Path Optimization for the Sign Problem in Field Theories using Neural Network

Akira Ohnishi[1], Yuto Mori[2], and Kouji Kashiwa[3]

[1] *Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto 606-8502, Japan*
[2] *Department of Physics, Faculty of Science, Kyoto University, Kyoto 606-8502, Japan*
[3] *Fukuoka Institute of Technology, Wajiro, Fukuoka 811-0295, Japan*

E-mail: ohnishi@yukawa.kyoto-u.ac.jp

(Received February 10, 2019)

We investigate the sign problem in field theories by using the path optimization method (POM) with use of the feedforward neural network (FNN). We utilize FNN to prepare and optimize the trial function specifying the integration path (manifold) in field theories in the framework of POM. POM with use of FNN has been applied to several field theories having the sign problem. Specifically, the 0+1 dimensional QCD is discussed. It is found that the average phase factor is enhanced significantly and we can reduce the statistical errors of observables.

KEYWORDS: Sign problem, Lattice QCD, Path Optimization Method

1. Introduction

At finite chemical potential μ in lattice field theories, the fermion determinant generally takes a complex number, then the statistical weight becomes complex. With the complex action, strong cancellation of the weight occurs at large volume. This *sign problem* is one of the serious problems in theoretical physics. A promising idea to attack the sign problem is to complexify the integration variables, and several complexified variable methods have been proposed, such as the Lefschetz thimble method [1, 2], the complex Langevin method [3], and the path optimization method [4–11].

In the path optimization method, parameterized trial functions specifying the integration path (manifold) are optimized to evade the sign problem. Provided that the complex Boltzmann weight is a holomorphic (complex analytic) function of the complexified variable, the Cauchy(-Poincare) theorem tells us that the partition function is independent of the integration path as long as the path is modified continuously from the real axis without going through the singular point or cuts. Then as in the saddle point integral, it would be possible to obtain the integral of complex valued Boltzmann weight efficiently by modifying the integration path from the real axis. In order to prepare and optimize the integration path in the field theories on the lattice having with many variables, it is useful to apply the neural network, which can handle many multi-variable functions.

The path optimization method with the neural network has been applied to the one-dimensional integral [5], 1+1 dimensional complex ϕ^4 theory [6], the Polyakov loop extended Nambu-Jona-Lasinio (PNJL) model [7], and the 0+1 dimensional QCD [8, 9], as discussed in the presentation at QNP 2018. In these cases, we find that the sign problem is weakened and observables are obtained correctly and efficiently. In this proceedings, we examine the validity and usefulness of the path optimization method using the neural network in field theories having the sign problem, and discuss results of the 0+1 dimensional QCD [8, 9].

2. Path optimization with use of neural network

Let us consider the Boltzmann weight $\exp(-S)$, which is a holomorphic function of the complexified field variables $z(x) = \{z_i = x_i + iy_i; i = 1, 2, \ldots N\}$, with N being the number of variables. The partition function can be written as

$$\mathcal{Z} = \int_{C_{\mathbb{R}}} d^N x \, \exp(-S(x)) = \int_{C_{\mathbb{C}}} d^N z \, \exp(-S(z)) = \int_{C_{\mathbb{R}}} d^N x \, J(z) \exp(-S(z)) \,, \tag{1}$$

where $J(z) = \det(\partial z_i/\partial x_j) = \det(\delta_{ij} + i\partial y_i/\partial x_j)$ is the Jacobian. We here assume that the imaginary parts are given as functions of the real parts, $y(x) = \{y_i(x); i = 1, 2, \ldots N\}$, which specify the integration path. In the path optimization method, we optimize $y(x)$ to reduce the cost function,

$$\mathcal{F}[y] = \mathcal{Z}_{\rm pq} - |\mathcal{Z}| = |\mathcal{Z}| \left(|\langle e^{i\theta} \rangle|^{-1} - 1 \right) \,, \tag{2}$$

where θ is the complex phase of the statistical weight, $W = J(z)\exp(-S(z))$, and $\mathcal{Z}_{\rm pq}$ is the phase quenched partition function. Since the partition function is independent of the path, the above cost function is a monotonically decreasing function of the average phase factor.

