Bayesian Approach to Spectral Function Reconstruction for Euclidean Quantum Field Theories

Yannis Burnier and Alexander Rothkopf
Albert Einstein Center for Fundamental Physics, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland
(Received 1 August 2013; published 31 October 2013)

We present a novel approach to the inference of spectral functions from Euclidean time correlator data that makes close contact with modern Bayesian concepts. Our method differs significantly from the maximum entropy method (MEM). A new set of axioms is postulated for the prior probability, leading to an improved expression, which is devoid of the asymptotically flat directions present in the Shanon-Jaynes entropy. Hyperparameters are integrated out explicitly, liberating us from the Gaussian approximations underlying the evidence approach of the maximum entropy method. We present a realistic test of our method in the context of the nonperturbative extraction of the heavy quark potential. Based on hard-thermal-loop correlator mock data, we establish firm requirements in the number of data points and their accuracy for a successful extraction of the potential from lattice QCD. Finally we reinvestigate quenched lattice QCD correlators from a previous study and provide an improved potential estimation at $T = 2.33T_C$.

DOI: 10.1103/PhysRevLett.111.182003 PACS numbers: 12.38.Aw, 11.15.Ha, 12.38.Gc, 12.38.Mh

The numerical solution of inverse problems is an active area of research with important applications in science and engineering. In the context of QCD physics, the estimation of spectral functions from Euclidean correlators is of particular interest. A reliable determination of ground and excited state properties of mesons and baryons at zero temperature [1] from nonperturbative Monte Carlo simulations (lattice QCD), e.g., represents an important bridge between field theory and experiment. At finite temperature, lattice spectral functions allow us to scrutinize the physics of the early universe by elucidating phenomena, such as heavy quarkonium melting [2] and the transport properties [3] of the quark-gluon plasma produced in relativistic heavy-ion collisions.

The most common approach to spectral function reconstruction deployed today, the maximum entropy method (MEM) [4], is based on Bayesian inference. Nevertheless, even after 20 years of application [2,5–7], the reliability of the MEM is still under discussion [8–10]. Here we introduce a novel Bayesian approach that addresses key issues affecting the MEM: slow convergence of the underlying optimization task, high computational cost for extended search spaces, scale dependence in the prior functional and the Gaussian approximation required in the hyperparameter estimation.

The Bayesian strategy [11] relies on an application of the multiplication law for the joint probability distribution of the spectral function of the system under investigation ρ, the measured data D, and any other prior information I,

$$P[\rho, D, I] \Rightarrow P[\rho|D, I] = \frac{P[D|\rho, I]P[\rho|I]}{P[D|I]}. \quad (1)$$

We specify in the likelihood probability $P[D|\rho, I]$ how the data is obtained, while the prior probability $P[\rho|I]$ encodes how prior information on ρ itself enters the posterior $P[\rho|D, I]$. Ultimately our interest lies in the maximum *a posteriori* estimate for ρ, which we refer to as the Bayesian solution to the inverse problem.

In the following, we aim at inverting the convolution,

$$D(\tau) = \int d\omega K(\omega, \tau)\rho(\omega), \quad (2)$$

which connects the spectral function $\rho(\omega) > 0$ through a known kernel function $K(\tau, \omega)$ to the correlation function $D(\tau)$. In practice the correlator is estimated at N_τ points $D(\tau_i) = D_i$ from a sample of Gaussian distributed measurements. After discretization of the frequencies along N_ω points spaced by $\Delta\omega_l = \omega_{l+1} - \omega_l$, we can compute the corresponding data for each spectrum $\rho(\omega_l) = \rho_l$,

$$D_i^\rho = \sum_{l=1}^{N_\omega} \Delta\omega_l K_{il}\rho_l. \quad (3)$$

According to the Gaussian assumption, we use the quadratic distance

$$L = \frac{1}{2}\sum_{ij}(D_i - D_i^\rho)C_{ij}^{-1}(D_j - D_j^\rho), \quad (4)$$

to assign a probability to the data given a test spectral function. Here C_{ij} denotes the covariance matrix of the data points. We know that, independently of N_ω, if $L/N_\tau \gg 1$, ρ does not reproduce the data points within their errorbars, while if $L/N_\tau \ll 1$ the spectrum will contain many unnatural structures arising from overfitting the noise in the data. If the data is obtained from Gaussian distributed measurements and we insert the correct underlying ρ in Eq. (3), it leads to $L \sim N_\tau$. Hence, the most neutral reconstruction will satisfy $L/N_\tau = 1$, which we impose as a constraint arising from prior knowledge. Our likelihood probability thus reads

$$P[D|\rho, I] = \exp[-L - \gamma(L - N_\tau)^2], \tag{5}$$

where the limit $\gamma \to \infty$ is taken numerically. Note that maximizing this expression alone is still ill defined, since the $N_\omega \gg N_\tau$ parameters ρ_l are not yet uniquely fixed.

