# Notes: A very brief introduction to topology in (L)QCD

Jesse R. Stryker

January 30, 2018

This is a transcript of notes I made for a talk on 2017-12-08 for the UW Particle Theory Journal Club. I scanned a number of papers in the course of preparing this talk but ultimately found that Gattringer and Lang's *Quantum Chromodynamics on the Lattice* was the best primary source material for this type of talk. I followed Sidney Coleman's *Aspects of Symmetry* for the introduction of instantons.

## 1 Origin of topology

In this section of the talk I seek to explain why topology is an important consideration when studying QCD on the lattice.

### 1.1 Some general features of the Monte Carlo algorithm

To set the context, we assume that one is tasked with the stochastic evaluation of the Euclidean path integral for some observable:

$$\langle\hat{\mathcal{O}}\rangle = \frac{1}{Z}\int (dU)\det D[U]e^{-S_E[U]}\mathcal{O}[U] \tag{1}$$

Not only is the background spacetime Euclidean, but it is also assumed to be compactified in all directions (i.e. periodic boundary conditions or PBC). The compactification in the time-like direction means we are studying properties of a 3-dimensional system in thermal equilibrium at some inverse temperature, $\beta$. Note further that path integration over the fermionic fields has been carried out, resulting in a factor in the integrand of the determinant of the Dirac operator; this determinant is a function of the gauge fields, since the Dirac operator is.

In the elementary scenario of a Markov chain Monte Carlo (MCMC) evaluation, one makes proposals by updating the gauge fields on the links via

$$U_\mu(n)' = XU_\mu(n) \qquad X\in G, \quad X\simeq \mathbb{1} \tag{2}$$

The stipulation that $X\simeq\mathbb{1}$ is a result of the observation that, near a minimum of the action, small changes of the action are necessary to get an acceptance rate that allows sampling the neighborhood. A 'point' along the Markov chain is a particular assignment of SU(3) matrices to every link on the lattice :

$$U = \{U_\mu(n)\}_{n\in\Lambda}^{\mu=1,2,3.4} \tag{3}$$

This is what is meant by a gauge field configuration.

Returning to the updates, the random matrices $X$ serve to transport the $U(n)_\mu$ through the group manifold. Algorithms do exist for updating many sites and/or making larger jumps via $X$ without destroying the acceptance rate, but I won't be getting in to these. As noted earlier, we expect the Markov chain to spend time around minima of the action (which are maxima of $e^{-S}$, wherein it should be understood that the logarithm of the Dirac determinant must be absorbed into the action). Our updates are necessary to both explore the configuration space around the minima *as well as transition among minima*. I often refer to volumes surrounding minima as pockets of high probability weight.

Consider, e.g., a probability density for a 1-D integral which looks like two widely separated Gaussians, symmetric about the origin. A proper sampling of the whole domain of integration requires large jumps, and proper sampling of the "fluctuations" requires small jumps. Without getting into the abundance of information existing on update algorithms, I simply want to point out that "trapping" is a problem likely to arise in generating proposals if one does not take steps to avoid it. (Of course, trapping should not be a problem in the limit of infinite statistics, if the Markov update procedure is ergodic.)

The whole point of this talk is that such trapping comes up in LQCD. When this happens, autocorrelation times (time here referring to Monte Carlo time) can be so long as to render a simulation impractical without some algorithm designed to overcome the trapping. In the literature this is referred to as "critical slowing down." S. Schaefer (2009) observes this with several MCMC algorithms. Moreover, this is can be found in pure gauge theory (where the Dirac contribution is absent) as well as with dynamical fermions. The significance of this last statement should be more clear after we get to know the topological properties of the path integral.

## 1.2 Topology of continuum QCD

At the end of the day what we want to learn about is continuum physics, so it is fitting to start first with topology of a continuum theory. We will see there is a nice physical picture of what is going on when a simulation suffers from trapping.

A simplified, specific example for building intuition is illustrated by evaluating the Euclidean propagator of a nonrelativistic particle in a double-well potential (adapted from *Aspects of Symmetry*). This system lacks the complications of a full-blown field theory, while retaining the conceptual kernel we're after. We'll define

$$K_{fi} \equiv \langle x_f |\, e^{-T\hat{H}/\hbar}\, | x_i \rangle = \mathcal{N} \int [dx]_{\mathrm{BC:}\substack{x(T/2)=x_f \\ x(-T/2)=x_i}} \exp\left\{ -\frac{1}{\hbar} \int_{-T/2}^{T/2} dt \left( \frac{m}{2}\dot{x}^2 + V(x) \right) \right\} \tag{4}$$

and consider the semi-classical limit where $\hbar \to 0$.[1] In this limit, $Z$ becomes dominated by paths of minimal action sastifying the BCs.

