Learned integration contour deformation for signal-to-noise improvement in Monte Carlo calculations

William Detmold[1,2] **Gurtej Kanwar**[3]
Yin Lin[1,2] **Phiala E. Shanahan**[1,2] **Michael L. Wagman**[4]
[1]Center for Theoretical Physics, Massachusetts Institute of Technology
[2]The NSF AI Institute for Artificial Intelligence and Fundamental Interactions
[3]Albert Einstein Center, Institute for Theoretical Physics, University of Bern
[4]Fermi National Accelerator Laboratory

Abstract

Calculations of the strong nuclear interactions, encoded in the theory of Quantum Chromodynamics (QCD), are extraordinarily computationally demanding. In particular, the Monte Carlo integration used in lattice field theory calculations in this context suffers from severe signal-to-noise challenges. Complexifying the integration manifold with the complex contour deformation method reduces the variances of observables while guaranteeing the exactness of the results. In this work, we use convolutional neural networks to parametrize the deformed manifolds and demonstrate orders-of-magnitude reduction in the variance of a key observable (the Wilson loop) in a simplified model of QCD in three spacetime dimensions.

1 Introduction

Quantum chromodynamics (QCD) is the theory of the strong force between quarks and gluons which underpins nuclear physics as we know it today. Solving the theory analytically in the low energy regime relevant for nuclear physics is infeasible. Instead, QCD can be formulated on discretized spacetime lattices—a method known as *Lattice QCD*—allowing one to extract predictions from QCD by numerically computing path integrals with Monte Carlo integration [17]. This is presently the only *ab initio* technique available to study QCD. To date, lattice QCD calculations have offered rich insights into the non-perturbative physics of QCD with many cutting-edge results available [9, 19].

In lattice QCD calculations, physical quantities are estimated from the Monte Carlo integration of path integrals with samples drawn from probability density distributions of discretized QCD fields. Despite the tremendous success of this paradigm, the notorious *signal-to-noise* problems in Monte Carlo estimates have set the limits on the precision of many lattice QCD results. More concretely, let $\langle\cdot\rangle$ denote the expectation value and $\mathrm{Var}[\cdot]$ the variance. The signal-to-noise ratios (S/N) of many physical observables $O(l)$ characterized by some geometric scale l are expected to decay exponentially as [20, 22]

$$S/N[O(l)] \coloneqq \frac{|\langle O(l)\rangle|}{\sqrt{\mathrm{Var}[O(l)]}} \sim e^{-cl}, \tag{1}$$

where c is a constant. Exponentially more Monte Carlo measurements are thus needed for physics observables at large l to reach the same statistical precision as at small l. Precisely constraining long-distance physics is one of the most significant challenges facing lattice calculations, including for quantities of key current importance such as the muon $g-2$ [10], in which tensions between experimental and theory results may provide a hint of new physics. Solving or alleviating the problem would thus open doors to many more precise calculations in the future.

Machine Learning and the Physical Sciences Workshop, NeurIPS 2023.

To alleviate the signal-to-noise problem in the Monte Carlo integration of path integrals, we apply the *complex contour deformation* technique based on the Cauchy theorem to deform the integration manifold to complex planes. In doing so, the expectation values remain unchanged and unbiased while the variances vary. Neural networks can then be trained by unsupervised learning to efficiently locate the manifolds that minimize the variances of desired observables, thereby improving the signal-to-noise ratio in equation (1).

In this work, we demonstrate for the first time how to construct and train neural networks to find the deformed integration manifolds that reduce the variances of Wilson loops by orders of magnitude in the three-dimensional $SU(2)$ lattice gauge theory. Wilson loops are key observables that probe the confinement nature of QCD, while the $SU(2)$ lattice gauge theory is a toy model that shares many characteristics with QCD. The method we present here strictly improves the previous method presented by Detmold et al. [14, 15] by relaxing the open boundary condition requirement and generalizing the method to higher spacetime dimensions. In addition, we present a transfer learning scheme that allows us to generalize the learned deformed manifolds from one lattice geometry to another which provides a pathway towards scaling the method for larger lattice volumes that are necessary for realistic lattice QCD calculations.

