^{3A}[x_0^2+\rho^2]^{3(N_f-A)}} \tag{6.67}$$

where C is a constant obtained from the non-zero modes in the instanton background, and x_0 and ρ are the translational and rotational collective coordinates, respectively.

The integral over x_0 may be performed by introducing Feynman parameters, yielding

$$\Delta(x) = C' \int d\alpha[\alpha^{3A-1}(1-\alpha)^{3(N_f-A)-1}]d\rho \frac{(\Lambda\rho)^{\frac{11}{3}N-\frac{2}{3}N_f}\rho^{3N_f-5}}{[x^2\alpha(1-\alpha)+\rho^2]^{3N_f-2}}. \tag{6.68}$$

For large ρ, this behaves as

$$\Delta \sim \int \frac{d\rho}{\rho}\rho^{\frac{11}{3}(N-N_f)}. \tag{6.69}$$

The integral converges for large ρ if $N_f > N$, exhibits a power law divergence for $N_f < N$, and diverges logarithmically for $N_f = N$. It is also free of ultraviolet divergences (which would appear at the endpoints of the α integration), and thus it behaves as guessed above.[1]

In the infrared divergent cases, the divergent part is identical to the (similarly ill-defined) instanton contribution to $\langle \mathcal{O}_1(0)\mathcal{O}_2(0)\rangle$. For $N_f < N$, the (cutoff) integral is non-singular for small x, corresponding to non-singular corrections to the coefficients of operators appearing in the OPE. For the case $N_f = N$, the expression also has a logarithmic singularity for small x, indicating the appearance of the unit operator in the OPE, with a coefficient function behaving as $\log(x)$. It is necessary to define the operators appearing in these expressions at a scale M, and this introduces a mass scale both into the matrix element and into the coefficient of the unit operator. We will discuss the scaling with M in more detail in section 6.4. Indeed, the question of whether we can compute the Green's function reliably is now a delicate one. The operators $\mathcal{O}^a_{3N_f}$ possess anomalous dimensions. As a result, the coefficient functions behave as powers

[1]Note it is important that we divided the fields into two groups as we did; otherwise, there would be ultraviolet divergences associated with the definition of the local operators.

of logarithms. We see that the coefficient of the unit operator is proportional to a single power of a logarithm. So at best we can hope that $C^{(0)}$ dominates by a power of a logarithm. We will explore this possibility in section 6.4.

6.4 The Special Case $N = N_f$

We have seen that in the case $N_f = N$, the operators with dimension $3N_f$ and the unit operator appearing in equation (6.62) in the OPE behave similarly with x, up to powers of logarithms. The best we might hope, then, is that the instanton wins (or loses) by a power of a logarithm. To assess this, we need to compute the anomalous dimensions of the various operators appearing in the OPE. This is a straightforward one loop computation. Consider, first, the case of $SU(2)$ with two flavors. Taking as the basis of dimension six operators

$$\mathcal{O}_1 = \bar{u}u\ \bar{d}d \quad \mathcal{O}_2 = \bar{u}\sigma^{\mu\nu}u\ \bar{d}\sigma_{\mu\nu}\bar{d} \tag{6.70}$$

the matrix of anomalous dimensions is:

$$\Gamma = A\begin{pmatrix} 15/2 & -2 \\ 0 & 3/2 \end{pmatrix} \tag{6.71}$$

where

$$A = \frac{g^2}{16\pi^2}\frac{2}{\epsilon}. \tag{6.72}$$

Γ has eigenvalues

$$\gamma_1 = \frac{15}{2}A \quad \gamma_2 = \frac{3}{2}A. \tag{6.73}$$

Correspondingly, given that the β function in this theory is 6, at small x, relative to the unit operator contribution, $\ln(x)$, the contributions of the first operator behaves as $(\log x)^{15/12}$, while those of the second behave as $(\log x)^{1/4}$. So it is necessary to choose Green's functions carefully so as to obtain just the contribution of the second eigenoperator.

In $SU(3)$, the matrix Γ is 3×3, and there are not simple analytic expressions for the eigenvalues, but one finds that, again, there is one operator which diverges more slowly at small x than the contribution of the unit operator. In order to isolate an instanton contribution, it is necessary to take a linear combination of operators in the Green's function, $\Delta(x)$, which projects on this. In practice, say in a lattice computation, this will be quite challenging, to say the least.

6.5 Lattice Gauge Theory in the Chiral Limit

A potential application of these ideas lies in tests of lattice gauge theory simulations. What we have just learned is that, for suitable numbers of flavors and colors, maximally chirality violating correlation functions can be computed systematically in a semiclassical approximation. Given that the study of the chiral limit of lattice gauge theories is subtle, while, at the same time, current analyses probe quite small quark masses [6], comparison of "experiment" and "theory" could be useful.

The realistic case, for which a significant amount of data exists (i.e. for which there already exists a large ensemble of gauge configurations), would be $N_f = N = 3$.

We have seen that this is the borderline situation for the semiclassical analysis; at the same time, the large N analysis is not reliable. At most, the instanton computation dominates by a small power of a logarithm. Still, it would be of interest to examine the data to see *if* the Green's function exhibits roughly logarithmic variation with x, of a suitable order of magnitude. The simplest case to study where a reliable computation is possible is the case $N = 2, N_f = 3$. While not nature, this should be accessible to detailed numerical study. Here

$$\langle \mathcal{O}_1(x)\mathcal{O}_2(0)\rangle = C\frac{\Lambda^{16/3}}{x^{11/3}}(1 + \mathcal{O}(\alpha_s(x))). \tag{6.74}$$

where the numerical coefficient, C, is presented below in a particular scheme. The question, then, is whether, with current lattice technology, such a computation is feasible; as we now explain, we believe that, while such a calculation is challenging, given the lattice spacings and quark masses currently being explored, the answer may well be yes. Here we content ourselves with some rough estimates.

In practice, one would compute the correlation function, dividing out by the vacuum to vacuum amplitude($\langle 1\rangle$). Current lattice computations achieve quite small lattice spacings and quark masses. For non-zero quark mass, perturbing in the quark masses, there are more singular contributions to the correlation function than those due to instantons. (The relevant parameters are m_q/Λ and $1/(\Lambda x)$, where Λ is the QCD scale).

First, we estimate more carefully the size of instanton effects. The needed functional determinants can be obtained from 't Hooft's original paper [31]. Here we will content ourselves with a brief summary of special cases in the $\overline{MS}$ scheme. For the case $N_f = 3, N = 2$, we obtain:

$$\Delta(x) \approx 9 \times 10^3 \mu^{16/3} g^{-8} e^{\frac{-8\pi^2}{g^2(\mu)}} x^{-11/3} \tag{6.75}$$

while the logarithmically singular term, for $N_f = 3, N = 3$ is:

$$\Delta(x) = -1.0 \times 10^7 \mu^9 g^{-12} e^{\frac{-8\pi^2}{g^2(\mu)}} \log(M\ x). \tag{6.76}$$

To estimate the effects on the lattice size, we will take the smallest lattice spacing and quark masses considered by the MILC collaboration [6]; roughly $a = (4\ \mathrm{GeV})^{-1}$, and $m_u, m_d \approx 10 - 20$ MeV. We will consider measuring the correlation function at $x = (1.5\ \mathrm{GeV})^{-1}$ (by these scales, we mean scales such that the $SU(2)$ gauge coupling has the value corresponding to the observed $SU(3)$ coupling at that scale). The leading terms are independent of x and quite sensitive to the cutoff, a; roughly they give a contribution, in the case of three flavors:

$$\Delta = \frac{m_u m_d m_s}{a^6}. \tag{6.77}$$

Using the expression for Δ, equation (6.75), this is overwhelmingly larger than the instanton contribution. This problem can be avoided by choosing operators which are odd under parity (the parity symmetry in the case of staggered fermions is discussed, for example, in [11]). In that case, the leading contribution arises from a four loop diagram. This behaves as:

$$\Delta = c \left(\frac{\alpha_s}{4\pi}\right)^2 \frac{m_u m_d m_s}{a^2 x^4}. \tag{6.78}$$

and should be much *smaller* than the instanton contribution.

One then can ask whether the instanton contribution is large enough that it is realistic to hope to extract it. We believe the answer is yes, but that the computation is likely to be challenging. A yardstick (fermistick?) for comparison is

$$\Delta' = S_F^3 \sim \left(\frac{1}{\pi^2}\right)^3 |x|^{-9}. \tag{6.79}$$

with some contraction of the indices on S_F. This is roughly comparable in size to Δ for $x^{-1} = 1.5$ GeV. So one might hope that the numerical computation will not be so noisy as to mask the instanton contribution. Similar statements apply to the $SU(3)$ computation (which, again, is subject to greater uncertainties, but is clearly of interest). Examining somewhat similar numerical computations reinforces our optimism that this is a challenging, but tractable, computation [12].

Part II

Inflation

Chapter 7

Introduction to Cosmology

7.1 Isotropy and Homogeneity

Most cosmological models are based on the principle that the universe is the same everywhere. This vague notion is based on two precise concepts: isotropy and homogeneity. Isotropy means that space looks the same in any direction. Homogeneity means the metric is the same throughout the manifold. In other words, a homogeneous space is translation invariant, and an isotropic space is rotationally invariant. Isotropy and homogeneity are not equivalent concepts. A manifold can be non-isotropic but homogeneous. Or, it can be non-homogeneous, yet isotropic at certain points. As an example, consider the manifold $\mathbf{R} \times S^2$. This space is clearly homogeneous because the metric is the same throughout. However, it is not isotropic. One direction looks different than the other two, namely, the $\mathbf{R}$ direction. This is an example of a manifold that is homogeneous, but nowhere isotropic. Now let's look at the opposite. Consider a cone.

A cone is isotropic everywhere except its apex. But, it is manifestly not homogeneous. In fact, it is a general theorem that if a manifold is everywhere isotropic, then it is also homogeneous.