Extremizing the action, as always, means solving the Euler-LaGrange equations. Only here, we are led to a Newton's second law

$$m\ddot{x} = +V'(x) \tag{5}$$

[1] With an eye toward LQCD, this corresponds to the continuum limit, since the Yang-Mills action comes with a factor $g_s^{-2}$ and $g \to 0$ due to asymptotic freedom.

where the sign on the potential has been flipped. Therefore, what we seek are solutions to the "familiar" classical equations of motion for a particle moving in the *inverted potential*, $-V(x)$. By the way, it is convenient to define $V(\pm a) \equiv 0$, where $\pm a$ are the locations of the two minima. We can easily characterize all the possible classical solutions (ignoring the BC for now):

1. The particle comes in from $\pm\infty$, passes over both hills, and continues off to $\mp\infty$. ($E > 0$)
2. The particle sits on a hilltop forever. ($E = 0$)
3. The particle comes in from $\pm\infty$ and asymptotically approaches a hilltop for the rest of its life. ($E = 0$).
4. The particle comes in from $\pm\infty$ but lacks sufficient energy to overcome the barrier, getting deflected back from where it came. ($E < 0$)
5. The particle sloshes back and forth forever between the hills. ($E < 0$)

A particularly interesting propagator to consider is $K(+a, -a) = K(-a, +a)$. Both locations $x = \pm a$ are global minima. This corresponds, by way of the WKB approximation, to a *tunneling* from one vacuum to another. Then, for the saddle point expansion, we are looking at motions from one hilltop over to the other. We consider $T$ to be large. The only exact solution obeying the boundary conditions will be one with a very small positive energy, corresponding to extremely slow velocity at the hilltops.

However, there are many paths which only violate the equations of motion by microscopic deviations, and these are plentiful in the path integral. To give meaning to the notion of "short" and "long" times, let us define

$$\omega^2 = V''(a) \tag{6}$$

and have $\omega^{-1}$ be the reference unit. A particle having exactly $E = 0$ approaching a hilltop will do so exponentially:

$$|a - x| \propto e^{-\omega t} \tag{7}$$

Approximate solutions to the equations of motion, then, are paths in which the particle is almost always sitting on one hilltop or the other, but every so often, it gets an infinitesimal fluctuation inward, causing it to roll to the other side. Of course, it has to be on the correct hilltops at $t = \pm T/2$, to obey the BC.

Because $T$ is taken to be very large, such transitions from one side to the other occur in an "instant" on the scale of $T$, and for this reason we call these transition paths *instantons* or *anti-instantons* (depending on the direction). On the scale of $T$, the instantons are like step functions. The point here, from the view of evaluating the path integral, is that there is but a single path in the sum over paths which exactly satisifies the BC, but there is a whole *continuum* of instanton paths which nearly solve the equations of motion—that continuum being a consequence of integrating over the temporal location of the instanton(s).

The notion of topology comes about because we can classify such approximate saddle-points in the path integral by their numbers of instantons and anti-instantons: $(n_+, n_-)$ where, of course, $|n_+ - n_-| = 1$ for the propagator we're studying. It should be easy to convince yourself that we cannot continuously deform, for example, a (5,4) path into a (4,3) path without severe intermediate deviations from the equations of motion; thus the "topological" nature. Even though only the smallest $(n_+, n_-)$ sectors dominate the path integral, all the sectors can be thought of as describing local probability pockets in 'path space.'

Having built some intuition from quantum mechanics, let's get back to QCD. Putting the pure gauge theory in a large box, we can follow a similar story in a semi-classical analysis. In this case, the relevant topological quantity is known as the *index* or *winding number*,

$$\nu = \frac{1}{16\pi^2} \int d^4x \, \mathrm{tr}(F\tilde{F}) \tag{8}$$

$$\tilde{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}F^{\alpha\beta} \tag{9}$$

the integrand is sometimes called the winding number density. Allowing for a CP-violating topological term in the lagrangian, $\mathscr{L}_\theta$, the vacuum-vacuum transition amplitude is

$$\langle\theta| \, e^{-T\hat{H}} \, |\theta\rangle = \sum_\nu e^{i\nu\theta} Z_\nu \tag{10}$$

$$Z_\nu = \int [dA]_\nu e^{-S} \tag{11}$$

The path integral is broken up into a sum over sectors labeled by the index, $\nu$. The gauge fields summed over for $Z_\nu$ are required to have field strength-tensors resulting in the index $\nu$. In this case, the solutions having $\nu = \pm 1$ are the instantons and anti-instantons.

