A new Bayesian approach to the reconstruction of spectral functions

Yannis Burnier

*Institut de Théorie des Phénomènes Physiques, Ecole Polytechnique Fédérale de Lausanne,
CH-1015, Lausanne, Switzerland,*
Albert Einstein Center, University of Bern, Sidlerstr. 5, CH-3012 Bern, Switzerland
E-mail: yannis.burnier@epfl.ch

Alexander Rothkopf[*]

Albert Einstein Center, University of Bern, Sidlerstr. 5, CH-3012 Bern, Switzerland
E-mail: rothkop@itp.unibe.ch

We present a novel approach for the reconstruction of spectra from Euclidean correlator data that makes close contact to modern Bayesian concepts. It is based upon an axiomatically justified dimensionless prior distribution, which in the case of constant prior function $m(\omega)$ only imprints smoothness on the reconstructed spectrum. In addition we are able to analytically integrate out the only relevant overall hyper-parameter α in the prior, removing the necessity for Gaussian approximations found e.g. in the Maximum Entropy Method. Using a quasi-Newton minimizer and high-precision arithmetic, we are then able to find the unique global extremum of $P[\rho|D]$ in the full $N_\omega \gg N_\tau$ dimensional search space. The method actually yields gradually improving reconstruction results if the quality of the supplied input data increases, without introducing artificial peak structures, often encountered in the MEM. To support these statements we present mock data analyses for the case of zero width delta peaks and more realistic scenarios, based on the perturbative Euclidean Wilson Loop as well as the Wilson Line correlator in Coulomb gauge.

---

[*]Speaker.

---

1. Motivation

Lattice QCD constitutes an important tool in elucidating the physics of the strong force from first principles. Its Monte-Carlo based simulation techniques in Euclidean time allow us to investigate a multitude of non-perturbative settings out of reach from any other method available today. To connect to phenomenology, real-time information however needs to be extracted, whose direct determination remains out of reach due to the presence of the notorious sign problem.

All information encoded in the two-point functions of a quantum field theory in Minkowski time can be obtained from the real-valued spectral function $\rho(\omega)$ of the system. I.e. it allows us to generate any propagator, corresponding to one of the possible operator orderings, by an integration in the complex plane. For the Euclidean correlator on the other hand the connection to $\rho(\omega)$ is established via a real-valued integral kernel

$$D(\tau) = \int d\omega K(\tau, \omega) \rho(\omega). \tag{1.1}$$

Lattice QCD simulations give us access to stochastic estimates of $D(\tau_i) = D_i$ at a finite number N_τ of points along the temporal axis. The task at hand is to give meaning to the inherently *ill-defined* problem of extracting from this limited and noisy data the rich features of $\rho(\omega_l) = \rho_l > 0$ along $N_\omega \gg N_\tau$ frequencies.

The most common approach to invert a relation such as (1.1) is based on Bayesian inference [2] (for a recent non-Bayesian proposal see e.g. [3]). Beginning with Ref. [4] the strategy of determining the most probable spectral functions based on Bayes theorem

$$P[\rho|D,I] = \frac{P[D|\rho,I]P[\rho|I]}{P[D|I]} \quad \Rightarrow \quad \left.\frac{\delta P[\rho|D,I]}{\delta\rho}\right|_{\rho=\rho^{Bayes}} = 0 \qquad \begin{array}{ll} P[D|\rho,I] & \text{likelihood probability} \\ P[\rho|I] & \text{prior probability} \\ P[D|I] & \text{evidence} \end{array}$$

has found its way into the lattice QCD community under the name of Maximum Entropy Method (MEM), where the prior probability $P_{MEM}[\rho|I] = \exp[S_{SJ}]$ is given by the Shannon-Jaynes entropy [2, 4]. While important insight on finite temperature QCD [5] has been gained using the MEM, several technical short-comings (which have afflicted our own studies in the past [1]) prevented us from utilizing the full potential of the Bayesian approach:

- The dimensionality of the search space in the standard implementation by Bryan is fully determined by the number of datapoints N_τ. If the structures encoded in the spectrum are spread over a wide frequency range it has been shown [6] that the limited number of basis functions can adversely impact the reconstruction result. One consequence is that comparisons of spectra obtained at different temperatures (i.e. from datasets with different N_τ) contain additional systematic uncertainty, as the underlying search space differs.

