

## Improving efficiency of the path optimization method for a gauge theory

Yusuke Namekawa<sup>1,2,3,\*</sup>, Kouji Kashiwa,<sup>4</sup> Hidefumi Matsuda,<sup>2,5</sup> Akira Ohnishi<sup>3</sup>, and Hayato Takase

<sup>1</sup>*Education and Research Center for Artificial Intelligence and Data Innovation, Hiroshima University, Hiroshima 730-0053, Japan*

<sup>2</sup>*Department of Physics, Faculty of Science, Kyoto University, Kyoto 606-8502, Japan*

<sup>3</sup>*Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto 606-8502, Japan*

<sup>4</sup>*Department of Computer Science and Engineering, Faculty of Information Engineering, Fukuoka Institute of Technology, Fukuoka 811-0295, Japan*

<sup>5</sup>*Department of Physics and Center for Field Theory and Particle Physics, Fudan University, Shanghai 200433, China*

 (Received 15 October 2022; accepted 1 February 2023; published 22 February 2023)

We investigate efficiency of a gauge-covariant neural network and an approximation of the Jacobian in optimizing the complexified integration path toward evading the sign problem in lattice field theories. For the construction of the complexified integration path, we employ the path optimization method. The two-dimensional U(1) gauge theory with the complex gauge-coupling constant is used as a laboratory to evaluate the efficiency. It is found that the gauge-covariant neural network, which is composed of the Stout-like smearing, can enhance the average phase factor, as the gauge-invariant input does. For the approximation of the Jacobian, we test the most drastic case in which we perfectly drop the Jacobian during the learning process. It reduces the numerical cost of the Jacobian calculation from  $\mathcal{O}(N^3)$  to  $\mathcal{O}(1)$ , where  $N$  means the number of degrees of freedom of the theory. The path optimization using this Jacobian approximation still enhances the average phase factor at expense of a slight increase of the statistical error.

DOI: [10.1103/PhysRevD.107.034509](https://doi.org/10.1103/PhysRevD.107.034509)

### I. INTRODUCTION

The Monte Carlo (MC) method is an important tool for investigating nonperturbative properties of quantum field theories in which we do not know the analytic results. The MC method with the probability weight is, however, difficult to perform in high precision when the sign problem arises. The probability distribution function that controls the update of the configuration becomes complex and/or seriously oscillating functions. Actually, quantum chromodynamics (QCD) at finite density is a well-known theory that has a serious sign problem (for example, see Ref. [1]). Recently, several new methods have been developed to control the sign problem at high densities. One of the famous examples is the complex Langevin method [2,3], which is based on a stochastic quantization with complexified dynamical variables and is free from the sign problem. Thanks to the low numerical cost of the complex Langevin method, it has already been applied to four-dimensional QCD at finite density [4–11]. However, the validity region of the complex

Langevin method is limited. The complex Langevin method is proved to provide correct results if the boundary term disappears [12], or equivalently, the distribution of the driving force decays exponentially or faster [13]. Another famous method is the tensor renormalization group method [14], which is based on a coarse graining non-MC algorithm using a tensor network. Although the computational cost is extremely high, it has been vigorously tested even in four-dimensional theoretical models [15–18] as well as non-Abelian theories [19–23]. The improved algorithms are proposed to reduce the enormous computational cost [24,25]. The Lefschetz thimble method [26] is a MC scheme that complexifies dynamical variables and determines the integration path by solving an antiholomorphic flow equation from fixed points such that the imaginary part of the action is constant. Cauchy's integral theorem ensures that the integral is independent of a choice of the integration path if the path is given as a result of continuous deformation from the original path [27], crosses no poles of the integrand, and the integral at infinity has no contribution. A numerical study has been started with Langevin algorithm [28], Metropolis algorithm [29], and hybrid MC algorithm [30]. The additional ergodicity problem and the high-computational cost are the main bottlenecks of this method, but the algorithm development is overcoming them [31,32]. The path optimization method (POM) [33,34],

\*namekawa@hiroshima-u.ac.jp

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.*

also known as the sign-optimized manifold [35,36], is an alternative approach that modifies the integration path using machine learning through neural networks. Machine learning finds the best path on which the sign problem is maximally weakened. The POM successfully works for the complex  $\lambda\phi^4$  theory [33], the Polyakov-loop extended Nambu–Jona-Lasinio model [37,38], the Thirring model [35,39], the  $(0+1)$ -dimensional Bose gas [40], the  $(0+1)$ -dimensional QCD [41], the two-dimensional  $U(1)$  gauge theory with complexified coupling constant [42], the  $(2+1)$ -dimensional XY model [43], as well as noise reduction in observables [44]. The recent progress of the complexified path approaches is reviewed in Ref. [45].