7.2 The Friedmann-Robertson-Walker Metric

When cosmologists say the universe is isotropic and homogeneous, they don't mean the entire spacetime manifold, but rather, just space itself. Typically, one breaks up spacetime as $\mathbf{R} \times \Sigma$, where Σ is assumed to be isotropic and homogeneous. The metric of spacetime then becomes

$$ds^2 = -dt^2 + a^2(t)d\sigma^2, \tag{7.1}$$

where $d\sigma^2$ represents the metric on the isotropic and homogeneous Σ, and $a(t)$, known as the *scale factor* represents the relative size of the spacelike hypersurfaces Σ at fixed t. $d\sigma^2$ is tightly constrained by the assumptions of isotropy and homogeneity. This leads to the Friedmann-Robertson-Walker metric:

$$ds^2 = -dt^2 + a^2(t)\left(\frac{dr^2}{1-kr^2} + r^2 d\theta^2 + r^2\sin^2\theta d\phi^2\right). \tag{7.2}$$

In the above equation, k represents the curvature of Σ and takes values in $\{-1, 0, 1\}$ for negatively curved, no curvature, and positively curved, respectively. The coordinates used in the FRW metric are called *comoving coordinates*. That is, the comoving coordinates (r, θ, ϕ) of an object stay fixed even as the universe expands, as long as no forces are acting on that object. The physical distance is then given by

$$d_{phys} = a(t)r. \tag{7.3}$$

7.3 Dynamics

To understand the dynamics of the universe, we will need to solve Einstein's equation

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{7.4}$$

We will assume the universe is a perfect fluid, in which case

$$T_{\mu\nu} = (\rho + p)U_\mu U_\nu + p g_{\mu\nu}. \tag{7.5}$$

In the above equation, U_μ represents the four-velocity of the fluid, ρ the energy density, and p the pressure. It is simplest to analyze the physics in a comoving frame, in which case the four-velocity becomes simply

$$U_\mu = (1, 0, 0, 0). \tag{7.6}$$

The energy-momentum tensor simplifies to

$$T^\mu_\nu = \mathrm{diag}(-\rho, p, p, p). \tag{7.7}$$

To solve Einstein's equation (7.4), it is sometimes useful to write it in a different form. If we take the trace of (7.4), we get

$$R_{\mu\nu} = 8\pi G \left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T \right). \tag{7.8}$$

The trace of the energy-momentum tensor (7.7) is

$$T = -\rho + 3p. \tag{7.9}$$

To solve Einstein's equation, we need to know $R_{\mu\nu}$. To this end, we first find the Christoffel symbols:

$$\Gamma^0_{11} = \frac{a\dot{a}}{1-kr^2}, \tag{7.10}$$

$$\Gamma^0_{22} = a\dot{a}r^2, \tag{7.11}$$

$$\Gamma^1_{01} = \Gamma^2_{02}, \tag{7.12}$$

$$\Gamma^1_{22} = -r(1-kr^2), \tag{7.13}$$

$$\Gamma^2_{12} = \Gamma^3_{13} = \frac{1}{r}, \tag{7.14}$$

$$\Gamma^2_{33} = -\sin\theta\cos\theta, \tag{7.15}$$

$$\Gamma^1_{11} = \frac{kr}{1-kr^2}, \tag{7.16}$$

$$\Gamma^0_{33} = a\dot{a}r^2\sin^2\theta, \tag{7.17}$$

$$\Gamma^3_{03} = \frac{\dot{a}}{a}, \tag{7.18}$$

$$\Gamma^1_{33} = -r(1-kr^2)\sin^2\theta, \tag{7.19}$$

$$\Gamma^3_{23} = \cot\theta. \tag{7.20}$$

The non-zero components of $R_{\mu\nu}$ are therefore

$$\begin{aligned} R_{00} &= -3\frac{\ddot{a}}{a}, & (7.21) \\ R_{11} &= \frac{a\ddot{a} + 2\dot{a}^2 + 2k}{1 - kr^2}, & (7.22) \\ R_{22} &= r^2(a\ddot{a} + 2\dot{a}^2 + 2k), & (7.23) \\ R_{33} &= r^2(a\ddot{a} + 2\dot{a}^2 + 2k)\sin^2\theta. & (7.24) \end{aligned}$$

We are now in a position to solve Einstein's equation. Looking at the 00 component of equation (7.8), we get

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p). \qquad (7.25)$$

Taking the ij component and doing some simplifications gives

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}. \qquad (7.26)$$

These two equations are known as the *Friedmann equations.* Next, we define the *Hubble constant* H, as

$$H \equiv \frac{\dot{a}}{a} = \frac{\partial \ln a}{\partial t}. \qquad (7.27)$$

The value of the Hubble constant at present times is denoted H_0 and is roughly equal to

$$H_0 = 70 \pm 10 \text{ km/sec/Mpc}. \qquad (7.28)$$

H_0 is often written in the form

$$H_0 = 100h \text{ km/sec/Mpc}, \qquad (7.29)$$

and h is taken to be 0.7. Typical cosmological scales are set by the *Hubble length*

$$d_H = \frac{c}{H_0} \approx 3 \times 10^3 h^{-1} \text{ Mpc}. \tag{7.30}$$

Typical cosmological times are set by the *Hubble time*

$$t_H = \frac{1}{H_0} \approx 9.78 \times 10^9 h^{-1} \text{ yr}. \tag{7.31}$$

Using the Hubble constant and setting $8\pi G = 1$, we can rewrite the Friedmann equations:

$$H^2 = \frac{1}{3}\rho - \frac{k}{a^2}, \tag{7.32}$$

$$\dot{H} + H^2 = -\frac{1}{6}(\rho + 3p). \tag{7.33}$$

Combining these two equations gives the continuity equation

$$\frac{d\rho}{dt} + 3H(\rho + p) = 0. \tag{7.34}$$

We now define the equation of state parameter, w,

$$w \equiv \frac{p}{\rho}. \tag{7.35}$$

In terms of w, the continuity equation becomes

$$\frac{\dot{\rho}}{\rho} = -3(1 + w)\frac{\dot{a}}{a} \tag{7.36}$$

which can be integrated to obtain

$$\rho \propto a^{-3(1+w)}. \tag{7.37}$$

Using this value of ρ in the Friedmann equation (7.32), we can solve for a.

$$a^{\frac{1}{2}(1+3w)} da \propto dt. \tag{7.38}$$

Performing the integration, the scale factor becomes

$$a(t) \propto \begin{cases} t^{\frac{2}{3(1+w)}} & : \quad w \neq -1 \\ e^{Ht} & : \quad w = -1 \end{cases} . \tag{7.39}$$

For a matter dominated Universe, $w = 0$, for radiation dominated, $w = \frac{1}{3}$, and for a Universe dominated by a cosmological constant, $w = -1$. These lead to scale factors of $a \propto t^{2/3}$, $a \propto t^{1/2}$, and $a \propto e^{Ht}$, respectively. So, we see that in order to get exponential expansion, we require $w = -1$, or $\rho = -p$.

Next, we define the critical density

$$\rho_{crit} \equiv 3H_0^2 \tag{7.40}$$

where the subscript 0 means evaluated at the present time. For each species i of matter (baryons, photons, dark matter, dark energy, etc.) we define

$$\Omega_i = \frac{\rho_0^i}{\rho_{crit}}. \tag{7.41}$$

Using equation (7.37), we get that

$$\rho_i = \rho_{crit}\Omega_i \left(\frac{a_0}{a}\right)^{3(1+w_i)}. \tag{7.42}$$

The Friedmann equation (7.32) now becomes

$$\left(\frac{H}{H_0}\right)^2 = \sum_i \Omega_i \left(\frac{a_0}{a}\right)^{3(1+w_i)} \tag{7.43}$$

where $\Omega_k \equiv -\frac{k}{a_0^2 H_0^2}$ and $w_k = -\frac{1}{3}$. Note that with these definitions,

$$\sum_i \Omega_i = 1. \tag{7.44}$$

Let Ω_M, Ω_R, Ω_k, and Ω_Λ denote the densities of matter, radiation, curvature, and cosmological constant, respectively. Current observations place $\Omega_M \approx 0.28$, $\Omega_R \approx 0$, $\Omega_k \approx 0$, and $\Omega_\Lambda \approx 0.72$.

7.4 The Age of the Universe

The age of the universe can be estimated from the Friedmann equation (7.43) and the estimates of Ω_i. Let $x \equiv a/a_0$. Then,

$$\left(\frac{\dot{x}}{x}\right)^2 = H_0^2 \left(\Omega_M x^{-3} + \Omega_\Lambda\right) \tag{7.45}$$

so that

$$t_0 = \frac{1}{H_0}\int_0^1 \frac{dx}{x\sqrt{\Omega_M x^{-3} + \Omega_\Lambda}} \tag{7.46}$$

$$\approx 0.982778 H_0^{-1} \tag{7.47}$$

$$\approx 13.7 \times 10^9 \text{yr}. \tag{7.48}$$

7.5 Horizons

There are two notions of horizons in cosmology: a *particle horizon* and an *event horizon.* Both the particle horizon and the event horizon are observer dependent. The particle horizon encompasses the observable universe and is the maximum distance particles could have traveled to the observer in the age of the universe. An event horizon is the largest distance from which light emitted now can ever reach the observer. We calculated the age of the universe to be approximately 13.7 billion years. Naively, one might conclude that the particle horizon is therefore 13.7 billion light years. However, this would only be true in a static, non-expanding universe. In an expanding universe, the particle horizon is much further.

Let us now calculate the physical distance to our particle horizon. Recall from equation (7.3), that the physical distance to an event is

$$d_{phys} = a(t)r \tag{7.49}$$

where r is the comoving distance. Therefore, in order to calculate the physical distance to our particle horizon, we first need an expression for the comoving distance. To that end, consider a photon moving on a radial trajectory. The photon satisfies

$$ds^2 = 0 \tag{7.50}$$

so that

$$-dt^2 + a^2 dr^2 = 0. \tag{7.51}$$

In integral form, this translates to

$$\Delta r = \int_{t_1}^{t_2} \frac{dt}{a(t)}. \tag{7.52}$$

Therefore, the distance to the particle horizon as a function of t is given by

$$d_p(t) = a(t) \int_0^t \frac{dt'}{a(t')}. \tag{7.53}$$

Current estimates place the particle horizon at a distance of approximately 47 billion light years.

As explained earlier, the cosmic event horizon is the largest distance from which light emitted now could ever reach the observer in the future. If we let t_0 represent now, the physical distance to our event horizon is therefore

$$d_e = \int_{t_0}^{t_{max}} \frac{dt'}{a(t')} \tag{7.54}$$

where t_{max} represents the end of the universe.

