Alleviating the sign problem in a chiral random matrix model with contour deformations

Matteo Giordano,[1] Attila Pásztor,[1,2] Dávid Pesznyák,[1] and Zoltán Tulipánt[1]

[1]*ELTE Eötvös Loránd University, Institute for Theoretical Physics, Pázmány Péter sétány 1/A, H-1117, Budapest, Hungary*

[2]*HUN-REN-ELTE Theoretical Physics Research Group, Pázmány Péter sétány 1/A, 1117, Budapest, Hungary*

(Received 24 April 2023; accepted 14 September 2023; published 17 November 2023)

We studied integration contour deformations in the chiral random matrix theory of Stephanov [Phys. Rev. Lett. **76**, 4472 (1996)] with the goal of alleviating the finite-density sign problem. We considered simple ansätze for the deformed integration contours and optimized their parameters. We find that optimization of a single parameter manages to considerably improve on the severity of the sign problem. We show numerical evidence that the improvement achieved is exponential in the degrees of freedom of the system, i.e., the size of the random matrix. We also compare the optimization method with contour deformations coming from the holomorphic flow equations.

DOI: 10.1103/PhysRevD.108.094507

I. INTRODUCTION

Euclidean quantum field theories at nonzero particle density (or chemical potential) generally suffer from a complex action problem: the weights in the path integral representation are complex, and thus cannot be interpreted as a joint probability density function on the space of field configurations (up to a proportionality factor). This prevents the use of importance sampling methods for the direct simulation of these theories. In QCD, this complex action problem severely hampers first-principles studies of dense matter in the core of neutron stars, in neutron star mergers, in core collapse supernovae, as well as in heavy ion collisions at certain collision energies.

In the presence of a complex action problem one can still (in principle) simulate a modified theory with real and positive weights, and then use reweighting methods to calculate observables in the theory of interest. If the target theory has field variables ϕ, path integral weights $w_t(\phi)$, and partition function $Z_t = \int \mathcal{D}\phi w_t(\phi)$, and the simulated theory has the same field variables, but different—real and positive—path integral weights $w_s(\phi)$ and partition function $Z_s = \int \mathcal{D}\phi w_s(\phi)$, we can obtain expectation values in the target theory via the formula

$$\langle \mathcal{O} \rangle_t = \frac{\left\langle \frac{w_t}{w_s} \mathcal{O} \right\rangle_s}{\left\langle \frac{w_t}{w_s} \right\rangle_s}, \qquad \langle \mathcal{O} \rangle_x = \frac{1}{Z_x} \int \mathcal{D}\phi w_x(\phi) \mathcal{O}(\phi), \tag{1}$$

where x may stand for t or s and $\mathcal{O}(\phi)$ is some physical observable of interest. The denominator in Eq. (1) gives the ratio of the partition functions in the target and simulated theories, i.e.,

$$\left\langle \frac{w_t}{w_s} \right\rangle_s = \frac{Z_t}{Z_s}. \tag{2}$$

This ratio is typically exponentially small in the physical volume, with the exponent given by the free energy difference between the target and simulated theories. This ratio is also a rough measure of the numerical difficulty of a given reweighting scheme, with a given simulated and target theory. In order for reweighting to be effective, one wants the target and simulated theories to be as close to each other as possible. Ideally, one should find a simulated theory with $Z_s \approx Z_t$.

Two simple choices of a simulated theory are the phase-quenched (PQ) theory, with simulated weights proportional to

$$w_s^{\mathrm{PQ}} \equiv |w_t(\phi)|, \tag{3}$$

or—assuming that the partition function Z_t is real—the sign-quenched (SQ) theory, with simulated weights proportional to

$$w_s^{\mathrm{SQ}} \equiv |\mathrm{Re} w_t(\phi)|. \tag{4}$$

For the first case (phase reweighting) the reweighting factors $w_t/w_s^{\mathrm{PQ}} \equiv e^{i\theta}$ are pure phases. For the second case (sign reweighting) the reweighting factors are $w_t/w_s^{\mathrm{SQ}} = e^{i\theta}/|\cos\theta|$. For certain observables, such as manifestly real

observables or observables with a conjugation ($\phi \to \bar{\phi}$) symmetry, one can substitute w_t/w_s^{PQ} with $\cos\theta$ and w_t/w_s^{SQ} with a pure sign $\cos\theta/|\cos\theta|$. For phase or sign reweighting, we can then say that the complex action problem becomes a sign problem: the cancellations between contributions with different signs of $\cos\theta$ lead to a small Z_t/Z_s ratio, and in turn to small signal-to-noise ratios in the expectation values of observables.

The sign-quenched ensemble always has a less severe sign problem, due to the inequality $Z_t < Z_s^{\mathrm{SQ}} < Z_s^{\mathrm{PQ}}$, which is a consequence of $\cos\theta \le |\cos\theta| \le 1$. However, in the limit of a severe sign problem—i.e., as the distribution of the argument θ tends to a uniform distribution on $[-\pi, \pi)$—the severity of the sign problem for these two reweighting schemes only differs by a constant factor [1], given by $(Z_s^{\mathrm{PQ}}/Z_s^{\mathrm{SQ}})^2 \to (\pi/2)^2$.

In QCD and in other (more or less) QCD-like models, describing the interactions of several "flavors" of fermions, the path integral weights can be written schematically as

$$w_t(\phi) = \det M_1(\phi, \mu_1) \cdots \det M_{N_f}(\phi, \mu_{N_f}) e^{-S_B(\phi)}, \quad (5)$$

where the fields ϕ are real bosonic variables and S_B is the corresponding bosonic part of the action, N_f is the number of fermion flavors in the model, $\det M_k$ is the fermionic determinant of the kth flavor, and μ_k is the corresponding chemical potential, for $k = 1, \ldots, N_f$. The source of the sign problem is the fermionic determinant, which at nonzero μ is generally a complex number. Moreover, an important feature of the sign problem in QCD and QCD-like theories is that it tends to get much worse in the ranges of μ where zeros of the determinant in the complex μ plane become dense [2].

Nonetheless, reweighting from the phase- and sign-quenched theories is starting to become feasible even in full QCD [1,3], which has recently led to the calculation of the equation of state of a hot-and-dense quark-gluon plasma in the region of chemical potentials covered by the RHIC Beam Energy Scan [4]. However, the range of practical applicability of such an approach is limited in both volume and chemical potential by the smallness of the ratio Z_t/Z_s. Lacking a solution of the sign problem, it is then desirable to develop methods that at least alleviate it, to extend the range of parameters that reweighting methods can practically reach.

