Introduction to Lattice Field Theory

Matthew Wingate

DAMTP, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom

E-mail: M.Wingate@damtp.cam.ac.uk

Abstract: Lecture notes prepared for BUSSTEPP at University of Nottingham, September 2025.
[This version was compiled 2 September 2025 at 12:34.]

Contents

1 Introduction **1**

2 Path integrals **2**

3 Euclidean spacetime **2**

4 Discretization of scalar fields **3**

5 Energy spectrum **4**

6 Gauge fields **5**
6.1 Discretization of gauge fields 7
6.2 Confinement at strong coupling 9

7 Renormalization and the continuum limit **11**

8 Markov Chain Monte Carlo **14**
8.1 Importance sampling 14

9 Lattice fermions **19**
9.1 Discretization 19
9.2 Doubling in 1+1 dimensions 20
9.3 Wilson fermions 21
9.4 Staggered fermions 22
9.5 Correlation functions 23
9.6 Chirally symmetric fermions 23

1 Introduction

Lattice field theory has become an important approach to solving problems in strongly coupled gauge theories, especially quantum chromodynamics (QCD). Here we introduce the main ideas.

These lectures are intended for students beginning a PhD in theoretical high energy physics. Some familiarity with path integrals, renormalization, and gauge theory is assumed. While some in the audience may be using lattice field theory in their research, most will not. Therefore, the lectures are intended to focus on the main ideas, of interest to nonpractitioners as well as future experts.

2 Path integrals

In quantum field theories, the basic quantities of interest are vacuum matrix elements of field operators. For example, in a scalar field theory, the LSZ reduction formula relates scattering amplitudes to matrix element; for m incoming particles scattering into n outgoing particles we have

$$\begin{aligned}\langle k_1' \cdots k_n' | k_1 \cdots k_m \rangle = (i)^{m+n} \int d^4x_1 \cdots d^4x_m \, d^4x_1' \cdots d^4x_n' \\ \times \exp\left[-i \sum_{j=1}^{m} k_j \cdot x_j + i \sum_{j=1}^{n} k_j' \cdot x_j' \right] \\ \times \prod_{j=1}^{m} (\partial_j^2 + m^2) \prod_{j=1}^{n} (\partial_{j'}^2 + m^2) \\ \times \langle \Omega | \mathcal{T} \phi(x_1) \cdots \phi(x_m) \phi(x_1') \cdots \phi(x_n') | \Omega \rangle \,. \end{aligned} \tag{2.1}$$

Here $|\Omega\rangle$ denotes the vacuum and $\mathcal{T}$ denotes time-ordering (e.g. see §5 of [1].)

Vacuum matrix elements can be expressed as path integrals, for example

$$\langle \Omega | \mathcal{T} \phi(x_1) \cdots \phi(x_n) | \Omega \rangle = \frac{1}{Z} \int \mathcal{D}\varphi \, \varphi(x_1) \cdots \varphi(x_n) \, e^{iS_M[\varphi]} \tag{2.2}$$

$S_M[\phi] = \int d^4x \, \mathcal{L}[\varphi]$ is the action in Minkowski spacetime (in which I've implicitly used the mostly-minus signature above) and $Z = \int \mathcal{D}\varphi \, e^{iS_M[\varphi]}$. Note that I have used a different font to distinguish the quantum field operator ϕ from the classical field integration variable φ.

3 Euclidean spacetime

The approach we take later will not work in Minkowski spacetime. We would like to approximate integrals like the right-hand side of (2.2). However, the complex phase $\exp(iS_M)$ makes the integrand highly oscillatory. Instead, we switch to an imaginary time $x_4 = ix_0$, with $x_4 > 0$ being forward in imaginary time. Then we seek Euclidean spacetime correlation functions

$$\langle \Omega | \mathcal{T} \phi(x_1) \cdots \phi(x_n) | \Omega \rangle = \frac{1}{Z_E} \int \mathcal{D}\varphi \, \varphi(x_1) \cdots \varphi(x_n) \, e^{-S_E[\varphi]} \,. \tag{3.1}$$

Being able to learn anything about field theory in Minkowski time requires analytic continuation from imaginary time back to real time. This is not always possible. The correlation functions we compute should not have any nonanalyticities, such as poles or branch cuts arising in many scattering amplitudes. This limits what can be computed this way. Nevertheless, there are many things we can learn from analytic correlation functions. For example, one can determine the low-lying spectrum of the theory and one can compute matrix elements which govern decays, for example the electromagnetic or weak decays of hadrons.

4 Discretization of scalar fields

Let us introduce a rectangular, 4-dimensional lattice with lattice spacing a, so that the lattice sites are located at the following points

$$x \;=\; a(n_1, n_2, n_3, n_4)\,, \qquad n_\mu \in [0, N_\mu - 1]\,. \tag{4.1}$$

(Generalizations to other dimensions are straightforward.) On the lattice, derivatives are replaced by finite difference operators. There is not a unique choice for the replacement as we shall see. Consider the forward difference operator

$$(\nabla^+_\mu \varphi)(x) := \frac{1}{a}\left[\varphi(x + a\hat{\mu}) - \varphi(x)\right] = \partial_\mu \varphi(x) \;+\; \frac{a}{2}\partial^2_\mu \varphi(x) \;+\; \frac{a^2}{6}\partial^3_\mu \varphi(x) \;+\; \ldots \tag{4.2}$$

In the second step, we have Taylor expanded the field $\varphi(x + a\hat{\mu})$ about $a = 0$. Therefore, *if* a continuum limit exists, then ∇^+_μ approaches ∂_μ, with corrections at finite a given by a Taylor-like expansion (more later).

As we remarked, we might equally well choose instead a backward difference operator

$$\nabla^-_\mu \varphi := \frac{1}{a}\left[\varphi(x) - \varphi(x - a\hat{\mu})\right] = \partial_\mu \varphi \;-\; \frac{a}{2}\,\partial^2_\mu \varphi \;+\; \frac{a^2}{6}\,\partial^3_\mu \varphi \;+\; \ldots \tag{4.3}$$

It is well-known in numerical analysis that one can subtract, order-by-order, discretization errors by taking combinations of difference operators. For example the symmetric difference operator

$$\nabla^\pm_\mu \varphi := \frac{1}{2}\left(\nabla^+_\mu + \nabla^-_\mu\right)\varphi = \partial_\mu \varphi \;+\; \frac{a^2}{6}\,\partial^3_\mu \varphi \;+\; \ldots \tag{4.4}$$

has no $O(a)$ discretization errors. There is a strategy for improving lattice operators in the manner, due to Symanzik, as we will discuss later. It is complicated by the fact that quantum effects renormalize the coefficients necessary to cancel terms order-by-order in a.

The Lagrangian for a real scalar field can be discretized as

$$\mathcal{L} = \frac{1}{2}\varphi\left(-\partial^2 + m^2\right)\varphi \;\to\; \frac{1}{2}\varphi\left(-\sum_\mu \nabla^-_\mu \nabla^+_\mu \;+\; m^2\right)\varphi\,. \tag{4.5}$$

The corresponding action $S = a^4 \sum_x \mathcal{L}$ is

$$S = -\frac{a^2}{2}\sum_{x,\mu} \varphi(x)[\varphi(x + a\hat{\mu}) + \varphi(x - a\hat{\mu})] \;+\; \frac{a^2}{2}(8 + m^2 a^2)\sum_x \varphi(x)\varphi(x)\,. \tag{4.6}$$

Exercise 1. *Derive an expression for the free lattice scalar propagator in momentum space, given the action (4.6).*

5 Energy spectrum

We have a lattice action in Euclidean spacetime. In this section, let us illustrate one of the main applications of lattice field theory methods, determining the spectrum of the theory. We begin by defining the so-called transfer matrix (e.g. see [2]).

