Analytic continuation via 'domain-knowledge free' machine learning

Hongkee Yoon[1], Jae-Hoon Sim[1], Myung Joon Han[1,*]

[1]*Department of Physics, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea*

We present a machine-learning approach to a long-standing issue in quantum many-body physics, namely, analytic continuation. This notorious ill-conditioned problem of obtaining spectral function from imaginary time Green's function has been a focus of new method developments for past decades. Here we demonstrate the usefulness of modern machine-learning techniques including convolutional neural networks and the variants of stochastic gradient descent optimiser. Machine-learning continuation kernel is successfully realized without any 'domain knowledge' and the outstanding performance is achieved for both insulating and metallic band structure. Our machine-learning-based approach not only provides the more accurate spectrum than the conventional methods in terms of peak positions and heights, but is also more robust against the noise which is the required key feature for any continuation technique to be successful. Furthermore, its computation speed is 10^4–10^5 times faster than maximum entropy method.

1 Introduction

Matsubara Green's function method is a useful theoretical tool for quantum many-body problems. While the calculation often becomes much more tractable in the imaginary time (or equivalently, frequency) domain, working with Matsubara function inevitably introduces other theoretical difficulties. One of the most typical cases happens when one tries to obtain a spectral function (or any other measurable quantity), which is defined in real space, from imaginary Green's function. This procedure is known as 'analytic continuation' and poses a notorious ill-conditioned inverse problem. The severe noise sensitivity significantly undermines the predictability and the usefulness of theoretical methods such as quantum Monte Carlo (QMC). Many different approaches have been suggested to solve this problem including Pade approximation [1,2], maximum entropy (MEM) [3–6], and stochastic method [7]. All these methods are based on the physical knowledge or utilize the pre-understanding of the problem which are expressed in their own assumptions and fitting parameters. In other words, all these methods heavily rely on 'domain knowledge'.

Machine-learning (ML) approach is based on a different philosophy. The ML procedure is to develop a machinery which can self-learn the governing rule(s) of a given system or problem through the massive dataset 'training' [8,9]. Due to the remarkable progress in both hardware and software engineering, ML technique overwhelms the state-of-the-art human-designed algorithms in many different areas [10,11]. Recently, it becomes more and more popular in physics research. ML proves its capability in many different fields ranging from materials science [12,13] and statistical physics [14,15] to quantum many-body problems [16–18] and quantum informations [19,20].

In this paper, we apply modern ML techniques [21] to the long-standing physics problem of analytic continuation. By using convolutional neural network (CNN) [11,22] and stochastic gradient descent based optimizer (*i.e.*, stochastic gradient descent, *Adadelta*, *Adagrad*) [23–25], we successfully construct the ML kernel which can generate the real space spectral function from imaginary Green's function. We emphasize that our method does not require any 'domain-knowledge' [26]. In comparison to the conventional techniques, ML-based algorithm demonstrates its superiority in terms of accuracy and computation speed. The spectral weights and peak positions are in better agreement, and the computation speed is 10^4–10^5 faster. Further, ML-based method is more robust against the noise which is inevitably introduced in Monte-Carlo calculation for example. Our results show that the domain-knowledge free ML approach can be a new promising way to solve the long-standing physics problem that has not been well understood based on the currently available techniques.

2 Result and Discussion

Description of the problem and machine learning Matsubara frequency Green's function $G(i\omega_n)$ is analytically continued to the real frequency $G(\omega)$. For a given $G(i\omega_n)$, the spectral function is $A(\omega) = -\frac{1}{\pi}\,\mathrm{Im}\,G(\omega + i0^+)$. Note that calculating the Green's function for a given spectral function is straightforward, not ill-conditioned. On the other hand, the spectral function is obtained by

inverting the integral equation

$$G(i\omega_n) = \int d\omega \frac{A(\omega)}{i\omega_n - \omega} \tag{1}$$

$$= \int d\omega \mathbf{K}(i\omega_n, \omega) A(\omega) \tag{2}$$

where the kernel $\mathbf{K}(i\omega_n, \omega)$ has different forms for different problems. This continuation process is an ill-posed problem, and the direct inverse $A = \mathbf{K}^{-1}G$ is hardly feasible due to the high condition number. The key question is how to deal with intrinsic noises.

