Analytical continuation of imaginary axis data for optical conductivity

O. Gunnarsson,[1] M. W. Haverkort,[1] and G. Sangiovanni[2]

[1]*Max-Planck-Institut für Festkörperforschung, D-70506 Stuttgart, Germany*
[2]*Institut für Festkörperphysik, Technische Universität Wien, Vienna, Austria*

We compare different methods for performing analytical continuation of spectral data from the imaginary time or frequency axis to the real frequency axis for the optical conductivity $\sigma(\omega)$. We compare the maximum entropy (MaxEnt), singular value decomposition (SVD), sampling, and Padé methods for analytical continuation. We also study two direct methods for obtaining $\sigma(0)$. For the MaxEnt approach we focus on a recent modification. The data are split up in batches, a separate MaxEnt calculation is done for each batch and the results are averaged. For the problems studied here, we find that typically the SVD, sampling, and modified MaxEnt methods give comparable accuracy while the Padé approximation is usually less reliable.

PACS number(s): 72.15.Eb, 02.70.Ss

I. INTRODUCTION

For strongly correlated systems analytical methods usually involve uncontrolled approximations. Therefore stochastical methods such as quantum Monte Carlo (QMC),[1] quantum cluster methods,[2] or continuous time methods[3] are often used. Apart from statistical errors, such methods can produce quite accurate results but the results are obtained on the imaginary axis. A major problem is then the analytically continuing of the results to the real axis, which is an ill-posed problem. Small changes in the data on the imaginary axis can lead to large changes on the real axis. Since the imaginary axis data contain statistical noise, the analytical continuation is very difficult.

There are different ways of regularizing this ill-posed problem. One method combines the Bayesian theory with the maximum entropy (MaxEnt) approach, which has been found to be an efficient method for analytical continuation.[4,5] Other regularizations are used in the singular value decomposition (SVD) (Ref. 6 and 7) or stochastic regularization[8] methods. An alternative is provided by making a Padé approximation to the data as a function of imaginary frequency and then analytically continue the Padé expression to real frequencies.[9,10] A rather different approach is to use sampling methods, where a large number of spectra are added, weighted by the probability that they correspond to the imaginary axis data. Such methods have been proposed for $T=0$ (Ref. 11) and finite T.[12] Finally, there are simple approximate methods for obtaining the optical conductivity at zero frequency, $\sigma(\omega=0)$, directly from imaginary time or frequency data.

Two-particle correlation functions, such as the dynamical spin or charge correlation functions or the optical conductivity, provide important information about a variety of properties of the system. These two-particle functions are much more difficult to calculate in QMC-type frameworks than the one-particle Green's function,[13] and therefore much of the interest has focused on the electron Green's function. Here we therefore instead treat a two-particle function, the optical conductivity. While we here focus on transformation of QMC data from imaginary space to real space, we note that there are also QMC methods giving results directly for real frequencies.[14]

In this paper we compare the Padé, SVD, sampling, and MaxEnt methods for obtaining the optical conductivity from imaginary axis data. We define a frequency-dependent optical conductivity, $\sigma(\omega)$, where ω is a real frequency. This we refer to as the "exact" result. This $\sigma(\omega)$ can easily be transformed to the imaginary axis since this is a well-behaved transformation that can be performed with a high accuracy. We add statistical noise to the data, which then simulate the output of a QMC calculation. The data are then transformed back to the real axis, using the various methods for analytical continuation. If the methods work well, we should essentially recover the starting $\sigma(\omega)$, the exact result. This way we can judge the accuracy of the different methods. It is important to compare with a known exact result since analytical continuation methods can give spurious structures due to noise in the data. If a certain method A gives more structures than another method B, it is hard to judge whether these additional structures are real and method A is better or they are due to noise and method B is better. This problem is avoided if exact results are known. Here we construct the exact $\sigma(\omega)$ using results for the two-dimensional (2D) Hubbard model as a guide for the general shape.

We find that the SVD, sampling, and MaxEnt methods tend to give comparable accuracy while the Padé approximation often gives worse results. In particular, the Padé approximation often overestimates $\sigma(0)$. One of the direct methods for estimating $\sigma(0)$ [based on Eq. (6) in Sec. II] underestimates $\sigma(0)$, in particular, for a narrow Drude peak while the other [extrapolating Eq. (5) in Sec. II to $\nu=0$] typically gives better results.

In Sec. II we present some general results for the optical conductivity. The different methods for analytical continuation are presented in Sec. III and the results are show in Sec. IV.

II. OPTICAL CONDUCTIVITY AND CURRENT-CURRENT CORRELATION FUNCTION

The optical conductivity $\sigma(\omega)$ is obtained from the current-current correlation function

$$\Pi(\tau) = \frac{1}{3N}\langle \mathbf{j}(\tau) \cdot \mathbf{j}(0)\rangle, \tag{1}$$

where N is the number of sites, $\mathbf{j}$ is the current operator, $\mathbf{j}(\tau) = \exp(H\tau)\mathbf{j}\exp(-H\tau)$, τ is imaginary time and $\langle \cdots \rangle$ is the thermodynamic average. We then have (setting $\hbar = k_B = 1$)

$$\Pi(\tau) = \int_{-\infty}^{\infty} K(\tau, \omega)\sigma(\omega)d\omega, \tag{2}$$

where

$$K(\tau, \omega) = \frac{1}{\pi}\frac{\omega e^{-\tau\omega}}{1 - e^{-\beta\omega}} \tag{3}$$

is a bosonic kernel and $\beta = 1/T$. Alternatively, we can relate $\sigma(\omega)$ to the Fourier transform $\Pi(\nu)$ of $\Pi(\tau)$,

$$\Pi(\nu) \equiv \int_0^\beta e^{i\nu\tau}\Pi(\tau)d\tau = -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\omega}{i\nu - \omega}\sigma(\omega)d\omega$$

$$= \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\omega^2}{\nu^2 + \omega^2}\sigma(\omega)d\omega, \tag{4}$$

where we have used that $\sigma(\omega) = \sigma(-\omega)$ and $\nu = \nu_i = i\nu_0$ is a multiple of $\nu_0 = 2\pi T$. For large ν we have that $\pi(\nu) \sim \nu^{-2}$. This result can also be rewritten as

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\nu}{\omega^2 + \nu^2}\sigma(\omega)d\omega = \frac{\Pi(0) - \Pi(\nu)}{\nu} \equiv \gamma(\nu). \tag{5}$$

This provides a convolution of the optical conductivity with a Lorentzian with the width ν. In particular, if $\sigma(\omega)$ has little variation over an energy range on the order of ν_0, $\gamma(\nu = \nu_0)$ provides an estimate of $\sigma(0)$. This estimate can be improved by extrapolating $\gamma(\nu)$ to $\nu = 0$, as discussed below. Alternatively, we can use[12]

$$\sigma(0) \approx \frac{\beta^2}{\pi}\Pi\left(\tau = \frac{\beta}{2}\right), \tag{6}$$

which is accurate if $\sigma(\omega)$ has little variation over an energy range on the order of $0.69(2\pi T)$.

