Stochastic method for analytic continuation of quantum Monte Carlo data

Anders W. Sandvik
Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, Illinois 61801
(Received 28 October 1997)

A method for analytic continuation of quantum Monte Carlo data is presented. The spectrum $A(\omega)$ is parametrized as a sum of δ functions, the weights A_i of which are sampled according to a distribution $p(A) \sim \exp(-\chi^2/\Theta)$. It is argued that the calculated entropy S provides a criterion for determining the Θ corresponding to the ''best'' averaged spectrum. The appearance of spurious structure is signaled by a sharp drop in $\langle S(\Theta)\rangle$, which in test cases is preceded by a local maximum. Results for the dynamic spin structure factor of a 16-site Heisenberg chain obtained at this maximum are in better agreement with exact results than ''classic'' maximum-entropy results. [S0163-1829(98)02717-9]

The analytic continuation of imaginary-time quantum Monte Carlo (QMC) data to real frequency is a challenging problem, for which the best solution is not known. The recently developed maximum-entropy (Max-Ent) method[1] has proved to be very useful in many cases, but is founded on a hypothesis for which only heuristic arguments have been given. The inclusion of an entropy term in the optimization of the spectrum leads to a bias towards spectra with large entropy. This bias is controlled using arguments of Bayesian logic, with the goal of obtaining the most probable spectrum given the QMC data. In particular, the Max-Ent method should produce results free from structure due to the statistical errors (''spurious structure''). However, it is likely that the bias also leads to distortions beyond simple broadening, and may reduce some of the ''real'' structure present in the data.

Here an alternative method is presented which is based on the assumption that all spectra that fit the QMC data equally well, in the χ^2 sense, are equally probable candidates for the result sought. In the absence of a nonbiased criterion favoring certain types of spectra, it is proposed that all spectra (which are typically not smooth) corresponding to the same χ^2 values should be averaged over. One then obtains a family of smooth averaged spectra as a function of χ^2. It is argued that the entropy versus χ^2 *calculated* for these spectra signals the appearance of spurious structure, and provides a natural criterion for determining the best spectrum.

The spectral function $A(\omega)$ for some operator $\hat{O}$ is given in terms of the eigenstates $|n\rangle$ of the Hamiltonian $\hat{H}$ and the corresponding energies E_n by

$$A(\omega)=\frac{1}{\pi}\sum_{m,n} e^{-\beta E_n}|\langle m|\hat{O}|n\rangle|^2\delta(\omega-[E_m-E_n]), \quad (1)$$

where β is the inverse temperature. The corresponding imaginary-time correlation function that can be calculated in QMC simulations is

$$G(\tau)=\langle \hat{O}(\tau)\hat{O}(0)\rangle, \quad (2)$$

where $\hat{O}(\tau)=e^{\tau\hat{H}}\hat{O}e^{-\tau\hat{H}}$.

For the numerical analytic continuation of $G(\tau)$, the spectrum $A(\omega)$ is parametrized as a sum of N δ functions on a fixed frequency grid $\omega_1,\ldots,\omega_N$:

$$A(\omega)=\sum_{i=1}^{N} A_i\delta(\omega-\omega_i). \quad (3)$$

The imaginary-time function $G(\tau)$ corresponding to a bosonic $A(\omega)$ (which is the case considered here) is related to the spectrum according to

$$G(\tau)=\int_0^\infty d\omega A(\omega)K(\omega,\tau)=\sum_{i=1}^{N} A_iK(\omega_i,\tau), \quad (4)$$

where the kernel is given by

$$K(\omega,\tau)=[e^{-\tau\omega}+e^{-(\beta-\tau)\omega}]/\pi. \quad (5)$$

The corresponding correlation function calculated using the QMC method is denoted $\bar{G}(\tau)$, and is available for a discrete set of times $\{\tau_i\}$. The statistical error of $\bar{G}(\tau_i)$ is denoted σ_i. The deviation of $G(\tau)$ from the QMC data for a given set of weights $\{A_1,\ldots,A_N\}$ is quantified by

$$\chi^2=\sum_i [G(\tau_i)-\bar{G}(\tau_i)]^2/\sigma_i^2. \quad (6)$$

In some cases it may be necessary to use the full covariance matrix in Eq. (6), instead of just its diagonal elements.[1] However, in the test case considered below, Eq. (6) is adequate (no significant differences are seen in Max-Ent results obtained with and without the off-diagonal elements).