A crucial issue of the POM in gauge theories is control of redundancy from the gauge symmetry. In  $(0+1)$ -dimensional QCD at finite density [41], the POM works with and without the gauge fixing. In higher dimensions, however, the gauge-fixing or gauge-invariant input is required for neural networks to find an improved integral path. The two-dimensional  $U(1)$  gauge theory with a complex coupling [42,46,47] is a good test ground to investigate the effect of gauge degrees of freedom on the sign problem. The sign problem originates from the imaginary part of the complex coupling. Since the analytic result is available [48–50], we can use it to verify the simulation results. It has been found that the average phase factor, an indicator of the sign problem, is not improved without the gauge fixing. This may be related to the insufficient performance of the considering neural network in Ref. [42] compared to the complexity of the gauge symmetry. In contrast, the gauge-invariant input successfully enhances the average phase factor without the gauge fixing [47]. The link variables are no longer a direct input to the neural network. Gauge-invariant quantities, plaquettes in this case, are chosen as the input. This treatment mitigates the difficulties induced by the gauge symmetry, even in a simple neural network. A similar idea is employed as a part of lattice-gauge equivariant convolutional neural networks [51].

To tackle the sign problem in more realistic systems, it is hoped to develop efficient optimization methods further. In this study, we try to implement the feasible ways to the path optimization method using two methods:

- (i) *Improvement 1*: gauge-covariant neural network. and
- (ii) *Improvement 2*: Approximation of the Jacobian calculation in the learning process.

The gauge-covariant neural network is an alternative neural network which respects the gauge symmetry [52]. In the method, the usual neural network is replaced by the Stout-like smearing functions. Since the smearing process does not break the gauge covariance by definition, we can exactly deal with the gauge symmetry of theories. It is important to evaluate the performance of the gauge-covariant neural network toward large scale simulations. It is also necessary to reduce the numerical cost to calculate

the Jacobian. The cost is  $\mathcal{O}(N^3)$  with  $N$  being the degrees of freedom. We test the most drastic approximation. We completely neglect the Jacobian in the learning part. The Jacobian is calculated only at the configuration generation and the final step of the POM for measurement.

This paper is organized as follows. In Sec. II, the current status of the path optimization method is summarized. Improvements in the path optimization method using the gauge-covariant neural network and the approximation of Jacobian are discussed in Sec. III. Section IV shows numerical results and Sec. V summarizes this paper. In the Appendix, we discuss improvements in the optimization using the link-variable input.

## II. PATH OPTIMIZATION METHOD

We compile the current status of the path optimization method. In the path optimization method, dynamical variables are complexified to control the sign problem. The path integral contour is deformed on the complexified dynamical variable plane. Because of the Cauchy’s integral theorem, such deformation provides the same result obtained on the original integral contour, as long as the modified path does not cross any poles of the integrand, the Boltzmann weight, and its infinities do not contribute to the result. Note that the singular points of the action are generally irrelevant. They appear from the zeros of the fermion determinant and result in the zeros, not poles, of the integrand.

The procedure of the path optimization method can be summarized as follows:

- (i) *Step 1*: Create configurations on the original integral path using the MC method.
- (ii) *Step 2*: Using the back-propagation method to optimize the network parameters to construct the modified integral path.
- (iii) *Step 3*: Create configurations on the modified integral path using the reweighting method.
- (iv) *Step 4*: Evaluate the average phase factor.
- (v) *Step 5*: Return to Step 2 if the growth of the average phase factor is not sufficient.

In addition, we occasionally need the replica exchange MC method [53–56], as needed in the Lefschetz thimble method [57–59], when the integral path becomes complicated (see Ref. [42]). Details of the path optimization method are explained below.

### A. Gauge-variant input

The simplest way is to directly complexify the dynamical variables  $A_\mu(n)$  in the path optimization method as

$$A_\mu(n) \rightarrow \mathcal{A}_\mu(n) = A_\mu(n) + iz_\mu(n), \quad (1)$$

where  $\mathcal{A}_\mu(n) \in \mathbb{C}$  and  $A_\mu(n), z_\mu(n) \in \mathbb{R}$  at the lattice site  $n$  in the direction of  $\mu$ . Complexified variables,  $\mathcal{A}_\mu(n)$ , represent the modification of the integral path. To construct

the best  $z_\mu(n)$  for reduction of the sign problem, we use the neural network, which obeys the universal approximation theorem [60,61]. The actual procedure then becomes

$$\underbrace{A_\mu(n)}_{\text{input layer}} \rightarrow \text{hidden layer} \rightarrow \underbrace{\mathcal{A}_\mu(n)}_{\text{output layer}}. \quad (2)$$

In the lattice simulation, it is convenient to transform the above dynamical quantities into the link variables at the input layer because the lattice action is composed of the link variables,

$$\underbrace{A_\mu(n)}_{\text{input layer}} \rightarrow \underbrace{U_\mu(n)}_{\text{input layer}} \rightarrow \text{hidden layer} \rightarrow \underbrace{\mathcal{U}_\mu(n)}_{\text{output layer}}, \quad (3)$$

where  $U_\mu(n) := \exp(ig_0 A_\mu(n))$  and  $\mathcal{U}_\mu(n) := \exp(ig_0 \mathcal{A}_\mu(n))$  are the original and modified (complexified) link variables, respectively, with the bare coupling constant  $g_0$ . Such compact variables make the path integral finite.

