Complex paths around the sign problem

Andrei Alexandru[*]
Department of Physics, The George Washington University, Washington, D.C. 20052, USA

Gökçe Başar[†]
Department of Physics, University of North Carolina, Chapel Hill, North Carolina 27599, USA

Paulo F. Bedaque[‡]
Department of Physics, University of Maryland, College Park, Maryland 20742, USA

Neill C. Warrington[§]
Institute for Nuclear Theory, University of Washington, Seattle, Washington 98195, USA

(published 9 March 2022)

The Monte Carlo evaluation of path integrals is one of a few general purpose methods to approach strongly coupled systems. It is used in all branches of physics, from QCD and nuclear physics to the correlated electron systems. However, many systems of great importance (dense matter inside neutron stars, the repulsive Hubbard model away from half filling, and dynamical and nonequilibrium observables) are not amenable to the Monte Carlo method as it currently stands due to the so-called sign problem. A new set of ideas recently developed to tackle the sign problem based on the complexification of field space and the Picard-Lefshetz theory accompanying it is reviewed. The mathematical ideas underpinning this approach, as well as the algorithms developed thus far, are described together with nontrivial examples where the method has already been proved successful. Directions of future work, including the burgeoning use of machine learning techniques, are delineated.

DOI: 10.1103/RevModPhys.94.015006

CONTENTS

[*]aalexan@gwu.edu
[†]basar@unc.edu
[‡]bedaque@umd.edu
[§]ncwarrin@umd.edu

I. THE SIGN PROBLEM

Monte Carlo methods have been used with great success to study problems ranging from classical systems of particles to studies of hadrons using lattice quantum chromodynamics. The usual setup is to formulate the problem (classical or quantum) in a way analogous to a classical statistical system. Observables are then given by multidimensional integrals involving a Boltzmann factor that is computed numerically by importance sampling. There are, however, important systems

that cannot yet be solved using standard Monte Carlo methods. These are the systems either where the statistical weights become complex or where the signs oscillate. Roughly speaking, we say that the system suffers from a *sign problem* when the phase fluctuations increase as the size of the system is increased. These fluctuations lead to delicate cancellations that preclude a stochastic evaluation of the integral. This occurs in the study of neutron matter found in neutron stars, the repulsive Hubbard model away from half filling, and all field theoretical–many-body observables in real time. Solving the sign problem is of central importance in many fields of physics, and a number of approaches have been proposed to either solve or alleviate this problem. Some are more generic and some are problem specific, but in spite of this progress important questions about the physics of the previously mentioned system remain unanswered.

In this review we focus on a novel set of related methods relying on the analytical properties of the configuration weights. The fundamental idea is to express the partition sum as an integral over real degrees of freedom and to complexify each variable. The partition sum is originally an integral over the real manifold in this enlarged configuration space; however, as we later discuss, we can deform the multidimensional integration contour (without changing the value of the partition function) to a manifold that has better numerical properties. In particular, the phase fluctuations are either eliminated or significantly reduced. We describe in the review the geometry of the complex field space, its critical points, and the algorithms used to both find suitable manifolds and integrate over them. All of these steps are exemplified in simple field theories, usually in a lower number of dimensions, that contain, however, all the properties of the theories of physical interest.

A. Field theory–many-body physics as a path integral

The expectation value of any observable $\mathcal{O}$ in field theory can be calculated using the path integral[1]

$$\langle O\rangle = \frac{1}{Z}\int D\phi e^{-S_E(\phi)}O(\phi), \qquad Z = \int D\phi e^{-S_E(\phi)}. \quad (1.1)$$

In Eq. (1.1) ϕ is the generic name of the fields in the theory and S_E is the Euclidean (imaginary-time) action evaluated over a Euclidean "time" β that is equal to the inverse temperature of the system.[2] The path integral in Eq. (1.1) is an integral over an infinite-dimensional space. To evaluate it numerically (and to properly define it), we consider a discretized version where spacetime is replaced by a finite lattice. After discretization, the path integral becomes a finite-dimensional integral, albeit one over a large number of dimensions, proportional to the number of spacetime points composing the lattice. This is equivalent to a classical statistical mechanics problem in four spatial dimensions, where the state of the system is described by the field ϕ defined on the entire four-dimensional grid, and the probability of each state is controlled by the Boltzmann factor $\exp[-S_E(\phi)]$. Using Monte Carlo methods, a set of n configurations $\{\phi^{(1)},\ldots,\phi^{(\mathsf{n})}\}$ is generated with the probability distribution $\exp[-S_E(\phi)]/Z$. The observables and their errors are then estimated using

$$\langle O\rangle = \frac{1}{\mathsf{n}}\sum_a O(\phi^{(a)}), \qquad \epsilon_O = \sqrt{\frac{1}{\mathsf{n}(\mathsf{n}-1)}\sum_a [O(\phi^{(a)})-\langle O\rangle]^2}. \quad (1.2)$$

Numerous algorithms have been developed to obtain configurations $\phi^{(a)}$ distributed according to $e^{-S_E[\phi]}$ in an efficient way. The cost of the sampling process increases with a moderate power of the spacetime volume V (between 1 and 3), despite the fact that the Hilbert space dimension of the corresponding quantum system grows exponentially with the space volume. This is the advantage of Monte Carlo methods over direct diagonalization procedures.

B. Physical systems with sign problems

Many theories of interest in theoretical physics have sign problems in all currently known formulations. In fact, systems that cannot be fully understood because a sign problem hinders the use of Monte Carlo simulations are pervasive in all subfields of physics (and chemistry). Among those some have become "holy grails" in their respective field, problems whose solutions would have a revolutionary impact.

For instance, in nuclear physics QCD at finite baryon density has a sign problem. This problem prevents the understanding from first principles of both neutron stars and supernovae. Extensive work has been expended on evading this sign problem; see the reviews by Karsch (2000), Muroya *et al.* (2003), Philipsen (2007), de Forcrand (2010), and Aarts (2016) and references therein. Quantum Monte Carlo (QMC) studies of nuclei using "realistic nucleon-nucleon interactions" also suffer from the sign problem (Koonin, Dean, and Langanke, 1997; Wiringa *et al.*, 2000; Carlson *et al.*, 2015; Lähde *et al.*, 2015). The "constrained path algorithm" (Zhang, Carlson, and Gubernatis, 1995, 1997) is a widely used approximate method to address these sign problems.[3] Lattice field theory studies of nuclei have similar behavior; sign problems appear in studies of nuclei with different proton and neutron numbers, and when repulsive forces become sufficiently large (Lee, Borasoy, and Schaefer, 2004; Lee, 2009; Epelbaum *et al.*, 2014; Elhatisari *et al.*, 2017). Furthermore, lattice and QMC studies of nuclear matter encountered in astrophysics suffer from the sign problem. This includes spin polarized neutron matter (Fantoni, Sarsa,

[1]Similar expressions are obtained for the partition function $Z = \mathrm{tr}e^{-\beta H}$ of nonrelativistic quantum systems by discretizing both space and time, then using the Trotter formula.

[2]There is no assumption that the theory is relativistic. In fact, nonrelativistic systems in the second quantized form are frequently studied within this formalism.

[3]The constrained path algorithm is a generalization of the "fixed-node approximation," a similar approximate technique for avoiding the sign problem (Anderson, 1975).

and Schmidt, 2001; Gezerlis, 2011; Gandolfi *et al.*, 2014)[4] and lattice effective field theory (EFT) studies of nuclear matter beyond leading order (Lu *et al.*, 2019a).[5]

Many cold atom systems, when formulated with lattice or QMC methods, exhibit sign problems as well. Both spin and mass imbalanced spin $1/2$ fermions have a sign problem (Braun *et al.*, 2013; Roscher, Braun, and Drut, 2014). This sign problem makes it prohibitively difficult to conclusively demonstrate the existence of a number of conjectured phases [like the Larkin-Ovchinnikov-Fulde-Ferrell phases (Fulde and Ferrell, 1964; Larkin and Ovchinnikov, 1964] in more than $1+1$ dimensions. Bosonic nonrelativistic systems exhibit sign problems as well; this includes bosons under rotation (Berger, Morrell, and Drut, 2020) and those coupled to spin-orbit interactions (Attanasio and Drut, 2020). For a review see Berger *et al.* (2019).

A wide variety of lattice-supersymmetric models suffer from a sign problem too; for a review see Schaich (2019). In particular, first-principles tests of the gauge-gravity duality conjecture, even in the simplest case of reproducing supergravity black hole thermodynamics from D0-brane quantum mechanics, can claim to be bona fide controlled tests of the duality only if the phase fluctuations are under control (Hanada *et al.*, 2011;Berkowitz *et al.*, 2016). Sign problems are also faced in astrophysics (Morinaga, 2021), radio astronomy (Feldbrugge, Pen, and Turok, 2019; Jow *et al.*, 2021), and quantum cosmology (Di Tucci and Lehners, 2019; Di Tucci *et al.*, 2019; Han *et al.*, 2021; Matsui, 2021).

Sign problems are found in condensed matter physics as well. A particularly well-known example is the Hubbard model away from half filling (Hubbard, 1963; White *et al.*, 1989; Loh *et al.*, 1990), which is thought to model essential characteristics of high T_c superconductors. Path integral formulations of fullerene exhibit the sign problem as well (Ostmeyer *et al.*, 2020). Furthermore, some models of frustrated magnetism on triangular and *kagome* lattices, of interest for their conjectured spin-liquid ground states, exhibit the sign problem (Sindzingre, Lecheminant, and Lhuillier, 1994; Lacroix, Mendels, and Mila, 2011; Mishchenko, Kato, and Motome, 2021). As a result, there is uncertainty in the zero-temperature properties of these models.

C. Reweighting and the sign problem

The standard workaround for sampling complex actions is to use *reweighting*. The idea is to split the integrand into a positive part that is used for Monte Carlo sampling, usually the absolute value of the integrand, and a fluctuating part that is included in observables. Using the absolute value as a sampling weight, we have the following identity:

$$\langle O\rangle = \frac{\langle O e^{-i\mathrm{Im}S_E(\phi)}\rangle_0}{\langle e^{-i\mathrm{Im}S_E(\phi)}\rangle_0}, \quad \langle O\rangle_0 = \int D\phi \frac{e^{-\mathrm{Re}S_E(\phi)}}{Z_0} O(\phi), \tag{1.3}$$

and $Z_0 \equiv \int D\phi e^{-\mathrm{Re}S_E(\phi)}$. The idea, then, is to use the *phase quenched* action $\mathrm{Re}S_E$ to sample configurations and take into account the imaginary part of the action when computing observables. From a numerical point of view, this procedure works when the phase fluctuations are mild and we can estimate the phase average $\langle e^{-i\mathrm{Im}S_E(\phi)}\rangle_0$ with enough accuracy; this means that the error estimate for this average should be significantly smaller than its mean. Since the magnitude of $e^{-i\mathrm{Im}S_E(\phi)}$ for each configuration is 1, to resolve the mean accurately we require a number of configurations $\mathsf{n} \gg 1/\langle e^{-i\mathrm{Im}S_E(\phi)}\rangle_0^2$. When the average phase is small, reweighting requires a large number of samples and becomes impractical. For many systems at finite density, the phase average goes to zero exponentially fast in the spatial volume or inverse temperature. This occurs because the phase average is the ratio of the following two partition functions:

$$\langle e^{-i\mathrm{Im}S_E(\phi)}\rangle_0 = \frac{Z}{Z_0} = \frac{e^{-\beta f V}}{e^{-\beta f_0 V}} = e^{-\beta V \Delta f}, \tag{1.4}$$

where $\Delta f = f - f_0 > 0$ is the difference in the free energy density between the original system and the phase quenched system. In this case, the numerical effort grows exponentially as we increase the volume and/or lower the temperature. This is what is usually defined as the sign problem. An even worse problem arises when calculating real-time correlation functions. In that case, we are interested in integrals of the form

$$\langle \mathcal{O}\rangle = \frac{1}{Z}\int D\phi\, e^{iS(\phi)}\mathcal{O}, \tag{1.5}$$

where S is the real-time Minkowski space action of the system.[6] Since there is no damping of the magnitude of the integrand and the value of the field $\phi(t,\mathbf{x})$ (for any t and $\mathbf{x}$) grows, the average phase is strictly zero, even for small sized systems. A similar argument applies to observables, like parton distribution functions, that are defined on the light cone [a detailed account of this problem and a proposal to use quantum computers to solve it was published recently (Lamm, Lawrence, and Yamauchi, 2020)].

Note that the existence of a sign problem does not necessarily preclude numerical study. There are cases where the sign problem is mild enough that most relevant information about the system in the region of interest can be extracted before the sign fluctuations become an obstacle. For example, when studying the phase diagram of a simple heavy-dense quark model for QCD (discussed later), the end point of the first order phase transition can be studied via reweighting for system sizes as large as 100^3 even though the model has a sign problem (Alford *et al.*, 2001). We mention this study to point

[4]Unpolarized neutron matter, however, can be formulated free of the sign problem (Chen and Kaplan, 2004; Lee and Schäfer, 2005).

[5]Wigner SU(4) symmetric approximations to pionless EFT have no phase oscillations and have been profitably used (Wigner, 1937; Lee, 2007; Lu *et al.*, 2019b). Similarly, gauge theories with pseudoreal representations, like SU(2) and G_2, also have no sign problem at finite density.

[6]In thermal equilibrium at nonzero temperature, real-time correllators can be computed from path integrals defined in the closed-time contour in complex time (Schwinger, 1961; Keldysh, 1964); see Sec. III.D.1.

out that, from a practical point of view, methods that merely reduce sign fluctuations, without completely eliminating them, are also important.

D. The absence of a general solution

It is of theoretical, if not practical, interest to know whether a generic solution to the sign problem exists. If one takes an exponentially vanishing average sign in the system size as the definition of the sign problem, then there are definitely models in which the sign problem can be solved. For instance, for many systems it is possible to rewrite the path integral using a different set of states and obtain an expression free of phase fluctuations. This was accomplished for the two-component scalar theory using dual variables (Endres, 2007; Gattringer and Kloiber, 2013), and by reorganizing the summation over configurations for the previously mentioned heavy-dense system (Alford *et al.*, 2001; Alexandru, Bergner *et al.*, 2018). Similarly, there is a class of fermionic models that, when formulated in terms of fermion bags (Chandrasekharan and Wiese, 1999; Alford *et al.*, 2001; Chandrasekharan, 2012, 2013; Huffman and Chandrasekharan, 2014, 2016, 2020; Hann, Huffman, and Chandrasekharan, 2017; Ayyar, Chandrasekharan, and Rantaharju, 2018), have strictly positive Boltzmann weights even though other formulations have a severe sign problem. For a recent review of sign problem–free methods in condensed matter physics, see Li and Yao (2019). As it turns out, however, a solution of this kind is unlikely to work for all systems.

There is an often-cited, general argument implying that a generic solution to the sign problem, applicable to all systems, is extremely unlikely to exist. It relies on the NP $\neq$ P conjecture from computational theory. NP decision problems are problems that can be solved on a *nondeterministic* Turing machine in a time that increases only polynomially with the system size, whereas P problems are the ones that can be solved in polynomial time in a *deterministic* way. While no proof exists, it is widely believed that there are NP problems that are not P. In connection to this question, an important subset of NP problems are the NP-hard or NP-complete problems. If any of these NP-hard problems can be solved in polynomial time on a classical computer, then all NP problems can, thus invalidating the conjecture. There are spin glass–like systems with a sign problem that can be mapped onto NP-hard problems (Troyer and Wiese, 2005). Using the previous chain of arguments, a generic solution to the sign problem that would solve this problem would imply NP = P, which is considered highly unlikely.

E. A survey of methods to deal with sign problems

As previously mentioned, some of the most physically interesting models in particle, nuclear, and condensed matter physics have sign problems. Given the interest in these problems, it is not surprising that a variety of approaches have been tried to either solve or circumvent the sign problem. In this review we focus on Lefschetz thimble inspired methods, but we now point out some approaches attempted through the years to understand the phase diagram of QCD and other relativistic theories.

A first set of methods uses simulations in the parameter region where the action is real; the result is then extrapolated to the region of interest. One version of this idea is to rely on results from imaginary chemical potentials. Monte Carlo simulations can be used directly either to infer features of the phase diagram for real chemical potentials or to compute observables and fit them using a polynomial ansatz or a Padé approximation and then analytically continue these functions to real values of μ (de Forcrand and Philipsen, 2002, 2003; D'Elia and Lombardo, 2003, 2004; Cea, Cosmai, and Papa, 2014; Bellwied *et al.*, 2015; Bonati *et al.*, 2015; Borsanyi *et al.*, 2020). Another approach is to compute the derivatives of thermodynamic observables with respect to μ at $\mu = 0$, then use Taylor expansions to extend these results to $\mu > 0$ (de Forcrand *et al.*, 2000; Miyamura, 2002; Endrodi *et al.*, 2011; Kaczmarek *et al.*, 2011; Bonati *et al.*, 2018; Bazavov *et al.*, 2019). Yet another method is to use multiparameter reweighting by combining simulations from different temperatures at $\mu = 0$ to determine the phase transition line and critical point in QCD (Fodor and Katz, 2002).

Another class of methods attempts to alleviate the sign problem using a rewriting of the path integral in terms of new variables. One possibility is to reorganize the sum over the configurations in subsets that have either only positive sign contributions to the partition function, thus solving the sign problem, or a much reduced sign problem (Rossi and Wolff, 1984; Karsch and Mutter, 1989; Chandrasekharan and Wiese, 1999; Alford *et al.*, 2001; Bloch, Bruckmann, and Wettig, 2013; Alexandru, Bergner *et al.*, 2018). Another direction is to reformulate the problem in terms of dual variables in which the sign problem is absent (Endres, 2007; Gattringer and Kloiber, 2013). It turns that for QCD the use of the canonical ensemble partition function (as opposed to the grand canonical ensemble) makes the sign fluctuations milder, and it can be used to investigate small enough systems (Barbour, Davies, and Sabeur, 1988; Hasenfratz and Toussaint, 1992; Alexandru *et al.*, 2005; Kratochvila and de Forcrand, 2005; de Forcrand and Kratochvila, 2006; Li *et al.*, 2010; Alexandru and Wenger, 2011; Li, Alexandru, and Liu, 2011; Nakamura, 2016). Finally, Fermi bags are enough to completely eliminate the sign problem in some low-dimensional models (Chandrasekharan and Wiese, 1999; Alford *et al.*, 2001; Chandrasekharan, 2012, 2013; Huffman and Chandrasekharan, 2014, 2016, 2020; Hann, Huffman, and Chandrasekharan, 2017; Ayyar, Chandrasekharan, and Rantaharju, 2018). These methods are model dependent and require insight to be applied in each new class of models.

Recently a proposal based on the density of states method was explored as a way to alleviate sign fluctuations (Fodor, Katz, and Schmidt, 2007; Langfeld and Lucini, 2014; Gattringer and Török, 2015; Garron and Langfeld, 2016, 2017).

Finally, there is a significant effort to simulate QCD at finite density using the complex Langevin approach (Klauder, 1983; Parisi, 1983),[7] which is based on the idea of stochastic quantization (Parisi and Wu, 1981). This method shares with the thimble methods its starting point: the configuration space

[7]See Berger *et al.* (2019) for a review of the complex Langevin approach.

of N real degrees of freedom is extended to an N-dimensional complex one. The important difference is that the complex Langevin approach sets up a stochastic process that moves freely in this enlarged space of $2N$ real degrees of freedom, whereas the methods that we discuss in this review sample an N-dimensional manifold. Results show that, while instabilities are present in complex Langevin QCD simulations, for heavy quark masses credible results can be obtained for temperatures above the deconfinement transition. In the hadronic phase, the simulations become unstable and unreliable (Aarts and Stamatescu, 2008; Aarts, 2009; Aarts, Seiler, and Stamatescu, 2010; Aarts *et al.*, 2011, 2013; Seiler, Sexty, and Stamatescu, 2013; Sexty, 2014, 2019; Fodor *et al.*, 2015; Scherzer, Sexty, and Stamatescu, 2020). This approach has also been applied to condensed matter cold atom systems (Berger *et al.*, 2019).

II. CAUCHY THEOREM, HOMOLOGY CLASSES, AND HOLOMORPHIC FLOW

A. Deformation of the domain of integration: A multidimensional Cauchy theorem

The well-known Cauchy theorem for functions of one complex variable states that for an analytic function $f(z)$ the integral over a closed loop vanishes as follows:

$$\oint_C f(z) = 0. \tag{2.1}$$

Equation (2.1) can be used to "deform" the contour of integration from, say, the real line to a different contour on the complex plane, as long as the initial and final points of the contours coincide. In many applications the contour starts and/or ends at a point at infinity and the issue becomes whether moving these ending points may cross a singularity of $f(z)$ at infinity. For instance, take the integral

$$\int d\phi e^{-\phi^4+\phi} \tag{2.2}$$

over different contours on the complex plane starting or ending at different points at infinity. Since there are no singularities at any finite values of z, Cauchy's theorem allows us to deform the contour of integration as long as no singularity "at infinity" is crossed. The integral in Eq. (2.2) is well defined (it converges) if and only if the initial and final asymptotic directions of the contour are in the regions $A, \ldots, D$ shown in Fig. 1. The integral over two different contours whose ends lie on the same regions have, on account of Cauchy's theorem, the same value. For instance, the real line, contour 1, is equivalent to contour 2 since both start in region A and end in region B. The integral over contour 3 is not even well defined as it diverges, while the value for the integral over contour 4 is different from the value on contour 1 or 2. In fact, imagine starting from the real line and continuously deforming it toward contour 4. At some point the integral will cease to be well defined as its end point leaves region B and the integral becomes divergent. As the end point enters region C the integral becomes finite again but acquires a different value than on the real line.

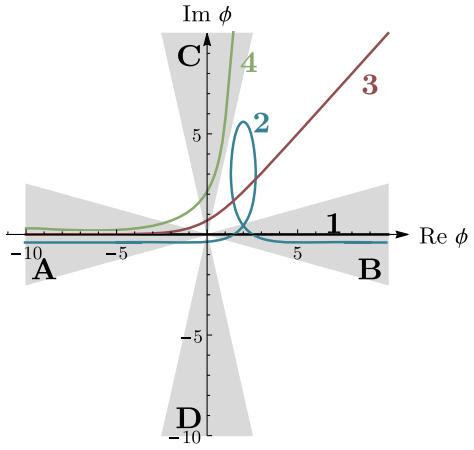


FIG. 1. Several contours of integration for the integral in Eq. (2.2). Contours 1 (the real line) and 2 produce the same result. Contour 4 yields a different result, while the integral over contour 3 is divergent. The gray areas show directions in the complex plane ("good" regions) where the integrand vanishes fast enough that the integral converges.

In fact, there are only three independent classes of contours (known as "homology classes") on which the integral in Eq. (2.2) may be evaluated: those that start in region A and end in region B, C, or D, denoted $A \to B$, $A \to C$, and $A \to D$, respectively. Any other contour with different asymptotic behavior, for instance, $B \to C$, can be expressed as a linear combination of contours (with integer coefficients) belonging to one of these three classes. Cauchy's theorem guarantees that any contour that lies in one of these classes can be smoothly deformed to some other contour in the same class without changing the value of the integral. In contrast, as previously explained, it cannot be deformed to a contour that lies in a different class. In short, all possible domains over which the integral [Eq. (2.2)] is well defined can be classified as a linear combinations of three discrete classes of contours. Each class contains a continuous family of "equivalent" contours that can be smoothly deformed to one another without changing the value of the integral. As we later see, the reason that there are three classes is that the function ϕ^4 in the exponent is a quartic polynomial that in general has three saddle points.

