PHYSICAL REVIEW D **94,** 114515 (2016)

Argument for justification of the complex Langevin method and the condition for correct convergence

Keitaro Nagata,[1,*] Jun Nishimura,[1,2,†] and Shinji Shimasaki[1,3,‡]

[1] *KEK Theory Center, High Energy Accelerator Research Organization, 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan*

[2] *Graduate University for Advanced Studies (SOKENDAI), 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan*

[3] *Research and Education Center for Natural Sciences, Keio University, Hiyoshi 4-1-1, Yokohama, Kanagawa 223-8521, Japan*

(Received 17 September 2016; published 27 December 2016)

The complex Langevin method is a promising approach to the complex-action problem based on a fictitious time evolution of complexified dynamical variables under the influence of a Gaussian noise. Although it is known to have a restricted range of applicability, the use of gauge cooling made it applicable to various interesting cases including finite density QCD in certain parameter regions. In this paper we revisit the argument for justification of the method. In particular, we point out a subtlety in the use of time-evolved observables, which play a crucial role in the previous argument. This requires that the probability of the drift term should fall off exponentially or faster at large magnitude. We argue that this is actually a necessary and sufficient condition for the method to be justified. Using two simple examples, we show that our condition tells us clearly whether the results obtained by the method are trustable or not. We also discuss a new possibility for the gauge cooling, which can reduce the magnitude of the drift term directly.

DOI: 10.1103/PhysRevD.94.114515

I. INTRODUCTION

Solving the negative sign problem or more generally the complex-action problem is one of the most important challenges in computational science. This is a problem that occurs in an attempt to apply the idea of importance sampling to multiple integration with a weight which fluctuates in sign or in its complex phase. The complex Langevin method (CLM) [1,2] is a promising approach, which can be applied to a variety of models with a complex weight albeit not all of them. For instance, it has been applied successfully to finite density QCD either with heavy quarks [3] or in the deconfined phase [4] using a new technique called gauge cooling [5]. Whether it is applicable also in the case with light quarks and in the confined phase is one of the hottest topics in this field [6–10].

The CLM may be viewed as a generalization of the stochastic quantization [11], which generates dynamical variables with a given probability by solving the Langevin equation that describes a fictitious time evolution of those variables under the influence of a Gaussian noise. (See Ref. [12] for a comprehensive review.) When one applies this idea to the calculation of expectation values of observables with a complex weight, one necessarily has to complexify the dynamical variables due to the complex drift term, which is derived from the complex weight. Correspondingly, the drift term and the observables should be extended to holomorphic functions of the complexified variables by analytic continuation. Then by measuring the observables for the complexified variables generated by the Langevin process and calculating their expectation values at sufficiently late times, one can obtain the expectation values of the observables for the original real variables with the complex weight.

It has been known for a long time that this method does not always work. Typically, the complex Langevin process reaches thermal equilibrium without any problem, but the results for the expectation values obtained in the way mentioned above turn out to be simply wrong in some cases. The reason for the failure was discussed in Refs. [13,14] starting from the complex Langevin equation with a continuous Langevin time. There, it was found that a subtlety exists in the integration by parts used in translating the time evolution of the probability distribution of the complexified variables into that of the observables. In order for the integration by parts to be valid, the probability distribution of the complexified variables should have appropriate asymptotic behaviors. By now, the following two conditions are recognized.

(1) The probability distribution should be suppressed strongly enough when the complexified variables take large values [13,14]. Typically, this becomes a problem when the complexified variables make long excursions in the imaginary directions during the Langevin simulation.

(2) The drift term can have singularities while it is otherwise a holomorphic function of the complexified variables. In that case, the probability distribution

*knagata@post.kek.jp
†jnishi@post.kek.jp
‡shinji.shimasaki@keio.jp

should be suppressed strongly enough near the singularities [15].

In fact, both these conditions are relevant in applying the CLM to finite density QCD. The condition 1 is an issue because the link variable, upon complexification, becomes an $\mathrm{SL}(3,\mathbb{C})$ matrix, which forms a noncompact manifold. Here, the idea of gauge cooling turned out to be useful [5]. It is based on the fact that the $\mathrm{SU}(3)$ gauge symmetry of the action and the observables is enhanced to the $\mathrm{SL}(3,\mathbb{C})$ gauge symmetry upon complexification of the dynamical variables. One can actually make a complexified gauge transformation after each Langevin step in such a way that the link variables stay close to the original $\mathrm{SU}(3)$ manifold during the Langevin simulation. Using this technique, the CLM became applicable to finite density QCD in the heavy dense limit [5,16,17] and in the deconfined phase [4,8].

The condition 2 is also an issue in finite density QCD because the drift term $\mathrm{Tr}[(D+m)^{-1}\partial(D+m)]$, which comes from the fermion determinant, has singularities corresponding to the appearance of zero eigenvalues of the Dirac operator $D+m$. This becomes a problem at low temperature when the mass is small as demonstrated clearly in the chiral random matrix theory (cRMT) [6]. In the case of cRMT, changing the integration variables in the original path integral to the polar coordinates was shown to solve the problem [7]. This is possible because the change of variables in the original path integral leads to an inequivalent complex Langevin process.[1] In our previous publication [10], we proposed that it should be possible to solve this problem also by the gauge cooling with different criteria for choosing the complexified gauge transformation. The results for the cRMT look promising.

The argument for justification of the CLM given in Refs. [13,14] has been extended to the case including the gauge cooling procedure recently [18]. This settles down various skepticism on the validity of gauge cooling. For instance, the gauge cooling uses the complexified gauge symmetry, which is not respected by the noise term in the complex Langevin equation. Despite such issues, the argument for justification goes through as is shown explicitly in Ref. [18].

In this paper, we revisit the argument for justification of the CLM with or without the gauge cooling procedure. In particular, we point out a subtlety in the use of time-evolved observables, which play a crucial role in the argument. In the previous argument, it was assumed implicitly that time-evolved observables can be used for infinitely long time. We argue that this assumption is too strong. In fact, we only need to use the time-evolved observables for a finite but nonzero time to complete the argument for justification. This still requires that the probability distribution of the drift term should be suppressed, at least, exponentially at large magnitude.

We also point out that the integration by parts, which was considered to be the main issue in justifying the CLM, requires a slightly weaker condition than the one we obtain above. This conclusion is reached by reformulating the argument starting with a discretized Langevin time[2] with the step-size ϵ. In this case, we can always define the time-evolution of observables in such a way that it is equivalent to the usual description with fixed observables and the time-dependent probability distribution of the complexified variables. However, an issue arises when one tries to take the $\epsilon \to 0$ limit. Thus the failure of the integration by parts can be understood as the failure of the $\epsilon \to 0$ limit for an expression involving the time-evolved observables. Based on this understanding, we find that the integration by parts can be justified if the probability distribution of the drift term falls of faster than any power-law at large magnitude. This is slightly weaker than the condition that the probability distribution of the drift term should be suppressed, at least, exponentially at large magnitude. Therefore, we may regard the latter as a necessary and sufficient condition for justifying the CLM. In the case of the real Langevin method [11], there is no need to consider the time-evolved observables in justifying the method, which implies that all the conditions encountered above are simply irrelevant.

We substantiate our argument by investigating two simple examples. The first one is a model studied in Ref. [15] to clarify the problem related to a singular drift, while the second one is a model studied in Ref. [19] to clarify the problem related to long excursions into the deeply imaginary regime. In both models, there are two parameter regions; the CLM works in one of them but fails in the other. We measure the probability distribution of the drift term and investigate its asymptotic behavior at large magnitude. It is found that the probability distribution is indeed exponentially suppressed when the CLM works, while it is only power-law suppressed when the CLM fails. Thus, our simple condition tells us clearly whether the results obtained by the method are trustable or not in a unified manner.

The rest of this paper is organized as follows. In Sec. II we discuss the justification of the CLM, and point out that the use of time-evolved observables can be subtle. This leads to our proposal of a necessary and sufficient condition for justifying the CLM. In Sec. III we investigate two models, in which the CLM was thought to fail for different reasons. In particular, we show that our new condition can

[1] The reason why it works in this case is rather trivial, though. After complexification of the polar coordinates, the chemical potential μ can be absorbed by shifting the imaginary part of the angular variables. Thus the complex Langevin equation reduces to that for $\mu = 0$, which does not have any problem. Also it is not obvious how one can extend this idea to finite density QCD.

[2] Some preliminary discussions for finite ϵ are given already in our previous publication [18] for the purpose of treating the case in which the gauge cooling transformation remains finite in the $\epsilon \to 0$ limit.

tell whether the results are trustable or not. In Sec. IV we extend the argument in Sec. II to the case of lattice gauge theory. We also discuss a new possibility for the gauge cooling, which can reduce the magnitude of the drift term directly. Section V is devoted to a summary and discussions.

II. THE CASE OF A 0-DIMENSIONAL MODEL

In this section we revisit the argument for justification of the CLM. In particular, we point out that the use of time-evolved observables, which play a crucial role in the argument, can be subtle, and this leads to a condition that the probability distribution of the drift term should fall off exponentially or faster at large magnitude. Our argument starts with a finite step-size ϵ for the discretized Langevin time, which is different from the previous argument [13,14], which starts from the complex Langevin equation with a continuous Langevin time. The purpose of this is to clarify the condition for the validity of the integration by parts, which was considered the main issue in the previous argument. In fact, we find that this condition is slightly weaker than the one we newly obtain. Therefore, the latter is actually a necessary and sufficient condition for the CLM to be justified. Here we discuss a 0-dimensional model for simplicity, but generalization to the lattice gauge theory is straightforward as we show explicitly in Sec. IV.

We include the gauge cooling procedure to keep our discussion as general as possible. This part is similar to what we have already done in our previous paper [18]. The readers who are not interested in the gauge cooling can omit the gauge cooling procedure by simply setting the transformation matrix g to identity in all the expressions below. In Ref. [18], we have also reviewed the previous argument for justification of the CLM, which may be compared with our new argument.

A. The complex Langevin method

Let us consider a system of N real variables x_k ($k = 1, \ldots, N$) given by the partition function[3]

$$Z = \int dx w(x) = \int \prod_k dx_k w(x), \tag{2.1}$$

where the weight $w(x)$ is a complex-valued function of the real variables x_k ($k = 1, \ldots, N$).

When one considers the Langevin equation for this system, the drift term

$$v_k(x) = \frac{1}{w(x)} \frac{\partial w(x)}{\partial x_k} \tag{2.2}$$

becomes complex, and therefore, one necessarily has to complexify the dynamical variables[4] as $x_k \mapsto z_k = x_k + i y_k$. Then, the discretized complex Langevin equation is given by

$$z_k^{(\eta)}(t+\epsilon) = z_k^{(\eta)}(t) + \epsilon v_k(z) + \sqrt{\epsilon}\eta_k(t), \tag{2.3}$$

where the drift term $v_k(z)$ is obtained by analytically continuing (2.2). The probabilistic variables $\eta_k(t)$ in (2.3) are, in general, complex

$$\eta_k(t) = \eta_k^{(\mathrm{R})}(t) + i\eta_k^{(\mathrm{I})}(t), \tag{2.4}$$

and obey the probability distribution $\propto \mathrm{e}^{-\frac{1}{4}\sum_t \{\frac{1}{N_\mathrm{R}}\eta_k^{(\mathrm{R})}(t)^2 + \frac{1}{N_\mathrm{I}}\eta_k^{(\mathrm{I})}(t)^2\}}$, where we have to choose

$$N_\mathrm{R} - N_\mathrm{I} = 1. \tag{2.5}$$

For practical purposes, one should actually use $N_\mathrm{R} = 1$, $N_\mathrm{I} = 0$, corresponding to real $\eta_k(t)$, to reduce the excursions in the imaginary directions, which spoil the validity of the method [13,14,19].