Chapter 8

Basics of Inflation

8.1 Motivation for Inflation

There are two puzzles of the Big Bang which point towards a more comprehensive theory. They are known as the flatness problem and the horizon problem.

Let us first describe the flatness problem. Consider the Friedmann equation in a universe without vacuum energy

$$H^2 = \frac{1}{3}\left(\rho_M + \rho_R\right) - \frac{k}{a^2}. \tag{8.1}$$

Recall now that, $\rho_M \sim a^{-3}$ and $\rho_R \sim a^{-4}$. Since the Plank epoch, a has increased by as much as 10^{30}. Therefore, one would expect the $\frac{k}{a^2}$ term to dominate. However, current estimates place the ratio $\frac{k/a^2}{\rho/3} \sim 1$.

The horizon problem is the fact that, to a high degree of precision, the CMB is isotropic. This is problematic because the CMB at last-scattering consisted of causally disconnected regions of space. We have showed that the comoving particle horizon is given by

$$\begin{aligned} r_p &= \int_0^t \frac{dt'}{a(t')} && (8.2) \\ &= \int_0^a \frac{da}{a^2 H} && (8.3) \\ &= \int_0^a d\ln a \left(\frac{1}{aH}\right) && (8.4) \\ &= \frac{1}{H_0}\int_0^a a^{\frac{1}{2}(1+w)} d\ln a. && (8.5) \end{aligned}$$

During Big Bang expansion, $w \geq 0$ and

$$r_p \propto a^{\frac{1}{2}(1+3w)}. \qquad (8.6)$$

Therefore, the comoving particle horizon distance is a monotonically increasing function. This implies that comoving scales entering the horizon today, were causally disconnected at the time of last scattering. Yet, they are observed to be at the same temperature, to a high degree of precision. This is the horizon problem.

8.2 Necessary Conditions for Inflation

To solve the horizon problem, we need all points of the CMB at last scattering to be in causal contact. However, this is inconsistent with a monotonically increasing comoving horizon, as is the case for non-inflationary Big Bang cosmology. Therefore, we must posit the existence of an epoch in the early universe in which the comoving horizon decreased. From equation (8.6), we see that this requires the *comoving Hubble radius*, $(aH)^{-1}$, to have decreased at some point in the early universe. Moreover, a decreasing comoving Hubble radius implies $\ddot{a} > 0$:

$$\begin{aligned} & \frac{d}{dt}(aH)^{-1} < 0 && (8.7) \\ \Rightarrow\quad & \frac{d}{dt}(\dot{a})^{-1} < 0 && (8.8) \\ \Rightarrow\quad & \frac{\ddot{a}}{\dot{a}^2} > 0 && (8.9) \\ \Rightarrow\quad & \ddot{a} > 0. && (8.10) \end{aligned}$$

The last line comes about since the universe is expanding, so $\dot{a} > 0$. This tells us that a necessary condition to solve the horizon problem is for the universe to have underwent a period of accelerated expansion in its early history. One can now ask, what stress-energy source can give rise to accelerated expansion? To answer this, we must look at the Friedmann equation

$$\frac{\ddot{a}}{a} = -\frac{1}{6}(\rho + 3p). \tag{8.11}$$

If $\ddot{a} > 0$, then $\rho + 3p < 0$. Necessary, and equivalent, conditions for inflation are therefore:

- A decreasing comoving horizon: $\frac{d}{dt}(aH)^{-1} < 0$.
- Accelerated expansion: $\ddot{a} > 0$.
- Negative pressure: $p < -\frac{1}{3}\rho$.
- $w < -\frac{1}{3}$.

The next question to ask is whether these conditions simultaneously solve the flatness problem. The answer is yes. In order for accelerated expansion to occur, the universe must be dominated by something other than radiation or matter. For simplicity, let's assume the accelerated expansion is due to vacuum energy. In this case, $\rho/3 \propto a^0$. Therefore, as a increases, ρ gets larger and larger compared to k/a^2. In other words, the universe gets flatter and Ω is drive to unity.

8.3 The Physics of Inflation

The most straightforward way of obtaining accelerated expansion is by positing the existence of a scalar field, known as the *inflaton*. The vacuum energy of the inflaton will be what drives inflation. Furthermore, we will assume that ϕ is homogeneous so that

$$\phi(t, \mathbf{x}_i) = \phi(t) \tag{8.12}$$

and

$$\partial_i \phi = 0. \tag{8.13}$$

The lagrangian of such a scalar field is given by

$$\begin{aligned} \mathcal{L} &= \frac{1}{2}\partial^\mu \phi \partial_\mu \phi - V(\phi) && (8.14) \\ &= \frac{1}{2}\dot{\phi}^2 - V(\phi). && (8.15) \end{aligned}$$

The energy momentum tensor is

$$T^{\mu\nu} = \partial^\mu \phi \partial^\nu \phi - \mathcal{L} g^{\mu\nu}. \tag{8.16}$$

Recall that for a perfect fluid,

$$T^\mu_\nu = \mathrm{diag}(-\rho, p, p, p) \tag{8.17}$$

which implies that

$$\rho = \frac{1}{2}\dot{\phi}^2 + V(\phi) \tag{8.18}$$

$$p = \frac{1}{2}\dot{\phi}^2 - V(\phi). \tag{8.19}$$

Note that in order to get $w < -1/3$ as required for inflation, we must have $V(\phi) > \dot{\phi}^2$. In other words, *the potential energy must dominate over the kinetic energy*. Next, we vary the action

$$S = \int d^4x \sqrt{-g}\mathcal{L} \tag{8.20}$$

to get the equation of motion:

$$\ddot{\phi} + 3H\dot{\phi} + \frac{\partial V}{\partial \phi} = 0. \tag{8.21}$$

This is the same equation that describes a ball rolling down a hill with the Hubble constant acting as a friction term. Since we require the potential energy to dominate over the kinetic energy, the ball must roll slowly. This necessitates a sufficiently shallow potential so that the friction term dominates over $\ddot{\phi}$.

8.4 Slow Roll Inflation

Let us now define what are known as the *slow roll parameters*:

$$\epsilon \equiv \frac{M_p^2}{2}\left(\frac{V'}{V}\right)^2 \tag{8.22}$$

$$\eta \equiv M_p^2 \frac{V''}{V}. \tag{8.23}$$

Satisfying the conditions

$$\begin{aligned} \dot{\phi}^2 &\ll V(\phi) && (8.24) \\ |\ddot{\phi}| &\ll |3H\dot{\phi}| && (8.25) \\ |\ddot{\phi}| &\ll \left|\frac{\partial V}{\partial \phi}\right| && (8.26) \end{aligned}$$

requires the smallness of the slow roll parameters:

$$\begin{aligned} \epsilon &\ll 1 && (8.27) \\ |\eta| &\ll 1. && (8.28) \end{aligned}$$

This can be checked for self-consistency fairly easily. In the slow roll regime, the Friedmann equation and the equation of motion reduce to

$$\begin{aligned} H^2 &\approx \frac{1}{3}V(\phi) \approx \text{constant} && (8.29) \\ \dot{\phi} &\approx -\frac{V'}{3H} && (8.30) \end{aligned}$$

Therefore,

$$\begin{aligned} \dot{\phi}^2 &\ll V(\phi) && (8.31) \\ \Rightarrow &\ \frac{\dot{\phi}^2}{V} \ll 1 && (8.32) \\ \Rightarrow &\ \left(\frac{V'}{V}\right)^2 \ll 1. && (8.33) \end{aligned}$$

Self-consistency of the η condition can be checked in a similar fashion. Violation of the slow roll conditions signals the end of inflation. During inflation, the universe undergoes exponential expansion

$$a(t) = e^{Ht}. \qquad (8.34)$$

One can then calculate the number of e-foldings, N, the universe undergoes:

$$\begin{aligned} N(\phi) &\equiv \ln \frac{a_{end}}{a} && (8.35) \\ &= \int_t^{t_{end}} H dt && (8.36) \\ &\approx \frac{1}{M_p^2} \int_{\phi_{end}}^{\phi} \frac{V}{V'} d\phi && (8.37) \\ &= \frac{1}{M_p^2} \int_{\phi_{end}}^{\phi} \frac{1}{\sqrt{2\epsilon}} d\phi && (8.38) \end{aligned}$$

8.5 An Example

As an example, consider the potential

$$V(\phi) = \frac{1}{2} m^2 \phi^2. \qquad (8.39)$$

Calculation of the slow roll parameters yields

$$\begin{aligned} \epsilon &= \frac{M_p^2}{2} \left(\frac{V'}{V} \right)^2 && (8.40) \\ &= \frac{2M_p^2}{\phi^2} && (8.41) \end{aligned}$$

and

$$\begin{aligned} \eta &= M_p^2 \frac{v''}{V} && (8.42) \\ &= \frac{2M_p^2}{\phi^2}. && (8.43) \end{aligned}$$

Inflation ends when $\epsilon(\phi_{end}) = \eta(\phi_{end}) = 1$, which implies

$$\phi_{end} = \sqrt{2} M_p. \qquad (8.44)$$

The number of e-folds is

$$N(\phi) = \frac{1}{M_p^2}\int_{\phi_{end}}^{\phi}\frac{V}{V'}d\phi \tag{8.45}$$

$$= \frac{1}{2M_p^2}\int_{\phi_{end}}^{\phi}\phi d\phi \tag{8.46}$$

$$= \frac{\phi^2}{4M_p^2} - \frac{1}{2}. \tag{8.47}$$

Let ϕ_{CMB} denote the value of ϕ when the CMB fluctuations are created, and let N_{CMB} denote the number of e-foldings from this point until the end of inflation. From the above equation, we see that

$$\phi_{CMB} \approx 2\sqrt{N_{CMB}}M_p. \tag{8.48}$$

The inflaton has moved

$$\Delta\phi \equiv \phi_{CMB} - \phi_{end} \tag{8.49}$$

$$= (2\sqrt{N_{CMB}} - \sqrt{2})M_p > M_p. \tag{8.50}$$

When $\Delta\phi > M_p$, it is called *large field inflation* and when $\Delta\phi < M_p$ it is called *small field inflation.*