In the neural network with the input represented as  $t_i$ , corresponding to the real and imaginary parts of  $U_\mu(n)$ , the variables on the hidden layer ( $y_j$ ) and the output ( $z_k$ ) are given as

$$\begin{aligned} y_j &= F(w_{ji}^{(1)} t_i + b_j^{(1)}), \\ z_k &= \omega_k F(w_{kj}^{(2)} y_j + b_k^{(2)}), \end{aligned} \quad (4)$$

where  $i = 1, \dots, N_{\text{input}}$ ,  $j = 1, \dots, N_{\text{hidden}}$ , and  $k = 1, \dots, N_{\text{output}}$  with the numbers of units in the input, hidden, and output layers,  $N_{\text{input}}$ ,  $N_{\text{hidden}}$ , and  $N_{\text{output}}$ . In the case of the gauge-variant input,  $N_{\text{input}} = N_{\text{output}} = 4N_{\text{vol}}$ , where  $N_{\text{vol}} = N_1 N_2$  with  $N_\mu$  being the lattice size in the  $\mu$ -direction.  $N_{\text{hidden}}$  is taken to be proportional to  $N_{\text{vol}}$ .  $w$ ,  $b$ , and  $\omega$  are parameters of the neural network and  $F$  is the so-called activation function. We employ a hyperbolic tangent function for the activation function.

### B. Gauge-invariant input

For the learning process in the path optimization method, the gauge invariance plays a central role and thus should be imposed on the neural network. One possible way is using the gauge-invariant input. The link variables are transformed into the plaquette on the input layer, which is the gauge-invariant quantity and the fundamental building block of the lattice Lagrangian density,

$$\underbrace{A_\mu(n)}_{\text{input layer}} \rightarrow \underbrace{U_\mu(n)}_{\text{input layer}} \rightarrow \underbrace{P_{\mu\nu}(n)}_{\text{input layer}} \rightarrow \text{hidden layer} \rightarrow \underbrace{\mathcal{U}_\mu(n)}_{\text{output layer}}, \quad (5)$$

where the plaquette  $P_{\mu\nu}$  is defined as

$$P_{\mu\nu}(n) := U_\mu(n) U_\nu(n + \hat{\mu}) U_\mu^{-1}(n + \hat{\nu}) U_\nu^{-1}(n), \quad (6)$$

where  $\hat{\mu}$  ( $\hat{\nu}$ ) is a unit vector in the  $\mu$  ( $\nu$ ) direction. In this case,  $N_{\text{input}} = 2N_{\text{vol}}$ .

It is applicable to any lattice gauge action including the plaquette. The path-optimized lattice action is represented by the modified plaquettes. With this procedure, the hidden layer has the gauge invariance by definition. This procedure corresponds to the data preprocessing known in the machine learning community. Its effectiveness is confirmed in the two-dimensional U(1) gauge theory [47].

### C. Cost function and observables

The cost function ( $\mathcal{F}$ ) is needed to optimize several parameters of the neural network in Eq. (4). The convenient form of the cost function for the POM is

$$\mathcal{F}[z] = \int d^{N_{\text{input}}} t |e^{i\theta(t)} - 1|^2 \times |J(t)e^{-S(t)}|, \quad (7)$$

where  $J(t)$  means the Jacobian of the input and complexified variables and  $S$  represents the action.  $\theta(t)$  is the total phase defined by  $e^{i\theta(t)} = J(t)e^{-S(t)} / |J(t)e^{-S(t)}|$ . Minimization of the cost function (7) provides a modified integral path, which usually makes the Boltzmann weight,  $e^{-S}$ , being complex. The phase reweighting is required as

$$\langle \mathcal{O} \rangle = \frac{\langle \mathcal{O} e^{i\theta} \rangle_{\text{pq}}}{\langle e^{i\theta} \rangle_{\text{pq}}}, \quad \langle \mathcal{O} \rangle_{\text{pq}} := \frac{1}{Z} \int \mathcal{D}\mathcal{U} [\mathcal{O} |J e^{-S}|]_{\mathcal{U} \in \mathcal{C}} \quad (8)$$

where  $\mathcal{O}$  represents any operator such as the plaquette.  $\langle \dots \rangle_{\text{pq}}$  means the phase-quenched expectation value and  $Z$  is the partition function. The denominator,  $\langle e^{i\theta} \rangle_{\text{pq}}$ , is so-called the average phase factor which dominates the statistical error of observables with the phase reweighting. The absolute value  $|J e^{-S}|$  in Eq. (8) is used as the Boltzmann weight. Since  $|J e^{-S}|$  is definitely real, we can perform the MC simulation exactly. Another way is to use  $\exp(-\text{Re}S)$  as the Boltzmann weight. In this case, the upper bound of the average phase factor  $\langle e^{-(S - \ln |J e^{-S}|)} \rangle_{\text{R}}$  is not necessarily 1 where  $\langle \dots \rangle_{\text{R}}$  means the expectation values with the Boltzmann weight,  $e^{-\text{Re}S}$ .

