SPECTRAL PROPERTIES FROM QUANTUM MONTE CARLO DATA: A CONSISTENT APPROACH

R. PREUSS, W. VON DER LINDEN* AND W. HANKE
Institut für Theoretische Physik, Universität Würzburg
Am Hubland, D-97074 Würzburg, Germany ‡
**) Max-Planck-Institut für Plasmaphysik, EURATOM Association*
D-85740 Garching b. München, Germany §

Abstract. Bayesian statistics in the frame of the maximum entropy concept has widely been used for inferential problems, particularly, to infer dynamic properties of strongly correlated fermion systems from Quantum-Monte-Carlo (QMC) imaginary time data. In current applications, however, a consistent treatment of the error-covariance of the QMC data is missing. Here we present a closed Bayesian approach to account consistently for the QMC-data.

Key words: Quantum Monte Carlo, Analytic Continuation, Spectral Properties, Covariance.

1. Introduction

Since the discovery of the high temperature superconductors (HTSC) in 1986 [1] these materials are a major topic in condensed matter physics. Not only the unique properties of the HTSC's attract attention, but also their most promising use in various fields of industrial application. But despite a tremendous effort in this area a consistent theory is still lacking.

The crystal structures of all HT_c-superconductors have one thing in common: planes consisting of copper and oxygen atoms, where the Cu-atoms span a square lattice, and the O-atoms are located between the copper sites. Shown by experiment the major physics happens to take place in these CuO-planes. Therefore in our theoretical approach only these planes are considered.

The CuO-planes can be mapped onto a so-called 3 band Hubbard Hamiltonian [2], describing the motion of the electrons traveling around in the plane, undergoing a Coulomb repulsion for two electrons on the same site and taking into account chemical properties like the energy difference in the copper and oxygen

‡Email: roland@physik.uni-wuerzburg.de
§Email: wvl@ibmop5.ipp-garching.mpg.de

orbitals. Being further simplified to only one kinetic and one potential term (giving the so-called 1 band Hubbard model) and describing now merely a plane of copper atoms, this model still cannot be treated neither analytically nor by using perturbation theory methods (both terms are of same order of magnitude). Only numerical methods are capable of revealing the unusual properties arising from the correlation effects of the interacting electrons. Among these are the exact diagonalization (ED) of a small sized system (up to 25 sites), which will be used in this paper for reasons of comparison, and the Quantum-Monte-Carlo (QMC) simulation of somewhat larger systems (up to 300 sites). We apply the later method to extract results on one-particle spectra, which can then be compared with data from photoemission and inverse photoemission experiments on the HTSC-materials.

The problem we have to deal with is that the QMC method provides information of dynamical quantities on the imaginary time axis only [3,4]. The necessary analytic continuation to the real frequency spectral weight function is done by inverting the spectral theorem being a Laplace transformation, which is an ill-posed problem. The difficulties inverting such a function are enforced by the availability of only a finite number of the QMC data, additionally blurred by statistical errors. A direct inversion of the Laplace transform would tremendously overfit the noise and the desired signal would be buried underneath it.

Here MaxEnt has become a standard and successful technique to perform this inversion (for a review see [5]). Originally been introduced to infer celestial images from incomplete and noisy radio-astronomic data [6,7], it has been applied successfully to various other data-analysis problems [3,8–10].

A complication arises, however, in the present application as the errors of the QMC-data are correlated. It has been proposed [3] to include the error-covariance matrix as exact data-constraints. This approach has been heavily debated as it leads to the following dilemma. For a standard QMC sample size, the off-diagonal elements of the covariance matrix are not negligible and ought to be taken into account. However, the errors of the covariance matrix are huge and the information provided by the QMC covariance matrix is useless and in many applications even disadvantageous. The dilemma obviously arises due to the neglect of the errors of the covariance matrix. Here we present a closed Bayesian approach to account fully for the noisy QMC data plus covariance matrix. This scheme allows to infer reliable results whenever only a small sample size is available.

2. Quantified Maximum Entropy

Bayes' theorem is used to determine the posterior probability $P(A|DCH)$ (*Posterior*) of the sought-for quantity A, given hypotheses H and QMC data D, which are related to A via $D_l^{ex} = D_l^{ex}(\{A\})$. The experimental data D_l deviate from the exact values by the statistical QMC errors η_l. Due to the QMC-algorithm, the errors are correlated and the information about the error-covariance matrix will be denoted by C. Bayes' theorem, $P(A|DCH) = P(D|ACH)P(A|CH)/P(D|CH)$, relates the *Posterior* to the likelihood function $P(D|ACH)$, which contains the error statistic of the QMC data, and the prior probability $P(A|CH)$. Our prior knowledge is part of the hypotheses H.