For shorthand, denote the 3-dimensional slice of the scalar field φ with time coordinate $t \equiv x_4/a$ as $\Phi_t := \{\varphi(x) | x_4 = at\}$ and write the action as

$$S[\phi] = \sum_t L[\Phi_{t+1}, \Phi_t] \,. \tag{5.1}$$

Then we define the *transfer matrix* T to be composed of elements

$$T[\Phi_{t+1}, \Phi_t] = \exp\left(-L[\Phi_{t+1}, \Phi_t]\right) \,. \tag{5.2}$$

This can be thought of as an imaginary-time evolution operator, acting on wave functions Ψ depending on the timeslice field Φ_t: $|\Psi_t\rangle$ or $\Psi(\Phi_t)$ as

$$\begin{aligned} |\Psi_{t+1}\rangle &= T|\Psi_t\rangle \\ \text{or } \Psi(\Phi_{t+1}) &= \int \left[\prod_{\vec{x}} d\varphi(\vec{x}, t)\right] T[\Phi_{t+1}, \Phi_t]\Psi(\Phi_t) \,. \end{aligned} \tag{5.3}$$

Correlation functions can then be expressed using the transfer matrix. For example the 2-point function

$$\begin{aligned} \langle\phi(x_1)\phi(x_2)\rangle &= \frac{1}{Z}\int \left[\prod_x d\varphi(x)\right] \varphi(x_1)\varphi(x_2) \prod_{t=0}^{\infty} T[\Phi_{t+1}, \Phi_t] \\ &= \lim_{N_t\to\infty} \frac{\mathrm{Tr}\left(T^{N_t-\Delta t}\,\phi(x_1)\,T^{\Delta t}\,\phi(x_2)\right)}{\mathrm{Tr}\; T^{N_t}} \end{aligned} \tag{5.4}$$

where $\Delta t = |(x_1)_4 - (x_2)_4|/a$ and, adopting notation from statistical mechanics, the trace is over all possible timeslice configurations $\{\Psi_t\}$.

The transfer matrix is the link from the Euclidean lattice fields to the Hilbert space. A lattice Hamiltonian H can be defined in terms of the transfer matrix

$$H := -\frac{1}{a}\log T \,. \tag{5.5}$$

Rewriting (5.4)

$$\langle\phi(x_1)\phi(x_2)\rangle = \lim_{N_t\to\infty} \frac{\mathrm{Tr}\left(e^{-aH(N_t-\Delta t)}\,\phi(x_1)\,e^{-aH\Delta t}\,\phi(x_2)\right)}{\mathrm{Tr}\; e^{-aHN_t}} \tag{5.6}$$

puts the correlation function in a useful form. In finite volume, the spectrum of states is discrete, and we take the trace to be over the eigenstates of H. Inserting a complete set of H eigenstates to the left of each $\phi(x)$, we have (letting $\vec{x}_1 = \vec{x}_2 = 0$ for simplicity)

$$\langle\phi(x_1)\phi(x_2)\rangle = \sum_{n\geqslant 0} |\langle 0|\phi(0)|n\rangle|^2\, e^{-aE_n\Delta t} \,. \tag{5.7}$$

where the energies E_n are eigenvalues of H with $E_1 < E_2 < \ldots$ and we take the vacuum energy to be 0. Considering the sum with only $n > 0$ gives the connected Green's function

$$\langle \phi(x_1)\phi(x_2)\rangle_c = \sum_{n>0} |\langle 0|\phi(0)|n\rangle|^2 \, e^{-aE_n \Delta t} \, . \tag{5.8}$$

For large Δt, the function tends to a single exponential, with a decay rate given by the excitation energy E_1 and amplitude $|\langle 0|\phi(0)|n\rangle|^2$ (which we assume to be nonzero)

$$\langle \phi(x_1)\phi(x_2)\rangle_c \sim |\langle 0|\phi(0)|1\rangle|^2 \, e^{-aE_1 \Delta t} \, . \tag{5.9}$$

This demonstrates the connection between two-point functions and the spectrum of the theory.

The energy difference $E_1 - E_0$ has a corresponding length scale $\xi = (E_1 - E_0)^{-1}$. Since it is the longest physical length scale, it acts as a correlation length, in the statistical field theoretic sense. In order for the lattice theory to approach a continuum limit, we require that $\xi/a \to \infty$. For those of you who have studied critical phenomena, this should remind you of a second order phase transition, which is characterized by a divergent correlation length.

6 Gauge fields

Consider a complex scalar field theory, where φ has N components. Let us integrate the continuum action by parts to give a more convenient expression

$$S = \int d^4x \, \varphi^\dagger(-\partial^2 + m^2)\varphi + V(\varphi^\dagger\varphi) = \int d^4x \, \left[(\partial_\mu\varphi)^\dagger(\partial^\mu\varphi) + m^2\varphi^\dagger\varphi + V(\varphi^\dagger\varphi)\right] \, . \tag{6.1}$$

Say S is invariant under global transformations of the form

$$\varphi(x) \mapsto \Lambda \, \varphi(x) \tag{6.2}$$

where Λ is an $N \times N$ complex matrix. This implies the existence of a group G of transformations which leave the action invariant. The corresponding set of matrices Λ form an N-dimensional representation of G.

Let us suppose G is a Lie group, such as $SU(n)$. If we want to gauge this symmetry, that is, if we want an action invariant under local transformations

$$\varphi(x) \mapsto \Lambda(x)\varphi(x) \, , \tag{6.3}$$

then we must modify the action; the kinetic term is no longer invariant. We want to find a gauge-covariant derivative D_μ such that $D_\mu\varphi$ transforms the same way as φ.

For inspiration, let us write the derivative in the direction of unit vector n^μ as

$$n^\mu \partial_\mu \varphi = \lim_{a\to 0} \frac{1}{a} \left[\varphi(x + an) - \varphi(x)\right] \, . \tag{6.4}$$

We will refer to this formal $a \to 0$ limit as the *classical continuum limit*. This is distinct from the physical continuum limit we will discuss in Section 7. Under (6.3), the difference on the righthand side of (6.4) transforms as

$$\varphi(x + an) - \varphi(x) \mapsto \Lambda(x + an)\varphi(x + an) - \Lambda(x)\varphi(x) \,. \tag{6.5}$$

Let us define a *Wilson line*, $U(y, x)$ going from x to y, with the transformation property

$$U(y, x) \mapsto \Lambda(y)\, U(y, x)\, \Lambda(x)^{-1} \,. \tag{6.6}$$

To be completely precise, we need to specify the particular path taken from x to y, say the oriented curve $\mathcal{C}_{yx}$ connects x to y. We require the following other properties to hold:

1. The U belong to the same representation as the Λ.

2. The trivial Wilson line, corresponding to a curve of zero length, is the identity:

$$U(0) = I_N \,. \tag{6.7}$$

3. If we write $\mathcal{C}_{xy}$ as the reverse orientation of $\mathcal{C}_{yx}$, then

$$U(\mathcal{C}_{xy}) = U(\mathcal{C}_{yx})^{-1} \,. \tag{6.8}$$

4. Given a curve $\mathcal{C}_{yx}$ from x to y and another curve $\mathcal{C}_{zy}$ from y to z, we can combine them to form the curve from x to z (through y), written as $\mathcal{C}_{zy} \circ \mathcal{C}_{yx}$. The corresponding Wilson lines satisfy

$$U(\mathcal{C}_{zy} \circ \mathcal{C}_{yx}) = U(\mathcal{C}_{zy})U(\mathcal{C}_{yx}) \,. \tag{6.9}$$

Below, it will usually be clear what path we mean, so we will usually just specify the endpoints.

We can use the Wilson line to construct our desired covariant derivative. By including a Wilson line in the second term,

$$n^\mu D_\mu = \lim_{a \to 0} \frac{1}{a} \left[\varphi(x + an) - U(x + an, x)\varphi(x)\right] \,, \tag{6.10}$$

we find that, at least before taking the limit $a > 0$, $D_\mu\varphi$ clearly transforms in the same way as φ (6.3). Therefore, $(D_\mu\varphi)^\dagger\, D^\mu\varphi$ is gauge invariant.

In mathematical jargon, the Wilson line is a parallel transporter: at each coordinate $\varphi(x)$ is a vector in a vector space $V_x \cong \mathbb{C}^N$, and U is a map $U(y, x) : V_x \to V_y$.

Let us examine the classical $a \to 0$ limit and make contact with standard treatments of gauge theory. Recalling that $U(0) = I_N$, near $a = 0$ we can we can use the exponential map from the Lie group G to its Lie algebra $\mathfrak{g}$. We can define $n^\mu A_\mu(x)$ as the field at the midpoint along a small Wilson line.

$$U(x + an, x) = \exp\left[igan^\mu A_\mu(x + \tfrac{a}{2}n)\right] \,. \tag{6.11}$$

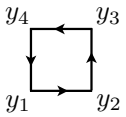

Figure 1. A square Wilson loop called a *plaquette*.