## III. IMPROVEMENTS

We explain two improvements to the path optimization; the gauge-covariant neural network to represent the modified integral path, and the approximation of the Jacobian in the learning process.

### A. Gauge-covariant neural network

One of the possible ways to suitably treat the gauge symmetry in the neural network is to use the gauge-covariant neural network [52]. The neural network is

represented by the gauge-covariant function using some smearing process. We employ the following Stout-like smearing,

$$\underbrace{A_\mu(n)}_{\text{input layer}} \rightarrow \underbrace{U_\mu(n)}_{\text{hidden layer}} \rightarrow \underbrace{\tilde{U}_\mu^{(1)}(n)}_{\text{hidden layer}} \rightarrow \cdots \rightarrow \underbrace{\mathcal{U}_\mu(n)}_{\text{output layer}}, \quad (9)$$

where  $U_\mu(n)$  denotes the original link variable,  $\tilde{U}_\mu(n)$  means the smeared (complexified) link variable defined as

$$\tilde{U}_\mu^{(l+1)}(n) = \exp\left[iW_\mu^{(l)}(n)\right]\tilde{U}_\mu^{(l)}(n), \quad (10)$$

with

$$W_\mu^{(l)} = \sum_{\nu \neq \mu} \left( \rho_+^{(l)} \mathcal{P}_{\mu\nu}^{(l)} + \rho_-^{(l)} \mathcal{P}_{\mu\nu}^{(l)-1} \right), \quad (11)$$

here  $l$  means the number of smearing steps in the hidden layer, and  $\rho_\pm^{(l)} \in \mathbb{C}$  are parameters optimized via the back-propagation method. We use the Stout-like smearing twice in the hidden layer to complexify the dynamical variables. The complexified plaquette  $\mathcal{P}^{(l)}$  in Eq. (11) is calculated from the link variable of that in the step  $l$ . The smearing process combines the link variables again and again as  $l$  increases and forms a gauge-covariant neural network. This network does not correspond to the fully connected layer, but to the residual connection. The link variables in the previous step is reflected in the total factor in Eq. (10), and thus may be robust against the vanishing gradient problem.

Note that the gauge-covariant neural network has a smaller number of parameters than those of the neural network with gauge-variant and gauge-invariant inputs. The number of the gauge-covariant neural network parameters is  $2N_{\text{vol}}$ , while that of the neural network with gauge-invariant input is  $(N_{\text{input}}N_{\text{hidden}} + N_{\text{hidden}}) + (2N_{\text{hidden}}N_{\text{vol}} + 2N_{\text{vol}}) + 2N_{\text{vol}}$ , where  $N_{\text{input}}, N_{\text{hidden}}$  are proportional to  $N_{\text{vol}}$ .

## B. Approximation of Jacobian

The calculation cost of the Jacobian in the modification of the path integral contour is rather high,  $\mathcal{O}(N^3)$  for the system with the degrees of freedom  $N$ . Reduction of the cost of the Jacobian is necessary for a large-scale simulation using the POM.

There are several studies for reducing the Jacobian calculation cost in the MC update; the Gaussian or real approximations [62] that reduce  $\mathcal{O}(N^3)$  to  $\mathcal{O}(N)$ , the implementation of the Grady algorithm in the holomorphic gradient flow [63] that reduces  $\mathcal{O}(N^3)$  to  $\mathcal{O}(N^2)$ , the diagonal ansatz of the Jacobian matrix [35] that reduces  $\mathcal{O}(N^3)$  to  $\mathcal{O}(N)$ , the nearest neighbor lattice site ansatz [40] that reduces  $\mathcal{O}(N^3)$  to  $\mathcal{O}(N)$ , the introduction of the world-volume approach [31] which requires only a combination

of the Jacobian and the vector, i.e.,  $(Jv)$  evaluated using the antiholomorphic flow without the explicit form of the Jacobian in the MC update and thus the cost is reduced from  $\mathcal{O}(N^3)$  to  $\mathcal{O}(N^2)$ , and the application of complex-valued affine coupling layers [64] that reduces  $\mathcal{O}(N^3)$  to  $\mathcal{O}(N)$ , in principle.

As a first attempt, we use the simplest treatment for the Jacobian calculation in the path optimization method. We completely neglect the Jacobian contribution in the learning part. The numerical cost for the Jacobian calculation is reduced from  $\mathcal{O}(N^3)$  to  $\mathcal{O}(1)$ . It should be noted that we need the Jacobian calculation in the evaluation part of observables. We will see that this drastic approximation in the path optimization method can work well compared to the full calculation.

## IV. NUMERICAL RESULT

We employ two-dimensional U(1) gauge theory with a complex gauge-coupling constant as a laboratory to test the improvements for the path optimization method. The action is

$$S = -\frac{\beta}{2} \sum_n (P_{12}(n) + P_{12}(n)^{-1}), \quad (12)$$

where  $\beta = 1/g_0^2$ . A complex value of  $\beta$  causes a sign problem. The analytic solution can be found in Refs. [48–50]. The theoretical setup is the same as that in Ref. [47]. We only change some parts of the POM for the improvements mentioned above. 50,000 gauge configurations are generated by the hybrid MC with its error estimation by the jackknife procedure. The Adam optimizer [65] is used for the neural network.