Hence, we continue by specifying the prior probability $P[\rho|I]$, which acts as a regulator and will allow us to select a unique Bayesian set of ρ_l's. The MEM utilizes the Shanon-Jaynes entropy S_{SJ} in its prior probability [4], which is constructed from four axioms. Similarly, we introduce our expression for $P[\rho|I] \propto \exp[S]$, after replacing two of these axioms.

Prior information is incorporated explicitly through a function $m(\omega_l) = m_l$, which, by definition, is the correct spectral function in the absence of data [4]. It is usually obtained through previous reconstructions with less accurate data or from independent estimates. We begin the construction of the functional S with the following:

Axiom I: subset independence. Let us consider two different subsets Ω_1 and Ω_2 along the frequency axis. If prior information imposes constraints on the spectrum ρ within each of these subsets, then the result of the reconstruction should not depend on treating these domains separately or in a combined fashion $S[\Omega_1, m(\Omega_1)] + S[\Omega_2, m(\Omega_2)] = S[\Omega_1 \cup \Omega_2, m(\Omega_1 \cup \Omega_2)]$. This relation is satisfied if S is written as an integral over frequencies,

$$S \propto \int d\omega s(\rho(\omega), m(\omega), \omega). \tag{6}$$

While this axiom coincides with the one used in the MEM, we continue by introducing the following new one:

Axiom II: scale invariance. In general $\rho(\omega)$ does not have to be a probability distribution. Indeed, depending on the observable the spectrum is associated with, its scaling can differ from $1/\omega$. We hence require that the choice of units for ρ and m shall not change the result of the reconstruction; i.e., we must construct our prior probability using ratios of ρ/m only,

$$S = \tilde{\alpha} \int d\omega s(\rho(\omega)/m(\omega)). \tag{7}$$

Now that the integrand s does not carry a dimension, we introduce the dimensionful hyperparameter $\tilde{\alpha}$ to also make the argument of the exponential dimensionless.

Axiom III: smoothness of the reconstructed spectra. The only certain information about the spectral function is that it is a positive definite and smooth function. Hence, we wish the prior functional to impose these traits on the reconstructed spectrum even if no further prior information is known; i.e., in the case of $m(\omega) = m_0$, a smooth spectrum shall be chosen independently of m_0.

The strategy towards this end relies on penalizing spectra which deviate between two adjacent frequencies ω_1 and ω_2. If changing the ratio $r_l = \rho_l/m_0$ at the two frequencies does not change the values of D^ρ beyond the errorbars, then S should favor $r_1 = r_2$. The penalty between the case where the same value exists at both frequencies $r_1 = r_2 = r$ and the case where they differ by a small amount $r_1 = r(1 + \epsilon)$, $r_2 = r(1 - \epsilon)$, hence, has to be independent of r and symmetric in whether $r_1 \gtrless r_2$,

$$2s(r) - s(r(1 + \epsilon)) - s(r(1 - \epsilon)) = \epsilon^2 C_2. \tag{8}$$

This is precisely the discretized expression for the differential equation $-r^2 s''(r) = C_2$, whose solution yields

$$S = \tilde{\alpha} \int d\omega \left(C_0 - C_1 \frac{\rho}{m} + C_2 \ln\left(\frac{\rho}{m}\right)\right). \tag{9}$$

The remaining axiom is identical to the MEM case, since it establishes the Bayesian meaning of $m(\omega)$.

Axiom IV: maximum at the prior. In the absence of data, S must become maximal at $\rho = m$. Conventionally, its value at this point is chosen to vanish,

$$S(r = 1) = 0, \qquad S'(r = 1) = 0, \qquad S''(r = 1) < 0. \tag{10}$$

The two first conditions fix the constants C_0, C_1 and C_2 up to an overall constant, which we absorb into the hyperparameter $\alpha \propto \tilde{\alpha}$. The last condition forces α to be positive. Our final result, hence, reads

$$S = \alpha \int d\omega \left(1 - \frac{\rho}{m} + \ln\left(\frac{\rho}{m}\right)\right). \tag{11}$$

