Path integral contour deformations for observables in $SU(N)$ gauge theory

William Detmold,[1,2] Gurtej Kanwar,[1,2] Henry Lamm,[3] Michael L. Wagman,[3] and Neill C. Warrington[4]

[1]*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

[2]*The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, Cambridge, Massachusetts 02139, USA*

[3]*Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA*

[4]*Institute for Nuclear Theory, University of Washington, Seattle, Washington 98195-1550*

(Received 23 February 2021; accepted 15 April 2021; published 24 May 2021)

Path integral contour deformations have been shown to mitigate sign and signal-to-noise problems associated with phase fluctuations in lattice field theories. We define a family of contour deformations applicable to $SU(N)$ lattice gauge theory that can reduce sign and signal-to-noise problems associated with complex actions and complex observables. For observables, these contours can be used to define deformed observables with identical expectation value but different variance. As a proof-of-principle, we apply machine learning techniques to optimize the deformed observables associated with Wilson loops in two dimensional $SU(2)$ and $SU(3)$ gauge theory. We study loops consisting of up to 64 plaquettes and achieve variance reduction of up to 4 orders of magnitude.

DOI: 10.1103/PhysRevD.103.094517

I. INTRODUCTION

In order to test the Standard Model and search for new physics in experiments involving hadrons and nuclei, precision calculations of Standard Model observables are required. To achieve this, Monte Carlo (MC) calculations of lattice-regularized path integrals for quantum chromodynamics (QCD) have been used to make precise predictions for many phenomenologically relevant quantities in the meson and nucleon sectors; for recent reviews see Refs. [1–5]. Lattice QCD calculations using MC methods are performed in Euclidean spacetime where the action $S(U)$ is typically a real function of the discretized gauge field $U_{x,\mu} \in SU(3)$ and an ensemble of gauge fields can be generated with probability distribution proportional to $e^{-S(U)}$. Predictions for the expectation values of observables $\mathcal{O}$ are then obtained by averaging the results for $\mathcal{O}(U)$ obtained for each gauge field configuration.

For correlation functions describing nucleons, nuclei, and most mesons, $\mathcal{O}(U)$ is complex and includes a gauge-field-dependent phase [6,7]. Phase fluctuations grow more rapid with increasing Euclidean time separation and lead to a "signal-to-noise (StN) problem" in which the StN for a fixed size statistical ensemble decreases exponentially with increasing time separation [6–12]. Alternatively, multinucleon systems can be probed by including nonzero quark chemical potential in the action. In this case, $e^{-S(U)}$ is not real and positive definite and cannot be interpreted as a probability distribution and the theory is said to have a "sign problem." If $e^{-\mathrm{Re}S(U)}$ is used to define a probability distribution in this case, $e^{-i\mathrm{Im}S(U)}$ leads to rapid phase fluctuations of path integrands and the appearance of a StN problem that is exponential in the spacetime volume [13–19].

A generic method for taming sign and StN problems in path integrals has recently emerged. This method involves deforming the manifold of integration of the path integral into a complexified field space. If the path integrand is a holomorphic function of the field variables, then a multidimensional version of Cauchy's integral theorem ensures that expectation values of the corresponding observable is unchanged by the manifold deformation. Manifold deformation may, however, change the values of observables on individual field configurations and therefore modify the severity of phase fluctuations and associated sign/StN problems. Several methods for finding useful manifolds have been proposed and successfully applied in lattice field theories as well as nonrelativistic quantum mechanical theories relevant for condensed matter physics [20–46]. For a recent review see Ref. [47]. Although most applications have focused on improving sign problems in theories with complex actions, an analogous method for improving StN problems for complex observables in theories with real actions was introduced in Ref. [48].

The focus of this work is on addressing sign/StN problems in QCD-like theories; in this setting, path integral contour deformations have so far been restricted to solving sign problems in simple contexts. Thermodynamic observables at nonzero quark chemical potential have been computed in $(0+1)$D QCD, a theory with a single link variable, using Lefschetz thimble methods [49] and the generalized thimble method [50], while preliminary results using neural-network manifolds were obtained in [51]. In higher dimensions, Lefschetz thimbles were used to analyze finite-density observables for one- and two-site systems in the heavy-dense limit in Ref. [52], and this study predicted that the number of relevant Lefschetz thimbles would grow exponentially with the number of lattice sites for larger systems. Very recently a method for computing in Yang-Mills theories with complex gauge couplings was proposed then applied to $(1+1)D$ $U(1)$ gauge theory [53]. The task of computing noisy observables in non-Abelian lattice gauge theories for larger spacetime volumes and higher spacetime dimensions has remained challenging.

This paper introduces a simple yet expressive family of complexified manifolds for taming sign/StN problems in $SU(N)$ gauge theory using path integral contour deformations. The Jacobians required for calculations using this family of deformations are shown to be triangular matrices whose determinants can be computed with a cost that scales linearly with the spacetime volume. This family of manifolds can be applied to all theories involving $SU(N)$ gauge fields, including gauge theories coupled to matter fields, although their practical utility for improving sign/StN problems is only explored here for pure gauge theory.

The deformed observable method introduced in Ref. [48] relates path integrals over deformed contours to path integrals written in terms of modified observables on undeformed contours, enabling improvement in the StN of observables without the need to modify MC sampling. We apply the method here to calculations of Wilson loops in $SU(2)$ and $SU(3)$ gauge theory, in which Wilson loops are known to have an exponentially severe StN problem and have been used to study other StN improvement methods [54]. Calculations are performed in $(1+1)$D as a proof of concept, as it is possible to compare with exact StN results derived analytically and to use specialized approaches for efficient Monte Carlo ensemble generation for $(1+1)$D gauge theories. Results are obtained for a range of Wilson loop areas and lattice spacings including areas of up to 64 lattice units at the finest lattice spacing. The variances of Wilson loops with largest areas are reduced by factors of $10^3 - 10^4$, demonstrating that deformed observables can dramatically improve StN problems in $SU(N)$ lattice gauge theory. The linear scaling with spacetime volume of these contour deformations suggests that it should be computationally feasible to explore the application of analogous contour deformations to $(3+1)$D lattice gauge theory in future work.

The remainder of this paper is organized as follows. Section II describes our approach to contour deformations for $SU(N)$ variables, including a family of complex manifolds for integration over sets of $SU(N)$ variables, and reviews the deformed observables method introduced in Ref. [48]. Section III presents analytical results for expectation values and variances of observables in $(1+1)$D $SU(N)$ lattice gauge theory. Results for MC calculations of deformed observables for Wilson loops are presented for $SU(2)$ gauge theory in Sec. IV and for $SU(3)$ gauge theory in Sec. V. A summary of results and consideration of future work is found in Sec. VI.

II. GENERAL FORMALISM

Cauchy's integral theorem implies that the contour of a complex line integral can be deformed without changing the integral value if the integrand is holomorphic in the intervening region and the endpoints are held fixed.[1] When multidimensional integration is performed, the full theorem can be generalized if the integral is describable as iterated complex line integrals or by a technical extension to the full multivariate setting [55]. For the purpose of contour deformations, however, only a weaker form of the theorem (equivalent to Stokes' theorem) is required. Specifically, a contour deformation from manifold $\mathcal{M}_A$ to $\mathcal{M}_B$ leaves the integral value unchanged if $\mathcal{M}_A \cup \mathcal{M}_B$ bounds a region in which the integrand is holomorphic; see Ref. [47] for a simple proof. To implement such contour deformations and confirm holomorphy of an integrand throughout the relevant region of configuration space, a coordinate parametrization is useful. We discuss such parametrizations and contour deformation for $SU(N)$ groups and $SU(N)$ gauge theory in the following sections.

A. Contour deformations of angular parameters

A general formalism for applying path integral contour deformations to $SU(N)$ group integrals can be obtained by using manifold coordinates that map subsets of $\mathbb{R}^{N^2-1}$ to $SU(N)$. For any N, the group manifold can be given explicit global coordinates using N^2-1 angular variables [56]. These variables can be divided into *azimuthal* angles $\phi_1,\ldots,\phi_J \in [0,2\pi]$ and *zenith* angles $\theta_1,\ldots,\theta_K \in [0,\pi/2]$, where $J = (N^2+N-2)/2$ and $K = (N^2-N)/2$.[2] The azimuthal angles are periodic, such that $\phi_i = 0$ is identified with $\phi_i = 2\pi$, while the zenith angles have distinct endpoints. We define the combined coordinate $\Omega \equiv (\phi_1,\ldots,\phi_J,\theta_1,\ldots,\theta_K)$.

[1] For periodic functions this condition on the endpoints can be relaxed, as discussed in Sec. II A.

[2] This is not the only possible assignment of angular coordinates to the manifold. For example, Appendix B explores an alternative parametrization for $SU(2)$.

A generic integral over group-valued variable $U \in SU(N)$ can be written as

$$\mathcal{I} = \int dU f(U), \tag{1}$$

where the Haar measure dU is defined to be the unique measure that satisfies $d(VU) = d(UV) = dU$ for $V \in SU(N)$. We choose the conventional normalization condition $\int dU = 1$. Using the coordinates introduced above, the integral can immediately be cast as integration over a subdomain of $\mathbb{R}^{N^2-1}$,

$$\mathcal{I} = \prod_{j=1}^{J}\left[\int_0^{2\pi} d\phi_j\right]\prod_{k=1}^{K}\left[\int_0^{\pi/2} d\theta_k\right] h(\Omega) f(U(\Omega)), \tag{2}$$

where $h(\Omega)$ is the Jacobian factor associated with the change of measure from dU to $d\Omega = \prod_j d\phi_j \prod_k d\theta_k$, and $U(\Omega)$ is the group element at the manifold coordinate Ω. The specific forms of $h(\Omega)$ and $U(\Omega)$ for the groups $SU(2)$ and $SU(3)$ are presented in Sec. IV A and V A.

From Eq. (2), it is clear that each θ_k can be deformed into the complex plane holding the endpoints 0 and $\pi/2$ fixed. Each ϕ_j can be deformed under the weaker constraint that the endpoints remain identified, as shown in Fig. 1. This can be seen by noting that the endpoints of the shifted contour can be connected to the original endpoints using a pair of segments parallel to the imaginary axis; these segments differ by a real 2π shift and a change of orientation so that integrals of periodic functions along these segments exactly cancel. This approach to deforming periodic variables with identified endpoints has previously been applied to path integrals involving $U(1)$ variables [48,57–59]. If the integrand $h(\Omega)f(U(\Omega))$ is a holomorphic function of all components of Ω, the value of the integral is unchanged by the deformations described above. We can define a deformed integration contour by the maps $\tilde{\phi}_j = \tilde{\phi}_j(\Omega) \in \mathbb{C}$ and $\tilde{\theta}_i = \tilde{\theta}_i(\Omega) \in \mathbb{C}$; for conciseness the collective set of deformed coordinates can be written as a function of original coordinates $\tilde{\Omega} \equiv (\tilde{\phi}_1, \ldots, \tilde{\phi}_J, \tilde{\theta}_1, \ldots, \tilde{\theta}_K) = \tilde{\Omega}(\Omega)$. The value of the integral is unchanged by this deformation,

$$\begin{aligned}\mathcal{I} &= \prod_{j=1}^{J}\left[\int_0^{2\pi} d\phi_j\right]\prod_{k=1}^{K}\left[\int_0^{\pi/2} d\theta_k\right] J(\Omega) h(\tilde{\Omega}) f(U(\tilde{\Omega})) \\ &= \prod_{j=1}^{J}\left[\int_0^{2\pi} d\phi_j\right]\prod_{k=1}^{K}\left[\int_0^{\pi/2} d\theta_k\right] h(\Omega) \\ &\quad \times \left\{ J(\Omega)\frac{h(\tilde{\Omega})}{h(\Omega)} f(U(\tilde{\Omega}))\right\} \\ &= \int dU J(U) f(\tilde{U}).\end{aligned} \tag{3}$$

Above, $\tilde{U} \equiv U(\tilde{\Omega})$ is the deformed group element, $J(\Omega) \equiv \det \frac{\partial \tilde{\Omega}_\alpha}{\partial \Omega_\beta}$ is the (complex) Jacobian factor relating the measure $d\Omega$ of the base contour to the measure $d\tilde{\Omega}$ of the deformed contour, and the Jacobian factor relating the Haar measure dU between the original and deformed contours is given by

$$J(U) \equiv J(\Omega) h(\tilde{\Omega})/h(\Omega). \tag{4}$$

For any concrete map $U(\Omega)$, the deformed group element is given by simply applying the map to the complexified coordinate $\tilde{\Omega}$. Any real Lie group has a unique complexification and the space in which $\tilde{U}$ lives is well understood. In particular, the complexification of $SU(N)$ is $SL(N, \mathbb{C})$ [60]. This is easy to see on intuitive grounds, as $SU(N)$ matrices are specified by unit-norm eigenvalues with determinant 1 and the effect of complexifying the group is to allow arbitrary nonzero eigenvalues while preserving the determinant constraint, resulting in the group $SL(N, \mathbb{C})$.

Though the deformation is defined in terms of the manifold coordinates, we can see in the last line of Eq. (3) that this new integral can be written independently of the coordinates, using the modified integrand $J(U)f(\tilde{U}(U))$. This form suggests a coordinate-independent definition of a holomorphic integrand and contour deformation. Such a general approach is beyond the scope of this work, but we note that a deformation can be applied in any coordinate system so long as the integrand can be shown to be holomorphic in *some* coordinate system. The angular coordinates for $SU(N)$ are sufficient to show that all components of the matrix representation of U are holomorphic (i.e., the Wirtinger derivatives $\partial U(\Omega)/\partial \bar{\Omega}_i$ are zero [55]). The components of U^{-1} are holomorphic whenever U is invertible, as can be seen by starting with the definition of the inverse, $U^{-1}U = 1$, applying Wirtinger

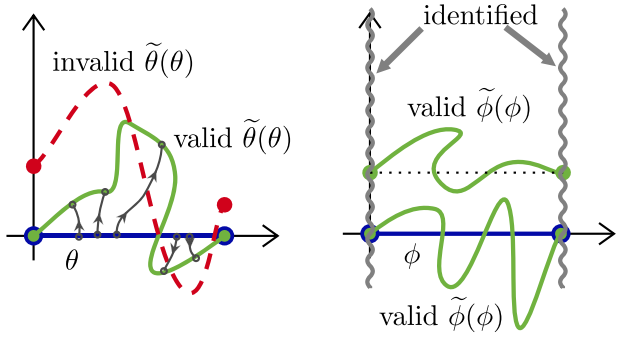

FIG. 1. Left: schematic depiction of valid and invalid contour deformations, defined by the mapping $\tilde{\theta}(\theta)$ from base coordinates to the manifold, when the original domain is a finite interval. Right: schematic depiction of additional allowed deformations (shifts) when endpoints are identified; these shifts are applicable to $U(1)$ variables or azimuthal angles ϕ in $SU(N)$ manifolds.

derivatives to both sides, $\partial(U^{-1}U)/\bar{\Omega}_i = 0$, and using holomorphy of U to obtain $\partial U^{-1}/\partial\bar{\Omega}_i = 0$.

The general construction so far is valid for collections of $SU(N)$ group variables, and is therefore applicable to $SU(N)$ lattice gauge theory. Standard pure-gauge actions for $SU(N)$ lattice gauge theory can be written as holomorphic functions of the variables U and their inverses, if we replace instances of $U^\dagger$ with U^{-1}, and are thus holomorphic throughout the $SL(N,\mathbb{C})$ domain of each complexified group variable. This fact has previously been recognized in complex Langevin approaches to extensive sign problems in lattice gauge theory [61–63]. Similarly, (unrooted) fermionic determinants can be written as polynomials in components of gauge links and are therefore holomorphic [36], and many observables can be analyzed in the same way.

B. Vertical contour deformations

We define a family of deformed manifolds for an $SU(N)$ variable in terms of the angular parametrization by

$$\tilde{\Omega} = \Omega + if(\Omega;\lambda,\chi), \tag{5}$$

where $f(\Omega;\lambda,\chi) \in \mathbb{R}^{N^2-1}$. This family of *vertical deformations* inspired by Refs. [57,64] is not fully general as it only includes contour deformations describable by vertically shifting each point purely in the imaginary direction, and in particular it does not include any deformations with $\mathrm{Re}\tilde{\Omega}(\Omega) = \mathrm{Re}\tilde{\Omega}(\Omega')$ for some $\Omega \neq \Omega'$. The Jacobian of any deformation in this family is straightforward to compute as

$$J(\Omega) = \det\frac{\partial\tilde{\Omega}_\alpha}{\partial\Omega_\beta} = \det\left[\delta_{\alpha\beta} + i\frac{\partial f_\alpha(\Omega;\lambda,\chi)}{\partial\Omega_\beta}\right], \tag{6}$$

where $\alpha,\beta = 1,\ldots,N^2-1$ index the angular parameters $\Omega = (\phi_1,\ldots,\phi_J,\theta_1,\ldots,\theta_K)$.

The function f can further be expanded in terms of Fourier modes,

$$f(\Omega;\lambda,\chi) = \sum_{\{n_i\}=0}^{\Lambda}\sum_{\{m_j\}=1}^{\Lambda}\lambda_I T_I(\Omega;\chi_I^1,\ldots,\chi_I^a), \tag{7}$$

where $I \equiv (n_1\ldots n_a, m_1\ldots m_b)$, and

$$T_I(\Omega;\chi^1\ldots\chi^a) \equiv \prod_{i=1}^{a}\sin(\phi_i n_i + \chi^i)\prod_{j=1}^{b}\sin(2\theta_j m_j), \tag{8}$$

which provides a complete basis for vertical contour deformations in the limit $\Lambda\to\infty$. Including successively more Fourier modes by increasing the Fourier cutoff Λ systematically improves the flexibility of the function f. In our applications to $SU(N)$ gauge theories in $(1+1)$D below, Fourier cutoffs $\Lambda \in \{0,1,2\}$ are explored. It is noteworthy that the sum over azimuthal Fourier modes includes the constant mode as well as phase offsets χ_i because azimuthal angles can be deformed without fixing their endpoints as discussed above. These constant modes are essential for the sign/StN problem reduction achieved in $(1+1)$D examples below.

