# An Introduction to Lattice Field Theory [1]

Uwe-Jens Wiese
Institut für Theoretische Physik
Universität Bern
Sidlerstrasse 5
CH-3012 Bern, Switzerland

August 30, 2009

[1]Background Material for Lectures presented at the Summer School for Graduate Students "Foundations and New Methods in Theoretical Physics" in Saalburg, September 7 – 11, 2009

These notes provide a brief introduction to the lattice regularization of quantum field theory. Classical field theory is introduced as a generalization of point mechanics to systems with infinitely many degrees of freedom. Quantum mechanics is formulated with path integrals first in real and then in Euclidean time. Field theories in Euclidean space-time resemble 4-d systems of classical statistical mechanics. Lattice fermions suffer from the doubling problem which can be solved by an appropriate definition of chiral symmetry on the lattice. Lattice gauge theories naturally explain confinement in the strong coupling limit. Monte Carlo simulations support the assumption that confinement persists in the weak coupling continuum limit.

# Contents

# Chapter 1

# Introduction and Motivation

The standard model of particle physics summarizes all we know about the fundamental forces of electromagnetism, as well as the weak and strong interactions (but not gravity). It has been tested in great detail up to energies in the hundred GeV range and has passed all these tests very well. The standard model is a relativistic quantum field theory that incorporates the basic principles of quantum mechanics and special relativity. Like quantum electrodynamics (QED) the standard model is a gauge theory, however, with the non-Abelian gauge group $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ instead of the simple Abelian $U(1)_{em}$ gauge group of QED. The gauge bosons are the photons mediating the electromagnetic interactions, the $W$- and $Z$-bosons mediating the weak interactions, as well as the gluons mediating the strong interactions. Gauge theories can exist in several phases: in the Coulomb phase with massless gauge bosons (like in QED), in the Higgs-phase with spontaneously broken gauge symmetry and with massive gauge bosons (e.g. the $W$- and $Z$-bosons), and in the confinement phase, in which the gauge bosons do not appear in the spectrum (like the gluons in quantum chromodynamics (QCD)). All these different phases are indeed realized in Nature and hence in the standard model that describes it.

## 1.1 The Need for a Non-perturbative Regularization

Field theories are systems with infinitely many degrees of freedom — a given number per space point. Their quantization is a subtle issue because too naive approaches lead to divergent results. In order to avoid meaningless divergent results, quantum field theories must be regularized by introducing an ultraviolet cut-off. In particular, just writing down the action of a classical field theory and saying "quantum mechanics will take care of the rest" is cheating. In order to properly define a quantum field theory one must also specify the integration measure of the fields in a path integral. One approach is to expand the path integral in powers of the coupling constant. The resulting Feynman diagrams are then regularized order by order in the coupling. This perturbative approach to field theory has led to impressive results in weakly interacting theories. For example, the anomalous magnetic moment of the electron derived from QED is the quantitatively best understood quantity in physics. Still, even at weak coupling the perturbative approach to field theory is not entirely satisfactory. It is known that perturbation theory is only an asymptotic expansion. The sum of all orders is divergent and thus does not define the theory beyond perturbation theory. Even more important, for strongly coupled theories, like QCD at low energies, the perturbative regularization is completely useless.

Confinement or the Higgs mechanism are non-perturbative phenomena. In order to study them from first principles one must first define the theory beyond perturbation theory. The lattice regularization provides a clean way of doing this by replacing the space-time continuum with a discrete mesh of lattice points. One should not view the lattice as an approximation to the continuum theory. It rather provides a definition of a theory that is undefined directly in the continuum. Of course, in order to recover the continuum limit, the theory must be renormalized by sending the lattice spacing to zero while adjusting the bare coupling constants appropriately. This requires the existence of a second order phase transition in the corresponding 4-d statistical mechanics system. In fact, the lattice is a beautiful regularization because it is local and it respects local gauge symmetries. The fact that it violates some space-time symmetries is less important, because these symmetries are automatically recovered in the continuum limit.

Maintaining chiral symmetry on the lattice is a subtle (but most important)

issue. Lattice fermions have several technical problems that have prevented the non-perturbative formulation of the standard model for many years. For example, chiral fermions — like neutrinos — suffer from the lattice fermion doubling problem. Every left-handed neutrino necessarily comes with a right-handed partner. Until recently, it was not known how to couple only the left-handed particles to an electroweak lattice gauge field. Thanks to a recent breakthrough in lattice gauge theory, the standard model is now consistently defined beyond perturbation theory. Even the perturbative definition of the standard model has been incomplete beyond one loop, due to ambiguities in treating $\gamma_5$ in dimensional regularization. All these ambiguities are now eliminated, thanks to the new lattice results. It is good to know that the standard model now stands on a firm mathematical basis and that the path integral expressions we write down to define it are completely well-defined even beyond perturbation theory.

## 1.2 An Unappreciated Hierarchy Problem

Physical phenomena arise over a vast range of energy scales. Attempts to unify gravity with the other fundamental forces suggest that the Planck scale

$$M_P = \frac{1}{\sqrt{G}} \approx 10^{19} \text{ GeV}, \tag{1.2.1}$$

which is constructed from Newton's constant $G$ (and from $\hbar$ and $c$ which we have put to 1) is the highest energy scale relevant to particle physics. On the other hand, ordinary matter receives almost all of its mass from protons and neutrons which have a mass $M \approx 1$ GeV. Can we understand why nucleons exist so far below the Planck scale? This is a typical hierarchy problem of which there are several in physics (also including the notorious cosmological constant problem). As Wilczek has pointed out, the large hierarchy between the Planck scale $M_P$ and the nucleon mass $M$ is responsible for the feebleness of gravity. To understand this, let us compare the strengths of the gravitational attraction and the electrostatic repulsion between two protons at some large distance $R$. The force of gravity is given by

$$F_g = G\frac{M^2}{R^2}, \tag{1.2.2}$$

while the electrostatic force is

$$F_e = \frac{e^2}{R^2}, \tag{1.2.3}$$

where $e$ is the proton's electric charge. The ratio of the two forces is thus

$$\frac{F_g}{F_e} = G\frac{M^2}{e^2} \approx 137\frac{M^2}{M_P^2} \approx 10^{-36}. \tag{1.2.4}$$

Hence, if we can understand why $M \ll M_P$, we can understand why gravity is a very weak force.

As Wilczek has pointed out, the nucleon mass $M$ is much smaller than the Planck scale $M_P$ partly due to the asymptotic freedom of QCD. At the classical level, QCD with massless quarks (QCD Lite as Wilczek calls it) has no dimensionful parameter at all. When the theory is quantized, a scale enters through the mechanism of dimensional transmutation. A non-zero nucleon mass arises even in massless QCD due to the confinement of colored quarks and gluons inside color-neutral hadrons. Thus, the nucleon mass $M$ is a non-perturbatively generated scale which cannot be understood using perturbation theory. The continuum theory, i.e. dimensional regularization and renormalization applied to the QCD Lagrangian, is not even defined at a non-perturbative level. The only non-perturbative definition of QCD is provided by lattice field theory in which continuous space-time is replaced by a hypercubic lattice. In particular, it should be pointed out that lattice QCD is not an approximation to any pre-existing non-perturbatively well-defined theory in the continuum. Of course, as in any other quantum field theory, one must ultimately remove the cut-off. On the lattice, the shortest physical distance is the lattice spacing $a$ which defines an ultraviolet momentum cut-off $1/a$. Removing the cut-off thus means taking the continuum limit $a \to 0$. The masses of hadrons $M = 1/\xi$ are the inverse of a correlation length $\xi$. Taking the continuum limit means that the physical mass $M$ must be much smaller than the cut-off, i.e.

$$M \ll \frac{1}{a} \;\Rightarrow\; \xi \gg a. \tag{1.2.5}$$

Hence, in the continuum limit the physical correlation length $\xi$ goes to infinity in units of the lattice spacing. In the language of classical statistical mechanics this corresponds to a second order phase transition.

Most of the time lattice QCD is used as a very powerful tool for numerical calculations of hadronic properties. However, the lattice can do a lot more than just that. To illustrate this, we will now use lattice QCD to explain why nucleons can exist naturally far below the Planck scale. Of course, it is well known that QCD is not valid up to the Planck scale. In particular, it is embedded in the

standard model which itself is an effective theory limited to energies below about 1 TeV. However, unlike the full standard model, thanks to asymptotic freedom QCD alone makes sense at arbitrarily high energy scales. Whatever replaces QCD and the standard model at ultra-short distances — be it string theory or some tiny wheels turning around at the Planck scale — Nature must have found a concrete way to regularize the QCD physics at ultra-short distances. Due to renormalizability and universality, only the symmetries but not the details of this regularization should matter at low energies. For simplicity, we will use lattice QCD (and not, for example, string theory) as an admittedly oversimplified model of Nature at ultra-short distances. In other words, we identify the lattice cut-off $1/a$ with the Planck scale $M_P$.

Using lattice QCD, how can we then understand why the nucleon mass $M$ is far below $M_P = 1/a$? As Wilczek pointed out, one key ingredient is asymptotic freedom. Thanks to asymptotic freedom, without any fine-tuning of the bare gauge coupling a non-Abelian lattice Yang-Mills theory produces a correlation length $\xi$ that is larger than the lattice spacing $a$ by a factor exponentially large in the inverse coupling. In particular, choosing a bare coupling that is not unnaturally small, one can easily generate a hierarchy of scales like $M_P/M = \xi/a \approx 10^{19}$. Interestingly, the situation is not as simple when one proceeds from a pure gluon Yang-Mills theory to full lattice QCD including quarks. In particular, unlike continuum QCD, lattice QCD does not naturally have a chiral symmetry that can protect the quark masses from running up to the "Planck scale" $1/a$. Indeed, for about two decades lattice field theorists have suffered from a hierarchy problem in the fermion sector. This problem first arose when Wilson removed the unwanted doubler fermions by breaking chiral symmetry explicitly. Recovering chiral symmetry in the continuum limit then requires a delicate fine-tuning of the bare fermion mass. In particular, if at ultra-short distances Nature would be a lattice gauge theory with Wilson fermions, without unnatural fine-tuning, quarks would have masses at the Planck scale and the lightest particles would be glueballs. In that case it would be very puzzling why ordinary matter consists not just of gluons, but also of light quarks. If one works in continuum QCD one often takes chiral symmetry for granted, and one may view this hierarchy puzzle just as a problem of the lattice formulation. However, one should not forget that continuum QCD is not even defined beyond perturbation theory. In addition, subtleties of the definition of $\gamma_5$ in the framework of dimensional regularization affect even the continuum theory, and are just another aspect of the same deep problem of chiral symmetry that is manifest on the lattice. Indeed, there is a se-

vere hierarchy problem for non-perturbative fermion dynamics that Nature must have solved somehow because it presents us with nucleons that exist far below the Planck scale.

Remarkably, the long-standing hierarchy problem of lattice fermions — and hence of the non-perturbative regularization of chiral symmetry — has recently been solved very elegantly. Using previous research of Callan and Harvey, Kaplan realized that massless four-dimensional lattice fermions arise naturally, i.e. without fine-tuning, as zero-modes localized on a domain wall embedded in a five-dimensional space-time. In particular, left- and right-handed fermions can be localized on a domain wall and an anti-wall. When the wall and the anti-wall are separated by a sufficiently large distance, the left- and right-handed modes cannot mix, simply because they are spatially separated. As a result, a Dirac fermion arises which is protected from picking up a large mass and which is thus naturally light. Remarkably, in contrast to four dimensions, a Wilson term in a five-dimensional lattice theory removes the doubler fermions without breaking the chiral symmetry of the light four-dimensional domain wall fermions.

When Kaplan proposed his idea of regulating chiral fermions using domain walls, Narayanan and Neuberger were developing independently another approach to regulating chiral fermions using an infinite number of flavors. Based on this approach they developed what is now referred to as overlap lattice fermions. Since the flavor-space can be viewed as an extra dimension, the overlap approach is closely related to the domain wall approach. When one separates the wall and the anti-wall by an infinite distance, domain wall fermions turn into overlap fermions. Overlap fermions have the advantage that they have an exact chiral symmetry, while the chiral symmetry of domain wall fermions is only approximate for a finite wall-anti-wall separation. Both overlap and domain wall fermions yield naturally light quarks, and both are naturally related to the physics of a higher-dimensional space-time.

Hasenfratz and Niedermayer have investigated non-perturbative renormalization group blocking transformations on the lattice. The fixed points of such transformations correspond to lattice actions which are completely free of cut-off effects — so-called perfect actions. Perfect actions for pure gauge theory, as well as for free Wilson and staggered fermions have also been investigated. In the process of these investigations Hasenfratz rediscovered an old paper by Ginsparg and Wilson. He also realized that what is now called the Ginsparg-Wilson relation is the key to understanding chiral symmetry on the lattice. The Ginsparg-Wilson

relation represents a general requirement on a lattice action which guarantees that it has good chiral properties. When Ginsparg and Wilson discovered this relation, it seemed impossible to explicitly construct lattice actions that obey it. By now it has been shown that classically perfect lattice actions can be approximated well enough, so that the Ginsparg-Wilson action is satisfied with high accuracy. From the point of view of practical lattice QCD calculations this represents very important progress. However, the explicit construction of perfect actions is a delicate problem that can be considered a very elaborate form of fine-tuning. Hence, it seems unnatural that Nature has chosen anything like a perfect action to regularize the strong interactions at ultra-short distances. Unlike perfect fermions, overlap fermions can describe massless quarks in QCD without fine-tuning. By integrating out the extra dimension, Neuberger has constructed lattice Dirac operators for massless quarks analytically and these Dirac operators do indeed satisfy the Ginsparg-Wilson relation exactly. Remarkably, both overlap as well as domain wall fermions, which naturally have a chiral symmetry without fine-tuning, are related to the physics in a higher-dimensional space-time. Hence, the existence of light four-dimensional fermions may be a concrete hint to the physical reality of extra dimensions.

The full strength of the Ginsparg-Wilson relation was realized by Lüscher who discovered that it suggests a natural definition of lattice chiral symmetry which reduces to the usual one in the continuum limit. Based on this insight, Lüscher achieved a spectacular breakthrough: the non-perturbative construction of lattice chiral gauge theories. Hence, not only QCD in which chiral symmetry is global, but also the standard model with its local chiral symmetry now stands on a solid non-perturbative basis. Even continuum perturbation theory can benefit from these developments. In particular, the ambiguities in the definition of $\gamma_5$ that arise in multi-loop calculations using dimensional regularization can be eliminated when one uses the lattice regularization. Still, there is a very long way to go from Lüscher's theoretical construction to practical numerical calculations in chiral gauge theories like the standard model.

The situation is a lot simpler, but still highly non-trivial, in applications of Ginsparg-Wilson fermions to simulations of QCD. Compared to the standard Wilson or staggered lattice fermions, which are already very difficult to treat fully dynamically, domain wall, overlap, or perfect fermions demand even much larger computing power. Hence, at present they are often used in the so-called quenched approximation in which the fermion determinant is ignored. If one does not want to wait a long time for even bigger computers, it will require an

algorithmic breakthrough to bring the theoretical developments of lattice chiral symmetry to fruition in fully dynamical simulations of lattice QCD.

If one imagines that Nature has used something like domain wall fermions to regularize the strong interactions, it is natural that nucleons (and not just glueballs) exist far below the Planck scale. However, it remains mysterious where the quark masses themselves come from. In the standard model the quark masses arise from Yukawa couplings to the Higgs field, but the values of these couplings are free parameters. Still, within the standard model the traditional gauge hierarchy problem arises: why is the electroweak scale so small compared to the Planck scale? Chiral symmetry can protect fermion masses from running to the ultimate high-energy cut-off, but it cannot protect scalars. A potential solution of the gauge hierarchy problem is provided by supersymmetry. Supersymmetry relates scalars to fermions and thus allows chiral symmetry to indirectly generate naturally light scalars as well. At a non-perturbative level, supersymmetry is as undefined as chiral symmetry was before the recent developments on the lattice. In the worst case, supersymmetry may just be a perturbative illusion which does not arise naturally at a non-perturbative level. Unfortunately, unlike for chiral symmetry, Nature has not yet provided us with experimental evidence for supersymmetry (except as an accidental symmetry in heavy nuclei). Hence, one cannot be sure that it is indeed possible to construct naturally light scalars at a non-perturbative level. Perhaps the many beautiful results obtained within supersymmetric continuum theories should make us optimistic that these theories actually exist at a rigorous level beyond perturbation theory. Again, Kaplan and his collaborators have taken very interesting steps towards constructing supersymmetric theories on the lattice. It remains to be seen whether these developments will lead to a repetition of the Ginsparg-Wilson revolution of lattice chiral symmetry.

# Chapter 2

# Quantum Field Theory

This chapter provides a brief summary of the mathematical structure of quantum field theory. Classical field theories are discussed as a generalization of point mechanics to systems with infinitely many degrees of freedom — a given number per space point. Similarly, quantum field theories are just quantum mechanical systems with infinitely many degrees of freedom. In the same way as point mechanics systems, classical field theories can be quantized with path integral methods. The quantization of field theories at finite temperature leads to path integrals in Euclidean time. This provides us with an analogy between quantum field theory and classical statistical mechanics.

## 2.1 From Point Mechanics to Classical Field Theory

Point mechanics describes the dynamics of classical non-relativistic point particles. The coordinates of the particles represent a finite number of degrees of freedom. In the simplest case — a single particle moving in one spatial dimension — we are dealing with a single degree of freedom: the $x$-coordinate of the particle. The dynamics of a particle of mass $m$ moving in an external potential $V(x)$ is described by Newton's equation

$$m\partial_t^2 x = ma = F(x) = -\frac{dV(x)}{dx}. \tag{2.1.1}$$

Once the initial conditions are specified, this ordinary second order differential equation determines the particle's path $x(t)$, i.e. its position as a function of time. Newton's equation results from the variational principle to minimize the action

$$S[x] = \int dt\ L(x, \partial_t x), \tag{2.1.2}$$

over the space of all paths $x(t)$. The action is a functional (a function whose argument is itself a function) that results from the time integral of the Lagrange function

$$L(x, \partial_t x) = \frac{m}{2}(\partial_t x)^2 - V(x). \tag{2.1.3}$$

The Euler-Lagrange equation

$$\partial_t \frac{\delta L}{\delta(\partial_t x)} - \frac{\delta L}{\delta x} = 0, \tag{2.1.4}$$

is nothing but Newton's equation.

