Fast and efficient stochastic optimization for analytic continuation

F. Bao,[1,2] Y. Tang,[1,3] M. Summers,[1,4] G. Zhang,[1] C. Webster,[1] V. Scarola,[3] and T. A. Maier[1,4]
[1]*Computer Science and Mathematics Division, Oak Ridge, Tennessee 37831, USA*
[2]*Department of Mathematics, University of Tennessee at Chattanooga, Chattanooga, Tennessee 37403, USA*
[3]*Department of Physics, Virginia Tech, Blacksburg, Virginia 24061, USA*
[4]*Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA*
(Received 15 June 2016; revised manuscript received 6 September 2016; published 28 September 2016)

The analytic continuation of imaginary-time quantum Monte Carlo data to extract real-frequency spectra remains a key problem in connecting theory with experiment. Here we present a fast and efficient stochastic optimization method (FESOM) as a more accessible variant of the stochastic optimization method introduced by Mishchenko *et al.* [Phys. Rev. B **62**, 6317 (2000)], and we benchmark the resulting spectra with those obtained by the standard maximum entropy method for three representative test cases, including data taken from studies of the two-dimensional Hubbard model. We generally find that our FESOM approach yields spectra similar to the maximum entropy results. In particular, while the maximum entropy method yields superior results when the quality of the data is strong, we find that FESOM is able to resolve fine structure with more detail when the quality of the data is poor. In addition, because of its stochastic nature, the method provides detailed information on the frequency-dependent uncertainty of the resulting spectra, while the maximum entropy method does so only for the spectral weight integrated over a finite frequency region. We therefore believe that this variant of the stochastic optimization approach provides a viable alternative to the routinely used maximum entropy method, especially for data of poor quality.

DOI: 10.1103/PhysRevB.94.125149

I. INTRODUCTION

Quantum Monte Carlo (QMC) methods provide numerically exact results for interacting quantum many-particle systems, and thus they are widely used to study their physics. An important drawback, however, is their inability to directly give real frequency results, a key limitation considering the large number of experiments that measure dynamic quantities. From the imaginary-time QMC data, the real frequency spectrum $A(\omega)$ has to be recovered through the process of analytic continuation, a highly ill-posed inverse problem that remains a key stumbling block in connecting theory with experiment.

To address this challenge, it has proven useful to employ a framework based on Bayesian statistical inference. The state-of-the-art and most widely used tool based on Bayesian statistics is the maximum entropy (MaxEnt) method [1] pioneered by Silver, Sivia, Jarrell, and Gubernatis [2–5] for applications in this area. It introduces an entropy-like regularization term that measures the deviation from a default spectrum, and then the most probable spectrum is obtained through a deterministic optimization process. Another method that uses explicit but adjustable regularization through the use of consistent constraints was recently introduced by Prokof'ev and Svistunov [6].

An alternative stochastic method was developed by Sandvik [7], in which a fictitious temperature is introduced to define the probability of a given spectrum by a Boltzmann weight. This allows for efficient Monte Carlo sampling of possible spectra from which the final spectrum is obtained as a weighted average. A refined version of this approach, which, similar to MaxEnt, uses a default model, was later introduced by Beach [8] and shown to become formally equivalent to the MaxEnt method if the fictitious system is treated at a mean-field level. In addition, Fuchs *et al.* [9] showed that the fictitious temperature introduced in this algorithm can be eliminated based on the principles of Bayesian statistical inference in a similar fashion as the regularization parameter of the MaxEnt approach is removed.

Mishchenko *et al.* [10] used a similar idea to set up a stochastic optimization method (SOM) that randomly samples solutions with a certain weight but without interpretation of the weights as a Boltzmann distribution. In this approach, one randomly samples a large enough number of possible solutions $A(\omega)$, each of which optimizes the deviation from the QMC data but allows for solutions with larger deviation to implicitly regularize the problem. One important feature of this approach is that it uses a different and much more complex parametrization of the spectrum that does not impose a rigid, discrete frequency grid and allows for overlapping rectangles from which the spectrum is composed. While this allows for more flexibility in the solution, it leads to a complex update algorithm and a very large search space that is difficult to manage.

Here, we introduce a fast and efficient stochastic optimization method (FESOM) as an accessible variant of Mishchenko's original SOM that is based on the same idea, i.e., a stochastic sampling of possible spectra. But instead of the complex parametrization introduced by Mishchenko *et al.*, it uses the usual parametrization of solutions $A(\omega)$ in terms of a discrete frequency grid, resulting in a more manageable algorithm. We apply this approach to a number of representative problems, and we compare the results against those obtained from standard MaxEnt calculations. We include two test cases of approximate spectral functions derived from the two-dimensional Hubbard model on a square lattice. In the following section, we review the state-of-the art MaxEnt method that we use to benchmark our approach. We then discuss our FESOM in Sec. IV, and we show the results of three different numerical examples in Sec. V.

II. ANALYTIC CONTINUATION AND BAYESIAN STATISTICS

The analytic continuation process involves an inversion of the integral,

$$G(i\omega_n) = \int d\omega\, K(i\omega_n,\omega)A(\omega). \tag{1}$$

Here, $G(i\omega_n)$ is an observable such as the single-particle Green's function measured in a QMC calculation as a function of discrete Matsubara frequencies ω_n on the imaginary axis, $A(\omega) = -1/\pi\, \mathrm{Im}\, G(\omega)$ is the spectral function and quantity of interest, and $K(i\omega_n,\omega)$ is the kernel. For the fermionic Green's function considered here, one has $\omega_n = (2n+1)\pi T$ for a temperature T, and the kernel takes the form

$$K(i\omega_n,\omega) = \frac{1}{i\omega_n - \omega}. \tag{2}$$

After discretization of the real frequency axis into L intervals, $\{\omega_l\}_{l=0}^{L}$, Eq. (1) is written in matrix-vector form,

$$G_n = \sum_{l=1}^{L} K_{nl} A_l, \tag{3}$$

with $K_{nl} \equiv \Delta\omega_l/(i\omega_n - \omega_l)$, $G_n \equiv G(i\omega_n)$, and $A_l \equiv A(\omega_l)$, and the frequency intervals $\Delta\omega_l = \omega_{l+1} - \omega_l$. The difficulty of inverting Eq. (1) arises from the small tails in the kernel function at large frequencies ω. In other words, the matrix K_{nl} is ill-conditioned, i.e., small changes or statistical errors in the QMC data G_n cause large errors in the quantity of interest A_l, and there are an infinite number of solutions.

