Interplay of the sign problem and the infinite volume limit: Gauge theories with a theta term

Yiming Cai[*] and Thomas Cohen[†]

Maryland Center for Fundamental Physics and the Department of Physics, University of Maryland, College Park, Maryland 20742, USA

Ari Goldbloom-Helzner

Montgomery Blair High School, Silver Spring, Maryland 20901, USA

Brian McPeak[‡]

Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA

QCD and related gauge theories have a sign problem when a θ term is included; this complicates the extraction of physical information from Euclidean-space calculations as one would do in lattice studies. The sign problem arises in this system because the partition function for configurations with fixed topological charge Q, $\mathcal{Z}_Q$, are summed weighted by $\exp(iQ\theta)$ to obtain the partition function for fixed θ, $\mathcal{Z}(\theta)$. The sign problem gets exponentially worse numerically as the space-time volume is increased. Here it is shown that, apart from the practical numerical issues associated with large volumes, there are some interesting issues of principle. A key quantity is the energy density as a function of θ, $\varepsilon(\theta) = -\log\left(\mathcal{Z}(\theta)\right)/V$. This is expected to be well defined in the large four-volume limit. Similarly, one expects the energy density for a fixed topological density $\tilde{\varepsilon}(Q/V) = -\log(\mathcal{Z}_Q)/V$ to be well defined in the limit of large four volumes. Intuitively, one might expect that if one had the infinite volume expression for $\tilde{\varepsilon}(Q/V)$ to arbitrary accuracy, then one could reconstruct $\varepsilon(\theta)$ by directly summing over the topological sectors of the partition function. We show here that there are circumstances where this is not the case. In particular, this occurs in regions where the curvature of $\varepsilon(\theta)$ is negative.

I. INTRODUCTION

A. θ dependence

Due its nonpertubative structure, the vacuum of quantum chromodynamics (QCD) and other non-Abelian gauge theories is complicated. Accordingly, it is important to understand this vacuum structure by studying how the vacuum responds when conditions are altered. This paper focuses on the effects of changing the so-called θ term. The QCD Lagrangian density has the form

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}G_{\mu\nu}G^{\mu\nu} - \frac{g^2}{32\pi^2}\theta\epsilon^{\alpha\beta\mu\nu}G_{\alpha\beta}G^{\mu\nu},$$
(1)

where $G_{\mu\nu}$ is the field strength tensor; the last term is often omitted. It is the so-called θ term; the θ parameter is sometimes referred to as the vacuum angle. In Euclidean space, it is associated with a winding number Q, called the topological charge, which is given by

[*]yiming@umd.edu
[†]cohen@physics.umd.edu
[‡]bmcpeak@umich.edu

$$Q = \int_V \frac{g^2}{32\pi^2}\epsilon^{\alpha\beta\mu\nu}G_{\alpha\beta}G_{\mu\nu}.$$
(2)

The integer Q equals the difference in the number of right-handed and left-handed zero modes of the Dirac operator according to the Atiyah-Singer index theorem [1]. The θ term violates CP, so the parameter θ measures the amount of CP violation in QCD and QCD-like theories. Since Q is quantized for any configuration, the θ dependence of any physical observable is periodic in 2π; thus, it is useful to restrict our attention to θ between $-\pi$ and π. The case of $\theta = \pi$ is particularly interesting since formally it is CP conserving—under CP transformations $\theta \to -\theta$ but π and $-\pi$ are 2π apart and by periodicity are equivalent.

The θ term is of both theoretical and experimental interest for many reasons, not the least of which because it breaks both P and CP. On the theoretical side, the study of θ dependance is important as it gives an important indication of how the theory responds to a P and CP violating probe. It is interesting to know, for example, whether CP is spontaneously broken in a particular theory at $\theta = \pi$. There are limiting cases where there are good reasons to believe that Dashen's phenomenon, the spontaneous breaking of CP symmetry at $\theta = \pi$ [2], can occur. This is believed to happen for example in pure Yang-Mills

in the large N_c limit [3,4], or in QCD with $N_f = 2$ degenerate light flavors with a mass small enough so that the leading term in chiral perturbation theory dominates [5].

A critical issue for standard model physics is that, while *CP* violation has been observed in the electroweak sector, no *CP* violating effects have been observed in strong interactions. Precise measurements of the electric dipole moment of the neutron have put the upper bound of the theta term at about 10^{-9} away from the *CP* conserving point [6–8]. The problem addressing why θ is so tiny so that *CP* violation is not observed is known as the strong *CP* problem [9]. Attempts to solve the strong *CP* problem by invoking physics beyond the standard model remains a central problem in contemporary physics; however, it is beyond the scope of this paper.

However, from the perspective of QCD itself, there is a related issue. While *CP* violation is known to be small in QCD, one still cannot rigorously rule out the possibility that $\theta \approx \pi$ rather than $\theta \approx 0$. The upper bound of $|\theta| < 10^{-9}$ is fixed since the value of θ is proportional to *CP* violation, which, in turn, fixes the neutron electric dipole moment experiment. The electric dipole moment is experimentally bounded. However, we know $\theta = \pi$ is also formally *CP* invariant, so $|\theta - \pi| < 10^{-9}$ may also not be in conflict with the experimental results. This seems quite unlikely since lattice calculations done with $\theta = 0$ appear to describe the world quite well. However, in the absence of lattice studies at $\theta = \pi$, as a logical matter one cannot rule out the possibility that the $\theta = \pi$ results for most observables are close enough to the $\theta = 0$ results that $\theta = \pi$ is not excluded.

One could easily rule out $\theta = \pi$ if *CP* is spontaneously broken at $\theta = \pi$ [2]. Thus, it would be very useful to know whether QCD (and other gauge theories) spontaneously breaks *CP* at $\theta = \pi$. If *CP* is spontaneously broken, then the energy density as a function of θ, $\varepsilon(\theta)$ will have a discontinuity in its slope at $\theta = \pi$, or, to be more precise, it will develop such a discontinuity in the limit that the infinite volume limit is taken. As noted above, there are regimes where spontaneous *CP* violation is expected to occur at $\theta = \pi$, such as a regime of infinite N_c or a sufficiently small quark mass with two or more flavors. We note, however, that while both of these cases act to suggest that QCD with three colors and physical quark masses spontaneously breaks *CP* at $\theta = \pi$, they are by no means definitive.

For example, Witten points out that in the large N_c limit, $\varepsilon(\theta)$ is parabolic as higher terms in the curvature are suppressed by factors of $1/N_c$, and the periodicity condition forces the function to be defined piecewise as $\min \sum_k (\theta - 2\pi k)^2$ [3,4]. This leads to a discontinuity at $\theta = \pi$. However, one could imagine the following scenario: at very large but finite volumes, the curve is very nearly a perfect parabola, except in a region of a size which goes to zero as N_c goes to infinity, where the curve rapidly turns

over. In such a scenario, the infinite N_c theory has a discontinuity and spontaneously breaks *CP* while for *any* finite N_c *CP* is unbroken. Similarly, near the chiral limit—where the first nontrivial term in the chiral expansion dominates—it has been shown that the energy density, periodic in 2π, is proportional to $(1 - \cos\frac{\theta}{N_f})$ in $\theta \in [-\pi, \pi]$, where N_f is identical to the number of degenerate light quark flavors [5]. This automatically yields Dashen's phenomenon, a discontinuity at $\theta = \pi$, when $N_f \geq 2$. This may seem to be compelling since, in the real world, m_q is small. However, one might worry that although small, m_q may not be negligibly small. This worry stems in part from the fact that the behaviors at large N_c and small m_q are qualitatively different, implying that the large N_c limit and the small m_q limits do not commute. This is hardly surprising, as there are many places in QCD where the large N_c and chiral limits do not commute [10]. The key point here is the fact that an N_c of 3 might be sufficiently large to push the system out of the small m_q regime as far as the behavior at $\theta = \pi$ is concerned.

To illustrate the issue, consider Fig. 1. In this figure the form of $\varepsilon(\theta)$ is given for the infinite N_c limit and for the leading nontrivial term in the chiral expansion. To simplify the comparison (and all comparisons in this paper) we give $\varepsilon(\theta)$ divided by the topological susceptibility, $\chi_0 = \frac{\partial^2 \varepsilon(\theta)}{\partial \theta^2}|_{\theta=0}$; we also set $\varepsilon(0)$ to zero. It is clear that both curves have discontinuities in the slope at $\theta = \pi$ and thus both break *CP* spontaneously. However, it is also clear that the two curves are quite different in the regime near $\theta = \pi$. Thus, in the regime where m_q is small and N_c is simultaneously large, one expects QCD to interpolate between these two in some manner that depends on how close the system is to the two limits. Now, while it is clear that each of these limits has a discontinuity in the slope at $\theta = \pi$, it is
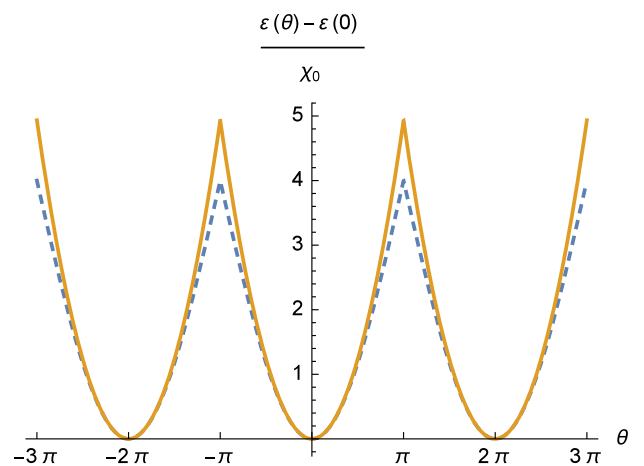

FIG. 1. $\varepsilon(\theta)$ in units of the topological susceptibility and with $\varepsilon(0)$ subtracted off. The dashed line is at the large N_c limit and the solid line is at leading nontrivial order in a chiral expansion for two degenerate light flavors.

not clear *a priori* that the interpolating function also does. It is easy to envision a scenario in which all of the interpolating curves are smooth everywhere with no discontinuity in the slope at $\theta = \pi$ but which have a curvature at π that increases as the limiting cases are approached and diverge at the limits, yielding the sharp forms seen.

