Machine learning Hadron Spectral Functions in Lattice QCD

Shi-Yang Chen,[a,*] **Heng-Tong Ding,**[a] **Fei-Yi Liu,**[a,b] **Gábor Papp**[b] **and Chun-Bin Yang**[a]

[a]*Key Laboratory of Quark and Lepton Physics (MOE) & Institute of Particle Physics,
Central China Normal University, Wuhan 430079, China*

[b]*Institute for Physics, Eötvös Loránd University, 1/A Pázmány P. sétány, H-1117, Budapest, Hungary
E-mail:* shiyang-chen@mails.ccnu.edu.cn, fyliu@mails.ccnu.edu.cn,
hengtong.ding@mail.ccnu.edu.cn, gabor.papp@ttk.elte.hu,
cbyang@ccnu.edu.cn

Hadron spectral functions carry all the information of hadrons and are encoded in the Euclidean two-point correlation functions. The extraction of hadron spectral functions from the correlator is a typical ill-posed inverse problem and infinite number of solutions to this problem exists. We propose a novel neural network (sVAE) based on the Variation Auto-Encoder (VAE) and Bayesian theorem. Inspired by the maximum entropy method (MEM) we construct the loss function of the neural work such that it includes a Shannon-Jaynes entropy term and a likelihood term. The sVAE is then trained to provide the most probable spectral functions. For the training samples of spectral function we used general spectral functions produced from the Gaussian Mixture Model. After the training is done we performed the mock data tests with input spectral functions consisting 1) only a free continuum, 2) only a resonance peak, 3) a resonance peak plus a free continuum and 4) a NRQCD motivated spectral function. From the mock data test we find that the sVAE in most cases is comparable to the maximum entropy method in the quality of reconstructing spectral functions and even outperforms the MEM in the case where the spectral function has sharp peaks with insufficient number of data points in the correlator. By applying to temporal correlation functions of charmonium in the pseudoscalar channel obtained in the quenched lattice QCD at $0.75\,T_c$ on $128^3 \times 96$ lattices and $1.5\,T_c$ on $128^3 \times 48$ lattices, we find that the resonance peak of η_c extracted from both the sVAE and MEM has a substantial dependence on the number of points in the temporal direction (N_τ) adopted in the lattice simulation and N_τ larger than 48 is needed to resolve the fate of η_c at $1.5\,T_c$.

---

*Speaker

1. Introduction

Hadron spectral function is a quantity of great importance, since it includes all the information about hadrons, such as the dissociation temperature or transport coefficients. However, it is not directly accessible from lattice QCD computations, and is encoded in the Euclidean correlators computable on the lattice as [1]

$$G(\tau, T) = \sum_{x,y,z} \left\langle J_H(0, \vec{0}) J_H^+(\tau, \vec{x}) \right\rangle_T = \int_0^\infty \frac{d\omega}{2\pi} K(\omega, \tau, T) \rho(\omega, T) , \quad (1)$$

where the $K(\omega, \tau, T)$ is an integral kernel expressed as,

$$K(\omega, \tau, T) = \frac{\cosh(\omega(\tau - \frac{1}{2T}))}{\sinh(\frac{\omega}{2T})} . \quad (2)$$

The extraction of the spectral functions from the correlators is a typical ill-posed inverse problem, since typically $O(10)$ data points for the correlator are available on the lattice, however, around $O(1000)$ data points in the spectral function are needed in the ω space for a proper reconstruction. Thus in principle infinite number of solutions exists. Various methods e.g., maximum entropy method (MEM) [2], modified MEM with a different entropy term [3], stochastic approaches [4], Backus Gilbert method [5] and Tikhonov regularization [6], have been adopted to tackle the inverse problem. Among these methods, the MEM, which is based on the Bayesian theorem, reconstructs the most probable spectral function and is proved to have an unique solution if it exists.

