

Gibbs sampling of complex-valued distributions

L. L. Salcedo*

Departamento de Física Atómica, Molecular y Nuclear and Instituto Carlos I de Física Teórica y Computacional, Universidad de Granada, E-18071 Granada, Spain

(Received 2 November 2015; published 11 October 2016)

A new technique is explored for the Monte Carlo sampling of complex-valued distributions. The method is based on a heat bath approach where the conditional probability is replaced by a positive representation of it on the complex plane. Efficient ways to construct such representations are also introduced. The performance of the algorithm is tested on small and large lattices with a $\lambda\phi^4$ theory with quadratic nearest-neighbor complex coupling. The method works for moderate complex couplings, reproducing reweighting and complex Langevin results and fulfilling various Schwinger-Dyson relations.

DOI: 10.1103/PhysRevD.94.074503

I. INTRODUCTION

In many physical problems, including statistical mechanics and field theory on the lattice, one has to deal with a large number of variables. Simple estimates show that, when confronted with a generic integral of high dimensionality, the Monte Carlo method is often the most efficient strategy [1]. In a typical application the configurations x of the physical system have a probability distribution $P(x)$ in the form of a Boltzmann weight, $P(x) = e^{-S(x)}$, where $S(x)$ is the action (or the Hamiltonian) of the configuration, and expectation values of the observables $A(x)$ come as averages weighted with $P(x)$. The standard approach is then to produce a sample of the distribution $P(x)$, and obtain an estimate of the expectation value of the observables from the arithmetic mean of that sample. Sampling a distribution means to produce points x with a frequency that in average equals $P(x)$. In this sense, when the weights of the distribution are negative or complex a direct sampling becomes meaningless. Such complex weights do appear in many instances, e.g., in the presence of fermions [2], in quantum chromodynamics with a baryon number chemical potential [3] or in real time path integral formulations of quantum mechanics (as opposed to Euclidean time ones). The impossibility of a direct sampling of the distribution of interest in those cases constitutes the well known sign (or phase) problem. To face this situation a number of ideas, specific or generic, have been proposed in the literature (see for instance Refs. [4–22]), yet, at present there is no efficient approach to deal with the problem of sampling generic complex distributions in a systematic way. Certainly nothing as universally valid and efficient as, e.g., the Metropolis algorithm is known for the complex case.

Since a straight sampling of a complex distribution is not possible some oblique approach is needed. Let us mention just two techniques. One is to sample a positive auxiliary

probability distribution $P_0(x)$, introducing a compensating weight $P(x)/P_0(x)$ in the observables. This is the reweighting technique. This method is very general, but it suffers from the well known overlap problem for large systems [10,23].

The other possibility we mention is based on relaxing the concept of sampling from configurations to observables. Instead of producing sequences of suitable distributed points (configurations) and then computing the observables for them, it is sufficient to have an algorithm producing, for each observable, a stochastic sequence of values in such a way that the arithmetic mean of those sequences reproduces (in average) the correct expectation values. The key point here is that the random values assigned to the observables need not, and in general will not, correspond to actual configurations of the system. Since there is a lot of freedom to do this (for each observable, many different distributions can be devised having the same expectation value) a practical way to proceed is by generating *complex* configurations (regarded as basic observables), for which the other observables are computed. This is the representation technique which is the main focus of this work.

The representation technique relies on constructing a suitable *real and positive* distribution, $\rho(z)$, defined on the complex plane \mathbb{C}^d , *representing* the original complex probability $P(x)$ defined on \mathbb{R}^d . This means that for any observable A , the expectation value of $A(x)$ with $P(x)$ equals the expectation value of $A(z)$ with $\rho(z)$, where $A(z)$ stands for the analytic continuation of $A(x)$ from the real to the complex manifold. Standard importance sampling is then applied to $\rho(z)$.

Note that in the representation approach $\langle 1 \rangle = 1$ automatically. This is in contrast to reweighting. There, barring the rare cases in which the normalizations of P and P_0 are known, the normalization ensuring $\langle 1 \rangle = 1$ has to be enforced *a posteriori*, using the very Monte Carlo calculation.

The idea of trading the complex distribution $P(x)$ by a positive representation of it is implicit in the complex

*salcedo@ugr.es

Langevin approach [4,24]. Unfortunately the complex Langevin approach is not the definitive answer to the complex sampling problem. For one reason, it is limited to actions $S(x)$ having an analytical extension on the complex plane, and for another, the algorithm is not always convergent, or even worse, sometimes it converges to wrong equilibrium solutions, that is, to $\rho(z)$ which are representations of complex distributions different from the target distribution $P(x)$ [7,25–29].

As alternatives to the complex Langevin approach, Söderberg first considered representations not directly based on a complex Langevin algorithm [30]. An explicit formulation and discussion of the concept of representations by themselves was presented in [9]. There, explicit representations were constructed for many complex probabilities, in particular, Gaussian times polynomial of any degree and any number of dimensions and arbitrary complex distributions with support at a single point (these are the sum of derivatives of Dirac deltas at the same point). Necessary and sufficient conditions for the existence of positive representations in \mathbb{R}^n were given in [31], as well as the proof of existence for $n = 1$. In [32] it was shown that essentially any complex probability on \mathbb{R}^n admits a positive representation, as do complex probabilities defined on arbitrary compact Lie groups (this covers the case of periodic distributions) and explicit constructive methods were presented. Recent work along the same lines of representations not based on complex Langevin can be found in [33,34].

Despite the results in [9,32] the practical problem of representing and sampling complex distributions is not solved for two reasons related to locality and uniqueness.

As is well known the Monte Carlo approach is more efficient than other approaches when the number of variables (degrees of freedom) involved, i.e., the number of dimensions of the configuration manifold, is large. However a good performance of the method requires the action $S(x)$ (and by extension, the probability) to be *local*. By local it is meant that the action is the sum of terms, each of them depending on a few variables and each variable appearing only in a few such terms. The requirement of locality is often essential to have a numerically efficient update procedure of the configurations in Monte Carlo calculations. For example, the coupling between too many variables is the reason why perfect actions [35] are not numerically favored in practice.

Therefore a key question to carry out a Monte Carlo sampling of a complex probability by means of a representation is whether $\rho(z)$ is local or not. The great virtue of the complex Langevin approach is that the algorithm retains the locality of the original complex action. On the contrary, the constructions of representations found in [9,32] are nonlocal even if the complex action is local.

The second problem is that of uniqueness. The representation of a given $P(x)$ is not unique. Many different $\rho(z)$

produce the same expectation values on the set of holomorphic observables [9]. Most such representations are useless since they go deeply into the complex plane, where analytically continued observables behave wildly and variances become large. This implies that in the representation technique (including complex Langevin) there is an analog of the overlap issue existing in the reweighting approach; a representation may be formally correct yet produce unacceptable fluctuations in the Monte Carlo estimates.

In view of these impediments to the direct construction of a representation for a given complex distribution $P(x)$ in the many-dimensional case, here we explore a heat bath approach. The update is carried out sequentially for each of the variables, keeping the remaining variables fixed to their current value. To do the update of a variable, we replace its conditional probability, a complex function, by a positive representation of it on the complex plane. The point of following this procedure is that, being the conditional probability a function of a single variable, obtaining a representation for it is relatively easy and inexpensive. The locality issue is bypassed since, if the complex action is local, the representation of the conditional probability will also depend locally on the remaining variables. The quality of the representation, regarding variances, can also be controlled more easily in the one-dimensional case.

Nevertheless, in the complex case we are not protected by the standard convergence theorems for Markov chains based on positive probabilities [1], and this can lead to difficulties, as already found in the complex Langevin approach. That problems may arise in a complex heat bath approach can be understood from the following consideration. Assume $P(x_1, x_2)$ is the complex probability to be sampled. One would be safe by constructing the positive two complex-dimensional representation $\rho(z_1, z_2)$. However, if a Gibbs approach is used, one needs to represent instead the conditional probability $P(x_1|z_2) = P(x_1, z_2)/P(z_2)$, where the marginal probability $P(z_2) = \int dx_1 P(x_1, z_2)$ is required at complex values of z_2 . The trouble is that, even if $P(x_2)$ is never zero on the real axis, it can have zeros on the complex plane.

In any case, in our view, presently lacking reliable and general sampling methods of complex probabilities, it seems worthwhile to explore and test new approaches to assess their performance. This work is organized as follows. In Sec. II we discuss the concept of representation (Sec. II A) and show that the complexness of $P(x)$ puts restrictions on how localized the representation can be on the complex plane, i.e., on the quality of the possible representations (Sec. II B). Also, in Sec. II C constructive representation techniques are presented, where the quality of the representations can be optimized, both in one dimension and in higher dimensions. Section III is devoted to discuss and analyze the complex heat bath approach. The method is introduced in Sec. III A and it is applied to a

quadratic action on a hypercubic lattice in Sec. III B. A deeper study of the performance of the algorithm is presented in Sec. III C. There a $\lambda\phi^4$ theory with complex nearest-neighbor complex coupling is analyzed. The Monte Carlo results of the complex heat bath algorithm are compared to those obtained from reweighting and complex Langevin. Conclusions are presented in Sec. IV.

II. COMPLEX PROBABILITIES AND REPRESENTATIONS

A. Concept of representation of a complex probability

We will call *complex probability* to any normalizable complex distribution $P(x)$. After normalization, it may take negative or complex values instead of being real and positive for all x . The expectation values of observables $A(x)$ are obtained as usual through

$$\langle A \rangle_P = \frac{\int d\mu(x) P(x) A(x)}{\int d\mu(x) P(x)}. \quad (2.1)$$

Here $d\mu(x)$ is the appropriate positive measure in the x -manifold. In what follows we often assume $x \in \mathbb{R}^n$, or a periodic version of \mathbb{R}^n , and $d\mu(x) = d^n x$.

A *representation* of P is an ordinary probability $\rho(z)$ (i.e., a normalizable, real and positive distribution) defined on the complexified manifold, \mathbb{C}^n , and such that it produces the same expectation values as P upon analytical extension of the observable, that is,

$$\langle A(x) \rangle_P = \langle A(z) \rangle_\rho, \quad (2.2)$$

where $A(z)$ refers to the analytical extension of $A(x)$ from \mathbb{R}^n to \mathbb{C}^n , and

$$\langle A \rangle_\rho = \frac{\int d^{2n} z \rho(z) A(z)}{\int d^{2n} z \rho(z)}. \quad (2.3)$$

We will often refer to the real manifold as the *real axis* of \mathbb{C}^n . In the periodic case $\rho(z)$ is normalized as (assuming a 2π period in each direction)

$$\int_{[0, 2\pi]^n} d^n x \int_{\mathbb{R}^n} d^n y \rho(x + iy) = 1, \quad (2.4)$$

so the complexified manifold is noncompact in the imaginary direction.

Strictly speaking, Eq. (2.2) expresses that ρ is a representation of P valid for the observable A . Generically we will say that ρ is a representation of P when it is valid for a large set of sufficiently well behaved observables or test functions. One can specify the set of test functions A as that of polynomials of z , or that of plane waves, $\exp(-ikz)$, $k \in \mathbb{R}^n$. Since the plane waves grow exponentially in the

complex plane the latter set is more restrictive in general, nevertheless, we show below that often the support of ρ can be chosen to remain within a finite strip along the real axis. This avoids any problem related to the exponential growth of $A(z)$. In the case of periodic distributions the set of Fourier modes $\exp(-ikz)$, $k \in \mathbb{Z}^n$ is a natural choice.

If the support of a representation fills \mathbb{C}^n , observables with singular points cannot be reproduced by it. However, representations are constructed below with support smaller than \mathbb{C}^n . In that case some observables with singularities on the complex plane can also be reproduced. This is guaranteed when the singularities lie outside some simply connected region containing both the support of the representation and the real axis. A one-dimensional example is $P(x) = \exp(-(x-i)^2/2)/\sqrt{2\pi}$, which can be represented by $\rho(x, y) = \delta(y-1) \exp(-x^2/2)/\sqrt{2\pi}$. This representation correctly reproduces $A(x) = 1/(x-z_0)$ provided $\text{Im}z_0$ is either negative or larger than one.

B. Conditions on the support of the representations

The application of Monte Carlo sampling to a complex distribution, by means of a representation of it, differs in an essential aspect from the standard case of positive distributions. In the standard case, one often makes a direct sampling of the distribution. Rarely a reweighting method is used since this tends to increase the variance in the expectation values (for generic observables). In the complex case one has to somehow construct the representation and such representation is by no means unique. Any target complex probability $P(x)$ admits many valid representations, all of them formally correct but vastly different as regards to performance. The nonuniqueness implies in particular that the expectation value of nonholomorphic observables, $B(z, z^*)$, can be different for different representations. An estimate of the variance, e.g., $B = |A|^2 - |\langle A \rangle|^2$ [where $A(z)$ is holomorphic], follows this rule and so, while $\langle A \rangle$ does not depend on the representation, its Monte Carlo estimate does.

For instance, if $\rho(z)$ is a representation, it is easy to show that the new distribution $\rho'(z)$ obtained by convolution of ρ with any positive distribution of the type $C(|z|^2)$ (i.e., rotationally symmetric) provides a new representation [9].¹ This new representation will be wider, more spatially extended, than the previous one, and this is an undesirable feature. In general one will want a representation $\rho(z)$ as localized and close to the real axis as possible. The reason is that most observables will grow, often exponentially, as one departs from the real axis, and as a consequence the statistical fluctuations also grow, rendering a Monte Carlo approach less efficient.

¹Many more new representations can be obtained by adding a *null representation*, ρ_0 , i.e., one with vanishing expectation value for any holomorphic observable, however, there is no guarantee that the sum will be positive definite.

The overlap problem of the reweighting technique for positive distributions exists, to some extent, also in the representation technique. In principle, the best representation is the most localized one, in the sense that variances of observables will be smaller. Incidentally, one sensible way to measure the localization is through the (relative) entropy,²

$$S(\rho) = \langle -\log(\rho/\rho_{\text{ref}}) \rangle_{\rho}, \quad (2.5)$$

rather than the variance matrix of z . To see this consider a $\rho(z)$ composed of two well-separated narrow Gaussian functions. The quantity $\langle |z - \langle z \rangle|^2 \rangle_{\rho}$ measures the typical separation between points in the support of $\rho(z)$ and this will be large if the two Gaussian functions are far from each other. On the other hand the entropy is independent of the separation (as long as the two Gaussians have a negligible overlap). One can obtain the expectation values of observables for each of the Gaussians with small variance (since they are narrow) and then combine the result with a final small error. As said, proximity to the real axis is convenient too, however, for a valid representation this cannot be controlled since it is dictated by $P(x)$ as we argue below.