The feedforward neural network (FNN) in a useful tool to prepare and optimize the integration path $y(x)$ in field theories. FNN consists of the input, hidden, and output layers, and each layer consists of units which mimic neurons in the brain. Units in one layer are connected with those in the previous and next layers, and the variables are modified by combining the linear and non-linear transformations. In FNN, any functions can be described in the large unit-number limit even in the case of a single hidden layer [12]. It should be noted that the present optimization of FNN parameters corresponds to the so-called *unsupervised learning*. Since we do not know the correct integration path in advance, we cannot prepare the teacher data and cannot perform the usual *supervised* learning.

We generate the Monte Carlo configurations by the hybrid Monte Carlo (HMC). We regard the real part of variables as the dynamical variable, where the molecular dynamics Hamiltonian is $H(x, p) = p^2/2 + \mathrm{Re}S(z(x))$. Jacobian effects are included at the Metropolis judgement, while they are ignored in the molecular dynamics evolution. We prepare $N_{\rm config}$ configurations, update parameters using $N_{\rm batch}$ configurations in each mini-batch training. After updating parameters $N_{\rm config}/N_{\rm batch}$ times, we generate $N_{\rm config}$ configurations using the updated neural network parameters. We repeat this update process $N_{\rm epoch}$ times.

3. Application to gauge theory: 0+1 dimensional QCD

We consider the 0+1 dimensional QCD (QCD_{0+1}) with one species of staggered Fermion at finite μ [14]. The lattice action is given as

$$S = \frac{1}{2} \sum_{\tau} \left(\bar{\chi}_\tau e^{\mu} U_\tau \chi_{\tau+\hat{0}} - \bar{\chi}_{\tau+\hat{0}} e^{-\mu} U_\tau^{-1} \chi_\tau \right) + m \sum_{\tau} \bar{\chi}_\tau \chi_\tau \,, \tag{3}$$

where χ_τ and U_τ are the staggered Fermion and the link variable, respectively, and m is the bare mass. QCD_{0+1} can be reduced to be a one link problem by using the gauge transformation, and the partition function is known to be obtained analytically as

$$\mathcal{Z} = \frac{\sinh[(N_c + 1)E/T]}{\sinh[E/T]} + 2\cosh(N_c \mu/T) \,, \tag{4}$$

where $E = \mathrm{arcsinh}\, m$ and $T = 1/N_\tau$ with N_τ being the temporal lattice size. While the QCD_{0+1} is a toy model, the origin of the sign problem, the temporal hopping term, is the same as that in the 3+1 dimensional QCD (QCD_{3+1}).

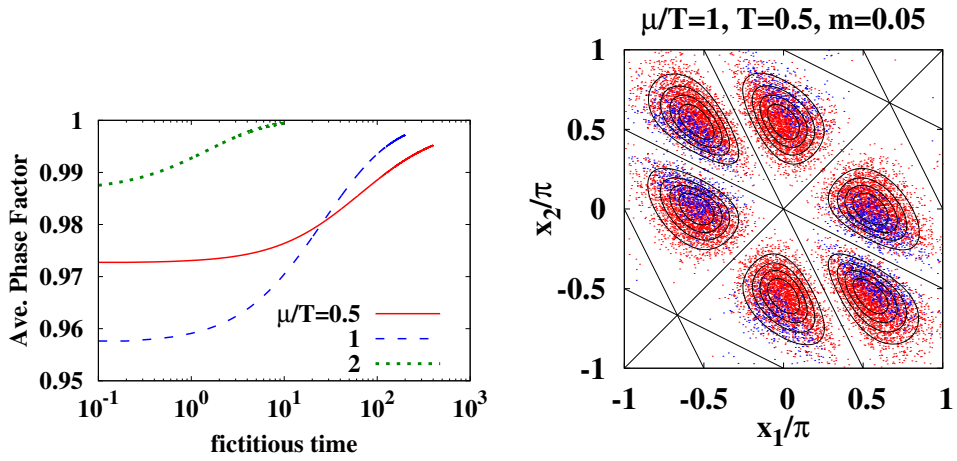

Fig. 1. The average phase factor and statistical weight distribution in 0+1 dimensional QCD at $T = 0.5$ are shown. In the left panel, we show the average phase factors at $\mu/T = 0.5$, 1 and 2 with diagonalized gauge fixing as functions of the fictitious time. In the right panel, we show the statistical weight after path optimization with and without diagonalized gauge fixing by contours and by dots, respectively.