Given the importance of understanding the energy dependence of the vacuum as a function of θ, it is natural to explore the prospects of calculating it directly from QCD using lattice methods. As is well known, there is no practical way to do this due to the so-called sign problem. This is generally considered as a practical problem associated with the natural algorithms to compute the energy density. In this paper, we note that there is in an interesting theoretical issue connected to the sign problem.

B. A sign problem

Before outlining the key issues, it is useful to define the quantities of interest. We do so in Euclidean space, the natural setup for lattice studies and for simplicity of discussion work in the continuum limit here. We note in passing that there are subtleties associated with topology when considering the continuum limit of a discrete lattice [11–14]. However, these are unrelated to the issues discussed in this paper. The energy density of the QCD θ vacuum can be given in terms of the QCD partition function in Euclidean space $\mathcal{Z}(\theta, V)$ as

$$\varepsilon(\theta) = -\lim_{V \to \infty} \frac{1}{V} \log \mathcal{Z}(\theta, V). \tag{3}$$

The Euclidean-space QCD partition function is given as a functional integral,

$$\mathcal{Z}(\theta, V) = \int [dA] \det[i\mathcal{D}[A] - M] \exp(-S_{\mathrm{YM}} + i\theta Q), \tag{4}$$

where S_{YM} and Q are functionals of the gluon field configurations on a space-time region of volume V.

We consider the Euclidean-space theory confined to a box of four-dimensional space-time volume $V = L_x L_y L_z L_t$. Ultimately we are interested in infinite volumes. However, for practical calculations on a lattice one must use a finite V and then take it to be large enough to reduce the effect of finite volume effects and estimate, with some accuracy, their size. The boundary conditions imposed on the lattice are often taken to be periodic for boson fields and antiperiodic for fermions.

It is well known that $\mathcal{Z}(\theta, V)$ can be written as a Fourier series over the partition function $\mathcal{Z}_Q(V)$ with fixed topological charge Q:

$$\mathcal{Z}(\theta, V) = \sum_{Q \in \mathbb{Z}} \mathcal{Z}_Q(V) e^{i\theta Q}$$
$$= \sum_{Q \in \mathbb{Z}} \mathcal{Z}_Q(V) \cos(\theta Q), \tag{5}$$

$$\mathcal{Z}_Q(V) = \frac{1}{\pi} \int_0^\pi \mathcal{Z}(\theta) e^{-i\theta Q} \mathrm{d}\theta$$
$$= \frac{1}{\pi} \int_0^\pi \mathcal{Z}(\theta) \cos(\theta Q) \mathrm{d}\theta, \tag{6}$$

where we use the fact that $\mathcal{Z}(\theta)$ is even and periodic in θ [15]. The fixed-topology partition function used here is [5]

$$\mathcal{Z}_Q(V) = \int [dA]_Q \det[i\mathcal{D}[A] - M] \exp(-S_{\mathrm{YM}}), \tag{7}$$

where the integration is performed over all field configurations on a lattice with a given topological charge Q.

The quantity $\mathcal{Z}_Q(V)$ allows us to define a new energy density distribution function:

$$\tilde{\varepsilon}(q, V) = -\frac{1}{V} \log(\mathcal{Z}_{(qV)}(V)), \tag{8}$$

where we define topological charge density $q \equiv \frac{Q}{V}$ with it being understood that qV is an integer, so $\varepsilon(\theta)$ and $\tilde{\varepsilon}(q, V)$ are related by

$$\varepsilon(\theta) = -\lim_{V \to \infty} \frac{1}{V} \log \left(\sum_{Q=-\infty}^{\infty} e^{-\tilde{\varepsilon}(\frac{Q}{V}, V)V} e^{i\theta Q} \right). \tag{9}$$

One expects that the energy density should be an intensive property that depends on the intensive quantity $q = \frac{Q}{V}$. Thus, one expects a well-defined infinite volume limit, so that we can define

$$\tilde{\varepsilon}(q) \equiv \lim_{V \to \infty} \tilde{\varepsilon}(q, V) \quad \text{with} \quad qV \text{ a positive integer}$$
$$= -\lim_{V \to \infty} \frac{1}{V} \log \left(\frac{1}{\pi} \int_0^\pi e^{-\varepsilon(\theta)V} e^{-i\theta qV} \mathrm{d}\theta \right). \tag{10}$$

Clearly, we would like to be able to compute $\varepsilon(\theta)$ directly from QCD. If we could, we could potentially rule out, definitively, the possibility that $\theta = \pi$ by showing explicitly that CP is spontaneously broken at $\theta = \pi$. Ideally, we could answer the question of whether this happens by doing lattice studies at a nonzero θ. However, lattice studies at a θ away from zero are not practical and as a result $\varepsilon(\theta)$ of real QCD remains unknown. The reason they are not is because of a so-called sign problem.

The core of the problem is the oscillatory nature of the functional integrand in Eq. (4) or equivalently the terms in the sum in Eq. (5). Such integrals or sums involve large cancellations which lead to the loss of a considerable

amount of accuracy. This problem also makes the standard Monte Carlo methods impractical: functional integrals with an oscillatory integrand suffer from an integration weight which is not necessarily positive, and in this case normal sampling methods are not practical. Indeed, for this problem the sign problem implies that the computation cost is exponentially large as a function of V.

It is worth recalling why the sign problem implies an exponential expense in terms of computer resources. For simplicity we consider the difference between $\varepsilon(\pi)$ and $\varepsilon(0)$. It is instructive to separate the $Q = 0$ term from the rest and rewrite Eq. (9) as

$$\varepsilon(\pi) - \varepsilon(0) = -\lim_{V \to \infty} \frac{1}{V} \log \left(\frac{A(V) - B(V)}{A(V)} \right)$$

$$\text{with} \quad A(V) \equiv \frac{\mathcal{Z}_{Q=0}(V)}{\mathcal{Z}(\theta = 0, V)}$$

$$\text{and} \quad B(V) \equiv -\frac{2 \sum_{Q=1}^{\infty} (-1)^Q \mathcal{Z}_Q(V)}{\mathcal{Z}(\theta = 0, V)}, \tag{11}$$

where the factor of 2 and the sum over a positive Q in the definition of $B(V)$ reflect the fact that CP invariance implies that $\tilde{\varepsilon}(\frac{-Q}{V}, V) = \tilde{\varepsilon}(\frac{Q}{V}, V)$. In deriving Eq. (11), we used the well-known fact [16] that

$$\lim_{V \to \infty} \mathcal{Z}_{Q=0}(V) = \lim_{V \to \infty} \mathcal{Z}(\theta = 0, V). \tag{12}$$

Equation (12) implies that $A(V)$ is a subexponential function of V. This in turn implies that in order to capture the difference between $\varepsilon(\pi)$ and $\varepsilon(0)$, $B(V)$ must cancel $A(V)$ to one part in $\exp(V(\varepsilon(\theta) - \varepsilon(0)))$. If one assumes that these differences in ε are of the order unity, one sees that to extract the energy dependence via a direct summation of the Fourier series requires cancellations due to the fluctuating sign that scale exponentially in the volume. This, in turn, implies that to get a meaningful result one would need to compute both A and B with an accuracy that also scales exponentially in the volume. But, in a Monte Carlo algorithm the accuracy scales as the square root of the resources, so to get sufficient accuracy in each Q sector and summing over sectors requires resources that scale exponentially with the volume.

This exponentially serious sign problem implies that, as a practical matter, simply using Eq. (9) to get $\varepsilon(\theta)$ is not practical except for very small systems that are well away from describing the infinite volume result in $3 + 1$ dimensions. A similar exponentially serious sign problem also occurs in QCD with a nonzero chemical potential. Some possible solutions have been proposed to evade these sign problems [17–26]. Among all of these proposals, using an imaginary chemical potential has generated significant attention [27,28]. Similarly, calculating an imaginary θ first and then analytically continuing it to a real θ in order to avoid a sign problem have been used to calculate

deconfinement temperature, electric dipole moment, and so forth [29–34]. In practice, analytic continuation from an imaginary θ to a real θ can be done for a real θ fairly near 0. This is because when θ is very small, using any reasonable expansion form of energy density, we can neglect higher order terms in the expansion and only take the several lowest order terms as an approximate analytical form in which to extrapolate to a real θ. However, because we lack knowledge of the exact form for energy density $\varepsilon(\theta)$ in real QCD, it is not practical to analytically continue an imaginary θ to an arbitrary real $\theta \sim \pi$, where we have no reason to expect high order terms to be negligible.

The sign problem is generally thought of as a practical difficulty that prevents practical calculations of $\varepsilon(\theta)$. However, there is also an underappreciated theoretical question associated with the sign problem that is the focus of this paper. The issue is the following: suppose that one is able to determine $\tilde{\varepsilon}(q)$ with arbitrary accuracy. Does this give us enough information, in principle, to reconstruct $\varepsilon(\theta)$ by summing over the topological sectors? Intuitively, it may seem obvious at first blush that the answer is yes. After all, both quantities are thought to be intensive and thus to be well defined in the infinite volume limit. Thus, it seems highly plausible that while the two intensive quantities depend on each other, neither should depend on finite volume corrections. However, as will be shown here, things are a bit more subtle than this.