All the previous observations generalize to higher dimensions. Instead of integrals over one-dimensional paths we consider integrals over N cycles, orientable manifolds with no boundary with real dimension N immersed in the $2N$-dimensional space. The integral over a cycle $\mathcal{M}$ is defined by[8]

$$\int_{\mathcal{M}} f(\phi) d\phi_1 \wedge \cdots \wedge d\phi_N = \int_{\mathbb{M}} f(\phi(\zeta)) \det J(\zeta) d\zeta_1 \cdots d\zeta_N, \tag{2.3}$$

[8]Readers not familiar with the formalism of differential forms may take the right-hand side of Eq. (2.3) as the definition of an integral over N-dimensional manifolds embedded in $\mathbb{C}^N$. We use this definition extensively in this review.

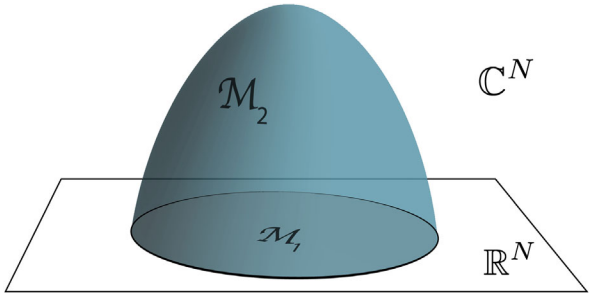


FIG. 2. Schematic of a multidimensional deformation. The original domain of integration ($\mathcal{M}_1 \subset \mathbb{R}^N$) is deformed to $\mathcal{M}_2 \subset \mathbb{C}^N$. This deformation sweeps out a manifold $\mathcal{B} \subset \mathbb{C}^N$, whose boundary is $\partial\mathcal{B} = -\mathcal{M}_1 \cup \mathcal{M}_2$.

where $\phi_i = \Phi_i(\zeta_1, \ldots, \zeta_N)$ is a parametrization of the N-dimensional manifold $\mathcal{M}$ by N real coordinates $\zeta_1, \ldots, \zeta_N$, $\mathbb{M}$ is the region of $\mathbb{R}^N$ used to parametrize $\mathcal{M}$, and $\det J(\zeta) = \partial(\phi_1, \ldots, \phi_N)/\partial(\zeta_1, \ldots, \zeta_N)$ is the determinant of the Jacobian of the parametrization, which is in general a complex number. ϕ stands for all $\phi_1, \ldots, \phi_N$ (the approach is similar for ζ).

Assume that we have two such cycles $\mathcal{M}_1$ and $\mathcal{M}_2$ that can be smoothly deformed into one another. The space swept by the deformation will be denoted with $\mathcal{B}$ and the two cycles form the boundary $\partial\mathcal{B} = \mathcal{M}_1 - \mathcal{M}_2$, where the minus sign indicates opposite orientation; see Fig. 2. By Stokes's theorem we have

$$\int_{\partial\mathcal{B}} f(\phi) d\phi_1 \wedge \cdots \wedge d\phi_N = \int_{\mathcal{B}} df(\phi) \wedge d\phi_1 \wedge \cdots \wedge d\phi_N, \tag{2.4}$$

where $df = (\partial f/\partial\phi_i) d\phi_i + (\partial f/\partial\bar{\phi}_i) d\bar{\phi}_i$ ($\bar{\phi}$ is the complex conjugate of ϕ). Since $f(\phi)$ is assumed to be holomorphic, we have $\partial f/\partial\bar{\phi}_i = 0$. In the sum $(\partial f/\partial\phi_1) d\phi_1 + \cdots + (\partial f/\partial\phi_N) d\phi_N$ every term is proportional to one of the terms in $d\phi_1 \wedge \cdots \wedge d\phi_N$, so $df \wedge d\phi_1 \wedge \cdots \wedge d\phi_N = 0$ since $d\phi_i \wedge d\phi_i = 0$. We arrive then at

$$\begin{aligned} &\int_{\partial\mathcal{B}} f(\phi) d\phi_1 \wedge \cdots \wedge d\phi_N \\ &\quad = \int_{\mathbb{M}_1 - \mathbb{M}_2} f(\Phi(\zeta)) \det J(\zeta) d\zeta_1 \cdots d\zeta_N = 0, \end{aligned} \tag{2.5}$$

which is the generalization of the Cauchy theorem that we are interested in.[9] This theorem can be used to deform the manifold of integration without altering the value of the integral just as we previously discussed for the one-dimensional case. In fact, our discussion of contour deformation readily generalizes to the multidimensional case. For manifolds approaching infinity along certain directions (in reality, N-dimensional planes), the integral is convergent and well defined ("good regions"); for others it is not. Furthermore, it can be shown, assuming that the integrand is well behaved in the sense discussed later, that the manifolds for which the integral converges are separated in discrete equivalence classes: those with the same asymptotic properties lead to the same integral. A continuous deformation of manifolds of integration from one equivalent class to another, that is, from one good region to another, necessarily goes through manifolds where the integral diverges. Such deformations are the analog of deformations crossing a "singularity at infinity" in the one-dimensional case. All this is in close analogy with the familiar one-dimensional case. A detailed discussion of the mathematical details was given by Pham (1983).

B. Holomorphic gradient flow

We are interested in deforming integrals from $\mathbb{R}^N$ (the real cycle) to some other N cycle without altering the value of the integral but while alleviating the sign problem in integrals of interest in field theory, which are typically of the form

$$\int_{\mathbb{R}^N} e^{-S(\phi)} \mathcal{O}(\phi) \prod_i d\phi_i, \tag{2.6}$$

where S is the action of the theory and $\mathcal{O}$ is some observable. One way of performing this deformation is with the help of the holomorphic flow. The holomorphic flow is defined for every action S by the the following differential equations:

$$\frac{d\phi_i}{dt} = \overline{\frac{\partial S}{\partial \phi_i}}. \tag{2.7}$$

For every point ϕ in $\mathbb{R}^N$ and a fixed flow time T, the solution of Eq. (2.7) with the initial condition $\phi(t = 0) = \zeta$ defines a point $\tilde{\phi} = \mathcal{F}_T(\zeta)$ in $\mathbb{C}^N$. By flowing all points of $\mathbb{R}^N$ in this manner, we obtain the flowed manifold $\mathcal{M}_T = \mathcal{F}_T(\mathbb{R}^N)$.[10]

The holomorphic flow has the following two important properties:

$$\frac{d}{dt} S_R = \frac{1}{2}\left[\frac{dS}{dt} + \overline{\frac{dS}{dt}}\right] = \frac{\partial S}{\partial \phi_i}\left(\overline{\frac{\partial S}{\partial \phi_i}}\right) \geq 0, \tag{2.8}$$

$$\begin{aligned} \frac{d}{dt} S_I &= \frac{1}{2i}\left[\frac{dS}{dt} - \overline{\frac{dS}{dt}}\right] \\ &= \frac{1}{2i}\left[\frac{\partial S}{\partial \phi_i}\left(\overline{\frac{\partial S}{\partial \phi_i}}\right) - \left(\overline{\frac{\partial S}{\partial \phi_i}}\right)\frac{\partial S}{\partial \phi_i}\right] = 0. \end{aligned} \tag{2.9}$$

That is, the imaginary part S_I is constant along the flow, while the real part of the action S_R increases monotonically [which is why Eq. (2.7) is also called *upward* flow].[11] The fact that S_R increases along the flow means that the integrand vanishes along asymptotic directions even faster in the flowed manifold $\mathcal{M}_T$ than in $\mathbb{R}^N$, leading to the convergence of the integral at all T. By the previously exposed arguments, this means that $\mathcal{M}_T$ is equivalent to $\mathbb{R}^N$ for the purpose of computing the

[9] We thank Scott Lawrence for a discussion on this point.

[10] Other flows to generate manifolds were proposed by Tanizaki, Nishimura, and Verbaarschot (2017).

[11] This can also be seen by noting that the holomorphic flow is the gradient flow of S_R and the Hamiltonian flow for the Hamiltonian S_I.

integral; that is, it is in the same homology class as $\mathbb{R}^N$, as in the one-dimensional example explained at the beginning of this section.

C. Lefschetz thimbles and Picard-Lefschetz theory

Even though $\mathcal{M}_T$ is equivalent to $\mathbb{R}^N$, evaluating the path integral on $\mathcal{M}_T$ rather than $\mathbb{R}^N$ is computationally advantageous in controlling the sign problem. Before we explain why this occurs, we introduce the necessary mathematical background (for a different perspective, see Appendix B).

We begin by focusing on the stationary points of the flow, namely, the critical points of the action ϕ^c where $\partial S/\partial\phi_i|_{\phi^c} = 0$. The *Lefschetz thimble* $\mathcal{T}$ attached to a critical point ϕ_c is defined as the set of initial conditions $\phi(0) \in \mathbb{C}^N$ for which the *downward* flow

$$\frac{d\phi_i}{dt} = -\overline{\frac{\partial S}{\partial \phi_i}} \tag{2.10}$$

asymptotically approaches the critical point. Similarly, the *dual thimble* $\mathcal{K}$ is the set of all points for which the upward flow asymptotes to ϕ_c. For a constructive definition of $\mathcal{T}$, we begin by linearizing the flow around ϕ^c as follows:

$$\frac{d\phi_i}{dt} = \overline{\frac{\partial^2 S}{\partial\phi_i\partial\phi_j}}\bigg|_{\phi=\phi^c} (\bar{\phi}_j - \bar{\phi}^c_j), \tag{2.11}$$

whose solution can be written as

$$\phi(t) - \phi^c = \sum_{a=1}^{N} c_a \rho^{(a)} e^{\lambda_a t}, \tag{2.12}$$

where c_a are real and $\rho^{(a)}$ are the solutions to the modified eigenvector problem [Takagi vectors (Takagi, 1924)]

$$H_{ij}(\phi_c)\rho^{(a)}_j = \lambda_a \bar{\rho}^{(a)}_i, \tag{2.13}$$

with $H_{ij} \equiv \partial^2 S/\partial\phi_i\partial\phi_j$. The modified eigenvalues λ_a can be chosen to be real, and the eigenvalues and eigenvectors then come in pairs $(\lambda_a, \rho^{(a)}), (-\lambda_a, i\rho^{(a)})$. The set of N vectors $\rho^{(a)}$ that define the directions around a critical point where the flow moves away from the critical point forms a basis (with real coefficients) for the tangent space of $\mathcal{T}$ at ϕ^c. Likewise, the set of N vectors $i\rho^{(a)}$ that define the directions around a critical point where the flow moves toward the critical point forms a basis for the tangent space of $\mathcal{K}$ at ϕ^c. These two tangent spaces together span the tangent space at ϕ^c in $\mathbb{C}^N$. With this knowledge, in the infinitesimal neighborhood of the critical point, we can solve for the *vanishing cycle* $v(\epsilon)$ as $S(\phi) - S(\phi^c) \approx z_i H_{ij} z_j = \epsilon$, which is an $(N-1)$-dimensional surface in the tangent space of $\mathcal{T}$. The thimble can be constructed by taking the vanishing cycle as the initial condition and flowing upward: $\mathcal{T} = \cup_{0\le T<\infty} \mathcal{F}_T(v(\epsilon))$ when $\epsilon \to 0$. In other words, we can build the thimble slice by slice by using the flow. We can further use the fact that the flow defines a one-to-one map between the initial point and the flowed point and instead can consider an infinitesimally small N-dimensional ball $\mathcal{B}$ in the tangent plane. $\mathcal{B}$ is already a small portion of the thimble near ϕ^c. If we take $\mathcal{B}$ as the initial condition, its image under upward flow with $T \to \infty$ is the thimble: $\mathcal{T} = \mathcal{F}_{T\to\infty}(\mathcal{B})$. This is the main idea behind the "contraction algorithm," which is a method to simulate path integrals on a given thimble; see Sec. III.A.1.

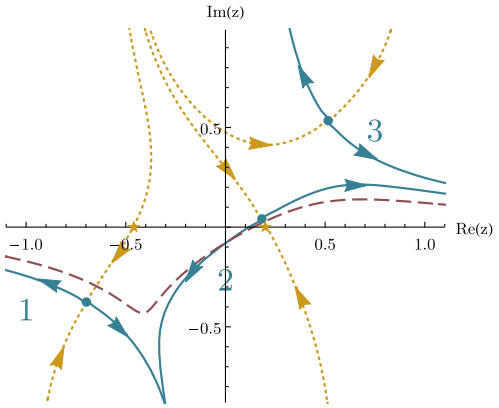


FIG. 3. Thimbles (solid blue lines), dual thimbles (dotted yellow lines), critical points (blue dots), their preimage under the flow (orange stars), and the flowed real line (dashed red line) for $G = 1.1e^{i0.05}$, $p = 1$, $\mu = 0.3$, and $m = i0.1$ for the action in Eq. (2.14). The arrows indicate the direction of the upward flow.

For a concrete illustration of these ideas, see Fig. 3, where the action is taken to be

$$S(\phi) = \phi^2/G - \log\left[(p^2 + i\mu)^2 + (\phi + m)^2\right]. \tag{2.14}$$

S can be thought of as a toy model for the action of a fermionic model coupled to an auxiliary field ϕ after the fermions have been integrated out. Notice that e^{-S} is a holomorphic function even though S is not; this is a feature common to theories with fermions. This theory has three critical points, attached to which are thimbles and dual thimbles. Only thimbles 1 and 2 contribute to the integral. The real line, evolved by the holomorphic flow by a time $T = 1.0$, is shown as the dashed red line. Notice how it approximates the union of the two contributing thimbles.

In Sec. II.A we stated that the domain of integration of an integral of the form of Eq. (2.3) is naturally identified by a set of equivalence classes of N cycles identified by their asymptotic behavior. The thimbles are representatives of these equivalence classes, with each thimble representing a different class.[12] More concretely, we assume that there are finitely many critical points ϕ^c_α indexed by α and $\mathrm{Im}S(\phi^c_\alpha) \neq \mathrm{Im}S(\phi^c_\beta)$ for $\alpha \neq \beta$.[13] Attached to each critical point is a thimble $\mathcal{T}_\alpha$ and a dual thimble $\mathcal{K}_\alpha$. As previously explained, different thimbles

[12]In this section we consider only those integration domains with no boundaries. The generalization of thimbles with boundaries were studied extensively by Delabaere and Howls (2002).

[13]These assumptions ensure that no two critical points are connected by flow since the flow conserves the imaginary part of the action. A situation where two critical points are connected by the flow is known as the *Stokes phenomenon*. We discuss Stokes phenomenon in Sec. III.

do not intersect with each other (they carry different values of ImS), and $\mathcal{T}_\alpha$ intersects with $\mathcal{K}_\beta$ if and only if $\alpha = \beta$. In other words, $\langle \mathcal{K}_\alpha, \mathcal{T}_\beta \rangle = \delta_{\alpha\beta}$, where $\langle \cdot, \cdot \rangle$ denotes the intersection number between two cycles. The intersection occurs at ϕ^c_α. Since ReS is bounded from below on a thimble, the integral (2.3) is guaranteed to be well defined when it is evaluated on a thimble $\mathcal{T}_\alpha$. In fact, the set of all thimbles forms a complete basis for the space of equivalence classes of "good domains" (i.e., the homology group), and any domain, say, $\mathcal{M}$, over which Eq. (2.3) is well defined is equivalent to a unique linear combination of thimbles (Pham, 1983):

$$\mathcal{M} \equiv \sum_\alpha n_\alpha(\mathcal{M}) \mathcal{T}_\alpha, \qquad n_\alpha(\mathcal{M}) = \langle \mathcal{K}_\alpha, \mathcal{M} \rangle. \tag{2.15}$$

In Eq. (2.15) the integer coefficients n_α are given by the number of intersections between $\mathcal{M}$ and the dual thimble $\mathcal{K}_\alpha$. The sign depends on the relative orientations of $\mathcal{K}_\alpha$ and $\mathcal{M}$. Notice that some of the n_α may vanish; it is said then that those thimbles do not contribute to the integral. A simple example of this is shown in Fig. 3.

Thimbles are the multidimensional generalization of the concept of "steepest descent" or "stationary phase" contour from the theory of complex functions of one variable. They are useful for studying the semiclassical expansion of path integrals in field theory (Cherman, Dorigoni, and Unsal, 2014; Dunne and Ünsal, 2016) and their asymptotic analysis; see (Aniceto, Başar, and Schiappa (2019) for a review of the new developments related to "resurgent transseries." Thimbles have also been used in attempts at defining ill-defined path integrals by defining the relevant partition function as an integral over one or more thimbles instead of over $\mathbb{R}^N$ (Harlow, Maltz, and Witten, 2011; Witten, 2010, 2011). For our purposes, the relevant property of the thimbles is that the imaginary part of the action, and consequently the phase of the integrand of the partition function, is constant on the thimble. Therefore, instead of evaluating the path integral on $\mathbb{R}^N$ where the phase is a rapidly oscillating function, evaluating it on the equivalent thimble decomposition where the phase is piecewise constant can provide a significant practical advantage. This fact by itself, however, is not quite enough to solve the sign problem. As can be seen from Eq. (2.3), the phase of the integrand depends also on the phase of the Jacobian (the "residual phase"). The Jacobian will have a rapidly oscillating phase if the shape of the manifold of integration oscillates quickly along real and imaginary directions. For theories in the semiclassical regime this does not happen, because the parts of the thimble with significant statistical weight are close to the critical point. Experience shows that the residual phase in many strongly coupled models introduces a mild sign problem (many examples follow in the review).[14]

An important question that naturally arises then is as follows: Which thimble, or combination of thimbles, is equivalent to $\mathbb{R}^N$? We can answer this question by considering

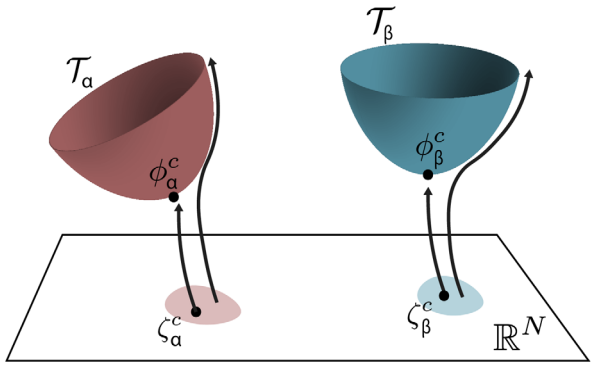


FIG. 4. The points $\zeta^c \in \mathbb{R}^N$ flow to the critical points $\phi^c \in \mathbb{C}^N$. The points in the neighborhood of each ζ^c approach the thimble but eventually veer off. Two such thimbles $\mathcal{T}_\alpha$ and $\mathcal{T}_\beta$ are shown for illustrative purposes.

the manifold $\mathcal{M}_T$ obtained by taking every point of $\mathbb{R}^N$ as an initial condition and flowing them by a time T. Since the real part of the action grows monotonically with T, the integral remains convergent at all T and, by the previous arguments, the value of the integral remains the same. Since $\mathbb{R}^N$ and the dual thimble of any critical point are N-dimensional spaces, they will generically intersect at isolated points (if they intersect at all). If we call each of those points ζ^c, we have $\phi^c = \mathcal{F}_{T\to\infty}(\zeta^c)$. When one starts from one of these intersection points ζ^c, the flow leads to the critical point on a trajectory lying on the dual thimble $\mathcal{K}$; see Figs. 3 and 4. The trajectory starting at points near ζ^c initially approaches the critical point but then veers along the unstable directions of the critical point slowly approaching the thimble; see Fig. 4. Points in $\mathbb{R}^N$ far from the intersection points take a more direct route toward infinity (or some other point where the action diverges). Therefore, all points in $\mathbb{R}^N$ flow, at large times, to points near a set of thimbles that together are equivalent to $\mathbb{R}^N$ (or to points where the action diverges). Furthermore, every thimble is counted as many times as there are intersection points between the corresponding dual thimble. Consequently, the thimble decomposition of $\mathbb{R}^N$ can explicitly be obtained as the following limit:

$$\mathcal{M}_{T\to\infty} = \sum_\alpha n_\alpha(\mathbb{R}^N) \mathcal{T}_\alpha, \tag{2.16}$$

where $\mathcal{M}_{T=0} = \mathbb{R}^N$. We stress that even though the thimble decomposition is obtained as the infinite flow time limit, the value of the integral remains unchanged during the deformation and $\mathcal{M}_T$ is equivalent to $\mathbb{R}^N$ for any finite value of T:

$$\begin{aligned} Z &= \int_{\mathbb{R}^N} d\phi\, e^{-S(\phi)} = \int_{\mathcal{M}_T} d\phi\, e^{-S(\phi)} \\ &= \sum_\alpha n_\alpha(\mathbb{R}^N) \int_{\mathcal{T}_\alpha} d\phi\, e^{-S(\phi)}. \end{aligned} \tag{2.17}$$

Note that, in theories where more than one thimble contributes to the partition function, there is a possibility that the contributions from different thimbles come with phases $\exp(-i\mathrm{Im}S_{\mathrm{eff}})$ (constant over each separate thimble) that

[14]One can construct examples of extremely strongly coupled theories where the residual phase introduces a severe sign problem (Lawrence, 2020).

induce a sign problem (sometimes called the "global sign problem"). This kind of sign problem is not helped by integrating over thimbles. However, for the contributions from different thimbles to nearly cancel an approximate symmetry, it is required to relate the contributions of different thimbles. Monte Carlo methods can be adapted to situations like that by sampling points related to the symmetry at the same time.

In field theories, where the dimensionality of the integral is large, it is extremely difficult to find the thimbles (it is in fact equivalent to classifying all *complex* solutions of the equations of motion), and it is even harder to find their intersection numbers n_α. The previous discussion is useful, however, in establishing an algorithm to solve this problem numerically and on the fly during a Monte Carlo run. It also clarifies the fact that there is nothing special about thimbles, as opposed to other manifolds obtained from flowing $\mathbb{R}^N$ by a finite time T. These other manifolds do not improve the sign problem as much as the thimbles do but still give the correct result for the integral and can be advantageous for numerical and algorithmic reasons.

III. ALGORITHMS ON OR NEAR THIMBLES

A. Single thimble methods

Early simulations using complex manifolds focused on sampling the path integral contribution from the "main" thimble, the thimble associated with the critical point with the smallest value of $S_R(\phi_c)$. This was based on the hope that in the relevant continuum and thermodynamic limits the path integral would be dominated by the contribution of a single thimble or that a regularization can be defined for relevant quantum field theories in terms of a single thimble path integral (Cristoforetti, Di Renzo, and Scorzato, 2012; Di Renzo, Singh, and Zambello, 2019). Although there is no evidence that this conjecture is valid, algorithms to sample a single thimble are stepping stones toward multithimble integration. We discuss in this section the algorithms proposed to sample the integral along a single thimble: the contraction algorithm, a Metropolis based algorithm (Alexandru, Başar, and Bedaque, 2015), a hybrid Monte Carlo algorithm (Fujii *et al.*, 2013), and the Langevin algorithm (Cristoforetti, Di Renzo, and Scorzato, 2012).

As discussed earlier, finding the thimble decomposition for the path integral is a hard problem that has been attempted only for quantum mechanical systems (Fujii, Kamata, and Kikukawa, 2015). However, in many cases it is feasible to find the main thimble even for realistic systems using the symmetry of the problem. The problem of finding the critical point is usually reduced to a "gap" equation to be solved analytically or numerically. For the algorithms discussed in this section, we assume that we have identified this critical point and we want to sample configurations on the corresponding thimble.