- If the search space is decoupled from the number of datapoints [6] and the number of available basis functions increased, we find that the convergence of the underlying optimization task slows down significantly, since there exist flat directions in the Shannon-Jaynes entropy functional ($\rho_l, m_l \ll 1/\alpha$ or $\rho_l \ll m_l$). In addition, a multitude of peaks appears that are not part of the actual encoded spectrum and it is not trivial to identify them as such.

- The MEM contains a hyperparameter α, which is treated self-consistently based on a Gaussian approximation [2]. This can be justified only if the default model $m(\omega)$ lies close to the correct spectrum. In practice, we usually do not possess reliable information about the peak structures we wish to reconstruct, so that a different treatment of α is called for.

Here we wish to report on a novel strategy [7] for spectral function reconstruction that attempts to remedy the above mentioned issues. In Section:

2. We introduce a modified likelihood function that contains additional knowledge on how neutral reconstructions can be achieved.

3. A prior distribution for ρ is constructed axiomatically, which in the presence of a constant default model $m(\omega) = m_0$ imprints only smoothness on the reconstructed spectrum.

4. Through the advantageous analytical structure of this new prior functional we are able to integrate out explicitly the hyperparameter α from the joint probability distribution.

5. We present spectral reconstructions from mock data containing a three delta peak spectrum as well as a challenging scenario based on the hard-thermal loop Wilson loop and Wilson lines in Coulomb gauge [8, 9]. Using the quasi-Newton LBFGS algorithm allows us to carry out the necessary optimization tasks in the full N_ω dimensional search space.

2. The likelihood

The datapoints D_i we use to reconstruct the spectrum are assumed to arise from the averaging of Gaussian distributed measurements. Consequently the probability of the data, given a test spectral function ρ can be quantified using the quadratic distance

$$L = \frac{1}{2}\sum_{ij}(D_i - D_i^\rho)C_{ij}^{-1}(D_j - D_j^\rho), \tag{2.1}$$

where C_{ij} denotes the correlation matrix of the data and D_i^ρ are the values obtained by plugging ρ into the discretized version of Eq. (1.1)

$$D_i^\rho = \sum_{l=1}^{N_\omega} \Delta\omega_l K_{il}\rho_l. \tag{2.2}$$

Note that in the function L the discretization of frequencies N_ω and $\Delta\omega_l = \omega_{l+1} - \omega_l$ does not enter explicitly, as it operates on the level of Euclidean time datapoints. Furthermore we know that if $L \gg N_\tau$, ρ does not reproduce the measured data within errorbars. Contrary, if $L \ll N_\tau$, it is to be expected that the inevitable noise in the data is overfitted and artificial distortions in the reconstruction ensue. Actually, if we were to know the correct spectral function underlying the Gaussian distributed measurements D_i and plug it into Eq. (2.1) via Eq. (2.2), we will obtain $L \approx N_\tau$. Hence we require for the most neutral reconstruction that $L = N_\tau$ by adding a constraint to the function L

$$P[D|\rho, I] = \exp[-L - \gamma(L - N_\tau)^2], \tag{2.3}$$

where we take $\gamma \to \infty$ numerically. We have not yet achieved to regularize the procedure. Indeed there still exist an infinite number of degenerate solutions maximizing Eq. (2.3). From Bayes theorem we know that taking into account prior knowledge is the key to a unique answer.

3. A new prior distribution

The prior distribution used in the MEM has been derived, based on very general axioms that were chosen with image reconstruction in mind [2]. In our case the focus is quite different. Besides emphasizing the faithful reconstruction of the structures actually encoded in the data, we aim at preventing the appearance of artificial peaks, which otherwise impede the physics interpretation of the obtained results. Hence we wish our prior probability to favor smooth functions if no other prior information is given[1].