To pose this issue mathematically let us define the quantity

$$\underline{\varepsilon}(\theta) = -\lim_{V \to \infty} \frac{1}{V} \log \left(\sum_Q e^{-\tilde{\varepsilon}(\frac{Q}{V})V} e^{i\theta Q} \right)$$

$$= -\lim_{V \to \infty} \lim_{\tilde{V} \to \infty} \frac{1}{V} \log \left(\sum_Q e^{-\tilde{\varepsilon}(\frac{Q}{V}, \tilde{V})V} e^{i\theta Q} \right). \tag{13}$$

The new notation $\underline{\varepsilon}(\theta)$ is used to distinguish it from the $\varepsilon(\theta)$ defined in Eq. (9). The definitions of $\underline{\varepsilon}(\theta)$ and $\varepsilon(\theta)$ differ only in an ordering of limits. In $\underline{\varepsilon}(\theta)$, $\tilde{V}$ is taken to infinity prior to taking V to infinity, while in $\varepsilon(\theta)$ the limits are taken simultaneously. The question, then, amounts to whether or not $\underline{\varepsilon}(\theta) = \varepsilon(\theta)$.

As noted above, it seems rather plausible that $\underline{\varepsilon}(\theta) = \varepsilon(\theta)$ since it relates one intensive quantity to another. It turns out that, in principle, one *can* fully reconstruct $\varepsilon(\theta)$, from $\tilde{\varepsilon}(q)$ via directly summing over topological sectors for cases in which ε curves upward [i.e. $\varepsilon''(\theta) > 0$] everywhere in the region $-\pi < \theta < \pi$. Such cases are interesting since they must spontaneously break CP at $\theta = \pi$ due to a discontinuity in the slope (as in the examples in Fig. 1). However, it turns out that *if* $\varepsilon(\theta)$ has regions for which $\varepsilon''(\theta) < 0$, then, as a result of the severe sign problem, even perfect knowledge of $\tilde{\varepsilon}(q)$ is insufficient to reconstruct those regions. Remarkably, in these cases, in order to fully

reconstruct the infinite volume behavior for $\varepsilon(\theta)$ one needs to understand the finite volume effects for $\tilde{\varepsilon}(q, V)$.

This paper is organized as follows. In the following section we demonstrate the phenomenon in the context of a simple "toy" problem that, while not being QCD, illustrates the issues. The toy problem is one for which a dilute instanton gas is a valid approximation. It is found that in the region where $\varepsilon''(\theta) > 0$, $\underline{\varepsilon}(\theta) = \varepsilon(\theta)$. The key analysis in understanding this behavior is analyzed in the subsequent section. The central point is that the relevant sums can be approximated as integrals that can be analyzed in terms of a saddle point approximation. The argument is generalized in the next section; it is shown that the phenomenon can be expected to occurs when $\varepsilon''(\theta) \leq 0$ with no restriction to the dynamics of the toy problem. The paper concludes with a discussion of these results.

II. A TOY PROBLEM: THE DILUTE INSTANTON GAS

A. $\varepsilon(\theta)$ versus $\underline{\varepsilon}(\theta)$

It is well known that neither Yang-Mills nor QCD can be approximated well by a dilute instanton gas [35]. Nevertheless, we are going to first consider a dilute instanton gas with instantons of fixed action and fixed size as a toy model to illustrate the underlying issues. We envision such a model as arising from some unspecified theory (not necessarily in $3 + 1$ dimensions), which has instanton and anti-instanton classical solutions and an analog of the θ term and, in some parametric limit, the theory becomes semiclassical and is dominated by widely spaced instantons of a fixed size.

In such models, it is well known that $\mathcal{Z}_Q(V)$ is obtained by summing over the effects of instantons and anti-instantons

$$\mathcal{Z}_Q(V) = \mathcal{Z}_0 \sum_{n=0}^{\infty} \frac{(\frac{1}{2}ce^{-S_0}V)^{|Q|+2n}}{n!(n+|Q|)!} = \mathcal{Z}_0 I_Q(ce^{-S_0}V), \quad (14)$$

where $\mathcal{Z}_0$ is a prefactor that sums up effects other than instantons, S_0 is the action of a single instanton, V is the space

time volume, and c is a constants with dimension 4 that includes the effects of fluctuations. The sum yields the modified Bessel function $I_Q(ce^{-S_0}V)$.

From this it is simple to identify $\varepsilon(\theta)$. For an integer Q,

$$I_Q(z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{z\cos\theta} \cos(Q\theta) d\theta, \quad (15)$$

which from Eq. (5) implies that

$$\mathcal{Z}(\theta) = \mathcal{Z}_0 \exp\left(ce^{-S_0}V\cos(\theta)\right), \quad (16)$$

from which the standard dilute instanton gas expression [36] for $\varepsilon(\theta)$ follows:

$$\varepsilon(\theta) = \varepsilon_0 + \chi_0(1 - \cos(\theta))$$

with $\quad \chi_0 \equiv ce^{-S_0}, \quad$ and

$$\varepsilon_0 \equiv \frac{-\log(\mathcal{Z}_0)}{V} - ce^{-S_0}. \quad (17)$$

Next let us consider $\tilde{\varepsilon}$, the energy density associated with the fixed topological sectors. Previously it was argued that this should be an intensive quantity which depends only on $q = Q/V$. Let us verify that this is true for the dilute instanton gas model. Start with the $\mathcal{Z}_Q$ given in Eq. (14) and exploit the series expansion of the modified Bessel function $I_\nu(\nu z)$ around the uniform limit $\nu \to \infty$ through positive real values,

$$I_\nu(\nu z) \sim \frac{e^{\nu\eta}}{(2\pi\nu)^{\frac{1}{2}}(1 + z^2)^{\frac{1}{4}}} \sum_{k=0}^{\infty} \frac{U_k(p)}{\nu^k} \quad \text{with}$$

$$\eta = (1 + z^2)^{\frac{1}{2}} + \log\left(\frac{z}{1 + (1 + z^2)^{\frac{1}{2}}}\right), \quad \text{and}$$

$$p = (1 + z^2)^{-\frac{1}{2}}, \quad (18)$$

where $U_k(p)$s are polynomials in p of degree $3k$ with $U_0(p)$ equal to unity [37]. This yields

$$\tilde{\varepsilon}(q, V) = \varepsilon_0 + \chi_0 + q\log\left(\frac{q + \sqrt{\chi_0^2 + q^2}}{\chi_0}\right) - \sqrt{\chi_0^2 + q^2} + \frac{\log\left((2\pi)^2 V^2 (q^2 + \chi_0^2)\right) - 4\log\left(\sum_k \frac{U_k\left(\frac{q}{\sqrt{q^2 + \chi_0^2}}\right)}{(qV)^k}\right)}{4V}. \quad (19)$$

Thus, in the limit $V \to \infty$ with a fixed q,

$$\tilde{\varepsilon}(q) = \lim_{V \to \infty} \tilde{\varepsilon}(q, V) = \varepsilon_0 + \chi_0 + q\log\left(\frac{q + \sqrt{\chi_0^2 + q^2}}{\chi_0}\right) - \sqrt{\chi_0^2 + q^2} = \varepsilon_0 + \chi_0 + q\sinh^{-1}\left(\frac{q}{\chi_0}\right) - \sqrt{\chi_0^2 + q^2}. \quad (20)$$

As advertised, $\tilde{\varepsilon}$ is an intensive quantity only dependent on q, a result previously derived in [16] for the analogous case of QCD with one light flavor which also has a cosinusoidal dependance of $\varepsilon(\theta)$.

Before proceeding, it will be useful to note that the form of $\tilde{\varepsilon}(q)$ as a mapping of real numbers to real numbers is unique. However, if we continue to the complex q plane, as we will have cause to do later, the functional form is multibranched with branch points at $\pm i\chi_0$ and branch cuts extending along the imaginary axis to infinity. The imaginary axis is free of a branch cut only for $-\chi_0 < iq < \chi_0$. The form of $\tilde{\varepsilon}(q)$ used here corresponds to the principal branch.

With knowledge of $\tilde{\varepsilon}$, we are in a position to investigate whether it is sufficient to determine $\tilde{\varepsilon}$. To do so we must calculate $\underline{\varepsilon}(\theta)$, defined in Eq. (13), and ask whether it agrees with $\varepsilon(\theta)$. The calculation of $\underline{\varepsilon}(\theta)$ involves evaluating $\sum_Q e^{-\tilde{\varepsilon}(\frac{Q}{V})V} e^{i\theta Q}$. Unfortunately, we know of no way to do this analytically, so the calculation of $\underline{\varepsilon}(\theta)$ was done numerically with the large V limit approximated by V being taken as large enough that $V\chi_0 \gg 1$ (the results were quite stable except in the immediate vicinity of $\theta = \pi/2$ by $V\chi_0 = 30$). In Fig. 2, $\varepsilon(\theta)$ and the numerically calculated $\underline{\varepsilon}(\theta)$ are plotted in units of the topological susceptibility.

There are several things to notice about Fig. 2. The first is that for $0 < \theta < \pi/2$ one sees that up to the quality of the numerics, $\varepsilon(\theta) = \underline{\varepsilon}(\theta)$. However, for $\pi/2 < \theta < \pi$, $\varepsilon(\theta) \neq \underline{\varepsilon}(\theta)$. This demonstrates quite clearly that the answer to the question of whether knowledge of $\tilde{\varepsilon}(q)$ is sufficient to reconstruct $\varepsilon(\theta)$ via direct summation over topological sectors is, it depends. The focus of the reminder of this paper is on what it depends upon. Before attacking this question, a couple of other observations are in order. The first is that to numerical accuracy, it appears that $\underline{\varepsilon}(\theta)$ is linear in the region $\pi > \theta > \frac{\pi}{2}$. Beyond that, $\theta = \pi$ is linear with the opposite slope, yielding a discontinuous slope.