Another challenge facing any algorithm for the Monte Carlo evaluation of integrals over thimbles is to restrict sampling to the thimble manifold. For most systems there is no known method that can identify points on the thimble based on the local behavior of the action. Rather, a point has to be transported through the reverse flow [Eq. (2.10)] to decide whether or not it approaches the critical point. The thimble attached to this point can then be constructed by integrating the upward flow equations starting in the neighborhood of the critical point. As the thimble on the neighborhood of the critical point is approximated by the tangent space spanned by the Takagi vectors with positive eigenvalues in Eq. (2.13), we can take points on the tangent plane (close enough to the critical point) as the initial conditions of the holomorphic flow equation (2.7) to find points lying on the thimble. This "backward-and-forward" procedure then allows us to find points on the thimble near other points on the thimble, as required by Monte Carlo procedures, at the expense of integrating the flow equations. This process provides a map between the N-dimensional neighborhood of the critical point and the thimble attached to it. It is an essential ingredient for all single thimble algorithms discussed here. For a given parametrization of the tangent space near the critical point ϕ^c,

$$\phi_n = \phi^c + \sum_{a=1}^{N} \zeta_a \rho^{(a)}, \qquad \zeta_a \in \mathbb{R}, \tag{3.1}$$

integrating the upward flow for a time T produces a map $\phi_n \to \phi_f = \mathcal{F}_T(\phi_n)$. In Eq. (3.1) ϕ_n is a point near ϕ^c and ϕ_f is moved far away by the flow. For large enough T, this will map a small neighborhood of the critical point onto a manifold close to the thimble, and the larger the value of T, the closer the manifold generated by the $\phi_n \to \phi_f$ mapping is to the thimble. As a practical method of determining an appropriate value for T, simulations can be carried out for increasing values of T until the results converge.

Having chosen an appropriate T, we now have the means to parametrize the thimble using the tangent plane close to the critical point. We can then approximate the integral over the thimble as

$$\int_{\mathcal{T}} d\phi_f e^{-S(\phi_f)} \approx \int_U d\phi_n \det J(\phi_n) e^{-S(\phi_f(\phi_n))}, \tag{3.2}$$

where $J_{ij} = \partial(\phi_f)_i/\partial(\phi_n)_j$ is the Jacobian of the map and U is the region around ϕ^c in the tangent plane that is mapped to the manifold approximating the region of the thimble that dominates the integral. For the special case where the tangent plane is in the same homology class as the thimble, the region U can be extended to the entire tangent plane and Eq. (3.2) becomes exact for all flow times T. For the case in which the tangent plane is not in the same homology class, the relation becomes exact only in the limit of large T. In practice the region U is generated implicitly in the simulations: we start in the neighborhood of the critical point, and the proposed updates move smoothly, or in small discrete steps, through the configuration space. The potential barriers force the simulation to stay in the relevant region. To fix terminology we refer to the region U in the tangent plane as the *parametrization manifold* and the image under the map $\mathcal{F}_T(U)$ as the *integration manifold*.

The goal of the algorithms presented here is to sample the integration manifold according to the Boltzmann factor $\exp(-S)$. Since the action and the integration measure are

complex, we need to use a modified Boltzmann factor for sampling. The probability density we will sample corresponds to

$$P_0(\phi_f)|d\phi_f| = \frac{1}{Z_0} e^{-\mathrm{Re}S(\phi_f)}|d\phi_f|, \qquad Z_0 \equiv \int_{\mathcal{T}} |d\phi_f| e^{-\mathrm{Re}S(\phi_f)}. \tag{3.3}$$

The final result for observables will have to include the phase

$$\langle \mathcal{O} \rangle = \frac{\langle \mathcal{O} e^{i\varphi} \rangle_0}{\langle e^{i\varphi} \rangle_0}, \qquad e^{i\varphi} \equiv e^{-i\mathrm{Im}S(\phi_f)} \frac{d\phi_f}{|d\phi_f|}. \tag{3.4}$$

Since we are sampling the configurations from a single thimble, or from a manifold that is close to it, the imaginary part of the action is constant (or nearly so.) The only fluctuations come from the residual phase associated with the phase of the measure $d\phi_f$. If we view this as an integral over the parametrization manifold, then the probability measure is

$$P_0(\phi_n) = \frac{1}{Z_0} e^{-\mathrm{Re}S_{\mathrm{eff}}(\phi_n)}, \qquad S_{\mathrm{eff}} = S(\phi_f(\phi_n)) - \ln \det J(\phi_n). \tag{3.5}$$

The complex phase in this case is $\exp(-i\mathrm{Im}S_{\mathrm{eff}})$, and the fluctuations of this phase are dominated by the Jacobian phase that corresponds to the residual phase. Note that to compute the effective action for a point ϕ_n in the parametrization space we have to integrate the upward flow differential equation with initial condition ϕ_n for a time T to get ϕ_f. $S(\phi_f)$ is then the action contribution. The other contribution comes from the Jacobian. As explained in Appendix A the Jacobian matrix can be computed by integrating the matrix differential equation as follows:

$$\frac{dJ}{dt} = \overline{H(\phi(t))J(t)}, \tag{3.6}$$

where $H(\phi(t))$ is the Hessian matrix of S along the flow and the initial condition $J(0)$ is a matrix whose columns form an orthonormal basis in the tangent to the parametrization space at ϕ_n. Equation (3.6) flows a basis in the tangent space at ϕ_n to a basis in the tangent space at ϕ_f. Since our parametrization space is a hyperplane, the basis for the tangent space at ϕ_n can be chosen to be the same at all points in U, for example, the positive Takagi vectors or any other basis spanning this tangent space.

Equation (3.6) can also be used to map a single infinitesimal displacement represented by a vector v_n in the tangent space at ϕ_n to a displacement represented by a vector v_f in the tangent space on the thimble at ϕ_f. In Eq. (3.6) $J(t)$ is then replaced by $v(t)$, the column vector representing the displacement. The initial condition is $v(0) = v_n$ and the final result $v(T) = v_f$ is a vector in the tangent space at ϕ_f. For this reason we sometimes call this equation the *vector flow* in the review.

1. Contraction algorithm

Several sampling algorithms are based on the mapping between the tangent plane and the approximate thimble. The most straightforward is the contraction algorithm (Alexandru, Başar, and Bedaque, 2015; Alexandru, Başar, Bedaque, Ridgway, and Warrington, 2016a), which generates configurations in the parametrization manifold based on the probability P_0 using the Metropolis method (Metropolis *et al.*, 1953) based on the effective action $\mathrm{Re}S_{\mathrm{eff}}$. The basic process is now detailed.

(1) After a critical ϕ^c point is identified, the tangent space of its thimble is computed by solving Eq. (2.13) and finding the $\rho^{(a)}$ value corresponding to positive $\lambda^{(a)}$.
(2) Start with a point $\phi_n = \phi^c + \sum_{a=1}^{N} \zeta_a \rho^{(a)}$ on the tangent space. Evolve ϕ_n by the holomorphic flow by a time T to find ϕ_f, compute the Jacobian $J(\phi_n)$ by integrating the flow equation for the basis, and then compute the action $S_{\mathrm{eff}}(\phi_n)$.
(3) Propose new coordinates $\zeta' = \zeta + \delta\zeta$, where $\delta\zeta$ is a random vector chosen with a symmetric probability function, that is, $P(\delta\zeta) = P(-\delta\zeta)$. Evolve $\phi'_n = \sum_a \zeta'_a \rho^{(a)}$ by the holomorphic flow by a time T to find ϕ'_f and compute $J(\phi'_n)$ and $S_{\mathrm{eff}}(\phi'_n)$.
(4) Accept or reject ζ' with probability $\min\{1, e^{-S'_{\mathrm{eff}}+S_{\mathrm{eff}}}\}$.
(5) Repeat steps 3 and 4 until a sufficient ensemble of configurations is generated.

To make the updating effective, we have to account for the fact that the map $\mathcal{F}_T$ is highly anisotropic. If we consider the flow close to the critical point, we see that displacements in the direction of the Takagi vector $\rho^{(a)}$ are mapped onto vectors that have their magnitude increased by $\exp(\lambda^{(a)}T)$. Even small differences in the eigenvalues $\lambda^{(a)}$ lead to large differences as T increases. If the parametrization space proposals $\delta\zeta$ are isotropic, then the update process becomes inefficient. Ideally we want to generate proposals that are isotropic on the integrations manifold, but since the map changes from point to point, this requires care to ensure that the detailed balance is preserved. It turns out that this is possible, but we refrain from discussing this point until later. An easy fix for this problem is to adjust the size of displacement for the proposal based on the flow around the critical point. The proposal is then $\delta\zeta_a = \exp(-\lambda^{(a)}T)\delta$, with δ a random variable chosen with uniform probability in the interval $[-\Delta, \Delta]$. The step size Δ is tuned to get reasonable acceptance rates. If the distortions induced by the map $\mathcal{F}_T$ vary only slightly from ϕ^c to the points sampled by the process, then this algorithm is effective.

By far the most computationally expensive part of the contraction algorithm (and most other thimble algorithms) is the computation of the Jacobian (even for most bosonic systems the cost scales with N^3 and N is proportional to the spacetime volume). Methods to deal with this problem are discussed in Sec. III.D.

Another Metropolis based method was proposed to sample single thimble configurations (Mukherjee, Cristoforetti, and Scorzato, 2013) and was tested for a single plaquette U(1) problem. In this proposal the Jacobian is not included in the sampling and is to be included via reweighting in the observable measurement. This reweighting will fail for most

systems that have more than a few degrees of freedom since for these systems the Jacobian fluctuates over many orders of magnitude.

2. The HMC method on thimbles

A more sophisticated algorithm based on the hybrid Monte Carlo (HMC) method (Duane *et al.*, 1987) was proposed and tested for the ϕ^4 model (Fujii, Kamata, and Kikukawa, 2015). In principle, a straightforward extension of the HMC method could be applied to the action $\mathrm{Re}S_{\mathrm{eff}}$ on the parametrization manifold. The problem with such an approach is that it would require the calculation of the derivatives of $\det J$ or some related quantity, which is cumbersome. This could be sidestepped by neglecting the Jacobian in the sampling (Ulybyshev, Winterowd, and Zafeiropoulos, 2020), but this requires reweighting it in the observables and fails for large systems. The proposal is then to use the HMC method as defined by the Hamiltonian in the larger $\mathbb{C}^N$ space, where the motion is confined to be on the thimble via forces of constraint (Fujii, Kamata, and Kikukawa, 2015). This has the advantage that the Jacobian is accounted for implicitly, but the algorithm requires solving implicit equations to project back to the thimble. For the cases where the thimble is relatively flat and smooth, these equations can be solved robustly via iteration, as is the case with the ϕ^4 system in the investigated parameter range.

The basic idea is to integrate the equations of motion generated by the Hamiltonian

$$\mathcal{H}(\pi,\phi_f)=\tfrac{1}{2}\pi^\dagger\pi+\mathrm{Re}S(\phi_f) \tag{3.7}$$

subject to the constraint that $\phi_f\in\mathcal{T}$. Forces of constraint perpendicular to the thimble keep the system confined on its surface. The momentum π is in the tangent space at ϕ_f, so it is a *real* linear combination of columns of $J(\phi_n)$. The perpendicular force has to be a real linear combination of the columns of $iJ(\phi_n)$ since this forms a basis in the space perpendicular (according to the scalar product $\langle v|w\rangle\equiv\mathrm{Re}v^\dagger w$) to the thimble.

For a practical implementation we need to provide an integrator for these equations of motion for finite time steps. A symplectic integrator for this problem is provided using the following method:

$$\begin{aligned}\pi_{1/2}&=\pi-\partial_{\phi_f}\mathrm{Re}S(\phi_f)\frac{\Delta t}{2}+iJ(\phi_n)\lambda,\\ \phi_f'&=\phi_f+\pi_{1/2}\Delta t,\\ \pi'&=\pi_{1/2}-\partial_{\phi_f}\mathrm{Re}S(\phi_f')\frac{\Delta t}{2}+iJ(\phi_n')\lambda'.\end{aligned} \tag{3.8}$$

The map $(\pi,\phi_f)\to(\pi',\phi_f')$ is symplectic and time reversible, thus satisfying the requirements for the HMC method. Note that this map requires the determination of λ and λ', two sets of N real numbers that encode the effect of the constraint forces acting perpendicularly on the thimble. λ is determined by the requirement that $\phi_f'\in\mathcal{T}$ and λ' by requiring that π' is in the tangent space at ϕ_f'. For small enough Δt, these requirements lead to unique "small" solutions (which vanish in the $\Delta t\to 0$ limit) for λ's. A solution for λ' can be computed in a straightforward way via the projection method we discuss later. Computing λ is more difficult and the current proposal is to use an iterative method (Fujii, Kamata, and Kikukawa, 2015). This iteration is guaranteed to converge for small enough Δt, but for a fixed size Δt no guarantees can be made, even for the existence of a solution.

With these ingredients in hand, the basic steps of the HMC method are the following:

(1) At the beginning of each "trajectory" an isotropic Gaussian momentum π is generated in the tangent space at ϕ_f, $P(\pi)\propto\exp(-\pi^\dagger\pi/2)$.
(2) The equations of motion are integrated by repeatedly iterating the previous integrator steps $t/\Delta t$ times, where t is the trajectory length.
(3) At the end of the trajectory the proposed (π',ϕ_f') are accepted with a probability determined by the change in Hamiltonian $P_{\mathrm{acc}}=\min\{1,\exp(-\mathcal{H}+\mathcal{H}')\}$.

One ingredient for this and other algorithms that we discuss later is the projection to the tangent space at ϕ_f. If we have the Jacobian matrix in hand $[J(\phi_n)]$, its columns form a real basis of the tangent space and the columns of $iJ(\phi_n)$ form a basis for the orthogonal space. Every vector $v\in\mathbb{C}^N$ can then be decomposed into its parallel $[P_\parallel(\phi_f)v]$ and perpendicular $[P_\perp(\phi_f)v]$ components using standard algebra. This step is required for finding λ' in the symplectic integrator. It can also be used to find the starting momentum at the beginning of the trajectory: we generate a random vector in $\mathbb{C}^N$ with probability $P(\tilde\pi)\propto\exp(-\tilde\pi^\dagger\tilde\pi/2)$ and then project it to the tangent plane $\pi=P_\parallel(\phi_f)\tilde\pi$.

The previously discussed projection can be readily implemented when we have the Jacobian matrix $J(\phi_n)$. However, calculating this matrix is an expensive operation that is likely to become a bottleneck for simulations of systems with a large number of degrees of freedom. One solution to this problem is the following (Alexandru, Başar, Bedaque, and Ridgway, 2017): we use the map $v\to J(\phi_n)v$, which maps the tangent space at ϕ_n on the parametrization manifold to the tangent space at ϕ_f on the thimble. This calculation can be implemented efficiently by solving the vector flow equation (3.6) for a single vector v. We extend this to arbitrary vectors that are not included in the tangent space. For a generic vector v we split it into $v_1=P_\parallel^{(0)}v$ and $iv_2=P_\perp^{(0)}v$. Here $P_\parallel^{(0)}$ is the projection on the tangent space of the parametrization manifold, the space spanned by the Takagi vectors, and $P_\perp^{(0)}$ is its orthogonal complement. Both v_1 and v_2 belong to the tangent space at ϕ_n, so $J(\phi_n)v_{1,2}$ can be computed using the vector flow equations. This then defines a map from any vector v to $J(\phi_n)v=J(\phi_n)v_1+iJ(\phi_n)v_2$, which requires two integrations of the vector flow. Using this map we can then compute $J^{-1}(\phi_n)v$ using an iterative method such as BiCGSTAB. It is then straightfoward to prove that $P_\parallel(\phi_f)v=J(\phi_n)P_\parallel^{(0)}J(\phi_n)^{-1}v$.

3. Langevin on thimbles

The Langevin algorithm was proposed as a possible sampling method for single thimble manifolds (Cristoforetti,

Di Renzo, and Scorzato, 2012; Cristoforetti *et al.*, 2013; Cristoforetti, Scorzato, and Di Renzo, 2013). The idea is to sample the thimble manifold $\mathcal{T}$ with a probability density proportional to $\exp(-\mathrm{Re}S)$ with respect to the Riemann measure induced by embedding $\mathcal{T}$ in $\mathbb{C}^N$. The residual phase of the measure is taken into account via reweighting. The imaginary part of the action is constant over the thimble and will not contribute to averages.

The Langevin process simulates the evolution of the system via a drift term due to the action and a Brownian motion term. The discretized version of the process is given by the following updates:

$$\phi'_f = \phi_f - \partial_{\phi_f}\mathrm{Re}S(\phi_f)\Delta t + \eta\sqrt{2\Delta t}, \tag{3.9}$$

where the vector η is a random N-dimensional vector in the tangent space at the thimble at ϕ_f.

Two details are important here: how the vector η is chosen and how the new configuration ϕ'_f is projected back to the thimble. The proposal is to choose η isotropically at ϕ_f by generating a Gaussian $\tilde{\eta}$ unconstrained in $\mathbb{C}^N$ and then project it to the tangent space at ϕ_f using a procedure similar to the projection outlined in Sec. III.A.2, $\eta = P_{\parallel}(\phi_f)\tilde{\eta}$. This ensures an isotropic proposal in the tangent space and the norm of the vector is adjusted such that it follows the χ^2 distribution with N degrees of freedom (Cristoforetti *et al.*, 2013).

At every step we start with ϕ_f on the thimble and we move along the tangent direction, since both the drift and the random vector lie in the tangent plane. Unless the thimble is a hyperplane, this shift will take us out of the thimble. A projection back to the thimble is required. The proposed methods rely on evolving the new configuration in the downward flow toward the critical point, projecting there to the thimble, and flowing back (Cristoforetti, Di Renzo, and Scorzato, 2012; Cristoforetti, Scorzato, and Di Renzo, 2013). This proposal was found to be unstable (Cristoforetti, Scorzato, and Di Renzo, 2013). The only simulations we are aware of that employ this algorithm involve simulations on the tangent plane to the thimble (Cristoforetti *et al.*, 2013). In this case the updates do not require any projection, since the manifold is flat. To make this algorithm practical for the general case, a robust projection method is needed.

We make a final note about the Langevin algorithm: for a finite Δt the method is not exact. Simulations have to be carried out for decreasing Δt and then extrapolated to $\Delta t = 0$ to remove the finite step-size errors. For other Langevin methods, an accept-reject step can be used to remove the finite step-size errors, but this has not been developed for thimble simulations.

While both the Langevin method and the HMC algorithm perform updates directly on ϕ_f with a drift (or force) term evaluated locally, it is worth emphasizing that the updates still require the integration of the flow equations. This is because the projection of the shift to the tangent plane to the thimble and the required projection back to the manifold after the update can currently be done only by connecting ϕ_j with its image under the flow ϕ_n in the infinitesimal neighborhood of the critical point. The advantage of these methods over the Metropolis algorithm, assuming that a practical projection method is available, is that the updates can lead to a large change in action resulting in small autocorrelation times in the Markov chain.

4. Case study: Bosonic gases

We presently consider the relativistic Bose gas at finite density for an application of these algorithms to bosonic systems with sign problems. The continuum Euclidean action of this system is

$$S = \int d^4x[\partial_0\phi^*\partial_0\phi + \nabla\phi^*\cdot\nabla\phi + (m^2-\mu^2)|\phi|^2 + \mu\underbrace{(\phi^*\partial_0\phi - \phi\partial_0\phi^*)}_{j_0(x)} + \lambda|\phi|^4], \tag{3.10}$$

where $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$ is a complex scalar field. This action encodes the properties of a two-component system of bosons with a contact interaction and an internal global U(1) symmetry that breaks spontaneously at high density. In Euclidean space, the current j_0 is complex and causes a sign problem.[15] This system was studied using a complex Langevin (Aarts, 2009) and in a different representation where the sign problem disappears (Endres, 2007; Gattringer and Kloiber, 2013).

This system was studied with the contraction algorithm given by Alexandru, Başar, Bedaque, Ridgway, and Warrington (2016a), the HMC method (Fujii *et al.*, 2013), and the Langevin process (Cristoforetti *et al.*, 2013). The following lattice discretization of Eq. (3.10) was used:

$$S = \sum_{x,a}\left[\left(4+\frac{m^2}{2}\right)\phi_{x,a}\phi_{x,a} - \sum_{\nu=1}^{3}\phi_{x,a}\phi_{x+\hat{\nu},a} - \cosh\mu\,\phi_{x,a}\phi_{x+\hat{0},a} + i\sinh\mu\,\epsilon_{ab}\phi_{x,a}\phi_{x+\hat{0},b} + \frac{\lambda}{4}(\phi_{x,a}\phi_{x,a})^2 - h(\phi_{x,1}+\phi_{x,2})\right], \tag{3.11}$$

where ϵ_{ab} is the antisymmetric tensor and $\epsilon_{12} = 1$. This lattice action is used in the remainder of this discussion. The final term must be included in the lattice theory to obtain a well-defined thimble decomposition and we take h as small.

To apply the contraction algorithm, it is first necessary to find critical points (extrema) of the action (3.11). Restricting attention to those critical points that are constant in spacetime, the following extremum condition is obtained:

$$(2+m^2)\phi - 2\cosh\mu\phi + 2\lambda|\phi|^2\phi = h. \tag{3.12}$$

Three extrema exist and we denote them as ϕ_0, ϕ_+, ϕ_-. The corresponding Lefschetz thimbles are denoted as $\mathcal{T}_0, \mathcal{T}_+, \mathcal{T}_-$. Depending on the parameters of the theory, different combinations of thimbles contribute to the path integral. To this end,

[15]This is most readily seen in Fourier space in the continuum: $\int d^4x\, j_0(x) = (2\pi)^{-4}\int d^4p(-2ip_0)|\phi(p)|^2$ is purely imaginary.

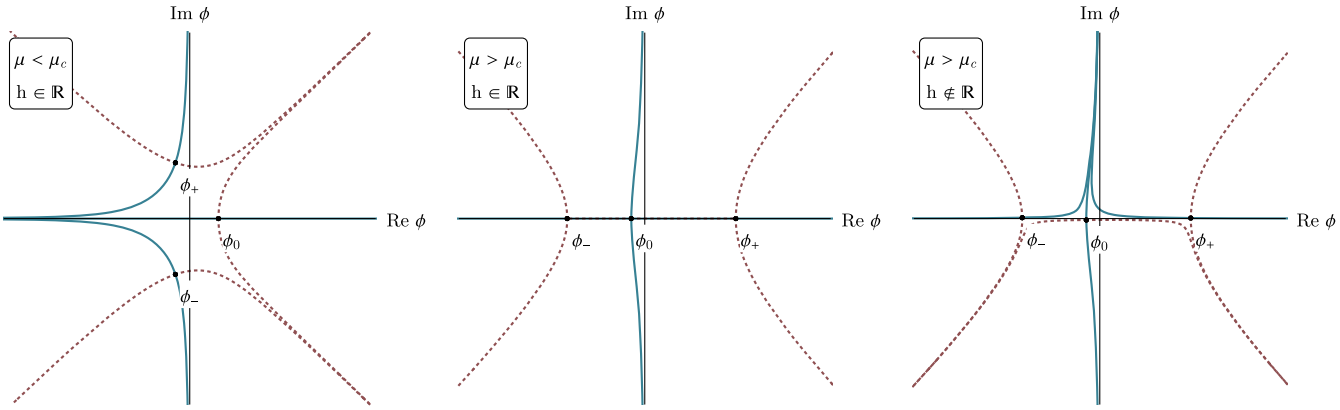


FIG. 5. Projections of the thimbles (solid blue lines) $\mathcal{T}_0, \mathcal{T}_+$, and $\mathcal{T}_-$ and the dual thimbles (dotted red lines) $\mathcal{K}_0, \mathcal{K}_+$, and $\mathcal{K}_-$ onto the one-complex-dimensional subspace of constant fields. The intersection of the original domain of integration with this subspace corresponds to the real line. The typical arrangement of thimbles varies with the chemical potential.

the one-dimensional projections of $\mathcal{T}_0, \mathcal{T}_+, \mathcal{T}_-$ depicted in Fig. 5 are useful.

For $\mu < \mu_c = \cosh^{-1}(1 + m^2/2)$, only $\mathcal{T}_0$ contributes to the path integral. This occurs because $S_R(\phi_\pm) < S_R(\phi)$ for any ϕ on the original integration manifold, and therefore no point can flow to $\phi_\pm$ by the upward flow. This is sufficient to eliminate $\mathcal{T}_\pm$ as contributing thimbles.