To construct $P[\rho|I] = \exp[S]$, we begin with the axiom *Subset independence* already used in the derivation of the Shannon Jaynes entropy, which tells us that S has to be written as an integral over frequencies. Our new axiom *Scale invariance* is related to the fact that depending on the dimensionality of the correlation function $D(\tau)$ in Eq. (1.1) the scaling of $\rho(\omega)$ also differs. Note that the default model $m(\omega)$ must exhibit the same scaling as the correct spectrum. We require that the choice of units does not influence the result of the reconstruction so that S has to be a function of the ratio ρ/m only

$$S = \tilde{\alpha} \int d\omega \, s\Big(\rho(\omega)/m(\omega)\Big). \tag{3.1}$$

A dimensionfull hyperparameter $[\alpha] = 1/[\omega]$ is introduced to make the overall expression dimensionless.

We proceed with our third axiom *Smoothness of the reconstructed spectra*. If the default model is constant $m(\omega) = m_0$, we require S to choose a smooth spectrum independent of m_0. This is achieved by penalizing spectra, whose values at adjacent frequencies ω_1 and ω_2 differ from each other. I.e. the penalty awarded by S between the situation where the ratios $r_l = \rho_l/m_l$ at the two frequencies are the same $r_1 = r_2 = r$ and where they differ as $r_1 = r(1+\varepsilon)$ and $r_2 = r(1-\varepsilon)$ has to be independent of r and must be the same regardless of $r_1 \gtrless r_2$. The corresponding discrete differential equation $2s(r) - s(r(1+\varepsilon)) - s(r(1-\varepsilon)) = \varepsilon^2 C_2$ hence leads us to

$$S = \tilde{\alpha} \int d\omega \left(C_0 - C_1 \frac{\rho}{m} + C_2 \ln\left(\frac{\rho}{m}\right) \right). \tag{3.2}$$

The final step is taken using the axiom *Maximum at the prior*, which we inherit from the MEM derivation. It establishes the Bayesian meaning of $m(\omega)$ by requiring that S is maximal for $\rho = m$. This fixes all the integration constants in Eq. (3.2) and we end up with the final expression

$$S = \alpha \int d\omega \left(1 - \frac{\rho}{m} + \ln\left(\frac{\rho}{m}\right) \right), \tag{3.3}$$

where $\alpha > 0$. Note that $S < 0$ for all $\rho \neq m(\omega)$ and that its behavior around the quadratic minimum $\rho = m$ is identical to S_{SJ}. Hence the proof of the uniqueness of the Bayesian reconstruction can be directly carried over from the MEM. The last remaining step requires us to treat the positive hyperparameter α in a Bayesian fashion.

---

[1]Reconstructing delta peaked spectra that arise at zero temperature is nevertheless possible. The prior only imprints its signature on those parts of the spectrum that are not constrained by the data. Indeed very good data (large N_τ and small ΔD) can and will override its influence (see e.g. Sec. 5).
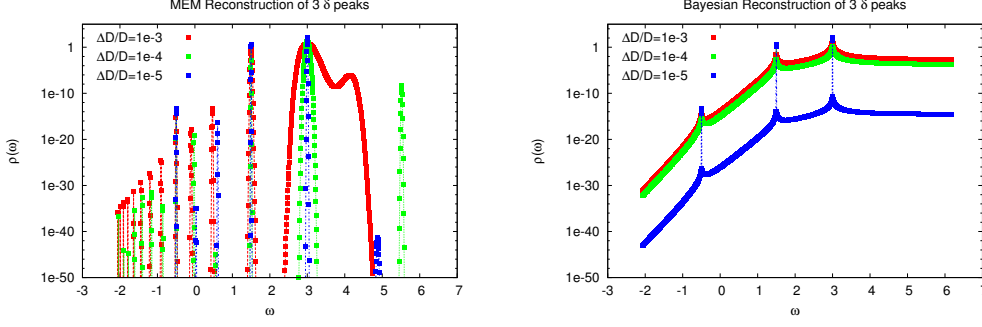
Figure 1: Bayesian reconstruction of a three δ-peak mock spectrum using MEM (left) and our Bayesian method (right). Both methods are supplied $N_\tau = 32$ ideal datapoints perturbed by Gaussian noise corresponding to errors of $\Delta D/D = 10^{-3}$ (red) 10^{-4} (green) 10^{-5} (blue). Note the complete absence of artificial peaked structures on the right. At $\Delta D/D = 10^{-5}$ there still remains a width of two points (right), since the used $\omega \in [-2, 6]$ with $N_\omega = 1200$ does not contain the exact values $\omega = -0.5, 1.5, 3$.