C. Path integral deformations for noisy observables

We next review the deformed observable approach presented in Ref. [48], in which contour deformations of lattice field theory path integrals are used to define modified observables with improved noise properties and unchanged expectation value. We focus here on $SU(N)$ lattice gauge theory path integrals, which are high-dimensional integrals over a collection of group-valued degrees of freedom $U_{x,\mu} \in SU(N)$. Here, x specifies a site on the discrete spacetime lattice and $\mu \in \{1,\ldots,N_d\}$ is any of the N_d spacetime directions on the lattice. The integrals under study take the form

$$\langle\mathcal{O}\rangle \equiv \frac{1}{Z}\int\mathcal{D}U\,\mathcal{O}(U)e^{-S(U)}, \tag{9}$$

where

$$Z = \int\mathcal{D}U\,e^{-S(U)} \tag{10}$$

and the action $S(U) \in \mathbb{R}$ is a function of all gauge links. Details on the construction of lattice gauge theory are presented in Sec. III.

It is possible to deform the integration contour of an $SU(N)$ lattice gauge theory path integral by individually deforming each group-valued variable $U_{x,\mu}$ using the formalism presented above. In principle the deformed link, $\tilde{U}_{x,\mu}$, could be a function of all other links on the lattice. However, evaluating the Jacobian factor arising from such an arbitrary deformation would require $O(V^3)$ operations, where V is the number of sites of the lattice. For state-of-the-art lattice field theory calculations, this is intractable. A similar obstacle is encountered in the application of normalizing flows to sampling probability distributions in image or lattice data analysis, see, e.g., Ref. [65] for a review, where it is avoided by explicitly restricting to triangular Jacobians for which the determinant factor is efficiently calculable from the diagonal elements. In this work, we similarly restrict to triangular Jacobians by allowing deformations of each variable to depend only on previous variables in a canonical ordering, described in detail for our particular studies of $SU(2)$ and $SU(3)$ in the following sections. Exploring other options is an interesting possibility for future work; for example, transformations built from a composition of multiple triangular transformations may allow more general spacetime dependence without significant increase in cost, whereas other

alternatives based on convolutions may scale supra-linearly as the volume is increased.

Given a deformed manifold $\mathcal{M}$ with tractable Jacobian factor, a deformed integral can be constructed to compute the same expectation value defined by Eq. (9),

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int_{\mathcal{M}} \mathcal{D}\tilde{U}\, \mathcal{O}(\tilde{U}) e^{-S(\tilde{U})}. \tag{11}$$

The above equality holds if the original manifold can be deformed to $\mathcal{M}$ without encountering any nonholomorphic regions of the integrand, i.e., if the integrand is holomorphic in the region bounded by the union of $\mathcal{M}$ and the original manifold. The abstract manifold $\mathcal{M}$ can be specified by a map $\tilde{U}(U)$, which then gives a concrete prescription for computing Eq. (11),

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathcal{D}U J(U) \mathcal{O}(\tilde{U}(U)) e^{-S(\tilde{U}(U))}. \tag{12}$$

In contrast to Eq. (9), the path integral in Eq. (12) involves a generically complex-valued action, $S(\tilde{U}(U)) \in \mathbb{C}$, and a different observable $J(U)\mathcal{O}(\tilde{U}(U))$ written in terms of the (efficiently computed) Jacobian factor $J(U)$. Cauchy's theorem implies that these modifications conspire to cancel and result in an identical expectation value.

It is useful to further rewrite Eq. (12) as a path integral with respect to the original action,

$$\begin{aligned} \langle \mathcal{O} \rangle &= \frac{1}{Z} \int \mathcal{D}U \{ J(U) \mathcal{O}(\tilde{U}(U)) e^{-S(\tilde{U}(U)) + S(U)} \} e^{-S(U)} \\ &\equiv \frac{1}{Z} \int \mathcal{D}U\, \mathcal{Q}(U) e^{-S(U)} = \langle \mathcal{Q} \rangle. \end{aligned} \tag{13}$$

In this rewriting, it is clear that the deformed path integral is still accessible by performing Monte Carlo sampling with respect to the original action $S(U)$ and estimating the sample mean of $\mathcal{Q}(U)$. These methods can therefore be applied at the measurement step, after an ensemble of field configurations has been sampled. In essence, the deformed path integral defines a new observable $\mathcal{Q}(U)$ that has provably identical expectation value to the original observable $\mathcal{O}(U)$. Notably, this new observable generically has very different structure from $\mathcal{O}$, as it may be nonlocal and depends on the structure of the action.

The variance of $\mathcal{Q}(U)$ can, however, be vastly different from the variance of $\mathcal{O}(U)$. For most observables, samples of $\mathcal{O}(U)$ are complex-valued, and the variance of the real and imaginary components are given by

$$\begin{aligned} \mathrm{Var}[\mathrm{Re}\,\mathcal{O}] &= \frac{1}{2}\langle |\mathcal{O}^2| \rangle + \frac{1}{2}\mathrm{Re}\langle \mathcal{O}^2 \rangle - [\mathrm{Re}\langle \mathcal{O} \rangle]^2, \\ \mathrm{Var}[\mathrm{Im}\,\mathcal{O}] &= \frac{1}{2}\langle |\mathcal{O}^2| \rangle - \frac{1}{2}\mathrm{Re}\langle \mathcal{O}^2 \rangle - [\mathrm{Im}\langle \mathcal{O} \rangle]^2. \end{aligned} \tag{14}$$

These are not generically identical to the variance of $\mathrm{Re}\mathcal{Q}$ and $\mathrm{Im}\mathcal{Q}$,

$$\begin{aligned} \mathrm{Var}[\mathrm{Re}\,\mathcal{Q}] &= \frac{1}{2}\langle |\mathcal{Q}^2| \rangle + \frac{1}{2}\mathrm{Re}\langle \mathcal{Q}^2 \rangle - [\mathrm{Re}\langle \mathcal{Q} \rangle]^2, \\ \mathrm{Var}[\mathrm{Im}\,\mathcal{Q}] &= \frac{1}{2}\langle |\mathcal{Q}^2| \rangle - \frac{1}{2}\mathrm{Re}\langle \mathcal{Q}^2 \rangle - [\mathrm{Im}\langle \mathcal{Q} \rangle]^2. \end{aligned} \tag{15}$$

The final term is unchanged because $\langle \mathcal{O} \rangle = \langle \mathcal{Q} \rangle$, but the first two terms in both lines of Eq. (15) are not generically equal to the corresponding terms in Eq. (14). Explicitly, those terms are given by

$$\begin{aligned} \langle |\mathcal{Q}^2| \rangle &= \frac{1}{Z} \int \mathcal{D}U |J(U)\mathcal{O}(\tilde{U}(U))|^2 e^{-2\,\mathrm{Re}\,S(\tilde{U}(U)) + S(U)}, \\ \langle \mathcal{Q}^2 \rangle &= \frac{1}{Z} \int \mathcal{D}U J(U)^2 \mathcal{O}(\tilde{U}(U))^2 e^{-2S(\tilde{U}(U)) + S(U)}. \end{aligned} \tag{16}$$

Extra factors of $S(U)$ persist in both expressions in Eq. (16), and a nonholomorphic absolute value appears for $\langle |\mathcal{Q}^2| \rangle$, preventing identification with the terms in Eq. (14). It can thus be fruitful to look for a modified observable $\mathcal{Q}$ for which the terms in Eq. (16) are minimized and the statistical noise is less than that of the original observable.

Given an explicit parametrization of the deformed contour, standard gradient-based optimization methods can be applied to find the parameters that minimize the terms in Eq. (16). Since the parameters only affect the observable itself (the sampling weight is always $e^{-S(U)}$), the gradient of the variance with respect to the parameters can be written as an expectation value under the original Monte Carlo sampling. In this work, minimization of Eq. (16) is performed using stochastic gradient descent over the deformed contour parameters, which approximately converges to a local minimum of the variance under mild assumptions [66–69], and starting at the original manifold ensures that the result improves relative to (or at worst is equivalent to) the original variance.

A potential obstacle to the deformed observable method is that some contour deformations could lead to a severe overlap problem between probability distributions proportional to $e^{-S(U)}$ and $e^{-\mathrm{Re}[S(\tilde{U}(U))]}$ that could make estimates of deformed observable variances from finite Monte Carlo ensembles unreliable. The possibility of constructing deformed observables with severe overlap problems can be mitigated, however, by constructing deformed observables that minimize the variance of $\mathcal{Q}$ since large fluctuations of $e^{-S(\tilde{U}(U)) + S(U)}$ will tend to increase the variance of $\mathcal{Q}$. Overlap problems could still arise due to overfitting to the sample variance of a Monte Carlo ensemble that does not sample the field configurations associated with large fluctuations of $e^{-S(\tilde{U}(U)) + S(U)}$; practical strategies to avoid such overfitting problems are discussed in Sec. IV C below.

The terms to minimize in Eq. (16) are specific to a particular observable $\mathcal{O}$. There is no reason to expect a

single manifold deformation to be optimal for all possible observables, but one does expect similar observables to be highly correlated, and thus to receive similar variance improvements from the same deformation. This suggests two useful practical improvements:

(1) Optimal manifolds can be found for a few representative observables and can be reused for other similar observables.
(2) When optimizing manifold parameters, the optimal parameters for a similar observable can be used to initialize the search.

In our studies of $SU(2)$ and $SU(3)$ lattice gauge theory in Sec. IV and Sec. V, the initialization approach was found to significantly reduce the number of steps required for optimization.

D. Rewriting observables before deformation

It is often the case that the expectation value of an observable $\mathcal{O}$ has multiple equivalent path integral representations, for example when a theory possesses a gauge or global symmetry that modifies $\mathcal{O}$ but leaves the action and expectation values of observables invariant. In addition to the freedom of choosing the parametrization of contour deformations for a given path integral discussed above, the application of contour deformations to observables includes the freedom of choosing which path integral representation to deform.

Simple examples of this freedom arise in two-dimensional $U(1)$ Euclidean lattice gauge theory. For instance, path integral representations of Wilson loops with unit area in this theory are proportional to

$$\int_{-\pi}^{\pi}\frac{d\phi}{2\pi}e^{i\phi}e^{\beta\cos\phi}=I_1(\beta), \tag{17}$$

which is an integral representation of the modified Bessel function of the first kind $I_n(\beta)$, with $n=1$, written in terms of the field variable ϕ. The integrand appearing can be interpreted as a product of an observable $\mathcal{O}=e^{i\phi}$ and an e^{-S} factor for the Euclidean action $S=-\beta\cos\phi$. This theory has a charge conjugation symmetry that acts on the integrand of Eq. (17) by $\phi\to-\phi$, which leaves the action invariant but modifies the observable $e^{i\phi}\to e^{-i\phi}$. An equally valid integral representation of $I_1(\beta)$ is obtained by averaging the observable and its charge conjugate as

$$I_1(\beta)=\int_{-\pi}^{\pi}\frac{d\phi}{2\pi}\cos(\phi)e^{\beta\cos\phi}. \tag{18}$$

The choice of integrand in Eq. (17) or Eq. (18) is irrelevant for Monte Carlo calculations using the integration contours shown because the Monte Carlo estimator used in the first case is $\mathrm{Re}e^{i\phi}=\cos\phi$. However, the ability of path integral contour deformations to reduce the variance of a Monte Carlo calculation does depend on the representation. In particular, the variance of a Monte Carlo evaluation of Eq. (17) can be significantly reduced by contour deformations while the variance of a Monte Carlo evaluation of Eq. (18) cannot, as discussed below and shown in Fig. 2.

Reference [48] demonstrated that the variance of two-dimensional $U(1)$ Wilson loops can be significantly reduced using contour deformations of the representation given in Eq. (17); we review the analytical derivation here. Denote averaging of $\mathcal{O}(\phi)$ with respect to the path integral of the theory with action $-\beta\cos\phi$ by

$$\langle\mathcal{O}(\phi)\rangle_\beta\equiv\int_{-\pi}^{\pi}\frac{d\phi}{2\pi I_0(\beta)}\mathcal{O}(\phi)e^{\beta\cos\phi}, \tag{19}$$

where the modified Bessel function with rank $n=0$ is used to normalize the distribution. In general, the modified Bessel functions of the first kind are given by $I_n(\beta)=\int_{-\pi}^{\pi}\frac{d\phi}{2\pi}e^{in\phi}e^{\beta\cos\phi}$ and will be used throughout the following

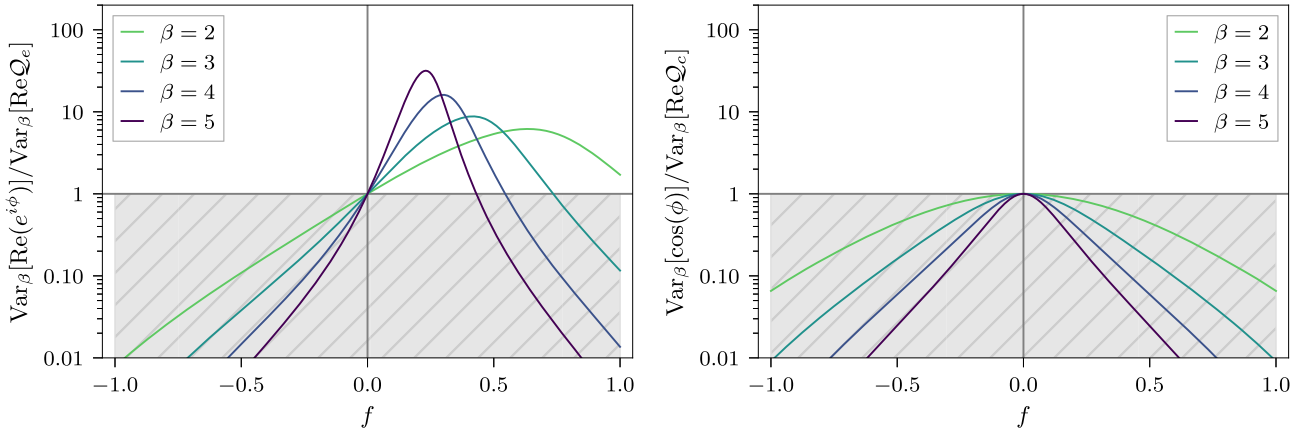

FIG. 2. Ratios of the variance of observable $\mathrm{Re}(e^{i\phi})=\cos(\phi)$ to the variance of deformed observables obtained by applying the transformation $\phi\to\phi+if$ to the integrand $e^{i\phi}$ (left) or $\cos(\phi)$ (right) in the one-dimensional path integral with Euclidean action $-\beta\cos\phi$ discussed in the main text. Gray hatching indicates the region in which the variance of the deformed observable is higher than the original (i.e., there is no improvement).

derivation. Further denoting the corresponding variance of $\mathcal{O}$ given β by $\mathrm{Var}_\beta[\mathcal{O}]$, the variance of $\mathrm{Re}\, e^{i\phi} = \cos\phi$ can be computed to be

$$\begin{aligned}\mathrm{Var}_\beta[\mathrm{Re}\, e^{i\phi}] &= \langle \cos^2(\phi)\rangle_\beta - \langle e^{i\phi}\rangle_\beta^2 \\ &= \frac{1}{2}\left[1 + \frac{I_2(\beta)}{I_0(\beta)}\right] - \left[\frac{I_1(\beta)}{I_0(\beta)}\right]^2. \end{aligned} \tag{20}$$

The constant vertical deformation

$$\phi \to \tilde{\phi}(\phi) = \phi + if \tag{21}$$

leads to a deformed observable

$$\mathcal{Q}_e(\phi) = e^{i\tilde{\phi}(\phi)} e^{\beta\cos(\tilde{\phi}(\phi)) - \beta\cos(\phi)} \tag{22}$$

satisfying $\langle \mathcal{Q}_e\rangle_\beta = \langle e^{i\phi}\rangle_\beta$. Using $\mathrm{Re}\mathcal{Q}_e$ as a Monte Carlo estimator instead of $\mathrm{Re}\, e^{i\phi}$ results in variance

$$\mathrm{Var}_\beta[\mathrm{Re}\mathcal{Q}_e] = e^{-2f}\left[\frac{R(\beta,f)}{2} + V(\beta,f)\right] - \left[\frac{I_1(\beta)}{I_0(\beta)}\right]^2, \tag{23}$$

where

$$\begin{aligned} V(\beta,f) &= \left(\frac{e^f - \frac{1}{2}}{e^{-f} - \frac{1}{2}}\right)\frac{I_2(\beta\sqrt{5 - 4\cosh(f)})}{2I_0(\beta)}, \\ R(\beta,f) &= \frac{I_0(\beta(2\cosh(f) - 1))}{I_0(\beta)}. \end{aligned} \tag{24}$$

For positive β, $\mathrm{Var}_\beta[\mathrm{Re}\mathcal{Q}_e]$ generically has a minimum at some strictly positive f that defines the optimal contour for reducing the variance of $\mathcal{Q}_e$. Using deformed observables associated with this optimal contour rather than the original contour leads to variance reduction whose significance is larger than an order of magnitude for $\beta \gtrsim 4$ as shown in Fig. 2. Larger variance reductions can be obtained for multidimensional generalizations of Eq. (17) associated with larger $U(1)$ Wilson loops [48].

Applying the same constant vertical deformation to the representation in Eq. (18) leads to an alternative deformed observable

$$\mathcal{Q}_c(\phi) = \cos(\tilde{\phi}(\phi)) e^{\beta\cos(\tilde{\phi}(\phi)) - \beta\cos(\phi)} \tag{25}$$

satisfying $\langle \mathcal{Q}_c\rangle_\beta = \langle e^{i\phi}\rangle_\beta$. The real part of this alternative deformed observable has variance

$$\begin{aligned}\mathrm{Var}_\beta[\mathrm{Re}\mathcal{Q}_c] = &\frac{1}{2}e^{-2f}V(\beta,f) + \frac{1}{2}e^{2f}V(\beta,-f) \\ &+ \frac{1}{2}R(\beta,f) - \left[\frac{I_1(\beta)}{I_0(\beta)}\right]^2. \end{aligned} \tag{26}$$

This expression is symmetric under $f \to -f$ and its gradient with respect to f therefore vanishes at $f = 0$, which can be verified to be the minimum of f. The deformed observable $\mathcal{Q}_c$ associated with integrating along the real axis is thus the choice with lowest variance, while increasing $|f|$ away from zero always leads to increased variance as illustrated for several choices of β in Fig. 2.