Approaches that address this problem can be formulated in terms of Bayesian statistical inference, in which one considers the Bayesian formula

$$P(A|G) \propto P(G|A)P(A). \tag{4}$$

Here, $P(A|G)$ is the posterior probability of the spectrum A given the data G, the prior probability $P(A)$ encodes prior information about A, and the likelihood function $P(G|A)$ measures the quality of the fit between G and KA. The problem of finding the most probable spectrum A given the data G is thereby converted into the much easier problem of optimizing the likelihood function and prior probability. One can then select the most probable spectrum A that maximizes $P(A|G)$ as in the case of the MaxEnt method, or obtain the final spectrum A from a weighted average over possible solutions,

$$\bar{A} = \int dA\, A\, p(A|G), \tag{5}$$

as in the case of the stochastic methods.

III. MAXIMUM ENTROPY

The MaxEnt approach [5] uses the Bayesian statistical inference formula, Eq. (4), to find the most probable spectrum A given the input data G. This is done by maximizing both the likelihood function $P(G|A)$ and the prior probability $P(A)$.

The likelihood function $P(G|A)$ is defined according to the central limit theorem as

$$P(G|A) = e^{-\chi^2/2}, \tag{6}$$

where

$$\chi^2[A] = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{G_n - \sum_l K_{nl}A_l}{\sigma_n}\right)^2 \tag{7}$$

encodes the quality of the fit of the data G by the spectrum A. Here $G_n = 1/N_s \sum_{i=1}^{N_s} G_n^i$ is obtained as the mean value of a number N_s of different QMC samples, with $G_n^i \equiv G^i(i\omega_n)$ the ith sample, and the variance

$$\sigma_n^2 = \frac{1}{N_s - 1}\sum_{i=1}^{N_s}\left(G_n^i - G_n\right)^2. \tag{8}$$

Note that this form assumes that no correlations between different frequencies $i\omega_n$ are present in the QMC data G_n. When there are correlations, the covariance matrix has to be diagonalized, and both the data G_n and the kernel K have to be rotated into this diagonal representation [11].

A simple minimization of χ^2 with a least-squares fit of the data G with KA leads to noisy and an infinite number of nonunique solutions. The MaxEnt method addresses this problem by regularization of the least-squares fit. It introduces a prior distribution

$$p(A) = e^{\alpha S[A]}, \tag{9}$$

with α a positive constant, the regularization parameter, and

$$\begin{aligned} S[A] &= -\int d\omega\left[A(\omega) - D(\omega) - A(\omega)\ln\frac{A(\omega)}{D(\omega)}\right] \\ &= -\sum_{l=1}^{L}\left[A(\omega_l) - D(\omega_l) - A(\omega_l)\ln\frac{A(\omega_l)}{D(\omega_l)}\right]\Delta\omega_l \end{aligned} \tag{10}$$

an entropy-like term defined relative to a positive-definite and normalized function $D(\omega)$, the default model. Thus, in order to maximize the posterior probability $p(A|G)$, the MaxEnt minimizes the function

$$Q[A] = \tfrac{1}{2}\chi^2[A] - \alpha S[A]. \tag{11}$$

The Bayesian inverse optimization of the posterior probability $p(A|G) \propto e^{-Q[A]}$ hence becomes a deterministic optimization for the regularized form $\frac{1}{2}\chi^2[A] - \alpha S[A]$ as a standard optimization problem. Here α mediates the competition between the χ^2 fit of the data and prior information contained in $S[A]$. It is the $\chi^2[A]$ term that ensures that the spectral function will give a good fit of the data, while the $S[A]$ term avoids overfitting of the data by guiding $A(\omega)$ toward a default model given by $D(\omega)$. The Bayesian inference formulation also allows us to eliminate the free parameter α by calculating the posterior probability of α, i.e., $P(\alpha|G)$. One can then perform the MaxEnt procedure for different values of α to give estimates for the spectrum A_α and then select the most probable A_α that corresponds to the maximum $p(\alpha|\bar{G})$. Here we use Bryan's method [12], in which one averages over all spectra A_α weighted by the posterior probability of α, to obtain

$$A = \int d\alpha\, P(\alpha|G)A_\alpha. \tag{12}$$

IV. FAST AND EFFICIENT STOCHASTIC OPTIMIZATION METHOD

The stochastic inference method introduced by Sandvik [7] and refined by Beach [8] and Fuchs *et al.* [9], and the stochastic optimization method developed by Mishchenko [10], are alternative numerical approaches to solving the analytic continuation problem. While these approaches can outperform the traditional MaxEnt method and yield spectra with more features and less regularization, they can be very numerically expensive. Fuchs *et al.* [9] commented that the necessity to perform calculations for a wide range of regularization parameters in their refined approach can lead to runtimes of 20 processor hours. Similarly, the complexity of the parametrization of the spectrum used in Mishchenko's SOM and the associated extensive search space seem comparably expensive. Here we discuss an efficient and more accessible variant of this stochastic optimization method that also uses a Bayesian framework for the analytic continuation problem with only minimal prior information on the spectrum.

In many situations, one has only minimal knowledge of the prior probability of A, i.e., $p(A)$. Therefore, we assume that the prior distribution $p(A)$ is uniform and the posterior distribution is equivalent to the likelihood, i.e., $p(A|G) = p(G|A)$. The most straightforward way to construct an empirical distribution is to use the Markov chain Monte Carlo (MCMC) sampling method [13]. However, the dimension of the distribution is the partition number L of the frequency ω, which is typically large. In this case, the MCMC sampling method becomes very inefficient, especially in simulating the statistically insignificant region due to the low acceptance rate. To overcome this problem in MCMC-type methods, we propose an efficient scalable numerical algorithm that constructs an empirical distribution for $p(G|A)$ with independent random samples.

The central idea of our algorithm is to build the target probability distribution $p(A|G)$ by running several parallel optimization procedures. Toward that end, we compute J realizations of optimal spectral functions A based on QMC data G in a stochastic manner, and we use the distribution of all J realizations of stochastically optimized spectral functions to be a representation of the distribution for A. For each realization, the stochastic optimization aims to minimize the χ^2 error, and the random optimal spectral function will be very noisy due to the fact that the analytic continuation problem is ill-posed. Also, different realizations have very different features. However, since all random samples of the spectral function have very small χ^2 error, statistically they capture the feature of the true spectral function A, and the mean value of all the samples will be a good estimate for the final spectral function A.

Specifically, for any given initial guess of the spectral function $A(\omega)$, which we denote by $D(\omega)$, we introduce an initial partition Π_0 of the frequency axis defined by

$$\Pi_0 := \{\omega_l | a = \omega_0 \leqslant \omega_1 \leqslant \omega_2 \leqslant \cdots \leqslant \omega_l \leqslant \omega_{L-1} \leqslant \omega_L = b\}.$$

Here a is the lower boundary of the test frequency region and b is the upper boundary. In many cases, one is more interested in resolving features in the low-frequency region. Therefore, we let the partition step size grow exponentially with increasing absolute value of the frequency. (For example, we chose the step size 0.1 for the interval $[-2,2]$, 0.2 for $[-4,-2]$ and $[2,4]$, 0.4 for $[-6,-4]$ and $[4,6]$, and so on.). Thus the frequency step size $\Delta\omega_l$ is small for small $|\omega|$ and grows with $|\omega|$, so that we have better resolution in the more important low-frequency region.