We have performed the path optimization in two ways, with and without the diagonalized gauge fixing. In the former, the link variables are diagonalized to be $U = \mathrm{diag}(e^{ix_1}, e^{ix_2}, e^{ix_3})$ with $x_1 + x_2 + x_3 = 0$ by using the residual gauge degrees of freedom, then the measure is given as $dU = Hdz_1dz_2$, where the Haar measure is given as $H = 8/3\pi^2 \prod_{a<b} \sin^2[(z_a - z_b)/2]$. We have only two variables in the diagonalized gauge, then it is possible to integrate on the two dimensional mesh points of (x_1, x_2). The variational parameters are the imaginary parts on the mesh points (y_1, y_2) themselves. The gradient descent equation reads $dy_i(x_1, x_2)/dt = -\partial\mathcal{F}/\partial y_i(x_1, x_2)$, where t is the fictitious time. By optimizing the path, the average phase factor (APF) goes above 0.99 easily as shown in the left panel of Fig. 1. While the APF on the original path is above 0.95 and already large, the reduction of $(1 - \mathrm{APF})$ is significant. In QCD_{3+1}, the action in QCD_{0+1} appears in each spatial point on the lattice, then APF in QCD_{3+1} on a $L^3 \times N_\tau$ lattice would be roughly estimated as $\mathrm{APF}_{3+1} \simeq (\mathrm{APF}_{0+1})^{L^3}$, provided that other action terms do not make it worse. Then $\mathrm{APF}_{0+1} = 0.95$ leads to $\mathrm{APF}_{3+1} \simeq 4 \times 10^{-12}$ on a $8^3 \times N_\tau$ lattice, while $\mathrm{APF}_{0+1} = 0.995$ gives $\mathrm{APF}_{3+1} \simeq 0.08$. This APF is not large but it would be enough to obtain some meaningful results.

With the diagonalized gauge fixing, the statistical distribution is separated in the six regions, as shown by the contour in the right panel of Fig. 1. This separation comes from the Haar measure, which is zero at $z_a = z_b$ $(a \neq b)$. The separated distribution does not cause trouble in the mesh point integral, but is problematic in HMC simulations.

Now we proceed to link variable sampling without diagonalized gauge fixing. The SU(3) link variable U is complexified to a SL(3) variable $\mathcal{U}$. We here adopt the complexification, $U \in \mathrm{SU}(3) \to \mathcal{U}(U) = U \prod_{a=1}^{N_c^2-1} e^{y_a\lambda_a/2} \in \mathrm{SL}(3)$. This form of the link variable is convenient to calculate the derivative of the matrix element with respect to the real part of variables (matrix elements of $U \in \mathrm{SU}(3)$) as well as the imaginary part, y_a. We adopt HMC for sampling, and utilize FNN for optimization. In HMC, we regard U as the dynamical variable of the molecular dynamics, and the Hamiltonian is taken to be $H = P^2/2 + \mathrm{Re}S(\mathcal{U}(U))$, where P is the conjugate momentum matrix of U. After optimization, we generate configurations in HMC using the optimized FNN. The SL(3) link variables are diagonalized by the similarity transformation, $\mathrm{diag}(e^{iz_1}, e^{iz_2}, e^{iz_3})$ with $z_1 + z_2 + z_3 = 0$. In the right panel of Fig. 1, we show the distribution of (x_a, x_b) $(a \neq b)$ with $x_a = \mathrm{Re}\, z_a (a = 1, 2, 3)$ by dots. We show the distribution with $(a, b) = (1, 2)$ with blue dots and other combinations of (a, b) with

red dots. This distribution is consistent with that in the probability distribution in the diagonalized gauge, and the six separated regions are found to be visited in the HMC sampling. We have also confirmed that we can reproduce the exact results of the quark condensate, quark number density, and the Polyakov loop in both of the treatment of link variables within the statistical errors. These results will be reported in the forthcoming paper [9].