The qualitative features of these results can be understood from the behaviors of magnitude and phase fluctuations of deformed observables. The magnitude of $e^{i\tilde{\phi}}$ is reduced for all ϕ by a constant vertical deformation with $f > 0$. To preserve the results of the holomorphic integral in Eq. (17) under this deformation, cancellations from phase fluctuations must correspondingly be less severe and sign/StN problems should therefore be improved. On the other hand, the magnitude of $\cos(\tilde{\phi})$ always satisfies $|\cos(\tilde{\phi})| \geq |\cos(\phi)|$ and therefore can only be increased by applying a vertical contour deformation. This suggests that phase fluctuations must become more severe to preserve deformation-invariant integral results and that sign/StN problems will be worsened. This argument applies to nonconstant vertical deformations and suggests that it is generically difficult to construct a contour deformation of Eq. (18) that leads to a deformed observable with variance comparable to the observable $\mathcal{Q}_e$ obtained by deforming Eq. (17).

In the non-Abelian gauge theories that are focus of this work, the traces of Wilson loops define gauge invariant observables related to the potential energies of static quark configurations analogous to $e^{i\phi}$ in $U(1)$ gauge theory. The eigenvalues of Wilson loops in $U(N)$ or $SU(N)$ gauge theories can be expressed as $e^{i\phi_j}$, where $\phi_j \in \mathbb{R}$ for $j = 1, \ldots, N$ and for the case of $SU(N)$ the phases satisfy $\sum_{j=1}^N \phi_j = 0 \bmod 2\pi$. The trace of the Wilson loop is given by $\sum_j e^{i\phi_j}$. For $SU(N)$ gauge theory, the unit determinant condition therefore results in analogous obstacles to improving the variance of the trace of Wilson loops if the observable is not first rewritten using symmetries.

For the case of $SU(2)$, there is a precise correspondence between Wilson loop traces and Eq. (18). The unit determinant condition requires the Wilson loop eigenvalues to be of the form $e^{i\phi}$ and $e^{-i\phi}$ and the trace appearing in both the observable and the Wilson action [70] (discussed below in Sec. IV) are proportional to $\cos(\phi)$. It is therefore similarly impossible to improve the variance of a unit area $SU(2)$ Wilson loop trace using constant vertical deformations, and the previous analysis suggests it is difficult to make significant variance improvements to traces of $SU(2)$ Wilson loops of general area using vertical contour deformations. However, this analysis also indicates that $e^{i\phi}$ could provide a suitable starting point for defining deformed observables. The gauge invariant component of the Wilson loop eigenvalue $e^{i\phi}$ is $\cos(\phi)$ making this a

suitable rewriting of the observable that does not affect the expectation value but enables variance improvements by contour deformation.[3]

In the case of $SU(N)$ with $N \geq 3$, complex eigenvalues are not guaranteed to come in complex conjugate pairs, but the unit determinant condition still guarantees that any vertical deformation of the eigenvalue phases will increase the magnitude of at least one $e^{i\phi_j}$ phase factor. In the sum over eigenvalues defining the trace of a Wilson loop, the largest eigenvalue magnitude will set the typical observable magnitude on generic gauge fields in the integration domain. Given that one eigenvalue is always boosted in magnitude, on average the deformed integrand has to oscillate faster than the undeformed integrand to ensure an expectation value which is overall unchanged (as guaranteed by Cauchy's theorem). What results is a StN problem even worse than before. If one instead chooses the nongauge-invariant integrand $e^{i\phi_1}$ to measure the same expectation value, the determinant constraint does not prevent exponentially decreasing the magnitude throughout the integration domain using contour deformations. For example, one can choose the shift

$$\begin{aligned}\phi_1 &\to \tilde{\phi}_1 = \phi_1 + if \\ \phi_2 &\to \tilde{\phi}_2 = \phi_2 - if/2 \\ \phi_3 &\to \tilde{\phi}_3 = \phi_3 - if/2\end{aligned} \tag{27}$$

which has the effect of reducing the observable magnitude by e^{-f} on every gauge field in the integration domain.

Rewriting Wilson loop observables based on eigenvalues requires diagonalization, and a more practical alternative is to use the (1,1) component of the (matrix-valued) Wilson loop as another nongauge-invariant function with the same expectation value as the trace divided by N. The phase of this (or any other) single color component of the Wilson loop is not constrained by the unit determinant condition and therefore one expects that a suitable parametrization can be found in which vertical deformations can be applied to the phase of the (1,1) component of the Wilson loop analogously to $e^{i\phi}$. Such parametrizations are given for $SU(2)$ in Sec. IV and for $SU(3)$ in Sec. V following Ref. [56] and are used as starting points for defining deformed observables with reduced variance in calculations of Wilson loop expectation values. An alternative parametrization of $SU(2)$ in which the real part of the (1,1) component of the Wilson loop is expressed as $\cos(\alpha/2)$ is explored in Appendix B; as expected, vertical contour deformations do not improve the variance of unit area Wilson loops and less (though still significant) variance reduction is found for larger area Wilson loops with this alternative parametrization.

[3]We could instead use linearity of the expectation value to start from an explicitly gauge-invariant observable and rewrite $\langle\cos(\phi)\rangle = \frac{1}{2}\langle e^{i\phi}\rangle + \frac{1}{2}\langle e^{-i\phi}\rangle$. Most generally, we could then apply distinct deformations to each expectation value on the right-hand-side of this expression, and in particular applying constant imaginary shifts with opposite signs will result in the same StN improvement for both terms. In this example, the two expectation values can be equated using charge conjugation symmetry, which allows the exact rewriting $\langle\cos(\phi)\rangle = \langle e^{i\phi}\rangle$, but the technique of splitting the expectation value using linearity and applying independent deforms may be useful where such a symmetry is not present.

III. NOISE PROBLEMS IN $SU(N)$ LATTICE GAUGE THEORY

A simple setting for analyzing $SU(N)$ lattice gauge theory is obtained by considering $(1+1)$D Euclidean spacetime with open boundary conditions. In this spacetime geometry, much like in $(3+1)$ dimensions, the theory features confinement of static test charges and an exponentially severe StN problem associated with static quark correlation functions, which can be identified with Wilson loops [70]. Numerical calculations of Wilson loop expectation values can be performed at much lower computational cost in $(1+1)$D than $(3+1)$D, facilitating a first exploration of path integral contour deformations applied to non-Abelian gauge theory observables on nontrivial lattices. Analytic results for $(1+1)$D observables such as Wilson loops are also known [71,72] and can be used to verify the correctness of numerical results. These results are extended to analytic results for the variances of $(1+1)$D Wilson loops below, which are then used in Secs. IV–V to verify the correctness and study the effectiveness of contour deformations applied to Wilson loops. In particular, analytic results can be used to determine the StN gains obtained by using deformed observables even when the corresponding undeformed observables are too noisy to be determined reliably.

A. $SU(N)$ lattice gauge theory in $(1+1)$D

Lattice gauge theory in $(1+1)$D is defined on a set $\mathcal{V}$ of Euclidean spacetime points x arranged in a discrete two-dimensional lattice, with vectors $\hat{1}$ and $\hat{2}$ giving the displacement in lattice units between neighboring lattice sites along the two Euclidean spacetime axes. The discretized gauge field is represented by group-valued variables on each link of the lattice, with $U_{x,\mu}$ denoting the variable associated with link $(x, x+\hat{\mu})$. The physical content of the theory is encoded in the (discretized) action. We consider the Wilson action for $SU(N)$ lattice gauge theory [70], given for a $(1+1)$D Euclidean spacetime volume by

$$S \equiv -\frac{1}{g^2}\sum_{x\in\mathcal{V}} \mathrm{tr}(P_x + P_x^{-1}), \tag{28}$$

where g is the bare gauge coupling and each *plaquette* $P_x \in SU(N)$ is defined as

$$P_x \equiv U_{x,1} U_{x+\hat{1},2} U^{-1}_{x+\hat{2},1} U^{-1}_{x,2}. \tag{29}$$

Writing the action and plaquettes using inversion rather than Hermitian conjugation allows the relevant integrands to be interpreted in the following sections as holomorphic functions of integration variables throughout the complexified domain. For $SU(N)$ elements these operations are equivalent, but analytically continuing the action to $SL(N,\mathbb{C})$ requires the use of the inverse [61–63].

Expectation values of operators $\mathcal{O}(U)$ in the lattice regularized theory are defined by specializing Eqs. (9)–(10) to the particular case of $SU(N)$ lattice gauge theory

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathcal{D}U \, \mathcal{O}(U) e^{-S(U)}, \tag{30}$$

where the Euclidean partition function Z is defined by

$$Z = \int \mathcal{D}U \, e^{-S(U)} \tag{31}$$

and $\mathcal{D}U = \prod_{x,\mu} dU_{x,\mu}$ in terms of the Haar measure $dU_{x,\mu}$ of $SU(N)$.

With open boundary conditions in $(1+1)$D, the partition function defined by this action factorizes into a product of independent integrals over each P_x. To exploit this factorization in $(1+1)$D, a gauge fixing prescription can be applied in which $U_{x,2} = 1$ for all x and $U_{x,1} = 1$ for sites with $x_2 = 0$ (a maximal tree gauge). In this gauge,

$$P_x = U_{x,1} U^{-1}_{x+\hat{2},1}, \tag{32}$$

which can be easily inverted to obtain

$$U_{x,1} = \left[\prod_{k=0}^{x_2-1} P_{x+k\hat{2}} \right]^{-1}. \tag{33}$$

The variables P_x are therefore in one-to-one correspondence with the remaining nongauge-fixed $U_{x,1}$. The Haar measure is invariant under this change of variables, and the path integral defining the partition function factorizes as

$$Z = \prod_{x \in \mathcal{V}'} z = z^{|\mathcal{V}'|}, \tag{34}$$

where $\mathcal{V}' \subset \mathcal{V}$ is the subset of lattice points with unconstrained $U_{x,1}$ in this gauge (those for which $x_2 \neq 0$) and z is the contribution to the partition function from a single plaquette,

$$z \equiv \int dP \, e^{\frac{1}{g^2} tr(P+P^{-1})}. \tag{35}$$

The calculations of z and similar single-variable $SU(N)$ integrals are presented in Appendix A.

Wilson loops are defined by the matrix-valued quantity

$$W_{\mathcal{A}} \equiv \prod_{x,\mu \in \partial\mathcal{A}} U_{x,\mu}, \tag{36}$$

where $\prod_{x,\mu\in\partial\mathcal{A}} U_{x,\mu}$ represents an ordered product of links along the boundary $\partial\mathcal{A}$ of the two-dimensional region $\mathcal{A}$ with area A. The expectation value of the gauge-invariant observable $\frac{1}{N}\mathrm{tr}(W_{\mathcal{A}})$ probes the interaction between a pair of static quarks if the region $\mathcal{A}$ is taken to be rectangular. Inserting Eq. (33) into Eq. (36) gives[4]

$$\frac{1}{N}\mathrm{tr}(W_{\mathcal{A}}) = \frac{1}{N}\mathrm{tr}\left(\prod_{x\in\mathcal{A}} P_x \right). \tag{37}$$

Using linearity of expectation values and factorization of path integrals analogous to Eq. (34), the expectation values of Wilson loops can be related to products of (matrix-valued) single-variable expectation values,

$$\left\langle \frac{1}{N}\mathrm{tr}(W_{\mathcal{A}}) \right\rangle = \frac{1}{N}\mathrm{tr}\left(\prod_{x\in\mathcal{A}} \langle P_x \rangle \right). \tag{38}$$

Each single-variable expectation value is given by $\langle P_x^{ab} \rangle = \langle \chi_1 \rangle \delta^{ab}$, allowing the traced Wilson loop to be written as a product of scalars,

$$\left\langle \frac{1}{N} tr(W_{\mathcal{A}}) \right\rangle = \prod_{x\in\mathcal{A}} \langle \chi_1 \rangle = \langle \chi_1 \rangle^A, \tag{39}$$

where we have introduced the single-variable normalized expectation value of the group character function $\chi_1(P) = \mathrm{tr}(P)$,

$$\langle \chi_1 \rangle \equiv \frac{1}{z} \int dP \frac{1}{N}\mathrm{tr}(P) e^{\frac{1}{g^2}\mathrm{tr}(P+P^{-1})}, \tag{40}$$

whose value is computed in Appendix A.

Equation (39) implies that Wilson loop expectation values follow area law scaling, $\langle \mathrm{tr}(W_{\mathcal{A}})/N \rangle \sim e^{-\sigma A}$, and $SU(N)$ gauge theory in $(1+1)$D confines for all values of the coupling, with a separation-independent force between static test charges given by the string tension

$$\sigma \equiv -\lim_{A\to\infty} \partial_A \ln W_{\mathcal{A}} = -\ln \langle \chi_1 \rangle. \tag{41}$$

Although $\langle \chi_1 \rangle$ is in general given by a convergent infinite series in Eq. (A6), in the case of $SU(2)$ a simpler form can be found in terms of modified Bessel functions,

[4]For simplicity we restrict to rectangular Wilson loops with one corner at the origin.

TABLE I. The couplings used in our numerical studies of $SU(2)$ and $SU(3)$ lattice gauge theory. The dimensionless quantity σV is fixed to 6.4 while σ and V are individually varied. The conventional Wilson action parameter $\beta = 2N/g^2$ is also reported.

		$SU(2)$		$SU(3)$	
σ	V	g	β	g	β
0.4	16	0.98	4.2	0.72	11.7
0.2	32	0.71	8.0	0.53	21.7
0.1	64	0.51	15.5	0.38	41.8

$$\sigma^{SU(2)} = \ln\left(\frac{I_1(4/g^2)}{I_2(4/g^2)}\right), \tag{42}$$

which goes to zero as $g^2 \to 0$. This observation can be generalized to all $SU(N)$ groups, and the lattice-units string tension goes to zero while the static quark correlation length grows to infinity in the limit of $g^2 \to 0$ in all cases. We can consider this to be the naive continuum limit of the theory, though the correlation lengths of dynamical quantities such as plaquettes or localized Wilson loops remain finite by the factorization of the path integral. When investigating the approach to the continuum in Sec. IV and V, we should decrease the coupling while fixing the dimensionless quantity σV, where V is the total number of plaquettes; the particular choices of couplings and V used in our numerical studies are reported in Table I. Results are plotted versus σA when comparing quantities at fixed physical separation is important.

B. Noise and sign problems in the Wilson loop

Although the expectation value $\langle \mathrm{tr}(W_\mathcal{A})/N\rangle$ is real, the integrand $\mathrm{tr}(W_\mathcal{A})/N$ has fluctuating signs (for $N = 2$) or fluctuating complex phases (for $N \geq 3$) across the domain of integration. These fluctuations result in a sign/StN problem for this observable. The sample mean of $\mathrm{Re\,tr}(W_\mathcal{A})/N$ gives an estimator for $\langle \mathrm{tr}(W_\mathcal{A})/N\rangle$, and the variance of this estimator can be directly computed,

$$\begin{aligned}\mathrm{Var}[\mathrm{Re\,tr}(W_\mathcal{A})/N] &= \frac{1}{N^2}\langle \mathrm{Re tr}(W_\mathcal{A})^2\rangle - \frac{1}{N^2}\mathrm{Re}\langle \mathrm{tr}(W_\mathcal{A})\rangle^2 \\ &= \frac{1}{2N^2}\langle |\mathrm{tr}(W_\mathcal{A})^2|\rangle + \frac{1}{2N^2}\langle \mathrm{tr}(W_\mathcal{A})^2\rangle \\ &\quad - \frac{1}{N^2}\mathrm{Re}\langle \mathrm{tr}(W_\mathcal{A})\rangle^2. \end{aligned} \tag{43}$$

The expectation values in the first and second terms in the variance can be factorized analogously to the Wilson loop expectation value, and are shown in Appendix A to be

$$\begin{aligned}\langle |\mathrm{tr}(W_\mathcal{A})^2|\rangle &= 1 + (N^2-1)\langle \chi_{1,-1}\rangle^A \\ \langle \mathrm{tr}(W_\mathcal{A})^2\rangle &= \frac{N(N+1)}{2}\langle \chi_2\rangle^A + \frac{N(N-1)}{2}\langle \chi_{1,1}\rangle^A, \end{aligned} \tag{44}$$

in terms of the single-site integrals $\langle \chi_{1,-1}\rangle$, $\langle \chi_2\rangle$, and $\langle \chi_{1,1}\rangle$, defined in Eqs. (A9) and (A13). In total, the variance is

$$\mathrm{Var}[\mathrm{Re\,tr}(W_\mathcal{A})/N] = \frac{1}{2N^2} + \frac{O(e^{-cA})}{2N^2} - e^{-2\sigma A}, \tag{45}$$

where c is a constant. The fact that $\langle \chi_r\rangle < 1$ for nontrivial irreps r (assuming that g^2 is finite) [73] implies that $c > 0$ and therefore that the variance is asymptotically constant as $A \to \infty$,

$$\mathrm{Var}[\mathrm{Re\,tr}(W_\mathcal{A})/N] \sim \frac{1}{2N^2}. \tag{46}$$

The signal-to-noise ratio for n samples can be written exactly in terms of Eqs. (43), (44), and (39), but for this analysis it is sufficient to identify the asymptotic behavior from Eqs. (45) and (39), giving

$$\begin{aligned}\mathrm{StN}[\mathrm{Re\,tr}(W_\mathcal{A})/N] &= \frac{\langle \frac{1}{N} tr(W_\mathcal{A})\rangle}{\sqrt{\frac{1}{n}\mathrm{Var}[\mathrm{Re}\,\frac{1}{N}\mathrm{tr}(W_\mathcal{A})]}} \\ &\sim N\sqrt{2n}e^{-\sigma A}, \end{aligned} \tag{47}$$

which degrades exponentially with area A. For the estimator $\mathrm{Re\,tr}(W_\mathcal{A})/N$, the analysis above shows that this can only be counteracted by exponentially increasing the number of samples n. Equations (44) and (45) also make clear that the leading asymptotic behavior of the variance is due to the typical magnitude-squared of the observable, $\langle |\mathrm{tr}(W_\mathcal{A})^2|/N^2\rangle$, which remains $O(1)$ for all areas. Cancellations due to phase fluctuations are required to reproduce the exponentially small Wilson loop expectation values predicted for large areas, confirming that the StN problem can be related to a sign problem in the Wilson loop observable.