Typical results for $\Pi(\nu)$ are shown in Fig. 1 for a 2D Hubbard model on a square lattice with nearest ($t = -0.4$ eV) and second nearest ($t' = 0.12$ eV) hopping. The Coulomb interaction is $U = 3.2$ eV and $\beta = 15$ eV^{-1}. This gives the occupancy 0.95. To obtain a conductivity, we have assumed that such 2D sheets are stacked on top of each other with a distance $c = 6.6$ Å appropriate for La$_{2-x}$Sr$_x$CuO$_4$. The data were obtained from a calculation in the dynamical cluster approximation (DCA) (Ref. 2) using a cluster with eight sites.

From Eq. (4) we can see that $\Pi(\nu) \sim \nu^{-2}$ for large ν. In Fig. 1 we show $\Pi(\nu)\nu^2$, which indeed saturates for large ν. From Eq. (4) we expect this to happen when ν is much larger than a typical energy scale of $\sigma(\omega)$, which in this case has peaks at $\omega = 0$ and $\omega \approx \pm U = \pm 3.2$ eV. In agreement with this, Fig. 1 shows saturation for ν on the order of 10 eV. For larger values of ν, $\Pi(\nu)$ essentially just gives information about $\int \omega^2 \sigma(\omega)d\omega$. The figure also shows results for $\Pi_0(\nu)$,
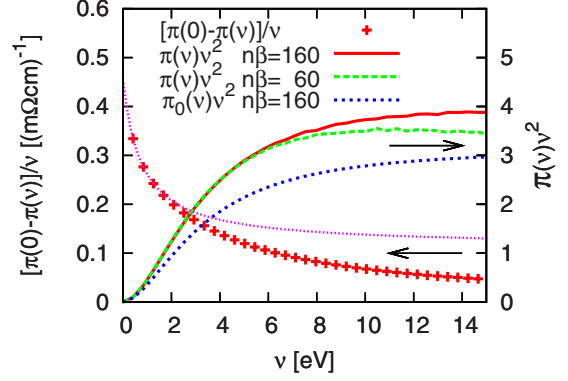

FIG. 1. (Color online) The current-current correlation function $\pi(\nu)$ multiplied by ν^2 as a function of ν as well as the function $\pi_0(\nu)$ calculated without vertex corrections. Results are shown for a large number $n\beta = 160$ time slices as well as for a small $n\beta = 60$, illustrating the resulting poor accuracy for large ν in the latter case. The figure also shows $\gamma(\nu) = [\pi(0) - \pi(\nu)]/\nu$ [Eq. (5)] and its extrapolation to $\nu = 0$ (thin dotted line), which provides an estimate of $\sigma(0)$.

which is calculated neglecting all vertex corrections. Π_0 is then obtained simply as a product (bubble) of two (dressed) Green's functions. Even for large ν, Π, and Π_0 are different. This can be understood from Eq. (4). Although both behave as ν^{-2}, the prefactor is different.

Figure 1 also shows $[\Pi(0) - \Pi(\nu)]/\nu$ [Eq. (5)], providing an estimate of $\sigma(0)$. To improve this estimate we extrapolate to $\nu = 0$. For small values of ν, $\Pi(\nu)$ depends mainly on $\sigma(w)$ for small ω. We allow for the possibility that $\sigma(\omega)$ has a Drude like peak at $\omega = 0$ by using the Ansatz

$$\sigma(\omega) = a + b\frac{\Gamma/\pi}{\omega^2 + \Gamma^2}, \tag{7}$$

where we have also added a constant a. In Fig. 1 we have fitted this form to the results for the lowest three nonzero values of ν. This extrapolation greatly improves the estimate, as can be seen from the examples below.

III. METHODS

A. Padé approximation

In the Padé approximation a function $f(z)$ in the complex plane, z, is described as the ratio between to polynomials $P(z)$ and $Q(z)$, $f(z) = P(z)/Q(z)$. The function is fitted to the output of a QMC calculation so that the results for certain imaginary frequencies ν_n are reproduced exactly. The analytical continuation is then performed by evaluating the function on the real axis. In the context of Green's functions this has, in particular, been used by Vidberg and Serene.[9] They fit to N data points, using a construction which for an even N leads to a polynomial Q which is one order higher than P so that P/Q behaves as $1/z$ for large z. This is appropriate for Green's functions, considered by them, but not necessarily for the response functions considered here, which behave as $1/z^2$ for large z. We have therefore constructed a Padé approximation where Q is two orders higher than P, which is

used in the following. This construction requires N to be odd. For the special case considered by Vidberg and Serene there are simple formulas for generating the polynomials[9] while this is somewhat more complicated in the general case.[10]

In fitting the P and Q to N data points, we have used data for one negative frequency, $-\nu_0 \equiv -2\pi T$, and the $N-1$ lowest nonnegative frequencies. This typically gives more stable results than using only nonnegative frequencies. On the other hand, using positive and negative frequencies symmetrically tends to put poles close to the real axis and gives poor spectra on the real axis. One negative frequency therefore often appears to be a good compromise.

