PROCEEDINGS OF SCIENCE

Exploration of Efficient Neural Network for Path Optimization Method

Yusuke Namekawa,[a,b,*] **Kouji Kashiwa,**[c] **Hidefumi Matsuda,**[d,e] **Akira Ohnishi**[e] **and Hayato Takase**

[a] *Education and Research Center for Artificial Intelligence and Data Innovation, Hiroshima University, Hiroshima 730-0053, Japan*

[b] *Department of Physics, Faculty of Science, Kyoto University, Kyoto 606-8502, Japan*

[c] *Department of Computer Science and Engineering, Faculty of Information Engineering, Fukuoka Institute of Technology, Fukuoka 811-0295, Japan*

[d] *Department of Physics and Center for Field Theory and Particle Physics, Fudan University, Shanghai 200433, China*

[e] *Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto 606-8502, Japan*

E-mail: namekawa@hiroshima-u.ac.jp

We present our attempts to control the sign problem by the path optimization method with an emphasis on the efficiency of the neural network. We found a gauge invariant neural network is successful in the 2-dimensional U(1) gauge theory with a complex coupling. We also investigate the possibility of improvement in the learning process.

The 39th International Symposium on Lattice Field Theory,
8th-13th August, 2022,
Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany

*Speaker

PoS(LATTICE2022)026

1. Motivation

The Monte Carlo method is an important tool for the investigation of the non-perturbative nature of quantum field theories with unknown analytical results. However, it is difficult to obtain a high accuracy result when the sign problem occurs. Several approaches have been proposed to evade the sign problem, as reviewed in Ref. [1].

The path optimization method (POM) [2, 3], also called the sign-optimized manifold [4, 5], utilizes machine learning to find the optimal path which maximally weakens the sign problem. The POM has been applied to several models [2, 4, 6–15]. The POM works well in a small system, such as finite-density $0+1$-dimensional QCD [10]. On the other hand, the neural network requires gauge-fixing or gauge-invariant inputs in a larger system to reduce the sign problem, as confirmed in two-dimensional $\mathrm{U}(1)$ gauge theory with complex coupling [13]. The average phase factor, an indicator of the sign problem, does not improve unless the gauge is fixed. In contrast, the gauge-invariant input, *i.e.*, the plaquette input successfully improves the average phase factor without gauge fixing. It is important to employ a neural network that respects the symmetry of the target theory. A similar idea has been adopted as a part of the lattice gauge equivariant convolutional neural network [16]. Although we have overcome the failure of the POM for a gauge theory with a large degree of freedom, further development is desirable to approach a realistic system with a large volume.

We perform a feasibility study for the POM, such as a gauge covariant neural network and approximation of the Jacobian computation in the learning process. A gauge covariant neural network respects the gauge symmetry [17], as the gauge invariant input does. In the gauge covariant neural network, a Stout-like smearing function is employed to construct the neural network, which is gauge covariant by definition. We also test the reduction of the numerical cost of the Jacobian, which is $O(N^3)$ for a system with the degrees of freedom N. The Jacobian is computed only in the final step of the configuration generation and the measurement, while ignored in the learning part. It significantly reduces the cost with less control of the neural network. We verify the approximation of the Jacobian in the learning part does not spoil the precision of the simulation result. Part of our results has been reported in Ref. [15].

2. Method

The path optimization method utilizes a neural network to suppress the sign problem. It deforms the contour of the path integral on the complexified variable plane, keeping the integral unchanged due to Cauchy's theorem. The complexification of the gauge field is $A_\mu(n) \to \mathcal{A}_\mu(n) = A_\mu(n) + iz_\mu(n)$, $A_\mu(n), z_\mu(n) \in \mathbb{R}$. The imaginary part of the complexified variable $z_\mu(n)$ is tuned via the neural network. The neural network consists of three layers: the input layer, the hidden layer, and the output layer. The variables on the input layer are denoted by t_i as a one-dimensional array. We employ a link variable of the gauge field $U_\mu(n) := \exp(ig_0 A_\mu(n))$ for t_i, which will be modified to the complexified links $\mathcal{U}_\mu(n) := \exp(ig_0\mathcal{A}_\mu(n))$ by the neural network. The variables on the hidden and output layers are denoted by y_j and z_k, which are determined using an activation function F such that