Attributing the StN problem to $O(1)$ magnitudes for individual samples of the Wilson loop observable at all areas also inspires our deformations of the Wilson loop observable in the following sections. The quantity $\langle |\mathrm{tr}(W_\mathcal{A})^2|/N^2\rangle$ can be written as an integral of a nonholomorphic integrand which will generically be modified by contour deformations of the path integral. If we choose contour deformations that reduce the average magnitude of the observable, this quantity, and thus the leading term of the variance, will be reduced. The observable mean is unchanged and the StN ratio will thus increase under such a deformation.

For $SU(2)$, the single-site integrals can be evaluated straightforwardly (see Appendix A) and the $SU(2)$ variance is

$$\mathrm{Var}[\mathrm{Re\,tr}(W_{\mathcal{A}}^{SU(2)})/N] = \frac{1}{4} + \frac{3}{4}\left(\frac{I_3(4/g^2)}{I_1(4/g^2)}\right)^A - e^{-2\sigma^{SU(2)}A}, \tag{48}$$

where $\sigma^{SU(2)}$ is given in Eq. (42). The StN of $SU(2)$ Wilson loops in $(1+1)$D can therefore be explicitly calculated,

$$\mathrm{StN}[\mathrm{Re\,tr}(W_{\mathcal{A}}^{SU(2)})/N] = \frac{2\sqrt{n}e^{-\sigma^{SU(2)}A}}{\sqrt{1+3(\frac{I_3(4/g^2)}{I_1(4/g^2)})^A - 4e^{-2\sigma^{SU(2)}A}}}. \tag{49}$$

Using numerical evaluation of the corresponding single-site integrals for $SU(N)$ Wilson loops yields theoretical curves for the variance and signal-to-noise for general N. In the studies below, we choose to deform the (1,1) component of the Wilson loop, $W_{\mathcal{A}}^{11}$, instead of $\mathrm{tr}W_{\mathcal{A}}/N$ following the reasoning of Sec. II D. The variance of $W_{\mathcal{A}}^{11}$ can be related to the variance of $\mathrm{tr}(W_{\mathcal{A}})/N$ and is compared to Monte Carlo results in the following sections.

IV. $SU(2)$ PATH INTEGRAL CONTOUR DEFORMATIONS

As a proof-of-principle, we apply path integral contour deformations to Wilson loop calculations in $SU(2)$ lattice gauge theory in $(1+1)$D with open boundary conditions. An identical setting with gauge group $SU(3)$ is investigated in the following section.

A. Gauge field parametrization

There are many possible parametrizations of the $SU(2)$ group manifold, any of which can be used to define valid path integral contour deformations. We argue above that it is advantageous to consider a single component of the Wilson loop, taken without loss of generality to be $W_{\mathcal{A}}^{11}$, as the observable whose path integral contour is deformed in order to calculate $\langle W_{\mathcal{A}}^{11}\rangle = \langle \mathrm{tr}(W_{\mathcal{A}})/N\rangle$. Contour deformations that reduce the magnitude of $W_{\mathcal{A}}^{11}$ in generic gauge field configurations while preserving $\langle W_{\mathcal{A}}^{11}\rangle$ can be expected to reduce phase fluctuations and therefore the variance of $W_{\mathcal{A}}^{11}$. The angular parametrization of each plaquette $P_x \in SU(2)$ is useful for this purpose, and is explicitly defined by

$$\begin{aligned} P_x^{11} &= \sin\theta_x e^{i\phi_x^1}, \\ P_x^{12} &= \cos\theta_x e^{i\phi_x^2}, \\ P_x^{21} &= -\cos\theta_x e^{-i\phi_x^2}, \\ P_x^{22} &= \sin\theta_x e^{-i\phi_x^1}, \end{aligned} \tag{50}$$

following the generalized $SU(N)$ angular parametrization given in Ref. [56]. The azimuthal angles satisfy $\phi_x^1, \phi_x^2 \in [0, 2\pi]$, with endpoints identified, while the angle θ_x spans the finite interval $[0, \pi/2]$. The normalized Haar measure can be written in these coordinates as

$$dP_x = h(\Omega_x)d\Omega_x \equiv \frac{1}{4\pi^2}\sin(2\theta_x)d\theta_x d\phi_x^1 d\phi_x^2. \tag{51}$$

We begin by considering the effects of simple deformations using these parameters. In the simplest case of a region $\mathcal{A}$ with area $A = 1$, the Wilson loop consists of a single plaquette, $W_{\mathcal{A}}^{11} = P_x^{11}$, where the loop starts and ends at site x. The magnitude of $W_{\mathcal{A}}^{11}$ can be reduced by $e^{-\lambda}$ by deforming $\phi_x^1 \to \phi_x^1 + i\lambda$ analogously to the approach described for $e^{i\phi}$ integrals above. In the case of $A = 2$, the Wilson loop can be written in terms of the product of two plaquettes, $W_{\mathcal{A}}^{11} = (P_x P_{x'})^{11}$. In the angular parametrization, the Wilson loop is a sum of two terms

$$(P_x P_{x'})^{11} = \sin\theta_x \sin\theta_{x'} e^{i\phi_x^1 + i\phi_{x'}^1} - \cos\theta_x \cos\theta_{x'} e^{i\phi_x^2 - i\phi_{x'}^2}. \tag{52}$$

The first term involves products of diagonal entries whose magnitude can be reduced by $e^{-\lambda}$ by taking $\phi_x^1 \to \phi_x^1 + i\lambda$ or $\phi_{x'}^1 \to \phi_{x'}^1 + i\lambda$ and the second term involves off-diagonal components whose magnitude can be reduced analogously by taking $\phi_x^2 - \phi_{x'}^2 \to (\phi_x^2 - \phi_{x'}^2) + i\lambda$. For $A > 2$, it can be seen similarly that shifting $\phi_x^1 \to \phi_x^1 + i\lambda$ and $(\phi_x^2 - \phi_{x+1}^2) \to (\phi_x^2 - \phi_{x+1}^2) + i\lambda$ for all x leads to suppression of the magnitudes of all terms appearing in $W_{\mathcal{A}}^{11}$.

A family of contour deformations capable of expressing these constant imaginary shifts to the phases of all elements of P_x can therefore be expected to reduce phase fluctuations and the variance of $W_{\mathcal{A}}^{11}$. Such a family of contour deformations is parametrized below as a subset of the vertical deformation expanded in a Fourier series in Eqs. (5)–(8).

An alternative parametrization of $SU(2)$ is explored in Appendix B, in which it is found that imaginary shifts along these lines are more difficult to express and orders of magnitude less variance reduction is achieved when applying the same optimization methods. This exploration suggests that a choice of parametrization that allows the observable to be expressed in the form $e^{i\phi}$ is important for variance reduction in observables afflicted with a sign problem.

It is also possible to directly parametrize the gauge field $U_{x,\mu} \in SU(2)$ using Eq. (50). This alternative parametrization may be useful in more than two spacetime dimensions, where Gauss' law constraints imply that not all plaquettes are independent and a path integral change of variables from $U_{x,\mu}$ to $P_x^{\mu\nu}$ cannot be performed straightforwardly.

B. Fourier deformation basis

In our study of $SU(2)$ gauge theory, we optimize over a family of vertical contour deformations expressed in terms

of a Fourier series truncated above a specific cutoff mode. To avoid a costly Jacobian calculation, each plaquette variable P_x is deformed conditioned on plaquettes P_y at sites earlier in the product defining $W_{\mathcal{A}}$ in Eq. (37), which we write as $y \leq x$. This family of deformations is given by

$$\begin{aligned}
\tilde{\theta}_x &\equiv \theta_x + i\sum_{y\leq x} f_\theta(\theta_y, \phi_y^1, \phi_y^2; \kappa^{xy}, \lambda^{xy}, \eta^{xy}, \chi^{xy}, \zeta^{xy}),\\
\tilde{\phi}_x^1 &\equiv \phi_x^1 + i\kappa_0^{x;\phi^1} + i\sum_{y\leq x} f_{\phi^1}(\theta_y, \phi_y^1, \phi_y^2; \kappa^{xy}, \lambda^{xy}, \eta^{xy}, \chi^{xy}, \zeta^{xy}),\\
\tilde{\phi}_x^2 &\equiv \phi_x^2 + i\kappa_0^{x;\phi^2} + i\sum_{y\leq x} f_{\phi^2}(\theta_y, \phi_y^1, \phi_y^2; \kappa^{xy}, \lambda^{xy}, \eta^{xy}, \chi^{xy}, \zeta^{xy}),
\end{aligned} \tag{53}$$

in terms of parameters κ^{xy}, λ^{xy}, η^{xy}, χ^{xy}, and ζ^{xy}. The functions f_θ, f_{ϕ^1}, and f_{ϕ^2} compute the shift in the imaginary direction of the angular parameters of P_x conditioned on P_y, and their decomposition in terms of Fourier modes is detailed below. For this ordered dependence on previous sites, the Jacobian determinant factorizes into a product of block determinants per lattice site

$$J = \prod_x j_x(\theta_x, \phi_x^1, \phi_x^2), \tag{54}$$

where

$$j_x(\theta_x, \phi_x^1, \phi_x^2) = \det\begin{pmatrix} \frac{\partial f_\theta}{\partial \theta_x} & \frac{\partial f_\theta}{\partial \phi_x^1} & \frac{\partial f_\theta}{\partial \phi_x^2} \\ \frac{\partial f_{\phi^1}}{\partial \theta_x} & \frac{\partial f_{\phi^1}}{\partial \phi_x^1} & \frac{\partial f_{\phi^1}}{\partial \phi_x^2} \\ \frac{\partial f_{\phi^2}}{\partial \theta_x} & \frac{\partial f_{\phi^2}}{\partial \phi_x^1} & \frac{\partial f_{\phi^2}}{\partial \phi_x^2} \end{pmatrix}. \tag{55}$$

The structure of the deformation in Eq. (53) therefore bypasses the need for expensive Jacobian determinant calculations involving matrices whose rank grows with the spacetime volume and is inspired by analogous methods to simplify Jacobian determinant calculations in normalizing flows [65]. Note that an absolute value is not taken over the determinant in Eq. (55).

The vertical deformation in Eq. (53) can be expanded in a multiparameter Fourier series as

$$\begin{aligned}
f_\theta &= \sum_{m=1}^{\Lambda} \kappa_m^{xy;\theta} \sin(2m\theta_y)\left\{1 + \sum_{n=1}^{\Lambda}\left[\eta_{mn}^{xy;\theta,\phi^1}\sin(n\phi_y^1 + \chi_{mn}^{xy;\theta,\phi^1}) + \eta_{mn}^{xy;\theta,\phi^2}\sin(n\phi_y^2 + \chi_{mn}^{xy;\theta,\phi^2})\right]\right\},\\
f_{\phi^1} &= \sum_{m=1}^{\Lambda} \kappa_m^{xy;\phi^1} \sin(m\phi_y^1 + \zeta_m^{xy;\phi^1})\left\{1 + \sum_{n=1}^{\Lambda}\left[\lambda_{mn}^{xy;\phi^1,\theta}\sin(2n\theta_y) + \eta_{mn}^{xy;\phi^1,\phi^2}\sin(n\phi_y^2 + \chi_{mn}^{xy;\phi^1,\phi^2})\right]\right\},\\
f_{\phi^2} &= \sum_{m=1}^{\Lambda} \kappa_m^{xy;\phi^2} \sin(m\phi_y^2 + \zeta_m^{xy;\phi^2})\left\{1 + \sum_{n=1}^{\Lambda}\left[\lambda_{mn}^{xy;\phi^2,\theta}\sin(2n\theta_y) + \eta_{mn}^{xy;\phi^2,\phi^1}\sin(n\phi_y^1 + \chi_{mn}^{xy;\phi^2,\phi^1})\right]\right\},
\end{aligned} \tag{56}$$

where Λ is a hyperparameter that sets the maximum Fourier mode to include and controls the total number of free parameters. As the zero modes have trivial y dependence, we have collected them in Eq. (53) into the y-independent terms $\kappa_0^{x;\phi^1}$ and $\kappa_0^{x;\phi^2}$. The included Fourier terms are defined to satisfy the constraints $\tilde{\theta}_x(0) = 0$, $\tilde{\theta}_x(\pi/2) = \pi/2$, $\tilde{\phi}_x^1(0) = \tilde{\phi}_x^1(2\pi)$, and $\tilde{\phi}_x^2(0) = \tilde{\phi}_x^2(2\pi)$, which together ensure that the endpoints of both the zenith and azimuthal integration domains are appropriately handled as described in Sec. II A. The derivatives needed for the Jacobian in Eqs. (54)–(55) can be calculated straightforwardly by differentiating Eq. (56). The additional factor describing the change in the Haar measure needed to compute the Jacobian of the group measure is given in these coordinates as

$$\prod_x \frac{h(\tilde{\Omega}_x)}{h(\Omega_x)} = \prod_x \left[\frac{\sin(2\tilde{\theta}_x)}{\sin(2\theta_x)}\right]. \tag{57}$$

Combining the results of Eq. (50) and Eqs. (53)–(57), the expectation value of any holomorphic observable $\mathcal{O}(\{P_x\})$ is equal to the expectation value of the deformed observable

$$Q(\{P_x\}) \equiv \mathcal{O}(\{\tilde{P}_x\}) \frac{e^{-S(\{\tilde{P}_x\})}}{e^{-S(\{P_x\})}} \prod_x j_x \left[\frac{\sin(2\tilde{\theta}_x)}{\sin(2\theta_x)}\right], \tag{58}$$

where

$$\tilde{P}_x = \begin{pmatrix} \sin\tilde{\theta}_x e^{i\tilde{\phi}_x^1} & \cos\tilde{\theta}_x e^{i\tilde{\phi}_x^2} \\ -\cos\tilde{\theta}_x e^{-i\tilde{\phi}_x^2} & \sin\tilde{\theta}_x e^{-i\tilde{\phi}_x^1} \end{pmatrix} \in SL(2,\mathbb{C}). \tag{59}$$

If the plaquettes are sampled in the matrix representation for Monte Carlo evaluation, computing the observable Q in Eq. (58) requires converting to the angular representation before deforming and evaluating. This conversion is given by

$$\theta_x = \arcsin(|P_x^{11}|),$$
$$\phi_x^1 = \arg(P_x^{11}),$$
$$\phi_x^2 = \arg(P_x^{12}), \tag{60}$$

and can be done when evaluating the observable $Q(\{P_x\})$. Though these functions are not entire, the conversion used here does not determine whether the integrand itself is a holomorphic function of these angular parameters.

C. Optimization procedure

This contour deformation expanded in a Fourier series provides a means of exploring deformed observables with potentially reduced variance. It is shown above that simple deformations within this family are possible to construct by hand and are already sufficient to reduce the average magnitude of Wilson loop observables. However, these deformations are restricted to zero modes of the Fourier expansion and rely on construction based on intuition. To maximize the variance reduction, we explore numerical optimization of the deformation parameters κ^{xy}, λ^{xy}, η^{xy}, χ^{xy}, and ζ^{xy} as discussed in Sec. II C. We are interested in $\mathrm{Re}W_{\mathcal{A}}^{11}$, for which the terms of Eq. (43) that can be modified by contour deformation are

$$\mathcal{L} \equiv \langle(\mathrm{Re}Q_{\mathcal{A}})^2\rangle = \frac{1}{2}\langle|Q_{\mathcal{A}}^2|\rangle + \frac{1}{2}\langle Q_{\mathcal{A}}^2\rangle, \tag{61}$$

where $Q_{\mathcal{A}}$ is the deformed observable associated with the $W_{\mathcal{A}}^{11}$. The first term in Eq. (61) is manifestly nonholomorphic due to the absolute value over a complex-valued observable, while the second term includes squared reweighting factors of the original and deformed action which prevent identification as a deformation of $\langle(W_{\mathcal{A}}^{11})^2\rangle$. These terms together define the *loss function* $\mathcal{L}$ that we aim to minimize as a function of the deformation parameters.

This loss function is written as an expectation value in terms of sampling from the original Monte Carlo weights $e^{-S(\{P_x\})}$, and its gradient can similarly be written as an expectation value,

$$\nabla\mathcal{L} = \langle 2\mathrm{Re}Q_{\mathcal{A}}\nabla\mathrm{Re}Q_{\mathcal{A}}\rangle. \tag{62}$$

The term $\nabla\mathrm{Re}Q_{\mathcal{A}}$ can be expanded using the explicit form of $Q_{\mathcal{A}}$ given in Eq. (58), as well as the forms of the observable $W_{\mathcal{A}}^{11}$ and the action S in terms of $\{P_x\}$ in Sec. III A. For this study, the gradient $\nabla\mathrm{Re}Q_{\mathcal{A}}$ was computed explicitly and cross-checked using automatic differentiation available in the JAX library [74]. Equation (62) can be stochastically estimated using an (undeformed) ensemble of n configurations $\{P_x^i\}$, $i \in [1, \ldots, n]$, drawn proportional to the weight $e^{-S(\{P_x\})}$,

$$\nabla\mathcal{L} \approx \frac{1}{n}\sum_{i=1}^{n}[2\mathrm{Re}Q_{\mathcal{A}}(\{P_x^i\})\nabla\mathrm{Re}Q_{\mathcal{A}}(\{P_x^i\})]. \tag{63}$$

In all experiments below, we used the Adam optimizer [75] to iteratively update parameters based on these stochastic gradient estimates. Each gradient estimate was computed using 1/100th of the generated ensemble; empirically, this small subset of the data was sufficient to learn useful manifold parameters with significant variance reduction. The optimizer was configured with default hyperparameters, except for a dynamically scheduled step size. Stochastic noise on gradient estimates and large optimizer step size can either slow convergence or result in unstable training dynamics, while step sizes that are too small waste computation as parameters fail to move quickly along precisely estimated gradients. We thus used a dynamic schedule that reduced the step size over time. In particular, our step size schedule started with an initial step size s_0 and then permanently reduced the step size by a factor of F (i.e., $s_{i+1} = s_i/F$) each time the loss function failed to improve over a window of W steps. The schedule halted training after the step size was reduced N_r times. We used the parameters $F = 10$, $W = 50$, $N_r = 2$, and $s_0 = 10^{-2}$ for both $SU(2)$ and $SU(3)$ gauge theory.