Related works Detmold et al. [14, 15] first applied the complex contour deformation technique with machine learning techniques to reduce the variances of Wilson loop observables on two dimensional scalar, $U(1)$, and $SU(N)$ lattice gauge theories. References [1–8, 12, 13, 21] applied similar ideas to reduce the severity of sign problems in many lattice models. Alexandru et al. [8] wrote a comprehensive review on the contour deformation technique and its applications to lattice calculations.

2 $SU(2)$ lattice gauge theory in three dimensions

The three-dimensional $SU(2)$ lattice gauge theory is defined on a three-dimensional regular grid with dimensions $L_0 \times L_1 \times L_2$. The degree of freedom in this theory is the gauge link $U_\mu(x)$, which is an $SU(2)$ matrix that is associated with a lattice point $x = (n_0, n_1, n_2)$, $n_i \in [0, 1, ..., L_i - 1]$, $i = 0, 1, 2$, and direction $\mu = 0, 1, 2$. Here we use periodic boundary conditions in all directions such that $U_\mu(x) = U_\mu(x + L_\nu \hat{\nu})$ for $\nu = 0, 1, 2$.

Wilson loops of side lengths τ, $W_{\mu\nu,\tau}(x) \in SU(2)$, encoding information about the confining potential, are defined as

$$W_{\mu\nu,\tau}(x) \coloneqq \prod_{i=0}^{\tau-1} U_\mu(x + i\hat{\mu}) \prod_{j=0}^{\tau-1} U_\nu(x + \tau\hat{\mu} + j\hat{\nu}) \prod_{k=0}^{\tau-1} U_\mu^{-1}(x + k\hat{\mu}) \prod_{l=0}^{\tau-1} U_\nu^{-1}(x + l\hat{\nu}). \tag{2}$$

The expectation values of any lattice observable $O[U]$ defined in terms of the set of gauge links U are given by the path integrals on the lattice as

$$\langle O \rangle \coloneqq \frac{1}{Z} \int d[U] e^{-S[U]} O[U] = \frac{1}{Z} \int \left(\prod_{x,\mu} dU_\mu(x) \right) e^{-S[U]} O[U], \;\; Z \coloneqq \int d[U] e^{-S[U]}. \tag{3}$$

The path integrals integrate over all gauge link degrees of freedom on a given lattice geometry, and the normalized Haar integration measure $dU_\mu(x)$ naturally defines integration on the $SU(2)$ group manifold of link $U_\mu(x)$. The action $S[U] \in \mathbb{R}$ encodes the dynamics of the theory by specifying the integration weights and is given by

$$S[U] \coloneqq -\frac{\beta}{4} \sum_x \sum_{\mu=0}^{2} \sum_{\nu=\mu+1}^{2} \mathrm{Tr} \left(P_{\mu\nu}(x) + P_{\mu\nu}^{-1}(x) \right), \tag{4}$$

where β is a constant and $P_{\mu\nu}(x) \coloneqq W_{\mu\nu,\tau=1}(x)$ is the one-by-one Wilson loop, or 'plaquette'. This high-dimensional integral is typically approached by Monte Carlo sampling over configurations of link variables according to the probability density $e^{-S[U]}/Z$. The sample mean of the observable measured on these Monte Carlo samples then gives an estimate of the expectation value in equation (3) with an associated sample variance that quantifies the error.

3 Model and loss function

Here we aim to decrease the variance of the observable $\langle \mathrm{Re}W_\tau \rangle \coloneqq \langle \mathrm{Re}(W_{01,\tau}(x=0))_{00} \rangle$ which contains all the information about the Wilson loop of size τ-by-τ. The subscript denotes the 00 component of the $SU(2)$ matrix. The signal-to-noise ratio of $\mathrm{Re}W_\tau$ decays exponentially as the area $A = \tau^2$ increases. To mitigate this issue with the complex contour deformation technique, we first introduce a parametrization of the $SU(2)$ matrices and rewrite the path integrals in terms of integration over real parameters. We use the parametrization of $SU(2)$ matrices introduced by Bronzan [11],