In the case of a positive, additive distribution function (PAD) – e.g. the spectral density – the most ignorant *Prior* is the entropic *Prior* [11]

$$P(A|CH) = \frac{1}{Z_S} \exp\Big(\alpha \underbrace{\int [A(\omega) - m(\omega) - A(\omega)\ln(\frac{A(\omega)}{m(\omega)})]d\omega}_{S}\Big) \quad . \tag{1}$$

Z_S is the normalization constant guarantying $\int P(A|CH)\mathcal{D}A = 1$. In the following we will assume, as indicated in Eq. 1 that $A = A(\omega)$ is a function of the frequency ω. The entropy S is measured relative to a default-model $m(\omega)$ which contains "weak prior assumptions" which can still be overruled by the data constraints.

Next we will turn to the determination of the *Likelihood*, the central topic of this paper. The *Likelihood* quantifies the probability for the realization of the specific data-values D measured in the experiment, supposing the exact function A were known. The *Likelihood* describes therefore the error statistics of the QMC-data

$$P(D|ACH) =: \rho(D^{\text{ex}} - D) = \rho(\eta) \quad \text{given C and H} \quad . \tag{2}$$

In the following, C stands for the error-covariance matrix C_{ij}

$$C_{ij} = \frac{1}{N-1}\langle \eta_i \eta_j \rangle = \frac{1}{N-1}\langle (D_i^{\text{ex}} - D_i)(D_j^{\text{ex}} - D_j)\rangle \quad , \tag{3}$$

measured by QMC. N is the number of data-values D_i.

Assuming that the exact values of the error-covariance are known one obtains for the *Likelihood* the following result, which is the ubiquitous normal distribution

$$\rho(\eta) = \frac{1}{\sqrt{\det(2\pi C)}} e^{-\frac{1}{2}\sum_{ij}\eta_i C_{ij}^{-1}\eta_j} \tag{4}$$

and simplifies to a Gaussian if the errors are uncorrelated $C_{ij} = \delta_{ij}\sigma_i^2$,

$$\rho(\eta) = \frac{1}{\sqrt{\prod(2\pi\sigma_i^2)}} e^{-\frac{1}{2}\sum_i \frac{\eta_i^2}{\sigma_i^2}} \tag{5}$$

Unfortunately, this handy result is only valid if the QMC error-covariance were known exactly, which is not the case. Setting the off-diagonal elements of C to zero, which means to assume no correlations in the data, may result in over-fitting the data [5]. But the errors of the covariance matrix are significant: in our data we had to discover an average of 30% relative error for the diagonal elements, and up to 100% for the off-diagonal elements in the small bin case, decreasing only to about an average of 20% for the diagonal elements in the large bin case.

Therefore, the errors of the covariance matrix have to be treated on the same footing as the errors of the QMC data D in the first place – by quantified MaxEnt [6].

3. Consideration of the error of the covariance

Again, the *Posterior* for $\rho(\eta)$, given the QMC error-covariance C_{ij}, and the statistical errors σ_{ij} of C_{ij} and all our hypotheses can be determined via Bayes' theorem

$$P(\rho|C\sigma H) = \frac{P(C|\rho\sigma H)P(\rho|H)}{P(C|H)} \qquad . \tag{6}$$

Superfluous conditions have been discarded. Again the entropic *Prior* is invoked. QMC simulations provide the statistical error σ_{ij} of the covariance matrix. Further information is not available by present QMC simulations. We therefore assume that the error σ_{ij} are uncorrelated and known, this is part of our hypotheses H. A generalization beyond this assumption is straightforward. The *Likelihood* reads

$$P(C|\rho\sigma H) = \exp\Big(-\frac{1}{2}\underbrace{\sum_{ij}\frac{(C_{ij}-\int\rho(\eta)\eta_i\eta_j\,d\eta)^2}{\sigma_{ij}^2}}_{\chi^2}\Big) \qquad . \tag{7}$$

The MaxEnt result is hence obtained upon maximizing the *Posterior*, or rather

$$\mathcal{L}(\rho, C) = \alpha S - \frac{1}{2}\chi^2 - \lambda_0 \sum_i \Big(\int \rho(\eta)d\eta - 1\Big) \qquad . \tag{8}$$

Introducing the Legendre transform $\tilde{\mathcal{L}}$ [12]

$$\alpha\lambda_{ij} := \frac{\partial \mathcal{L}}{\partial C_{ij}} = \frac{C_{ij}^{\mathrm{QMC}} - C_{ij}}{\sigma_{ij}^2} \tag{9}$$

$$\tilde{\mathcal{L}}(\rho, \lambda) = \mathcal{L}(\rho, C) - \alpha\sum_{ij}\lambda_{ij}C_{ij} \qquad , \tag{10}$$

we obtain for the PAD

$$\rho(\eta) = \frac{1}{Z}\exp \eta^T\lambda\eta \qquad , \qquad Z = \int e^{-\eta^T\lambda\eta}d^N\eta = \frac{\pi^{N/2}}{\sqrt{\det(\lambda)}} \qquad . \tag{11}$$