Analytic continuation based on minimizing χ^2 was attempted by Schüttler and Scalapino, who also noted that the positive definiteness of $A(\omega)$ greatly helps in obtaining reasonable results.[2] However, a straightforward χ^2 minimization works well only if a small N is used in Eq. (3), and can then of course only roughly indicate the distribution of spectral weight. For larger N, there are typically many spectra with different appearances which have very similar χ^2 values, and those representing the closest fits likely contain structure strongly affected by the statistical errors of the QMC data. Attempts have been made to include a term enforcing smoothness, thereby eliminating spurious peaks.[3]

The quantity to be minimized is then $\chi^2+\alpha D$, where D is a measure of smoothness, for which $D=\Sigma_i(A_{i+1}-A_i)^2$ was used.[3] One problem with this and related[4] methods is that the outcome depends on the parameter α, and there is no obvious way of determining its optimum value (corresponding to the ''best'' spectrum).

The Max-Ent method[1] is based on the ansatz that the likelihood p of a given spectrum in the presence of the QMC data is given by

$$p(A)\sim\exp(-\chi^2/2+\alpha S), \tag{7}$$

where S is the entropy,

$$S=-\int_{-\infty}^{\infty} d\omega A(\omega)\ln[A(\omega)/m(\omega)], \tag{8}$$

and $m(\omega)$ is a default model which defines the zero of the entropy and to which the spectrum reduces in the absence of data (or $\alpha\to\infty$). Starting from this ansatz, Bayesian logic can be used to determine an optimum value of α, so that there are in principle no free parameters in the problem, except for the default model. It has been argued that the default model provides for an opportunity to include various prior knowledge, such as known frequency moments of $A(\omega)$. However, one can also argue that a flat default model is the most appropriate. There are also several variants of the method that differ in the way the parameter α is determined. The ''classic'' Max-Ent method determines α self-consistently from the spectrum, whereas other variants integrate over certain distributions of α. Apart from these remaining ambiguities, there are no rigorous arguments for using the postulated likelihood function (7). Nevertheless, the Max-Ent method has proved very useful in practical calculations, and represents a definite step forward.

There is, however, nothing that guarantees that the Max-Ent method resolves the full amount of ''real'' structure present in the imaginary-time QMC data. It is also clear that the entropy term leads to some distortions of the spectrum. The bias towards high-entropy curves, which typically have Gaussian-like features, leads to difficulties with spectra that have features such as sharp edges or flat portions. Hence, exploring alternative methods is still important. As already briefly discussed above, the method introduced in this paper is based on the assumption that all spectra with equal χ^2 values are equally probable, and that no further bias should be introduced that favors certain types of spectra. The problem is then to find some way to characterize the spectra, so that a particular χ^2 can be identified at which the averaging should be done. It will be argued here that the entropy

$$S=-\sum_{i=1}^{N} A_i\ln(A_i)K(\omega_i,0), \tag{9}$$

[corresponding to Eq. (8) with a constant default] serves this purpose.