Let us define the expectation value $\langle \cdots \rangle_\eta$ with respect to η as

$$\langle \cdots \rangle_\eta = \frac{\int \mathcal{D}\eta \cdots \mathrm{e}^{-\frac{1}{4}\sum_t \{\frac{1}{N_\mathrm{R}}\eta_k^{(\mathrm{R})}(t)^2 + \frac{1}{N_\mathrm{I}}\eta_k^{(\mathrm{I})}(t)^2\}}}{\int \mathcal{D}\eta \mathrm{e}^{-\frac{1}{4}\sum_t \{\frac{1}{N_\mathrm{R}}\eta_k^{(\mathrm{R})}(t)^2 + \frac{1}{N_\mathrm{I}}\eta_k^{(\mathrm{I})}(t)^2\}}}. \tag{2.6}$$

With this notation, we have, for instance,

$$\begin{aligned} \langle \eta_k^{(\mathrm{R})}(t_1)\eta_l^{(\mathrm{R})}(t_2)\rangle_\eta &= 2N_\mathrm{R}\delta_{kl}\delta_{t_1,t_2}, \\ \langle \eta_k^{(\mathrm{I})}(t_1)\eta_l^{(\mathrm{I})}(t_2)\rangle_\eta &= 2N_\mathrm{I}\delta_{kl}\delta_{t_1,t_2}, \\ \langle \eta_k^{(\mathrm{R})}(t_1)\eta_l^{(\mathrm{I})}(t_2)\rangle_\eta &= 0. \end{aligned} \tag{2.7}$$

When the system (2.1) has a symmetry under

$$x'_j = g_{jk} x_k, \tag{2.8}$$

where g is a representation matrix of a Lie group, we can use the symmetry to apply gauge cooling. Upon complexifying the variables $x_k \mapsto z_k$, the symmetry property of the drift term and the observables naturally enhances from (2.8) to

[3] In many examples, the weight is given by $w(x) = \mathrm{e}^{-S(x)}$ in terms of the action $S(x)$, but we prefer not to use the action in our discussion to avoid any ambiguities arising from taking the log of the complex weight [6,7,20].

[4] In this respect, there is a closely related approach based on the so-called Lefschetz thimble [21,22], which has attracted much attention recently. See Refs. [23–27] and references therein. There is also a new proposal [28] for generalizing this approach to overcome a few important problems in the original idea.

$$z'_j = g_{jk} z_k, \tag{2.9}$$

where g is an element of the Lie group that can be obtained by complexifying the original Lie group. The discretized complex Langevin equation including the gauge cooling is given by

$$\tilde{z}_k^{(\eta)}(t) = g_{kl} z_l^{(\eta)}(t), \tag{2.10}$$

$$z_k^{(\eta)}(t+\epsilon) = \tilde{z}_k^{(\eta)}(t) + \epsilon v_k(\tilde{z}^{(\eta)}(t)) + \sqrt{\epsilon}\eta_k(t). \tag{2.11}$$

Equation (2.10) represents the gauge cooling, where g is an element of the complexified Lie group chosen appropriately as a function of the configuration $z^{(\eta)}(t)$ before cooling. We regard (2.10) and (2.11) as describing the t-evolution of $z_k^{(\eta)}(t)$ and treat $\tilde{z}_k^{(\eta)}(t)$ as an intermediate object. The basic idea is to determine g in such a way that the modified Langevin process does not suffer from the problem of the original Langevin process (2.3).

We consider observables $\mathcal{O}(x)$, which are invariant under (2.8) and admit holomorphic extension to $\mathcal{O}(x+iy)$. Note that the symmetry of the observables also enhances to (2.9). Its expectation value can be defined as

$$\begin{aligned}\Phi(t) &= \langle \mathcal{O}(x^{(\eta)}(t) + iy^{(\eta)}(t)) \rangle_\eta \\ &= \int dxdy \mathcal{O}(x+iy) P(x,y;t),\end{aligned} \tag{2.12}$$

where we have defined the probability distribution of $x_k^{(\eta)}(t)$ and $y_k^{(\eta)}(t)$ by

$$P(x,y;t) = \left\langle \prod_k \delta(x_k - x_k^{(\eta)}(t)) \delta(y_k - y_k^{(\eta)}(t)) \right\rangle_\eta. \tag{2.13}$$

Under certain conditions, we can show that

$$\lim_{t\to\infty} \lim_{\epsilon\to 0} \Phi(t) = \frac{1}{Z} \int dx \mathcal{O}(x) w(x), \tag{2.14}$$

which implies that the CLM is justified.

B. The t-evolution of the expectation value

Let us first discuss the t-evolution of the expectation value $\Phi(t)$, which is given by

$$\begin{aligned}\Phi(t+\epsilon) &= \langle \mathcal{O}(x^{(\eta)}(t+\epsilon) + iy^{(\eta)}(t+\epsilon)) \rangle_\eta \\ &= \int dxdy \mathcal{O}(x+iy) P(x,y;t+\epsilon).\end{aligned} \tag{2.15}$$

Note that the t-evolution of $P(x,y;t)$ can be readily obtained from the complex Langevin equation (2.10) and (2.11) as

$$\begin{aligned}P(x,y;t+\epsilon) &= \frac{1}{\mathcal{N}} \int d\eta \mathrm{e}^{-\frac{1}{4}\{\frac{1}{N_\mathrm{R}}\eta_k^{(\mathrm{R})2} + \frac{1}{N_\mathrm{I}}\eta_k^{(\mathrm{I})2}\}} \\ &\quad \times \int d\tilde{x} d\tilde{y} \delta(x - \tilde{x} - \epsilon \mathrm{Re} v(\tilde{z}) - \sqrt{\epsilon}\eta^{(\mathrm{R})}) \delta(y - \tilde{y} - \epsilon \mathrm{Im} v(\tilde{z}) - \sqrt{\epsilon}\eta^{(\mathrm{I})}) \tilde{P}(\tilde{x},\tilde{y};t) \\ &= \frac{1}{\epsilon\mathcal{N}} \int d\tilde{x} d\tilde{y} \exp\left[-\left\{ \frac{(x - \tilde{x} - \epsilon \mathrm{Re} v(\tilde{z}))^2}{4\epsilon N_\mathrm{R}} + \frac{(y - \tilde{y} - \epsilon \mathrm{Im} v(\tilde{z}))^2}{4\epsilon N_\mathrm{I}} \right\} \right] \tilde{P}(\tilde{x},\tilde{y};t),\end{aligned} \tag{2.16}$$

where $\mathcal{N} = 2\pi\sqrt{N_\mathrm{R} N_\mathrm{I}}$ is just a normalization constant, and we have defined the probability distribution for $\tilde{z}^{(\eta)}(t)$ in (2.10) as

$$\tilde{P}(\tilde{x},\tilde{y};t) = \int dxdy \delta(\tilde{x} - \mathrm{Re}(z^{(g)})) \delta(\tilde{y} - \mathrm{Im}(z^{(g)})) P(x,y;t), \tag{2.17}$$

$$z_k^{(g)} = g_{kl}(x,y) z_l. \tag{2.18}$$

Using (2.16) in (2.15), we obtain

$$\Phi(t+\epsilon) = \int dxdy \mathcal{O}(x+iy) \int d\tilde{x} d\tilde{y}\, \tilde{P}(\tilde{x},\tilde{y};t) \frac{1}{\epsilon\mathcal{N}} \exp\left[-\left\{ \frac{(x - \tilde{x} - \epsilon \mathrm{Re} v(\tilde{z}))^2}{4\epsilon N_\mathrm{R}} + \frac{(y - \tilde{y} - \epsilon \mathrm{Im} v(\tilde{z}))^2}{4\epsilon N_\mathrm{I}} \right\} \right]. \tag{2.19}$$

Here we make an important assumption. Let us note that the convergence of the integral (2.12) or (2.19) is not guaranteed because the observable $|\mathcal{O}(x+iy)|$ can become infinitely large, and therefore it is possible that the expectation value of $\mathcal{O}(x+iy)$ is ill-defined. We restrict the observables to those for which the integral (2.12) converges absolutely at any $t \geq 0$. This is legitimate since we are concerned with a situation in which one obtains a finite result, but it is wrong in the sense that (2.14) does not hold.

Under the above assumption, we can exchange the order of integration in (2.19) due to Fubini's theorem, and rewrite it as

$$\Phi(t+\epsilon) = \int dxdy \mathcal{O}_\epsilon(x+iy)\tilde{P}(x,y;t), \tag{2.20}$$

where we have defined

$$\begin{aligned}\mathcal{O}_\epsilon(z) &= \frac{1}{\epsilon\mathcal{N}}\int d\tilde{x}d\tilde{y}\exp\left[-\left\{\frac{(\tilde{x}-x-\epsilon\mathrm{Re}v(z))^2}{4\epsilon N_\mathrm{R}} + \frac{(\tilde{y}-y-\epsilon\mathrm{Im}v(z))^2}{4\epsilon N_\mathrm{I}}\right\}\right]\mathcal{O}(\tilde{x}+i\tilde{y}) \\ &= \frac{1}{\mathcal{N}}\int d\eta \mathrm{e}^{-\frac{1}{4}\{\frac{1}{N_\mathrm{R}}\eta_k^{(\mathrm{R})2}+\frac{1}{N_\mathrm{I}}\eta_k^{(\mathrm{I})2}\}}O(z+\epsilon v(z)+\sqrt{\epsilon}\eta).\end{aligned} \tag{2.21}$$

Note that if $\mathcal{O}(z)$ and $v_k(z)$ are holomorphic, so is $\mathcal{O}_\epsilon(z)$. When we say "holomorphic," we admit the case in which the function has singular points.

In order to proceed further, we expand (2.21) with respect to ϵ and perform the integration over η. After some algebra, we get (See Appendix A of Ref. [18] for derivation)

$$\mathcal{O}_\epsilon(z) = :\mathrm{e}^{\epsilon L}:\mathcal{O}(z), \tag{2.22}$$

where the expression $\mathrm{e}^{\epsilon L}$ is a short-hand notation for

$$\mathrm{e}^{\epsilon L} \equiv \sum_{n=0}^{\infty}\frac{1}{n!}\epsilon^n L^n, \tag{2.23}$$

and the operator L is defined by

$$L = \left(\mathrm{Re}v_k(z) + N_\mathrm{R}\frac{\partial}{\partial x_k}\right)\frac{\partial}{\partial x_k} + \left(\mathrm{Im}v_k(z) + N_\mathrm{I}\frac{\partial}{\partial y_k}\right)\frac{\partial}{\partial y_k}. \tag{2.24}$$

The symbol $:\ldots:$ in (2.22) implies that the operators are ordered in such a way that derivative operators appear on the right; e.g., $:(f(x)+\partial)^2: = f(x)^2 + 2f(x)\partial + \partial^2$.

Since $\mathcal{O}(z)$ is a holomorphic function of z, we have

$$\begin{aligned}L\mathcal{O}(z) &= \left(\mathrm{Re}v_k(z) + N_\mathrm{R}\frac{\partial}{\partial z_k}\right)\frac{\partial\mathcal{O}}{\partial z_k} + \left(\mathrm{Im}v_k(z) + iN_\mathrm{I}\frac{\partial}{\partial z_k}\right)\left(i\frac{\partial\mathcal{O}}{\partial z_k}\right) \\ &= \left(v_k(z) + (N_\mathrm{R}-N_\mathrm{I})\frac{\partial}{\partial z_k}\right)\frac{\partial\mathcal{O}}{\partial z_k} \\ &= \tilde{L}\mathcal{O}(z),\end{aligned} \tag{2.25}$$

where we have used (2.5) and defined

$$\tilde{L} = \left(\frac{\partial}{\partial z_k} + v_k(z)\right)\frac{\partial}{\partial z_k}. \tag{2.26}$$

Hence we can rewrite (2.22) as

$$\mathcal{O}_\epsilon(z) = :\mathrm{e}^{\epsilon\tilde{L}}:\mathcal{O}(z). \tag{2.27}$$

Plugging (2.27) in (2.20), we formally obtain

$$\begin{aligned}\Phi(t+\epsilon) &= \sum_{n=0}^{\infty}\frac{1}{n!}\epsilon^n\int dxdy(:\tilde{L}^n:\mathcal{O}(z))\tilde{P}(x,y;t) \\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\epsilon^n\int dxdy(:\tilde{L}^n:\mathcal{O}(z))|_{z^{(g)}}P(x,y;t) \\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\epsilon^n\int dxdy(:\tilde{L}^n:\mathcal{O}(z))P(x,y;t).\end{aligned} \tag{2.28}$$

In the third equality, we have used the fact that $:\tilde{L}^n:\mathcal{O}(z)$ are invariant under the complexified symmetry transformation (2.9). Thus we find [18] that the effect of the gauge cooling represented by g disappears in the t-evolution of observables invariant under the symmetry transformation (2.9), although the t-evolution of the probability distribution $P(x,y;t)$ is affected nontrivially by the gauge cooling as in (2.16).