B. Singular value decomposition

A widely used technique for inverse problems is the SVD.[6,7] Here we essentially follow Creffield et al.,[7] except that we work in imaginary frequency space rather than in imaginary time space, for reasons discussed Sec. IV. In the SVD method, the real frequency space is spanned by a set of eigenvectors. The kernel in Eq. (4) is discretized, giving

$$\Pi(\nu_i) = \sum_{j=1}^{N_\omega} K_{ij}\sigma(\omega_j), \quad i = 1, N_\nu. \tag{8}$$

If the data for different imaginary frequencies ν_i are uncorrelated, as is the case here, we introduce the eigenfunctions of the operator $KK^\dagger$,

$$\sum_{j=1}^{N_\omega}\sum_{l=1}^{N_\nu} K_{ij}K_{lj}^* v_l^k = \alpha_k^2 v_i^k, \quad i = 1, N_\nu. \tag{9}$$

We introduce vectors u_k

$$K^\dagger v^k = \alpha_k u^k, \tag{10}$$

which satisfy

$$Ku^k = \alpha_k v^k. \tag{11}$$

The spectral function can then be expanded as

$$\sigma(\omega_j) = \sum_{k=1}^{N_\nu} \frac{1}{\alpha_k} u_j^k \sum_{i=1}^{N_\nu} (v_i^k)^* \Pi(\nu_i). \tag{12}$$

This expansion is very ill behaved since some of the eigenvalues are very small. The expansion is therefore truncated so that only eigenvalues are considered for which

$$\alpha^k/\alpha_1 > \sigma_0, \tag{13}$$

where α_1 is the largest eigenvalue and σ_0 is the accuracy of the data. In this way we only consider the n_ν eigenvectors with the largest eigenvalues. To further improve the method, the kernel K is multiplied by a "support" function, which is equal to one in the range where $\sigma(\omega)$ is expected to be large and vanishes smoothly outside this region. Here we have used the function $1/[1+(\omega/\omega_0)^8]$, where $\omega_0 = 5$ was used.

C. Maximum entropy

A popular method for analytical continuation is the MaxEnt method.[4] This method is based on Bayes's theorem[15]

$$\Pr[\sigma,\Pi] = \Pr[\sigma|\Pi]\Pr[\Pi] = \Pr[\Pi|\sigma]\Pr[\sigma], \tag{14}$$

where $\Pr[\sigma,\Pi]$ is the joint probability that the spectral function is $\sigma(\omega)$ and that the QMC calculation gives the correlation function $\Pi(\nu)$. While the MaxEnt method usually is formulated for imaginary time τ, we here formulate it for imaginary frequency ν, for reasons discussed in Sec. IV. $\Pr[\sigma|\Pi]$ is the conditional probability that the spectral function is $\sigma(\omega)$ provided that the correlation function $\Pi(\nu)$ was obtained from the QMC calculation. From this one obtains[16]

$$\Pr[\sigma|\Pi] = \frac{\Pr[\Pi|\sigma]\Pr[\sigma]}{\Pr[\Pi]}. \tag{15}$$

This rewriting converts the ill-posed problem of determining $\sigma(\omega)$ given $\Pi(\nu)$ into the much easier problem of determining $\Pi(\nu)$ given $\sigma(\omega)$. $\Pr[\Pi]$ is a normalization factor, which is independent of $\sigma(\omega)$, and therefore is no complication. The remaining issue is then how to choose $\Pr[\sigma]$, which represents our prior knowledge about $\sigma(\omega)$. If we put this probability to a constant and then maximize the likelihood function $\Pr[\Pi|\sigma]$ the result is typically very bad, resulting in a saw-tooth type of spectra.[16] In MaxEnt one therefore defines the prior probability in terms of a maximum entropy function

$$S = \int d\omega \left\{ \sigma(\omega) - m(\omega) - \sigma(\omega)\ln\frac{\sigma(\omega)}{m(\omega)} \right\}, \tag{16}$$

where $m(\omega)$ is a default model. Other definitions are also possible.[17] In the MaxEnt method the quantity

$$\Pr[\Pi|\sigma]e^{\alpha S} \tag{17}$$

is maximized, using an appropriate value for α.[4] Here we have chosen α according to the classic MaxEnt method, using a flat prior for α.[4]

We sometimes find that this approach leads to unphysical oscillations in $\sigma(\omega)$. We have shown that the reason is that the MaxEnt method sometimes chooses an α which attaches too much significance to the noise in the data. This problem can be avoided by using a modification[18] of the MaxEnt method. We split the data for $\Pi(\nu)$ in several batches and perform a MaxEnt calculation for each batch. These results are then averaged. Typically, but not always, this leads to better results than averaging the data sets and then performing just one MaxEnt calculation for the average.[18]

The choice of default model can influence the outcome substantially. Here we have chosen a "reasonable" but structureless model (see Sec. IV). Using a model more similar to the actual spectrum improves the result. If the spectrum is calculated for several T, the result for a higher T can be used as the default model for a lower T. This can improve the results without introducing undue bias. Since we only consider one T here, we have not followed that approach.

D. Sampling method

The MaxEnt method avoids the saw-tooth problem but the definition of entropy requires the introduction of a default model, which can bias the output. Instead we can average[16] over $\Pr[\sigma|\Pi]$,

$$\langle \sigma \rangle = \int \sigma \, \mathrm{Pr}[\sigma | \Pi] D\sigma, \tag{18}$$

where $D\sigma$ indicates a functional integral over all $\sigma(\omega)$. $\mathrm{Pr}[\sigma | \Pi]$ is given by Eq. (15), where we furthermore put $\mathrm{Pr}[\sigma] \equiv \mathrm{const}$ for all non-negative $\sigma(\omega)$. Thus we assume that this is our only prior knowledge of $\sigma(\omega)$. In Ref. 12 we worked in imaginary time space. For reasons discussed in Sec. IV we here work in imaginary frequency space. Then the likelihood function is given by[16]

$$\mathrm{Pr}[\Pi | \sigma] = \frac{1}{\Pi_{i=1}^{n_\nu}(2\pi\tilde{\sigma}_i)}$$
$$\times \exp\left\{ -\sum_{i=1}^{n_\nu} [\Pi(\nu_i) - \Pi_\sigma(\nu_i)]^2 / (2\tilde{\sigma}_i^2) \right\}, \tag{19}$$

where $\tilde{\sigma}_i$ is the accuracy of the data $\Pi(\nu_i)$, and $\Pi_\sigma(\nu_i)$ is the transformation of $\sigma(\omega)$ to imaginary frequencies.