Recently machine learning has also been applied in the reconstruction of spectral functions [7–12]. Variational auto-encoder (VAE) [13] is based on the Bayesian theorem and includes two parts: one is the encoder or recognition model and the other is the decoder or generative model. These two support each other and are connected via the distribution over the latent space $\mathcal{Z}$. The former encodes the information of the input data x into a latent space $\mathcal{Z}$ and passes a conditional probability $Q(z|x)$ into a bottleneck architecture, while the latter decodes the latent space $\mathcal{Z}$ that was encoded in the bottleneck layer by the encoder to regenerate the probability $P(x|z)$ of the inputs. These results backpropagate from the neural network in the form of the loss function, during which the evidence lower bound consisting of a marginal likelihood of the input data and the closeness to the true posterior $Q_{gt}(z|x)$

In this proceedings, we present a novel neural network based on the VAE, i.e. sVAE, to reconstruct the hadron spectral function from Euclidean temporal correlation functions, and the detailed version of the current proceedings can be seen in [14]. Inspired by the MEM, the loss function in the sVAE is constructed such that it includes a Shannon-Jaynes entropy term and a likelihood function. The 's' in the name of 'sVAE' is then attributed to the entropy term in the loss function. The sVAE is trained such that it learns to obtain the most probable spectral function by balancing the closeness to prior information of spectral function encoded in the entropy and the correlator data in the likelihood. In section 2 we present the details of the loss function in the sVAE, the topology of the neural network and the training samples of spectral functions. In section 3 we show the mock data tests of the sVAE. In section 4 we show the application of the sVAE to the correlators computed in the quenched lattice QCD. We finally summarize our results in section 5.

2. sVAE

In this section, we present the details on the sVAE, including its loss function, topology and the used training samples of spectral functions. To avoid the clutter we suppress the arguments τ and T of the temporal correlator G, and ω and T of the spectral function ρ.

2.1 The loss function and topology of the sVAE

The loss function of the sVAE is expressed as follows [14],

$$\mathcal{L} = -E_{Q(z|\rho_{gt},G_{gt})}\left[\log \frac{P(\rho|z)P(G|\rho,z)}{P(G|z)}\right] + KL\Big(Q(z|\rho_{gt},G_{gt})\|P(z|G)\Big), \tag{3}$$

where $Q(z|\rho_{gt},G_{gt})$ is the distribution of latent code z given ρ_{gt} and G_{gt} as the "ground-truth" values of spectral and correlation function, respectively. $P(z|G)$ gives the probability of z given the correlator G. These two probabilities are estimated by two networks of "Encoder 1" and "Encoder 2" (see Fig. 1), respectively. "Encoder 1" encodes the (ρ_{gt},G_{gt}) pair into latent space $\mathcal{Z}$, while "Encoder 2" maps the stochastic correlation functions to the same latent space $\mathcal{Z}$

$$KL\Big(Q(z|G_{gt},\rho_{gt})\|P(z|G)\Big) = \int dz\, Q(z|G_{gt},\rho_{gt}) \log\left(\frac{Q(z|G_{gt},\rho_{gt})}{P(z|G)}\right). \tag{4}$$

The prior probability $P(\rho|z)$ in Eq. (3) is parameterized in terms of the Shannon-Jaynes entropy S as [15, 16],

$$P(\rho|z) = \frac{1}{Z_S}e^S, \qquad \text{where} \quad S = \sum_l^{N_\omega}\left(\rho_l(z) - \rho_{gt,l} - \rho_l(z)\log\left(\frac{\rho_l(z)}{\rho_{gt,l}}\right)\right). \tag{5}$$

Here $Z_S \approx (2\pi)^{\frac{N_\omega}{2}}$ is normalization factor, and N_ω is the number of data points used to sample the frequency space ω. $P(G|\rho,z)$ gives the probability to reconstruct the measured correlator G parameterized in terms of the likelihood function [15, 16],

$$P(G|z,\rho) = \frac{1}{Z_L}e^{-L} \qquad \text{with} \quad L = \sum_{j=\tau_{min}}^{N_\tau/2} L_j = \sum_{j=\tau_{min}}^{N_\tau/2} \frac{\left(\hat{G}_j[\rho(z)] - G_j\right)^2}{2\alpha_j^2(z)G_j^2} \tag{6}$$