Figure 1 shows a comparison of the average phase factor by several inputs to the POM as functions of the neural-network iteration. The POM using the gauge-covariant neural-network and gauge-invariant inputs successfully enhances the average phase factor up to 0.99, while that using the link input does not. It clearly shows the importance of the gauge symmetry in a neural network. The neural network respecting the gauge symmetry is efficient to control the sign problem by the POM.

The volume dependence of the average phase factor is plotted in Fig. 2. Both the gauge-covariant neural network and the gauge-invariant input lead to enhancement of the average phase factor and milder volume dependence. Lower enhancement of the average phase factor by the gauge-covariant neural network may be related to the fact that the gauge-covariant neural network has fewer network parameters than the neural network with the gauge-invariant input.

Figure 3 presents a comparison of the average phase factor by the POM using the gauge-invariant input with and without the Jacobian calculation in the neural network.



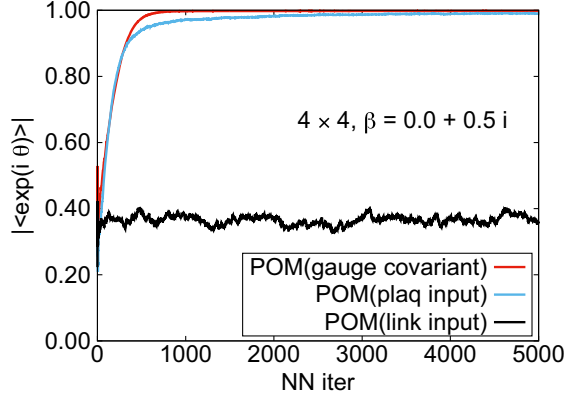


FIG. 1. Comparison of the average phase factor by the gauge-covariant neural network with those by the gauge-invariant and gauge-variant link inputs to the POM [47] as functions of the neural-network iteration at  $\beta = 0.5i$  on  $4 \times 4$  lattice.

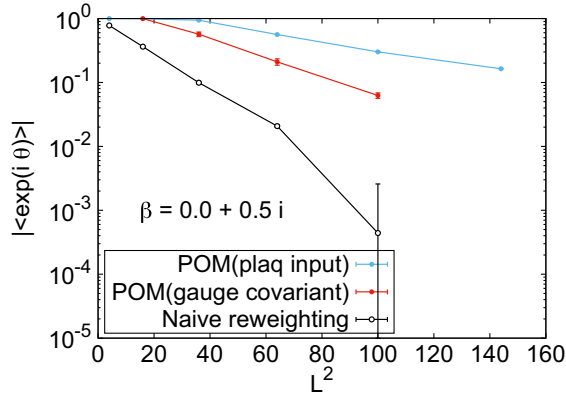


FIG. 2. Volume dependence of the average phase factor by the gauge-covariant neural network as well as the gauge-invariant input to the POM and the naive reweighting method [47] at  $\beta = 0.5i$  on  $4 \times 4$ — $12 \times 12$  lattices.

The POM without the Jacobian gives less enhancement of the average phase factor. However, the decrease is tiny in our setup. We also compare the expectation values of the imaginary part of the plaquette in Fig. 4. All results are consistent with the analytic value. The naive reweighting gives a huge error, whereas the POM with and without the Jacobian gives a much smaller error, indicating that the sign problem is under control. The difference between errors of the POM with and without the Jacobian is 1%. It suggests that the  $J = 1$  approximation in the neural network works well.

## V. SUMMARY

We explored the improvement in the path optimization method for the gauge theory by the following two methods:

- (1) Gauge-covariant neural network.
- (2) Approximation of the Jacobian in the learning process.

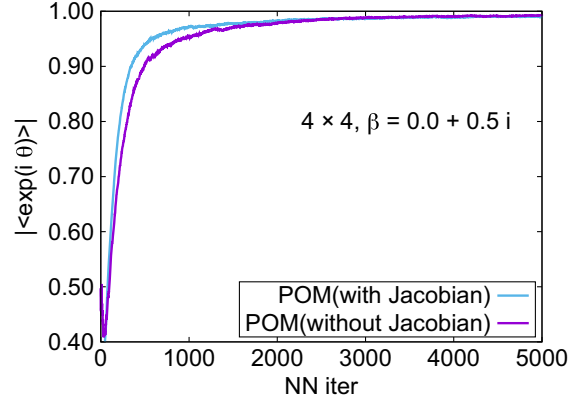


FIG. 3. Comparison of the POM using the gauge-invariant input with and without Jacobian in the neural network at  $\beta = 0.5i$  on  $4 \times 4$  lattice.