This new prior distribution is strictly concave ($S \leq 0$) and exhibits a similar quadratic behavior around the maximum $\rho = m$ as the Shanon-Jaynes entropy S_{SJ}. Hence, the uniqueness of its maximum can be established analogously to the MEM [5]. In the case where ρ_l, $m_l \ll 1/\alpha$ or $\rho_l \ll m_l$, their contribution to S and to the variation $\delta S/\delta\rho$ is not suppressed; thus, we avoid the asymptotic flatness inherent in S_{SJ}. To obtain a proper probability density, we still need to normalize e^S, which leads us to

$$P[\rho|\alpha, m] = e^S \Big/ \prod_{i=1}^{N_\omega} e^{\alpha\Delta\omega_i} (\alpha\Delta\omega_i)^{-\alpha\Delta\omega_i} m_i \Gamma(\alpha\Delta\omega_i). \tag{12}$$

By construction, our prior distribution contains the positive hyperparameter α, which we need to treat in a Bayesian fashion. In the MEM, several spectra ρ^α for different values of α are reconstructed and ultimately averaged over [4], weighted by the values of the evidence $P[D|\alpha, m]$. The calculation of the evidence relies however on a Gaussian approximation, whose validity is not guaranteed.

Here we take a different route [12] and integrate out α explicitly from the joint probability distribution $P[\rho, D, \alpha, m, I]$; i.e., we take into account the influence of all possible prior distributions in the resulting posterior probability $P[\rho|D, m, I]$ on which we base the reconstruction. As the prior information I entering Eq. (5) is independent from the hyperparameter, we carry it implicitly for the moment. Starting from the multiplication law for

$$P[\rho, D, \alpha, m] = P[D|\rho, \alpha, m]P[\rho|\alpha, m]P[\alpha|m]P[m]$$
$$= P[\alpha|\rho, D, m]P[\rho|D, m]P[D|m]P[m],$$

we note that m and α are independent. The likelihood $P[D|\rho, \alpha, m]$ does not contain any knowledge on the hyperparameters underlying the given ρ. Hence only $P[\alpha|\rho, D, m]$, $P[\rho|\alpha, m]$ and $P[\alpha]$ carry an α dependence.

Assuming no further knowledge, i.e ($P[\alpha] = 1$), we integrate the joint probability with respect to α and arrive at the α-independent

$$P[\rho|D, m, I] = \frac{P[D|\rho, I]}{P[D|m, I]} \int d\alpha P[\rho|\alpha, m]. \quad (13)$$

In the above expression the I labels are restored, $P[D|\rho, I]$ is given by Eq. (5), $P[\rho|\alpha, m]$ by Eq. (12), and $P[D|m, I]$ is an irrelevant constant. For large values of S, we approximate the integral over α through a next-to-leading order resummation of logarithms, while for small S a numerical evaluation is possible.

After integration, no dependence on α remains. The presence of m is not problematic, for its constant values do not influence the reconstruction result if ρ is properly normalized. Hence, $P[\rho|D, m, I]$ does not contain any meaningful external parameters, and we can proceed to find its maximum numerically. To this end we deploy the quasi-Newton memory limited Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm, which allows us to approach $\delta P[\rho|D, m, I]/\delta\rho = 0$ by varying each of the N_ω parameters ρ_l individually. An inversion of the Hessian matrix at intermediate steps is not required, which leads to a significant reduction in computational cost compared to the usual Levenberg-Marquardt approach. Note that in contrast to the MEM with S_{SJ}, now without asymptotically flat directions, we successfully locate the global extremum of $P[\rho|D, m, I]$ and do not need to stop the algorithm at an artificial cutoff in step size.

One application is the static potential between two heavy quarks, which is related to the spectral structure of the rectangular Wilson loop $W_\square(r, \tau)$ and possibly the Wilson line correlator in Coulomb gauge $W_{||}(r, \tau)$ [7]. The extraction of such spectra with $K(\omega, \tau) = e^{-\omega\tau}$ has been shown to pose a severe challenge to the MEM [10].

Here we benchmark our approach through reconstruction of known spectra from cutoff regularized ($\Lambda = 5\pi$) Euclidean correlators, calculated in hard thermal loop (HTL) resummed perturbation theory [10]. Subsequently, we attempt to extract from them the known HTL interquark potential [13] by fitting the position and width of the lowest- lying peak [7]. The ideal HTL data points are perturbed by Gaussian noise with variance $\sigma_i^2 = (\eta D_i)^2$, leading to constant relative errors $\Delta D/D = \eta$. In preparation for lattice QCD, we assume no prior knowledge on the spectrum and supply a constant prior $m(\omega) = 1$.