In practice, it is more convenient to average over some narrow distribution of χ^2, instead of a fixed value (corresponding to a δ-function distribution). A natural choice for the distribution is

$$p(A)\sim\exp(-\chi^2/\Theta), \tag{10}$$

and this then defines a family of spectra parametrized by Θ [or, alternatively, the corresponding $\langle\chi^2(\Theta)\rangle$]. The averaging can be carried out using Monte Carlo simulation in the weight space $\{A_1,\ldots,A_N\}$, starting from a large Θ which is gradually reduced as the simulation proceeds (''simulated annealing''[5]). When Θ is large, $\langle\chi^2\rangle$ is large and also has large fluctuations. As Θ is lowered, the fluctuations diminish and χ^2 approaches its global minimum, provided that the system is carefully ''annealed'' (i.e., the simulation is started at a sufficiently large Θ, and the reduction is done sufficiently slowly). The entropy, Eq. (9), is a measure of the amount of structure in the spectrum. It is therefore likely that the appearance of spurious peaks will be accompanied by a drop in the average entropy $\langle S\rangle$ at the Θ where such structure becomes significant. As the global χ^2 minimum is approached $\langle S\rangle$ of course becomes independent of Θ. A sharp drop in the entropy before the approach to a constant value has indeed been observed in all cases tested so far. Furthermore, there is often a clear maximum at some $\Theta=\Theta^*$ preceding the drop. It is then natural to postulate that Θ^* is the optimum value at which to average the spectrum.

Next results illustrating these ideas are presented. The test case is the dynamic structure factor $S(q,\omega)$ [Eq. (1) with $\hat{O}=S^z_q$] of the spin-1/2 Heisenberg chain, defined by the Hamiltonian

$$\hat{H}=J\sum_{i=1}^{L}\mathbf{S}_i\cdot\mathbf{S}_j \quad (J>0). \tag{11}$$

A small size, $L=16$, is considered, so that exact finite-temperature results for comparison can be obtained by exact diagonalization. The imaginary-time correlation function corresponding to $S(q,\omega)$ is

$$G(q,\tau)=\frac{1}{L}\sum_{j=1}^{L}\sum_{k=1}^{L} e^{iq(j-k)}\langle S^z_k(\tau)S^z_j(0)\rangle. \tag{12}$$

For a small system, the exact $S(q,\omega)$ at $T=0$ contains only a small number of δ functions. Including some reasonable broadening gives a continuous spectrum, which nevertheless has structure on a scale that cannot be expected to be resolved using any numerical analytic continuation of QMC data. At finite T, the distribution of δ functions is quite dense for $L=16$, and at high enough T the spectrum is relatively smooth on a reasonable frequency scale ($\Delta\omega\sim0.1J$). Here $q=\pi/4$ and $q=\pi/2$ at $T/J=0.5J$ are considered. The QMC results were obtained using a ''stochastic series expansion'' algorithm,[6] which is free from systematic errors. The relative statistical errors of the $\bar{G}(q,\tau_i)$ used here are less than 10^{-3} for every τ_i.

It is convenient to normalize the imaginary-time data so that $\bar{G}(q,0)=1$, corresponding to a normalization of the spectrum, $\int d\omega A(\omega)=\pi$ [the spectrum is easily renormalized after the continuation, by multiplying with the original $\bar{G}(q,0)$]. The annealing and sampling of the spectra are done using updates of the weights A_i in Eq. (3) in such a way that the first few frequency moments are conserved. The nth moment is given by

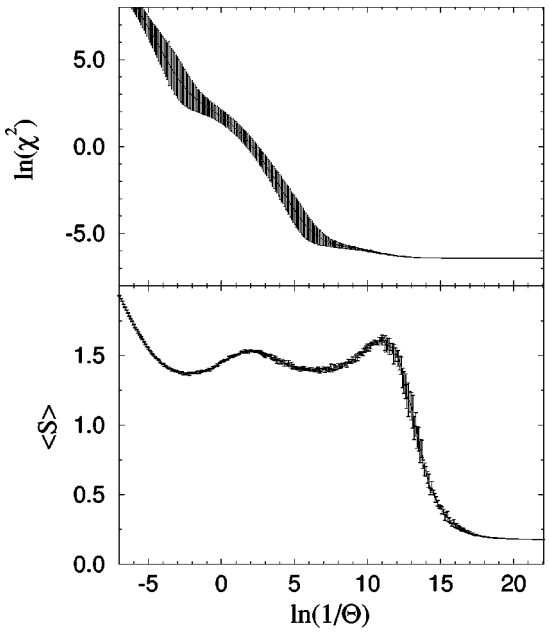

FIG. 1. The behavior of χ^2 (top) and the entropy (bottom) for $S(q/2,\omega)$ at $T/J=0.5$. The error bars in the top graph indicate the average $\pm$ one standard deviation of the distribution of χ^2 during the simulation. In the entropy graph, the error bars indicate the standard deviation of individual averages obtained from six independent annealings.