For $\mu > \mu_c$, the contributing thimbles changes. As seen in the center of Fig. 5, when $h \in \mathbb{R}$ there are flow trajectories connecting both ϕ_- and ϕ_+ to ϕ_0. This feature, called the Stokes phenomenon, introduces complications into the decomposition of the path integral into an integer linear combination of thimbles. We avoid the Stokes phenomenon altogether by simply introducing a complex h; for a detailed discussion of our procedures see Alexandru, Başar, Bedaque, Ridgway, and Warrington (2016a).

Since our purpose is to illustrate the contraction algorithm, we consider only the $\mu > \mu_c$ case here. As an example, let $m = \lambda = 1.0$, $h = 0.1(1 + i/10)$, and $\mu = 1.3$. With these choices, $\mathcal{T}_+$ contributes most to the path integral. The results obtained on flowed manifolds are plotted in Fig. 6. The variance of S_I decreases (up to statistical errors) as a function of flow time; this demonstrates that the integral over $\mathcal{T}_+$ indeed has reduced phase fluctuations relative to $\mathbb{R}^N$. Furthermore, the convergence of observables as a function of flow time strongly suggests convergence to $\mathcal{T}_+$.

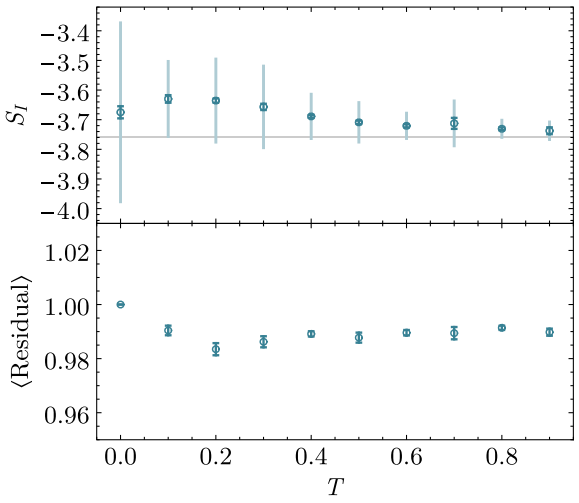


FIG. 6. The imaginary part of the action and the residual phase computed on $\mathcal{F}_T(\mathcal{T}_+)$ using the contraction algorithm. The horizontal line denotes the value of $\mathrm{Im}S(\phi_+)$.

B. Generalized thimble method

The main limitation of the methods discussed thus far is that they are capable of computing the integral over only one thimble. However, the integral over the real variables is generically equivalent to the integral over a collection of thimbles. Finding this collection of thimbles is a daunting process; integrating over all of them is an even harder task. However, there is a way of bypassing this difficulty based on what we learned in Sec. II: the generalized thimble method.

Recall that if every point of $\mathbb{R}^N$ (the integration region of the path integral) is taken to be the initial condition for Eq. (2.7) that is then integrated for a time T, we obtain a manifold $\mathcal{M}_T = \mathcal{F}_T(\mathbb{R}^N)$ that is equivalent to the initial $\mathbb{R}^N$ manifold (in the sense that the path integral over $\mathbb{R}^N$ and $\mathcal{M}_T$ are the same). In addition, for large enough values of T, $\mathcal{M}_T$ approaches exactly the combination of thimbles equivalent to $\mathbb{R}^N$. It is important to understand how the thimbles are approached. In the large T limit an isolated set of points in $\mathbb{R}^N$, we call each of them ζ^c, approach the critical points ϕ^c of the relevant thimbles. Points near them initially approach the critical points but, when close to them, move along the unstable directions, which are almost parallel to the thimble but slowly approach it; see Fig. 4. Points far from ζ^c run toward a point when the action diverges, either at infinity or at a finite distance (in fermionic theories thimbles meet at points where the action diverges, as exemplified by the Thirring model discussed later). This means that the correct combination of thimbles equivalent to the original path integral can be parametrized by points in $\mathbb{R}^N$. This is an advantage over the contraction method, where only one thimble at a time can be parametrized. We then have

$$\int_{\mathbb{R}^N} d\phi\, e^{-S(\phi)} = \int_{\mathcal{M}_T} d\tilde{\phi}\, e^{-S(\tilde{\phi})} \det J(\tilde{\phi}) = \int_{\mathbb{R}^N} d\zeta\, e^{-S[\mathcal{F}_T(\zeta)]} \det J(\zeta). \tag{3.13}$$

The generalized thimble method consists of using a Metropolis algorithm on $\mathbb{R}^N$ with the action $\mathrm{Re}S_{\mathrm{eff}}$, where $S_{\mathrm{eff}}(\zeta) = S[\mathcal{F}_T(\zeta)] - \log\det J(\zeta)$.

1. Generalized thimble algorithm (GTA)

(1) Start with a point ζ in $\mathbb{R}^N$. Evolve it by the holomorphic flow by a time T to find $\phi_f = \mathcal{F}_T(\zeta)$.
(2) Propose new coordinates $\zeta' = \zeta + \delta\zeta$, where $\delta\zeta$ is a random vector drawn from a symmetric distribution. Evolve it by the holomorphic flow by a time T to find $\phi'_f = \mathcal{F}_T(\zeta')$.
(3) Accept ζ' with probability $P_{\mathrm{acc}} = \min\{1, e^{-\Delta \mathrm{Re}S_{\mathrm{eff}}}\}$.
(4) Repeat from step 2 until a sufficient ensemble of configurations is generated.

Methods to speed up (or bypass) the frequent computation of the Jacobian J are an improvement of the method and are discussed later; see Sec. III.D.

While the previous algorithm is exact, the practical applicability of the GTA depends on the landscape induced by $\exp(-\mathrm{Re}S_{\mathrm{eff}})$ on $\mathcal{M}_T$. At large T, the points ζ that are mapped to the statistically significant parts of $\mathcal{M}_T$ lie on small, isolated regions. This explains why the phase of the integrand fluctuates less on $\mathcal{M}_T$ than on $\mathbb{R}^N$. The imaginary part of $S[\mathcal{F}_T(\zeta)]$ on points on $\mathcal{M}_T$ are the same as the imaginary parts of the action $S(\zeta)$ in a small region around ζ^c, the only region with significant statistical weight $\exp\{-\mathrm{Re}S_{\mathrm{eff}}[\mathcal{F}_T(\zeta)]\}$.

Between the regions around the different ζ^c lie areas with small statistical weight $\exp\{-\mathrm{Re}S_{\mathrm{eff}}[\mathcal{F}_T(\zeta)]\}$ that are mapped to points where the action nearly diverges, as we discussed in Sec. II.C. A probability landscape of this form may trap the Monte Carlo chain in one of the high probability regions, thereby breaking ergodicity. A trapped Monte Carlo chain is effectively sampling only one of the thimbles contributing to the integral (more precisely, it is an approximation to a one thimble computation). This problem can be alleviated by making T small. In that case $\mathcal{M}_T$ will be farther away from the thimbles, the phase oscillations are larger, and the original sign problem may not be controlled. The usefulness of the GTA relies then on being able to find a value of T such that the sign problem is sufficiently ameliorated while the trapping of the Monte Carlo chain is not a problem. In several examples discussed later, over a large swatch of parameter space, it is not difficult to find a range of values of T for which the GTA is useful. Still, one should perform due diligence and try to diagnose trapping signs in every calculation, as is always the case in Monte Carlo calculations.

2. Case study: 0 + 1D Thirring model

We now use the finite density and temperature Thirring model in $0+1$, $1+1$, and $2+1$ spacetime dimensions to illustrate several of the techniques discussed in this review. The Thirring model was initially formulated as an example of a solvable model in $1+1$ dimensions (Thirring, 1958), and it describes fermions with a contact vector-vector interaction and is described by the Lagrangian density

$$\mathcal{L} = \bar{\psi}^a(i\partial\!\!\!/ + m + \mu\gamma^0)\psi^a + \frac{g^2}{2N_F}\bar{\psi}^a\gamma_\mu\psi^a\bar{\psi}^a\gamma^\mu\psi^a, \tag{3.14}$$

where ψ is a spinor for the appropriate spacetime dimension and a indexes the N_F different flavors of fermions. This theory is, in $1+1$ dimensions, asymptotically free. The $N_F = 1$ case is identical to the Gross-Neveu model and its ground state breaks a discrete symmetry spontaneously and, in this respect, resembles QCD. For $N_F > 1$ the $\langle\bar{\psi}\psi(0)\bar{\psi}\psi(x)\rangle$ two-point function exhibits power law decay, the closest behavior to long-range order possible in one spatial dimension (Witten, 1978).

We will use two discretizations of the Thirring model, one using staggered fermions and the other using Wilson fermions. The lattice action in d dimensions is

$$S = \sum_{x,\nu}\frac{N_F}{g^2}[1-\cos A_\nu(x)] + \sum_{x,y}\bar{\psi}^a(x)D_{xy}\psi^a(y), \tag{3.15}$$

with

$$D^W_{xy} = \delta_{xy} - \kappa\sum_{\nu=0,1}[(1-\gamma_\nu)e^{iA_\nu(x)+\mu\delta_{0\nu}}\delta_{x+\nu,y} + (1+\gamma_\nu)e^{-iA_\nu(x)-\mu\delta_{0\nu}}\delta_{x,y+\nu}], \tag{3.16}$$

with $1/\kappa = 2m + 4d$ for Wilson fermions or

$$D^{\mathrm{KS}}_{xy} = m\delta_{xy} + \frac{1}{2}\sum_{\nu=0,1}[\eta_\nu(x)e^{iA_\nu(x)+\mu\delta_{0\nu}}\delta_{x+\nu,y} - \eta^\dagger_\nu(x)e^{-iA_\nu(x)-\mu\delta_{0\nu}}\delta_{x,y+\nu}], \tag{3.17}$$

with $\eta_0(x) = 1$, $\eta_1 = (-1)^{x_0}$, and $\eta_2 = (-1)^{x_0+x_1}$ for Kogut-Susskind staggered fermions. The flavor index goes from 1 to N_F in the Wilson fermion case but from 1 to $N_F/2$ in the staggered case. Integrating over the bosonic field $A_\nu(x)$ leads to a discretized version of Eq. (3.14), showing their equivalence. Integration over the fermion fields leads to purely bosonic action more amenable to numerical calculations as follows:

$$S = N_F\left(\frac{1}{g^2}\sum_{x,\nu}[1-\cos A_\nu(x)] - \gamma\log\det D(A)\right), \tag{3.18}$$

with $\gamma = 1$ (Wilson) or $\gamma = 1/2$ (staggered). Both of these actions describe N_F Dirac fermions in the continuum. The presence of the chemical potential μ renders the fermion determinant complex and is the origin of the sign problem in this model.

The $(0+1)$-dimensional case can be solved exactly with the lattice action in Eq. (3.18), and it has been used as a check on several methods designed to handle sign problems (Pawlowski and Zielinski, 2013; Li, 2016; Fujii, Kamata, and Kikukawa, 2017; Di Renzo, Singh, and Zambello, 2021; Di Renzo and Zambello, 2021). Its thimble structure is known. In the $A_0(x) = \mathrm{const}$ sector, it is shown in Fig. 7. The "main

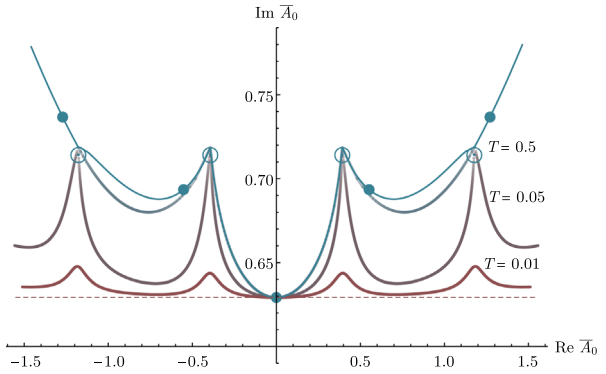


FIG. 7. Complex $\bar{A}_0 = (1/V)\sum_x A_0(x)$ plane for the Thirring model. The filled circles are critical points, the open circles are singular points of the action, and the thin lines are thimbles. The dashed line is the tangent space to the main thimble, while the other solid lines are the manifolds $\mathcal{M}_T$ obtained by flowing the tangent space by $T = 0.01$, 0.05, and 0.5. Notice how $\mathcal{M}_T$ approaches the correct combination of thimbles as T is increased.

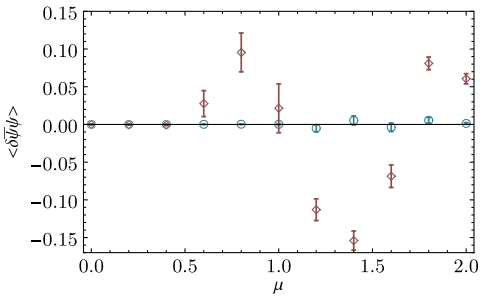


FIG. 8. The difference in the value of the chiral condensate between the exact result and the one obtained using the contraction method (with $T = 2$) shown in red diamonds and the generalized thimble method (with $T = 0$, that is, integration over the tangent space). This is for the $0 + 1$ Thirring model with parameters $N = 8$, $m = 1$, and $g^2 = 1/2$ (lattice units).

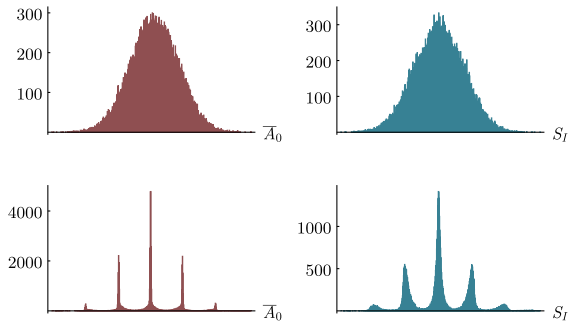


FIG. 9. Histogram of the average field $\bar{A}_0$ (left panels) and the imaginary part of the action (right panels) in a Monte Carlo sampling in the 1D Thirring model using the GTM with the $T = 0$ (top line) and $T = 0.5$ (bottom line), $g^2a^2 = 1/2$, and $N = 32$ calculations. In the $T = 0$ calculation the phase e^{-iS_I} fluctuates too wildly and the result has large uncertainties. In the $T = 0.5$ calculation on the bottom line the phase fluctuates much less. It is also evident that regions on the tangent space corresponding to several thimbles are being sampled. The multimodal distribution in the $T = 0.5$ calculation indicates that larger flow values could lead to trapping of the Monte Carlo chain in a region corresponding to only one thimble.

critical point" is the critical point with the smallest real part of the action (the one with $\mathrm{Re}A_0 = 0$ for this system). Therefore, in the semiclassical limit it should dominate the path integral. Thimbles touch each other at points where the fermion determinant vanishes and the effective bosonic action diverges (shown as empty circles in Fig. 7). The tangent space to the main thimble ($\mathbb{T}$) is just the real space shifted in the imaginary direction (the dashed red line in Fig. 7). The integration over the tangent space is no more expensive than the one over the real space since no flowing is required and the Jacobian of the transformation is 1. The tangent space lying parallel to the real space has the same asymptotic behavior as $\mathbb{R}^N$ and is equivalent to it for the computation of the integral. The figure also shows the result of "flowing" the tangent space by different values of T: the larger the value of T, the closer the resulting manifold(s) get to the thimbles. Starting from the tangent space and using a flow time $T = 2$, the obtained manifold $\mathcal{F}_T(\mathbb{T})$ is nearly indistinguishable from the thimbles.

Alexandru, Başar, and Bedaque (2015) studied the model using the contraction algorithm. The results shown in Fig. 8 indicate that the fermion condensate, for instance, is close to the exact result but does not agree with it, particularly for certain values of μ near the transition from $\langle\bar{\psi}\psi\rangle = 0$ to $\langle\bar{\psi}\psi\rangle \neq 0$. The size of the discrepancy is consistent with a semiclassical estimate of the contributions of other thimbles (besides the main thimble). Similar behavior was seen in a one-site model of fermions (Tanizaki, Hidaka, and Hayata, 2016). The integration over the tangent space, however, gives the correct result. The average sign on the tangent space is smaller than the one obtained with the contraction method. For temperatures that are not too low, the sign fluctuation is, however, small enough to allow for the computation to be done on the tangent plane. But as the temperature is lowered, the sign fluctuations grow and it becomes difficult to sample the correct distribution, as predicted by general arguments; see Eq. (1.4). One can then use the generalized thimble method and integrate on the manifold $\mathcal{F}_T(\mathbb{T})$ for a suitable value of T. With a T that is too small, the sign fluctuation is too large; a T that is too large is essentially an integration over one thimble and the wrong results are obtained. It is interesting to understand how the transition between these two behaviors occur. In Fig. 9 histograms of the imaginary part of the effective action are shown for both $T = 0$ and $T = 0.5$. It is clear that for $T = 0.5$ the fields sampled are concentrated around the preimage of a few (five) critical points, while with $T = 0$ (no flow) the distribution is broader. Consequently, the values of the phase $\exp(-i\mathrm{Im}S)$ fluctuate less when there is flow and the sign problem is minimized. On the other hand, for a large enough flow time, the probability distribution $\exp(-\mathrm{Re}S)$ becomes multimodal and the trapping of Monte Carlo chains can prevent proper sampling; see Fig. 11. Thus, the generalized thimble method (GTM) trades the sign problem for the problem of sampling a multimodal distribution. This trade is not without profit: in many cases one can find values of T such that the sign problem is sufficiently alleviated but trapping has not set in yet. These values of T can be determined by trial and error. As T is increased trapping occurs quite suddenly, and it

is not difficult to detect it by noticing a jump on the values of the observables. In addition, there are well studied ways to deal with trapping, as explained in Sec. IV. Still, trapping is a source of concern in GTM calculations and other, more general techniques have been developed to avoid it; see Sec. IV).

3. Case study: (1 + 1)D Thirring model

The lessons learned in applying the generalized thimble method to the $(0+1)$-dimensional Thirring model carry over to the more interesting $(1+1)$-dimensional case. Extensive calculations on the finite density and temperature $(1+1)$-dimensional Thirring model with two flavors were made over a range of parameters in the strong coupling region (Alexandru, Başar, Bedaque, Ridgway, and Warrington, 2017) with both Wilson and staggered fermions. The thimble structure of the $1+1$ models is more complex than the $0+1$ case. Still, all critical points and thimbles present in the $(0+1)$-dimensional case have analogs in $1+1$ dimensions [which indicates that many others are without a $(1+1)$-dimensional analog]. It is still true that the critical point closest to the real space (the main critical point) is a constant shift of $A_0(x)$ by an imaginary amount and that its tangent space is just a translation of $\mathbb{R}^N$ by an imaginary amount; see Fig. 7. The path integration over $\mathbb{R}^N$ has a bad sign problem for all values of the chemical potential larger than the fermion renormalized mass ($\mu > m_f$), that is, for all values of μ for which there is an appreciable number of fermion-antifermion unbalance.[16] The integration over the tangent space of the main thimble can be accomplished at no extra cost by simply shifting the variables of integration by a constant imaginary amount. This step by itself improves the sign problem considerably. The reason is that the tangent space is a rough approximation of the main thimble, especially the region near the critical point that dominates the path integral in the semiclassical regime. Still, for larger volumes, smaller temperatures, and higher chemical potential, the shift to the tangent space is not enough to control the sign fluctuation. It was determined that flow times of the order of $T = 0.4$ are sufficient to drastically reduce the sign fluctuation and, at the same time, not cause problems with the trapping and ergodicity of the Monte Carlo chain. Some of the results are summarized in Fig. 10. Alexandru, Başar, Bedaque, Ridgway, and Warrington (2017) demonstrated that the same method also works well as the continuum and thermodynamic limits are approached.

C. Trapping and tempered algorithms

The landscape induced by $\exp(-\mathrm{Re}S_{\mathrm{eff}})$ on the parametrization manifold changes as a function of the flow time T. For small T the landscape is typically flat, while for larger T the landscape is steeper. When the sign problem is severe enough to require large flow times, the landscape of $\exp(-\mathrm{Re}S_{\mathrm{eff}})$ has

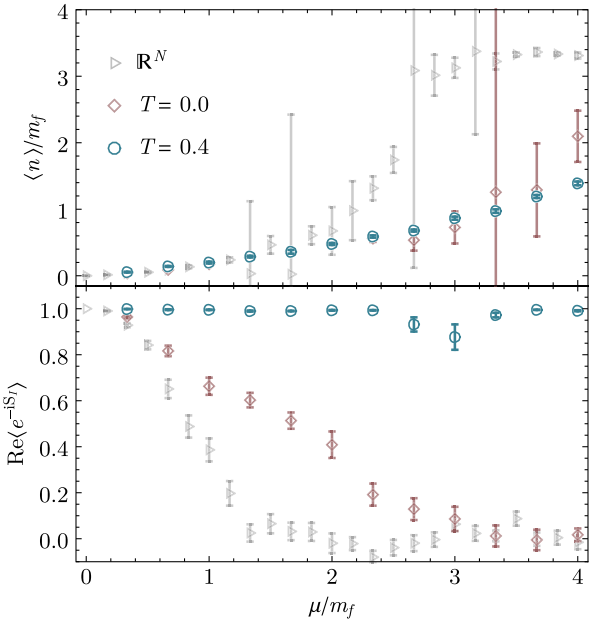


FIG. 10. Fermion density (top panel) and average sign (bottom panel) of the $(1+1)$-dimensional Thirring model on a 10×10 lattice, $g = 1$, and $m = -0.25$ (lattice units) with Wilson fermions. The sign problem is strongly suppressed, and one moves the path integration from $\mathbb{R}^N$ to the tangent plane $\mathbb{T}$ and from that to the flowed manifold $\mathcal{F}_T(\mathbb{T})$, thus allowing for precise measurements of the density and other observables (Alexandru, Başar, Bedaque, Ridgway, and Warrington, 2017).

high peaks and low valleys and the probability distribution can become multimodal. The purpose of this section is to detail several algorithms addressing this difficulty.

We first discuss the method of *tempered transitions* (Neal, 1996). Designed to combat trapping, a tempered proposal is a composite proposal assembled from small steps that, taken together, more rapidly cover phase space than a standard proposal does. A tempered proposal is constructed as follows. First, let $p_0(\phi), p_1(\phi), \ldots, p_n(\phi)$ be a sequence of increasingly relaxed probability distributions such that $p_0(\phi) \equiv p(\phi)$ is the distribution of interest and $p_n(\phi)$ is significantly more uniform. Next, for every i let $\hat{T}_i$ be a transition probability satisfying the detailed balance with respect to p_i, that is,

$$p_i(\phi)\hat{T}_i(\phi \to \phi') = p_i(\phi')\hat{T}_i(\phi' \to \phi). \tag{3.19}$$

A tempered update $\hat{T}$ is then executed by first generating a sequence of $2n$ configurations

$$\phi_0 \to \phi_1 \to \cdots \to \phi_n \equiv \phi_n' \to \phi_{n-1}' \to \cdots \to \phi_0' \tag{3.20}$$

using the transition probabilities $\hat{T}_1, \hat{T}_2, \ldots, \hat{T}_n, \hat{T}_n, \ldots, \hat{T}_1$, followed by the following accept-reject step with a probability:

$$P_{\mathrm{acc}}(\phi_0 \to \cdots \to \phi_0') = \min\{1, F(\phi)/F(\phi')\}, \tag{3.21}$$

where

[16]We note here that, contrary to single thimble calculations (Tanizaki, Hidaka, and Hayata, 2016), the generalized thimble method reproduces the "silver blaze" phenomena, the fact that the system is trivial at small temperatures and chemical potentials smaller than the mass of the lightest fermionic excitation (Cohen, 2003).

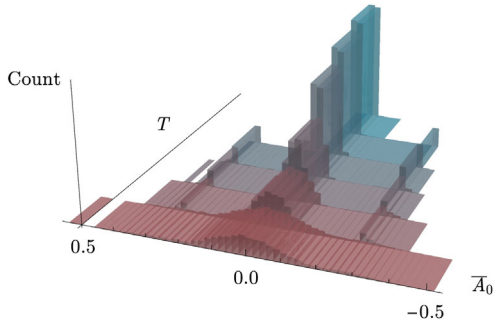


FIG. 11. The evolution of field space sampled in the $(0+1)$-dimensional Thirring model as a function of flow time. At small flow times the distribution is relatively uniform and much of phase space is sampled. The distribution sharpens as the flow time increases, and at sufficiently large flow times, the shoulder thimbles centered around ± 0.3 cease to be sampled.