having a form of a likelihood function, and $Z_L = \prod_j^{N_\tau}\sqrt{2\pi\alpha_j^2(z)G_j^2}$. Here $\hat{G}[\rho(z)]$ is the correlator reconstructed from $\rho(z)$ according to the relation shown in Eq. 1, and $\alpha(z)$ is a weight parameter to adjust the weight between the entropy term and likelihood term. Through the "Encoder 1" and "Encoder 2" we obtain the probability of z, i.e. $Q(z|G_{gt},\rho_{gt})$ and $P(z|G)$, respectively, we then construct a "Decoder" network (cf. Fig. 1). Through the "Decoder" the spectral function $\rho(z)$ and the corresponding weight $\alpha(z)$ are obtained for the evaluation of the loss function.

During the training the first term in Eq. (3) aims to minimize $S - L$ by balancing the closeness of ρ to the ground truth value of spectral function ρ_{gt} and to correlator data G. This resembles the MEM. While the second term, the KL term, enforces the z distribution obtained from "Encoder 2" to approach the one from "Encoder 1". After the training, we drop "Encoder1", and only the "Encoder 2"-"Decoder" chain is used for the reconstruction as shown in Fig. 2. Since "Encoder
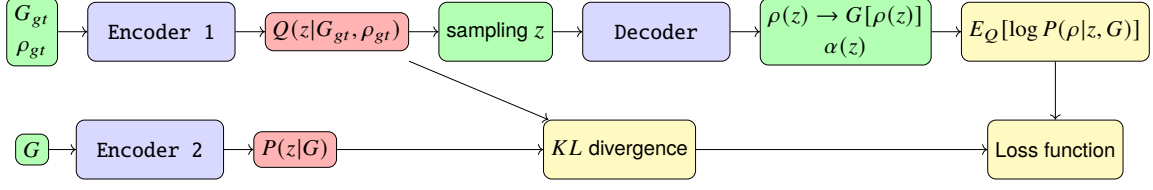
Figure 1: Training process involving the encoders and decoder [14].

Figure 2: Reconstruction process [14].

2" results in a probability distribution of the latent variable z given the stochastic correlator G, the reconstructed spectral function can be obtained with N_s samples as follows [14],

$$\tilde{\rho} = \int dz \, \rho(z) P(G|\rho, z) P(z|G) \rightarrow \sum_{n=1}^{N_s} \frac{\rho(z) P(G|\rho, z_n) P(z_n|G)}{\sum\limits_{k=1}^{N_s} P(G|\rho, z_k) P(z_k|G)} . \tag{7}$$

2.2 Training samples

The training spectral function is constructed using a Guassian mixture model. In practice the spectral function consists of a smoothed sum of many Gaussian peaks with randomly chosen width and height parameter, and also a continuum (constant) part [14]. The peak locations of Gaussian peaks are chosen uniformly within [0,4] in the ω space with number of points $N_\omega = 10000$. Five sampled training spectral functions and the covered region are illustrated in Fig. 3.
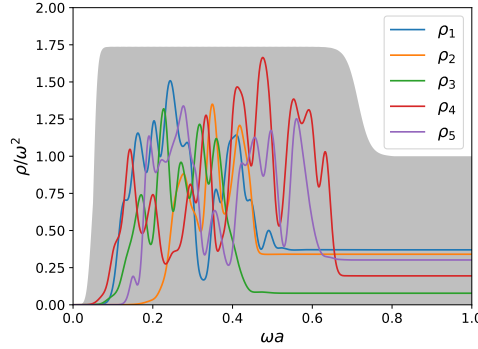

Figure 3: 5 samples of the training spectral function ρ. The grey shadow area shows the full sampled range [14].