Classical field theories are a generalization of point mechanics to systems with infinitely many degrees of freedom — a given number for each space point $\vec{x}$. In this case, the degrees of freedom are the field values $\phi(\vec{x})$, where $\phi$ is some generic field. In case of a neutral scalar field, $\phi$ is simply a real number representing one degree of freedom per space point. A charged scalar field, on the other hand, is described by a complex number and hence represents two degrees of freedom per space point. The scalar Higgs field $\phi^a(\vec{x})$ (with $a \in \{1, 2\}$) in the standard model is a complex doublet, i.e. it has four real degrees of freedom per space point. An Abelian gauge field $A_i(\vec{x})$ (with a spatial direction index $i \in \{1, 2, 3\}$) — for example, the photon field in electrodynamics — is a neutral vector field with 3 real degrees of freedom per space point. One of these degrees of freedom is redundant due to the $U(1)_{em}$ gauge symmetry. Hence, an Abelian gauge field has two physical degrees of freedom per space point which correspond to the two polarization states of the massless photon. Note that the time-component $A_0(\vec{x})$ does not represent a physical degree of freedom. Instead, it is a Lagrange multiplier field that enforces the Gauss law. A non-Abelian gauge field $A_i^a(\vec{x})$ is charged and has an additional index $a$. For example, the gluon field in chromodynamics with a color index $a \in \{1, 2, ..., 8\}$ represents $2 \times 8 = 16$ physical degrees of freedom per space point, again because of redundancy due to the $SU(3)_c$ color gauge symmetry. The field that represents the $W$- and $Z$-bosons in the standard model has an index $a \in \{1, 2, 3\}$ and transforms under the gauge group $SU(2)_L$. Thus, it represents $2 \times 3 = 6$ physical degrees of freedom.

However, in contrast to the photon, the $W$- and $Z$-bosons are massive due to the Higgs mechanism and have three (not just two) polarization states. The extra degree of freedom is provided by the Higgs field.

The analog of Newton's equation in field theory is the classical field equation of motion. For example, for a neutral scalar field this is the Klein-Gordon equation

$$\partial_\mu \partial^\mu \phi = -\frac{dV(\phi)}{d\phi}. \tag{2.1.5}$$

Again, after specifying appropriate initial conditions it determines the classical field configuration $\phi(x)$, i.e. the values of the field $\phi$ at all space-time points $x = (t, \vec{x})$. Hence, the role of time in point mechanics is played by space-time in field theory, and the role of the point particle coordinates is now played by the field values. As before, the classical equation of motion results from minimizing the action

$$S[\phi] = \int d^4x \ \mathcal{L}(\phi, \partial_\mu \phi). \tag{2.1.6}$$

The integral over time in eq.(2.1.2) is now replaced by an integral over space-time and the Lagrange function of point mechanics gets replaced by the Lagrange density function (or Lagrangian)

$$\mathcal{L}(\phi, \partial_\mu \phi) = \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - V(\phi). \tag{2.1.7}$$

A simple interacting field theory is the $\phi^4$ theory with the potential

$$V(\phi) = \frac{m^2}{2}\phi^2 + \frac{\lambda}{4}\phi^4. \tag{2.1.8}$$

Here $m$ is the mass of the scalar field and $\lambda$ is the coupling strength of its self-interaction. Note that the mass term corresponds to a harmonic oscillator potential in the point mechanics analog, while the interaction term corresponds to an anharmonic perturbation. As before, the Euler-Lagrange equation

$$\partial_\mu \frac{\delta L}{\delta(\partial_\mu \phi)} - \frac{\delta L}{\delta \phi} = 0, \tag{2.1.9}$$

is the classical equation of motion, in this case the Klein-Gordon equation. The analogies between point mechanics and field theory are summarized in table 2.1.

| Point Mechanics | Field Theory |
|---|---|
| time $t$ | space-time $x = (t, \vec{x})$ |
| particle coordinate $x$ | field value $\phi$ |
| particle path $x(t)$ | field configuration $\phi(x)$ |
| action $S[x] = \int dt\ L(x, \partial_t x)$ | action $S[\phi] = \int d^4x\ \mathcal{L}(\phi, \partial_\mu \phi)$ |
| Lagrange function $L(x, \partial_t x) = \frac{m}{2}(\partial_t x)^2 - V(x)$ | Lagrangian $\mathcal{L}(\phi, \partial_\mu \phi) = \frac{1}{2}\partial_\mu \phi \partial^\mu \phi - V(\phi)$ |
| equation of motion $\partial_t \frac{\delta L}{\delta(\partial_t x)} - \frac{\delta L}{\delta x} = 0$ | field equation $\partial_\mu \frac{\delta L}{\delta(\partial_\mu \phi)} - \frac{\delta L}{\delta \phi} = 0$ |
| Newton's equation $\partial_t^2 x = -\frac{dV(x)}{dx}$ | Klein-Gordon equation $\partial_\mu \partial^\mu \phi = -\frac{dV(\phi)}{d\phi}$ |
| kinetic energy $\frac{m}{2}(\partial_t x)^2$ | kinetic energy $\frac{1}{2}\partial_\mu \phi \partial^\mu \phi$ |
| harmonic oscillator potential $\frac{m}{2}\omega^2 x^2$ | mass term $\frac{m^2}{2}\phi^2$ |
| anharmonic perturbation $\frac{\lambda}{4}x^4$ | self-interaction term $\frac{\lambda}{4}\phi^4$ |

Table 2.1: *The dictionary that translates point mechanics into the language of field theory.*

## 2.2 The Path Integral in Real Time

The quantization of field theories is most conveniently performed using the path integral approach. Here we first discuss the path integral in quantum mechanics — quantized point mechanics — using the real time formalism. A mathematically more satisfactory formulation uses an analytic continuation to so-called Euclidean time. This will be discussed in the next section.

The real time evolution of a quantum system described by a Hamilton operator $H$ is given by the time-dependent Schrödinger equation

$$i\hbar \partial_t |\Psi(t)\rangle = H|\Psi(t)\rangle. \tag{2.2.1}$$

For a time-independent Hamilton operator the time evolution operator is given by

$$U(t', t) = \exp(-\frac{i}{\hbar} H(t' - t)), \tag{2.2.2}$$

such that

$$|\Psi(t')\rangle = U(t', t)|\Psi(t)\rangle. \tag{2.2.3}$$

Let us consider the transition amplitude $\langle x'|U(t',t)|x\rangle$ of a non-relativistic point particle that starts at position $x$ at time $t$ and arrives at position $x'$ at time $t'$. Using

$$\langle x|\Psi(t)\rangle = \Psi(x,t) \tag{2.2.4}$$

we obtain

$$\Psi(x',t') = \int dx\ \langle x'|U(t',t)|x\rangle \Psi(x,t), \tag{2.2.5}$$

i.e. $\langle x'|U(t',t)|x\rangle$ acts as a propagator for the wave function. The propagator is of physical interest because it contains information about the energy spectrum. When we consider propagation from an initial position $x$ back to the same position we find

$$\begin{aligned} \langle x|U(t',t)|x\rangle &= \langle x|\exp(-\frac{i}{\hbar}H(t'-t))|x\rangle \\ &= \sum_n |\langle x|n\rangle|^2 \exp(-\frac{i}{\hbar}E_n(t'-t)). \end{aligned} \tag{2.2.6}$$

We have inserted a complete set, $\sum_n |n\rangle\langle n| = \mathbb{1}$, of energy eigenstates $|n\rangle$ with

$$H|n\rangle = E_n|n\rangle. \tag{2.2.7}$$

Hence, according to eq.(2.2.6), the Fourier transform of the propagator yields the energy spectrum as well as the energy eigenstates $\langle x|n\rangle$.

Inserting a complete set of position eigenstates we arrive at

$$\begin{aligned} \langle x'|U(t',t)|x\rangle &= \langle x'|\exp(-\frac{i}{\hbar}H(t'-t_1+t_1-t))|x\rangle \\ &= \int dx_1 \langle x'|\exp(-\frac{i}{\hbar}H(t'-t_1))|x_1\rangle \\ &\times\ \langle x_1|\exp(-\frac{i}{\hbar}H(t_1-t))|x\rangle \\ &= \int dx_1 \langle x'|U(t',t_1)|x_1\rangle\langle x_1|U(t_1,t)|x\rangle. \end{aligned} \tag{2.2.8}$$

It is obvious that we can repeat this process an arbitrary number of times. This is exactly what we do in the formulation of the path integral. Let us divide the time interval $[t,t']$ into $N$ elementary time steps of size $\varepsilon$ such that

$$t'-t = N\varepsilon. \tag{2.2.9}$$

Inserting a complete set of position eigenstates at the intermediate times $t_i, i \in \{1, 2, ..., N-1\}$ we obtain

$$\begin{aligned} \langle x'|U(t',t)|x\rangle &= \int dx_1 \int dx_2 ... \int dx_{N-1} \langle x'|U(t',t_{N-1})|x_{N-1}\rangle ... \\ &\times \langle x_2|U(t_2,t_1)|x_1\rangle\langle x_1|U(t_1,t)|x\rangle. \end{aligned} \tag{2.2.10}$$

In the next step we concentrate on one of the factors and we consider a single non-relativistic point particle moving in an external potential $V(x)$ such that

$$H = \frac{p^2}{2m} + V(x). \tag{2.2.11}$$

Using the Baker-Campbell-Haussdorff formula and neglecting terms of order $\varepsilon^2$ we find

$$\begin{aligned} \langle x_{i+1}|U(t_{i+1},t_i)|x_i\rangle &= \langle x_{i+1}| \exp(-\frac{i\varepsilon p^2}{2m\hbar}) \exp(-\frac{i\varepsilon}{\hbar}V(x))|x_i\rangle \\ &= \frac{1}{2\pi}\int dp \langle x_{i+1}| \exp(-\frac{i\varepsilon p^2}{2m\hbar})|p\rangle\langle p| \exp(-\frac{i\varepsilon}{\hbar}V(x))|x_i\rangle \\ &= \frac{1}{2\pi}\int dp \exp(-\frac{i\varepsilon p^2}{2m\hbar}) \exp(-\frac{i}{\hbar}p(x_{i+1}-x_i)) \\ &\times \exp(-\frac{i\varepsilon}{\hbar}V(x_i)). \end{aligned} \tag{2.2.12}$$

The integral over $p$ is ill-defined because the integrand is a very rapidly oscillating function. To make the expression well-defined we replace the time step $\varepsilon$ by $\varepsilon - ia$, i.e. we go out into the complex time plane. After doing the integral we take the limit $a \to 0$. Still, one should keep in mind that the definition of the path integral required an analytic continuation in time. One finds

$$\langle x_{i+1}|U(t_{i+1},t_i)|x_i\rangle = \sqrt{\frac{m}{2\pi i\hbar\varepsilon}} \exp(\frac{i}{\hbar}\varepsilon[\frac{m}{2}(\frac{x_{i+1}-x_i}{\varepsilon})^2 - V(x_i)]). \tag{2.2.13}$$

Inserting this back into the expression for the propagator we obtain

$$\langle x'|U(t',t)|x\rangle = \int \mathcal{D}x \exp(\frac{i}{\hbar}S[x]). \tag{2.2.14}$$

The action has been identified in the time continuum limit as

$$\begin{aligned} S[x] &= \int dt\ [\frac{m}{2}(\partial_t x)^2 - V(x)] \\ &= \lim_{\varepsilon\to 0}\sum_i \varepsilon[\frac{m}{2}(\frac{x_{i+1}-x_i}{\varepsilon})^2 - V(x_i)]. \end{aligned} \tag{2.2.15}$$

The integration measure is defined as

$$\int \mathcal{D}x = \lim_{\varepsilon \to 0} \sqrt{\frac{m}{2\pi i\hbar\varepsilon}}^{N-1} \int dx_1 \int dx_2 ... \int dx_{N-1}. \tag{2.2.16}$$

This means that we integrate over the possible particle positions for each intermediate time $t_i$. In this way we integrate over all possible paths of the particle starting at $x$ and ending at $x'$. Each path is weighted with an oscillating phase factor $\exp(\frac{i}{\hbar}S[x])$ determined by the action. The classical path of minimum action has the smallest oscillations, and hence the largest contribution to the path integral. In the classical limit $\hbar \to 0$ only that contribution survives.

## 2.3 The Path Integral in Euclidean Time

As we have seen, it requires a small excursion into the complex time plane to make the path integral mathematically well-defined. Now we will make a big step into that plane and actually consider purely imaginary so-called Euclidean time. The physical motivation for this, however, comes from quantum statistical mechanics. Let us consider the quantum statistical partition function

$$Z = \mathrm{Tr} \exp(-\beta H), \tag{2.3.1}$$

where $\beta = 1/T$ is the inverse temperature. It is mathematically equivalent to the time interval we discussed in the real time path integral. In particular, the operator $\exp(-\beta H)$ turns into the time evolution operator $U(t', t)$ if we identify

$$\beta = \frac{i}{\hbar}(t' - t). \tag{2.3.2}$$

In this sense the system at finite temperature corresponds to a system propagating in purely imaginary (Euclidean) time. By dividing the Euclidean time interval into $N$ time steps, i.e. by writing $\beta = Na/\hbar$, and again by inserting complete sets of position eigenstates we now arrive at the Euclidean time path integral

$$Z = \int \mathcal{D}x \exp(-\frac{1}{\hbar}S_E[x]). \tag{2.3.3}$$

The action now takes the Euclidean form

$$\begin{aligned} S_E[x] &= \int dt\ [\frac{m}{2}(\partial_t x)^2 + V(x)] \\ &= \lim_{a \to 0} \sum_i a[\frac{m}{2}(\frac{x_{i+1} - x_i}{a})^2 + V(x_i)]. \end{aligned} \tag{2.3.4}$$

In contrast to the real time case the measure now involves $N$ integrations

$$\int \mathcal{D}x = \lim_{a\to 0} \sqrt{\frac{m}{2\pi\hbar a}}^N \int dx_1 \int dx_2 ... \int dx_N. \tag{2.3.5}$$

The extra integration over $x_N = x'$ is due to the trace in eq.(2.3.1). Note that there is no extra integration over $x_0 = x$ because the trace implies periodic boundary conditions in the Euclidean time direction, i.e. $x_0 = x_N$.

The Euclidean path integral allows us to evaluate thermal expectation values. For example, let us consider an operator $\mathcal{O}(x)$ that is diagonal in the position state basis. We can insert this operator in the path integral and thus compute its expectation value

$$\langle \mathcal{O}(x) \rangle = \frac{1}{Z} \mathrm{Tr}[\mathcal{O}(x) \exp(-\beta H)] = \frac{1}{Z} \int \mathcal{D}x \; \mathcal{O}(x(0)) \exp(-\frac{1}{\hbar} S_E[x]). \tag{2.3.6}$$

Since the theory is translation invariant in Euclidean time one can place the operator anywhere in time, e.g. at $t = 0$ as done here. When we perform the low temperature limit, $\beta \to \infty$, the thermal fluctuations are switched off and only the quantum ground state $|0\rangle$ (the vacuum) contributes to the partition function, i.e. $Z \sim \exp(-\beta E_0)$. In this limit the path integral is formulated in an infinite Euclidean time interval, and describes the vacuum expectation value

$$\langle \mathcal{O}(x) \rangle = \langle 0|\mathcal{O}(x)|0\rangle = \lim_{\beta\to\infty} \frac{1}{Z} \int \mathcal{D}x \; \mathcal{O}(x(0)) \exp(-\frac{1}{\hbar} S_E[x]). \tag{2.3.7}$$

It is also interesting to consider 2-point functions of operators at different instances in Euclidean time

$$\begin{aligned} \langle \mathcal{O}(x(0))\mathcal{O}(x(t)) \rangle &= \frac{1}{Z} \mathrm{Tr}[\mathcal{O}(x) \exp(-Ht) \mathcal{O}(x) \exp(Ht) \exp(-\beta H)] \\ &= \frac{1}{Z} \int \mathcal{D}x \; \mathcal{O}(x(0))\mathcal{O}(x(t)) \exp(-\frac{1}{\hbar} S_E[x]). \end{aligned} \tag{2.3.8}$$

Again, we consider the limit $\beta \to \infty$, but we also separate the operators in time, i.e. we also let $t \to \infty$. Then the leading contribution is $|\langle 0|\mathcal{O}(x)|0\rangle|^2$. Subtracting this, and thus forming the connected 2-point function, one obtains

$$\lim_{\beta,t\to\infty} \langle \mathcal{O}(x(0))\mathcal{O}(x(t)) \rangle - |\langle \mathcal{O}(x) \rangle|^2 = |\langle 0|\mathcal{O}(x)|1\rangle|^2 \exp(-(E_1 - E_0)t). \tag{2.3.9}$$

Here $|1\rangle$ is the first excited state of the quantum system with an energy $E_1$. The connected 2-point function decays exponentially at large Euclidean time

separations. The decay is governed by the energy gap $E_1 - E_0$. In a quantum field theory $E_1$ corresponds to the energy of the lightest particle. Its mass is determined by the energy gap $E_1 - E_0$ above the vacuum. Hence, in Euclidean field theory particle masses are determined from the exponential decay of connected 2-point correlation functions.

## 2.4 Spin Models in Classical Statistical Mechanics

So far we have considered quantum systems both at zero and at finite temperature. We have represented their partition functions as Euclidean path integrals over configurations on a time lattice of length $\beta$. We will now make a completely new start and study classical discrete systems at finite temperature. We will see that their mathematical description is very similar to the path integral formulation of quantum systems. Still, the physical interpretation of the formalism is drastically different in the two cases. In the next section we will set up a dictionary that allows us to translate quantum physics language into the language of classical statistical mechanics.