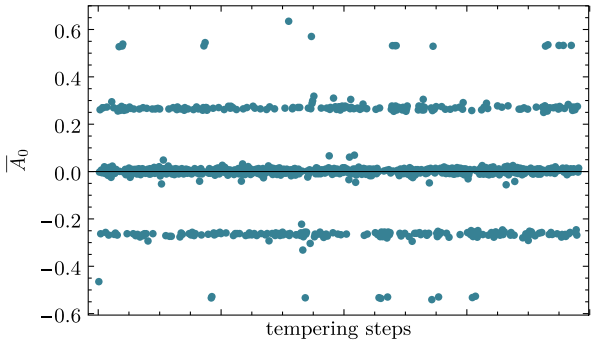


FIG. 12. $\bar{A}_0 = N_t^{-1} \sum_t A_{0,t}$ is plotted after each tempered transition. The five heavily visited positions in field space correspond to five thimbles contributing to the path integral. Compare this distribution to the sharpest distribution in Fig. 11, where only one thimble is sampled.

$$F(\phi) \equiv \frac{p_1(\phi_0)}{p_0(\phi_0)} \frac{p_2(\phi_1)}{p_1(\phi_1)} \cdots \frac{p_{n-1}(\phi_{n-2})}{p_{n-2}(\phi_{n-2})} \frac{p_n(\phi_{n-1})}{p_{n-1}(\phi_{n-1})}. \tag{3.22}$$

What is gained by using tempered proposals is enhanced ergodicity. Since the distributions p_i are increasingly uniform, the corresponding transition probabilities $\hat{T}_i$ may grow in support without decreasing the acceptance probability. To apply this general framework to simulations trapped by holomorphic gradient flow, suppose the flow time T is large enough that the probability distribution of interest

$$p(\zeta) = p_0(\zeta) = \frac{e^{-\mathrm{Re}S_{\mathrm{eff}}(\zeta)}}{Z} \tag{3.23}$$

is multimodal. Consider a sequence of flow times $T_0 < T_1 < \cdots < T_n$ such that $T_0 = T$ and $T_n \ll T_0$. This defines a sequence of probability distributions $p_0(\zeta), p_1(\zeta), \ldots, p_n(\zeta)$ that are decreasingly multimodal; we use this sequence to perform tempered proposals.

When this method was applied to the $(0+1)$-dimensional Thirring model at finite density (Alexandru, Başar, Bedaque, and Warrington, 2017), severely trapped simulations were liberated. Certain thermodynamic parameters exist for which at least five thimbles contribute non-negligibly to the path integral. Trapping to a single thimble, however, can become arbitrarily severe: for example, at $T = 0.5$, the multimodality of $p_0(\zeta)$ is so severe that over the course of a Metropolis chain with 10^7 steps not a single transition occurred. Tempered proposals free these trapped MCs, however; this is demonstrated in Fig. 12, where a proper sampling of the $T = 0.5$ probability distribution is achieved. In this case, five separate thimbles are sampled over the course of 2000 tempered proposals. Even though tempered proposals cost more than standard proposals, the improvement in ergodicity renders the added effort worthwhile.

A similar method, *parallel tempering*, was proposed to help sample from such multimodal distributions (Swendsen and Wang, 1986; Geyer, 1991; Earl and Deem, 2005). Parallel tempering involves simulating n replicas of the system of interest, each having a particular value of the tempering parameter. Each stream evolves separately and swaps between replicas are added satisfying the detailed balance. The swapping of configurations between adjacent replicas leads to enhanced ergodicity relative to the single chain case. Fukuma and Umeda (2017) and Fukuma, Matsumoto, and Umeda (2019b) developed the tempered Lefschetz thimble method (TLTM), an application of parallel tempering to multimodal distributions generated by flow.[17] As with tempered transitions, in this method the flow time is chosen as a tempering parameter. The TLTM method has been successfully applied to the $(0+1)$-dimensional Thirring model (Fukuma and Umeda, 2017), where trapping due to flow times as large as $T = 2.0$ have been solved.[18] Fukuma, Matsumoto, and Umeda (2018) also studied how to pick the flow times optimally and devised a geometric method for this optimization. More recently the TLTM was applied to the Hubbard model away from half filling on small lattices (Fukuma, Matsumoto, and Umeda, 2019a).

D. Algorithms for the Jacobian

The most computationally expensive part of many algorithms involving deformation of contours in field space (such as the contraction method or the generalized thimble method) is the calculation of the Jacobian J related to the parametrization of the manifold of integration. For bosonic systems where the Hessian can be computed efficiently, the calculation time is dominated by the matrix multiplication in the flow equation and its computation complexity is $\mathcal{O}(N^3)$, where N is proportional to the spacetime volume of the theory. The calculation of $\det J$ has similar computational complexity. This prohibitive cost prevents the study of all but the smallest models.

There are ways to bypass this large cost. Cristoforetti *et al.* (2014) introduced the following stochastic estimator to compute the phase: $\Phi(\phi_n) = \arg\det J(\phi_n)$. The main idea

[17] A modified TLTM method, called the "world volume hybrid Monte Carlo method," was recently proposed (Fukuma and Matsumoto, 2021; Fukuma, Matsumoto, and Namekawa, 2021).

[18] Because the thermodynamic parameters used by Fukuma and Umeda (2017) do not match those used by Alexandru, Başar, Bedaque, and Warrington (2017), it is currently not possible to compare the efficacy of tempered transitions and the TLTM. A comparison would, however, be useful.

stems from the observation that the Jacobian can be expressed as $J = UR$ for some unitary matrix U and some real, upper-triangular matrix R, a property that follows from the fact that $J^\dagger J \in \mathbb{R}$; therefore, $\arg\det J = \arg\det U$. Note that since J and U are related by a real matrix, this corresponds to a change in basis in the tangent plane, so the columns of U form a basis of the tangent space too: an orthonormal basis. Moreover, U satisfies $d\log\det U(t)/dt = -i\mathrm{Im}\mathrm{Tr}[U^T(t)H(t)U(t)]$. The trace can be estimated stochastically by using random vectors $\xi \in \mathbb{R}^N$, with $\langle \xi_i \xi_j \rangle = \delta_{ij}$, where the average is taken over the random source. If we generate N_R vectors we have

$$\mathrm{Tr}[U^T(t)H(t)U(t)] \approx \frac{1}{N_R}\sum_{r=1}^{N_R}(\xi^{(r)})^T U^T(t)H(t)U(t)\xi^{(r)}. \tag{3.24}$$

Now $\eta^{(r)} = U\xi^{(r)}$ is a random vector in the tangent plane that is isotropically distributed, and its length with respect to the real Euclidean metric satisfies $\langle\!\langle \eta^{(r)}|\eta^{(r)}\rangle\!\rangle = N$. We can generate such vectors without computing U: we generate a random vector $\tilde{\eta}$ isotropically in $\mathbb{C}^N$ with $\langle \tilde{\eta}^\dagger \tilde{\eta}\rangle = 2N$, for example, by using a Gaussian distribution $P(\tilde{\eta}) \propto \exp(-\tilde{\eta}^\dagger\tilde{\eta}/2)$, and then project it to the tangent space $\eta = P_\parallel(\phi_f)\tilde{\eta}$ using the same procedure presented when we discussed the HMC algorithm. Using $i\Phi(t) = \log\det U(t)$, one can then estimate the phase from

$$\Phi(T) \approx \Phi(0) - \mathrm{Im}\int_0^T dt \frac{1}{N_R}\sum_{r=1}^{N_R}\mathrm{Tr}[\eta^{(r)}(t)^T H(t)\eta^{(r)}(t)], \tag{3.25}$$

whose computational cost scales as $\mathcal{O}(N \times N_R)$. By comparing this stochastic estimation algorithm by explicit computation for a complex ϕ^4 theory, Cristoforetti *et al.* (2014) presented numerical evidence that this algorithm indeed provides a nontrivial speedup for the computation of the residual phase in relatively large systems. However, its applicability is limited to the phase of the Jacobian; the GTM also requires the magnitude.

For methods that require the Jacobian, we can substitute computationally cheap *estimators* for them. The idea is to use the estimators during the generation of configurations and correct for the difference when computing the observables. Two estimators for $\log\det J$ were introduced by Alexandru, Başar, Bedaque, Ridgway, and Warrington (2016b). They are given by

$$W_1 = \int_0^T dt \sum_i \rho^{(i)\dagger}\overline{H(t)\rho^{(i)}}, \qquad W_2 = \int_0^T dt\,\mathrm{Tr}\bar{H}(\tau), \tag{3.26}$$

where $\rho^{(i)}$ are the Takagi vectors of $H_{ij}(0)$ with positive eigenvalues. The first estimator W_1 is equal to $\log\det J$ for quadratic actions. The second estimator is equal to $\ln\det J$ when the Jacobian is real along the flow. As such, it is expected to be a good estimator for Jacobians that are mostly real. The bias introduced by the use of estimators instead of the Jacobian is corrected by reweighting the difference between them when computing observables with the help of

$$\langle\mathcal{O}\rangle = \frac{\langle \mathcal{O}e^{-\Delta S}\rangle_{\mathrm{Re}S'_{\mathrm{eff}}}}{\langle e^{-\Delta S}\rangle_{\mathrm{Re}S'_{\mathrm{eff}}}}, \tag{3.27}$$

where $S'_{\mathrm{eff}} = S - W_{1,2}$ and $\Delta S = S_{\mathrm{eff}} - \mathrm{Re}S'_{\mathrm{eff}}$. The estimator is useful when ΔS has small fluctuations over the sampled field configurations, that is, if $W_{1,2}$ "tracks" $\log\det J$ well.

For theories where the Hessian can be computed efficiently, such as bosonic theories with local actions, the W_1 estimator has a computational cost of $\mathcal{O}(N^2)$ and W_2 has $\mathcal{O}(N)$ complexity, a significant improvement over $\mathcal{O}(N^3)$ for the full Jacobian. To use Eq. (3.27) the correct Jacobian J needs to be computed. This has to be done, however, only on field configurations used in the average in Eq. (1.2). Typically, configurations obtained in subsequent Monte Carlo steps are correlated and only one configuration out of tens or hundreds of steps is used in Eq. (3.27). The idea is then to use the cheaper Jacobian estimators like W_1 and W_2 during the collection of configurations and to compute the expensive Jacobian J only when making measurements, which cheapens the calculation by orders of magnitude. This strategy was used in the ϕ^4 model in $3+1$ dimensions (Alexandru, Başar, Bedaque, Ridgway, and Warrington, 2016a) and the Thirring model in $1+1$ dimensions (Alexandru, Başar, Bedaque, Ridgway, and Warrington, 2017), both times at finite density. However, for other classes of problems, such as real-time systems, the estimators W_1 and W_2 do not provide a significant improvement.

A more robust algorithm for the Jacobian was introduced by Alexandru, Başar, Bedaque, and Ridgway (2017). The key idea is to modify the proposal mechanism in such a way as to incorporate the Jacobian as part of the effective action. As an added bonus, the procedure leads to isotropic proposals on the integration manifold. As in the contraction algorithm, the goal is to generate a distribution on the parametrization manifold with a probability proportional to $\exp[-\mathrm{Re}S_{\mathrm{eff}}(\phi_n)]$. This is a Metropolis method, so we need to make a proposal and then accept or reject it. For update proposals, we generate a random complex vector in the tangent plane at ϕ_f, uniformly distributed with normal distribution $P(\eta) \propto \exp(-\eta^\dagger\eta/\delta^2)$. The parameter δ controls the step size and is tuned to optimize the acceptance rate. The vector η is generated using the projection discussed earlier: a $\tilde{\eta} \in \mathbb{C}^N$ sampled from a Gaussian distribution and then $\eta = P_\parallel(\phi_f)\tilde{\eta}$ using the vector flow projection. The update in the parametrization space is $\phi'_n = \phi_n + \epsilon$, where $\epsilon = J^{-1}(\phi_n)\eta$ is a vector in the tangent space at ϕ_n. Here we take advantage of the fact that the parametrization space is flat and ϕ'_n does not need to be projected.

Since the proposals are not symmetric, the accept-reject step has to be slightly modified to satisfy the detailed balance. The added factor does not cancel the Jacobian, unless the proposal satisfies an implicit equation that is not easy to solve. A better alternative is based an algorithm by Grady (1985): the ratio of Jacobians is taken into account implicitly using a

stochastic generated vector. The vector is generated with probability $P(\xi) \propto \exp[-\xi^\dagger (J'^\dagger J')\xi]$, where $J' = J(\phi'_n)$ and the proposal is accepted with probability (Alexandru, Başar, Bedaque, and Ridgway, 2017)

$$P_{\text{acc}} = \min\{1, e^{-\text{Re}[S'-S]+\xi^\dagger \Delta J \xi - \epsilon^\dagger \Delta J \epsilon}\}, \tag{3.28}$$

where $\Delta J = (J'^\dagger J') - (J^\dagger J)$. We stress that ξ is a complex random vector with $2N$ independent components, whereas ϵ has only N independent components.

The highlight of this method is that by construction it samples the probability distribution $e^{-\text{Re}S}|\det J|$ without an explicit computation of $|\det J|$. It requires only the computation of $J^{-1}\eta$ and $J\epsilon$, both of which scale as $\mathcal{O}(N)$ for most bosonic theories.

A simplified algorithm that may lead to further computational speedup can be achieved when instead of $J(\phi_n)$ we approximate it with $J(\phi^c)$. The Jacobian is then required only to compute the displacements ϵ and the accept or reject is done simply based on the change of the action since $\Delta J = 0$. For this method $J(\phi^c)^{-1}$ can be computed once at the start of the simulation. The difference between $J(\phi_n)$ and $J(\phi^c)$ has to be included by reweighting the observables as was done with $W_{1,2}$. This method should work well when the fluctuations of $J(\phi_n)$ are mild. Alexandru, Başar, Bedaque, and Ridgway (2017) showed this to be the case for the real-time study of a $(1+1)$-dimensional ϕ^4 theory even in the strongly coupled regime.

1. Case study: Real-time field theory

The generalized thimble method and the entire machinery used to deal with the computational cost of the Jacobian was applied to one of the most challenging sign problems: the calculation of real-time correlators in field theory. These correlators are the building blocks for the computation of transport coefficients like diffusivity, conductivity, and viscosities and are of great importance in a variety of physical contexts. Similar methods can also be used in fully nonequilibrium situations. At the same time the available theoretical tools to study this problem are limited. Even perturbation theory requires complicated resummations (Braaten and Pisarski, 1990; Jeon and Yaffe, 1996), and in the strongly coupled regime the conventional lattice methods are not applicable, as later detailed. Alternatively, stochastic quantization (or the complex Langevin method) has been utilized but it seems to converge to the wrong result if the time separation $t - t'$ is more than the inverse temperature β (Berges *et al.*, 2007).

The central objects of interest here are *time dependent* correlation functions of the form

$$\langle \mathcal{O}_1(t)\mathcal{O}_2(t')\rangle_\beta = \text{Tr}[\hat{\rho}\mathcal{O}_1(t)\mathcal{O}_2(t')], \tag{3.29}$$

where $\hat{\rho}$ is the density matrix that reduces to the familiar Boltzmann factor $e^{-\beta H}/\text{Tr}(e^{-\beta H})$ in equilibrium. Time dependent correlation functions can be generated from the Schwinger-Keldysh (SK) path integral (Schwinger, 1961; Keldysh, 1964) as follows:

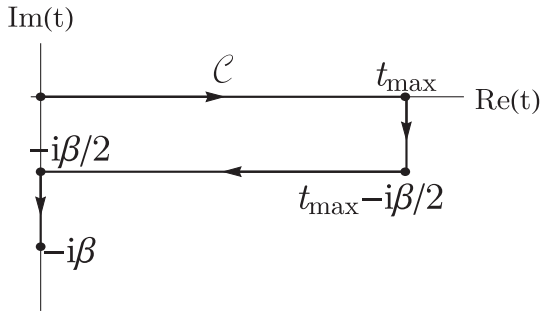


FIG. 13. The Schwinger-Keldysh contour in a complex time plane. The real part corresponds to forward and backward time evolution, and the imaginary part corresponds to the insertion of the equilibrium density matrix.

$$\begin{aligned}\langle \mathcal{O}_1(t)\mathcal{O}_2(t')\rangle_\beta &= \text{Tr}[\mathcal{O}_1(0)e^{-iH(t-t')}\mathcal{O}_2(0)e^{iH[i\beta/2+(t-t')+i\beta/2]}] \\ &= \frac{1}{Z}\int \mathcal{D}\phi\, e^{iS_{\text{SK}}[\phi]}\mathcal{O}_1(t)\mathcal{O}_2(t'),\end{aligned} \tag{3.30}$$

where the SK action is obtained by integrating the Lagrangian over a complex contour, as shown in Fig. 13. The real part corresponds to forward and backward time evolution and the imaginary part corresponds to the insertion of the equilibrium density matrix $e^{-\beta\hat{H}}/\text{Tr}e^{-\beta\hat{H}}$. For instance, a discretized Schwinger-Keldysh action for a scalar theory reads[19]

$$\begin{aligned}S(\phi) &= \sum_{t,\vec{x}} a_t a\left[\frac{1}{2}\frac{(\phi_{t+1,\vec{x}}-\phi_{t,\vec{x}})^2}{a_t^2}\right. \\ &\quad \left. + \frac{1}{2}\sum_{\hat{i}}\frac{(\phi_{t,\vec{x}+\hat{i}}-\phi_{t,\vec{x}})^2}{a^2} + V(\phi_{t,\vec{x}})\right], \\ a_t &= \begin{cases} ia \text{ for } 0 \le t < N_t, \\ -ia \text{ for } N_t \le t < 2N_t, \\ a \text{ for } 2N_t \le t < 2N_t + N_\beta, \end{cases}\end{aligned} \tag{3.31}$$

from which the following correlators result:

$$\langle \phi_{t_1,\vec{x}_1}\phi_{t_2,\vec{x}_2}\rangle = \frac{\int(\prod_{t,\vec{x}} d\phi_{t,\vec{x}})e^{-S(\phi)}\phi_{t_1,\vec{x}_1}\phi_{t_2,\vec{x}_2}}{\int(\prod_{t,\vec{x}} d\phi_{t,\vec{x}})e^{-S(\phi)}}. \tag{3.32}$$

The Boltzmann weight of the Minkowski part of the SK contour ($0 \le t < 2N_t$) is purely imaginary, as expected from the real-time evolution, and leads to a severe sign problem. In fact, owing to the fact of its pure phase with no damping term, it is impossible to define a phase quenched measure and reweigh the phase. For this reason, conventional lattice methods do not work for real-time problems even if

[19]For simplicity, we consider the bosonic case but the formalism can be generalized to the fermionic case in a straightforward fashion. We also include an overall factor i so that the associated Boltzmann weight is e^{-S}.

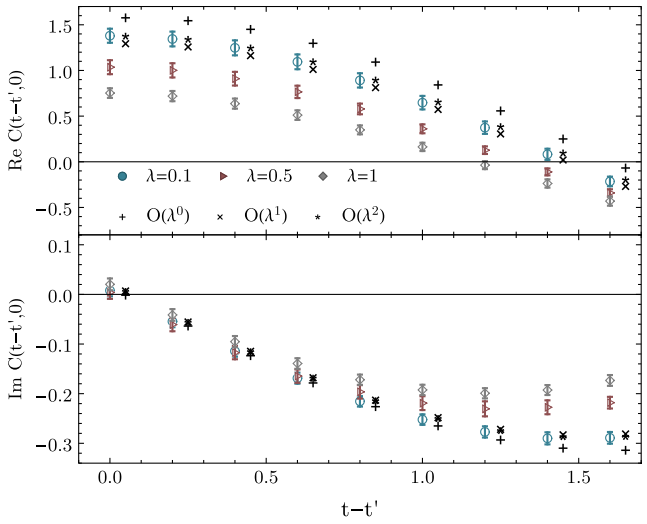


FIG. 14. Monte Carlo computation of the time order correlation function defined in Eq. (3.33) for $p = 0$ and $\lambda = 0.1, 0.5, 1$. The $\lambda = 0.1$ result is compared to the analytical perturbation theory calculations at $\mathcal{O}(\lambda^0)$, $\mathcal{O}(\lambda^1)$, and $\mathcal{O}(\lambda^2)$, which are offset along the x axis for visual clarity. The $\lambda = 0.5, 1.0$ results are outside of the validity of perturbation theory (see Fig. 16) and the perturbation theory calculations for these values are not shown here.

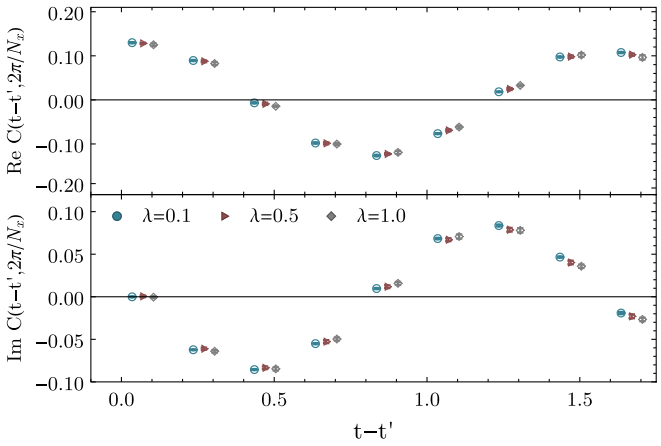


FIG. 15. Monte Carlo computation of the time order correlation function defined in Eq. (3.33) for $p = 2\pi/N_x$ and $\lambda = 0.1, 0.5, 1$.

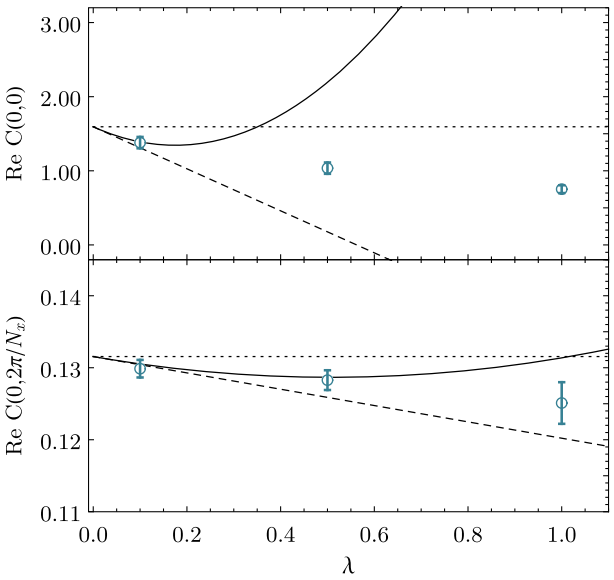


FIG. 16. Comparison between the perturbative calculation and the Monte Carlo result.The dotted, dashed, and solid lines denote the $\mathcal{O}(\lambda^0)$, $\mathcal{O}(\lambda^1)$, and $\mathcal{O}(\lambda^2)$ calculations, respectively.

unlimited computational power is available.[20] By contrast, on any manifold $\mathcal{M}$ that is obtained by flowing from $\mathbb{R}^N$ by some fixed flow time, $\mathrm{Re}S > 0$, and the action provides a damping factor making the real-time path integral well defined. The generalized thimble method has been successful in computing time dependent correlation functions in $(0+1)$-dimensional (quantum mechanics) (Alexandru, Başar, Bedaque, Vartak, and Warrington, 2016; Mou, Saffin, and Tranberg, 2019) and $(1+1)$-dimensional bosonic field theories with $V(\phi) = \lambda\phi^4/4!$ potential. In Figs. 14 and 15 the two lowest spatial Fourier modes of the time-ordered correlator

$$C(t-t',p) = T\langle\phi(t,p)\phi(t',p)^\dagger\rangle_\beta, \tag{3.33}$$

where

$$\phi(t,p) = \frac{1}{N_x}\sum_{x=0}^{N_x-1} e^{ipx}\phi_{tx} \tag{3.34}$$

are plotted for different values of λ (Alexandru, Başar, Bedaque, and Ridgway, 2017). To ensure the validity of the method, the weak coupling ($\lambda = 0.1$) Monte Carlo result is compared with the analytically performed zeroth, first, and second order lattice perturbation theory calculations. In the strong coupling regime that lies outside of the domain of perturbation theory (see Fig. 16) the method works as well as it does in the weak coupling regime without any problems. In the quantum mechanical case a similar cross-check was performed that showed agreement between the Monte Carlo results and the exact result obtained by numerically solving the Schrödinger equation (Alexandru, Başar, Bedaque, Vartak, and Warrington, 2016). Mou, Saffin, and Tranberg (2019) and Mou *et al.* (2019) studied the $(1+1)$-dimensional model with a nonequilibrium density matrix.

