ON THE VALIDITY OF COMPLEX LANGEVIN METHOD FOR PATH INTEGRAL COMPUTATIONS*

ZHENNING CAI†, XIAOYU DONG‡, AND YANG KUANG†

Abstract. The complex Langevin (CL) method is a classical numerical strategy to alleviate the numerical sign problem in the computation of lattice field theories. Mathematically, it is a simple numerical tool to compute a wide class of high-dimensional and oscillatory integrals. However, it is often observed that the CL method converges but the limiting result is incorrect. The literature has several unclear or even conflicting statements, making the method look mysterious. By an in-depth analysis of a model problem, we reveal the mechanism of how the CL result turns biased as the parameter changes, and it is demonstrated that such a transition is difficult to capture. Our analysis also shows that the method works for any observables only if the probability density function generated by the CL process is localized. To generalize such observations to lattice field theories, we formulate the CL method on general groups using rigorous mathematical languages for the first time, and we demonstrate that such localized probability density function does not exist in the simulation of lattice field theories for general compact groups, which explains the unstable behavior of the CL method. Fortunately, we also find that the gauge cooling technique creates additional velocity that helps confine the samples, so that we can still see localized probability density functions in certain cases. Thereby, the gauge cooling method significantly broadens the application of the CL method. The limitations of gauge cooling are also discussed. In particular, we prove that gauge cooling has no effect for Abelian groups, and we provide an example showing that biased results still exist when gauge cooling is insufficient to confine the probability density function.

Key words. complex Langevin method, gauge cooling, lattice field theory

AMS subject classification. 65C05

DOI. 10.1137/20M1363224

1. Introduction. Quantum field theory (QFT) is a fundamental theoretical framework in particle physics and condensed matter physics, which has achieved great success in explaining and discovering elementary particles in history. Although QFT still lacks a rigorous mathematical foundation, there have already been numerous approaches to carrying out computations in QFT, based on either perturbative or nonperturbative approaches. Perturbative approaches can be applied when the coupling constant, which appears in the coupling term in the Lagrangian describing the interaction between particles, is relatively small, so that the asymptotic expansions with respect to the coupling constant, often denoted by Feynman diagrams, can be adopted as approximations. In quantum chromodynamics (QCD), which studies the interaction between quarks and gluons, the perturbative approaches work in the case of large momentum transfers. However, when studying QCD at small momenta or energies (less than 1GeV), due to renormalization, the coupling constant is comparable to 1 and the perturbative theory is no longer accurate [22]. Therefore, one has to

†Department of Mathematics, National University of Singapore, 119076, Singapore (matcz@nus.edu.sg, matkuan@nus.edu.sg).

‡State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, People's Republic of China (dongxiaoyu@lsec.cc.ac.cn).

resort to nonperturbative approaches, typically lattice QCD calculations, to obtain reliable approximations of the observables.

Lattice QCD is formulated based on the path integral quantization of the classical gauge field theory. In general, the expectation of any observable O can be computed by evaluating the discrete path integral

$$\langle O \rangle = \frac{1}{Z} \int_\Omega O(x) e^{-S(x)} \, dx, \qquad Z = \int_\Omega e^{-S(x)} \, dx,$$

where Ω is a space whose number of dimensions is often larger than 10^4, and $S(x)$ is the action. This problem is of broad interest in statistical mechanics [21], quantum mechanics [15], and string theories [9]. Due to the high dimensionality, one has to apply Monte Carlo methods to compute this integral. However, when the chemical potential is nonzero, the action is no longer real-valued, so that $e^{-S(x)}$ is no longer positive, which then causes a severe *numerical sign problem* in the computation [19]. Here numerical sign problem refers to the phenomenon that the variance of a stochastic quantity is way larger than its expectation, resulting in significant difficulty in reducing the relative error in Monte Carlo simulations. The large variance is usually due to strong oscillation of the stochastic quantity, causing significant cancellation of its positive and negative contributions. Such a problem typically occurs in quantum Monte Carlo simulations, including both lattice field theory [27] and real-time dynamics [38]. In lattice QCD, the imaginary part of $S(x)$ contributes to the high oscillation, so that the "partition function" Z itself already has a small value and is difficult to compute accurately.

In general, there is no universal approach to solving the numerical sign problem. For example, the inchworm Monte Carlo method [15, 14], which takes the idea of partial resummation, has been proposed to mitigate the numerical sign problem for real-time dynamics of the impurity model or open quantum systems; the Lefschetz thimble method [17, 16], which uses Morse theory to change the integral path, has been applied in lattice QCD computations. In this work, we are interested in another approach to taming the numerical sign problem, known as the complex Langevin (CL) method, whose basic idea is to search for a positive probability function in a higher-dimensional space that is equivalent to the "complex probability density function" $Z^{-1}e^{-S(x)}$, so that we can apply the Monte Carlo method in this higher-dimensional space, which is free of numerical sign problems [42]. In the CL method, such a space is established by complexifying all the variables, meaning to replace all real variables by complex variables, and extending all functions defined on $\mathbb{R}$ to functions on $\mathbb{C}$ by analytic continuation. Thus the number of dimensions is doubled. The equivalent probability density function in this complexified space is sampled by complexifying the Langevin method. Such a method is also known as stochastic quantization [31]. We refer the readers to [39] for a recent review of this method.

Since the CL method was proposed in [24, 33], it has never been fully understood theoretically. In numerical experiments, it is frequently seen that this method generates biased results, and its validity has therefore been questioned by a number of researchers [20, 7, 34]. Its application had been rarely seen until one breakthrough of the CL method, called the gauge cooling technique, was introduced in [41]. Such a strategy utilizes the redundant degrees of freedom in the gauge field theory to stabilize the method. With this improvement, the CL method has been successfully applied to finite density QCD [5, 43, 25] and the computations in the superstring theory [8, 32]. Recently, it has also been used in the computation of spin-orbit coupling [10].

Despite the success, failure of the CL method still occurs, and researchers are still working hard on understanding the algorithm by formal analysis and some particular examples [3, 29, 34], hoping to make further improvements. For instance, the recent work [36] analyzes a specific one-dimensional case, where the authors quantified the bias by relating it to the boundary terms when performing integration by parts. The results therein have been further applied in [37] to correct the CL results. In [6], the authors analyze the role of poles in the Langevin drift and find that although the converged CL results satisfy the Schwinger–Dyson equation, the integrals may still be incorrectly predicted. This phenomenon is further studied in [35] in a rigorous way for the one-dimensional case. The issue revealed in [36] shows that the correctness of the CL method depends highly on the decay rate of the probability density function and the growth rate of the observable at infinity, which is difficult to predict before the computation, especially in the high-dimensional case. The only case in which the convergence can be guaranteed for any observable is when the probability density function is localized, which has been studied in [3] for a one-dimensional model problem.

In this paper, we will carry out a deeper study of this specific case and clarify some unclear statements and some misunderstanding in previous studies. Later, we will show that in lattice QCD simulations, the localized probability density function can occur only after gauge cooling is introduced, although it is not always effective. This reveals why this technique is essential to the CL method.

To begin with, we will provide a brief review of the CL method and introduce the model problem studied in [3].

1.1. Review of the complex Langevin method. To sketch the basic idea of the CL method in the lattice field theory, we can simply consider the one-dimensional integration problem which aims to find

$$(1.1) \qquad \langle O \rangle = \frac{1}{Z} \int_{\mathbb{R}} O(x) e^{-S(x)} \, dx, \quad Z = \int_{\mathbb{R}} e^{-S(x)} \, dx.$$

When $S(x)$ is real, we can regard Z as the partition function, so that the integral can be evaluated by the Langevin method. Specifically, we can approximate (1.1) by

$$\langle O \rangle \approx \frac{1}{N} \sum_{k=1}^{N} O(X_k),$$

where X_k are samples generated by simulating the Langevin equation

$$(1.2) \qquad dx = K(x) \, dt + dw, \quad K(x) = -S'(x),$$

and we choose $X_k = x(T + k\Delta T)$ for a sufficiently large T and sufficiently long time difference ΔT. In (1.2), $w(t)$ is the standard Brownian motion satisfying $dw^2 = 2dt$. The method converges when the stochastic process is ergodic. We refer the readers to [28] for more details about the theory of ergodicity.

When $S(x)$ is complex, the theory of the above method breaks down, since Z is no longer a partition function. Interestingly, the above algorithm can still be carried out, at least formally, if the functions $O(x)$ and $S(x)$ can be extended to the complex plane. Now we assume that both $O(x)$ and $S(x)$ are analytic and can be extended to $\mathbb{C}$ holomorphically. By naming the new functions as $O(z)$ and $S(z)$, we can still carry

out the above process by the replacement $x \to z$ and $X_k \to Z_k$. More specifically, if we denote z by $x + iy$, $x, y \in \mathbb{R}$, then the integral (1.1) is approximated by

$$(1.3) \qquad \langle O \rangle \approx \frac{1}{N} \sum_{k=1}^{N} O(X_k + iY_k),$$

where the samples X_k and Y_k are generated by simulating the following CL equation

$$(1.4) \qquad \begin{cases} dx = K(x,y)\,dt + dw, & K(x,y) = \mathrm{Re}(-S'(x+iy)), \\ dy = J(x,y)\,dt, & J(x,y) = \mathrm{Im}(-S'(x+iy)), \end{cases}$$

and choosing $X_k = x(T + k\Delta T)$ and $Y_k = y(T + k\Delta T)$. For the initial condition, we require that $y(0) = 0$ and $x(0)$ can be an arbitrary real number. This is the CL method.

The validity of the CL method is usually studied using the dual Fokker–Planck (FP) equation, which describes the evolution of the probability density function of $x(t)$ and $y(t)$ for the SDE (1.4). Using $P(x,y;t)$ to denote the joint probability of $x(t)$ and $y(t)$, we can derive from (1.4) that

$$(1.5) \qquad \partial_t P = L^T P, \qquad P(x,y;0) = p(x)\delta(y),$$

where $p(x)$ is a probability density function on the real axis, and L^T represents the FP operator:

$$(1.6) \qquad L^T P = \partial_{xx} P - \partial_x(KP) - \partial_y(JP).$$

Thus the right-hand side of (1.3) converges to the quantity

$$(1.7) \qquad \lim_{t \to +\infty} \int_{\mathbb{R}} \int_{\mathbb{R}} O(x+iy)P(x,y;t)\,dx\,dy.$$

The justification of the CL method requires us to check whether the above quantity equals $\langle O \rangle$.

Such equivalence has been shown in [7, 36] under certain conditions. Here we would like to restate the result as a rigorous theorem, which requires the following assumptions on the observable function $O(x)$ and the "complex probability function":
(H1) Let $\mathcal{O}(x,y;t)$ be the solution of the backward Kolmogorov equation

$$\partial_t \mathcal{O} = L\mathcal{O}, \qquad \mathcal{O}(x,y;0) = O(x+iy),$$

where $L = \partial_{xx} + K\partial_x + J\partial_y$. It holds that

$$(1.8) \quad \int_{\mathbb{R}} \int_{\mathbb{R}} \mathcal{O}(x,y;\tau)P(x,y;t-\tau)\,dx\,dy = \int_{\mathbb{R}} \int_{\mathbb{R}} O(x+iy)P(x,y;t)\,dx\,dy$$

for any $0 \le \tau \le t$.
(H2) The "forward Kolmogorov equation" for the complex-valued function $\rho(x;t)$

$$(1.9) \qquad \partial_t \rho = \partial_x(S'(x)\rho) + \partial_{xx}\rho, \qquad \rho(x;0) = p(x)$$

has the unique steady-state solution

$$\lim_{t \to +\infty} \rho(x;t) = \rho_\infty(x) = \frac{1}{Z}e^{-S(x)}.$$

(H3) For any $0 \leq \tau \leq t$, it holds that

$$\lim_{X \to \infty} [\mathcal{O}(X, 0; \tau)\partial_x \rho(X; t - \tau) - \rho(X; t - \tau)\partial_x \mathcal{O}(X, 0; \tau)]$$
$$= \lim_{X \to \infty} S'(X)\mathcal{O}(X, 0; \tau)\rho(X; t - \tau) = 0.$$

Based on the above assumptions, the following theorem implies the validity of the CL method.

THEOREM 1.1. *Assume that the conditions* (H1) *and* (H3) *hold. Then for any* $t > 0$,

$$(1.10) \qquad \int_{\mathbb{R}} \rho(x, t)O(x)\, dx = \int_{\mathbb{R}} \int_{\mathbb{R}} P(x, y; t)O(x + iy)\, dx\, dy.$$

In the above theorem, we can take the limit $t \to +\infty$ on both sides of (1.10). By the assumption (H2), one sees that (1.7) equals $\langle O \rangle$, which justifies the CL method. For comprehensiveness, we provide the proof of the theorem in Appendix A. The CL method computes the right-hand side of (1.10), which no longer includes rapidly sign-changing functions as long as the observable $O(x + iy)$ is not oscillatory. Thereby the numerical sign problem is mitigated.

Unfortunately, this theorem is far from satisfactory since the hypotheses are difficult to justify. In fact, these hypotheses are often too ideal such that they are often violated, resulting in some mysterious behaviors of the CL method. In applications, we often find the CL method fails to work due to divergence. Even worse, sometimes the algorithm appears to be convergent, but as the number of samples N increases, the right-hand side of (1.3) does not converge to its left-hand side, leading to a biased numerical result. This means that the conditions of Theorem 1.1, which look reasonable for the Langevin method, are often too strong in the application of the CL method. An example will be given in the next subsection.

Remark 1. In applications of the CL method in field theories, the function $\exp(-S)$ is usually provided by the product of a real part $\exp(-S_0)$ and the fermionic determinant. However, the fermionic determinant may contain zeros, and as a result, the "complex drift velocity"

$$K(x, y) + iJ(x, y) = \frac{1}{\exp(-S(x + iy))} \frac{\mathrm{d}}{\mathrm{d}x} \exp(-S(x + iy))$$

includes poles. In this case, Theorem 1.1 still holds if the poles do not appear on $\mathbb{R}$. In this paper, we temporarily restrict ourselves to the simpler holomorphic case, and the difference will be revealed in Remark 2 at the end of section 2.3. Note that Theorem 1.1 can be generalized to the multidimensional case without difficulty.

1.2. Failure of the CL method. The failure of the CL method has been demonstrated in a number of previous works [34, 29, 36]. Here we adopt the example used in [3, 29] to demonstrate such a phenomenon. We choose the observable as $O(x) = x^2$ and the complex action as

$$(1.11) \qquad S(x) = \frac{1}{2}(1 + iB)x^2 + \frac{1}{4}x^4, \quad B \in \mathbb{R}.$$

It has been found in [29] that the CL method converges to the correct expectation values when B is small, while for large B, the CL method may still converge, but the
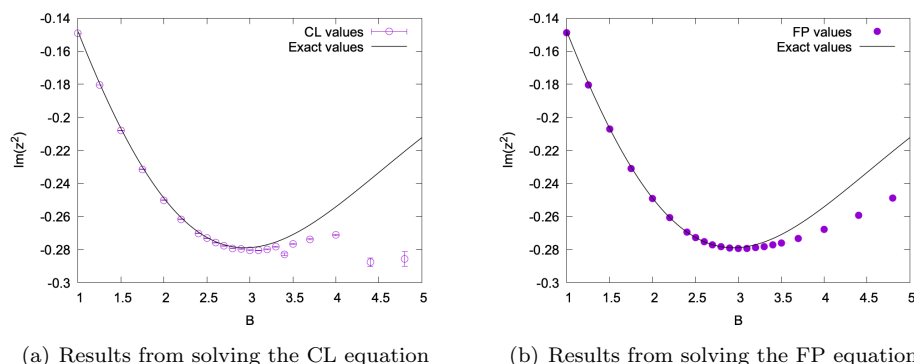
(a) Results from solving the CL equation   (b) Results from solving the FP equation

FIG. 1. *The imaginary part of the expectation value $O(z) = z^2$.*

limiting value differs from the integral (1.1). We have redone the numerical experiment by simulating the stochastic equation (1.4) for $B = 1$ to 5, and the results are given in Figure 1(a), showing the same behavior as in [29]. Furthermore, we find that for $B > 4$, the simulation becomes unstable in the sense that arithmetic overflow often appears, and for stable simulations, the CL result deviates from the exact integral significantly. To better confirm the phenomenon, we solve the FP equation (1.5) numerically using the method to be introduced in subsection 2.1 (see Figure 1(b)). The disagreement between the two figures for large B also implies the lack of reliability for the CL simulation.

