

PAPER • OPEN ACCESS

## Noise enhanced neural networks for analytic continuation

To cite this article: Juan Yao *et al* 2022 *Mach. Learn.: Sci. Technol.* **3** 025010

View the [article online](#) for updates and enhancements.

### You may also like

- [A first-principles study on the quantum tunneling of methylhydroxycarbene isomerization in various solvents](#)  
E D Susanti, F Rusydi, I Puspitasari et al.
- [Research on the Law of Stress and Strain Response of Surrounding Rock of Underground Powerhouse Cavity Excavated by Layered Cutting](#)  
Xue Lin, Yang Gao, Yanbo Jiang et al.
- [Application of Glass Fiber Concrete under Thin-Walled Middle Partition Wall Construction in Tunnel](#)  
Yuanyuan Tian, Yigong Wang, Haixiang Nie et al.



## PAPER

# Noise enhanced neural networks for analytic continuation

## OPEN ACCESS

Juan Yao<sup>1,2,3</sup> , Ce Wang<sup>4</sup>, Zhiyuan Yao<sup>4</sup> and Hui Zhai<sup>4,\*</sup>

RECEIVED  
4 May 2022

ACCEPTED FOR PUBLICATION  
12 May 2022

PUBLISHED  
26 May 2022

<sup>1</sup> Shenzhen Institute for Quantum Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, People's Republic of China

<sup>2</sup> International Quantum Academy, Shenzhen 518048, People's Republic of China

<sup>3</sup> Guangdong Provincial Key Laboratory of Quantum Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, People's Republic of China

<sup>4</sup> Institute for Advanced Study, Tsinghua University, Beijing 100084, People's Republic of China

\* Author to whom any correspondence should be addressed.

E-mail: [huizhai.physics@gmail.com](mailto:huizhai.physics@gmail.com)

**Keywords:** machine learning, analytic continuation, neural networks

Original Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



## Abstract

Analytic continuation maps imaginary-time Green's functions obtained by various theoretical/numerical methods to real-time response functions that can be directly compared with experiments. Analytic continuation is an important bridge between many-body theories and experiments but is also a challenging problem because such mappings are ill-conditioned. In this work, we develop a neural network (NN)-based method for this problem. The training data is generated either using synthetic Gaussian-type spectral functions or from exactly solvable models where the analytic continuation can be obtained analytically. Then, we applied the trained NN to the testing data, either with synthetic noise or intrinsic noise in Monte Carlo simulations. We conclude that the best performance is always achieved when a proper amount of noise is added to the training data. Moreover, our method can successfully capture multi-peak structure in the resulting response function for the cases with the best performance. The method can be combined with Monte Carlo simulations to compare with experiments on real-time dynamics.

## 1. Introduction

Analytic continuation is an important problem in computational quantum many-body physics [1, 2]. For quantum many-body systems, a variety of methods, ranging from perturbation approximations to quantum Monte Carlo algorithms, are performed in imaginary time and only produce imaginary-time Green's function [3–5]. Nevertheless, to compare with real-time responses measured in experiments, we need the response function or the closely related spectral function  $\mathcal{A}(\Omega)$  from which the response function can be calculated. On the other hand, the spectral function and the Matsubara frequency representation of the imaginary-time Green's function,  $\mathcal{G}(i\omega_n)$ , are related by [6, 7]

$$\mathcal{G}(i\omega_n) = \int \frac{\mathcal{A}(\Omega)}{i\omega_n - \Omega} d\Omega. \quad (1)$$

Hence this relation is crucial to analytically continue the imaginary-time Green's function to real time that connects numerical results to experimental observations.

However, extracting  $\mathcal{A}(\Omega)$  from the above relation, equation (1), is a difficult problem. Although the mapping from  $\mathcal{A}(\Omega)$  to  $\mathcal{G}(i\omega_n)$  is linear, the coefficients decrease and approach zero as  $\Omega$  increases, making the inverse of such mapping ill-conditioned [8–10]. That is to say, a small noise embedded in  $\mathcal{G}(i\omega_n)$  can be significantly amplified and lead to huge errors in the evaluated  $\mathcal{A}(\Omega)$ . Since  $\mathcal{G}(i\omega_n)$  obtained by various numerical methods inevitably contains noise, thus obtained  $\mathcal{A}(\Omega)$  always has large errors, which makes the comparison with experimental measurements unreliable. This is the intrinsic difficulty in the analytic continuation [8–10].

Many methods have been developed to address this problem. One of the most popular approaches is the maximum entropy (MaxEnt) method which aims to find the most probable  $A(\Omega)$  that maximizes an entropy functional [11–18]. However, the method is biased toward Gaussian type spectral functions, and sharp peaks or edges in the spectral function can be easily missed. Several stochastic methods have then been proposed to recover these singular features [19–25]. Nevertheless, these stochastic methods are usually time-consuming [28]. Many of the existing methods also rely on prior physical information. For instance, relying on the conjecture of holographic duality and the insight from gravity physics, the analytic continuation has also been applied to study quantum critical dynamics [26].

The fast development of applying machine learning methods to physics problems offer a new route to address this problem. The neural network (NN) is a useful tool to express functional mapping that can also tolerant errors. Nevertheless, the neural-network-based methods have only been discussed in few recent works on the analytic continuation problem [27–30]. While these method offer new promise, lots of issues remain open. Especially, since the difficulty of the analytic continuation roots in the amplification of the noise, one basis question is whether we should add noise into the training data and, if so, what is the proper amplitude of the noise. To the best of our knowledge, this issue has not been well studied before. In this work, we systematically study the noise effect in the training dataset. The key finding is that the NN's performance can be significantly enhanced when a proper amount of noise is added to the data.