Here we note that many techniques to handle this kind of ill-posed problems have been actively developed in the ML field of research for more than last two decades [27,28]. Crucially required are the efficient data representation in high-dimensional space and the practical algorithm to optimize massive variables in deep neural networks. In spite of the challenging features of the problems, notable achievements can be found in many different areas [29].

In the current study, we adopted 'fully connected layer (FCL)' and CNN [11,17,22,29], and try to perform analytic continuation within high-dimensional space. In order to systematically check the input noise dependence, we considered several different sets of random noise inputs and examined the output spectra. The uniformly distributed random noise $N(iw_n) \in [-\lambda, \lambda]$ are used for our main presentation and the noise strength λ is varied from 0 to 0.5. The noised input is then defined as $G(i\omega_n)^{\rm in}_{\rm noise} = G(i\omega_n) + \lambda(N(i\omega_n))$. We also considered the other types of noise character. In particular, the frequency-dependent $\lambda(i\omega_n)$ has been carefully checked since it is often the case of QMC-DMFT (dynamical mean-field theory) calculations. Gaussian-type noise is also tested. While we mainly present the uniform random noise, any part of our conclusion is not changed by

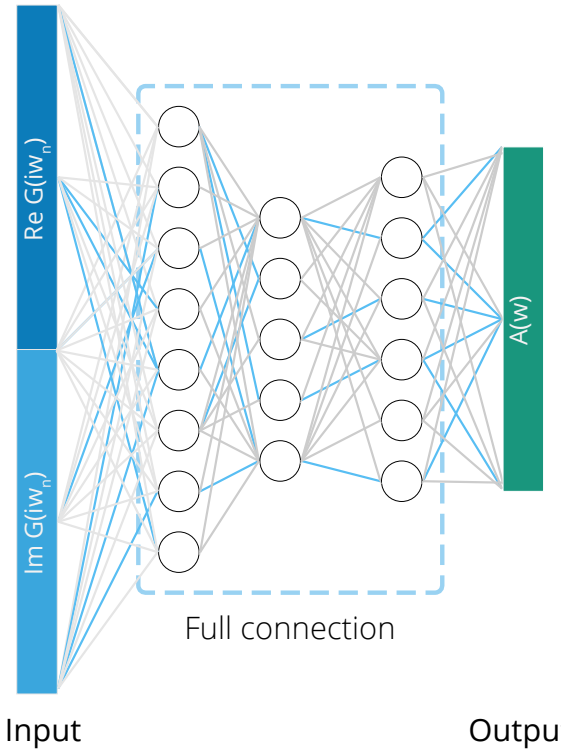

Figure 1: **FCL neural network.** An illustration of our neural network architecture which consists only of FCLs. The input $G(i\omega_n)$ is an array of complex numbers. $\mathrm{Re}[G(i\omega_n)]$ and $\mathrm{Im}[G(i\omega_n)]$ are arranged as a 1D array to be inserted into the neural-net input. The dropout layers are located in between all the FCLs to improve the overfitting of neural networks (not shown). The green box represents the output layer of $A(\omega)$. The blue neural network lines are the schematic representation of activated connections.

this choice of noise type.

Fully connected layers As the first step toward ML-based analytic continuation, we consider the neural network composed of FCLs which may be regarded as an early-stage ML approach [30].