For simplicity, let us concentrate on simple classical spin models. Here the word spin does not mean that we deal with quantized angular momenta. All we do is work with classical variables that can point in specific directions. The simplest spin model is the Ising model with classical spin variables $s_x = \pm 1$. (Again, these do not represent the quantum states up and down of a quantum mechanical angular momentum 1/2.) More complicated spin models with an $O(N)$ spin rotational symmetry are the XY model ($N = 2$) and the Heisenberg model ($N = 3$). The spins in the XY model are 2-component unit-vectors, while the spins in the Heisenberg model have three components. In all these models the spins live on the sites of a $d$-dimensional spatial lattice. The lattice is meant to be a crystal lattice (so typically $d = 3$) and the lattice spacing has a physical meaning. This is in contrast to the Euclidean time lattice that we have introduced to make the path integral mathematically well-defined, and that we finally send to zero in order to reach the Euclidean time continuum limit. The Ising model is characterized by its classical Hamilton function (not a quantum Hamilton operator) which simply specifies the energy of any configuration of

spins. The Ising Hamilton function is a sum of nearest neighbor contributions

$$\mathcal{H}[s] = J \sum_{\langle xy \rangle} s_x s_y - \mu B \sum_x s_x, \tag{2.4.1}$$

with a ferromagnetic coupling constant $J < 0$ that favors parallel spins, plus a coupling to an external magnetic field $B$. The classical partition function of this system is given by

$$Z = \int \mathcal{D}s \ \exp(-\mathcal{H}[s]/T) = \prod_x \sum_{s_x = \pm 1} \exp(-\mathcal{H}[s]/T). \tag{2.4.2}$$

The sum over all spin configurations corresponds to an independent summation over all possible orientations of individual spins. Thermal averages are computed by inserting appropriate operators. For example, the magnetization is given by

$$\langle s_x \rangle = \frac{1}{Z} \prod_x \sum_{s_x = \pm 1} s_x \exp(-\mathcal{H}[s]/T). \tag{2.4.3}$$

Similarly, the spin correlation function is defined by

$$\langle s_x s_y \rangle = \frac{1}{Z} \prod_x \sum_{s_x = \pm 1} s_x s_y \exp(-\mathcal{H}[s]/T). \tag{2.4.4}$$

At large distances the connected spin correlation function typically decays exponentially

$$\langle s_x s_y \rangle - \langle s_x \rangle \langle s_y \rangle \sim \exp(-|x-y|/\xi), \tag{2.4.5}$$

where $\xi$ is the so-called correlation length. At general temperatures the correlation length is typically just a few lattice spacings. When one models real materials, the Ising model would generally be a great oversimplification, because real magnets, for example, not only have nearest neighbor couplings. Still, the details of the Hamilton function at the scale of the lattice spacing are not always important. There is a critical temperature $T_c$ at which $\xi$ diverges and universal behavior arises. At this temperature a second order phase transition occurs. Then the details of the model at the scale of the lattice spacing are irrelevant for the long range physics that takes place at the scale of $\xi$. In fact, at their critical temperatures some real materials behave just like the simple Ising model. This is why the Ising model is so interesting. It is just a very simple member of a large universality class of different models, which all share the same critical behavior. This does not mean that they have the same values of their critical temperatures. However, their magnetization goes to zero at the critical temperature with the same power of $T_c - T$, i.e. their critical exponents are identical.

## 2.5 Quantum Mechanics versus Statistical Mechanics

We notice a close analogy between the Euclidean path integral for a quantum mechanical system and a classical statistical mechanics system like the Ising model. The path integral for the quantum system is defined on a 1-dimensional Euclidean time lattice, just like an Ising model can be defined on a $d$-dimensional spatial lattice. In the path integral we integrate over all paths, i.e. over all configurations $x(t)$, while in the Ising model we sum over all spin configurations $s_x$. Paths are weighted by their Euclidean action $S_E[x]$ while spin configurations are weighted with their Boltzmann factors depending on the classical Hamilton function $\mathcal{H}[s]$. The prefactor of the action is $1/\hbar$, and the prefactor of the Hamilton function is $1/T$. Indeed $\hbar$ determines the strength of quantum fluctuations, while the temperature $T$ determines the strength of thermal fluctuations. The kinetic energy $\frac{1}{2}((x_{i+1} - x_i)/a)^2$ in the path integral is analogous to the nearest neighbor spin coupling $s_x s_{x+1}$, and the potential term $V(x_i)$ is analogous to the coupling $\mu B s_x$ to an external magnetic field. The magnetization $\langle s_x \rangle$ corresponds to the vacuum expectation value of an operator $\langle \mathcal{O}(x) \rangle$ and the spin-spin correlation function $\langle s_x s_y \rangle$ corresponds to the 2-point correlation function $\langle \mathcal{O}(x(0))\mathcal{O}(x(t)) \rangle$. The inverse correlation length $1/\xi$ is analogous to the energy gap $E_1 - E_0$ (and hence to a particle mass in a Euclidean quantum field theory). Finally, the Euclidean time continuum limit $a \to 0$ corresponds to a second order phase transition where $\xi \to \infty$. The lattice spacing in the path integral is an artifact of our mathematical description which we send to zero while the physics remains constant. In classical statistical mechanics, on the other hand, the lattice spacing is physical and hence fixed, while the correlation length $\xi$ goes to infinity at a second order phase transition. All this is summarized in the dictionary of table 2.2.

## 2.6 The Transfer Matrix

The analogy between quantum mechanics and classical statistical mechanics suggests that there is an analog of the quantum Hamilton operator in the context of classical statistical mechanics. This operator is the so-called transfer matrix. The Hamilton operator induces infinitesimal translations in time. In classical statistical mechanics, on the other hand, the analog of continuous time is a 1-

| Quantum mechanics | Classical statistical mechanics |
|---|---|
| Euclidean time lattice | $d$-dimensional spatial lattice |
| elementary time step $a$ | crystal lattice spacing |
| particle position $x$ | classical spin variable $s$ |
| particle path $x(t)$ | spin configuration $s_x$ |
| path integral $\int \mathcal{D}x$ | sum over configurations $\prod_x \sum_{s_x}$ |
| Euclidean action $S_E[x]$ | classical Hamilton function $\mathcal{H}[s]$ |
| Planck's constant $\hbar$ | temperature $T$ |
| quantum fluctuations | thermal fluctuations |
| kinetic energy $\frac{1}{2}(\frac{x_{i+1}-x_i}{a})^2$ | neighbor coupling $s_x s_{x+1}$ |
| potential energy $V(x_i)$ | external field energy $\mu B s_x$ |
| weight of a path $\exp(-\frac{1}{\hbar}S_E[x])$ | Boltzmann factor $\exp(-\mathcal{H}[s]/T)$ |
| vacuum expectation value $\langle \mathcal{O}(x) \rangle$ | magnetization $\langle s_x \rangle$ |
| 2-point function $\langle \mathcal{O}(x(0))\mathcal{O}(x(t)) \rangle$ | correlation function $\langle s_x s_y \rangle$ |
| energy gap $E_1 - E_0$ | inverse correlation length $1/\xi$ |
| continuum limit $a \to 0$ | critical behavior $\xi \to \infty$ |

Table 2.2: *The dictionary that translates quantum mechanics into the language of classical statistical mechanics.*

dimensional spatial lattice. Hence, the transfer matrix cannot induce infinitesimal space translations. Instead it induces translations by the smallest possible distance — namely by one lattice spacing. For a quantum mechanical system the transfer matrix transports us by one lattice spacing in Euclidean time, and it is given by

$$T = \exp(-\frac{a}{\hbar}H). \tag{2.6.1}$$

Now we want to construct the transfer matrix for the 1-dimensional Ising model without an external magnetic field. The corresponding partition function is given by

$$Z = \prod_x \sum_{s_x=\pm 1} \exp(\beta J \sum_x s_x s_{x+1}). \tag{2.6.2}$$

The transfer matrix obeys

$$Z = \mathrm{Tr} T^N, \tag{2.6.3}$$

where $N$ is the number of lattice points, and its matrix elements are given by the Boltzmann factor corresponding to a nearest neighbor pair by

$$\langle s_{x+1}|T|s_x\rangle = \exp(\beta J s_x s_{x+1}). \tag{2.6.4}$$

This is a $2 \times 2$ matrix. The eigenvalues of the transfer matrix can be written as $\exp(-E_0)$ and $\exp(-E_1)$. The energy gap then determines the inverse correlation length as

$$1/\xi = E_1 - E_0. \tag{2.6.5}$$

It is instructive to compute $\xi$ as a function of $\beta$ to locate the critical point of the 1-d Ising model.

Here we will do the corresponding calculation for the 1-d XY model. In the XY model the spins are unit vectors $(\cos\varphi_x, \sin\varphi_x)$ in the XY-plane that are attached to the points $x$ of a $d$-dimensional lattice. Here we consider $d = 1$, i.e. we study a chain of XY-spins. The standard Hamilton function of the XY model is given by

$$\mathcal{H}[\varphi] = J \sum_{\langle xy\rangle} (1 - \cos(\varphi_{x+1} - \varphi_x)). \tag{2.6.6}$$

In complete analogy to the Ising model the transfer matrix is now given by

$$\langle \varphi_{x+1}|T|\varphi_x\rangle = \exp(-\beta J(1 - \cos(\varphi_{x+1} - \varphi_x))), \tag{2.6.7}$$

which is a matrix with an uncountable number of rows and columns, because there is a continuum of values for $\varphi_x$ and $\varphi_{x+1}$. Still, we can ask about the eigenvalues

of this matrix. For this purpose we consider the Fourier representation

$$\langle\varphi_{x+1}|T|\varphi_x\rangle = \sum_{m\in\mathbf{Z}} \langle\varphi_{x+1}|m\rangle \exp(-\beta J) I_m(\beta J) \langle m|\varphi_x\rangle, \tag{2.6.8}$$

where

$$\langle\varphi_x|m\rangle = \exp(im\varphi_x), \tag{2.6.9}$$

are the eigenvectors of the transfer matrix. The eigenvalues are given in terms of modified Bessel functions

$$\exp(-E_m) = \exp(-\beta J) I_m(\beta J). \tag{2.6.10}$$

The energy gap between the ground state and an excited state is given by

$$E_m - E_0 = \log \frac{I_0(\beta J)}{I_m(\beta J)}, \tag{2.6.11}$$

which is non-zero for finite $\beta$. In the zero temperature limit $\beta \to \infty$ we have

$$\frac{I_0(\beta J)}{I_m(\beta J)} \sim 1 + \frac{m^2}{2\beta J}, \tag{2.6.12}$$

such that

$$\xi = 1/(E_1 - E_0) \sim 2\beta J \to \infty. \tag{2.6.13}$$

Hence, there is a critical point at zero temperature. In the language of quantum mechanics this implies the continuum limit of a Euclidean lattice theory corresponding to a quantum mechanical problem. In the continuum limit the energies corresponding to the eigenvalues of the transfer matrix take the form

$$E_m - E_0 \sim \frac{m^2}{2\beta J}. \tag{2.6.14}$$

These energies are in lattice units (the lattice spacing was put to 1). Hence, to extract physics we need to consider energy ratios and we find

$$\frac{E_m - E_0}{E_1 - E_0} \sim m^2. \tag{2.6.15}$$

These are the appropriate energy ratios of a quantum rotor — a particle that moves on a circle. Indeed the XY-spins describe an angle, which can be interpreted as the position of the quantum particle. Also the eigenvectors of the

transfer matrix are just the energy eigenfunctions of a quantum rotor. Hence, we just solved the Schrödinger equation with a discrete Euclidean time step using the transfer matrix instead of the Hamilton operator. The fact that energy ratios approach physically meaningful constants in the continuum limit is known as scaling. Of course, the discretization introduces an error as long as we are not in the continuum limit. For example, at finite $\beta$ the energy ratio is

$$\frac{E_m}{E_1} = \frac{\log(I_0(\beta J)/I_m(\beta J))}{\log(I_0(\beta J)/I_1(\beta J))}, \tag{2.6.16}$$

which is different from the continuum answer $m^2$. This cut-off effect due to a finite lattice spacing is known as a scaling violation.

There are formulations of the path integral that are free of cut-off effects and hence exhibit perfect scaling. The actions in these path integrals are known as perfect actions. As an example of a perfect action let us consider the 1-d XY model with the so-called Villain action. Then the transfer matrix is modified to

$$\langle \varphi_{x+1}|T|\varphi_x\rangle = \sum_{n\in\mathbf{Z}} \exp(-\frac{\beta J}{2}(\varphi_{x+1} - \varphi_x + 2\pi n)^2). \tag{2.6.17}$$

By a Fourier transformation we now find

$$\langle \varphi_{x+1}|T|\varphi_x\rangle = \sum_{m\in\mathbf{Z}} \langle \varphi_{x+1}|m\rangle \exp(-\frac{m^2}{2\beta J})\langle m|\varphi_x\rangle, \tag{2.6.18}$$

which leads to the same eigenvectors but to new eigenvalues

$$E_m = \frac{m^2}{2\beta J}, \tag{2.6.19}$$

which lead to the exact continuum ratio $(E_m - E_0)/(E_1 - E_0) = m^2$ even at finite temperature, i.e. even away from the continuum limit. Perfect actions are of great importance, because they yield exact continuum physics although the lattice spacing is still finite. In general, it is difficult to construct perfect actions for interacting field theories. Later we will see how perfect actions can be constructed for simple free field theories.

## 2.7 Lattice Scalar Field Theory

So far we have restricted ourselves to quantum mechanical problems and to classical statistical mechanics. The former were defined by a path integral on a 1-d

Euclidean time lattice, while the latter involved spin models on a $d$-dimensional spatial lattice. When we quantize field theories on the lattice, we formulate the theory on a $d$-dimensional space-time lattice, i.e. usually the lattice is 4-dimensional. Just as we integrate over all configurations (all paths) $x(t)$ of a quantum particle, we now integrate over all configurations $\phi(x)$ of a quantum field defined at any Euclidean space-time point $x = (\vec{x}, x_d)$. Again the weight factor in the path integral is given by the action. Let us illustrate this for a free neutral scalar field $\phi(x) \in R$. Its Euclidean action is given by

$$S_E[\phi] = \int d^d x \, [\frac{1}{2}\partial_\mu \phi \partial_\mu \phi + \frac{m^2}{2}\phi^2]. \tag{2.7.1}$$

Interactions can be included, for example, by adding a $\frac{\lambda}{4}\phi^4$ term to the action. The Feynman path integral for this system is formally written as

$$Z = \int \mathcal{D}\phi \exp(-S_E[\phi]). \tag{2.7.2}$$

(Note that we have put $\hbar = c = 1$.) The integral is over all field configurations, which is a divergent expression if no regularization is imposed. One can make the expression mathematically well-defined by using dimensional regularization of Feynman diagrams. This approach is, however, limited to perturbation theory. The lattice allows us to formulate field theory beyond perturbation theory, which is very essential for strongly interacting theories like QCD, but also for the standard model in general. For example, due to the heavy mass of the top quark, the Yukawa coupling between the Higgs and top quark field is rather strong. The above free scalar field theory, of course, does not really require a non-perturbative treatment. We use it only to illustrate the lattice quantization method in a simple setting. On the lattice the continuum field $\phi(x)$ is replaced by a lattice field $\Phi_x$, which is restricted to the points $x$ of a $d$-dimensional space-time lattice with spacing $a$. The above continuum action can be approximated by discretizing the continuum derivatives such that

$$S_E[\Phi] = a^d \sum_{x,\mu} \frac{1}{2a}(\Phi_{x+\hat{\mu}} - \Phi_x)^2 + a^d \sum_x \frac{m^2}{2}\Phi_x^2. \tag{2.7.3}$$

Here $\hat{\mu}$ is a vector of length $a$ in the $\mu$-direction. The integral over all field configurations now becomes a multiple integral over all values of the field at all lattice points

$$Z = \prod_x \int_{-\infty}^{\infty} d\Phi_x \exp(-S_E[\Phi]). \tag{2.7.4}$$

For a free field theory the partition function is just a Gaussian integral. In fact, one can write the lattice action as

$$S_E[\Phi] = \frac{1}{2}\sum_{x,y} \Phi_x \mathcal{M}_{xy} \Phi_y, \tag{2.7.5}$$

where the matrix $\mathcal{M}$ describes the couplings between lattice points. Diagonalizing this matrix by a unitary transformation $\mathcal{U}$ one has

$$\mathcal{M} = \mathcal{U}^\dagger \mathcal{D} \mathcal{U}. \tag{2.7.6}$$

Introducing

$$\Phi'_x = \mathcal{U}_{xy}\Phi_y \tag{2.7.7}$$

one obtains

$$Z = \prod_x \int d\Phi'_x \exp(-\frac{1}{2}\sum_x \Phi'_x \mathcal{D}_{xx} \Phi'_x) = (2\pi)^{N/2} \mathrm{det}\mathcal{D}^{-1/2}, \tag{2.7.8}$$

where $N$ is the number of lattice points.