From the point of view of the localization, the best choice would be $\rho(z) = P(x)\delta(y)$, where y refers to the imaginary axis coordinate. Unfortunately such representation is not positive for complex P . In fact, there is a kind of uncertainty principle implying that the less positive P is, the wider ρ should be on the complex plane. The width here refers to the extension of the support of ρ in the imaginary direction.

To make this principle more precise let us consider the one-dimensional case, $x \in \mathbb{R}$, although similar considerations hold for any number of dimensions. Let P and ρ be normalized, and

$$\tilde{P}(k) := \int dx e^{-ikx} P(x) = \langle e^{-ikx} \rangle_{\rho} \quad (2.6)$$

then, using $z = x + iy$,

$$|\tilde{P}(k)| = |\langle e^{-ikx} \rangle_{\rho}| = |\langle e^{-ikz} \rangle_{\rho}| \leq |\langle |e^{-ikz}| \rangle_{\rho}| = \langle e^{ky} \rangle_{\rho}. \quad (2.7)$$

Therefore, if the support of ρ lies in the region $y \leq Y_1$, it follows that $\langle e^{ky} \rangle_{\rho} \leq e^{kY_1}$ for all positive k , and hence $|\tilde{P}(k)| \leq e^{kY_1} \forall k > 0$. Likewise, if the support of ρ lies in $y \geq Y_2$, necessarily $|\tilde{P}(k)| \leq e^{kY_2}$ for all negative k . It follows that

$$\begin{aligned} Y_1 \geq Y_+ &\equiv \max_{k>0} \left(\frac{1}{k} \log |\tilde{P}(k)| \right), \\ Y_2 \leq Y_- &\equiv \min_{k<0} \left(\frac{1}{k} \log |\tilde{P}(k)| \right). \end{aligned} \quad (2.8)$$

²In Eq. (2.5), ρ_{ref} is some fixed reference distribution. ρ and ρ_{ref} are positive and normalized.

In other words, any representation ρ must have some support in the region $y \geq Y_+$ as well as in the region $y \leq Y_-$. If the support of ρ falls in a strip $Y_2 \leq y \leq Y_1$ the width of the strip is restricted by the conditions $Y_1 \geq Y_+$ and $Y_2 \leq Y_-$. These formulas extend immediately to the case of periodic distributions.

As an example, consider the one-dimensional action

$$S(x) = \frac{\beta}{4} x^4 + iqx, \quad \beta > 0. \quad (2.9)$$

For $\beta = 0.5$ and $q = 2$ one finds $Y_+ = 4.29$. Thus a proper representation of this complex action must have some support above $y = 4.29$. As it turns out, the complex Langevin algorithm applied to this action with q positive will produce an equilibrium distribution entirely located *below* the real axis. This is because once the random walk goes below the real axis it can never get above it.³ Therefore we know, without doing the actual calculation, that the complex Langevin algorithm converges to the wrong equilibrium distribution in this case. Actually, this action is not at all pathological and it admits a valid representation of the two-branch type described below, in Sec. II C 3.

These bounds can be easily generalized to other observables and any number of dimensions as follows: Let $P(x)$ be a complex probability in \mathbb{R}^n and $\rho(z)$ a representation of P , and let $A(z)$ be a test function. Then

$$|\langle A(x) \rangle_{\rho}| = |\langle A(z) \rangle_{\rho}| \leq \max\{|A(z)|, z \in \text{supp}(\rho)\}. \quad (2.10)$$

This inequality puts conditions on the support of ρ . Indeed, given an observable A , its expectation value a does not depend on the choice of ρ . Then one can define the set \mathcal{A} of points where $|A(z)| \geq |a|$. The inequality implies that the support of any proper representation of P must have some overlap with \mathcal{A} , and this for all test functions A .

For instance, again for the action of Eq. (2.9) with $\beta = 0.5$ and $q = 2$, one finds that $\langle 1/(x-i) \rangle = -2.82i$. Thus, the set \mathcal{A} is the disk of radius 0.35 centered at $z = i$. Since the lower half plane has no overlap with this disk, any representation without some support above the real axis can automatically be ruled out.⁴

³We always refer to the standard implementation of the complex Langevin method which has only horizontal noise and with no kernel.

⁴A tricky point is that the argument works because the singularity at $z = i$ would lie beyond the support of any such representation and so for them $1/(x-i)$ would be an acceptable test function. The two-branch representations (discussed below) for this action have support above the singularity and so for them $1/(x-i)$ is not an acceptable test function, and in fact, their support has no overlap with the disk $|z-i| < 0.35$.

C. Construction of explicit representations

Barring those of [9,32], the representations of complex probabilities existing in the literature are limited to a few cases: quadratic actions, which can be obtained analytically, complex Langevin constructions, which have only an empirical basis, and some representations constructed by first choosing a suitable $\rho(z)$ and then finding to which complex probability it corresponds. This can be done, e.g., by means of the projection [7]

$$\begin{aligned}\langle A(z) \rangle_\rho &= \int d^n x d^n y \rho(x, y) A(x + iy) \\ &= \int d^n x \left(\int d^n y \rho(x - iy, y) \right) A(x) \\ &=: \int d^n x P(x) A(x) = \langle A(x) \rangle_P.\end{aligned}\quad (2.11)$$

In this section we discuss some new explicit constructions of representations for generic complex probabilities. Further constructions, including the case of complex probabilities defined on a compact Lie group can be found in [32].

Note that the need of an analytical extension of $P(x)$ itself (in addition to that of the observables) is not a general requirement to have a representation. It is an idiosyncrasy of the complex Langevin algorithm [which actually requires a holomorphic $\log P(x)$] and of some other approaches (e.g., the one-branch representations, and the complex Gibbs sampling discussed below) while other methods apply to generic normalizable complex distributions.

1. An explicit representation in one dimension

The one-dimensional distribution

$$Q(x) = \delta(x) + \delta'(x) \quad (2.12)$$

admits the representation

$$q(z) = \frac{1}{8\pi} \left| 1 - \frac{z}{2} \right|^2 e^{-|z|^2/4} \quad (2.13)$$

as is readily verified by checking that $\langle x^n \rangle_Q = \langle z^n \rangle_q$ for all non-negative integers n .

This basic representation can then be used to construct representations of arbitrary complex probabilities in any number of dimensions [32]. Here we show this for a one-dimensional $P(x)$ which is assumed to be normalized. Let us decompose $P(x)$ as

$$\begin{aligned}P(x) &= P_0(x) + F'(x), \\ P_0(x) &> 0, \quad \int dx P_0(x) = 1.\end{aligned}\quad (2.14)$$

$P_0(x)$ can be chosen in many ways and this choice fixes $F(x)$: Since both P and P_0 are normalized, the function

$$F(x) = \int_{-\infty}^x (P(x') - P_0(x')) dx' \quad (2.15)$$

vanishes for large x . This F can be written as $F(x) = P_0(x)h(x)$ where $h(x)$ will be a complex function in general. In this case, a representation of $P(x)$ is provided by

$$\begin{aligned}\rho(z) &= \int dx d^2 z' P_0(x) q(z') \delta(z - x - h(x)z') \\ &= \int dx P_0(x) \frac{q((z-x)/h(x))}{|h(x)|^2}.\end{aligned}\quad (2.16)$$

[$\delta(z - z_0)$ refers to the two-dimensional delta on the complex plane.] To verify that this $\rho(z)$ is really a normalized representation of $P(x)$ let us apply it to a generic observable [using the first form in Eq. (2.16)]

$$\langle A \rangle_\rho = \int dx d^2 z' P_0(x) q(z') A(x + h(x)z'). \quad (2.17)$$

Now, because $q(z)$ is a representation of $Q(x)$ and $A(z)$ is analytical, it follows that

$$\begin{aligned}\langle A \rangle_\rho &= \int dx dx' P_0(x) Q(x') A(x + h(x)x') \\ &= \int dx P_0(x) (A(x) - h(x)A'(x)) \\ &= \int dx P(x) A(x) \\ &= \langle A(x) \rangle_P\end{aligned}\quad (2.18)$$

[where $A'(x)$ denotes the derivative of $A(x)$].

The formula in Eq. (2.17) already indicates how to carry out a sampling of the complex probability $P(x)$, namely, sample x with $P_0(x)$ and z with $q(z)$ and average the values of the observable computed at $x + h(x)z$.

2. One-branch representations in one dimension

The Monte Carlo method suggested in Eq. (2.17) can be extended to any number of dimensions, however, the determination of the function $\vec{h}(x)$ (which is far from unique) is not so straightforward as in the one-dimensional case [32]. Even in the one-dimensional case the construction presented above has the problem that the support of $\rho(z)$ will be spatially more extended than necessary. This problem is common to constructive approaches of generic type. Essentially, the procedure to obtain Eq. (2.16) has been to rewrite $P(x)$ in the form

$$\begin{aligned}
P(x) &= P_0(x) + (P_0(x)h(x))' \\
&= \int dx' P_0(x') (\delta(x' - x) - h(x')\delta'(x' - x)) \quad (2.19)
\end{aligned}$$

and proceed to replace $\delta(x' - x) - h(x')\delta'(x' - x)$ by a representation of it. The width of the representation of $\delta(x) - h\delta'(x)$ on the complex plane increases with $|h|$, with a coefficient which depends on the phase of h . In the worst cases, $Y_{\pm} = \pm|h|$, for $h = \pm i|h|$, respectively [Y_{\pm} were introduced in Eq. (2.8)].

As a rule, the approaches such as that in Eq. (2.19), based on writing the target complex probability P as

$$\begin{aligned}
P(x) &= \sum_n w_n p_n(x), \\
\rho(z) &= \sum_n w_n \rho_n(z), \quad (2.20)
\end{aligned}$$

where $p_n(x)$ is some fixed basis of normalized complex probabilities with known representations $\rho_n(z)$ and the w_n are some positive weights, allow for an immediate representation of $P(x)$ but the width of $\rho(z)$ along the real axis will be fixed by that of the $\rho_n(z)$ (in fact, by the worst case), so in general $\rho(z)$ will be wider than necessary.

As it turns out, better representations, with width closer to the bounds Y_{\pm} discussed above, can be obtained by adapting the representation to the target complex probability $P(x)$. An obvious example is a probability of the type

$$P(x) = P_0(x - a), \quad (2.21)$$

where $P_0(x)$ is positive (for real x) and the constant $a = a_R + ia_I$ may be complex. In this case $\rho(z) = P_0(x - a_R)\delta(y - a_I)$ ($z = x + iy$) is a valid representation, and in fact the best one. The support lies on a straight line parallel to the real axis. We will refer to representations with support on a single line as one-branch representations.⁵

Not all complex probabilities admit a one-branch representation. For $P(x)$ defined on \mathbb{R} , this requires the existence of a path $z(t)$ on the complex plane connecting $x = \pm\infty$, thus ensuring that integration along the path is equivalent to integration along the real axis, such that $P(z)dz$ stays positive, ensuring a positive weight [30]. In this case $P(x)$ can be sampled by using a standard inversion method, namely, for $u \sim U(0, 1)$, the value of z is determined by the condition $u = \int_{-\infty}^z P(z')dz'$. For periodic $P(x)$, the path must also be periodic. An example is shown in Fig. 1 for $P(\phi) = Ne^{-ae^{i\phi} - be^{-i\phi}}$ with $a = 0.5$ and $b = 0.15i$.

⁵Of course, if $P(x)$ happens to have the form in Eq. (2.21) the representation $\rho(z) = P_0(x - a_R)\delta(y - a_I)$ works in any number of dimensions.

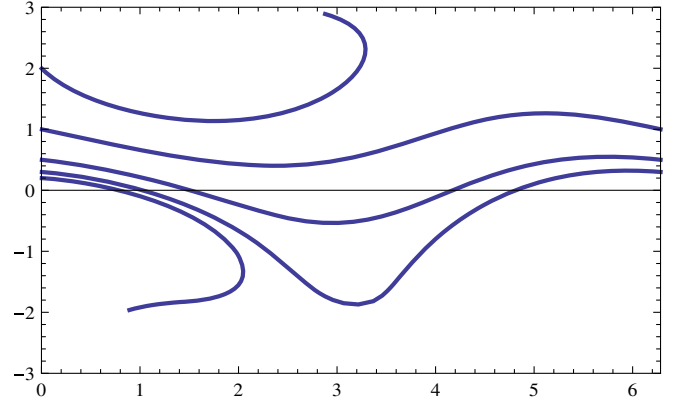


FIG. 1. One-branch paths on the complex plane ϕ , parametrized by $0 \leq u \leq 1$, for the periodic complex probability with action $S(\phi) = ae^{i\phi} + be^{-i\phi}$, with $a = 0.5$ and $b = 0.15i$. From bottom to top, the paths start at $\phi = 0.2i, 0.3i, 0.5i, i, 2i$. The three middle paths are periodic and so any of them provides a one-branch representation of the complex probability.

Although one-branch representations do not always exist for a given $P(x)$ a possible device is to split $P(x)$ as $P = P_1 + P_2$ with P_1 chosen in such a way that a representation is known for it and $P_2 = P - P_1$ admits a one-branch representation. The idea here is that a common P_1 can be used for a family of P , and only the concrete one-branch representation of P_2 has to be constructed in each case, e.g., by integration of $dz/du = 1/P_2(z)$.

3. Two-branch representations in one dimension

Representations with support on two lines, called here two-branch representations, exist quite generally for periodic and nonperiodic complex probabilities. The ones we consider here are of the type

$$\begin{aligned}
\rho(z) &= Q_1(x)\delta(y - Y_1) + Q_2(x)\delta(y - Y_2), \\
Q_{1,2}(x) &\geq 0 \quad Y_1 \geq Y_2, \quad (2.22)
\end{aligned}$$

so the support is along the lines $y = Y_1$ and $y = Y_2$ parallel to the real axis. Here $Q_1(x)$ and $Q_2(x)$ are two suitable positive functions. For definiteness we have chosen $Y_1 \geq Y_2$. Note that requiring this $\rho(z)$ to be a representation of $P(x)$ is equivalent to imposing the relation

$$P(x) = Q_1(x - iY_1) + Q_2(x - iY_2) \quad (2.23)$$

upon analytical extension of functions $Q_{1,2}(z)$ which are positive on the real axis. Indeed,

$$\begin{aligned}
&\int d^2z \rho(z) A(z) \\
&= \int dx (Q_1(x)A(x + iY_1) + Q_2(x)A(x + iY_2)) \\
&= \int dx (Q_1(x - iY_1) + Q_2(x - iY_2))A(x). \quad (2.24)
\end{aligned}$$

In fact for two given distinct $Y_{1,2}$, the *real* functions $Q_{1,2}$ are unique in the noncompact case. For periodic probability distributions the only ambiguity is an additive constant which can be moved between the two functions.