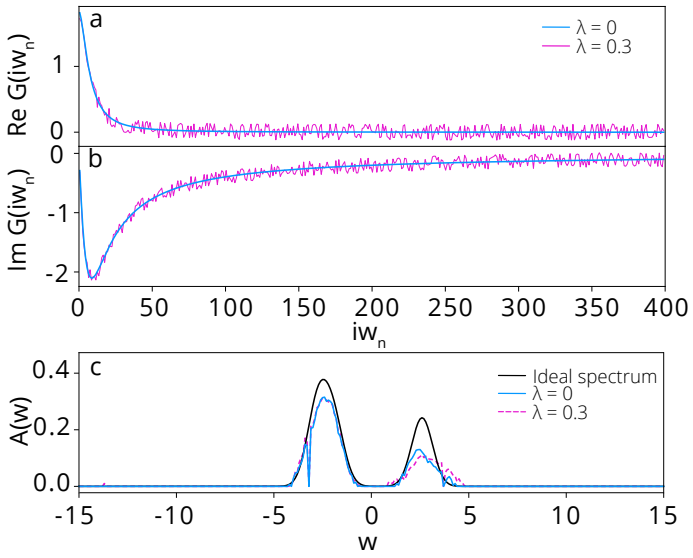

Figure 2: **Analytic continuation result of FCL.** The input Green's functions and the output spectra calculated by the FCL neural network kernel (without CNN layer). (a, b) $\mathrm{Re}[G(i\omega_n)]$ and $\mathrm{Im}[G(i\omega_n)]$ are presented in (a) and (b), respectively. The blue curves are generated from the ideal spectrum shown by the black line in (c). The magenta lines show the noised input $G(i\omega_n)^{\mathrm{in}}_{\mathrm{noise}}$ with $\lambda = 0.3$ (see the main text for more details). (c) The calculated spectral functions are presented by blue-solid (λ=0; noise-free) and magenta-dashed curve (λ=0.3) in comparison to the ideal spectrum (black-solid line).

Roughly, the use of single FCL can be regarded as one multiplication process of an inversion matrix to the input Green's function [31]. Having more FCLs thus corresponds to the increased number of matrix multiplications to represent the inversion. Practically it is not expected to achieve a notable improvement just by increasing the number of hidden layers [27,32]. After testing many different numbers of hidden layer sets, we indeed found that the performance is not much enhanced. Thus, in the below, we focus on the results of three layers (Fig.1).

Figure 2 presents the result of analytic continuation by using FCLs neural network. The black line in Fig. 2(c) is the spectrum from which imaginary Green's functions of Fig. 2(a) and (b) (blue lines) are generated. Therefore, if the continuation procedure is perfect, the continued spectrum should be identical with the black line in Fig. 2(c). Note that the process of obtaining $G(i\omega_n)$ from $A(\omega)$ is not ill-conditioned. Once $G(i\omega_n)$ is calculated, one can perform the continuation and compare the result of $A(\omega)$ with the original one, namely the ideal spectrum.

The FCL continuation results are shown in Fig. 2(c). The blue-solid and magenta-dashed line corresponds to λ=0 and λ=0.3, respectively. It is clearly noted that the continued spectra are not smooth and significantly deformed in comparison to the ideal black line. This result demonstrates the challenging nature of the problem. At the same time, however, we also note that the overall shape of spectrum is captured by our FCL neural network although the unexpected wriggles are found, and they become worse as the noise level increases. We emphasize that this level of performance is hardly achievable through the direct matrix inversion of Eq. (2). This promising aspect is largely attributed to the 'dropout' and the regularization procedure which prevent overfittings [33]. In this regard, while not satisfactory at all, our FCL result shows a possibility of neural network approach for the analytic continuation.

Convolutional neural network Many techniques have been suggested to overcome the deficiency of FCL. One key idea is to identify the essential features of a problem and to reconstruct them in a higher dimensional space [28]. Principal component analysis (PCA) [34] is an example which proved to be powerful for data compression and dimensionality reduction. Unfortunately, however, PCA