$$U_\mu(x) = \begin{pmatrix} \sin\big(\theta_\mu(x)\big)e^{i\phi_{1,\mu}(x)} & \cos\big(\theta_\mu(x)\big)e^{i\phi_{2,\mu}(x)} \\ -\cos\big(\theta_\mu(x)\big)e^{-i\phi_{2,\mu}(x)} & \sin\big(\theta_\mu(x)\big)e^{-i\phi_{1,\mu}(x)} \end{pmatrix}, \tag{5}$$

where $\theta_\mu(x) \in [0, \pi/2]$ and $\phi_{i,\mu}(x) \in [0, 2\pi)$, $i = 1, 2$. In this parametrization, the normalized Haar measure takes the form $dU_\mu(x) = (1/4\pi) d\theta_\mu(x) d\phi_{1,\mu}(x) d\phi_{2,\mu}(x) \sin(2\theta_\mu(x))$ and the path integral can be written as an iterated integral over these real parameters on their respective domains.

The general idea for using the complex contour deformation to improve signal-to-noise ratios is as follows: given an integral over some real manifold $\mathcal{M} \subset \mathbb{R}^n$ for some n, we can continuously deform $\mathcal{M}$ to $\widetilde{\mathcal{M}} \subset \mathbb{C}^n$ by complexifying the variables. Then according to the Cauchy theorem, the integral value evaluated on $\widetilde{\mathcal{M}}$ remains unchanged if the integrand is holomorphic and the boundaries of $\mathcal{M}$ and $\widetilde{\mathcal{M}}$ are compatible. A proof is detailed by Alexandru et al. [8]. $\langle \mathrm{Re}W_\tau \rangle$ is unchanged under the deformation since the integrand in equation (3) is holomorphic and taking the real part commutes with taking the expectation value; on the other hand, the variance $\mathrm{Var}[\mathrm{Re}W_\tau] = \langle (\mathrm{Re}W_\tau)^2 \rangle - \langle \mathrm{Re}W_\tau \rangle^2$ is not holomorphic and can be different on different manifolds.

In this context, the role of the machine-learned model is to define a deformed manifold as a map $\tilde{U}[U]$ between the original and complexified variables. Here we investigate models that generate *constant deformations* [14, 15]

$$\tilde{\theta}_\mu(x) = \theta_\mu(x), \tag{6}$$

$$\tilde{\phi}_{i,\mu}(x) = \phi_{i,\mu}(x) + i\Delta_{i,\mu}(x), \quad (i = 1, 2). \tag{7}$$

Only the ϕ angles are deformed because periodic parameters can be shifted in the imaginary direction by constants without violating the Cauchy theorem as shown by Detmold et al. [14, 15]. The new deformed integral can be evaluated by computing $\langle \mathrm{Re}Q_\tau \rangle$, where $Q_\tau[U] \coloneqq e^{-S[\tilde{U}[U]]+S[U]} W_\tau[U]$ is the deformed observable.

The shift field $\Delta_{i,\mu}(x) \in \mathbb{R}$ can be optimized to minimize the variances of $\mathrm{Re}Q_\tau$. In particular, the loss function for Wilson loops with a side length τ is the unbiased estimator of the non-holomorphic part $\langle (\mathrm{Re}Q_\tau)^2 \rangle$ of the variance of Q_τ,

$$L_\tau = \frac{1}{n-1} \sum_{i=1}^{n} (\mathrm{Re}Q_\tau[U^{(i)}])^2, \tag{8}$$

where $\{U^{(1)}, \cdots, U^{(n)}\}$ is the mini-batch of uncorrelated U's sampled according to the probability density defined in equation (3).

4 Experiment details and results

The goal of the unsupervised learning task is to find the shift field $\Delta_{i,\mu}(x)$ that minimizes the variance of $\mathrm{Re}Q_\tau$. However, due to the limited expressivity of the constant deformation to represent a deformed manifold, we make use of a key symmetry of the action (gauge symmetry) in equation (4) to reformulate the problem in a way that is suitable for constant deformation. In particular, a set of gauge links can always be fixed to the identity matrix $\mathbb{I}$ under suitable gauge symmetry transformations as long as those links do not form any closed loops. We exploit this freedom to fix the gauge links according to

$$\begin{aligned} U_0(n_0 < L_0 - 1, n_1 = 0, n_2 = 0) &= U_1(n_0, n_1 < L_1 - 1, n_2 = 0) \\ &= U_2(n_0, n_1, n_2 < L_2 - 1) = \mathbb{I}, \end{aligned} \tag{9}$$