## 2. Framework

Below we first introduce our general framework that includes both the training data preparation and the training scheme.

To prepare training data, we need to include a variety of spectral functions  $\mathcal{A}(\Omega)$ . For each given type of  $\mathcal{A}(\Omega)$ , the corresponding imaginary-frequency Green's function  $\mathcal{G}(i\omega_n)$  can be generated according to equation (1). Each pair of  $\{\mathcal{G}(i\omega_n), \mathcal{A}(\Omega)\}$  contributes one sample in the dataset. Since both the input and the output of NN are discrete, we shall first discretize  $\mathcal{G}(i\omega_n)$  and  $\mathcal{A}(\Omega)$ . We discretize  $\Omega$  into  $N_{\text{out}}$  number of points  $\Omega_j$ ,  $\{j = 1, \dots, N_{\text{out}}\}$  in a proper range of  $\Omega$ , and we normalize  $\mathcal{A}(\Omega_j)$  to  $\bar{\mathcal{A}}(\Omega_j)$  such that  $\sum_{j=1}^{N_{\text{out}}} \bar{\mathcal{A}}(\Omega_j) = 1$ . Then, the output spectral function  $\bar{\mathcal{A}}(\Omega)$  is represented by a discretized vector  $\{\bar{\mathcal{A}}(\Omega_j) | j = 1, \dots, N_{\text{out}}\}$ . For the input Green's function,  $\omega_n = (2n + 1)\pi/\beta$  where  $n = 1, \dots, N_{\text{in}}$  and  $\beta$  is set as unity.  $\mathcal{G}(i\omega_n)$  is generally a complex number. To mimic the inevitable noise in the calculated  $\mathcal{G}(i\omega_n)$  from various numerical methods, we add a randomly sampled noise individually for each point  $\omega_n$ , i.e.

$$\tilde{\mathcal{G}}(i\omega_n) = \mathcal{G}(i\omega_n)(1 + \delta + i\delta'), \quad (2)$$

where noise  $\delta$  and  $\delta'$  satisfy a Gaussian distribution with zero mean and standard deviation  $\eta$ . The input Green's function  $\mathcal{G}(i\omega_n)$  is then represented by a discretized vector  $\{\text{Re}[\tilde{\mathcal{G}}(i\omega_n)], \text{Im}[\tilde{\mathcal{G}}(i\omega_n)] | n = 1, \dots, N_{\text{in}}\}$ .

The multi-layer fully connected NN used in this work is depicted in figure 1. The input layer is the discretized  $\mathcal{G}(i\omega_n)$ , and the output layer is the discretized  $\mathcal{A}(\Omega)$ . The training process is carried out with the Adam optimizer [31], and the Kullback-Leibler divergence is used as the training loss function, which is defined as

$$\text{KLD} = \sum_t \sum_j \bar{\mathcal{A}}^t(\Omega_j) \log [\bar{\mathcal{A}}^t(\Omega_j) / \bar{\mathcal{A}}_{\text{NN}}^t(\Omega_j)], \quad (3)$$