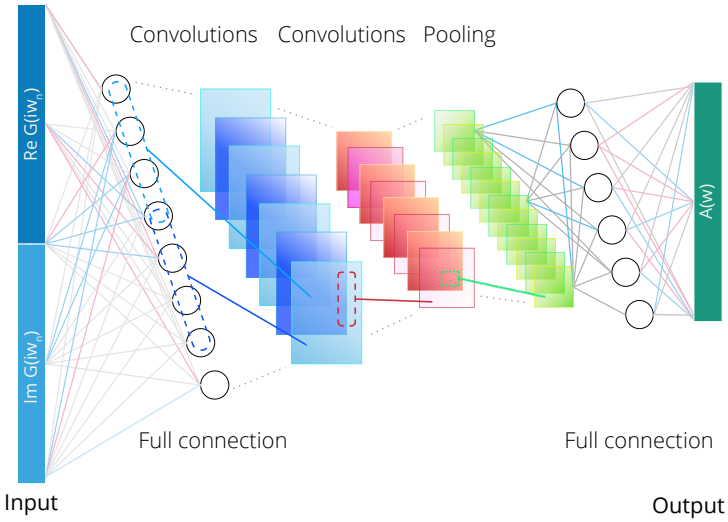

Figure 3: **CNN neural network.** An illustration of the neural network architecture which consists of both FCLs and CNN. Two convolution layers (blue and red squares) and one pooling layer (green squares) are placed in between the FCLs. Colored neural network lines are the schematic representations of activated connections.

can only be used in rank 1 (vector) and rank 2 (matrix) for most of the cases. While some techniques for tensor PCA have been proposed, they seem to need further developments [35,36]. A typical fundamental limitation of PCA is that each principal component is given by a linear combination of original variables whereas non-linearity is essential for ill-posed problems [37]. In this regard, CNN is a useful advanced technique leading the modern machine-learning era [11,29,38–41]. The performance of CNN image processing surpasses the human-designed algorithms based on 'domain knowledge' [29,38]. Due to its outstanding feature selection in tensor space, CNN is widely adopted by high-dimensional noise filters for auto-encoder and sound/video data [42].

In analytic continuation, input/output data are represented by a certain set of numbers. Thus

it can be regarded as an inverse problem that has to be performed within a dimension corresponding to those numbers. With this observation, we applied CNN technique to the long-standing ill-posed problem of analytic continuation. Figure 3 shows our neural network structure. We aim to create a minimal model with the smallest possible number of layers. Thus our neural network is designed to contain CNN layers in between two FCLs since we learned in the above that three FCLs could capture the basic features of spectra. While it is conventional to have CNN layers just next to the input layer in the image processing (*e.g.*, AlexNet [29], VGG [39], GoogleNet [40], and ResNet [41]), we take a different strategy of inserting the CNN layer after the matrix operation through FCL. It is because the full information of input Green's function needs to be utilized in our problem. The total number of parameters in our neural network is $\sim$600,000 and $\sim$500,000 for including and excluding CNN, respectively. It is noted that the network size is not much increased by having CNN layers. We have adopted modern optimization algorithm, namely '*Adadelta*' [23], to optimize this large number of neural networks parameters.

Figure 4 shows the continuation result of using CNN. The model spectrum (black line in Fig. 4(c)) is designed to mimic a Mott-Hubbard insulator consisted of two distinct Hubbard bands with different peak heights. The outstanding performance of CNN can clearly be seen from that the continuation results are significantly improved in comparison to the FCL-only data in Fig. 2. The overall shape, peak positions and relative peak heights are well reproduced without any undesirable wriggle. Importantly, the reproducibility remains quite robust against the noise even if the deviation from the ideal spectrum (black) becomes noticeable as the noise level increases (from light-blue-solid lines to dark-blue-dashed and to red-dashed).