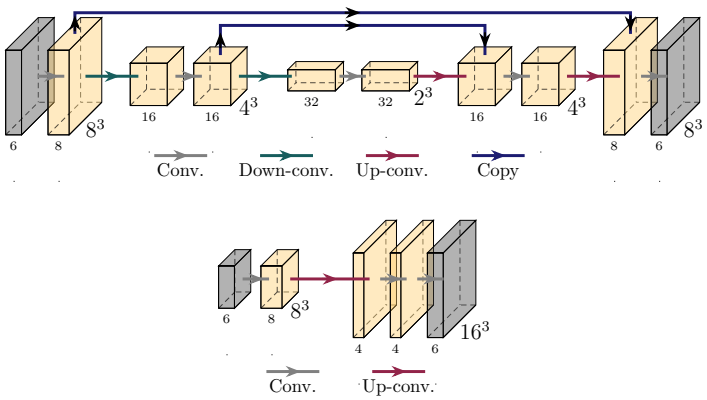

Figure 1: Schematic plots of the U-net (top) and the convolutional neural network for transfer learning (bottom). The widths of boxes show the channel dimensions. “Conv” includes the convolution with a 3-by-3 kernel, batch normalization layer, and ReLU activation function; “Down-conv” down-samples the lattice with a stride-2 convolution and a 3-by-3 kernel; and "Up-conv" up-samples the lattices with a stride-2 transposed convolution and a 2-by-2 kernel.

which empirically yields in the best results for the variance of $\mathrm{Re}Q_\tau$ in the following.

In this work, we perform experiments on three different lattice geometries: 8^3, 16^3, and 32^3. For the 8^3 lattice, we parametrize the shift field $\Delta_{i,\mu}(x)$ with U-nets introduced by Ronneberger et al. [23]. The input to the U-net is the concatenation of two three-channel binary masks: the first mask denotes which links $U_\mu(x)$ ($\mu = 0, 1, 2$) form the Wilson loops on the lattice and the second mask denotes which links are fixed to $\mathbb{I}$ in equation (9). Together they form the six-channel input. The output also has six channels representing the $i = 1, 2$ and $\mu = 0, 1, 2$ indices of $\Delta_{i,\mu}(x)$. The details of the U-net are depicted in figure 1.

For the 16^3 and 32^3 lattices, we generate the shift fields with another convolutional neural network (CNN) that takes the trained shift fields for $\mathrm{Re}Q_\tau$ on lattices of size $L \times L \times L$ as inputs to generate shift fields for $\mathrm{Re}Q_{2\tau}$ on lattices of size $2L \times 2L \times 2L$. This is because training directly on 32^3 lattices with U-nets requires a large number of training samples and iterations. We can instead first train this CNN to transfer shift fields from 8^3 to 16^3 lattices, and then fine-tune this pretrained network to transfer shift fields from 16^3 to 32^3 lattices. We empirically find that this training scheme dramatically decreases the number independent samples and iterations needed for convergence. The details of this network are depicted in figure 1.

All Monte Carlo samples are uncorrelated and generated using the Hamiltonian Monte Carlo algorithm with `Chroma` [16]. We trained these CNNs respectively using 1.5×10^5, 1.1×10^5, and 2.4×10^4 Monte Carlo samples for the 8^3, 16^3, and 32^3 lattice geometries, and evaluated the performance on 500 Monte Carlo samples in each case. We fix $\beta = 3.75$ defining the action $S[U]$. Training is done with the Adam optimizer [18] using one compute node with eight NVIDIA A100 GPUs. The training time for each network is few node-hours. The results are shown in figure 2. The variances of the deformed Wilson loops are improved by orders of magnitude for large areas.