Let us first analyze the noncompact case, that is, $P(x)$ defined on \mathbb{R} . We further assume $P(x)$ to be normalized. Taking a Fourier transform

$$\begin{aligned}\tilde{P}(k) &= \int dx P(x) e^{-ikx} \\ &= \int d^2z \rho(z) e^{-ikz} \\ &= \tilde{Q}_1(k) e^{kY_1} + \tilde{Q}_2(k) e^{kY_2}.\end{aligned}\quad (2.25)$$

Imposing now the condition that $Q_{1,2}(x)$ are real, and hence $\tilde{Q}_{1,2}^*(k) = \tilde{Q}_{1,2}(-k)$, one obtains

$$\begin{aligned}\tilde{Q}_1(k) &= \frac{e^{-kY_2} \tilde{P}(k) - e^{kY_2} \tilde{P}^*(-k)}{2 \sinh(k(Y_1 - Y_2))}, \\ \tilde{Q}_2(k) &= \frac{e^{-kY_1} \tilde{P}(k) - e^{kY_1} \tilde{P}^*(-k)}{2 \sinh(k(Y_2 - Y_1))}.\end{aligned}\quad (2.26)$$

The functions $\tilde{Q}_{1,2}(k)$ are well behaved at $k=0$ since $\tilde{P}(0) = 1$ (any real normalization would do as well) and

$$\begin{aligned}\tilde{Q}_1(0) &= \frac{\text{Im}\langle x \rangle - Y_2}{Y_1 - Y_2}, \\ \tilde{Q}_2(0) &= \frac{\text{Im}\langle x \rangle - Y_1}{Y_2 - Y_1}.\end{aligned}\quad (2.27)$$

So a real solution exists and is unique. An obvious necessary condition for $Q_{1,2}(x)$ to be non-negative is $\tilde{Q}_{1,2}(0) \geq 0$, hence

$$Y_2 \leq \text{Im}\langle x \rangle \leq Y_1.\quad (2.28)$$

Once again this relation shows that a complex $P(x)$ requires a representation with a minimum width around the real axis.⁶

The formulas of $\tilde{Q}_{1,2}(k)$ can be reexpressed as convolutions in x -space,

$$\begin{aligned}Q_1(x) &= -\frac{1}{2(Y_1 - Y_2)} \text{Im} \left(\tanh \left(\frac{\pi(x + iY_2)}{2(Y_1 - Y_2)} \right) * P(x) \right), \\ Q_2(x) &= \frac{1}{2(Y_1 - Y_2)} \text{Im} \left(\tanh \left(\frac{\pi(x + iY_1)}{2(Y_1 - Y_2)} \right) * P(x) \right).\end{aligned}\quad (2.29)$$

⁶This relation is just the condition $Y_1 \geq \log(|\tilde{P}(k)|)/k$ for small positive k , and similarly for Y_2 , so it is weaker than Eq. (2.8).

(These formulas rely on the assumption $Y_1 > Y_2$.) Note that, despite the presence of \tanh , the functions $Q_{1,2}(x)$ vanish for large x thanks to the condition $\int dx \text{Im} P(x) = 0$.

The construction guarantees that the functions $Q_{1,2}(x)$ are real. Although lacking a detailed proof (the proof exists for the compact case, below), one empirically finds that $Y_{1,2}$ can be chosen so that these functions become non-negative. This happens for Y_1 and $-Y_2$ sufficiently large. Obviously, from our discussion in Sec. II B, it follows that this requires $Y_1 \geq Y_+$ and $Y_2 \leq Y_-$ [and hence also the weaker condition in Eq. (2.28)].

We have not observed a significant improvement by optimizing $Y_{1,2}$ separately, so for simplicity we will adopt the symmetric choice

$$Y \equiv Y_1 = -Y_2 > 0.\quad (2.30)$$

Except in the trivial case of P positive, Q_1 and/or Q_2 have negative regions for too small values of Y . What is found is that as Y grows, the minima of Q_1 and Q_2 grow as well until Y reaches a critical value. From then on $Q_{1,2}(x)$ are non-negative and the minima jump to $x = \pm\infty$ where these functions vanish. So the critical value of Y is such that $Q_{1,2} \geq 0$ and Q_1 or Q_2 vanish at some finite point.

From the numerical point of view the critical value of Y is the optimal one, since, as a rule, the closer the support of ρ to the real axis the smaller the variance in the observables. This rule holds in all approaches based on representing complex probabilities, including complex Langevin. In the two-branch case, probabilities which require relatively large values of Y can be considered as numerically hard, while those admitting small values are soft.

Examples of two-branch representations are displayed in Fig. 2. One such example is the complex action $S(x) = \beta x^2/2$ with $\beta = 1 + i$ using $Y = 0.7$. Another is $P(x) = \delta(x) + \delta'(x)$, using the optimal value $Y = 1.58$. Of course, in this latter case the convolutions in Eq. (2.29) can be obtained analytically. The size of the representation $q(z)$ in Eq. (2.13) can be estimated from $\langle |z|^2 \rangle_q = 6$. This is larger than that of the two-branch one $\langle |z|^2 \rangle_\rho \approx 5$. Actually in the two-branch case y is fixed to known values, either Y or $-Y$, with equal weight, so this variable does not add to the variance. In this view, recalling our discussion on the entropy in Sec. II B, the true uncertainty (understood as uncontrolled fluctuation) would be better estimated from $\langle x^2 \rangle_{Q_1} \approx 2.5$. This is particularly clear in the present example due to $Q_1 = Q_2$, because $\langle A \rangle = \langle A(x + iY) + A(x - iY) \rangle_{Q_1}$ exactly, with no added fluctuation from y .

Let us turn now to the case of a one-dimensional *periodic* complex probability $P(x)$. We assume P to be normalized and with period 2π , $\int_0^{2\pi} P(x) dx = 1$. In order to construct a two-branch representation, Eqs. (2.22) and (2.23) apply. Decomposing in Fourier modes

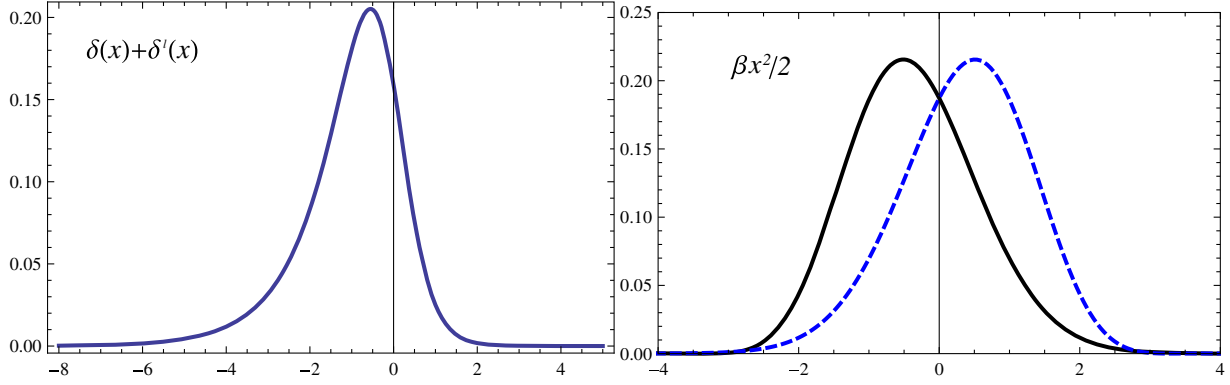


FIG. 2. Left: Function $Q_1(x) = Q_2(x)$ for $P(x) = \delta(x) + \delta'(x)$, using the optimal value $Y = 1.58$. Right: Functions $Q_1(x)$ (solid line) and $Q_2(x) = Q_1(-x)$ (dashed line) for the two-branch representation of the action $S(x) = \beta x^2/2$ with $\beta = 1 + i$ using the optimal value $Y = 0.7$.

$$P(x) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} e^{ikx} \tilde{P}_k, \quad \tilde{P}_0 = 1, \quad (2.31)$$

and requiring $Q_{1,2}(x)$ to be real periodic functions, one obtains

$$\begin{aligned} \tilde{Q}_{1,k} &= \frac{e^{-kY_2} \tilde{P}_k - e^{kY_2} \tilde{P}_{-k}^*}{2 \sinh(k(Y_1 - Y_2))}, \\ \tilde{Q}_{2,k} &= \frac{e^{-kY_1} \tilde{P}_k - e^{kY_1} \tilde{P}_{-k}^*}{2 \sinh(k(Y_2 - Y_1))}, \quad k \neq 0. \end{aligned} \quad (2.32)$$

In the periodic case the zero modes of $\tilde{Q}_{1,2}$ are not determined by $P(x)$, instead one has only the conditions

$$\tilde{Q}_{1,0} + \tilde{Q}_{2,0} = 1, \quad \tilde{Q}_{1,0}, \tilde{Q}_{2,0} \geq 0. \quad (2.33)$$

Let $\hat{Q}_{1,2}(x)$ denote the functions $Q_{1,2}(x)$ (reconstructed from their Fourier modes) without including the zero mode, i.e., $\hat{Q}_r = Q_r - \tilde{Q}_{r,0}/(2\pi)$, $r = 1, 2$. It is readily shown that whenever the following conditions are met

$$\begin{aligned} \min_x \hat{Q}_1 &\geq -1, & \min_x \hat{Q}_2 &\geq -1, \\ \min_x \hat{Q}_1 + \min_x \hat{Q}_2 &\geq -1, \end{aligned} \quad (2.34)$$

(by choosing $Y_{1,2}$ sufficiently large) suitable $\tilde{Q}_{1,0}, \tilde{Q}_{2,0}$ can be added so that $Q_{1,2}(x)$ are non-negative, thus providing a two-branch representation.

The solutions $Q_{1,2}(x)$ can also be expressed as convolutions (for simplicity we take the symmetric case $Y_1 = -Y_2 = Y$)

$$\begin{aligned} Q_r(x) &= \frac{1}{2\pi} \tilde{Q}_{r,0} \\ &+ \int \frac{d\phi}{2\pi} (C(\phi, Y) P_R(x - \phi) - \sigma_r S(\phi, Y) P_I(x - \phi)), \\ r = 1, 2, \quad \sigma_{1,2} &\equiv \pm 1. \end{aligned} \quad (2.35)$$

Here $P_{R,I}(x)$ stand for the real and imaginary parts of $P(x)$, and

$$\begin{aligned} C(x, Y) &\equiv \sum_{n=1}^{\infty} \frac{\cos(nx)}{\cosh(nY)}, \\ S(x, Y) &\equiv \sum_{n=1}^{\infty} \frac{\sin(nx)}{\sinh(nY)}. \end{aligned} \quad (2.36)$$

Clearly as Y increases the functions C and S go uniformly to zero. This ensures that eventually the conditions (2.34) on $\hat{Q}_{1,2}(x)$ are met so that $Q_{1,2}(x) \geq 0$. The functions $C(x, Y)$ and $S(x, Y)$ are displayed in Fig. 3.

Examples of two-branch representations for the periodic case are shown in Fig. 4 for the action

$$S(x) = \beta \cos(x) + imx, \quad \beta \in \mathbb{C}, \quad m \in \mathbb{Z}, \quad (2.37)$$

using optimal values of Y . The numerical solutions have been obtained from the Fourier modes which are Bessel functions.

The case $m = 0$ and β imaginary is particularly interesting, since this is one of the few cases (the other being a quadratic action, or a real action) in which the equilibrium Fokker-Planck equation of the complex Langevin algorithm can be solved analytically in $\rho(z)$. Unfortunately, that solution turns out to be wrong, giving $\langle e^{-ikz} \rangle_{\text{CL}} = \delta_{k,0}$, instead of the correct Fourier modes of P .

The existence of two-branch representations for very general complex probabilities proves in particular that representations exist which are valid for observables which

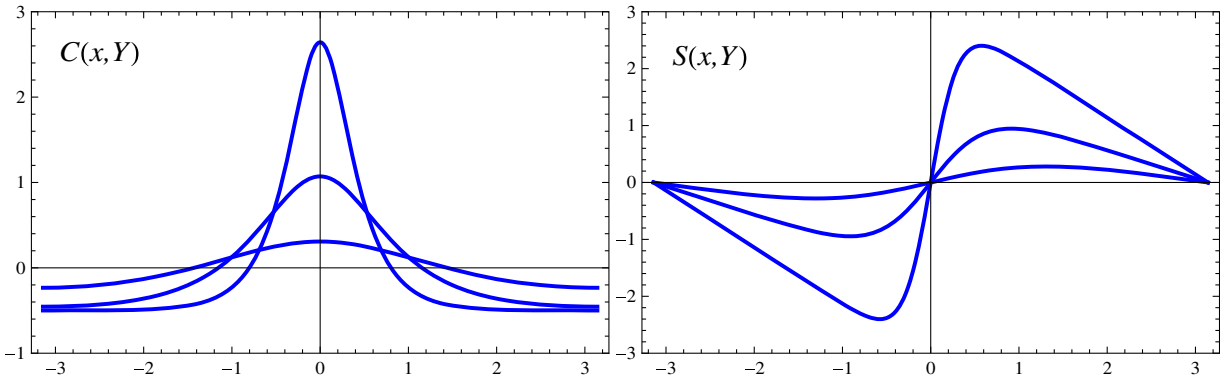


FIG. 3. Functions $C(x, Y)$ and $S(x, Y)$ for $Y = 0.5$ (larger amplitude), 1 and 2 (smaller amplitude).

are entire functions, no matter how wildly behaved they are at infinity in the imaginary direction, provided they are convergent at infinity in the real direction (in the non-compact case). On the other hand, representations that fill the complex plane would be problematic for those wild observables; since the integral of $A(z)\rho(z)$ would not be absolutely convergent the variance would diverge, rendering a Monte Carlo approach useless.