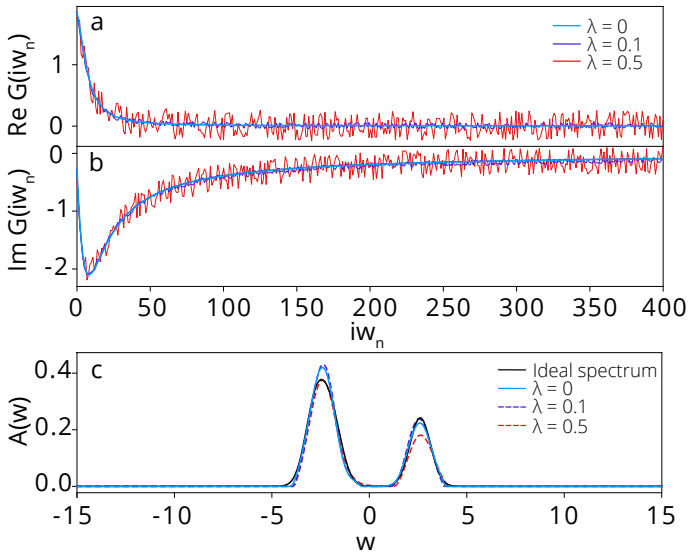

Figure 4: **Analytic continuation result of CNN.** (a, b) $\mathrm{Re}[G(i\omega_n)]$ and $\mathrm{Im}[G(i\omega_n)]$ are presented in (a) and (b), respectively. The light-blue curves are generated from the ideal spectrum shown by the black line in (c). Three different noise levels are presented in light-blue (λ=0), dark-blue (λ=0.1) and red lines (λ=0.5). (c) Analytic continuation result of CNN-ML kernel for λ=0 (noise-free; light-blue-solid), λ=0.1 (dark-blue-dashed), and λ=0.5 (red-dashed) along with the ideal spectrum (black-solid line).

The robustness against the input noise is a crucially required feature for the reliable analytic continuation since the noise is unavoidably present in stochastic approaches. As shown in Fig. 4(c), the overall features and the detailed shapes of the spectrum are well maintained even for the case of significant noise levels. This result shows the powerfulness of ML-based analytic continuation kernel.

The performance of ML kernel is further demonstrated by the comparison to the conventional

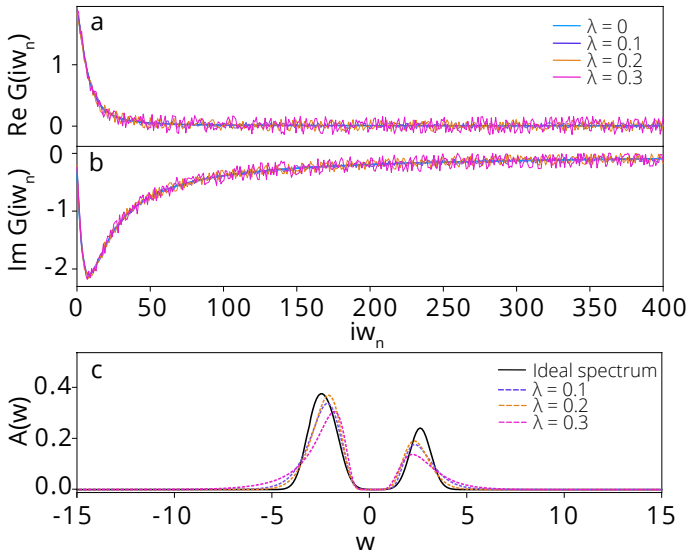

Figure 5: **Analytic continuation result of MEM.** (a, b) Input Green's function of $\mathrm{Re}[G(i\omega_n)]$ and $\mathrm{Im}[G(i\omega_n)]$ are presented in (a) and (b), respectively. The light blue curves are generated from the ideal spectrum shown by the black line in (c). Four different noise levels are presented in light-blue (λ=0), dark-blue (λ=0.1), yellow (λ=0.2), and magenta lines (λ=0.3). (c) Analytic continuation result of conventional MEM for λ=0.1 (dark-blue-dashed), λ=0.2 (yellow-dashed), and λ=0.3 (magenta-dashed) along with the ideal spectrum (black-solid line).

continuation techniques. Fig. 5 shows the result of MEM which is one of the most widely-used methods for analytic continuation [3,5,6]. It is clearly noted that, even at a significantly lower noise level, the MEM result is markedly deviated from the ideal spectrum in terms of peak position and height. It is in a sharp contrast to the ML-based result of Fig. 4 in which the spectrum is well preserved even at $\lambda = 0.5$.

Figure 6 shows the result of ML kernel for metallic spectrum having coherent as well as