5 Conclusion and outlook

In this work, we have shown for the first time how to construct, train, and scale CNNs to parametrize constant manifold deformations to enhance the signal-to-noise ratios of Wilson loop observables by orders of magnitude in a lattice gauge theory in three dimensions. These results were possible despite restricting to the simple constant deformations described in equation (7). It would be natural to further improve these results by consider a larger class of deformations, for example by allowing field-dependent functions $\Delta_{i,\mu}(x)$ or allowing the $\theta_\mu(x)$ variables to be deformed. Exploring alternate or partial gauge-fixing schemes beyond the choices considered during this study may also significantly further reduce the observable variances.

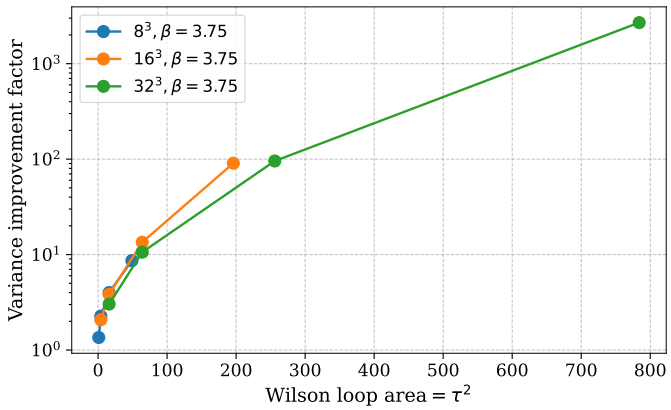

Figure 2: Variance improvement factors as a function of Wilson loop areas evaluated on the test dataset. The variance improvement factor is defined as $\mathrm{Var}(\mathrm{Re}W_\tau)/\mathrm{Var}(\mathrm{Re}Q_\tau)$ where $\mathrm{Re}W_\tau$ is the original Wilson loop observable and $\mathrm{Re}Q_\tau$ is the machine-learned observable. The results demonstrate significant variance reductions of the machine-learned observables compared to the original one. For example, the improvement factor for Wilson loops with an area of $28 \times 28 = 784$ (the largest Wilson loops shown in this figure) is about 3×10^3. This means that for a given target statistical precision of the $\langle \mathrm{Re}W_\tau \rangle$ estimate, we need 3×10^3 times more uncorrelated Monte Carlo samples to achieve the same precision if we use the $\langle \mathrm{Re}W_\tau \rangle$ estimate.

This method can be easily applied to $SU(3)$ lattice gauge theory in four dimensions as is necessary for QCD calculations, and its success in this setting is currently under investigation. While the total efficiency of this method depends on the tradeoff between the optimization cost and cost to directly generate samples, if results with a similar magnitude of improvement are achieved for QCD calculations, it would almost certainly result in a net improvement in the efficiency of many measurements. As statistical precision limits all aspects of such calculations, improving the precision of available results will likely unlock significant new physics insights.

Acknowledgments

WD, YL and PES are supported in part by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, under grant Contract Number DE-SC0011090 and by the National Science Foundation under Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, `http://iaifi.org/`). WD and PES are additionally supported by the U.S. Department of Energy SciDAC5 award DE-SC0023116. PES is additionally supported by the Early Career Award DE-SC0021006 and by the Simons Foundation grant 994314 (Simons Collaboration on Confinement and QCD Strings). GK is supported by the Schweizerischer Nationalfonds through grant agreement no. 200020_200424. MLW is supported by Fermi Research Alliance, LLC under under grant Contract Number DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

References

[1] G. Aarts. Lefschetz thimbles and stochastic quantization: Complex actions in the complex plane. *Phys. Rev. D*, 88(9):094501, 2013. doi: 10.1103/PhysRevD.88.094501.

[2] A. Alexandru, G. Basar, and P. Bedaque. Monte Carlo algorithm for simulating fermions on Lefschetz thimbles. *Phys. Rev. D*, 93(1):014504, 2016. doi: 10.1103/PhysRevD.93.014504.

[3] A. Alexandru, G. Basar, P. F. Bedaque, G. W. Ridgway, and N. C. Warrington. Sign problem and Monte Carlo calculations beyond Lefschetz thimbles. *JHEP*, 05:053, 2016. doi: 10.1007/JHEP05(2016)053.