To extract the energy values of the corresponding quantum Hamilton operator we need to study the 2-point function of the lattice field

$$\langle \Phi_x \Phi_y \rangle = \frac{1}{Z}\int \mathcal{D}\Phi \ \Phi_x \Phi_y \exp(-S_E[\Phi]). \tag{2.7.9}$$

This is most conveniently done by introducing a source field in the partition function, such that

$$Z[J] = \int \mathcal{D}\Phi \exp(-S_E[\Phi] + a^d \sum_x J_x \Phi_x). \tag{2.7.10}$$

Then the connected 2-point function is given by

$$\langle \Phi_x \Phi_y \rangle - \langle \Phi \rangle^2 = \frac{\partial^2 \log Z[J]}{\partial J_x \partial J_y}|_{J=0}. \tag{2.7.11}$$

The Boltzmann factor characterizing the problem with the external sources is given by the exponent

$$\frac{1}{2}\Phi \mathcal{M} \Phi - J\Phi = \frac{1}{2}\Phi' \mathcal{M} \Phi' - \frac{1}{2} J \mathcal{M}^{-1} J. \tag{2.7.12}$$

Here we have introduced

$$\Phi' = \Phi - \mathcal{M}^{-1}J. \tag{2.7.13}$$

Integrating over $\Phi'$ in the path integral we obtain

$$Z[J] = (2\pi)^{N/2}\det\mathcal{D}^{-1/2}\exp(\frac{1}{2}J\mathcal{M}^{-1}J), \tag{2.7.14}$$

and hence

$$\langle\Phi_x\Phi_y\rangle = \mathcal{M}^{-1}_{xy}. \tag{2.7.15}$$

It is instructive to invert the matrix $\mathcal{M}$ by going to Fourier space, i.e. by writing

$$\Phi_x = \frac{1}{(2\pi)^d}\int_B d^dp\ \Phi(p)\exp(ipx). \tag{2.7.16}$$

The momentum space of the lattice is given by the Brillouin zone $B = ]-\pi/a, \pi/a]^d$. For the 2-point function in momentum space one then finds

$$\langle\Phi(-p)\Phi(p)\rangle = [\sum_\mu(2\sin(p_\mu a/2))^2 + m^2]^{-1}. \tag{2.7.17}$$

This is the lattice version of the continuum propagator

$$\langle\Phi(-p)\Phi((p)\rangle = (p^2 + m^2)^{-1}. \tag{2.7.18}$$

From the lattice propagator we can deduce the energy spectrum of the lattice theory. For this purpose we construct a lattice field with definite spatial momentum $\vec{p}$ located in a specific time slice

$$\Phi(\vec{p})_t = \sum_x \Phi_{\vec{x},t}\exp(-i\vec{p}\cdot\vec{x}), \tag{2.7.19}$$

and we consider its 2-point function

$$\langle\Phi(-\vec{p})_0\Phi(\vec{p})_t\rangle = \frac{1}{2\pi}\int_{-\pi/a}^{\pi/a} dp_d\langle\Phi(-p)\Phi(p)\rangle\exp(ip_dt). \tag{2.7.20}$$

Inserting the lattice propagator of eq.(2.7.17) one can perform the integral. One encounters a pole in the propagator when $p_d = iE$ with

$$(2\sinh(Ea/2))^2 = \sum_i(2\sin(p_ia/2))^2 + m^2. \tag{2.7.21}$$

The 2-point function then takes the form

$$\langle \Phi(-\vec{p})_0 \Phi(\vec{p})_t \rangle = C \exp(-Et), \tag{2.7.22}$$

i.e. it decays exponentially with slope $E$. This allows us to identify $E$ as the energy of the lattice scalar particle with spatial momentum $\vec{p}$. In general, $E$ differs from the correct continuum dispersion relation

$$E^2 = \vec{p}^2 + m^2. \tag{2.7.23}$$

Only in the continuum limit, i.e. when $E$, $\vec{p}$ and $m$ are small in lattice units, the lattice dispersion relation agrees with the one of the continuum theory. Again, we observe scaling violations, i.e. deviations from continuum results, as long as we are not in the continuum limit.

# Chapter 3

# Symmetries of the Strong Interactions

In this chapter we review some aspects of symmetries in the continuum formulation of QCD with an emphasis on chiral symmetry.

## 3.1 $SU(N_c)$ Yang-Mills Theory in the Continuum

Let us consider an anti-Hermitean non-Abelian $SU(N_c)$ gauge field

$$A_\mu(x) = igA_\mu^a(x)T^a, \tag{3.1.1}$$

which (for $N_c = 3$) describes the gluons of QCD. Here $g$ is the gauge coupling, $A_\mu^a(x)$ (with $a \in \{1, 2, ..., N_c^2 - 1\}$) is the real-valued non-Abelian vector potential at the Euclidean space-time point $x$, and the $T^a$ (which obey $\mathrm{Tr}(T^aT^b) = \frac{1}{2}\delta_{ab}$) are the Hermitean generators of the $SU(N_c)$ algebra. The algebra-valued field strength takes the form

$$F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) + [A_\mu(x), A_\nu(x)], \tag{3.1.2}$$

and the corresponding Euclidean Yang-Mills action is given by

$$S_{YM}[A] = \int d^4x \ \frac{1}{2g^2}\mathrm{Tr}(F_{\mu\nu}F_{\mu\nu}). \tag{3.1.3}$$

The action is invariant under group-valued gauge transformations $\Omega(x) \in SU(N_c)$,

$$A'_\mu(x) = \Omega(x)(A_\mu(x) + \partial_\mu)\Omega(x)^\dagger, \tag{3.1.4}$$

under which the field strength transforms as

$$F_{\mu\nu}(x) = \Omega(x)F_{\mu\nu}(x)\Omega(x)^\dagger. \tag{3.1.5}$$

The quantum theory is defined by a functional integral over all gluon fields

$$Z = \int \mathcal{D}A \ \exp(-S_{YM}[A]), \tag{3.1.6}$$

which is a formal expression before it is properly regularized. In perturbation theory this is possible using standard dimensional regularization techniques. Through the regularization, a scale is introduced into the quantum theory which explicitly breaks the scale invariance of the classical Yang-Mills theory. This anomaly in the scale invariance is responsible for the phenomenon of dimensional transmutation: in the quantum theory the dimensionless coupling constant $g$ of the classical theory is traded for a dimensionful scale. In the modified minimal subtraction renormalization scheme this scale is $\Lambda_{\overline{MS}}$ which is defined in the framework of perturbation theory. We will soon define the theory beyond perturbation theory by regularizing it on a space-time lattice. In a non-perturbative context, a natural scale is the dynamically generated mass gap $M$ — the energy of the lowest state above the vacuum. In a Yang-Mills theory this state is the lightest glueball. The $SU(N_c)$ Yang-Mills theory is a quantum theory without any free parameter. For example, the dimensionless ratio $M/\Lambda_{\overline{MS}}$ is a pure number predicted by the theory. The relation of $M$ or $\Lambda_{\overline{MS}}$ to units like GeV, on the other hand, is, of course, not predicted by the theory. Such man-made mass units are related to the kilogram, defined by the arbitrary amount of platinum-iridium alloy deposited near Paris a long time ago.

Another quantity of physical interest is the topological charge

$$Q[A] = -\frac{1}{32\pi^2}\int d^4x \ \varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}(F_{\mu\nu}F_{\rho\sigma}) \in \Pi_3[SU(N_c)] = \mathbf{Z}, \tag{3.1.7}$$

which takes integer values in the third homotopy group of the gauge group. The topological charge gives rise to an additional parameter, the vacuum angle $\theta$, in the Yang-Mills functional integral

$$Z(\theta) = \int \mathcal{D}A \ \exp(-S_{YM}[A] + i\theta Q[A]). \tag{3.1.8}$$

For $\theta \neq 0$ or $\pi$ the $\theta$-term explicitly breaks parity as well as CP. The bound $|\theta| < 10^{-9}$ derived from the measurement of the electric dipole moment of the neutron suggests that $\theta = 0$ in Nature. This result is puzzling because in the Standard Model CP is already explicitly broken by the complex phase of the Cabbibo-Kobayashi-Maskawa matrix. The puzzle to understand why $\theta = 0$ is known as the strong CP problem.

## 3.2 QCD with $N_f$ Quark Flavors

In the next step we add $N_f$ massless quarks to the pure gluon theory. Quarks and anti-quarks are described by anti-commuting Dirac spinor fields $\psi(x)$ and $\overline{\psi}(x)$. In Euclidean space-time these two fields represent independent Grassmann degrees of freedom. Under a non-Abelian gauge transformation the quark and anti-quark fields transform in the fundamental representations $\{N_c\}$ and $\{\overline{N_c}\}$, respectively, i.e.

$$\psi(x)' = \Omega(x)\psi(x),\ \overline{\psi}(x)' = \overline{\psi}(x)\Omega(x)^\dagger. \tag{3.2.1}$$

The fermionic part of the Euclidean action of massless QCD takes the form

$$S_F[\overline{\psi}, \psi, A] = \int d^4x\ \overline{\psi}\gamma_\mu(\partial_\mu + A_\mu)\psi, \tag{3.2.2}$$

which is gauge invariant by construction. The Euclidean Dirac matrices are Hermitean and obey the anti-commutation relations

$$\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu},\ \{\gamma_\mu, \gamma_5\} = 0,\ \gamma_5 = \gamma_1\gamma_2\gamma_3\gamma_4. \tag{3.2.3}$$

We now decompose the quark fields into left- and right-handed components

$$\begin{aligned} &\psi_L(x) = P_L\psi(x),\ \psi_R(x) = P_R\psi(x),\ \psi(x) = \psi_L(x) + \psi_R(x), \\ &\overline{\psi}_L(x) = \overline{\psi}(x)P_R,\ \overline{\psi}_R(x) = \overline{\psi}(x)P_L,\ \overline{\psi}(x) = \overline{\psi}_L(x) + \overline{\psi}_R(x). \end{aligned} \tag{3.2.4}$$

The chiral projectors are given by

$$P_R = \frac{1+\gamma_5}{2},\ P_L = \frac{1-\gamma_5}{2}. \tag{3.2.5}$$

Inserting the decomposed spinors into the fermionic part of the action one obtains

$$S_F[\overline{\psi}, \psi, A] = \int d^4x\ \left[\overline{\psi}_L\gamma_\mu(\partial_\mu + A_\mu)\psi_L + \overline{\psi}_R\gamma_\mu(\partial_\mu + A_\mu)\psi_R\right], \tag{3.2.6}$$

i.e. the action decouples into two contributions from left- and right-handed quarks.

As a result, the action of massless QCD is invariant against $U(N_f)_L \otimes U(N_f)_R$ chiral transformations

$$\begin{aligned} \psi'_L(x) &= L\ \psi_L(x),\ \overline{\psi}'(x) = \overline{\psi}_L(x)L^+,\ L \in U(N_f)_L, \\ \psi'_R(x) &= R\ \psi_R(x),\ \overline{\psi}'(x) = \overline{\psi}_R(x)R^+,\ R \in U(N_f)_R. \end{aligned} \tag{3.2.7}$$

Due to an anomaly in the axial $U(1)_A$ symmetry, the symmetry of the quantum theory is reduced to $SU(N_f)_L \otimes SU(N_f)_R \otimes U(1)_B$ where the $U(1)_B = U(1)_{L=R}$ symmetry represents baryon number conservation.

Chiral symmetry is only approximate in Nature, because the quark mass terms couple left- and right-handed fermions. The mass terms in the QCD action take the form

$$S_M[\overline{\psi}, \psi] = \int d^4x\ \left[\overline{\psi}_R \mathcal{M} \psi_L + \overline{\psi}_L \mathcal{M}^\dagger \psi_R\right], \tag{3.2.8}$$

which is again gauge invariant but no longer chirally invariant. The quark mass matrix takes the form

$$\mathcal{M} = \mathrm{diag}(m_u, m_d, m_s, ..., m_{N_f}). \tag{3.2.9}$$

If all quark masses are equal, i.e. if $\mathcal{M} = m\mathbb{1}$, the mass term is invariant only against simultaneous transformations $L = R$. Hence, chiral symmetry is then explicitly broken down to

$$SU(N_f)_{L=R} \otimes U(1)_{L=R} = SU(N_f)_F \otimes U(1)_B, \tag{3.2.10}$$

which corresponds to the flavor and baryon number symmetry. In Nature the quark masses are all different, and the symmetry is in fact explicitly broken down to

$$\prod_{f=1}^{N_f} U(1)_f = U(1)_u \otimes U(1)_d \otimes U(1)_s \otimes ... \otimes U(1)_{N_f}. \tag{3.2.11}$$

The physical up and down quark masses are a lot smaller than $\Lambda_{\overline{MS}}$ while the strange quark mass is of the order of $\Lambda_{\overline{MS}}$. Consequently, $SU(2)_L \otimes SU(2)_R$ is a very good approximate global symmetry, while $SU(3)_L \otimes SU(3)_R$ is broken more strongly. It should be noted that the actual values of the quark masses are reasonably well known from comparison with experiment, but are at present not at all understood theoretically. In particular, we don't know why there are three light quark flavors. Before one understands the relevant physics beyond the

Standard Model, the origin of the chiral symmetry of QCD remains mysterious and the symmetry itself seems accidental.

The total action of QCD is simply given by

$$S_{QCD}[\overline{\psi},\psi,A] = S_{YM}[A] + S_F[\overline{\psi},\psi,A] + S_M[\overline{\psi},\psi], \tag{3.2.12}$$

and the corresponding QCD functional integral is

$$Z = \int \mathcal{D}\overline{\psi}\mathcal{D}\psi\mathcal{D}A \ \exp(-S_{QCD}[\overline{\psi},\psi,A]). \tag{3.2.13}$$

Again, this is a formal mathematical expression before it is properly regularized. In the continuum this can be done only perturbatively. We will soon discuss the lattice regularization which defines QCD non-perturbatively.

## 3.3 The Axial Anomaly and the Index Theorem

The $U(1)_A$ symmetry of the classical action of massless QCD is explicitly broken by quantum effects. As a consequence of this anomaly the flavor-singlet axial current

$$j^5_\mu(x) = \overline{\psi}(x)\gamma_\mu\gamma_5\psi(x), \tag{3.3.1}$$

which is conserved at the classical level, has a non-zero divergence

$$\partial_\mu j^5_\mu(x) = -\frac{N_f}{32\pi^2}\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[F_{\mu\nu}(x)F_{\rho\sigma}(x)], \tag{3.3.2}$$

due to instantons (and other topological charge carriers) in the quantum theory. In particular, the variation of the axial charge $Q^5(t) = \int d^3x \ j_0(\vec{x},t)$ is given by

$$Q^5(t=\infty) - Q^5(t=-\infty) = N_f \ Q[A], \tag{3.3.3}$$

where $Q[A]$ is the topological charge of eq.(3.1.7). Only in the $N_c \to \infty$ limit the anomaly vanishes and the chiral symmetry of massless QCD is enhanced to the full $U(N_f)_L \otimes U(N_f)_R$ group.

The axial anomaly is deeply connected with the Atiyah-Singer index theorem, which relates the zero-modes of the massless Dirac operator $D[A] = \gamma_\mu(\partial_\mu + A_\mu)$

to the topological charge $Q[A]$. The eigenvalues of the Dirac operator are purely imaginary and come in complex conjugate pairs. Only the zero eigenvalues are not paired. Since the Dirac operator anti-commutes with $\gamma_5$, the eigenvectors of the zero-modes (which obey $D[A]\psi = 0$) have a definite handedness, i.e. $\gamma_5\psi = \pm\psi$. The index theorem states that

$$\text{index}(D[A]) = n_+ - n_- = N_f \; Q[A], \tag{3.3.4}$$

i.e. the index of the operator $D[A]$, which is defined as the difference between the number of left- and right-handed zero-modes, is given by the topological charge.

As a consequence of the index theorem, topologically non-trivial gluon field configurations (with $Q[A] \neq 0$) necessarily induce zero-modes in the Dirac operator and thus lead to a vanishing fermion determinant $\text{det}D[A] = 0$. As a function of the vacuum angle $\theta$, the functional integral of massless QCD takes the form

$$\begin{aligned} Z(\theta) &= \int \mathcal{D}\overline{\psi}\mathcal{D}\psi\mathcal{D}A \; \exp(-S_{QCD}[\overline{\psi}, \psi, A] + i\theta Q[A]) \\ &= \int \mathcal{D}A \; \text{det}D[A] \exp(-S_{YM}[A] + i\theta Q[A]) \\ &= \int \mathcal{D}A \; \text{det}D[A] \exp(-S_{YM}[A]) = Z(0). \end{aligned} \tag{3.3.5}$$

Since $\text{det}D[A] = 0$ when $Q[A] \neq 0$, there are no $\theta$-vacuum effects in massless QCD. This would "solve" the strong CP problem (why is $\theta = 0$ ?) if, for example, the up quark would be massless. Of course, this would leave us with the "up quark problem": why should $m_u = 0$ ? In any case, $m_u = 0$ seems not to be realized in Nature and the strong CP problem remains puzzling.

## 3.4 Spontaneous Chiral Symmetry Breaking

Due to the approximate chiral symmetry of QCD one would expect corresponding near degeneracies in the spectrum of strongly interacting particles. Indeed, hadrons can be classified as isospin multiplets. The isospin transformations act on left- and right-handed fermions simultaneously, i.e. $SU(2)_I = SU(2)_{L=R}$. The $SU(3)_F = SU(3)_{L=R}$ flavor symmetry is more approximate but is still clearly visible in the spectrum. The full $SU(N_f)_L \otimes SU(N_f)_R \otimes U(1)_B$ chiral symmetry, on the other hand, is not manifest in the spectrum at all. In particular, one

does not observe mass-degenerate parity doublets of hadrons, as one should if chiral symmetry was manifest in the spectrum. Furthermore, one observes very light pseudo-scalar particles — the pions $\pi^+$, $\pi^0$, and $\pi^-$ — as well as somewhat heavier pseudo-scalars — the four kaons $K^+$, $K^0$, $\overline{K^0}$, $K^-$ and the $\eta$-meson.

From the experimental evidence one concludes that chiral symmetry must be spontaneously broken. Indeed, when a continuous global symmetry breaks spontaneously, massless Goldstone bosons appear in the spectrum. According to Goldstone's theorem, the number of massless bosons is given by the difference of the number of generators of the full symmetry group $G$ and the subgroup $H$ that remains unbroken. In massless QCD the full chiral symmetry group is

$$G = SU(N_f)_L \otimes SU(N_f)_R \otimes U(1)_B, \tag{3.4.1}$$

while the unbroken subgroup is the flavor symmetry

$$H = SU(N_f)_{L=R} \otimes U(1)_B. \tag{3.4.2}$$

Hence, in this case one expects $N_f^2 - 1$ massless Goldstone bosons. For $N_f = 2$ these are the three pions, while for $N_f = 3$ there are eight Goldstone bosons — the pions, the kaons, and the $\eta$-meson. In Nature these particles are not exactly massless, because chiral symmetry is explicitly broken by the quark masses. The masses of the up and down quarks are much smaller than the QCD scale $\Lambda_{\overline{MS}}$ which leads to the very small pion mass. The mass of the strange quark, on the other hand, is of the order of $\Lambda_{\overline{MS}}$, thus leading to larger masses of the kaons and the $\eta$-meson. Still, their masses are small enough to identify these particles as pseudo-Goldstone bosons.