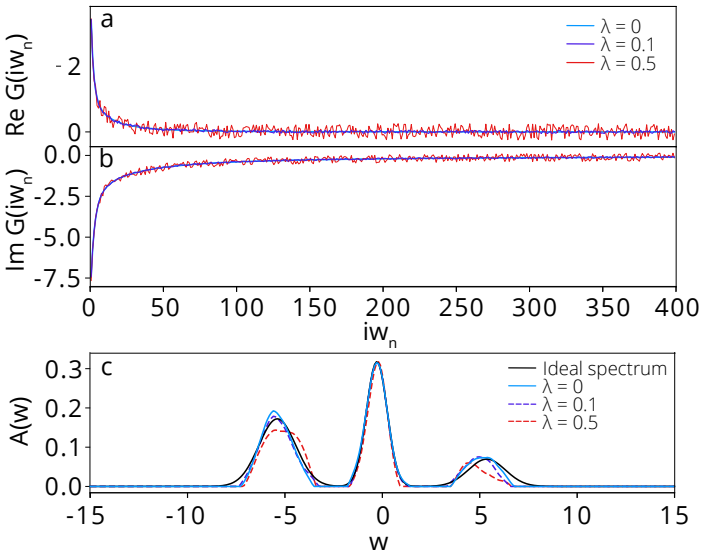

Figure 6: **Analytic continuation result of CNN for metallic spectrum.** (a, b) Input Green's function of $\mathrm{Re}[G(i\omega_n)]$ and $\mathrm{Im}[G(i\omega_n)]$ are presented in (a) and (b), respectively. The light-blue curves are generated from the ideal spectrum shown by the black line in (c). Three different noise levels are presented in light-blue (λ=0), dark-blue (λ=0.1) and red lines (λ=0.5). (c) Analytic continuation result of CNN-ML kernel for λ=0 (noise-free; light-blue-solid), λ=0.1 (dark-blue-dashed), and λ=0.5 (red-dashed) along with the ideal spectrum (black-solid line).

incoherent peaks. Once again, our machine-learning kernel well reproduces the original spectrum. The robustness against noise is also excellent as in the insulating case. In particular, the coherent peak is considerably well reproduced while the incoherent states are moderately affected by the noises. It is a good indication for predicting the phase from a given Green's function.

Finally, we emphasize the efficiency of ML-based analytic continuation. Once the ML kernel is well trained, the continuation process can be performed at a speed of $\sim$10,000 Green's

functions/sec, which is at least 10^4–10^5 faster than the conventional MEM.

3 Discussion

Modern ML technique proves its usefulness for a long-standing physics problem of analytic continuation. Its superiority over the conventional technique is demonstrated in the accuracy, speed, and the robustness against noise. For both insulating and metallic spectrum, our CNN-based ML kernel gives better agreement with the ideal spectrum in terms of peak position and height. Up to the high level of random noises, at which MEM fails to produce the reliable results, ML technique retains its accuracy. In terms of computation speed, the trained kernel is 10^4–10^5 times faster than the conventional method. Our result suggestes that ML approach can be promising for many of long-standing physics problem in which the conventional 'domain knowledge' does not properly work.

Methods

We constructed the ML-based analytic continuation kernel by using widely-adopted open-source deep-learning framework, namely, '*keras*' [43] with '*tensorflow*' [44] backend. For continuation problem, training process is straightforward since the calculation of $G(i\omega_n)$ from a given $A(\omega)$ is not ill-conditioned. Our training sets consist of $\sim$105,000 different combinations of peak number, heights and positions. Among these various types of spectra, more training sets were used for the case of single-, double-, and triple-peak spectral functions since they are more relevant to the applications. It is straightforward to extend the number of training sets to an arbitrary number. The validation

set consists of 10,000 different types of peaks with different random sequences. The energy range, peak height and peak width varies from -10 to 10, 0.2 to 1 and 0.3 to 1.2, respectively. Neural networks are trained by using '*Adadelta*' [23] optimizer. We found that '*stochastic gradient decent (SGD)*' and even '*RMSprop*' [45] optimizer quite often suffer from the 'gradient vanishing problem'; *i.e.*, all variables of a neural net are quickly set to zero. On the other hand, the recently-developed adaptive stochastic variant optimizers (such as '*Adadelta*', '*Adagrad*' [25], '*Adam*' [24], and '*Adamax*' [24]) produce the reliable results. We eventually chose '*Adadelta*' as it clearly exhibits the best performance. For the activation function, we chose a combination of rectified linear unit (ReLU) and scaled exponential linear unit (SeLU). It is found that $\sim$8000 epochs are mostly enough for neural network training.

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education(2018R1A2B2005204).