The training sample ρ_{gt} and resulting temporal correlator $G_{gt}(\tau)$ are fed to "Encoder 1" (cf. Fig. 1). Only $\tau \le N_\tau/2$ is used due to the symmetry (Eq. (2)), and $\tau \ge 4$ is set due to lattice cutoff effects. For each $G_{gt}(\tau)$, a Gaussian noise is added to construct the stochastic data $G(\tau)$ so as to mimic the error from computations in lattice QCD. Thus G is sampled from a Gaussian distribution

with its mean value G_{gt} and its variance having the following form,

$$\sigma_{lat}(\tau)/G_{lat}(\tau) \times G_{gt}(\tau), \tag{8}$$

where $\sigma_{lat}(\tau)/G_{lat}(\tau)$ is the lattice noise ratio level at τ (typically 1.5% at $\tau = N_\tau/2$). The resulting stochastic data $G(\tau)$ is fed to "Encoder 2" (cf. Fig. 1).
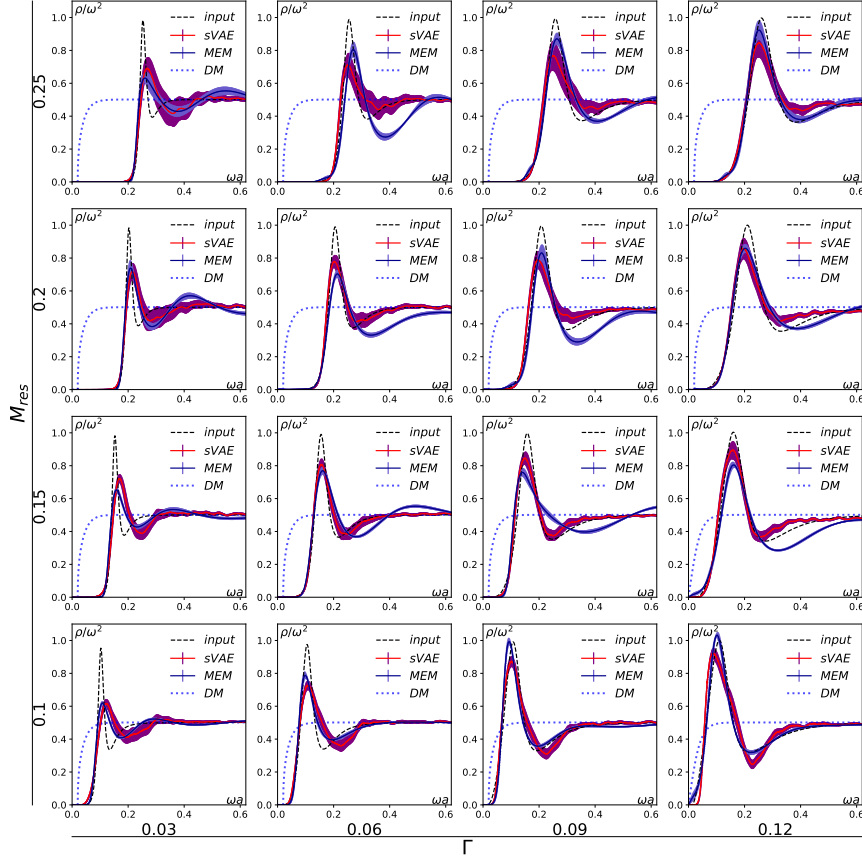
3. Mock data tests

After the training with the training samples of spectral functions illustrated in Fig. 3 is done, we first make two mock data tests with only a free continuum spectral function and only a resonance peak in the input spectral function. The results from both the sVAE and MEM in these two cases are shown in the left and right plots of Fig. 4, respectively. It can be seen that the results from the sVAE and MEM in these two cases are comparable, and sVAE seems outperform MEM in the reconstruction of the peak location of the resonance peak shown in the right plot of Fig. 4.
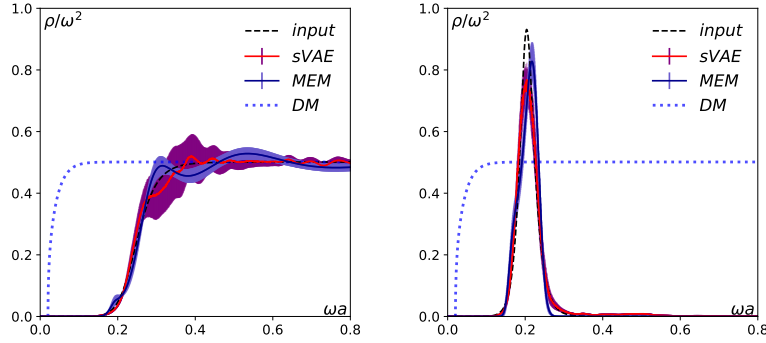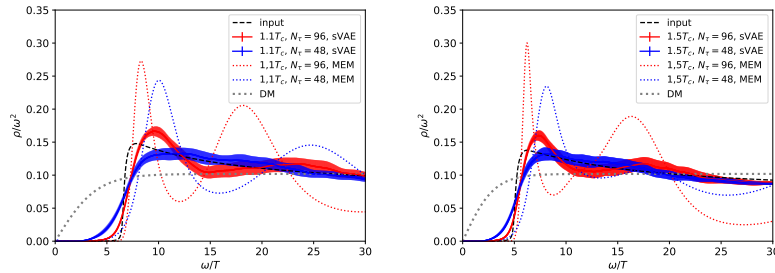