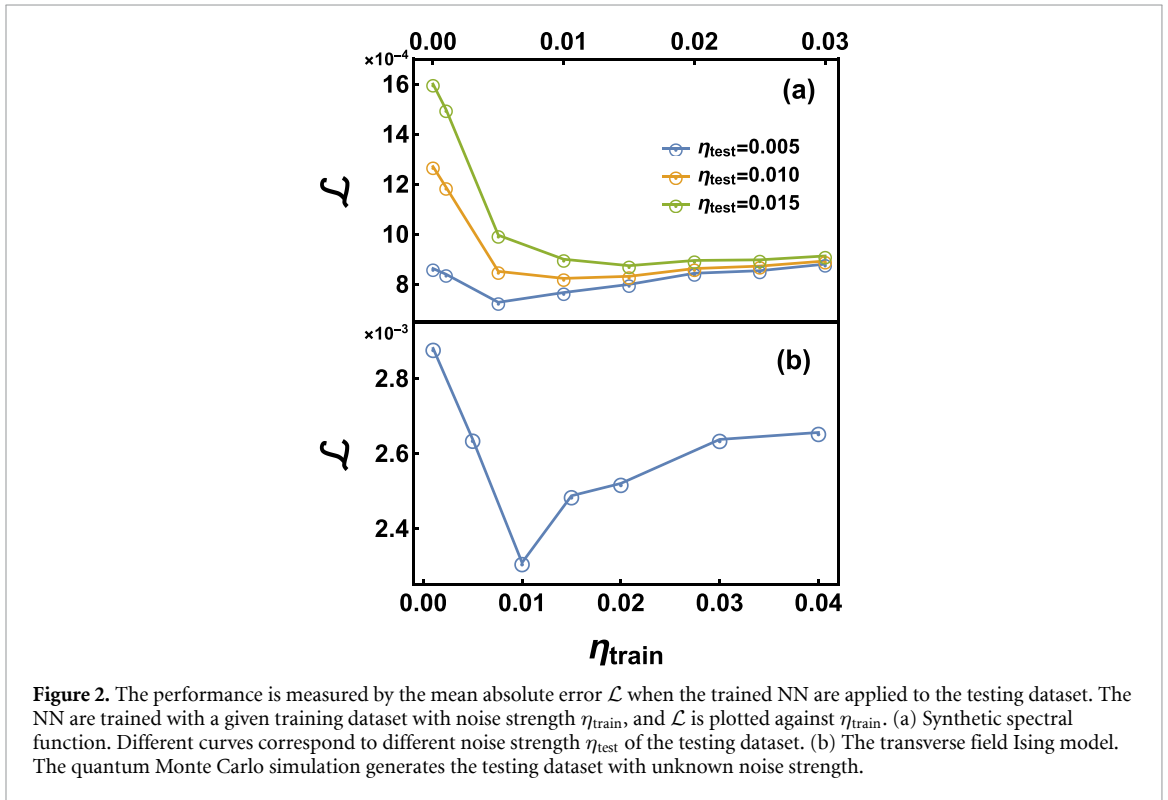
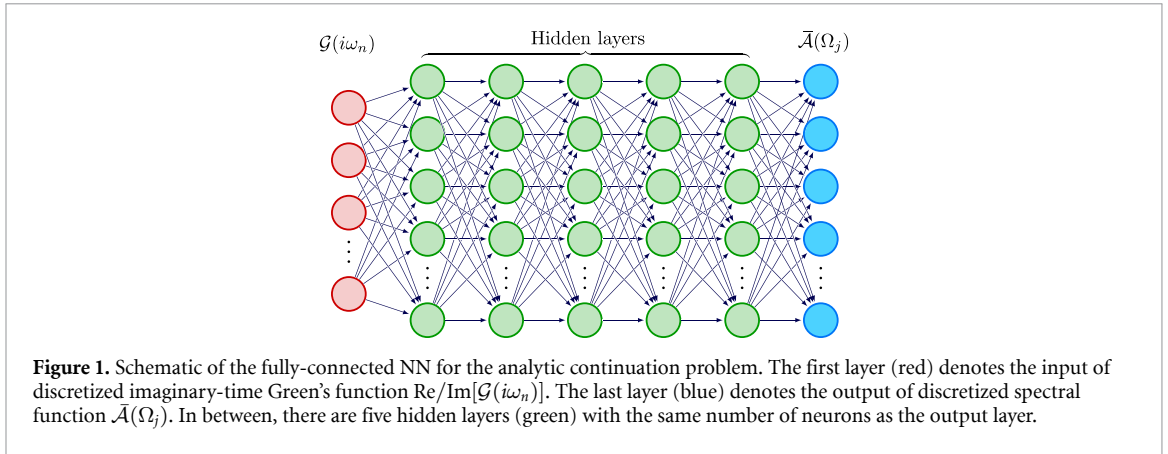
where  $t$  labels samples in the dataset.  $\{\bar{\mathcal{A}}(\Omega_j)\}$  are the labels and  $\{\bar{\mathcal{A}}_{\text{NN}}(\Omega_j)\}$  are the outputs of the NN. After training, the mean absolute error on the testing set is chosen to evaluate the performance, which is defined as

$$\mathcal{L} = \frac{1}{\mathcal{N}} \sum_{t,j} |\bar{\mathcal{A}}_{\text{NN}}^t(\Omega_j) - \bar{\mathcal{A}}^t(\Omega_j)|. \quad (4)$$

Here the summation runs over all samples in the testing dataset, and  $\mathcal{N}$  is the product of the dimension of the output and the number of samples in the testing dataset. For all training processes,  $10^5$  number of training samples and  $5 \times 10^3$  testing samples are prepared.

## 3. Results

Below we will present our results on two cases, one on synthetic spectral functions and the other on the transverse field Ising model.



### 3.1. Synthetic spectral functions

We generate the synthetic spectral function as a summation of Gaussian distributions.