8.6 Connecting Inflation to Observation

Knowing the number of e-folds necessary for inflation is important for model building. We can estimate it as follows. Let the superscripts i, f, r, and 0 denote the beginning of inflation, the end of inflation, the beginning of the radiation dominated era, and today, respectively. We will assume that $\Omega_k^{(i)}$ was of order unity at the beginning of inflation. Recall,

$$\Omega_k = \frac{k}{a^2H^2}. \tag{8.51}$$

Suppose $a(t)$ increased by a factor e^N during inflation. Then,

$$\Omega_k^{(f)} = \frac{k}{a_f^2H_f^2} \tag{8.52}$$

$$\sim e^{-2N}\Omega_k^{(i)} \tag{8.53}$$

$$\sim e^{-2N}. \tag{8.54}$$

The curvature today is given by

$$\Omega_k^{(0)} = \frac{k}{a_0^2H_0^2} \tag{8.55}$$

$$= \Omega_k^{(f)}\left(\frac{a_fH_f}{a_0H_0}\right)^2 \tag{8.56}$$

$$= e^{-2N}\left(\frac{a_fH_f}{a_0H_0}\right)^2. \tag{8.57}$$

To solve the flatness problem, we must have

$$e^N > \frac{a_fH_f}{a_0H_0}. \tag{8.58}$$

We are now going to make the assumption that the scale factor and expansion rate stay roughly constant between the end of inflation and the beginning of the radiation dominated era. In other words, we are assumming

$$a_fH_f \approx a_rH_r. \tag{8.59}$$

Next, we define a_{eq} and H_{eq}, the scale factor and expansion rate at matter-radiation equality:

$$a_{eq} = \frac{\Omega_R}{\Omega_M} a_0 \tag{8.60}$$

$$H_{eq} = H_0 \sqrt{2\Omega_M} \left(\frac{a_0}{a_{eq}} \right)^{3/2}. \tag{8.61}$$

Over the whole radiation-matter era, the expansion rate can be expressed as

$$H = \frac{H_{eq}}{\sqrt{2}} \sqrt{\left(\frac{a_{eq}}{a}\right)^3 + \left(\frac{a_{eq}}{a}\right)^4}. \tag{8.62}$$

We can use the approximation $a \approx a_r$ so that $a \ll a_{eq}$ and

$$H_r \approx \frac{H_{eq}}{\sqrt{2}} \left(\frac{a_{eq}}{a_r} \right)^2. \tag{8.63}$$

This gives

$$e^N > \Omega_R^{1/4} \sqrt{\frac{H_r}{H_0}} \tag{8.64}$$

$$= \left(\Omega_R \frac{\rho_r}{\rho_{0,crit}} \right)^{1/4} \tag{8.65}$$

$$\approx 8 \times 10^{26}. \tag{8.66}$$

This gives a lower bound on the number of e-folds of roughly

$$N_{CMB} > 62. \tag{8.67}$$

We have been assuming that the universe is homogeneous and isotropic. Of course, this is just an approximation. Among the many exciting features of inflation, is a prediction of density fluctuations and the formation of structure. The details are beyond the scope of this dissertation, but the basic idea is as follows. We consider a perturbation of the metric away from the isotropic and homogeneous FRW metric. We must also therefore consider a perturbation to the energy tensor. This induces a perturbation, $\delta\rho$, in the energy density. We then expand this in Fourier modes:

$$\delta\rho = F(t) \int d^3q \alpha(\mathbf{q}) \Delta(q) e^{i\mathbf{q}\cdot\mathbf{x}}, \tag{8.68}$$

where

$$\Delta(q) \sim q^{n_s - 1}. \tag{8.69}$$

n_s is known as the *spectral index.* It can also be shown that

$$\frac{\delta\rho}{\rho} \sim \frac{V^{3/2}}{V' M_p^3}. \tag{8.70}$$

Current observations place

$$\frac{V^{3/2}}{V' M_p^3} \approx 5.15 \times 10^{-4} \tag{8.71}$$

$$n_s \approx 0.96. \tag{8.72}$$

Any successful model of inflation needs to roughly predict these values.

8.7 Hybrid Inflation

We now turn our attention to a specific model of inflation, known as hybrid inflation [15, 16, 19, 20]. Hybrid inflation involves two scalar fields, ϕ and σ. The σ field will denote the inflaton, and the ϕ field is termed the waterfall field. The effective potential is given by

$$V(\sigma, \phi) = \left(\kappa\phi^2 - \mu^2\right)^2 + \frac{1}{2}m^2\sigma^2 + \kappa^2\sigma^2\phi^2. \tag{8.73}$$

We wish to find the absolute minimum of V, so we look at the partial derivatives:

$$\frac{\partial V}{\partial \sigma} = m^2\sigma + 2\kappa^2\sigma\phi^2 \tag{8.74}$$

$$\frac{\partial V}{\partial \phi} = 4\kappa\phi^3 - 4\mu^2\phi + 2\kappa^2\sigma^2\phi. \tag{8.75}$$

This yields 3 extrema for V:

$$(\sigma, \phi) = (0,0), \ \left(0, \pm\frac{\mu}{\sqrt{\kappa}}\right). \tag{8.76}$$

However, the absolute minimum of V is

$$(\sigma_0, \phi_0) = \left(0, \pm\frac{\mu}{\sqrt{\kappa}}\right). \tag{8.77}$$

Now, let's fix σ and ask what value of ϕ minimizes V. Once again, we look at the partial derivative

$$\frac{\partial V}{\partial \phi} = 4\kappa\phi^3 - 4\mu^2\phi + 2\kappa^2\sigma^2\phi = 0. \tag{8.78}$$

Solving for ϕ gives

$$\phi^2_{min} = \begin{cases} 0 & , \quad \text{if } \sigma^2 > \frac{2\mu^2}{\kappa} \\ \frac{2\mu^2 - \kappa\sigma^2}{2\kappa} & , \quad \text{if } \sigma^2 \leq \frac{2\mu^2}{\kappa} \end{cases}. \tag{8.79}$$

Let us define the critical point σ_c by

$$\sigma_c^2 \equiv \frac{2\mu^2}{\kappa}. \tag{8.80}$$

Then we see that for $\sigma > \sigma_c$, the minimum of ϕ is 0. Now, put $\sigma = x\sigma_c$ so that the potential becomes

$$V(\phi) = \kappa^2\phi^4 + 2\kappa\mu^2(x^2-1)\phi^2 + \mu^4 + \frac{x^2m^2\mu^2}{\kappa}. \tag{8.81}$$

The global properties of $V(\phi)$ depend critically on the value of x. For $x > 1$, V has one a global minimum at $\phi = 0$ and there are no other extrema. However, for $x < 1$, $\phi = 0$ becomes a local maximum and there are now two degenerate global minima at

$$\phi_{min} = \pm\sigma_c\sqrt{\frac{1-x^2}{2}} \tag{8.82}$$
$$= \pm\sqrt{\frac{\mu^2(1-x^2)}{\kappa}}. \tag{8.83}$$

The hybrid inflation model posits that the initial value of the inflaton field was much greater than σ_c. Moreover, the curvature of V in the ϕ direction is greater than the σ direction. Thus, during the early universe, the waterfall field ϕ quickly rolled to its minimum $\phi_{min} = 0$ and σ was able to remain large for a much longer time. The inflationary epoch takes place in this regime where σ is large and $\phi \approx 0$. We will assume that $m^2 \ll H^2$. Define a parameter

$$T \equiv \frac{m^2 M_p^2}{\mu^4}. \tag{8.84}$$

From the inflationary equation of motion

$$H^2 \approx \frac{1}{3}\frac{V}{M_p^2}, \tag{8.85}$$

we conclude

$$T \ll 1. \tag{8.86}$$

To verify that conditions are right for inflation, let us calculate the slow roll parameters ϵ and η in this regime.

$$
\begin{aligned}
\epsilon &= \frac{M_p^2}{2}\left(\frac{V'}{V}\right)^2 && (8.87)\\
&= \frac{M_p^2}{2}\frac{m^4\sigma^2}{\mu^8} && (8.88)\\
&< T^2 && (8.89)\\
&\ll 1 && (8.90)\\
\eta &= M_p^2\left(\frac{V''}{V}\right) && (8.91)\\
&= T && (8.92)\\
&\ll 1. && (8.93)
\end{aligned}
$$

Let's now analyze what happens to the fields σ and ϕ once $\sigma = \sigma_c$. From the inflation equation of motion we get

$$\Delta\sigma = \frac{m^2}{3H}\sigma\Delta t. \tag{8.94}$$

Moreover,

$$H^2 \approx \frac{V}{3M_p^2}. \tag{8.95}$$

Therefore, in a time $\Delta t = H^{-1}$ the field σ decreases from σ_c by

$$\Delta\sigma = \frac{m^2M_p^2}{\mu^4}\sigma_c. \tag{8.96}$$

The value of σ after the time interval Δt is therefore

$$\sigma = \left(1 - \frac{m^2M_p^2}{\mu^4}\right)\sigma_c. \tag{8.97}$$

Once σ passes σ_c, ϕ no longer remains at 0 and begins to roll to the absolute minimum. We can then calculate its slow roll parameter, η_ϕ, and see how small it is.

$$\begin{aligned}
\eta_\phi &= M_p^2 \frac{V''}{V} && (8.98) \\
&\approx \frac{M_p^2}{\mu^4}\left(12\kappa^2\phi^2 - 4\kappa\mu^2 + 2\kappa^2\sigma^2\right) && (8.99) \\
&\approx \frac{M_p^2}{\mu^4}\left(-4\kappa\mu^2 + 2\kappa^2\sigma^2\right) && (8.100) \\
&\approx \frac{M_p^2}{\mu^4}\left(-4\kappa\mu^2 + 2\kappa^2\left(1 - \frac{m^2 M_p^2}{\mu^4}\right)^2 \sigma_c^2\right) && (8.101) \\
&\approx -\frac{8\kappa m^2 M_p^4}{\mu^6} && (8.102)
\end{aligned}$$

If $\mu^3 \ll \kappa m M_p^2$, then $\eta \gg 1$ and the slow roll conditions are violated. Thus, ϕ rapidly rolls to its minimum and inflation ends. Motion in the σ direction becomes very fast as well, and σ rapidly approaches its minimum value of 0. Thus, in the given hybrid inflation model, once σ reaches the σ_c threshold, inflation ends almost instantaneously.