The sign problem in the real-time problem gets more severe when the time interval between the operators $|t-t'|$ is increased in units of inverse temperature. This is because the real part of $\mathcal{C}$ that generates the pure phase contribution to the path integral becomes larger. Therefore, a larger flow time is needed to handle a larger $|t-t'|$. Recently, with the help of the algorithms for the previously described Jacobian, $|t-t'| = 4\beta$ was achieved on a lattice with $N_t = 8$, $N_\beta = 2$, and $N_x = 8$. Extensions to larger time separations seem to be hindered by trapping in a local minima in the Monte Carlo

[20] In principle, it is possible to extract the real-time correlator (3.30) from a purely Euclidean time correlator by analytic continuation. The extrapolation is, however, numerically unstable and requires exponentially accurate precision in Euclidean time.

(Metropolis-Hastings) evolution, which calls for alternative sampling methods to be utilized. For alternate methods to tame real-time sign problems, see Kanwar and Wagman (2021) and Lawrence and Yamauchi (2021).

E. Gauge theories

The thimble structure of gauge theories is more complicated due to the fact that critical points are not isolated points but instead continuous manifolds formed as gauge orbits. The thimbles attached to the critical points carry the same degeneracy due to the gauge symmetry. One might envision instead of working with the degenerate field space $\mathcal{M}$ fixing the gauge and working with the quotient space $\mathcal{M}/\mathcal{G}$, where the critical points would be isolated and the Picard-Lefschetz theory can be used as usual. As we later discuss, this is possible for Abelian gauge fields, but not for non-Abelian gauge fields. The reason for this is that some critical points have nontrivial stabilizers and become singular points on $\mathcal{M}/\mathcal{G}$.[21] In this case, Picard-Lefschetz theory has to be modified to accommodate these complications, which was discussed in the context of Chern-Simons theory by Witten (2011).

Lattice gauge theory remains largely unexplored from the perspective of Picard-Lefschetz theory at the moment. We review a few exploratory examples from the literature in Secs. III.E.1 and III.E.2, studies that focused on simulating small lattice systems. We note that QCD inspired models were also studied: a $0+1$ SU(3) system that reduces to a one-link model (Di Renzo and Eruzzi, 2018) and a pure-gauge U(1) model at imaginary coupling (Pawlowski *et al.*, 2020, 2021). Before discussing the examples in detail, we discuss some generalities. In lattice gauge theory the fundamental degrees of freedom are gauge links U_i, where $i \equiv (x,\mu)$ is the collective index for the link variable $U_{x\mu} \equiv U(x+\hat{\mu},x) \equiv U_i$. The derivative with respect to the link variable is defined as

$$\mathcal{D}_i^a f(U) \equiv \frac{\partial}{\partial t} f(e^{itT^a} U_i)|_{t=0}. \tag{3.35}$$

As usual, we consider the complexification of the Lie group where the link variables can be parametrized as $U = e^{i\xi_a T^a}$, where T^a are the group generators and ξ_a are complex variables. For example, the complexification of SU(N) leads to SL(N). The holomorphic flow equation reads

$$\frac{dU_i}{d\tau} = i\sum_a [T^a \overline{\mathcal{D}_i^a S(U)}] U_i, \tag{3.36}$$

and it satisfies the desired properties $d\mathrm{Re}S(U)/dt = |\mathcal{D}_i^a S(U)|^2 > 0$ and $d\mathrm{Im}S(U)/dt = 0$. Unlike ordinary derivatives, the group derivatives do not commute:

$$[\mathcal{D}_i^a, \mathcal{D}_j^b] = -f^{abc}\delta_{ij}\mathcal{D}_j^c, \qquad [\bar{\mathcal{D}}_i^a, \bar{\mathcal{D}}_j^b] = -f^{abc}\delta_{ij}\bar{\mathcal{D}}_j^c,$$
$$[\mathcal{D}_i^a, \bar{\mathcal{D}}_j^b] = 0, \tag{3.37}$$

where f^{abc} are the structure constants of the gauge group such that $[T^a, T^b] = if^{abc}T^c$. Therefore, the flow equation for the tangent space generated by $e_i^a = \overline{\mathcal{D}_i^a S}$ is modified as

$$\frac{de_i^a}{d\tau} = \overline{e_j^b \mathcal{D}_j^b \mathcal{D}_i^a S} - f^{abc} e_i^b \overline{\mathcal{D}_i^a S}. \tag{3.38}$$

1. Case study: Heavy-dense QCD

Some exploratory work toward implementing the thimble method in QCD has been done within the so-called heavy-dense QCD, which is QCD with heavy quarks in the high-density limit (Zambello and Di Renzo, 2019). We present the details later but note here that the effective action for this model is not actually a gauge action, but rather a spin model. As opposed to the heavy mass limit where the quarks decouple from the theory, in the simultaneous high-mass, high-density limit

$$m_0 \to \infty, \qquad \mu \to \infty, \qquad e^\mu/m_0 : \text{fixed}, \tag{3.39}$$

the quarks remain in the picture and the theory has a nontrivial phase structure controlled by μ. Just like QCD, heavy-dense QCD exhibits a sign problem. At the same time it is not as computationally demanding as full QCD, which makes it a fruitful arena for testing new approaches to the sign problem (Aarts *et al.*, 2016; Zambello and Di Renzo, 2019). In this limit the fermion determinant simplifies dramatically as (Bender *et al.*, 1992; Blum, Hetrick, and Toussaint, 1996)

$$\det D_f \to \prod_{\vec{x}} \det(1+\gamma P_{\vec{x}})^2 \det(1+\tilde{\gamma} P_{\vec{x}}^{-1})^2, \tag{3.40}$$

where $\gamma \equiv (2e^{\mu}/m_0)^{N_t}$, $\tilde{\gamma} \equiv (2e^{-\mu}/m_0)^{N_t}$, and $P_{\vec{x}} = \prod_{t=0}^{N_t-1} U_0(\vec{x},t)$ is the Polyakov loop. Equation (3.40) has a simple physical interpretation: in the infinite mass limit, quarks are pinned to their spacial location and do not move. Therefore, a quark (antiquark) at a spatial point $\vec{x}$ is simply described by the Polyakov loop $P_{\vec{x}}$ ($P_{\vec{x}}^{-1}$). Furthermore, owing to the high-density limit, the antiquark contribution is negligible (i.e., $\gamma \gg \tilde{\gamma}$), and one can neglect the second determinant on the right-hand side of Eq. (3.40). Since the fermion determinant has no dependence on the spatial links ($U_{\mu\neq 0}$), one can obtain the effective action for heavy-dense QCD by integrating out the spatial degrees of freedom in the QCD path integral as follows:

$$Z_{\mathrm{QCD}} = \int DU_\mu e^{(\beta/2N_c)\sum_p (\mathrm{Tr}U_p + \mathrm{Tr}U_p^\dagger)} \prod_{f=1}^{N_f} \det D_f$$
$$\to \int DU_0 e^{-S_{\mathrm{HD}}(U_0)}, \tag{3.41}$$

where we assume that all N_f quarks are heavy and have identical chemical potentials for simplicity. The effective action for the heavy-dense QCD in this case is

$$S_{\mathrm{HD}} \equiv S_{\mathrm{gauge}} - 2N_f \sum_{\vec{x}} \log[\det(1+\gamma P_{\vec{x}})],$$
$$S_{\mathrm{gauge}} \approx -\left(\frac{\beta}{18}\right)^{N_t} \sum_{\langle \vec{x}\vec{y}\rangle} (\mathrm{Tr}P_{\vec{x}}\mathrm{Tr}P_{\vec{y}}^{-1} + \mathrm{Tr}P_{\vec{y}}\mathrm{Tr}P_{\vec{x}}^{-1}), \tag{3.42}$$

[21] For example, the zero field configuration is such a critical point stabilized by the entire gauge group.

where $\langle\cdot\rangle$ denotes the nearest neighbors. The leading order pure-gauge action, including the coefficient $(\beta/18)^{N_t}$, follows from the character expansion of the original gauge action (Langelage, Neuman, and Philipsen, 2014). In the low temperature limit where $N_t \gg 1$ and $\beta \sim \mathcal{O}(1)$, the pure-gauge contribution S_{gauge} can be further neglected and S_{HD} simply reduces to

$$S_{\text{HD}} \approx -2N_f \sum_{\vec{x}} \log\det(1+\gamma P_{\vec{x}}). \tag{3.43}$$

In particular, for $N_c = 3$ the determinant over the gauge group reduces to

$$\det(1+\gamma P_{\vec{x}}) = 1 + \gamma \mathrm{Tr} P_{\vec{x}} + \gamma^2 \mathrm{Tr} P_{\vec{x}}^{-1} + \gamma^3. \tag{3.44}$$

Higher order corrections to S_{HD} are given in powers of the hopping parameter $\kappa \equiv 2/m_0$ and were given by Zambello and Di Renzo (2019). Furthermore, Zambello and Di Renzo (2019) focused on $\mu \approx \mu_c \equiv m = -\log(2\kappa)$, where the nuclear phase transition occurs at zero temperature. It is possible and convenient to work in the temporal gauge, which eliminates all the links in $P_{\vec{x}}$ but one for a fixed $\vec{x}$ (say, $t = 0$) such that $P_{\vec{x}} = U_0(\vec{x}, t=0) \equiv U_{\vec{x}}$. The holomorphic gradient flow equation (3.36) in the temporal gauge reads

$$\frac{dU_{\vec{x}}}{dt} = i\sum_a (T^a \overline{\mathcal{D}^a S_{\text{HD}}[U]}) U_{\vec{x}} = -2\gamma \sum_a T^a \overline{\frac{\mathrm{Tr}(T^a U_{\vec{x}})}{\det(1+\gamma U_{\vec{x}})}} U_{\vec{x}}. \tag{3.45}$$

The critical points satisfy $\mathrm{Tr}(T^a U_x^{\text{cr}}) = 0$ and therefore are elements of the center:

$$U_{\vec{x}}^{\text{cr}} = e^{i\omega_{\vec{x}}}, \qquad \omega_x \in \left\{ \frac{2\pi n}{N_c} \middle| n = 0, \ldots, N_c - 1 \right\}. \tag{3.46}$$

Since ω_x can take one of these three values at each lattice site, the number of critical points exponentially grows with the volume as $(N_c)^V$. However, they contribute to the path integral with different weights. Zambello and Di Renzo (2019) studied this model with $N_c = 3$ in small spatial volumes up to $3^3 - 4^3$ and in a parameter range where only a few critical points, and hence thimbles, contribute significantly to the path integral. They estimated the critical points' semiclassical weights $e^{-S_{HD}[U^{cr}]}$ using importance sampling. Furthermore, they performed the Monte Carlo computations of the charge density $\langle n\rangle$ and the Polyakov loop $\langle P\rangle = (1/N)\sum_{\vec{x}} \mathrm{Tr}\langle U_{\vec{x}}\rangle$ over the thimbles with one and two lattice sites. The results show the expected behavior in the cold limit near $\mu = \mu_{\text{cr}}$, namely, $\langle n\rangle$ sharply changing[22] from 0 to 1 [i.e., the silver blaze behavior (Cohen, 2003)] and $\langle P\rangle$ having a narrow peak around μ_{cr}. Furthermore, the contribution of three thimbles is necessary to obtain this expected result.

[22]The saturation of the fermion density is due to the finite volume.

2. Case study: 2D QED

Another example of a gauge theory, two-dimensional QED with the lattice action

$$S = \frac{1}{g^2}\sum_r (1 - \cos P_r) - \sum_a \ln\det D^{(a)}, \tag{3.47}$$

with

$$D_{xy}^{(a)} = m_a \delta_{xy} + \frac{1}{2}\sum_{\nu\in\{0,1\}} [\eta_\nu e^{iQ_a A_\nu(x) + \mu\delta_{\nu 0}} \delta_{x+\hat{\nu},y} - \eta_\nu e^{-iQ_a A_\nu(x) - \mu\delta_{\nu 0}} \delta_{x,y+\hat{\nu}}] \tag{3.48}$$

was studied by Alexandru, Başar *et al.* (2018), who used the generalized thimble method. Here $D^{(a)}$ denotes the Kogut-Susskind fermionic matrix[23] for flavor a, and P_r denotes the plaquette

$$P_r \equiv A_1(r) + A_0(r+\hat{x}) - A_1(r+\hat{t}) - A_0(r), \tag{3.49}$$

and $\hat{t}$ and $\hat{x}$ are the unit vectors in the time and space directions. In the case of Abelian gauge theories, it is convenient to work with the complexified gauge field $A_\mu(x) \in \mathbb{C}^{2N}$, where N is the number of lattice sites, instead of the gauge links. This way degeneracies due to gauge redundancy can be addressed in a straightforward fashion. For any point x, the gauge orbit is generated using the gauge transformation $A_\mu(x) \to A_\mu(x) + \alpha(x+\hat{\mu}) - \alpha(x)$, which forms an $(N-1)$-dimensional[24] subspace. The original real field space can therefore be locally expressed as a direct product $\mathbb{R}^{2N} = \mathcal{M}_0 \times \mathcal{G}$, where $\mathcal{M}_0$ is the space of gauge inequivalent field configurations and $\mathcal{G}$ is the gauge orbit. The flow leaves $\mathcal{G}$ invariant because the gradient $\bar{\partial}S/\partial A$ is orthogonal to the gauge orbits and it commutes with the gauge transformations. Therefore, the middle-dimensional manifold (i.e., a manifold with a real dimension N) obtained by flowing $\mathbb{R}^{2N}$ by an amount T can be decomposed as $\mathcal{M}_T \times \mathcal{G}$, where $\mathcal{M}_T$ is the result of flowing $\mathcal{M}_0$ by T. Furthermore, the critical points on the gauge fixed slice are now isolated and the thimble decomposition follows straightforwardly as the limit $T \to \infty$. This argument illustrates conceptually how the generalized thimble method works in the presence of an Abelian gauge field. For the actual lattice computations, however, there is no need to fix the gauge, as the Markov chain will randomly sample $\mathcal{G}$, which has no effect on the results as long as only gauge invariant observables are evaluated.

Alexandru, Başar *et al.* (2018) put the generalized thimble method into action in a U(1) gauge theory with three fermion species with charges 1, 1, and −2. Their charges are chosen in

[23]To have neutral excitations a three flavor model with charges $(2,-1,-1)$ was studied by Alexandru, Başar *et al.* (2018). The two flavor model with equal charges has no sign problem due to charge conjugation symmetry.

[24]Note that $\alpha = \text{const}$ is not a gauge transformation.

such a way that a state with finite fermion density does not necessarily have a net charge, which would render the energy of the state infinite in the thermodynamics limit. It is shown that a computational speedup can be achieved by computing the equation of state compared to a conventional real space computation. Even though the cold limit, which shows the silver blaze behavior (on a 14×10 lattice), can be achieved this way, faster algorithms are needed for going to larger lattices. Finally, two-dimensional QED in the continuum was studied in the mean field approximation given by Tanizaki and Tachibana (2017), where the exact thimble decomposition was worked out for $N_f = 1, 2, 3$ and it was shown that the sign problem can be eliminated by deforming the path integral into the thimble decomposition.

IV. OTHER MANIFOLDS AND THE ALGORITHMS THAT CAN FIND THEM

A. Well beyond thimbles

We have seen that deforming the integration from $\mathbb{R}^N$ to a proper combination of thimbles is not always desirable from a numerical point of view. The generalized thimble method, for instance, uses a rough approximation of thimbles that, while having a smaller average sign, has better ergodic properties. There is no reason, however, to be limited to manifolds close to the thimbles. The condition that S_I is constant is only one constraint in a $2N$-dimensional space, and there presumably are many manifolds of integration where the sign of the integrand is fixed. In this section we consider a few methods to search for other manifolds unrelated to thimbles that both alleviate the sign problem and are numerically convenient.

An ideal manifold of integration would (i) ameliorate the sign problem significantly both because the action is nearly real on it and because the residual phase is small, (ii) be computationally cheap to find, and (iii) be parametrized in such a way that the associated Jacobian is also computationally cheap. These restrictions are hard to satisfy all at the same time, and only the first steps in this direction were taken. It seems that insight into particular models will be essential for exploiting this general idea profitably. We show next that, in cases where the thimble method generates some of this insight, it is not difficult to improve it by allowing for more general manifolds. In other theories it is an open problem to find a way to capitalize on the freedom of picking more general manifolds.

B. Learnifolds

Suppose a number of points on the thimble(s) (or some approximation of them) are obtained using the computationally costly holomorphic flow equations. It is reasonable to expect that the sign fluctuations on a manifold that interpolates between the original manifold $\mathbb{R}^N$ and the thimble will be small. A point found using the flow (in reality a complex field configuration or an element of $\mathbb{C}^N$) can be viewed as a map connecting its real part to the corresponding imaginary part; that is, for $\phi \in \mathcal{M}_T$ the map f takes $\mathrm{Re}\phi \to f(\mathrm{Re}\phi) = \mathrm{Im}\phi$. Here we assume that the manifold does not "fold"; i.e., every real value of the field corresponds to a unique imaginary part.

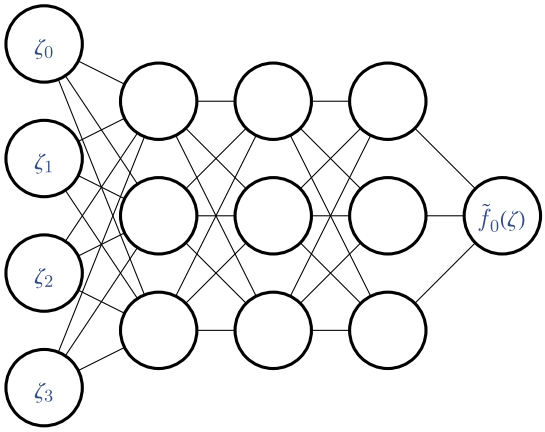


FIG. 17. Topology of a feed-forward network with five layers: one input layer with four nodes, three intermediate layers with three nodes each, and one output layer with one node. The inputs in the incoming layer (shown on the left) are the real values of the field. The output is the imaginary value of the coordinate of the first point of the lattice $\tilde{f}_0(\zeta)$.

We seek a manifold with a simple parametrization that "interpolates" the configurations sampled on $\mathcal{M}_T$. The result of the interpolation problem can be thought of as a map that approximates f, the map that connects the given real parts of coordinates to their imaginary parts. Thus, finding the interpolation of the points obtained by the flow can be formulated as learning a general rule from a set of examples. This is a typical problem studied by the artificial intelligence community, and we can borrow some of their techniques to bear on it.

More concretely, suppose that we have a "training set" $\mathcal{S}$ that is a number of field configurations ϕ_i^a lying on the manifold $\mathcal{M}_T$, where $i = 1, \ldots, N$ indexes their components and $a = 1, \ldots, |\mathcal{S}|$, with $|\mathcal{S}|$ the size of the training set. We parametrize the interpolating manifold $\mathcal{L}_\mathcal{S}$, an approximation to $\mathcal{M}_T$, as

$$\phi_i = \zeta_i + i\tilde{f}_i(\zeta), \tag{4.1}$$

where $\zeta_i \in \mathbb{R}^N$ and $\tilde{f}_i$ is a real function approximating f. The function $\tilde{f}_i$ is represented by a feed-forward network of the type depicted in Fig. 17. The nodes on the left layer represent the input values, in our case the values of ζ_i. The results are combined on the second layer by making linear combinations of them, adding a bias, and feeding them to a nonlinear function $\sigma(x)$ that we take to be of the form $\sigma(x) = \log(1 + e^x)$. The result is

$$v_j = \sigma\left(b_j + \sum_i w_{ij}\zeta_i\right), \tag{4.2}$$

where j indexes the nodes of the second layer. These results are then combined again, piped through $\sigma(x)$, and fed to the next layer. At the end all results are combined in a single number that represents $\tilde{f}_{i=0}(\zeta)$. By translation invariance, the values of $\tilde{f}_i(\zeta)$ for other $i \neq 0$ can be obtained by translating the inputs ζ_i. The feed-forward network is parametrized by the weights (w's) of every link and biases b's of every node.

These parameters are chosen as follows in order to minimize the discrepancy between the training set and the results of the network:

$$C(w,b) = \frac{1}{|\mathcal{S}|}\sum_{a=1}^{|\mathcal{S}|} |\tilde{f}_{w,b}(\mathrm{Re}\phi^a) - \mathrm{Im}\phi^a|, \tag{4.3}$$

where $\tilde{f}_{w,b}(\mathrm{Re}\phi^a)$ is the result of applying the network, with parameters w_{ij} and b_j, to $\mathrm{Re}\phi^a$.[25] To minimize $C(w,b)$ a gradient descent algorithm is used. The computation of the gradient is efficiently done using the back-propagation algorithm. In fact, the existence of this simple algorithm is an important motivation to use feed-forward networks, as opposed to networks with more complicated topologies. The minimization process is sped up tremendously by using the adaptive moment estimate algorithm (ADAM) (Kingma and Ba, 2014) [other methods were discussed by Ruder (2016)], another process borrowed from the artificial intelligence literature. Since the manifold $\mathcal{L}_\mathcal{S}$ is defined by a network that learned how to approximate $\mathcal{M}_T$, we call $\mathcal{L}_\mathcal{S}$ the "learnifold." An example of the practical use of this method is described in Sec. IV.B.1.

Three comments are worth making at this point. First, while the usefulness of this method can be gauged on a case-by-case basis, its correctness is guaranteed by construction. In fact, any network will define a manifold of integration in the same homology class as $\mathbb{R}^N$ since the mapping

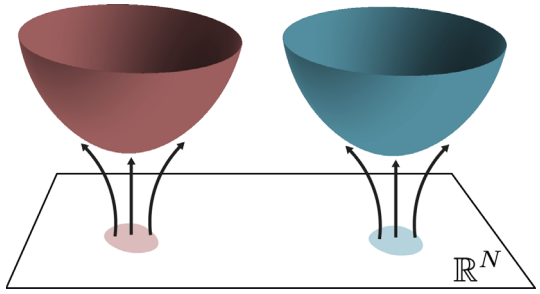
$$\phi_i(\zeta) = \zeta_i + is\tilde{f}_i(\zeta), \tag{4.4}$$

for $0 \le s \le 1$, defines a one-parameter family of manifolds interpolating between $\mathbb{R}^N$ and $\mathcal{L}_\mathcal{S}$. Care must be taken on the asymptotic behavior of $\mathcal{L}_\mathcal{S}$, determined by the function $\tilde{f}_i(\zeta)$, in order for $\mathcal{L}_\mathcal{S}$ to be in the same class as $\mathcal{M}_T$. However, in the case of periodic ζ, as in the following applications, this is automatic.