One possible route to do this is the use of contour deformations in the path integral (see Ref. [5] for a recent review). If the path integral weights $w_t(\phi)$ are holomorphic functions of the field variables,[1] the multivariate Cauchy theorem guarantees that complexified integration manifolds in the same homology class as the original one yield the same partition function. However, the phase- and sign-quenched integrands are not holomorphic, and therefore the phase- and sign-quenched partition functions are not invariant under such deformations. It may then be possible to bring the ratios Z_t/Z_s closer to unity, thus making reweighting more effective.

There are different ways to deform integration contours. Historically, methods based on Lefschetz thimbles appeared first [5,8–13]. Lefschetz thimbles are the disjoint components of the integration contour defined by requiring that the imaginary part of the classical action is constant in each component. The thimble structure of theories with a fermionic determinant is usually quite complicated [14–18]. Simple toy models reveal the following features: (i) cancellations between competing thimbles are very important for getting the correct results, and (ii) the thimbles themselves are not smooth at the zeros of the fermionic determinant. Thus, the use of thimbles might be impractical for such theories. However, Lefschetz thimbles are, in general, not the numerically optimal integration contours [19]; i.e., they are not necessarily the contours with the largest Z_t/Z_s, so there is no need to concentrate solely on them.

A second class of methods is based on numerical optimization. The main idea here is to parametrize the integration manifold by a finite number of parameters, which are then optimized to make the sign problem as mild as possible. Such methods were applied to a one-dimensional integral [20], the $0+1$D scalar theory [21], the $0+1$D Polyakov-improved Nambu-Jona-Lasinio model [22], $0+1$D QCD [23], the $1+1$D scalar field theory [24], the $1+1$D Thirring model [25], the $2+1$D Thirring model [26], Bose gases of several dimensions [27], the $1+1$D U(1) gauge theory with a complex coupling constant [28], and the $2+1$D XY model at finite density [29]. Here, we apply contour optimization methods to a fermionic toy model that shares relevant technical features with finite chemical potential QCD: the chiral random matrix model proposed by Stephanov in Ref. [30].

Since it is an exactly solvable model with a sign problem, the Stephanov model is a very useful test bed for methods aimed at solving or alleviating the sign problem. This model has been studied with the complex Langevin approach [31–33], which fails for this particular model [34] even with the introduction of gauge cooling [33]. There are also preliminary results for this model with the tempered Lefschetz thimble method [35] which is based on parallel tempering [36] in the flow time of the holomorphic flow [10,37]. This method—similar to other flow-based methods—produces a weaker sign problem, albeit at the cost of substantially increasing the per-configuration cost of generating the ensemble compared to ordinary phase reweighting.

In this paper we study the Stephanov model with optimization methods. There are, roughly speaking, two approaches to such an optimization: one can look for the

[1]A notable exception is lattice QCD with rooted staggered fermions [6,7].

optimum using either a very general ansatz with a large number of parameters or a very specific ansatz tailored for the model at hand, and with a small number of parameters. The first approach has clearly the potential to find a good optimum, e.g., using machine learning techniques, but it also has some disadvantages. In fact, for such a general approach the number of optimization parameters has to be increased as one increases the number of degrees of freedom of the system. This means that the cost of finding good contours might turn out to be prohibitive, similar to what happens with methods based on Lefschetz thimbles. In this exploratory study we follow the second, *ad hoc* approach, and optimize ansätze with only a few parameters. Moreover, the number of these parameters is kept independent of the number of degrees of freedom of the system. We can then be sure that the optimization itself is numerically cheap, and that the per-configuration cost of generating the ensembles is essentially as low as on the original contours. Obviously, the drawback of this approach is that to write down an ansatz with only a few parameters that produces a substantial improvement in the severity of the sign problem, some physical or mathematical insight is needed.

For the toy model studied in this paper, the insight required to use the *ad hoc* approach is available, and so we can write down appropriate ansätze. We will then show that a quite cheap numerical optimization procedure leads one to contours with a reduced sign problem. We will also present numerical evidence that the reduction in the severity of the sign problem is exponential: while the sign problem on the optimized contours is still exponential in the number of degrees of freedom, the corresponding exponent is reduced. This conclusion is similar to what some of us have shown in Ref. [29] for a purely bosonic model [the $(2+1)$-dimensional XY model at a nonzero chemical potential]. Notably, such an exponential reduction can be achieved without changing the number of optimization parameters with the system size.

In this work we will only consider phase-quenched simulations, for simplicity. Similar arguments and methods should, however, also apply to the sign-quenched case [29].

The plan of the paper is the following: In Sec. II we introduce the model discussed in this work. In Sec. III we provide details on the different contour deformation procedures we tested. In Sec. IV we illustrate the chemical potential and volume dependence of the achieved improvement and also compare our results with a method based on Lefschetz thimbles: the holomorphic flow of Ref. [10]. We summarize our conclusions in Sec. V.

II. THE CHIRAL RANDOM MATRIX MODEL

Throughout this paper we will only consider $N_f = 2$ with $\mu_1 = \mu_2 \equiv \mu$ for simplicity. The random matrix model of Stephanov [30] for N_f degenerate flavors of quarks is then defined by the partition function

$$\mathcal{Z}_N^{N_f} = e^{N\mu^2} \int \mathrm{d}W \mathrm{d}W^\dagger (\det(D+m))^{N_f} e^{-N\mathrm{Tr}WW^\dagger}, \quad (6)$$

where the massless Dirac matrix is

$$D = \begin{pmatrix} 0 & iW+\mu \\ iW^\dagger+\mu & 0 \end{pmatrix}, \quad (7)$$

m is the quark mass, and W is a general $N \times N$ complex matrix. The model has no concept of physical volume. The number of degrees of freedom of the model scales with N^2.

The two observables we will study in this paper are the chiral condensate

$$\Sigma = \frac{1}{2N} \frac{\partial \log \mathcal{Z}_N^{N_f}}{\partial m} \quad (8)$$

and the quark density

$$n = \frac{1}{2N} \frac{\partial \log \mathcal{Z}_N^{N_f}}{\partial \mu}. \quad (9)$$

An important feature of the model is that it can be solved analytically, both in the $N \to \infty$ limit where the integral is dominated by a saddle point and at finite N where it reduces to the calculation of moments of Gaussian integrals. Thus, in this particular model we will be able to compare numerical results with exact analytic solutions.