From what I understand, most of the non-zero winding number paths contributing to the path integral can be viewed as the result of patching together many isolated instantons and anti-instantons. Just like in the quantum mechanics example, we can't expect that these are exact solutions of the equation of motion, but they are close enough that they can have significant weight in the path integral. There are also exact solutions with non-zero winding number that can't be decomposed like this. As the distance between individual intstantons is taken to infinity, such multi-instanton solutions must become solutions to the equations of motion. And, importantly, pair creation of instantons doesn't change the overall index.

## 1.3 Critical slowing down

Knowing that the continuum partition function has disconnected topological sectors, we expect this feature to emerge in a lattice simulation as $a \to 0$. One way we could circumvent this would be to employ *open* boundary conditions, rather than the usual PBC. The reason this is not desirable is because we lose discrete translational invariance if we remove periodicity (for a finite volume). So, we retain PBC.

When the gauge fields are updated, we can expect that a very large jump through configuration space is required to change the global topological charge. **Critical slowing down** refers to severe autocorrelation times in LQCD calculations of observables that can be attributed to such trapping in topological charge sectors.

## 2 Notions of topology on the lattice

For dynamical improved Wilson fermions, one can define (using the Clover field-strength tensor)

$$Q = \frac{1}{16\pi^2} a^4 \mathrm{tr}[F\tilde{F}] \tag{12}$$

This lattice quantity does not in general give an integer value, but it can be tracked to examine trapping in the continuum limit.

- Note that the notion of instanton is also only approximate. The "bath" of instantons is a nice intuitive picture, but the saddle point configurations are in reality discrete jumps of the configuration $U$ at different time slices.
- Fluctuations in the continuum smaller than the lattice spacing are also invisible to the lattice simulation.
- Moreover, QCD has fermions built in, which we have left out entirely so far in these considerations.

### 2.1 Chiral symmetry and topology; index theorem

- We know that (approximate) chiral symmetry of continuum QCD is an essential feature for understanding QCD phenomenology. We've already discussed this quarter various aspects of implementing chiral symmetry on the lattice. One way of stating that chiral symmetry is present at the lagrangian level is that $\{\gamma_5, D\} = 0$.
- Let us therefore assume that our discretization of the Dirac operator is not only $\gamma_5$-hermitian, i.e.

  $$\boxed{D^\dagger = \gamma_5 D \gamma_5,} \tag{13}$$

  but that it satisfies the **Ginsparg-Wilson equation**:

  $$\boxed{\{\gamma_5, D\} = aD\gamma_5 D.} \tag{14}$$

  This breaks chiral symmetry explicitly, but restores it in the continuum limit.

The aim of the following discussion will be to obtain an understanding of the "index theorem" for LQCD, **which is going to connect the topology of the gauge fields to the counting of chiral modes of the (Ginsparg-Wilson) Dirac operator**.

Let us begin by exploring the consequences of our choice of Dirac operator. The first two points concern the eigenvalues of $D$:

$$\boxed{\text{i. } \textbf{The eigenvalues of } D \textbf{ are real, or else come in complex conjugate pairs}} \tag{15}$$

I mention this result first because it has been discussed already in a previous talk. We will return to this for proof of the index theorem later on.

We will also want the following result concerning matrix elements of $\gamma_5$:

$$\boxed{\text{ii. } \textbf{If } Dv_\lambda = \lambda v_\lambda \textbf{ , then } \lambda \textbf{ is real or } (v_\lambda, \gamma_5 v_\lambda) = 0} \tag{16}$$

The proof is very simple:

$$\begin{aligned}\lambda(v_\lambda, \gamma_5 v_\lambda) &= (v_\lambda, \gamma_5 D v_\lambda) \\ &= (v_\lambda, D^\dagger \gamma_5 v_\lambda) \qquad \text{(using } \gamma_5\text{-hermiticity)} \\ &= (D v_\lambda, \gamma_5 v_\lambda) \\ &= \lambda^*(v_\lambda, \gamma_5 v_\lambda) \qquad \square\end{aligned} \tag{17}$$