With the initial guess $D(\omega_l)$ and the partition Π_0, we initialize R realizations of spectral functions A, denoted by $\tilde{A}_0^r$, $r = 1, \ldots, R$, with

$$\tilde{A}_0^r(\omega_l) = D(\omega_l), \quad l = 0,1,\ldots,L.$$

Each realization is initialized with the same $D(\omega_l)$, for which we typically choose a Gaussian in the absence of external information. For each realization starting with $\tilde{A}_0^r$, we run an independent stochastic optimization procedure to minimize the χ^2 error and update the simulated spectral function $\tilde{A}_i^r$ from iteration step i to $i+1$, where $i = 0,1,2,\ldots$. We find that if we run enough iteration steps, the final result does not depend on the initial function D.

Suppose we have the rth realization of the simulated spectral function at iteration step i, i.e., $\tilde{A}_i^r$. To find an optimal solution, we add a Gaussian process, denoted by the random vector of length L, $\lambda_i^r := (\lambda_i^r(\omega_1), \lambda_i^r(\omega_2), \ldots, \lambda_i^r(\omega_L))$, to $\tilde{A}_i^r$ and get a proposed spectral function

$$\tilde{A}_{i+\frac{1}{2}}^r = \frac{1}{I}\left(\tilde{A}_i^r + \lambda_i^r\right). \tag{13}$$

Here the constant I is chosen so that the spectrum $\tilde{A}_{i+\frac{1}{2}}^r$ is normalized, i.e., it satisfies

$$\sum_{l=1}^{L} \tilde{A}_{i+\frac{1}{2}}^r(\omega_l)\Delta\omega_l = 1. \tag{14}$$

In principle, the only constraint we impose on the random variables λ_i^r is that the proposal spectrum is positive-definite, i.e., $\tilde{A}_{i+\frac{1}{2}}^r \geqslant 0$. However, in order to allow for implicit regularization and to improve efficiency, we normally set the Gaussian process λ_i^r to a multivariate Gaussian random variable with mean zero and covariance C, which determines the smoothness of the noise λ_i^r as a function of frequency ω_l. If the correlation is strong (large C), the noise we add is smooth; for small C, the noise fluctuates strongly between neighboring frequencies. Thus, the covariance C may be considered a smoothing factor, which provides an implicit regularization. Since the partition step size on the frequency axis restricts the resolution of possible features in the spectral function, we let the covariance function C depend on the partition of the frequency. There are many choices of the covariance function, including constant, linear, squared exponential, Ornstein-Uhlenbeck, rational quadratic, or other forms. Here, we choose an Ornstein-Uhlenbeck form, i.e.,

$$C(\omega_{l_1}, \omega_{l_2}) = \exp(-\alpha|l_1 - l_2|), \tag{15}$$

where α is a positive constant. A popular choice of α is provided by the "maximum posteriori estimate," which is a mode of the posterior distribution [14]. Note that we let the noise correlations depend on the number of intermediate partition steps, $|l_1 - l_2|$, instead of the frequency directly. This means that the effective correlation between frequencies changes with frequency since the resolution of our frequency grid changes. It is small in the low-frequency region where the step size is small and the resolution is high, while the effective

correlation is high in the larger frequency region where the step size is large. This frequency adaptive noise correlation is consistent with the idea that finer structures are to be resolved in $A(\omega)$ in the more important low-frequency region, while stronger smoothing can take place in the higher-frequency region. In general, larger values of C will impose more smoothing on individual realizations $\tilde{A}^r(\omega_l)$ and thus reduce the number of realizations needed to obtain a smooth average $\bar{A}(\omega)$. Thus, the correlation parameter C may be used as a tuning parameter to balance the gain in details in $A(\omega)$ against an increase in computer time. Given the covariance matrix $C(\omega_{\ell_1},\omega_{\ell_2})$ in Eq. (8), we then generate the random vector λ_i^r from the L-dimensional joint normal distribution $N(0,C)$ with mean 0 and covariance C.

If the proposed spectral function $\tilde{A}^r_{i+\frac{1}{2}}$ fits the data better than the previous $\tilde{A}^r_i$, i.e., if $\chi^2[\tilde{A}^r_{i+\frac{1}{2}}] < \chi^2[\tilde{A}^r_i]$, we accept the update and set $\tilde{A}^r_{i+1} = \tilde{A}^j_{i+\frac{1}{2}}$. Otherwise, the update is rejected and $\tilde{A}^r_{i+1} = \tilde{A}^r_i$. Thus, the χ^2 error between the simulated Green's function $\tilde{G}$ and the QMC experimental data G will decrease monotonically. In our implementation, the optimization process is stopped in the jth iteration if $\chi^2(\tilde{A}^r_j) \leqslant \epsilon$ for a fixed threshold ϵ. $\tilde{A}^r := \tilde{A}^r_j$ then denotes the final spectral function for realization r. From a number R of independent stochastic optimization procedures, we obtain a set of random optimal spectral functions, i.e., $\{\tilde{A}^r\}_{r=1}^R$, which forms an empirical distribution for the spectral function A, denoted by $P_0(\tilde{A}|G)$. We note that the stochastic optimizations for different realizations are independent, which makes the algorithm scalable in the stochastic optimization procedure.

The threshold ϵ is a user-defined positive constant, which should be chosen according to the complexity of the problem. In practice, we keep ϵ of the same order as the variance of the QMC data in order to avoid overfitting of the data and to keep the efficiency of the optimization process high. In contrast to the SOM used by Mishchenko *et al.* [10], we do not allow for updates that increase χ^2. The fact that we use a random, global process to update the spectral function, however, provides a means to get out of local minima with high χ^2. In spite of this, the optimization process slows down at very small χ^2, because the probability of finding a better solution is small. As we observe in practice, keeping ϵ of similar size to the QMC errors ensures that the process does not become prohibitively inefficient.

The random optimal spectral function that results from a single realization may not capture all the important features in the true spectral function and will be noisy due to the fact that the problem is ill-posed. The weighted average, Eq. (5), of the different realizations, however, will be smooth if the number of realizations is large enough. Since we stop each optimization when χ^2 reaches the same value ϵ, the weights in Eq. (5) are all identical and the final spectrum $\bar{A}(\omega_n)$ is obtained from a simple average,

$$\bar{A}(\omega_l) = \frac{1}{R}\sum_{r=1}^{R} \tilde{A}^r(\omega_l). \tag{16}$$

Note that, in practice, we chose a maximum number S of update steps. If a particular optimization procedure for realization r does not reach $\chi^2 \leqslant \epsilon$ in S steps, the update process is stopped and the spectrum in step S, $\tilde{A}^r_S$, is used as the final result $\tilde{A}^r$ for this realization. In this case, we still use Eq. (16) to compute the final spectrum $\bar{A}(\omega_l)$ and assume that it is accurate enough. For the examples we considered in Sec. V, however, we find that χ^2 always reaches ϵ before S updates, so this is not an issue. We generally choose the number R of realizations large enough to get a smooth final solution for the spectrum. R is generally problem-specific and also depends on the parameter α in the correlation between neighboring frequencies, as will be discussed in Sec. V, Example 1.