Figure 4: Mock data tests with an input spectral function containing only a continuum part (left) and with only a resonance peak (right). The input mock spectral function is denoted by the black dashed line, and the mean value and uncertainty of the output spectral function from the trained sVAE are represented by the red solid line and purple band, respectively. The blue line and blue band are the mean value and uncertainty of the output spectral function from MEM, respectively. The corresponding blue dotted line is the default model of MEM. The results are obtained with $N_\tau = 96$ and are taken from Ref. [14].

We make further mock data tests with a more complicated input spectral function. The input spectral function now consists of a continuum part and one resonance (Breit-Wigner) peak. In this test we study the dependence of the reconstruction quality on the peak location and width of the resonance peak in the input spectral function. In Fig. 5 we increase the peak width of the input spectral function (Γ) from left to right horizontally, and increase the peak location (M_{res}) from bottom to top vertically. As seen from Fig. 5 both the sVAE and MEM cannot reconstruct the spectral function in a satisfactory way when the peak width is small, e.g. $\Gamma = 0.03$. With $\Gamma \geq 0.06$ both the sVAE and MEM can reconstruct the spectral function reasonably well and yield comparable results.

As understood from the tests shown in Fig. 5 the reconstruction quality crucially depends on how rapidly the peak rises or how small the peak width is, we make a further mock data test

Figure 5: Mock data tests with each input spectral function containing a resonance peak and a continuum part with $N_\tau = 96$ [14]. The legends are the same as those in Fig. 4.


Figure 6: Mock data tests with input spectral functions obtained from 2-loop perturbative NRQCD at T=1.1T_c (left) and 1.5T_c (right) [14]. In both plots the input spectral function is denoted by the black dashed line, and the red solid line and blue solid line with surrounding bands denote the results obtained from the sVAE for $N_\tau = 96$ and $N_\tau = 48$, respectively. The dotted lines with the same color denote results obtained from the MEM with the default model denoted as grey dotted lines.

with a NRQCD motiviated spectral function as the input spectral function. The NRQCD spectral function increases rapidly around the threshold ($\omega/T \sim 6$) and then slowly decreases. We show the reconstructed spectral functions obtained from both the sVAE and MEM in Fig 6. It turns out that both the sVAE and MEM cannot reproduce the input spectral function, and the MEM even produce fake peaks which do not exist in the input spectral functions.