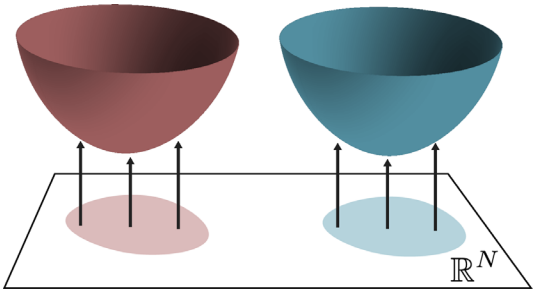
Second, the parametrization of Eq. (4.1) helps one to avoid the "trapping" of Monte Carlo chains as compared to the parametrization through the flow used in the generalized thimble method. This is explained by Fig. 18(b), where we can see that the same manifold is parametrized by a small region of $\mathbb{R}^N$ (in the generalized thimble method) or a much larger region using Eq. (4.1). Regions of $\mathbb{R}^N$ with large statistical weights are then less separated by low probability regions, which facilitates the Monte Carlo sampling. Similarly, the Jacobian of the parametrization of Eq. (4.1) fluctuates less than the Jacobian of the flow parametrization.

Finally, it is unlikely that the learnifold $\mathcal{L}_\mathcal{S}$ is better at controlling the sign problem than the manifold obtained by flow (of which the training set is taken). The usefulness of the method relies on the possibility of sampling the learnifold at a cost orders of magnitude cheaper than flowing (and computing or estimating the Jacobian). The hope is that the increase in statistics allowed by the speed of the process compensates for the smaller average sign.

[25]Other cost functions can be used instead of the $C(w,b)$ one used here.



(a) Parametrization using the holomorphic flow: small regions of $\mathbb{R}^N$ map into large regions of thimbles (or $\mathcal{M}_T$).



(b) Parametrization of Eq. 4.1: regions of $\mathbb{R}^N$ mapping into large regions of $\mathcal{L}_\mathcal{S}$ are larger and with smaller gaps between them, which helps to prevent trapping of the Monte Carlo's Markov chain.

FIG. 18. (a) Parametrization using the holomorphic flow. Small regions of $\mathbb{R}^N$ map onto large regions of thimbles (or $\mathcal{M}_T$). (b) Parametrization of Eq. (4.1). Regions of $\mathbb{R}^N$ mapping onto large regions of $\mathcal{L}_\mathcal{S}$ are larger and have smaller gaps between them, which helps to prevent trapping of the Monte Carlo Markov chain.

1. Case study: The (1+1)D Thirring model revisited

The application of the learnifold method to the $(1+1)$-dimensional Thirring model was discussed by Alexandru, Bedaque *et al.* (2017). The first step is to collect a number of complex configurations $\phi^a = \mathcal{F}_T(\zeta^a)$ obtained by evolving real configurations ζ^a by a time T according to the flow equations (2.7) in order to form the training set $\mathcal{S}$. This step is identical to what is done in the generalized thimble method. One wants to approximate the manifold $\mathcal{M}_T$ particularly well in the region that is going to be sampled the most. We also want to sample some of the field configurations on $\mathcal{M}_T$ toward the large field value region to make sure that the profile of the learnifold matches $\mathcal{M}_T$ in this region too. Therefore, some configurations are collected by running a Metropolis chain with weight e^{-S} and some others with weight $e^{-S/\tau}$, with $\tau > 1$. This way $\mathcal{M}_T$ is sufficiently sampled, so the interpolation manifold $\mathcal{L}_\mathcal{S}$ approximates it well in the statistically important regions and it is not radically wrong at asymptotically distant regions. Notice that there is no particular reason for the Monte Carlo chain to be thermalized while collecting these configurations; all that is required is to have a good enough sampling of the statistically important regions of $\mathcal{M}_T$ and some sampling of the other regions. After these configurations are generated they are used as the training set for the feed-forward network. To enforce translation invariance, all

translations of them are added to the training set, resulting in a larger set, typically of the order of 10^5 elements. This set is too large to be used in the minimization process, so different subsets ("minibatches") are used at different steps of the gradient descent (or ADAM step). Details were given by Alexandru, Bedaque *et al.* (2017). A comparison of the computational costs in the learnifold and generalized thimble methods is not straightforward, because the cost of the learnifold method is divided into a "fixed" cost related to the generation of the training set and the minimization of the cost function on the one hand and the running of the Monte Carlo given the optimal manifold. The second part is faster than the flowing required by the thimble method by orders of magnitude, but the first part may dominate the total costs. Calculations of the kind done by Alexandru, Başar, Bedaque, Ridgway, and Warrington (2017) can be performed more effectively using the learnifold method.

Similar methods were recently applied to the solution of the Hubbard model away from half filling (Ulybyshev, Dorozhinskii, and Pavlovskii, 2020)

C. Path optimization

An even more radical departure from thimbles is embodied in the path optimization method. The idea here is to consider a family of manifolds $\mathcal{M}_\lambda$ parametrized by a set of parameters (collectively denoted by λ) and to maximize the average sign within that family. The result of the maximization defines a manifold $\mathcal{M}_{\lambda_0}$ that is the best, within that family, at ameliorating the sign problem. The viability of the method rests on the observation that the gradient of the average sign in parameter space can be calculated with a (usually short) *sign problem–free* Monte Carlo calculation. Indeed, the average sign on a manifold $\mathcal{M}$ is given by

$$\langle\sigma\rangle_\lambda = \frac{\int_{\mathcal{M}_\lambda} d\phi e^{-S(\phi)}}{\int_{\mathcal{M}_\lambda} |d\phi| e^{-\mathrm{Re}S(\phi)}} = \frac{\int_{\mathbb{R}^N} d\zeta e^{-S_{\mathrm{eff}}(\zeta)}}{\int_{\mathbb{R}^N} d\zeta e^{-\mathrm{Re}S_{\mathrm{eff}}(\zeta)}}, \quad (4.5)$$

where $S_{\mathrm{eff}}(\zeta) = S[\phi(\zeta)] - \ln\det J(\zeta)$ includes the determinant of the Jacobian of the $\mathcal{M}_\lambda$ parametrization $\phi_i = \phi_i(\zeta)$. The numerator of Eq. (4.5) is independent of λ due to Cauchy's theorem. The denominator, however, as an integral of a nonholomorphic function, does depend on λ. In fact,

$$\frac{\nabla_\lambda\langle\sigma\rangle}{\langle\sigma\rangle} = \frac{\int_{\mathbb{R}^N} d\zeta e^{-\mathrm{Re}S_{\mathrm{eff}}(\zeta)}[-\nabla_\lambda \mathrm{Re}S + \mathrm{ReTr}(J^{-1}\nabla_\lambda J)]}{\int_{\mathbb{R}^N} d\zeta e^{-\mathrm{Re}S_{\mathrm{eff}}(\zeta)}}. \quad (4.6)$$

The average sign does not affect the direction of the vector $\nabla_\lambda\langle\sigma\rangle$ and can be neglected during the maximization process, while the right-hand side term of Eq. (4.6), the average $\langle -\nabla_\lambda \mathrm{Re}S + \mathrm{ReTr}(J^{-1}\nabla_\lambda J)\rangle_{\mathrm{Re}S_{\mathrm{eff}}}$, can be computed by the Monte Carlo method without encountering a sign problem. Knowledge of the gradient $\nabla_\lambda\langle\sigma\rangle$ (to be more precise, knowledge of its direction in λ space) allows a maximization routine like ADAM (Kingma and Ba, 2014) to find the values of λ leading to the largest possible sign within the parametrized family of manifolds.

A few facts make this scheme practical. First, a rough computation of the gradient is usually enough; some stochastic noise is actually useful in avoiding local minima that could otherwise trap the maximization process. Second, the last configurations obtained with one value of λ is close to being thermalized as λ is changed to a nearby value during the maximization process, bypassing the need for long thermalization periods at each step of the minimization. Finally, it is imperative that the family of manifolds considered (i) includes only manifolds in the same homology class as $\mathbb{R}^N$, (ii) contain manifolds where the sign problem is sufficiently ameliorated, and (iii) are parametrized in such a way the computation of the Jacobian J is cheap. Condition (i) is relatively easy to satisfy, but there is tension between conditions (ii) and (iii). Theoretical insight into the specific model of interest is required for the successful application of this method, and it is currently sorely missed in most theories of physical significance.

We note that contour deformations can be applied to any theory with a holomorphic path integrand. This includes theories with real actions but complex observables. This fact, combined with path optimization, has been used to tame the *signal to noise* problem encountered in calculations of correlation functions in simple field theories (Detmold *et al.*, 2020, 2021).

1. Case study: The $(2+1)$D Thirring model

The path optimization method was applied to a one-dimensional integral by Mori, Kashiwa, and Ohnishi (2017), the $(1+1)$D Thirring model at finite density by Alexandru, Bedaque, Lamm, and Lawrence (2018), the $(1+1)$D ϕ^4 model by Mori, Kashiwa, and Ohnishi (2018) (with a neural network parametrization of the manifolds similar to the one discussed in Sec. IV.B), the Polyakov-Nambu-Jona-Lasino model (known as the PNJL model) in $(0+1)$D (Kashiwa, Mori, and Ohnishi, 2019a, 2019b), $(0+1)$D QCD (Mori, Kashiwa, and Ohnishi, 2019), and $(1+1)$D ϕ^4 (Bursa and Kroyter, 2018). Here we discuss its application to the 3D Thirring model (Alexandru, Bedaque, Lamm, Lawrence, and Warrington, 2018).

The action defining the $(2+1)$D Thirring model is in Eq. (3.14). The family of manifolds considered is given by

$$\begin{aligned} A_0(x) &= \zeta_0(x) + i\{\lambda_0 + \lambda_1\cos\zeta_0(x) + \lambda_2\cos[2\zeta_0(x)]\},\\ A_1(x) &= \zeta_1(x),\\ A_2(x) &= \zeta_2(x), \end{aligned} \quad (4.7)$$

where λ_0, λ_1, and λ_2 are real numbers parametrizing the manifolds. The ansatz selection in Eq. (4.7) is motivated by the following considerations. First, the determinant of the Jacobian $J = \partial A_\nu/\partial\zeta_i$ is trivial to compute (the cost scales with the spacetime volume V, as opposed to V^3) since the value of $A_\mu(x)$ depends only on $\zeta(x)$ evaluated at the same spacetime point x:

$$\det J = \prod_x \{1 - \lambda_1\sin\zeta_0(x) - 2\lambda_2\sin[2\zeta_0(x)]\}. \quad (4.8)$$

Second, in the limit $\mu\to\infty$ the partition function

$$\lim_{\mu\to\infty} Z \approx \left[\int d^3A\, e^{(1/g^2)(\sum_\nu \cos A_\nu) + (i/2)A_0}\right]^{\beta V} \quad (4.9)$$

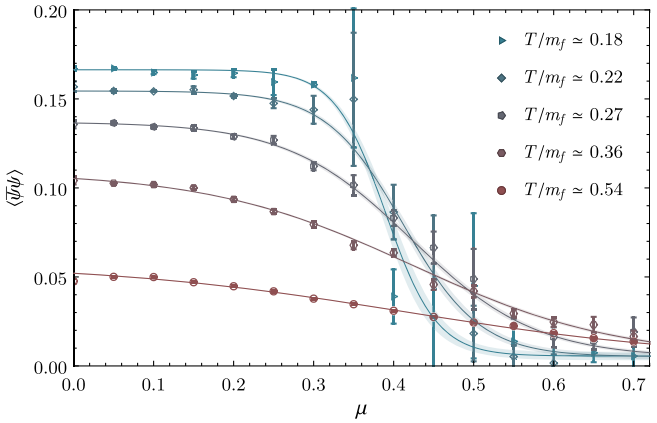


FIG. 19. $\langle\bar{\psi}\psi\rangle$ as a function of μ for the $(2+1)$-dimensional Thirring model in a $\beta \times 6^2$ lattice showing the melting of the chiral condensate as the density is increased. The solid lines are fit to the functional form $\langle\bar{\psi}\psi\rangle = A\tanh[\beta(\mu - \mu_c)]$.

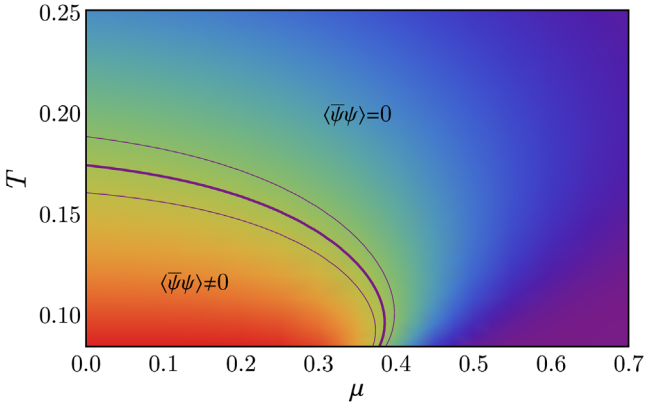


FIG. 20. Phase diagram of the 3D Thirring model (Alexandru, Bedaque, Lamm, Lawrence, and Warrington, 2018). The thick central line shows the location where $\langle\bar{\psi}\psi\rangle_{\mu,T} = 0.5\langle\bar{\psi}\psi\rangle_{0,0}$, and its width represents the statistical errors. The thinner lines indicate $\langle\bar{\psi}\psi\rangle_{\mu,T} = (0.5 \pm 0.05)\langle\bar{\psi}\psi\rangle_{0,0}$ to help gauge the sharpness of the transition.

factorizes into a separate integral at every spacetime point, and the sign problem arises entirely from A_0. The ansatz in Eq. (4.7) reflects this. Third, in the weak coupling limit $g^2 \to 0$ we expect the functional integral to be dominated by the saddle point with the smallest action that, as discussed in Sec. III.B.2, has the form $A_0 = i\alpha$, $A_1 = A_2 = 0$ for a real constant α. The ansatz in Eq. (4.7) contains manifolds that approach this thimble near its critical point. Finally, the variables A_ν are periodic variables with a period 2π [so they belong to $(S^1)^N$, not $\mathbb{R}^N$], and the question of whether the manifolds defined by Eq. (4.7) have a different asymptotic behavior is not present. Furthermore, by varying s from 1 to 0 in $A_0(x) = \zeta_0(x) + is\{\lambda_0 + \lambda_1\cos\zeta_0(x) + \lambda_2\cos[2\zeta_0(x)]\}$, we see that every member of the family of manifolds can be smoothly deformed to $(S^1)^N$, guaranteeing the applicability of the Cauchy theorem.

This method was used by Alexandru, Bedaque, Lamm, Lawrence, and Warrington (2018) in lattices of sizes up to 10^3 and action parameters near the continuum. It is interesting to examine how the maximization process proceeds. The parameter λ_0 quickly acquires a nonzero value close to the position of the critical point $A_0 = i\alpha$, $A_1 = A_2 = 0$. The corresponding manifold does not go through exactly through the critical point and has a larger average sign than the space tangent to that critical point. Afterward, λ_1 and λ_2 settle on their preferred values, giving a small curvature to the manifold. More complicated functions of ζ_0 in Eq. (4.7) do not seem to improve the average sign. It seems that one is required to go beyond the factorized form in Eq. (4.7) for further improvements.

The results obtained by this method show a clear transition (technically a crossover, as the fermion mass breaks chiral symmetry explicitly) between a phase with a large chiral condensate and another, at higher temperatures and densities, where chiral symmetry is restored and the chiral condensate is small; see Fig. 19. The resulting phase diagram is shown in Fig. 20. The computation shown in the figure is done in finite volume where the transition between regions of small and large chiral condensate is a smooth rather than a sharp phase transition. The sharpness of the phase transition is quantified in the figure by the two thinner lines.

V. CONCLUSION AND PROSPECTS

A bird's eye view of the developments described here reveal some broad lessons that should not be lost amid the technical details. The first is that the main idea that the thimble approach to the sign problem is based on is sound and that no fundamental flaw has been revealed, either conceptual or practical. This is not to say one can currently use the method to solve any sign problem. Indeed, the arguments of Troyer and Wiese (2005) give us a reason to suspect that a general solution to all sign problems is likely nonexistant. But the set of ideas explored in this review provide a new setting where the simulation of many models can and should be attempted. This is not a trivial statement. There is a common perception among nonpractitioners that there is a "conservation of difficulty" and that any approach to solve the sign problem will reveal, at closer inspection, a simulation cost as large as the naive attempts. This is demonstrably untrue, as the examples discussed in this review show.

The second foundational lesson is that the integration over all the relevant thimbles is both necessary and possible, with several algorithms already proposed and tested, usually in small scale simulations. These simulations are not without challenges: larger lattices generally produce more severe sign problems, the holomorphic gradient flow is numerically expensive, and for correctness the Jacobian must always be computed. However, algorithms aimed at addressing these difficulties have quickly developed. In fact, the rapid algorithmic development in the last few years generated a problem (or an opportunity) as most calculations were aimed at demonstrating the algorithm correctness and scaling properties and not focused on the physics of the problem. For instance, current technology should be able to clarify the phase diagram of a variety of $(1+1)$-dimensional models at finite temperature and density. At a larger computational cost, $(1+2)$-dimensional models can also be studied now. In fact, recent papers have begun setting up the path toward a solution to the repulsive Hubbard model away from half filling, a result that would be a game changer in the field (Hubbard, 1963;

Mukherjee and Cristoforetti, 2014; Saito, 2017; Fukuma, Matsumoto, and Umeda, 2019a, 2020; Ulybyshev, Winterowd, and Zafeiropoulos, 2019; Ostmeyer *et al.*, 2020).

An important insight arising from the research on thimble-related methods was that deformation of the integration manifold to other manifolds is both possible and profitable. This observation, as simple as it is, has vast consequences. Indeed, the condition that the imaginary part of the effective action is to be constant is only one constraint in a $2N$-dimensional space. This leaves a $2N - N - 1 = N - 1$ parameter family of possible directions of the tangent space of the integration manifold to choose from while still solving the sign problem. This freedom is not explored by holomorphic flow methods to deform contours of integration.[26] A few ideas exist on how to explore this newfound freedom. One is to use information about the model obtained elsewhere to devise parametrized families of integration manifolds suitable for that particular model. This approach provides a way of bringing physical insight into a Monte Carlo calculation that is sometimes characterized as a brute force method. Whatever insight is brought to the model, obtained by rigorous or intuitive, approximate methods, can then be used to speed up a calculation (hopefully exponentially) that is guaranteed to converge to the correct answer using the Monte Carlo method. A surprising recent development is that the physical insight into a model can be substituted by systematic machine learning techniques. We expect the near future to bring many more developments in this direction.

ACKNOWLEDGMENTS

We thank Hank Lamm, Scott Lawrence, Greg Ridgway, and Sohan Vartak, who collaborated with us on this topic. We also thank Tom Cohen, Gerald Dunne, Tarun Grover, Sergei Gukov, Jay Sau, Christian Schmidt, Luigi Scorzato, and Mithat Ünsal for conversations on this topic over the last few years. A. A. gratefully acknowledges the hospitality of the Physics Department at the University of Maryland, where part of this work was carried out. This work was supported in part by the U.S. DOE under Contracts No. DE-FG02-93ER-40762 No. DE-SC0021143, and No. DE-FG02-95ER40907 and the University of North Carolina at Chapel Hill Junior Faculty Development Award.

APPENDIX A: COMPUTATION OF THE JACOBIAN

The evolution by the holomorphic flow equation (2.7) by time T maps initial conditions ζ onto $\Phi(\zeta)$. The Jacobian of this transformation is derived in this appendix.

Begin by considering two infinitesimally close coordinates ζ and ζ', and let

$$v = \phi(\zeta', 0) - \phi(\zeta, 0) \tag{A1}$$

denote the corresponding difference vector between them. Flowing for time Δt, both $\phi(\zeta', 0)$ and $\phi(\zeta, 0)$ move, changing the difference vector. We denote this time dependent difference as $v(\Delta t)$; see Fig. 21. In the limit $\Delta t \to 0$,

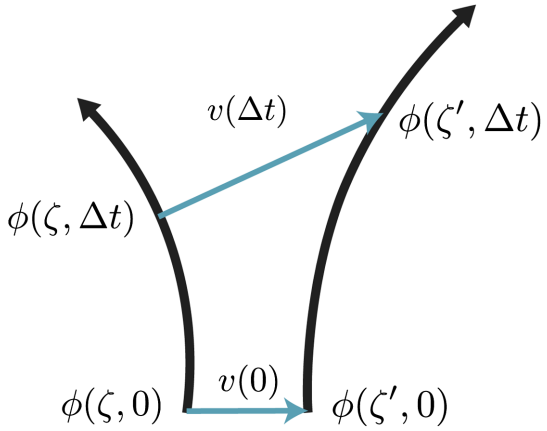


FIG. 21. Two nearby points evolving by the holomorphic flow. Their difference vector, shown in blue, evolves according to Eq. (A3).

$$\begin{aligned} v_a(\Delta t) &\equiv \phi_a(\zeta', \Delta t) - \phi_a(\zeta, \Delta t) \\ &= \left[\phi_a(\zeta', 0) + \Delta t \overline{\frac{\partial S}{\partial \phi_a}[\phi_a(\zeta', 0)]}\right] \\ &\quad - \left[\phi_a(\zeta, 0) + \Delta t \overline{\frac{\partial S}{\partial \phi_a}[\phi_a(\zeta, 0)]}\right] \\ &= [\phi_a(\zeta', 0) - \phi_a(\zeta, 0)] \\ &\quad + \Delta t \overline{\frac{\partial^2 S}{\partial \phi_a \partial \phi_b}[\phi_a(\zeta, 0)][\phi_b(\zeta', 0) - \phi_b(\zeta, 0)]} \\ &= v_a(0) + \Delta t \overline{H_{ab}[\phi_a(\zeta, 0)] v_b(0)}. \end{aligned} \tag{A2}$$

In other words, a vector evolves along a flow trajectory according to the following differential equation:

$$\frac{dv_a}{dt} = \overline{H_{ab}[\phi(\zeta, t)] v_b(t)}. \tag{A3}$$

We can use Eq. (A3) to evolve a set of N vectors forming an orthonormal basis. Packaging these vectors in the columns of a matrix $J(0) = \mathbb{1}$, we see that $J(t)$ obeys

$$\frac{dJ}{dt} = \overline{HJ}. \tag{A4}$$

APPENDIX B: ANOTHER DEFINITION FOR THIMBLES

In this appendix we give a different perspective on Lefschetz thimbles. We begin by focusing on the stationary points of the flow, namely, the critical points of the action ϕ^c, such that $\partial S/\partial \phi_i|_{\phi^c} = 0$. Around a critical point[27] it is always

[26] Notice that, contrary to the multidimensional case, in the familiar case of a single complex variable the condition $\mathrm{Im} S_{\mathrm{eff}} = 0$ defines a unique contour.

[27] In our analysis we consider only isolated, quadratic (non-degenerate) critical points. A degenerate critical point where the Hessian determinant of ϕ vanishes can be split into μ nondegenerate critical points with a small deformation, with μ being the Milnor number of the critical point. A similar analysis presented in this section follows (Pham, 1983).

possible to find local coordinates $\{z_i = x_i + iy_i\}$, with $i = 1, \ldots, N$, such that

$$S(\phi) - S(\phi^c) = z_1^2 + \cdots + z_N^2 = (x_1^2 + \cdots + x_N^2) - (y_1^2 + \cdots + y_N^2) + 2i(x_1 y_1 + \cdots + x_N y_N), \tag{B1}$$

whose existence is guaranteed by the Morse lemma. Now consider the $(N-1)$-dimensional surface $v(s)$ defined by $x_1^2 + \cdots + x_N^2 = s$ and $y_1 = \cdots = y_N = 0$. This surface is known as the vanishing cycle since it vanishes at the critical point. It can be viewed as the level set of the action around the critical point $S^{-1}(s + s_c)$, where $s_c = S(\phi^c)$. We now move the vanishing cycle by varying s. This can be done by taking the vanishing cycle around the critical point $v(\epsilon)$ and then flowing it. As s runs from 0 to ∞, the vanishing cycle sweeps a real N-dimensional surface. This N-dimensional surface, defined as the union of vanishing cycles on the half line $0 \le s < \infty$, $\mathcal{T} = \cup_s v(s)$, is known as the Lefschetz thimble associated with the critical point ϕ^c. Similarly, we define an $(N-1)$-dimensional "dual" cycle, $v^D(s)$ by $x_1 = \cdots = x_N = 0$ and $y_1^2 + \cdots + y_N^2 = s$. We call the union of these dual cycles on the half line $0 \le s < \infty$, $\mathcal{K} = \cup_s v^D(s)$, the dual thimble.[28]

[28]Note that what we call the thimble and the dual thimble are referred as downward (upward) cycles referring to the fact that the weight $e^{-\mathrm{Re}S}$ monotonically decreases (increases) over them with flow (Witten, 2011).