Chapter 9

Studies in Small Field Inflation

9.1 Introduction

There is good evidence that the universe underwent a period of inflation early in its history. Yet it is probably fair to say that there do not exist completely reliable, calculable microscopic theories of inflation. Slow roll inflation provides a simple phenomenology; many, if not most, microscopic scenarios for inflation, involving branes, extra dimensions and the like, admit such a description. Indeed, slow roll inflation makes clear why it is nearly impossible, at present, to formulate a compelling microscopic theory. Planck scale effects are *necessarily* important, and this requires a full understanding of issues like dynamics of moduli and supersymmetry breaking, even within a consistent theory of gravity (i.e. a string model). Even as a phenomenology, there are a number of models for slow roll inflation. We follow the review of Lyth [22] in dividing these into "large field", "medium field" and "small field" types, where large, medium or small here refers to field variations much larger than, comparable, or much smaller than the Planck mass. Almost by definition, large or medium field inflation is difficult to describe in a systematic fashion, without a complete theory of quantum gravity. Small field inflation, however, is another matter. Here one should be able to characterize inflation in terms of a low energy effective action for some number of light fields, with a limited set of relevant parameters.

Our goal in this section is to characterize inflation in terms of a low energy effective action for some number of light fields with a limited set of relevant parameters. limited set of relevant parameters.[1,2] We will see that with some very mild assumptions about genericity, we can characterize small field inflation quite simply:

1. The effective theory should exhibit an approximate (global) supersymmetry.
2. The effective theory should obey a discrete R symmetry.
3. Supersymmetry is spontaneously broken in the effective theory.
4. The (approximate) goldstino may or may not lie in a multiplet with the inflaton.
5. The effective theory exhibits an approximate, continuous R symmetry.
6. Terms allowed by such a symmetry break the continuous global symmetry and spoil inflation, unless the inflationary scale (the square of the Goldstino decay constant) is sufficiently small.

[1]Expansions of this type have been considered by various authors; an early discussion appears in [23].

[2]Here we are not using "relevant" in the conventional renormalization group sense; but instead referring to their relevance to inflation; the correspondence to the usual terminology will be clear shortly.

7. If the requirement above is satisfied, there are further requirements on the Kahler potential in order to obtain slow roll inflation with adequate e-foldings. This sets an *irreducible* minimum amount of fine tuning necessary to achieve acceptable inflation. This tuning grows in severity with the number of Hubble mass fields.

8. In order that inflation ends, the inflaton must couple to other light degrees of freedom, or must have appreciable self-couplings in the final ground state. The coupling to this extra field, or the self couplings, are fixed by $\frac{\delta\rho}{\rho}$ and the inflationary scale. In the case of an extra field, the resulting structure is necessarily what is called "hybrid inflation" [8, 9, 20, 21, 23]. In the latter, which we will call "R breaking inflation" (RBI), further fine tuning is required. In either case, the spectral index is less than one.

Many of these points have been made before, but perhaps not in the systematic fashion discussed here; in particular, the inevitability of this structure for small field inflation does not seem to be appreciated. Similarly, most models of hybrid inflation invoke supersymmetry and R symmetries [5, 8–10, 21, 23, 28, 30]. However, the approximate supersymmetry of the effective action has not been stressed; more important, the R symmetries have generally been taken to be continuous, and the consequences of discreteness, particularly the upper bound on the scale of inflation, have not been considered before, to our knowledge. The restrictions on the Kahler potential have been noted in early work [8, 21], but then seem frequently ignored; their role in determining the irreducible level of fine tuning and the possible number of e-foldings, particularly sharp in light of the observations about R symmetry, does not seem to have been appreciated.

Given the role of supersymmetry in small field inflation, it is natural to investigate embedding this structure into theories of low energy breaking. This has the potential to expose connections between scales of inflation and scales of low energy physics, as we discuss. Ideas concerning metastable, dynamical supersymmetry breaking raise the prospect of sharpening these connections further.

In the following sections, we elaborate on each of these points. In section 9.2, we explain the requirement for supersymmetry and the structure of the effective action. In sections 9.3,9.4, we relate the inflationary observables to the parameters of the effective action. In section 9.6, we obtain an upper bound on the scale of inflation from interactions which violate the continuous R symmetry. In section 9.7, we consider possible alternative structures which might result from relaxing our assumptions; non-hybrid models can arise if we admit some additional fine tuning. In our concluding section, we briefly consider large field inflation and provide an assessment.

9.2 Inflation as Spontaneous Supersymmetry Breaking

If we impose some fine tuning or genericity constraints, we can characterize small field inflation quite simply. First we require a field which is light compared to the Hubble constant during inflation, H_0. Scalar fields with mass *of order* H_0 are already unnatural, unless the theory, at least at these scales, is supersymmetric[3]. As we will review, scalar fields with mass much smaller than H_0, even with supersymmetry, are unnatural. So having one field light compared to H_0 represents a fine tuning; it would seem unlikely that there is more than one such field. For $H_0 \ll M_p$, we can write a supergravity effective action in an approximately flat space. More precisely, given our assumption that the inflaton is small compared to M_p, the system is described, approximately, by a globally supersymmetric effective action which exhibits *spontaneous* supersymmetry breaking. This may seem obvious, but it is perhaps worth pointing out that it follows from the assumption of small field excursions. Consider, for simplicity, a single light field, S. For fields small compared to M_p, we can write:

$$K \approx K_0 + S^\dagger S + \alpha \frac{(S^\dagger S)^2}{M_p^2} + \ldots \tag{9.1}$$

Similarly, we can expand W in a power series in S.

$$W = W_0 + \mu^2 S + \frac{m}{2} S^2 + \frac{\lambda}{3} S^3 + \ldots . \tag{9.2}$$

Because of the small field assumption, W_0 cannot dominate during inflation, so

$$W_0 < H_0 M_p^2. \tag{9.3}$$

Then $\mu^2 \sim H_0 M_p$ so that

$$W_0 \ll \mu^2 M_p. \tag{9.4}$$

We next examine the slow roll parameter η. First, however, recall that

$$V = \left| \frac{\partial W}{\partial S} \right|^2 . \tag{9.5}$$

[3]Ordinary dynamical symmetry breaking is problematic, since one needs a vast mismatch between the associated energy scale and the mass scale of the excitations; Goldstone excitations, as in "natural" inflation, require decay constants much greater than the Planck scale.

Therefore, modulo order 1 constants, η contains the terms

$$\eta \ni \frac{m^2 M_p^2}{\mu^4}, \ \frac{M_p^2 \lambda^4 S^6}{\mu^8}. \tag{9.6}$$

Slow roll then implies

$$m \ll \mu \left(\frac{\mu}{M_p} \right); \quad \lambda \ll \frac{\mu^2}{M_p^2}. \tag{9.7}$$

There are also constraints on the Kahler potential parameter α which we will discuss shortly. Formulated in this way, the fermionic component of S is the Goldstino, and its scalar component is the inflaton. The possibility that there is another chiral field, whose scalar component is the inflaton will be considered further when we discuss the Kahler potential constraints.

The absence (smallness) of terms W_0, S^2, S^3 is most readily accounted for if the theory possesses an R symmetry. This can be understood as a version of the Nelson-Seiberg theorem [24], which requires an (approximate) continuous R symmetry for (meta-)stable supersymmetry breaking. Because we do not expect continuous global symmetries, we will assume that the underlying R symmetry is discrete (and we will, in general, take it to be Z_N, $N > 2$), while the continuous symmetry is an accident. In a moment, we will see that there are further reasons that an R symmetry seems necessary for successful inflation.

At the end of inflation, supersymmetry must be restored, and the cosmological constant vanish. One might try to model this without adding additional degrees of freedom. Once the R symmetry is understood as a discrete symmetry, one expects higher order terms in the S superpotential, and supersymmetric vacua at large fields.

We will see that understanding inflation in terms of flow towards such a minimum is possible, but adds additional complications. So we will first add additional degrees of freedom coupled to S. For the moment we will suppose that there is one such field, ϕ. Any additional light fields coupled to S are likely to be quite light; we will discuss this issue later. We are lead, then, to write

$$W = S(\kappa\phi^2 - \mu^2) + \text{ non} - \text{renormalizable terms.} \tag{9.8}$$

The supersymmetric minimum occurs where

$$\frac{\partial W}{\partial S} = 0 \tag{9.9}$$

$$\frac{\partial W}{\partial \phi} = 0. \tag{9.10}$$

Therefore, the system has a supersymmetric minimum at $\kappa\phi^2 = \mu^2,\ S = 0$. Classically, however, it has a moduli space with

$$|\kappa S|^2 > |\kappa\mu^2|. \tag{9.11}$$

This pseudomoduli space will be lifted both by radiative corrections in κ, quantum and Planck-suppressed terms in the Kahler potential, and the non-renormalizable terms in the superpotential. As we will see, all are necessarily relevant if inflation occurs in the model. Inflation takes place on this pseudomoduli space; ϕ is effectively pinned at zero during inflation.

The R symmetry now explains the absence of additional dangerous couplings, such as ϕ^2. Perhaps most strikingly, though, it forbids a constant in the superpotential, guaranteeing that at the end of inflation, when supersymmetry is (nearly) restored, the vacuum energy (nearly) vanishes. We will explore, in section 9.7, non-hybrid models in which the superpotential must be suitably tuned.

Three points should be noted:

1. The assumption that the symmetry is discrete means couplings like $\frac{S^{N+1}}{M_p^{N-2}}$ are permitted, and, as we will soon see, they significantly constraint inflation.

2. There are additional conditions, as we will shortly enumerate, on the Kahler potential in order that one obtain inflation with an adequate number of e-foldings. These constitute at least one fine tuning needed to obtain an inflaton with mass small compared to the Hubble scale.

3. This structure is not unique; the inflaton need not lie in a supermultiplet with the Goldstino. If there are several multiplets with non-zero R charges, it is possible to tune parameters so that the scalar component of one of these other multiplets is light, while the partner of the Goldstino is heavy. As an example, one can contemplate another field, I. If I couples simply to ϕ^2, this is problematic; inflation does not end. So it is necessary to introduce a field ϕ', and take for the superpotential

$$W = S(\kappa\phi^2 - \mu^2) + \lambda I \phi \phi' + \ldots \tag{9.12}$$

Note that at the minimum of the potential (assumed to be at $S = I = \phi' = 0$), all fields are massive.