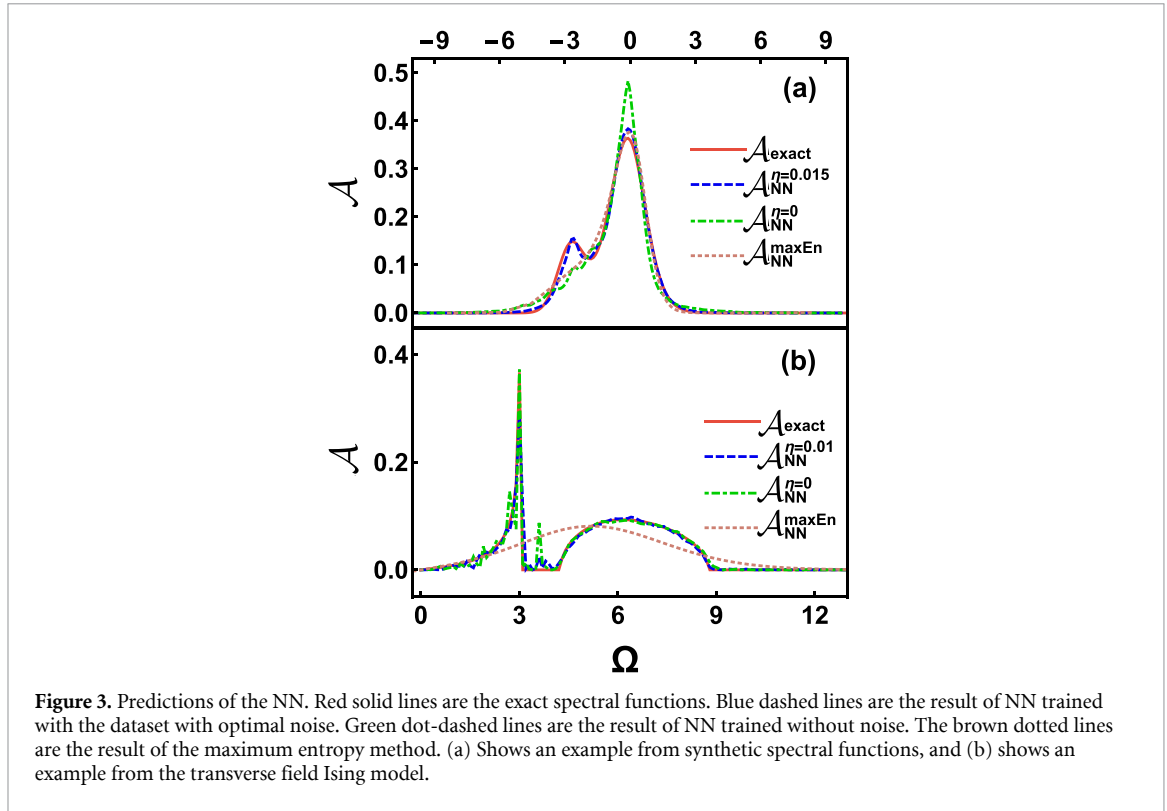
$$\mathcal{A}(\Omega) = \sum_i \lambda_i \mathcal{N}(\Omega|\mu_i, \sigma_i). \tag{5}$$

Here  $\mathcal{N}(\Omega|\mu, \sigma)$  denotes a Gaussian distribution centered at  $\mu$  with width  $\sigma$ , that is

$$\mathcal{N}(\Omega|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(\Omega - \mu)^2\right\}. \tag{6}$$

The coefficients  $\lambda_i$  take random value in  $[0, 1]$  under the constraint  $\sum_i \lambda_i = 1$ , which ensures the normalization condition  $\int d\Omega \mathcal{A}(\Omega) = 1$ . Here the maximum number of Gaussian components is three. The mean value  $\mu$  of each Gaussian distribution ranges from  $[-6, 6]$  and the width  $\sigma$  ranges from  $[0.1, 4]$ . Since these synthetic spectral functions are made of Gaussian components, their corresponding imaginary-frequency Green's function  $\mathcal{G}(i\omega_i)$  can be exactly calculated according to equation (1).

After training, the performance of the NN is measured by the mean absolute value  $\mathcal{L}$  on the testing dataset. Here we add noise with amplitude  $\eta_{\text{test}}$  into the testing dataset in order to mimic the inevitable noise in numerical simulations.  $\mathcal{L}$  is plotted in figure 2(a) for different values of  $\eta_{\text{test}}$ . Smaller of  $\mathcal{L}$  indicates better



**Figure 3.** Predictions of the NN. Red solid lines are the exact spectral functions. Blue dashed lines are the result of NN trained with the dataset with optimal noise. Green dot-dashed lines are the result of NN trained without noise. The brown dotted lines are the result of the maximum entropy method. (a) Shows an example from synthetic spectral functions, and (b) shows an example from the transverse field Ising model.

performance of the trained NN. Figure 2(a) shows that, if  $\eta_{\text{train}}$  is below  $\eta_{\text{test}}$ ,  $\mathcal{L}$  becomes large and the NN does not perform well.  $\mathcal{L}$  approaches the optimal performance when  $\eta_{\text{train}} \approx \eta_{\text{test}}$ . Then,  $\mathcal{L}$  slowly increases when  $\eta_{\text{train}} > \eta_{\text{test}}$ .

Figure 3(a) shows a concrete typical example in the testing dataset with noise amplitude  $\eta_{\text{test}} = 0.015$ . Here we show the NN results with and without noise in the training data, and we also show results from the MaxEnt method for comparison. As one can see, both the MaxEnt method (brown dotted line) and the NN results without noise (green dot-dashed line) miss the first peak. The MaxEnt method tends to generate a smooth and simple distribution with minimal singularities. In contrast, the results can be significantly improved when noise is added into the training data and when the noise amplitude is comparable with that in the test data. The blue dashed line shows that the NN result with  $\eta_{\text{train}} = \eta_{\text{test}} = 0.015$  can successfully capture both positions, amplitudes, and widths of two peaks in the spectral function, with tiny error compared with the exact results (solid red line). Hence, we show that our NN can successfully predict the spectral function, and the performance of the NN can be enhanced by adding noise properly into the training dataset.