IV. RESULTS

To study methods of analytical continuation, we choose a model of $\sigma(\omega)$ on the real frequency axis, using calculations for a 2D Hubbard model as a guide. Using Eq. (4), the corresponding $\Pi(\nu)$ can easily be calculated. This is a well behaved and stable transformation. We generate results for the 60 smallest non-negative frequencies. We add random noise to this calculated $\Pi(\nu)$,

$$\Pi_\mu(\nu_i) = \Pi(\nu_i)(1 + r_{\mu,i}), \tag{20}$$

where $r_{\mu,i}$ has a Gaussian distribution with the width σ_0. This simulates the data that may be obtained from a QMC calculation by solving the Bethe-Salpeter equation. We generate ten different sets of data using different random numbers for each set.

In the DCA approach, $\Pi(\nu)$ is Fourier transformed to obtain $\Pi(\tau)$. This may require knowledge of $\Pi(\nu)$ for frequencies where the calculation is not very accurate. Although this problem can usually be circumvented by using the asymptotic behavior of $\Pi(\nu)$ for large ν, it then seems easier to work directly in ν space. Then if necessary, we can decide to use fewer values of ν than is needed to converge the Fourier transform and only use values which we believe are accurate. Specifically for the present calculation, a Fourier transform to τ space would lead to additional complications. Although by construction the present $\Pi_\mu(\nu_i)$ has a perfectly Gaussian noise which is uncorrelated for different values of ν_i, the Fourier transformed data would have correlation between different τ points. Methods working in τ space and methods working in ν space would then have data of different quality. To be able to compare all methods on an equal footing, we have therefore formulated them in imaginary frequency space, which essentially involves using kernels appropriate for this space.

The $\Pi_\mu(\nu_i)$ data are then analytically continued back to the real axis, using methods of interest. Since we know the exact result, namely the $\sigma(\omega)$ we started from, we can test the accuracy of the methods.

For the MaxEnt method we analytically continued each data set and then took the average.[18] As discussed above, the reason is that the MaxEnt method tends to attach too much significance to the noise. The batching method reduces the importance of the noise at the cost of using data with a lower accuracy. For the SVD [with the condition in Eq. (13)] and sampling methods we have not noticed any tendency to over-emphasizing the noise. Therefore we averaged the data before doing the analytical continuation to get data with the highest possible accuracy. For the Pade method with many data points on the imaginary axis there is a strong tendency to overemphasize the noise. However, we have not noticed any general improvement by "batching" the data, and therefore also for the Pade approximation we averaged the data before doing the analytical continuation.

The optical conductivity $\sigma(\omega)$ typically has peaks at $\omega = 0$ and at approximately $\omega = \pm U$, where U is the Hubbard on-site Coulomb interaction. We therefore use the real axis $\sigma(\omega)$,

$$\sigma(\omega) = \left\{ \frac{W_1}{1 + (\omega/\Gamma_1)^2} + \frac{W_2}{1 + [(\omega - \epsilon)/\Gamma_2]^2} \right. $$
$$\left. + \frac{W_2}{1 + [(\omega + \epsilon)/\Gamma_2]^2} \right\} \frac{1}{1 + (\omega/\Gamma_3)^6}. \tag{21}$$

Here $\Gamma_3 \gg (\Gamma_1, \Gamma_2)$ cuts off $\sigma(\omega)$ for large ω. Otherwise $\Pi(\nu)$ would not decay as ν^{-2}, as it should. Here we let ω and Γ_i have the unit eV and σ the unit $(\mathrm{m}\Omega\,\mathrm{cm})^{-1}$. Since the smallest nonzero frequency $\nu = \nu_0 \equiv 2\pi T$, we expect structures on an energy scale much smaller than ν_0 to be described very poorly. Here we use $T = 1/15$, giving $\nu_0 = 0.42$. We then choose two different models with $\Gamma_1 = 0.30$ and 0.6, respectively. For both models we use $\Gamma_2 = 1.2$, $\Gamma_3 = 4$, and $\epsilon = 3$. We use the weights $W_1 = 0.3$ and $W_2 = 0.2$.

Figure 2 shows results for $\Gamma_1 = 0.6$ and data with relatively good accuracy $\sigma_0 = 0.01$. Figure 2(a) shows results according to the Padé approximation for different numbers ($\nu_{\max}$) of frequencies. For $\nu_{\max} = 5$ the spectrum is rather structureless and the peak at $\omega = 3$ is not well described. Since $\nu_{\max} = 5$ corresponds to an imaginary frequency 1.7, smaller than the energy scale for the structures on the real axis, this is not surprising. As $\nu_{\max}$ is increased and more information is added, this peak is formed, although at too small energy. The peak at $\omega = 0$ is also not very well described.

Figure 2(b) shows results according to the SVD method. The eigenvalues α_k in Eq. (9) are in this case 0.50, 0.12, 0.048, 0.020, 0.0067, 0.0020, 0.00055, 0.00015,... The optimal value of n_ν according to the criterion in Sec. III B and $\sigma_0 = 0.01$ is then 5 and the corresponding results are shown by the thick line. Results are also shown by thin lines for $n_\nu = 3$, 4, 6, and 7. $n_\nu = 3$ is too small, and misses most of the structures. The values $n_\nu = 4$ gives similar results as $n_\nu = 5$ while $n_\nu = 6$ gives some unphysical oscillations and $n_\nu = 7$ puts in large spurious structure, giving too much weight to the noise.