4. Summary

We have discussed the sign problem in field theories by using the path optimization method with use of the neural network. In the path optimization method, we complexify integration variables as $x_i \to z_i = x_i + iy_i$, and optimize the imaginary parts as functions of real parts, $y_i(x)$, to evade the sign problem. For the preparation and optimization of the integration path specified by $y(x)$, we adopt the feedforward neural network, which can describe any functions in the large unit-number limit. We have demonstrated that the average phase factor is enhanced significantly in the 0+1 dimensional QCD at finite μ.

It is desired to apply the present framework to more realistic theories. For this purpose, one of the obstacles is the numerical cost to calculate the Jacobian and its derivatives. In order to reduce the cost, it is helpful to assume the function form of the imaginary part [10, 11]. These assumptions have been demonstrated to work well in enhancing the average phase factor, and we can save the numerical cost for a sparse Jacobian matrix.

This work is supported in part by the Grants-in-Aid for Scientific Research from JSPS (Nos. 15H03663, 16K05350, 18K03618), and by the Yukawa International Program for Quark-hadron Sciences (YIPQS).

References

[1] E. Witten, AMS/IP Stud. Adv. Math. **50** (2011) 347;
M. Cristoforetti *et al.* [AuroraScience Collaboration], Phys. Rev. D **86** (2012) 074506;
H. Fujii, D. Honda, M. Kato, Y. Kikukawa, S. Komatsu, T. Sano, JHEP **1310** (2013) 147.
[2] A. Alexandru, G. Basar, P. F. Bedaque, G. W. Ridgway, N. C. Warrington, JHEP **1605** (2016) 053.
[3] G. Parisi and Y. s. Wu, Sci. Sin. **24** (1981) 483; J. R. Klauder, Acta Phys. Austriaca Suppl. **25** (1983) 251; G. Aarts, E. Seiler, I. O. Stamatescu, Phys. Rev. D **81** (2010) 054508; K. Nagata, J. Nishimura, S. Shimasaki, Phys. Rev. D **94** (2016) 114515; E. Seiler, D. Sexty, I. O. Stamatescu, Phys. Lett. B **723** (2013) 213; Y. Ito and J. Nishimura, JHEP **1612** (2016) 009;
[4] Y. Mori, K. Kashiwa, A. Ohnishi, Phys. Rev. D **96** (2017) 111501.
[5] A. Ohnishi, Y. Mori, K. Kashiwa, EPJ Web Conf. **175** (2018) 07043.
[6] Y. Mori, K. Kashiwa, A. Ohnishi, PTEP **2018** (2018) 023B04.
[7] K. Kashiwa, Y. Mori, A. Ohnishi, Phys. Rev. D **99** (2019), 014033.
[8] A. Ohnishi, Y. Mori and K. Kashiwa, arXiv:1812.11506 [hep-lat].
[9] Y. Mori, K. Kashiwa, A. Ohnishi, in preparation.
[10] A. Alexandru, P. F. Bedaque, H. Lamm, S. Lawrence, Phys. Rev. D **97** (2018) 094510.
[11] F. Bursa and M. Kroyter, JHEP **1812** (2018) 054.
[12] G. Cybenko, Mathematics of control, signals and systems (MCSS) **2** (1989) 303;
K. Hornik, M. Stinchcombe, H. White, Neural Networks **2** (1989) 359.
[13] M. D. Zeiler, arXiv:1212.5701.
[14] N. Bilic and K. Demeterfi, Phys. Lett. B **212** (1988) 83; L. Ravagli and J. J. M. Verbaarschot, Phys. Rev. D **76** (2007) 054506; G. Aarts and K. Splittorff, JHEP **1008** (2010) 017; J. Bloch, J. Phys. Conf. Ser. **432** (2013) 012023; C. Schmidt and F. Ziesché, PoS LATTICE **2016** (2017) 076; F. Di Renzo, and G. Eruzzi, Phys. Rev. D **97** (2018) 014503.