$$y_j = F\left(w^{(1)}_{ji} t_i + b^{(1)}_j\right),$$
$$z_k = \omega_k F\left(w^{(2)}_{kj} y_j + b^{(2)}_k\right), \tag{1}$$

where the indices i, j, k run from 1 to N_{input}, N_{hidden}, and N_{output}, respectively. N_{hidden} is taken to be proportional to N_{vol}. The activation function is the hyperbolic tangent in our case. w, b, and ω are neural network parameters, optimized using the loss function. Our loss function $\mathcal{F}$ quantifies enhancement of the average phase factor such that

$$\mathcal{F}[z] = \int d^{N_{\text{input}}} t \, |e^{i\theta(t)} - 1|^2 \, |J(t)\, e^{-S(t)}|, \tag{2}$$

where $\theta(t)$ is the phase originated from the Jacobian $J(t)$ of the input and complexified variables and the action S: $\exp(i\theta(t)) = J(t)\exp(-S(t)) \,/\, |J(t)\exp(-S(t))|$. The expectation value of the observable O is given by

$$\langle O \rangle = \frac{\langle O e^{i\theta} \rangle_{\text{pq}}}{\langle e^{i\theta} \rangle_{\text{pq}}}, \tag{3}$$
$$\langle O \rangle_{\text{pq}} = \frac{1}{Z} \int \mathcal{D}U \, \left[O \, |J\, e^{-S}| \right]_{\mathcal{U} \in C}, \tag{4}$$

which has the same mean value as that without the path optimization, and has a reduced statistical error if the sign problem is improved by the path optimization.

The path optimization always works in principle, because the neural network can approximate any function in arbitrary precision proved by the universal approximation theorem [18, 19]. In practice, however, it sometimes requires an enormous computer time. We confirmed it in the two-dimensional U(1) gauge theory with a complex coupling [13]. While the neural network with the link variable input never improved the sign problem, the neural network with the gauge invariant input, instead of the gauge-variant input of the link variables, can improve the sign problem significantly. For the gauge invariant object, we construct the plaquette on the input layer and pass it to the hidden layer. The hidden layer, then, respects the gauge symmetry by construction in (1).

As an alternative approach, we investigated the efficiency of the gauge-covariant neural network [17], which also respects the gauge symmetry. It utilizes the Stout-like smearing as the gauge-covariant function.

$$\mathcal{U}^{(l+1)}_\mu(n) = e^{iW^{(l)}_\mu(n)} \, \mathcal{U}^{(l)}_\mu(n), \tag{5}$$
$$W^{(l)}_\mu(n) = \sum_{\nu \neq \mu} \left(\rho^{(l)}_+ \mathcal{P}^{(l)}_{\mu\nu}(n) + \rho^{(l)}_- \mathcal{P}^{(l)}_{\mu\nu}{}^{-1}(n) \right), \tag{6}$$
$$\mathcal{P}_{\mu\nu}(n) = \mathcal{U}_\mu(n)\, \mathcal{U}_\nu(n+\hat{\mu})\, \mathcal{U}^{-1}_\mu(n+\hat{\nu})\, \mathcal{U}^{-1}_\nu(n), \tag{7}$$

where $l = 1, \ldots, N_{\text{smear}}$. $\rho^{(l)}_\pm \in \mathbb{C}$ is a neural network parameter. Notice that this gauge covariant neural network has a residual connection. It may have better control of the vanishing gradient problem.

The bottleneck of these approaches in the POM is the Jacobian calculation, which costs $O(N^3)$ for the degrees of freedom N. We tested the efficiency of an approximation $J = 1$ in the learning

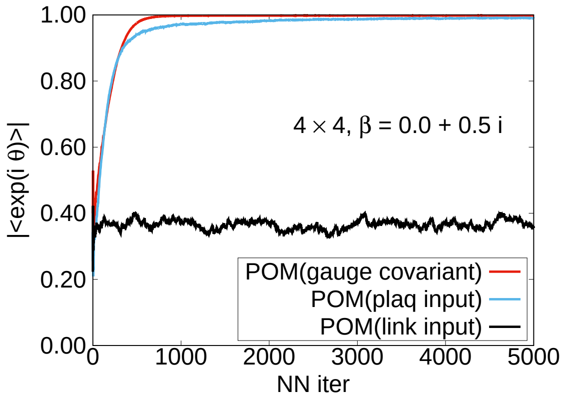

Figure 1: The average phase factor by the gauge-covariant neural network at $\beta = 0.5i$ on 4×4 lattice as a function of the neural network iteration [15]. Those by the gauge-invariant plaquette input and the gauge-variant link input to the POM are also plotted [13].