4. Integration of hyperparameters

Since we do not have any additional information on the hyperparameter α, we wish to integrate it out explicitly from the joint probability distribution $P[\rho, D, \alpha, m, I]$. For notational clarity we will refrain for the moment from writing out the label I for the prior knowledge entering Eq. (2.3), as it is completely independent from α. The joint probability can be expanded as

$$P[\rho, D, \alpha, m] = P[D|\rho, \alpha, m]P[\rho|\alpha, m]P[\alpha|m]P[m] = P[\alpha|\rho, D, m]P[\rho|D, m]P[D|m]P[m], \quad (4.1)$$

where a closer inspection reveals that not only m and α are independent but also $P[D|\rho, \alpha, m]$ does not depend on the hyperparameter, since ρ itself is one of the given quantities. Thus the only α dependence is found in $P[\rho|\alpha, m], P[\alpha|\rho, D, m]$ and $P[\alpha]$. Introducing the integrals with respect to α and setting $P[\alpha] = 1$ one obtains

$$P[D|\rho] \int d\alpha P[\rho|\alpha, m] = \int d\alpha P[\alpha|\rho, D, m]P[\rho|D, m]P[D|m]. \quad (4.2)$$

By definition $\int d\alpha P[\alpha|\rho, D, m] = 1$ and we can rearrange the terms to get

$$P[\rho|D, m, I] = \frac{P[D|\rho, I]}{P[D|m, I]} \int d\alpha P[\rho|\alpha, m], \quad (4.3)$$

where the labels I have been restored. This expression constitutes the basis for our reconstruction. We carry out the single remaining α integral for large values of S through a next-to-leading order resummation of logarithms, while at small S a numerical integration can be performed.

5. Numerical implementation and mock data tests

To find the maximum of Eq. (4.3) we need to vary the $N_\omega \gg N_\tau$ components ρ_l individually. To do so we use the LBFGS algorithm, which approximates the inverse Hessian matrix of the Levenberg-Marquardt scheme iteratively. This significantly reduces the computational cost and we can actually perform reconstructions with $N_\omega \sim \mathcal{O}(1000)$ without problem. In the following we demonstrate the performance of our method for the kernel $K(\tau, \omega) = \exp[-\omega\tau]$ by mock data
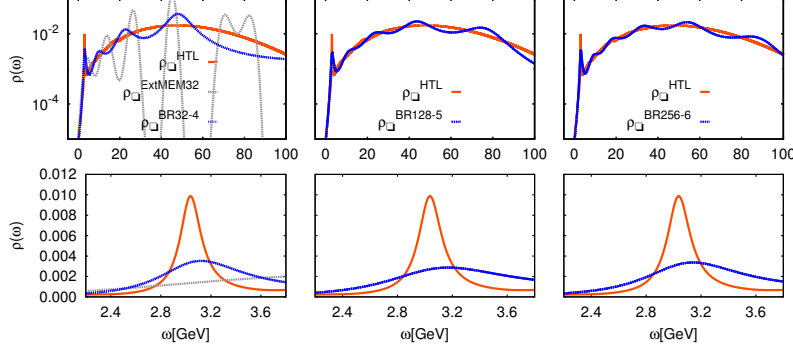
Figure 2: Bayesian reconstruction of the $T = 2.33 \times 270$MeV hard-thermal loop Wilson loop spectrum (orange) from $N_\tau = 32, 128, 256$ Euclidean datapoints with $\Delta D/D = 10^{-4}, 10^{-5}, 10^{-6}$ (left-to-right, blue) at $r = 0.264$fm. (left column) A reference MEM reconstruction from [8] is shown in gray