In Fig. 1 we present reconstructions [$N_\omega = 1000$, $\omega \in [-126, 189]$ (left), $N_\omega = 1200$, $\omega \in [-15.7, 15.7]$ GeV (right)] from mock data at $T = 2.33T_C$ for qualitative comparison. The reference MEM [9] (ρ^{MEM32}) based on $N_\tau = 32$ ideal data points [10] fails to reproduce even the Lorentzian shape of the lowest peak in the HTL spectrum (ρ^{HTL}). $\rho_\square^{\mathrm{HTL}}$ contains a peak and a large background, both of which our method is able to capture with $N_\tau = 32$ at $\Delta D/D = 10^{-4}$. In $\rho_{||}^{\mathrm{HTL}}$ a single peak is dominant for which $N_\tau = 32$ at $\Delta D/D = 10^{-2}$ suffices. To showcase possible improvements for future lattice QCD studies, we also present the results for $N_\tau = 128$ data points and $\Delta D/D = 10^{-5}$ (BR128-5).

As a quantitative check, we reproduce the known HTL interquark potential from the lowest spectral peak. Our method hence needs to yield the correct position (Re[V]) and width (Im[V]) of a skewed Lorentzian. In the top panel

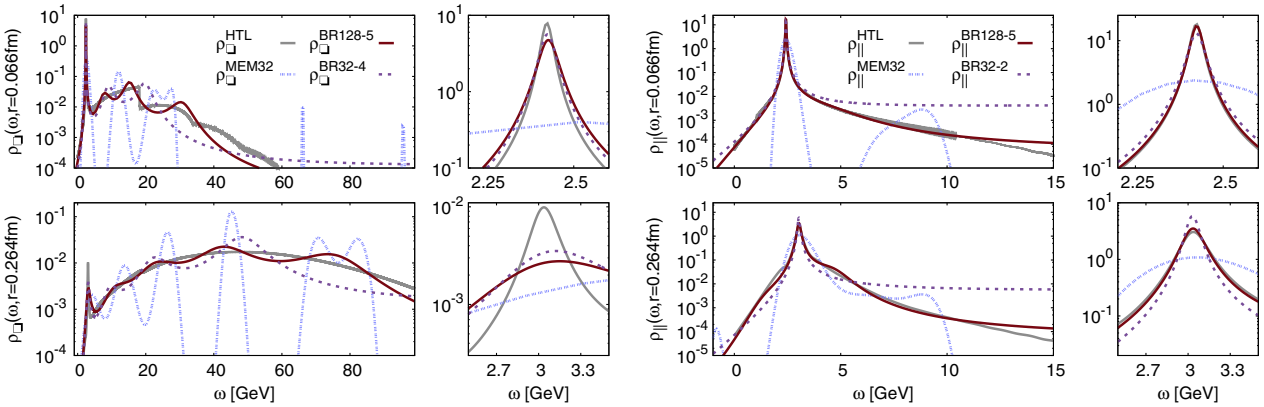

FIG. 1 (color online). Wilson loop (left) and Wilson line correlator (right) spectra from HTL mock data [10] at $r = 0.066$ fm (top) and 0.264 fm (bottom). As reference we show the HTL spectra ρ^{HTL} and the MEM ρ^{MEM32} [9] with extended search space $N_{\mathrm{ext}} = 80$ based on $N_\tau = 32$ ideal data points. Both peak and background in $\rho_\square^{\mathrm{HTL}}$ are captured using $N_\tau = 32$ and $\Delta D/D = 10^{-4}$ (BR32-4), while the single peak in $\rho_{||}^{\mathrm{HTL}}$ requires only $\Delta D/D = 10^{-2}$ (BR32-2). BR128-5 shows the result for a still realistic scenario of $N_\tau = 128$ with mock error $\Delta D/D = 10^{-5}$.