Another observation is that the use of more than two branches does not seem to introduce any improvement regarding localization of the support of the representation.

4. Representations in higher dimensions

The construction of representations by means of convolutions, either in the compact or noncompact cases, expresses that

$$Q_r = C_r * P + C_r^* * P^*, \quad r = 1, 2 \quad (2.38)$$

involving two known universal functions $C_{1,2}(x, Y_1, Y_2)$. Unfortunately, this simple scheme does not immediately extend to higher dimensions, although representations based on the Fourier modes exist in the periodic case [32]. Counting degrees of freedom one can expect that regardless the number of dimensions, a complex

distribution can always be traded by two real (and hopefully positive) distributions, however unless the construction of $Q_{1,2}$ is local, that is, updating one coordinate depends on only a few other coordinates, the method will not be useful in the many dimensional case, that is, when a Monte Carlo approach is needed.

Nevertheless, it is of interest to discuss the construction of optimal representations in higher dimensions. We consider the two-dimensional case, as the ideas involved can be extrapolated to the general case. Also we take the periodic case which is simpler.

Let $P(x_1, x_2)$ be a periodic two-dimensional normalized complex probability,

$$\iint_0^{2\pi} P(x_1, x_2) dx_1 dx_2 = 1. \quad (2.39)$$

One way to proceed is by reducing the dimension. Let $P(x_2)$ be the marginal distribution of x_2 , and $P(x_1|x_2)$ the conditional probability of x_1

$$P(x_2) = \int_0^{2\pi} P(x_1, x_2) dx_1, \quad P(x_1|x_2) = \frac{P(x_1, x_2)}{P(x_2)}. \quad (2.40)$$

Now, let $\rho(z_2)$ be a representation of $P(x_2)$ and $\rho(z_1|z_2)$ a representation of $P(x_1|z_2)$ regarded as a function of x_1 ,

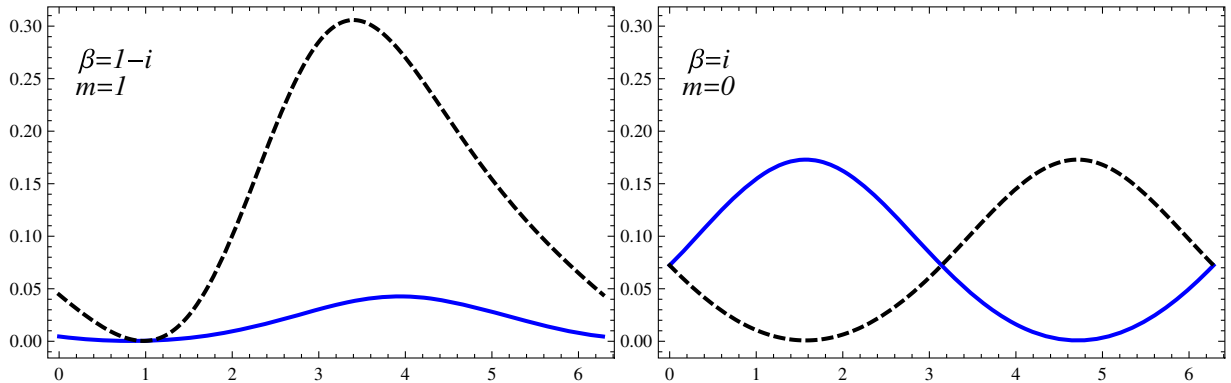


FIG. 4. Functions $Q_1(x)$ (solid line) and $Q_2(x)$ (dashed line) for the two-branch representation of the periodic action $S(x) = \beta \cos(x) + imx$. Left: $\beta = 1 - i$ and $m = 1$, with $Y = 1.17$. Right: $\beta = i$, $m = 0$, with $Y = 0.93$.

and where $P(x_1|z_2)$ refers to the analytical extension of $P(x_1|x_2)$. These are regular one-dimensional distributions that we know how to sample. Then $\rho(z_1, z_2) \equiv \rho(z_1|z_2)\rho(z_2)$ is a representation of $P(x_1, x_2)$. Indeed, for any observable $A(x_1, x_2)$

$$\begin{aligned} \langle A \rangle_\rho &= \int A(z_1, z_2) \rho(z_1|z_2) \rho(z_2) d^2 z_1 d^2 z_2 \\ &= \int A(x_1, z_2) P(x_1|z_2) \rho(z_2) dx_1 d^2 z_2 \\ &= \int A(x_1, x_2) P(x_1|x_2) P(x_2) dx_1 dx_2 \\ &= \langle A \rangle_P. \end{aligned} \quad (2.41)$$

One drawback with the approach just presented is that the analytical extension of $P(x_1, x_2)$ is needed, and also that $P(z_2)$ might have zeros on the complex plane. These problems can be solved by means of the following trick. Write P as

$$P(x_1, x_2) = \frac{1}{2\pi} (P(x_2) - P'(x_2)) + P'(x_1, x_2), \quad (2.42)$$

where $P'(x_2)$ is any positive distribution with normalization less than one. In this case $P(x_2) - P'(x_2)$ is one dimensional, normalizable and easily representable. Also, the marginal probability of the remainder $P'(x_1, x_2)$ is just $P'(x_2)$. Since this marginal probability is already positive one can apply the previous method to $P'(x_1, x_2)$ and a representation of $P'(x_2)$ is not needed. Regarding the choice of the auxiliary probability $P'(x_2)$, we note that a nice (i.e., localized) representation of $P'(x_1|x_2)$ favors a $P'(x_2)$ as large as possible, while the representation of $P(x_2) - P'(x_2)$ favors a small $P'(x_2)$, so a compromise has to be taken.

Another approach mimics the two-branch method discussed for the one-dimensional case. The natural proposal is (already taking a symmetric choice, this is unessential)

$$P(\vec{x}) = \sum_{r=1,2} Q_r(\vec{x} - i\sigma_r \vec{Y}), \quad (2.43)$$

where $Q_r(\vec{x})$ are positive functions, $\vec{Y} = (Y_1, Y_2)$ and $\sigma_r = \pm 1$ for $r = 1, 2$. Introducing Fourier modes

$$P(\vec{x}) = \frac{1}{(2\pi)^2} \sum_{\vec{k}} e^{i\vec{k}\cdot\vec{x}} \tilde{P}_{\vec{k}}, \quad (2.44)$$

one readily obtains the solution

$$\tilde{Q}_{r,\vec{k}} = \sigma_r \frac{e^{\sigma_r \vec{k}\cdot\vec{Y}} \tilde{P}_{\vec{k}} - e^{-\sigma_r \vec{k}\cdot\vec{Y}} \tilde{P}_{-\vec{k}}^*}{2 \sinh(2\vec{k}\cdot\vec{Y})}. \quad (2.45)$$

An obstruction arises here for the modes $\vec{k}\cdot\vec{Y} = 0$. For them the formula is only consistent if $\tilde{P}_{\vec{k}} = \tilde{P}_{-\vec{k}}^*$ and in this case $\tilde{Q}_{r,\vec{k}} = \tilde{P}_{\vec{k}}/2$.

There is no obstruction when \vec{Y} is chosen so that $\vec{k}\cdot\vec{Y} \neq 0$ unless $\vec{k} = \vec{0}$ (similarly to the one-dimensional case, $\vec{k} = \vec{0}$ poses no problem due to $\tilde{P}_{\vec{0}} = 1$). For example, consider the complex distribution

$$\begin{aligned} P(\vec{x}) &= Ng(x_1)g(x_2)g(x_1 - x_2), \\ g(x) &\equiv 1 + 2a \cos(x), \quad a \in \mathbb{C}. \end{aligned} \quad (2.46)$$

The only relevant Fourier modes are $k_{1,2} = 0, \pm 1, \pm 2$, so the choice $\vec{Y} = (Y, 3Y)$ guarantees that $\vec{k}\cdot\vec{Y} = 0$ only for $\vec{k} = \vec{0}$. The complex probability proposed in Eq. (2.46) can be represented with positive $Q_{1,2}(\vec{x})$ by taking a sufficiently large value of $Y > 0$. This is displayed in Fig. 5 for $a = i$. There we have taken $\tilde{Q}_{r,\vec{0}} = \frac{1}{2}$ for $r = 1, 2$, and automatically $Q_2(\vec{x}) = Q_1(-\vec{x})$. Similar results are obtained for $Y_2/Y_1 = \sqrt{2}$, which obviously also guarantees $\vec{k}\cdot\vec{Y} \neq 0$. It is interesting that this probability distribution would be beyond a complex Langevin approach, as $P(z_1, z_2)$ has zeros on \mathbb{C}^2 .

When many Fourier modes are involved it is not possible to prevent small values of $\vec{k}\cdot\vec{Y}$ and a more direct solution has to be adopted. Without loss of generality, let us assume that our choice is $\vec{Y} = (Y, 0)$ hence the problematic modes are those with $k_1 = 0$. This choice implies that only x_1 is moved to the complex plane,

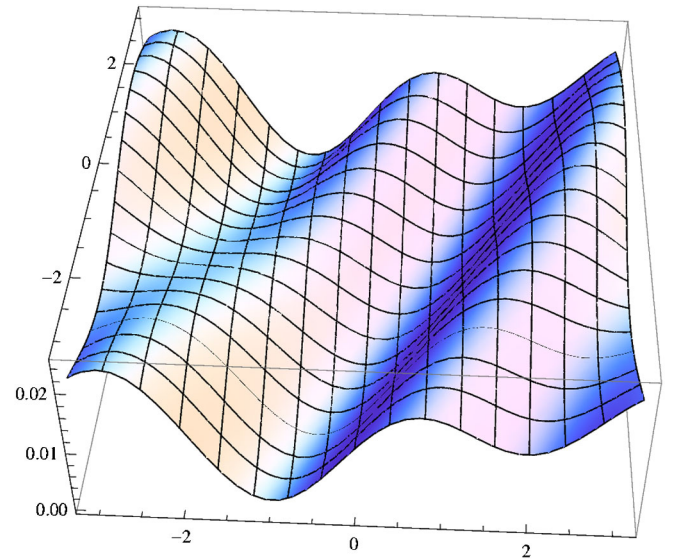


FIG. 5. Representation of $P \propto g(x_1)g(x_2)g(x_1 - x_2)$ for $g(x) = 1 + 2a \cos(x)$ with $a = i$, using $Y_2 = 3Y_1$. The function $Q_1(x_1, x_2)$ is displayed on $[-\pi, \pi] \times [-\pi, \pi]$ for $Y_2 = 4.43$. This is the optimal value, that is, $\min Q_1 = 0$.

$$P(x_1, x_2) = \sum_{r=1,2} Q_r(x_1 - i\sigma_r Y, x_2). \quad (2.47)$$

Clearly, this equation is only consistent when the marginal probability $P(x_2)$ is positive. If this is not the case, the solution is to use the trick described above, namely, to use Eq. (2.42) choosing a positive $P'(x_2)$. Now the two-dimensional version of the two-branch method works for $P'(x_1, x_2)$, and the one-dimensional version works for $P(x_2) - P'(x_2)$. Using this technique for the example in Eq. (2.46) with $a = i$, a value $Y = 2$ suffices for the two-dimensional representation and $Y = 2.45$ for the one-dimensional one. Variances are also smaller than using $\bar{Y} = (Y, 3Y)$.

As is well known, the real Langevin approach to sampling ordinary probability distributions is affected by the segregation problem: the random walk cannot cross the submanifold $P(x) = 0$. When $P(x)$ is real but with negative and positive regions, the same problem is inherited by the complex version of the algorithm, the random walk gets trapped in one of the positive or negative connected regions [36]. Such a problem does not exist here. For $a = 1$, the probability distribution $P(x_1, x_2)$ of Eq. (2.46) is real and changes sign whenever the argument of any of the factors $g(x)$ takes the values $x = \pm \arccos(-1/2)$. In Fig. 6 a representation of this $P(x_1, x_2)$ is presented, using $Y_2 = 3Y_1 = 2.17$.

III. COMPLEX HEAT BATH

As already noted the Monte Carlo method is required when the number of degrees of freedom is large. Further, the sampling algorithm must be local, in the sense defined in the Introduction, in order for the implementation not to be prohibitively expensive. A local implementation can be done for particular actions or Hamiltonians which admit

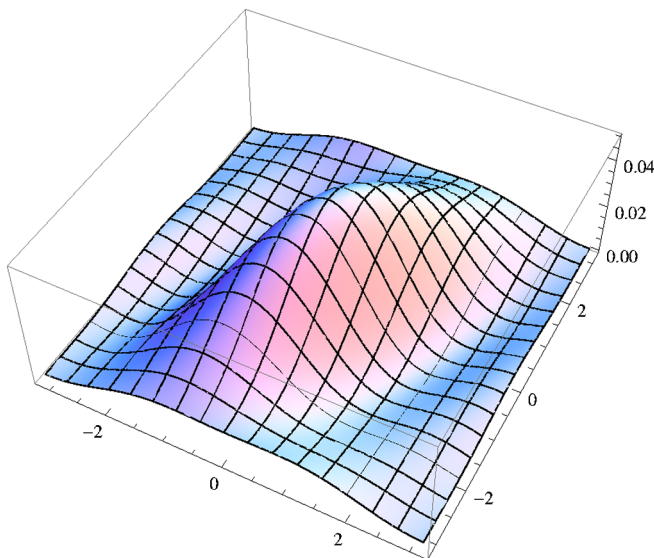


FIG. 6. Same as Fig. 5 with $a = 1$ and $Y_2 = 3Y_1 = 2.17$.

some kind of short-cut allowing to turn the complex probability problem into a standard one. Another local approach for complex probabilities is that of reweighting, by sampling a local and positive auxiliary probability distribution. Often $|P(x)|$ is used for this purpose. As is well known, this approach works nicely for small systems but becomes inefficient for larger ones where the expectation value of the phase of P gets close to zero [6]. Among the approaches based on constructing a representation of $P(x)$, only the complex Langevin algorithm preserves locality. This makes the algorithm very attractive. Unfortunately, this method lacks a solid mathematical basis, and in fact it does not always produce correct results [7,28].