The model shares with QCD the feature that the phase-quenched theory corresponds to an isospin chemical potential and has an analog of the pion condensation transition at some $\mu = \mu_c^{\mathrm{PQ}}$. For chemical potentials exceeding μ_c^{PQ} the sign problem of the model is severe. From the point of view of the Dirac spectrum, for $\mu = 0$ the eigenvalues are purely imaginary, while for $\mu \neq 0$ the eigenvalues of D acquire a real part and are distributed inside a strip of width μ^2 in the real direction. When the quark mass is inside this strip, the model has a severe sign problem. This roughly corresponds to the analog of the pion condensed phase in the phase-quenched theory. Due to these similarities, this model has been considered several times in the literature as a good toy model for the sign problem in QCD [34,35].

In this model, unlike in QCD, the expectation value of the average phase does not always tend to zero in the limit of an infinite system. Rather, it only goes to zero in a given range of chemical potentials bounded by the solutions to the equation [38]:

$$0 = 1 - \mu^2 + \frac{m^2}{\mu^2 - m^2} - \frac{m^2}{4(\mu^2 - m^2)^2}. \quad (10)$$

Using a quark mass of $m = 0.2$, the two solutions of this equation are $\mu = 0.35 = \mu_c^{\mathrm{PQ}}$ and 1.02. This is the regime where the sign problem in the model is strongest.

III. CONTOUR DEFORMATION METHODS

A. Optimization method

We will restrict ourselves to ansätze with simple, analytically calculable Jacobians with $\mathcal{O}(N^0)$ computational cost and a small number of parameters, independent of the number of degrees of freedom.

Let $A = \mathrm{Re}W$ and $B = \mathrm{Im}W$. These two real matrices will be deformed to complex matrices α and β. Thus,

$$W = A + iB \to X = \alpha + i\beta,$$
$$W^\dagger = A^{\mathrm{T}} - iB^{\mathrm{T}} \to Y = \alpha^{\mathrm{T}} - i\beta^{\mathrm{T}}. \tag{11}$$

Here the complex matrices α and β will be parametrized by the same number of real parameters as A and B. After applying such a deformation $X^\dagger \neq Y$. The severity of the sign problem is then given by

$$\langle e^{i\theta}\rangle = \left\langle \left[\frac{\det(D+m)\det\mathcal{J}}{|\det(D+m)\det\mathcal{J}|}\right]^{N_f} e^{-iN\mathrm{Im}\mathrm{Tr}XY}\right\rangle, \tag{12}$$

where the Jacobian determinant is

$$\det\mathcal{J} = \left|\frac{\partial(\alpha,\beta)}{\partial(A,B)}\right|. \tag{13}$$

B. Holomorphic flow

Using the holomorphic flow (or generalized thimble method) of Ref. [10] for the complexified action of the Stephanov model,

$$S = -N\mu^2 - N_f \log\det(D+m) + N\mathrm{Tr}(XY), \tag{14}$$

we deform the integration manifold by evolving the original one with the differential equation

$$\frac{\mathrm{d}Y_{ij}}{\mathrm{d}t} = \overline{\frac{\partial S}{\partial Y_{ij}}} = N\bar{X}_{ji} - N_f[(\overline{XG})_{ji} + i\mu\bar{G}_{ji}], \tag{15}$$

where the overbar denotes complex conjugation, t is the flow parameter, and

$$G = [m^2 - \mu^2 - i\mu(X+Y) + YX]^{-1}. \tag{16}$$

Solving this system of equations with initial conditions $X_0 = W$, $Y_0 = W^\dagger$ for a fixed flow time t_{f} we obtain a deformed manifold $\mathcal{M}_{t_{\mathrm{f}}}$. We parametrize each point on the flowed manifold by the real matrices A and B. That is, we parametrize the flowed manifold by the initial conditions of the flow equation.

The computation of expectation values requires the Jacobian of the holomorphic flow,

$$\det J = \left|\frac{\partial(X,Y)}{\partial(A,B)}\right|, \tag{17}$$

as well. Denoting the Hessian with H, the Jacobian matrix J is obtained as the solution of the equation

$$\frac{\mathrm{d}J}{\mathrm{d}t} = \overline{HJ}, \tag{18}$$

with initial conditions

$$J_{X_{ij},A_{ij}} = 1, \quad J_{X_{ij},B_{ij}} = i, \quad J_{Y_{ij},A_{ji}} = 1, \quad J_{Y_{ij},B_{ji}} = -i. \tag{19}$$

Computing the Jacobian directly is numerically expensive, so we estimate it [39] with

$$W = \exp\left[\int_0^{t_{\mathrm{F}}} \mathrm{d}t\,\mathrm{Tr}\,\overline{H(t)}\right]. \tag{20}$$

The difference between W and $\det J$ is taken into account by reweighting when computing observables,

$$\langle\mathcal{O}\rangle = \frac{\langle\mathcal{O}e^{-\Delta S}\rangle_{S'_{\mathrm{eff}}}}{\langle e^{-\Delta S}\rangle_{S'_{\mathrm{eff}}}}, \tag{21}$$

where $S'_{\mathrm{eff}} = S - \ln W$, $\Delta S = S_{\mathrm{eff}} - \mathrm{Re}S'_{\mathrm{eff}}$, and $\langle.\rangle_{S'_{\mathrm{eff}}}$ is the average with respect to $e^{-\mathrm{Re}S'_{\mathrm{eff}}}$. This way, we needed to compute $\det J$ exactly only for the configurations used for measurements.

In the large flow time limit, the flowed manifold tends toward the Lefschetz thimbles. At smaller flow times, it still reduces the sign problem, although less than a complete thimble decomposition would.

IV. NUMERICAL RESULTS

A. Simple ansätze

As a rule, all of our ansätze have been parametrized such that the undeformed integration manifold is at value zero for all optimizable parameters.

1. Ansatz-1

From the definition in Eq. (7) it is easy to see that the sign problem can be removed from the quark determinant by a simple shift of the form $\alpha = A + i\mu\mathbf{1}$. This, however, introduces a sign problem in the Gaussian term $e^{-N\mathrm{Tr}(XY)}$. By finding a trade-off between the two terms, the severity of the sign problem may be optimized. This motivates our first ansatz, with two real parameters k_1 and k_2 defined by