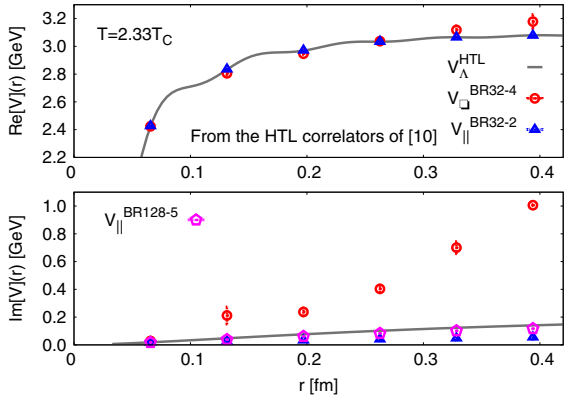

FIG. 2 (color online). Reconstruction of the HTL ($T=2.33T_C$) potential (solid line) based on the inferred spectra from the HTL Wilson loop $V_\square(r)$ (circle) and the HTL Wilson line correlator $V_{||}(r)$ (triangle) of [10]. (top) Re$[V](r)$ is reproduced with sub 10% and sub 1% deviation from $N_\tau = 32$ data points with $\Delta D/D = 10^{-4}$ and 10^{-2}, respectively. On the other hand, Im$[V](r)$ (bottom) requires at least $N_\tau = 128$ with $\Delta D/D = 10^{-5}$ for 20% accuracy (pentagon).

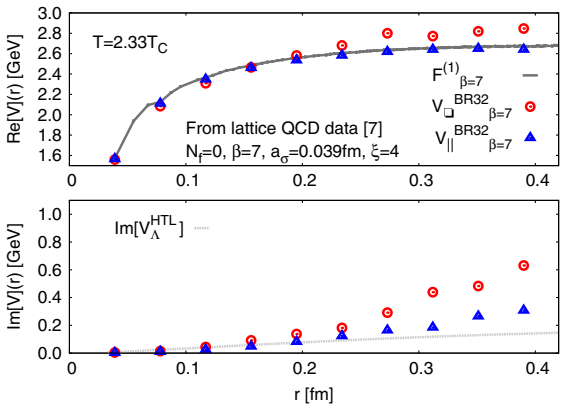

FIG. 3 (color online). Improved estimate for the heavy quark potential from quenched lattice QCD [7] Wilson loops $V_\square(r)$ (circle) and Wilson line correlators $V_{||}(r)$ (triangle). (top) At the small values of $r < 0.4$ fm shown, Re$[V](r)$ appears to lie close to the Coulomb gauge color singlet free energies $F^{(1)}$ (solid line). (bottom) At $T = 2.33T_C$ we find an Im$[V](r)$ that is of the same order as in HTL perturbation theory (dashed line).

of Fig. 2 we show the real part (solid line) obtained from the Wilson loop $V_\square(r)$ (circle) and Wilson line correlators $V_{||}(r)$ (triangle). Error bars are estimated from three reconstructions with different mock noise of equal strength. With $N_\tau = 32$ the correct real part is reproduced with 10% and 1% accuracy, respectively; especially the strong artificial rise in Re$[V_\square]$ observed in the MEM in Ref. [10] is absent.

With our method it is also possible to reproduce the width from the Wilson line correlators. To this end we utilize $N_\tau = 128$ and $\Delta D/D = 10^{-5}$, which is still realistic in quenched lattice QCD. The resulting reconstruction of Im$[V](r)$ with sub 20% deviation is shown in the lower panel of Fig. 2 (pentagon). (For $N_\tau = 128$ we only show Im$[V_{||}](r)$, since the background from cusp divergences in $\rho_\square$ [10] will necessitate even better data.)

With these limitations in mind, we apply our method to the Wilson loop and Wilson lines in quenched lattice QCD [7] at $T = 2.33T_C$ ($N_\tau = 32$, $\beta = 7$, $a_\sigma = 0.0039$ fm, $\xi = 4$). The improved estimate in Fig. 3 for both Re$[V_\square]$ and Re$[V_{||}]$ shows that, as expected at the small distances treated here [14], their values lie close to the color singlet free energies in Coulomb gauge $F^{(1)}$. While we do not expect the width, i.e., Im$[V]$ (bottom), to be captured reliably yet at $N_\tau = 32$, it is interesting to note that its values appear to be of the same order of magnitude as in the HTL calculation at this temperature.

We have introduced a novel Bayesian approach to spectral function reconstruction. It cures the conceptual and practical issues affecting the MEM by introducing an improved dimensionless prior distribution devoid of asymptotically flat directions in Eq. (11). In the case of a constant prior function $m(\omega) = m_0$ and normalization of ρ, no external parameter needs to be adjusted, since we integrate out explicitly the hyperparameter α as shown in Eq. (13). Combined with the LBFGS optimizer algorithm, which varies each of the individual N_ω parameters ρ_l, we achieve a significant improvement in the reconstruction of spectra as demonstrated in Figs. 1 and 2, enabling in turn a reliable determination of the $T > 0$ heavy-quark potential (Fig. 3). Hence we look forward to further applications in lattice QCD and beyond.