This field is an element of the algebra representation corresponding to the group representation of the Wilson line U. That is, we can write

$$A_\mu(x) = A^a_\mu(x) T^a \tag{6.12}$$

where the implicit sum over the index a goes from 1 up to the dimension of $\mathfrak{g}$, and the $N \times N$ matrices T^a are Hermitian generators of $\mathfrak{g}$ in the appropriate representation. For example, if $G = SU(2)$, and $N = 2$, then the $T^a = \frac{1}{2}\sigma^a$ are the Pauli matrices; for larger N, then the T^a are a set of 3 linearly independent, traceless, Hermitian $N \times N$ matrices.

Taking the classical $a \to 0$ limit of (6.10) we have

$$D_\mu \varphi(x) = \left[\partial_\mu + igA_\mu(x)\right] \varphi(x)\,. \tag{6.13}$$

We see that we now arrive at the familiar gauge-covariant derivative in the continuum.

6.1 Discretization of gauge fields

In lattice gauge theory, we follow the steps described above, stopping short of taking the formal $a \to 0$ limit. The gauge degrees of freedom are the shortest Wilson lines, i.e. those living on the links connecting neighbouring lattice sites. When combining gauge fields with scalar matter (or similarly with fermionic matter) we use the finite difference operator in (6.10).

As in the continuum, we need to include the gauge field's kinetic energy and any self-couplings in the theory's action. Once again, we consider Wilson lines, this time closed Wilson lines, called *Wilson loops*.

Let us consider the smallest nontrivial loop on the lattice: the *plaquette*, a square composed of four straight Wilson lines of length a (Fig. 1). For example, consider a plaquette in the x_1 - x_2 plane can be written as

$$P_{12}(x) = U(y_1, y_4) U(y_4, y_3) U(y_3, y_2) U(y_2, y_1)\,, \tag{6.14}$$

with $y_1 = x$, $y_2 = x + a\vec{e}_1$, $y_3 = x + a(\vec{e}_1 + \vec{e}_2)$, $y_4 = x + a\vec{e}_2$. In lattice literature, more common notation is to write $U(x + a\hat{\mu}, x) \equiv U_\mu(x)$, so that

$$P_{\mu\nu}(x) = U_\nu(x)^\dagger U_\mu(x + a\hat{\nu})^\dagger U_\nu(x + a\hat{\mu}) U_\mu(x)\,, \tag{6.15}$$

Expanding about small a, after some work we find

$$P_{12}(x) = I + iga^2 F_{12}(x) + O(a^2)\,. \tag{6.16}$$

where $F_{12}(x)$ is the field strength tensor in the x_1 - x_2 plane

$$F_{\mu\nu}(x) = \nabla^+_\mu A_\nu(x) - \nabla^+_\nu A_\mu(x) + [A_\mu, A_\nu]\,. \tag{6.17}$$

In nonabelian gauge theories, a closed line is not gauge-invariant on its own. For a non-trivial closed curve $\mathcal{C}$, $U_\mathcal{C}(y,y) \neq I$ and does not commute with general gauge transformation $\Lambda(x)$

$$U_\mathcal{C}(y,y) \mapsto \Lambda(y) U_\mathcal{C}(y,y) \Lambda^\dagger(y)\,. \tag{6.18}$$

We define a gauge-invariant *Wilson loop* to be the trace over matrix indices

$$\mathrm{Tr}\,\Lambda(y) U_\mathcal{C}(y,y) \Lambda^\dagger(y) = \mathrm{Tr}\, U_\mathcal{C}(y,y)\,. \tag{6.19}$$

Wilson's plaquette action for $SU(n)$ gauge theory, with $\beta = 2n/g_0^2$, is written as

$$S_\Box[U] \;=\; \beta \sum_{x,\mu,\nu>\mu} \left(1 \,-\, \frac{1}{2n} \mathrm{Tr}\,[P_{\mu\nu}(x) \,+\, P^\dagger_{\mu\nu}(x)]\right) \tag{6.20}$$

with $P_{\mu\nu}(x)$ as in (6.15), taking the links to act on the fundamental representation. That is, the links are elements of $SU(n)$.

Exercise 2. *Show that the plaquette action (6.20) has the correct classical continuum limit.*

With the action in hand, we can consider expectation values of observables, e.g a Wilson loop

$$\langle W(\ell) \rangle \;=\; \frac{1}{Z} \int [dU]\, \mathrm{Tr} \prod_{\mathcal{C}\in\ell} U(\mathcal{C})\, e^{-S_\Box[U]}\,. \tag{6.21}$$

For simplicity, we assume from now on that the U are acting on the fundamental representation.

To make the integral over the group finite and well-defined, we need a measure. Requiring the measure to be linear, positive, invariant and normalized essentially determines this Haar measure [3]. Integrating over a single element of $SU(3)$, with group indices explicit, we have

$$\int dU \;=\; 1 \tag{6.22}$$

$$\int dU\, U^{ab} \;=\; 0 \tag{6.23}$$

$$\int dU\, U^{ab} U^{cd} \;=\; 0 \tag{6.24}$$

$$\int dU\, U^{ab} (U^\dagger)^{cd} \;=\; \tfrac{1}{3}\, \delta^{ad}\delta^{bc} \tag{6.25}$$

$$\int dU\, U^{ab} U^{cd} U^{ef} \;=\; \tfrac{1}{3!}\, \epsilon^{ace}\epsilon^{bdf} \tag{6.26}$$

The reason the last two integrals do not vanish for the indices indicated, is that the integrand is equal to 1. For unitary groups, $UU^\dagger = 1$, and for $SU(3)$ the product of 3 group elements

contains a singlet component. Denoting irreducible representations of a group by their dimensions (and including a bar for complex conjugation) the preceding statements can be understood from $\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{1} \oplus \mathbf{8}$ and $\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{1} \oplus \mathbf{8} \oplus \mathbf{10}$. These facts will be useful in the strong coupling expansion.

Since we have a gauge theory formulated on a finite lattice using elements of the compact gauge group, path integrals are finite. In particular, there is no requirement to fix to a particular gauge; the integration averages over all gauges. Expectation values of gauge non-invariant operators vanish. For example, let O be a gauge non-invariant (local) operator, so that under a gauge transformation $\Lambda(x)$, $O[U'](x) = \Lambda O[U](x)$. Let us concentrate on $x = 0$ and perform a gauge transformation with $\Lambda(0) = \exp(2\pi i n)\, I_n$ an element of $Z(n)$, the centre of $SU(n)$, leaving the $\Lambda(x), x \neq 0$ arbitrary. The measure and action are gauge invariant, thus after the gauge transformation

$$\langle O \rangle \;=\; \frac{1}{Z} \int [dU] e^{2\pi i/n} O[U] e^{-S_\Box[U]} \;=\; e^{2\pi i/n} \langle O \rangle \,.$$

Repeating the transformation again n times yields

$$\langle O \rangle \;=\; \sum_{k=1}^{n} e^{2\pi i k/n} \langle O \rangle \;=\; 0 \,. \tag{6.27}$$

A general proof of the vanishing of non-invariant expectation values was given by Elitzur.

6.2 Confinement at strong coupling

The behavior of Wilson loops is an indicator of the phenomenon we call as confinement. Actually the definition of confinement is a subtle business which we do not have time to treat properly. A good review is given by [4].

Area law vs. perimeter law: Consider a Wilson loop in the shape of a $R \times T$ rectangle lying in the $(0, 0, x_3, x_4)$ plane, say. In the limit of very large T, one can interpret the dependence on R as a potential $V(R)$ between 2 static test charges, a heavy quark and anti-quark.

$$\langle W(R,T) \rangle \;\sim\; e^{-V(R)T} \tag{6.28}$$

Linear confinement implies for large R

$$V(R) \;\sim\; kR \tag{6.29}$$

thus a general Wilson loop will have an expectation value decaying exponential with the area enclosed by the rectangle

$$\langle W(R,T) \rangle \;\sim\; e^{-kRT} \,. \tag{6.30}$$

If the theory is not confining then $V(R)$ could approach a constant, 2 times a free quark's self-energy, say M, and the Wilson loop expectation value will have a exponential decay characterized only by the temporal extent of the Wilson loop

$$\langle W(R,T) \rangle \;\sim\; e^{-2MT} \,. \tag{6.31}$$

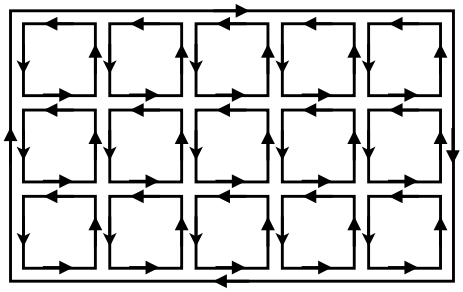

Figure 2. Example of a leading-order strong-coupling diagram for an $SU(n)$ Wilson loop.