$$\alpha = A + ik_1\mathbf{1}, \tag{22}$$

$$\beta = B + ik_2\mathbf{1}. \tag{23}$$

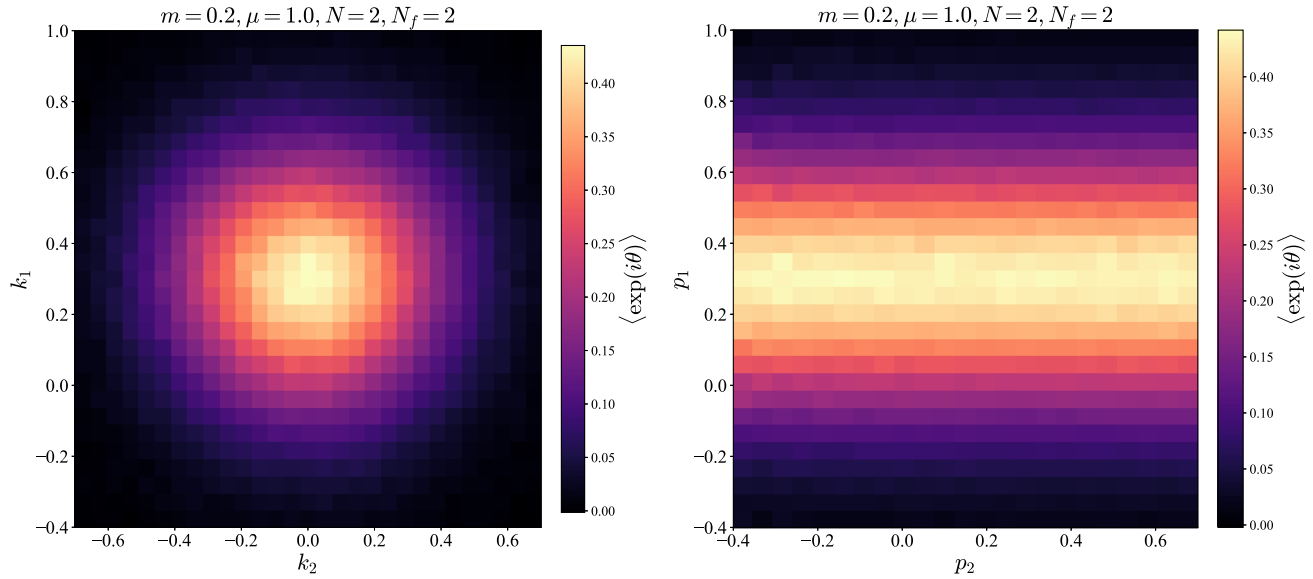

FIG. 1. Left: the average phase with Ansatz-1 as a function of k_1 and k_2. There is a local minimum at $k_2 \approx 0$ and $k_1 > 0$. Right: the average phase with Ansatz-2 as a function of p_1 and p_2. There is an apparent saddle parallel to the $p_2 = 0$ line at $p_1 = k_1 > 0$. The plot is cut above $p_2 = -1/2$, since the Jacobian matrix becomes singular at that value.

The Jacobian determinant for this ansatz is simply unity. The parameter k_2 is introduced on a whim, as the matrices A and B do not have to be treated symmetrically. The results for the average phase in a scan in these two parameters for $N = 2$, $m = 0.2$, and $\mu = 1.0$ is shown in Fig. 1 (left). While there is a clearly nonzero optimal value for k_1, the optimal value of the k_2 parameter is near zero. This remains true for all values of the parameters N, μ, and m we simulated.

2. Ansatz-2

When we introduce a shift $A \to A + ik\mathbf{1}$, the argument of the Gaussian term changes according to

$$\mathrm{Tr}(XY) = \mathrm{Tr}(AA^{\mathrm{T}} + BB^{\mathrm{T}}) - Nk^2 + 2ik\mathrm{Tr}A. \quad (24)$$

This motivates our second ansatz, with two real parameters p_1 and p_2 defined by

$$\alpha = A + ip_1\mathbf{1} + p_2\mathrm{Tr}A\mathbf{1}, \quad (25)$$

$$\beta = B. \quad (26)$$

The p_1 parameter of this ansatz is identical to the k_1 parameter of the previous ansatz. The Jacobian determinant for this ansatz is simply $\det \mathcal{J} = 1 + Np_2$, i.e., configuration-independent, and can be ignored. The results for the average phase in a scan in these two parameters for $N = 2$, $m = 0.2$, and $\mu = 1.0$ can be seen in Fig. 1 (right). While there is a clearly nonzero optimal value for $p_1 = k_1$, the p_2 parameter only appears to move on a saddle.

3. Ansatz-3

We now move on to a more complicated ansatz with 10 complex (or 20 real) parameters $a, b, c, d, e, f, g, h, j, k$ defined by

$$\alpha = (a + b\mathrm{Tr}A + c\mathrm{Tr}B)\mathbf{1} + (1 + d)A + eB, \quad (27)$$

$$\beta = (f + g\mathrm{Tr}A + h\mathrm{Tr}B)\mathbf{1} + jA + (1 + k)B. \quad (28)$$

The Jacobian determinant for this ansatz is

$$\begin{aligned}\det \mathcal{J} = {} & ((1 + d)(1 + k) - ej)^{N^2-1} \\ & \times [((1 + d) + Nb)((1 + k) + Nh) \\ & - (e + Nc)(j + Ng)]. \end{aligned} \quad (29)$$

The severity of the sign problem was then optimized via the AdaDelta method [40], with the objective function

$$-\log\langle e^{i\theta}\rangle = -\log\frac{\mathcal{Z}}{\mathcal{Z}_{\mathrm{PQ}}} = -\log\mathcal{Z} + \log\mathcal{Z}_{\mathrm{PQ}}, \quad (30)$$

where we suppressed the N and N_f indices for the partition function. The gradient with respect to the deformation parameters is given by

$$\nabla \log \mathcal{Z}_{\mathrm{PQ}} = -\langle \nabla S^{\mathrm{a}}_{\mathrm{eff}} \rangle, \quad (31)$$

where

$$S^{\mathrm{a}}_{\mathrm{eff}} = N\mathrm{ReTr}XY - N_f \log|\det M| - \log|\det \mathcal{J}| \quad (32)$$

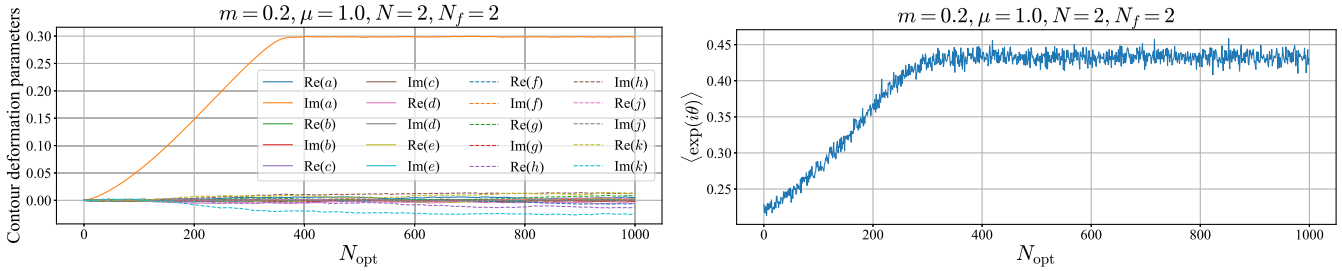

FIG. 2. Left: parameters as a function of the optimization step for Ansatz-3. Right: the average phase as a function of the optimization step for Ansatz-3.