For this specific example, it is known by the analysis in [3] that for $B \leq \sqrt{3}$, there exists $y_- > 0$ such that $P(x, y; t) = 0$ for any x and t if $|y| > y_-$. Furthermore, the authors of [3] proposed the decay rate $\exp(-ax^4)$ for the marginal probability density function of x. The fast decay of the probability density function on the complex plane implies that the conditions (H1)–(H3) hold true, so that Theorem 1.1 guarantees the validity of the CL method. In this situation, following [3], we consider the probability density function as "localized." In the literature, the localization of the probability may refer to a sufficient decay of the probability density function such that the samples that drift far away from the real axis can be ignored. In this sense, as will be discussed in the next section, the admissible decay rate depends on the increasing rate of the observable, which will introduce significant difficulty to our analysis. Therefore in this paper, we only consider the localization of the probability in a strong sense, i.e., the probability density function is said to be localized if and only if there exists $Y \in \mathbb{R}^+$ such that $P(x, y) = 0$ for any $x \in \mathbb{R}$ and $|y| > Y$. Here we focus mainly on the localization in the y-direction since in the applications of the CL method (mainly QCD), the x-direction is usually a compact group, as will be detailed in section 3.

However, the behavior of the solution for $B > \sqrt{3}$ remains unclear. In [3], the authors predicted a power decay of the probability function when $B > \sqrt{3}$, but they left only a vague comment saying that this is "an important signal of failure." However, in [29], the authors argued by numerical experiment that the probability density function falls off exponentially for $B \leq 2.6$, implying again that Theorem 1.1 holds and the CL method is unbiased, and when $B > 2.8$, power decay is observed, indicating the failure of the CL method.

These conflicting results for this model problem also imply the lack of understanding of the CL method. In this work, we will begin our discussion from this toy example and try to answer the following questions:

- How does the power decay of the probability affect the numerical value of the expectation? How is it related to the conditions of Theorem 1.1?
- What is the critical value of B from which the CL result deviates from the exact integral?
- What occurs to the distribution function when B passes this critical value? Is it a smooth transition?

After getting more mathematical insights into this model problem, we will further study the phenomenon of localized probability density functions in the lattice field theory. The rest of the paper is organized as follows. Section 2 is devoted to a comprehensive study of the model problem (1.11). The formulation of the CL method in lattice field theories will be provided in section 3, where we will also demonstrate the non-existence of localized probability density functions. In section 4, we study the effect of the gauge cooling technique in localizing the probability density functions. Finally, some concluding remarks are given in section 5.

2. A study of the model problem. This section is devoted to a closer look at the integration problem with action (1.11), for which the drift velocities are

$$(2.1) \quad K(x,y) = -\left(x - By + x^3 - 3xy^2\right), \qquad J(x,y) = -\left(y + Bx + 3x^2y - y^3\right).$$

Following [29], we study the complexified observable $O(z) = z^2 = x^2 - y^2 + i2xy$. To understand the phenomenon shown in Figure 1, we will first resolve the ambiguity in the literature about the decay rate of the probability density function; this will be done by solving the FP equation (1.5) numerically.

2.1. Numerical method for the Fokker–Planck equation. The FP equation for the CL equation has been numerically solved in [3, 36]. It is found in [36] that according to the CFL condition, very small time steps need to be adopted by explicit schemes due to the large values of velocities when $|x|$ and $|y|$ are large. Therefore in our study, we take the numerical method of characteristics to achieve higher efficiency.

For any given $x(0)$ and $y(0)$, the characteristic curve for (1.5) starting from this point is given by the following equations:

$$(2.2) \quad x'(t) = K(x,y), \quad y'(t) = J(x,y).$$

It can be derived from (1.5) and (2.2) that on these characteristic curves, the probability density function P satisfies the following differential equation:

$$(2.3) \quad \frac{d}{dt}\left(e^{\int_0^t f(x(s),y(s))ds}P(x(t),y(t);t)\right) = e^{\int_0^t f(x(s),y(s))ds}\partial_{xx}P(x(t),y(t);t),$$

where $f(x,y) = \partial_x K(x,y) + \partial_y J(x,y)$. Based on such a form, we apply the backward Euler method to obtain the following semidiscretization of (2.3):

$$P(x(t_{k+1}),y(t_{k+1});t_{k+1}) - \Delta t\,\partial_{xx}P(x(t_{k+1}),y(t_{k+1});t_{k+1})$$
$$(2.4) \qquad = e^{-\int_{t_k}^{t_{k+1}} f(x(s),y(s))ds}P(x(t_k),y(t_k);t_k),$$

where $\Delta t = t_{k+1} - t_k$ denotes the time step, and $(x(\cdot),y(\cdot))$ denotes a single characteristic curve. Therefore if we want to obtain $P(x_l,x_j;t_{k+1})$ for some specific point (x_l,x_j), we need to solve the characteristic curve between t_k and t_{k+1} by finding the solution of the backward system of ordinary differential equations

$$(2.5) \qquad \begin{cases} x'(t) = K(x,y), & x(t_{k+1}) = x_l, \\ y'(t) = J(x,y), & y(t_{k+1}) = y_j, \end{cases}$$

so that in the last term of (2.4), the values of $x(t_k)$ and $y(t_k)$ can be determined, and the integral of f can be computed. In our scheme, we solve the backward ODE system (2.5) by the classic Runge–Kutta scheme and integrate f using Simpson's rule. In fact, since K and J are independent of t, for any given (x_l, x_j), the point $(x(t_k), y(t_k))$ and the integral of f do not change if Δt does not change. Therefore, in our implementation, we just choose a fixed time step so that these quantities need to be computed only once for each spatial grid point (x_l, x_j). Numerically, the integral of $P(x, y; t)$ may deviate from one, and we scale the whole function after each time step by multiplying a constant to restore this property.

For the spatial discretization in our simulations, we adopt the finite difference method and the Fourier spectral method in different cases. The Fourier spectral method provides good accuracy for the derivatives, which are needed in the asymptotic expansions to be studied in subsection 2.4, while in the study of the decay of the probability, we adopt the finite difference method to avoid aliasing error appearing when periodizing the domain in the Fourier spectral method, which may ruin the tail of the probability density function. In what follows, we provide some details of these two methods.

2.1.1. Finite difference scheme. We adopt the uniform grid with $N \times M$ cells, each of which has the size of Δx and Δy in the x- and y-directions, respectively. Suppose that at time t_k the probability is $P(x, y; t_k)$, and we want to determine P at time t_{k+1} with a central difference scheme to approximate ∂_{xx}. Then the full discretization of (2.3) at point (x_l, y_j) is given as

(2.6)
$$P(x_l, y_j; t_{k+1}) - \frac{\Delta t}{(\Delta x)^2} \left[P(x_{l+1}, y_j; t_{k+1}) - 2P(x_l, y_j; t_{k+1}) + P(x_{l-1}, y_j; t_{k+1}) \right]$$
$$= \lambda_{l,j}(\Delta t) P(\tilde{x}_l, \tilde{y}_j; t_k),$$

where $(\tilde{x}_l, \tilde{y}_j) = (x(t_k), y(t_k))$ is obtained by solving (2.5), and $\lambda_{l,j}(\Delta t)$ is the exponential of the integral of f in (2.4). For any points located outside the computational domain, we set the value of P to be zero. In general, the point $(\tilde{x}_l, \tilde{y}_j)$ is not on the grid point, and the value of $P(\tilde{x}_l, \tilde{y}_j; t_k)$ is obtained from the bilinear interpolation of $P(x, y; t_k)$. Defining

$$\mathbf{P}_j^{(k+1)} = [P(x_1, y_j; t_{k+1}), P(x_2, y_j; t_{k+1}), \ldots, P(x_N, y_j; t_{k+1})]^\top, \quad j = 1, \ldots, M,$$

by (2.6), we are required to solve the linear systems $S\mathbf{P}_j^{(k+1)} = \mathbf{b}_j^{(k+1)}, j = 1, \ldots, M,$ with $S \in \mathbb{R}^{N \times N}$ being a tridiagonal matrix and $\mathbf{b}_j^{(k+1)}$ corresponding to the right-hand side of (2.6). These tridiagonal linear systems can be efficiently solved by the Thomas algorithm.

2.1.2. Fourier spectral method. To employ the Fourier spectral method, the probability is assumed to be periodic in both real and imaginary directions. This is reasonable if we choose a sufficiently large computational domain such that the probability is sufficiently small on the boundary. Suppose the domain is $[-L_x/2, L_x/2] \times [-L_y/2, L_y/2]$ and the number of Fourier coefficients is N, M in the x-, y-directions, respectively. The probability density function is approximated by

(2.7)
$$P(x, y; t) = \frac{1}{NM} \sum_{n=-N/2}^{N/2-1} \sum_{m=-M/2}^{M/2} \hat{P}_{n,m}(t) e^{\frac{2\pi i}{L_x} nx} e^{\frac{2\pi i}{L_y} my}.$$

Thus we can write (2.4) for $x(t_{k+1}) = x_l$ and $y(t_{k+1}) = y_j$ as

$$\frac{1}{NM} \sum_{n=-N/2}^{N/2-1} \sum_{m=-M/2}^{M/2} \left(1 + \frac{4\pi^2 n^2}{L_x^2} \Delta t\right) \hat{P}_{n,m}(t_{k+1}) e^{\frac{2\pi i}{L_x} n x_l} e^{\frac{2\pi i}{L_y} m y_j}$$
$$= \lambda_{l,j}(\Delta t) P(\tilde{x}_l, \tilde{y}_j; t_k).$$

Applying the discrete inverse Fourier transform on both sides, we get the scheme
(2.8)

$$\hat{P}_{n,m}(t_{k+1}) = \left(1 + \frac{4\pi^2 n^2}{L_x^2} \Delta t\right)^{-1} \sum_{l=1}^{N} \sum_{j=1}^{M} \lambda_{l,j}(\Delta t) P(\tilde{x}_l, \tilde{y}_j; t_k) e^{-\frac{2\pi i}{L_x} n x_l} e^{-\frac{2\pi i}{L_y} m y_j}.$$

Note that the computational cost of the above scheme is $O(M^2 N^2)$ since $\tilde{x}_l$ and $\tilde{y}_j$ are not collocation points.

2.2. Decay of the distribution. To resolve the ambiguity about the decay rate of steady-state probability density function $P(x,y) := P(x, y; \infty)$, we solve the FP equation for a sufficiently long time until the steady state is attained. For the purpose of visualization, we consider the marginal probability density functions

(2.9) $$P_x(x) = \int_{-\infty}^{\infty} P(x, y)\, dy, \quad P_y(y) = \int_{-\infty}^{\infty} P(x, y)\, dx,$$

and we focus mainly on the cases with B close to $\sqrt{3}$ where conflicting results are observed in the literature as stated in subsection 1.2. The quantities (2.9) are plotted in Figure 2 for B from 1.5 to 2.3. From the figure, we observe the following phenomena: (i) $P_y(y)$ drops rapidly for $B \leq 1.7$; (ii) when B surpasses 1.8, the tail of $P_y(y)$ starts to rise up; (iii) for $B \geq 2.1$, both $P_y(y)$ and $P_x(x)$ show the power-like decay. These results contradict the statement in [30] that the power decay shows up only when B is greater than 2.6. In our results, such decay is obvious as early as $B = 2.0$. In fact, we conjecture that such power decay appears immediately when B exceeds $\sqrt{3}$. It is not obvious in the numerical experiments only because of the small coefficient in front of the power decay. Such an argument can be supported by some analysis of the FP equation, which will be clarified in the following two parts.

2.2.1. Possibility of the exponential decay ($B \leq \sqrt{3}$). In the steady-state FP equation $\partial_x(KP) + \partial_y(JP) = \partial_{xx}P$, both K and J are polynomials, which inspires us to conjecture that the solution $P(x,y)$ may behave like the exponential of a polynomial when x or y is large, so that $P(x,y)$ decays exponentially. Specifically,
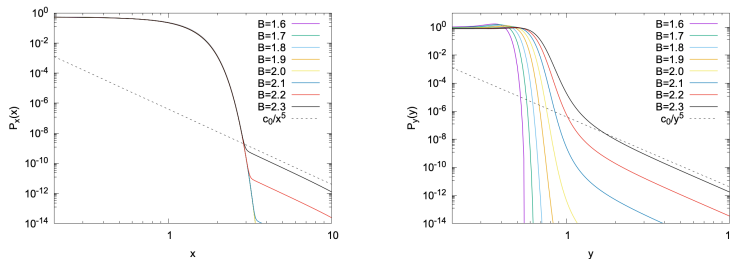

FIG. 2. $P_x(x)$ and $P_y(y)$ for different B.

we can write $P(x, y)$ as

$$P(x, y) \approx \exp(-Ar^\beta)g(\theta) \qquad \text{for } r \gg 1,$$

where (r, θ) is the polar coordinates of (x, y). Straightforward computation yields

$$
\begin{aligned}
L^T P \approx \Big(& \frac{g(\theta)}{2}[\beta^2 A r^{\beta-2}(Ar^\beta - 1) + (\beta^2 A^2 r^{2\beta-2} - 2\beta A r^{\beta+2} - (\beta-2)\beta A r^{\beta-2} \\
& + 12r^2)\cos 2\theta - 2\beta A r^\beta + 4] + g'(\theta)[(\beta A r^{\beta-2} + r^2 + r^{-2}) + B]\sin 2\theta \\
& + g''(\theta)r^{-2}\sin^2\theta \Big) \exp(-Ar^\beta) \qquad \text{for } r \gg 1.
\end{aligned}
$$

When $r \to +\infty$, the term with slowest decay behaves like

$$\varphi(r) = r^{\max(\beta+2, 2\beta-2)}\exp(-Ar^\beta).$$

By focusing on this leading order term, we have

$$
\frac{L^T P(x, y)}{\varphi(r)} \approx
\begin{cases}
-\beta A g(\theta)\cos 2\theta & \text{if } \beta < 4, \\
-\beta A g(\theta)\cos 2\theta + \beta^2 A^2 g(\theta)\cos^2\theta & \text{if } \beta = 4, \\
\beta^2 A^2 g(\theta)\cos^2\theta & \text{if } \beta > 4.
\end{cases}
$$

Since $P(x, y)$ is the steady-state solution of the FP equation, the above quantity must equal zero. For any α, this requires that $g(\theta)$ be zero for almost every θ.

If $\beta < 4$, the value of $g(\theta)$ can be nonzero for $\theta = \pm\pi/4$ and $\theta = \pm 3\pi/4$. This means that $P(x, y)$ can be nonzero in two strips parallel to the lines $y = \pm x$. However, such strips cannot be formed due to the diffusion in the x direction. Similarly, if $\beta > 4$, $P(x, y)$ can have nonzero values in the strip perpendicular to the x-axis, which is also not allowed because of the Brownian motion. This excludes the choices $\beta < 4$ and $\beta > 4$.

In fact, such an FP equation only allows the localizing strip to be parallel to the x-axis, which corresponds to $\theta = 0$ and $\theta = \pi$. When $\beta = 4$, choosing $A = 1/4$ allows us to have $g(\theta) \neq 0$ when either $\theta = 0$ or $\theta = \pi$ holds. As a summary, such analysis shows that if $P(x, y)$ has exponential decay, the only possible choice of β is 4, and in this case, the support of $P(x, y)$ must be confined in a strip-like domain parallel to the x-axis. This occurs when $B \leq \sqrt{3}$, as can be demonstrated in the following proposition.

PROPOSITION 2.1. *Suppose* $0 \leq B \leq \sqrt{3}$. *There exists a constant* $\alpha > 0$ *such that* $J(x, y)$ *defined in* (2.1) *satisfies the following conditions:*
   (i) $J(x, \alpha) \leq 0$ *for all* $x \in \mathbb{R}$.
   (ii) $J(x, -\alpha) \geq 0$ *for all* $x \in \mathbb{R}$.