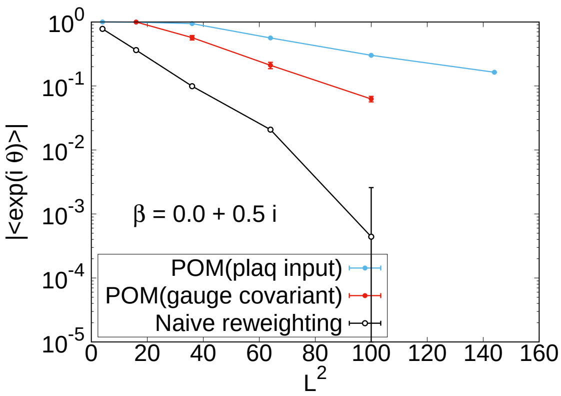

Figure 2: Comparison of the volume scaling of the average phase factor by the gauge-covariant neural network [15] with those of the gauge invariant plaquette input to the POM and the naive reweighting method [13] at $\beta = 0.5i$ on $4 \times 4 - 12 \times 12$ lattices.

process of the neural network, keeping the exact Jacobian calculation in the configuration generation and measurements. It provides a drastic reduction of the numerical cost for the neural network, $O(N^3)$ to $O(1)$. The disadvantage of this approximation is less control of the sign problem. The enhancement of the average phase factor is lowered, and the statistical error of the observable increases. It is necessary to confirm if the $J = 1$ approximation in the learning process does not significantly affect the observable error.

3. Results

We test the improvements for the POM, i.e., the gauge covariant neural network and the approximation of the Jacobian in the learning process by the two-dimensional U(1) gauge theory with a complex gauge coupling $\beta = 1/g_0^2 \in \mathbb{C}$,

$$S = -\frac{\beta}{2} \sum_n \left(P_{12}(n) + P_{12}(n)^{-1} \right). \tag{8}$$

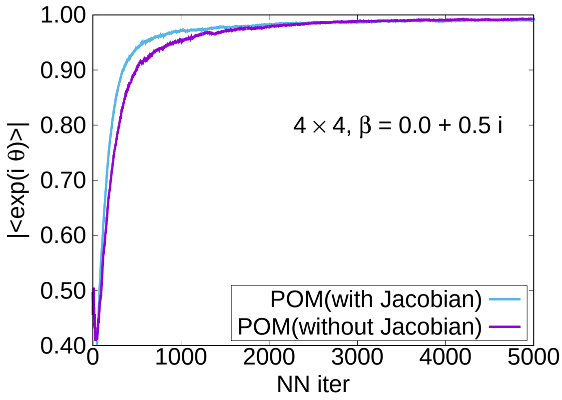

Figure 3: The average phase factor by the POM using the gauge invariant input with and without Jacobian in the learning process at $\beta = 0.5i$ on 4×4 lattice [15].

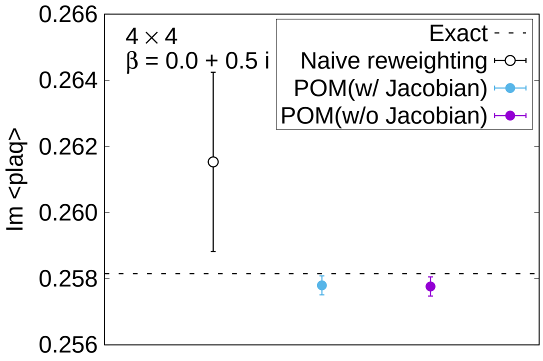

Figure 4: Imaginary part of plaquette expectation values by the POM with and without Jacobian in the neural network as well as by the naive reweighting method at $\beta = 0.5i$ on 4×4 lattice [15].