with gradient

$$\nabla S_{\text{eff}}^{\text{a}} = N\text{ReTr}[(\nabla X)Y + X(\nabla Y)] - \frac{N_f}{2}\text{Tr}[M^{-1}(\nabla M) + \bar{M}^{-1}(\nabla \bar{M})] - \text{Re}\left[\frac{\nabla \det \mathcal{J}}{\det \mathcal{J}}\right]. \tag{33}$$

Note that for Ansatz-3 the Jacobian is independent of the configuration, and the last term can be dropped from Eq. (33). For Ansatz-4, to be discussed below, the Jacobian will depend on the configuration, and thus the last term is needed. An example of such an optimization run is shown in Fig. 2. As with the previous two ansätze, only a single parameter emerges $k_1 = p_1 = \text{Im}a$.

4. Ansatz-4

Experiments with the first three ansätze revealed only one parameter of interest, which can be thought of as a simple one-parameter imaginary shift of the trace of the matrix A. One might wonder whether more general deformations of the trace could lead to a better improvement. Thus we look at nonlinear deformations of the trace $\tau = \text{Tr}A$ of the matrix A with an undeformed B matrix. The integral measure is given by

$$\prod_{i,j=1}^{N} \mathrm{d}A_{ij} = \mathrm{d}\tau \prod_{\substack{i,j=1 \\ (i,j)\neq(N,N)}}^{N} \mathrm{d}A_{ij} = \mathrm{d}\tau \prod_{\substack{i,j=1 \\ i\neq j}}^{N} \mathrm{d}A_{ij} \prod_{k=1}^{N} \mathrm{d}\left(A_{kk} - \frac{\tau}{N}\right) \tag{34}$$

The deformed matrix α is obtained from A as

$$A = \frac{\tau}{N}\mathbf{1} + \left(A - \frac{\tau}{N}\mathbf{1}\right) = \frac{\tau}{N}\mathbf{1} + \tilde{A} \rightarrow \alpha = \frac{\tilde{\tau}}{N}\mathbf{1} + \tilde{A}, \tag{35}$$

where $\text{Tr}\tilde{A} = 0$ and

$$\tilde{\tau} = \tau + if(\tau; \ldots), \tag{36}$$

for some function f that depends on τ and possibly other parameters. For simplicity, we choose f to be piecewise linear,

$$f(\tau; x_{k(\tau)}, x_{k(\tau)+1}, y_{k(\tau)}, y_{k(\tau)+1}) = \frac{y_{k(\tau)}(x_{k(\tau)+1} - \tau)}{x_{k(\tau)+1} - x_{k(\tau)}} + \frac{y_{k(\tau)+1}(\tau - x_{k(\tau)})}{x_{k(\tau)+1} - x_{k(\tau)}}. \tag{37}$$

The parameters to optimize are the y_i, while the node points x_i of the linear interpolation are fixed parameters, and chosen with regular spacing, $x_{l+1} - x_l = \Delta$ for all l, and

$$k(\tau) = \text{floor}\left[\frac{\tau - x_0}{\Delta}\right]. \tag{38}$$

By numerical experimentation we have found that the choice of the node points is not important, as long as the full interpolation range is large enough to cover the most probable values of TrA on the original contours and Δ is small enough. If these conditions are met, optimal contours with ansätze with different node points appear to be piecewise approximations of the same smooth curve. The Jacobian is

$$\det \mathcal{J} = 1 + i\frac{y_{k(\tau)+1} - y_{k(\tau)}}{\Delta}. \tag{39}$$

The parameters are then optimized as with Ansatz-3. A comparison of the results from this ansatz with the constant shift found using ansätze 1 to 3 is shown in Fig. 3 for several values of the chemical potential. For highly probable values of TrA the two ansätze roughly agree, while for the highly improbably values of TrA, the optimization does not move the ansatz away from the original contour, as there are no configurations to use for the optimization of that part of the contour. These two asymptotic regimes are smoothly connected. The measured sign problem on this contour is identical to the one measured with ansätze 1 to 3, up to statistical errors—not surprisingly since deviations of f from a constant happen on unimportant configurations.

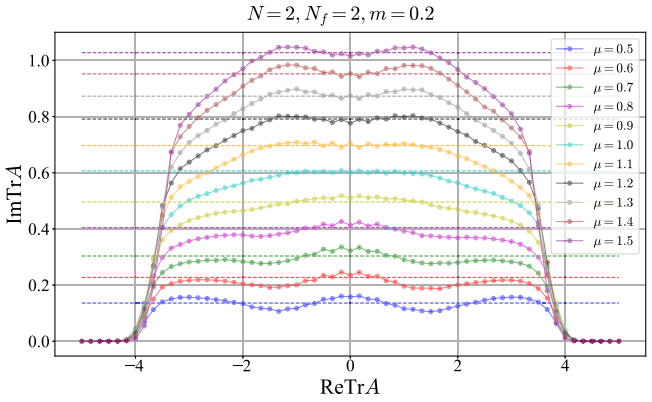

FIG. 3. Ansatz-4 (piecewise optimization of the trace: full points) compared to Ansatz-1 (imaginary constant shift of A proportional to the unit matrix: dashed line) for several values of the chemical potential. The two procedures find very similar contours. The differing tails are at large values of $|\mathrm{Tr}A|$ and have a small statistical weight, while the small wiggles in the middle around the average value $k_1 = p_1$ do not substantially change the severity of the sign problem.

B. Chemical potential and matrix size dependence

Now that we have discovered a good contour deformation parameter, let us look at what kind of improvements can be achieved by such a one-parameter deformation. From here on out we show results with Ansatz-1, with k_2 set to zero.

The matrix size and chemical potential dependence of the average phase for the original and optimized contours are shown in Fig. 4. The "volume," i.e., matrix size dependence at a fixed chemical potential in the left panel, reveals an improvement on the sign problem that is exponential in the matrix size: while the severity of the sign problem is roughly linear on a logarithmic plot for both the original and optimized contours, the slopes are quite different. The right panel shows the chemical potential dependence for several values of N. Apparently, contour optimization improves on the sign problem the most in the regime where it is the most severe.