If the ϵ-expansion (2.28) is valid, we can truncate the infinite series for sufficiently small ϵ as

$$\Phi(t+\epsilon) = \Phi(t) + \epsilon\int dxdy\{\tilde{L}\mathcal{O}(z)\}P(x,y;t) + O(\epsilon^2), \tag{2.29}$$

which implies that the $\epsilon \to 0$ limit can be taken without any problem, and we get

$$\frac{d}{dt}\Phi(t) = \int dxdy\{\tilde{L}\mathcal{O}(z)\}P(x,y;t). \tag{2.30}$$

However, it is known from the previous argument [13,14] using a continuous Langevin time that there are cases in which (2.30) does not hold due to the failure of the integration by parts. In the present argument, the reason why (2.30) can be violated should be attributed to the possible breakdown of the expression (2.28). Note that the operator $\tilde{L}^n$ involves the nth power of the drift term $v_k(z)$ in (2.26), which may become infinitely large. Therefore, the integral that appears in (2.28) may be divergent for large enough n.

We emphasize here that what we have done in this section is just an alternative presentation of the known problem that (2.30) can be violated. In particular, the previous argument using a continuous Langevin time is absolutely correct since the discretized complex Langevin equation approaches smoothly the continuum one in the $\epsilon \to 0$ limit. Note also that the problem under discussion cannot be solved by using a sufficiently small ϵ or an adaptive step-size [29]. The advantage of our argument using a discretized Langevin time is that we can interpret the failure of the integration by parts in the previous argument as the breakdown of the ϵ-expansion (2.28) due to the appearance of a large drift term. This makes it possible to compare the condition required for the validity of the expression (2.30) with the one discussed in the next section.

C. Subtlety in the use of time-evolved observables

In this section we assume that the problem discussed in the previous section does not occur and that (2.30) holds. Repeating this argument for $\tilde{L}^n \mathcal{O}(z)$, we obtain

$$\left(\frac{d}{dt}\right)^n \Phi(t) = \int dxdy \{\tilde{L}^n \mathcal{O}(z)\} P(x,y;t). \tag{2.31}$$

Therefore, a finite time-evolution can be written formally as[5]

$$\Phi(t+\tau) = \sum_{n=0}^{\infty} \frac{1}{n!} \tau^n \int dxdy \{\tilde{L}^n \mathcal{O}(z)\} P(x,y;t), \tag{2.32}$$

which is similar to (2.28). In order for this expression to be valid for a finite τ, however, it is not sufficient to assume that the integral that appears in (2.32) is convergent for arbitrary n. What matters is the convergence radius of the infinite series (2.32). In the previous argument, the proof of the key identity (2.14) was given assuming implicitly that the convergence radius is infinite. This is actually a too strong assumption, which is not satisfied even in cases where the CLM is known to give correct results (See, e.g., our results in Sec. III.). Below we show that we can modify the proof slightly so that we only have to assume that the convergence radius $\tau_{\rm conv}(t)$, which depends on t in general, is bounded from below as $\tau_{\rm conv}(t) \geq \tau_0 > 0$ for $0 \leq t < \infty$.

In order to show (2.14), we first prove the lemma

$$\int dxdy \{\tilde{L}^n \mathcal{O}(x+iy)\} P(x,y;t) = \int dx \{(L_0)^n \mathcal{O}(x)\} \rho(x;t) \tag{2.33}$$

for arbitrary integer n and arbitrary $t \geq 0$, where the operator L_0 is defined by

$$L_0 = \left(\frac{\partial}{\partial x_k} + v_k(x)\right) \frac{\partial}{\partial x_k}, \tag{2.34}$$

and the complex valued function $\rho(x;t)$ is defined as the solution to the Fokker-Planck (FP) equation

$$\frac{\partial \rho}{\partial t} = (L_0)^\top \rho = \frac{\partial}{\partial x_k}\left(\frac{\partial}{\partial x_k} - v_k(x)\right)\rho, \tag{2.35}$$

$$\rho(x;0) = \rho(x). \tag{2.36}$$

Here the symbol $L_0^\top$ is defined as an operator satisfying $\langle L_0 f, g\rangle = \langle f, L_0^\top g\rangle$, where $\langle f, g\rangle \equiv \int f(x) g(x) dx$, assuming that f and g are functions that allow integration by parts. The initial condition is assumed to be

$$P(x,y;0) = \rho(x)\delta(y), \tag{2.37}$$

where $\rho(x) \geq 0$ and $\int dx \rho(x) = 1$, so that (2.33) is trivially satisfied at $t = 0$.

The proof of (2.33) is then given by induction with respect to t. Let us assume that (2.33) holds at $t = t_0$. Then we obtain

$$\int dxdy \{e^{\tau \tilde{L}} \mathcal{O}(x+iy)\} P(x,y;t_0) = \int dx \{e^{\tau L_0} \mathcal{O}(x)\} \rho(x;t_0), \tag{2.38}$$

where τ should be smaller than the convergence radius of the τ-expansion (2.32) at $t = t_0$. [The τ-expansion on the right-hand side of (2.38) is expected to have no problems due to the properties of the complex weight $\rho(x;t_0)$ obtained by solving the FP equation (2.35) for a well-defined system.] Since taking the derivative with respect to τ does not alter the convergence radius, we obtain

$$\int dxdy \{e^{\tau \tilde{L}} \tilde{L}^n \mathcal{O}(x+iy)\} P(x,y;t_0) = \int dx \{e^{\tau L_0} (L_0)^n \mathcal{O}(x)\} \rho(x;t_0) \tag{2.39}$$

for arbitrary n. Note that

$$\text{l.h.s of Eq. (2.39)} = \int dxdy \{\tilde{L}^n \mathcal{O}(x+iy)\} P(x,y;t_0+\tau), \tag{2.40}$$

where we have used a relation like (2.32) for the observable $\tilde{L}^n \mathcal{O}(x+iy)$, and

[5] Subtlety of Eq. (2.32) for finite τ at $t = 0$ was discussed in Ref. [30] in a one-variable case with a complex quartic action. We thank M. Niedermaier for bringing our attention to this work.

$$\text{r.h.s of Eq. (2.39)} = \int dx\{(L_0)^n \mathcal{O}(x)\} e^{\tau (L_0)^\top} \rho(x;t_0) = \int dx\{(L_0)^n \mathcal{O}(x)\}\rho(x;t_0+\tau), \tag{2.41}$$

where we have used integration by parts[6] in the first equality, and (2.35) in the second equality. Thus we find that (2.33) holds at $t = t_0 + \tau$, which completes the proof of (2.33) for arbitrary $t \geq 0$.

In order to show (2.14), we only need to consider the $n = 0$ case in (2.33), which reads

$$\int dxdy \mathcal{O}(x+iy)P(x,y;t) = \int dx \mathcal{O}(x)\rho(x;t). \tag{2.42}$$

Note that Eq. (2.35) has a t-independent solution

$$\rho_{\text{time-indep}}(x) = \frac{1}{Z} w(x). \tag{2.43}$$

According to the argument given in Ref. [15], the solution to (2.35) asymptotes to (2.43) at large t if (2.42) holds and $P(x,y;t)$ converges to a unique distribution in the $t \to \infty$ limit. Hence, (2.14) follows from (2.42).

D. The condition for correct convergence

Let us discuss the condition for the validity of the ϵ-expansion (2.28) and the condition for the τ-expansion (2.32) to have a finite convergence radius. In fact, it is the latter that is stronger. As we mentioned in Sec. II B, these conditions are related to the behavior of the probability distribution for such configurations (x,y) that make the drift term $v_k(z)$ large. More precisely, we are concerned with the magnitude of the drift term, which may be defined as

$$u(z) = \max_g \max_{1\leq i \leq N} |v_i(z^{(g)})|, \tag{2.44}$$

where g represents a symmetry transformation (2.8) of the original theory.[7] Note that $u(z)$ thus defined is invariant under (2.8). The integral that appears in (2.28) and (2.32) for each n involves

$$\int dxdy u(z)^n P(x,y;t) = \int_0^\infty du u^n p(u;t) \tag{2.45}$$

as the most dominant contribution, where we have defined the probability distribution of the magnitude $u(z)$ by

$$p(u;t) \equiv \int dxdy \delta(u(z)-u) P(x,y;t). \tag{2.46}$$

If $p(u;t)$ is only power-law suppressed at large u, the integral (2.45) is divergent for sufficiently large n. Therefore, in order for (2.45) to be convergent for arbitrary n, $p(u;t)$ should fall off faster than any power law. This is required for the ϵ-expansion (2.28) or the τ-expansion (2.32) to be valid.

Here we consider the case in which $p(u;t)$ is exponentially suppressed as $p(u;t) \sim e^{-\kappa u}$ at large u. Then, the integral (2.45) can be estimated as

$$\int_0^\infty du u^n p(u;t) \sim \frac{n!}{\kappa^{n+1}}. \tag{2.47}$$

Plugging this into (2.32), we find that the convergence radius of the infinite series can be estimated as $\tau \sim \kappa$. This implies that $p(u;t)$ has to fall off exponentially or faster in order for the convergence radius of the τ-expansion (2.32) to be nonzero, which is important in our argument given in Sec. II C.

Let us discuss the subtlety of the ϵ-expansion (2.28) in more detail. Note that $\Phi(t+\epsilon)$ defined by (2.19) is a finite well-defined quantity for a finite ϵ under the assumption made below Eq. (2.19). Nevertheless, the ϵ-expansion (2.28) can be ill-defined. This can happen because the expansion parameter ϵ is multiplied to the drift term in (2.19), which can become infinitely large in the integral. In order to illustrate this point, let us consider a simple integral

$$I = \int_{-1}^{1} dx e^{-\epsilon/x^2}, \tag{2.48}$$

which is clearly well-defined for arbitrary $\epsilon \geq 0$. However, if we expand the integrand with respect to ϵ, we get

$$I = \sum_{n=0}^{\infty} \frac{1}{n!}\epsilon^n (-1)^n \int_{-1}^{1} dx \frac{1}{x^{2n}}, \tag{2.49}$$

which is invalid because we obtain divergent terms for $n \geq 1$.

We can evaluate (2.48) as follows. Changing the integration variable $t = \sqrt{\epsilon}/x$, we get

$$I = 2\sqrt{\epsilon}\int_{\sqrt{\epsilon}}^{\infty} dt \frac{1}{t^2} e^{-t^2} = 2\{e^{-\epsilon} - \sqrt{\pi\epsilon}(1-\mathrm{Erf}(\sqrt{\epsilon}))\}, \tag{2.50}$$

where we have performed integration by parts in the first equality, and Erf is the error function. Expanding (2.50) with respect to ϵ, we obtain $O(\epsilon^{n/2})$ terms, which are absent in the formal expression (2.49).

[6] This is expected to be valid, as stated also in Refs. [13,14], due to the properties of the complex weight $\rho(x)$ obtained by solving the FP equation (2.35) for a well-defined system.

[7] In the case of $O(N)$ symmetry $g \in O(N)$, for instance, the definition (2.44) is equivalent to $u(z) = \max_{\vec{n}} |\vec{n}\cdot\vec{v}(z)|$, where the maximum is taken with respect to a unit vector $\vec{n}$ in $\mathbb{R}^N$.

E. Some comments on the previous argument

In this subsection, we clarify the relationship of our new argument and the previous one. Here we omit the gauge cooling for simplicity. In Refs. [13,14], the quantity

$$F(t,\tau) \equiv \int dxdy\mathcal{O}(z;\tau)P(x,y;t-\tau) \qquad (2.51)$$

was introduced with the time-evolved observable $\mathcal{O}(z;\tau) = \mathrm{e}^{\tau\tilde{L}}\mathcal{O}(z)$, and it was shown to be τ-independent for $0 \le \tau \le t$ by using the integration by parts

$$\int dxdy\mathcal{O}(z;\tau)L^{\top}P(x,y;t-\tau) = \int dxdy\{L\mathcal{O}(z;\tau)\}P(x,y;t-\tau). \qquad (2.52)$$

Note, however, that the quantity (2.51) has to be evaluated as

$$F(t,\tau) = \sum_{n=0}^{\infty}\frac{1}{n!}\tau^n \int dxdy\{\tilde{L}^n\mathcal{O}(z)\}P(x,y;t-\tau), \qquad (2.53)$$

where the infinite series on the right-hand side may have a finite convergence radius $\tau = \tau_{\mathrm{conv}}$. In that case, (2.53) is ill-defined for $\tau > \tau_{\mathrm{conv}}$. Our argument in Sec. II C avoids this problem by using (2.38) only for $\tau < \tau_{\mathrm{conv}}$ and employing the induction with respect to t instead.