Chiral symmetry breaking has not yet been derived analytically from the QCD Lagrangian. In particular, spontaneous chiral symmetry breaking is a non-perturbative phenomenon whose understanding requires a formulation of QCD beyond perturbation theory. Such a formulation is provided by lattice field theory which will be discussed below. Indeed, numerical simulations in lattice QCD confirm that chiral symmetry is spontaneously broken. For example, one detects spontaneous chiral symmetry breaking by investigating the chiral order parameter

$$\langle\overline{\psi}\psi\rangle = \langle 0|\overline{\psi}(x)\psi(x)|0\rangle = \langle 0|\overline{\psi}_R(x)\psi_L(x) + \overline{\psi}_L(x)\psi_R(x)|0\rangle. \tag{3.4.3}$$

The order parameter is invariant against simultaneous transformations $R = L$, but not against general chiral rotations. If chiral symmetry would be intact the chiral condensate would vanish. When the symmetry is spontaneously broken, on the other hand, $\langle\overline{\psi}\psi\rangle$ is non-zero.

# Chapter 4

# Free Lattice Fermions

In this chapter we begin to formulate QCD on a space-time lattice which serves as an ultraviolet regulator. We replace Euclidean space-time by a hypercubic lattice of points $x$ with lattice spacing $a$. The lattice provides an ultraviolet momentum cut-off $1/a$. The continuum limit is reached when $a \to 0$. One has a lot of freedom in writing down a lattice regularized theory. In order to ensure that one reaches the desired theory in the continuum limit one must pay attention to the relevant symmetries. The most important symmetry of QCD is the $SU(N_c)$ gauge invariance. It is an important strength of the lattice regularization that it manifestly respects gauge invariance. The fact that space-time symmetries are explicitly broken down to discrete translations and the hypercubic rotation group of the lattice is not a severe problem. In particular, in QCD the hypercubic symmetry is powerful enough to ensure that the full Poincaré symmetry of the continuum is automatically recovered as $a \to 0$. Discrete symmetries like parity and charge conjugation are also easy to maintain on the lattice. This review concentrates on the question of how to realize chiral symmetry on the lattice. In this chapter we consider lattice theories of free quarks only. Gluon fields will be added in the next chapter.

## 4.1 Fermionic Path Integrals and Grassmann Algebras

We have defined the path integral by using the classical action. Theories with fermions have no immediate classical limit, and the definition of the path integral needs special care. The first step is to define a so-called Grassmann algebra, which works with anticommuting classical variables $\eta_i$ with $i \in 1, 2, ..., N$. A Grassmann algebra is characterized by the anticommutation relations

$$\{\eta_i, \eta_j\} = \eta_i \eta_j + \eta_j \eta_i = 0. \tag{4.1.1}$$

An element of the Grassmann algebra is a polynomial in the generators

$$f(\eta) = f + \sum_i f_i \eta_i + \sum_{ij} f_{ij} \eta_i \eta_j + \sum_{ijk} f_{ijk} \eta_i \eta_j \eta_k + ... \tag{4.1.2}$$

The $f_{ij...l}$ are ordinary complex (or sometimes real) numbers, which are antisymmetric in $i, j, ..., l$. One also defines formal differentiation and integration procedures. The differentiation rules are

$$\frac{\partial}{\partial \eta_i} \eta_i = 1, \ \frac{\partial}{\partial \eta_i} \eta_i \eta_j = \eta_j, \ \frac{\partial}{\partial \eta_i} \eta_j \eta_i = -\eta_j, \tag{4.1.3}$$

and integration is defined by

$$\int d\eta_i = 0, \ \int d\eta_i \ \eta_i = 1, \ \int d\eta_i d\eta_j \ \eta_i \eta_j = -1. \tag{4.1.4}$$

These integrals are formal expressions. One should not ask over which range of $\eta_i$ we actually integrate.

The Grassmann algebra we use to define fermion fields is generated by Grassmann numbers $\Psi_x$ and $\bar{\Psi}_x$, which are completely independent. The index $x$ runs over all space-time points as well as over all spin, flavor or color indices. Let us consider the simplest (completely unrealistic) case of just two degrees of freedom $\Psi$ and $\bar{\Psi}$, and let us perform the Gaussian integral

$$\int d\bar{\Psi} d\Psi \ \exp(-m\bar{\Psi}\Psi) = \int d\bar{\Psi} d\Psi \ (1 - m\bar{\Psi}\Psi) = m. \tag{4.1.5}$$

Note that the expansion of the exponential terminates because $\Psi^2 = \bar{\Psi}^2 = 0$. When we enlarge the Grassmann algebra to an arbitrary number of elements the

above formula generalizes to

$$\prod_x \int d\bar{\Psi}_x d\Psi_x \ \exp(-\bar{\Psi}_x \mathcal{M}_{xy} \Psi_y) = \int \mathcal{D}\bar{\Psi}\mathcal{D}\Psi \ \exp(-\bar{\Psi}\mathcal{M}\Psi) = \det\mathcal{M}. \quad (4.1.6)$$

In the two variable case we have

$$\int d\bar{\Psi} d\Psi \ \bar{\Psi}\Psi \exp(-m\bar{\Psi}\Psi) = 1, \quad (4.1.7)$$

which generalizes to

$$\int \mathcal{D}\bar{\Psi}\mathcal{D}\Psi \ \bar{\Psi}_x \Psi_y \exp(-\bar{\Psi}\mathcal{M}\Psi_y) = \mathcal{M}^{-1}_{ij} \det\mathcal{M}. \quad (4.1.8)$$

## 4.2 Naive Lattice Fermions and the Doubling Problem

In the continuum the Euclidean action of a free Dirac fermion in $d$ space-time dimensions is given by

$$S[\overline{\psi}, \psi] = \int d^d x \ \overline{\psi}(\gamma_\mu \partial_\mu + m)\psi, \quad (4.2.1)$$

and the functional integral takes the form

$$Z = \int \mathcal{D}\overline{\psi}\mathcal{D}\psi \ \exp(-S[\overline{\psi}, \psi]). \quad (4.2.2)$$

On the lattice the continuum fermion field $\overline{\psi}(x), \psi(x)$ is replaced by Grassmann variables $\bar{\Psi}_x, \Psi_x$ which live on the lattice points $x$. The continuum derivative can be discretized by a finite difference, such that

$$S[\bar{\Psi}, \Psi] = a^d \ \bar{\Psi} D \Psi = a^d \sum_{x,\mu} \frac{1}{2a} (\bar{\Psi}_x \gamma_\mu \Psi_{x+\hat{\mu}} - \bar{\Psi}_{x+\hat{\mu}} \gamma_\mu \Psi_x) + a^d \sum_x m \bar{\Psi}_x \Psi_x. \quad (4.2.3)$$

Here $\hat{\mu}$ is a vector of length $a$ in the $\mu$-direction. In the continuum limit $a \to 0$ the lattice sum $a^d \sum_x$ becomes the continuum integral $\int d^d x$ over space-time. The corresponding lattice Dirac operator which is a matrix in the Dirac- and space-time indices takes the form

$$D_{x,y} = \sum_\mu \frac{1}{2a} (\gamma_\mu \delta_{x+\hat{\mu},y} - \gamma_\mu \delta_{x-\hat{\mu},y}) + m\delta_{x,y}. \quad (4.2.4)$$

The lattice functional integral can be written as

$$Z = \int \mathcal{D}\bar{\Psi}\mathcal{D}\Psi \ \exp(-a^d \ \bar{\Psi} D \Psi) = \prod_x \int d\bar{\Psi}_x d\Psi_x \ \exp(-S[\bar{\Psi}, \Psi]). \tag{4.2.5}$$

In particular, the fermionic Grassmann integration measure is completely regularized explicitly.

The momentum space of the lattice theory is a $d$-dimensional Brillouin zone $B = [-\pi/a, \pi/a]^d$ with periodic boundary conditions. Going to momentum space, the naive fermion action from above gives rise to the lattice fermion propagator

$$\langle \bar{\Psi}(-p)\Psi(p)\rangle = [i \sum_\mu \gamma_\mu \frac{1}{a} \sin(p_\mu a) + m]^{-1}. \tag{4.2.6}$$

By performing a Fourier transform in the Euclidean energy $p_d$ one obtains the fermion 2-point function

$$\langle \bar{\Psi}(-\vec{p}, 0)\Psi(\vec{p}, x_d)\rangle = \frac{1}{2\pi} \int_{-\pi/a}^{\pi/a} dp_d \ \langle \bar{\Psi}(-p)\Psi(p)\rangle \exp(ip_d x_d) \sim \exp(-E(\vec{p})x_d). \tag{4.2.7}$$

At large Euclidean time separation $x_d$ the 2-point function decays exponentially with the energy $E(\vec{p})$ of a fermion with spatial momentum $\vec{p}$. For the naive fermion action the lattice dispersion relation takes the form

$$\sinh^2(E(\vec{p})a) = \sum_i \sin^2(p_i a) + (ma)^2. \tag{4.2.8}$$

The continuum dispersion relation $E(\vec{p})^2 = \vec{p}^{\ 2} + m^2$ is indeed recovered in the continuum limit $a \to 0$. However, besides $\vec{p} = 0$ there are other momenta $\vec{p}$ for which $E(\vec{p})$ becomes small. These are located at the corners of the Brillouin zone where the components of the momentum vector take the values $p_i = 0$ or $\pi/a$, such that $\sin(p_i a) = 0$. As a consequence, the lattice dispersion relation leads to additional states in the spectrum which are absent in the continuum theory and which do not disappear in the continuum limit. Hence, the naive lattice fermion action does not lead to the correct continuum theory. The extra states appearing in the lattice dispersion relation show up as additional physical particles — the so-called doubler fermions. Fermion doubling is a manifestation of a deep fundamental problem of lattice regularized fermionic theories with a chiral symmetry. The fermion doubling problem leads to a multiplication of fermion species. The lattice fermion propagator of eq.(4.2.6) has $2^d$ poles instead of just

one as in the continuum. The origin of the doubling problem is deeply connected with chiral symmetry and can be traced back to the axial anomaly. The doubler fermions pose a severe problem in lattice field theory. Without removing them we cannot describe Nature's QCD (which has 3 and not $2^d = 2^4 = 16$ light quark flavors).

## 4.3 The Nielsen-Ninomiya Theorem

Before we try to eliminate the doubler fermions let us prove a general theorem due to Nielsen and Ninomiya: a chirally invariant free fermion lattice action, which is local, translation invariant, and real necessarily has fermion doubling. The theorem is based on topology. It holds because the lattice momentum space (the Brillouin zone $B$) is a torus. A general chirally symmetric and translationally invariant lattice action for free fermions takes the form

$$S[\bar{\Psi}, \Psi] = a^d \sum_{x,y} \bar{\Psi}_x \gamma_\mu \rho_\mu (x - y) \Psi_y. \tag{4.3.1}$$

The function $\rho_\mu(x - y)$ determines the strength of the coupling between the fermion field values $\bar{\Psi}_x$ and $\Psi_y$ at two points $x$ and $y$ which may be separated by an arbitrarily large distance. Locality of the lattice action does not mean that the points $x$ and $y$ must be nearest neighbors. It only means that $\rho_\mu(x - y)$ decays exponentially at large separations $x - y$. Going to momentum space, locality implies that in Fourier space $\rho_\mu(p)$ is a regular function (without poles) over the Brillouin zone. The corresponding lattice fermion propagator takes the form

$$\langle \bar{\Psi}(-p) \Psi(p) \rangle = [i \sum_\mu \gamma_\mu \rho_\mu(p)]^{-1}. \tag{4.3.2}$$

Reality and translation invariance of the lattice action imply that $\rho_\mu(p)$ is a real-valued periodic function over the Brillouin zone.

Poles of the propagator — and hence physical or doubler fermions — correspond to zeros of $\rho_\mu(p)$, i.e. to points $p$ with $\rho_\mu(p) = 0$ for all $\mu$. The Nielsen-Ninomiya theorem states that a regular, real-valued, and periodic function $\rho_\mu(p)$ necessarily vanishes at more than just one point. It is trivial to prove this for $d = 1$. In that case, there is a single regular periodic function $\rho_1(p)$ which should at least have one zero in order to describe the physical fermion pole. The function is positive on one side of the zero and negative on the other side. Hence,

it must go through zero again in order to satisfy periodicity, thus leading to a doubler fermion pole in the lattice propagator. In higher dimensions the proof is analogous. For example, for $d = 2$ there are two functions $\rho_1(p)$ and $\rho_2(p)$. The zeros of $\rho_1(p)$ lie on a closed curve in the two-dimensional Brillouin zone. This curve may be closed via the periodic boundary conditions. The zeros of $\rho_2(p)$ lie on another closed curve that intersects the first one in the pole position of the physical fermion. Due to the periodic boundary conditions of the Brillouin zone, the two curves must necessarily also intersect somewhere else. The curves cannot just touch each other because this would lead to an incorrect dispersion relation for the physical fermion. In $d$ dimensions the zeros of $\rho_\mu(p)$ (with $\mu = 1, 2, ..., d$) lie on $d$ closed $(d-1)$-dimensional surfaces. Again, those cannot intersect in just one point. If they intersect once they necessarily intersect also somewhere else. This proves lattice fermion doubling for a chirally symmetric, translation invariant, real-valued lattice action. It should be noted that the theorem does not specify the number of doubler fermions. It is indeed possible to reduce the number of doublers from $2^d - 1$ to 1, but it is impossible to eliminate the doubler fermions completely.

One may try to evade the theorem by violating one of its basic assumptions. Giving up translation invariance or the reality of the action has not led to acceptable solutions of the fermion doubling problem. Giving up locality is probably the last thing one should do in field theory. For example, the early idea of SLAC fermions turned out to be unacceptable for this reason.

## 4.4 Wilson Fermions

In his work on lattice gauge theory Wilson removed the doubler fermions in a direct and radical way by breaking chiral symmetry explicitly. Then the Nielsen-Ninomiya theorem is evaded because the propagator contains additional terms without $\gamma_\mu$. The so-called Wilson term gives the fermion doublers a mass of the order of the cut-off while the physical fermion remains massless. Hence, in the continuum limit chiral symmetry is recovered in the physical sector. Wilson's modification of the naive fermion action takes the form of a discretized second derivative

$$S[\bar{\Psi}, \Psi] \;=\; a^d \sum_{x,\mu} \frac{1}{2a} (\bar{\Psi}_x \gamma_\mu \Psi_{x+\hat{\mu}} - \bar{\Psi}_{x+\hat{\mu}} \gamma_\mu \Psi_x) + a^d \sum_x m \bar{\Psi}_x \Psi_x$$

$$+ \quad a^d \sum_{x,\mu} \frac{1}{2a}(2\bar{\Psi}_x\Psi_x - \bar{\Psi}_x\Psi_{x+\hat{\mu}} - \bar{\Psi}_{x+\hat{\mu}}\Psi_x). \tag{4.4.1}$$

Then the lattice propagator takes the form

$$\langle\bar{\Psi}(-p)\Psi(p)\rangle = [i\sum_\mu \gamma_\mu \frac{1}{a}\sin(p_\mu a) + m + \sum_\mu \frac{2}{a}\sin^2(\frac{p_\mu a}{2})]^{-1}. \tag{4.4.2}$$

The Wilson term acts as a momentum-dependent mass term. For small momenta it vanishes quadratically, and hence it does not affect the dispersion of the physical fermion, at least in the continuum limit. For the doubler fermions, on the other hand, the Wilson term is non-zero, and gives them a mass of the order of the cut-off $1/a$. In the continuum limit the doubler fermions are hence eliminated from the spectrum of the theory. Unfortunately, in lattice QCD this leads to a variety of complications. In particular, recovering chiral symmetry in the continuum limit requires unnatural fine-tuning of the bare fermion mass.

## 4.5 Perfect Lattice Fermions

In this section we relate the continuum theory of free fermions to a corresponding lattice theory by an exact renormalization group transformation. This is achieved by defining lattice fermion fields as block averages of continuum fields integrated over hypercubes. The resulting lattice theory is in all respects equivalent to the underlying continuum theory, i.e. it is completely free of lattice artifacts. For example, it has the same energy-momentum dispersion relation as the continuum theory. Even more important, it has an exact chiral symmetry (which may, however, be hidden). Lattice actions with these properties are known as perfect actions.