In preliminary investigations, we found that naive manifold optimization resulted in overtraining, i.e., overfitting parameters to the specific training data available [76–78]. In the context of manifold optimization, this corresponds to minimizing a finite-ensemble variance estimator rather than minimizing the true variance of $\mathrm{Re}Q_{\mathcal{A}}$. In practice this produced deformed observables with higher variance when measured on a different ensemble than the training set.

To mitigate overtraining in the final results, we reserved a "test set" of configurations, independent of the training data, on which the loss function was periodically measured [79]; the reserved test set of configurations was chosen to have the same size as the training set. The step size schedule was configured to use loss measurements based on this test set, ensuring that training was slowed and halted before significantly overfitting. We further used a minibatching technique, in which a set of $m \leq n$ configurations are resampled from the training set to estimate Eq. (63), as a means of avoiding overtraining [69]. The minibatch size was chosen to be equal to the size of the full training set (i.e., $m = n$), thus minibatch evaluation in this context was just a resampling operation, giving variation in gradient estimates over multiple evaluations. The fluctuations in these gradient estimates are on the order of the uncertainty of the variance estimator, preventing overfitting below this uncertainty. For each observable $W_{\mathcal{A}}^{11}$ we also found it important to restrict to deforming only the plaquettes within the region $\mathcal{A}$. Though including extra degrees of freedom cannot make the optimal variance any higher (the optimization steps could always leave those plaquettes

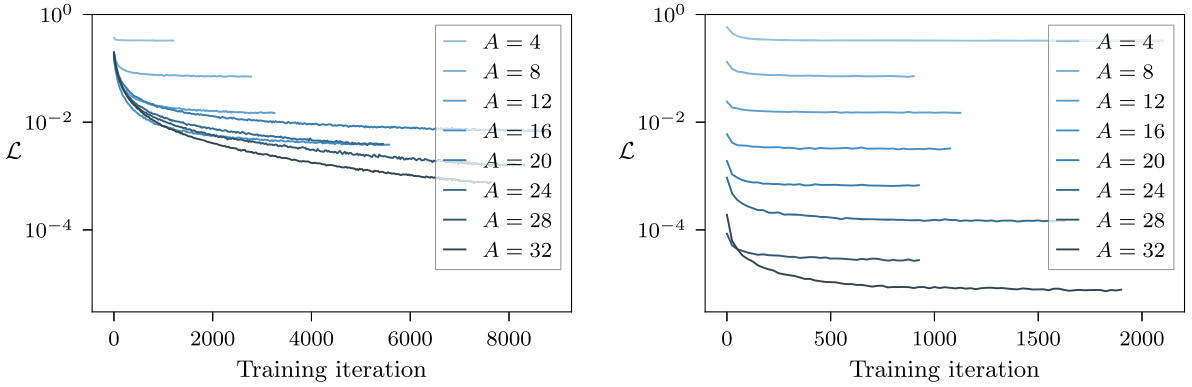

FIG. 3. $SU(2)$ Wilson loop loss $\mathcal{L}$ plotted with respect to training iterations for optimization of Wilson loops with $g = 0.71$ and area A starting with the undeformed manifold (left) or starting with the deformed manifold calculated for area $A - 1$ (right). The loss curves are averaged over blocks of 25 iterations for clarity. In each case, the training is halted based on the plateau criterion described in the main text, resulting in traces of different lengths. For Wilson loops with larger areas, the final loss value is substantially lower when initialized from optimal manifold parameters at a $A - 1$, despite using nearly four times fewer training iterations.

undeformed), in practice we found that including such degrees of freedom allowed the deformed manifold to rapidly overtrain, resulting in worse performance overall. Appendix C details further possible approaches to avoiding overfitting and overlap problems using a regularizing term added to the loss function. These approaches were found to be unnecessary for our final results.

Finally, making a good choice of initial manifold parameters yielded practical improvement in training time and quality. On one hand, initializing the manifold parameters to zero ensures that gradient descent starts from the original manifold, and the optimization procedure should find a local minimum with variance no higher than the original manifold (up to noise from stochastic gradient estimates). However, correlations in sign and magnitude fluctuations of observables with similar structure, such as Wilson loops $W^{11}_{\mathcal{A}}$ and $W^{11}_{\mathcal{A}'}$ with overlapping areas $\mathcal{A}$ and $\mathcal{A}'$, suggest that the variance reduction from contour deformations will be correlated as well. Though the optimal manifold for one observable will not generically be optimal for the other, it can serve as a better starting point than the original manifold. In our study of Wilson loops, we initialized manifold parameters for each Wilson loop of area A using the optimal parameters for the Wilson loop of area $A - 1$, when the Fourier cutoff $\Lambda = 0$, or using the optimal parameters for the Wilson loop of area A and cutoff $\Lambda - 1$, when $\Lambda \neq 0$. Figure 3 shows the improvement in optimization time and quality using this method for manifold deformations with $\Lambda = 0$. While this approach sacrifices the guarantee that the local minimum obtained corresponds to a deformed observable with variance no higher than the original observable (in the limit of infinitely precise gradient estimates), in practice we find that this property is not violated. We also note that this property can be easily checked after optimization, and if the variance were found to increase with respect to the original observable training could instead be started from the original manifold to recover the guarantee.

D. Monte Carlo calculations

We investigated the practical performance of these contour deformations on the three sets of $SU(2)$ parameters detailed in Table I. These parameters correspond to variation in the string tension by a factor of 4. At each choice of parameters, $n = 32000$ configurations were generated, exploiting the factorization of the path integral to draw samples of each plaquette P_x independently from the marginal distribution,

$$p(P_x) \propto e^{-\frac{1}{g^2}\mathrm{tr}(P_x + P_x^{-1})}. \tag{64}$$

Plaquette samples were generated using Hybrid Monte Carlo [80] with parameters tuned to maintain autocorrelation times of order 1–2,[5] and these individual plaquettes were then arranged into lattices consisting of V sites each. A random shuffle was applied to the collection of plaquettes prior to this assignment to avoid spurious spatial correlations, ensuring an asymptotically vanishing bias in the limit of infinite ensemble size.

On each ensemble, we performed a study of deformations with Fourier cutoff Λ fixed to zero. For all three choices of parameters, training on a subset of 320 configurations was sufficient to converge to manifolds with variance reduction up to four orders of magnitude when comparing the largest deformed observables to the original Wilson loop observables. An additional subset of 320 configurations was reserved as the test set during optimization, and the remaining 31360 configurations were used to measure results. Measurements of $Q_{\mathcal{A}}$ were found to be consistent with the exact results given in Sec. III A and with the Monte Carlo results for $W^{11}_{\mathcal{A}}$ in the region

[5]For $SU(2)$ gauge theory it is also possible to apply a heat bath algorithm to directly draw samples. However, more complicated variants are required for $SU(3)$ [81–83] and hybrid Monte Carlo was selected for simplicity.

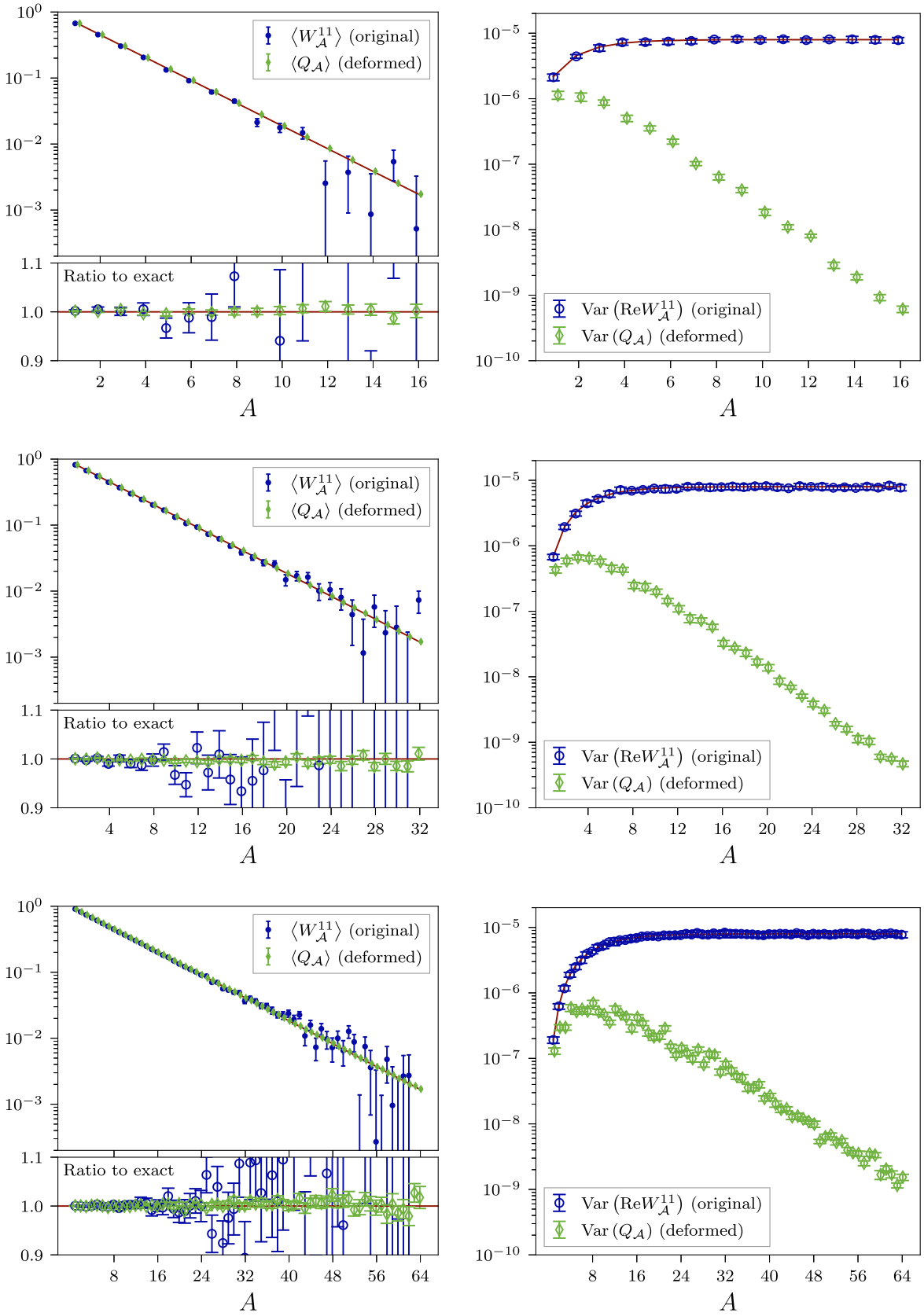

FIG. 4. $SU(2)$ Wilson loop expectation values and variances for ensembles with three different values of the gauge coupling $g = 0.98$, 0.71, 0.51 (top to bottom). Red lines indicate analytical results for $\langle W_{\mathcal{A}}^{11} \rangle = \langle \mathrm{tr}(W_{\mathcal{A}})/2 \rangle$ (left plots) and for $\mathrm{Var}(\mathrm{Re} W_{\mathcal{A}}^{11})$ (right plots).

where the estimates of $W_{\mathcal{A}}^{11}$ were reliable. The variance of $\mathrm{Re}W_{\mathcal{A}}^{11}$ is dominated by $\langle(\mathrm{Re}W_{\mathcal{A}}^{11})^2\rangle$, which is positive-definite for all gauge field configurations. It was therefore possible to precisely measure the variance without a sign/StN problem, and the expected agreement with analytical results was reproduced. In particular, $O(1)$ scaling at large area can be clearly seen. In contrast, we found that the variance of $Q_{\mathcal{A}}$ for manifolds optimized as above falls exponentially at large area, giving exponential improvements in the signal-to-noise. These results are presented for all three ensembles in Fig. 4.

Improvements from contour deformation were also found to be similar for Wilson loops with fixed physical area σA across the three values of the lattice spacing. Figure 5 compares the improvement in the Wilson loop variance versus the dimensionless scale σA across all three ensembles, and the three curves can be seen to nearly collapse at small areas, though there are differences of roughly a factor of 4 at the largest areas. Despite this variation, the variance of the Wilson loop with largest area is reduced by more than 10^3 even for the finest ensemble. Analogous results were observed for $(1+1)$D $U(1)$ gauge theory in Ref. [48].

For $\Lambda = 0$, there are few enough parameters that it is possible to investigate the optimal parameters found by the optimization procedure. Figure 6 depicts the values of $\kappa_0^{x;\phi^1}$ and $\kappa_0^{x;\phi^2}$ when optimized to reduce variance of $Q_{\mathcal{A}}$ at three choices of the loop area. It is argued in Sec. IV A that the magnitude of $Q_{\mathcal{A}}$ can be reduced on each sample if ϕ_x^1 and the differences $\phi_x^2 - \phi_{x+1}^2$ are shifted by a positive imaginary constant. This manifold is approximately discovered by the optimization procedure: the final parameters are a positive, nearly constant $\kappa_0^{x;\phi^1}$, corresponding to a positive imaginary shift applied to ϕ_x^1, and a decreasing $\kappa_0^{x;\phi^2}$, corresponding to a positive imaginary shift applied to each difference $\phi_x^2 - \phi_{x+1}^2$. Only these relative differences between neighboring ϕ_x^2 have an effect on the value of $Q_{\mathcal{A}}$, thus the overall shift on the collection of $\kappa_0^{x;\phi^2}$ is irrelevant.

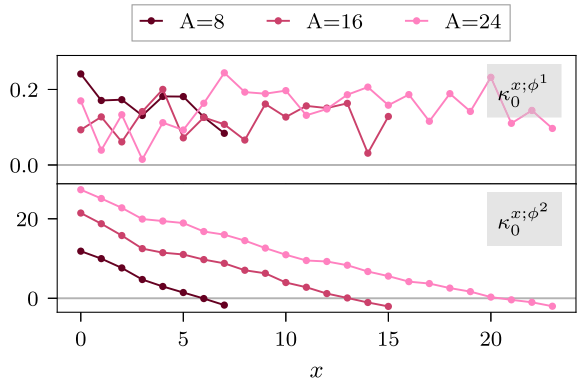

FIG. 6. The manifold parameters found by optimizing the variance of the deformed Wilson loop observable $Q_{\mathcal{A}}$ at three different choices of area A on the ensemble with total volume $V = 32$ and $\beta = 8.0$. Optimization at each A was initialized using the parameters found for the observable with area $A-1$, as detailed in the main text.

We further optimized manifold parameters using Fourier cutoffs $\Lambda = 1, 2$ on the ensemble with coupling $g = 0.71$ to investigate gains from including higher Fourier modes. Including Fourier modes larger than the constant term enables more complex deformations of each angular parameter, and introduces possible dependence on plaquettes at sites $y \le x$ when deforming P_x. Despite this increased expressivity, these additional parameters did not provide significantly larger StN improvements compared to using only constant terms, as shown in Fig. 7. In some

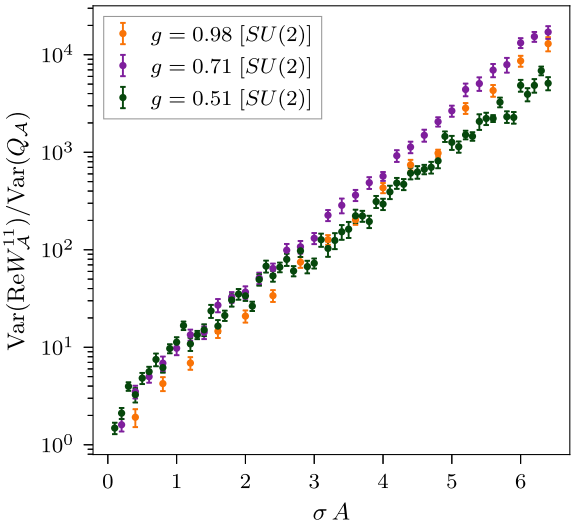

FIG. 5. $SU(2)$ Wilson loop variance ratios of standard observables to deformed observables for ensembles with three different values of the gauge coupling g, corresponding to three different values of lattice spacing.

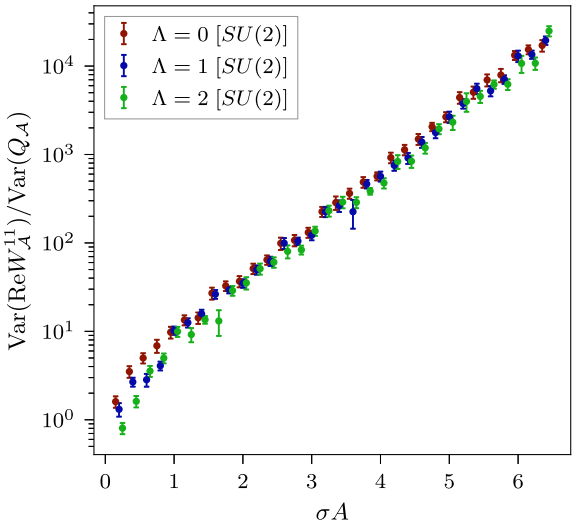

FIG. 7. $SU(2)$ Wilson loop variance ratios of standard observables to deformed observables for ensembles with $g = 0.71$ and three different values of the manifold parametrization cutoff.

cases, the optimized manifold with larger cutoff resulted in higher variance (lower variance ratio) than the optimized manifold with cutoff $\Lambda = 0$. The manifolds with larger cutoffs include all possible manifolds with smaller cutoffs, thus this is necessarily a training effect, likely due to noisier gradients and less stable training dynamics. We did not pursue noise reduction and alternative approaches to training (such as iteratively including higher Λ) as these manifolds with higher cutoffs did not produce significant improvements at any value of the area.