A. Complex Gibbs sampling approach

These limitations make it worthwhile to explore alternative approaches. The one we analyze here is to extend the standard heat bath approach to the case of complex probabilities. In the standard Gibbs sampling the configuration follows a Markovian walk and each variable is updated in turn using its conditional probability, keeping the other variables fixed in that step. In our complex version, the configuration follows a Markovian walk on the complex manifold. Each variable is updated in turn using a *representation* of its conditional probability, conditional to the values of the other variables, which in general will be complex. This relies on the analytical extension of the complex probability to the complex manifold, in the same way as in the complex Langevin algorithm.

Let us remark that the need of an analytical extension of $P(x)$ is a limitation of the complex Gibbs sampling approach. Representations exist for any (or very general) complex probabilities. Having one such representation, any standard sampling method could be used, although the issue of localization on the complex manifold would still persist. Unfortunately, such representations in high dimensions are not easy to obtain through any useful (local) approach. The virtue of a complex Gibbs sampling is that one does not need to obtain representations of multidimensional complex distributions, since the conditional probability depends just on the one variable to be updated.

Let us show that this procedure is formally correct. To make our point it is sufficient to consider just two variables. Let $P(x_1, x_2)$ be the complex probability in \mathbb{R}^2 , and let $\rho(z_1, z_2)$ be a representation of it, a positive distribution on \mathbb{C}^2 . We want to verify that the updated distribution is also a representation. If the variable z_1 is updated to a new value z'_1 , keeping z_2 fixed, the new distribution on \mathbb{C}^2 will be

$$\rho'(z'_1, z_2) = \rho_{\text{rep}}(z'_1|z_2)\rho(z_2). \quad (3.1)$$

Here $\rho(z_2)$ is the marginal distribution of z_2 and $\rho_{\text{rep}}(z'_1|z_2)$ is a representation of the conditional probability $P(x'_1|x_2)$ (regarded as a function of x'_1 only):

$$\begin{aligned}\rho(z_2) &= \int d^2 z_1 \rho(z_1, z_2), \\ P(x'_1 | z_2) &= \frac{P(x'_1, z_2)}{P(z_2)}.\end{aligned}\quad (3.2)$$

Likewise, $P(z_2) = \int dx_1 P(x_1, z_2)$ where $P(z_1, z_2)$ refers to the complex probability analytically extended to the complex plane. $P(z_2)$ can also be regarded as the analytical extension of the marginal probability $P(x_2)$.

In order to show that $\rho'(z'_1, z_2)$ is also a representation of P let us consider a generic (holomorphic) observable, $A(z_1, z_2)$, then

$$\begin{aligned}& \int d^2 z'_1 d^2 z_2 \rho'(z'_1, z_2) A(z'_1, z_2) \\ &= \int d^2 z'_1 d^2 z_2 \rho_{\text{rep}}(z'_1 | z_2) \rho(z_2) A(z'_1, z_2) \\ &= \int dx'_1 d^2 z_2 P(x'_1 | z_2) \rho(z_2) A(x'_1, z_2) \\ &= \int dx'_1 \int d^2 z_1 d^2 z_2 \frac{P(x'_1, z_2)}{P(z_2)} \rho(z_1, z_2) A(x'_1, z_2) \\ &= \int dx'_1 \int dx_1 dx_2 \frac{P(x'_1, x_2)}{P(x_2)} P(x_1, x_2) A(x'_1, x_2) \\ &= \int dx'_1 dx_2 P(x'_1, x_2) A(x'_1, x_2).\end{aligned}\quad (3.3)$$

That is, $\langle A \rangle_{\rho'} = \langle A \rangle_P$ and ρ' is a representation of P . In the second equality it has been used that $\rho_{\text{rep}}(z'_1 | z_2)$ is a representation of $P(x'_1 | z_2)$ with respect to z'_1 . In the fourth equality it has been used that $\rho(z_1, z_2)$ is a representation of $P(x_1, x_2)$ and the other factors depend analytically on z_1 and z_2 .

Note that Eq. (3.1) is not a standard heat bath sampling of $\rho(z_1, z_2)$: there z_1 would be updated using $\rho(z'_1 | z_2) = \rho(z'_1, z_2) / \rho(z_2)$ instead of $\rho_{\text{rep}}(z'_1 | z_2)$. In the standard sampling $\rho(z_1, z_2)$ is unchanged under updates whereas in the complex Gibbs sampling all one shows is that the new distribution is still a representation, but not necessarily the same as before the update.

Of course, a standard Gibbs sampling of $\rho(z)$ would be completely correct (and in fact preferable), but the problem is that $\rho(z)$ is not available, we only know how to construct representations of one-dimensional or (at any rate) low-dimensional complex distributions such as $P(x_i | \{z_{j \neq i}\})$, and hope that the Markovian chain converges to the correct result. We have shown that the property of being a representation is preserved, however, the powerful theorems that apply for Markovian chains of positive probability distributions are not guaranteed to work in the complex case. In this sense, the complex Gibbs sampling lacks a sound mathematical basis, as is also the case for other approaches, like the complex Langevin algorithm.

An specific way in which our complex heat bath sampling may find trouble is related to the need of analytical extension: even if the marginal probability $P(\{x_{j \neq i}\})$ is never zero on the real manifold, the actual algorithm depends on the marginal probability on the complex manifold through analytical extension, $P(\{z_{j \neq i}\})$. Zeros in this function are potentially problematic; proximity to a zero implies a highly nonpositive definite distribution and this requires going deeply into the complex plane.

B. Complex Gaussian action

Preliminary tests on simple distributions with few variables indicate that the approach may work, however, the situation might be different for large systems. In order to test the complex Gibbs sampling proposal in many-variable settings, we have studied a d -dimensional hypercubic lattice of size N in each direction. First we consider a quadratic action with nearest-neighbor complex coupling.

The partition function is

$$Z(\beta) = \int e^{-S[\phi]} \prod_x d\phi_x. \quad (3.4)$$

It is defined through integration over the $V = N^d$ variables ϕ_x taking real values, although during the Monte Carlo simulation the ϕ_x will become complex in general. The action is given by

$$S[\phi] = \sum_x \left(\phi_x^2 + \beta \phi_x \sum_{\mu=1}^d \phi_{x+\hat{\mu}} \right), \quad \beta \in \mathbb{C}, \quad (3.5)$$

and we adopt periodic boundary conditions. The action is complex by allowing β to be complex.

This case is simple enough to have analytic expressions of the expectation values of typical observables. Also, explicit representations of the V -dimensional distribution $P(x)$ can be obtained in the present case. For this action the implementation of the complex heat bath algorithm is straightforward. Indeed, the conditional probability of the variable ϕ_x is

$$\begin{aligned}P(\phi_x | \{\phi_{x' \neq x}\}) &\propto \exp(-\phi_x^2 - \beta \phi_x \hat{\phi}_x) \\ &\propto \exp\left(-\left(\phi_x + \frac{\beta}{2} \hat{\phi}_x\right)^2\right), \\ \hat{\phi}_x &\equiv \sum_{\mu=1}^d (\phi_{x+\hat{\mu}} + \phi_{x-\hat{\mu}}).\end{aligned}\quad (3.6)$$

Thus the update of the variable ϕ_x takes the simple form

$$\phi_x = \xi - \frac{\beta}{2} \hat{\phi}_x, \quad (3.7)$$

where the random variable ξ is distributed according to $e^{-\xi^2}$.

We have checked through numerical experiments that whenever this Monte Carlo simulation converges it does so to the correct expectation values. Clearly, $\text{Im}\beta$ does not affect $|P(\phi)|$, so the action is well behaved, meaning that the integrals involved are absolutely convergent, provided $\text{Re}\beta$ stays within the appropriate limits (for even N this is simply $|\text{Re}\beta| < 1/d$) regardless of the value of $\text{Im}\beta$. However, in our complex version of the Gibbs sampling, convergence takes place only for suitably bounded values of $\text{Im}\beta$. This is shown in Table I for a 2^4 lattice. The lack of convergence in otherwise well-posed problems reflects the fact that the standard Markov chain convergence theorems do not immediately apply to the complex case, and in particular for our complex Gibbs procedure.

C. Complex $\lambda\phi^4$ action

In order to further check the approach we have considered a nonlinear version of the previous action, by adding a $\lambda\phi^4$ term. Specifically, we study the action

$$S[\phi] = \sum_x \left(\phi_x^4 + \phi_x^2 + \beta\phi_x \sum_{\mu=1}^d \phi_{x+\hat{\mu}} \right),$$

$$\beta \in \mathbb{C}, \quad (3.8)$$

again with hypercubic geometry $V = N^d$ and periodic boundary conditions.

The action enjoys some obvious symmetries (translations, lattice isotropy and spatial reflection, as well as parity under $\phi_x \rightarrow -\phi_x$). In addition when N is an even number, a change $\beta \rightarrow -\beta$ can be compensated by the transformation $\phi_x \rightarrow \pm\phi_x$, with a plus/minus sign for even/odd sites. Thus for even values of N , $Z(\beta)$ is an even function of β . Also $Z(\beta)^* = Z(\beta^*)$, hence, when N is even Z is real for purely imaginary β .

As observables to be estimated by Monte Carlo we take the following ones:

$$O_1 = \frac{\beta}{V} \sum_x \phi_x \hat{\phi}_x, \quad O'_1 = \frac{1}{V} \sum_x \phi_x \tilde{\phi}_x,$$

$$O_2 = \frac{\beta}{2Vd} \sum_x \hat{\phi}_x^2, \quad O'_2 = \frac{1}{2Vd} \sum_x \tilde{\phi}_x \hat{\phi}_x, \quad (3.9)$$

where the auxiliary fields are defined as

$$\hat{\phi}_x \equiv \sum_{\mu=1}^d (\phi_{x+\hat{\mu}} + \phi_{x-\hat{\mu}}), \quad \tilde{\phi}_x \equiv 4\phi_x^3 + 2\phi_x. \quad (3.10)$$

TABLE I. For a 2^4 lattice, and for several values of $\text{Re}\beta$, maximum values of $|\text{Im}\beta|$ for which the complex bath algorithm converges, for the Gaussian action of Eq. (3.5).

$\text{Re}\beta$	0.0	0.1	0.2	0.3	0.4	0.5
$ \text{Im}\beta $	0.70	0.65	0.58	0.48	0.35	0.0

The observables O_1 and O_2 are quadratic while O'_1 and O'_2 are quartic, and so they have larger variance than the former ones.

1. Strong coupling expansion

In a strong coupling expansion in powers of β , $\log Z(\beta)$ comes as a sum over closed paths on the lattice, formed by consecutive links. Terms of order β^n correspond to paths of length n . For even N all closed paths have even length, so $Z(-\beta) = Z(\beta)$. For odd N , contractile closed paths have even length but homotopically nontrivial paths (winding through the periodic boundary conditions) can have an odd length, hence introducing odd powers of β in $\log Z(\beta)$. These start at order β^N .

Within the strong coupling expansion, the expectation values of O_1 and O_2 come as a series of powers of β with real coefficients. The leading order contributions are⁷

$$\langle O_1 \rangle = -2d\beta^2 \langle \phi^2 \rangle_0^2 + O(\beta^3),$$

$$\langle O_2 \rangle = \beta \langle \phi^2 \rangle_0 + O(\beta^2),$$

$$\langle \phi^2 \rangle_0 \equiv \langle \phi^2 \rangle_{\beta=0} = 0.234. \quad (3.11)$$

Actually, O_1 relates sites with opposite parity, so for even (or infinite) N its expectation value has only even powers of β . In this case $\langle O_1 \rangle$ will be real for purely imaginary β . For odd N , odd powers start at order β^N . These contributions come from noncontractile closed paths.⁸ Likewise, $\langle O_2 \rangle$ has only odd powers of β and it is imaginary for imaginary β , except for odd N , in which case even powers start at $O(\beta^{N-1})$. Inspection of $\langle O'_1 \rangle$ shows that in this regard, this quantity behaves as $\langle O_1 \rangle$, likewise $\langle O'_2 \rangle$ behaves as $\langle O_2 \rangle$. The leading contributions from topologically nontrivial paths are

$$\langle O_1 \rangle_{\text{top}} = -2d(-\beta)^N \langle \phi^2 \rangle_0^N + O(\beta^{N+2}),$$

$$\langle O_2 \rangle_{\text{top}} = -(-\beta)^{N-1} \langle \phi^2 \rangle_0^{N-1} + O(\beta^{N+1}). \quad (3.12)$$

These are the leading terms in $\text{Im}\langle O_1 \rangle$ and $\text{Re}\langle O_2 \rangle$ for imaginary β and odd N .

⁷This is for $N > 2$. For $N = 2$ the terms in Eq. (3.12) are of the same order and they have to be added to these.

⁸Of course, an identical conclusion follows from the relation $\langle O_1 \rangle = -2\beta/V \partial \log Z / \partial \beta$.

Since the nonlinear action has no exact solution, we have monitored the accuracy of the complex Gibbs sampling algorithm by using different checks. A simple one is the fulfillment of the reality conditions on $\langle O_1 \rangle$ and $\langle O_2 \rangle$ as well as consistency with the strong coupling results for small β .

2. Transfer matrix

Another check has been made by using a transfer matrix approach for $d = 1$ (for $d > 1$ this technique becomes prohibitive). In this approach $Z = \text{Tr}(\hat{T}^N)$ where the transfer matrix \hat{T} is an operator in $L^2(\mathbb{R})$ with kernel

$$\langle \phi' | \hat{T} | \phi \rangle = e^{-\frac{1}{2}(\phi'^4 + \phi^4 + \phi'^2 + \phi^2 + 2\beta\phi'\phi)}. \quad (3.13)$$

The expectation value of observables of the type $A = F(\phi_{x+1}, \phi_x)$, such as O_1 , O'_1 and O'_2 , can then be obtained as follows:

$$\begin{aligned} \langle A \rangle &= \frac{\sum_n \lambda_n^N \langle A \rangle_n}{\sum_n \lambda_n^N}, \\ \langle A \rangle_n &\equiv \lambda_n^{-1} \int d\phi' d\phi \bar{\psi}_n^*(\phi') \psi_n(\phi) \langle \phi' | \hat{T} | \phi \rangle F(\phi', \phi). \end{aligned} \quad (3.14)$$

Here $\langle \bar{\psi}_n |$ and $|\psi_n \rangle$ are left and right eigenvectors of \hat{T} with eigenvalue λ_n and normalized as $\langle \bar{\psi}_n | \psi_m \rangle = \delta_{nm}$. As it turns out, when β^2 is a real number, \hat{T} is a normal operator and it admits a real and orthonormal eigenbasis. The eigenvalues are real for real β . For imaginary β , half of the eigenvalues are real and the other half are imaginary.