This case is referred to as a perimeter-law behaviour, in distinction to the area law (6.30) – this terminology will more sense after some examples. Another possibility is realized for theories in a Coulomb phase where one would find

$$\langle W(R,T)\rangle \;\sim\; e^{-\frac{\alpha T}{4\pi R}}\,. \tag{6.32}$$

In the strong coupling limit of $SU(n)$ lattice gauge theory we can show analytically that confinement occurs, at least as indicated by area-law behaviour of Wilson loops. We expand about $\beta = 2n/g_0^2 = 0$. (Recalling the analogy with statistical physics, this is sometimes interpreted as a high temperature expansion, although this is not the same as the physical high temperature limit of the gauge theory, which occurs as aN_t becomes small compared to physical length scales.) Let us take the plaquette action (6.20) again, this time ignoring the constant term which cancels in the numerator and denominator of expectation values,

$$S_\Box[U] \;=\; -\frac{\beta}{2n}\sum_{x,\mu,\nu>\mu}\mathrm{Tr}\,(P_{\mu\nu}(x)+P_{\mu\nu}(x)^\dagger) \;\equiv\; -\beta Q[U]\,. \tag{6.33}$$

We can expand

$$e^{\beta Q[U]} \;=\; \sum_{n=0}^{\infty}\frac{(\beta Q[U])^n}{n!}\,. \tag{6.34}$$

Strong coupling implies small β, so we seek the leading contribution. Does the power of β scale like the area of the Wilson loop, the temporal extent, or as a Coulomb law?

Let us consider a Wilson loop which is a rectangle in the $(0,0,x_3,x_4)$ plane, say following a path which connects the corners $(x_3,x_4) = (0,0)$, $(0,R)$, (R,T), $(0,T)$, in that orientation of the large rectangle in Figure 2. Thus take

$$\begin{aligned} W(R,T) \;=\; \mathrm{Tr}\Big[&U_3^\dagger(0)\,U_3^\dagger(a\vec{e}_3)\cdots U_3^\dagger(j\,a\vec{e}_3)\cdots U_3^\dagger((R-1)a\vec{e}_3)\\ &\times\; U_4^\dagger(Ra\vec{e}_3)\cdots U_4^\dagger(Ra\vec{e}_3+(T-1)a\vec{e}_4)\\ &\times\; U_3((R-1)a\vec{e}_3+Ta\vec{e}_4)\cdots U_3(Ta\vec{e}_4)\\ &\times\; U_4((R-1)a\vec{e}_3)\cdots U_4(a\vec{e}_4)\,U_4(0)\Big]\,. \end{aligned} \tag{6.35}$$

Its expectation value is the integral

$$\langle W(R,T)\rangle \;=\; \frac{1}{Z}\int\left[\prod_{x,\mu}dU_\mu(x)\right]W(R,T)\sum_{n=0}^{\infty}\frac{(\beta Q[U])^n}{n!}\,. \tag{6.36}$$

As the integrals (6.22)–(6.26) suggest, the leading order term in the strong coupling expansion of (6.36) will have to have an integrand equal to 1. That is, all the link variables composing the Wilson loop will need to be canceled by terms in the action. For every link in $W(Z,T)$ (6.35), we will need the corresponding inverse link from a plaquette in the action. These plaquettes will involve new links, which integrate to zero if unmatched by their inverse. The lowest order term in β which has each link and its inverse appearing together is the one depicted in Figure 2, the term where the interior of the large Wilson loop is tiled with plaquettes. Therefore, we see that

$$\langle W(R,T)\rangle \sim \beta^{RT} \simeq e^{-kRT}\,. \tag{6.37}$$

Large $SU(n)$ Wilson loops have area-law behaviour, therefore the potential $V(R)$ we infer is linear in R, as in (6.30).

Exercise 3. *In the strong coupling limit of $U(1)$ gauge theory, explicitly show that Wilson loops obey an area law. Explain why the next-to-leading order term is $O(\beta^{RT+4})$. We know continuum QED is not a confining theory, i.e. static charges have a Coulomb potential $V(R) \propto 1/R$, not linear $V(R) \propto R$. How can we reconcile this fact with the strong coupling lattice result?*

Exercise 4. *[Advanced] In the case of $SO(3)$ gauge theory, show that the strong coupling expansion for the Wilson loop obeys a perimeter law. [Recall that $SO(3)$ is the group of rotations in 3 dimensions and that its irreducible representations are of dimension $2\ell+1$, where $\ell = 0, 1, 2, \ldots$ and represents angular momentum in some contexts. The key fact is that, in this case, there is a singlet contribution to $\mathbf{3}\otimes\mathbf{3}$ and $\mathbf{3}\otimes\mathbf{3}\otimes\mathbf{3}$ representations. There is no distinct $\bar{\mathbf{3}}$ representation, so Wilson lines do not need arrows in the way they do for $SU(3)$.]*

7 Renormalization and the continuum limit

We introduced the classical continuum limit above as simply expanding about $a = 0$ and taking the $a \to 0$ limit. This is not enough for meaningful physics calculations. What we mean by the physical continuum limit, is that a becomes arbitrarily small compared to any physical length scale, say r. From the perspective of the lattice, then, physical length scales must diverge in the continuum limit. In a statistical field theory, this corresponds to taking the limit of diverging correlation length $\xi = r/a$, that is, tuning the parameters of the lattice theory to a second-order phase transition. Whether we view the continuum limit as taking the ultraviolet cutoff a to zero while keeping the physics fixed, or as a critical phenomenon on the lattice, the tool we need is the renormalization group [3].

Let us focus on $SU(n)$ gauge theory, where the only parameter is the coupling, which we now denote g_0. Take $\Theta(r, a, g_0(a))$ to be a dimensionless observable we can compute on the lattice. This might depend on some physical length scale r, as well as the lattice spacing and the coupling. We write $g_0(a)$ explicitly to indicate that we will need to vary this coupling as we take the continuum limit. For example, $\Theta(r, a, g_0(a))$ could be a Wilson loop (rectangle) where we wish to keep $r = Ra$ fixed as $a \to 0$. Another example comes

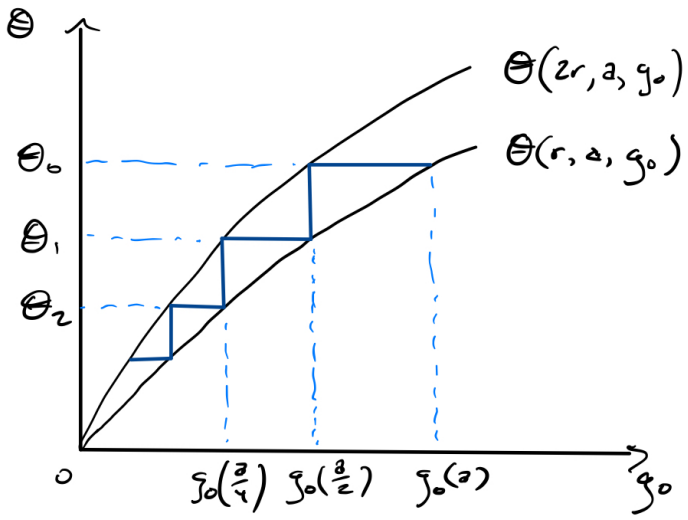

Figure 3. Step scaling procedure to determine the running coupling, case of asymptotic freedom.

from defining a force $F(r)$ between static charges, which can be made dimensionless as $\Theta(r, a, g_0(a)) = r^2 F(r)$. Finally, we could take Θ to be a ratio of masses in the theory, say the first and second excited state masses. This is often done in computations, but for the argument below, we wish to work with quantities with well-defined explicit length scale which we can vary.