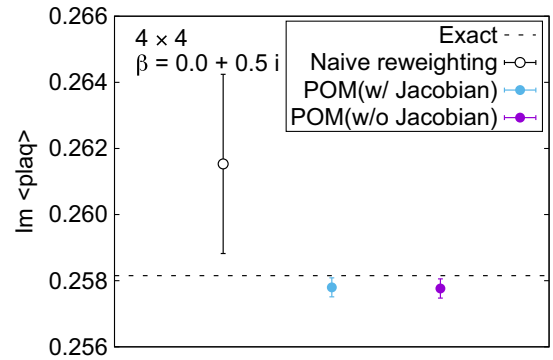


FIG. 4. Comparison of the imaginary part of plaquette expectation values by naive reweighting and POM with and without Jacobian in the neural network at  $\beta = 0.5i$  on  $4 \times 4$  lattice.

Using Method 1, the neural network is gauge covariant by definition. Due to this property of the network, we do not need the gauge fixing to enhance the average phase factor, unlike Ref. [42]. The enhancement is not as impressive as that of the gauge-invariant input [47] probably due to the fewer number of network parameters. However, the gauge-covariant neural network has an advantage that it can be applied to the non-Abelian theory in a straightforward manner. Similar performance evaluation in the non-Abelian theory is desirable, especially in QCD at finite density.

Using Method 2, we can significantly reduce the numerical cost. Neglecting the Jacobian contributions in the learning part reduces the cost from  $\mathcal{O}(N^3)$  to  $\mathcal{O}(1)$ . Even with this drastic approximation, the path optimization can proceed well, and the average phase factor is sufficiently increased. The drawback is a larger statistical error, but the increase is small, 1% in our calculation. We can extend Method 2 by first repeating the learning process without Jacobian contributions and then performing the learning process with the exact Jacobian calculation. This way has the similarity with the pretraining known in the

machine learning community and may accelerate the learning of networks.

By using the above two improvements, we can control the gauge symmetry and reduce the numerical cost significantly in the path optimization method at least for the two-dimensional U(1) gauge theory with complex coupling constant. These improvements may lead us to an exploration of the non-Abelian theory, such as SU(2) and SU(3), which has more serious sign problem and requires a higher numerical cost. We hope to report this attempt in the future.

### ACKNOWLEDGMENTS

Our code is based in part on LTKf90 [66]. This work is supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grants No. JP19H01898, No. JP21H00121, No. JP21K03553, and No. JP22H05112 and by National Natural Science Foundation of China (NSFC) Grant No. 12075061.

### APPENDIX: IMPROVEMENTS WITH LINK-VARIABLE INPUT

We revisit efficiency of the link-variable input to the neural network. As already shown in Fig. 1, the average phase factor does not increase with the link-variable input in our setup. According to the universal approximation theorem [60,61], however, any continuous function can be expressed in a neural network with a large number of units in hidden layers. The failure of the optimization implies lack of the expressive power in the present neural network, or lack of enough number of training data for the network to learn the gauge symmetry.

We examine the improvements in the expression power of the network. The default network parameters (hyperparameters) are as follows [47]. The batch size (number of configurations used in calculating the derivatives in the stochastic gradient method) is  $N_{\text{batch}} = 10$ , the number of the hidden layers is  $N_{\text{lay}} = 1$ , and the number of units in the hidden layer is  $N_{\text{unit}} = 10$  on a  $2 \times 2$  lattice. These hyperparameters are increased to explore enhancement of the average phase factor.

We also test enlargement of the gauge configurations by use of the gauge transformation. After generating gauge configurations in the standard way, we adopt the gauge transformation to the configurations. It produces additional configurations though equivalent under the

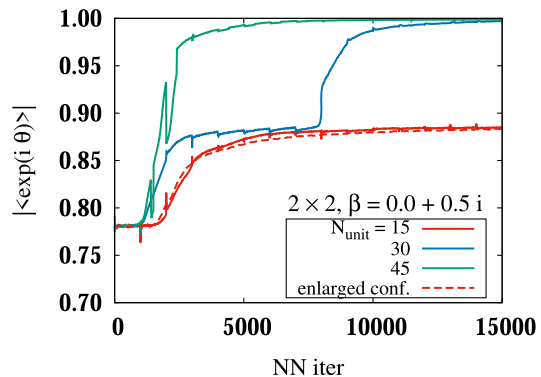


FIG. 5. The average phase factor by the POM with the gauge-link input as a function of the neural-network iteration at  $\beta = 0.5i$  on a  $2 \times 2$  lattice. The number of units in the hidden layer is  $N_{\text{unit}} = 15, 30, 45$ . The result of the enlarged configurations using  $N_{\text{unit}} = 15$  is also plotted.

gauge symmetry. This approach is similar to the data augmentation known in the machine learning community. Through the data augmentation, we can increase the training data and avoid the overtraining. The neural network with the enlarged gauge-transformed configurations is expected to learn the symmetry of the input data better. With this expectation, we have performed the optimization using the enlarged configurations.