The authors thank T. Hatsuda, S. Sasaki, O. Kaczmarek, S. Kim, P. Petreczky, and H. T. Ding for fruitful discussions on the MEM, C. A. Rothkopf for insight on Bayesian inference, and the DFG-Heisenberg group of Y. Schröder at Bielefeld University for computing resources. This work was partly supported by the Swiss National Science Foundation (SNF) under Grant No. 200021-140234.

[1] T. Yamazaki *et al.*, Phys. Rev. D **65**, 014501 (2001); K. Sasaki, S. Sasaki, and T. Hatsuda, Phys. Lett. B **623**, 208 (2005); G. S. Bali, S. Collins, and C. Ehmann, Phys. Rev. D **84**, 094506 (2011); Z. Fodor and C. Hoelbling, Rev. Mod. Phys. **84**, 449 (2012).

[2] T. Umeda, K. Nomura, and H. Matsufuru, Eur. Phys. J. C **39**, 9 (2005); M. Asakawa and T. Hatsuda, Phys. Rev. Lett. **92**, 012001 (2004); S. Datta, F. Karsch, P. Petreczky, and I. Wetzorke, Phys. Rev. D **69**, 094507 (2004); A. Jakovac, P. Petreczky, K. Petrov, and A. Velytsky, Phys. Rev. D **75**, 014506 (2007); G. Aarts, C. Allton, M. B. Oktay, M. Peardon, and J.-I. Skullerud, Phys. Rev. D **76**, 094513 (2007); H. Ohno, S. Aoki, S. Ejiri, K. Kanaya, Y. Maezawa, H. Saito, and T. Umeda, Phys. Rev. D **84**, 094504 (2011); G. Aarts, C. Allton, S. Kim, M. P. Lombardo, M. B. Oktay, S. M. Ryan, D. K. Sinclair, and J.-I. Skullerud, J. High Energy Phys. 11 (2011) 103; H. T.

Ding, A. Francis, O. Kaczmarek, F. Karsch, H. Satz, and W. Soeldner, Phys. Rev. D **86**, 014509 (2012).
[3] G. Aarts, C. Allton, J. Foley, S. Hands, and S. Kim, Phys. Rev. Lett. **99**, 022002 (2007); H.-T. Ding, A. Francis, O. Kaczmarek, F. Karsch, E. Laermann, and W. Soeldner, Phys. Rev. D **83**, 034504 (2011).
[4] J. Skilling and S. F. Gull, Lect. Notes Phys., Monogr. **20**, 341 (1991); M. Jarrell and J. E. Gubernatis, Phys. Rep. **269**, 133 (1996);
[5] M. Asakawa, T. Hatsuda, and Y. Nakahara, Prog. Part. Nucl. Phys. **46**, 459 (2001);
[6] D. Nickel, Ann. Phys. (N.Y.) **322**, 1949 (2007).
[7] A. Rothkopf, T. Hatsuda, and S. Sasaki, Proc. Sci., LAT2009 (2009) 162 [arXiv:0910.2321]; Phys. Rev. Lett. **108**, 162001 (2012); Y. Burnier and A. Rothkopf, Phys. Rev. D **86**, 051503 (2012); A. Rothkopf, Mod. Phys. Lett. A **28**, 1330005 (2013).
[8] O. Gunnarsson, M. W. Haverkort, and G. Sangiovanni, Phys. Rev. B **82**, 165125 (2010); A. Rothkopf, Proc. Sci., LATTICE 2012 (2012) 100 [arXiv:1208.5162].
[9] A. Rothkopf, J. Comput. Phys. **238**, 106 (2013).
[10] Y. Burnier and A. Rothkopf, Phys. Rev. D **87**, 114019 (2013).
[11] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2007), 2nd ed.
[12] D. J. C. MacKay, Neural Comput. **11**, 1035 (1999).
[13] M. Laine, O. Philipsen, P. Romatschke, and M. Tassler, J. High Energy Phys. 03 (2007) 054; A. Beraudo, J.-P. Blaizot, and C. Ratti, Nucl. Phys. **A806**, 312 (2008); N. Brambilla, J. Ghiglieri, A. Vairo, and P. Petreczky, Phys. Rev. D **78**, 014017 (2008).
[14] Y. Burnier, M. Laine, and M. Vepsalainen, J. High Energy Phys. 01 (2010) 054 and references therein.