If a continuum limit exists, then the a-dependence of $\Theta(r, a, g_0(a))$ should become negligible for small enough a. In this so-called "scaling regime," we could halve a and obtain the same result, as long as g_0 is suitably adjusted

$$\Theta(r, \tfrac{a}{2}, g_0(\tfrac{a}{2})) = \Theta(r, a, g_0(a)) + O(a^2)\,. \tag{7.1}$$

This is referred to as "cutoff independece". Since Θ is a dimensionless function, we can rescale (halve) our length unit on the lefthand side leaving Θ and g_0 unchanged, so that

$$\Theta(2r, a, g_0(\tfrac{a}{2})) = \Theta(r, a, g_0(a)) + O(a^2)\,. \tag{7.2}$$

Iterating these two steps n times gives

$$\Theta(2r, a, g_0(\tfrac{a}{2^{n+1}})) = \Theta(r, a, g_0(\tfrac{a}{2^n})) + O(a^2)\,. \tag{7.3}$$

We can use this "step scaling" procedure, which is a discretized renormalization group transformation, to determine how $g_0(a)$ must vary as we take the cutoff away. Suppose that we know (or can determine) the g_0 dependence of Θ. Consider the two functions $\Theta(r, a, g_0)$ and $\Theta(2r, a, g_0)$, where r and a are fixed, as in Figure 3. In what follows, we assume that we can neglect terms $O(a^2)$. Choose some initial point along the curve $\Theta(r, a, g_0(a))$, say

$$\Theta_0 := \Theta(r, a, g_0(a))\,, \tag{7.4}$$

and find the corresponding point on the second curve, i.e. determine the g_0 such that

$$\Theta(2r, a, g_0(\tfrac{a}{2})) = \Theta_0 \,. \tag{7.5}$$

Once we have $g_0(\frac{a}{2})$, we return to the first curve and identify

$$\Theta_1 := \Theta(r, a, g_0(\tfrac{a}{2})) \,. \tag{7.6}$$

In turn, we can determine $g_0(\frac{a}{4})$ through

$$\Theta(2r, a, g_0(\tfrac{a}{4})) = \Theta_1 \,. \tag{7.7}$$

We can iterate this

$$\Theta_n := \Theta(r, a, g_0(\tfrac{a}{2^n})) = \Theta(2r, a, g_0(\tfrac{a}{2^{n+1}})) \,. \tag{7.8}$$

Thus we have found how the coupling runs with the cutoff, albeit discretely. Note that this procedure works whether g_0 is large or small; our procedure has been fully nonperturbative.

A (physical) continuum limit is said to exist, if this iterative procedure "flows" to a fixed point in coupling constant space, here just in g_0. Figure 3 illustrates one possibility, that $g_0 \to 0$. This is what happens in Yang–Mills theory and in QCD and is called asymptotic freedom. Other possibilities are the g_0 flows to a nontrivial fixed point $g_0 \to g_* > 0$, or that $g_0 \to \infty$. In the latter case, we would say that no continuum limit exists.

Exercise 5. *Sketch versions of the step scaling procedure (as in Figure 3) for the cases of (a) a nontrivial fixed point and (b) the nonexistence of a continuum limit.*

Exercise 6. *In the weak coupling limit of Yang–Mills theory, the following perturbative expression for the potential between a static quark–antiquark pair can be obtained*

$$V(r, g_0, a) = \frac{C}{4\pi r}\left(g_0^2 \;+\; \frac{22}{16\pi^2}\, g_0^4 \log\frac{r}{a} \;+\; O(g_0^6)\right)$$

where C is a constant. Determine what $g_0(a)$ must be in order for the potential is independent of a (assume a is small).

8 Markov Chain Monte Carlo

8.1 Importance sampling

Now we turn to the practical side of lattice field theory: nonperturbative calculation. The equivalence of Euclidean lattice path integrals and traces over microstates in statistical mechanics points the way toward developing numerical methods.

We wish calculate expectation values of gauge invariant objects like the Wilson loop $\langle W(R,T)\rangle$ and Green's functions like $\langle\phi(x_1)\cdots\phi(x_n)\rangle$. Generically let us say we wish to calculate

$$\langle X\rangle \;=\; \frac{1}{Z}\int [dU]\, X[U]\, e^{-S[U]}\,. \tag{8.1}$$

Even for a simple group such as $U \in \mathbb{Z}_2$ where the integral is actually a discrete sum, there are too many possible field configurations to carry out an explicit, exact calculation. (Sometimes it is instructive to study lattice models with discrete gauge groups. These will not approach continuum gauge theories in the usual sense.) Instead we turn to statistical methods. Assuming the action is real and bounded from below, we can treat part of the integrand as a probability density in the space of allowed field configurations

$$\rho[U] \;=\; \frac{1}{Z}\, e^{-S[U]}\,. \tag{8.2}$$

If we could find the field configurations with the largest $\rho[U]$ then we could approximate (8.1). The fact that this distribution has the form of a Boltzmann distribution correctly suggests that only a tiny proportion of field configurations actually give non-negligible contributions to integrals such as (8.1).

We will now discuss how to generate a set of field configuration which are approximately distributed according to the distribution (8.2). Let us label a specific configuration of the gauge (or scalar, etc.) fields by $\{U\}_i$. If we have $N_{\rm conf}$ properly distributed configurations

$$\langle X\rangle \simeq \frac{1}{N_{\rm conf}}\sum_i X[\{U\}_i]\,. \tag{8.3}$$

The right-hand side approaches the left-hand side in the $N_{\rm conf}\to\infty$ limit.

The procedure of generating field configurations with a distribution approaching the physical one is called *importance sampling*. We use a *Markov process* to generate a sequence (or *Markov chain*) of configurations; given a $\{U\}_i$ the process tells us how to generate a $\{U\}_{i+1}$. To make the notation simpler, I will shortly be writing $C_i \equiv \{U\}_i$. In the lattice field theory literature, we also call this our *updating algorithm*. The defining property which makes such an updating process "Markov" is that the probability of going from any C_i to any particular C_{i+1} depends only on C_i and C_{i+1}. There is no dependence on the history of how C_i was obtained, i.e. no dependence on any C_k, with $k \leqslant i-1$.

There are further conditions which the Markov process must meet if we are to obtain the correct distribution. If an updating process satisfies the following conditions, then two theorems guarantee a valid updating algorithm [5].

C1. Irreducible: Starting from any configuration C_i it is possible to reach any another configuration C_j in a finite number of steps. That is, for any pair of configurations, there exists an $N < \infty$ such that

$$P^{(N)}(C_i \to C_j) \;\equiv\; P_{ij}^{(N)} \;=\; \sum_{\{k_\ell\}} P_{ik_1} \cdots P_{k_{N-1}j} \;>\; 0\,. \tag{8.4}$$

A Markov chain satisfying this condition is said to be ergodic, or irreducible. (I believe these terms are interchangeable, or at least coincide for the examples we care about.)

C2. Aperiodic: There is a nonzero probability that an initial configuration will be revisited after any number of steps. That is,

$$P_{ii}^{(N)} = \sum_{\{k_\ell\}} P_{ik_1} \cdots P_{k_{N-1}i} \;>\; 0\,, \quad \forall_{N>0}\,. \tag{8.5}$$

C3. Positive recurrent: The theorems also require all the states (i.e. the configurations) of our Markov chain to be finite. Define $p_{ii}^{(N)}$ to be the probability of the system returning to state C_i from C_i in N steps *without* returning to C_i in any intermediate step. (This requirement does not hold for $P_{ii}^{(N)}$ above.) We define the mean recurrence time of state C_i to be

$$\tau_i := \sum_{N=1}^{\infty} N p_{ii}^{(N)}\,. \tag{8.6}$$

A Markov chain is said to be positive recurrent if the mean recurrence time of all states C_i are finite.

A Markov chain for which these conditions hold is said to be **ergodic**.

Let us consider a few simple examples of Markov chains which break some of the conditions above. Take the integers as our states: $C_i = n \in \mathbb{Z}$ and consider some random walks.

RW1. From any "state" n, we take $P_{n,n-2} = P_{n,n+2} = \frac{1}{2}$ with all other probabilities zero. This is a reducible chain since we have no chance to switch parity; if we start with even/odd n, then the algorithm will never reach an odd/even state.