Let us discuss the validity of the integration by parts (2.52). Expanding (2.16) with respect to ϵ, we obtain

$$P(x,y;t+\epsilon) = (:\mathrm{e}^{\epsilon L}:)^{\top}P(x,y;t). \qquad (2.54)$$

In the $\epsilon \to 0$ limit, we obtain the FP-like equation

$$\frac{\partial}{\partial t}P(x,y;t) = L^{\top}P(x,y;t). \qquad (2.55)$$

Using this, we obtain

$$\frac{\partial}{\partial t}F(t,\tau) = \int dxdy\mathcal{O}(z;\tau)\frac{\partial}{\partial t}P(x,y;t-\tau) = \int dxdy\mathcal{O}(z;\tau)L^{\top}P(x,y;t-\tau), \qquad (2.56)$$

which is the left-hand side of (2.52). On the other hand, our argument given before (2.30) implies that

$$\frac{\partial}{\partial t}F(t,\tau) = \int dxdy\{\tilde{L}\mathcal{O}(z;\tau)\}P(x,y;t-\tau) \qquad (2.57)$$

may or may not hold depending on the validity of the ϵ-expansion like (2.28). Note that the right-hand side of (2.57) is nothing but the right-hand side of (2.52) due to (2.25). From (2.56) and (2.57), we therefore find that the validity of the integration by parts (2.52) is equivalent to the validity of (2.57), which requires that the probability distribution of the drift term falls off faster than any power-law at large magnitude. This condition is slightly weaker than the one from the validity of the use of time-evolved observables for a finite time. Note that a function $f(x) = e^{-\sqrt{x}}$, for instance, falls off faster than any power-law at large x, and yet it is not suppressed exponentially at large x. Therefore, we consider that a necessary and sufficient condition for justifying the CLM is that the probability distribution of the drift term falls off exponentially or faster at large magnitude.

In the previous work [13,14], it was recognized that the probability distribution of the complexified dynamical variables should fall off fast enough at large absolute values to make sure that the integration by parts used in the argument is valid. However, the rate of the fall-off required to justify the CLM was not clear. This was also the case with the singular-drift problem [15]. How fast the probability distribution should fall off near the singularity was not clear. For this reason, while it was possible to understand the failure of the CLM found by comparison with correct results available from other methods, it was not possible to tell whether the results of the CLM are trustable or not without knowing the correct results in advance. The advantage of our condition based on the probability distribution of the drift term is that we can clearly state that it is the exponential fall-off that is required for justification of the CLM. This condition ensures not only the validity of the integration by parts used in the argument but also the validity of the use of time-evolved observables for a finite nonzero time. As we demonstrate in Sec. III, the condition indeed tells us clearly whether the results of the CLM are trustable or not.

Let us also comment on the property

$$\lim_{t\to\infty}\int dxdy\{\tilde{L}\mathcal{O}(z)\}P(x,y;t) = 0, \qquad (2.58)$$

which was proposed as a necessary condition for justifying the CLM [14]. From the viewpoint of our new argument, (2.58) follows from (2.30), which is true if the ϵ-expansion is valid. However, the quantity on the left-hand side of (2.58) is difficult to evaluate since the history of the observable $\tilde{L}\mathcal{O}(z)$ typically has spikes with different phase factors, and huge cancellations occur among configurations. This limits the usefulness of (2.58) as a necessary condition.

F. The case of the real Langevin method

In order to appreciate better the situation in the complex Langevin method, let us here consider the case of the real Langevin method [11], which is a standard method for a real-action system based on importance sampling. In this case, there is no need to complexify the dynamical variables, and the probability distribution $P(x;t)$ and the weight $\rho(x;t)$ are identical. The discussion in Sec. II C is not needed, and therefore the expressions like (2.28) and (2.32) do not have to make sense. Thus the issues concerning the time-evolved observables become totally irrelevant.

All we need to justify the method is to show that the discretized t-evolution of $P(x;t)$ like (2.16) reduces to the FP equation (2.35) in the $\epsilon \to 0$ limit. Note that the ϵ-expansion of (2.16) gives (2.54), and the FP equation is obtained if the expansion can be truncated at the order of ϵ. The problem occurs in the region of x, where the drift term $v_k(x)$ becomes large. However, the integral of $P(x;t)$ in that region is typically small, and it is expected to vanish in the $\epsilon \to 0$ limit. Therefore, we may expect that (2.16) reduces to the FP equation (2.35) in the $\epsilon \to 0$ limit. In order to confirm this, we have studied a system

$$Z = \int dx |x|^{-1/2} e^{-x^2/2}, \tag{2.59}$$

where x is a real variable. The drift term is given by $v(x) = -\frac{1}{2x} - x$, which diverges at $x = 0$. The probability distribution of the drift term is only power-law suppressed at large magnitude, but the distribution of x in the thermal equilibrium approaches $w(x) = |x|^{-1/2} e^{-x^2/2}$ as the step-size ϵ is reduced.

Applying the same argument to the case of the CLM, the FP-like equation (2.55) should be obtained in the $\epsilon \to 0$ limit. However, the ϵ-expansion (2.28) can still be subtle, and that is precisely the reason why the integration by parts (2.52) can be invalid.

III. DEMONSTRATION OF OUR CONDITION

In this section, we demonstrate our condition in Sec. II D, which is required to justify the CLM. For this purpose, we investigate two simple examples, in which the CLM was thought to fail due to the singular-drift problem and the excursion problem, respectively, in some parameter region. According to our new argument, however, these failures should be attributed to the appearance of a large drift term. We measure the probability distribution of the drift term and show that it is only power-law suppressed at large magnitude when the CLM fails, whereas it is exponentially suppressed when the CLM works. Thus the failures of the CLM can be understood in a unified manner. Our condition is also of great practical importance since it tells us clearly whether the obtained results are trustable or not.

A. A model with a singular drift

As a model with a singular drift, we consider the partition function [15]

$$Z = \int dx w(x), \qquad w(x) = (x + i\alpha)^p e^{-x^2/2}, \tag{3.1}$$

where x is a real variable and α and p are real parameters. For $\alpha \neq 0$ and $p \neq 0$, the weight $w(x)$ is complex, and the sign problem occurs.

We apply the CLM to (3.1). Since there is no symmetry that can be used for gauge cooling, we do not introduce the gauge cooling procedure (2.10) or the probability distribution (2.17) for the transformed variables. Otherwise, all the equations in the previous section apply to the present case by just setting the number of variables to $N = 1$. The drift term in this model is given by

$$v(z) = \frac{p}{z + i\alpha} - z, \tag{3.2}$$

which is singular at $z = -i\alpha$.

The complex Langevin simulation is performed for $p = 4$ with various values of α using the step-size $\epsilon = 10^{-5}$. The initial configuration is chosen to be $z = 0$, and the first 3×10^5 steps are discarded for thermalization. After that, we make 10^{10} steps and perform measurement every 10^3 steps. In Fig. 1 we plot the real part of the expectation value of $\mathcal{O}(z) = z^2$ against α. It is found that the CLM gives the correct results for $\alpha \gtrsim 3.7$.

In Fig. 2 we show the scatter plot of configurations obtained after thermalization for $\alpha = 5$ (Left) and $\alpha = 3$

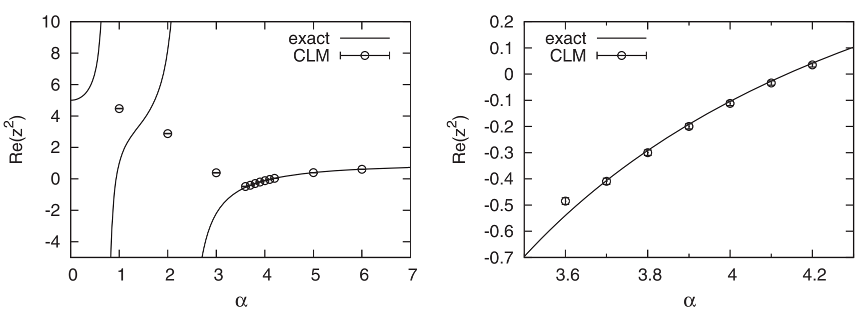

FIG. 1. (Left) The real part of the expectation value of $\mathcal{O}(z) = z^2$ obtained by the CLM for $p = 4$ is plotted against α. The solid line represents the exact result. (Right) Zoom-up of the same plot in the region $3.6 \leq \alpha \leq 4.2$.

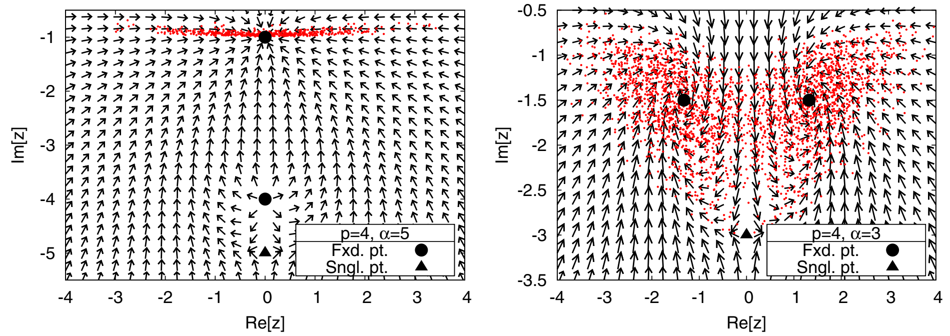

FIG. 2. The scatter plot of thermalized configurations (red dots) and the flow diagram (arrows) are shown for $\alpha = 5$ (Left) and $\alpha = 3$ (Right) with $p = 4$. Filled circles represent the fixed points, and the filled triangles represent the singular points.

(Right). The data points appear near the singular point $z = -i\alpha$ for $\alpha = 3$ but not for $\alpha = 5$. This change of behavior can be understood from the flow diagram in the same figure, which shows the normalized drift term $v(z)/|v(z)|$ by an arrow at each point. The fixed points of the flow diagram can be readily obtained by solving $v(z) = 0$. For $\alpha > 2\sqrt{p}$, there are two fixed points at

$$(x, y) = \left(0, -\frac{\alpha \pm \sqrt{\alpha^2 - 4p}}{2}\right), \qquad (3.3)$$

one of which ($-$) is attractive and the other ($+$) is repulsive. Since we adopt a real noise in the complex Langevin equation (2.3), the thermalized configurations appear near the horizontal line stemming from the attractive fixed point, and that is why no configuration appears near the singular point.

For $\alpha = 2\sqrt{p}$, the two fixed points merge into one at $(0, -\alpha/2)$, and for $\alpha < 2\sqrt{p}$, there are two fixed points at

$$(x, y) = \left(\pm\sqrt{p - \frac{\alpha^2}{4}}, -\frac{\alpha}{2}\right), \qquad (3.4)$$

which are vortex-like. In fact, there is a flow on the imaginary axis toward the singular point, which makes the thermalized configurations appear near it. Thus the property of the flow diagram changes qualitatively at $\alpha = 2\sqrt{p}$, which corresponds to $\alpha = 4$ in our case. This is indeed close to the critical value of α found by comparison with the exact result in Fig. 1 (Right).