It is important to point out that the original frequency partition Π_0 is not informed by the data G and thus is not adaptive to the features in the spectral function. Because of the stochastic nature of the SOM procedure, one has a representation for the data-informed distribution $\tilde{P}_0(A|G)$ in addition to the approximate spectral function $\bar{A}$. From the distribution, one can get the standard deviation for every single frequency in Π_0 and from that construct a confidence band for the estimated spectral function $\bar{A}$. A wide confidence band indicates large fluctuations in the different realizations, which may point to possible fine structure in the true spectral function A. Based on the width of the confidence band, one can then modify the frequency partitioning to allow the algorithm to resolve more detail in the solution. If the confidence band is wide in a certain frequency region, we use more partition points in that region, and conversely, if the confidence band is narrow, we use fewer partition points. We then rerun the stochastic optimization procedure with the modified frequency partition. The complete algorithm for fixed frequency partitioning is summarized in Algorithm 1.

Algorithm 1: Stochastic optimization method

Initialize Choose partition Π_0 for frequency grid ω_l, initial guess of the spectral function $D(\omega_l)$, sample size R, χ^2 error threshold ϵ, and optimization update step number S.

for $r = 1, 2, \ldots, R$
  Let $\tilde{A}^r_0 = D$ on partition Π_0
  while $1 \leqslant i < S$ **do**
    $i = i + 1$
    Propose $\tilde{A}^r_{i+\frac{1}{2}} = \tilde{A}^r_i + \lambda^r_i$
    Compute $\chi^2(\tilde{A}^r_i)$ and $\chi^2(\tilde{A}^r_{i+\frac{1}{2}})$
    if $\chi^2(\tilde{A}^r_{i+\frac{1}{2}}) \geqslant \chi^2(\tilde{A}^r_i)$
      $\tilde{A}^r_{i+1} = \tilde{A}^r_i$
    else
      $\tilde{A}^r_{i+1} = \tilde{A}^r_{i+\frac{1}{2}}$
    end if
    if $\chi^2(\tilde{A}^r_{i+\frac{1}{2}}) \leqslant \epsilon$
      $\tilde{A}^r = \tilde{A}^r_{i+1}$
      Break
    end if
  end while
end for

Approximate the empirical distribution $P(\tilde{A})$

Compute $\bar{A}$ from $P(\tilde{A})$ according to Eq. (16).

V. NUMERICAL EXAMPLES

In this section, we discuss the results of applying our FESOM variant to three different numerical examples to assess its effectiveness and compare with results obtained from the standard MaxEnt procedure. The first two examples are cases for which the exact spectrum $A(\omega)$ is known and different samples for the data $G(i\omega_n)$ are generated synthetically. The third case is a problem for which $G(i\omega_n)$ is generated from a QMC simulation of a single-band Hubbard model and the true spectrum $A(\omega)$ is unknown. For all three examples in this paper, we choose the FESOM regularization parameter for the noise correlations $\alpha = 0.5$.

A. Example 1

In this example, we consider a synthetic problem for which we make up a spectrum $A(\omega)$ with features similar to those expected for the electronic spectral function of a metallic system with a pronounced quasiparticle peak at the Fermi energy $\omega = 0$ (black line in Fig. 1). From this spectrum $A(\omega)$ we generate the input data $G(i\omega_n)$ using the Hilbert transform in Eq. (1) and setting the temperature $T = 0.1$. We then generate 1000 samples of $G(i\omega_n)$ by adding noise, i.e., the ith sample G_n^i is obtained as $G_n^i = [\sigma_{\text{noise}} N(0,1) + 1]G_n$, where $N(0,1)$ is noise drawn from a normal distribution with mean 0 and standard deviation 1, and σ_{noise} is the noise amplitude.

We first illustrate the behavior of individual realizations of the stochastically optimized spectral function $\tilde{A}(\omega)$ for input data with relatively low quality, which we generated using a noise amplitude $\sigma_{\text{noise}} = 0.1$. We have set the χ^2 threshold $\epsilon = 0.05$. In the top panel of Fig. 1, we compare the true spectrum $A(\omega)$ (black line) with five different realizations of $\tilde{A}(\omega)$ (blue dashed curves). As one can see, all realizations capture the central peak in the true spectral function, but different realizations have different features in the higher-frequency region. The bottom panel of Fig. 1 shows 100 different realizations of the stochastic optimal spectral function samples, again compared with the true spectral function $A(\omega)$. Here one sees again that the center peak is well captured by all samples with very little difference between the samples. In addition, the plots show that statistically, the samples capture the peak on the left as well as the fluctuations on the right.

As discussed in Sec. IV, we use a Gaussian process, in which the added noise is correlated between adjacent frequencies, to propose updates to the spectral function in the stochastic optimization procedure. This leads to proposals that are significantly smoother than what one would get if the noise added to different frequency points was uncorrelated. This improves the efficiency of the algorithm since fewer realizations are needed to get a smooth average $\bar{A}(\omega)$. This benefit, however, comes at the cost of losing possible fine-structure details, which are potentially flattened out by the correlated noise.

To illustrate the effect of this implicit regularization, we compare in Fig. 2 the simulated spectral function obtained by using correlated noise proposals (blue dash-dotted line) with that obtained from using noncorrelated noise proposals (magenta dash-dotted line). We have found that in the case of correlated noise proposals, 100 realizations are sufficient