Let us derive a perfect fermion action by blocking from the continuum. For this purpose we average the continuum fermion field $\psi(y)$ over hypercubes $c_x$ of size $a^d$ centered at the points $x$ of a $d$-dimensional Euclidean lattice

$$\Psi_x = \frac{1}{a^d}\int_{c_x} d^d y\ \psi(y),\ \ \bar{\Psi}_x = \frac{1}{a^d}\int_{c_x} d^d y\ \overline{\psi}(y), \tag{4.5.1}$$

which in momentum space corresponds to

$$\Psi(p) = \sum_{l\in\mathbf{Z}^d} \psi(p+2\pi l/a)\Pi(p+2\pi l/a),\ \bar{\Psi}(-p) = \sum_{n\in\mathbf{Z}^d} \overline{\psi}(-p-2\pi n/a)\Pi(p+2\pi n/a), \tag{4.5.2}$$

Note that the lattice fermion field is periodic over the Brillouin zone. The Fourier transform of the blocking kernel is given by

$$\Pi(p) = \prod_{\mu=1}^{d} \frac{2\sin(p_\mu a/2)}{p_\mu a}. \tag{4.5.3}$$

The lattice fermion propagator is related to the continuum propagator by

$$\begin{aligned} \langle \bar{\Psi}(-p)\Psi(p)\rangle &= \sum_{l\in\mathbf{Z}^d} \langle \overline{\psi}(-p-2\pi l/a)\psi(p+2\pi l/a)\rangle \Pi(p+2\pi l/a)^2 \\ &= \sum_{l\in\mathbf{Z}^d} [i\gamma_\mu(p_\mu + 2\pi l_\mu/a) + m]^{-1} \Pi(p+2\pi l/a)^2. \end{aligned} \tag{4.5.4}$$

For $m = 0$ the lattice propagator corresponds to a lattice action

$$S[\bar{\Psi}, \Psi] = a^d \sum_{x,y} \bar{\Psi}_x \gamma_\mu \rho_\mu(x-y) \Psi_y, \tag{4.5.5}$$

with couplings $\rho_\mu(x-y)$ calculable by a Fourier transformation. This lattice action is perfect by construction, i.e. its spectrum is identical with the one of the continuum theory. Hence, there should be no fermion doubling. On the other hand, the action is manifestly chirally invariant. This seems to contradict the Nielsen-Ninomiya theorem. However, the theorem is evaded because the action turns out to be non-local. Its couplings $\rho_\mu(x-y)$ do not decay exponentially at large distances. Instead for $d \geq 2$ they decay only power-like. As a consequence, in momentum space $\rho_\mu(p)$ is not regular (it actually has poles) and therefore the topological arguments behind the Nielsen-Ninomiya theorem do not apply. The non-locality can be seen easily for $d = 1$. Then $\gamma_1 = 1$ and the sum in eq.(4.5.4) can be performed analytically, resulting in a massless propagator that takes the form

$$\langle \bar{\Psi}(-p)\Psi(p)\rangle = \sum_{l\in\mathbf{Z}} [i(p+2\pi l/a)]^{-1} \Pi(p+2\pi l/a)^2 = \frac{a}{2i}\cot(\frac{pa}{2}). \tag{4.5.6}$$

This implies

$$\rho_1(p) = \frac{2}{a}\tan(\frac{pa}{2}), \tag{4.5.7}$$

which is singular at the edge of the Brillouin zone ($p = \pm\pi/a$). The corresponding coupling in coordinate space,

$$\rho_1(x-y) = \frac{1}{a}(-1)^{(x-y)/a}, \tag{4.5.8}$$

does not decay at all at large distances $x-y$ and thus describes an extremely non-local action. For $d \geq 2$ the chirally symmetric perfect action remains non-local with a power-law decay of the couplings at large distances.

Although the non-locality of the perfect action arose naturally by blocking the theory from the continuum, from a practical point of view it is very inconvenient. For example, in a numerical simulation it would be very demanding to include couplings to far away neighbors. It follows from the Nielsen-Ninomiya theorem that, in order to obtain a local perfect action, one must break chiral symmetry explicitly. This can be done by modifying the above way of blocking from the continuum which was chirally covariant. If one chooses to break chiral symmetry explicitly in the blocking procedure, the resulting perfect lattice action is not manifestly chirally invariant, but it is local. This can be achieved by constructing a perfect lattice action $S[\bar{\Psi}, \Psi]$ as

$$\begin{aligned} \exp(-S[\bar{\Psi}, \Psi]) &= \int \mathcal{D}\overline{\psi}\mathcal{D}\psi \exp\{-\frac{1}{(2\pi)^d}\int d^d p\ \overline{\psi}(-p)[i\gamma_\mu p_\mu + m]\psi(p)\} \\ &\times \exp\{-\frac{1}{c}\frac{1}{(2\pi)^d}\int_B d^d p \\ &\times [\bar{\Psi}(-p) - \sum_{n\in\mathbf{Z}^d} \overline{\psi}(-p-2\pi n/a)\Pi(p+2\pi n/a)] \\ &\times [\Psi(p) - \sum_{l\in\mathbf{Z}^d} \psi(p+2\pi l/a)\Pi(p+2\pi l/a)]. \end{aligned} \tag{4.5.9}$$

The coefficient $c$ is a source of explicit chiral symmetry breaking, which is injected into the theory via the renormalization group transformation that maps the continuum theory to the lattice theory. For $c \to 0$ we recover the chirally invariant but non-local perfect lattice action from before. In general one obtains

$$\langle \bar{\Psi}(-p)\Psi(p)\rangle = \sum_{l\in\mathbf{Z}^d} [i\gamma_\mu(p_\mu + 2\pi l_\mu/a) + m]^{-1}\Pi(p+2\pi l/a)^2 + c, \tag{4.5.10}$$

which corresponds to a local perfect action as long as $c \neq 0$.

Let us vary $c$ in order to optimize the locality of the perfect action. For this purpose we again consider $d = 1$. Then the sum in eq.(4.5.10) can be performed analytically and the fermion propagator takes the form

$$\langle \bar{\Psi}(-p)\Psi(p)\rangle = \frac{1}{m} - \frac{2}{m^2 a}[\coth(\frac{ma}{2}) - i\cot(\frac{pa}{2})]^{-1} + c. \tag{4.5.11}$$

If we choose

$$c = \frac{\exp(ma) - 1 - ma}{m^2 a}, \tag{4.5.12}$$

the propagator reduces to

$$\langle \bar{\Psi}(-p)\Psi(p)\rangle = \left(\frac{\exp(ma)-1}{ma}\right)^2 \left[i\frac{1}{a}\sin(pa) + \frac{\exp(ma)-1}{a} + \frac{2}{a}\sin^2(\frac{pa}{2})\right]^{-1}. \tag{4.5.13}$$

This corresponds to the standard Wilson fermion action except that the mass $m$ is now replaced by $(\exp(ma)-1)/a$. Hence, for the above choice of $c$, in one dimension the perfect action is ultralocal, i.e. it has only nearest-neighbor interactions. In the massless limit $m = 0$ the optimal choice for locality is $c = a/2$. When we go to more than one dimension the action remains local, but it is no longer ultralocal.

Next we derive the energy-momentum dispersion relation of perfect lattice fermions. The fermion 2-point function takes the form

$$\begin{aligned}
\langle \bar{\Psi}(-\vec{p},0)\Psi(\vec{p},x_d)\rangle &= \frac{1}{2\pi}\int_{-\pi/a}^{\pi/a} dp_d\ \langle \bar{\Psi}(-p)\Psi(p)\rangle \exp(ip_d x_d) \\
&= \frac{1}{2\pi}\int_{-\pi/a}^{\pi/a} dp_d\ \Big\{ \sum_{l\in\mathbf{Z}^d} [i\gamma_\mu(p_\mu + 2\pi l_\mu/a) + m]^{-1} \\
&\times\ \Pi(p+2\pi l/a)^2 + c\Big\} \exp(ip_d x_d) \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} dp_d \sum_{l\in\mathbf{Z}^{d-1}} \frac{m}{(\vec{p}+2\pi\vec{l}/a)^2 + p_d^2 + m^2} \\
&\times \prod_{i=1}^{d-1}\left(\frac{2\sin(p_i a/2)}{p_i a + 2\pi l_i}\right)^2 \left(\frac{2\sin(p_d a/2)}{p_d a}\right)^2 \exp(ip_d x_d) + c\ \delta_{x_d,0} \\
&= \sum_{\vec{l}\in\mathbf{Z}^{d-1}} C(\vec{p}+2\pi\vec{l}/a)\exp(-E(\vec{p}+2\pi\vec{l}/a)x_d) + c\ \delta_{x_d,0}.
\end{aligned} \tag{4.5.14}$$

The sum over $l_d$ has been combined with the integral of $p_d$ over $[-\pi/a, \pi/a]$ to an integral over the momentum space of the continuum theory. The sum over the spatial $\vec{l} \in \mathbf{Z}^{d-1}$ leads to infinitely many poles of the integrand, and hence to infinitely many states that contribute an exponential to the 2-point function. The energies of these states are given by the location of the poles, $E(\vec{p}+2\pi\vec{l}/a) = -ip_d$, with

$$E(\vec{p}+2\pi\vec{l}/a)^2 = -p_d^2 = (\vec{p}+2\pi\vec{l}/a)^2 + m^2. \tag{4.5.15}$$

Hence, the energy-momentum dispersion relation of perfect lattice fermions is exactly the same as in the continuum. In particular, as a result of exact blocking from the continuum, there are no lattice artifacts. Furthermore, the form of the renormalization group blocking transformation has no effect on the physical spectrum. In particular, the explicit chiral symmetry breaking term proportional to $c$ only leads to a contact term $c\ \delta_{x_d,0}$ in the 2-point function. Hence, it has no effect on the spectrum which is extracted from the 2-point function at large Euclidean time separations $x_d$. Remarkably, the spectrum of the lattice theory displays the consequences of Poincaré invariance despite the fact that the lattice action only has the discrete lattice symmetries.

Chiral symmetry is hidden in a similar way. Due to the explicit chiral symmetry breaking parameter $c$ in the renormalization group blocking transformation, even for $m = 0$ the perfect lattice action is not manifestly chirally invariant. Still, all physical consequences of chiral symmetry are correctly reproduced by the perfect action. As we will see later, this is the due to the by now famous Ginsparg-Wilson relation,

$$\{D^{-1}, \gamma_5\} = a\gamma_5. \tag{4.5.16}$$

Here $D$ is the lattice Dirac operator and $D^{-1}$ is the lattice fermion propagator. Indeed, using the optimal parameter $c = a/2$ for the perfect action one obtains

$$\{D^{-1}, \gamma_5\} = \{\sum_{l \in \mathbf{Z}^d} [i\gamma_\mu (p_\mu + 2\pi l_\mu/a)]^{-1} \Pi(p + 2\pi l/a)^2 + c, \gamma_5\} = 2c\gamma_5 = a\gamma_5. \tag{4.5.17}$$

The Ginsparg-Wilson relation is the key to understanding chiral symmetry on the lattice. In the continuum, chiral symmetry implies $\{D^{-1}, \gamma_5\} = 0$. If one insists on this relation also on the lattice, i.e. if one insists on manifest chiral symmetry for a lattice action, the Nielsen-Ninomiya theorem implies fermion doubling (or, even worse, a violation of locality). The Ginsparg-Wilson relation $\{D^{-1}, \gamma_5\} = a\gamma_5$ reduces to the relation $\{D^{-1}, \gamma_5\} = 0$ in the continuum limit $a \to 0$. Still, at finite lattice spacing $a$, the right-hand side of the Ginsparg-Wilson relation implies an explicit breaking of chiral symmetry. In the case of the perfect action the explicit breaking is due to the parameter $c = a/2$ in the renormalization group blocking transformation. This minimal explicit violation of chiral symmetry is sufficient to evade the Nielsen-Ninomiya theorem, and thus to avoid fermion doubling. Still, as we have seen explicitly for the perfect action, the physics (for example, the spectrum) remains the same as in the continuum. We will see later that the Ginsparg-Wilson relation leads to a natural definition

of chiral symmetry on the lattice which reduces to the usual one in the continuum limit.

# Chapter 5

# Lattice QCD

While many properties of lattice chiral symmetry can be studied in the free theory, one certainly also needs to understand it in the interacting theory. Hence, it is now time to endow the quarks with their non-trivial QCD dynamics by coupling them to the gluon field. However, before doing so, we first discuss lattice Yang-Mills theory without quarks.

## 5.1 Lattice Yang-Mills Theory

Maintaining manifest gauge invariance is essential when gauge theories are regularized on the lattice. In the continuum, gauge transformations involve space-time derivatives of group-valued functions $\Omega(x)$. On the lattice there are no infinitesimally close points, and continuum derivatives are usually simply replaced by finite differences. However, in order to maintain gauge invariance, one must proceed more carefully. Wegner and Wilson, as well as Smit, independently introduced the concept of a parallel transporter $U_{x,\mu} \in SU(N_c)$ connecting neighboring lattice points $x$ and $x + \hat{\mu}$. The parallel transporter is related to an underlying continuum gauge field $A_\mu(x) = igA^a_\mu(x)T^a$ by

$$U_{x,\mu} = \mathcal{P} \exp \int_0^a dt \; A_\mu(x + \hat{\mu}t), \tag{5.1.1}$$

where $\mathcal{P}$ denotes path-ordering. Under a non-Abelian gauge transformation the parallel transporter transforms as

$$U'_{x,\mu} = \Omega_x U_{x,\mu} \Omega^\dagger_{x+\hat{\mu}}. \tag{5.1.2}$$

Wilson has constructed the Yang-Mills part of a simple lattice QCD action by multiplying parallel transporters around an elementary plaquette. The standard Wilson action is constructed as a sum over all plaquettes

$$S_{YM}[U] = -a^4 \sum_{x,\mu,\nu} \frac{1}{g^2 a^2} \mathrm{Tr}[U_{x,\mu} U_{x+\hat{\mu},\nu} U^\dagger_{x+\hat{\nu},\mu} U^\dagger_{x,\nu} + U_{x,\nu} U_{x+\hat{\nu},\mu} U^\dagger_{x+\hat{\mu},\nu} U^\dagger_{x,\mu}]. \tag{5.1.3}$$

It reduces to the continuum Yang-Mills action in the limit $a \to 0$.

To fully define the path integral we must also consider the measure. The lattice functional integral is obtained as an integral over all configurations of parallel transporters $U_{x,\mu}$, i.e.

$$Z = \prod_{x,\mu} \int_{SU(N_c)} dU_{x,\mu} \ \exp(-S_{YM}[U]). \tag{5.1.4}$$

One integrates independently over all link variables using the local Haar measure $dU_{\mu,x}$ for each parallel transporter. The Haar measure is a left- and right-invariant measure, i.e.

$$\int_{SU(N_c)} dU \ f(\Omega U) = \int_{SU(N_c)} dU \ f(U\Omega) = \int_{SU(N_c)} dU \ f(U), \tag{5.1.5}$$

for any function $f(U)$ and for any $SU(N_c)$ matrix $\Omega$. It is convenient to normalize the measure such that

$$\int_{SU(N_c)} dU = 1. \tag{5.1.6}$$

For compact groups like $SU(N_c)$ the integration is over a finite domain. This makes it unnecessary to fix the gauge in lattice QCD because the functional integral is finite even without gauge fixing. This is another important advantage of the formulation using parallel transporters.

The Yang-Mills functional integral from above contains a single parameter — the bare gauge coupling $g$. When one wants to perform the continuum limit, one must search for values of $g$ for which the correlation length of the lattice theory diverges in lattice units. In the language of statistical mechanics one is looking for

a second order phase transition. Due to asymptotic freedom, in lattice QCD one expects a second order phase transition at $g \to 0$. To analyze the phase structure of a gauge theory one needs to study order parameters. A simple local order parameter like $\langle U_{x,\mu} \rangle$ is not useful. This follows from Elitzur's theorem which states that gauge-variant observables simply vanish. A useful order parameter in a gauge theory must be gauge invariant and, in addition, non-local. In a pure gluon theory a good order parameter was suggested independently by Wegner and Wilson as

$$W_{\mathcal{C}} = \mathrm{Tr} \prod_{(x,\mu)\in\mathcal{C}} U_{x,\mu}. \tag{5.1.7}$$

For a rectangular curve $\mathcal{C}$ with side lengths $R$ and $T$ the Wilson loop describes the instantaneous creation and annihilation of a static quark-anti-quark pair at distance $R$ which then exists for a time $T$. The Wilson loop is related to the static quark-anti-quark potential $V(R)$ by

$$\lim_{T\to\infty} \langle W_{\mathcal{C}} \rangle \sim \exp(-V(R)T). \tag{5.1.8}$$

In QCD we expect quarks and anti-quarks to be confined to one another by a potential rising linearly at large separations $R$, i.e.

$$\lim_{R\to\infty} V(R) \sim \sigma R, \tag{5.1.9}$$

where $\sigma$ is the string tension. In a confinement phase the Wilson loop hence shows an area law

$$\lim_{R,T\to\infty} \langle W_{\mathcal{C}} \rangle \sim \exp(-\sigma RT). \tag{5.1.10}$$

Confinement is indeed verified very accurately in numerical simulations of lattice Yang-Mills theories.

## 5.2 Confinement in the Strong Coupling Limit

In lattice gauge theory it is straightforward to prove confinement for large values of the bare gauge coupling $g$. In the strong coupling region, however, the correlation length is small and we cannot take the continuum limit. In fact, due to asymptotic freedom we expect the continuum limit to be at $g \to 0$. It is an open question if confinement persists in the continuum limit. However, there is a lot of positive numerical evidence that this is indeed the case. Assuming that

there is no phase transition between the strong and weak coupling regions, the derivation of confinement in the strong coupling regime would carry over to the continuum limit. In the strong coupling expansion we expand in powers of $1/g$ around $g = \infty$. To leading order the pure gluon action is then simply zero. The Wilson loop operator takes the form

$$W_{\mathcal{C}} = U_{1ij}U_{2jk}U_{3kl}...U_{Nmi}, \tag{5.2.1}$$

where $N = 2(R + T)$ is the number of links along the loop. Using the group integration formula

$$\int_{SU(N_c)} dU\ U_{ij} = 0,\ \int_{SU(N_c)} dU\ U_{ij}U_{kl} = \frac{1}{N_c}\delta_{jk}\delta_{il}, \tag{5.2.2}$$

then immediately implies $W_{\mathcal{C}} = 0$. The second formula of eq.(5.2.2) suggests that to get a non-zero result one needs to have the product of two matrix elements for each link variable on the Wilson loop. This can be achieved by expanding the Boltzmann factor of the action to higher orders in $1/g$. The lowest non-zero contribution comes from tiling the Wilson loop with plaquettes that result from expanding

$$\exp(-\frac{1}{2g^2}\text{Re Tr}(1 - U_{\mu\nu,x})) = \exp(-\frac{N_c}{2g^2})(1 + \frac{1}{2g^2}\text{Re Tr}U_{\mu\nu,x}) \tag{5.2.3}$$

for the plaquettes in the interior of the Wilson loop. Taking the $U_{\mu\nu,x}$ term for all these plaquettes and using the second formula of eq.(5.2.2) gives

$$\langle W_{\mathcal{C}}\rangle = \frac{1}{(g^2)^{RT}}. \tag{5.2.4}$$

Indeed we find an area law and we read off the string tension as

$$\sigma = -\log(\frac{1}{g^2}). \tag{5.2.5}$$

Higher order corrections arise from deformations of the simple tiling from above. The leading correction comes from eliminating one of the $RT$ plaquettes of the original tiling, and replacing it with five plaquettes at the surface of an elementary lattice cube, such that the resulting surface bounded by the Wilson loop has no holes. The cube can be attached above and below the plane of the Wilson loop, and it can go out in the two possible orthogonal directions. This results in $4RT$

contributions, which all have four more plaquettes than the leading term, and are hence suppressed by $1/(g^2)^4$. Up to that order one finds

$$\langle W_{\mathcal{C}} \rangle = \frac{1}{(g^2)^{RT}}(1 + 4RT\frac{1}{(g^2)^4}). \tag{5.2.6}$$

The last two terms are the first term in an exponential, and hence we can read off the corrected string tension as

$$\sigma = -\log(\frac{1}{g^2}) + 4\frac{1}{(g^2)^4}. \tag{5.2.7}$$

The string tension has been computed to higher orders. Still, one is unable to go to high enough orders to reach the physically interesting scaling region in which we can perform the continuum limit. At present, only numerical simulation techniques are powerful enough to enter that regime. Consequently, the above result for the string tension cannot yet be compared with experimental results. It is instructive to derive the glueball mass in a similar manner. For this purpose one considers the correlation function of two elementary plaquette operators separated in Euclidean time. Using the strong coupling expansion one can identify the exponential decay, and extract the correlation length whose inverse is the massgap or glueball mass.