The statistical improvement factor, defined as the square of the ratio of the phases obtained for the deformed and for the original contour, $(\langle e^{i\theta}\rangle_{\mathrm{orig}}/\langle e^{i\theta}\rangle_{\mathrm{def}})^2$, is shown on the left panel of Fig. 5 for $N = 2$, 4, and 6. For larger matrices, $\langle e^{i\theta}\rangle$ was zero within statistical errors on the original

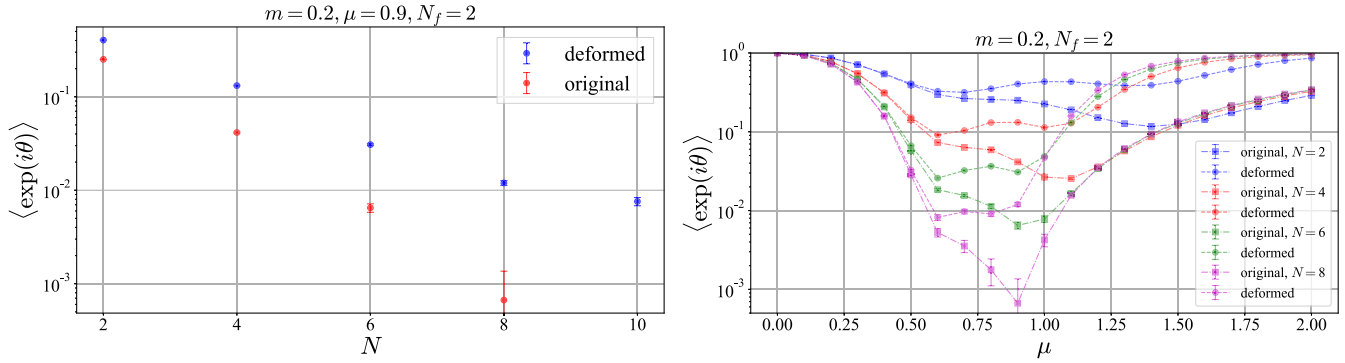

FIG. 4. Left: dependence of the average phase on the size of the random matrix for the original and optimized contours. Right: dependence of the average phase on the chemical potential for the original and optimized contours.

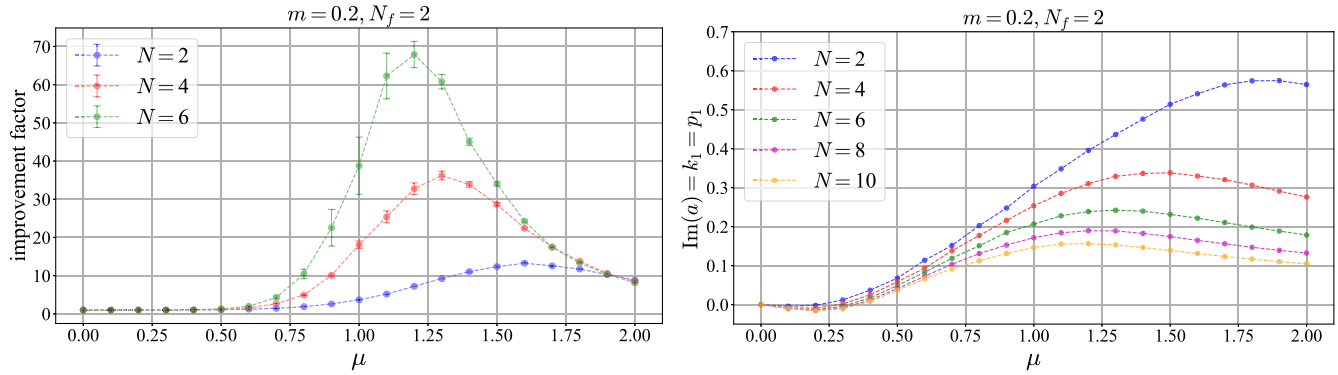

FIG. 5. Left: dependence of the statistical improvement factor $(\langle e^{i\theta}\rangle_{\mathrm{orig}}/\langle e^{i\theta}\rangle_{\mathrm{def}})^2$, achieved by contour optimization as a function of μ for different matrix sizes. Right: dependence of the optimal contour parameter $k_1 = p_1 = \mathrm{Im}a$ on μ for different matrix sizes.

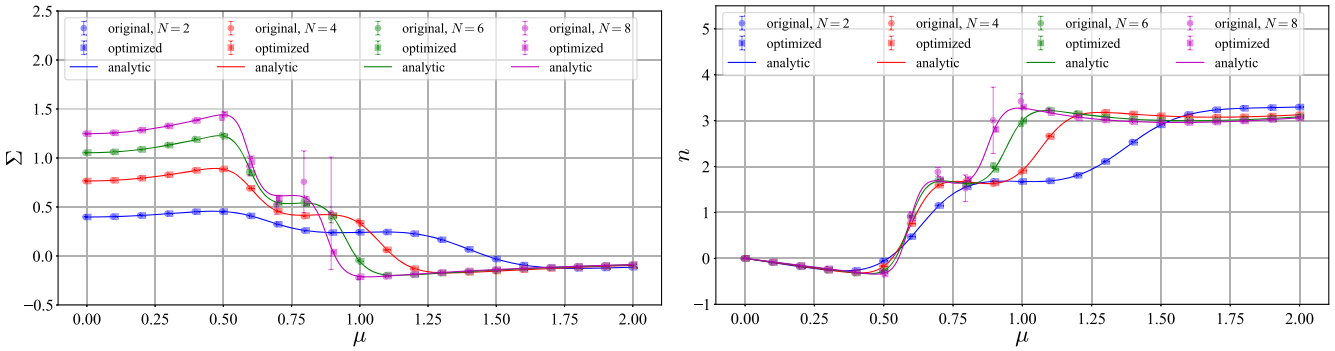

FIG. 6. The chiral condensate (left) and the quark number (right) as a function of μ for several values of the matrix size N. Results from simulations on the original and on the improved contours are compared with analytic results.

contours, and this ratio could not be calculated. We see that the ratio monotonically increases with N, and as a function of μ it is maximal close to the value of μ where the sign problem is the strongest. The optimal values for the deformation parameter $k_1 = p_1 = \mathrm{Im}a$ for different values of μ and N are shown in the right panel of Fig. 5.

As a sanity check, we also calculated the expectation value of the chiral condensate and of the quark number on both the original and the optimized contours, and compared them to the analytic results; see Fig. 6. They both show excellent agreement, but the optimized contours have significantly smaller error bars.