Proof. Here we only show that $J(x, \alpha) \leq 0$ for all $x \in \mathbb{R}$, and the proof of the other part is almost identical. For simplicity, we define

$$Q(x) := J(x, \alpha) = -3\alpha x^2 - Bx + \alpha^3 - \alpha.$$

The function $Q(x)$ is a quadratic polynomial, whose maximum value can be obtained as

$$(2.10) \qquad \max_{x \in \mathbb{R}} Q(x) = \frac{B^2}{12\alpha} + \alpha^3 - \alpha = \frac{1}{\alpha}\left[\left(\alpha^2 - \frac{1}{2}\right)^2 + \frac{B^2}{12} - \frac{1}{4}\right].$$

When $0 \leq B \leq \sqrt{3}$, we can choose

$$(2.11) \qquad \alpha = \frac{1}{\sqrt{2}} \sqrt{1 - \sqrt{1 - \frac{B^2}{3}}}$$

so that $\max_{x \in \mathbb{R}} Q(x)$ is exactly zero, meaning that $Q(x)$ is always nonpositive, which completes the proof of the proposition. $\square$

This proposition shows that when $0 \leq B \leq \sqrt{3}$, the solution of (1.4) satisfies $y(t) \in [-\alpha, \alpha]$ if the initial condition $y(0) \in [-\alpha, \alpha]$. This can be illustrated by plotting the velocity field $(K(x, y), J(x, y))$, which shows that on the lines $y = \pm\alpha$, all the velocities point toward the horizontal axis. Thus $(x(t), y(t))$ can never drift out of the strip between these two lines, causing zero values of $P(x, y)$ for all $|y| > \alpha$. In other words, the distribution $P(x, y)$ has a compact support $[-\alpha, \alpha]$ in the imaginary direction, and our previous analysis shows that the decay rate in the x-direction is like $\exp(-x^4/4)$. The right panel of Figure 2 also validates the existence of such a strip.

2.2.2. Possibility of the power decay ($B > \sqrt{3}$). For completeness, we apply a similar analysis to demonstrate the rate of the power decay. Such analysis has already been done in [3], while the angular function is not included. Here we will carry out a more rigorous analysis by assuming

$$P(x, y) \approx r^{-\beta} g(\theta) \qquad \text{for } r \gg 1.$$

Thus

$$L^T P \approx r^{-\beta-2} \left(g''(\theta) \sin^2(\theta) + g'(\theta) \left(Br^2 + (\beta + r^4 + 1) \sin 2\theta \right) \right.$$

$$\left. + \frac{1}{2} g(\theta) \left[\beta^2 + \left(\beta(\beta + 2) - 2(\beta - 6)r^4 \right) \cos(2\theta) - 2(\beta - 2)r^2 \right] \right)$$

$$\approx [(6 - \beta)g(\theta) \cos 2\theta + g'(\theta) \sin 2\theta] r^{2-\beta}.$$

Therefore $(6 - \beta)g(\theta) \cos 2\theta + g'(\theta) \sin 2\theta = 0$, whose solution is $g(\theta) = C(\sin 2\theta)^{\beta/2-3}$ for any $C \in \mathbb{R}$. Only when $\beta = 6$, the positivity of $P(x, y)$ can be guaranteed. Thus we conclude that

$$P(x, y) \approx \frac{C}{(x^2 + y^2)^3} \qquad \text{for } x^2 + y^2 \gg 1,$$

which is clearly not a localized probability density function and should correspond to any B exceeding the critical value $\sqrt{3}$. As a result, the decay of the marginal probability density functions (2.9) behaves like

$$P_x(x) \propto x^{-5} \qquad \text{for } |x| \gg 1,$$
$$P_y(y) \propto y^{-5} \qquad \text{for } |y| \gg 1,$$

which has been numerically verified as shown in Figure 2. The analysis further confirms that the absence of power decay in the numerical results of $B = 1.8$ and 1.9 is due to the smallness of C. As we will see later in section 2.4.2, the values of the probability density function may have dropped below the machine epsilon when the power decay shows up, so that the numerical method is not able to capture such a decay rate.

2.3. Effect of the fat-tailed distribution. Knowing that $B = \sqrt{3}$ separates the two types of decay rates, we would like to study how this affects the observables. When B is large, since the CL method no longer converges to the exact integral, we conclude that at least one of the assumptions (H1)–(H3) is violated. Among the three conditions, the only one that may be related to the tail of $P(x,y)$ is (H1). In fact, in [7], instead of given as a condition, the assumption (H1) is derived as follows:

$$
\begin{aligned}
&\frac{\partial}{\partial \tau} \int_{\mathbb{R}} \int_{\mathbb{R}} \mathcal{O}(x,y;\tau) P(x,y;t-\tau) \, dx \, dy \\
(2.12) \quad &= \int_{\mathbb{R}} \int_{\mathbb{R}} \left[\frac{\partial \mathcal{O}(x,y;\tau)}{\partial \tau} P(x,y;t-\tau) + \mathcal{O}(x,y;\tau) \frac{\partial P(x,y;t-\tau)}{\partial \tau} \right] dx \, dy \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} [P(x,y;t-\tau) L \mathcal{O}(x,y;\tau) - \mathcal{O}(x,y;\tau) L^T P(x,y;t-\tau)] \, dx \, dy,
\end{aligned}
$$

which equals zero due to the formal mutual adjointness of L and L^T. However, it has also been pointed out in [7, 36] that the above quantity may not vanish if $P(x,y;t)$ does not have sufficient decay. Specifically, in order that (2.12) equals zero, we need the following limits:

(2.13)

$$
\lim_{X \to \infty} \int_{\mathbb{R}} \frac{\partial \mathcal{O}(X,y;\tau)}{\partial x} P(X,y;t-\tau) \, dy = \lim_{X \to \infty} \int_{\mathbb{R}} \frac{\partial P(X,y;\tau)}{\partial x} \mathcal{O}(X,y;t-\tau) \, dy = 0,
$$

(2.14)

$$
\lim_{X \to \infty} \int_{\mathbb{R}} K(X,y) \mathcal{O}(X,y;\tau) P(X,y;t-\tau) \, dy
$$

$$
= \lim_{Y \to \infty} \int_{\mathbb{R}} J(x,Y) \mathcal{O}(x,Y;t-\tau) P(x,Y;\tau) \, dx = 0.
$$

Only when these limits hold for all t and τ can we ensure that integration by parts without boundary terms can be carried out to show that (2.12) equals zero.

We focus on the second integral in (2.14) and the other limits can be considered in a similar way. Since the limit (2.14) must hold for all t and τ, a necessary condition for the validity of the CL method can be obtained by setting $t = \tau$ and letting τ tend to infinity, which yields

(2.15)

$$
\lim_{y \to \infty} E(y) = 0, \quad E(y) := \int_{\mathbb{R}} J(x,y) \mathcal{O}(x,y;0) P(x,y;\infty) \, dx.
$$

Here $P(x,y;\infty)$ is just the function $P(x,y)$ as stated in the beginning of section 2.2, and now it is clear that (2.15) holds only when $P(x,y)$ has sufficient decay when $y \to \infty$. In our case, when $B > \sqrt{3}$ and $|y|$ is large,

$$
E(y) \approx -\int_{\mathbb{R}} (y + Bx + 3x^2 y - y^3)(x^2 - y^2 + i2xy) \frac{C}{(x^2 + y^2)^3} \, dx = -\frac{C\pi(iB + 4y^2 - 1)}{4y^2},
$$

whose real part does not vanish as y tends to infinity. Similarly, the other limit in (2.14) does not hold either. As a result, biased results are generated. Such a phenomenon has also been numerically validated in Figure 3. When $B \le \sqrt{3}$, due to the fast decay of $P(x,y)$, we observe unbiased results in the simulations.

Remark 2. If the velocities K and J contain poles, then the conditions (2.13)–(2.14) must be supplemented by the corresponding boundary conditions at poles. We refer the readers to a recent paper [40] for some discussions on such cases.
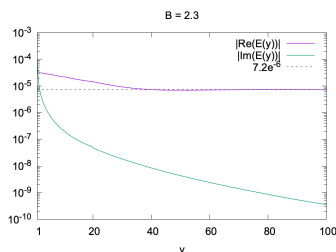
FIG. 3. *The decay of $E(y)$ for $B = 2.3$.*

2.4. Asymptotics near $B = \sqrt{3}$. By this example, we would like to reveal more on what happens at the critical point $B = \sqrt{3}$. Figure 1 shows that the two curves, which initially coincide, eventually separate from each other as B increases, meaning that at least one of the curves is not analytic at point $B = \sqrt{3}$. In this section, we would like to study the asymptotics around this point and explain how the analyticity fails.

The most straightforward idea to study the asymptotics is to set $B^\epsilon = \sqrt{3} - \epsilon$ and expand the associated probability density function $P^\epsilon(x, y)$ by

$$(2.16) \qquad P^\epsilon(x, y) = P_0(x, y) + \epsilon P_1(x, y) + \epsilon^2 P_2(x, y) + \cdots .$$

By setting $\epsilon = 0$, we see that $P^0(x, y)$ corresponds to the probability density function for $B = \sqrt{3}$, and it has been shown in the proof of Proposition 2.1 that

$$(2.17) \qquad \mathrm{supp}_y P^\epsilon = [-\alpha^\epsilon/\sqrt{2}, \alpha^\epsilon/\sqrt{2}], \quad \alpha^\epsilon = \sqrt{1 - \sqrt{1 - (B^\epsilon)^2/3}}.$$

Unfortunately, such an expansion does not converge due to the following proposition.

PROPOSITION 2.2. *Let $P^\epsilon(y)$ be a class of functions defined for every $\epsilon \in (0, \delta)$, and the functions satisfy $\mathrm{supp}\, P^\epsilon(y) = [-\alpha^\epsilon, \alpha^\epsilon]$. Then there does not exist a sequence of functions $\{P_0(y), P_1(y), \cdots\}$ such that the infinite series*

$$(2.18) \qquad \sum_{n=0}^{+\infty} \epsilon^n P_n(y)$$

converges pointwisely to $P^\epsilon(y)$ for any $\epsilon \in (0, \delta)$.

Proof. Note that α^ϵ monotonically decreases as ϵ increases. Suppose the series (2.18) converges to $P^\epsilon(y)$ for any $\epsilon \in (0, \delta)$; we know that for $\epsilon \in (\delta/2, \delta)$ and $y \in (\alpha^{\delta/2}, \alpha^0)$, it holds that

$$P^\epsilon(y) = \sum_{n=0}^{+\infty} \epsilon^n P_n(y) = 0.$$

Regarding the series as the power series with respect to ϵ, we know that for $y \in (\alpha^{\delta/2}, \alpha^0)$, the value of $P^\epsilon(y)$ must be zero for any $\epsilon \in (0, \delta)$. This contradicts the assumption that $\mathrm{supp}\, P^\epsilon = [-\alpha^\epsilon, \alpha^\epsilon]$ for $\epsilon \in (0, \delta/2)$. $\qquad \square$

The above result shows that in Figure 1, the curve of observable predicted by the CL method is likely to be nonanalytic. Below we will provide a legitimate asymptotic expansion for $P^\epsilon(x, y)$ based on the knowledge of its support.

2.4.1. The asymptotic expansion. To avoid the divergence problem arising from the variation of the support, we are going to scale the variable y to align the support of $P^\epsilon(x, y)$ for any ϵ. Precisely, we let $a^\epsilon = 1/\alpha^\epsilon$ and construct a new function $\tilde{P}^\epsilon$ as

$$(2.19) \qquad \qquad \tilde{P}^\epsilon(x, y) = P^\epsilon(x, a^\epsilon y)$$

such that $\mathrm{supp}_y \tilde{P}^\epsilon = [-1/\sqrt{2}, 1/\sqrt{2}]$ for any ϵ, and the original probability density function can be reconstructed as $P^\epsilon(x, y) = \tilde{P}^\epsilon(x, \alpha^\epsilon y)$. Thus the governing equation of $P^\epsilon(x, y)$ is

$$(2.20) \qquad \frac{\partial}{\partial t}\tilde{P}^\epsilon + \frac{\partial}{\partial x}(\tilde{K}^\epsilon \tilde{P}^\epsilon) + \alpha^\epsilon \frac{\partial}{\partial y}(\tilde{J}^\epsilon \tilde{P}^\epsilon) = \frac{\partial^2}{\partial x^2}\tilde{P}_\epsilon,$$

in which $\tilde{K}^\epsilon(x, y) := K(x, a^\epsilon y)$ and $\tilde{J}^\epsilon(x, y) := J(x, a^\epsilon y)$. Note that the Taylor expansion of a^ϵ is

$$(2.21) \qquad a^\epsilon = 1 + a_1 \epsilon^{\frac{1}{2}} + a_2 \epsilon + \cdots, \qquad a_1 = \frac{\sqrt{2}}{\sqrt[4]{3}}, \quad a_2 = \frac{2}{\sqrt{3}}, \quad \ldots.$$

Therefore, we should expand all the quantities with respect to $\sqrt{\epsilon}$ instead of ϵ. Let

$$(2.22) \qquad \tilde{P}^\epsilon = \sum_{k=0}^{+\infty} \epsilon^{\frac{k}{2}} \tilde{P}_{\frac{k}{2}}, \quad K^\epsilon = \sum_{k=0}^{+\infty} \epsilon^{\frac{k}{2}} \tilde{K}_{\frac{k}{2}}, \quad J^\epsilon = \sum_{k=0}^{+\infty} \epsilon^{\frac{k}{2}} \tilde{J}_{\frac{k}{2}}.$$

One can figure out all the terms $\tilde{K}_{\frac{k}{2}}$ and $\tilde{J}_{\frac{k}{2}}$ by the analytical expressions of $K(x, y)$ and $J(x, y)$. Then by balancing the terms with various orders of ϵ in (2.20), we can obtain the equations for $\tilde{P}_{\frac{k}{2}}$:

$$\mathcal{O}(1): \quad \frac{\partial}{\partial t}\tilde{P}_0 + \frac{\partial}{\partial x}(\tilde{K}_0 \tilde{P}_0) + \frac{\partial}{\partial y}(\tilde{J}_0 \tilde{P}_0) = \frac{\partial^2}{\partial x^2}\tilde{P}_0;$$

$$\mathcal{O}(\epsilon^{1/2}): \quad \frac{\partial}{\partial t}\tilde{P}_{\frac{1}{2}} + \frac{\partial}{\partial x}\left(\tilde{K}_0 \tilde{P}_{\frac{1}{2}}\right) + \frac{\partial}{\partial y}\left(\tilde{J}_0 \tilde{P}_{\frac{1}{2}}\right) + \frac{\partial}{\partial x}\left(\tilde{K}_{\frac{1}{2}} \tilde{P}_0\right)$$

$$(2.23) \qquad \qquad + \frac{\partial}{\partial y}\left(\tilde{J}_{\frac{1}{2}} \tilde{P}_0 - a_{\frac{1}{2}} \tilde{J}_0 \tilde{P}_0\right) = \frac{\partial^2}{\partial x^2}\tilde{P}_{\frac{1}{2}};$$

$$\mathcal{O}(\epsilon): \quad \frac{\partial}{\partial t}\tilde{P}_1 + \frac{\partial}{\partial x}(\tilde{K}_0 \tilde{P}_1) + \frac{\partial}{\partial y}(\tilde{J}_0 \tilde{P}_1) + \frac{\partial}{\partial x}\left(\tilde{K}_{\frac{1}{2}} \tilde{P}_{\frac{1}{2}} + \tilde{K}_1 \tilde{P}_0\right)$$

$$+ \frac{\partial}{\partial y}\left(\tilde{J}_1 \tilde{P}_0 + \tilde{J}_{\frac{1}{2}} \tilde{P}_{\frac{1}{2}} - a_{\frac{1}{2}} \tilde{J}_0 \tilde{P}_{\frac{1}{2}} + \left(a_{\frac{1}{2}}^2 - a_1\right) \tilde{J}_0 \tilde{P}_0\right) = \frac{\partial^2}{\partial x^2}\tilde{P}_1;$$

$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots.$$

Thus we can obtain each $\tilde{P}_{k/2}$ by solving these equations. Due to the appearance of $\epsilon^{1/2}$, when this expansion is extended to negative ϵ, the function $\tilde{P}^\epsilon(x, y)$ no longer represents a probability density function, as implies the nonanalyticity of the solution provided by the CL method.

The convergence of the series (2.22) can be verified numerically. Instead of solving (2.23) directly, we adopt an alternative way to find the functions $\tilde{P}_{k/2}$. In fact, these quantities are related to the formal expansion (2.16), whose terms satisfy the equations
(2.24)
$$\frac{\partial}{\partial t}P_l + \frac{\partial}{\partial x}(K_0 P_l) + \frac{\partial}{\partial y}(J_0 P_l) - \frac{\partial}{\partial x}(y P_{l-1}) + \frac{\partial}{\partial y}(x P_{l-1}) = \frac{\partial^2}{\partial x^2}P_l, \quad l = 1, 2, \ldots,$$

which can be derived by inserting (2.16) into the FP equation (1.5) and balancing the terms with the same orders of ϵ. It can be verified that by setting

$$(2.25) \quad \tilde{P}_0 = P_0, \qquad \tilde{P}_{\frac{1}{2}} = a_{\frac{1}{2}} y \frac{\partial P_0}{\partial y}, \qquad \tilde{P}_1 = P_1 + a_1 y \frac{\partial P_0}{\partial y} + \frac{1}{2} a_{\frac{1}{2}}^2 y^2 \frac{\partial^2 P_0}{\partial y^2}, \dots,$$

we can obtain the solutions to (2.23). The method we adopt is to solve (2.24) numerically, and then use (2.25) to convert the results to $\tilde{P}_{k/2}$. In this process, high-order derivatives of the solutions are needed in (2.25), which requires us to adopt the Fourier spectral method described in subsection 2.1.2 to find the numerical solutions.