RW2. Instead take $P_{n,n-1} = P_{n,n+1} = \frac{1}{2}$ with all other probabilities zero. This is irreducible; any integer can eventually be reached in a finite number of steps (however large). However, this chain is periodic. If we start in state n, we can only return to n after an even number of steps: $P_{nn}^{2K+1} = 0$ for any K.

RW3. Finally consider $P_{n,n-1} = P_{nn} = P_{n,n+1} = \frac{1}{3}$. This chain is irreducible and aperiodic, however it turns out to be "null recurrent" rather than positive recurrent. The τ_i diverge.

Random walks over infinite spaces are not ergodic. However, we could carry out random walks on finite space such as a circle. In field theory, we will be interested in infinite spaces, but with much more sharply peaked probability distributions. We will fine algorithsm such that C1-C3 hold. Then it can be proved that

T1. The system will approach an equilibrium distribution, independent of initial configuration.

T2. Averages over the Markov chain will approach the desired ensemble averages with statistical error of order $1/\sqrt{N}$.

We want the equilibrium distribution of the process to be proportional to $e^{-S(C)}$ (8.2). A sufficient, but not necessary, condition to guarantee this occurs is known as *detailed balance*. This says that the probabilities for going between configurations C and C' should satisfy

$$e^{-S(C)}\, P(C \to C') \;=\; e^{-S(C')}\, P(C' \to C)\,. \tag{8.7}$$

To understand this, imagine configuration C gives an action $S(C)$ close to its minimum and C' gives a significantly larger action. In this case $e^{-S(C)} \gg e^{-S(C')}$. Detailed balance (8.7) then implies that the transition from $C \to C'$ is highly unlikely compared to the transition $C' \to C$. The algorithm allows the occasional excursion into configurations with large action, but most of the configurations sampled will be near the minimum of S, as expected from the sharply peaked Boltzmann distribution (8.2).

Perhaps the most intuitive process is the *heatbath algorithm*. Making the identification of the probability density (8.2) with the Boltzmann weight $\exp(-H/T)$, we can imagine sweeping through the lattice, visiting each link $U_\mu(x)$ and choosing a new value for that particular variable, ignoring the current value of $U_\mu(x)$. For example, if a specific link is Ising-like, $U_\mu(x) = \pm 1$, (this is $\mathbb{Z}_2$ gauge theory) then we can choose the updated spin value to be $+1$ with probability

$$\rho(+1) \;=\; \frac{e^{-S(+1)}}{e^{-S(+1)} \,+\, e^{-S(-1)}}\,. \tag{8.8}$$

For simplicity here we suppress the dependence of ρ and S on all other link variables, which are held fixed as we update the aforementioned specific link.

We implement this by choosing a uniformly distributed random number $r \in [0,1]$ and accepting the value $+1$ only if $r < \rho(+1)$. Due to this random component of the updating process, these are referred to as *Monte Carlo* algorithms, after the famous casino area in Monaco. In the Ising case, if the proposed new value is rejected, then we know the new value must be -1. For larger groups, one needs to test other possibilities, e.g. $\pm i$ in a $\mathbb{Z}_4$ model. For continuous groups, heatbath proceeds rather slowly on its own, as it can take many tries to find an accepted update, especially as the lattice spacing decreases compared to physical length scales (i.e. as the correlation length diverges in lattice units).

If heatbath is used, it is done so in combination with other updating steps, such as one called overrelaxation. In an overrelaxation step, one changes each link variable ensuring that the action is kept constant. This is easy to implement if we employ symmetry transformations. This change is always accepted, but it does not fully sample the required probability distribution (8.2). Therefore, overrelaxation can only be used in combination with an ergodic algorithm such as heatbath.

Better algorithms proceed by proposing a single change to the configuration, accepting the change with some probability, and, if rejected, simply moving on to the next site/link

with previous site/link unchanged. Let us take $W(U \to U')$ to be the probabilistic weight of a proposal to change a particular variable with value U in configuration C to the value U' resulting in configuration C', and $P_{\rm acc}(C \to C')$ is the probability of accepting this proposal:

$$P(C \to C') \;=\; W(U \to U')\, P_{\rm acc}(C \to C')\,. \tag{8.9}$$

Metropolis, et al. proposed this type of algorithm in 1953, and so straightforward applications of this idea are called the *Metropolis algorithms*, although more complicated algorithms simply include a *Metropolis accept/reject step* at the end. Here we assign certain probabilistic weights for proposing a new value for a degree-of-freedom, say U' given the old value U. The weights should be symmetric, i.e.

$$W(U \to U') = W(U' \to U)\,. \tag{8.10}$$

We will give a few examples soon.

Having decided on the transition weights, we next apply the accept/reject step by comparing the action of the old configuration C with the action of the proposed configuration C', $\Delta S = S(C') - S(C)$. We accept the change with probability

$$P_{\rm acc}(C \to C') = \begin{cases} 1 & \text{if } \Delta S \leqslant 0 \\ e^{-\Delta S} & \text{if } \Delta S > 0 \end{cases}. \tag{8.11}$$

We implement the second case by generating a random number $r \in [0,1]$ with uniform likelihood and accepting the change if $r < \exp(-\Delta S)$.

Application to the Ising or $\mathbb{Z}_2$ model is simple. We may always propose a change of state: that is we take $W(-1 \to 1) = W(1 \to -1) = 1$ and $W(-1 \to -1) = W(1 \to 1) = 0$.

In a $U(1)$ model, we might propose to change $U \to \exp(i\theta)U$ with θ randomly chosen in the interval $(-\varepsilon, \varepsilon)$ with uniform probability. ε can be tuned for efficiency. One needs to balance the desire to make big changes, better sampling the desired distribution, while keeping the acceptance rates from becoming too small. In one of the exercises you are to find $P(C \to C')$ given this information and demonstrate detailed balance is satisfied.

These ideas can be generalized to nonabelian gauge theories. An element $V \in SU(2)$ can be parametrized by 4 real numbers (3 independent)

$$V \;=\; v_0 I \;+\; i\vec{v}\cdot\vec{\sigma} \tag{8.12}$$

with $v_0^2 + |\vec{v}|^2 = 1$. If we choose 4 random numbers $r_\rho \in (-\frac{1}{2}, \frac{1}{2})$ we can propose $U \to U' = VU$ with

$$v_0 \;=\; \text{sign}(r_0)\sqrt{1-\varepsilon^2} \;\text{ and }\; \vec{v} \;=\; \varepsilon\,\frac{\vec{r}}{|\vec{r}|}\,, \tag{8.13}$$

where $0 < \varepsilon < 1$ can be tuned for efficiency. For $SU(N)$, $N \geqslant 3$, one can build generate a new group element out of products of $SU(2)$ rotations, subgroup by subgroup. In this case, care should be taken that the weight of the proposed change V is the same as that for the inverse change $V^{-1} = V^\dagger$, in order to satisfy detailed balance.

As might be imagined, some work has been done in the last 70 years to improve upon the original Metropolis idea. It is beyond the scope of these lectures to elaborate, but anyone wishing to carry out their own calculations should read further [6, 7].

Exercise 7. *Let $\rho_{eq}(C)$ be the desired "equilibrium" distribution and let $\rho_0(C)$ be an initial non-equilibrium distribution. Given an updating algorithm which has transition probability density $P(C \to C')$, let us consider the case where C and C' differ by only one link variable, say $U_\mu(x)$ vs $U'_\mu(x)$. Then we can write everything in terms of U and U'; in particular for this single-variable update, define the update probability $p(U \to U') = P(C \to C')$. If we repeatedly update only that one variable, the distribution evolves from Monte Carlo step N to step $N+1$*

$$\rho_{N+1}(U') \;=\; \int dU\, \rho_N(U)\, p(U \to U') \qquad (*)$$

(with x-dependence suppressed). Explain why $\int dU'\, p(U \to U') = 1$. Show that detailed balance is sufficient to ensure $\rho_{eq}(U)$ is a fixed point of $(*)$. (In our case $\rho_{eq}(U) = e^{-S(U)}$, but this is not relevant here.) This argument can be generalized to updates which involve all variables.