The point that separates the region for which $\varepsilon(\theta) = \underline{\varepsilon}(\theta)$ from the region for which $\varepsilon(\theta) \neq \underline{\varepsilon}(\theta)$ appears from the numerics to be $\theta = \pi/2$—or something very close to it.
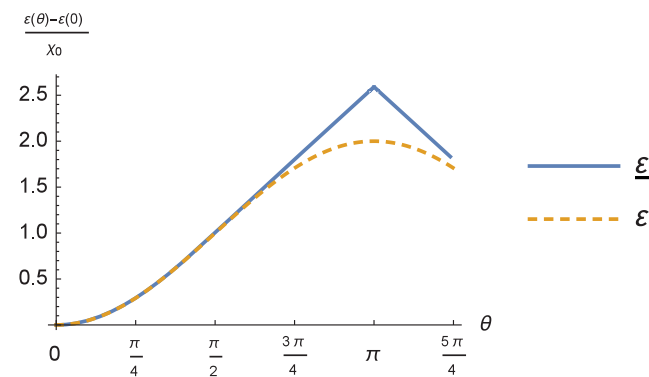

FIG. 2. $\varepsilon(\theta)$ and $\underline{\varepsilon}(\theta)$ in units of χ_0 for a dilute instanton gas model. The two functional forms are clearly different for $\theta > \frac{\pi}{2}$, indicating that in this region knowledge of $\tilde{\varepsilon}(q)$ is insufficient to reproduce $\varepsilon(\theta)$ by direct summation even though they are both intensive quantities of interest. The numerical evaluation of $\underline{\varepsilon}(\theta)$ was done with a large, but finite, four-dimensional volume. $V\chi_0 = 350$ in the calculations.

Assuming that the point really is exactly at $\theta = \pi/2$ gives rise to the issue of what makes that point special. An obvious conjecture is that $\theta = \pi/2$ is a point of inflection. As will be seen in the course of this paper, there is very strong evidence that this conjecture is correct.

B. The severity of the sign problem

One thing illustrated quite clearly by the different energy densities in Fig. 2 for our toy problem, the dilute instanton gas, is just how serious the sign problem can be. As one approaches the infinite volume limit the fractional difference between $\tilde{\varepsilon}(q)$ and $\tilde{\varepsilon}(q, V)$ clearly goes to zero in the infinite volume limit. However, even as this difference becomes vanishingly small at large volumes, their associated energy functions of θ, $\underline{\varepsilon}(\theta)$, and $\varepsilon(\theta)$ respectively become very different for $\theta > \frac{\pi}{2}$. Before turning to more realistic situations it is worth exploring why this is so for the toy problem.

To understand how this vanishingly small difference in the dilute instanton gas results in order unity differences between $\underline{\varepsilon}(\theta)$ and $\varepsilon(\theta)$, it is important to first recognize that the key quantities in the calculations are not the energy densities but the generating functions, $\mathcal{Z} \sim \exp(-V\varepsilon)$. Note, that the correction term of the order $1/V$ in the expansion of $\tilde{\varepsilon}(q, V)$ in Eq. (20) leads to an order unity shift in $\mathcal{Z}_Q$. One might be tempted to ascribe the order unity differences between $\underline{\varepsilon}(\theta)$ and $\varepsilon(\theta)$ to these differences. However, the underlying cause is more subtle than this.

To see this, it is instructive to compute the θ dependence using various approximations to $\tilde{\varepsilon}(q, V)$ including various orders of correction in $1/V$ and truncating beyond it. We define $\tilde{\varepsilon}_n(q, V)$ as an approximation to $\tilde{\varepsilon}(q, V)$ that includes all terms up to $\mathcal{O}(V^{-n})$ and truncates the rest. Thus,

$$\tilde{\varepsilon}_0(q, V) = \tilde{\varepsilon}(q)$$

$$\tilde{\varepsilon}_1(q, V) = \tilde{\varepsilon}(q) + \frac{\log\left((2\pi)^2 V^2 (q^2 + \chi_0^2)\right)}{4V}$$

$$\tilde{\varepsilon}_2(q, V) = \tilde{\varepsilon}_1(q, V) - \frac{U_1\left(\frac{q}{\sqrt{q^2 + \chi_0^2}}\right)}{qV^2} \quad \ldots \quad (21)$$

This allows us to define $\varepsilon_n(\theta)$ as

$$\varepsilon_n(\theta) = -\lim_{V \to \infty} \frac{1}{V} \log\left(\sum_{Q=-\infty}^{\infty} e^{-\tilde{\varepsilon}_n(\frac{Q}{V}, V)V} e^{i\theta Q}\right), \quad (22)$$

that is, it is the standard definition except that it replaces $\tilde{\varepsilon}(\frac{Q}{V}, V)$ by $\tilde{\varepsilon}_n(\frac{Q}{V}, V)$ and thereby includes some fixed level of $1/V$ corrections. The $\varepsilon_n(\theta)$ can be computed numerically using a large but finite value for the volume.

If the explanation for the difference between $\underline{\varepsilon}(\theta)$ and $\varepsilon(\theta)$ were due to the order unity difference in the generating functions arising from the $1/V$ corrections in $\tilde{\varepsilon}(q, V)$, one would find that up to numerical accuracy $\varepsilon_n(\theta)$ would be
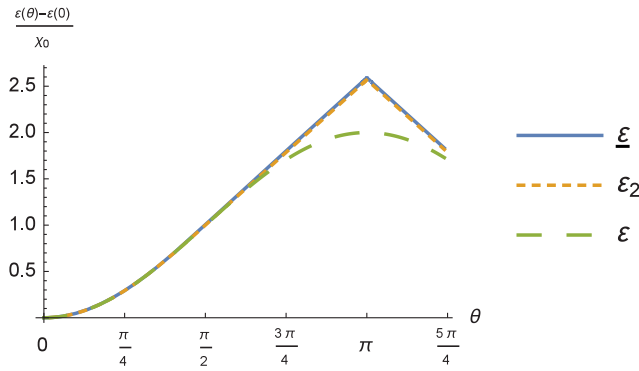
FIG. 3. A comparison of $\underline{\varepsilon}(\theta)$, $\varepsilon_2(\theta)$, and $\varepsilon(\theta)$.

equal to $\varepsilon(\theta)$ for all cases where $n \geq 1$. However, this is not the case. In Fig. 3, $\underline{\varepsilon}(\theta)$, $\varepsilon_2(\theta)$, and $\varepsilon(\theta)$ are plotted. It is immediately apparent that $\varepsilon_2(\theta) \neq \varepsilon(\theta)$ differs substantially from $\varepsilon(\theta)$, coinciding only in the region $0 < \theta < \pi/2$; indeed it coincides with $\underline{\varepsilon}(\theta)$. Why, then, does $\underline{\varepsilon}(\theta)$ differ from $\varepsilon(\theta)$?.

Ultimately the answer resides in the discussion associated with Eq. (11). Note that to obtain $\varepsilon(\theta)$ accurately, one requires cancellations which are exponentially sensitive with respect to the volume. Thus, it is not that surprising that inclusion of any given level of power law correction in the volume might be insufficient. However, if this is the case, then why does $\underline{\varepsilon}(\theta)$ equal $\varepsilon(\theta)$ in the regime where $|\theta| < \pi/2$? The next section is aimed at answering that question in the context of this toy model.

III. SADDLE POINT APPROXIMATION FOR THE INSTANTON GAS MODEL

A. A useful identity

To proceed, we will exploit a powerful identity relating the limit of integrals and sums which holds for a wide class of functions $f(x)$ that goes to positive infinity as $x \to \pm\infty$ along the real axis:

$$\lim_{\lambda \to 0} \frac{\int_{-\infty}^{\infty} dx \exp\left(-\lambda f\left(\frac{x}{\lambda}\right)\right) \exp(i\theta x)}{\sum_{n=-\infty}^{\infty} \exp\left(-\lambda f\left(\frac{n}{\lambda}\right)\right) \exp(i\theta n)} = 1, \quad (23)$$

which holds for a large class of functions, f, for $-\pi < \theta < \pi$. The importance of this identity to our problem is that at large V the sum $\sum_Q e^{-\tilde{\varepsilon}(\frac{Q}{V})V} e^{i\theta Q}$ can be replaced by an integral over Q, which is far easier to study analytically. The identity is not trivial. Of course, one expects that a sum will converge to an integral in situations where the limit of the sum yields Reimann's construction for integrals. However, that does not happen here. Note that the cases of interest to us here are ones in which the first term in the sum is exponentially larger than the total, even as the limit is approached.

The derivation of the identity is straightforward but is worth sketching here as it illustrates some key features of the problem that we will exploit later. The first step is to notice that $\sum_{n=-\infty}^{\infty} \delta(x - n) = \sum_{k=-\infty}^{\infty} \exp(i2\pi kx)$, from which it follows that

$$\sum_{n=-\infty}^{\infty} \exp\left(-\lambda f\left(\frac{n}{\lambda}\right)\right) \exp(i\theta n)$$

$$= \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} dx \exp\left(-\lambda f\left(\frac{x}{\lambda}\right)\right) \exp(i\theta x) \exp(i2\pi kx)$$

$$= \sum_{k=-\infty}^{\infty} \lambda \int_{-\infty}^{\infty} dy \exp\left(-\lambda(f(y) - i(\theta + 2\pi k)y)\right), \quad (24)$$

assuming that the various integrals and sums all converge. The last form of Eq. (24) is instructive. Integrals of this form in the limit of large λ are typically accurately approximated via the saddle point approximation assuming appropriate analyticity properties [38]. Up to power law factors in λ, the kth integral is given by

$$\int_{-\infty}^{\infty} dy \exp(-\lambda(f(y) - i(\theta + 2\pi k)y)) \sim \max_k \exp(-\lambda g_k)$$

where $g_k = \min_j(g_k^j)$ with $g_k^j = (f(y_k^j) - i(\theta + 2\pi k)y_k^j)$, (25)

where y_k^j is the jth saddle point for the function $(f(y) - i(\theta + 2\pi k)y)$ and thus satisfies

$$\left. \frac{d(f(y(t)) - i(\theta + 2\pi k)y(t))}{dt} \right|_{t=t_k^j} = 0, \quad (26)$$

where $y(t)$ is a contour in the complex plane and $y_k^j \equiv y(t_k^j)$. Note that the derivative of both the real and imaginary parts of $f(y(t)) - i(\theta + 2\pi k)y(t)$ have to vanish at the saddle point. The key point is that each integral is exponentially dominated by its minimum saddle point in the complex plane as λ becomes large.