V. $SU(3)$ PATH INTEGRAL CONTOUR DEFORMATIONS

We further investigated the ability of contour deformations to reduce the variance of Wilson loops in $SU(3)$ gauge theory in $(1+1)$D with open boundary conditions. This setting is identical to the previous section, except for the use of $SU(3)$ rather than $SU(2)$ gauge field variables. Suitable parametrizations for contour deformations of $SU(3)$ gauge fields are discussed below.

A. Gauge field parametrization and contour deformation

For the $SU(3)$ gauge group, we use the angular parametrization constructed in Ref. [56]. The components of a single plaquette $P_x \in SU(3)$ are parametrized as

$$\begin{aligned}
P_x^{11} &= \cos\theta_x^1 \cos\theta_x^2 e^{i\phi_x^1},\\
P_x^{12} &= \sin\theta_x^1 e^{i\phi_x^3},\\
P_x^{13} &= \cos\theta_x^1 \sin\theta_x^2 e^{i\phi_x^4},\\
P_x^{21} &= \sin\theta_x^2 \sin\theta_x^3 e^{-i(\phi_x^4+\phi_x^5)}\\
&\quad - \sin\theta_x^1 \cos\theta_x^2 \cos\theta_x^3 e^{i(\phi_x^1+\phi_x^2-\phi_x^3)},\\
P_x^{22} &= \cos\theta_x^1 \cos\theta_x^3 e^{i\phi_x^2},\\
P_x^{23} &= -\cos\theta_x^2 \sin\theta_x^3 e^{-i(\phi_x^1+\phi_x^5)}\\
&\quad - \sin\theta_x^1 \sin\theta_x^2 \cos\theta_x^3 e^{i(\phi_x^2-\phi_x^3+\phi_x^4)},\\
P_x^{31} &= -\sin\theta_x^1 \cos\theta_x^2 \sin\theta_x^3 e^{i(\phi_x^1-\phi_x^3+\phi_x^5)}\\
&\quad - \sin\theta_x^2 \cos\theta_x^3 e^{-i(\phi_x^2+\phi_x^4)},\\
P_x^{32} &= \cos\theta_x^1 \sin\theta_x^3 e^{i\phi_x^5},\\
P_x^{33} &= \cos\theta_x^2 \cos\theta_x^3 e^{-i(\phi_x^1+\phi_x^2)}\\
&\quad - \sin\theta_x^1 \sin\theta_x^2 \sin\theta_x^3 e^{-i(\phi_x^3-\phi_x^4-\phi_x^5)},
\end{aligned} \tag{65}$$

in terms of the three zenith angles $0 \le \theta_x^1, \theta_x^2, \theta_x^3 \le \pi/2$ and the five azimuthal angles $0 \le \phi_x^1, \ldots, \phi_x^5 \le 2\pi$ for each plaquette. We collect these angles into a variable $\Omega_x = (\theta_x^1, \ldots, \theta_x^3, \phi_x^1, \ldots, \phi_x^5)$ for ease of notation. The Haar measure is related to the measure of Ω by [56]

$$\begin{aligned}
dP_x = \frac{1}{2\pi^5} &\sin\theta_x^1 (\cos\theta_x^1)^3 \sin\theta_x^2 \cos\theta_x^2 \sin\theta_x^3 \cos\theta_x^3\\
&\times d\theta_x^1 d\theta_x^2 d\theta_x^3 d\phi_x^1 \ldots d\phi_x^5.
\end{aligned} \tag{66}$$

To compute deformed observables from Monte Carlo samples in the matrix representation, an inverse map of (65) is needed and for example can be specified by

$$\begin{aligned}
\theta_x^1 &= \arcsin(|P_x^{12}|),\\
\theta_x^2 &= \arccos(|P_x^{11}|/\cos(\theta_x^1)),\\
\theta_x^3 &= \arccos(|P_x^{22}|/\cos(\theta_x^1)),\\
\phi_x^1 &= \arg(P_x^{11}),\\
\phi_x^2 &= \arg(P_x^{22}),\\
\phi_x^3 &= \arg(P_x^{12}),\\
\phi_x^4 &= \arg(P_x^{13}),\\
\phi_x^5 &= \arg(P_x^{32}).
\end{aligned} \tag{67}$$

An $SU(3)$ field configuration in $(1+1)$D with open boundary conditions is defined by a collection of angular variables Ω_x associated all plaquettes P_x on the lattice. The ath component of the deformed angles at site x is denoted by $(\tilde{\Omega}_x)_a$, and for vertical deformations expanded in Fourier modes is specified by

$$\begin{aligned}
\tilde{\phi}_x^a &= \phi_x^a + i\kappa_0^{x;\phi^a} + i\sum_{y\le x} f_{\phi^a}(\Omega_y; \kappa^{xy}, \lambda^{xy}, \chi^{xy}, \zeta^{xy}),\\
\tilde{\theta}_x^a &= \theta_x^a + i\sum_{y\le x} f_{\theta^a}(\Omega_y; \kappa^{xy}, \lambda^{xy}, \chi^{xy}).
\end{aligned} \tag{68}$$

where

$$\begin{aligned}
f_{\theta^a} &= \sum_{m=1}^{\Lambda} \kappa_m^{xy;a} \sin(2m\theta_y^a)\left\{1 + \sum_{n=1}^{\Lambda}\left[\sum_{\substack{r\neq a\\ r=1}}^{3} \lambda_{mn}^{xy;ar}\sin(2n\theta_y^r) + \sum_{s=1}^{5}\eta_{mn}^{xy;as}\sin(n\phi_y^s + \chi_{mn}^{xy;as})\right]\right\},\\
f_{\phi^a} &= \sum_{m=1}^{\Lambda} \kappa_m^{xy;a} \sin(m\phi_y^a + \zeta_m^{xy;a})\left\{1 + \sum_{n=1}^{\Lambda}\left[\sum_{r=1}^{3} \lambda_{mn}^{xy;ar}\sin(2n\theta_y^r) + \sum_{\substack{s\neq a\\ s=1}}^{5}\eta_{mn}^{xy;as}\sin(n\phi_y^s + \chi_{mn}^{xy;as})\right]\right\}.
\end{aligned} \tag{69}$$

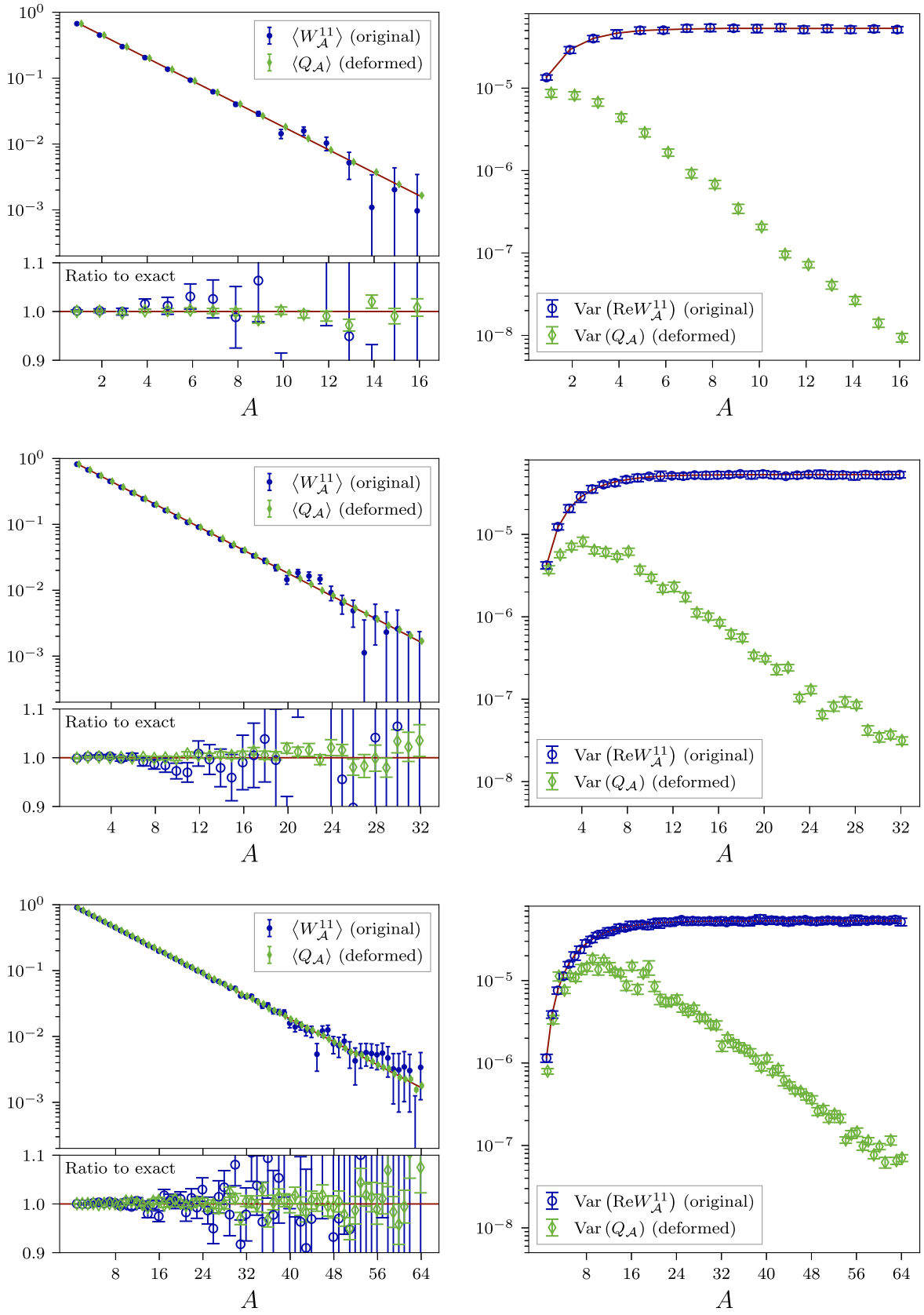

FIG. 8. $SU(3)$ Wilson loop expectation values and variances for ensembles with three different values of the gauge coupling that from top to bottom correspond to $g = 0.72, 0.53, 0.38$. Red lines correspond to analytical results for $\langle W_{\mathcal{A}}^{11} \rangle = \langle \mathrm{tr}(W_{\mathcal{A}})/3 \rangle$ (left plots) and $\mathrm{Var}(\mathrm{Re} W_{\mathcal{A}}^{11})$ (right plots).

Deformed observables analogous to Eq. (58) can be constructed for the $SU(3)$ case using this parametrization and ratios of the deformed and undeformed Haar measure factors obtained from Eq. (66).

B. Results

Practical performance of these deformations was investigated by optimizing Wilson loop variance using the three sets of $SU(3)$ parameters detailed in Table I as in the $SU(2)$ case. The couplings were tuned to match the string tensions used for $SU(2)$ gauge theory and correspond to lattice spacings varying by a factor of 4. For each choice of parameters, an ensemble of $n = 32000$ configurations was generated using the factorized HMC method discussed in Sec. IV D. Figure 8 shows variance reduction for Wilson loops of all possible sizes for the three lattice spacings studied. At the largest loop areas, we found variance reduction of greater than three orders of magnitude. Across all three ensembles, analytical results for the Wilson loop expectation values and variances were reproduced by the undeformed Monte Carlo estimates. The expectation value of the deformed observable is consistent with the analytical and original Monte Carlo results, while the variance of the deformed observable exponentially decreases with increasing σA.

Figure 9 compares the variance reduction achieved at all three lattice spacings versus physical loop area σA. We found approximately equivalent improvement in the variance at the two coarser lattice spacings ($g = 0.72$ and $g = 0.53$), and a small, yet significant, decrease in the variance improvement achieved at the finest lattice spacing ($g = 0.38$). Despite this, the variance was reduced by three orders of magnitude at the largest area on the finest lattice by using an optimized deformed observable, and at all three couplings variance improvements are consistent with exponential in the physical loop area. The number of parameters to be optimized grows with the volume in lattice units, and the analogous results observed for $SU(2)$ gauge theory suggest that the results at finer lattice spacings could be partially explained by increased difficulty in training the larger number of parameters.

The $\Lambda = 0$ manifold parametrization involves few enough parameters that it is possible to investigate the structure of the optimal parameters similarly to the case of $SU(2)$ gauge theory. As shown in Fig. 10, we found that the optimized values of $\kappa_0^{x;3}$ and $\kappa_0^{x;4}$ decrease with x, while $\kappa_0^{x;1}$ and $\kappa_0^{x;2}$ appear to converge toward approximately constant positive and negative values, respectively. The final parameter $\kappa_0^{x;5}$ fluctuates in both the positive and negative directions. These results can be qualitatively explained by expanding the (1,1) component of the Wilson loop for small area. For $A = 1$ the Wilson loop is equivalent to P_x, for which the (1,1) component is

$$P_x^{11} = \cos\theta_x^1 \cos\theta_x^2 e^{i\phi_x^1}. \tag{70}$$

The magnitude of this quantity can be reduced by shifting $\phi^1 \to \tilde{\phi}_x^1 = \phi_x^1 + i\lambda$, which is consistent with the positive

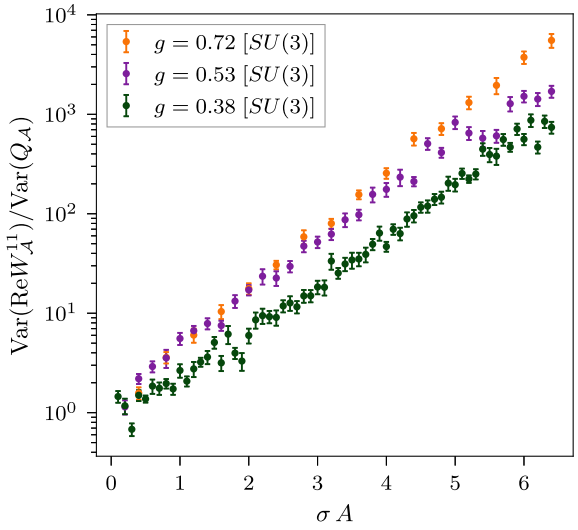

FIG. 9. $SU(3)$ Wilson loop variance ratios of standard observables to deformed observables for ensembles with three different values of the gauge coupling that correspond to $g = 0.72$, 0.53, 0.38 (top to bottom).

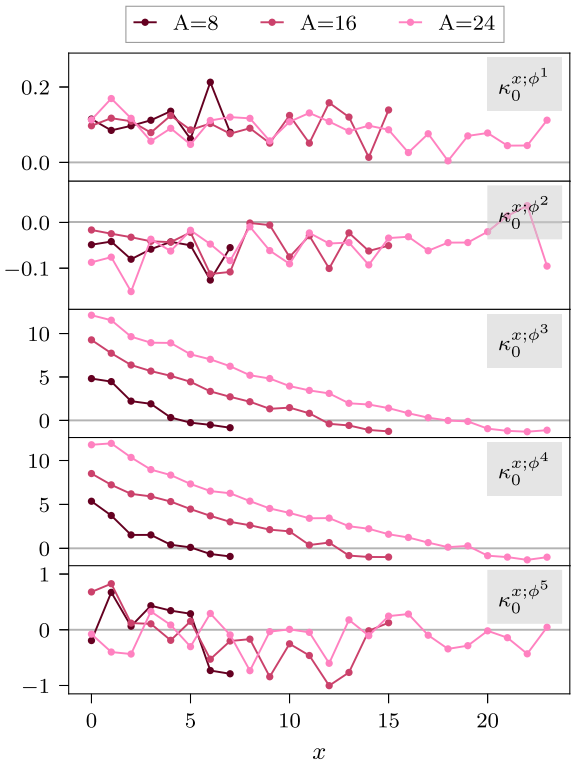

FIG. 10. The parameters of the optimal manifold for Wilson loop $Q_{\mathcal{A}}$ at three different areas A, as determined on the ensemble with total volume $V = 32$ and $g = 0.53$. Optimization for each $Q_{\mathcal{A}}$ was initialized using the parameters for optimal $Q_{\mathcal{A}'}$ with region $\mathcal{A}'$ of area $A - 1$, as detailed in the main text.

$\kappa_0^{x;1}$ values obtained after optimization shown in Fig. 10. Extending the analysis to the $A = 2$ Wilson loop, the (1,1) component is given by

$$\begin{aligned}(P_x P_{x'})^{11} = & e^{i\phi_x^1 + i\phi_{x'}^1} \cos\theta_x^1 \cos\theta_{x'}^1 \cos\theta_x^2 \cos\theta_{x'}^2 \\ & - e^{i\phi_{x'}^1 + i\phi_{x'}^2 + i(\phi_x^3 - \phi_{x'}^3)} \cos\theta_{x'}^2 \cos\theta_{x'}^3 \sin\theta_x^1 \sin\theta_{x'}^1 \\ & - e^{-i\phi_{x'}^2 + i(\phi_x^4 - \phi_{x'}^4)} \cos\theta_x^1 \cos\theta_{x'}^3 \sin\theta_x^2 \sin\theta_{x'}^2 \\ & - e^{i\phi_x^4 + i\phi_{x'}^1 - i\phi_{x'}^3 + i\phi_{x'}^5} \cos\theta_x^1 \cos\theta_{x'}^2 \sin\theta_{x'}^1 \sin\theta_x^2 \sin\theta_{x'}^3 \\ & + e^{i\phi_x^3 - i\phi_{x'}^4 - i\phi_{x'}^5} \sin\theta_x^1 \sin\theta_{x'}^2 \sin\theta_{x'}^3. \end{aligned} \tag{71}$$

The magnitude of the first, second, and fourth terms are reduced by shifting $\phi_x^1 \to \phi_x^1 + i\lambda$ and $\phi_{x'}^1 \to \phi_{x'}^1 + i\lambda$ with $\lambda > 0$. The magnitude of the second and third terms can be reduced by shifting $\phi_x^3 - \phi_{x'}^3 \to \phi_x^3 - \phi_{x'}^3 + i\delta$ and $\phi_x^4 - \phi_{x'}^4 \to \phi_x^4 - \phi_{x'}^4 + i\delta$ with $\delta > 0$; this is also consistent with a positive imaginary shift of $i\delta$ in $\phi_x^3 - \phi_{x'}^4$ and $\phi_x^4 - \phi_{x'}^3$, reducing the magnitude of the fourth and fifth terms. These deformations result in reduced magnitude and correspondingly lower variance. Deformations with these qualitative features are reproduced in the optimized manifolds found for $\Lambda = 0$. Finally, we note that $\phi_{x'}^2$ appears in the exponent with opposite signs in the second and third terms, and similarly for $\phi_{x'}^5$ in the fourth and fifth terms, so there is no constant vertical deformation of these terms that will reduce the overall magnitude.