Exercise 8. *Show that the Metropolis algorithm satisfies detailed balance (8.7) for the case of $U(1)$ gauge theory.*

9 Lattice fermions

9.1 Discretization

The discretization of the Dirac action which we would naively write down would be something like

$$S_f[\psi,\bar{\psi},U] \;=\; a^4 \sum_x \bar{\psi}_x[(m \,+\, \gamma\cdot\nabla^{\pm})\psi]_x \tag{9.1}$$

where we have promoted the difference operator to be gauge covariant. Writing things out more explicitly, the naive fermion action is

$$S_f[\psi,\bar{\psi},U] \;=\; a^4 \sum_x \bar{\psi}(x)\Big\{ m\,\psi(x) \;+\; \frac{1}{2a}\sum_{\mu=1}^{4}\gamma_\mu\Big[U_\mu^\dagger(x)\,\psi(x+a\hat{\mu}) \\ - U_\mu(x-a\hat{\mu})\psi(x-a\hat{\mu})\Big]\Big\}\,. \tag{9.2}$$

A convenient notation for the kernel of the action is the fermion matrix M

$$S_f[\psi,\bar{\psi},U] \;=\; \bar{\psi}_i\, K_{ij}[U]\,\psi_j \tag{9.3}$$

where the indices run over spacetime sites, color, and spin degrees-of-freedom. K is sometimes called the "fermion matrix."

Before discussing the naive action in more detail, let's pause to think about how to include the fermion matrix in Monte Carlo calculations. In order to include fermion effects in the importance sampling, we have to formally integrate out the Grassmann-valued fields. Luckily we know how to do this for bilinear actions:

$$\int [d\psi\, d\bar{\psi}\, dU]\, e^{-S_g[U]\,-\,\bar{\psi}K[U]\psi} \;=\; \int [dU]\;\det K[U]\, e^{-S_g[U]} \tag{9.4}$$

The algorithmic problem how to include the fermion determinant in the importance sampling. No matter what, its presence requires the algorithm to be nonlocal: a change in a single link variable requires a computation which scales like the volume of the lattice. Since explicit computation of the determinant would scale like $(N_x N_y N_z N_t)^3$, the most widely used algorithms re-exponentiate the determinant by introducing pseudofermions, fields which are complex-valued.

$$\int [d\zeta\, d\bar{\zeta}\, dU]\, e^{-S_g[U]\,-\,\bar{\zeta}\,K[U]^{-1}\zeta}\,. \tag{9.5}$$

Computing the inverse is still a nonlocal calculation, but at least it scales linearly with the 4-volume. There is a further difficulty, the matrix K has eigenvalues which are proportional to the quark mass m; for physical values of the u and d quarks, the matrix is very nearly singular. This is a consequence of chiral symmetry breaking; Banks and Casher showed that a nonzero chiral condensate requires a nonzero density of zero modes (in the $m = 0, V \to \infty$ limit). Therefore the inversion becomes increasingly difficult for physically small values of m and accounts for the bulk of computational effort.

The free naive propagator is obtained by setting all links $U_\mu(x) = 1$ and transforming to momentum space

$$\Delta(p) = \frac{m - i\sum_\mu \gamma_\mu \bar{p}_\mu}{m^2 + \bar{p}^2} \tag{9.6}$$

where we used the shorthand

$$\bar{p}_\mu := \frac{1}{a} \sin p_\mu a \,. \tag{9.7}$$

In the $a \to 0$ limit, the massless propagator (9.6) has a pole not only at $p = (0, 0, 0, 0)$, but also if any momentum component is π/a.

The Nielsen-Ninomiya theorem says that doubling will occur for any lattice action which is chirally symmetric, i.e. for any operator D (the massless part of $K = \gamma_\mu D_\mu + m$) which satisfies

$$\{\gamma_5, D\} \;=\; 0\,. \tag{9.8}$$

9.2 Doubling in 1+1 dimensions

It is instructive to consider free massless fermions in $1 + 1$ dimensions. Following Ref. [8], take our Minkowski metric to be $\eta = \mathrm{diag}(1, -1)$ and Dirac matrices

$$\gamma^0 = \sigma^1\,, \quad \gamma^2 = i\sigma^2\,, \quad \text{and} \quad \gamma^5 = -\gamma^0\gamma^1 = \sigma^3 \tag{9.9}$$

where these are Pauli's σ matrices. The action of a massless Dirac fermion field is

$$S = \int d^2x \; i\bar{\psi}\partial\!\!\!/\,\psi\,. \tag{9.10}$$

We can decompose the Dirac fermion into two Weyl fermions, eigenvectors of γ^5 as

$$\psi = \begin{pmatrix} \chi_+ \\ \chi_- \end{pmatrix} \quad \text{with action} \quad S = \int d^2x \; i\left[\chi_+^\dagger \partial_- \chi_+ + \chi_-^\dagger \partial_+ \chi_-\right] \tag{9.11}$$

where $\partial_\pm = \partial_t \pm \partial_x$.

The Weyl field χ_+ satisfies the equations of motion $\partial_-\chi_+ = 0$ which tells us χ_+ depends on the single variable $t + x$. An excitation in the χ_+ at the origin would move toward increasingly negative values of x as time evolves, hence these states are referred to as "left-movers." Similarly χ_- excitations are "right-movers."

Notice that the action (9.11) is chirally invariant, that we can transform χ_+ and χ_- independently:

$$\chi_+ \mapsto e^{i\alpha_+}\chi_+ \quad \text{and} \quad \chi_- \mapsto e^{i\alpha_-}\chi_- \tag{9.12}$$

with α_+ and α_- independent.

[FIGURES]

Consider the dispersion relation of these excitations in the continuum. These are just the lines $E = \pm p$. In the picture of Dirac's sea, states with $E < 0$ are filled and those with

$E > 0$ are empty. Right-mover excitations are on the $E = p$ branch and left-movers on the $E = -p$ branch.

When we discretize the theory on a lattice, momentum becomes bounded in a Brillouin zone

$$p \in (-\frac{\pi}{a}, \frac{\pi}{a}] \tag{9.13}$$

with periodicity, $p = -\pi/a$ is identified with π/a. The continuum dispersion must be deformed to respect this periodicity. There are two ways to do this: the left-moving and right-moving branches can be joined up or kept separate.

Naive fermions keep the branches separate. The calculation here is essentially the same as in 4-dimensions. For simplicity let's just consider discretizing the spatial dimension. The dispersion relation becomes

$$E = \pm\frac{1}{a}\sin pa\,. \tag{9.14}$$

The energy vanishes as $p \to \pm\pi/a$, enforcing periodicity while keeping the branches separate. However, the dispersion relation near the point $p = \pi/a$ has the opposite slope. Excitations near this point on the χ_- branch look like low-energy left-movers rather than high-energy right-movers. This is fermion doubling. Any interactions will then allow a coupling between left- and right-moving degrees of freedom.

An alternative possibility is that the branches of the continuum dispersion relation join up. In this case we can avoid any $E = 0$ modes away from $p = 0$. However the joining of the two branches is a coupling of the χ_+ and χ_- degrees-of-freedom. The only way for an action to have this type of dispersion relation is to include a term coupling these fields. Such a term will break chiral invariance, since $\chi+$ and χ_- will not be able to transformed independently of one another.

See Tong's *Gauge Theory* notes [8] for much more about lattice fermions and their connection to chiral anomalies.

9.3 Wilson fermions

The simplest remedy for the doubling problem is to add a second derivative operator to the action

$$S_f[\psi,\bar\psi,U] \;=\; a^4\sum_x \bar\psi_x\left[\left(m \,+\, \gamma\cdot\nabla^{\pm} \,-\, \frac{a}{2}\,\nabla^+\cdot\nabla^-\right)\psi\right]_x . \tag{9.15}$$

Then the free propagator is

$$\Delta(p) \;=\; \frac{m + a^{-1}\sum_\mu(1-\cos(p_\mu a)) + i\sum_\mu \gamma_\mu \bar p_\mu}{[m + a^{-1}\sum_\mu(1-\cos(p_\mu a))]^2 + \bar p^2}\,. \tag{9.16}$$

Effectively this gives the doublers a mass proportional to $1/a$.

This term is higher dimension, so it is irrelevant in the continuum limit. At finite lattice spacing, however, it breaks chiral symmetry explicitly, giving the quark mass an additive renormalization of $O(a^{-1})$. One must fine-tune to the chiral limit at each lattice spacing.