The sign problem is caused by the imaginary part of β. We generated 50000 gauge configurations by the Hybrid Monte Carlo method. The Adam optimizer [20] with the learning rate of 10^{-3} and the decay rate of 0.99 is used for the neural network. The batch normalization [21] with the batch size of 1000 is also employed to stabilize the learning process. We set the number of the hidden layer being proportional to the lattice volume. The statistical error is estimated by the Jackknife method, binning 250 trajectory data.

Figure 1 represents the average phase factor of the path optimization with the gauge-covariant neural network [15]. For comparison, we also plot the results of the POM using the gauge-invariant plaquette input and the gauge-variant link input [13]. As the neural network iteration proceeds, the average phase factor by the gauge covariant neural network quickly increases, indicating the sign problem is reduced well. A similar result is obtained by the POM using the gauge invariant input, though that using the gauge variant input does not enhance the average phase factor. Our results suggest the gauge symmetry plays a crucial role in the neural network.

Figure 2 shows the volume dependence of the average phase factors. The naive reweighting method gives a steep decrease in the average phase factor toward the thermodynamic limit. It becomes milder by use of the gauge covariant neural network and the POM with the gauge invariant input. Notice that the number of neural network parameters of the gauge-covariant neural network is smaller than that of the POM with the gauge invariant input, which seems to cause a lower enhancement of the average phase factor by the gauge-covariant neural network.

Figure 3 is the average phase factor obtained by the POM using the gauge invariant input with and without $J = 1$ approximation in the learning process. The POM with the exact Jacobian calculation leads to faster enhancement of the average phase factor than that with $J = 1$ approximation. However, the difference becomes negligible after the 2000 neural network iterations. The tininess of the difference is also confirmed for an observable. Figure 4 presents the imaginary part of the plaquette expectation value. Although all results agree with the exact value [22–24], the POM with the gauge invariant input gives a well suppressed statistical error, compared with the naive reweighting method. The use of $J = 1$ approximation does not significantly increase the statistical error. The difference is 1% in this case. Our result quantitatively shows the $J = 1$ approximation in the learning process is efficient and does not increase the statistical error significantly in our setup.

4. Conclusion

We explored the gauge-covariant neural network [17] and approximation of the Jacobian in the learning process for the path optimization of the gauge theory. For the gauge-covariant neural network, we find a large enhancement of the average phase factor, though the enhancement does not reach that of the gauge invariant input [13]. It is still important, however, to perform a similar test in the non-abelian theory, for which the gauge covariant neural network is applicable easily. The approximation of the Jacobian in the learning process leads to a significant reduction of the numerical cost without spoiling the data quality of the result. The statistical error becomes larger due to less control of the sign problem, but the increase is 1% in our case. We may have better control by a combination of the learning process with and without the exact Jacobian.

We apply these improvements to the POM for non-abelian theories, such as SU(2) and SU(3). Our improvements are helpful for control of the sign problem with acceptable costs. It is also interesting to test our improvements for the path optimization of the other theoretical models.

Acknowledgments

Our code is based in part on LTKf90 [25]. This work is supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers JP19H01898, JP21H00121, JP21K03553, JP22H05112 and by National Natural Science Foundation of China (NSFC) Grant Number 12075061.

References

[1] K. Nagata, *Finite-density lattice QCD and sign problem: current status and open problems*, *Soryusironkenkyu* **31** (2020), no. 1 1–156, [arXiv:2108.12423].

[2] Y. Mori, K. Kashiwa, and A. Ohnishi, *Toward solving the sign problem with path optimization method*, *Phys. Rev. D* **96** (2017), no. 11 111501, [arXiv:1705.05605].

PoS(LATTICE2022)026

[3] Y. Mori, K. Kashiwa, and A. Ohnishi, *Application of a neural network to the sign problem via the path optimization method*, *PTEP* **2018** (2018), no. 2 023B04, [arXiv:1709.03208].

[4] A. Alexandru, P. F. Bedaque, H. Lamm, and S. Lawrence, *Finite-Density Monte Carlo Calculations on Sign-Optimized Manifolds*, *Phys. Rev. D* **97** (2018), no. 9 094510, [arXiv:1804.00697].