## 5.3 Standard Wilson Action for Lattice QCD

We still need to couple the quarks to the gluons. First we do this by gauging $SU(N_c)$ in the action of free Wilson fermions of eq.(4.4.1)

$$\begin{aligned} S_{QCD}[\overline{\Psi}, \Psi, U] &= a^4 \sum_{x,\mu} \frac{1}{2a}(\bar{\Psi}_x \gamma_\mu U_{x,\mu} \Psi_{x+\hat{\mu}} - \bar{\Psi}_{x+\hat{\mu}} \gamma_\mu U^\dagger_{x,\mu} \Psi_x) + a^4 \sum_x m \bar{\Psi}_x \Psi_x \\ &+ a^4 \sum_{x,\mu} \frac{1}{2a}(2\bar{\Psi}_x \Psi_x - \bar{\Psi}_x U_{x,\mu} \Psi_{x+\hat{\mu}} - \bar{\Psi}_{x+\hat{\mu}} U^\dagger_{x,\mu} \Psi_x) \\ &- a^4 \sum_{x,\mu,\nu} \frac{1}{g^2 a^2} \mathrm{Tr}[U_{x,\mu} U_{x+\hat{\mu},\nu} U^\dagger_{x+\hat{\nu},\mu} U^\dagger_{x,\nu} + U_{x,\nu} U_{x+\hat{\nu},\mu} U^\dagger_{x+\hat{\mu},\nu} U^\dagger_{x,\mu}]. \end{aligned} \tag{5.3.1}$$

In order to eliminate the doubler fermions we have introduced the Wilson term which breaks chiral symmetry explicitly. The lattice regularized functional inte-

gral takes the form

$$Z = \prod_x \int d\bar{\Psi}_x d\Psi_x \prod_{x,\mu} \int_{SU(N_c)} dU_{x,\mu} \exp(-S_{QCD}[\bar{\Psi}, \Psi, U]). \tag{5.3.2}$$

It depends on two parameters — the bare gauge coupling $g$ and the bare quark mass $m$. Due to asymptotic freedom, in order to reach the continuum limit one must take $g \to 0$. When one puts $m = 0$ for free Wilson fermions one reaches the chiral limit. In the interacting theory, on the other hand, this is no longer the case. In particular, since chiral symmetry is explicitly broken, the bare quark mass $m$ must be fine-tuned in order to reach a massless limit. This fine-tuning is very unnatural from a theoretical point. In particular, following the discussion in the introduction, if one imagines Wilson's lattice QCD as an oversimplified model for the short-distance physics at the Planck scale, one could not understand at all why there are light fermions in Nature. The fine-tuning of $m$ is also inconvenient from a practical point of view. For example, in a numerical simulation of Wilson's lattice QCD one must fine-tune $m$ to many digits accuracy in order to make the pion massless. If one does this at relatively large $g$, i.e. before one reaches the continuum limit $g \to 0$, this massless "pion" is not even a proper Goldstone boson of a spontaneously broken chiral symmetry. For Wilson fermions an exact chiral symmetry that can break spontaneously does not exist at finite lattice spacing. It emerges only in the continuum limit after a delicate fine-tuning of $m$.

## 5.4 Ginsparg-Wilson Relation and Lattice Chirality

In the discussion of the perfect free fermion action we have encountered the Ginsparg-Wilson relation eq.(4.5.16). Following Lüscher, we will now use this relation to construct a version of chiral symmetry that is natural for a lattice theory and reduces to the usual one in the continuum limit. For this purpose we consider a lattice fermion action

$$S[\bar{\Psi}, \Psi, U] = a^4 \ \bar{\Psi} D[U] \Psi = a^4 \sum_{x,y} \bar{\Psi}_x D[U]_{x,y} \Psi_y, \tag{5.4.1}$$

which is defined in terms of the lattice Dirac operator $D[U]$. This operator should be local (i.e. it should decay exponentially at large distances $x - y$) but

not ultralocal. The lattice Dirac operator obeys the Ginsparg-Wilson relation if the corresponding fermion propagator $D[U]^{-1}$ satisfies

$$\{D[U]^{-1}, \gamma_5\} = D[U]^{-1}\gamma_5 + \gamma_5 D[U]^{-1} = a\gamma_5. \tag{5.4.2}$$

Alternatively, the Ginsparg-Wilson relation can be written as

$$\gamma_5 D[U] + D[U]\gamma_5 = aD[U]\gamma_5 D[U]. \tag{5.4.3}$$

It is non-trivial to construct lattice actions that obey the Ginsparg-Wilson relation. Until now we have seen that the perfect action for massless free fermions indeed satisfies this relation. In the next section we will see that the same is true for overlap fermions. For the moment we don't worry about the concrete form of $D[U]$, we just assume that it obeys eq.(5.4.3).

Let us first consider an infinitesimal chiral rotation of the form familiar from the continuum

$$\begin{aligned} \Psi' &= \Psi + \delta\Psi = (1 + i\varepsilon^a T^a \gamma_5)\Psi, \\ \bar{\Psi}' &= \bar{\Psi} + \delta\bar{\Psi} = \bar{\Psi}(1 + i\varepsilon^a T^a \gamma_5). \end{aligned} \tag{5.4.4}$$

Here $T^a$ (with $a \in \{1, 2, ..., N_f^2 - 1\}$) are the generators of $SU(N_f)$ and $\varepsilon^a$ is a small parameter. In order to discuss flavor-singlet axial transformations with an infinitesimal parameter $\varepsilon^0$ we also define $T^0 = \mathbb{1}$. If the lattice action is local and has no fermion doubling, the Nielsen-Ninomiya theorem implies that it cannot be invariant under the above chiral rotations. On the other hand, the lattice fermion measure is invariant under the full chiral symmetry $U(N_f)_L \otimes U(N_f)_R$. This is very different from massless QCD in the continuum. In the continuum the action is invariant under $U(N_f)_L \otimes U(N_f)_R$ chiral transformations, while the measure is invariant only under $SU(N_f)_L \otimes SU(N_f)_R \otimes U(1)_{L=R}$. In particular, due to the anomaly the measure of the continuum theory is not invariant under flavor-singlet axial transformations, while the measure of the lattice theory is invariant.

Next we consider Lüscher's modification of the standard chiral transformation

$$\begin{aligned} \Psi' &= \Psi + \delta\Psi = \left(1 + i\varepsilon^a T^a \gamma_5 (1 - \frac{a}{2} D[U])\right) \Psi, \\ \bar{\Psi}' &= \bar{\Psi} + \delta\bar{\Psi} = \bar{\Psi} \left(1 + i\varepsilon^a T^a (1 - \frac{a}{2} D[U])\gamma_5\right). \end{aligned} \tag{5.4.5}$$

Through $D[U]$ Lüscher's lattice version of a chiral transformation depends on the gluon field. Still, in the continuum limit $a \to 0$ it reduces to the standard

chiral symmetry of the continuum theory. It is remarkable that eq.(5.4.5) is a symmetry of any lattice action that obeys the Ginsparg-Wilson relation eq.(5.4.3). This follows from

$$\begin{aligned}\bar{\Psi}' D[U] \Psi' &= \bar{\Psi}\left(1 + i\varepsilon^a T^a (1 - \frac{a}{2} D[U])\gamma_5\right) D[U] \left(1 + i\varepsilon^a T^a \gamma_5 (1 - \frac{a}{2} D[U])\right) \Psi \\ &= \bar{\Psi} D[U] \Psi + \bar{\Psi}\left(i\varepsilon^a T^a [\gamma_5 D[U] + D[U]\gamma_5 - a D[U]\gamma_5 D[U]]\right) \Psi \\ &+ \mathcal{O}(\varepsilon^2) \\ &= \bar{\Psi} D[U] \Psi + \mathcal{O}(\varepsilon^2). \end{aligned} \tag{5.4.6}$$

Similarly, the variation of the lattice fermion measure takes the form

$$\begin{aligned}\mathcal{D}\bar{\Psi}' \mathcal{D}\Psi' &= \mathcal{D}\bar{\Psi}\left(1 + i\varepsilon^a T^a (1 - \frac{a}{2} D[U])\gamma_5\right) \left(1 + i\varepsilon^a T^a \gamma_5 (1 - \frac{a}{2} D[U])\right) \mathcal{D}\Psi \\ &= \mathcal{D}\bar{\Psi}\mathcal{D}\Psi \left(1 + i\varepsilon^a \mathrm{Tr}[T^a \gamma_5 (2 - a D[U])]\right) + \mathcal{O}(\varepsilon^2) \\ &= \mathcal{D}\bar{\Psi}\mathcal{D}\Psi \left(1 - i\varepsilon^0 a \mathrm{Tr}[\gamma_5 D[U]]\right) + \mathcal{O}(\varepsilon^2). \end{aligned} \tag{5.4.7}$$

Hence, while any Ginsparg-Wilson fermion action is invariant under Lüscher's lattice chiral symmetry, the lattice fermion measure is not. Exactly as in the continuum, the fermionic measure of the lattice theory changes under flavor-singlet axial transformations, while it is invariant under $SU(N_f)_L \otimes SU(N_f)_R \otimes U(1)_{L=R}$. We will see later that the non-invariance of the fermionic measure under flavor-singlet axial transformations indeed gives rise to the correct axial anomaly.

## 5.5 Domain Wall and Overlap Fermions

In the early nineties Kaplan proposed a novel method to preserve chirality on the lattice. The idea was to use the fact that chiral fermions become trapped on domain walls. Kaplan used a Wilson-Dirac operator in five dimensions with a mass term that is a function of the fifth direction. In particular, the mass term changes sign creating a four-dimensional domain wall at the points where it vanishes. A four-dimensional chiral fermion is then trapped on the domain wall. In the meantime Narayanan and Neuberger were developing an idea of using an infinite number of regulator "flavor" fields to preserve chirality. They realized that Kaplan's construction was equivalent to their idea since the fifth dimension is analogous to a flavor space. They used their interpretation and argued that the determinant of a chiral fermion in the background of a gauge field is equivalent

to the overlap of two many-body fermionic ground states. Initially, it seemed that the overlap was a reliable technique to regulate even chiral gauge theories on a lattice. Unfortunately, it was soon realized that the fermions with opposite chirality (which originate from an anti-wall) could not be easily decoupled.

Although it was not clear whether the domain wall and overlap approach gave a completely satisfactory formulation of lattice chiral gauge theories involving no doublers of opposite chirality, something highly non-trivial had been achieved. It was possible to construct a lattice theory with a fermion with opposite chirality and use it effectively to preserve chiral symmetry in a vector-like gauge theory. An elegant way to use a five-dimensional fermion action to represent quarks in lattice QCD was first proposed by Shamir and elaborated further by Furman and Shamir. This fermion is commonly referred to as a domain wall fermion and is used extensively in lattice simulations. Its action is constructed on a five-dimensional space-time lattice with coordinates $(x, x_5)$, where $x$ refers to the usual four dimensions and $x_5 \in \{a_5, 2a_5, ..., L_5\}$ refers to the fifth direction of finite extent $L_5$. Since the fifth direction is physically different from the other directions we have introduced a new lattice spacing $a_5$ in that direction. The domain wall fermion action is given by

$$S_F[\bar{\Psi}, \Psi, U] = a^4 a_5 \sum_{x,x_5,y,y_5} \bar{\Psi}_{x,x_5} D_{DW}[U]_{x,x_5;y,y_5} \Psi_{y,y_5}. \tag{5.5.1}$$

The domain wall Dirac operator is given by

$$\begin{aligned}
D_{DW}[U]_{x,x_5;y,y_5} &= \delta_{x_5,y_5} D^{\parallel}[U]_{x,y} + \delta_{x,y} D^{\perp}[U]_{x_5,y_5}, \\
D^{\parallel}[U]_{x,y} &= M\delta_{x,y} + \sum_{\mu} \frac{1}{2a} \left( \gamma_\mu U_{x,\mu} \delta_{x+\hat{\mu},y} - \gamma_\mu U^{\dagger}_{x-\hat{\mu},\mu} \delta_{x-\hat{\mu},y} \right) \\
&- \sum_{\mu} \frac{1}{2a} \left( 2\delta_{x,y} - U_{x,\mu} \delta_{x+\hat{\mu},y} - U^{\dagger}_{x-\hat{\mu},\mu} \delta_{x-\hat{\mu},y} \right), \\
D^{\perp}[U]_{x_5,y_5} &= \begin{cases} (P_R\, \delta_{2a_5,y_5} - \delta_{a_5,y_5})/a_5 - mP_L\, \delta_{L_5,y_5} & \text{for } x_5 = 1, \\ (P_R\, \delta_{x_5+a_5,y_5} + P_L\, \delta_{x_5-a_5,y_5} - \delta_{x_5,y_5})/a_5 & \text{for } a_5 < s < L_5, \\ (P_L\, \delta_{L_5-a_5,y_5} - \delta_{L_5,y_5})/a_5 - mP_R\, \delta_{a_5,y_5} & \text{for } x_5 = L_5. \end{cases}
\end{aligned} \tag{5.5.2}$$

Here $P_R$ and $P_L$ are the chiral projection operators defined in eq.(3.2.5). In the above action the parameter $M$ is not the mass of the quark that is bound to the wall. By comparing with Wilson fermions one sees that the sign of the Wilson term has changed. In order to produce massless quarks one should set

$0 \leq M \leq 2/a_5$ at tree level and take $L_5 \to \infty$. There is a technical problem that needs to be taken into account. When $L_5$ becomes infinite there are only $N_f$ flavors of four-dimensional massless quarks bound to the wall, but there is an infinite number of modes at the cut-off. This may cause spurious effects at low energies. Hence one needs to use bosonic (Pauli-Villars type fields) to cancel the contribution of these high-energy modes.

There is a close connection between the domain wall approach and the overlap formula developed by Narayanan and Neuberger. Neuberger realized that it is possible to find an analytic formula for an effective Dirac operator that describes the massless chiral mode of the domain wall fermion. Using his insight on the overlap formula for vector-like gauge theories, he found a simple and elegant formula for the four-dimensional Dirac operator, which is referred to as the overlap Dirac operator and which is given by

$$D_O[U] = \frac{1}{2a}\Big[1 + \gamma_5 \frac{H[U]}{\sqrt{H[U]^2}}\Big] \tag{5.5.3}$$

where $H[U] = \gamma_5 D^{\parallel}[U]$ and $D^{\parallel}[U]$ is the operator we defined above in the context of the domain wall fermion. In order to obtain massless quarks one needs to set $0 \leq Ma \leq 2/a_5$ as before. In fact, it is possible to find an analytic formula for an effective Dirac operator that represents the chiral massless modes of the domain wall fermion even for finite $L_5$ . This operator takes the form

$$D_{L_5}[U] = \frac{1}{a}\Big[1 + \gamma_5 \frac{(1+\tilde{H}[U])^{L_5/a_5} - (1-\tilde{H}[U])^{L_5/a_5}}{(1+\tilde{H}[U])^{L_5/a_5} + (1-\tilde{H}[U])^{L_5/a_5}}\Big], \tag{5.5.4}$$

where

$$\tilde{H}[U] = \gamma_5 \tilde{X}[U],\ \ \tilde{X}[U] = \frac{a_5 D^{\parallel}[U]}{2 + a_5 D^{\parallel}[U]}. \tag{5.5.5}$$

In the limit of $L_5 \to \infty$ one obtains

$$\lim_{L_5 \to \infty} D_{L_5}[U] = D_{DWO}[U] = \frac{1}{2a}\Big[1 + \gamma_5 \frac{\tilde{H}[U]}{\sqrt{\tilde{H}[U]^2}}\Big], \tag{5.5.6}$$

which reduces to the overlap Dirac operator $D_O[U]$ when $a_5 \to 0$.

Although today we know that the Ginsparg-Wilson relation leads to an exact chiral symmetry on the lattice, this connection was not appreciated until recently.

After Ginsparg and Wilson discovered this interesting relation they found it difficult to explicitly construct a local operator that satisfies it. The relation was soon forgotten. After the discovery of perfect and overlap fermions the relation was rediscovered by Hasenfratz. It is straightforward to check that $D_O[U]$ and $D_{DWO}[U]$ indeed satisfy the Ginsparg-Wilson relation. These new Dirac operators couple every pair of sites on the lattice. It is possible to show that if one wants to benefit from good chiral properties of Ginsparg-Wilson fermions, one has to give up the notion of ultralocal actions. However, close to the continuum limit the couplings in the overlap Dirac operator fall off exponentially with the distance. In this sense these new Dirac operators are still local. Unfortunately, the closeness to the continuum limit is quite important to maintain both the chiral and local properties of the Dirac operator. Recently, a physical picture based on the locality of zero modes of the Dirac operator was used to map out the regions in coupling constant space where $D_O[U]$ and $D_{DWO}[U]$ lead to a good regularization of massless quarks.