The computational details are stated as follows. The computational domain is set to be $[-5, 5] \times [-1, 1]$, which is sufficiently large since $P^\epsilon(x, y)$ decays fast in the x-direction and has a compact support in the y-direction. We choose $N = 480$ and $M = 240$ in (2.7), and after solving P_0 from the original FP equation, we compute P_1 to P_5 successively by solving (2.24). All the equations are solved until a steady state is attained. Based on these functions, one can compute $\tilde{P}_{k/2}$ up to $k = 11$. Then the results are inserted back into (2.22) with the infinite series truncated. In Figure 4, we plot the function $P^\epsilon(x, y)$ with $\epsilon = 0.02$ approximated by different truncations, which shows the convergence of the series of $\tilde{P}^\epsilon$ given in (2.22). In particular, we observe that some oscillations appearing in the early truncations are suppressed as we increase the number of terms.

Now we study the analyticity of the curves shown in Figure 1. The expectation value for the observable can be related to the scaled function $\tilde{P}^\epsilon(x, y)$ by

$$\langle O \rangle^\epsilon = \int_{\mathbb{R}} \int_{\mathbb{R}} O(x + iy) P^\epsilon(x, y) \, dx \, dy = a^\epsilon \int_{\mathbb{R}} \int_{\mathbb{R}} O(x + ia^\epsilon y) \tilde{P}^\epsilon(x, y) \, dx \, dy.$$

The expansion of $O(x + ia^\epsilon y)$ can be obtained by substituting the expansion of a^ϵ (2.21) to the observable function $O(x + iy) = (x + iy)^2$, which then yields an expansion of $\langle O \rangle^\epsilon$,

$$(2.26) \qquad \langle O \rangle^\epsilon = \sum_{k=0}^{+\infty} \epsilon^{\frac{k}{2}} \langle O \rangle_{\frac{k}{2}},$$
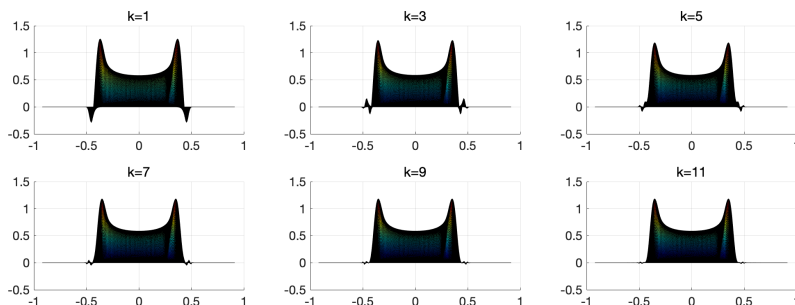

FIG. 4. *The distribution constructed by expansion* (2.22).

where the coefficients $\langle O \rangle_{k/2}$ can be evaluated by

$$\langle O \rangle_0 = \int_{\mathbb{R}} \int_{\mathbb{R}} O_0 \tilde{P}_0 \, dx \, dy, \qquad \langle O \rangle_{\frac{1}{2}} = \int_{\mathbb{R}} \int_{\mathbb{R}} (O_0 \tilde{P}_{\frac{1}{2}} + O_{\frac{1}{2}} \tilde{P}_0 + a_{\frac{1}{2}} O_0 \tilde{P}_0) \, dx \, dy,$$

$$\langle O \rangle_1 = \int_{\mathbb{R}} \int_{\mathbb{R}} \left(O_0 \tilde{P}_1 + O_{\frac{1}{2}} \tilde{P}_{\frac{1}{2}} + O_1 \tilde{P}_0 + a_{\frac{1}{2}} (O_0 \tilde{P}_{\frac{1}{2}} + O_{\frac{1}{2}} \tilde{P}_0) + a_1 O_0 P_0 \right) \, dx \, dy, \; \cdots .$$

Some numerical values of $\langle O \rangle_{k/2}$ are tabulated in Table 1, from which one can see that $\langle O \rangle_{k/2}$ with half-integer orders are all very small. Actually, using the relations (2.25), one can show that

$$\langle O \rangle_{\frac{k}{2}} = \begin{cases} \int_{\mathbb{R}} \int_{\mathbb{R}} O(x + iy) P_{\frac{k}{2}}(x, y) \, dx \, dy & \text{if } k \text{ is even}, \\ 0 & \text{if } k \text{ is odd}. \end{cases}$$

This indicates that the asymptotic expansion for the expectation $\langle O \rangle^\epsilon$ contains only integer orders, as allows us to extend the expansion to negative ϵ, which matches the exact values of the integral (see Figure 5). Unfortunately, the solution of the FP equation cannot adopt the form of the series (2.22) for $\epsilon < 0$, causing failure of the CL computation.

2.4.2. Transition to the global support. Despite the nonanalyticity of the observable predicted by the CL method, we expected that the transition is smooth when the global support appears. For $\epsilon < 0$, assume that

$$P^\epsilon(x, y) \approx \frac{C^\epsilon}{(x^2 + y^2)^3} \qquad \text{for } x^2 + y^2 \gg 1.$$

We conjecture that the coefficient C^ϵ has the form $\exp(\alpha_0/\epsilon)$, with α_0 being a constant. Such a form allows a C^∞ transition from zero to a nonzero value, leading to

TABLE 1
Coefficients of $\langle O \rangle_{k/2}$ in (2.26).

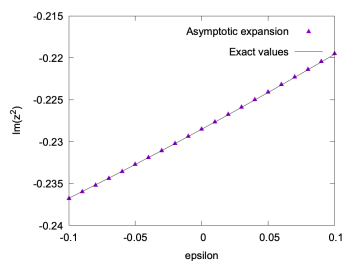| $\langle O \rangle_0$	$0.3579 - 0.2284i$	$\langle O \rangle_{\frac{1}{2}}$	$-2.4 \times 10^{-14} + 2.3 \times 10^{-15}i$
$\langle O \rangle_1$	$0.1136 + 0.0864i$	$\langle O \rangle_{\frac{3}{2}}$	$3.1 \times 10^{-10} + 2.1 \times 10^{-12}i$
$\langle O \rangle_2$	$-0.0165 + 0.0360i$	$\langle O \rangle_{\frac{5}{2}}$	$-3.6 \times 10^{-7} + 2.0 \times 10^{-9}i$
$\langle O \rangle_3$	$-0.0089 - 0.0030i$	$\langle O \rangle_{\frac{7}{2}}$	$-1.9 \times 10^{-4} + 4.7 \times 10^{-9}i$

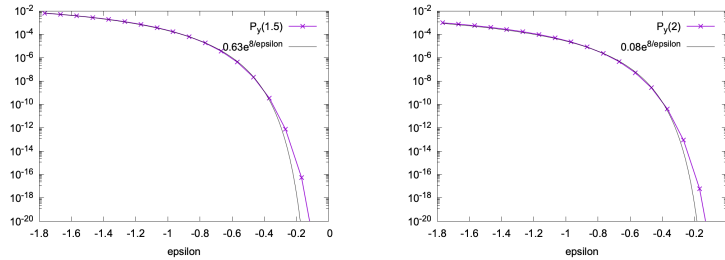FIG. 5. *The expectations obtained from expansion* (2.22) *versus ϵ.*

FIG. 6. P_y at point $y = 1.5$ and 2 for different ϵ.

the smoothness of the value of $\langle O \rangle^\epsilon$ predicted by the CL method. We verify this numerically by fitting the marginal probability density function $P_y(y)$ (defined in (2.9)) for $\epsilon < 0$. It can be expected that $P_y(y) \propto C(y) \exp(\alpha_0/\epsilon)$ for sufficiently large y. Here we pick $y = 1.5$ and $y = 2$ and do the curve fitting in Figure 6. Note that when ϵ is close to zero, the value of $P^\epsilon(x, y)$ is very small, so that the accuracy is affected by the round-off error. However, for $\epsilon < -0.4$, the numerical solution perfectly fits our conjecture. This indicates the C^∞ transition from local support to global support.

2.5. Implications of the model problem. By a careful analysis of the model problem with action (1.11), we have gained better understanding of some properties of the CL method. In general, the CL method looks quite fragile. For every CL result that looks convergent, we have to analyze the decay of the probability density function to show its validity. This can be done for some simple cases. For example, in the one-dimensional case with $S(x)$ dominated by the monomial x^k, we can use the same method as in section 2.2 to show that the decay rate is $P(x, y) \approx (x^2 + y^2)^{-(k-1)}$. However, for the multidimensional case, such analysis will become more nontrivial, and this becomes one of obstacles in the application of the CL method. When the probability density function does not decay sufficiently fast, biased results or even arithmetic overflow may occur. Recently, some numerical techniques to fix such issues have been proposed in [11, 37], which require further numerical analysis to understand their numerical error.

Without additional fixes, the CL method may still be valid for a wide range of observables if the probability density function is localized. Such existence usually depends on the parameters in the action. Unfortunately, as the parameter changes, the localization of the support may vanish in an unnoticeable way, which also introduces difficulty in judging the legitimacy of the numerical solution. Even worse, in the field theories, we can prove that such localized probability does not exist if the CL dynamics is not intervened. This will be detailed in the next section.

3. Nonexistence of localized probability in lattice field theories. In lattice field theories, the variables in the integral are a collection of group elements defined on the lattice. Specifically, we denote lattice nodes in the $(1 + d)$-dimensional spacetime by the indices

$$x = (t, x_1, \ldots, x_d), \ t = 0, \ldots, N_0 - 1, \quad x_1 = 0, \ldots, N_1 - 1, \ldots, x_d = 0, \ldots, N_d - 1.$$

Here we have assumed that all the lattice nodes are indexed by integers. For simplicity, we let $\mathcal{X} = (\mathbb{Z}/N_0\mathbb{Z}) \times (\mathbb{Z}/N_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/N_d\mathbb{Z})$ be the range of x, which also indicates the periodic boundary condition in our assumption. Below we are going to formulate the CL method for lattice field theories with general groups. We emphasize here that

this is the first complete mathematical formulation of the CL method since it was proposed.

For each $x \in \mathcal{X}$ and each $\mu = 0, 1, \ldots, d$, we define a "link variable" $U_{x,\mu} \in G$, where G is a compact Lie group with identity element e and its Lie algebra being $\mathfrak{g}$. Let $\{U\} \in G^{(d+1)N_0 N_1 \cdots N_d}$ be the collection of all these link variables $U_{x,\mu}$. Then both the observable $O(\cdot)$ and the action $S(\cdot)$ are functions of $\{U\}$, and the expectation of the observable is given by

$$\langle O \rangle = \frac{1}{Z} \int_{G^N} O(\{U\}) \exp\left(-S(\{U\})\right) d\{U\}, \qquad Z = \int_{G^N} \exp\left(-S(\{U\})\right) d\{U\},$$

where the integral is defined by the Haar measure of G, and we have used the shorthand $N = (d+1)N_0 N_1 \cdots N_d$ for simplicity.

To apply the CL method to this problem when $S(\cdot)$ is complex, some additional assumptions need to be imposed:

(A1) The group G is equipped with a Riemannian metric $\langle \cdot, \cdot \rangle_g$ for every $g \in G$, and the metric is bi-invariant.

(A2) The group G has a complexification

$$(3.1) \qquad\qquad\qquad G_{\mathbb{C}} = G \cdot \exp(i\mathfrak{g}),$$

whose Lie algebra is $\mathfrak{g}_{\mathbb{C}} = \mathfrak{g} \oplus i\mathfrak{g}$.

(A3) Both $O(\cdot)$ and $S(\cdot)$ can be extended to $G_{\mathbb{C}}^N$ as holomorphic functions.

By (A1), we can assume that $\{X^1, X^2, \ldots, X^m\}$ is an orthonormal basis of $\mathfrak{g}$ under the metric $\langle \cdot, \cdot \rangle_e$. The metric on $G_{\mathbb{C}}$ is chosen as the right-invariant metric:

$$(3.2) \quad \begin{aligned} &\langle X_1 + iX_1', X_2 + iX_2' \rangle_e = \langle X_1, X_2 \rangle_e + \langle X_1', X_2' \rangle_e && \forall X_1, X_2, X_1', X_2' \in \mathfrak{g}, \\ &\langle Z_1, Z_2 \rangle_g = \langle (dR_{g^{-1}})_g(Z_1), (dR_{g^{-1}})_g(Z_2) \rangle_e && \forall g \in G_{\mathbb{C}} \text{ and } Z_1, Z_2 \in T_g G_{\mathbb{C}}, \end{aligned}$$

where $R_h : g \mapsto gh$ is the right translation operator, so that $(dR_h)_g$ is the map from the tangent space $T_g G_{\mathbb{C}}$ to the tangent space $T_{gh} G_{\mathbb{C}}$. Note that when G is non-Abelian, this metric on $G_{\mathbb{C}}$ is in general not bi-invariant. The metric on $G_{\mathbb{C}}^N$ can then be naturally defined by summing up the metrics for all the components. Since each element in $\mathfrak{g}$ can be viewed as a right-invariant vector field on G, we will use the notation $\mathcal{L}_{X_{x,\mu}^a}$ to denote the Lie derivative with respect to the link variable $U_{x,\mu}$ along the right-invariant vector field X^a and use $\mathcal{L}_{Y_{x,\mu}^a}$ to denote the Lie derivative with respect to the link variable $U_{x,\mu}$ along $Y^a = iX^a$. Thus by (A3), we know that S satisfies the Cauchy–Riemann equations

$$\mathcal{L}_{X_{x,\mu}^a} \operatorname{Re} S = \mathcal{L}_{Y_{x,\mu}^a} \operatorname{Im} S, \qquad \mathcal{L}_{X_{x,\mu}^a} \operatorname{Im} S = -\mathcal{L}_{Y_{x,\mu}^a} \operatorname{Re} S.$$

The equations for $O(\cdot)$ are similar. Let $K_{x,\mu}^a = -\mathcal{L}_{X_{x,\mu}^a} \operatorname{Re} S$ and $J_{x,\mu}^a = -\mathcal{L}_{X_{x,\mu}^a} \operatorname{Im} S$. We can then write the CL equation:

$$dU_{x,\mu} = \sum_{a=1}^m (dR_{U_{x,\mu}})_e \left(\left[K_{x,\mu}^a(\{U\}) \, dt + dw_{x,\mu}^a \right] X^a + \left[J_{x,\mu}^a(\{U\}) \, dt \right] Y^a \right),$$

$$(3.3) \qquad x \in \mathcal{X}, \quad \mu = 0, 1, \ldots, d,$$

where the Brownian motions $w_{x,\mu}^a$ are independent of each other for different x, μ, a, and we take the Stratonovich interpretation of the stochastic differential equation in

(3.3). Since S is complex-valued, the link variable $U_{x,\mu}$ is generally in $G_{\mathbb{C}}$. Thus the CL method approximates the observable by

$$\langle O \rangle \approx \frac{1}{N} \sum_{k=1}^{N} O\Big(\{U(T + k\Delta T)\}\Big)$$

for sufficiently large T and sufficient time difference ΔT. Numerically, (3.3) is solved following

(3.4)

$$U_{x,\mu}(t + \Delta t)$$
$$\approx \exp\left(\sum_{a=1}^{m} \left[\left(\Delta t\, K_{x,\mu}^{a}(\{U(t)\}) + \sqrt{2\Delta t}\,\eta_{x,\mu}^{a}\right) X^{a} + \Delta t\, J_{x,\mu}^{a}(\{U(t)\}) Y^{a}\right]\right) U_{x,\mu}(t),$$

where $\exp(\cdot)$ is the exponential map from $\mathfrak{g}_{\mathbb{C}}$ to $G_{\mathbb{C}}$, and each $\eta_{x,\mu}^{a}$ is a normally distributed random variable with mean zero and standard deviation one, generated at each time step.