analyses. In the first a three delta peak spectrum $\rho(-0.5) = 10^{-14}, \rho(1.5) = \frac{1}{26}, \rho(3) = \frac{25}{26}$ is encoded via Eq. (2.2) in $N_\tau = 32$ datapoints to which Gaussian noise is added. For the reconstruction we discretize $\omega \in [-2,6]$ with $N_\omega = 1200$ points and use 512 bit arithmetic. Assuming no knowledge of the peak positions, we provide a flat prior. As shown in Fig. 1, the results from the MEM (left) and our method (right) are quite distinct. For different errors in the data $\Delta D/D = 10^{-3}$ (red) 10^{-4} (green) 10^{-5} (blue) we consistently obtain smaller reconstructed widths and more importantly manage to completely avoid artificial peak structures that show up in the MEM.

Where the data constrains the spectrum and overrides the prior, sharp peaks arise. Everywhere else, our prior favors a smooth functional form, which becomes parallel to $m(\omega)$ at large ω. Another important point is that with our new method we find a true maximum of Eq. (2.3). I.e. even for $\Delta D/D = 10^{-5}$ the optimizer step size diminishes down to $\Delta = 10^{-60}$ within several hours, while in the MEM values of $\Delta < 10^{-9}$ are not encountered even after one week of running.

A more demanding test (see also [7, 9]) is based on the analytically calculated hard-thermal loop Wilson loop $W_\square^{HTL}(r,\tau)$ and the Wilson line correlator in Coulomb gauge $W_\parallel^{HTL}(r,\tau)$[8]. These play a central role in determining the heavy quark potential at finite temperature [1, 7, 8, 9]. In particular the position and width of the lowest lying peak in $\rho(r,\omega)$ at a certain spatial extend r encodes the real- and imaginary part of the $V_{\bar{Q}Q}(r)$ respectively. It is known [8] that for the Wilson loop this peak is quite small and embedded in a large background from the cusp divergences of the closed rectangular path. The Wilson lines feature a much smaller background.

We show in Fig. 2 the reconstruction of $\rho_\square^{HTL}(r,\omega)$ (orange) from $W_\square^{HTL}(r,\tau)$ ($N_\omega = 1000$, $\omega \in [-126, 189]$GeV) at $r = 0.264$fm ($T = 2.33T_C$) for improving input data $N_\tau = 32, 128, 256$ and $\Delta D/D = 10^{-4}, 10^{-5}, 10^{-6}$ (left-to-right, blue). More points and lower error enable us to better capture the higher frequencies, while the physically important peak remains challenging. Its position at best lies within 10% of the correct value but the width is reproduced only poorly (see also [9]).