REFERENCES

Aarts, G., 2009, Phys. Rev. Lett. **102**, 131601.
Aarts, G., 2016, J. Phys. Conf. Ser. **706**, 022004.
Aarts, G., F. Attanasio, B. Jäger, and D. Sexty, 2016, J. High Energy Phys. 09, 087.
Aarts, G., L. Bongiovanni, E. Seiler, D. Sexty, and I.-O. Stamatescu, 2013, Eur. Phys. J. A **49**, 89.
Aarts, G., F. A. James, E. Seiler, and I.-O. Stamatescu, 2011, Eur. Phys. J. C **71**, 1756.
Aarts, G., E. Seiler, and I.-O. Stamatescu, 2010, Phys. Rev. D **81**, 054508.
Aarts, G., and I.-O. Stamatescu, 2008, J. High Energy Phys. 09, 018.
Alexandru, A., G. Başar, and P. Bedaque, 2015, arXiv:1510.03258.
Alexandru, A., G. Başar, P. F. Bedaque, H. Lamm, and S. Lawrence, 2018, Phys. Rev. D **98**, 034506.
Alexandru, A., G. Başar, P. F. Bedaque, and G. W. Ridgway, 2017, Phys. Rev. D **95**, 114501.
Alexandru, A., G. Başar, P. Bedaque, G. W. Ridgway, and N. C. Warrington, 2016a, Phys. Rev. D **94**, 045017.
Alexandru, A., G. Başar, P. F. Bedaque, G. W. Ridgway, and N. C. Warrington, 2016b, Phys. Rev. D **93**, 094514.
Alexandru, A., G. Başar, P. F. Bedaque, G. W. Ridgway, and N. C. Warrington, 2017, Phys. Rev. D **95**, 014502.
Alexandru, A., G. Başar, P. F. Bedaque, S. Vartak, and N. C. Warrington, 2016, Phys. Rev. Lett. **117**, 081602.
Alexandru, A., G. Başar, P. F. Bedaque, and N. C. Warrington, 2017, arXiv:1703.02414.
Alexandru, A., P. F. Bedaque, H. Lamm, and S. Lawrence, 2017, Phys. Rev. D **96**, 094505.
Alexandru, A., P. F. Bedaque, H. Lamm, and S. Lawrence, 2018, arXiv:1804.00697.
Alexandru, A., P. F. Bedaque, H. Lamm, S. Lawrence, and N. C. Warrington, 2018, Phys. Rev. Lett. **121**, 191602.
Alexandru, A., G. Bergner, D. Schaich, and U. Wenger, 2018, Phys. Rev. D **97**, 114503.
Alexandru, A., M. Faber, I. Horvath, and K.-F. Liu, 2005, Phys. Rev. D **72**, 114513.
Alexandru, A., and U. Wenger, 2011, Phys. Rev. D **83**, 034502.
Alford, M. G., S. Chandrasekharan, J. Cox, and U. J. Wiese, 2001, Nucl. Phys. **B602**, 61.
Anderson, J. B., 1975, J. Chem. Phys. **63**, 1499.
Aniceto, I., G. Başar, and R. Schiappa, 2019, Phys. Rep. **809**, 1.
Attanasio, F., and J. E. Drut, 2020, Phys. Rev. A **101**, 033617.
Ayyar, V., S. Chandrasekharan, and J. Rantaharju, 2018, Phys. Rev. D **97**, 054501.
Barbour, I. M., C. T. H. Davies, and Z. Sabeur, 1988, Phys. Lett. B **215**, 567.
Bazavov, A., *et al.* (HotQCD Collaboration), 2019, Phys. Lett. B **795**, 15.
Bellwied, R., S. Borsanyi, Z. Fodor, J. Günther, S. D. Katz, C. Ratti, and K. K. Szabo, 2015, Phys. Lett. B **751**, 559.
Bender, I., T. Hashimoto, F. Karsch, V. Linke, A. Nakamura, M. Plewnia, I. Stamatescu, and W. Wetzel, 1992, Nucl. Phys. B, Proc. Suppl. **26**, 323.
Berger, C. E., K. J. Morrell, and J. E. Drut, 2020, arXiv:2004.02344.
Berger, C. E., L. Rammelmüller, A. C. Loheac, F. Ehmann, J. Braun, and J. E. Drut, 2019, arXiv:1907.10183.
Berges, J., S. Borsanyi, D. Sexty, and I. O. Stamatescu, 2007, Phys. Rev. D **75**, 045007.
Berkowitz, E., E. Rinaldi, M. Hanada, G. Ishiki, S. Shimasaki, and P. Vranas, 2016, Phys. Rev. D **94**, 094501.
Bloch, J., F. Bruckmann, and T. Wettig, 2013, J. High Energy Phys. 10, 140.
Blum, T. C., J. E. Hetrick, and D. Toussaint, 1996, Phys. Rev. Lett. **76**, 1019.
Bonati, C., M. D'Elia, M. Mariti, M. Mesiti, F. Negro, and F. Sanfilippo, 2015, Phys. Rev. D **92**, 054503.
Bonati, C., M. D'Elia, F. Negro, F. Sanfilippo, and K. Zambello, 2018, Phys. Rev. D **98**, 054510.
Borsanyi, S., Z. Fodor, J. N. Guenther, R. Kara, S. D. Katz, P. Parotto, A. Pasztor, C. Ratti, and K. K. Szabo, 2020, arXiv:2002.02821.
Braaten, E., and R. D. Pisarski, 1990, Nucl. Phys. **B337**, 569.
Braun, J., J.-W. Chen, J. Deng, J. E. Drut, B. Friman, C.-T. Ma, and Y.-D. Tsai, 2013, Phys. Rev. Lett. **110**, 130404.
Bursa, F., and M. Kroyter, 2018, J. High Energy Phys. 12, 054.
Carlson, J., S. Gandolfi, F. Pederiva, S. C. Pieper, R. Schiavilla, K. Schmidt, and R. Wiringa, 2015, Rev. Mod. Phys. **87**, 1067.
Cea, P., L. Cosmai, and A. Papa, 2014, Phys. Rev. D **89**, 074512.
Chandrasekharan, S., 2012, Phys. Rev. D **86**, 021701.
Chandrasekharan, S., 2013, Eur. Phys. J. A **49**, 90.
Chandrasekharan, S., and U.-J. Wiese, 1999, Phys. Rev. Lett. **83**, 3116.
Chen, J.-W., and D. B. Kaplan, 2004, Phys. Rev. Lett. **92**, 257002.
Cherman, A., D. Dorigoni, and M. Unsal, 2014, arXiv:1403.1277.
Cohen, T. D., 2003, Phys. Rev. Lett. **91**, 222001.
Cristoforetti, M., F. Di Renzo, G. Eruzzi, A. Mukherjee, C. Schmidt, L. Scorzato, and C. Torrero, 2014, Phys. Rev. D **89**, 114505.
Cristoforetti, M., F. Di Renzo, A. Mukherjee, and L. Scorzato, 2013, Phys. Rev. D **88**, 051501.

Cristoforetti, M., F. Di Renzo, and L. Scorzato (AuroraScience Collaboration), 2012, Phys. Rev. D **86**, 074506.
Cristoforetti, M., L. Scorzato, and F. Di Renzo, 2013, J. Phys. Conf. Ser. **432**, 012025.
de Forcrand, P., 2010, arXiv:1005.0539.
de Forcrand, P., and S. Kratochvila, 2006, Nucl. Phys. B, Proc. Suppl. **153**, 62.
de Forcrand, P., and O. Philipsen, 2002, Nucl. Phys. **B642**, 290.
de Forcrand, P., and O. Philipsen, 2003, Nucl. Phys. **B673**, 170.
de Forcrand, P., *et al.* (QCD-TARO Collaboration), 2000, Nucl. Phys. B, Proc. Suppl. **83–84**, 408.
Delabaere, E., and C. J. Howls, 2002, Duke Math. J. **112**, 199.
D'Elia, M., and M.-P. Lombardo, 2003, Phys. Rev. D **67**, 014505.
D'Elia, M., and M.-P. Lombardo, 2004, Phys. Rev. D **70**, 074509.
Detmold, W., G. Kanwar, H. Lamm, M. L. Wagman, and N. C. Warrington, 2021, Phys. Rev. D **103**, 094517.
Detmold, W., G. Kanwar, M. L. Wagman, and N. C. Warrington, 2020, arXiv:2003.05914.
Di Renzo, F., and G. Eruzzi, 2018, Phys. Rev. D **97**, 014503.
Di Renzo, F., S. Singh, and K. Zambello, 2019, arXiv:2002.00472.
Di Renzo, F., S. Singh, and K. Zambello, 2021, Phys. Rev. D **103**, 034513.
Di Renzo, F., and K. Zambello, 2021, arXiv:2109.02511.
Di Tucci, A., J. Feldbrugge, J.-L. Lehners, and N. Turok, 2019, Phys. Rev. D **100**, 063517.
Di Tucci, A., and J.-L. Lehners, 2019, Phys. Rev. Lett. **122**, 201302.
Duane, S., A. Kennedy, B. Pendleton, and D. Roweth, 1987, Phys. Lett. B **195**, 216.
Dunne, G. V., and M. Ünsal, 2016, Proc. Sci. LATTICE2015, 010 [arXiv:1511.05977].
Earl, D. J., and M. W. Deem, 2005, Phys. Chem. Chem. Phys. **7**, 3910.
Elhatisari, S., E. Epelbaum, H. Krebs, T. A. Lähde, D. Lee, N. Li, B.-n. Lu, U.-G. Meißner, and G. Rupak, 2017, Phys. Rev. Lett. **119**, 222505.
Endres, M. G., 2007, Phys. Rev. D **75**, 065012.
Endrodi, G., Z. Fodor, S. D. Katz, and K. K. Szabo, 2011, J. High Energy Phys. 04, 001.
Epelbaum, E., H. Krebs, T. A. Lähde, D. Lee, U.-G. Meißner, and G. Rupak, 2014, Phys. Rev. Lett. **112**, 102501.
Fantoni, S., A. Sarsa, and K. E. Schmidt, 2001, Phys. Rev. Lett. **87**, 181101.
Feldbrugge, J., U.-L. Pen, and N. Turok, 2019, arXiv:1909.04632.
Fodor, Z., and S. D. Katz, 2002, J. High Energy Phys. 03, 014.
Fodor, Z., S. D. Katz, and C. Schmidt, 2007, J. High Energy Phys. 03, 121.
Fodor, Z., S. D. Katz, D. Sexty, and C. Török, 2015, Phys. Rev. D **92**, 094516.
Fujii, H., D. Honda, M. Kato, Y. Kikukawa, S. Komatsu, and T. Sano, 2013, J. High Energy Phys. 10, 147.
Fujii, H., S. Kamata, and Y. Kikukawa, 2015, J. High Energy Phys. 11, 078.
Fujii, H., S. Kamata, and Y. Kikukawa, 2017, arXiv:1710.08524.
Fukuma, M., and N. Matsumoto, 2021, Prog. Theor. Exp. Phys. 023B08.
Fukuma, M., N. Matsumoto, and Y. Namekawa, 2021, arXiv:2107.06858.
Fukuma, M., N. Matsumoto, and N. Umeda, 2018, J. High Energy Phys. 11, 060.
Fukuma, M., N. Matsumoto, and N. Umeda, 2019a, Phys. Rev. D **100**, 114510.
Fukuma, M., N. Matsumoto, and N. Umeda, 2019b, arXiv:1912.13303.
Fukuma, M., N. Matsumoto, and N. Umeda, 2020, arXiv:2001.01665.
Fukuma, M., and N. Umeda, 2017, Prog. Theor. Exp. Phys. 073B01.
Fulde, P., and R. A. Ferrell, 1964, Phys. Rev. **135**, A550.
Gandolfi, S., A. Lovato, J. Carlson, and K. E. Schmidt, 2014, Phys. Rev. C **90**, 061306(R).
Garron, N., and K. Langfeld, 2016, Eur. Phys. J. C **76**, 569.
Garron, N., and K. Langfeld, 2017, Eur. Phys. J. C **77**, 470.
Gattringer, C., and T. Kloiber, 2013, Nucl. Phys. **B869**, 56.
Gattringer, C., and P. Törek, 2015, Phys. Lett. B **747**, 545.
Geyer, C. J., 1991, Interface Foundation of North America, retrieved from the University of Minnesota Digital Conservancy, https://hdl.handle.net/11299/58440.
Gezerlis, A., 2011, Phys. Rev. C **83**, 065801.
Grady, M., 1985, Phys. Rev. D **32**, 1496.
Han, M., Z. Huang, H. Liu, D. Qu, and Y. Wan, 2021, Phys. Rev. D **103**, 084026.
Hanada, M., J. Nishimura, Y. Sekino, and T. Yoneya, 2011, J. High Energy Phys. 12, 020.
Hann, C. T., E. Huffman, and S. Chandrasekharan, 2017, Ann. Phys. (Amsterdam) **376**, 63.
Harlow, D., J. Maltz, and E. Witten, 2011, J. High Energy Phys. 12, 071.
Hasenfratz, A., and D. Toussaint, 1992, Nucl. Phys. **B371**, 539.
Hubbard, J., 1963, Proc. R. Soc. A **276**, 238, http://www.jstor.org/stable/2414761.
Huffman, E., and S. Chandrasekharan, 2016, Phys. Rev. E **94**, 043311.
Huffman, E., and S. Chandrasekharan, 2020, Phys. Rev. D **101**, 074501.
Huffman, E. F., and S. Chandrasekharan, 2014, Phys. Rev. B **89**, 111101.
Jeon, S., and L. G. Yaffe, 1996, Phys. Rev. D **53**, 5799.
Jow, D. L., F. Lin, XI, E. Tyhurst, and U.-L. Pen, 2021, arXiv:2103.08687.
Kaczmarek, O., F. Karsch, E. Laermann, C. Miao, S. Mukherjee, P. Petreczky, C. Schmidt, W. Soeldner, and W. Unger, 2011, Phys. Rev. D **83**, 014504.
Kanwar, G., and M. L. Wagman, 2021, Phys. Rev. D **104**, 014513.
Karsch, F., 2000, Nucl. Phys. B, Proc. Suppl. **83–84**, 14.
Karsch, F., and K. H. Mutter, 1989, Nucl. Phys. **B313**, 541.
Kashiwa, K., Y. Mori, and A. Ohnishi, 2019a, Phys. Rev. D **99**, 114005.
Kashiwa, K., Y. Mori, and A. Ohnishi, 2019b, Phys. Rev. D **99**, 014033.
Keldysh, L., 1964, Zh. Eksp. Teor. Fiz. **47**, 1515.
Kingma, D. P., and J. Ba, 2014, arXiv:1412.6980.
Klauder, J. R., 1983, in *Recent Developments in High-Energy Physics*, Acta Physica Austriaca Supplementum Vol. 25, edited by H. Mitter and C. B. Lang (Springer-Verlag, Vienna), p. 251.
Koonin, S., D. Dean, and K. Langanke, 1997, Phys. Rep. **278**, 1.
Kratochvila, S., and P. de Forcrand, 2005, Proc. Sci. LAT2005, 167 [arXiv:hep-lat/0509143].
Lacroix, C., P. Mendels, and F. Mila, 2011, *Introduction to Frustrated Magnetism: Materials, Experiments, Theory*, Springer Series in Solid-State Sciences Vol. 164 (Springer-Verlag, Berlin).
Lähde, T. A., T. Luu, D. Lee, U.-G. Meißner, E. Epelbaum, H. Krebs, and G. Rupak, 2015, Eur. Phys. J. A **51**, 92.
Lamm, H., S. Lawrence, and Y. Yamauchi (NuQS Collaboration), 2020, Phys. Rev. Research **2**, 013272.
Langelage, J., M. Neuman, and O. Philipsen, 2014, J. High Energy Phys. 09, 131.
Langfeld, K., and B. Lucini, 2014, Phys. Rev. D **90**, 094502.

Larkin, A., and Y. Ovchinnikov, 1964, Zh. Eksp. Teor. Fiz. **47**, 1136.
Lawrence, S., 2020, arXiv:2006.03683.
Lawrence, S., and Y. Yamauchi, 2021, Phys. Rev. D **103**, 114509.
Lee, D., 2007, Phys. Rev. Lett. **98**, 182501.
Lee, D., 2009, Prog. Part. Nucl. Phys. **63**, 117.
Lee, D., B. Borasoy, and T. Schaefer, 2004, Phys. Rev. C **70**, 014007.
Lee, D., and T. Schäfer, 2005, Phys. Rev. C **72**, 024006.
Li, A., A. Alexandru, and K.-F. Liu, 2011, Phys. Rev. D **84**, 071503.
Li, A., A. Alexandru, K.-F. Liu, and X. Meng, 2010, Phys. Rev. D **82**, 054502.
Li, D., 2016, arXiv:1605.04623.
Li, Z.-X., and H. Yao, 2019, Annu. Rev. Condens. Matter Phys. **10**, 337.
Loh, E. Y., J. E. Gubernatis, R. T. Scalettar, S. R. White, D. J. Scalapino, and R. L. Sugar, 1990, Phys. Rev. B **41**, 9301.
Lu, B.-N., N. Li, S. Elhatisari, D. Lee, J. E. Drut, T. A. Lähde, E. Epelbaum, and U.-G. Meißner, 2019a, arXiv:1912.05105.
Lu, B.-N., N. Li, S. Elhatisari, D. Lee, E. Epelbaum, and U.-G. Meißner, 2019b, Phys. Lett. B **797**, 134863.
Matsui, H., 2021, arXiv:2102.09767.
Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953, J. Chem. Phys. **21**, 1087.
Mishchenko, P. A., Y. Kato, and Y. Motome, 2021, arXiv:2106.07937.
Miyamura, O. (QCD-TARO Collaboration), 2002, Nucl. Phys. **A698**, 395.
Mori, Y., K. Kashiwa, and A. Ohnishi, 2017, Phys. Rev. D **96**, 111501.
Mori, Y., K. Kashiwa, and A. Ohnishi, 2018, Prog. Theor. Exp. Phys. 023B04.
Mori, Y., K. Kashiwa, and A. Ohnishi, 2019, Prog. Theor. Exp. Phys. 113B01.
Morinaga, T., 2021, Phys. Rev. D **103**, 083014.
Mou, Z.-G., P. M. Saffin, and A. Tranberg, 2019, J. High Energy Phys. 11, 135.
Mou, Z.-G., P. M. Saffin, A. Tranberg, and S. Woodward, 2019, J. High Energy Phys. 06, 094.
Mukherjee, A., and M. Cristoforetti, 2014, Phys. Rev. B **90**, 035134.
Mukherjee, A., M. Cristoforetti, and L. Scorzato, 2013, Phys. Rev. D **88**, 051502.
Muroya, S., A. Nakamura, C. Nonaka, and T. Takaishi, 2003, Prog. Theor. Phys. **110**, 615.
Nakamura, A., S. Oka, and Y. Taniguchi, 2016, J. High Energy Phys. 02, 054.
Neal, R., 1996, Stat. Comput. **6**, 353.
Ostmeyer, J., E. Berkowitz, S. Krieg, T. A. Lähde, T. Luu, and C. Urbach, 2020, arXiv:2005.11112.
Parisi, G., 1983, Phys. Lett. **131B**, 393.
Parisi, G., and Y.-s. Wu, 1981, Sci. Sin. (Engl. Ed.) **24**, 483.
Pawlowski, J. M., M. Scherzer, C. Schmidt, F. P. G. Ziegler, and F. Ziesché, 2020, arXiv:2001.09767.
Pawlowski, J. M., M. Scherzer, C. Schmidt, F. P. G. Ziegler, and F. Ziesché, 2021, Phys. Rev. D **103**, 094505.
Pawlowski, J. M., and C. Zielinski, 2013, Phys. Rev. D **87**, 094503.
Pham, F., 1983, in *Singularities* Proceedings of Symposia in Pure Mathematics Vol. 40, Part 2, edited by P. Orlik (American Mathematical Society, Providence).
Philipsen, O., 2007, Eur. Phys. J. Special Topics **152**, 29.
Roscher, D., J. Braun, and J. E. Drut, 2014, Phys. Rev. A **89**, 063609.
Rossi, P., and U. Wolff, 1984, Nucl. Phys. **B248**, 105.
Ruder, S., 2016, arXiv:1609.04747.
Saito, H., 2017, J. Phys. Soc. Jpn. **86**, 093001.
Schaich, D., 2019, Proc. Sci. LATTICE2018, 005[arXiv:1810.09282].
Scherzer, M., D. Sexty, and I.-O. Stamatescu, 2020, Phys. Rev. D **102**, 014515.
Schwinger, J. S., 1961, J. Math. Phys. (N.Y.) **2**, 407.
Seiler, E., D. Sexty, and I.-O. Stamatescu, 2013, Phys. Lett. B **723**, 213.
Sexty, D., 2014, Phys. Lett. B **729**, 108.
Sexty, D., 2019, Phys. Rev. D **100**, 074503.
Sindzingre, P., P. Lecheminant, and C. Lhuillier, 1994, Phys. Rev. B **50**, 3108.
Swendsen, R. H., and J.-S. Wang, 1986, Phys. Rev. Lett. **57**, 2607.
Takagi, T., 1924, Jpn. J. Math. **1**, 83.
Tanizaki, Y., Y. Hidaka, and T. Hayata, 2016, New J. Phys. **18**, 033002.
Tanizaki, Y., H. Nishimura, and J. J. M. Verbaarschot, 2017, J. High Energy Phys. 10, 100.
Tanizaki, Y., and M. Tachibana, 2017, J. High Energy Phys. 02, 081.
Thirring, W. E., 1958, Ann. Phys. (N.Y.) **3**, 91.
Troyer, M., and U.-J. Wiese, 2005, Phys. Rev. Lett. **94**, 170201.
Ulybyshev, M., C. Winterowd, and S. Zafeiropoulos, 2019, arXiv:1906.02726.
Ulybyshev, M., C. Winterowd, and S. Zafeiropoulos, 2020, Phys. Rev. D **101**, 014508.
Ulybyshev, M. V., V. I. Dorozhinskii, and O. V. Pavlovskii, 2020, Phys. Part. Nucl. **51**, 363.
White, S. R., D. J. Scalapino, R. L. Sugar, E. Y. Loh, J. E. Gubernatis, and R. T. Scalettar, 1989, Phys. Rev. B **40**, 506.
Wigner, E., 1937, Phys. Rev. **51**, 106.
Wiringa, R. B., S. C. Pieper, J. Carlson, and V. R. Pandharipande, 2000, Phys. Rev. C **62**, 014001.
Witten, E., 1978, Nucl. Phys. **B145**, 110.
Witten, E., 2010, arXiv:1009.6032.
Witten, E., 2011, in *Chern-Simons Gauge Theory: 20 Years After*, AMS/IP Studies in Advanced Mathematics Vol. 50, edited by J. E. Andersen, H. U. Boden, A. Hahn, and B. Himpel (American Mathematical Society, Providence), p. 347.
Zambello, K., and F. Di Renzo, 2019, Proc. Sci. LATTICE2018, 148 [arXiv:1811.03605].
Zhang, S., J. Carlson, and J. E. Gubernatis, 1995, Phys. Rev. Lett. **74**, 3652.
Zhang, S., J. Carlson, and J. E. Gubernatis, 1997, Phys. Rev. B **55**, 7464.