Figure 2(c) shows results from the sampling method. For small values of $\nu_{\max}$ the structures are poorly described and, in particular, the Hubbard peak is placed at a too low energy. As discussed for the Padé approximation, this is not surpris-
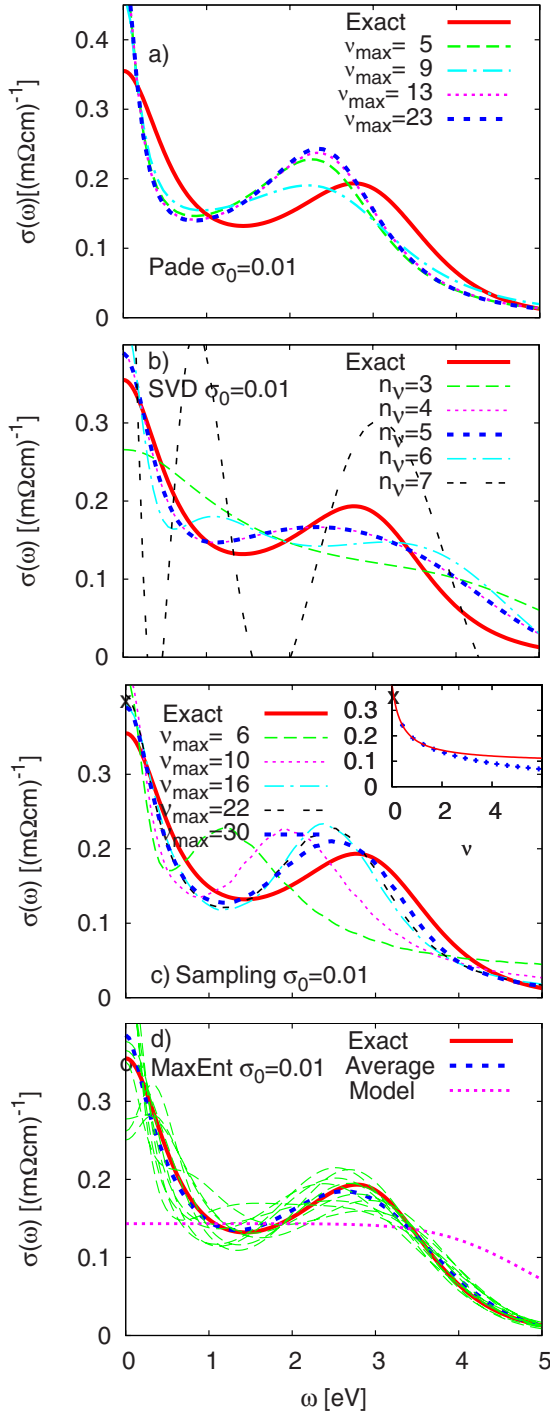
FIG. 2. (Color online) The optical conductivity for model [Eq. (21)] using $\Gamma_1=0.60$ and $\sigma_0=0.01$ according to (a) the Padé, (b) the SVD method, (c) the sampling, and (d) the MaxEnt methods compared with the exact results. (a) and (c) show results for different values of the maximum frequency ν_{max} considered and (b) for different values of n_ν. (d) shows results both for each individual sample (thin lines) and the average over all ten samples as well as the model used. The thick line in (b) indicates the optimum value of $n_\nu=5$ and the thick line in (c) the largest value of $\nu_{max}=30$ considered here. The x in the main part of (c) shows the estimate of $\sigma(0)$ by extrapolating $[\Pi(0)-\Pi(\nu)]/\nu$ to zero. This is illustrated by the inset in (c), where the symbol $\times$ gives the exact value of $\sigma(0)$. The symbol o in (d) is the estimate of $\sigma(0)$ based on $\Pi(\beta/2)$ [Eq. (6)].
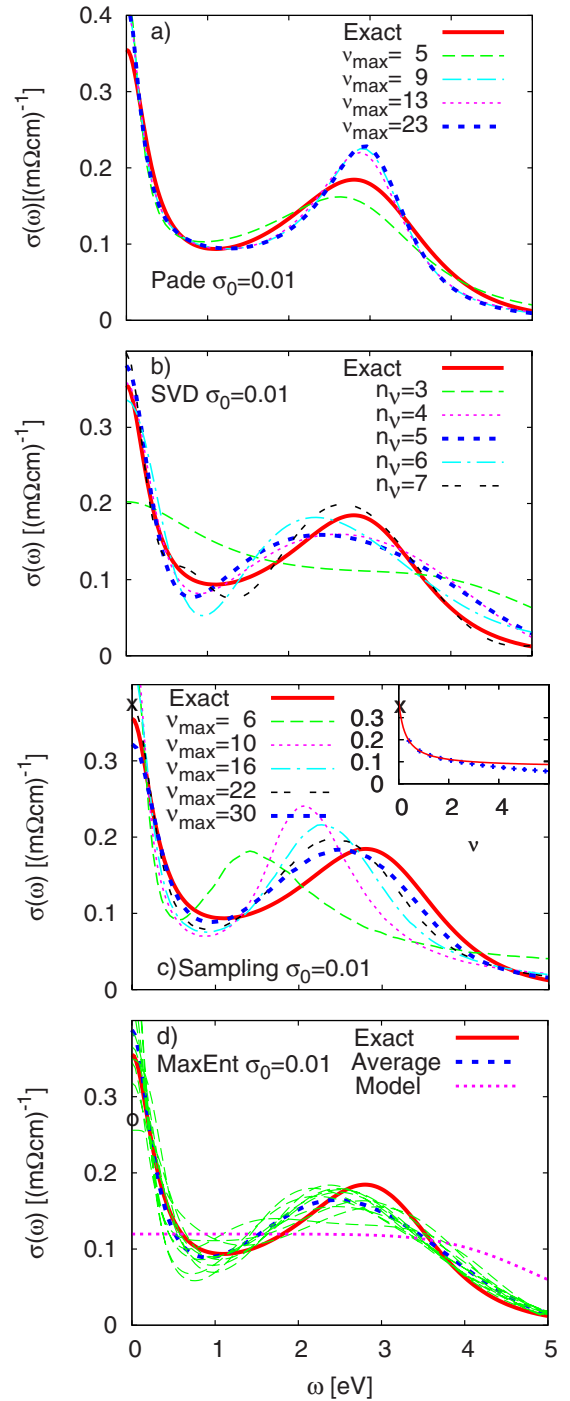
FIG. 3. (Color online) The same as Fig. 2 but for $\Gamma_1=0.3$ ($\sigma_0=0.01$).

ing since only information for small imaginary frequencies is used. As ν_{max} is increased the description improves. For a large $\nu_{max}=30$, shown by the thick line, the description is rather good. The inset shows the quantity $\gamma(\nu)$ in Eq. (5) and the extrapolation to $\nu=0$, giving an estimate of $\sigma(0)$. The symbol $\times$ in the inset gives the exact result and the $\times$ in the main Fig. 2(c) shows the result estimated from this extrapolation. This estimate is in this case somewhat too large.