4. When we consider the constraints arising from points (1) and (2) above, and the values of the cosmological parameters, we will see that the basic hybrid model of

eqn. 9.8 does not produce suitable inflation unless the N of the Z_N is very large. The model of eqn. 9.12, on the other hand, *does* produce successful inflation for suitable values of the parameters and modest N.

5. We will see that these structures can be embedded in a model of low energy supersymmetry breaking. This is not required, but would seem elegant and economical.

9.3 The Structure of the Effective Action

In this section, we assume supersymmetry at scales above the Hubble constant of inflation, and the presence of a discrete Z_N R-symmetry. We focus, for now, on the single field model. Our considerations will generalize immediately to the multi-field case. For slow roll, it is crucial that $\eta \ll 1$. This translates into the statement that the curvature of the S potential, during inflation, must be smaller than the Hubble constant.

$$\eta \ll 1 \tag{9.13}$$

$$\Rightarrow M_p^2 \frac{V''}{V} \ll 1 \tag{9.14}$$

$$\Rightarrow V'' \ll \frac{V}{M_p^2}. \tag{9.15}$$

$$\tag{9.16}$$

However, $V \approx \mu^4$ and during inflation, $H \approx \frac{1}{3} V \approx \mu^4$. Therefore,

$$V'' \ll \frac{\mu^4}{M_p^2}, \frac{H}{M_p^2}. \tag{9.17}$$

This is a strong condition, and as we will describe, requires tuning the parameters of the potential for S. An even stronger condition arises from the requirement that the field actually flows towards the minimum at the origin.

We have already argued that an R symmetry is a necessary ingredient in successful small-field inflation. Because we do not expect continuous global symmetries in nature, the R symmetry must be discrete; we will take it to be Z_N. So the superpotential has the form of eqn. 9.8, but with additional terms, which we assume to be Planck suppressed[4]:

$$W = S(\kappa\phi^2 - \mu^2) + W_R; \quad W_R = \frac{\lambda}{2(N+1)} \frac{S^{N+1}}{M_p^{N-2}} + \mathcal{O}\left(\frac{S^{2N+1}}{M_p^{2N-2}}\right). \tag{9.18}$$

We have called the S^{N+1} term W_R because it breaks the would-be continuous R symmetry.

The Kahler potential is critical to obtaining a mass for S much smaller than the Hubble constant and successful inflation. Expanding in powers of S, and exploiting the assumption of R symmetry, it takes the form

$$K = S^\dagger S + \phi^\dagger\phi - \frac{\alpha}{4M_p^2}(S^\dagger S)^2 + \ldots. \tag{9.19}$$

We are assuming that, apart from μ, the Planck scale is the only relevant scale; if this is not the case, the fine-tuning problems we discuss below will be more severe. With this assumption, the neglected terms will not be important when the fields are small. We have assumed that other physics is controlled by the Planck scale.

[4]Such terms in inflationary models have been considered in [18], where they also play an important role; the scales assumed there, and the detailed picture of inflation, are quite different, though they resemble some of our discussion in section 9.7.

For large values of the fields, the supergravity contributions to the potential, arising from the quartic terms in the Kahler potential, dominate the potential for S. Most importantly, they give rise to a quadratic term [8, 21]:

$$V_{SUGRA} = \alpha\mu^4 \frac{S^\dagger S}{M_p^2}. \tag{9.20}$$

For lower values of S, the quantum corrections arising from integrating out ϕ, can be important. In particular, in the regime where $|\kappa^2 S^2| \gg |\kappa\mu^2|$ (i.e. on the pseudo moduli space, eqn. 9.11), these corrections are easily computed [9, 10, 21]:

$$\delta K_{quant}(S, S^\dagger) = \frac{\kappa^2}{16\pi^2} S^\dagger S \log(S^\dagger S). \tag{9.21}$$

This Kahler potential is appropriate to the description of the theory at scales below κS, and at times when S is slowly varying. We now need to compute the corresponding quantum correction to the potential. In the regime $|S| > S_c$, ϕ vanishes and F_S is nonzero:

$$F_S = \frac{\partial W}{\partial S} = \mu^2. \tag{9.22}$$

The one-loop vacuum energy is

$$\sum_i (-1)^F \int \frac{d^3k}{(2\pi)^3} \frac{1}{2} \sqrt{\vec{k}^2 + m_i^2}, \tag{9.23}$$

where the sum is over all helicity states, and the $(-1)^F$ is $+1$ for bosons and -1 for fermions. If supersymmetry were unbroken, the bosonic and fermionic contributions would cancel. However, the nonzero F_S term breaks supersymmetry, and therefore there is a logarithmic divergence. Expanding the square root in powers of m^2/k^2, we see that the leading term is independent of m^2 and therefore vanishes. Moreover, the next term vanishes by the sum rule

$$\sum(-1)^F m_i^2 = 0. \tag{9.24}$$

Therefore, at one loop, the potential is

$$V(S) \approx \sum \frac{(-1)^F}{64\pi^2} m_i^4 \ln\left(\frac{m_i^2}{\Lambda^2}\right) \tag{9.25}$$

and Λ denotes a renormalization mass. Along the trajectory $S > S_c$, $\phi = 0$ the one loop effective corrected potential becomes

$$V_{eff}(S) \approx \mu^4 + \frac{\kappa^2}{32\pi^2}\left[2\mu^4 \ln\left(\frac{\kappa^2|S|^2}{\Lambda^2}\right) + (\kappa S^2 - \mu^2)^2 \ln\left(1 - \frac{\mu^2}{\kappa S^2}\right)\right. \tag{9.26}$$

$$\left. + (\kappa S^2 + \mu^2)^2 \ln\left(1 + \frac{\mu^2}{\kappa S^2}\right)\right]. \tag{9.27}$$

In the regime $S \gg S_c$, the quantum correction to the potential reduces to:

$$V_{quant} = \frac{\kappa^2}{16\pi^2}\mu^4 \log(S^\dagger S). \tag{9.28}$$

The quantum corrections will dominate on the pseudo moduli space if the scale, S_{quant}, below which the quantum contribution to the potential are larger than the supergravity contribution, lies on the moduli space. S_{quant} is obtained by comparing the second derivative of V_{SUGRA} with that of V_{quant}:

$$|S_{quant}^2| = \frac{1}{2\alpha}\frac{\kappa^2}{16\pi^2}M_p^2. \tag{9.29}$$

The structure we have outlined above constitutes what is usually called *supersymmetric hybrid inflation* [5, 8–10, 21, 23, 28, 30], but with the modification W_R. Again we see that with the assumption of small field excursions (compared to M_p), and some modest assumptions about naturalness, hybrid inflation is *almost* inevitable. After further studies of this structure, we will subject the various assumptions to closer scrutiny, and ask whether some may be relaxed.

So far we have neglected W_R. This term creates an additional potential on the pseudo moduli space (indeed it is not sensible to speak of such a moduli space in anything but an approximate sense, even neglecting supersymmetry breaking). The leading correction to the potential behaves as

$$\frac{\lambda\mu^2 S^N}{M_p^{N-2}}. \tag{9.30}$$

For sufficiently large field, this overwhelms both V_{SUGRA} and V_{quant}. The potential, in this regime, is not flat enough to inflate. As we will see, this constrains μ.

9.4 Inflation in the Single Field model

In this section, we attempt to implement inflation in the single field model. We will encounter difficulties, finding that the model is not compatible with facts of cosmology except for very large N. But the model will be illustrative and is readily modified to accommodate astrophysical observations.

Let us first suppose that W_R is sufficiently suppressed that it can be ignored during inflation. We will quantify this in the next section. Our interest is in inflation on the pseudo moduli space. The S potential arises from the Kahler potential. The slow roll conditions are:

$$\eta = \frac{V''}{V} M_p^2 \ll 1; \quad \epsilon = \frac{1}{2}\left(\frac{V'}{V}\right)^2 M_p^2 \ll 1. \tag{9.31}$$

If V_{SUGRA} dominates, both conditions are satisfied if $\alpha \ll 1$, and if $|S| \ll M_p$; these are minimal conditions for successful inflation in any case.

If there is a region where V_{quant} dominates, i.e. if quantum corrections dominate over the supergravity potential before reaching the "waterfall regime" ($|\kappa t S| \gg |\mu|$), then inflation may end before S reaches the waterfall regime; $\eta \approx 1$ for [9]:

$$S_f^2 = \frac{\kappa^2}{8\pi^2} M_p^2. \tag{9.32}$$

Alternatively, inflation may end when one enters the waterfall region,

$$S_{wf} = \frac{1}{\sqrt{\kappa}}\mu \tag{9.33}$$

We will see now that the requirement that κ be suitable to lead to sufficient inflation yields

$$S_{qu} \ll S_{wf} \tag{9.34}$$

So inflation, if it occurs at all, takes place in the supergravity regime. This is problematic, if nothing else, from the point of view of the spectral index, n_s,

$$n_s = 1 + 2\eta. \tag{9.35}$$

In the supergravity regime, this is greater than one, which appears inconsistent with results from WMAP.