Next, we'll add in the Ginsparg-Wilson equation, with the result

$$\boxed{\text{iii. } \textbf{The eigenvalues of } D \textbf{ all live on a circle in } \mathbb{C} \textbf{ touching the origin.}} \tag{18}$$

The proof here is also easy, but requires some cleverness.

$$\begin{aligned}&\{\gamma_5, D\} = aD\gamma_5 D \\ \Rightarrow\, &\gamma_5 D + D\gamma_5 = a(\gamma_5 D^\dagger)D \\ \Rightarrow\, &D + \gamma_5 D \gamma_5 = aD^\dagger D \\ \Rightarrow\, &D + D^\dagger = aD^\dagger D\end{aligned} \tag{19}$$

And a similar argument shows that

$$D^\dagger + D = aDD^\dagger \tag{20}$$

If we take matrix elements of the former equation with respect to an eigenvector $v_\lambda$, we find that

$$\begin{aligned}&\lambda + \lambda^* = a\lambda^*\lambda \\ \Rightarrow &\left(\text{Re }\lambda - \frac{1}{a}\right)^2 + (\text{Im }\lambda)^2 = \frac{1}{a^2}.\end{aligned} \tag{21}$$

I've omitted a couple lines of algebra. This shows all the eigenvalues live on a circle in the complex plane of radius $a^{-1}$ and centered at $a^{-1}$.

So far in our journey to the index theorem, we've extracted a few results just from the properties of the Dirac operator. All we need is one more important result:

$$\boxed{\text{iv. } \textbf{The zero modes of } D \textbf{ can be chosen to have definite chirality.}} \tag{22}$$

Clearly vanishing eigenvalues are consistent with the above analysis. Let us suppose then that $v_0$ is such a zero mode. Then

$$\begin{aligned}&Dv_0 = 0 \\ \Rightarrow\, &\gamma_5 D v_0 = 0\end{aligned} \tag{23}$$

But from the Ginsparg-Wilson condition,

$$\begin{aligned}&D\gamma_5 + \gamma_5 D = aD\gamma_5 D \\ \Rightarrow\, &\gamma_5 D = aD\gamma_5 D - D\gamma_5 \\ \Rightarrow\, &\gamma_5 D v_0 = aD\gamma_5 D v_0 - D\gamma_5 v_0 \\ \Rightarrow\, &0 = D\gamma_5 v_0\end{aligned} \tag{24}$$

The point of the last two results is that

$$[D, \gamma_5] v_0 = 0 \tag{25}$$

i.e. *$D$ and $\gamma_5$ commute when restricted to the zero-mode subspace.* Hence, we reach the conclusion that a basis for the zero-mode subspace can be chosen such that each eigenvector has definite chirality:

$$\gamma_5 v_0 = \pm v_0 \qquad \square \tag{26}$$

We can finally state the theorem. Define a *topological charge density*, $q(n)$, by

$$q(n) = \frac{1}{2a^3} \mathrm{tr}_{\mathrm{CD}} \left[\gamma_5 D(n|n)\right] \tag{27}$$

so that the (lattice) **topological charge**, $Q_{\mathrm{top}}$, is given by

$$Q_{\mathrm{top}} = a^4 \sum_{n \in \Lambda} q(n) \tag{28}$$

Furthermore, let $n_{-,+}$ denote the number of left- and right-handed zero modes of the Dirac operator. The lattice equivalent of the Atiyah-Singer theorem is the statement that

$$\boxed{Q_{\mathrm{top}} = n_- - n_+} \tag{29}$$

Several comments are now in order.

1. First, the proof. Note that, as a consequence of our earlier results (19) and (20), we have that

$$[D, D^\dagger] = 0, \tag{30}$$

implying $D$ is a *normal* operator; therefore its eigenstates form a complete orthonormal basis. With this in hand, let us evaluate the topological charge:

$$\begin{aligned} Q_{\mathrm{top}} &= a^4 \sum_{n \in \Lambda} \frac{1}{2a^3} \mathrm{tr}_{\mathrm{CD}} \left[\gamma_5 D(n|n)\right] \\ &= \frac{1}{2} \mathrm{tr} \left[\gamma_5 a D\right] \\ &= \frac{1}{2} \mathrm{tr} \left[\gamma_5 (aD - 2)\right] \qquad \text{(using } \gamma_5 \text{ tracelessness)} \\ &= \frac{1}{2} \sum_\lambda (v_\lambda, \gamma_5 (aD - 2) v_\lambda) \qquad \text{(using that } v_\lambda \text{ are a basis )} \\ &= \frac{1}{2} \sum_\lambda (v_\lambda, \gamma_5 (a\lambda - 2) v_\lambda) \end{aligned} \tag{31}$$