[4] A. Alexandru, G. Basar, P. F. Bedaque, S. Vartak, and N. C. Warrington. Monte Carlo Study of Real Time Dynamics on the Lattice. *Phys. Rev. Lett.*, 117(8):081602, 2016. doi: 10.1103/PhysRevLett.117.081602.

[5] A. Alexandru, G. Basar, P. F. Bedaque, and G. W. Ridgway. Schwinger-Keldysh formalism on the lattice: A faster algorithm and its application to field theory. *Phys. Rev. D*, 95(11):114501, 2017. doi: 10.1103/PhysRevD.95.114501.

[6] A. Alexandru, P. F. Bedaque, H. Lamm, and S. Lawrence. Deep Learning Beyond Lefschetz Thimbles. *Phys. Rev. D*, 96(9):094505, 2017. doi: 10.1103/PhysRevD.96.094505.

[7] A. Alexandru, P. F. Bedaque, H. Lamm, and S. Lawrence. Finite-Density Monte Carlo Calculations on Sign-Optimized Manifolds. *Phys. Rev. D*, 97(9):094510, 2018. doi: 10.1103/PhysRevD.97.094510.

[8] A. Alexandru, G. Basar, P. F. Bedaque, and N. C. Warrington. Complex paths around the sign problem. *Rev. Mod. Phys.*, 94(1):015006, 2022. doi: 10.1103/RevModPhys.94.015006.

[9] Y. Aoki et al. FLAG Review 2021. *Eur. Phys. J. C*, 82(10):869, 2022. doi: 10.1140/epjc/s10052-022-10536-1.

[10] T. Aoyama et al. The anomalous magnetic moment of the muon in the Standard Model. *Phys. Rept.*, 887:1–166, 2020. doi: 10.1016/j.physrep.2020.07.006.

[11] J. B. Bronzan. Parametrization of SU(3). *Phys. Rev. D*, 38:1994, 1988. doi: 10.1103/PhysRevD.38.1994.

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

[13] M. Cristoforetti, F. Di Renzo, A. Mukherjee, and L. Scorzato. Monte Carlo simulations on the Lefschetz thimble: Taming the sign problem. *Phys. Rev. D*, 88(5):051501, 2013. doi: 10.1103/PhysRevD.88.051501.

[14] W. Detmold, G. Kanwar, M. L. Wagman, and N. C. Warrington. Path integral contour deformations for noisy observables. *Phys. Rev. D*, 102(1):014514, 2020. doi: 10.1103/PhysRevD.102.014514.

[15] W. Detmold, G. Kanwar, H. Lamm, M. L. Wagman, and N. C. Warrington. Path integral contour deformations for observables in $SU(N)$ gauge theory. *Phys. Rev. D*, 103(9):094517, 2021. doi: 10.1103/PhysRevD.103.094517.

[16] R. G. Edwards and B. Joo. The Chroma software system for lattice QCD. *Nucl. Phys. B Proc. Suppl.*, 140:832, 2005. doi: 10.1016/j.nuclphysbps.2004.11.254.

[17] C. Gattringer and C. B. Lang. *Quantum chromodynamics on the lattice*, volume 788. Springer, Berlin, 2010. ISBN 978-3-642-01849-7, 978-3-642-01850-3. doi: 10.1007/978-3-642-01850-3.

[18] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In Y. Bengio and Y. LeCun, editors, *ICLR 2015*, 2015. URL `http://arxiv.org/abs/1412.6980`.

[19] A. S. Kronfeld et al. Lattice QCD and Particle Physics. 7 2022.

[20] G. P. Lepage. The Analysis of Algorithms for Lattice Field Theory. In *Theoretical Advanced Study Institute in Elementary Particle Physics*, 6 1989.

[21] Y. Mori, K. Kashiwa, and A. Ohnishi. Application of a neural network to the sign problem via the path optimization method. *PTEP*, 2018(2):023B04, 2018. doi: 10.1093/ptep/ptx191.

[22] G. Parisi. The Strategy for Computing the Hadronic Mass Spectrum. *Phys. Rept.*, 103:203–211, 1984. doi: 10.1016/0370-1573(84)90081-4.

[23] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI 2015: 18th International Conference*, pages 234–241. Springer, 2015.