Figure 2(d) shows the MaxEnt results. Results are shown for each of the ten data sets and also the average of the
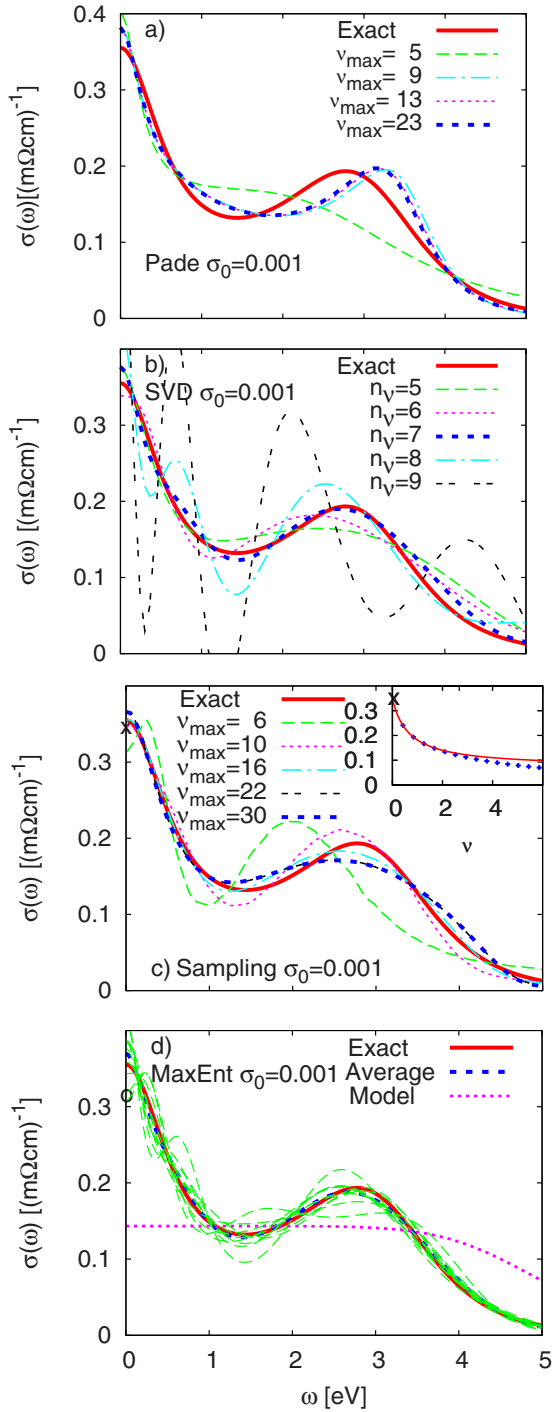
FIG. 4. (Color online) The same as Fig. 2 ($\Gamma_1 = 0.6$) but for $\sigma_0 = 0.001$.

FIG. 5. (Color online) The same as Fig. 2 but for $\Gamma_1 = 0.3$ and $\sigma_0 = 0.001$.

results is shown. Each MaxEnt spectrum shows rather large spurious oscillations due to the method giving too much weight to the noise. The average of these spectra, however, is rather good. The symbol o in Fig. 2(d) also shows the estimate in Eq. (6) of $\sigma(0)$. This estimate is accidentally quite good, although the spectrum has substantial variations over the range $|\omega| \lesssim 0.692\pi T$, and the requirement for Eq. (6) is not well satisfied. The reason is that the noise happens to make this estimate accurate while in Fig. 4 with more accurate data the estimate is less good.
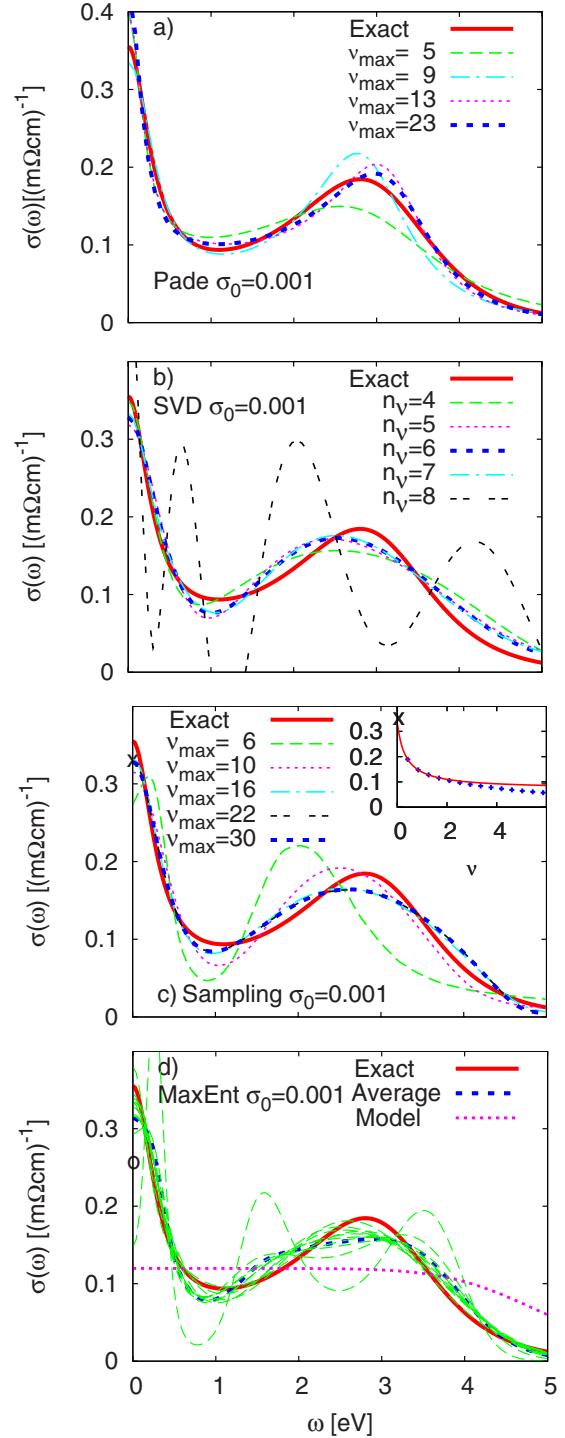
Figure 3 shows results for $\Gamma_1 = 0.3$, i.e., a narrower peak at $\omega = 0$. The SVD, sampling, and MaxEnt methods give comparable accuracy as in Fig. 2. The accuracy of the estimates of $\sigma(0)$ from Eq. (6) [$\bigcirc$ in Fig. 3(d)] is worse than in Fig. 2 since the peak at $\omega = 0$ is narrower and assumption behind Eq. (6) is less well satisfied. The estimate from Eqs. (5) and (7) [$\times$ in Fig. 3(c)] is of comparable accuracy as in Fig. 2.