According to our new argument given in the previous section, the appearance of thermalized configurations near the singularity of the drift term invalidates the CLM because the drift term can become large with a probability that is not suppressed exponentially. This is confirmed in Fig. 3, which shows the probability distribution for the magnitude of the drift term for various α within $3.6 \leq \alpha \leq 4.2$ in the semi-log (Left) and log-log (Right) plots. We find that the distribution falls off faster than exponential for $\alpha \geq 3.8$ and that its dependence on α in this region is very small. For $\alpha \leq 3.7$, the distribution follows the same behavior as those for $\alpha \geq 3.8$ at small u, but it starts to deviate from it at larger u. From the log-log plot, we find that the fall-off at large u is consistent with a power law. This change of behavior occurs near the value of α, where the CLM starts to give wrong results as shown in Fig. 1 (Right). In fact, at $\alpha = 3.7$, we cannot tell only from

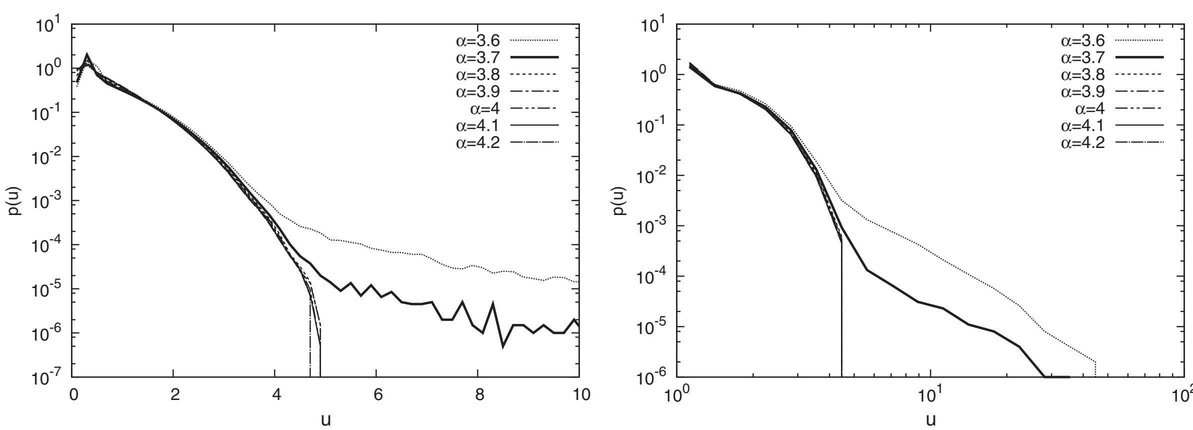

FIG. 3. The probability distribution $p(u)$ for the magnitude $u = |v|$ of the drift term is shown for various α within $3.6 \leq \alpha \leq 4.2$ in the semi-log (Left) and log-log (Right) plots.

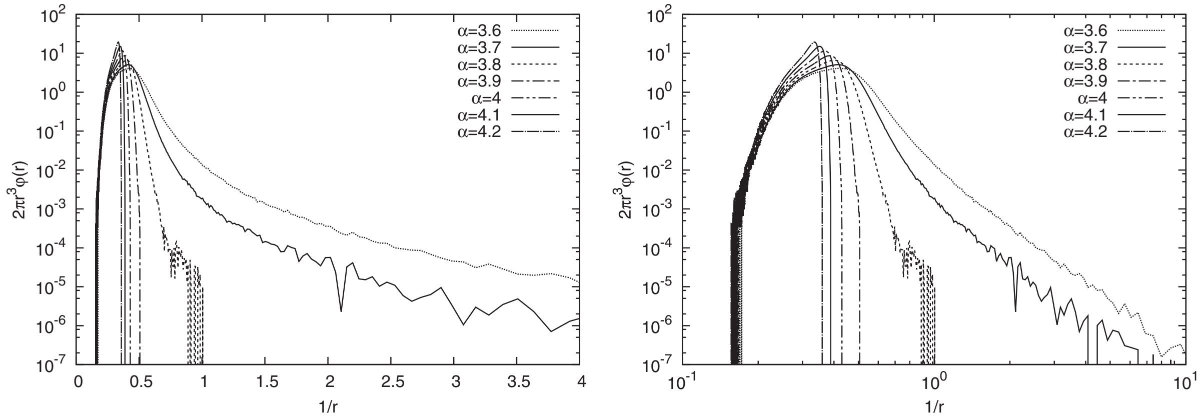

FIG. 4. The quantity $2\pi r^3\varphi(r)$, where $\varphi(r)$ is the radial distribution defined by (3.5), is shown as a function of $1/r$ for various α within $3.6 \le \alpha \le 4.2$ in the semi-log (Left) and log-log (Right) plots.

the expectation values of observables that the CLM is giving wrong results presumably because the discrepancies are too small to be measured. We consider this as a good feature of our condition.

In Ref. [15], the radial distribution

$$\varphi(r) = \frac{1}{2\pi r}\int P(x,y,\infty)\delta\Big(\sqrt{x^2+(y+\alpha)^2}-r\Big)dxdy \tag{3.5}$$

around the singular point $(x,y) = (0,-\alpha)$ was introduced to investigate the singular-drift problem. Since the magnitude of the drift term is given by $u \sim 1/r$, the probability distribution of the drift term is given by $p(u) \sim 2\pi r^3\varphi(r)$ at small r. In Fig. 4, we therefore show $2\pi r^3\varphi(r)$ as a function of $1/r$ in the semi-log (Left) and log-log (Right) plots. We observe a clear power-law tail for $\alpha \le 3.7$. Thus, the problem of the large drift term can also be detected by the radial distribution around the singularity if it is plotted in this way.

B. A model with a possibility of excursions

As a model with a possibility of excursions, we consider the partition function [19]

$$Z = \int dx w(x), \qquad w(x) = \mathrm{e}^{-\frac{1}{2}(A+iB)x^2-\frac{1}{4}x^4}, \tag{3.6}$$

where x is a real variable and A and B are real parameters. For $B \neq 0$, the weight $w(x)$ is complex and the sign problem occurs.

We apply the CLM to the model (3.6). The drift term is given by

$$v(z) = -(A+iB)z - z^3, \tag{3.7}$$

which can be decomposed into the real and imaginary parts as

$$\begin{aligned}\mathrm{Re}v(z) &= -(Ax - By + x^3 - 3xy^2),\\ \mathrm{Im}v(z) &= -(Ay + Bx + 3x^2y - y^3).\end{aligned} \tag{3.8}$$

Note that each component of the drift term can become infinitely large with both positive and negative signs at large $|x|$ and $|y|$, which means that there is a potential danger of excursions (or even runaways) in this model.

The complex Langevin simulation is performed for $A = 1$ with various values of B. The simulation parameters are the same as those in Sec. III A except that here we replace the step-size $\epsilon = 10^{-5}$ by $\epsilon = 0.01/|v(z)|$ when the magnitude of the drift term $|v(z)|$ exceeds 10^3. The use of such an adaptive step-size [29] is needed[8] to avoid the runaway problem that occurs at $B \ge 3$. In Fig. 5 we plot the imaginary part[9] of the expectation value of $\mathcal{O}(z) = z^2$. We find that the CLM gives correct results for $B \lesssim 2.8$.

In Fig. 6 we show the scatter plot of configurations obtained after thermalization for $B = 2$ (Left) and $B = 4$ (Right). The data points spread out in the large $|y|$ region for $B = 4$ but not for $B = 2$. This change of behavior can be understood from the flow diagram in the same figure. In fact, it was shown [19] that for $B < \sqrt{3}$, there is a striplike region $|y| \le C$ in which $\mathrm{Im}v(z) \le 0$ for $y > 0$ and $\mathrm{Im}v(z) \le 0$ for $y < 0$. In that case, the thermalized configurations are strictly restricted to $|y| \le C$ as far as a real noise is used in the complex Langevin equation (2.3). For $B > \sqrt{3}$, this does not occur. In fact, it was found that the distribution in the large $|x|$ and $|y|$ region is suppressed only by a power law [19] at sufficiently large B.

According to our new argument, this slow fall-off of the probability distribution of x and y invalidates the CLM because the drift term can become large with the probability

[8] The probability of $|v(z)|$ exceeding 10^3 is less than 10^{-4} even for the largest $B = 5$ we studied.

[9] The real part shows similar behaviors, but the discrepancies from the exact result at $B \gtrsim 3$ is less clear.

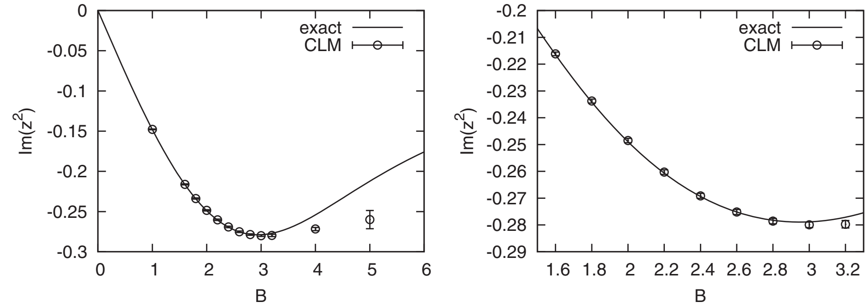

FIG. 5. (Left) The imaginary part of the expectation value of $\mathcal{O}(z) = z^2$ is plotted against B for $A = 1$. The solid line represents the exact result. (Right) Zoom-up of the same plot in the region $1.6 \leq B \leq 3.2$.

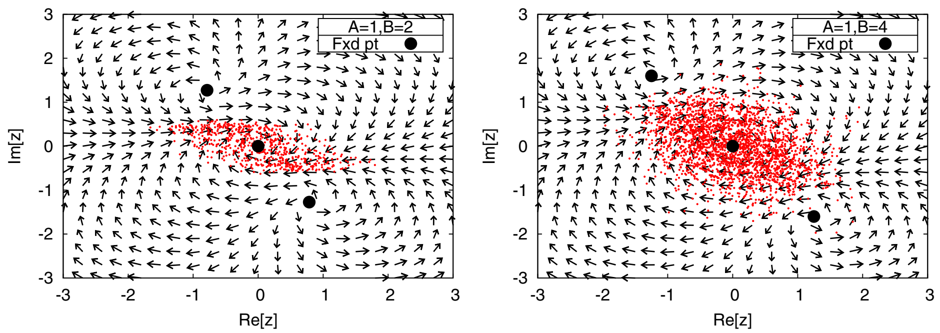

FIG. 6. The scatter plot of thermalized configurations (red dots) and the flow diagram (arrows) are shown for $B = 2$ (Left) and $B = 4$ (Right) with $A = 1$ in both cases. Filled circles represent the fixed points. There is no singular point in this model.

that is not suppressed exponentially. This is confirmed in Fig. 7, where we show the probability distribution for the magnitude of the drift term for various B within $1.6 \leq B \leq 3.2$ in the semi-log (Left) and log-log (Right) plots. We find that the distribution falls off exponentially for $B \leq 2.6$ and that its dependence on B in this region is small. For $B \geq 2.8$, the distribution follows the same behavior as those for $B \leq 2.6$ at small u, but it starts to deviate from it at larger u. From the log-log plot, we find that the fall-off at large u is consistent with a power law. This change of behavior occurs near the value of B, where the CLM starts to give wrong results as shown in Fig. 5 (Right). In fact, at $B = 2.8$, we cannot tell only from the expectation values of observables that the CLM is giving wrong results presumably because the discrepancies are too small to be measured.

Since the drift term is given by (3.8) as a function of x and y, it is clear that the large $|x|$ and large $|y|$ regions are responsible for the slow fall-off of the probability distribution of the drift term. In Fig. 8, we therefore show the y-distribution for various B within $1.6 \leq B \leq 3.2$ in the semi-log (Left) and log-log (Right) plots. We observe a

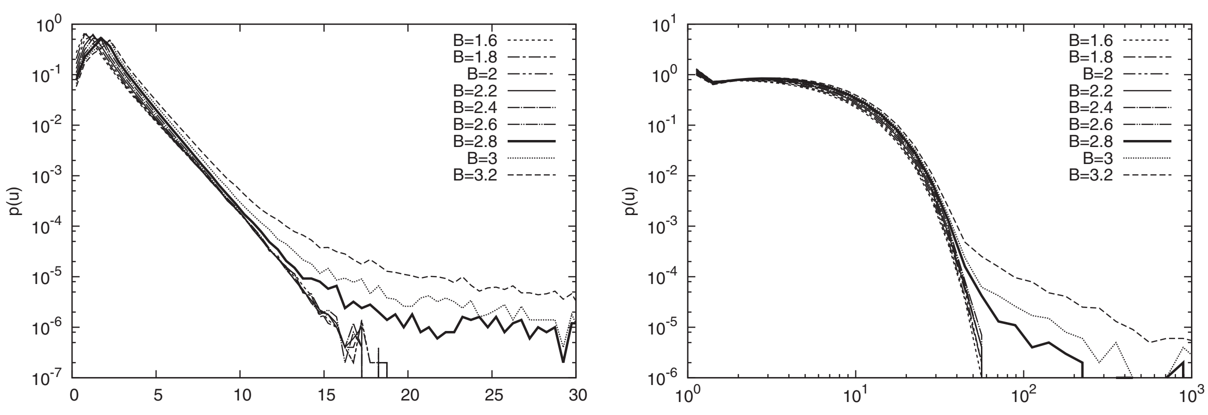

FIG. 7. The probability distribution $p(u)$ for the magnitude $u = |v|$ of the drift term is shown for various B within $1.6 \leq B \leq 3.2$ in the semi-log (Left) and log-log (Right) plots.