Analogously to the case of $SU(2)$ gauge theory, no significant StN improvements were obtained by including higher Fourier modes. Figure 11 directly compares optimized manifolds for $\Lambda \leq 2$ at the coarsest lattice spacing,

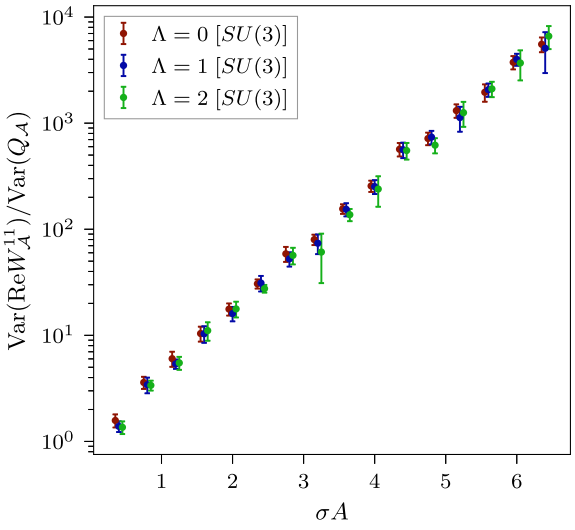

FIG. 11. $SU(3)$ Wilson loop variance ratios of standard observables to deformed observables for ensembles with $g = 0.53$ and three different values of the manifold parametrization cutoff.

showing that the variance improvement was unchanged by including higher Fourier modes. Constant vertical deformations therefore appear to saturate the StN benefits achieved by general Fourier series parametrizations for deformed Wilson loops in $(1+1)$D. More complicated parametrizations of contour deformations may prove more useful in $(3+1)$D; however, numerical studies of deformed observables in $(3+1)$D lattice gauge theory are left to future work.

VI. CONCLUSIONS

In this work, we have defined a family of complex manifolds for path integral deformations in $SU(N)$ lattice field theories. The manifolds introduced here are described in terms of an angular parametrization of each $SU(N)$ variable, with dependence between variables at differing spacetime sites restricted to enforce a triangular Jacobian. We find that choosing a parametrization in which the observable of interest can be written as $\mathcal{O} = e^{i\theta}X$ is a useful practical choice, allowing constant shift deformations in the parameter θ to make substantial progress in reducing noise. Choosing a spacetime dependence of the deformation that gives rise to a triangular Jacobian is key to ensuring the deformed integral can be computed efficiently, as the Jacobian determinant can then be evaluated with cost linear in the number of spacetime lattice sites.

This manifold parametrization can be combined with the method of deformed observables introduced in Ref. [48] to reduce noise in observables. This approach is applicable when the action is real and the Boltzmann weight e^{-S} can be treated as a probability measure. We stress that this method of deformed observables does not require changing the Monte Carlo sampling, despite being based on an analysis of contour deformation of the entire path integral, and can be thought of as an approach to analytically relate observables with identical expectation values and different variance. Keeping the Monte Carlo weights unchanged allows manifold optimization using estimates of the deformed observable variance computed with respect to a fixed Monte Carlo ensemble. There is a tradeoff between the cost of optimizing manifold parameters and the statistical precision gained. In practice, we find that initializing manifold parameters from optimal parameters for similar observables significantly reduces the associated cost.

This method was shown in Sec. IV and V to improve the variance of Wilson loop observables in $(1+1)$D $SU(2)$ and $SU(3)$ lattice gauge theory by orders of magnitude. For the original Wilson loop observables, the signal-to-noise ratio decreases exponentially with area. The deformed observables mitigate this StN problem, and in particular we find that the improvements are consistent with an exponential in the physical Wilson loop area, with the most significant reduction in noise for Wilson loops of the largest area. The improvement in variance was empirically found to be similar at three different lattice spacings,

though less of an improvement was seen at the finest lattice spacing for $SU(3)$; the achieved improvements in the continuum limit and for other theories is thus an interesting subject of future investigation. However, we stress that making any gains at finite lattice spacing is still a significant step forward due to the convenience of the method: optimizing a deformation on a fixed ensemble quickly gives new observables that encode the same physical content while having significantly reduced noise.

In demonstrating the method, we focused here on a particular deformation of the angular parameters based on a Fourier series expansion and shift in the imaginary direction only. Writing the observable phase fluctuations in terms of these periodic angular parameters that can be shifted by a constant in the imaginary direction led to deformed observables with significant StN improvement. The surprising result that zero-mode terms alone significantly reduce noise, with neither dependence between plaquettes at different spacetime sites nor dependence on the values of the angular parameters themselves, suggests that the majority of the StN problem in these $(1+1)$D theories arises from independent local fluctuations of $SU(N)$ angular parameters.

Complications are expected in higher dimensions, as Gauss' law implies that plaquettes at differing spacetime locations and orientations must satisfy many independent constraints. Deformations thus cannot independently address fluctuations in each plaquette included in a Wilson loop, or more generally in each fundamental degree of freedom included in an observable. It is therefore an interesting line of future work to determine how best to incorporate this spacetime dependence in higher dimensional applications of path integral contour deformations. The approach employed here is one of many possible approaches to creating expressive transformations with efficiently computable Jacobians. This issue has been explored in some depth in normalizing flows for sampling in many contexts, including image generation and ensemble generation for lattice field theory; see Refs. [65,84] for recent reviews. Similar techniques may prove more fruitful in future applications of path integral contour deformations to observables of phenomenological interest in $(3+1)$D lattice gauge theories.

ACKNOWLEDGMENTS

W. D. and G. K. are supported in part by the U.S. DOE Grant No. DE-SC0011090. W. D. is also supported by the SciDAC4 Award No. DE-SC0018121. H. L. is supported by a Department of Energy QuantiSED grant. N. C. W. is supported by the U.S. DOE under Grant No. DE-FG02-00ER41132. This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

APPENDIX A: SINGLE-VARIABLE $SU(N)$ INTEGRALS

The calculation of z for $SU(N)$ can be performed analogously to the $U(N)$ case considered in Refs. [71,72] with further details for the $SU(N)$ case given in Ref. [73]. In the eigenbasis of P, the Haar measure is given by

$$dP = \frac{1}{N!}\prod_{I=1}^{N}\left[\frac{d\theta_I}{2\pi}\prod_{J<I}|e^{i\theta_I}-e^{i\theta_J}|^2 \times \sum_{n=-\infty}^{\infty} 2\pi\delta\left(\sum_{K=1}^{N}\theta_K + 2\pi n\right)\right], \tag{A1}$$

where the δ-function enforces the unit determinant condition of $SU(N)$. Using the Fourier series representation of the δ-function for the compact variable $\sum_{K=1}^{N}\theta_K$, this can be expressed as

$$dP = \frac{1}{N!}\prod_{I=1}^{N}\left[\frac{d\theta_I}{2\pi}\prod_{J<I}|e^{i\theta_I}-e^{i\theta_J}|^2\sum_{q=-\infty}^{\infty}e^{iq\theta_I}\right]. \tag{A2}$$

The product of $e^{i\theta_I}-e^{i\theta_J}$ factors can be expressed in terms of the determinant of a Vandermonde matrix [71,72]. The $SU(N)$ Haar measure is given by a sum of similar determinants, and z can be expressed as [73]

$$z = \sum_{q=-\infty}^{\infty}\det(\mathcal{Z}^q), \tag{A3}$$

where the entries of the matrix $\mathcal{Z}^q$ are given by

$$\mathcal{Z}^q_{IJ} \equiv \int\frac{d\theta}{2\pi}e^{i[q+I-J]\theta}e^{\frac{2}{g^2}\cos(\theta)} = I_{q+I-J}\left(\frac{2}{g^2}\right), \tag{A4}$$

where $I_n(x)$ is a modified Bessel function. For example, in the $SU(2)$ case z is given explicitly by

$$z^{SU(2)} = \sum_{q=-\infty}^{\infty}\left[I_q\left(\frac{2}{g^2}\right)\right]^2 - I_{q+1}\left(\frac{2}{g^2}\right)I_{q-1}\left(\frac{2}{g^2}\right). \tag{A5}$$

A simpler but equivalent form $z^{SU(2)} = \frac{g^2}{2}I_1(4/g^2)$ can also be derived using the parametrization introduced in Sec. IV. From this, we can derive an expression for $\langle\chi_1\rangle$ from taking derivatives of z

$$\frac{\partial}{\partial(2N/g^2)}\log z = -\frac{g^3}{4N}\frac{\partial}{\partial g}\log z = \langle\chi_1\rangle = e^{-\sigma}. \tag{A6}$$

Similar to $\langle\mathrm{tr}(W_\mathcal{A})/N\rangle$, we can factorize $\langle|\mathrm{tr}(W_\mathcal{A})^2|/N^2\rangle$ and $\langle\mathrm{tr}(W_\mathcal{A})^2/N^2\rangle$ into integrals of character functions over single-elements of $SU(N)$ using an identity [85]

$$\int d\Omega \Omega_{i_1 j_1} \Omega^\dagger_{k_1 l_1} \Omega_{i_2 j_2} \Omega^\dagger_{k_2 l_2} = \frac{1}{N^2-1}(\delta_{i_1 l_1}\delta_{j_1 k_1}\delta_{i_2 l_2}\delta_{j_2 k_2} + \delta_{i_1 l_2}\delta_{j_1 k_2}\delta_{i_2 l_1}\delta_{j_2 k_1}) - \frac{1}{N(N^2-1)}(\delta_{i_1 l_2}\delta_{j_1 k_1}\delta_{i_2 l_1}\delta_{j_2 k_2} + \delta_{i_1 l_1}\delta_{j_1 k_2}\delta_{i_2 l_2}\delta_{j_2 k_1}). \tag{A7}$$

From Table II, we recognize that $\langle|\mathrm{tr}(W_\mathcal{A})^2/N^2|\rangle$ is related to $\chi_{1,-1}(P_x)$. Thus, applying Eq. (A7) we can derive that

$$\begin{aligned}\int d\Omega \chi_{1,-1}(A\Omega P \Omega^\dagger B) &= \mathrm{tr}(AB)\mathrm{tr}(B^\dagger A^\dagger)\langle\chi_{1,-1}\rangle \\ &\quad + \mathrm{tr}(AA^\dagger B^\dagger B)\frac{1}{N}(1-\langle\chi_{1,-1}\rangle) - 1 \\ &= [\mathrm{tr}(AB)tr(B^\dagger A^\dagger) - 1]\langle\chi_{1,-1}\rangle \\ &= \chi_{1,-1}(AB)\langle\chi_{1,-1}\rangle \end{aligned} \tag{A8}$$

where

$$\langle\chi_{1,-1}\rangle \equiv \frac{1}{z}\int dP \frac{1}{N^2-1}\chi_{1,-1}(P) e^{\frac{1}{g^2}\mathrm{tr}(P+P^\dagger)}. \tag{A9}$$

Iterating this identity within $\langle|\mathrm{tr}(W_\mathcal{A})^2|\rangle$ gives

$$\begin{aligned}&\langle|\mathrm{tr}(W_\mathcal{A})^2|\rangle \\ &= N^2\langle\chi_{1,-1}\rangle^A + (1-\langle\chi_{1,-1}\rangle)\sum_{k=0}^{A-1}\left(\frac{\langle\chi_{1,-1}\rangle - 1}{N^2}\right)^k \\ &= 1 + (N^2-1)\langle\chi_{1,-1}\rangle^A. \end{aligned} \tag{A10}$$

Since $\mathrm{tr}(W_\mathcal{A})^2$ is not a character χ_r, attempting to factorize it generates new terms. Specifically, in addition to $\mathrm{tr}(W_{\mathcal{A}'})^2$ one finds $\mathrm{tr}(W^2_{\mathcal{A}'})$ for $\mathcal{A}' \subset \mathcal{A}$. Instead, a basis involving $\chi_2(P)$ and $\chi_{1,-1}(P)$ can be constructed from linear combinations of these traces and satisfies

$$\int d\Omega \chi_2(A\Omega P\Omega^\dagger B) = \frac{2}{N(N+1)}\chi_2(AB)\langle\chi_2\rangle, \tag{A11}$$

and

$$\int d\Omega \chi_{1,1}(A\Omega P\Omega^\dagger B) = \frac{2}{N(N-1)}\chi_{1,1}(AB)\langle\chi_{1,1}\rangle, \tag{A12}$$

where as in Eq. (A9) we have factorized using the expectation values of the characters

$$\begin{aligned}\langle\chi_2\rangle &\equiv \frac{1}{z}\int dP \frac{2}{N(N+1)}\chi_2(P) e^{\frac{1}{g^2}\mathrm{tr}(P+P^\dagger)}, \\ \langle\chi_{1,1}\rangle &\equiv \frac{1}{z}\int dP \frac{2}{N(N-1)}\chi_{1,1}(P) e^{\frac{1}{g^2}\mathrm{tr}(P+P^\dagger)}. \end{aligned} \tag{A13}$$

TABLE II. Properties of group representations: the dimension d_r and the character χ_r. We have followed the normalizations in Table 14 of Ref. [73].

r	d_r	χ_r
$\{1\}$	N	$\mathrm{tr}U$
$\{2\}$	$\frac{N(N+1)}{2}$	$\frac{1}{2}(\mathrm{tr}^2U + \mathrm{tr}U^2)$
$\{1,1\}$	$\frac{N(N-1)}{2}$	$\frac{1}{2}(\mathrm{tr}^2U - \mathrm{tr}U^2)$
$\{1,-1\}$	N^2-1	$\lvert\mathrm{tr}U\rvert^2 - 1$

Putting these together and iterating gives

$$\langle\mathrm{tr}(W_\mathcal{A})^2\rangle = \frac{N(N+1)}{2}\langle\chi_2\rangle^A + \frac{N(N-1)}{2}\langle\chi_{1,1}\rangle^A. \tag{A14}$$

Combining Eqs. (A10) and (A14) gives the general expression for variance of the $SU(N)$ Wilson loop

$$\begin{aligned}&\mathrm{Var}[\mathrm{Retr}(W^{SU(N)}_\mathcal{A})/N] \\ &= \frac{1}{2N^2}\langle|\mathrm{tr}(W_\mathcal{A})^2|\rangle + \frac{1}{2N^2}\langle\mathrm{tr}(W_\mathcal{A})^2\rangle - \frac{1}{N^2}\langle\mathrm{tr}(W_\mathcal{A})\rangle^2 \\ &= \frac{1}{2N^2}\left[1 + (N^2-1)\langle\chi_{1,-1}\rangle^A + \frac{N(N+1)}{2}\langle\chi_2\rangle^A \right. \\ &\quad \left. + \frac{N(N-1)}{2}\langle\chi_{1,1}\rangle^A\right] - e^{-2\sigma^{SU(N)}A}. \end{aligned} \tag{A15}$$

The character expectation values can be obtained for general $SU(N)$ gauge groups by numerically evaluating the integrals in Eq. (A13).

For $SU(2)$, it is straightforward to analytically evaluate the character expectation values appearing in Eq. (A13). Not all representations of $SU(2)$ are independent using the enumeration in Table II, and in particular $\chi_2(P) = \chi_{1,-1}(P)$ and $\chi_{1,1}(P) = \chi_0(P) = 1$. After removing the redundant characters, the remain integrals $\chi_j(P)$ can be solved using expression of the characters in terms of angles [73]

$$\chi_j = \frac{\sin([j+1/2]\alpha)}{\sin(\alpha/2)} \tag{A16}$$

where $j = 0, \frac{1}{2}, 1\cdots$ index the unique characters of $SU(2)$ and $r = \{1,-1\}$ equals $j = 1$. With these, one has

$$\frac{1}{z}\int DP \frac{1}{d_j}\chi_j(P) e^{\frac{1}{g^2}\mathrm{Tr}(P+P^\dagger)} = \frac{I_{2j+1}(4/g^2)}{I_1(4/g^2)}. \tag{A17}$$

From which we derive

$$\mathrm{Var}[\mathrm{Re}\,\mathrm{tr}(W^{SU(2)}_\mathcal{A})/N] = \frac{1}{4} + \frac{3}{4}\left(\frac{I_3(4/g^2)}{I_1(4/g^2)}\right)^A - e^{-2\sigma^{SU(2)}A}, \tag{A18}$$

APPENDIX B: ALTERNATIVE $SU(2)$ COORDINATES

As an alternative to the parametrization presented in Sec. IV A, plaquettes $P_x \in SU(2)$ can be represented as

$$P_x = \exp\left(\frac{i\alpha_x}{2}\hat{n}_x \cdot \vec{\sigma}\right) = \cos\left(\frac{\alpha_x}{2}\right) + i\sin\left(\frac{\alpha_x}{2}\right)\hat{n}_x \cdot \vec{\sigma}. \tag{B1}$$

The $\mathfrak{su}(2)$ unit vector $\hat{n}_x$ can be further parametrized as

$$\hat{n}_x = (\cos\phi_x \sin\theta_x, \sin\phi_x \sin\theta_x, \cos\theta_x), \tag{B2}$$

and a general $SU(2)$ group element P_x can be parametrized in terms of the three angles

$$0 \le \alpha_x < 2\pi, \qquad 0 \le \phi_x < 2\pi, \qquad 0 \le \theta_x < \pi. \tag{B3}$$

The Haar measure is given in these coordinates as

$$dP_x = \frac{1}{4\pi^2}\sin^2\left(\frac{\alpha_x}{2}\right) d\alpha_x \sin\theta_x d\theta_x d\phi_x, \tag{B4}$$

The inverse map needed to obtain these angular parameters for an $SU(2)$ matrix P_x is given by

$$\begin{aligned} \alpha_x &= 2\arccos\left[\frac{1}{2}(P_x^{11} + P_x^{22})\right] \\ \theta_x &= \arccos\left[\frac{P_x^{11} - P_x^{22}}{2i\sin(\alpha_x/2)}\right] \\ \phi_x &= \frac{1}{2}\arg\left[\frac{P_x^{21}}{P_x^{12}}\right]. \end{aligned} \tag{B5}$$

As with Eq. (60), these are not entire functions of P_x but this is not an obstacle for contour deformation because it is only the parametrization given by Eqs. (B1) and (B2) that determines whether path integrands can be interpreted as holomorphic functions of the angles $\{\alpha_x, \theta_x, \phi_x\}$ associated with P_x.