In this presentation, G can be regarded as the counterpart of the real axis, and then $G_{\mathbb{C}}$ is the counterpart of the complex plane. The verification of the CL method in the lattice field theory is similar to Theorem 1.1. We first define the nonnegative probability density function as $P(\{U\}; t)$ for all $\{U\} \in G_{\mathbb{C}}^{N}$. The evolution equation of $P(\{U\}; t)$ is

(3.5)

$$\frac{\partial P}{\partial t} + \sum_{x \in \mathcal{X}} \sum_{\mu=0}^{d} \sum_{a=1}^{m} \left(\mathcal{L}_{X_{x,\mu}^{a}}(K_{x,\mu}^{a} P) + \mathcal{L}_{Y_{x,\mu}^{a}}(J_{x,\mu}^{a} P)\right) = \sum_{x \in \mathcal{X}} \sum_{\mu=0}^{d} \sum_{a=1}^{m} \mathcal{L}_{X_{x,\mu}^{a}} \mathcal{L}_{X_{x,\mu}^{a}} P,$$

where

$$K_{x,\mu}^{a} = -\mathcal{L}_{X_{x,\mu}^{a}} \operatorname{Re} S, \qquad J_{x,\mu}^{a} = -\mathcal{L}_{X_{x,\mu}^{a}} \operatorname{Im} S.$$

Similarly to (1.9), we define the complex-valued function $\rho(\{U\}; t)$ for $\{U\} \in G^{N}$ as the solution of

(3.6) $$\frac{\partial \rho}{\partial t} + \sum_{x \in \mathcal{X}} \sum_{\mu=0}^{d} \sum_{a=1}^{m} \mathcal{L}_{X_{x,\mu}^{a}}[(K_{x,\mu}^{a} + iJ_{x,\mu}^{a})\rho] = \sum_{x \in \mathcal{X}} \sum_{\mu=0}^{d} \sum_{a=1}^{m} \mathcal{L}_{X_{x,\mu}^{a}} \mathcal{L}_{X_{x,\mu}^{a}} \rho.$$

The initial condition of (3.6) is $\rho(\{U\}; 0) = p(\{U\})$, where $p(\{U\})$ is a probability density function on G^{N}. To describe the initial condition for (3.5), we need to use the Cartan decomposition (3.1). In fact, the map $G \times \mathfrak{g} \to G_{\mathbb{C}}$ defined by (3.1) is a diffeomorphism. Therefore for every $U_{x,\mu} \in G_{\mathbb{C}}$, there exist unique $V_{x,\mu} \in G$ and $W_{x,\mu} \in \exp(i\mathfrak{g})$ such that $U_{x,\mu} = V_{x,\mu} W_{x,\mu}$. Thus we can define the initial condition of (3.5) as

(3.7) $$P(\{U\}; t) = p(\{V\}) \prod_{x \in \mathcal{X}} \prod_{\mu=0}^{d} \delta_{e}(W_{x,\mu}), \qquad \{U\} \in G_{\mathbb{C}}^{N},$$

where $\delta_{e}(\cdot)$ is the Dirac function defined on $\exp(i\mathfrak{g})$ whose value is infinity at the identity element. Now we are ready to state the theorem describing the validity of the CL method.

THEOREM 3.1. *Let $P(\{U\}; t)$ be the unique steady-state solution of* (3.5) *with the initial condition* (3.7), *and let $\rho(\{U\}; t)$ be the solution of* (3.6) *with the initial condition $\rho(\{U\}; 0) = p(\{U\})$. We further suppose that $\mathcal{O}(\{U\}; t)$ with $\{U\} \in G_{\mathbb{C}}^N$ and $t \geq 0$ satisfies the backward Kolmogorov equation*

$$\frac{\partial \mathcal{O}}{\partial t} = \sum_{x \in \mathcal{X}} \sum_{\mu=0}^{d} \sum_{a=1}^{m} \left(K_{x,\mu}^a \mathcal{L}_{X_{x,\mu}^a} \mathcal{O} + J_{x,\mu}^a \mathcal{L}_{Y_{x,\mu}^a} \mathcal{O} \right)$$

$$+ \sum_{x \in \mathcal{X}} \sum_{\mu=0}^{d} \sum_{a=1}^{m} \mathcal{L}_{X_{x,\mu}^a} \mathcal{L}_{X_{x,\mu}^a} \mathcal{O}, \quad \mathcal{O}(\{U\}; 0) = O(\{U\}),$$

and it holds that

$$(3.8) \qquad \int_{G_{\mathbb{C}}^N} \mathcal{O}(\{U\}; \tau) P(\{U\}; t - \tau) \, d\{U\} = \int_{G_{\mathbb{C}}^N} O(\{U\}) P(\{U\}; t) \, d\{U\}$$

for any $t > 0$ and $\tau \in [0, t]$. Then for any $t > 0$,

$$(3.9) \qquad \int_{G^N} O(\{U\}) \rho(\{U\}; t) \, d\{U\} = \int_{G_{\mathbb{C}}^N} \mathcal{O}(\{U\}) P(\{U\}; t) \, d\{U\}.$$

In the above theorem, the equality (3.8) corresponds to the assumption (H1) in subsection 1.1. The assumption (H3) is no longer needed due to the compactness of G. If we further assume that (3.6) has a steady-state solution

$$\lim_{t \to +\infty} \rho(\{U\}; t) = \rho_\infty(\{U\}) = \frac{1}{Z} \exp\left(-S(\{U\}) \right), \qquad \{U\} \in G^N,$$

then we can take the limit $t \to +\infty$ of (3.9) to validate the CL method. The proof of this theorem is completely parallel to the proof of Theorem 1.1, and we omit its details.

Again, this theorem only gives us an unsatisfactory result due to the strong assumption (3.8). To ensure that (3.8) holds, we again need to have conditions similar to (2.14). Note that (2.13) is not necessary again due to the compactness of G. The corresponding property holds for any observable O only if the support of $P(\cdot; t)$ is compact for large t. Unfortunately, such localized probability density function does not exist, as will be proven in the following subsections. To begin with, we study a simple case where $G = U(1)$.

3.1. Analysis for $U(1)$ theories. Due to its simplicity, the $U(1)$ theory is often employed to understand the properties of the CL method that are observed in other group theories [7, 34]. Here we also use this simple case to demonstrate our claims without involving the heavy notations in the group theory. When $G = U(1) = \{\exp(i\theta) \mid \theta \in \mathbb{R}\}$, its Lie algebra $\mathfrak{g}$ is the imaginary axis, and the metric can just be defined by

$$\langle X, Y \rangle_e = X^\dagger Y \qquad \forall X, Y \in \mathfrak{g},$$

where $\dagger$ denotes the complex conjugate. The complexification of G is $G_{\mathbb{C}} = \{\exp(i\theta) \mid \theta \in \mathbb{C}\} = \mathbb{C} \backslash \{0\}$. Therefore the action S, as a function on $G_{\mathbb{C}}^N$, can be written as

$$S(e^{i\theta_1}, e^{i\theta_2}, \ldots, e^{i\theta_N}) = S\left(e^{i(x_1 + iy_1)}, e^{i(x_2 + iy_2)}, \ldots, e^{i(x_N + iy_N)} \right),$$

where we have assumed that $\theta_k = x_k + iy_k$ for $k = 1, \ldots, N$. Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)^T$, $\boldsymbol{y} = (y_1, y_2, \ldots, y_N)^T$ and define

$$\bar{S}(\boldsymbol{x}, \boldsymbol{y}) = S\left(e^{i(x_1+iy_1)}, e^{i(x_2+iy_2)}, \ldots, e^{i(x_N+iy_N)}\right).$$

Then $\bar{S}(\boldsymbol{x}, \boldsymbol{y})$ is periodic with respect to $x_1, x_2, \ldots, x_N$, and the period is 2π for each variable. With such notations, the CL equation can be written similarly to (1.4):

$$(3.10) \qquad \begin{cases} d\boldsymbol{x} = \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{y})\, dt + d\boldsymbol{w}, & \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{Re}(-\nabla_{\boldsymbol{x}} \bar{S}(\boldsymbol{x}, \boldsymbol{y})), \\ d\boldsymbol{y} = \boldsymbol{J}(\boldsymbol{x}, \boldsymbol{y})\, dt, & \boldsymbol{J}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{Im}(-\nabla_{\boldsymbol{x}} \bar{S}(\boldsymbol{x}, \boldsymbol{y})), \end{cases}$$

where $\boldsymbol{w} = (w_1, \ldots, w_N)^T$ with each w_k being an independent Brownian motion. The initial condition satisfies $\boldsymbol{y}(0) = 0$. Now we are going to study (3.10), which has much simpler notations.

In order to localize the probability density function, we need to find a bounded, simply connected domain $\Omega \subset \mathbb{R}^N$ such that $\boldsymbol{J}(\boldsymbol{x}, \boldsymbol{y}) \cdot \boldsymbol{n}(\boldsymbol{y}) \leq 0$ for all $\boldsymbol{x} \in [0, 2\pi)^N$ and $\boldsymbol{y} \in \partial\Omega$. Here $\boldsymbol{n}(\boldsymbol{y})$ denotes the outer unit normal vector of Ω at point $\boldsymbol{y}$. A two-dimensional case is illustrated in Figure 7. Note that due to the Brownian motion in the $\boldsymbol{x}$ direction, the domain Ω must be the same for every $\boldsymbol{x}$. Thus once the probability density function is completely attracted into $[0, 2\pi)^N \times \overline{\Omega}$, it will forever be confined therein. Here $\overline{\Omega}$ refers to the closure of Ω. Since $\bar{S}$ is periodic with respect to $\boldsymbol{x}$ and $\boldsymbol{J}$ is the partial gradient of $\mathrm{Im}\, \bar{S}$ with respect to $\boldsymbol{x}$, by the divergence theorem, we have

$$(3.11) \qquad \int_{[0,2\pi)^N} \boldsymbol{J}(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{x} = \boldsymbol{0}$$

for any $\boldsymbol{y}$. Since $\boldsymbol{n}$ only depends on $\boldsymbol{y}$, it follows that $\int_{[0,2\pi)^N} \boldsymbol{J}(\boldsymbol{x}, \boldsymbol{y}) \cdot \boldsymbol{n}(\boldsymbol{y})\, d\boldsymbol{x} = 0$. Therefore $\boldsymbol{J}(\boldsymbol{x}, \boldsymbol{y}) \cdot \boldsymbol{n}(\boldsymbol{y}) \leq 0$ for any $\boldsymbol{x} \in [0, 2\pi)^N$ implies $\boldsymbol{J}(\boldsymbol{x}, \boldsymbol{y}) \cdot \boldsymbol{n}(\boldsymbol{y}) = 0$ for any $\boldsymbol{x} \in [0, 2\pi)^N$. By Cauchy–Riemann equations, $\boldsymbol{J}(\boldsymbol{x}, \boldsymbol{y}) = \nabla_{\boldsymbol{y}}(\mathrm{Re}\, \bar{S})$. Therefore we conclude that the existence of localized probability density function requires

$$\nabla_{\boldsymbol{y}}(\mathrm{Re}\, \bar{S}) \cdot \boldsymbol{n} = 0 \qquad \forall \boldsymbol{x} \in [0, 2\pi)^N \text{ and } \boldsymbol{y} \in \partial\Omega.$$

The holomorphism of $\bar{S}$ indicates that $\mathrm{Re}\, \bar{S}$ is a harmonic function. Therefore its periodicity with respect to $\boldsymbol{x}$, together with the above Neumann boundary condition, shows that $\mathrm{Re}\, \bar{S}$ is a constant. As a consequence, $\bar{S}$ is a constant everywhere, as corresponds to the trivial case of the CL equation.

By the above analysis, we know that for an irreducible action S (meaning that the essential number of independent variables equals N), if the probability density
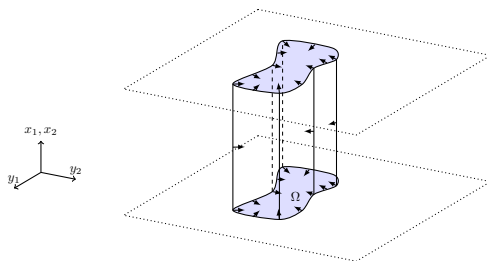

FIG. 7. *Localization of the distribution. The arrows denote the velocity field.*

function is localized, it must be localized to $[0, 2\pi)^N \times \{\boldsymbol{y_0}\}$ for some $\boldsymbol{y_0} \in \mathbb{R}^N$, corresponding to the case where $\bar{S}(\boldsymbol{x}, \boldsymbol{y_0})$ is real, which is equivalent to a translated real action. This is a pessimistic result, which indicates that when applying the CL method, one always needs to check whether the condition (3.8) holds, which is highly nontrivial since $P(\{U\}; t)$ is not directly available. More precisely, due to the periodicity of $O(x)$, one can see from its Fourier expansion that $O(x + iy)$ is expected to grow at least exponentially in the imaginary direction. To cover this class of observables, we need the probability density function to have a decay rate faster than any exponential functions, which looks like a strong assumption. Such possibility will be discussed in our future works. For a given observable, we refer the readers to [37] for a recent work on the estimation of the boundary terms. The same phenomenon also exists for general group G, as will be detailed in the next subsection.

Remark 3. In the above discussion, one missing case is that Ω is an m-dimensional bounded submanifold of $\mathbb{R}^N$ with $(m-1)$-dimensional boundary, where $0 < m < N$. In this case, we can reduce the $2N$-dimensional harmonic equation to an $2m$-dimensional problem by Cauchy–Riemann equations. Then using the same technique, we can show that Ω is bounded only if S is a constant on Ω. In the directions perpendicular to Ω, the action may be "real." An example is $\bar{S}(x_1, x_2, y_1, y_2) = i + \cos(x_1 + iy_1)$, for which we can have $\Omega = \{0\} \times [a, b]$ for any a and b since $\bar{S}$ does not depend on x_2 and y_2. In such cases, the action is essentially still a translated real action. Here we omit the details of the proof due to the length constraints of the paper.

3.2. Analysis for lattice gauge theories. For lattice gauge theories, the derivation follows the same idea as the $U(1)$ theories. However, when G is non-Abelian, the separation of the "real part" and the "imaginary part" becomes non-trivial. To this aim, we define two operators on $G_\mathbb{C}$ for any $g \in G_\mathbb{C}$:

$$L_g : h \mapsto gh, \qquad \Psi_g : h \mapsto ghg^{-1},$$

where L_g is the left translation operator, and Ψ_g is known as the inner automorphism. It is obvious that $L_g = R_g \Psi_g$. Then the following proposition holds.

PROPOSITION 3.2. *Let U_τ be a curve in $G_\mathbb{C}$ parametrized by $\tau \in \mathbb{R}$. Suppose $U_\tau = V_\tau W_\tau$ for $V_\tau \in G$ and $W_\tau \in \exp(i\mathfrak{g})$, and*

$$(3.12) \qquad \frac{dV_\tau}{d\tau} = (dR_{V_\tau})_e(X_\tau), \qquad \frac{dW_\tau}{d\tau} = (dR_{W_\tau})_e(Y_\tau)$$

for some $X_\tau \in \mathfrak{g}$ and $Y_\tau \in i\mathfrak{g}$. Then

$$(3.13) \qquad \frac{dU_\tau}{d\tau} = (dR_{U_\tau})_e\Big(X_\tau + (d\Psi_{V_\tau})_e Y_\tau\Big).$$

Inversely, if (3.13) holds, then (3.12) holds. Furthermore, if the vector $n_\tau \in T_{W_\tau} \exp(i\mathfrak{g})$ satisfies

$$\left\langle \frac{dW_\tau}{d\tau}, n_\tau \right\rangle_{W_\tau} = 0,$$

then

$$(3.14) \qquad \left\langle \frac{dU_\tau}{d\tau}, (dL_{V_\tau})_{W_\tau}(n_\tau) \right\rangle_{U_\tau} = 0.$$

The proof of this proposition will be deferred to Appendix B. From this result, we can rewrite the CL equation (3.3) by separating the "real" and "imaginary" parts. The details are given in the following corollary.

COROLLARY 3.3. *In the CL equation* (3.3), *if* $U_{x,\mu} = V_{x,\mu}W_{x,\mu}$ *for* $V_{x,\mu} \in G$ *and* $W_{x,\mu} \in \exp(i\mathfrak{g})$, *then*

$$dV_{x,\mu} = \sum_{a=1}^{m} \left(K_{x,\mu}^a(\{U\}) \, dt + dw_{x,\mu}^a \right) (dR_{V_{x,\mu}})_e(X^a),$$

$$dW_{x,\mu} = \sum_{a=1}^{m} \left(J_{x,\mu}^a(\{U\}) \, dt \right) (dR_{W_{x,\mu}})_e \left((d\Psi_{V_{x,\mu}^{-1}})_e(Y^a) \right).$$

This corollary is a straightforward result of the equivalence between (3.12) and (3.13), and we omit its proof. It shows that the Brownian motion allows $V_{x,\mu}$ to explore everywhere in G. Therefore if $P(\{U\}; t)$ has a compact support for all t, the support has the form $\Omega = G^N \cdot \Omega_I$, where Ω_I is a domain in $[\exp(i\mathfrak{g})]^N$. Like in the $U(1)$ theory, such Ω_I exists only when the action S is a constant. To show this, we need the following lemma, which is the counterpart of (3.11) in the $U(1)$ theory.