$$\rho_n=\sum_{i=1}^{N}\omega_i^n A_i[1\pm\exp(-\beta\omega_i)], \tag{13}$$

with $+$ and $-$ corresponding to even and odd n, respectively. An update conserving $\rho_0,\ldots,\rho_m$ amounts to randomly selecting a set of $m+2$ weights $A_{R(1)},\ldots,A_{R(m+2)}$ $[R(i)\neq R(j)]$ and changing these, $A_{R(i)}\rightarrow A'_{R(i)}$, such that $\rho'_k=\rho_k$ $(k=0,\ldots,m)$. The positive definite weights satisfying the $m+1$ constraints define a finite line segment in an $(m+2)$-dimensional hypercube. A point on this line is chosen at random, and the new spectrum is accepted or rejected according to the standard Metropolis algorithm, with the probability distribution given by Eq. (10). Updates that do not conserve moments lead to a too slow random walk in the parameter space, and the spectrum for a given Θ does not converge within a reasonable time. Different updates conserving between one and four moments were used in the work discussed here. A sequence of $\sim N/(m+1)$ of each of these updates defines one sampling ''step.'' Using a logarithmic scale, the annealing is done by periodically increasing $\ln(1/\Theta)$ by a fixed amount. For all the results shown here an increment $\Delta\ln(1/\Theta)=0.1$ was used.

Figure 1 shows the behavior of χ^2 and the entropy in a simulated annealing for $S(\pi/2,\omega)$, using $N=50$ frequencies with spacing $\Delta\omega=0.1$ and an imaginary-time grid with $\Delta\tau=0.2$. The number of simulation steps at each Θ ranged from 5×10^4 to 2×10^5, and six independent annealings were carried out. Looking at $\ln(\chi^2)$, three different regimes can be identified: For large Θ, there is a rapid drop vs $\ln(1/\Theta)$. There is then a crossover [at $\ln(1/\Theta)\approx8$] to a regime where the change is much slower, and the fluctuations decrease rapidly. Finally, a constant value is approached as the simulation settles at the global χ^2 minimum.[7] Intuitively, it seems natural that the spectrum should be averaged in the intermediate regime, where χ^2 is close to its minimum but still fluctuates some. However, this regime is rather wide, and one cannot easily define a particular ''best'' value of Θ just from the behavior of χ^2. The entropy, on the other hand, shows a clear structure in this regime.[8] The sharp drop as the global χ^2 minimum is approached can be understood as due to a rapid increase in the amount of structure (sharp peaks). In order to understand the local maximum preceding the entropy drop, it is useful to examine the appearance of spectra averaged at different values of Θ.