Figure 4 shows results for $\Gamma_1 = 0.60$, i.e., a broader peak as in Fig. 2 but for very accurate data, $\sigma_0 = 0.001$. The accuracy
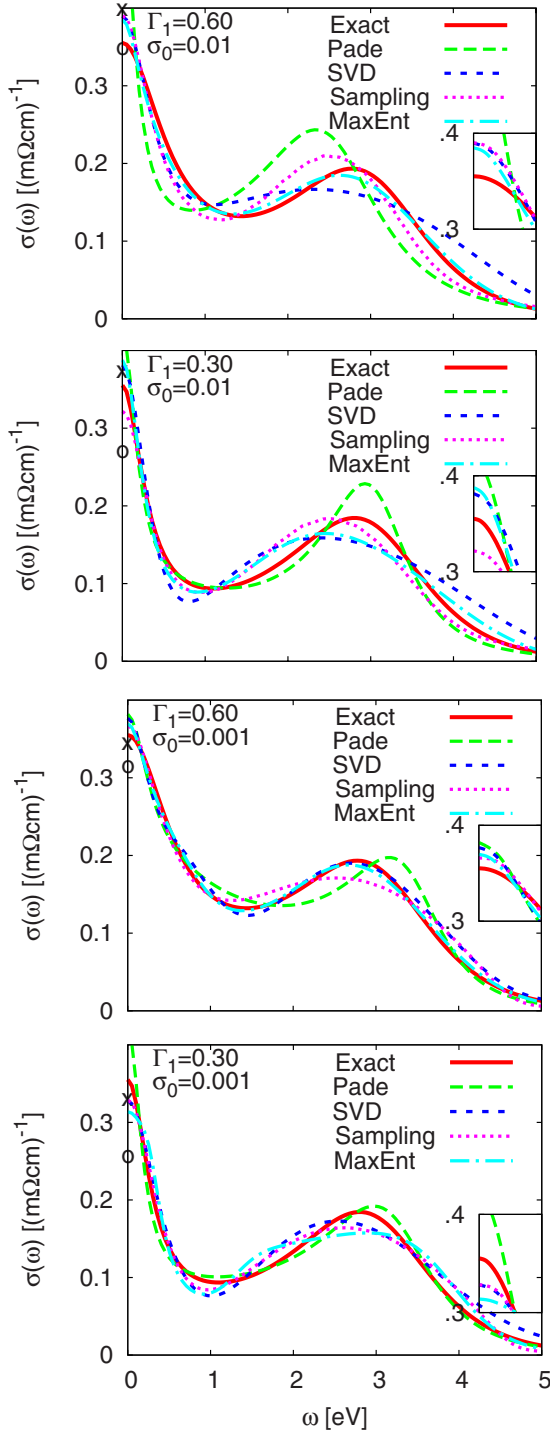
FIG. 6. (Color online) Comparison of the different methods for $\Gamma_1 = 0.6$ and 0.3 and for $\sigma = 0.01$ and 0.001. The insets show a magnified view in the range $0 \leq \omega \leq 0.25$.

FIG. 7. (Color online) Difference between the spectrum calculated using one of the methods and the exact spectrum for the parameters $\Gamma_1 = 0.6$ and 0.3 and for $\sigma = 0.01$ and 0.001.

of the Padé approximations is now improved, as expected. Because of the higher accuracy of the data, the optimum n_ν has increased from 5 to 7 for the SVD method. This leads to an improvement compared with Fig. 2(b), although there is a small unphysical oscillation at $\omega \sim 1$. There is a reduced spread of the thin curves in Fig. 4(d), representing the MaxEnt result for each individual data set. The average is only
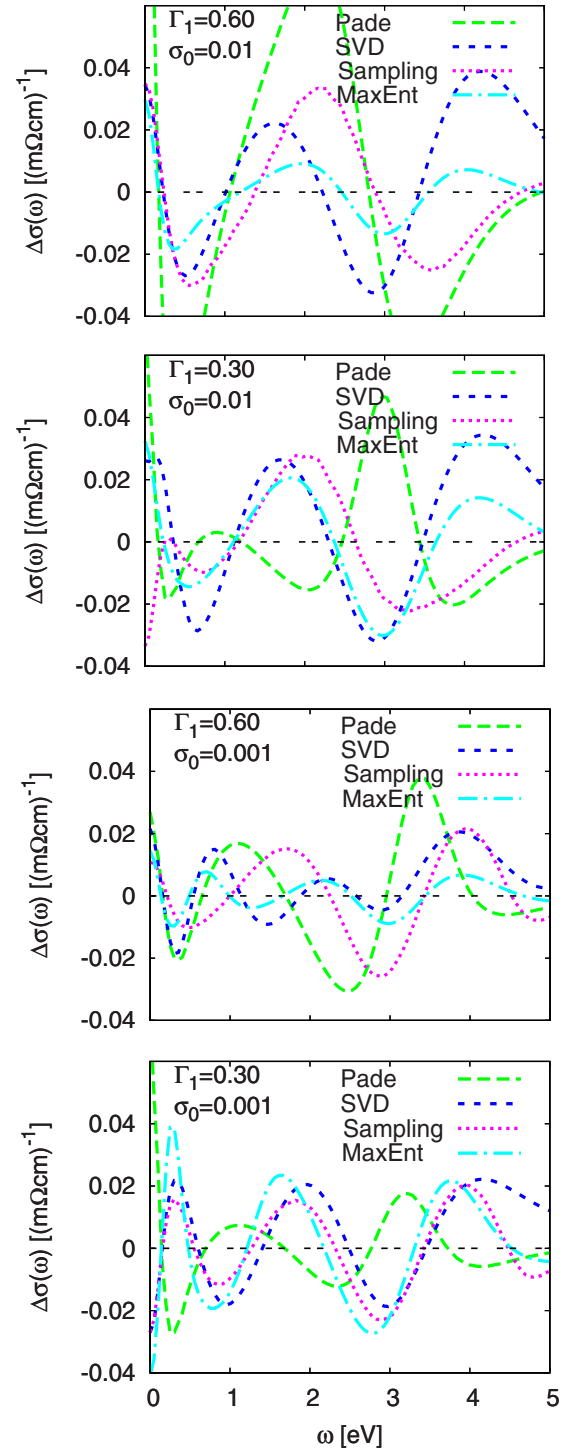
marginally improved. Figure 5 shows high accuracy data for a narrow peak, $\Gamma_1 = 0.3$.