LEMMA 3.4. *Let* f *be a differentiable function on* $G_{\mathbb{C}}$, *and* $n \in T_W \exp(i\mathfrak{g})$ *for some* $W \in \exp(i\mathfrak{g})$. *Then*

$$\int_G \left\langle (dR_{VW})_e \left(\sum_{a=1}^{m} \mathcal{L}_{X^a} f(VW) Y^a \right), (dL_V)_W(n) \right\rangle_{VW} dV = 0,$$

where $\mathcal{L}_{X^a}$ *denotes the Lie derivative along the right translationally invariant vector field generated by* X^a.

Proof. Suppose $n = (dR_W)_e(Y)$ for some $Y \in i\mathfrak{g}$. Then

$$(dL_V)_W(n) = (dL_V)_W((dR_W)_e(Y)) = (dR_W)_V((dL_V)_e(Y)).$$

According to the right translational invariance of the inner product, we have

$$\left\langle (dR_{VW})_e \left(\sum_{a=1}^{m} \mathcal{L}_{X^a} f(VW) Y^a \right), (dL_V)_W(n) \right\rangle_{VW}$$
$$= \left\langle (dR_V)_e \left(\sum_{a=1}^{m} \mathcal{L}_{X^a} f(VW) Y^a \right), (dL_V)_e(Y) \right\rangle_V.$$

By further assuming $Y = iX$ for $X \in \mathfrak{g}$ and using the definition of the metric (3.2), we can rewrite the above expression as

$$\left\langle (dR_{VW})_e \left(\sum_{a=1}^{m} \mathcal{L}_{X^a} f(VW) Y^a \right), (dL_V)_W(n) \right\rangle_{VW}$$
$$= \left\langle (dR_V)_e \left(\sum_{a=1}^{m} \mathcal{L}_{X^a} \tilde{f}(V) X^a \right), (dL_V)_e(X) \right\rangle_V,$$

where $\tilde{f}(V) = f(VW)$. Thus we can fix W and consider all terms on the right-hand side of the equation as objects on G. From this point of view, the right-hand side is in fact the Lie derivative of $\tilde{f}$ along the left-invariant vector field generated by $X \in \mathfrak{g}$. According to [26, Theorem 14.35, Corollary 16.13], its integral equals zero. $\square$

Now we are ready to state and prove our main result.

THEOREM 3.5. *Suppose there exists a bounded, simply connected domain* $\Omega_I \subset [\exp(i\mathfrak{g})]^N$ *such that for* $\Omega = G^N \cdot \Omega_I \subset G_{\mathbb{C}}^N$,

$$(3.15) \qquad \langle Z(\{U\}), n(\{U\})\rangle_{\{U\}} \leq 0 \qquad \forall \{U\} \in \partial\Omega,$$

where n *denotes the outer unit normal of* Ω, *and* $Z(\{U\})$ *is the velocity field*

$$(3.16) \qquad Z(\{U\}) = \bigoplus_{x,\mu} \sum_{a=1}^{m} (dR_{U_{x,\mu}})_e \left(K_{x,\mu}^a(\{U\})X^a + J_{x,\mu}^a(\{U\})Y^a\right).$$

Then S *is a constant everywhere.*

Proof. For $\{W\} \in \partial\Omega_I$, we can write the outer unit normal of Ω_I in the following form:

$$n(\{W\}) = \bigoplus_{x,\mu} n_{x,\mu}(\{W\}), \qquad n_{x,\mu}(\{W\}) \in T_{W_{x,\mu}}\exp(i\mathfrak{g}).$$

Since $\partial\Omega = G^N \cdot \partial\Omega_I$, for any $\{U\} \in \partial\Omega$, we can find $\{V\} \in G^N$ and $\{W\} \in [\exp(i\mathfrak{g})]^N$ such that $U_{x,\mu} = V_{x,\mu}W_{x,\mu}$. Then according to (3.14), the outer normal vector of Ω at $\{U\} \in \partial\Omega$ is

$$n_{\{U\}} = \bigoplus_{x,\mu} (dL_{V_{x,\mu}})_{W_{x,\mu}}(n_{x,\mu}(\{W\})),$$

and thus

$$\langle Z(\{U\}), n(\{U\})\rangle_{\{U\}}$$
$$= \sum_{x,\mu} \left\langle (dR_{U_{x,\mu}})_e \left(\sum_{a=1}^{m} J_{x,\mu}^a(\{U\})Y^a\right), (dL_{V_{x,\mu}})_{W_{x,\mu}}(n_{x,\mu}(\{W\}))\right\rangle_{U_{x,\mu}}.$$

Since $J_{x,\mu}^a(\{U\})$ denotes the derivative of $-\operatorname{Im} S$ along $R_{V_{x,\mu}}(X^a)$, we see from Lemma 3.4 that

$$\int_{G^N} \langle Z(\{U\}), n(\{U\})\rangle_{\{U\}} \, d\{V\} = 0.$$

By (3.15), the value of $\langle Z(\{U\}), n(\{U\})\rangle_{\{U\}}$ must be zero for every $\{U\} \in \partial\Omega$, which can be considered as the homogeneous Neumann boundary condition of $\operatorname{Re} S$ on $\partial\Omega$ due to the fact that $J_{x,\mu}^a$ can also be regarded as the derivative of $\operatorname{Re} S$ along $R_{U_{x,\mu}}(Y^a)$, and $K_{x,\mu}^a$ plays no role in the inner product $\langle Z(\{U\}), n(\{U\})\rangle_{\{U\}}$. Furthermore, the holomorphism of $S(\{U\})$ indicates that $\operatorname{Re} S$ is harmonic. Therefore by the uniqueness of the solutions to elliptic equations [44, Proposition 7.6], we know that the real part of S is a constant, and thus S is also a constant. □

4. Localized probability density functions with gauge cooling technique. The analysis in section 3 reveals the reason why the application of the CL method is highly delicate. As mentioned in section 1, the method became successful mainly after the method of gauge cooling was proposed [41]. Before this work, the idea of gauge cooling already existed in some literature [12, 4], known as gauge fixing, which is used to study some simple models. In [13], the study of gauge fixing

is extended to the analysis of gauge cooling for the one-dimensional $SU(n)$ theory, and it is found that localized probability density functions can sometimes be obtained after applying the gauge cooling technique. However, this is not guaranteed by gauge cooling, and in [13], the authors also showed some cases where the probability density functions are global even after gauge cooling is applied. In such cases, it appears in the numerical experiments in [13] that some legitimate results can also be generated. According to our observation in section 2, such results may also be biased but only with small deviations from the exact integrals. In this work, we will restudy these examples and get a deeper understanding of the gauge cooling technique. To begin with, we will briefly review how gauge cooling works in the lattice field theory.

4.1. Review of the gauge cooling technique. The gauge cooling technique is developed based on the gauge invariance of the lattice field theory. For the complexified gauge field $\{U\} \in G_{\mathbb{C}}^N$, we introduce the gauge transformation by

$$(4.1) \qquad \widetilde{U}_{x,\mu} = g_x^{-1} U_{x,\mu} g_{x+\hat{\mu}} \qquad \forall x \in \mathcal{X}, \quad \forall \mu = 0, 1, \dots, d,$$

where $g_x \in G_{\mathbb{C}}$ is defined for every $x \in \mathcal{X}$, and $\hat{\mu}$ refers to the canonical unit vector $(0, \dots, 0, 1, 0, \dots, 0)^T$ whose μth component is 1. Here we remind the readers that the periodic boundary condition is used so that $x + \hat{\mu}$ is always well defined. After the transformation, a new field $\{\widetilde{U}\}$ is formed by the link variables $\widetilde{U}_{x,\mu}$. In the gauge theory, both the observable and the action are invariant under gauge transformation:

$$O(\{\widetilde{U}\}) = O(\{U\}), \qquad S(\{\widetilde{U}\}) = S(\{U\}).$$

As a result, we can apply any gauge transformation at any time during the evolution of the CL equation, which does not introduce any biases. Therefore, after each time step (3.4), we can choose suitable $g_x \in G_{\mathbb{C}}$ for each $x \in \mathcal{X}$ and apply the gauge transformation (4.1) to $\{U(t)\}$ so that the dynamics can hopefully be stabilized. To choose g_x appropriately, we let $F(\{U\})$ for all $\{U\} \in G_{\mathbb{C}}^N$ be the distance between $\{U\}$ and the submanifold G^N, and then solve the following minimization problem:

$$(4.2) \qquad \operatorname*{argmin}_{g_x \in G_{\mathbb{C}} \text{ for all } x \in \mathcal{X}} F(\{\widetilde{U}\}).$$

Gauge cooling refers to the gauge transformation (4.1) using the solution of the optimization problem (4.2). We hope that such a choice of g_x can help pull the sample field closer to G^N, causing a faster decay of the probability density function, so that the boundary terms can be reduced or eliminated. A formal justification of this method can be found in [30], which shows that the gauge cooling is unbiased. Numerically, the optimization problem (4.2) is solved by the gradient descent method with a sufficient number of iterations [41, 2].

Once $F(\cdot)$ is chosen, we can define the submanifold

$$M = \left\{ \{U\} \,\Big|\, F(\{U\}) \le F(\{\widetilde{U}\}) \text{ for any } g_x, \; x \in \mathcal{X} \right\},$$

which is the set of fields with "optimal guage" that minimizes the distance to G^N. The gauge cooling technique ensures that the field stays on M in the CL method. Therefore the governing equation can be formulated by mapping the right-hand side of (3.3) to the tangent space of M, and this mapping is linear but depends on the choice of the distance F.

By restricting the dynamics on the manifold M, we may have a chance to localize the probability density function since the argument in the previous section no longer holds. Deeper analysis of the gauge cooling technique requires the detailed form of $F(\{U\})$. One general choice of the distance function is

$$(4.3) \qquad F(\{U\}) = \frac{1}{2} \sum_{x \in \mathcal{X}} \sum_{\mu=0}^{d} \langle Y_{x,\mu}, Y_{x,\mu} \rangle_e.$$

Here $Y_{x,\mu} \in i\mathfrak{g}$ appears in the Cartan decomposition $U_{x,\mu} = V_{x,\mu} \exp(Y_{x,\mu})$. Interestingly, with such a distance function, we can find that gauge cooling does not have any effect when G is an Abelian group, including the simplest $U(1)$ theory. The result will be proven in the following subsection.

4.2. Gauge cooling for Abelian groups. Gauge cooling takes effect only when the gauge field does not lie on the manifold M. However, when G is Abelian (so that $G_{\mathbb{C}}$ is also Abelian), the CL dynamics automatically ensures that the field always stays on M. The details are given in the following theorem.

THEOREM 4.1. *Suppose G is an Abelian group and the distance function $F(\{U\})$ is defined by (4.3). Then the CL equation (3.3) guarantees that $\{U(t)\} \in M$ for all t if the initial condition $\{U(0)\} \in M$.*

Proof. Since G is an Abelian group, we can assume that $g_x = \exp(h_x)$ and $U_{x,\mu} = V_{x,\mu} \exp(Y_{x,\mu})$ for $h_x, Y_{x,\mu} \in i\mathfrak{g}$, so that the gauge transformation (4.1) becomes

$$(4.4) \qquad \widetilde{U}_{x,\mu} = V_{x,\mu} \exp(Y_{x,\mu} - h_x + h_{x+\hat{\mu}}) \qquad \forall x \in \mathcal{X}, \quad \forall \mu = 0, 1, \ldots, d.$$

Here it suffices to choose $g_x \in \exp(i\mathfrak{g})$ since adding a factor in G does not change the distance function $F(\{\widetilde{U}\})$. Thus the distance function turns out to be

$$F(\{\widetilde{U}\}) = \frac{1}{2} \sum_{x \in \mathcal{X}} \sum_{\mu=0}^{d} \langle Y_{x,\mu} - h_x + h_{x+\hat{\mu}}, Y_{x,\mu} - h_x + h_{x+\hat{\mu}} \rangle_e.$$

This is a quadratic function in the linear space $i\mathfrak{g}$, so that the minimization problem (4.2) can be solved by solving the first-order optimality condition:

$$(4.5) \qquad \sum_{\mu=0}^{d} (2h_x - h_{x-\hat{\mu}} - h_{x+\hat{\mu}}) = \sum_{\mu=0}^{d} (Y_{x,\mu} - Y_{x-\hat{\mu},\mu}) \qquad \forall x \in \mathcal{X}.$$

This equation has the form of a discrete Poisson equation with periodic boundary condition, and therefore the solution is unique up to a constant, which does not change the gauge transformation. This also indicates that the manifold M can be described by

$$M = \left\{ \{U\} \middle| \prod_{\mu=0}^{d} U_{x,\mu} U_{x-\hat{\mu},\mu}^{-1} \in G \, \forall x \in \mathcal{X} \right\},$$

since $\{U\} \in M$ indicates that the right-hand side of (4.5) is zero.

We assume that the initial value of $\{U(0)\}$ lies on M. Since G is Abelian, the inner automorphism Ψ_g is the identity operator. According to Corollary 3.3, we can derive the following equation for $Y_{x,\mu}$:

$$dY_{x,\mu} = \sum_{a=1}^{m} J_{x,\mu}^a(\{U\})Y^a \, dt.$$

If we can show that

$$(4.6) \qquad \sum_{\mu=0}^{d}(J_{x,\mu}^a - J_{x-\hat{\mu},\mu}^a) = 0 \qquad \forall x \in \mathcal{X}, \quad \forall a = 1,\ldots,m,$$

then it is clear that the right-hand side of (4.5) will remain zero if its initial value is zero. The Equations (4.6) can be seen from the gauge invariance of the action $S(\{U\})$. For given $x \in \mathcal{X}$ and $a = 1,\ldots,m$, consider the gauge transformation

$$\tilde{U}_{x,\mu}(\tau) = \exp(-\tau X^a)U_{x,\mu}, \qquad \tilde{U}_{x-\hat{\mu},\mu}(\tau) = U_{x-\hat{\mu},\mu}\exp(\tau X^a) \qquad \forall \mu = 0, 1,\ldots,d,$$

and other components of $\{\tilde{U}\}$ stay unchanged. Let $\tilde{S}(\tau) = S(\{\tilde{U}(\tau)\})$ for $\tau \in \mathbb{R}$. By chain rule, it is straightforward to verify that

$$\operatorname{Im}\frac{d\tilde{S}}{d\tau} = \sum_{\mu=0}^{d}\left(J_{x,\mu}^a - J_{x-\hat{\mu},\mu}^a\right).$$

The gauge invariance of S shows that $\tilde{S}$ is a constant. Therefore the above derivative is zero, meaning that (4.6) holds. $\qquad\square$

The above theorem shows that for the exact CL dynamics, gauge cooling does not change the field. However, this only refers to the exact dynamics, meaning that no finite-digit arithmetic is applied in the computation. Numerically, due to the inexact computer arithmetic, the field may deviate from the manifold M. In this case, the gauge cooling technique can act as a projection to keep the field on M, which also helps stabilize the dynamics. This is used in a recent work [23], where the $U(1)$ gauge theory is considered. In fact, even in the real Langevin dynamics, in which we are sure that gauge cooling has no effect, applying such a technique also helps avoid the possible instability produced by computer arithmetic [41]. Next, we are going to focus on the $SU(n)$ theory, which is non-Abelian so that gauge cooling is expected to be effective.

Remark 4. The distance function used in [23] is

$$F(\{U\}) = \sum_{x,\mu}\left[\exp\left(2\sqrt{\langle Y_{x,\mu}, Y_{x,\mu}\rangle_e}\right) + \exp\left(-2\sqrt{\langle Y_{x,\mu}, Y_{x,\mu}\rangle_e}\right) - 2\right],$$

which differs slightly from our definition (4.3). We expect that these two functions have similar effects since their leading-order term agrees after Taylor expansion. Further studies are needed to understand the effects of different distance functions.