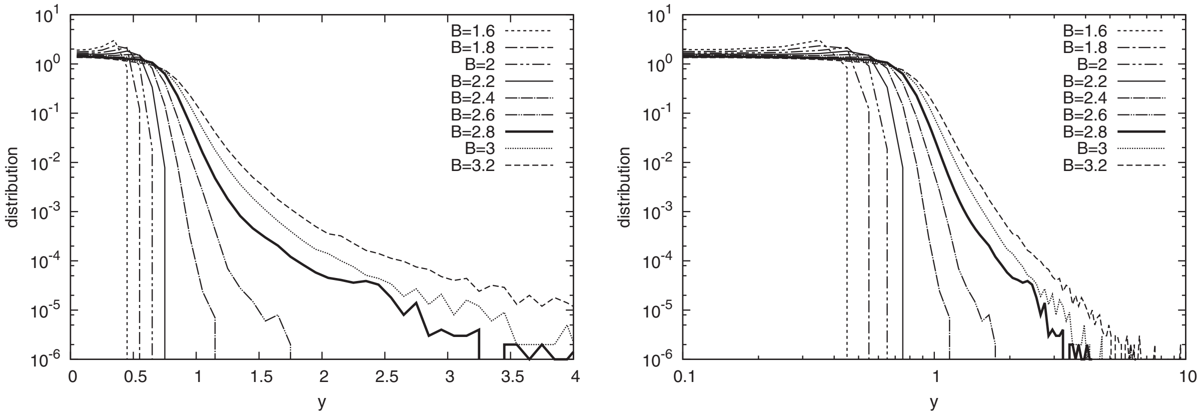

FIG. 8. (Left) The y-distribution of the thermalized configurations of $z = x + iy$ is shown for various B within $1.6 \leq B \leq 3.2$ in the semi-log (Left) and log-log (Right) plots.

slow fall-off consistent with a power law for $B \geq 2.8$. Thus, the problem of the large drift term can also be detected by the y-distribution. However, the change of behavior is clearer in the probability distribution $p(u)$ for the drift term.

IV. GENERALIZATION TO LATTICE GAUGE THEORY

In this section, we discuss the generalization of our argument in Sec. II to lattice gauge theory, which is defined by the partition function

$$Z = \int dU w(U) = \int \prod_{n\mu} dU_{n\mu} w(U), \tag{4.1}$$

where the weight $w(U)$ is a complex-valued function of the configuration $U = \{U_{n\mu}\}$ composed of link variables $U_{n\mu} \in \mathrm{SU}(3)$, and the integration measure $dU_{n\mu}$ represents the Haar measure for the SU(3) group. The only complication compared with the case discussed in Sec. II comes from the fact that the dynamical variables take values on a group manifold. The Langevin equation in such a case with a real action is discussed intensively in Refs. [31–35]. Using this formulation, we can easily generalize our discussions to the case of lattice gauge theory. In Sec. IV D we discuss a new possibility for the gauge cooling, which can reduce the magnitude of the drift term directly.

A. The complex Langevin method

In the Langevin equation, the drift term is given by

$$v_{an\mu}(U) = \frac{1}{w(U)} D_{an\mu} w(U), \tag{4.2}$$

where we have defined the derivative operator $D_{an\mu}$, which acts on a function $f(U)$ of the unitary gauge configuration as

$$D_{an\mu} f(U) = \frac{\partial}{\partial x} f(\mathrm{e}^{ixt_a} U_{n\mu}) \Big|_{x=0} \tag{4.3}$$

with t_a being the generators of the SU(3) group normalized by $\mathrm{tr}(t_a t_b) = \delta_{ab}$. When the weight $w(U)$ is complex, the drift term (4.2) becomes complex, and therefore, the link variables evolve into $\mathrm{SL}(3, \mathbb{C})$ matrices (i.e., 3×3 general complex matrices with the determinant one) even if one starts from a configuration of SU(3) matrices. Let us therefore complexify the link variables as $U_{n\mu} \mapsto \mathcal{U}_{n\mu} \in \mathrm{SL}(3, \mathbb{C})$. Then, the discretized complex Langevin equation is given by

$$\begin{aligned} &\mathcal{U}_{n\mu}^{(\eta)}(t+\epsilon) \\ &\quad = \exp\left\{ i \sum_a (\epsilon v_{an\mu}(\mathcal{U}) + \sqrt{\epsilon}\eta_{an\mu}(t)) t_a \right\} \mathcal{U}_{n\mu}^{(\eta)}(t), \end{aligned} \tag{4.4}$$

where the drift term $v_{an\mu}(\mathcal{U})$ is obtained by analytically continuing (4.2). The probabilistic variables $\eta_{an\mu}(t)$ are defined similarly to (2.4).

The lattice gauge theory is invariant under the SU(3) gauge transformation

$$U'_{n\mu} = g_n U_{n\mu} g_{n+\hat{\mu}}^{-1}, \tag{4.5}$$

where $g_n \in \mathrm{SU}(3)$. When one complexifies the variables $U_{n\mu} \mapsto \mathcal{U}_{n\mu} \in \mathrm{SL}(3, \mathbb{C})$, the symmetry property of the drift term and the observables naturally enhances to the $\mathrm{SL}(3, \mathbb{C})$ gauge symmetry that can be obtained by complexifying the original Lie group. Thus, instead of (4.5), one obtains

$$\mathcal{U}'_{n\mu} = g_n \mathcal{U}_{n\mu} g_{n+\hat{\mu}}^{-1} \tag{4.6}$$

with $g_n \in \mathrm{SL}(3, \mathbb{C})$.

The gauge cooling [5] modifies the complex Langevin equation (4.4) into

$$\tilde{\mathcal{U}}_{n\mu}^{(\eta)}(t) = g_n \mathcal{U}_{n\mu}^{(\eta)}(t) g_{n+\hat{\mu}}^{-1}, \tag{4.7}$$

$$\mathcal{U}_{n\mu}^{(\eta)}(t+\epsilon) = \exp\left\{ i \sum_a (\epsilon v_{an\mu}(\tilde{\mathcal{U}}) + \sqrt{\epsilon}\eta_{an\mu}(t)) t_a \right\} \tilde{\mathcal{U}}_{n\mu}^{(\eta)}(t), \tag{4.8}$$

where g_n is an element of the complexified Lie group chosen appropriately as a function of the configuration $\mathcal{U}^{(\eta)}(t)$ before cooling. We regard (4.7) and (4.8) as describing the t-evolution of $\mathcal{U}_{n\mu}^{(\eta)}(t)$ and treat $\tilde{\mathcal{U}}_{n\mu}^{(\eta)}(t)$ as an intermediate object. The basic idea is to determine g in such a way that the modified Langevin process does not suffer from the problem of the original Langevin process (4.4).

We consider observables $\mathcal{O}(U)$, which are gauge invariant and admit holomorphic extension to $\mathcal{O}(\mathcal{U})$. Note that the symmetry of the observables also enhances to (4.6). Its expectation value can be defined as

$$\Phi(t) = \langle \mathcal{O}(\mathcal{U}^{(\eta)}(t)) \rangle_\eta = \int d\mathcal{U} \mathcal{O}(\mathcal{U}) P(\mathcal{U}; t), \tag{4.9}$$

where we have defined the probability distribution of $\mathcal{U}^{(\eta)}(t)$ by

$$P(\mathcal{U}; t) = \left\langle \prod_{n\mu} \delta(\mathcal{U}_{n\mu}, \mathcal{U}_{n\mu}^{(\eta)}(t)) \right\rangle_\eta, \tag{4.10}$$

using the delta function defined by

$$\int d\mathcal{U} f(\mathcal{U}) \delta(\mathcal{U}_{n\mu}, \tilde{\mathcal{U}}_{n\mu}) = f(\tilde{\mathcal{U}}) \tag{4.11}$$

for any function $f(\mathcal{U})$. The integration measure $d\mathcal{U}$ for the complexified link variables is given by the Haar measure for the $SL(3,\mathbb{C})$ group normalized appropriately. Under certain conditions, we can show that

$$\lim_{t\to\infty} \lim_{\epsilon\to 0} \Phi(t) = \frac{1}{Z} \int d\mathcal{U} \mathcal{O}(\mathcal{U}) w(\mathcal{U}), \tag{4.12}$$

which implies that the CLM is justified.

B. The *t*-evolution of the expectation value

Let us first discuss the t-evolution of the expectation value $\Phi(t)$, which is given by

$$\Phi(t+\epsilon) = \langle \mathcal{O}(\mathcal{U}^{(\eta)}(t+\epsilon)) \rangle_\eta = \int d\mathcal{U} \mathcal{O}(\mathcal{U}) P(\mathcal{U}; t+\epsilon). \tag{4.13}$$

Note that the t-evolution of $P(\mathcal{U}; t)$ can be readily obtained from the complex Langevin equation (4.7) and (4.8) as[10]

$$\begin{aligned} P(\mathcal{U}; t+\epsilon) = \frac{1}{\mathcal{N}} \int d\eta \mathrm{e}^{-\frac{1}{4}\{\frac{1}{N_\mathrm{R}}\eta_{an\mu}(t)^{(\mathrm{R})2} + \frac{1}{N_\mathrm{I}}\eta_{an\mu}(t)^{(\mathrm{I})2}\}} \\ \times \int d\tilde{\mathcal{U}} \delta\Big(\mathcal{U}, \exp\Big\{ i \sum_a (\epsilon v_{an\mu}(\tilde{\mathcal{U}}) \\ + \sqrt{\epsilon}\eta_{an\mu}(t)) t_a \Big\} \tilde{\mathcal{U}}_{n\mu} \Big) \tilde{P}(\tilde{\mathcal{U}}; t), \end{aligned} \tag{4.14}$$

where $\mathcal{N} = 2\pi\sqrt{N_\mathrm{R} N_\mathrm{I}}$ is just a normalization constant, and we have defined the probability distribution for $\tilde{\mathcal{U}}^{(\eta)}(t)$ in (4.7) as

$$\tilde{P}(\tilde{\mathcal{U}}; t) = \int d\mathcal{U} \delta(\tilde{\mathcal{U}}, \mathcal{U}^{(g)}) P(\mathcal{U}; t), \tag{4.15}$$

$$\mathcal{U}_{n\mu}^{(g)} = g_n \mathcal{U}_{n\mu} g_{n+\hat{\mu}}^{-1}. \tag{4.16}$$

Using (4.14) in (4.13), we obtain

$$\begin{aligned} \Phi(t+\epsilon) = \frac{1}{\mathcal{N}} \int d\eta \mathrm{e}^{-\frac{1}{4}\{\frac{1}{N_\mathrm{R}}\eta_{an\mu}^{(\mathrm{R})2} + \frac{1}{N_\mathrm{I}}\eta_{an\mu}^{(\mathrm{I})2}\}} \int d\mathcal{U} \mathcal{O}(\mathcal{U}) \int d\tilde{\mathcal{U}} \\ \times \delta\Big(\mathcal{U}, \exp\Big\{ i \sum_a (\epsilon v_{an\mu}(\tilde{\mathcal{U}}) + \sqrt{\epsilon}\eta_{an\mu}) t_a \Big\} \tilde{\mathcal{U}}_{n\mu} \Big) \\ \times \tilde{P}(\tilde{\mathcal{U}}; t). \end{aligned} \tag{4.17}$$

Here we make an important assumption. Let us note that the convergence of the integral (4.9) or (4.17) is not guaranteed because the observable $|\mathcal{O}(\mathcal{U})|$ can become infinitely large, and therefore it is possible that the expectation value of $\mathcal{O}(\mathcal{U})$ is ill-defined. We restrict the observables to those for which the integral (4.9) converges absolutely at any $t \geq 0$. This assumption is legitimate since we are concerned with a situation in which one obtains a finite result, but it is wrong in the sense that (4.12) does not hold.