### 3.2. The transverse field Ising model

For synthetic spectral functions, we know precisely the amount of noise we add into the testing dataset. However, noise is always unavoidably generated from numerical simulations and its amplitude is generally not precisely known. To address such situations, we consider the one-dimensional transverse field Ising model, whose Hamiltonian reads

$$\hat{H} = -J \sum_l \hat{\sigma}_l^z \hat{\sigma}_{l+1}^z - g \sum_l \hat{\sigma}_l^x, \quad (7)$$

where  $\hat{\sigma}_l$  are the standard Pauli operators on site  $l$ . Here  $g$  is the transverse field strength,  $J$  is the coupling along  $z$ -direction which will be set as a unit below, and the boundary condition is taken to be periodic. Since this model can be solved exactly by mapping to free fermions using the Jordan–Wigner transformation, closed analytic forms of both  $\mathcal{G}(i\omega_n)$  and  $\mathcal{A}(\Omega)$  can be obtained (see appendix A for details). Because the system is translational invariant, the calculation is mostly easily done in momentum space. Nevertheless, to keep our notation clean and consist, the dependences of physical quantities on the momentum quantum number  $q$  and the parameter  $g$  have been suppressed. In this way, a different choice of  $q$  and  $g$  generates a separate set of  $\{\mathcal{G}(i\omega_n), \mathcal{A}(\Omega)\}$ , which contributes one sample to the training dataset. For the testing dataset, the imaginary-time Green's function  $\mathcal{G}(i\omega_n)$  is computed using a Worm-type quantum Monte Carlo method (see appendix B for details) [32], where noise with unknown strength is intrinsically embedded. As labels,

their corresponding  $\mathcal{A}(\Omega)$  are still obtained by the exact mapping to free fermions under the same parameters  $q$  and  $g$ .

Figure 2(b) shows  $\mathcal{L}$  as a function of  $\eta_{\text{train}}$ . In this case, it also shows a clear minimum around  $\eta_{\text{train}} \approx 0.01$ . This result means that the most proper noise strength also exists, even though the intrinsic numerical noise in the testing dataset may not be Gaussian type. With the optimum noise strength, the trained NN outputs the best prediction for the testing dataset produced by the quantum Monte Carlo simulations. One example is shown in figure 3(b). Compared with the synthetic spectral function shown in figure 3(a), the spectral function of the transverse field Ising model is much more singular and contains sharp edges. Moreover, a dark continuum [33], a finite region with vanishing  $\mathcal{A}(\Omega)$ , exists between the two peaks. With the optimal noise strength, the trained NN correctly resolves both the positions and height of the two peaks, the locations of the sharp edges and the dark continuum. In comparison, the MaxEnt method fails to capture these features of the spectral function. The NN trained without noise also produces two peaks, but the fluctuations are much stronger, especially in the dark continuum.

## 4. Summary and outlook

In summary, we have developed a NN-based method for analytic continuation. The training data is generated either from synthetic Gaussian type spectral functions or by the exact solution of the transverse field Ising model, where the analytic continuation can be done analytically. The validity of this method is demonstrated with data obtained from quantum Monte Carlo simulations. The main finding of this work is that a proper amount of noise has to be added to the training data in order to reach optimal performance. Our NN-based method can be easily applied to perform analytic continuation of Monte Carlo results from other models, such as the Bose–Hubbard model and the anisotropic Heisenberg models, and compare the analytic continuation results with experimental measurements of real-time dynamics, such as quantum simulation experiments with ultracold atomic gases.

## Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://github.com/jy19Phy/Noise-Enhanced-Neural-Networks-for-Analytic-Continuation>.

## Acknowledgments

We thank Lei Wang for helpful discussions. This work is supported by NSFC Grant No. 11904190 (J Y), 11734010 (H Z), and Beijing Outstanding Young Scientist Program (H Z).

## Appendix A. Exact solution for the transverse Ising model

For the transverse Ising model, it is possible to obtain the analytic expressions of both the imaginary-time Green's function and the spectral function. The first step is to rotate the spins along the  $y$ -axis for negative 90 degrees which defines the following new spin operators  $\hat{\pi}_l$ ,

$$\hat{\pi}_l^x = \hat{\sigma}_l^z, \quad \hat{\pi}_l^y = \hat{\sigma}_l^y, \quad \hat{\pi}_l^z = -\hat{\sigma}_l^x. \quad (\text{A1})$$

The next step is to perform the standard Jordan–Wigner transformation that maps spin operators to fermion operators  $\hat{a}_l^\dagger$  and  $\hat{a}_l$

$$\begin{cases} \hat{\pi}_l^+ = (\hat{\pi}_l^x + i\hat{\pi}_l^y)/2 = (-1)^{\sum_{m<l} \hat{n}_m} \hat{a}_l^\dagger, \\ \hat{\pi}_l^- = (\hat{\pi}_l^x - i\hat{\pi}_l^y)/2 = (-1)^{\sum_{m<l} \hat{n}_m} \hat{a}_l, \\ \hat{\pi}_l^z = 2\hat{a}_l^\dagger \hat{a}_l - 1, \end{cases} \quad (\text{A2})$$

where  $\hat{n}_m = \hat{a}_m^\dagger \hat{a}_m$  is the fermion occupation operator at site  $m$ . In terms of these fermionic operators, the resulting Hamiltonian of equation (7) is quadratic,

$$\hat{H} = -J \sum_l \left( \hat{a}_l^\dagger - \hat{a}_l \right) \left( \hat{a}_{l+1} + \hat{a}_{l+1}^\dagger \right) + g \sum_l (2\hat{n}_l - 1). \quad (\text{A3})$$