Thus, for a wide class of functions, one expects that, at large λ, the sum over k to be exponentially dominated by the term with the smallest value of g_k. For many typical cases one expects this to be the $k = 0$ term for $-\pi < \theta < \pi$. In these cases, as θ goes beyond π, one expects the $k = -1$ term to take over as the dominant term, as θ exceeds 3π for the $k = -2$ term to take over, and so forth. If one focuses on the region $-\pi < \theta < \pi$ for these typical cases, the sum at large λ is exponentially well approximated by the $k = 0$ term. However, from the middle line of Eq. (24), the $k = 0$ term is precisely the integral in the numerator of Eq. (23) which establishes Eq. (23) for this class of function.

There is an important subtlety. The saddle point approximation is obtained by distorting the path of the integral from along the real line to some other path with the same

end points in the complex plane. The integrals will coincide provided the integrand is analytic everywhere in the region enclosed by the two paths. Distorting the path through saddle points along the paths with the steepest decent allows one to show that the region near the saddle point dominates the integral exponentially—at least locally—and thereby justifies the approximation. Saddle points are not the only points that can exponentially dominate the integrals. Suppose the function, f, is not analytic everywhere, but has a branch cut singularity. In that case, the branch point can act in a manner quite analogous to a saddle point and the argument given above goes through with minor changes. In particular, it is easy to see that circumstances can arise such that integration paths that are distorted to go around a branch point in f and are arbitrarily close to the branch cut on each side can be exponentially dominated at large λ by the contribution in the immediate vicinity of the branch point. When this happens the branch point plays the same basic role as a saddle point in that the value of the function at the branch point can determine the value of the integral up to subexponential factors. Note that there may be more than one branch point and the relevant branch point may be determined by the need to close contours at infinity.

Assuming that Eq. (23) holds, it can be exploited in computing $\underline{\varepsilon}(\theta)$. We can replace the sum over Q by an integral in computing since we are taking the infinite volume limit. Thus,

$$
\begin{aligned}
\underline{\varepsilon}(\theta) &= -\lim_{V\to\infty} \frac{1}{V} \log\left(\sum_Q e^{-\tilde{\varepsilon}(\frac{Q}{V})V} e^{i\theta Q} \right) \\
&= -\lim_{V\to\infty} \frac{1}{V} \log\left(\int dQ e^{-\tilde{\varepsilon}(\frac{Q}{V})V} e^{i\theta Q} \right) \\
&= -\lim_{V\to\infty} \frac{1}{V} \log\left(V \int dq e^{-V(\tilde{\varepsilon}(q)-i\theta q)} \right) \\
&= \tilde{\varepsilon}(q_\theta^{\mathrm{sp}}) - i\theta q_\theta^{\mathrm{sp}},
\end{aligned}
\tag{27}
$$

where the saddle point approximation is invoked in the last equality where q_θ^{sp} is the dominant saddle point associated with $\tilde{\varepsilon}(q) - i\theta q$.

On physical grounds one expects dominant saddle points to be on the imaginary axis in q: in Euclidean space, $\frac{1}{V}\frac{\partial \log(\mathcal{Z}(\theta))}{\partial \theta} = iq$, which means that a real θ is associated with an imaginary q. Note, moreover, that $\tilde{\varepsilon}(q)$ is an even function. This means that wherever $\tilde{\varepsilon}(q)$ is analytic along the imaginary axis $(\tilde{\varepsilon}(q) - i\theta q)$ will be real along the imaginary axis—i.e. have a constant phase of zero. Thus, the condition for a saddle point at $q = iq_0$ is

$$
\frac{\partial (\tilde{\varepsilon}(ix) - i\theta \cdot ix)}{\partial x}\bigg|_{x=q_0} = 0.
\tag{28}
$$

The upshot of this, plus the last equality in Eq. (27), is that in the infinite volume limit, one expects $\tilde{\varepsilon}$ to be related to $\underline{\varepsilon}$ by an analog of a Legendre transformation—but with a critical factor of i:

$$
\tilde{\varepsilon}(q(\theta)) = \underline{\varepsilon}(\theta) + i\theta q(\theta) \quad \text{with} \quad q(\theta) = i\frac{\partial \underline{\varepsilon}(\theta)}{\partial \theta}
$$

$$
\underline{\varepsilon}(\theta(q)) = \tilde{\varepsilon}(q) - i\theta(q)q \quad \text{with} \quad \theta(q) = -i\frac{\partial \tilde{\varepsilon}(q)}{\partial q}.
\tag{29}
$$

Equation (29) is central to the analysis. One key point is that, by its structure, Eq. (29) requires $\underline{\varepsilon}(\theta)$ and $\tilde{\varepsilon}(q)$ to be continued into the complex plane. It is also important to recall the limitations of Eq. (29). Its validity requires (i) that $\tilde{\varepsilon}(q)$ is analytic for at least some region along the imaginary axis, (ii) that the $k = 0$ integral in Eq. (24) is dominated by a saddle point along the imaginary axis, and (iii) that the identity in Eq. (23) holds for $\tilde{\varepsilon}(q) - i\theta q$, which follows if the sum on k in Eq. (24) is exponentially dominated by the $k = 0$ term at large V.

B. The saddle point approximation for the dilute instanton gas for $|\theta| < \pi/2$

Let us return to the dilute instanton gas and focus on what the saddle point approximation tells us about the relationship of $\underline{\varepsilon}(\theta)$ to $\varepsilon(\theta)$ for the regime $|\theta| < \pi/2$. We will start by assuming that the conditions justifying Eq. (29) hold and see what that implies. Subsequently, we will argue that these conditions should hold.

Let us test the hypothesis that in the regime $|\theta| < \pi/2$, $\underline{\varepsilon}(\theta) = \varepsilon(\theta)$, which for the dilute instanton approximation is given by $\varepsilon(\theta) = \varepsilon_0 + \chi_0(1 - \cos(\theta))$. If this hypothesis is correct, then $\frac{\partial \varepsilon}{\partial \theta} = \chi_0 \sin(\theta)$, and the first form of Eq. (29) becomes

$$
\tilde{\varepsilon}(i\chi_0 \sin(\theta)) = \varepsilon_0 + \chi_0(1 - \cos(\theta)) - \chi_0 \sin(\theta)\theta.
\tag{30}
$$

Note that we have shown previously that for the dilute instanton gas, $\tilde{\varepsilon}(q) = \varepsilon_0 + \chi_0 - \sqrt{\chi_0^2 + q^2} + q\sinh^{-1}(\frac{q}{\chi_0})$. Inserting $q(\theta) = i\frac{\partial \varepsilon}{\partial \theta} = i\chi_0 \sin(\theta)$ in this yields

$$
\begin{aligned}
&\tilde{\varepsilon}(i\chi_0 \sin(\theta)) \\
&= \varepsilon_0 + \chi_0\left(1 - \sqrt{1 - \sin(\theta)^2} - \chi_0 \sin(\theta)\sin^{-1}(\sin(\theta))\right) \\
&= \varepsilon_0 + \chi_0(1 - \cos(\theta)) - \chi_0 \sin(\theta)\theta,
\end{aligned}
\tag{31}
$$

where the second equality holds for $|\theta| < \pi/2$, where $\sin^{-1}(\sin(\theta)) = \theta$ and $\sqrt{1 - \sin(\theta)^2} = \cos(\theta)$. Thus, in the domain $|\theta| < \pi/2$, the left-hand side of Eq. (30) is indeed equal to the right and the hypothesis that $\underline{\varepsilon}(\theta) = \varepsilon(\theta)$ is consistent—providing the assumptions underlying Eq. (29) hold.

This demonstration of consistency in hardly surprising: it justifies the empirical observation in Sec. II based on the numerical results of direct summation with large volumes that $\underline{\varepsilon}(\theta) = \varepsilon(\theta)$ in the regime $|\theta| < \pi/2$. Of course, the demonstration here depends on the assumption that the three conditions justifying Eq. (29) hold. Condition (i), the existence of a region along the imaginary axis where $\tilde{\varepsilon}(q)$ is analytic, clearly holds for the form derived for the dilute instanton gas: $\tilde{\varepsilon}(q)$ is analytic from $-i\chi_0$ to $i\chi_0$. It is also highly plausible the condition (ii) holds. There is a saddle point at $q = i\chi_0 \sin(\theta)$.

Condition (iii), that the $k = 0$ integral dominates the sum over k in Eq. (24), requires a bit of care. It is easy to see that there are no saddle points along the imaginary axis for any of the integrals associated with $k \neq 0$. If saddle points exist for $k \neq 0$, then the second form of Eq. (29) with the substitution $\theta \to \theta + 2\pi k$ would determine them. Thus,

$$\theta + 2\pi k = -i \frac{\partial \tilde{\varepsilon}(q)}{\partial q}\bigg|_{q=q^{\mathrm{sp}}}. \tag{32}$$

Using $\tilde{\varepsilon}(q) = \varepsilon_0 + \chi_0 - \sqrt{\chi_0^2 + q^2} + q\sinh^{-1}\left(\frac{q}{\chi_0}\right)$, this means that

$$\theta + 2\pi k = -\sin^{-1}\left(i\frac{q^{\mathrm{sp}}}{\chi_0}\right). \tag{33}$$

Recall that we are considering q^{sp} on the imaginary axis so that iq^{sp} is real and that $\tilde{\varepsilon}(q)$ corresponds to the principal branch of the function, which forces the derived $\sin^{-1}$ to also correspond to the principal branch and thus to take the value from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$. This in turn means that for $-\pi < \theta < \pi$ there are only saddle points on the imaginary axis for $k = 0$ and $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$.