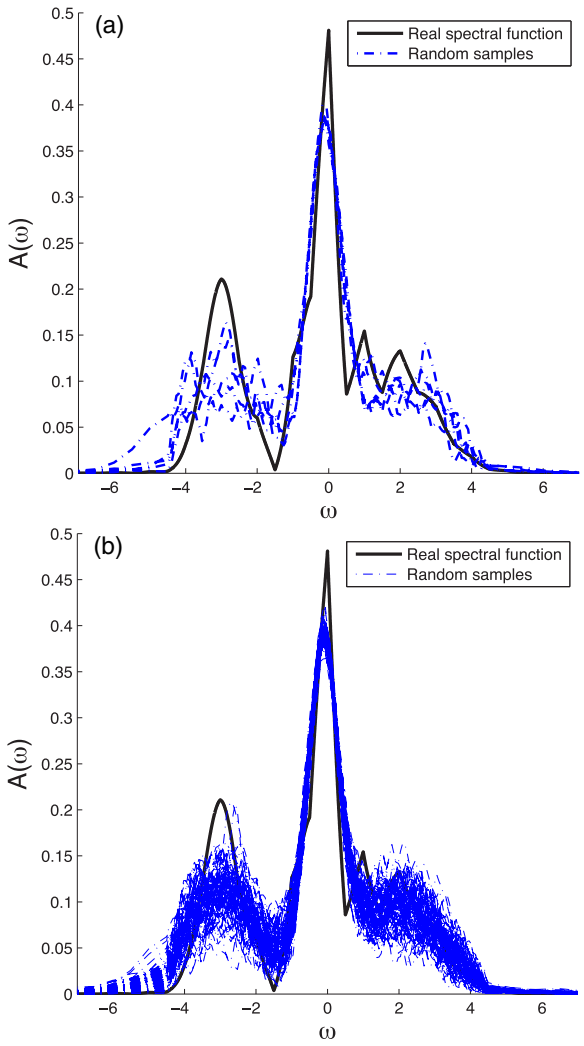

FIG. 1. Example 1. The synthetic spectral function $A(\omega)$ (black line) is used to generate different samples for the input data $G(i\omega_n)$ using Eq. (1) and compared to the results of 5 (a) and 100 (b) independent realizations of stochastic optimal spectral function samples (blue dashes curves). Here we have used a χ^2 threshold $\epsilon = 0.05$.

to give a smooth final average $\bar{A}(\omega)$, while the case of noncorrelated noise proposals required 500 realizations. As one can see from the green dash-dotted line, 100 realizations are not sufficient to give a smooth result when the noise proposals are uncorrelated. With correlated noise proposals, however, 100 realizations provide a smooth result for $\bar{A}(\omega)$, which is very similar to the result with uncorrelated noise and 500 realizations as well as to the true spectrum $A(\omega)$. It is also clear that in this case, the correlations in the noise do not result in any loss of detail in $\bar{A}(\omega)$.

To benchmark our FESOM variant against the state of the art, we compare in Fig. 3 the results of our approach with correlated noise and 100 realizations (blue dash-dotted line) with the spectrum obtained from the MaxEnt procedure described in Sec. III (red dash-dotted line). Here we have used the same 1000 generated samples of the input data $G(i\omega_n)$ in both FESOM and MaxEnt calculations and a Gaussian default

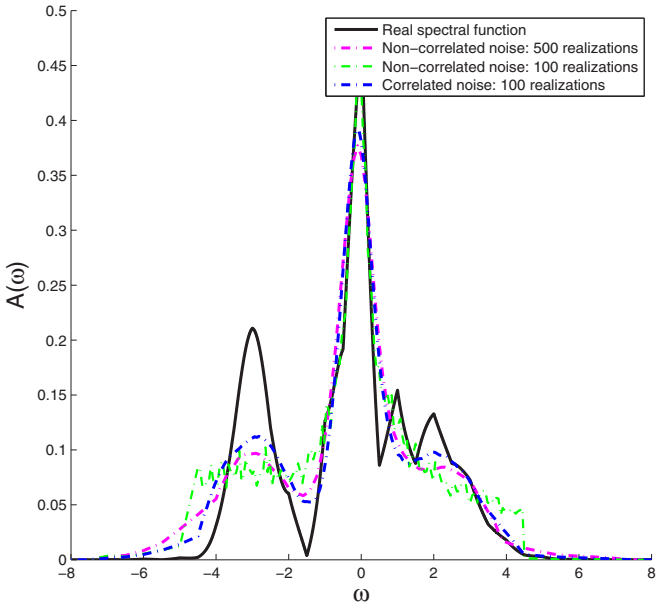

FIG. 2. Example 1. Averaged spectrum $\bar{A}(\omega)$ obtained from the stochastic optimization with 500 realizations and noncorrelated noise proposals ($\alpha = 0$, magenta dash-dotted line) compared with the result from 100 realizations and noncorrelated noise proposals ($\alpha = 0$, green dash-dotted line) and 100 realizations and correlated noise proposals ($\alpha = 0.5$, blue dash-dotted line). Here we have used a χ^2 threshold $\epsilon = 0.05$.

model for the MaxEnt. For this particular case of low-quality data, one sees that the MaxEnt result only captures the central peak, while the peaks at higher frequency on either side are washed out. Due to the large σ_{noise} of the data, the MaxEnt underfits the data and puts more weight on the entropy term $S[A]$ in order to minimize the deviation from the Gaussian

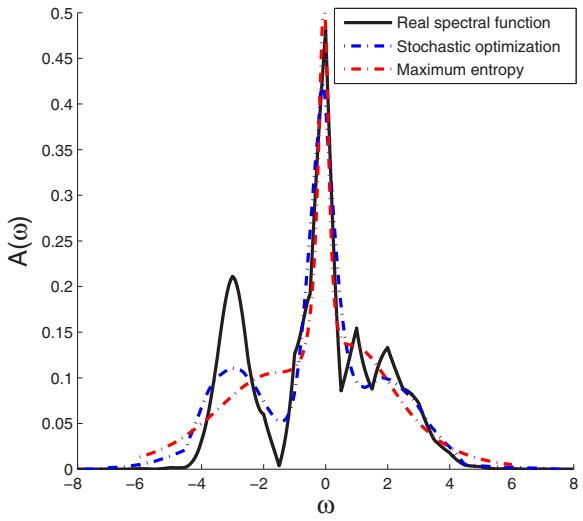

FIG. 3. Example 1, case 1. Comparison between the true spectrum $\bar{A}(\omega)$ (black solid curve) with the results of FESOM (blue dash-dotted line), using a χ^2 threshold $\epsilon = 0.05$, and MaxEnt (red dash-dotted line) for low-quality data. Here we have used a noise amplitude of 0.1 to generate the 1000 samples for the input data.

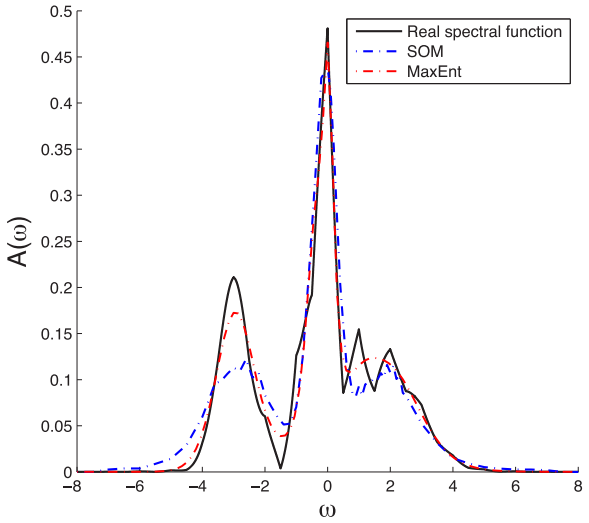

FIG. 4. Example 1, case 2. Comparison between the true spectrum $\bar{A}(\omega)$ (black solid curve) with the results of FESOM (blue dash-dotted line), using a χ^2 threshold $\epsilon = 0.05$, and MaxEnt (red dash-dotted line) for high-quality data. Here we have used a noise amplitude of 0.001 to generate the 1000 samples for the input data.

default model. In contrast, the FESOM method is able to resolve the higher-frequency structure reasonably well despite the low quality of the data.