Following [21], we treat S as real (this implies no loss of generality provided slow roll is valid and for small fields, i.e. when W_R is negligible), and define $\sigma = \sqrt{2}S$. The number of e-foldings in the supergravity regime is:

$$\mathcal{N}_{SUGRA} = \int_{\sigma_0}^{\sigma_i} d\sigma \frac{V}{V' M_p^2} \tag{9.36}$$

where σ_0 is the larger of $\sigma_{quant}, \sigma_{wf}$. This yields:

$$\mathcal{N}_{SUGRA} = \frac{1}{2\alpha} \ln \frac{\sigma_i}{\sigma_0}, \tag{9.37}$$

The total number of e-foldings in the would-be quantum regime is

$$\mathcal{N}_{quant} = \int_{\sigma_f}^{\sigma_{quant}} d\sigma \frac{V}{V' M_p^2} \tag{9.38}$$

$$= \frac{1}{4\alpha}. \tag{9.39}$$

σ_i, the initial value of the field, is not yet constrained by any of our considerations. For now we see that significant inflation requires that α is small. Generically, this is a tuning, of order $\frac{1}{N}$ (possibly modulo a logarithm). *This is irreducible.*

If there are 60 or so e-foldings in the quantum regime, we can determine the values of the slow roll parameters purely in terms of known quantities. $\frac{\delta\rho}{\rho}$ is determined in terms of μ and κ:

$$V^{3/2}/V' = 5.15 \times 10^{-4} M_p^3. \tag{9.40}$$

This expression determines μ in terms of κ (or vice versa):

$$\frac{V_{quant}^{3/2}}{V'_{quant}} = \frac{8\pi^2\mu^2\sigma_{60}}{\kappa^2} = 5.15 \times 10^{-4} M_p^3. \tag{9.41}$$

From equation (9.38) we can solve for σ_{60}:

$$\mathcal{N} = \int_0^{\sigma_N} d\sigma \frac{V}{V' M_p^2} \tag{9.42}$$

$$\Rightarrow \quad N = \frac{4\pi^2}{\kappa^2 M_p^2}\sigma_N^2 \tag{9.43}$$

$$\Rightarrow \quad \sigma_N = \frac{\kappa M_p}{2\pi}\sqrt{N}. \tag{9.44}$$

We can now solve for κ as a function of μ:

$$\frac{8\pi^2\mu^2}{\kappa^2}\left(\frac{\kappa M_p}{2\pi}\sqrt{60}\right) = \frac{V^{3/2}}{V'} \tag{9.45}$$

so that

$$\kappa = 4\pi\mu^2 M_p\sqrt{60}\left(\frac{V^{3/2}}{V'}\right)^{-1} \tag{9.46}$$

$$= 0.17 \times \left(\frac{\mu}{10^{15}\mathrm{GeV}}\right)^2 \tag{9.47}$$

$$= 7.1 \times 10^5 \times \left(\frac{\mu}{M_p}\right)^2. \tag{9.48}$$

Substituting our result for κ, the condition that one not have already entered the waterfall regime is:

$$\frac{\sigma_{60}}{\sigma_{wf}} = \frac{\kappa^{3/2}\sqrt{60}}{4\pi}\left(\frac{M_p}{\mu}\right) = 3.7 \times 10^8 \left(\frac{\mu}{M_p}\right)^2 \tag{9.49}$$

The requirement that this be much greater than one yields:

$$\frac{\mu}{M_p} \gg 5 \times 10^{-5}. \tag{9.50}$$

This is a rather high scale.

We will see in section 9.6 that there is an *upper* bound on the scale of inflation, μ (depending on N). Only for very large N or small λ are scales as large as those of equation (9.58) achievable.

9.5 Inflation in the Two Field Model

The difficulty we have encountered in the single field model can be resolved by invoking the model of eqn. 9.12, with two Hubble-mass fields. In this model, as we noted, the inflation is not the partner of the Goldstino. The quantum potential is now:

$$\delta K_{quant}(S, S^\dagger, I, I^\dagger) = \frac{\kappa^2}{16\pi^2} S^\dagger S \log(I^\dagger I). \tag{9.51}$$

The condition on κ required to obtain a suitable fluctuation spectrum is essentially as before, with S replaced by I; similarly for the formula for the number of e-foldings. But the condition that $I_{qu} \gg I_{wf}$ is now much different, since

$$I_{wf} = \frac{\kappa \mu^2}{\lambda} \tag{9.52}$$

and λ can be of order one. Indeed, in this model, inflation ends when I is sufficiently small that $\eta \approx 1$, which occurs well before reaching the waterfall regime.

In the limit in which we study, in which supergravity corrections are unimportant, η and n_s are universal, and ϵ is small.

$$\eta = -\frac{1}{\mathcal{N}};\ n_s = 1 + 2\eta. \tag{9.53}$$

So one expects, quite generally, that if inflation occurs in the quantum regime, $n_s \approx 0.98$. Again, we note that if inflation occurs in the supergravity regime, one predicts $n_s > 1$, which appears to be ruled out by current CMBR observations.

9.6 Constraints from W_R

In the presence of W_R, the system has supersymmetric minima.

$$\frac{\partial W}{\partial S} = \kappa\phi^2 - \mu^2 + \frac{\lambda}{2}\frac{S^N}{M_p^{N-2}} \tag{9.54}$$

$$\frac{\partial W}{\partial \phi} = 2S\kappa\phi \tag{9.55}$$

Therefore, the supersymmetric minimum is at

$$S^N = \frac{2\mu^2 M_p^{N-2}}{\lambda}; \quad \phi = 0. \tag{9.56}$$

At large S, the potential includes terms

$$V_R = 2\lambda\mu^2 \frac{S^N}{M_p^{N-2}} \tag{9.57}$$

If these terms dominate, the system will be driven towards the supersymmetric minimum. So if we insist that the system is driven to the R symmetric stationary point, we must require that these terms are small, and this in turn places limits on the scale μ (or, through equation 9.48, the coupling κ), as well as σ_i.

The analysis of the previous section goes through provided that $\sigma_i > \sigma_{quant}$, and that $V_R''(\sigma_i) \ll V_{quant}''(\sigma_i)$. The real constraint on the underlying model comes from the requirement that $V_R''(\sigma_{quant}) \ll V_{quant}''(\sigma_{quant})$. This translates into a restriction on μ, or equivalently κ:

$$\left(\frac{\mu}{10^{15} GeV}\right)^{2N-6} \ll \frac{0.34(69)^{N-2}\alpha^{N/2}}{\lambda N(N-1)} \times 10^{-6} \tag{9.58}$$

For $N = 4 - 6$, this yields for the maximal scale:

$$N = 4: \;\; \mu \approx 1.2 \times 10^{11} (\alpha \times 100) \text{ GeV} \tag{9.59}$$

$$N = 5: \;\; \mu \approx 1.6 \times 10^{13} (\alpha \times 100)^{5/8} \text{ GeV} \tag{9.60}$$

$$N = 6: \;\; \mu \approx 8.0 \times 10^{13} (\alpha \times 100)^{1/2} \text{ GeV} \tag{9.61}$$

Even for larger N, the scale is not extremely large; e.g. for $N = 12$, it is only of order 8×10^{14}.

For $N = 3$, there is no choice of μ for which W_R does not dominate. One can try to resolve this by including higher order terms in the Kahler potential and considering supergravity corrections to the *potential* of the form:

$$\delta V = \beta \mu^4 \frac{|S|^4}{M_p^2}. \tag{9.62}$$

However, in this case, all of the activity occurs, for $\beta \sim 1$, for $S \sim \sqrt{\alpha} M_p$, which does not seem consistent with the idea of small field inflation, unless α is tuned to be *extremely* small. Alternatively, the coefficient of the operator appearing in W_R might be very small. Calling

$$W_R = \frac{\lambda}{2(N+1)} \frac{S^{N+1}}{M_p^{N-1}}, \tag{9.63}$$

in the case $N = 3$, we require

$$\lambda \ll 10^{-5} \alpha^{3/2}. \tag{9.64}$$

Indeed, we could consider, for any N, the possibility that μ is larger than implied by eqn. 9.58, and λ is small. The general condition is:

$$\lambda \ll \left(\frac{\mu}{M_p} \right)^{6-2N} \alpha^{N/2} \frac{2(6.65 \times 10^{-5})^{N-2}}{N(N-1)}. \tag{9.65}$$

Thus λ has to be extremely small for reasonable values of N. Small μ is arguably more plausible.

In the two-field model, one has similar constraints on the inflationary scale. A coupling like that of W_R, with S^{N+1} replaced by SI^N, for example, has essentially identical effects.

Note that we now have now enumerated three types of constraints/tunings:

1. α must be small, comparable, up to a logarithmic factor, to $1/\mathcal{N}$, one over the number of e-foldings.

2. μ and κ must be small, in order that the superpotential corrections not drive S to a large field regime.

3. The initial conditions for S are constrained. S must lie in a range small enough that, at least for a time, the quantum potential is dominant, and large enough that inflation can take place.

9.7 Non-Hybrid Scenarios

So far, we have insisted on an unbroken, discrete R symmetry at the end of inflation. But we might relax this. For example, consider a model with superpotential

$$W = \mu^2 S - \frac{\lambda}{2(N+1)} \frac{S^{N+1}}{M_p^{N-2}}. \tag{9.66}$$

Here we might try to arrange that the field, during inflation, rolls towards the supersymmetric minimum at

$$S_0^N = \frac{2M_p^{N-2}\mu^2}{\lambda}. \tag{9.67}$$

As we will see, the conditions for slow roll inflation can be satisfied for a range of initial field values, and adequate e-foldings and fluctuation spectrum obtained. Inflation ends as the field moves towards the minimum. It is important, however, that the energy not be negative at the end of inflation, and this requires a small constant in the superpotential. This constant breaks the discrete R symmetry. The breaking effect, however, is small. Indeed one might want to introduce a spurion for this breaking, with size $S_0\mu^2/M_p$, and charge 2 under the discrete R symmetry. This would not significantly perturb the potential. While the choice of W_0 represents a fine tuning, arguably it is part of the same fine tuning required to understand the present value of the cosmological constant, and so it is required in any case.

To begin, it is worth noting why inflation can work in such a model:

1. Near the minimum, the field is massive compared to the Hubble scale, μ^2/M_p, so inflation can end.

2. At lower values of the field, S is lighter. Indeed, one can find a range of S such that

$$|\eta| \ll 1; \quad \eta < 0. \tag{9.68}$$

3. Inflation ends (η becomes of order one) before reaching an inflection point of the

potential.

4. A suitable number of e-foldings can be obtained, connected to n_s, but limited, again, by the value of the parameter α in the superpotential.

5. $\frac{\delta\rho}{\rho}$ constrains the inflationary energy scale.

Let's elaborate on these points. Parametrically, the inflection point in the potential is of order S_0. Inflation ends at σ_f, defined by the condition

$$V''(\sigma_f) \approx \frac{\mu^4}{M_p^2}. \tag{9.69}$$

σ_f is parametrically smaller than S_0 and the inflection point.