C. Comparison with the holomorphic flow

As experiments with simple ansätze so far revealed only a single important contour deformation parameter, it is a natural question to ask whether Lefschetz-thimble-based methods also "find" this deformation or not, and whether by utilizing such methods it is possible to improve the sign problem further compared to such a one-parameter deformation. For this reason, we performed the holomorphic flow on our $N = 2$ random matrices and obtained an estimate of the k_1 parameter from the flowed variables via $k_1^{\mathrm{flow}} = \mathrm{Im}\langle \mathrm{Tr}(\alpha(t_f) - A)\rangle/N$. This k_1 can then be substituted back in the one-parameter ansatz $\alpha = A + ik_1\mathbf{1}$, and the severity of the sign problem can be compared to the one found with the properly flowed manifold.

In Fig. 7 we show the sign problem as a function of mu for the original contour, for the optimized contour, for the flowed contour, and for the contour with k_1 extracted from the flow. We also compare two different values of the flow time. A few observations can be drawn from this figure. For smaller chemical potentials, the flow performs better than the optimization, which does not noticeably improve the sign problem. In contrast, for larger chemical potentials, optimization outperforms the flow. Interestingly, while the full flow at small chemical potentials gives a slightly weaker sign problem compared to the ansatz with k_1^{flow}, at larger chemical potentials the situation is reversed: the sign problem is slightly weaker with k_1^{flow} than with the solution of the full flow equation. The trends for the larger flow time $t_f = 0.14$ are the same, but the magnitude of the difference at lower chemical potentials is larger: the flow outperforms the simple ansatz for low to intermediate values of μ more, while at larger μ the one-parameter ansatz with k_1 extracted from the flow still outperforms the full flow toward the Lefschetz thimbles. While the simple ansatz outperforming the full flow (for any value of μ) may be somewhat surprising at first, it is not in contradiction with what we already know about contour deformations: The flow goes toward the Lefschetz thimbles, which are not generally the numerically optimal contours, and thus there is no reason for the full flow curve to be always above the curve with the simple ansatz with k_1^{flow}.

For larger matrix sizes, the holomorphic flow method runs into ergodicity problems, related to the zeros of the fermion determinant. In principle one could attempt to solve these ergodicity issues, e.g., with a combination of regularized flow equations [41] and parallel tempering [37]. This, however, makes the flow equations even more expensive. Since our main focus is on ameliorating the sign problem without blowing up the per-configuration cost of simulations, we have not gone in this direction.

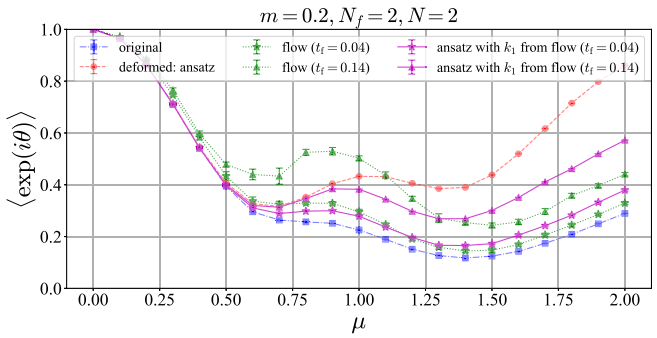

FIG. 7. The severity of the sign problem for $N = 2$ and $m = 0.2$ as a function of μ on the original contours, the optimized contours, the flowed contours, and the contours where the k_1 parameter of the ansatz is extracted from the flow for two different flow times t_f.

V. SUMMARY AND DISCUSSION

We have discussed contour deformations in the chiral random matrix model of Stephanov as a way to alleviate its sign problem. Using simple *ad hoc* ansätze we identified a single important deformation parameter, which allowed for an exponential reduction in the severity of the sign problem as a function of the matrix size.

Our results are quite encouraging, as they show that a simple one-parameter optimization can lead to an exponential reduction of the sign problem even in a fermionic theory, where the thimble decomposition is complicated and contour deformation approaches based on them might not be numerically effective. The fermionic nature of the matter fields does not appear to be a fundamental obstruction in the construction of contours that improve exponentially on the sign problem.

Furthermore, the phase diagram of the random matrix model is similar to what we expect in full QCD: the chiral phase transition is "hidden behind" the pion condensation phase in the phase-quenched theory. Hence, this bulk thermodynamic feature—the existence of a phase transition in the phase-quenched theory—also does not appear to be a fundamental obstruction.

The results and the ansätze in this paper, however, cannot be used directly to construct a good optimization ansatz in full QCD, as the toy model studied here and QCD differ on an important technical aspect. Concretely, in the Stephanov model there are contour deformations that can remove the sign problem from the fermion determinant for a single flavor (so from the full determinant when all chemical potentials are equal)—albeit at the cost of reintroducing it somewhere else in the Boltzmann weights. There are no such deformations in full QCD. The complexification of the SU(3) gauge group is the $\mathrm{SL}(3,\mathbf{C})$ group, which still requires a unit determinant for the link variables. To remove the chemical potential from a single quark determinant the timelike links would have to be deformed to $\mathrm{GL}(3,\mathbf{C})$ matrices, with a nonunit determinant, which lie outside the complexified gauge group. In the future it will therefore be important to work with more realistic toy models of QCD or even full QCD itself, as the choice of a suitable sign problem improving ansatz appears to be strongly dependent on the exact symmetries and exact matter content of a given theory.

ACKNOWLEDGMENTS

This work was supported by the NKFIH Grant No. KKP-126769. D. P. is supported by the ÚNKP-22-3 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund.