4.3. One-dimensional $SU(n)$ theory. Now we focus on the $SU(n)$ theory, which is non-Abelian for all $n > 1$ and is most commonly used in lattice QCD. When $G = SU(n)$, its Lie algebra $\mathfrak{g}$ is the space of all traceless skew-Hermitian matrices of order n, whose dimension $m = n^2 - 1$. The metric on $SU(n)$ is defined by

$$\langle X_1, X_2\rangle_U = \frac{1}{2}\operatorname{tr}\left(X_1^\dagger X_2\right) \qquad \forall U \in SU(n), \quad \forall X_1, X_2 \in T_U G,$$

and the complexification of $SU(n)$ is the special linear group $SL(n,\mathbb{C})$. In this case, choosing the distance function as (4.3) may be inconvenient since the calculation of Cartan decomposition is not straightforward. Therefore we follow [39], which defines $F(\{U\})$ by

$$(4.7) \qquad F(\{U\}) = \sum_{x \in \mathcal{X}} \sum_{\mu=0}^{d} \left[\mathrm{tr}\left(U_{x,\mu}^{\dagger} U_{x,\mu}\right) - n \right].$$

It can be shown that the above quantity equals zero only when all $U_{x,\mu}$ are in $SU(n)$. In what follows, we will study the one-dimensional case ($d=0$) as an extension of the work in [13], which also gives some insight into the multidimensional case, as will be commented on at the end of this section.

For $d=0$ with periodic boundary conditions, according to the study in [13], the manifold M can be characterized as

$$(4.8) \qquad M = \{(\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{C}^n \mid \lambda_1 \lambda_2 \cdots \lambda_n = 1\}.$$

In fact, the one-dimensional case is highly similar to the one-link case, whose formulation has been given in [4]. In the case of N links, the stochastic differential equation (defined by Itô calculus) on M has been derived in [13, eq. (4.9)]:[1]

$$(4.9)$$
$$\frac{1}{N} d\lambda_j = -\frac{1}{\sqrt{N}} \sum_{a=1}^{m} \lambda_j X_{jj}^a \, dw^a$$
$$- \left[\sum_{a=1}^{m} [(K^a + iJ^a) X_{jj}^a + 2e_j^T X^a \Omega_j X^a e_j] + \frac{2(n^2-1)}{n} \right] \lambda_j \, dt, \quad j = 1, \ldots, n.$$

Here $X^1, \ldots, X^m$ are the orthonormal basis of $\mathfrak{g}$, and X_{jj}^a is the jth diagonal element of X^a. The quadratic Casimir invariant is

$$(4.10) \qquad -\sum_{a=1}^{m} X^a X^a = \frac{2(n^2-1)}{n} I,$$

where I is the identity matrix. For example, when $n = 2$, the basis can be chosen as $X^1 = i\sigma_x$, $X^2 = i\sigma_y$, $X^3 = i\sigma_z$, where $\sigma_{x,y,z}$ are Pauli matrices; similarly, when $n = 3$, the basis can be chosen as i times Gell-Mann matrices. In (4.9), the matrix Ω_j is

$$\Omega_j = \mathrm{diag}\left\{ \frac{\lambda_1}{\lambda_1 - \lambda_j}, \ldots, \frac{\lambda_{j-1}}{\lambda_{j-1} - \lambda_j}, 0, \frac{\lambda_{j+1}}{\lambda_{j+1} - \lambda_j}, \ldots, \frac{\lambda_N}{\lambda_N - \lambda_j} \right\},$$

and K^a, J^a, and w^a are defined by

$$(4.11) \qquad K^a = \frac{1}{N} \sum_{x \in \mathcal{X}} K_{x,0}^a, \qquad J^a = \frac{1}{N} \sum_{x \in \mathcal{X}} J_{x,0}^a, \qquad w^a = \frac{1}{\sqrt{N}} \sum_{x \in \mathcal{X}} w_{x,0}^a.$$

Without repeating the details of the derivation in [13], we just mention here that (4.9) is derived by solving the minimization problem (4.2) analytically, and then coupling the solution into the CL dynamics.

---

[1]Compared with the notations in [13], we have changed the definition of w^a so that w^a defined in (4.11) corresponds to the standard Brownian motion. Therefore, compared with equation (4.9) in [13], an additional coefficient $1/\sqrt{N}$ is seen in front of the first term on the right-hand side of (4.9).

We would now like to study whether it is possible to localize the probability density function on M. To clarify the effect of gauge cooling, we first assume that $J^a = 0$, which singles out the effect of gauge cooling. In this case, we have the following result.

THEOREM 4.2. *If $J^a = 0$ for all $a = 1, \ldots, m$ in (4.9), we have*

$$\frac{d}{dt} \sum_{j=1}^{n} |\lambda_j|^2 \leq 0$$

for any initial value $\lambda_1(0), \ldots, \lambda_n(0)$.

Proof. By replacing the term $2(n^2 - 1)/n$ in (4.9) using (4.10), we can rewrite (4.9) as

$$d\lambda_j = N\left[-\frac{1}{\sqrt{N}} \sum_{a=1}^{m} \lambda_j X_{jj}^a \, dw^a - \sum_{a=1}^{m} \left(K^a X_{jj}^a + e_j^T X^a D_j X^a e_j - \sum_{a=1}^{m} (X_{jj}^a)^2\right) \lambda_j \, dt\right],$$

where $D_j = \text{diag}\{\frac{\lambda_1 + \lambda_j}{\lambda_1 - \lambda_j}, \ldots, \frac{\lambda_{j-1} + \lambda_j}{\lambda_{j-1} - \lambda_j}, 0, \frac{\lambda_{j+1} + \lambda_j}{\lambda_{j+1} - \lambda_j}, \ldots, \frac{\lambda_N + \lambda_j}{\lambda_N - \lambda_j}\}$. Since X^a is skew-Hermitian, it follows that

$$d\bar{\lambda}_j = N\left[\frac{1}{\sqrt{N}} \sum_{a=1}^{m} \bar{\lambda}_j X_{jj}^a \, dw^a - \sum_{a=1}^{m} \left(-K^a X_{jj}^a + e_j^T X^a \bar{D}_j X^a e_j - \sum_{a=1}^{m} (X_{jj}^a)^2\right) \bar{\lambda}_j \, dt\right].$$

Therefore by Itô calculus,

$$d(|\lambda_j|^2) = \bar{\lambda}_j \, d\lambda_j + \lambda_j \, d\bar{\lambda}_j - 2N \sum_{a=1}^{m} |\lambda_j|^2 \left(X_{jj}^a\right)^2 \, dt = -2N \sum_{a=1}^{m} e_j^T X^a (\text{Re} \, D_j) X^a e_j |\lambda_j|^2 \, dt.$$

Summing up the above equation for all $j = 1, \ldots, n$, we obtain

$$\begin{aligned}
\frac{d}{dt} \sum_{j=1}^{n} |\lambda_j|^2 &= -2N \sum_{j=1}^{n} \sum_{a=1}^{m} e_j^T X^a (\text{Re} \, D_j) X^a e_j |\lambda_j|^2 \\
&= 2N \sum_{j=1}^{n} \sum_{k \neq j}^{n} \sum_{a=1}^{m} |\lambda_j|^2 |X_{jk}^a|^2 \, \text{Re} \frac{\lambda_k + \lambda_j}{\lambda_k - \lambda_j} \\
&= 2N \sum_{j=1}^{n} \sum_{k \neq j}^{n} \sum_{a=1}^{m} |\lambda_k|^2 |X_{kj}^a|^2 \, \text{Re} \frac{\lambda_j + \lambda_k}{\lambda_j - \lambda_k} \\
&= N \sum_{j=1}^{n} \sum_{k \neq j}^{n} \sum_{a=1}^{m} |X_{jk}^a|^2 (|\lambda_j|^2 - |\lambda_k|^2) \, \text{Re} \frac{\lambda_k + \lambda_j}{\lambda_k - \lambda_j} \\
&= N \sum_{j=1}^{n} \sum_{k \neq j}^{n} \sum_{a=1}^{m} |X_{jk}^a|^2 (|\lambda_j|^2 - |\lambda_k|^2) \frac{|\lambda_k|^2 - |\lambda_j|^2}{|\lambda_k - \lambda_j|^2} \leq 0,
\end{aligned}$$

which completes the proof. □

This theorem indicates that gauge cooling has introduced additional velocity which helps confine the probability density function. It can then be expected that when J^a is small, the probability density function can be localized. A special case for the $SU(2)$ theory with $N = 1$ and $S(U) = -(A + iB) \, \text{tr} \, U$ for $A, B \in \mathbb{R}$ has been

considered in a number of references including [12, 4, 13]. When $n = 2$, the manifold M defined in (4.8) is identical to $\mathbb{C}\backslash\{0\}$, which turns out to be similar to the complexified $U(1)$ one-link theory. Mimicking the transformations in subsection 3.1, we can write down the CL dynamics as

(4.12)

$$dx = K \, dt + dw, \quad K = 2\left(-A \cosh y \sin x + B \sinh y \cos x + \frac{\sin 2x}{\cosh 2y - \cos 2x}\right),$$

$$dy = J \, dt, \qquad J = -2\left(A \sinh y \cos x + B \cosh y \sin x + \frac{\sinh 2y}{\cosh 2y - \cos 2x}\right),$$

where x is considered to be periodic with period 2π. Let $z = x + iy$; then the expectation value of interest is $O(z) = e^{iz} + e^{-iz}$. In Theorem 4.1 of [13], it was shown that when (A, B) locates in a certain region, the probability function is localized and the CL method produces the correct result. For example, in their numerical experiments with $A = 1$, when $B = 0.2$, (A, B) is inside the region, and the CL result gives correct results for all observables. However, when A, B are chosen such that the support of probability density function is not compact, [13] also shows that the CL method may produce a result that is close to the exact integral, but it is unclear whether the error comes from the bias or the stochastic noise. More precisely, three cases $(A, B) = (1, 2), (5, 1)$, and $(5, 10)$ are tested in [13], among which only the case $(A, B) = (1, 2)$ shows a clear bias, while no conclusion is drawn for the other two sets of parameters.

By our analysis of the model problem in section 2, it can be expected that for $A = 5$, the results may also be biased due to the global support of the probability density function. To confirm this, we carry out the analysis of the decay rate in a way similar to subsection 2.2. Assume the steady-state probability function generated by the stochastic differential equation (4.12) is of the form

$$P(x, y) \approx c(x)e^{-\beta y},$$

at large $y > 0$. Substituting this ansatz into the associated FP equation and considering only the leading-order terms, we obtain

$$L^T P \approx e^{y - \beta y}\left(c'(x)(B \cos(x) - A \sin(x)) + (\beta - 2)c(x)(A \cos(x) + B \sin(x))\right).$$

Equating this expression to zero, we can solve $c(x)$ as

$$c(x) = C(B \cos(x) - A \sin(x))^{\beta - 2}, \quad C \in \mathbb{R}.$$

Due to the fact that $P(x, y)$ must be positive, the parameter β can only take the value 2. To check whether the solution is biased, we follow the condition (2.14) to check the growth rate of J and O. When y is large, $J(x, y) \propto e^y$, $O(x + iy) \propto e^y$, whose product exactly cancels the decay of $P(x, y)$, meaning that the CL method fails to produce an unbiased result when $P(x, y)$ is not localized.

The behavior of the one-dimensional $SU(2)$ theory turns out to be very similar to the model problem studied in section 2: when the parameter exceeds a certain threshold, the distribution of samples is no longer localized, and then a biased result is produced. To verify this, we again use the numerical method introduced in subsection 2.1 to solve the FP equation. When $A = 1$, the marginal probability density function $P_y(y)$ is plotted in Figure 8, where one can clearly observe that when B increases, the support of the probability density function turns global at a certain
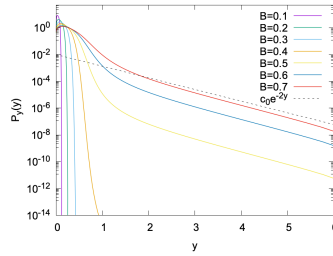
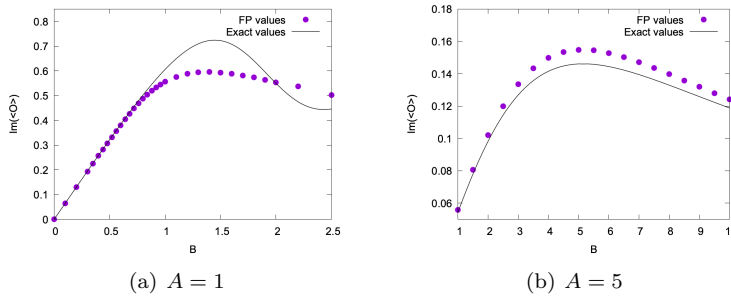FIG. 8. *The decay of P_y for $A = 1$.*

(a) $A = 1$

(b) $A = 5$

FIG. 9. *The imaginary part of the observables for $A = 1$ and $A = 5$.*

point, and the decay rate agrees with our theoretical analysis. Based on the computed probability density function, we compute the observables and show the results in Figure 9(a), from which one can observe that the CL results deviate from the exact values smoothly. We have also done the same numerical tests for $A = 5$, for which the support of the probability density function is the whole domain for any $B > 0$. The results shown in Figure 9(b) confirm the existence of the bias for all $A = 5$ and $B > 0$.

In conclusion, for the one-dimensional $SU(n)$ theory, gauge cooling can localize the probability density function for certain parameters, which stabilizes the CL method in some cases. However, when the parameters are set so that the velocity pushing samples away from the unitary field is large, the nonvanishing boundary term may still create bias in the numerical result.

Remark 5. In the multidimensional case, a very similar phenomenon can be observed. In [37], the boundary term of the heavy-dense QCD is studied, which includes two parameters β and μ, denoting the gauge coupling and the chemical potential, respectively. These parameters have roles similar to A and B in the above example. By numerical experiments, it is demonstrated in [37] that when β is large, although the tail of the probability density shrinks, the bias persists. The results in [1] for the same example show that smaller μ leads to smaller values of $F(\{U\})$ and more compact distributions of the samples, which also agrees with our observation in Figure 8. According to the theoretical study in the one-dimensional case, we also expect a range of β and μ where the CL dynamics with gauge cooling generates correct results, which need to be further explored in future works.

5. Conclusion and future works. This paper is devoted to the underlying mechanism of the CL method and the gauge cooling technique. By studying a controversial one-dimensional example, we have made a conclusion on the validity of the CL

method with respect to its parameters. Meanwhile, it is demonstrated by this example that the use of the CL method needs to be extremely careful even for simple cases. As pointed out in [7, 36], the decay rate of the probability density function must be carefully monitored in the numerical simulation. A numerical approach to monitoring this decay has been proposed in [37]. The only situation in which the method can provide an unbiased result for any observables is that the probability density function is localized. This occurs when all the velocities on the boundary of a certain bounded domain point inward, which may appear when the parameter controlling the magnitude of the imaginary part of the action is small. When the parameter exceeds a certain threshold, the bias in the result may arise smoothly, making it difficult to identify the appearance of the numerical failure.

For the lattice field theory, the situation is even worse since the localized probability density function does not exist. This reveals the importance of the gauge cooling technique, which introduces additional velocity that points toward the unitary field. When gauge cooling is applied, the probability density function may again be localized for certain parameters, so that the method is applicable for any observables. However, some limitations of gauge cooling, including its inability for Abelian groups and its failure in essentially suppressing the tails, is also uncovered by theoretical analysis.

This work also provides possible ideas to further develop the CL method, especially on the improvement of the dynamical stabilization proposed in [11]. This method regularizes the CL method by artificially introducing an velocity that pulls the samples back to the unitary field. This work may shed some light on the selection of the additional velocity, which should either create a velocity field that localizes the probability density function or essentially suppress the tail. This will be considered in our future works.

Appendix A. Proof of Theorem 1.1.

Proof. By condition (H1), we have

$$\frac{\partial}{\partial t}\left(\frac{\partial \mathcal{O}}{\partial y} - i\frac{\partial \mathcal{O}}{\partial x}\right) = \frac{\partial^2}{\partial x^2}\left(\frac{\partial \mathcal{O}}{\partial y} - i\frac{\partial \mathcal{O}}{\partial x}\right) + K_x\frac{\partial}{\partial x}\left(\frac{\partial \mathcal{O}}{\partial y} - i\frac{\partial \mathcal{O}}{\partial x}\right) + K_y\frac{\partial}{\partial y}\left(\frac{\partial \mathcal{O}}{\partial y} - i\frac{\partial \mathcal{O}}{\partial x}\right).$$

By the uniqueness of the solution of the advection-diffusion equation [18], we conclude that $\partial_y\mathcal{O} = i\partial_x\mathcal{O}$ for any x, y, and t since the initial value $\mathcal{O}(x, y; 0) = O(x + iy)$ satisfies the Cauchy–Riemann equations.