For sufficiently small S, the Kahler potential corrections dominate over the higher dimension superpotential corrections; this occurs for

$$S^{N-2} = \frac{\mu^2 M_p^{N-4} \alpha}{N(N-1)\lambda}. \tag{9.70}$$

So, similar to the hybrid case, we can distinguish two epochs of inflation; one in which the Kahler potential corrections dominate, and one in which the superpotential corrections dominate. We call these the Kahler epoch and the W epoch, respectively. If we call σ_i the initial field, σ_{kahler} the value of σ at the transition between the Kahler and W epochs, and σ_f the value of the field when inflation ends, we have

$$\sigma_{kahler} = \left(\frac{\alpha\mu^2 M_p^{N-4}}{N(N-1)\lambda}\right)^{\frac{1}{N-2}} \qquad \sigma_f = \left(\frac{\mu^2 M_p^{M-4}}{2N(N-1)\lambda}\right)^{\frac{1}{N-2}}. \tag{9.71}$$

If the system is to flow to the supersymmetric minimum S_0, we require $\sigma_i > \sigma_{quant}$. Note that we also require that α be small and negative if $\sigma_{quant} < \sigma_i < \sigma_{kahler}$. The number of e-foldings is

$$\mathcal{N} = \mathcal{N}_{kahler} + \mathcal{N}_W = \frac{1}{2\alpha}\int_{\sigma_i}^{\sigma_{kahler}} \frac{d\sigma}{\sigma} + \frac{\mu^2 M_p^{N-4}}{2N\lambda}\int_{\sigma_{kahler}}^{\sigma_f} \frac{d\sigma}{\sigma^{N-1}}. \tag{9.72}$$

If we assume the last$\mathcal{N}$ e-foldings of inflation happen in the regime dominated by the superpotential corrections, the spectral index is given by

$$n_s = 1 - \frac{2(N-2)}{\mathcal{N}(N-2) + N - 1} \tag{9.73}$$

For $\mathcal{N} = 60$ and $N = 4$ this gives $n_s \approx 0.97$.

When inflation ends, S oscillates about S_0. It is important that there be a mechanism to dissipate the energy of oscillation. This does require coupling to additional fields. For example, adding again a 5 and $\bar{5}$ of $SU(5)$:

$$\delta W = \kappa S \bar{5} 5, \tag{9.74}$$

S can decay to gauge boson pairs. Note, however, that for sufficiently small S, the quantum contributions to the potential will dominate over those we have considered up to now. For initial configurations in this regime, the field will flow towards the origin. This places a *lower* limit on σ_i, which depends on κ.

9.8 Conclusions

So we have seen that small field inflation is likely to require supersymmetry, and that conventional notions of naturalness also lead to the inevitable requirement of an R symmetry. This leaves two classes of models: hybrid and RBI. In the former, we have seen that the requirement that the R symmetry be discrete places an *upper* bound on the scale of inflation, which makes observations of tensor modes in the CMB extremely unlikely. Also inevitably, $n_s < 1$, typically about 0.98. In the RBI case, there are also constraints on scales, and one requires some sort of soft breaking of the R symmetry, describable, perhaps, by spurion fields.

We have seen that there is a simple effective field theory description of these types of inflation, and that one can use the language of global supersymmetry, perturbed slightly (but in critically important ways) by coupling to supergravity. In this framework, the simplest models have an inflaton which lies in a supermultiplet with the gravitino, but this is not necessary. Indeed, we have seen that once one considers higher dimension operators, there is an upper bound on the energy scale of inflation, and that models with at least one additional field more readily lead to successful inflation. One could go further than we have here in applying the language of [17] to this problem.

We have not discussed the problem of initial conditions at any length. Certainly there are constraints on the initial values of the fields, their velocities, and the degree of homogeneity required. These issues look different in different contexts, and we will leave them for further study. One striking feature of this framework is the nature of the cosmological moduli problem. In the models discussed here, the minimum of the potential is a point with an approximate R symmetry. Moduli which are charged under the symmetry naturally sit near the origin, and are not particularly light. Neutral fields (such as the field X in the retrofitted models) naturally sit near the origin as a result of the *accidental, continuous* R symmetry. There seems to be no moduli problem in this context. This issues will be explored further elsewhere.

Finally, as we noted at the beginning, almost by definition it is hard to make general statements about large field inflation. One could attempt this as a limit of the analysis we described above, but then the effective lagrangian has many more “relevant” parameters, and one can probably, at best, simply state constraints consistent with observations on combinations of these. It is hard to see how predictions can emerge without a detailed microscopic understanding of the underlying gravity (supergravity) theory.

Bibliography

[1] Ian Affleck. Testing the Instanton Method. *Phys.Lett.*, B92:149, 1980.

[2] Nathan Andrei and David J. Gross. The Effect of Instantons on the Short Distance Structure of Hadronic Currents. *Phys.Rev.*, D18:468, 1978.

[3] T. Appelquist and R. Shankar. A General Scheme for Computation of Instanton Effects. *Phys.Lett.*, B78:468–472, 1978.

[4] C.A. Baker, D.D. Doyle, P. Geltenbort, K. Green, M.G.D. van der Grinten, et al. An Improved experimental limit on the electric dipole moment of the neutron. *Phys.Rev.Lett.*, 97:131801, 2006.

[5] M. Bastero-Gil, S.F. King, and Q. Shafi. Supersymmetric Hybrid Inflation with Non-Minimal Kahler potential. *Phys.Lett.*, B651:345–351, 2007.

[6] A. Bazavov et al. Results from the MILC collaboration's SU(3) chiral perturbation theory analysis. *PoS*, LAT2009:079, 2009.

[7] Claude W. Bernard. Gauge Zero Modes, Instanton Determinants, and QCD Calculations. *Phys.Rev.*, D19:3013, 1979.

[8] Edmund J. Copeland, Andrew R. Liddle, David H. Lyth, Ewan D. Stewart, and David Wands. False vacuum inflation with Einstein gravity. *Phys.Rev.*, D49:6410–6433, 1994.

[9] G.R. Dvali, Q. Shafi, and Robert K. Schaefer. Large scale structure and supersymmetric inflation without fine tuning. *Phys.Rev.Lett.*, 73:1886–1889, 1994.

[10] W. Fischler and W. Tangarife Garcia. Hierarchies of Susy Splittings and Invisible Photinos as Dark Matter. *JHEP*, 1101:025, 2011.

[11] Maarten F.L. Golterman and Jan Smit. Selfenergy and Flavor Interpretation of Staggered Fermions. *Nucl.Phys.*, B245:61, 1984.

[12] Eric B. Gregory, Alan C. Irving, Chris M. Richards, and Craig McNeile. Methods for Pseudoscalar Flavour-Singlet Mesons with Staggered Fermions. *Phys.Rev.*, D77:065019, 2008.

[13] David J. Gross, Robert D. Pisarski, and Laurence G. Yaffe. QCD and Instantons at Finite Temperature. *Rev.Mod.Phys.*, 53:43, 1981.

[14] D.J. Gross and Frank Wilczek. Ultraviolet Behavior of Nonabelian Gauge Theories. *Phys.Rev.Lett.*, 30:1343–1346, 1973.

[15] Hardy M. Hodges, George R. Blumenthal, Lev A. Kofman, and Joel R. Primack. Nonstandard Primordial Fluctuations From a Polynomial Inflaton Potential. *Nucl.Phys.*, B335:197, 1990.

[16] L.A. Kofman and Andrei D. Linde. Generation of Density Perturbations in the Inflationary Cosmology. *Nucl.Phys.*, B282:555, 1987.

[17] Zohar Komargodski and Nathan Seiberg. From Linear SUSY to Constrained Superfields. *JHEP*, 0909:066, 2009.

[18] Kazuya Kumekawa, Takeo Moroi, and Tsutomu Yanagida. Flat potential for inflaton with a discrete R invariance in supergravity. *Prog.Theor.Phys.*, 92:437–448, 1994.

[19] Andrei D. Linde. Axions in inflationary cosmology. *Phys.Lett.*, B259:38–47, 1991.

[20] Andrei D. Linde. Hybrid inflation. *Phys.Rev.*, D49:748–754, 1994.

[21] Andrei D. Linde and Antonio Riotto. Hybrid inflation in supergravity. *Phys.Rev.*, D56:1841–1844, 1997.

[22] David H. Lyth. Particle physics models of inflation. *Lect.Notes Phys.*, 738:81–118, 2008.

[23] David H. Lyth and Antonio Riotto. Particle physics models of inflation and the cosmological density perturbation. *Phys.Rept.*, 314:1–146, 1999.

[24] Ann E. Nelson and Nathan Seiberg. R symmetry breaking versus supersymmetry breaking. *Nucl.Phys.*, B416:46–62, 1994.

[25] V.A. Novikov, Mikhail A. Shifman, A.I. Vainshtein, and Valentin I. Zakharov.

Operator Expansion in Quantum Chromodynamics Beyond Perturbation Theory. *Nucl.Phys.*, B174:378, 1980.

[26] Michael E. Peskin and Daniel V. Schroeder. An Introduction to quantum field theory. 1995.

[27] H.David Politzer. Reliable Perturbative Results for Strong Interactions? *Phys.Rev.Lett.*, 30:1346–1349, 1973.

[28] Mansoor Ur Rehman, Qaisar Shafi, and Joshua R. Wickman. Supersymmetric Hybrid Inflation Redux. *Phys.Lett.*, B683:191–195, 2010.

[29] Albert S. Schwarz. Quantum field theory and topology. 1994.

[30] Vedat Nefer Senoguz and Q. Shafi. Reheat temperature in supersymmetric hybrid inflation models. *Phys.Rev.*, D71:043514, 2005. 15 pages, 8 figures, revtex4. v2: minor changes, matches published version.

[31] Gerard 't Hooft. Computation of the Quantum Effects Due to a Four-Dimensional Pseudoparticle. *Phys.Rev.*, D14:3432–3450, 1976.

[32] Gerard 't Hooft. Gauge Theories of the Forces Between Elementary Particles. (In French). 1980.

[33] Edward Witten. Instantons, the Quark Model, and the 1/n Expansion. *Nucl.Phys.*, B149:285, 1979.

[34] Chen-Ning Yang and Robert L. Mills. Conservation of Isotopic Spin and Isotopic Gauge Invariance. *Phys.Rev.*, 96:191–195, 1954.