The boundary condition becomes  $\hat{a}_{L+1}^\dagger = -\hat{\eta}\hat{a}_1^\dagger$  where  $L$  is the number of sites and  $\hat{\eta} = (-1)^{\sum_l \hat{n}_l}$  is the fermion parity operator. After a phase rotation of the operator  $\hat{a}_l$  by  $\pi/4$ , in momentum space the Hamiltonian assumes the form

$$\hat{H} = \sum_k' \begin{bmatrix} \hat{a}_k^\dagger & \hat{a}_{-k} \end{bmatrix} \begin{bmatrix} 2g - 2J\cos(kd) & 2J\sin(kd) \\ 2J\sin(kd) & 2J\cos(kd) - 2g \end{bmatrix} \begin{bmatrix} \hat{a}_k \\ \hat{a}_{-k}^\dagger \end{bmatrix} \tag{A4}$$

where the prime means summation over different pairs of  $k$  and  $-k$  except for  $k = 0$ ,  $d$  is the lattice spacing, and  $\hat{a}_k = e^{i\frac{\pi}{4}} \frac{1}{\sqrt{L}} \sum_{l=1}^L e^{ilkd} \hat{a}_l$  with  $L$  being the lattice size. One can then easily diagonalize the Hamiltonian by the standard Bogoliubov transformation, resulting in a free fermion Hamiltonian

$$\hat{H} = \sum_k' \xi_k \left( \hat{f}_k^\dagger \hat{f}_k + \hat{f}_{-k}^\dagger \hat{f}_{-k} \right), \tag{A5}$$

where  $\hat{f}_k^\dagger$  ( $\hat{f}_k$ ) is the new fermion creation (annihilation) operator of momentum mode  $k$  and  $\xi_k = 2\sqrt{J^2 + g^2 - 2gJ\cos(kd)}$  is the corresponding quasi-particle energy. The explicit expression of the Bogoliubov transformation reads

$$\begin{bmatrix} \hat{a}_k \\ \hat{a}_{-k}^\dagger \end{bmatrix} = \begin{bmatrix} \cos\theta_k & -\sin\theta_k \\ \sin\theta_k & \cos\theta_k \end{bmatrix} \begin{bmatrix} \hat{f}_k \\ \hat{f}_{-k}^\dagger \end{bmatrix} \tag{A6}$$

where  $\theta_k$  is determined by  $\tan 2\theta_k = J\sin(kd)/[g - J\cos(kd)]$ .

We choose to study the imaginary-time spin-spin correlation function along the  $x$ -direction,

$$\mathcal{G}(l, \tau) = -\frac{1}{4} \langle \mathcal{T}_\tau \hat{\sigma}_l^x(\tau) \hat{\sigma}_0^x(0) \rangle, \tag{A7}$$

where  $\hat{\sigma}_l^x(\tau) = e^{\tau\hat{H}} \hat{\sigma}_l^x e^{-\tau\hat{H}}$  is the imaginary-time Heisenberg operator and  $\mathcal{T}$  denotes the time-ordering operator. This correlator corresponds to the density-density correlation function in terms of the Jordan–Wigner fermion  $\hat{a}_l$ . Since the Hamiltonian equation (A3) or (A5) is quadratic, the correlator can be evaluated using Wick’s theorem. After a tedious but straightforward calculation, the momentum and frequency representation of the correlator  $\mathcal{G}(q, i\omega_n)$ ,

$$\mathcal{G}(q, i\omega_n) = \frac{1}{L} \sum_l e^{-iqld} \int_0^\beta \mathcal{G}(l, \tau) e^{i\omega_n\tau} d\tau, \tag{A8}$$

which takes the following explicit form

$$\begin{aligned} \mathcal{G}(q \neq 0, i\omega_n) &= \sum_k (f_{k+q} + f_k - 1) \sin(\theta_{k+q} + \theta_k) \\ &\times \frac{i\omega_n \sin(\theta_k - \theta_{k+q}) + (\xi_{k+q} + \xi_k) \sin(\theta_k + \theta_{k+q})}{\omega_n^2 + (\xi_{k+q} + \xi_k)^2} \\ &+ \sum_k (f_{k+q} - f_k) \cos(\theta_{k+q} + \theta_k) \\ &\times \frac{i\omega_n \cos(\theta_k - \theta_{k+q}) + (\xi_{k+q} - \xi_k) \cos(\theta_k + \theta_{k+q})}{\omega_n^2 + (\xi_{k+q} - \xi_k)^2}. \end{aligned} \tag{A9}$$

Here  $f_k$  is the Fermi–Dirac distribution function  $f_k = 1/(e^{\beta\xi_k} + 1)$  and  $\omega_n = 2\pi n/\beta$  is the even Matsubara frequency where  $n = 1, \dots, N_{in}$ . With the analytical formula of  $\mathcal{G}(q \neq 0, i\omega_n)$ , we can then analytically perform the analytic continuation and read out the spectral function,

$$\begin{aligned} \mathcal{A}(q \neq 0, \Omega) &= -\frac{1}{\pi} \text{Im} \mathcal{G}(q \neq 0, i\omega_n \rightarrow \Omega + i0^+) \\ &= \sum_k (1 - f_{k+q} - f_k) \\ &\times \left[ \sin(\theta_{k+q} + \theta_k) \cos\theta_{k+q} \sin\theta_k \delta(\Omega - \xi_k - \xi_{k+q}) \right. \\ &\quad \left. - \sin(\theta_{k+q} + \theta_k) \sin\theta_{k+q} \cos\theta_k \delta(\Omega + \xi_k + \xi_{k+q}) \right] + \sum_k (f_k - f_{k+q}) \\ &\times \left[ \cos(\theta_{k+q} + \theta_k) \cos\theta_{k+q} \cos\theta_k \delta(\Omega + \xi_k - \xi_{k+q}) \right. \end{aligned}$$

$$+ \cos(\theta_{k+q} + \theta_k) \sin \theta_{k+q} \sin \theta_k \delta(\Omega - \xi_k + \xi_{k+q})]. \quad (\text{A10})$$