To study the dependence on data quality in more detail, we have generated a second set of higher-quality input data for the same problem by setting the noise amplitude $\sigma_{\text{noise}} = 0.001$. As one sees from Fig. 4, here, the MaxEnt gives a much better result with very good resolution of the structures at higher frequencies. The smaller σ_{noise} forces a better χ^2 fit of the input data and less similarity with the default model. Figure 4 also shows the FESOM result for this case, for which we have used the same χ^2 threshold $\epsilon = 0.05$ as in the case of the lower-quality data displayed in Fig. 3. One sees that the FESOM result is almost identical to the case with lower quality. (We have also tried to run the FESOM optimization with a smaller threshold $\epsilon = 0.001$, but we found that this leads to prohibitively long runtimes.) We stress that the inferiority of our FESOM result relative to the MaxEnt in the case of high-quality data does not necessarily reflect a general disadvantage of the traditional SOM framework, but it is likely a result of the simplifications we introduced in our variant to make the algorithm more efficient. Returning to the case of weak data in Fig. 3, we conclude that for cases of low-quality data, for which the MaxEnt procedure tends to underfit the data in the absence of a good default model, the FESOM approach can provide results that capture the true spectral function in much more detail.

As noted, one strength of the stochastic optimization is that one has information of the confidence interval for all frequencies, while MaxEnt only allows us to determine the uncertainty of the solution integrated over a finite frequency interval [11]. In Fig. 5, we plot the FESOM simulated spectral function for the low-quality input data with its 95% confidence region. We can see from the figure that in the frequency region [0.5,3] there are two small peaks in the true spectral function, and neither FESOM nor MEM could resolve both peaks well.

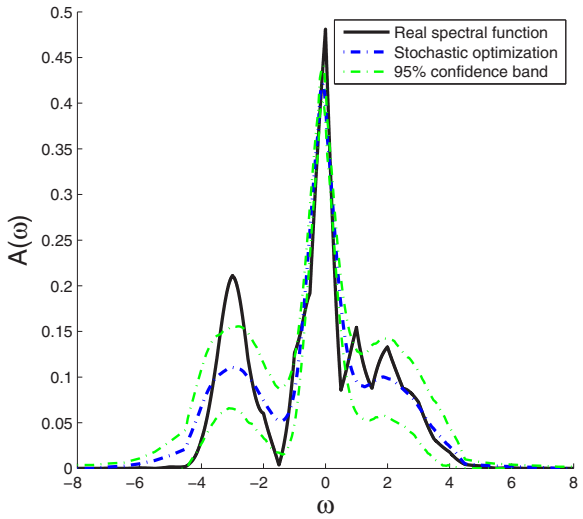

FIG. 5. Example 1. True spectrum $\bar{A}(\omega)$ (black solid line), FESOM simulated spectrum $\bar{A}(\omega)$ (blue dash-dotted line), and 95% confidence band in FESOM (green dash-dotted line). Here we have used a χ^2 threshold $\epsilon = 0.05$.

However, the large width of the FESOM confidence band in this region indicates strong fluctuations in the different realizations, which in turn could be a signal of possible fine structure in the true solution.

B. Example 2

We now consider an example for which the true spectral function $A(\omega)$ is known and generated from a simulation of a microscopic model, and the input data $G(i\omega_n)$ are again calculated from $A(\omega)$ through Eq. (1). Specifically, we consider a two-dimensional Hubbard model on a square lattice with nearest-neighbor hopping t and Coulomb repulsion U described by the Hamiltonian

$$H = -t \sum_{\langle ij \rangle} c^{\dagger}_{i\sigma} c_{j\sigma} + U \sum_{i} n_{i\uparrow} n_{i\downarrow}. \tag{17}$$

Here, $c^{\dagger}_{i\sigma}$ creates and $c_{i\sigma}$ destroys an electron with spin $\sigma = \uparrow, \downarrow$ on site i, and $n_{i\sigma} = c^{\dagger}_{i\sigma} c_{i\sigma}$ is the corresponding number operator. We use the dynamical mean-field theory (DMFT) [15] together with a noncrossing approximation (NCA) [16] to obtain the local spectral function $A(\omega)$ in the antiferromagnetic state. For the majority spin, the local spectral function $A(\omega)$ we obtain is shown as the black line in Fig. 6. Here we have used $U = 16t$ and set the filling to $\langle n \rangle = 0.95$ and temperature $T = 0.29$. One sees the lower and Hubbard bands at negative and positive frequencies, respectively, as well as the fine structure with multiple peaks in the lower Hubbard band. These resonances reflect the bound states of a hole propagating in an antiferromagnetic background [17].

From this $A(\omega)$, we again generated 1000 samples of the input data $G(i\omega_n)$ via Eq. (1) as in the previous example by adding noise with standard deviation 0.001. The same samples were then used in MaxEnt and FESOM to calculate an estimate of $A(\omega)$. The χ^2 threshold we have used for the FESOM simulation was set to $\epsilon = 0.001$.

Figure 6 compares the MaxEnt result (left panel) and the FESOM result (middle panel) with the true spectrum. Here one sees that both approaches capture the lower and upper Hubbard bands equally well. It is obvious that the MaxEnt has difficulty resolving the fine structure in the lower Hubbard band at negative frequencies. It captures the first dominant peak at $\omega = 0$, but it fails to reproduce the multiple peaks at lower frequencies. In comparison, the FESOM estimate also has the leading peak, but in addition it shows fluctuations at lower (negative) frequencies, reminiscent to some extent of the multipeak structure in the true $A(\omega)$. These fluctuations are also seen in the FESOM 95% confidence band plotted in the right panel, indicating their presence in a large fraction of the FESOM realizations. Furthermore, the large width of the confidence band in this region is a further sign of the fine structure that is present in the true solution. At higher negative frequencies, however, the FESOM algorithm finds an artificial peak near $\omega = -9$, while MaxEnt correctly predicts a smooth result in this region.

C. Example 3

We now turn to a real data problem, for which the input data $G(i\omega_n)$ are generated in a QMC simulation and the true spectral function $A(\omega)$ is not known. We again solve the 2D Hubbard model of Eq. (17), but instead of DMFT with an NCA impurity solver we now use a dynamic cluster approximation (DCA) QMC algorithm [18,19], which allows for the inclusion of nonlocal correlations in addition to the local correlations treated in the DMFT. This is accomplished by mapping the lattice model onto an effective cluster problem embedded in a dynamic mean-field host that is designed to represent the rest of the system and determined self-consistently. To solve the effective cluster problem, we use the continuous-time auxiliary-field QMC algorithm by Gull *et al.* [20].