4. Analysis on lattice QCD data

In this section, we apply the sVAE to reconstructing spectral functions from Euclidean two-point correlation functions computed in quenched lattice QCD. We focus on the charmonium correlator in the pseudo-scalar channel. The correlator data analyzed here comes from [17], which were computed on $128^3 \times 96$ and $128^3 \times 48$ in quenched lattice QCD corresponding to temperatures $0.75\ T_c$ and $1.5\ T_c$ with inverse lattice spacing a^{-1} = 18.97 GeV, respectively. To avoid lattice cut-off effects τ_{min} is set to 4, and the frequency space has been cut at $\omega_{max} a = 4$ due to the lattice cutoff in the free spectral function [18, 19].
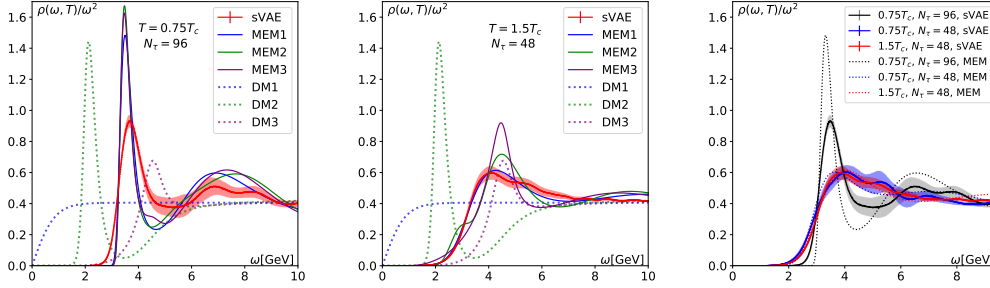
Figure 7: Left: Spectral functions at $0.75\ T_c$ obtained from the sVAE and MEM with $N_\tau = 96$. Middle: Same as the left plot but obtained at $1.5\ T_c$ with $N_\tau = 48$. Right: A combined plot of the left and right ones. Here the MEM results obtained only using 'DM1' are shown. In addition spectral functions obtained from the sVAE and MEM at $0.75\ T_c$ with $N_\tau = 48$ are also shown.

In the left plot of Fig. 7 we show the reconstructed spectral functions from the sVAE at 0.75 T_c. For comparison results obtained using the MEM are also shown with three different default models (DM). It can be observed that the results from the MEM are almost independent of DMs and the reconstructed peak location lies at $\omega \approx 3.34$ GeV. The mass of η_c as read-off from this peak location is consistent with the pole mass extracted from the spatial correlation function at 0.75 T_c [17]. On the other hand, the peak location obtained from the sVAE is about 5% larger than that from the MEM. And the peak height obtained from the sVAE is always smaller than that from the MEM. Despite these differences, the resonance peak structure obtained from the sVAE and MEM is comparable.

At $1.5\ T_c$ the reconstruction becomes harder since the number of data points become half of that at $0.75T_c$. As seen from the middle plot of Fig. 7 the spectral function obtained from the sVAE does not possess a remark peak structure and is only comparable with the MEM result obtained using a featureless DM ('DM1'). For the MEM results obtained using 'DM 2' and 'DM 3' the peak location shifts to a much larger $\omega \approx 4.26$ GeV.

To make direct comparison between the results shown in the left and right plots of Fig. 7 we summarize them in the right plot of Fig. 7 with MEM results only from 'DM1'. As mentioned in the previous paragraph it seems that significant thermal effects are attributed to the change in the spectral function at 1.5 T_c from 0.75 T_c. However, we also need to check the N_τ dependence in the reconstruction of the spectral functions from both methods. We thus looked at the reconstructed correlator at 0.75 T_c. This reconstructed correlator encodes the same spectral function at 0.75 T_c but has a smaller number of points in the temporal direction, i.e. $N_\tau = 48$. It can be clearly seen that there exits a large N_τ dependence in both sVAE and MEM during the extraction of $\rho(\omega, 0.75T_c)$. It is thus hard to tell the fate of η_c at 1.5 T_c based on the current lattice QCD data.

5. Summary and conclusion

In this proccedings, we presented a novel neural network (sVAE) based on the the variational auto-encoder and Bayesian theorem. The sVAE is trained such that it provides the most probable spectral function by balancing the prior and the likelihood in the loss function. For the training samples we used general spectral functions constructed using the Gaussian mixture model. In the mock data tests, we found that results from the sVAE are comparable to those from the MEM, and in some cases sVAE even outperforms MEM. In the application to the lattice QCD correlator data, the peak locations of the results at 0.75 T_c from sVAE and MEMs are consistent with each other, and the peak height in the output spectral function from the sVAE is always smaller than that from the MEM. At 1.5 T_c, the N_τ dependence of spectral functions still imposes a challenge to extract a reliable spectral function. Thus, to determine the fate of η_c at 1.5 T_c, N_τ larger than 48 is probably needed.

References

[1] H.-T. Ding, F. Karsch, and S. Mukherjee, "Thermodynamics of strong-interaction matter from Lattice QCD," *Int. J. Mod. Phys. E*, vol. 24, no. 10, p. 1530007, 2015.