Using this approach,⁹ we find perfect agreement with the complex heat bath results (whenever the latter method converges) for the cases studied, $\beta = 0.5i$, $0.7i$ and i , with $N = 20$. The complex heat bath algorithm crashes for $\beta = 2i$.

3. Virial relations

For higher dimensional lattices, a more useful check comes from the (generalized) virial relations, also referred to as Schwinger-Dyson equations or equations of motion. These exact relations state that for any (regular) observable $A[\phi]$, and any site x ,

$$\left\langle \frac{\partial A}{\partial \phi_x} \right\rangle = \left\langle A \frac{\partial S}{\partial \phi_x} \right\rangle. \quad (3.15)$$

⁹To do this we have discretized ϕ along the same lines explained below for constructing representations of the conditional probability.

The observables O_1 , O_2 , O'_1 and O'_2 arise naturally by choosing $A = \phi_x$ and $A = \phi_{x+\hat{\mu}}$. For our action, the virial relations imply that

$$\begin{aligned} \langle I_1 \rangle &= \langle I_2 \rangle = 0, \\ I_1 &\equiv O_1 + O'_1 - 1, \\ I_2 &\equiv O_2 + O'_2. \end{aligned} \quad (3.16)$$

4. Complex Gibbs sampling implementation

In addition, we compare the results obtained with the complex heat bath approach with those obtained with standard reweighting. We sample $e^{-\text{Re}(S)}$ using a Metropolis algorithm and include the phase in the observables. This technique is practical only for small lattices and/or small $\text{Im}\beta$ due to the overlap problem. Also, we compare with the complex Langevin equation results since this has become a rather standard practice in the context of complex probabilities.

In the three types of Monte Carlo calculations, reweighting (RW), complex Langevin (CL), and complex heat bath (CHB), the expectation values and their error are extracted from 20 independent runs with cold and hot starts. In RW and CHB 10^5 sweeps are applied, arranged in 100 batches of 1000 iterations each, to monitor the thermalization. In CL each run has duration 10^4 (in Langevin time units), also arranged in 100 batches of duration 10^2 each. The CL step size is controlled so that $\Delta t \leq 0.001$ and $|\Delta\phi_x| \leq 0.001$ [37,38]. In all three versions the ten first batches have been dropped as this was deemed sufficient to reach thermalization.

Since the CHB approach is new, some relevant details are in order. The conditional probability needed in the complex Gibbs sampling takes the following form:

$$P(\phi_x | \{\phi_{x' \neq x}\}) \propto \exp(-\phi_x^4 - \phi_x^2 - \beta\phi_x \hat{\phi}_x). \quad (3.17)$$

For complex β this distribution cannot be sampled directly and one has to resort to representations with two branches, as explained in Sec. II C 3. In order to study the performance of the method in this exploratory work, we have given priority to keeping things simple and under control, leaving improvements in technical details of the implementation to future work. The actual representation has been obtained by making use of Eq. (2.26) and relying on a fast Fourier transform algorithm, applied back and forth, to carry out the convolutions. To do this, $\text{Re}\phi$ is restricted to a box $[-L/2, L/2]$ and discretized with a step of size $h = L/2^K$. Typical values are $L = 20, 40$, and $K = 8, 10, 12$. We have checked, by comparing with a standard Metropolis calculation, that for real β no sizable error is introduced due to the finiteness of L and K .

A representation, i.e., a pair $Q_{1,2}(x)$, is obtained for each site to be updated. The minimal value of the parameter Y

[see Eq. (2.30)] is determined by taking steps ΔY starting at $|\text{Im}\langle x \rangle| + \Delta Y$. For $\Delta Y = 0.1$, this starting value is already sufficient in more than 99.99% of the cases. In average, 1.4 more steps are needed if $\Delta Y = 0.01$, with no noticeable gain in the accuracy of the expectation values.

During the Monte Carlo simulation one finds conditional probabilities (controlled by the value of the environmental parameter $\hat{\phi}_x$ there) with wildly different degrees of difficulty regarding its representation. For $|\text{Im}\beta| \leq 0.5$ the soft cases are overwhelmingly predominant and for these Y remains small. The hard cases are rare but unavoidable, since the parameter $\hat{\phi}_x$ can approach a zero of the normalization of the conditional probability. The closest zero is at $\hat{\phi}_x = 4.60i$. For these hard cases, Y can attain huge values which could spoil the simulation or make it crash. To prevent this we introduce a control parameter Y_s such that the site is not updated when a Y larger than Y_s would be required. Since the decision of skipping the site (on that sweep only) is taken *ex post facto*, some bias is introduced in the simulation.

For a too small value of Y_s (i.e., a hard cutoff) the bias shows as a violation of the virial relations $\langle I_{1,2} \rangle = 0$. For too large values, the variances increase and moreover the simulation may abort. For $\beta = 0.25i$ the value of Y_s is not crucial. For $\beta = 0.5i$, values of Y_s ranging from 2 to 20 allow to fulfill the virial relations with a controlled noise. For $\beta = 0.7i$, hard events become too frequent, making the whole approach inviable: a too small Y_s would be required, hence introducing an unacceptable bias in the results, at least for the observables I_1 and I_2 . The situation resembles that of an asymptotic expansion, e.g., $\sum_{n=0}^{\infty} (-1)^n n! x^n$: the series must be truncated because at some point the oscillatory terms start to grow. For small enough x (in our case β) many terms can be included and a good accuracy can be attained, but the exact unbiased result is never obtained, unless some resummation technique is applied.

Besides Y_s , a further regulator must be introduced. In principle the optimal Y is such that $\min_x \{Q_1(x), Q_2(x)\} = 0$. In the soft cases this poses no problem, however, for hard events such a strict requirement would

TABLE II. Expectation values (scaled by 10^3) of $O_{1,2}$ and $I_{1,2}$ for several settings. RW, CL and CHB refer to reweighting, complex Langevin and complex heat bath methods, respectively. Standard deviations of the means are indicated in parentheses; they affect the last digits shown. Each value is extracted from 20 runs with hot and cold starts and 10^5 sweeps (10^4 time units for CL) of which the last 90000 are used in the averages. Default parameters for CHB are $Y_s = 5$, $\Delta Y = 0.1$, and $\epsilon = 10^{-6}$. The mark * indicates that $Y_s = \infty$ there. The mark ** indicates that data lying beyond 8 standard deviations from the mean have been removed (see text).

β	N^d	L	K	$10^3 \times \langle O_1 \rangle$		$10^3 \times \langle O_2 \rangle$		$10^3 \times \langle I_1 \rangle$		$10^3 \times \langle I_2 \rangle$		Method
0.25i	3 ³			19.783(28)	-i1.134(33)	2.984(12)	+i56.017(19)	0.49(33)	-i0.07(9)	-0.22(10)	-i0.05(8)	RW
0.25i	3 ³			19.740(21)	-i1.120(99)	3.009(17)	+i55.978(51)	0.97(73)	-i0.09(8)	-0.10(30)	+i0.02(4)	CL
0.25i	3 ³	20	10	19.745(14)	-i1.099(28)	2.996(11)	+i55.967(16)	3.39(115)	+i0.09(23)	-0.12(9)	-i0.14(9)	CHB
0.25i	8 ³			19.985(137)	-i0.153(179)	-0.074(47)	+i55.958(44)	0.59(63)	+i1.16(68)	-0.55(53)	-i0.67(42)	RW
0.25i	8 ³			19.746(5)	-i0.021(20)	0.004(4)	+i55.977(12)	0.29(24)	-i0.02(2)	-0.03(7)	+i0.01(1)	CL
0.25i	8 ³	20	10	19.749(4)	+i0.001(8)	0.000(2)	+i55.969(5)	4.04(28)	+i0.04(5)	-0.02(3)	-i0.13(2)	CHB
0.5i	3 ³			71.642(116)	-i6.146(74)	8.487(49)	+i99.851(47)	-0.39(40)	+i0.53(28)	0.08(14)	-i0.15(21)	RW
0.5i	3 ³			71.628(76)	-i5.985(127)	8.567(49)	+i100.073(98)	0.16(75)	-i0.07(15)	0.42(22)	+i0.03(9)	CL
0.5i	3 ³	20	10	71.578(41)	-i6.173(49)	8.510(32)	+i99.882(32)	4.70(125)	+i2.78(135)	0.45(22)	+i0.21(29)	CHB
0.5i	8 ³			-3.0(37)	+i2.2(34)	1.8(38)	+i117.1(22)	-03.(18)	-i01.(17)	1.1(76)	+i118.(6)	RW
0.5i	8 ³			71.332(17)	+i0.076(32)	-0.019(18)	+i99.579(18)	0.24(18)	+i0.06(4)	0.08(6)	-i0.01(3)	CL
0.5i	8 ³	20	10	71.330(13)	-i0.003(13)	0.006(7)	+i99.568(10)	4.59(29)	-i0.01(20)	-0.01(5)	-i0.13(5)	CHB
0.5i	8 ³	20	10	71.352(11)	+i0.006(13)	-0.008(5)	+i99.570(8)	6.95(212)	-i1.22(264)	-0.06(12)	-i0.16(13)	CHB*
0.5i	8 ³	20	8	71.308(10)	+i0.013(12)	0.002(7)	+i99.571(7)	4.94(37)	-i0.11(20)	0.04(5)	-i0.07(5)	CHB
0.5i	8 ³	40	8	71.337(13)	+i0.011(14)	-0.001(7)	+i99.572(7)	1.40(66)	+i0.08(16)	0.04(6)	-i0.03(4)	CHB
0.5i	8 ³	40	8	71.337(13)	+i0.011(14)	0.000(7)	+i99.571(7)	0.10(14)	+i0.03(15)	0.02(4)	-i0.01(2)	CHB**
0.5i	8 ³	40	12	71.305(8)	-i0.014(15)	0.004(9)	+i99.545(9)	2.44(85)	+i0.17(19)	0.00(5)	-i0.01(5)	CHB
0.5i	8 ³	40	12	71.305(8)	-i0.014(15)	0.005(9)	+i99.544(9)	0.14(20)	+i0.02(13)	0.00(3)	+i0.03(3)	CHB**
0.5i	16 ³			71.303(7)	-i0.014(18)	-0.016(7)	+i99.569(7)	0.11(7)	+i0.02(2)	-0.03(3)	+i0.03(1)	CL
0.5i	16 ³	20	10	71.328(5)	-i0.005(6)	0.000(3)	+i99.567(4)	5.15(9)	-i0.09(7)	-0.01(2)	-i0.19(3)	CHB
0.5i	8 ⁴			90.652(9)	-i0.013(12)	0.000(4)	+i94.365(9)	0.06(9)	-i0.02(3)	0.00(2)	+i0.03(1)	CL
0.5i	8 ⁴	20	10	90.670(5)	+i0.005(8)	0.000(3)	+i94.369(5)	3.19(20)	-i0.16(16)	0.02(2)	+i0.14(3)	CHB

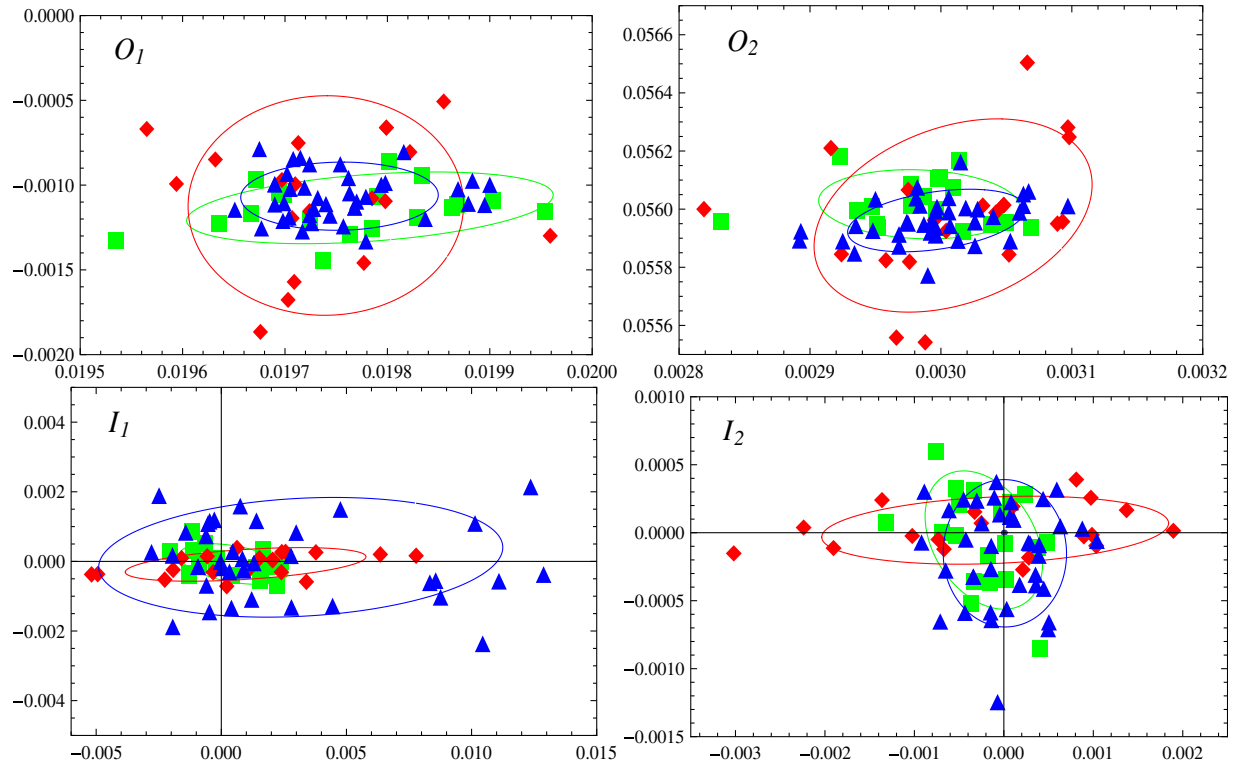


FIG. 7. Monte Carlo estimates of O_1 , O_2 , I_1 and I_2 for a 3^3 lattice with $\beta = 0.25i$, from RW (green squares), CL (red rhombuses) and CHB (blue triangles). Each point represents one of the 20 independent runs. To guide the eye ellipses are extracted from mean values and variance matrices of each cloud. They are scaled to enclose 68% of the Gaussian probability.

result in too large values of Y . In fact, for moderate values of Y the functions $Q_{1,2}(x)$ are positive where they are sizable, but may present small negative tails in the region of large x . Removal of these small tails is what forces Y to be large. To address this problem we relax the positivity requirement to $\min_x \{Q_1(x), Q_2(x)\} \geq -\epsilon$. Values $\epsilon \leq 10^{-8}$ are too small, but 10^{-7} or 10^{-6} already yield good results. In fact, even the larger value we have tried, $\epsilon = 10^{-3}$, turned out to be equally acceptable.