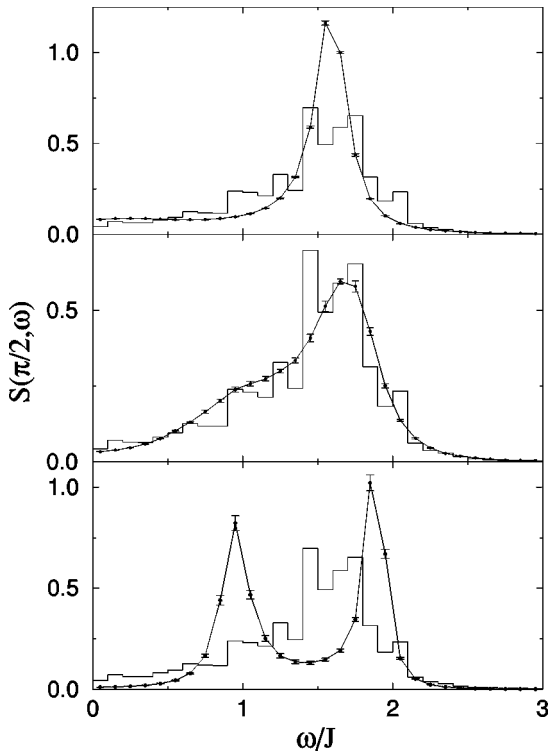

FIG. 2. Spectra averaged at three different values of Θ, for the same situation as in Fig. 1 (connected points with error bars, calculated on the basis of six different simulations). The values of $\ln(1/\Theta)$ are, from top to bottom, 6.5, 10.8, and 12.5. The histograms represent the exact result.

Figure 2 shows spectra averaged approximately at the local entropy maximum Θ^*, as well as before and after this point. The exact diagonalization result for $S(\pi/2,\omega)$ is also shown, in the form of histograms with a bin width 0.1J. For $\ln(1/\Theta)=6.5<\ln(1/\Theta^*)$, the sampled spectrum has a narrow peak located in the regime where the actual spectrum has the largest weight. As Θ approaches Θ^*, the peak broadens and becomes more asymmetric, before a second peak appears for $\ln(1/\Theta)>\ln(1/\Theta^*)$. One can now qualitatively understand the local entropy maximum as due to a flattening out of the spectrum prior to the appearance of a second sharp peak.

The entropy also exhibits another local maximum at lower $\ln(1/\Theta)$, and of course grows rapidly as $\Theta\rightarrow\infty$. The rather complex behavior certainly deserves more detailed studies (How general is this type of behavior? What do many local minima and maxima in $\langle S\rangle$ signify?), but here the main point is the existence of a local maximum before a sharp drop to a constant value.

In order to accurately determine Θ^*, it is necessary to carry out long simulations. If the annealing is performed too quickly (too few steps per Θ value), the calculated entropy curve exhibits a broader maximum than what is seen in Fig. 1. The left side of the peak is quite stable with respect to the annealing rate, but the location of the rapid drop is shifted towards higher $\ln(1/\Theta)$. Apparently, the simulation easily gets ''trapped'' at the local entropy maximum. Hence, in cases where the maximum is broad and its exact position is hard to determine, it may for practical purposes be better to estimate Θ^* as a point slightly to the left of the peak center (which within error bars could be the actual maximum). Figure 2 (and more detailed studies of the dependence on Θ) also shows that the change in the spectrum before the entropy peak is much less dramatic than right after, where sharp peaks rapidly emerge. A strategy of slightly underestimating $\ln(1/\Theta^*)$ also conforms with the general notion that too little structure is better than too much.

Next, results for both $q=\pi/2$ and $q=\pi/4$ are compared with spectra obtained using the ''classic'' Max-Ent method[1] (with a flat default). The point Θ^* used for the stochastic method was determined as discussed above, as a point slightly before the center of the last local entropy maximum, where a clear increase with $\ln(1/\Theta)$ has ceased [e.g., for the case $q=\pi/2$ discussed above, $\ln(1/\Theta)=10.6$ was used]. Figure 3 shows the results, along with histograms representing the exact spectra. The new method clearly reproduces the exact spectra better than the Max-Ent method, although the Max-Ent results do also represent reasonable broadened averages.

It should be stressed that although the entropy is used in the method proposed here, the underlying philosophy differs fundamentally from standard Max-Ent methods, where the inclusion of the entropy in the optimization explicitly affects the shape of the spectrum. In the stochastic method, a family of spectra is obtained based only on the QMC data, and the entropy is used only to single out one spectrum. Hence, any structure in the spectrum obtained is due solely to the QMC data.