In Fig. 5 we show the average phase factor using the link input with increased hyperparameters  $N_{\text{batch}} = 1000$ ,  $N_{\text{lay}} = 5$ , and  $N_{\text{unit}} = 15, 30, 45$  on a  $2 \times 2$  lattice. The enhancement of the average phase factor with larger  $N_{\text{batch}}$ ,  $N_{\text{lay}}$ , and  $N_{\text{unit}}$  is observed, as expected from the universal approximation theorem. We also find the gauge configurations enlarged by a factor of 100 using the gauge transformation give no additional enhancement, and hence less effective. With  $N_{\text{unit}} = 30$  and 45, the average phase factor go through a sudden increase once or twice and reaches 0.99 after 12,000 and 5,000 iterations, respectively. The  $N_{\text{unit}} = 15$  case seems to require a larger number of the neural-network iterations for the average phase factor of 0.99.

From the numerical results, we can clearly see that even a simple neural network using the link input can learn the gauge symmetry, though the cost of the optimization is significantly higher than that of the gauge-invariant input and the gauge-covariant neural network.

[1] K. Nagata, *Prog. Part. Nucl. Phys.* **127**, 103991 (2022).  
 [2] J. R. Klauder, *Phys. Rev. A* **29**, 2036 (1984).  
 [3] G. Parisi, *Phys. Lett.* **131B**, 393 (1983).

[4] D. Sexty, *Phys. Lett. B* **729**, 108 (2014).  
 [5] G. Aarts, E. Seiler, D. Sexty, and I.-O. Stamatescu, *Phys. Rev. D* **90**, 114505 (2014).