Recall our result (i.) that for any given $\lambda$, either $\lambda \in \mathbb{R}$ or else it has a vanishing $\gamma_5$ matrix element. Therefore the sum reduces to a sum over only the real eigenvalues. But these are only 0 or $2/a$. And those with $\lambda = 2/a$ end up canceling due to the factor $(a\lambda - 2)$. So all we are left with is

$$\begin{aligned} Q_{\mathrm{top}} &= \frac{1}{2} \sum_{\{\lambda :\, \lambda = 0\}} (v_\lambda, \gamma_5 (0 - 2) v_\lambda) \\ &= (-1)(n_+ - n_-) \\ &= n_- - n_+ \qquad \square \end{aligned} \tag{32}$$

2. We have called $q(n)$ a topological charge density without any justification for its connection to topology. The continuum definition is

$$q_{\text{cont}}(x) = \frac{1}{32\pi^2}\epsilon_{\mu\nu\rho\sigma}\text{tr}_{\text{C}}\left[F_{\mu\nu}(x)F_{\rho\sigma}(x)\right] \tag{33}$$

I find it necessary to defer to Gattringer and Lang §7.3.2 and the references therein for justification that $q(n) = q_{\text{cont}}(x) + \mathcal{O}(a^2)$ as $a \to 0$ (given certain smoothness conditions). To delve into this would go beyond the scope of my talk.

3. The novelty of the lattice index theorem is that the definition of $Q_{\text{top}}$ is a complicated function of the gauge field configuration, yet it takes integer values. And that number is directly related to a counting of the zero modes of the Dirac operator. In the continuum limit, $Q$ makes no reference to the Dirac operator, so the connection I want to emphasize is this intimate relation between topology of the gauge fields and left- and right-handed zero modes of $D$.

# 3 Lessons

I close my talk with some lessons to be taken away from all this discussion of topology in LQCD.

## 3.1 A couple applications: $\eta'$, axion masses

The *topological susceptibility* is defined in terms of $Q_{\text{top}}$ via

$$\chi_{\text{top}} \equiv \frac{1}{V}\left\langle Q_{\text{top}}^2\right\rangle \tag{34}$$

The *Witten-Veneziano formula* is a prediction for the mass of the $\eta'$ meson which depends on $m_K$, $m_\eta$, $N_f$, and $\chi_{\text{top}}$. The evaluation of the topological susceptibility is especially sensitive to trapping for the obvious reason it is a moment of $Q_{\text{top}}$. For the stochastic evaluation of such an observable, algorithms are needed to efficiently move between sectors. Also, the axion mass is related to the topological susceptibility by

$$m_a(T)^2 = \frac{\chi_{\text{top}}(T)}{f_a^2} \tag{35}$$

This relationship shouldn't come as a surprise since the axion field is coupled to the topological term of $\mathcal{L}_{\text{QCD}}$. Knowing the axion mass is a necessary ingredient for predicting present-day abundances of axion dark matter.

## 3.2 Recapitulation

- We've seen that topology is intimately tied to how traditional Monte Carlo algorithms sample the QCD path integral.
- For QCD($\theta$), the field solutions are interpreted as tunnelings between $\theta$ vacua. They also dominate the semiclassical limit vacuum fluctuations.

- We've seen that chiral symmetry is connected to a good notion of toplogy on the lattice. More precisely, our choice of a Ginsparg-Wilson Dirac operator **explicitly breaks chiral symmetry**, but the anticommutator of $D$ and $\gamma_5$ does vanish in the continuum limit:
$$\{\gamma_5, D\} = aD\gamma_5 D \tag{36}$$
- Furthermore, we saw a deep connection between the global properties of gauge field configurations and the zero modes of the Dirac operator. The index theorem equates $n_- - n_+$ to the topological charge.
- Topological observables are necessarily impacted by trapping and other observables can be too. A proper sampling of the $\theta = 0$ vacuum requires the ability to (in principle) sample any of the topological sectors of the path integral.