In conclusion, both the Green's function or the correlator  $\mathcal{G}(q, i\omega_n)$  and the spectral function  $\mathcal{A}(q, \Omega)$  can be evaluated exactly according to equations (A9) and (A10) for each  $g$ , with  $J$  being the energy unit. For preparation of the training dataset, setting temperature  $\beta = 2.0/J$  and interaction strength  $g$  ranging from  $0.5J$  to  $2.0J$  with interval  $0.1J$ . For each parameter  $g$ , the momentum  $q$  will be discretized into  $(0, 0.01, 2\pi]$  and setting  $d$  as the unit length. Here in this transverse problem,  $\mathcal{A}$  does not obey the normalization relation  $\int d\Omega \mathcal{A}(\Omega) = 1$ . Thus the mean-square error is adopted for the training process.

## Appendix B. Quantum Monte Carlo simulation of the transverse Ising model

We closely follow reference [32] to map spin operators into hard-core boson operators such that the resulting partition function can be efficiently sampled by the Worm algorithm [34, 35]. We first perform a spin rotation around the  $y$ -axis for 90 degrees and the spin rotational operators  $\hat{\pi}_l$  is defined as

$$\hat{\pi}_l^x = -\hat{\sigma}_l^z, \quad \hat{\pi}_l^y = \hat{\sigma}_l^y, \quad \hat{\pi}_l^z = \hat{\sigma}_l^x. \quad (\text{B1})$$

For the negative rotation, the corresponding boson Hamiltonian in this time can be written as a sum of the kinetic term  $\hat{K}$  and the potential energy term  $\hat{U}$ ,

$$\hat{H} = \hat{K} + \hat{U}, \quad (\text{B2})$$

where

$$\hat{K} \equiv \hat{K}_1 + \hat{K}_2 \quad (\text{B3})$$

$$\equiv -J \sum_l (\hat{a}_l^\dagger \hat{a}_{l+1} + \text{h.c.}) - J \sum_l (\hat{a}_l^\dagger \hat{a}_{l+1}^\dagger + \text{h.c.}) \quad (\text{B4})$$

and

$$\hat{U} = -2g \sum_l \hat{a}_l^\dagger \hat{a}_l. \quad (\text{B5})$$

Here we have omitted unimportant constants. Employing the expansion

$$\begin{aligned} e^{-\beta \hat{H}} &= e^{-\beta \hat{U}} \mathcal{T}_\tau \exp \left[ - \int_0^\beta \hat{K}(\tau) d\tau \right] \\ &= e^{-\beta \hat{U}} \sum_{n=0}^{\infty} (-1)^n \int_0^\beta d\tau_n \cdots \int_0^{\tau_n} d\tau_1 (\hat{K}_1(\tau_n) + \hat{K}_2(\tau_n)) \cdots (\hat{K}_1(\tau_1) + \hat{K}_2(\tau_1)), \end{aligned}$$

where  $\mathcal{T}_\tau$  is the imaginary time ordering operator and  $\hat{K}(\tau) = e^{\tau \hat{U}} \hat{K} e^{-\tau \hat{U}}$ . In the Fock basis a typical term (out of  $2^n$  terms) in the  $n$ th order expansion of the partition function reads

$$J^n \left( \prod_{j=1}^n \int d\tau_j \right) \cdot \exp \left[ - \int_0^\beta U(\tau) d\tau \right], \quad (\text{B6})$$

since the matrix element of the both hopping and pair creation/annihilation is one. Here  $U(\tau)$  is the potential energy of the hard-core bosons at imaginary time  $\tau$ . As is clear from equation (B6), the weight of the configuration is positive and can be sampled by the Monte Carlo method. Following the detailed update procedures in [32], the partition function can be efficiently sampled by Worm-type algorithms. The imaginary-time Green's function  $\mathcal{G}(l-l', \tau)$  defined in equation (A7) just corresponds to  $-\langle \mathcal{T}_\tau \hat{n}_l(\tau) \hat{n}_{l'}(0) \rangle$  asides from unimportant constants. Since  $\hat{n}_l(\tau)$  is diagonal in the Fock basis, the statistics can be easily accumulated in Monte Carlo simulations. We perform the simulation with system size  $L = 100$  at various values of transverse field  $g$  to prepare the training and testing datasets.