Under the above assumption, we can exchange the order of integration in (4.17) due to Fubini's theorem, and rewrite it as

$$\Phi(t+\epsilon) = \int d\mathcal{U} \mathcal{O}_\epsilon(\mathcal{U}) \tilde{P}(\mathcal{U}; t), \tag{4.18}$$

where we have defined

[10] In the present case of lattice gauge theory, we cannot perform the integration over η explicitly as is done in the second equality of (2.16). The same comment applies also to Eqs. (4.17) and (4.19). Clearly, this is just a matter of expressions, which does not cause any practical problems.

$$\mathcal{O}_\epsilon(\mathcal{U}) = \frac{1}{\mathcal{N}}\int d\eta \mathrm{e}^{-\frac{1}{4}\{\frac{1}{N_\mathrm{R}}\eta_k^{(\mathrm{R})2}+\frac{1}{N_\mathrm{I}}\eta_k^{(\mathrm{I})2}\}} \times \mathcal{O}\left(\exp\left\{i\sum_a(\epsilon v_{an\mu}(\mathcal{U})+\sqrt{\epsilon}\eta_{an\mu})t_a\right\}\mathcal{U}_{n\mu}\right). \tag{4.19}$$

Note that if $\mathcal{O}(\mathcal{U})$ and $v_{an\mu}(\mathcal{U})$ are holomorphic, so is $\mathcal{O}_\epsilon(\mathcal{U})$. When we say "holomorphic," we admit the case in which the function has singular points.

In order to proceed further, we expand (4.19) with respect to ϵ and perform the integration over η. After some algebra, we get

$$\mathcal{O}_\epsilon(\mathcal{U}) = :\mathrm{e}^{\epsilon L}:\mathcal{O}(\mathcal{U}), \tag{4.20}$$

where the operator L is defined by

$$L = (\mathrm{Re}v_{an\mu}(\mathcal{U})+N_\mathrm{R}\mathcal{D}^{(\mathrm{R})}_{an\mu})\mathcal{D}^{(\mathrm{R})}_{an\mu} + (\mathrm{Im}v_{an\mu}(\mathcal{U})+N_\mathrm{I}\mathcal{D}^{(\mathrm{I})}_{an\mu})\mathcal{D}^{(\mathrm{I})}_{an\mu}. \tag{4.21}$$

In Eq. (4.21), we have defined the derivative operators

$$\mathcal{D}^{(\mathrm{R})}_{an\mu}f(\mathcal{U}) = \frac{\partial}{\partial x}f(\mathrm{e}^{ixt_a}\mathcal{U}_{n\mu})\Big|_{x=0}, \tag{4.22}$$

$$\mathcal{D}^{(\mathrm{I})}_{an\mu}f(\mathcal{U}) = \frac{\partial}{\partial y}f(\mathrm{e}^{-yt_a}\mathcal{U}_{n\mu})\Big|_{y=0}, \tag{4.23}$$

where $f(\mathcal{U})$ are functions on the complexified group manifold, which are not necessarily holomorphic, and x and y are real parameters. These derivative operators may be regarded as analogues of $\frac{\partial}{\partial x_k}$ and $\frac{\partial}{\partial y_k}$ used in Sec. II. For later convenience, let us also define

$$\mathcal{D}_{an\mu} = \frac{1}{2}(\mathcal{D}^{(\mathrm{R})}_{an\mu} - i\mathcal{D}^{(\mathrm{I})}_{an\mu}), \tag{4.24}$$

$$\overline{\mathcal{D}}_{an\mu} = \frac{1}{2}(\mathcal{D}^{(\mathrm{R})}_{an\mu} + i\mathcal{D}^{(\mathrm{I})}_{an\mu}), \tag{4.25}$$

which are analogues of $\frac{\partial}{\partial z_k} = \frac{1}{2}(\frac{\partial}{\partial x_k} - i\frac{\partial}{\partial y_k})$ and $\frac{\partial}{\partial \bar{z}_k} = \frac{1}{2}(\frac{\partial}{\partial x_k} + i\frac{\partial}{\partial y_k})$, respectively. Note that for a holomorphic function $f(\mathcal{U})$, we have $\overline{\mathcal{D}}_{an\mu}f(\mathcal{U}) = 0$, and hence

$$\mathcal{D}^{(\mathrm{R})}_{an\mu}f(\mathcal{U}) = \mathcal{D}_{an\mu}f(\mathcal{U}), \qquad \mathcal{D}^{(\mathrm{I})}_{an\mu}f(\mathcal{U}) = i\mathcal{D}_{an\mu}f(\mathcal{U}). \tag{4.26}$$

Since $\mathcal{O}(\mathcal{U})$ is a holomorphic function of $\mathcal{U}$, we have

$$\begin{aligned} L\mathcal{O}(\mathcal{U}) &= (\mathrm{Re}v_{an\mu}(\mathcal{U})+N_\mathrm{R}\mathcal{D}_{an\mu})\mathcal{D}_{an\mu}\mathcal{O}(\mathcal{U}) \\ &\quad + (\mathrm{Im}v_{an\mu}(\mathcal{U})+iN_\mathrm{I}\mathcal{D}_{an\mu})i\mathcal{D}_{an\mu}\mathcal{O}(\mathcal{U}) \\ &= \{v_{an\mu}(\mathcal{U})+(N_\mathrm{R}-N_\mathrm{I})\mathcal{D}_{an\mu}\}\mathcal{D}_{an\mu}\mathcal{O}(\mathcal{U}) \\ &= \tilde{L}\mathcal{O}(\mathcal{U}), \end{aligned} \tag{4.27}$$

where we have used (2.5) and defined

$$\tilde{L} = (\mathcal{D}_{an\mu}+v_{an\mu}(\mathcal{U}))\mathcal{D}_{an\mu}. \tag{4.28}$$

Hence we can rewrite (4.20) as

$$\mathcal{O}_\epsilon(\mathcal{U}) = :\mathrm{e}^{\epsilon\tilde{L}}:\mathcal{O}(\mathcal{U}). \tag{4.29}$$

Plugging (4.29) in (4.18), we formally obtain

$$\begin{aligned} \Phi(t+\epsilon) &= \sum_{n=0}^{\infty}\frac{1}{n!}\epsilon^n\int d\mathcal{U}(:\tilde{L}^n:\mathcal{O}(\mathcal{U}))\tilde{P}(\mathcal{U};t) \\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\epsilon^n\int d\mathcal{U}(:\tilde{L}^n:\mathcal{O}(\mathcal{U}))|_{\mathcal{U}^{(g)}}P(\mathcal{U};t) \\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\epsilon^n\int d\mathcal{U}(:\tilde{L}^n:\mathcal{O}(\mathcal{U}))P(\mathcal{U};t). \end{aligned} \tag{4.30}$$

In the third equality, we have used the fact that $:\tilde{L}^n:\mathcal{O}(\mathcal{U})$ are invariant under the $\mathrm{SL}(3,\mathbb{C})$ transformation. Thus we find [18] that the effect of the gauge cooling represented by g disappears in the t-evolution of the $\mathrm{SL}(3,\mathbb{C})$ invariant observables, although the t-evolution of the probability distribution $P(\mathcal{U};t)$ is affected nontrivially by the gauge cooling as in (4.14).

If the ϵ-expansion (4.30) is valid, we can truncate the infinite series for sufficiently small ϵ as

$$\Phi(t+\epsilon) = \Phi(t) + \epsilon\int d\mathcal{U}\{\tilde{L}\mathcal{O}(\mathcal{U})\}P(\mathcal{U};t) + O(\epsilon^2), \tag{4.31}$$

which implies that the $\epsilon \to 0$ limit can be taken without any problem, and we get

$$\frac{d}{dt}\Phi(t) = \int d\mathcal{U}\{\tilde{L}\mathcal{O}(\mathcal{U})\}P(\mathcal{U};t). \tag{4.32}$$

As we discussed in Sec. II B, Eq. (4.32) can be violated because of the possible breakdown of the expression (4.30). Note that the operator $\tilde{L}^n$ involves the nth power of the drift term $v_{an\mu}(\mathcal{U})$ in (4.28), which may become infinitely large. Therefore, the integral that appears in (4.30) may be divergent for large enough n.

C. Subtlety in the use of time-evolved observables

In this section we assume that the problem discussed in the previous section does not occur and that (4.32) holds. Repeating this argument for $\tilde{L}^n\mathcal{O}(\mathcal{U})$, we obtain

$$\left(\frac{d}{dt}\right)^n\Phi(t) = \int d\mathcal{U}\{\tilde{L}^n\mathcal{O}(\mathcal{U})\}P(\mathcal{U};t). \tag{4.33}$$

Therefore, a finite time-evolution can be written formally as

$$\Phi(t+\tau) = \sum_{n=0}^{\infty} \frac{1}{n!}\tau^n \int d\mathcal{U}\{\tilde{L}^n \mathcal{O}(\mathcal{U})\}P(\mathcal{U};t), \qquad (4.34)$$

which is similar to (4.30). In order for this expression to be valid for a finite τ, however, it is not sufficient to assume that the integral that appears in (4.34) is convergent for arbitrary n. What matters is the convergence radius of the infinite series (4.34). Below we provide a proof of the key identity (4.12) assuming that the convergence radius $\tau_{\mathrm{conv}}(t)$, which depends on t in general, is bounded from below as $\tau_{\mathrm{conv}}(t) \geq \tau_0 > 0$ for $0 \leq t < \infty$.

In order to show (4.12), we first prove the lemma

$$\int d\mathcal{U}\{\tilde{L}^n \mathcal{O}(\mathcal{U})\}P(\mathcal{U};t) = \int dU\{(L_0)^n \mathcal{O}(U)\}\rho(U;t) \qquad (4.35)$$

for arbitrary integer n and arbitrary $t \geq 0$, where the operator L_0 is defined by

$$L_0 = (D_{an\mu} + v_{an\mu}(U))D_{an\mu}, \qquad (4.36)$$

and the complex valued function $\rho(U;t)$ is defined as the solution to the FP equation

$$\frac{\partial}{\partial t}\rho(U;t) = L_0^{\top}\rho(U;t) = D_{an\mu}(D_{an\mu} - v_{an\mu}(U))\rho(U;t), \qquad (4.37)$$

$$\rho(U;0) = \rho(U). \qquad (4.38)$$

Here the symbol $L_0^{\top}$ is defined as an operator satisfying $\langle L_0, g\rangle = \langle f, L_0^{\top} g\rangle$, where $\langle f,g\rangle \equiv \int f(U)g(U)dU$, assuming that f and g are functions that allow integration by parts. The initial condition is assumed to be

$$P(\mathcal{U},;0) = \int dU\rho(U;0)\prod_{n\mu}\delta(\mathcal{U}_{n\mu}, U_{n\mu}) \qquad (4.39)$$

with $\rho(U;0) \geq 0$ and $\int dU\rho(U) = 1$, so that (4.35) is trivially satisfied at $t = 0$.

The proof of (4.35) is then given by induction. Let us assume that (4.35) holds at $t = t_0$. Then we obtain

$$\int d\mathcal{U}\{e^{\tau\tilde{L}}\mathcal{O}(\mathcal{U})\}P(\mathcal{U};t_0) = \int dU\{e^{\tau L_0}\mathcal{O}(U)\}\rho(U;t_0), \qquad (4.40)$$

where τ should be smaller than the convergence radius of the τ-expansion (4.34) at $t = t_0$. [The τ-expansion on the right-hand side of (4.40) is expected to have no problems due to the properties of the complex weight $\rho(U;t_0)$ obtained by solving the FP equation (4.37) for a well-defined system.] Since taking the derivative with respect to τ does not alter the convergence radius, we obtain

$$\int d\mathcal{U}\{e^{\tau\tilde{L}}\tilde{L}^n\mathcal{O}(\mathcal{U})\}P(\mathcal{U};t_0) = \int dU\{e^{\tau L_0}(L_0)^n\mathcal{O}(U)\}\rho(U;t_0) \qquad (4.41)$$

for arbitrary n. Note that

$$\text{l.h.s of Eq. (4.41)} = \int d\mathcal{U}\{\tilde{L}^n\mathcal{O}(\mathcal{U})\}P(\mathcal{U};t_0+\tau), \qquad (4.42)$$

where we have used a relation like (4.34), and

$$\text{r.h.s of Eq. (4.41)} = \int dU\{(L_0)^n\mathcal{O}(U)\}e^{\tau(L_0)^{\top}}\rho(U;t_0) = \int dU\{(L_0)^n\mathcal{O}(U)\}\rho(U;t_0+\tau), \qquad (4.43)$$

where we have used integration by parts, which is valid because the link variables $U_{n\mu}$ take values on the compact SU(3) manifold. In the second equality, we have used (4.37). Thus we find that (4.35) holds at $t = t_0 + \tau$, which completes the proof of (4.35) for arbitrary $t \geq 0$.