# Chapter 6

# The Monte Carlo Method

A powerful numerical technique to solve problems in statistical mechanics or lattice field theory is the so-called Monte Carlo method. Here we discuss the application of the method in the context of classical spin models. The generalization to other statistical mechanics systems and to lattice field theory is straightforward. The idea behind the Monte Carlo method is to compute expectation values by generating spin configurations numerically. Of course, the partition function is an extremely large sum, such that doing it with numerical brute force is completely hopeless. In the Monte Carlo method predominantly those spin configurations are generated that have the largest contribution to the partition function. In fact, the Boltzmann factor $\exp(-\beta\mathcal{H}[s])$ is used as the probability to generate the spin configuration $[s]$.

## 6.1 The Concept of a Markov Chain

In a Monte Carlo simulation one generates a sequence of spin configurations

$$[s^{(1)}] \to [s^{(2)}] \to ... \to [s^{(N)}], \tag{6.1.1}$$

which form a so-called Markov chain, by applying an algorithm that turns the configuration $[s^{(i)}]$ into $[s^{(i+1)}]$. The initial configuration $[s^{(1)}]$ is either picked at random or selected otherwise. Ultimately, nothing should depend on this choice. After a (possibly large) number $M$ of Monte Carlo iterations (applications of

the algorithm) an equilibrium is reached, and the system has forgotten about the initial configurations. Only the configurations generated after equilibration are used in the actual calculation. To estimate the expectation value of some observable one averages its values over all configurations of the Monte Carlo sample

$$\langle \mathcal{O} \rangle = \lim_{N\to\infty} \frac{1}{N-M} \sum_{i=M+1}^{N} \mathcal{O}[s^{(i)}]. \tag{6.1.2}$$

In the limit $N \to \infty$ the calculation becomes exact. At finite $N-M$ one makes a calculable statistical error that decreases in proportion to $1/\sqrt{N-M-1}$. Hence, to increase the numerical accuracy by a factor of two one must run the Monte Carlo algorithm four times as long. The Boltzmann factor $\exp(-\beta\mathcal{H}[s])$ is not explicitly included in the above sum. It is implicitly included, because the configurations in the Markov chain occur with probability $\exp(-\beta\mathcal{H}[s])$.

## 6.2 Ergodicity and Detailed Balance

To demonstrate that a particular Monte Carlo algorithm converges to the correct equilibrium distribution it is sufficient to show that it is ergodic and obeys detailed balance. Ergodicity means that starting from an arbitrary initial configuration the algorithm can in principle reach any other spin configuration. This condition is obviously necessary, because the correct value for the expectation value can be obtained only if all spin configurations are included. Detailed balance means that

$$\exp(-\beta\mathcal{H}[s])w[s,s'] = \exp(-\beta\mathcal{H}[s'])w[s',s]. \tag{6.2.1}$$

Here $w[s,s']$ is the transition probability for the algorithm to turn the configuration $[s]$ into $[s']$. A Monte Carlo algorithm is completely characterized by the corresponding $w[s,s']$. Since the algorithm definitely generates a new configuration the proper normalization is

$$\sum_{[s']} w[s,s'] = 1. \tag{6.2.2}$$

When the Monte Carlo algorithm converges to an equilibrium distribution $p[s]$ of spin configurations, this distribution is an eigenvector of $w[s,s']$ with eigenvalue 1

$$\sum_{[s]} p[s]w[s,s'] = p[s']. \tag{6.2.3}$$

Now we want to show that the canonical Boltzmann distribution

$$p[s] = \exp(-\beta \mathcal{H}[s]) \tag{6.2.4}$$

is indeed an eigenvector of $w[s, s']$ if the algorithm obeys detailed balance. We find

$$\begin{aligned} \sum_{[s]} \exp(-\beta \mathcal{H}[s]) w[s, s'] &= \sum_{[s]} \exp(-\beta \mathcal{H}[s']) w[s', s] \\ &= \exp(-\beta \mathcal{H}[s']) \sum_{[s]} w[s', s] \\ &= \exp(-\beta \mathcal{H}[s']). \end{aligned} \tag{6.2.5}$$

Assuming ergodicity one can show that only one eigenvector with eigenvalue 1 exists, and that the equilibrium distribution is therefore unique.

## 6.3 The Metropolis Algorithm

A simple example of an algorithm that is ergodic and obeys detailed balance is the so-called Metropolis algorithm. In this algorithm a new configuration $[s']$ is randomly chosen based on the old configuration $[s]$. If the energy of the new configuration is smaller than the energy of the old configuration, the new configuration is accepted, i.e.

$$\mathcal{H}[s'] < \mathcal{H}[s] \Rightarrow w[s, s'] = 1. \tag{6.3.1}$$

On the other hand, if the new energy is larger, the new configuration is accepted only with a certain probability, i.e.

$$\mathcal{H}[s'] > \mathcal{H}[s] \Rightarrow w[s, s'] = \exp(-\beta(\mathcal{H}[s'] - \mathcal{H}[s])). \tag{6.3.2}$$

Otherwise the old configuration is kept. This algorithm obeys detailed balance. Let us consider two configurations $[s]$ and $[s']$. We can assume that $\mathcal{H}[s'] < \mathcal{H}[s]$ such that $w[s, s'] = 1$. Then of course, $\mathcal{H}[s] > \mathcal{H}[s']$ such that $w[s', s] = \exp(-\beta(\mathcal{H}[s] - \mathcal{H}[s']))$, and hence

$$\begin{aligned} \exp(-\beta \mathcal{H}[s]) w[s, s'] &= \exp(-\beta \mathcal{H}[s]) \\ &= \exp(-\beta \mathcal{H}[s']) \exp(-\beta(\mathcal{H}[s] - \mathcal{H}[s'])) \\ &= \exp(-\beta \mathcal{H}[s']) w[s', s]. \end{aligned} \tag{6.3.3}$$

We still need to specify how a new configuration is proposed. In the Ising model one visits the spins one by one and proposes to flip them. The resulting change of the energy is calculated by investigating the neighboring spins. Then following the Metropolis algorithm, it is decided if a given spin is flipped or not. When all spins on the lattice have been updated in this way one has completed one Metropolis sweep. It is obvious that any spin configuration can, at least in principle, be reached in this way, i.e. the Metropolis algorithm is indeed ergodic. A typical Monte Carlo simulation consists of a large number of sweeps, say 1 million, for example.

## 6.4 Error Analysis

Since any practical Monte Carlo simulation has a finite length, the results are not exact but are affected by statistical errors. Hence, an important part of every Monte Carlo calculation is the error analysis. An ideal Monte Carlo algorithm (which doesn't exist in practice) would generate a Markov chain of statistically independent configurations. If the Monte Carlo data for an observable $\mathcal{O}$ are Gaussian distributed, the standard deviation from their average (i.e. their statistical error) is given by

$$\Delta\mathcal{O} = \frac{1}{\sqrt{N-M-1}}\langle(\mathcal{O}-\langle\mathcal{O}\rangle)^2\rangle = \frac{1}{\sqrt{N-M-1}}(\langle\mathcal{O}^2\rangle - \langle\mathcal{O}\rangle^2). \qquad (6.4.1)$$

In order to reduce the statistical error by a factor of two, the number of independent equilibrated configurations $N-M$ must hence be increased by a factor of four.

Practical Monte Carlo algorithms (like the Metropolis algorithm) are not ideal, i.e. they do not generate statistically independent configurations. In particular, the Metropolis algorithm is rather simple, but not very efficient. Since the new configuration is generated from the previous configuration in the Markov chain, subsequent configurations are correlated. This implies that the actual statistical error is larger than the above naive estimate of the standard deviation. In order to detect the autocorrelation of the Monte Carlo data it is useful to bin these data. For this purpose one averages a number $N_b$ of subsequent measurements and treats this average as a statistically independent result. One then computes the standard deviation based on the $(N-M)/N_b$ statistically independent averages. Of course, if the bin size $N_b$ is too small, the averages are still correlated

and the corresponding standard deviation still underestimates the true statistical error. When one increases the bin size $N_b$, the corresponding standard deviation increases until subsequent bin averages are indeed statistically independent. Once the standard deviation has reached a plateau (by increasing $N_b$), one has obtained a reliable estimate of the true statistical error.

In order to estimate the number $\tau$ of Monte Carlo iterations that separate statistically independent spin configuration, it is also useful to determine the autocorrelation function of some observable $\mathcal{O}$

$$\langle \mathcal{O}^{(i)} \mathcal{O}^{(i+t)} \rangle = \lim_{N \to \infty} \frac{1}{N - M - t} \sum_{i=M+1}^{N-t} \mathcal{O}[s^{(i)}] \mathcal{O}[s^{(i+t)}] \propto \exp(-t/\tau). \qquad (6.4.2)$$

The autocorrelation time $\tau$ of the Metropolis algorithm actually increases when one approaches a second order phase transition. At a second order phase transition the correlation length $\xi$ diverges. One finds so-called critical slowing down

$$\tau \propto \xi^z, \qquad (6.4.3)$$

where $z$ is a dynamical critical exponent characterizing the efficiency of a Monte Carlo algorithm. For the Metropolis algorithm one finds $z \approx 2$, which leads to a very bad critical slowing down behavior.

# Chapter 7

# Conclusions

These lectures have only covered the most basic features of field theory on the lattice. Recent lattice developments have put chiral symmetry on a solid theoretical basis at a non-perturbative level. As discussed in the introduction, lattice QCD can now explain non-perturbatively why nucleons can exist naturally, i.e. without fine-tuning, far below the Planck scale, provided that space-time has additional hidden dimensions. This in turn explains why gravity is so weak, a non-trivial result one obtains from lattice QCD without doing any numerical work. Even chiral gauge theories like the standard model have now been constructed rigorously beyond perturbation theory. This is a very substantial step forward in the theoretical formulation of the basic laws of Nature. The development of Ginsparg-Wilson lattice fermions is also beginning to revolutionize practical lattice QCD simulations. In particular, if new algorithmic developments go hand in hand with the recent theoretical insights, Ginsparg-Wilson fermions may lead to substantial progress towards an accurate numerical solution of QCD. As usual, many new questions arise based on the new insights. For example, supersymmetry still waits to be put on rigorously solid grounds beyond perturbation theory. Also many practical numerical calculations with Ginsparg-Wilson fermions still need to be done. These notes may provide a certain basis for newcomers to enter this very active field of current research.

There is a vast literature of research articles on lattice field theory and there are some text books as well. Michael Creutz's book is still a good elementary introduction to the subject. The book by Gernot Münster and Istvan Montvay

is more complete and can be highly recommended. However, it is already a number of years old and does not cover the more recent developments, especially concerning chiral symmetry on the lattice. A more recent very good book that incorporates the breakthrough in the understanding of lattice chiral symmetry is the one authored by Thomas DeGrand and Carleton DeTar. Martin Lüscher has written a number of beautiful reviews on various advanced aspects of lattice field theory, including chiral gauge theories. These are perhaps the best pedagogical texts available at this level. However, the most exciting way to learn lattice field theory is probably to pick an interesting non-perturbative research problem and try to solve it using this very powerful technique.

# Acknowledgements

I like to thank Andreas Wipf for the invitation to lecture in Saalburg. Large parts of these lecture notes are taken from a review article written together with Shailesh Chandrasekharan. I like to thank him for an inspiring collaboration for over more than a decade.

# Chapter 8

# Exercises

The following simple exercises may be useful to get started working on lattice field theory problems.

## 1) Path Integral for a Free Particle

A free non-relativistic particle of mass $m$ moves on an infinite line. Use Feynman's path integral in real time to evaluate the transition amplitude for going from the point $x$ at time $t = 0$ to the point $x'$ at time $t = T$.

## 2) Free Particle on a Circle

A free non-relativistic particle of mass $m$ moves on a circle of circumference $L$. Use Feynman's path integral in real time to evaluate the transition amplitude for returning to the point $x$ after a time $T$. Extract the energy spectrum from this result.

## 3) Transfer Matrix of the 1-dimensional Ising Model

Consider the $\mathbf{Z}(2)$-symmetric Ising model with spin variables $s_x = \pm 1$ on a 1-dimensional periodic lattice with $N$ points and with the nearest-neighbor Hamilton function

$$\mathcal{H}[s] = -J \sum_{\langle xy \rangle} s_x s_y.$$

Evaluate the partition function

$$Z = \prod_{x=1}^{N} \sum_{s_x = \pm 1} \exp(-\beta \mathcal{H}[s]) = \mathrm{Tr} \mathcal{T}^N,$$

and analyze the $N$-dependence in order to extract the spectrum of the transfer matrix

$$\mathcal{T}(s_x, s_y) = \exp(\beta J s_x s_y).$$

Determine the correlation length as a function of $\beta$. Is there a second order phase transition, and if so at what temperature?

## 4) Spectrum of the 1-dimensional XY-Model

Consider the $O(2)$-symmetric XY-model with variables $\varphi_x \in [-\pi, \pi]$ on a 1-dimensional lattice with the nearest-neighbor Hamilton function

$$\mathcal{H}[\varphi] = -J \sum_{\langle xy \rangle} \cos(\varphi_x - \varphi_y).$$

Construct the transfer matrix $\mathcal{T}(\varphi_x, \varphi_y)$ and determine its spectrum. Determine the correlation length as a function of $\beta$. Is there a second order phase transition, and if so at what temperature?

## 5) Lattice Scalar Field

Consider a real-valued free scalar field $\Phi$ on a 4-dimensional space-time lattice. Derive the 2-point function $\langle \Phi(-p) \Phi(p) \rangle$ in momentum space. Use this result in order to obtain the lattice dispersion relation between energy $E$ and momentum $p_i$.

# 6) Fermionic Lattice Path Integral

Consider Grassmann variables $\Psi_x$, $\bar{\Psi}_x$, $\Psi_y$ and $\bar{\Psi}_y$ on a lattice with two points $x$ and $y$. The fermionic Euclidean action is given by

$$S[\bar{\Psi}, \Psi] = \bar{\Psi}_x \Psi_y + \bar{\Psi}_y \Psi_x + m(\bar{\Psi}_x \Psi_x + \bar{\Psi}_y \Psi_y).$$

Evaluate the partition function

$$Z = \int \mathcal{D}\bar{\Psi}\mathcal{D}\Psi \ \exp(-S[\bar{\Psi}, \Psi]),$$

using the rules for Grassmann integration. Determine the values of the 2-point function $\langle \bar{\Psi}_x \Psi_x \rangle$ and $\langle \bar{\Psi}_x \Psi_y \rangle$.

# 7) From the Group to the Algebra

Assume that $\Omega(x)$ is an $SU(N_c)$ gauge transformation, and prove that the expression $i\Omega(x)\partial_\mu \Omega(x)^\dagger$ is in the $su(N_c)$ algebra.

# 8) Gauge Covariance of a Parallel Transporter

Consider a parallel transporter $U_\mathcal{C} = \mathcal{P} \exp \int_\mathcal{C} dx_\mu A_\mu(x)$ constructed from a non-Abelian $SU(N)$ gauge field $A_\mu(x)$ along an open curve $\mathcal{C}$ from $x_0$ to $x_1$. Show that the gauge transformation $A_\mu(x)' = \Omega(x)(A_\mu(x) + \partial_\mu)\Omega(x)^\dagger$ induces the transformation $U'_\mathcal{C} = \Omega(x_0) U_\mathcal{C} \Omega(x_1)^\dagger$.

# 9) Metropolis Algorithm for the 1-d Ising Model

The purpose of this exercise is to develop a Metropolis Monte Carlo code for the 1-d Ising model. It would be useful to have a MAIN program from which appropriate subroutines are called. In the MAIN program one could define a

field (e.g. SPIN(N)) that contains the spin value for each lattice point N. It is also useful to define fields that contain the left and right neighbors of a given lattice point (e.g. ILEFT(N) and IRIGHT(N)). In this way periodic boundary conditions are easily implemented.

The MAIN program should call a subroutine START which sets an initial spin configuration. In principle, it does not matter which initial configuration one picks. However, it will be a good check of the procedure that the final answers should be statistically independent of the initial configuration. Hence, it would be useful to allow for different initial configurations, e.g. a completely ordered one with all spins parallel, or a random one. In order to assign a random value to an Ising spin, one can put it to 1 if a random number is larger than $1/2$ and to $-1$ otherwise.

The program should then contain a subroutine METROPOLIS in which one Metropolis sweep through the lattice is performed. In one sweep each spin is subjected to one Metropolis update, i.e. its neighbors are examined and the change of the energy under spin flip is calculated. Then the spin is flipped or not, according to the rules of the Metropolis algorithm. If the spin should be flipped with probability $p$, one can pick a random number $x$ and flip the spin if $x < p$.

There could also be a subroutine MAGNETIZATION which measures the magnetization of a given configuration. This value is then used in the computation of the susceptibility. An even more efficient method is to keep track of the magnetization after each individual spin update.

An important part of each Monte Carlo routine is the error analysis. When the Monte Carlo data are written into a file, the error analysis can be performed with a separate program.

Finally, we like to compare the results of our Monte Carlo simulation with the exact analytic results. For this purpose we need a separate routine that evaluates the exact result numerically. One can then compare Monte Carlo and analytic results and decide if they agree within the statistical errors. At first they will usually not and one must start the sometimes painful process of debugging the code.

# 10) Metropolis Algorithm for the 2-d Ising Model

The Monte Carlo program for the 2-d Ising model can be based on the one for the 1-d model and requires only minor modifications. In particular, the neighbor lists must be modified and the change of the energy under spin flip must take into account the interactions of a spin with all four neighbors.

One can now measure the susceptibility and show that it peaks near the analytically known critical temperature. It is also useful to monitor the statistical errors as one approaches the critical temperature. One will then see that the Metropolis algorithm becomes rather inefficient. It suffers from critical slowing down, i.e. the number of sweeps between statistically independent spin configurations increases dramatically when one approaches criticality.