Author contribution H. Y. performed ML calculations and J.-H. S. performed MEM under the supervision of M. J. H. All authors contributed to the development of idea and the discussion. H. Y and M. J. H. wrote the manuscript.

Competing interests The authors declare no competing interests.

Correspondence Correspondence and requests for materials should be addressed to M.J.H. (email: mj.han@kaist.ac.kr).

1. Vidberg, H. J. & Serene, J. W. Solving the Eliashberg equations by means ofN-point Padé approximants. *J Low Temp Phys* **29**, 179–192 (1977). 1

2. Gunnarsson, O., Haverkort, M. W. & Sangiovanni, G. Analytical continuation of imaginary axis data for optical conductivity. *Phys. Rev. B* **82**, 165125 (2010). 1

3. Jarrell, M. & Gubernatis, J. E. Bayesian inference and the analytic continuation of imaginary-time quantum Monte Carlo data. *Physics Reports* **269**, 133–195 (1996). 1, 2

4. Haule, K., Yee, C.-H. & Kim, K. Dynamical mean-field theory within the full-potential methods: Electronic structure of ${\text{CeIrIn}}_{5}$, ${\text{CeCoIn}}_{5}$, and ${\text{CeRhIn}}_{5}$. *Phys. Rev. B* **81**, 195107 (2010). 1

5. Gunnarsson, O., Haverkort, M. W. & Sangiovanni, G. Analytical continuation of imaginary axis data using maximum entropy. *Phys. Rev. B* **81** (2010). ArXiv: 1001.4351. 1, 2

6. Bergeron, D. & Tremblay, A.-M. S. Algorithms for optimized maximum entropy and diagnostic tools for analytic continuation. *Phys. Rev. E* **94**, 023303 (2016). 1, 2

7. Sandvik, A. W. Stochastic method for analytic continuation of quantum Monte Carlo data. *Phys. Rev. B* **57**, 10287–10290 (1998). 1

8. Rifai, S., Dauphin, Y. N., Vincent, P., Bengio, Y. & Muller, X. The manifold tangent classifier. In *Advances in Neural Information Processing Systems*, 2294–2302 (2011). 1

9. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (The MIT Press, Cambridge, Massachusetts, 2016). 1

10. Silver, D. *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016). 1

11. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015). 1, 2, 2

12. Faber, F. A., Lindmaa, A., von Lilienfeld, O. A. & Armiento, R. Machine Learning Energies of 2 Million Elpasolite $(AB{C}_{2}{D}_{6})$ Crystals. *Phys. Rev. Lett.* **117**, 135502 (2016). URL `https://link.aps.org/doi/10.1103/PhysRevLett.117.135502`. 1

13. Kolb, B., Lentz, L. C. & Kolpak, A. M. Discovering charge density functionals and structure-property relationships with PROPhet: A general framework for coupling machine learning and first-principles methods. *Scientific Reports* **7**, 1192 (2017). 1

14. Torlai, G. & Melko, R. G. Learning Thermodynamics with Boltzmann Machines. *Physical Review B* **94** (2016). URL `http://arxiv.org/abs/1606.02718`. ArXiv: 1606.02718. 1

15. Carrasquilla, J. & Melko, R. G. Machine learning phases of matter. *Nat Phys* **13**, 431–434 (2017). 1

16. Arsenault, L.-F., Lopez-Bezanilla, A., von Lilienfeld, O. A. & Millis, A. J. Machine learning for many-body physics: The case of the Anderson impurity model. *Phys. Rev. B* **90** (2014). ArXiv: 1408.1143. 1

17. Carleo, G. & Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **355**, 602–606 (2017). 1, 2

18. Ch'ng, K., Carrasquilla, J., Melko, R. G. & Khatami, E. Machine Learning Phases of Strongly Correlated Fermions. *Physical Review X* **7** (2017). URL http://arxiv.org/abs/1609.02552. ArXiv: 1609.02552. 1

19. Cai, X.-D. *et al.* Entanglement-Based Machine Learning on a Quantum Computer. *Phys. Rev. Lett.* **114**, 110504 (2015). URL https://link.aps.org/doi/10.1103/PhysRevLett.114.110504. 1