Since physically one expects any dominant saddle point to be on the imaginary axis and there are no saddle points for $k \neq 0$ along the imaginary axis, it is highly plausible that the integrals $k \neq 0$ are dominated by regions near branch points. Thus for these cases

$$\varepsilon_k(\theta) \equiv -\lim_{V \to \infty} \frac{1}{V} \log \left(V \int dq e^{-V(\tilde{\varepsilon}(q) - i(\theta + 2\pi k)q)} \right)$$
$$= \tilde{\varepsilon}(q^{\mathrm{bp}}) - (\theta + 2\pi k)(iq^{\mathrm{bp}}), \tag{34}$$

where q^{bp} is the branch point and for convenience we have introduced the symbol $\varepsilon_k(\theta)$. For the dilute instanton gas at $q^{\mathrm{bp}} = \pm i\chi_0$. The relevant branch point is determined by the sign of $\theta + 2\pi k$, since one wants to close contours at infinity. Thus, for $k \neq 0$ with the dilute instanton gas,

$$\varepsilon_k(\theta) = \varepsilon_0 + \chi_0 + \chi_0 \left(|\theta + 2\pi k| - \frac{\pi}{2} \right). \tag{35}$$

Note that in the domain $|\theta| < \frac{\pi}{2}$, $\varepsilon_k(\theta)$ for $k \neq 0$ is always greater than $\underline{\varepsilon}(\theta) = \varepsilon_0 + \chi_0(1 - \cos(\theta))$ (which was determined from the saddle point of the $k = 0$ integral). This means the $k \neq 0$ integrals are exponentially suppressed at large volumes compared to the $k = 0$ ones and can be neglected, which establishes condition (iii).

To summarize, for $|\theta| < \frac{\pi}{2}$ it has been shown self-consistently that at large volumes for the dilute instanton gas, the sum over Q can be replaced with an integral, this integral is dominated by a saddle point on the imaginary axis and that saddle point approximation to the integral yields $\underline{\varepsilon}(\theta) = \varepsilon(\theta)$. This is consistent with what was observed empirically by direct numerical summation. It also agrees with the naive intuition that $\varepsilon(\theta)$ ought to be directly obtainable from $\tilde{\varepsilon}(q)$ via direct summation over topological sectors given that both functions are intensive and independent of the volume.

C. The saddle point approximation for the dilute instanton gas for $\pi/2 < |\theta| < \pi$

Let us consider what happens when we try to extend the analysis based on the saddle point approximation to the $k = 0$ integral for the dilute instanton gas in the regime $\pi/2 < |\theta| < \pi$. It should be clear that the approximation breaks down. The easiest way to see this is to repeat the analysis at the beginning of Sec. III B. That is, one can assume that $\underline{\varepsilon}(\theta)$ does equal $\varepsilon(\theta)$. As in the $|\theta| < \pi/2$ case, one obtains Eq. (30) and the first equality of Eq. (31). However, the second equality in Eq. (31) fails for $\pi/2 < |\theta| < \pi$ since $\sin^{-1}(\sin(\theta)) = \pm(\pi - \theta)$ (where $\pm$ is chosen to be the same as the sign of θ) rather than θ and $\sqrt{1 - \sin(\theta)^2} = -\cos(\theta)$ rather than $\cos(\theta)$. Thus it implies that

$$\tilde{\varepsilon}(i\chi_0 \sin(\theta)) = \varepsilon_0 + \chi_0(1 + \cos(\theta)) \pm \chi_0 \sin(\theta)(\pi - \theta)$$
$$\neq \varepsilon_0 + \chi_0(1 - \cos(\theta)) - \chi_0 \sin(\theta)\theta, \tag{36}$$

where the second form is $\tilde{\varepsilon}(i\chi_0 \sin(\theta))$ as given in Eq. (30). This is a clear inconsistency and implies the starting assumption that $\underline{\varepsilon}(\theta) = \varepsilon(\theta)$ does not hold in this regime. This analysis shows why in the regime $\pi/2 < |\theta| < \pi$, the intuitive notion that $\underline{\varepsilon}(\theta)$ obtained from direct summation $\tilde{\varepsilon}(q)$ in the infinite volume limit should be equal to $\varepsilon(\theta)$ fails due to an extremely severe sign problem.

There is an equivalent way to see this by working directly from Eq. (33) which holds for $k = 0$ as well as $k \neq 0$. By the same logic that the $k \neq 0$ integrals were shown not to have saddle points in the regime $|\theta| < \pi/2$, we can see that there are no saddle points for $k = 0$ when $\pi/2 < |\theta| < \pi$. Equation (33) tells us that we are at saddle point on the imaginary axis for the $k = 0$ integral for the dilute instanton gas when $\theta = -\sin^{-1}(i\frac{q^{\mathrm{sp}}}{\chi_0})$ and the sine inverse is in its principle branch. Since the principal branch of the inverse sign has a range from $-\pi/2$ to $\pi/2$, this can

never be satisfied when $|\theta| > \pi/2$. Thus, there are no saddle points on the imaginary axis in this regime.

In the absence of saddle points one expects the integrals to be dominated by the appropriate branch points. The analysis of these is the same as for the $k \neq 0$ integrals in the previous subsection in the discussion accompanying Eqs. (34) and (35). As in that case, the branch points occur at $q^{bp} = \pm i\chi_0$ and the appropriate branch point is fixed by the need to close contours at infinity. The result in this regime is as given in Eq. (35), namely that $\underline{\varepsilon}(\theta) = \varepsilon_0 + \chi_0 + \chi_0|\theta| - \frac{\pi}{2}$. This confirms that $\underline{\varepsilon}(\theta) \neq \varepsilon(\theta)$ in this regime and is also consistent with what was observed numerically: $\underline{\varepsilon}$ is a linear function of θ in this regime. It is easy to see, by the same type of analysis as done in Sec. III B, that in this regime $\varepsilon_k(\theta)$ for $k \neq 0$ is always greater than $\underline{\varepsilon}(\theta)$ and thus that the $k = 0$ integral is exponentially dominant at large V.

To summarize, for $\pi/2 < |\theta| < \pi$ it has been shown that there are no saddle points on the imaginary axis and the dominant $k = 0$ integral is dominated by a branch point. This yields a linear relationship: $\underline{\varepsilon}(\theta) = \varepsilon_0 + \chi_0 + \chi_0|\theta| - \frac{\pi}{2} \neq \varepsilon(\theta)$. This result is consistent with what was observed empirically by direct numerical summation. Moreover, it shows why the naive intuition that $\varepsilon(\theta)$ ought to be directly obtainable from $\tilde{\varepsilon}(q)$ via direct summation over topological sectors given that both functions are intensive and independent of the volume fails.

D. Analytic considerations

Note that the sign problem at large volume can be viewed as an issue in analytic continuation. In the region $|\theta| < \pi/2$ where we can reconstruct $\varepsilon(\theta)$ from $\tilde{\varepsilon}(q)$ via direct summation, the calculation of $\varepsilon(\theta)$ amounts to analytically continuing $\tilde{\varepsilon}(q)$ to the imaginary axis and identifying the saddle points. In this context, the sign problem can be viewed as thought of as the difficulty in doing this analytical continuation numerically based entirely on exponentially sensitive numerical knowledge $\tilde{\varepsilon}(q)$ on the real axis.

Issues associated with analyticity are also at the heart of an apparent puzzle. Note that the analysis of the previous subsection and the numerical studies of the previous section both indicate that knowledge of $\tilde{\varepsilon}(q)$—even if perfect—is insufficient to fully reconstruct $\varepsilon(\theta)$ in the regime $\pi/2 < \theta < \pi$ via direct summation over topological sectors for the dilute instanton gas. Despite this, the question of whether the exact knowledge of $\tilde{\varepsilon}(q)$ is sufficient to reconstruct the full function $\varepsilon(\theta)$ for all θ's is somewhat subtle. Recall that the numerical evidence and the analysis based on the saddle point approximation indicate that for $|\theta| < \pi/2$, the knowledge of $\tilde{\varepsilon}(q)$ *is* sufficient to reconstruct $\varepsilon(\theta)$. If we are able to fully reconstruct $\varepsilon(\theta)$ for $|\theta| < \pi/2$ from $\tilde{\varepsilon}(q)$, one could in principle use the functional form obtained for $\varepsilon(\theta)$ in the regime $|\theta| < \pi/2$ to analytically continue into the region $|\theta| > \pi/2$. Moreover, for the dilute instanton gas, the

function is known to be analytic along the real axis and there is no obstruction to analytically continuing from $\varepsilon(\theta)$ into the regime $\pi/2 < |\theta| < \pi$. Thus it seems as though the information contained in $\tilde{\varepsilon}(q)$ should be sufficient to reconstruct the full function $\varepsilon(\theta)$.

How can one reconcile the inability to reconstruct $\varepsilon(\theta)$ from $\tilde{\varepsilon}(q)$ via direct summation with the fact that $\tilde{\varepsilon}(q)$ contains enough information to reconstruct it? The answer lies in the analytic structure of $\tilde{\varepsilon}(q)$ in the complex plane. Recall that the functional form of $\tilde{\varepsilon}(q)$ given in Eq. (20), while a uniquely defined function if viewed as a mapping from reals to reals, is multibranched as a function from complex q to complex $\tilde{\varepsilon}$. Indeed, the branch points of this function played in a key role in the study of the regime $\pi/2 < |\theta| < \pi$. The form we found, $\tilde{\varepsilon}(q)$, by studying it along the real axis corresponds to the principal branch of this function. However, the functional form $\tilde{\varepsilon}(q)$ "knows" about all of its branches since one can analytically continue from one branch to the next.