For this example, we have set the Coulomb interaction $U = 8t$, the site filling $\langle n \rangle = 0.95$, and the temperature $T = 0.08t$, and we have used a four-site 2×2 cluster for the DCA calculation. From previous calculations, it is known that these parameters give a local spectral function $A(\omega)$ with a pseudogap [21]. $A(\omega)$ is partially suppressed at $\omega = 0$, reminiscent of the normal-state pseudogap phase of the underdoped cuprate superconductors [22].

After the mean-field host is converged, we performed one additional iteration in which measurements of $G(i\omega_n)$ are performed and partitioned into 100 bins with a bin size of 100 000 measurements each. For the MaxEnt procedure, we diagonalized the covariance matrix and rotated the data and the kernel into a diagonal frame. Moreover, we used the annealing technique [11], in which the MaxEnt is performed for a set of decreasing temperatures and the resulting spectrum is used as a default model for the next lower temperature. The same set of 100 samples of $G(i\omega_n)$ is then used in both the MaxEnt and FESOM to determine an estimate of the spectral function $A(\omega)$. For the FESOM analytic continuation, we have set the χ^2 threshold $\epsilon = 0.001$.

In the left panel of Fig. 7, we compare the simulated spectral function $A(\omega)$ obtained from the MaxEnt (red solid line) with that of the FESOM calculation (blue dashed line). One again sees the two Hubbard bands centered below and above

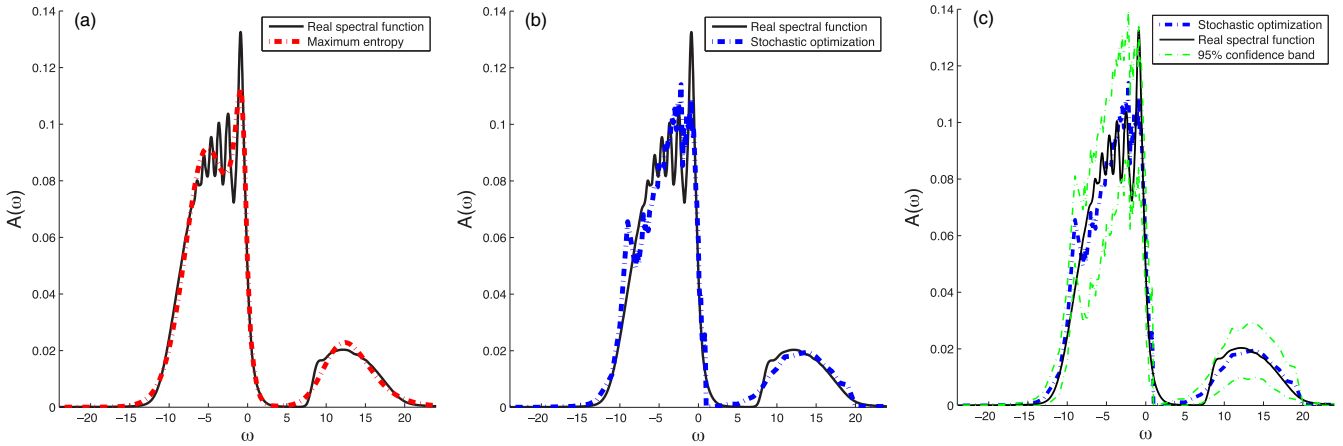

FIG. 6. Example 2. Comparisons between the true spectrum (black solid line) and the results of MaxEnt (a) and FESOM (b), for which we have used a χ^2 threshold $\epsilon = 0.001$. Panel (c) shows the 95% confidence band obtained in the FESOM. The real spectral function is obtained from a DMFT/NCA calculation of a 2D Hubbard model in the antiferromagnetic state with $U = 16t$, $\langle n \rangle = 0.95$, and $T = 0.29$.

$\omega = 0$ and split by $\sim U = 8t$. For this case, the MaxEnt result clearly shows more structure in the lower Hubbard band. Both the MaxEnt and the FESOM resolve the pseudogap feature, manifested as the dip in $A(\omega)$ at $\omega = 0$. But it is much better developed in the MaxEnt than in the FESOM $A(\omega)$. In addition, the MaxEnt result displays a shoulder at $\omega \sim -4$, which is not present in the FESOM result.

The right panel of Fig. 7 displays the 95% confidence band obtained from the FESOM simulation. The band is unusually wide even at small $|\omega|$, indicating large fluctuations in the different realizations. However, once again one sees that the confidence band follows the same trend as the mean spectrum $\bar{A}(\omega)$. This shows that the pseudogap feature is present in a large fraction of the FESOM realizations and is therefore likely a feature of the true spectral function.

We also tried the annealing technique for the FESOM simulation. We did find faster convergence of the optimization procedure in the last step of the annealing procedure. However, there was no change in the resulting spectrum. Considering that a separate optimization has to be carried out for each temperature, the annealing method is less efficient than just running a single optimization at the lowest temperature, and, in contrast to the MaxEnt, it does not provide any improvement in the solution.

VI. SUMMARY AND CONCLUSIONS

To summarize, we have introduced, analyzed, and benchmarked against maximum entropy a fast and efficient variant of the stochastic optimization method introduced by Mishchenko *et al.* [10], called FESOM, for the analytical continuation of imaginary frequency QMC data $G(i\omega_n)$, an ill-posed problem that remains a challenging barrier in connecting theory with experiment. In contrast to the routinely used maximum entropy method, which introduces a default model to regularize the problem, the stochastic optimization method only uses minimal prior information for the quantity of