## ORCID iD

Juan Yao  <https://orcid.org/0000-0001-7450-3987>



## References

- [1] Gubernatis J, Kawashima N and Werner P 2016 *Quantum Monte Carlo Methods: Algorithms for Lattice Models* (Cambridge: Cambridge University Press)
- [2] Pavarini E, Koch E, Anders F and Jarrell M 2012 *Correlated Electrons: From Models to Materials Modeling and Simulation* vol 2 (Jülich: Verlag des Forschungszentrum)
- [3] Suzuki M 1993 *Quantum Monte Carlo Methods in Condensed Matter Physics* (Singapore: World Scientific)
- [4] Gull E, Millis A J, Lichtenstein A I, Rubtsov A N, Troyer M and Werner P 2011 Continuous-time Monte Carlo methods for quantum impurity models *Rev. Mod. Phys.* **83** 349
- [5] Bertaina G, Galli D E and Vitali E 2017 Statistical and computational intelligence approach to analytic continuation in quantum Monte Carlo *Adv. Phys. X* **2** 302
- [6] Mahan G D 2000 *Many-Particle Physics* (Boston, MA: Springer US)
- [7] Altland A and Simons B 2010 *Condensed Matter Field Theory* 2nd edn (Cambridge: Cambridge University Press)
- [8] Ramm A G 2004 *Inverse Problems: Mathematical and Analytical Techniques with Applications to Engineering* (Berlin: Springer)
- [9] Kabanikhin S I 2008 Definitions and examples of inverse and ill-posed problems *J. Inverse Ill-Posed Problems* **16** 317
- [10] Kabanikhin S I 2011 *Inverse and Ill-Posed Problems: Theory and Applications* (Berlin: De Gruyter)
- [11] Jarrell M and Gubernatis J E 1996 Bayesian inference and the analytic continuation of imaginary-time quantum Monte Carlo data *Phys. Rep.* **269** 133
- [12] Bergeron D and Tremblay A-M S 2016 Algorithms for optimized maximum entropy and diagnostic tools for analytic continuation *Phys. Rev. E* **94** 023303
- [13] Sim J-H and Han M J 2018 Maximum quantum entropy method *Phys. Rev. B* **98** 205102
- [14] Kraberger G J, Triebel R, Zingl M and Aichhorn M 2017 Maximum entropy formalism for the analytic continuation of matrix-valued Green's functions *Phys. Rev. B* **96** 155128
- [15] Silver R N, Sivia D S and Gubernatis J E 1990 Maximum-entropy method for analytic continuation of quantum Monte Carlo data *Phys. Rev. B* **41** 2380
- [16] Gunnarsson O, Haverkort M W and Sangiovanni G 2010 Analytical continuation of imaginary axis data using maximum entropy *Phys. Rev. B* **81** 155107
- [17] Reymbaut A, Bergeron D and Tremblay A-M S 2015 Maximum entropy analytic continuation for spectral functions with nonpositive spectral weight *Phys. Rev. B* **92** 060509(R)
- [18] Rumetshofer M, Bauernfeind D and von der Linden W 2019 Bayesian parametric analytic continuation of Green's functions *Phys. Rev. B* **100** 075137
- [19] Sandvik A W 1998 Stochastic method for analytic continuation of quantum Monte Carlo data *Phys. Rev. B* **57** 10287
- [20] Fuchs S, Pruschke T and Jarrell M 2010 Analytic continuation of quantum Monte Carlo data by stochastic analytical inference *Phys. Rev. E* **81** 056701
- [21] Sandvik A W 2016 Constrained sampling method for analytic continuation *Phys. Rev. E* **94** 063308
- [22] Mishchenko A S, Prokof'ev N, Sakamoto A and Svistunov B 2000 Diagrammatic quantum Monte Carlo study of the Fröhlich polaron *Phys. Rev. B* **62** 6317
- [23] Gouliko O, Mishchenko A S, Pollet L, Prokof'ev N and Svistunov B 2017 Numerical analytic continuation: answers to well-posed questions *Phys. Rev. B* **95** 014102
- [24] Ghanem K and Koch E 2020 Average spectrum method for analytic continuation: efficient blocked-mode sampling and dependence on the discretization grid *Phys. Rev. B* **101** 085111
- [25] Ghanem K and Koch E 2020 Extending the average spectrum method: grid point sampling and density averaging *Phys. Rev. B* **102** 035114
- [26] Witzak-Krempa W, Sørensen E S and Sachdev S 2014 The dynamics of quantum criticality revealed by quantum Monte Carlo and holography *Nat. Phys.* **10** 361
- [27] Arsenaault L-F, Neuberger R, Hannah L A and Millis A J 2017 Projected regression method for solving Fredholm integral equations arising in the analytic continuation problem of quantum physics *Inverse Problems* **33** 115007
- [28] Yoon H, Sim J-H and Han M J 2018 Analytic continuation via domain knowledge free machine learning *Phys. Rev. B* **98** 245101
- [29] Fournier R, Wang L, Yazyev O V and Wu Q 2020 Artificial neural network approach to the analytic continuation problem *Phys. Rev. Lett.* **124** 056401
- [30] Huang D and Yang Y 2021 Learned optimizers for analytic continuation ( arXiv:2107.13265)
- [31] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (Cambridge, MA: MIT Press)
- [32] Huang C-J, Liu L, Jiang Y and Deng Y 2020 Worm-algorithm-type simulation of the quantum transverse-field Ising model *Phys. Rev. B* **102** 094101
- [33] Gouliko O, Mishchenko A S, Prokof'ev N and Svistunov B 2016 Dark continuum in the spectral function of the resonant Fermi polaron *Phys. Rev. A* **94** 051605
- [34] Prokof'ev N, Svistunov B and Tupitsyn I S 1998 Worm algorithm in quantum Monte Carlo simulations *Phys. Lett. A* **238** 253
- [35] Prokof'ev N, Svistunov B and Tupitsyn I S 1998 Exact, complete and universal continuous-time worldline Monte Carlo approach to the statistics of discrete quantum systems *J. Exp. Theor. Phys.* **87** 310