[1] S. Borsanyi, Z. Fodor, M. Giordano, S. D. Katz, D. Nogradi, A. Pasztor, and C. H. Wong, Lattice simulations of the QCD chiral transition at real baryon density, Phys. Rev. D **105**, L051506 (2022).
[2] K. Nagata, Finite-density lattice QCD and sign problem: Current status and open problems, Prog. Part. Nucl. Phys. **127**, 103991 (2022).
[3] M. Giordano, K. Kapás, S. D. Katz, D. Nógrádi, and A. Pásztor, New approach to lattice QCD at finite density; results for the critical end point on coarse lattices, J. High Energy Phys. 05 (2020) 088.
[4] S. Borsanyi, Z. Fodor, M. Giordano, J. N. Guenther, S. D. Katz, A. Pasztor, and C. H. Wong, Equation of state of a hot-and-dense quark gluon plasma: Lattice simulations at real μB vs extrapolations, Phys. Rev. D **107**, L091503 (2023).
[5] A. Alexandru, G. Basar, P. F. Bedaque, and N. C. Warrington, Complex paths around the sign problem, Rev. Mod. Phys. **94**, 015006, 2022.
[6] M. Golterman, Y. Shamir, and B. Svetitsky, Breakdown of staggered fermions at nonzero chemical potential, Phys. Rev. D **74**, 071501, 2006.
[7] M. Giordano, K. Kapás, S. D. Katz, D. Nógrádi, and A. Pásztor, Radius of convergence in lattice QCD at finite μ_B with rooted staggered fermions, Phys. Rev. D **101**, 074511 (2020).
[8] M. Cristoforetti, F. Di Renzo, and L. Scorzato, New approach to the sign problem in quantum field theories: High density QCD on a Lefschetz thimble, Phys. Rev. D **86**, 074506 (2012).
[9] 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).
[10] A. Alexandru, G. Başar, P. F. Bedaque, G. W. Ridgway, and N. C. Warrington, Sign problem and Monte Carlo calculations beyond Lefschetz thimbles, J. High Energy Phys. 05 (2016) 053.
[11] M. Fukuma and N. Matsumoto, Worldvolume approach to the tempered Lefschetz thimble method, Prog. Theor. Exp. Phys. **2021**, 023B08 (2021).
[12] F. Di Renzo, S. Singh, and K. Zambello, Taylor expansions on Lefschetz thimbles, Phys. Rev. D **103**, 034513 (2021).
[13] F. Di Renzo and K. Zambello, Solution of the Thirring model in thimble regularization, Phys. Rev. D **105**, 054501 (2022).
[14] T. Kanazawa and Y. Tanizaki, Structure of Lefschetz thimbles in simple fermionic systems, J. High Energy Phys. 03 (2015) 044.
[15] Y. Tanizaki, Y. Hidaka, and T. Hayata, Lefschetz-thimble analysis of the sign problem in one-site fermion model, New J. Phys. **18**, 033002 (2016).

[16] F. Di Renzo and G. Eruzzi, One-dimensional QCD in thimble regularization, Phys. Rev. D **97**, 014503 (2018).
[17] K. Zambello and F. Di Renzo, Towards Lefschetz thimbles regularization of heavy-dense QCD, Proc. Sci., LATTICE2018 (**2018**) 148.
[18] M. Ulybyshev, C. Winterowd, and S. Zafeiropoulos, Lefschetz thimbles decomposition for the Hubbard model on the hexagonal lattice, Phys. Rev. D **101**, 014508 (2020).
[19] S. Lawrence, Beyond thimbles: Sign-optimized manifolds for finite density, Proc. Sci., LATTICE2018 (**2018**) 149.
[20] Y. Mori, K. Kashiwa, and A. Ohnishi, Toward solving the sign problem with path optimization method, Phys. Rev. D **96**, 111501 (2017).
[21] F. Bursa and M. Kroyter, A simple approach towards the sign problem using path optimisation, J. High Energy Phys. 12 (2018) 054.
[22] K. Kashiwa, Y. Mori, and A. Ohnishi, Controlling the model sign problem via the path optimization method: Monte Carlo approach to a QCD effective model with Polyakov loop, Phys. Rev. D **99**, 014033 (2019).
[23] Y. Mori, K. Kashiwa, and A. Ohnishi, Path optimization in $0 + 1$D QCD at finite density, Prog. Theor. Exp. Phys. **2019**, 113B01 (2019).
[24] Y. Mori, K. Kashiwa, and A. Ohnishi, Application of a neural network to the sign problem via the path optimization method, Prog. Theor. Exp. Phys. **2018**, 023B04 (2018).
[25] A. Alexandru, P. F. Bedaque, H. Lamm, and S. Lawrence, Finite-density Monte Carlo calculations on sign-optimized manifolds, Phys. Rev. D **97**, 094510 (2018).
[26] A. Alexandru, F. Bedaque, H. Lamm, S. Lawrence, and N. C. Warrington, Fermions at Finite Density in $2 + 1$ Dimensions with Sign-Optimized Manifolds, Phys. Rev. Lett. **121**, 191602 (2018).
[27] F. Bursa and M. Kroyter, Optimisation of complex integration contours at higher order, J. High Energy Phys. 04 (2021) 181.
[28] K. Kashiwa and Y. Mori, Path optimization for $U(1)$ gauge theory with complexified parameters, Phys. Rev. D **102**, 054519, 2020.
[29] M. Giordano, K. Kapas, S. D. Katz, A. Pasztor, and Z. Tulipant, Exponential reduction of the sign problem at finite density in the $2 + 1$D XY model via contour deformations, Phys. Rev. D **106**, 054512, 2022.
[30] M. A. Stephanov, Random Matrix Model of QCD at Finite Density and the Nature of the Quenched Limit, Phys. Rev. Lett. **76**, 4472 (1996).
[31] G. Parisi, On complex probabilities, Phys. Lett. **131B**, 393 (1983).
[32] G. Aarts, E. Seiler, and I.-O. Stamatescu, The complex Langevin method: When can it be trusted?, Phys. Rev. D **81**, 054508 (2010).
[33] E. Seiler, D. Sexty, and I.-O. Stamatescu, Gauge cooling in complex Langevin for QCD with heavy quarks, Phys. Lett. B **723**, 213 (2013).
[34] J. Bloch, J. Glesaaen, J. J. M. Verbaarschot, and S. Zafeiropoulos, Complex Langevin simulation of a random matrix model at nonzero chemical potential, J. High Energy Phys. 03 (2018) 015.
[35] M. Fukuma, N. Matsumoto, and Y. Namekawa, Numerical sign problem and the tempered Lefschetz thimble method, Proc. Sci., CORFU2021 (**2022**) 254.
[36] R. H. Swendsen and J.-S. Wang, Replica Monte Carlo Simulation of Spin-Glasses, Phys. Rev. Lett. **57**, 2607 (1986).
[37] M. Fukuma and N. Umeda, Parallel tempering algorithm for integration over Lefschetz thimbles, Prog. Theor. Exp. Phys. **2017**, 073B01 (2017).
[38] J. Han and M. A. Stephanov, A random matrix study of the QCD sign problem, Phys. Rev. D **78**, 054507 (2008).
[39] A. Alexandru, G. Basar, P. F. Bedaque, G. W. Ridgway, and N. C. Warrington, Fast estimator of Jacobians in the Monte Carlo integration on Lefschetz thimbles, Phys. Rev. D **93**, 094514 (2016).
[40] M. D. Zeiler, ADADELTA: An adaptive learning rate method, arXiv:1212.5701.
[41] Y. Tanizaki, H. Nishimura, and J. J. M. Verbaarschot, Gradient flows without blow-up for Lefschetz thimbles, J. High Energy Phys. 10 (2017) 100.