[5] S. Lawrence, H. Oh, and Y. Yamauchi, *Lattice Scalar Field Theory At Complex Coupling*, arXiv:2205.12303.

[6] K. Kashiwa, Y. Mori, and A. Ohnishi, *Controlling the model sign problem via the path optimization method: Monte Carlo approach to a QCD effective model with Polyakov loop*, *Phys. Rev. D* **99** (2019), no. 1 014033, [arXiv:1805.08940].

[7] A. Alexandru, P. F. Bedaque, H. Lamm, S. Lawrence, and N. C. Warrington, *Fermions at Finite Density in 2+1 Dimensions with Sign-Optimized Manifolds*, *Phys. Rev. Lett.* **121** (2018), no. 19 191602, [arXiv:1808.09799].

[8] F. Bursa and M. Kroyter, *A simple approach towards the sign problem using path optimisation*, *JHEP* **12** (2018) 054, [arXiv:1805.04941].

[9] K. Kashiwa, Y. Mori, and A. Ohnishi, *Application of the path optimization method to the sign problem in an effective model of QCD with a repulsive vector-type interaction*, *Phys. Rev. D* **99** (2019), no. 11 114005, [arXiv:1903.03679].

[10] Y. Mori, K. Kashiwa, and A. Ohnishi, *Path optimization in* $0+1D$ *QCD at finite density*, *PTEP* **2019** (2019), no. 11 113B01, [arXiv:1904.11140].

[11] K. Kashiwa and Y. Mori, *Path optimization for* $U(1)$ *gauge theory with complexified parameters*, *Phys. Rev. D* **102** (2020), no. 5 054519, [arXiv:2007.04167].

[12] 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** (2021), no. 9 094517, [arXiv:2101.12668].

[13] Y. Namekawa, K. Kashiwa, A. Ohnishi, and H. Takase, *Gauge invariant input to neural network for path optimization method*, *Phys. Rev. D* **105** (2022), no. 3 034502, [arXiv:2109.11710].

[14] M. Giordano, K. Kapas, S. D. Katz, A. Pasztor, and Z. Tulipant, *Exponential reduction of the sign problem at finite density in the 2+1D XY model via contour deformations*, *Phys. Rev. D* **106** (2022), no. 5 054512, [arXiv:2202.07561].

[15] Y. Namekawa, K. Kashiwa, H. Matsuda, A. Ohnishi, and H. Takase, *Improving efficiency of the path optimization method for a gauge theory*, arXiv:2210.05402.

[16] M. Favoni, A. Ipp, D. I. Müller, and D. Schuh, *Lattice Gauge Equivariant Convolutional Neural Networks*, *Phys. Rev. Lett.* **128** (2022), no. 3 032003, [arXiv:2012.12901].

[17] A. Tomiya and Y. Nagai, *Gauge covariant neural network for 4 dimensional non-abelian gauge theory*, arXiv:2103.11965.

[18] G. Cybenko, *Approximation by superpositions of a sigmoidal function*, *Mathematics of Control, Signals, and Systems (MCSS)* **2** (1989), no. 4 303–314.

[19] K. Hornik, *Approximation capabilities of multilayer feedforward networks*, *Neural Networks* **4** (1991), no. 2 251–257.

[20] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 12, 2014. arXiv:1412.6980.

[21] S. Ioffe and C. Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift*, *CoRR* **abs/1502.03167** (2015) [arXiv:1502.03167].

[22] U. J. Wiese, *Numerical Simulation of Lattice θ Vacua: The 2-d U(1) Gauge Theory as a Test Case*, *Nucl. Phys. B* **318** (1989) 153–175.

[23] B. E. Rusakov, *Loop averages and partition functions in U(N) gauge theory on two-dimensional manifolds*, *Mod. Phys. Lett. A* **5** (1990) 693–703.

[24] C. Bonati and P. Rossi, *Topological susceptibility of two-dimensional $U(N)$ gauge theories*, *Phys. Rev. D* **99** (2019), no. 5 054503, [arXiv:1901.09830].

[25] S. Choe, S. Muroya, A. Nakamura, C. Nonaka, T. Saito, and F. Shoji, *Lattice tool kit in Fortran90*, *Nucl. Phys. B Proc. Suppl.* **106** (2002) 1037–1039.

PoS(LATTICE2022)026