A recent innovation is to add a *twisted mass* term to the Wilson action; the twist is in the SU(2) flavor space occupied by the up and down quarks.

$$S_f[\chi,\bar\chi,U] \;=\; a^4\sum_x \left\{\bar\chi_x\left(mI_2 \,+\, i\mu\gamma_5\tau^3 \,+\, \gamma\cdot\nabla^\pm \,-\, \frac{a}{2}\nabla^+\cdot\nabla^-\right)\chi_x\right\}. \tag{9.17}$$

The parameter μ is the twisted mass. One defines a new mass $M^2 = m^2+\mu^2$ and an angle through $\tan\alpha = \mu/m$. Then the mass term in the action above is $M\exp(i\alpha\gamma_5\tau^3)$. In the noninteracting case, $\alpha = \pi/2$ corresponds to $m = 0$ (and $\mu > 0$); this is referred to as *maximal twist*. Interactions renormalize the value of α at which maximal twist is achieved. Defining new fields $\psi = \exp(i\alpha\gamma_5\tau^3)\chi$ and $\bar\psi = \bar\chi\exp(i\alpha\gamma_5\tau^3)$ yields an action

$$\begin{aligned} S_f[\psi,\bar\psi,U] = a^4\sum_x \Bigg\{ &\bar\psi\left(\frac{4}{a}\,e^{-i\alpha\gamma_5\tau^3} \,+\, MI_2\right)\psi_x \\ &-\frac{1}{2a}\sum_\mu \bar\psi_x\Big[\left(e^{-i\alpha\gamma_5\tau^3}-\gamma_\mu\right)U_\mu^\dagger(x)\psi_{x+a\hat\mu} \\ &+\left(e^{-i\alpha\gamma_5\tau^3}+\gamma_\mu\right)U_\mu(x-a\hat\mu)\psi_{x-a\hat\mu}\Big]\Bigg\}. \end{aligned} \tag{9.18}$$

So we see that the twisting actually modifies the Wilson term. The benefit of this procedure is that $\mu > 0$ sets a lower bound on eigenvalues of K while being chirally symmetric. Thus, by tuning $m = m_{\rm crit}$ (equivalently, by tuning α to maximal twist) one can restore chiral symmetry, in the sense of removing chirality-violating mixings. The cost is that flavor symmetry is broken in a similar way that the doubling symmetry for staggered fermions is broken by gauge interactions. Nevertheless, the twisted fermion approach is a promising and active avenue for large scale LQCD calculations.

9.4 Staggered fermions

[In this section, I write the imaginary time component as the 0th component for consistency with the literature.]

The naive discretization of the Dirac action (9.2) has an extra set of symmetries, referred to as the *doubling symmetries*. For example the action is invariant under

$$\psi(x)\to e^{i\pi x_1/a}(i\gamma_5\gamma_1)\psi(x)\,,\qquad \bar\psi(x)\to\bar\psi(x)e^{i\pi x_1/a}(i\gamma_5\gamma_1)^\dagger\,. \tag{9.19}$$

There are 16 such transformations (1 trivial) corresponding to the 16 corners of the corners of the Brillouin zone, the example is for $(0,\pi/a,0,0)$.

The naive action can be diagonalized in spin by multiplying the fields by a product of γ-matrices, depending on the parity of the site in each dimension. That is, take

$$\Omega(x) \;\equiv\; (\gamma_0)^{x_0/a}(\gamma_1)^{x_1/a}(\gamma_2)^{x_2/a}(\gamma_3)^{x_3/a} \tag{9.20}$$

and defining new fields

$$\psi(x) \;=\; \Omega(x)\,\xi(x)\,,\qquad \bar\psi(x) \;=\; \bar\xi(x)\,\Omega^\dagger(x)\,. \tag{9.21}$$

Since the resulting action is diagonal in spin, it is sufficient to study a single spin component. We introduce a staggered fermion field $\chi(x)$ whose action is the same as for $\xi(x)$

$$S \;=\; a^4 \sum_x \bar{\chi}(x) \left[\sum_{\mu=0}^{3} \eta_\mu(x) \nabla_\mu \;+\; m \right] \chi(x) \tag{9.22}$$

where

$$\eta_\mu(x) = (-1)^{(x_0 + \ldots + x_{\mu-1})/a} \,. \tag{9.23}$$

The naive fermion propagator (even in the fully interacting theory) can be reconstructed from the staggered propagator using the identity

$$G_\psi(x;y) \;=\; \Omega(x)\, \Omega^\dagger(y) \;\times\; G_\chi(x;y) \,. \tag{9.24}$$

Staggered fermions are also referred to as Kogut-Susskind fermions.

In the importance sampling, a hypothesis is needed to include fermions which have a degeneracy other than a multiple of 4. This is referred to as the *fourth root hypothesis*: one uses the following as the weight in the integrand of path integrals

$$(\det K[U])^{N_f/4} e^{-S_g[U]} \,. \tag{9.25}$$

It is an open theoretical problem to prove this hypothesis will give the correct theory in the continuum limit. There are arguments suggesting this is the case (and a few arguments against, with corresponding rebuttals). The hypothesis is correct perturbatively. Also staggered chiral perturbation theory can be formulated which has its own equivalent hypothesis. The agreement between Monte Carlo data and staggered chiral perturbation theory fits is taken as further evidence that this procedure is valid.

9.5 Correlation functions

Consider a local fermion bilinear operator $O = \bar{\psi}^{f_1} \Gamma \psi^{f_2}$, where $\Gamma \in \{I, \gamma_\mu, \gamma_\mu\gamma_\nu, \gamma_5\gamma_\mu, \gamma_5\}$ and we have put flavor indices on the fields. If we were interested in computing the kaon mass, for example, we would take $O = \bar{s}\gamma_5 u$, where s and u are shorthand for ψ^s and ψ^u. The correlation function we compute

$$\begin{aligned} \langle O^\dagger(x) O(0) \rangle \;&=\; \langle \bar{u}(x)\gamma_5 s(x)\, \bar{s}(0)\gamma_5 u(0) \rangle \\ &=\; \mathrm{Tr}\left(\gamma_5 G^s(x;0) \gamma_5 G^u(0;x) \right) \,. \end{aligned} \tag{9.26}$$

The factors $G^q(x;y) = q(x)\bar{q}(y)$ are obtained by inverting the fermion matrix, on a given set of link values $G^q(x;y) = (K^{-1}[U])_{xy}$. This is a full, nonperturbative (gauge non-invariant) fermion propagator.

9.6 Chirally symmetric fermions

These lectures will not do justice to the developments of the 1990's which showed how to have light fermions with the full $\mathrm{SU}(N_f) \times \mathrm{SU}(N_f)$ chiral symmetry.

We do not have time to discuss the elegant formulation of domain wall and overlap fermions. This topic is at least a lecture in its own right. Tong's lectures [8] and the texts [6, 9] are good entries with references to the literature.

These formulations do not violate the Nielsen-Ninomiya theorem since the Dirac operator (in the case of domain wall fermions, the effective 4-dimensional operator governing the light modes) does not exactly anticommute with γ_5, but instead satisfies what is called the Ginsparg-Wilson relation

$$\{\gamma_5, D\} \;=\; aD\gamma_5 D\,. \tag{9.27}$$

Assuming D^{-1} exists, the relation is equivalent to

$$\{\gamma_5, D^{-1}(y-x)\} \;=\; a\gamma_5\,\delta(y-x)\,. \tag{9.28}$$

That is, the propagator anticommutes with γ_5 except for a contact term.

References

[1] M. Srednicki, *Quantum Field Theory* (Cambridge University Press, 2007).

[2] I. Montvay and G. Münster, *Quantum Fields on a Lattice* (Cambridge University Press, 1994).

[3] M. Creutz, *Quarks, Gluons, and Lattices* (Cambridge University Press, 1983).

[4] J. Greensite, Prog. Part. Nucl. Phys. **51**, 1 (2003), arXiv:hep-lat/0301023.

[5] H. J. Rothe, *Lattice Gauge Theories: An Introduction* (World Scientific, 1992).

[6] T. DeGrand and C. DeTar, *Lattice Methods for Quantum Chromodynamics* (World Scientific, 2006).

[7] F. Knechtli, M. Günther, and M. Peardon, *Lattice Quantum Chromodynamics: Practical Essentials* (Springer, 2017).

[8] D. Tong, Gauge Theory lecture notes available at https://www.damtp.cam.ac.uk/user/tong/gaugetheory.html.

[9] C. Gattringer and C. Lang, *Quantum Chromodynamics on the Lattice* (Springer, 2010).