The method has here been demonstrated only for a relatively simple test case. Results obtained for other models and dynamic quantities indicate that the behavior of the entropy vs Θ found here is typical for spectra with one broad continuous (on some reasonable frequency scale) structure with a single maximum. Good continued spectra are then obtained using $\Theta=\Theta^*$, and for $\Theta<\Theta^*$ two sharp peaks typically start to emerge. In cases where the actual spectrum has two peaks, there is also a sharp entropy drop as the global χ^2 minimum is approached. However, the entropy maximum

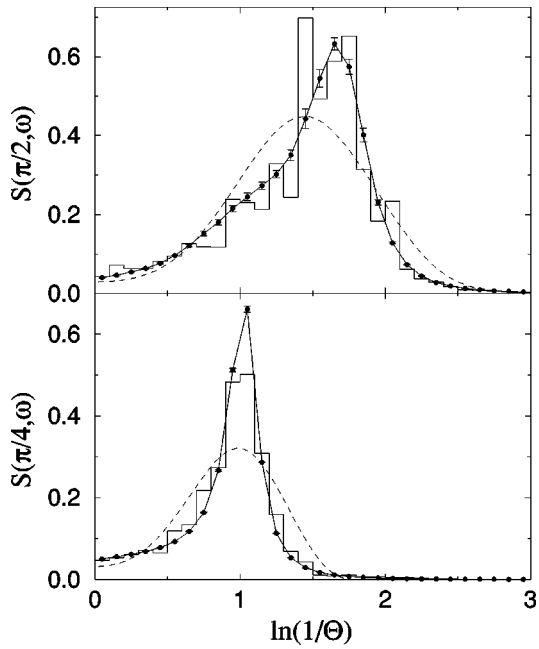

FIG. 3. The dynamic structure factor at two different wave numbers, obtained by averaging spectra at the respective $\Theta=\Theta^*$ (connected points with error bars). The Max-Ent results are shown as dashed curves, and the exact diagonalization results are represented by histograms.

associated with the appearance of the two peaks then occurs at quite high χ^2 values (if the data is sufficiently good) and is then clearly not the preferred point for sampling. For QMC data of very high accuracy one would presumably have a final entropy maximum before the sharp drop associated with the emergence of additional peaks, and one could then again use this to determine the optimum Θ for sampling. In typical cases the data may, however, be compatible with just two δ functions, and then it is difficult to determine a Θ^* (this is then also an indication that the data are not of sufficient accuracy for a reliable analytic continuation). Clearly more work is needed to clarify the general behavior of the entropy before the method can be applied to more complicated spectra than the single-maximum case considered here. A problem for practical use of the method is that the sampling needed for an accurate determination of Θ^* as well as the averaging needed to obtain a final result are quite time consuming. The good agreement with the exact results obtained here should motivate further work along these lines.

Support from the NSF under Grant Nos. DMR-95-27304 and DMR-97-12765 is acknowledged.

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

[2] H.-B. Schüttler and D. J. Scalapino, Phys. Rev. Lett. **55**, 1204 (1985); Phys. Rev. B **34**, 4744 (1986).

[3] S. R. White, D. J. Scalapino, R. L. Sugar, and N. E. Bickers, Phys. Rev. Lett. **63**, 1523 (1989).

[4] M. Jarrell and O. Biham, Phys. Rev. Lett. **63**, 2504 (1989).

[5] S. Kirkpatrick, S. Gelatt, and M. P. Vecchi, Science **220**, 671 (1983); S. Kirkpatrick, J. Stat. Phys. **34**, 975 (1984).

[6] A. W. Sandvik and J. Kurkijärvi, Phys. Rev. B **43**, 5950 (1991); A. W. Sandvik, J. Phys. A **25**, 3667 (1992).

[7] The very small minimum χ^2 reflects the presence of covariance in the QMC data.

[8] Note that $\langle S\rangle$ is the entropy averaged over the sampled spectra. The entropy of the average spectrum has a very similar behavior, but depends to a certain extent on the length of the simulation (longer simulation → smoother average → higher entropy). *A priori*, it is not clear which definition is preferable.