In order to show (4.12), we only need to consider the $n = 0$ case in (4.35), which reads

$$\int d\mathcal{U}\mathcal{O}(\mathcal{U})P(\mathcal{U};t) = \int dU\mathcal{O}(U)\rho(U;t). \qquad (4.44)$$

Note that Eq. (4.37) has a t-independent solution

$$\rho_{\mathrm{time-indep}}(U) = \frac{1}{Z}w(U). \qquad (4.45)$$

According to the argument given in Ref. [15], the solution to (4.37) asymptotes to (4.45) at large t if (4.44) holds and $P(\mathcal{U};t)$ converges to a unique distribution in the $t \to \infty$ limit. Hence, (4.12) follows from (4.44).

D. The magnitude of the drift term

Let us consider how to define the magnitude of the drift term, which is important in our condition for correct convergence discussed in Sec. II D. Corresponding to (2.44), we may define it as

$$u(\mathcal{U}) = \max_g \max_{an\mu} |v_{an\mu}(\mathcal{U}^{(g)})|, \qquad (4.46)$$

where g represents an SU(3) gauge transformation (4.5) of the original theory. Note that $u(\mathcal{U})$ thus defined is invariant under (4.5). This definition is not very useful, however, because taking the maximum with respect to the gauge transformation is not easy to perform. We would therefore like to propose an alternative one below, which is similar to (4.46) but much easier to deal with.

First we note that (4.46) can be rewritten as

$$u(\mathcal{U}) = \sqrt{\max_{g} \max_{an\mu} |v_{an\mu}(\mathcal{U}^{(g)})|^2}. \tag{4.47}$$

Next we replace the maximum with respect to the index a by the summation over it and define

$$\begin{aligned} \tilde{u}(\mathcal{U}) &= \sqrt{\max_{g} \max_{n\mu} \sum_{a=1}^{8} |v_{an\mu}(\mathcal{U}^{(g)})|^2} \\ &= \sqrt{\max_{n\mu} \sum_{a=1}^{8} |v_{an\mu}(\mathcal{U})|^2}, \end{aligned} \tag{4.48}$$

where the maximum with respect to the SU(3) gauge transformation can be omitted because the sum is gauge invariant. Since $u(\mathcal{U}) \le \tilde{u}(\mathcal{U}) \le 2\sqrt{2}u(\mathcal{U})$ holds, $\tilde{u}(\mathcal{U})$ may be considered a reasonable approximation to $u(\mathcal{U})$ for our purposes. If the probability distribution of $\tilde{u}(\mathcal{U})$ is suppressed exponentially at large magnitude, so is the probability distribution of $u(\mathcal{U})$, and vice versa.

The magnitude of the drift term defined by (4.46) or (4.48) is not invariant under the complexified $SL(3,\mathbb{C})$ gauge transformation. Therefore, we may try to make it smaller by the gauge cooling. In fact, the components of the drift term transform as an adjoint representation under the gauge transformation. Namely, if we define a 3×3 matrix $v_{n\mu}(\mathcal{U}) = \sum_{a=1}^{8} v_{an\mu}(\mathcal{U})t^a$, it transforms as

$$v_{n\mu}(\mathcal{U}^{(g)}) = g_n v_{n\mu}(\mathcal{U}) g_n^{-1}, \tag{4.49}$$

where $g_n \in SL(3,\mathbb{C})$. Therefore, we can use the gauge cooling to reduce the magnitude of the drift term associated with each site n defined as

$$u_n(\mathcal{U}) = \max_{\mu} \mathrm{tr}(v_{n\mu}^{\dagger}(\mathcal{U}) v_{n\mu}(\mathcal{U})). \tag{4.50}$$

Note that this can be done site by site unlike the gauge cooling with the unitarity norm [5], for instance, because of the transformation property (4.49) of the drift term.

V. SUMMARY AND DISCUSSIONS

In this paper we revisited the argument for justification of the CLM given originally in Refs. [13,14] and extended recently to the case including the gauge cooling procedure in Ref. [18]. In particular, we pointed out that the use of time-evolved observables, which are assumed to be justified for infinitely long time in the previous argument [13,14,18], can be subtle. In fact, we only have to use the time-evolved observables for a finite but nonzero time if we employ the induction with respect to the Langevin time in the argument. This still requires that the probability distribution of the drift term should be suppressed, at least, exponentially at large magnitude.

We also clarified the condition for the validity of the integration by parts, which was considered the main issue in the previous argument. Starting with a finite step-size ϵ for the discretized Langevin time, we found that the integration by parts is valid if the probability distribution of the drift term falls off faster than any power law at large magnitude. Since this is weaker than the condition obtained from the use of time-evolved observables for a finite time, we consider that the latter gives a necessary and sufficient condition for justifying the CLM.

Our condition based on the probability distribution of the drift term was demonstrated in two simple examples, in which the CLM was thought to fail due to the singular-drift problem and the excursion problem, respectively. We showed that the probability distribution is suppressed only by a power law when the method fails, whereas it is suppressed exponentially when the method works. Thus, our condition provides a simple way to judge whether the results obtained by the method are trustable or not.

The gauge cooling procedure can be included in our argument as we did in this paper extending our previous work [18]. Originally the gauge cooling was proposed to avoid the excursion problem [5], and recently it was used to solve the singular-drift problem by adopting different criteria for choosing the complexified gauge transformation [10]. Since the two problems are now understood as the problem of a large drift term in a unified manner, we may also choose the complexified gauge transformation in such a way that the magnitude of the drift term is reduced. In the lattice gauge theory, such gauge cooling can be done site by site due to the transformation property of the drift term. It would be interesting to see if the new type of gauge cooling, possibly combined with the previous ones, is effective in reducing the problem of a large drift term.

To conclude, we consider that the present work establishes the argument for justification of the CLM with or without gauge cooling. The crucial point for the success of the CLM turns out to be extremely simple. The probability of the drift term should be suppressed exponentially at large magnitude. Now that we have such a simple understanding of the method, we may also think of a new technique other than gauge cooling, which enables us to enlarge the range of applicability of the CLM further.

ACKNOWLEDGMENTS

The authors would like to thank J. Bloch, K. Fukushima, D. Sexty and Y. Tanizaki for valuable discussions. K. N. was supported by Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research (Kakenhi) Grants No. 26800154, MEXT SPIRE and JICFuS. J. N. was supported in part by Grant-in-Aid for Scientific

Research (No. 23244057 and 16H03988) from Japan Society for the Promotion of Science. S. S. was supported by the MEXT-Supported Program for the Strategic Research Foundation at Private Universities "Topological Science" (Grant No. S1511006).

Note added.—The present version of the paper has been changed significantly from the first version put on the arXiv, where we stated that the zero step-size limit is subtle. Through discussions with other people, we noticed at some point that this subtlety actually occurs only in the expression for time-evolved observables but *not* in the Fokker-Planck-like equation. We reached this understanding after reconsidering the case of the real Langevin method, in which the correct Fokker-Planck equation with a continuous Langevin time can be obtained even if the probability distribution of the drift term is suppressed only by a power law. These points are emphasized in Secs. II B and II F.

[1] G. Parisi, On complex probabilities, Phys. Lett. **131B**, 393 (1983).
[2] J. R. Klauder, Coherent state Langevin equations for canonical quantum systems with applications to the quantized Hall effect, Phys. Rev. A **29**, 2036 (1984).
[3] 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**, 114505 (2014).
[4] D. Sexty, Simulating full QCD at nonzero density using the complex Langevin equation, Phys. Lett. B **729**, 108 (2014).
[5] E. Seiler, D. Sexty, and I. O. Stamatescu, Gauge cooling in complex Langevin for lattice QCD with heavy quarks, Phys. Lett. B **723**, 213 (2013).
[6] A. Mollgaard and K. Splittorff, Complex Langevin dynamics for chiral Random Matrix Theory, Phys. Rev. D **88**, 116007 (2013).
[7] A. Mollgaard and K. Splittorff, Full simulation of chiral random matrix theory at nonzero chemical potential by complex Langevin, Phys. Rev. D **91**, 036007 (2015).
[8] Z. Fodor, S. D. Katz, D. Sexty, and C. Török, Complex Langevin dynamics for dynamical QCD at nonzero chemical potential: A comparison with multiparameter reweighting, Phys. Rev. D **92**, 094516 (2015).
[9] D. K. Sinclair and J. B. Kogut, Exploring complex-Langevin methods for finite-density QCD, *Proc. Sci.*, LATTICE2015 (2016) 153 [arXiv:1510.06367].
[10] K. Nagata, J. Nishimura, and S. Shimasaki, Gauge cooling for the singular-drift problem in the complex Langevin method: a test in Random Matrix Theory for finite density QCD, J. High Energy Phys. 07 (2016) 073.
[11] G. Parisi and Y. s. Wu, Perturbation theory without gauge fixing, Scientia Sinica **24**, 483 (1981).
[12] P. H. Damgaard and H. Huffel, Stochastic quantization, Phys. Rep. **152**, 227 (1987).
[13] G. Aarts, E. Seiler, and I. O. Stamatescu, The complex Langevin method: When can it be trusted?, Phys. Rev. D **81**, 054508 (2010).
[14] G. Aarts, F. A. James, E. Seiler, and I. O. Stamatescu, Complex Langevin: Etiology and diagnostics of its main problem, Eur. Phys. J. C **71**, 1756 (2011).
[15] J. Nishimura and S. Shimasaki, New insights into the problem with a singular drift term in the complex Langevin method, Phys. Rev. D **92**, 011501 (2015).
[16] G. Aarts, L. Bongiovanni, E. Seiler, D. Sexty, and I. O. Stamatescu, Controlling complex Langevin dynamics at finite density, Eur. Phys. J. A **49**, 89 (2013).
[17] 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. 09 (2016) 087.
[18] K. Nagata, J. Nishimura, and S. Shimasaki, Justification of the complex Langevin method with the gauge cooling procedure, Prog. Theor. Exp. Phys. (**2016**) 013B01.
[19] G. Aarts, P. Giudice, and E. Seiler, Localised distributions and criteria for correctness in complex Langevin dynamics, Ann. Phys. (Berlin) **337**, 238 (2013).
[20] J. Greensite, Comparison of complex Langevin and mean field methods applied to effective Polyakov line models, Phys. Rev. D **90**, 114507 (2014).
[21] E. Witten, Analytic continuation of Chern-Simons theory, AMS/IP Stud.Adv.Math. **50**, 347 (2011).
[22] 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**, 074506 (2012).
[23] 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**, 051501 (2013).
[24] H. Fujii, D. Honda, M. Kato, Y. Kikukawa, S. Komatsu, and T. Sano, Hybrid Monte Carlo on Lefschetz thimbles: A study of the residual sign problem, J. High Energy Phys. 10 (2013) 147.
[25] A. Mukherjee and M. Cristoforetti, Lefschetz thimble Monte Carlo for many-body theories: A Hubbard model study, Phys. Rev. B **90**, 035134 (2014).
[26] F. Di Renzo and G. Eruzzi, Thimble regularization at work: From toy models to chiral random matrix theories, Phys. Rev. D **92**, 085030 (2015).
[27] K. Fukushima and Y. Tanizaki, Hamilton dynamics for the Lefschetz thimble integration akin to the complex Langevin method, Prog. Theor. Exp. Phys. **2015**, 111A01 (2015).
[28] A. Alexandru, G. Basar, P. F. Bedaque, G. W. Ridgway, and N. C. Warrington, Sign problem and Monte Carlo

calculations beyond Lefschetz thimbles, J. High Energy Phys. 05 (2016) 053.

[29] G. Aarts, F. A. James, E. Seiler, and I.-O. Stamatescu, Adaptive stepsize and instabilities in complex Langevin dynamics, Phys. Lett. B **687**, 154 (2010).

[30] A. Duncan and M. Niedermaier, Temporal breakdown and Borel resummation in the complex Langevin method, Ann. Phys. (Berlin) **329**, 93 (2013).

[31] J. Alfaro and B. Sakita, Reports No. CCNY-HEP-82-16, No. C82-09-09.1.

[32] I. T. Drummond, S. Duane, and R. R. Horgan, The stochastic method for numerical simulations: Higher order corrections, Nucl. Phys. **B220**, 119 (1983).

[33] A. Guha and S. C. Lee, Stochastic quantization of matrix and lattice gauge models, Phys. Rev. D **27**, 2412 (1983).

[34] M. B. Halpern, Constrained quenched master field for continuum QCD, Nucl. Phys. **B228**, 173 (1983).

[35] G. G. Batrouni, H. Kawai, and P. Rossi, Coordinate independent formulation of the Langevin equation, J. Math. Phys. (N.Y.) **27**, 1646 (1986).