For $\tau \in [0, t]$, define

$$(A.1) \qquad F(t, \tau) = \int_{\mathbb{R}} \rho(x; t - \tau)\mathcal{O}(x, 0; \tau)\, dx.$$

We would like to show that $F(t, \tau)$ is independent of τ, which can be done by calculating the partial derivative:

$$\frac{\partial}{\partial \tau}F(t, \tau) = \int_{\mathbb{R}}\left[\rho(x; t - \tau)\frac{\partial}{\partial t}\mathcal{O}(x, 0; \tau) - \mathcal{O}(x, 0; \tau)\frac{\partial}{\partial t}\rho(x; t - \tau)\right] dx$$

$$= \int_{\mathbb{R}} \rho(x; t - \tau)\left(\frac{\partial^2}{\partial x^2}\mathcal{O}(x, 0; \tau) + K_x(x, 0)\frac{\partial}{\partial x}\mathcal{O}(x, 0; \tau) + K_y(x, 0)\frac{\partial}{\partial y}\mathcal{O}(x, 0; \tau)\right) dx$$

$$- \int_{\mathbb{R}} \mathcal{O}(x, 0; \tau)\left(\frac{\partial^2}{\partial x^2}\rho(x; t - \tau) + \frac{\partial}{\partial x}[S'(x)\rho(x; t - \tau)]\right) dx.$$

The holomorphism of $\mathcal{O}$ indicates $\partial_y \mathcal{O} = i\partial_x \mathcal{O}$, inserting which into the above equation yields

$$
\frac{\partial}{\partial \tau} F(t, \tau) = \int_{\mathbb{R}} \rho(x; t - \tau) \left(\frac{\partial^2}{\partial x^2} \mathcal{O}(x, 0; \tau) + S'(x) \frac{\partial}{\partial x} \mathcal{O}(x, 0; \tau) \right) dx
$$
$$
- \int_{\mathbb{R}} \mathcal{O}(x, 0; \tau) \left(\frac{\partial^2}{\partial x^2} \rho(x; t - \tau) + \frac{\partial}{\partial x} [S'(x)\rho(x; t - \tau)] \right) dx.
$$

According to (H3), we can apply integration by parts to the above result and conclude that $\partial_\tau F(t, \tau) = 0$. Therefore $F(t, 0) = F(t, t)$, meaning that

$$
\int_{\mathbb{R}} \rho(x; t)\mathcal{O}(x, 0; 0) \, dx = \int_{\mathbb{R}} \rho(x; 0)\mathcal{O}(x, 0; t) \, dx.
$$

Applying the initial conditions of ρ and $\mathcal{O}$, we get

$$
\int_{\mathbb{R}} \rho(x; t)O(x) \, dx = \int_{\mathbb{R}} p(x)\mathcal{O}(x, 0; t) \, dx.
$$

By comparing this equation with (1.10) and (1.8), we see that it remains only to show

$$
(A.2) \qquad \int_{\mathbb{R}} p(x)\mathcal{O}(x, 0; t) \, dx = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathcal{O}(x, y; \tau)P(x, y; t - \tau) \, dx \, dy
$$

for some τ. This can be done by setting $\tau = t$, so that the right-hand side of the above equation becomes

$$
\int_{\mathbb{R}} \int_{\mathbb{R}} \mathcal{O}(x, y; t)P(x, y; 0) \, dx \, dy = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathcal{O}(x, y; t)p(x)\delta(y) \, dx \, dy = \int_{\mathbb{R}} \mathcal{O}(x, 0; t)p(x) \, dx,
$$

which is clearly identical to the left-hand side of (A.2). $\qquad \square$

Appendix B. Proof of Proposition 3.2.

Proof. Given (3.12), we can compute $\frac{dU_\tau}{d\tau}$ by

$$
\frac{dU_\tau}{d\tau} = (dL_{V_\tau})_{W_\tau} \left(\frac{dW_\tau}{d\tau} \right) + (dR_{W_\tau})_{V_\tau} \left(\frac{dV_\tau}{d\tau} \right)
$$
$$
(B.1) \qquad = (dL_{V_\tau})_{W_\tau} \left((dR_{W_\tau})_e(Y_\tau) \right) + (dR_{W_\tau})_{V_\tau} \left((dR_{V_\tau})_e(X_\tau) \right)
$$
$$
= (dR_{W_\tau})_{V_\tau} \left((dL_{V_\tau})_e(Y_\tau) \right) + (dR_{U_\tau})_e(X_\tau).
$$

Using $L_{V_\tau} = R_{V_\tau} \circ \Psi_{V_\tau}$, one sees that $(dL_{V_\tau})_e(Y_\tau) = (dR_{V_\tau})_e\left((d\Psi_{V_\tau})_e(Y_\tau) \right)$, which can be inserted into (B.1) and yield (3.13) by $(dR_{W_\tau})_{V_\tau} \circ (dR_{V_\tau})_e = (dR_{U_\tau})_e$. Conversely, if (3.13) is given, then (3.12) holds due to the uniqueness of the Cartan decomposition.

To show (3.14), we write $n_\tau \in T_{W_\tau} \exp(i\mathfrak{g})$ as $n_\tau = (dR_{W_\tau})_e(n_\tau^e)$ for some $n_\tau^e \in i\mathfrak{g}$. Then by the right translational invariance of the inner product, it can be shown that

$$
(B.2) \qquad \left\langle \frac{dW_\tau}{d\tau}, n_\tau \right\rangle_{U_\tau} = \langle Y_\tau, n_\tau^e \rangle_e,
$$

$$
\left\langle \frac{dU_\tau}{d\tau}, (dL_{V_\tau})_{W_\tau}(n_\tau) \right\rangle_{U_\tau} = \langle X_\tau + (d\Psi_{V_\tau})_e(Y_\tau), (d\Psi_{V_\tau})_e(n_\tau^e) \rangle_e
$$
$$
(B.3) \qquad\qquad\qquad\qquad = \langle (d\Psi_{V_\tau})_e(Y_\tau), (d\Psi_{V_\tau})_e(n_\tau^e) \rangle_e
$$
$$
= \langle (dL_{V_\tau})_e(Y_\tau), (dL_{V_\tau})_e(n_\tau^e) \rangle_e,
$$

where the second equality of (B.3) uses the fact that $(d\Psi_{V_\tau})_e$ maps $i\mathfrak{g}$ to $i\mathfrak{g}$. Note that the metric is bi-invariant on G, meaning that the left translation $(dL_{V_\tau})_e$ does not change the value of the inner product due to $V_\tau \in G$. Therefore (B.2) and (B.3) are equal, which completes the proof. □

REFERENCES

[1] G. Aarts, F. Attanasio, B. Jäger, and D. Sexty, *The QCD phase diagram in the limit of heavy quarks using complex Langevin dynamics*, J. High Energy Phys., 20 (2016), 87, https://doi.org/10.1007/JHEP09(2016)087.

[2] G. Aarts, L. Bongiovanni, E. Seiler, D. Sexty, and I.-O. Stamatescu, *Controlling complex Langevin dynamics at finite density*, Eur. Phys. J. A, 49 (2013), 89, https://doi.org/10.1140/epja/i2013-13089-4.

[3] G. Aarts, P. Giudice, and E. Seiler, *Localised distributions and criteria for correctness in complex Langevin dynamics*, Ann. Phys., 337 (2013), pp. 238–260, https://doi.org/10.1016/j.aop.2013.06.019.

[4] G. Aarts, F. A. James, J. M. Pawlowski, E. Seiler, D. Sexty, and I.-O. Stamatescu, *Stability of complex Langevin dynamics in effective models*, J. High Energy Phys., 3 (2013), 73, https://doi.org/10.1007/JHEP03(2013)073.

[5] G. Aarts, E. Seiler, D. Sexty, and I.-O. Stamatescu, *Simulating QCD at nonzero baryon density to all orders in the hopping parameter expansion*, Phys. Rev. D, 90 (2014), 114505, https://doi.org/10.1103/PhysRevD.90.114505.

[6] G. Aarts, E. Seiler, D. Sexty, and I.-O. Stamatescu, *Complex Langevin dynamics and zeroes of the fermion determinant*, J. High Energy Phys., 2017 (2017), 44, https://doi.org/10.1007/JHEP05(2017)044.

[7] G. Aarts, E. Seiler, and I.-O. Stamatescu, *Complex Langevin method: When can it be trusted?*, Phys. Rev. D, 81 (2010), 054508, https://doi.org/10.1103/PhysRevD.81.054508.

[8] K. N. Anagnostopoulos, T. Azuma, Y. Ito, J. Nishimura, and S. K. Papadoudis, *Complex Langevin analysis of the spontaneous symmetry breaking in dimensionally reduced super Yang-Mills models*, J. High Energy Phys., 2018 (2018), 151, https://doi.org/10.1007/JHEP02(2018)151.

[9] K. N. Anagnostopoulos and J. Nishimura, *New approach to the complex-action problem and its application to a nonperturbative study of superstring theory*, Phys. Rev. D, 66 (2002), 106008, https://doi.org/10.1103/PhysRevD.66.106008.

[10] F. Attanasio and J. E. Drut, *Thermodynamics of spin-orbit-coupled bosons in two dimensions from the complex Langevin method*, Phys. Rev. A, 101 (2020), 033617, https://doi.org/10.1103/PhysRevA.101.033617.

[11] F. Attanasio and B. Jäger, *Dynamical stabilisation of complex Langevin simulations of QCD*, Eur. Phys. J. C, 79 (2019), 16, https://doi.org/10.1140/epjc/s10052-018-6512-7.

[12] J. Berges and D. Sexty, *Real-time gauge theory simulations from stochastic quantization with optimized updating*, Nuclear Phys. B, 799 (2008), pp. 306–329, https://doi.org/10.1016/j.nuclphysb.2008.01.018.

[13] Z. Cai, Y. Di, and X. Dong, *How does gauge cooling stabilize complex Langevin?*, Commun. Comput. Phys., 27 (2020), pp. 1344–1377, https://doi.org/10.4208/cicp.OA-2019-0126.

[14] H.-T. Chen, G. Cohen, and D. R. Reichman, *Inchworm Monte Carlo for exact non-adiabatic dynamics*. I. *Theory and algorithms*, J. Chem. Phys., 146 (2017), 054105, https://doi.org/10.1063/1.4974328.

[15] G. Cohen, E. Gull, D. R. Reichman, and A. J. Millis, *Taming the dynamical sign problem in real-time evolution of quantum many-body problems*, Phys. Rev. Lett., 115 (2015), 266802, https://doi.org/10.1103/PhysRevLett.115.266802.

[16] M. Cristoforetti, F. Di Renzo, A. Mukherjee, and L. Scorzato, *Monte Carlo simulations on the Lefschetz thimble: Taming the sign problem*, Phys. Rev. D, 88 (2013), 051501, https://doi.org/10.1103/PhysRevD.88.051501.

[17] M. Cristoforetti, F. Di Renzo, and L. Scorzato, *New approach to the sign problem in quantum field theories: High density QCD on a Lefschetz thimble*, Phys. Rev. D, 86 (2012), 074506, https://doi.org/10.1103/PhysRevD.86.074506.

[18] L. C. Evans, *Partial Differential Equation*, Grad. Stud. Math. 19, AMS, Providence, RI, 1999.

[19] C. Gattringer and C. Lang, *Quantum Chromodynamics on the Lattice*, Springer-Verlag, Berlin, 2010, https://doi.org/10.1007/978-3-642-01850-3.

[20] H. Gausterer, *Complex Langevin: A numerical method?*, Nuclear Phys. A, 642 (1998), pp. c239–c250, https://doi.org/10.1016/S0375-9474(98)00522-3.

[21] J. W. GIBBS, *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics*, Camb. Libr. Collect. Math., Cambridge University Press, Cambridge, UK, 2010, https://doi.org/10.1017/CBO9780511686948.

[22] W. GREINER, S. SCHRAMM, AND E. STEIN, *Quantum Chromodynamics*, Springer-Verlag, Berlin, 2007, https://doi.org/10.1007/978-3-540-48535-3.

[23] M. HIRASAWA, A. MATSUMOTO, J. NISHIMURA, AND A. YOSPRAKOB, *Complex Langevin Analysis of 2D U(1) Gauge Theory on a Torus with a θ Term*, https://arxiv.org/abs/2004.13982, 2020.

[24] J. KLAUDER, *A Langevin approach to fermion and quantum spin correlation functions*, J. Phys. A, 16 (1983), L317, https://doi.org/10.1088/0305-4470/16/10/001.

[25] J. B. KOGUT AND D. K. SINCLAIR, *Applying complex Langevin simulations to lattice QCD at finite density*, Phys. Rev. D, 100 (2019), 054512, https://doi.org/10.1103/PhysRevD.100.054512.

[26] J. M. LEE, *Introduction to Smooth Manifolds*, 2nd ed., Springer, New York, 2013, https://doi.org/10.1007/978-1-4419-9982-5.

[27] E. Y. LOH, J. E. GUBERNATIS, R. T. SCALETTAR, S. R. WHITE, D. J. SCALAPINO, AND R. L. SUGAR, *Sign problem in the numerical simulation of many-electron systems*, Phys. Rev. B, 41 (1990), pp. 9301–9307, https://doi.org/10.1103/PhysRevB.41.9301.

[28] J. C. MATTINGLY, A. M. STUART, AND D. J. HIGHAM, *Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise*, Stochastic Process. Appl., 101 (2002), pp. 185–232, https://doi.org/10.1016/S0304-4149(02)00150-3.

[29] K. NAGATA, J. NISHIMURA, AND S. SHIMASAKI, *Argument for justification of the complex Langevin method and the condition for correct convergence*, Phys. Rev. D, 94 (2016), 114515, https://doi.org/10.1103/PhysRevD.94.114515.

[30] K. NAGATA, S. SHIMASAKI, AND J. NISHIMURA, *Justification of the complex Langevin method with the gauge cooling procedure*, Prog. Theor. Exp. Phys., 2016 (2016), 013B01, https://doi.org/10.1093/ptep/ptv173.

[31] M. NAMIKI, *Stochastic Quantization*, Springer-Verlag, Berlin, 1992, https://doi.org/10.1007/978-3-540-47217-9.

[32] J. NISHIMURA AND A. TSUCHIYA, *Complex Langevin analysis of the space-time structure in the Lorentzian type IIB matrix model*, J. High Energy Phys., 2019 (2019), 77, https://doi.org/10.1007/JHEP06(2019)077.

[33] G. PARISI, *On complex probabilities*, Phys. Lett. B, 131 (1983), pp. 393–395, https://doi.org/10.1016/0370-2693(83)90525-7.

[34] L. L. SALCEDO, *Does the complex Langevin method give unbiased results?*, Phys. Rev. D, 94 (2016), 114505, https://doi.org/10.1103/PhysRevD.94.114505.

[35] L. L. SALCEDO AND E. SEILER, *Schwinger–Dyson equations and line integrals*, J. Phys. A, 52 (2018), 035201, https://doi.org/10.1088/1751-8121/aaefca.

[36] M. SCHERZER, E. SEILER, D. SEXTY, AND I.-O. STAMATESCU, *Complex Langevin and boundary terms*, Phys. Rev. D, 99 (2019), 014512, https://doi.org/10.1103/PhysRevD.99.014512.

[37] M. SCHERZER, E. SEILER, D. SEXTY, AND I.-O. STAMATESCU, *Controlling complex Langevin simulations of lattice models by boundary term analysis*, Phys. Rev. D, 101 (2020), 014501, https://doi.org/10.1103/PhysRevD.101.014501.

[38] M. SCHIRÓ, *Real-time dynamics in quantum impurity models with diagrammatic Monte Carlo*, Phys. Rev. B, 81 (2010), 085126, https://doi.org/10.1103/PhysRevB.81.085126.

[39] E. SEILER, *Status of complex Langevin*, EPJ Web Conf., 175 (2018), 01019, https://doi.org/10.1051/epjconf/201817501019.

[40] E. SEILER, *Complex Langevin: Boundary terms at poles*, Phys. Rev. D, 102 (2020), 094507, https://doi.org/10.1103/PhysRevD.102.094507.

[41] E. SEILER, D. SEXTY, AND I.-O. STAMATESCU, *Gauge cooling in complex Langevin for lattice QCD with heavy quarks*, Phys. Lett. B, 723 (2013), pp. 213–216, https://doi.org/10.1016/j.physletb.2013.04.062.

[42] E. SEILER AND J. WOSIEK, *Beyond complex Langevin equations: Positive representation of a class of complex measures*, EPJ Web Conf., 175 (2018), 11004, https://doi.org/10.1051/epjconf/201817511004.

[43] D. SEXTY, *Simulating full QCD at nonzero density using the complex Langevin equation*, Phys. Lett. B, 729 (2014), pp. 108–111, https://doi.org/10.1016/j.physletb.2014.01.019.

[44] M. TAYLOR, *Partial Differential Equations* I: *Basic Theory*, App. Math. Sci. 115, Springer, New York, 2011, https://doi.org/10.1007/978-1-4419-7055-8.