In the absence of a divergence induced background, the reconstruction of $\rho_\parallel^{HTL}(r,\omega)$ in Fig. 3 (orange) performs much better. We show the results from $W_\parallel^{HTL}(r,\tau)$ ($N_\omega = 1200$, $\omega \in [-15.7, 15.7]$GeV) for $N_\tau = 16, 32, 128$ and $\Delta D/D = 10^{-2}, 10^{-4}, 10^{-5}$ (left-to-right, blue). The position of the peak is already captured satisfactorily with the lowest quality data. Here more of the overall datapoints contribute to the low frequency peak. Thus with $N_\tau = 128$ and $\Delta D/D = 10^{-5}$, which is still attainable in quenched QCD, we manage to reproduce the width with sub 20% deviation. These results, together with the performance shown in the reconstruction of delta peak structures in Fig. 1, give
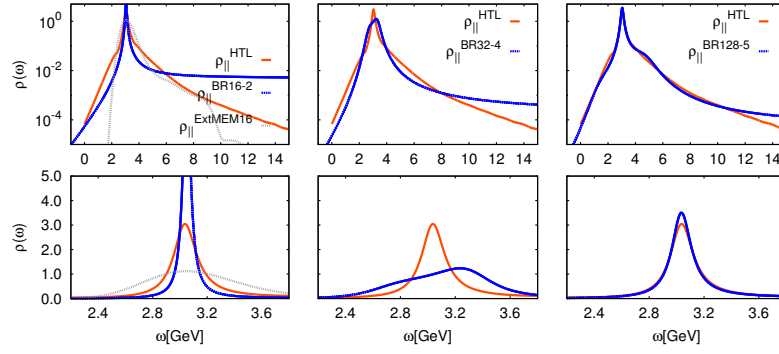
Figure 3: Reconstruction of the $T = 2.33 \times 270$MeV hard-thermal loop Wilson line spectrum (orange) from $N_\tau = 32, 128, 256$ Euclidean datapoints with $\Delta D/D = 10^{-2}, 10^{-4}, 10^{-5}$ (left-to-right, blue) at $r = 0.264$fm. (left column) A MEM reconstruction with the settings of [8], based on $\Delta D/D = 10^{-2}$ data, is shown in gray.

us confidence that our method can not only contribute to an improved determination of the heavy-quark potential (see [9]) but also to the reconstructions of spectral features from lattice simulations in general (see e.g. S. Kim's contribution to this conference [10]).

References

[1] A. Rothkopf, T. Hatsuda and S. Sasaki, PoS LAT **2009**, 162 (2009); Phys. Rev. Lett. **108**, 162001 (2012); Y. Burnier,A. Rothkopf, Phys. Rev. D **86**, 051503 (2012); A. Rothkopf, Mod. Phys. Lett. A **28**, 1330005 (2013).

[2] J. Skilling, S.F. Gull, Lecture Notes-Monograph Series 20, 341 (1991); M. Jarrell and J.E. Gubernatis, Physics Reports, **269**, 133-195, (1996).

[3] G. Cuniberti, E. De Micheli and G. A. Viano, Commun. Math. Phys. **216** (2001) 59. Y. Burnier, M. Laine, L. Mether, Eur. Phys. J. **C71** (2011) 1619; Y. Burnier, M. Laine, Eur. Phys. J. **C72** (2012) 1902.

[4] M. Asakawa, T. Hatsuda and Y. Nakahara, Prog. Part. Nucl. Phys. **46**, 459 (2001).

[5] T. Umeda, K. Nomura and H. Matsufuru, Eur. Phys. J. C **39S1**, 9 (2005); M. Asakawa, 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 *et al.*, Phys. Rev. D **84**, 094504 (2011); G. Aarts et al., JHEP **1111**, 103 (2011); H. T. Ding, A. Francis, O. Kaczmarek, F. Karsch, H. Satz and W. Soeldner, Phys. Rev. D **86**, 014509 (2012). G. Aarts et al., Phys. Rev. Lett. **99**, 022002 (2007); H. -T. Ding et. al., Phys. Rev. D **83**, 034504 (2011).

[6] A. Rothkopf, J. Comput. Phys. **238**, 106 (2013); A. Rothkopf, PoS LATTICE **2012**, 100 (2012);

[7] Y. Burnier and A. Rothkopf, arXiv:1307.6106 [hep-lat].

[8] Y. Burnier and A. Rothkopf, Phys. Rev. D **87**, 114019 (2013).

[9] Y. Burnier and A. Rothkopf, PoS LATTICE **2013**, 491 (2013).

[10] S. Kim, P. Petreczky and A. Rothkopf, PoS LATTICE **2013**, 169 (2013).