Deformed observables starting with the (1,1) component of $SU(2)$ Wilson loops can be defined using this parametrization. A family of vertical deformations for $\{\alpha_x, \theta_x, \phi_x\}$ can be defined analogously to the deformation described in Sec. IV B. Since α_x and θ_x have fixed (nonidentified) integration contour endpoints, a constant vertical deformation can only be applied to ϕ_x. For $A = 1$ in particular, $\mathrm{Tr}(P_x) = \cos(\alpha_x/2)$, and the trace is independent of the only constant vertical deformation that can be applied. Neither this constant vertical deformation nor nonconstant vertical deformations corresponding to Fourier basis cutoffs $\Lambda = 1, 2$ lead to statistically significant variance reduction with $A = 1$. As shown in Fig. 12, for $A > 1$ deformed observable results using this parametrization with $\Lambda = 0$ do lead to significant variance reduction when compared to undeformed contour results. However, orders of magnitude less variance reduction is obtained for large area Wilson loops using optimized deformed observables with this parametrization when compared to results using the parametrization explored in Sec. IV A. The fact that the parametrization in Eq. (B1) leads to less variance reduction than the parametrization in Sec. IV A can be intuitively explained by the inability of constant vertical deformations to decrease the magnitudes of the (1,1) components of (products of) $SU(2)$ matrices using the parametrization Eq. (B1). The significance of the difference between the results demonstrates the utility of rewriting observables before deformation in achieving practical StN improvements, as discussed in Sec. II D.

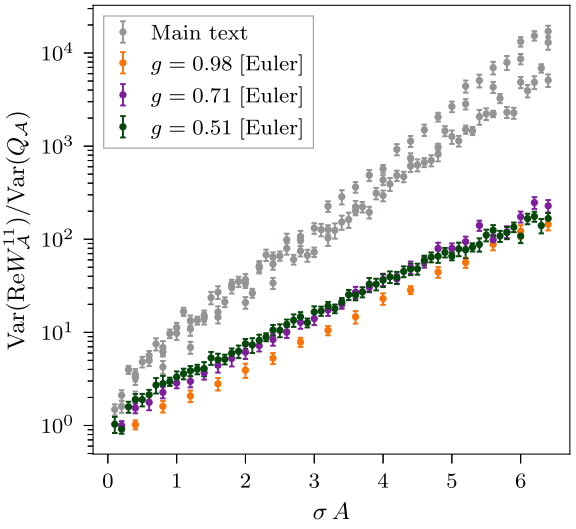

FIG. 12. Colored points $SU(2)$ Wilson loop variance ratios using the alternative gauge field parametrization defined in Eq. (B1). Gray points show analogous variance ratios using the parametrization defined in Sec. IV A for comparison and are identical to the results in Fig. 5.

APPENDIX C: REGULARIZATION TERMS TO AVOID OVERTRAINING AND OVERLAP PROBLEMS

When $\mathrm{Re}\,\tilde{S}$ is significantly different from S, we can encounter an overlap problem for training and evaluation; both processes involve factors of $e^{-\mathrm{Re}\tilde{S}+S}$ that can have very large magnitude fluctuations in this situation. To mitigate this problem, it is helpful to include regularization terms in the loss function $\mathcal{L}$. These terms may bias the exact loss minimum away from the optimal, but allow closer convergence to that optimal solution given finite statistics estimates of $\mathcal{L}$. The strength of these terms is controlled by a small parameter ϵ.

We discuss two possible terms here. First, an L2 regularizer [86] may be used, which simply ensures the

parameters controlling the deformation all remain close to zero. Generically labeling those parameters as λ_i, this loss term can be written

$$\mathcal{L}_{\mathrm{L2}} \equiv \epsilon \sum_i |\lambda_i|^2. \tag{C1}$$

In the limit of $\epsilon \to \infty$, the parameters λ_i are forced to zero and the optimization procedure must remain at the original manifold. A smaller choice of ϵ mildly biases the optimization procedure toward the original manifold, such that the loss function and gradients remain feasible to estimate with finite statistics. An alternate approach is to directly penalize distance between $\mathrm{Re}\tilde{S}$ and S using a regularization term,

$$\mathcal{L}_{\mathrm{act}} \equiv \epsilon \frac{1}{Z} \int dx\, e^{-S(x)} |S(x) - \mathrm{Re}\tilde{S}(x)|. \tag{C2}$$

This term is minimized when $S = \mathrm{Re}\,\tilde{S}$, providing a bias toward remaining close to the original manifold. Though written as a path integral, this quantity can be estimated using the original samples, much like the main loss function and gradients. Both of these regularizer terms were explored, however no severe overlap problem was observed during training when deformation parameters were restricted to those contained within the target Wilson loop observables. This is indicated in Fig. 13, where the effective sample size is shown for the intermediate lattice spacings in $SU(2)$ and $SU(3)$. Final results are based on training without either regularizer term.

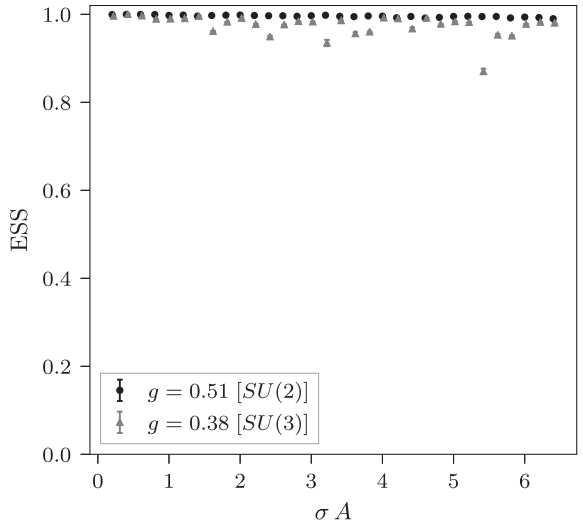

FIG. 13. Plot of the effective sample size (ESS) for the intermediate lattice spacing for both the $SU(2)$ and $SU(3)$ gauge groups. The ESS remaining close to unity indicates that no substantial overlap problem is induced upon manifold deformation.

[1] S. Aoki *et al.* (Flavour Lattice Averaging Group), Eur. Phys. J. C **80**, 113 (2020).
[2] W. Detmold, R. G. Edwards, J. J. Dudek, M. Engelhardt, H.-W. Lin, S. Meinel, K. Orginos, and P. Shanahan (USQCD Collaboration), Eur. Phys. J. A **55**, 193 (2019).
[3] C. Lehner *et al.* (USQCD Collaboration), Eur. Phys. J. A **55**, 195 (2019).
[4] V. Cirigliano, Z. Davoudi, T. Bhattacharya, T. Izubuchi, P. E. Shanahan, S. Syritsyn, and M. L. Wagman (USQCD Collaboration), Eur. Phys. J. A **55**, 197 (2019).
[5] T. Aoyama *et al.*, Phys. Rep. **887**, 1 (2020).
[6] M. L. Wagman and M. J. Savage, Phys. Rev. D **96**, 114508 (2017).
[7] M. L. Wagman, Statistical angles on the lattice QCD signal-to-noise problem, Ph.D. thesis, U. Washington, Seattle (main), 2017.
[8] G. Parisi, Phys. Rep. **103**, 203 (1984).
[9] G. P. Lepage, in *Boulder TASI 1989:97-120* (World Scientific Publishing Company, Boulder, 1989), pp. 97–120.
[10] S. R. Beane, W. Detmold, T. C. Luu, K. Orginos, A. Parreño, M. J. Savage, A. Torok, and A. Walker-Loud, Phys. Rev. D **79**, 114502 (2009).
[11] S. R. Beane, W. Detmold, K. Orginos, and M. J. Savage, J. Phys. G **42**, 034022 (2015).
[12] Z. Davoudi, W. Detmold, K. Orginos, A. Parreño, M. J. Savage, P. Shanahan, and M. L. Wagman, Phys. Rep. **900**, 1 (2021).
[13] P. E. Gibbs, Phys. Lett. B **182**, 369 (1986).
[14] T. D. Cohen, Phys. Rev. Lett. **91**, 222001 (2003).
[15] T. D. Cohen, Phys. Rev. Lett. **91**, 032002 (2003).
[16] K. Splittorff and J. J. M. Verbaarschot, Phys. Rev. Lett. **98**, 031601 (2007).
[17] K. Splittorff and J. J. M. Verbaarschot, Phys. Rev. D **75**, 116003 (2007).
[18] P. de Forcrand, *Proc. Sci.*, LAT2009 (2009) 010 [arXiv:1005.0539].
[19] A. Alexandru, C. Gattringer, H. P. Schadler, K. Splittorff, and J. J. M. Verbaarschot, Phys. Rev. D **91**, 074501 (2015).
[20] M. Cristoforetti, F. Di Renzo, and L. Scorzato (AuroraScience Collaboration), Phys. Rev. D **86**, 074506 (2012).
[21] G. Aarts, Phys. Rev. D **88**, 094501 (2013).
[22] M. Cristoforetti, F. Di Renzo, A. Mukherjee, and L. Scorzato, Phys. Rev. D **88**, 051501 (2013).
[23] A. Mukherjee, M. Cristoforetti, and L. Scorzato, Phys. Rev. D **88**, 051502 (2013).
[24] G. Aarts, L. Bongiovanni, E. Seiler, and D. Sexty, J. High Energy Phys. 10 (2014) 159.

[25] M. Cristoforetti, F. Di Renzo, G. Eruzzi, A. Mukherjee, C. Schmidt, L. Scorzato, and C. Torrero, Phys. Rev. D **89**, 114505 (2014).
[26] A. Alexandru, G. Başar, and P. Bedaque, Phys. Rev. D **93**, 014504 (2016).
[27] A. Alexandru, G. Başar, P. F. Bedaque, G. W. Ridgway, and N. C. Warrington, J. High Energy Phys. 05 (2016) 053.
[28] A. Alexandru, G. Başar, P. F. Bedaque, S. Vartak, and N. C. Warrington, Phys. Rev. Lett. **117**, 081602 (2016).
[29] H. Fujii, S. Kamata, and Y. Kikukawa, J. High Energy Phys. 12 (2015) 125; 09 (2016) 172(E).
[30] Y. Tanizaki, Y. Hidaka, and T. Hayata, New J. Phys. **18**, 033002 (2016).
[31] A. Alexandru, P. F. Bedaque, H. Lamm, and S. Lawrence, Phys. Rev. D **96**, 094505 (2017).
[32] A. Alexandru, G. Başar, P. F. Bedaque, and G. W. Ridgway, Phys. Rev. D **95**, 114501 (2017).
[33] Y. Mori, K. Kashiwa, and A. Ohnishi, Prog. Theor. Exp. Phys. **2018**, 023B04 (2018).
[34] Y. Tanizaki, H. Nishimura, and J. J. M. Verbaarschot, J. High Energy Phys. 10 (2017) 100.
[35] A. Alexandru, P. F. Bedaque, and N. C. Warrington, Phys. Rev. D **98**, 054514 (2018).
[36] A. Alexandru, G. Başar, P. F. Bedaque, H. Lamm, and S. Lawrence, Phys. Rev. D **98**, 034506 (2018).
[37] A. Alexandru, P. F. Bedaque, H. Lamm, and S. Lawrence, Phys. Rev. D **97**, 094510 (2018).
[38] A. Alexandru, P. F. Bedaque, H. Lamm, S. Lawrence, and N. C. Warrington, Phys. Rev. Lett. **121**, 191602 (2018).
[39] K. Kashiwa, Y. Mori, and A. Ohnishi, Phys. Rev. D **99**, 014033 (2019).
[40] M. Fukuma, N. Matsumoto, and N. Umeda, arXiv:1912.13303.
[41] M. Fukuma, N. Matsumoto, and N. Umeda, Phys. Rev. D **100**, 114510 (2019).
[42] K. Kashiwa, Y. Mori, and A. Ohnishi, Phys. Rev. D **99**, 114005 (2019).
[43] Z.-G. Mou, P. M. Saffin, and A. Tranberg, J. High Energy Phys. 11 (2019) 135.
[44] M. Ulybyshev, C. Winterowd, and S. Zafeiropoulos, Phys. Rev. D **101**, 014508 (2020).
[45] S. Lawrence, arXiv:2006.03683.
[46] S. Lawrence and Y. Yamauchi, arXiv:2101.05755.
[47] A. Alexandru, G. Basar, P. F. Bedaque, and N. C. Warrington, arXiv:2007.05436.
[48] W. Detmold, G. Kanwar, M. L. Wagman, and N. C. Warrington, Phys. Rev. D **102**, 014514 (2020).
[49] F. Di Renzo and G. Eruzzi, Phys. Rev. D **97**, 014503 (2018).
[50] C. Schmidt and F. Ziesché, *Proc. Sci.*, LATTICE2016 (2017) 076 [arXiv:1701.08959].
[51] A. Ohnishi, Y. Mori, and K. Kashiwa, J. Phys. Soc. Jpn. Conf. Proc. **26**, 024011 (2019).
[52] K. Zambello and F. D. Renzo, *Proc. Sci.*, LATTICE2018 (2018) 148 [arXiv:1811.03605].
[53] J. M. Pawlowski, M. Scherzer, C. Schmidt, F. P. G. Ziegler, and F. Ziesché, arXiv:2101.03938.
[54] M. Lüscher and P. Weisz, J. High Energy Phys. 09 (2001) 010.
[55] R. M. Range, *Holomorphic Functions and Integral Representations in Several Complex Variables*, Vol. 108 (Springer Science & Business Media, New York, 2013).
[56] J. Bronzan, Phys. Rev. D **38**, 1994 (1988).
[57] A. Alexandru, P. F. Bedaque, H. Lamm, and S. Lawrence, Phys. Rev. D **97**, 094510 (2018).
[58] K. Kashiwa and Y. Mori, Phys. Rev. D **102**, 054519 (2020).
[59] A. Alexandru, G. Basar, P. F. Bedaque, G. W. Ridgway, and N. C. Warrington, Phys. Rev. D **95**, 014502 (2017).
[60] D. Bump, Complexification, in *Lie Groups* (Springer, New York, NY, 2013), pp. 205–211.
[61] G. Aarts and I.-O. Stamatescu, J. High Energy Phys. 09 (2008) 018.
[62] D. Sexty, Nucl. Phys. **A931**, 856 (2014).
[63] E. Seiler, EPJ Web Conf. **175**, 01019 (2018).
[64] A. Alexandru, P. F. Bedaque, H. Lamm, S. Lawrence, and N. C. Warrington, Phys. Rev. Lett. **121**, 191602 (2018).
[65] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, arXiv:1912.02762.
[66] H. Robbins and S. Monro, Ann. Math. Stat. **22**, 400 (1951).
[67] V. S. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint*, Vol. 48 (Springer, New York, 2009), Chap. 2.
[68] J. Chee and P. Toulis, arXiv:1710.06382.
[69] L. Bottou, F. E. Curtis, and J. Nocedal, arXiv:1606.04838.
[70] K. G. Wilson, Phys. Rev. D **10**, 2445 (1974).
[71] S. R. Wadia, arXiv:1212.2906.
[72] D. Gross and E. Witten, Phys. Rev. D **21**, 446 (1980).
[73] J.-M. Drouffe and J.-B. Zuber, Phys. Rep. **102**, 1 (1983).
[74] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, JAX: Composable transformations of Python + NumPy programs (2018).
[75] D. P. Kingma and J. Ba, arXiv:1412.6980.
[76] M. Hardt, B. Recht, and Y. Singer, arXiv:1509.01240.
[77] J. Lin, R. Camoriano, and L. Rosasco, in *Proceedings of Machine Learning Research*, Vol. 48, edited by M. F. Balcan and K. Q. Weinberger (PMLR, New York, New York, USA, 2016), pp. 2340–2348.
[78] C. Zhang, O. Vinyals, R. Munos, and S. Bengio, arXiv:1804.06893.
[79] S. Raschka, arXiv:1811.12808.
[80] S. Duane, A. Kennedy, B. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216 (1987).
[81] M. Creutz, Phys. Rev. D **21**, 2308 (1980).
[82] E. Pietarinen, Nucl. Phys. **B190**, 270 (1981).
[83] P. de Forcrand and O. Jahn, in *Proceedings of the 3rd International Workshop on Numerical Analysis and Lattice QCD* (Lecture Notes in Computational Science and Engineering, Edinburgh, 2005), pp. 67–73.
[84] I. Kobyzev, S. Prince, and M. Brubaker, IEEE Trans. Pattern Anal. Mach. Intell. 1 (2020).
[85] M. Creutz, J. Math. Phys. (N.Y.) **19**, 2043 (1978).
[86] A. Krogh and J. A. Hertz, in *Proceedings of the 4th International Conference on Neural Information Processing Systems, NIPS'91* (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1991), pp. 950–957.