20. Torlai, G. & Melko, R. G. Neural Decoder for Topological Codes. *Phys. Rev. Lett.* **119**, 030501 (2017). URL https://link.aps.org/doi/10.1103/PhysRevLett.119.030501. 1

21. Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H. & others. Greedy layer-wise training of deep networks. *Advances in neural information processing systems* **19**, 153 (2007). 1

22. Kalchbrenner, N., Grefenstette, E. & Blunsom, P. A Convolutional Neural Network for Modelling Sentences. *arXiv:1404.2188 [cs]* (2014). ArXiv: 1404.2188. 1, 2

23. Zeiler, M. D. ADADELTA: An Adaptive Learning Rate Method. *arXiv:1212.5701 [cs]* (2012). ArXiv: 1212.5701. 1, 2, 3

24. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]* (2014). ArXiv: 1412.6980. 1, 3

25. Duchi, J., Hazan, E. & Singer, Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *Journal of Machine Learning Research* **12**, 2121–2159 (2011). 1, 3

26. Arsenault, L.-F., Neuberg, R., Hannah, L. A. & Millis, A. J. Projected regression method for solving Fredholm integral equations arising in the analytic continuation problem of quantum physics. *Inverse Problems* **33**, 115007 (2017). 1

27. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**, 2278–2324 (1998). 2, 2

28. Pillonetto, G., Dinuzzo, F., Chen, T., De Nicolao, G. & Ljung, L. Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica* **50**, 657–682 (2014). 2, 2

29. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. In Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q. (eds.) *Advances in Neural Information Processing Systems 25*, 1097–1105 (Curran Associates, Inc., 2012). 2, 2, 2

30. Frean, M. The upstart algorithm: A method for constructing and training feedforward neural networks. *Neural computation* **2**, 198–209 (1990). 2

31. Kung, S.-Y. Kernel Approaches to Unsupervised and Supervised Machine Learning. In *Advances in Multimedia Information Processing - PCM 2009*, Lecture Notes in Computer Science, 1–32 (Springer, Berlin, Heidelberg, 2009). 2

32. Dean, J. *et al.* Large scale distributed deep networks. In *Advances in neural information processing systems*, 1223–1231 (2012). 2

33. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* **15**, 1929–1958 (2014). 2

34. Shlens, J. A Tutorial on Principal Component Analysis. *arXiv:1404.1100 [cs, stat]* (2014). ArXiv: 1404.1100. 2

35. Inoue, K. Generalized Tensor PCA and Its Applications to Image Analysis. In *Applied Matrix and Tensor Variate Data Analysis*, SpringerBriefs in Statistics, 51–71 (Springer, Tokyo, 2016). 2

36. Zhang, A. & Xia, D. Tensor SVD: Statistical and Computational Limits. *arXiv:1703.02724 [cs, math, stat]* (2017). ArXiv: 1703.02724. 2

37. Archambeau, C. & Bach, F. R. Sparse probabilistic projections. In Koller, D., Schuurmans, D., Bengio, Y. & Bottou, L. (eds.) *Advances in Neural Information Processing Systems 21*, 73–80 (Curran Associates, Inc., 2009). 2

38. Ciresan, D. C., Meier, U., Gambardella, L. M. & Schmidhuber, J. Convolutional Neural Network Committees for Handwritten Character Classification. In *2011 International Conference on Document Analysis and Recognition*, 1135–1139 (2011). 2

39. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556 [cs]* (2014). ArXiv: 1409.1556. 2

40. Szegedy, C. *et al.* Going Deeper with Convolutions. *arXiv:1409.4842 [cs]* (2014). ArXiv: 1409.4842. 2

41. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. *arXiv:1512.03385 [cs]* (2015). ArXiv: 1512.03385. 2

42. Karpathy, A. *et al.* Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 1725–1732 (2014). 2

43. Chollet, F. *et al.* Keras. `https://github.com/keras-team/keras` (2015). 3

44. Abadi, M. *et al.* TensorFlow: A system for large-scale machine learning (2016). 3

45. Tieleman, T. & Hinton, G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning* **4**, 26–31 (2012). 3