- [6] Z. Fodor, S. D. Katz, D. Sexty, and C. Török, *Phys. Rev. D* **92**, 094516 (2015).
- [7] K. Nagata, J. Nishimura, and S. Shimasaki, *Phys. Rev. D* **98**, 114513 (2018).
- [8] J. B. Kogut and D. K. Sinclair, *Phys. Rev. D* **100**, 054512 (2019).
- [9] D. Sexty, *Phys. Rev. D* **100**, 074503 (2019).
- [10] M. Scherzer, D. Sexty, and I. O. Stamatescu, *Phys. Rev. D* **102**, 014515 (2020).
- [11] Y. Ito, H. Matsufuru, Y. Namekawa, J. Nishimura, S. Shimasaki, A. Tsuchiya, and S. Tsutsui, *J. High Energy Phys.* **10** (2020) 144.
- [12] G. Aarts, E. Seiler, and I.-O. Stamatescu, *Phys. Rev. D* **81**, 054508 (2010).
- [13] K. Nagata, J. Nishimura, and S. Shimasaki, *Phys. Rev. D* **94**, 114515 (2016).
- [14] M. Levin and C. P. Nave, *Phys. Rev. Lett.* **99**, 120601 (2007).
- [15] S. Akiyama, Y. Kuramashi, T. Yamashita, and Y. Yoshimura, *Phys. Rev. D* **100**, 054510 (2019).
- [16] S. Akiyama, D. Kadoh, Y. Kuramashi, T. Yamashita, and Y. Yoshimura, *J. High Energy Phys.* **09** (2020) 177.
- [17] S. Akiyama, Y. Kuramashi, T. Yamashita, and Y. Yoshimura, *J. High Energy Phys.* **01** (2021) 121.
- [18] S. Akiyama, Y. Kuramashi, and Y. Yoshimura, *Phys. Rev. D* **104**, 034507 (2021).
- [19] A. Bazavov, S. Catterall, R. G. Jha, and J. Unmuth-Yockey, *Phys. Rev. D* **99**, 114507 (2019).
- [20] M. Asaduzzaman, S. Catterall, and J. Unmuth-Yockey, *Phys. Rev. D* **102**, 054510 (2020).
- [21] M. Fukuma, D. Kadoh, and N. Matsumoto, *Prog. Theor. Exp. Phys.* **2021**, 123B03 (2021).
- [22] M. Hirasawa, A. Matsumoto, J. Nishimura, and A. Yosprakob, *J. High Energy Phys.* **12** (2021) 011.
- [23] T. Kuwahara and A. Tsuchiya, *Prog. Theor. Exp. Phys.* **2022**, 093B02 (2022).
- [24] D. Kadoh and K. Nakayama, [arXiv:1912.02414](https://arxiv.org/abs/1912.02414).
- [25] D. Kadoh, H. Oba, and S. Takeda, *J. High Energy Phys.* **04** (2022) 121.
- [26] E. Witten, *AMS/IP Stud. Adv. Math.* **50**, 347 (2011).
- [27] A. Alexandru, G. Basar, P. F. Bedaque, G. W. Ridgway, and N. C. Warrington, *J. High Energy Phys.* **05** (2016) 053.
- [28] M. Cristoforetti, F. Di Renzo, and L. Scorzato (Aurora Science Collaboration), *Phys. Rev. D* **86**, 074506 (2012).
- [29] A. Mukherjee, M. Cristoforetti, and L. Scorzato, *Phys. Rev. D* **88**, 051502 (2013).
- [30] H. Fujii, D. Honda, M. Kato, Y. Kikukawa, S. Komatsu, and T. Sano, *J. High Energy Phys.* **10** (2013) 147.
- [31] M. Fukuma and N. Matsumoto, *Prog. Theor. Exp. Phys.* **2021**, 023B08 (2021).
- [32] M. Fukuma, N. Matsumoto, and Y. Namekawa, *Prog. Theor. Exp. Phys.* **2021**, 123B02 (2021).
- [33] Y. Mori, K. Kashiwa, and A. Ohnishi, *Phys. Rev. D* **96**, 111501 (2017).
- [34] Y. Mori, K. Kashiwa, and A. Ohnishi, *Prog. Theor. Exp. Phys.* **2018**, 023B04 (2018).
- [35] A. Alexandru, P. F. Bedaque, H. Lamm, and S. Lawrence, *Phys. Rev. D* **97**, 094510 (2018).
- [36] S. Lawrence, H. Oh, and Y. Yamauchi, *Phys. Rev. D* **106**, 114503 (2022).
- [37] K. Kashiwa, Y. Mori, and A. Ohnishi, *Phys. Rev. D* **99**, 014033 (2019).
- [38] K. Kashiwa, Y. Mori, and A. Ohnishi, *Phys. Rev. D* **99**, 114005 (2019).
- [39] A. Alexandru, P. F. Bedaque, H. Lamm, S. Lawrence, and N. C. Warrington, *Phys. Rev. Lett.* **121**, 191602 (2018).
- [40] F. Bursa and M. Kroyter, *J. High Energy Phys.* **12** (2018) 054.
- [41] Y. Mori, K. Kashiwa, and A. Ohnishi, *Prog. Theor. Exp. Phys.* **2019**, 113B01 (2019).
- [42] K. Kashiwa and Y. Mori, *Phys. Rev. D* **102**, 054519 (2020).
- [43] M. Giordano, K. Kapas, S. D. Katz, A. Pasztor, and Z. Tulipant, *Phys. Rev. D* **106**, 054512 (2022).
- [44] W. Detmold, G. Kanwar, H. Lamm, M. L. Wagman, and N. C. Warrington, *Phys. Rev. D* **103**, 094517 (2021).
- [45] A. Alexandru, G. Basar, P. F. Bedaque, and N. C. Warrington, *Rev. Mod. Phys.* **94**, 015006 (2022).
- [46] J. M. Pawłowski, M. Scherzer, C. Schmidt, F. P. G. Ziegler, and F. Ziesché, *Phys. Rev. D* **103**, 094505 (2021).
- [47] Y. Namekawa, K. Kashiwa, A. Ohnishi, and H. Takase, *Phys. Rev. D* **105**, 034502 (2022).
- [48] U. J. Wiese, *Nucl. Phys.* **B318**, 153 (1989).
- [49] B. E. Rusakov, *Mod. Phys. Lett. A* **05**, 693 (1990).
- [50] C. Bonati and P. Rossi, *Phys. Rev. D* **99**, 054503 (2019).
- [51] M. Favoni, A. Ipp, D. I. Müller, and D. Schuh, *Phys. Rev. Lett.* **128**, 032003 (2022).
- [52] A. Tomiya and Y. Nagai, [arXiv:2103.11965](https://arxiv.org/abs/2103.11965).
- [53] R. H. Swendsen and J.-S. Wang, *Phys. Rev. Lett.* **57**, 2607 (1986).
- [54] C. J. Geyer, *Comput. Sci. Stat.* **23**, 156 (1991).
- [55] E. Marinari and G. Parisi, *Europhys. Lett.* **19**, 451 (1992).
- [56] K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
- [57] M. Fukuma and N. Umeda, *Prog. Theor. Exp. Phys.* **2017**, 073B01 (2017).
- [58] M. Fukuma, N. Matsumoto, and N. Umeda, *Phys. Rev. D* **100**, 114510 (2019).
- [59] M. Fukuma, N. Matsumoto, and N. Umeda, [arXiv:1912.13303](https://arxiv.org/abs/1912.13303).
- [60] G. Cybenko, *Math. Control. Signal. Syst. (MCSS)* **2**, 303 (1989).
- [61] K. Hornik, *Neural Netw.* **4**, 251 (1991).
- [62] A. Alexandru, G. Basar, P. F. Bedaque, G. W. Ridgway, and N. C. Warrington, *Phys. Rev. D* **93**, 094514 (2016).
- [63] A. Alexandru, G. Basar, P. F. Bedaque, and G. W. Ridgway, *Phys. Rev. D* **95**, 114501 (2017).
- [64] M. Rodekamp, E. Berkowitz, C. Gäntgen, S. Krieg, T. Luu, and J. Ostmeier, *Phys. Rev. B* **106**, 125139 (2022).
- [65] D. P. Kingma and J. Ba, in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*(2014), [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [66] S. Choe, S. Muroya, A. Nakamura, C. Nonaka, T. Saito, and F. Shoji, *Nucl. Phys. B, Proc. Suppl.* **106**, 1037 (2002).