Consider what happens to the function $\tilde{\varepsilon}(q)$ if one follows it from a point on the real axis along a path through the complex plane around the branch point at $q = i\chi_0$, a single time and back to the same point on the real axis evolving the function along the path via the Cauchy-Riemann equations. One will not obtain the same value for the function—its value on the principal branch—but rather the value of the function along a different branch. More generally, as soon as the path crosses the branch cut, one moves onto a new branch. We will denote the functional form for the branch we obtain in this process as $\tilde{\tilde{\varepsilon}}(q)$; it is given by

$$\tilde{\tilde{\varepsilon}}(q) = \varepsilon_0 + \chi_0 + \sqrt{\chi_0^2 + q^2} + q\left(i\pi - \sinh^{-1}\left(\frac{q}{\chi_0}\right)\right),$$

(37)

where the inverse hyperbolic sine is taken in its principal branch; the shift to a new branch of the inverse hyperbolic sine is reflected by a flipped sign and the presence of the factor of $i\pi$. It is worth noting that while $\tilde{\varepsilon}(q)$ was obtained by studying the behavior along the real q axis and produced a real function, this new branch is not real for a real q. Thus, it cannot be obtained directly by taking the large V limit of $\tilde{\varepsilon}(q, V)$ with q being real (as it will be in lattice studies); it requires an analytic continuation.

It is easy to verify that if one uses this branch instead of the principal branch $\tilde{\varepsilon}(q)$ when computing the stationary phase integral for the regime $\pi/2 < \theta < \pi$, then the results are indeed self-consistent and one reconstructs $\varepsilon(\theta) = \varepsilon_0 + \chi_0(1 - \cos(\theta))$. One can similarly verify numerically that directly summing over topological sectors using $\tilde{\tilde{\varepsilon}}(q)$ at large volume also reproduces $\varepsilon(\theta) = \varepsilon_0 + \chi_0(1 - \cos(\theta))$ with high accuracy. This indicates the sense in which the functional form $\tilde{\varepsilon}(q)$ contains the information to fully

reconstruct $\varepsilon(\theta)$. It is clear that fully reconstructing it for some region of θ depends on using a particular branch of the function. To reconstruct the full function one needs to use multiple branches—matching the appropriate branch to a given region of θ. On the other hand, the only way we know how to extract these branches is via analytic continuation. The bottom line, then, is that while $\tilde{\varepsilon}(q)$ contains information to fully reconstruct this information, it is in a form that is inaccessible when doing a direct sum over topological sectors.

IV. THE GENERAL CASE

Up to this point, we have focused on a toy model—the dilute instanton gas. We have done so since it was solvable and hence it is possible to verify one's analytic conclusions against direct numerical evaluations at large volume. The real importance of the toy model is that the analysis of it can serve as a paradigm for how to study cases which are not solvable in closed form.

The key point is that the analysis given in the preceding section holds far more generally than for the dilute instanton gas. It was shown there that when there exists a saddle point along the imaginary q axis for $\tilde{\varepsilon}(q) - iq\theta$, then $\underline{\varepsilon}(\theta) = \varepsilon(\theta)$—i.e. the direct summation over topological sectors using the infinite volume extrapolation for the energy density as a function of winding number density—gives the correct infinite volume extrapolation of the energy density as a function of θ. This will hold generally:

$$\underline{\varepsilon}(\theta) = -\lim_{V\to\infty}\sum_{k=-\infty}^{\infty}\frac{1}{V}\log\left(V\int dq e^{-V(\tilde{\varepsilon}(q)-i(\theta+2\pi k)q)}\right),$$
(38)

$$\varepsilon(\theta) = -\lim_{V\to\infty}\sum_{k=-\infty}^{\infty}\frac{1}{V}\log\left(V\int dq e^{-V(\tilde{\varepsilon}(q,V)-i(\theta+2\pi k)q)}\right).$$
(39)

When the integral in Eq. (38) has a saddle point q^{sp} for a particular range of θ, according to Eq. (29), one should have $\underline{\varepsilon}(\theta(q^{sp})) = \tilde{\varepsilon}(q^{sp}) - i\theta(q^{sp})q^{sp}$ and $\theta(q^{sp}) = -i\frac{\partial \tilde{\varepsilon}(q^{sp})}{\partial q^{sp}}$. Since $\tilde{\varepsilon}(q) = \lim_{V\to\infty}\tilde{\varepsilon}(q,V)$, $\theta(q^{sp}) = -i\lim_{V\to\infty}\frac{\partial \tilde{\varepsilon}(q^{sp},V)}{\partial q^{sp}}$ is also true, which in turns means Eq. (39) also has a saddle point at the same value of q as Eq. (38) at the infinite volume limit. Then $\varepsilon(\theta(q^{sp})) = \lim_{V\to\infty}\tilde{\varepsilon}(q^{sp},V) - i\theta(q^{sp})q^{sp} = \underline{\varepsilon}(\theta(q^{sp}))$, so that $\underline{\varepsilon}(\theta) = \varepsilon(\theta)$ is true in this general case if one assumes there is a saddle point for the integral in Eq. (38).

When there is no saddle point for the integral in Eq. (38), we know the branch point will play a similar role as the saddle point, but for Eq. (39), the previous argument cannot be used here because there is no branch point.

The Legendre-like relations of Eqs. (29) will, in general, hold if a saddle point exists, regardless of the other details of the system. This means that for a general θ dependence we can use Eq. (29) to probe the conditions for which a saddle point exists and hence determine conditions for which $\tilde{\varepsilon}(q)$ can be used to reconstruct $\varepsilon(\theta)$ via direct summation over topological sectors.

The principal result is that for cases where $\varepsilon(\theta)$ is analytic for a real θ between $-\pi$ and π, one generically has

$$\underline{\varepsilon}(\theta) = \varepsilon(\theta) \quad \text{if } |\theta| < \theta_{\max},$$
(40)

where $\theta_{\max}$ is the smallest positive value of θ satisfying either of the following conditions:

(1) $\frac{d^2\varepsilon(\theta)}{d\theta^2}\big|_{\theta=\theta_{\max}} = 0$.

(2) $\theta_{\max} = \pi$.

Moreover, if condition 1 is satisfied, i.e. there is a point of inflection, in general $\underline{\varepsilon}(\theta) \neq \varepsilon(\theta)$ for some region of $|\theta| > \theta_{\max}$ (typically extending either to $\theta = \pi$ or the next point of inflection, while if condition 2 is satisfied, then $\varepsilon(\theta)$ is not analytic at $\theta = \pi$ (typically with a discontinuous first derivative as in Fig. 1). This means that whenever there is a point of inflection in $\varepsilon(\theta)$, the sign problem is so severe that even exact knowledge of $\underline{\varepsilon}(\theta)$ is insufficient to reconstruct $\varepsilon(\theta)$ direct summation over the full range of θ's.

Note that this behavior is precisely what is seen for the dilute instanton gas, where $\underline{\varepsilon}(\theta)$ equals $\varepsilon(\theta)$ before the point of inflection at $\theta = \pi/2$ and ceases to be equal beyond the point of inflection.

These conditions can be obtained by requiring self-consistency. One starts by assuming that $\underline{\varepsilon}(\theta) = \varepsilon(\theta)$ and then determines where this relation fails. We start by noting that if $\underline{\varepsilon}(\theta) = \varepsilon(\theta)$ holds for a small $|\theta|$, then we expect that the integral determining $\underline{\varepsilon}(\theta)$ is fixed by a saddle point and not a branch point. Were it a branch point, then $\varepsilon(\theta)$ would be linear at small θ but since it is even would have to be proportional to $|\theta|$. This is inconsistent with the hypothesis that $\varepsilon(\theta)$ is analytic from $-\pi$ to π. Thus, one expects $\underline{\varepsilon}(\theta) = \varepsilon(\theta)$ to continue to hold so long as the saddle point approximation remains valid and to fail when the approximation breaks down.

This breakdown occurs at a branch point beyond which the saddle point ceases to exist. The branch points are easy to identify; $\tilde{\varepsilon}(q)$ ceases to be analytic along the imaginary axis at the branch points.

A key identity allows us to identify the branch points and establish the conditions where $\underline{\varepsilon}(\theta) = \varepsilon(\theta)$. The identity is that whenever $\underline{\varepsilon}(\theta) = \varepsilon(\theta)$ and there exists a saddle point of $\tilde{\varepsilon}(q) - iq\theta$ along the imaginary i axis, then

$$\frac{\partial^2 \varepsilon(\theta)}{\partial \theta^2}\frac{\partial^2 \tilde{\varepsilon}(q)}{\partial q^2}\bigg|_{q_\theta^{sp}} = 1,$$
(41)

which can be derived directly from Eq. (29). When a saddle point exists the first line of Eq. (29) implies that $\frac{dq^{sp}}{d\theta} = i\frac{\partial^2\varepsilon(\theta)}{\partial\theta^2}$ [assuming $\underline{\varepsilon}(\theta) = \varepsilon(\theta)$]. On the other hand, $\frac{1}{\frac{dq^{sp}}{d\theta}} = \frac{d\theta}{dq^{sp}}$ is fixed by the second line of Eq. (29): $\frac{d\theta}{dq^{sp}} = -i\frac{\partial^2\tilde{\varepsilon}(q)}{\partial q^2}|_{q=q^{sp}}$. Together they yield Eq. (41).

Equation (41) implies that $\frac{\partial^2\tilde{\varepsilon}(q)}{\partial q^2}|_{q_\theta^{sp}}$ diverges as one approaches a point of inflection of $\varepsilon(\theta)$ from below. This divergence signals that $\tilde{\varepsilon}(q)$ ceases to be analytic at this point—precisely as one expects if a branch point is encountered there. Although the branch point represents nonanalytic behavior in $\tilde{\varepsilon}$ rather than $\underline{\varepsilon}$, it is clearly a point at which $\underline{\varepsilon}(\theta)$ becomes nonanalytic since the function becomes linear beyond the branch point. This necessarily spoils the equality between $\varepsilon(\theta)$ and $\underline{\varepsilon}(\theta)$ since, by hypothesis, $\varepsilon(\theta)$ is analytic between $-\pi$ and π.

One might worry that, in principle, a branch point could exist without $\frac{\partial^2\tilde{\varepsilon}(q)}{\partial q^2}|_{q_\theta^{sp}}$ diverging as the branch point is approached below. If this happened, one could have a breakdown of the condition $\underline{\varepsilon}(\theta) = \varepsilon(\theta)$ without an inflection point in $\varepsilon(\theta)$. The concern stems from the possibility that higher derivative divergences could signal the non-analyticity even if the second derivative remains finite. One would expect such behavior, if for example $\tilde{\varepsilon}(q)$ had a contribution proportional to $\chi_0^{3/2}(1 + \frac{q^2}{\chi_0^2})^{5/2}$; the second derivative remains finite as $q \to i\chi_0$ but the third derivative diverges as do all higher derivatives. However, we can rule out this possibility if $\varepsilon(\theta)$ is analytic between $-\pi$ and π.