[2] M. Asakawa and T. Hatsuda, "J / psi and eta(c) in the deconfined plasma from lattice QCD," *Phys. Rev. Lett.*, vol. 92, p. 012001, 2004.

[3] Y. Burnier and A. Rothkopf, "Bayesian Approach to Spectral Function Reconstruction for Euclidean Quantum Field Theories," *Phys. Rev. Lett.*, vol. 111, p. 182003, 2013.

[4] H.-T. Ding, O. Kaczmarek, S. Mukherjee, H. Ohno, and H. T. Shu, "Stochastic reconstructions of spectral functions: Application to lattice QCD," *Phys. Rev. D*, vol. 97, no. 9, p. 094503, 2018.

[5] G. Backus and F. Gilbert, "The Resolving Power of Gross Earth Data," *Geophysical Journal International*, vol. 16, pp. 169–205, 10 1968.

[6] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Soviet Math. Dokl.*, vol. 4, pp. 1035–1038, 1963.

[7] L. Kades, J. M. Pawlowski, A. Rothkopf, M. Scherzer, J. M. Urban, S. J. Wetzel, N. Wink, and F. P. G. Ziegler, "Spectral Reconstruction with Deep Neural Networks," *Phys. Rev. D*, vol. 102, no. 9, p. 096001, 2020.

[8] S. Offler, G. Aarts, C. Allton, J. Glesaaen, B. Jäger, S. Kim, M. P. Lombardo, S. M. Ryan, and J.-I. Skullerud, "News from bottomonium spectral functions in thermal QCD," *PoS*, vol. LATTICE2019, p. 076, 2019.

[9] M. Zhou, F. Gao, J. Chao, Y.-X. Liu, and H. Song, "Application of radial basis functions neutral networks in spectral functions," *Phys. Rev. D*, vol. 104, no. 7, p. 076011, 2021.

[10] J. Horak, J. M. Pawlowski, J. Rodríguez-Quintero, J. Turnwald, J. M. Urban, N. Wink, and S. Zafeiropoulos, "Reconstructing QCD Spectral Functions with Gaussian Processes." arXiv:2107.13464, 2021.

[11] S. Shi, K. Zhou, J. Zhao, S. Mukherjee, and P. Zhuang, "Heavy Quark Potential in QGP: DNN meets LQCD." arXiv:2105.07862, 2021.

[12] L. Wang, S. Shi, and K. Zhou, "Reconstructing spectral functions via automatic differentiation." arXiv:2111.14760, 2021.

[13] D. P. Kingma and M. Welling, "An introduction to variational autoencoders," *Foundations and Trends® in Machine Learning*, vol. 12, no. 4, pp. 307–392, 2019.

[14] S. Y. Chen, H. T. Ding, F. Y. Liu, G. Papp, and C. B. Yang, "Machine learning spectral functions in lattice qcd." arXiv:2110.13521, 2021.

[15] M. Jarrell and J. Gubernatis, "Bayesian inference and the analytic continuation of imaginary-time quantum monte carlo data," *Physics Reports*, vol. 269, no. 3, pp. 133–195, 1996.

[16] M. Asakawa, T. Hatsuda, and Y. Nakahara, "Maximum entropy analysis of the spectral functions in lattice QCD," *Prog. Part. Nucl. Phys.*, vol. 46, pp. 459–508, 2001.

[17] H. T. Ding, A. Francis, O. Kaczmarek, F. Karsch, H. Satz, and W. Soeldner, "Charmonium properties in hot quenched lattice QCD," *Phys. Rev. D*, vol. 86, p. 014509, 2012.

[18] F. Karsch, E. Laermann, P. Petreczky, and S. Stickan, "Infinite temperature limit of meson spectral functions calculated on the lattice," *Phys. Rev. D*, vol. 68, p. 014504, 2003.

[19] G. Aarts and J. M. Martinez Resco, "Continuum and lattice meson spectral functions at nonzero momentum and high temperature," *Nucl. Phys. B*, vol. 726, pp. 93–108, 2005.