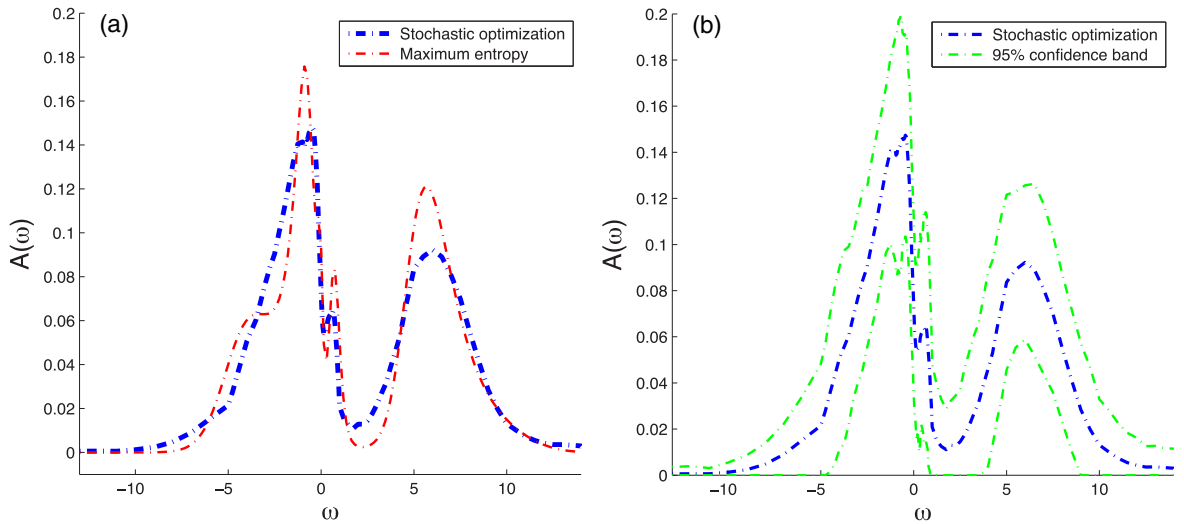

FIG. 7. Example 3. Spectral function $A(\omega)$ obtained from MaxEnt (a) and FESOM (b) (using a χ^2 threshold of $\epsilon = 0.001$) analytic continuation of DCA QMC data for a 2D Hubbard model with $U = 8t$, $\langle n \rangle = 0.95$, and $T = 0.08t$.

interest, the spectral function $A(\omega)$, and it does not introduce a default model. The basic idea of the SOM approach is to use several parallel optimization procedures to obtain a large set of equally likely estimates of the spectrum and determine the final spectral function $A(\omega)$ as an average over these samples. The optimization procedure minimizes the χ^2 misfit between the QMC data and the modeled data by sequentially and randomly proposing global changes to a test spectral function. A combination of three characteristics of our FESOM implementation results in a more efficient and less complex algorithm than the previous implementation: (i) It uses a fixed frequency grid just like MaxEnt. (ii) It only allows for proposal updates that lower χ^2, and it does not permit temporary increases. (iii) It uses a Gaussian process to update the spectral function, in which the noise added to adjacent frequencies is correlated. Characteristic (iii) can be viewed as an implicit regularization that results in much smoother individual estimates and therefore a smaller number of realizations required to get a smooth average.

We have applied this algorithm to three representative test-case problems and compared the results with those obtained from MaxEnt. For two of these problems, the true spectral function $A(\omega)$ was known and used to generate a noisy set of input data $G(i\omega_n)$. For the third case, we used QMC data for $G(i\omega_n)$ obtained from DCA simulations of a 2D Hubbard model. For these problems, our FESOM algorithm generally yielded similar spectra to those obtained from MaxEnt. For good-quality data with weak noise, we found that the MaxEnt procedure gives much better results than the FESOM method, while for poor-quality data the situation is reversed. In this case, the MaxEnt tends to underfit the data, while the FESOM procedure gives a much better result. Generally, we found that in contrast to MaxEnt, the quality of the FESOM depends very little on the quality of the input data. In addition, the FESOM provides information on the confidence of the resulting spectral function $A(\omega)$ for each frequency ω, in contrast to MaxEnt, which only gives this information for a finite interval in frequency.

For the test-case problems we have studied, the stochastic optimization required on average about 1–2 min on a single core (2.2 GHz Intel Core i7) to optimize a single realization and a total runtime of $\sim$2 core hours to produce the final spectrum as the average of 100 realizations. While the total runtime is about an order of magnitude longer than that of the MaxEnt procedure with annealing, per realization it is roughly of the same order. Trivial parallelization of the stochastic optimization over different realizations will therefore result in similar runtimes. We therefore believe that our implementation of the stochastic optimization technique provides a viable alternative to the MaxEnt procedure for the analytic continuation of QMC data, especially for cases with poor data quality.

ACKNOWLEDGMENTS

This research was sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the US Department of Energy. V.W.S. acknowledges support from AFOSR (Grant No. FA9550-15-1-0445) and ARO (Grant No. W911NF-16-1-0182).

[1] S. F. Gull and J. Skilling, IEEE Conf. Proc. **131F**, 646 (1984).
[2] R. N. Silver, J. E. Gubernatis, D. S. Sivia, and M. Jarrell, Phys. Rev. Lett. **65**, 496 (1990)
[3] R. N. Silver, D. S. Sivia, and J. E. Gubernatis, Phys. Rev. B **41**, 2380 (1990).
[4] J. E. Gubernatis, M. Jarrell, R. N. Silver, and D. S. Sivia, Phys. Rev. B **44**, 6011 (1991).
[5] M. Jarrell and J. E. Gubernatis, Phys. Rep. **269**, 133 (1996).
[6] N. V. Prokof'ev and B. V. Svistunov, JETP Lett. **97**, 649 (2013).
[7] A. W. Sandvik, Phys. Rev. B **57**, 10287 (1998).
[8] K. Beach, arXiv:cond-mat/0403055.
[9] S. Fuchs, T. Pruschke, and M. Jarrell, Phys. Rev. E **81**, 056701 (2010).
[10] A. S. Mishchenko, N. V. Prokofév, A. Sakamoto, and B. V. Svistunov, Phys. Rev. B **62**, 6317 (2000); A. S. Mishchenko in *Correlated Electrons: From Models to Materials*, edited by E. Pavarini, E. Koch, F. Anders, and M. Jarrell (Forschungszentrum Jülich, Jülich, 2012), Chap. 14.
[11] M. Jarrell, in *Correlated Electrons: From Models to Materials*, edited by E. Pavarini, E. Koch, F. Anders, and M. Jarrell (Forschungszentrum Jülich, Jülich, 2012), Chap. 13.
[12] R. K. Bryan, Eur. Biophys. J. **18**, 165 (1990).
[13] W. R. Gilks, Encyclopedia of Biostatistics 4 (2005).
[14] H. W. Sorenson, *Parameter Estimation — Principles and Problems* (Marcel Dekker, New York and Basel, 1980).
[15] A. Georges, W. Krauth, and M. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).
[16] N. E. Bickers, D. L. Cox, and J. W. Wilkins, Phys. Rev. B **36**, 2036 (1987).
[17] R. Strack and D. Vollhardt, Phys. Rev. B **46**, 13852 (1992).
[18] M. H. Hettler, A. N. Tahvildar-Zadeh, M. Jarrell, T. Pruschke, and H. R. Krishnamurthy, Phys. Rev. B **58**, R7475 (1998).
[19] T. Maier, M. Jarrell, T. Pruschke, and M. Hettler, Rev. Mod. Phys. **77**, 1027 (2005).
[20] E. Gull, P. Werner, and O. Parcollet, Europhys. Lett. **82**, 57003 (2008).
[21] M. Jarrell, T. Maier, M. H. Hettler, and A. Tahvildarzadeh, Europhys. Lett. **56**, 563 (2001).
[22] M. R. Norman, D. Pines, and C. Kallin, Adv. Phys. **54**, 715 (2005).