It is easy to show from Eq. (41) that $\frac{\partial^n\tilde{\varepsilon}(q)}{\partial q^n}|_{q_\theta^{sp}}$ for $n > 2$ is given by

$$\frac{\partial^n\tilde{\varepsilon}(q)}{\partial q^n}\bigg|_{q_\theta^{sp}} = \sum_{k_2,k_3,\cdots k_n}\delta_{n,(2+k_2+k_3+\cdots k_n)}\frac{c_{k_2,\ldots,k_n}\prod_{j=2,n}(\frac{\partial^j\varepsilon(\theta)}{\partial\theta^j})^{k_j}}{(\frac{\partial^2\varepsilon(\theta)}{\partial\theta^2})^{2n-3}},$$

(42)

where $c_{k_1,k_2,\ldots,k_n}$ are calculable coefficients and the k_j non-negative integers. The structure in Eq. (42) is significant since, by hypothesis, the system is in a regime in which $\varepsilon(\theta)$ is analytic and thus $\frac{\partial^j\varepsilon(\theta)}{\partial\theta^j}$ is finite for all j's. Therefore, the only way that $\frac{\partial^n\tilde{\varepsilon}(q)}{\partial q^n}|_{q_\theta^{sp}}$ can diverge is if $\frac{\partial^2\varepsilon(\theta)}{\partial\theta^2} = 0$—i.e. the system is at a point of inflection.

The conclusion of this analysis is that one expects that when $\varepsilon(\theta)$ has a point of inflection somewhere in the domain $-\pi < \theta < \pi$, for part of the domain the sign problem is so severe that one cannot obtain $\varepsilon(\theta)$ by directly summing over topological sectors using $\tilde{\varepsilon}(q)$, the finite volume limit of $\tilde{\varepsilon}(q, V)$. We tested this conclusion numerically by constructing numerous hypothetical forms $\varepsilon(\theta)$ that contained points of inflection for which we could obtain $\tilde{\varepsilon}(q)$. Using these extracted $\tilde{\varepsilon}(q)$'s we found in all cases that there were regions of θ beyond the point of inflection for which we could not reconstruct $\varepsilon(\theta)$ by summing over topological sectors.

V. CONCLUSION

This paper explored a subtlety in the relationship between the infinite volume limit and the sign problem in the context of theories with a θ term. It was shown that there exist circumstances for which the sign problem is so severe that for some values of θ one cannot obtain the correct infinite volume $\varepsilon(\theta)$ by summing over topological sectors using the exact infinite volume form for $\tilde{\varepsilon}(q)$. This occurs when $\varepsilon(\theta)$ has a point of inflection between $-\pi$ and π. This can be taken as an illustration of just how serious sign problems can be.

However, it is also worth stressing that there are regions in θ where $\tilde{\varepsilon}(q)$ is sufficient to obtain $\varepsilon(\theta)$ by direct summation. In some ways this is quite remarkable. After all, there are power law differences in the volume between $\tilde{\varepsilon}(q, V)$ and $\tilde{\varepsilon}(q)$, which translates into order one errors in $\mathcal{Z}_Q$, while one requires the sum over Q to yield cancellations which are accurate up to exponential accuracy in V since individual terms $\mathcal{Z}_Q$ are exponentially larger than the sum $\mathcal{Z}(\theta)$. Thus, one requires some type of conspiracy for these order one errors not to spoil the cancellations. Ultimately, the reason why $\tilde{\varepsilon}(q)$ turns out to be sufficient in these cases is that the sum over topological sectors can be rewritten as a sum of integrals, which at large V are exponentially dominated by a single integral which in turn may be well approximated using the saddle point method.

One interesting mathematical fact that emerges from the analysis concerns the analytic structure of the infinite volume functions $\varepsilon(\theta)$, $\tilde{\varepsilon}(q)$: unless both functions are trivially constant, at least one of these is not analytic over the entire complex plane. Recall that θ and q have something of a conjugate relationship with each other up to a key factor of i. We found that in cases where $\varepsilon(\theta)$ had a point of inflection that $\tilde{\varepsilon}(q)$ had a place where it was not analytic. The branch point in $\tilde{\varepsilon}(q)$ was approached as the point of inflection was approached in $\varepsilon(\theta)$. Thus, the only way that $\tilde{\varepsilon}(q)$ can be analytic everywhere from $-\pi$ to π is for $\varepsilon(\theta)$ to have no inflection points in the entire region; i.e. to have positive curvature. However, if this is the case, then periodicity in θ implies discontinuity of the slope at $\theta = n\pi$, which is clearly nonanalytic.

While this paper focused on a theoretical issue, the analysis may prove of value in extracting information about $\varepsilon(\theta)$ from practical lattice calculations. Note that given the sign problem, current techniques are not sufficient to directly compute $\varepsilon(\theta)$. However, one might hope that the determination of whether or not $\varepsilon(\theta)$ is in some qualitative class of functions might be accessible via a study of $\tilde{\varepsilon}(q)$ or equivalently $\varepsilon(i\theta)$. This possibility will be explored in future work.

ACKNOWLEDGMENTS

---

[1] M. F. Atiyah and I. M. Singer, Bull. Am. Math. Soc. 69, 422 (1963).

[2] R. F. Dashen, Phys. Rev. D 3, 1879 (1971).

[3] E. Witten, Ann. Phys. (N.Y.) 128, 363 (1980).

[4] E. Witten, Nucl. Phys. B160, 57 (1979).

[5] R. Brower, S. Chandrasekharan, J. W. Negele, and U. J. Wiese, Phys. Lett. B 560, 64 (2003).

[6] K. F. Smith et al., Phys. Lett. B 234, 191 (1990).

[7] I. S. Altarev et al., Phys. Lett. B 276, 242 (1992).

[8] J. P. Jacobs, W. M. Klipstein, S. K. Lamoreaux, B. R. Heckel, and E. N. Fortson, Phys. Rev. Lett. 71, 3782 (1993).

[9] M. Dine, in Flavor Physics for the Millennium: TASI 2000, edited by J. L. Rosner (World Scientific, Singapore, 2001), p. 349.

[10] T. D. Cohen, Rev. Mod. Phys. 68, 599 (1996).

[11] D. A. Smith and M. J. Teper, Phys. Rev. D 58, 014505 (1998).

[12] M. Gockeler, M. L. Laursen, G. Schierholz, and U. J. Wiese, Commun. Math. Phys. 107, 467 (1986).

[13] M. Luscher, Commun. Math. Phys. 85, 39 (1982).

[14] P. Woit, Phys. Rev. Lett. 51, 638 (1983).

[15] E. Witten, Phys. Rev. Lett. 81, 2862 (1998).

[16] H. Leutwyler and A. V. Smilga, Phys. Rev. D 46, 5607 (1992).

[17] T. Sasaki, J. Takahashi, Y. Sakai, H. Kouno, and M. Yahiro, Phys. Rev. D 85, 056009 (2012).

[18] T. Sasaki, H. Kouno, and M. Yahiro, Phys. Rev. D 87, 056003 (2013).

[19] S. Gupta, Pramana J. Phys. 63, 1211 (2004).

[20] G. Aarts, Phys. Rev. Lett. 102, 131601 (2009).

[21] G. Aarts and K. Splittorff, J. High Energy Phys. 08 (2010) 017.

[22] J. C. Osborn, K. Splittorff, and J. J. M. Verbaarschot, Phys. Rev. D 78, 105006 (2008).

[23] J. C. Osborn, K. Splittorff, and J. J. M. Verbaarschot, in Proceedings of the 8th Workshop on Continuous Advances in QCD (CAQCD-08), Minneapolis, 2008, edited by M. Peloso (Singapore, Hackensack, NJ, 2008), p. 135.

[24] L. Ravagli and J. J. M. Verbaarschot, Phys. Rev. D 76, 054506 (2007).

[25] D. Sexty, Nucl. Phys. A931, 856 (2014).

[26] J. Bloch, Phys. Rev. Lett. 107, 132002 (2011).

[27] M. G. Alford, A. Kapustin, and F. Wilczek, Phys. Rev. D 59, 054502 (1999).

[28] P. de Forcrand, Proc. Sci., LAT2009 (2009) 010 [arXiv: 1005.0539].

[29] M. D'Elia and F. Negro, Phys. Rev. Lett. 109, 072001 (2012).

[30] B. Alles and A. Papa, Phys. Rev. D 77, 056008 (2008).

[31] B. Alles, M. Giordano, and A. Papa, Phys. Rev. B 90, 184421 (2014).

[32] R. Horsley, T. Izubuchi, Y. Nakamura, D. Pleiter, P. E. L. Rakow, G. Schierholz, and J. Zanotti, arXiv:0808.1428.

[33] H. Panagopoulos and E. Vicari, J. High Energy Phys. 11 (2011) 119.

[34] V. Azcoiti, G. Di Carlo, A. Galante, and V. Laliena, Phys. Rev. Lett. 89, 141601 (2002).

[35] A. I. Vainshtein, V. I. Zakharov, V. A. Novikov, and M. A. Shifman, Usp. Fiz. Nauk 136, 553 (1982) [Sov. Phys. Usp. 25, 195 (1982)].

[36] S. Coleman, Aspects of Symmetry: Selected Erice Lectures (Cambridge University Press, Cambridge, England, 1988).

[37] F. W. J. Olver, Phil. Trans. R. Soc. A 247, 328 (1954).

[38] G. B. Arfken, H. J. Weber, and F. E. Harris, Mathematical Methods for Physicists: A Comprehensive Guide (Academic Press, New York, 2011).