5. Monte Carlo estimates

Monte Carlo results for $\langle O_{1,2} \rangle$ and $\langle I_{1,2} \rangle$ for purely imaginary β , obtained through our complex heat bath (CHB) method, are displayed in Table II. Reweighting (RW) and complex Langevin (CL) results are shown for comparison. One can see that CHB and CL produce consistent results (with exception of $\text{Re}\langle I_1 \rangle$). RW also concurs on the same values, when our calculation using this method can be trusted (three first RW rows in Table II).

As illustration, results for a 3^3 lattice with coupling $\beta = 0.25i$ are shown in Fig. 7. For the four observables, the points represent the estimate obtained for each of the 20 independent Monte Carlo runs, for each of the three versions. To guide the eye, we have analyzed the cloud

of points assuming a Gaussian probability distribution. The ellipses are scaled to enclose 68% of that Gaussian probability.

For $\beta = 0.5i$ and 8^3 , the RW estimates presented in Table II are not only noisy but also wrong, most clearly for $\text{Im}\langle I_2 \rangle$ which ought to be zero. By incorrect we mean that the cloud of points are well separated from the correct results (see Fig. 8) and so the deviations cannot be merely attributed to the dispersion of the points. Remarkably, this implies that consistent (yet incorrect) results are obtained starting from quite different initial conditions (cold or hot), a condition which is often invoked as a check of convergence of the Markovian chain. That the convergence is actually metastable is signaled however by the lack of fulfillment of the virial relations. Since such relations, Eq. (3.15), are always available for continuous degrees of freedom, they prove to be a rather useful (necessary although not sufficient) test to check convergence.

We emphasize that the RW method itself has no bias, and the trouble comes entirely from using a too short Markovian chain (in our case 10^5 sweeps). If a sufficiently large number of sweeps were used, RW would yield correct estimates with some noise. The latter can only be reduced by using even more sweeps.

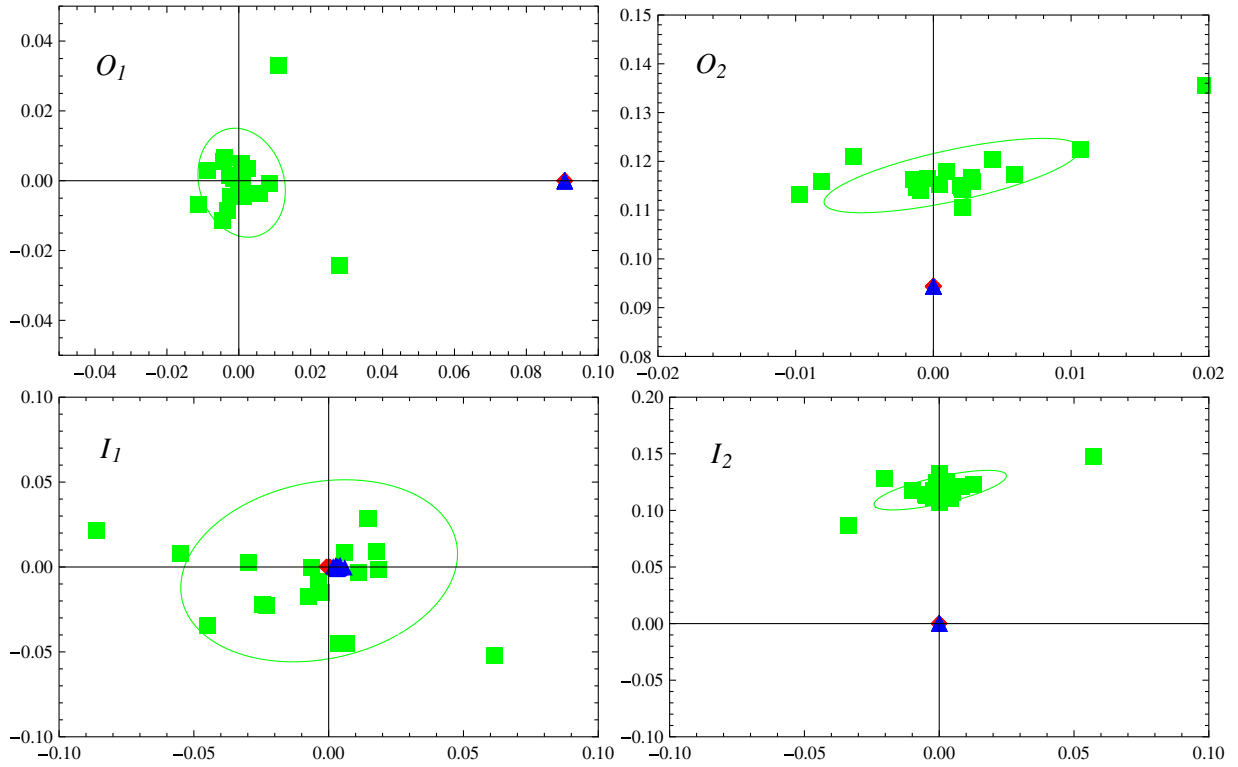


FIG. 8. Metastability of the RW calculation. Monte Carlo estimates of O_1 , O_2 , I_1 and I_2 for an 8^4 lattice with $\beta = 0.5i$, from RW (green squares), CL (red rhombuses) and CHB (blue triangles). For this lattice and coupling the RW estimate displays a net deviation (as well as enhanced dispersion). The incorrect RW estimates displayed actually reproduce the expectation values of the auxiliary action $\text{ReS}[\phi]$ (see text). Correct RW results would follow from using a sufficiently large number of sweeps.

Essentially¹⁰ the problem is that in RW, the sampling points (i.e., field configurations in our case) follow an auxiliary distribution $P_0(x)$ instead of the target distribution $P(x)$, and the weight proportional to P/P_0 is included in the observables:

$$\langle A \rangle_P = \frac{\langle wA \rangle_{P_0}}{\langle w \rangle_{P_0}}, \quad w(x) \equiv \frac{P(x)}{P_0(x)}. \quad (3.18)$$

In a proper implementation of RW, with sufficient (independent) sampling points, most points lie where P_0 is important, not P . The very few points which lie where P is important saturate all the weight in the averages, and the host of P -unimportant points has a negligible relative weight. This procedure provides an unbiased estimate but with much noise, since the actual number of important points employed in the averages is small. However, if not enough sampling points (sweeps) are used, the Markovian chain never gets to visit the region where P (but not P_0) is

¹⁰The discussion that follows is more literally suited to the case where the target distribution $P(x)$ is positive. When the weight is a phase, i.e., $|P/P_0| = 1$, one has to rely on cancellation of phases (i.e., destructive interference) where they change rapidly or noncancellation (constructive interference) where they change slowly, as in the stationary phase method.

important. In the absence of truly important points, the relative weight of the P_0 -distributed points is no longer small, and one obtains a wrong estimate which is closer to $\langle A \rangle_{P_0}$ than to $\langle A \rangle_P$ [assuming the variation of $w(x)$ can be neglected in the region where P_0 is important]. Actually, this expectation is verified by the results displayed in Table II for $\beta = 0.5i$ and 8^3 with 10^5 sweeps: For the action

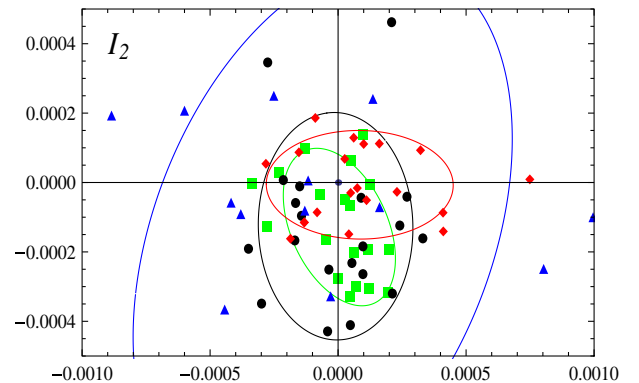


FIG. 9. Influence of the parameter Y_s . The points represent estimates of I_2 from 20 runs with 68% ellipses, for an 8^3 lattice with $\beta = 0.5i$: CHB with $Y_s = 2$ (green squares), CHB with $Y_s = 5$ (black disks), CHB with $Y_s = \infty$ (blue triangles), and CL (red rhombuses).

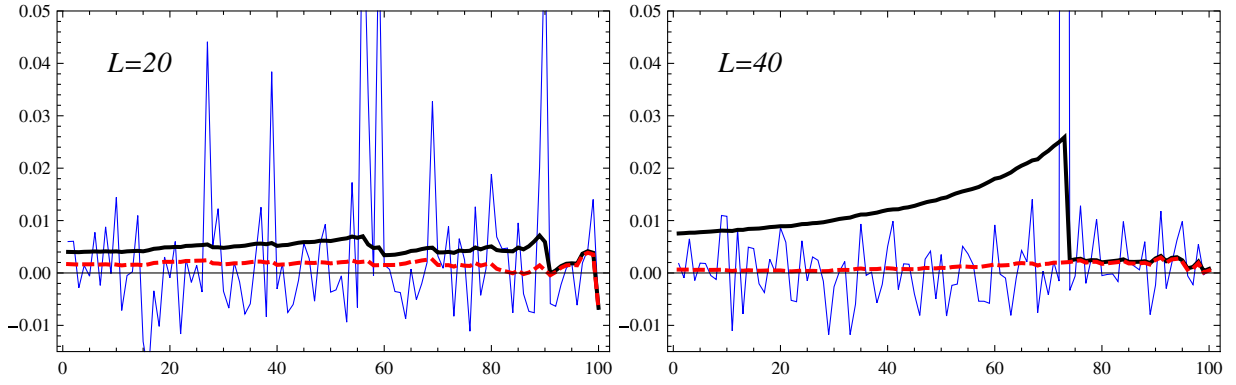


FIG. 10. Typical runs for $L = 20$ and $L = 40$ with $\beta = 0.5i$ in an 8^3 lattice. The 10^5 sweeps are arranged in 100 batches of 1000 sweeps each. The noisy thin solid (blue) line represents $\text{Re}\langle I_1 \rangle$ for each batch. The thick solid (black) line represents the average over all batches to the right (and so computed at later simulation times). The thick dashed (red) line shows the effect of removing batches which are off by 4σ ($L = 20$) or 8σ ($L = 40$). The high peak in “ $L = 40$ ” reaches 0.66.

in Eq. (3.8), P_0 corresponds to retaining only $\text{Re}\beta$ in the action. This is zero in our case and the action becomes ultralocal. An easy calculation then shows that

$$\begin{aligned} \langle O_1 \rangle_{P_0} &= \langle I_1 \rangle_{P_0} = 0, \\ \langle O_2 \rangle_{P_0} &= \langle I_2 \rangle_{P_0} = \beta \langle \phi^2 \rangle_0 \quad (\text{Re}\beta = 0). \end{aligned} \quad (3.19)$$

The value $\beta \langle \phi^2 \rangle_0 = i117 \times 10^{-3}$ is quite consistent with the RW results quoted in Table II.

This pathology, i.e., the increasing numerical effort required in RW to gain an overlap with the target distribution, worsens for larger lattices and/or couplings and prevents us from using RW in those cases. One way to improve the RW calculation would be to take a better

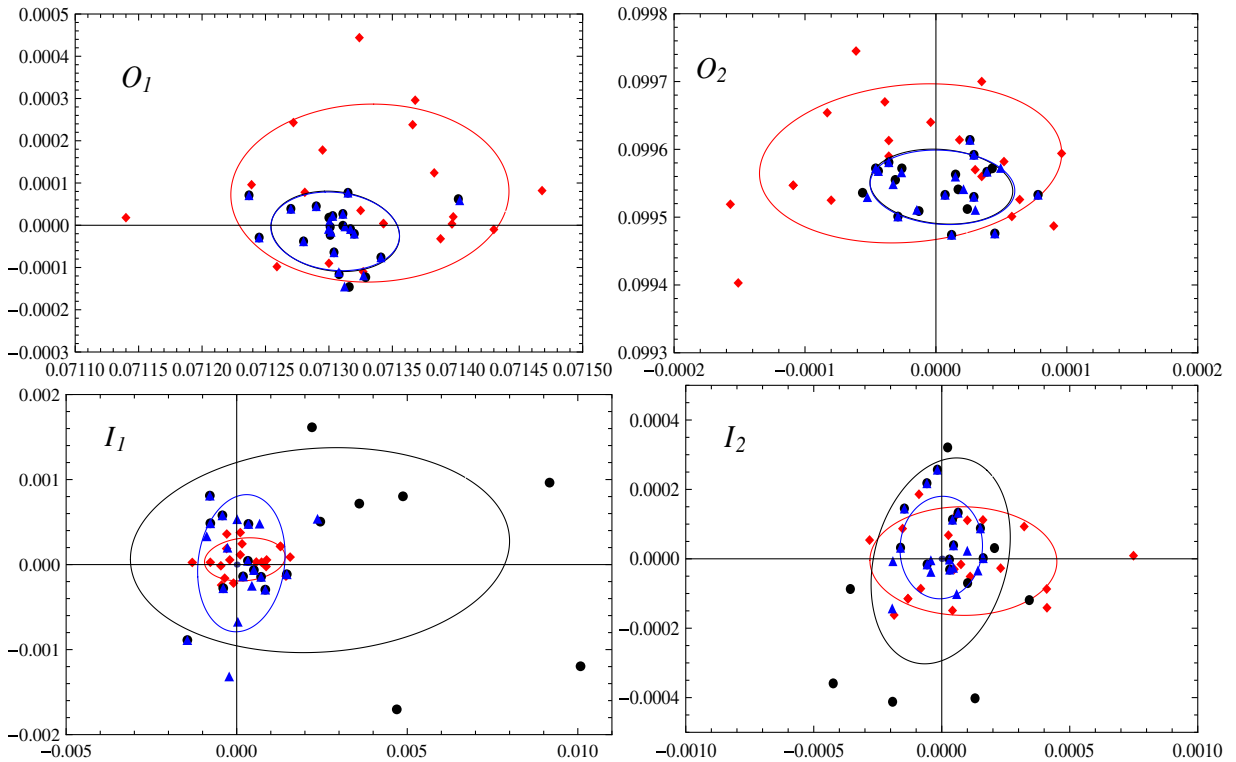


FIG. 11. Monte Carlo estimates of O_1 , O_2 , I_1 , and I_2 for an 8^3 lattice with $\beta = 0.5i$, using $L = 40$ and $K = 12$. CL is represented by red rhombuses, and CHB by black disks. The blue triangles are from CHB upon removal of batches (subsets of 1000 sweeps) lying beyond 8σ for $\text{Re}\langle I_1 \rangle$ (i.e., CHB** in Table II). As can be seen, the impact of the removal is small on O_1 and O_2 , while I_1 and I_2 are brought closer to zero.

auxiliary distribution $P_0(x)$, typically, by using effective values of the couplings of quartic and quadratic operators, and of β , paralleling the same approach used to study QCD at finite density [12].