In Fig. 6 we compare the different methods for the two different spectra ($\Gamma = 0.3$ and 0.6) and for the two accuracies ($\sigma_0 = 0.01$ and 0.001) considered here. Since the value of $\sigma(0)$ is of particular interest, we show results for small ω (≤ 0.25) in the insets. Typically the SVD, sampling, and

MaxEnt methods are of comparable accuracy, while the Pade approximation tends to overestimate $\sigma(0)$. The differences between the results of these methods and the exact result are shown in Fig. 7.

V. CORRELATION IN IMAGINARY TIME

We have so far generated data for imaginary frequencies and then added Gaussian noise. The noise for different frequencies is uncorrelated and the covariant matrix

$$C_{ik} = \frac{1}{M(M-1)} \sum_{j=1}^{M} [\bar{\Pi}(\nu_i) - \Pi^{(j)}(\nu_i)][\bar{\Pi}(\nu_k) - \Pi^{(j)}(\nu_k)] \quad (22)$$

is approximately diagonal. Here $\bar{\Pi}(\nu_i)$ is the average over the M samples $\Pi^{(j)}(\nu_i)$. If the data are obtained from a QMC calculation, C is in general not diagonal. There is then a need to make a transformation to a diagonal covariant matrix. Here we follow Jarrell and Gubernatis.[4] A matrix U is found such that

$$C' = U^{-1}CU \quad (23)$$

is diagonal. The data and kernel are then transformed to the new representation

$$K' = U^{-1}K \quad \Pi' = U^{-1}\Pi \quad (24)$$

and the diagonal elements of C' are used to define a new likelihood function. The result is that some of the diagonal elements of the covariant matrix are now larger, implying less accurate data than one might have thought. This does not, however, change the general conclusions above.

VI. CONCLUSIONS

We have compared different methods for analytically continuation of imaginary axis data to real frequencies for the optical conductivity. We transform spectra from the real frequency axis to the imaginary axis and add statistical noise. These data are then transformed back to the real axis using the different analytical continuation methods. By comparing with the original spectrum, we can compare the accuracy of these methods. Typically, these methods have problems if the spectra have features on a much smaller energy range than $2\pi T$. Due to the thermal broadening of physical spectra, this may not be a serious problem in many cases. Here we have focused on two cases where the relevant energy scale, Γ_1 is 0.3 or 0.6 compared with $2\pi T = 0.42$.

We also considered two methods for obtaining $\sigma(0)$ directly, Eq. (6) and extrapolation of $\gamma(\nu)$ in Eq. (5) to $\nu = 0$. The method based on Eq. (6) tends to underestimate $\sigma(0)$, in particular, if $\sigma(\omega)$ has a narrow Drude peak while the extrapolation of Eq. (5) is typically more accurate.

Calculations for the cases considered in this paper as well as for results from DCA typically gives larger values for $\sigma(0)$ in the Padé approximation than from the SVD, sampling, and MaxEnt approaches. The Padé approximation generally tends to give somewhat less accurate results than the other three methods. Sometimes unphysical results are obtained due to poles close to the real axis. The other three methods tend to give results of comparable accuracy. We nevertheless find it very useful to use all three methods. This provides cross checks and gives a somewhat better idea about what the true spectrum may look like.

[1] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar, Phys. Rev. D **24**, 2278 (1981).

[2] T. Maier, M. Jarrell, T. Pruschke, and M. H. Hettler, Rev. Mod. Phys. **77**, 1027 (2005).

[3] A. N. Rubtsov and A. I. Lichtenstein, JETP Lett. **80**, 61 (2004); P. Werner and A. J. Millis, Phys. Rev. B **74**, 155107 (2006).

[4] M. Jarrell and J. E. Gubernatis, Phys. Rep. **269**, 133 (1996).

[5] R. N. Silver, D. S. Sivia, and J. E. Gubernatis, Phys. Rev. B **41**, 2380 (1990); J. E. Gubernatis, M. Jarrell, R. N. Silver, and D. S. Sivia, *ibid.* **44**, 6011 (1991); W. von der Linden, Appl. Phys. A: Mater. Sci. Process. **60**, 155 (1995).

[6] M. Bertero, C. De Mol, and E. R. Pike, Inverse Probl. **1**, 301 (1985); M. Bertero and E. R. Pike, *Handbook of Statistics*, edited by N. K. Bose and C. R. Rao (Elsevier Sience, New York, 1993) Vol. 10; M. Bertero, P. Branzi, E. R. Pike, and L. Rebolia, Proc. R. Soc. London, Ser. A **415**, 257 (1988).

[7] C. E. Creffield, E. G. Klepfish, E. R. Pike, and S. Sarkar, Phys. Rev. Lett. **75**, 517 (1995).

[8] I. Krivenko and A. Rubtsov, arXiv:cond-mat/0612233 (unpublished).

[9] H. J. Vidberg and J. W. Serene, J. Low Temp. Phys. **29**, 179 (1977).

[10] G. A. Baker, Jr., *Essentials of Padé Approximants* (Academic, New York, 1975), p. 100.

[11] A. S. Mishchenko, N. V. Prokof'ev, A. Sakamoto, and B. V. Svistunov, Phys. Rev. B **62**, 6317 (2000).

[12] K. Vafayi and O. Gunnarsson, Phys. Rev. B **76**, 035115 (2007).

[13] S. Hochkeppel, F. F. Assaad, and W. Hanke, Phys. Rev. B **77**, 205103 (2008).

[14] F. Lin, M. A. Morales, K. T. Delaney, C. Pierleoni, R. M. Martin, and D. M. Ceperley, Phys. Rev. Lett. **103**, 256401 (2009).

[15] R. T. Cox, *The Algebra of Probable Inference* (John Hopkins University Press, Baltimore, Maryland, 1961); D. S. Sivia, *Data Analysis–A Bayesian Tutorial* (Clarendon Press, Oxford, 1996).

[16] See, e.g., J. Skilling, J. Microsc. **190**, 28 (1998).

[17] S. R. White, Phys. Rev. B **44**, 4670 (1991).

[18] O. Gunnarsson, M. W. Haverkort, and G. Sangiovanni, Phys. Rev. B **81**, 155107 (2010).