Returning to our main focus, the CHB method, we observe that the quality of the results remains stable for larger lattices (16^3 and 8^4). Also the parameter K (we have tried $K = 8, 10, 12$) does not seem to be crucial. Likely, this is because the error introduced by a finite step h is exponentially suppressed for boundaryless integrations of smooth functions.

The CHB results for the observables O_1 and O_2 are quite stable against changes in the regulators and their standard deviations are small. On the other hand the fluctuations are larger in I_1 and I_2 . This follows from the fact that $O_{1,2}$ are quadratic while $O'_{1,2}$ involve quartic operators. The worst case displayed in Table II corresponds to removing the regulator Y_s , which results in relatively large fluctuations in $I_{1,2}$ (also, one of the 20 runs aborted). The influence of the value of Y_s on $\langle I_2 \rangle$ is shown in Fig. 9.

A conspicuous feature in CHB for $L = 20$ is the small but systematic violation of $\langle I_1 \rangle = 0$. The real part of this observable displays a bias at the level of 0.005 that persists for all lattice sizes.¹¹ Since the bias is largely reduced for $L = 40$ (to the level of 0.0014) it would seem that $L = 20$ is simply too small a box for the action of Eq. (3.8) with $|\beta| = 0.5$, however, no bias exists in similar simulations when β is real.¹² Closer inspection reveals that for $L = 20$ the various runs have little dispersion around the biased result. On the contrary, for $L = 40$ most runs are unbiased (at the level of 0.0001) but for a few of them $\text{Re}\langle I_1 \rangle$ drifts as far as 0.010. We show in Fig. 10 a typical run with $L = 20$ and one of the troubled runs with $L = 40$. Moderately hard events are more frequent for the smaller box. For $L = 40$, these events become exceptional and far more violent. Typically, for runs with $Y_s = 5$, the ratio of sites whose update is skipped drops from 3×10^{-7} for $L = 20$ to less than 10^{-8} for $L = 40$. As said, in our analysis the runs were arranged in 100 batches of 1000 sweeps each. For $L = 40$, removing batches in which $\text{Re}\langle I_1 \rangle$ is off by 8 standard deviations introduces a considerable improvement in the fulfillment of the virial relations without disturbing $\langle O_{1,2} \rangle$. This can be seen in Table II, results marked with CHB**, and also in Fig. 11, where we display the result of runs with and without removal of eccentric batches. While no significant difference is appreciated in $O_{1,2}$, the estimates of $I_{1,2}$ are brought closer to zero after the removal. A quite

noteworthy feature, illustrated in Fig. 10, is that the Markovian chain quickly recovers after a troubled batch, leaving no long-lived distortion on the subsequent batches.

IV. SUMMARY AND CONCLUSIONS

In this work we have introduced a new approach aimed at the importance sampling of complex valued distributions through the use of representations.

We start by showing that, quite generally, the more complex (as opposed to positive definite) a distribution is, the farther into the complex plane must go any representation of it. Explicit bounds are given on how narrow a strip parallel to the real axis can be to contain the support of a representation of a given complex distribution.

Then we develop new techniques to construct representations of complex probabilities, alternative to the usual complex Langevin. Since the quality of the representation, as measured by its distance to the real axis, is essential to have controlled fluctuations in the Monte Carlo estimates, the methods presented intend to be optimal in this regard. In particular, for the purposes of illustration, we show representations of complex distributions for which the complex Langevin approach fails. We also show that the constructions can be extended for manifolds of dimension higher than one, although with increasing difficulty.

The specific new proposal for the sampling of complex valued distributions on manifolds of large dimension, i.e., with many degrees of freedom, is the heat bath approach. We have not devised a complex version of the Metropolis algorithm. In its complex version, the Gibbs sampling is implemented by using a representation of the conditional probability. The formal justification of the procedure is given in Sec. III A. This is formal since it is not guaranteed that the chain of updates should converge to a correct representation of the target complex distribution, it is only shown that such complex distribution is a fixed point of the algorithm.

To assess the performance of the proposal we consider first a field on a hypercubic lattice with periodic boundary conditions subjected to a quadratic action. This makes the update procedure particularly simple. We find that the method works correctly in this case, provided the imaginary part of the coupling is not too large.

To do a more thorough analysis we have added a $\lambda\phi^4$ term to the previous action, thus transforming the problem into a nonlinear (interacting) one. The representation of the conditional probability is no longer straightforward and the new construction methods developed in previous sections have to be applied. This is needed to achieve representations of sufficient quality. We show that a Monte Carlo calculation is possible and obtain various results for different geometries and couplings. Suitable observables are considered so that generalized virial relations apply. This is used to check the accuracy of the Monte Carlo results. Also we have compared with reweighting and with complex Langevin calculations. Due to overlap problems, our

¹¹A completely similar bias was observed in CL in preliminary calculations using $\Delta t \leq 0.01$ and $|\Delta\phi_x| \leq 0.01$, together with a noticeable shift in O_1 and O_2 . The bias in I_1 disappears with the finer steps, although one of the 16^3 runs crashed.

¹²Also, RW calculations putting the system in a similar discretized box show no bias (for lattices and couplings for which our RW calculation is reliable).

reweighting calculation is only reliable for 3^3 lattices, and with more noise for 8^3 with $\beta = 0.25i$.

For $d = 1$ lattices, good results are obtained for β as large as i . For $d = 3, 4$, we find that for moderate couplings, $\beta = 0.25i$ and $0.5i$, the observables O_1 and O_2 and I_2 are well reproduced, using reweighting and/or complex Langevin results and the Schwinger-Dyson relations as benchmarks. The results tend to deteriorate for larger couplings, as the random walk makes more frequent visits to regions farther apart from the real axis.

The quantity $\text{Re}\langle I_1 \rangle$ displays a small systematic bias which depends on the size of the box used to construct the representations. We have analyzed in some detail how, for large boxes, this bias is introduced by very specific contributions which are easy to identify and remove, without altering the other observables.

Although not free from obstacles, it seems clear that the approach opens alternative routes for the complex sampling problem, and at least some of the technical limitations founds could eventually be overcome, in particular, allowing to treat harder cases (larger β). In [32] it is shown that representations exist and can be constructed for complex distributions defined on compact Lie groups, so there is no impediment of principle to extend the approach to gauge models.

On a far more hypothetical note, simple variable counting indicates that a complex distribution can be traded by two positive distributions. We have illustrated this point with our two-branch representations. In this view, perhaps it could be possible to replace the standard lattice QCD formulation with chemical potential, a complex distribution, by a two-branch version, modeled to be positive, local and having the correct symmetries, as well as a parameter representing the chemical potential, and letting universality considerations in the continuum limit to identify it with the standard formulation. Such an approach would avoid altogether the need to deal with complex distributions and its sampling.

ACKNOWLEDGMENTS

The author thanks C. García-Recio for help with the fast Fourier transform code, and the Centro de Servicios de Informática y Redes de Comunicaciones (CSIRC), Universidad de Granada, for providing computing time. This work was supported by Spanish Ministerio de Economía y Competitividad and European FEDER funds (Grant No. FIS2014-59386-P), and by the Agencia de Innovación y Desarrollo de Andalucía (Grant No. FQM225).

-
- [1] N. Madras, *Lectures on Monte Carlo methods, Volume 16 of Fields Institute Monographs* (American Mathematical Society, Providence, 2002).
 - [2] M. Troyer and U.J. Wiese, Computational Complexity and Fundamental Limitations to Fermionic Quantum Monte Carlo Simulations, *Phys. Rev. Lett.* **94**, 170201 (2005).
 - [3] P. Hasenfratz and F. Karsch, Chemical potential on the lattice, *Phys. Lett.* **125B**, 308 (1983).
 - [4] G. Parisi, On complex probabilities, *Phys. Lett.* **131B**, 393 (1983).
 - [5] H. W. Hamber and H. c. Ren, Complex probabilities and the Langevin equation, *Phys. Lett.* **159B**, 330 (1985).
 - [6] G. Bhanot, A. Gocksch, and P. Rossi, On simulating complex actions, *Phys. Lett. B* **199**, 101 (1987).
 - [7] L. L. Salcedo, Spurious solutions of the complex Langevin equation, *Phys. Lett. B* **305**, 125 (1993).
 - [8] T. D. Kieu and C. J. Griffin, Monte Carlo simulations with indefinite and complex valued measures, *Phys. Rev. E* **49**, 3855 (1994).
 - [9] L. L. Salcedo, Representation of complex probabilities, *J. Math. Phys. (N.Y.)* **38**, 1710 (1997).
 - [10] I. M. Barbour, S. E. Morrison, E. G. Klepfish, J. B. Kogut, and M. P. Lombardo, Results on finite density QCD, *Nucl. Phys. A, Proc. Suppl.* **60**, 220 (1998).
 - [11] S. Chandrasekharan and U.J. Wiese, Meron Cluster Solution of a Fermion Sign Problem, *Phys. Rev. Lett.* **83**, 3116 (1999).
 - [12] Z. Fodor and S. D. Katz, A new method to study lattice QCD at finite temperature and chemical potential, *Phys. Lett. B* **534**, 87 (2002).
 - [13] N. Prokof'ev and B. Svistunov, Worm Algorithms for Classical Statistical Models, *Phys. Rev. Lett.* **87**, 160601 (2001).
 - [14] S. A. Baurle, Method of Gaussian Equivalent Representation: A New Technique for Reducing the Sign Problem of Functional Integral Methods, *Phys. Rev. Lett.* **89**, 080602 (2002).
 - [15] A. G. Moreira, S. A. Baurle, and G. H. Fredrickson, Global Stationary Phase and the Sign Problem, *Phys. Rev. Lett.* **91**, 150201 (2003).
 - [16] V. Azcoiti, G. Di Carlo, A. Galante, and V. Laliena, Finite density QCD: A new approach, *J. High Energy Phys.* **12** (2004) 010.
 - [17] J. Berges and I.-O. Stamatescu, Simulating Nonequilibrium Quantum Fields with Stochastic Quantization Techniques, *Phys. Rev. Lett.* **95**, 202003 (2005).
 - [18] G. Aarts, Can Stochastic Quantization Evade the Sign Problem? The Relativistic Bose gas at Finite Chemical Potential, *Phys. Rev. Lett.* **102**, 131601 (2009).

- [19] D. Banerjee and S. Chandrasekharan, Finite size effects in the presence of a chemical potential: A study in the classical nonlinear $O(2)$ sigma model, *Phys. Rev. D* **81**, 125007 (2010).
- [20] J. Bloch, Evading the Sign Problem in Random Matrix Simulations, *Phys. Rev. Lett.* **107**, 132002 (2011).
- [21] 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).
- [22] R. Rota, J. Casulleras, F. Mazzanti, and J. Boronat, Quantum Monte Carlo estimation of complex-time correlations for the study of the ground-state dynamic structure function, *J. Chem. Phys.* **142**, 114114 (2015).
- [23] P. R. Crompton, Composite reweighting the Glasgow method for finite density QCD, *Nucl. Phys.* **B619**, 499 (2001).
- [24] G. Parisi and Y. s. Wu, Perturbation theory without gauge fixing, *Scientia Sinica* **24**, 483 (1981).
- [25] J. Bloch, J. Mahr, and S. Schmalzbauer, Complex Langevin in low-dimensional QCD: The good and the not-so-good, *Proc. Sci.*, LATTICE2015 (2016) 158 [arXiv:1508.05252].
- [26] C. Pehlevan and G. Guralnik, Complex Langevin equations and Schwinger-Dyson equations, *Nucl. Phys.* **B811**, 519 (2009).
- [27] J. Ambjorn, M. Flensburg, and C. Peterson, Langevin simulations of configurations with static charges, *Phys. Lett.* **159B**, 335 (1985).
- [28] 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).
- [29] J. Ambjorn and S. K. Yang, Numerical problems in applying the Langevin equation to complex effective actions, *Phys. Lett.* **165B**, 140 (1985).
- [30] B. Soderberg, On the complex Langevin equation, *Nucl. Phys.* **B295**, 396 (1988).
- [31] D. Weingarten, Complex Probabilities on \mathbb{R}^{**N} as Real Probabilities on \mathbb{C}^{**N} and an Application to Path Integrals, *Phys. Rev. Lett.* **89**, 240201 (2002).
- [32] L. L. Salcedo, Existence of positive representations for complex weights, *J. Phys. A* **40**, 9399 (2007).
- [33] J. Wosiek, Beyond complex Langevin equations I: Two simple examples, arXiv:1511.09083.
- [34] J. Wosiek, Beyond complex Langevin equations: From simple examples to positive representation of Feynman path integrals directly in the Minkowski time, *J. High Energy Phys.* **04** (2016) 146.
- [35] W. Bietenholz and U. J. Wiese, Perfect lattice actions for quarks and gluons, *Nucl. Phys.* **B464**, 319 (1996).
- [36] K. Fujimura, K. Okano, L. Schulke, K. Yamagishi, and B. Zheng, On the segregation phenomenon in complex Langevin simulation, *Nucl. Phys.* **B424**, 675 (1994).
- [37] J. Flower, S. W. Otto, and S. Callahan, Complex Langevin equations and lattice gauge theory, *Phys. Rev. D* **34**, 598 (1986).
- [38] 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).