The covariance is now

$$C_{ij}^{\mathrm{QMC}} = \int \rho(\eta)\eta_i\eta_j\,d\eta = -\frac{\partial}{\partial\lambda_{ij}}\ln(Z) = \frac{1}{2}[\lambda^{-1}]_{ij} \qquad . \tag{12}$$

The solution can be cast into the same functional form (Eq. 4) as in the case of exact data-constraints [10,12], merely the determination of the Lagrange parameters is modified due to the presence of noise to

$$\lambda = \frac{1}{2}[C + \alpha M]^{-1} \qquad , \qquad M_{ij} = \sigma_{ij}^2\lambda_{ij} \qquad . \tag{13}$$

Hence the *Likelihood* $P(D|ACH)$ remains a normal distribution, merely the covariance is not the QMC error-covariance, it rather has to be determined via Eq. 13. The regularization parameter α entering Eq. 13 can be determined either self-consistently upon maximizing the marginal *Posterior* $P(\alpha|C\sigma H)$ [6,7], or via the historic condition $\chi^2 = N$. We employ the historic approach since we know that the number of good degrees of freedom is small and both stopping criteria will yield essentially the same result[7].

4. Application to the spectral properties of strongly correlated electrons

As a typical and topical problem we study the dynamic properties of the Hubbard model which is presently the subject of intense analytical and numerical studies. The detailed understanding of the dynamic properties of strongly correlated electrons is essential for the theoretical description of the high temperature superconductors. The Hubbard model reads

$$H = -t \sum_{<i,j>,\sigma} \left(c^\dagger_{i,\sigma} c_{j,\sigma} + h.c. \right) + U \sum_i n_{i\uparrow} n_{i\downarrow} \tag{14}$$

with the hopping matrix element t between two adjacent sites. $c^{(\dagger)}_{i,\sigma}$ destroys (creates) an electron of spin σ on site i, $< i,j >$ denotes nearest neighbors, U is a Coulomb repulsion for two electrons of opposite spin on the same site and $n_{i,\sigma} = c^\dagger_{i,\sigma} c_{i,\sigma}$.

Unfortunately, dynamic properties cannot be measured directly by QMC simulations. Dynamical information is provided by the one-particle Matsubara Greens function $D_l = - < T_\tau c(\tau_l) c^\dagger(0) >$ for discrete τ_l-values on the imaginary time axis, where $l = n * \beta / L,\ n = 0, ..., L$ and L is the number of time slices. In order to determine the spectral density $A(\omega)$ for real frequencies ω the spectral theorem is applied:

$$D_l = - \int A(\omega) \frac{e^{-\tau_l \omega}}{1 + e^{-\beta\omega}} d\omega \tag{15}$$

which is, as already stated above, an inverse Laplace transformation problem and pathologically ill-posed.

Further information on the spectrum is provided by making use of the lowest order moments of $A(\omega)$, $\mu_m = \int \omega^m A(\omega) d\omega$, which are given by commutation relations, $\mu_m =< [c, H]_m, c^\dagger >$, and are of simple shape for m=1,2 [13].

5. Results

In order to compare the QMC/MaxEnt-data with exact results, we consider a chain of $N = 12$ sites, which is still accessible by exact diagonalization (ED) techniques. The QMC-simulations were done for an inverse temperature of $\beta t = 20$ ($T = 0.05t$), where ground state behavior is achieved for this system size and a comparison with the $T = 0$-ED results is possible. After obtaining thermal equilibrium, up to 640000 sweeps through the space-time Ising-fields were performed.

The ED-spectra were artificially broadened with a Gaussian distribution to get the shape normally observed by experiment, as is standard procedure on handling these results. Therefore the comparison persists on the overall distribution of spectral weight rather than looking for the redisplaying of each (infinitely high) peak of the ED-data.

Data from consecutive measurements are highly correlated, even for the same statistical variable D_l. A study of the skewness (third moment) and the kurtosis (fourth moment) of the data showed, that to get Gaussian behavior at least 200 measurements, each separated by 4 sweeps, have to be accumulated to form one bin. Then the results of a certain number of bins are used for the inversion process.

But binning the data does not suffice to get rid of all the correlations. Still one has to consider the correlations in imaginary time τ (i.e. between D_l and $D_{l'}$). In Fig. 1 the QMC-result for one single bin is compared to the final shape of the Greens function for $N^{bin} = 800$. Instead of being distributed 'at random' around the average the data for this bin are systematically lower than the average for $\tau < \beta/2$ and systematically higher for $\tau > \beta/2$. These correlations may be reduced by forming larger and larger bins (i.e. using more and more computation time). But it is the aim of this paper to show that this is not a sensible thing to do and that one can do better by taking into account the correlations [5] and particularly by accounting for the statistical errors of the covariance matrix.

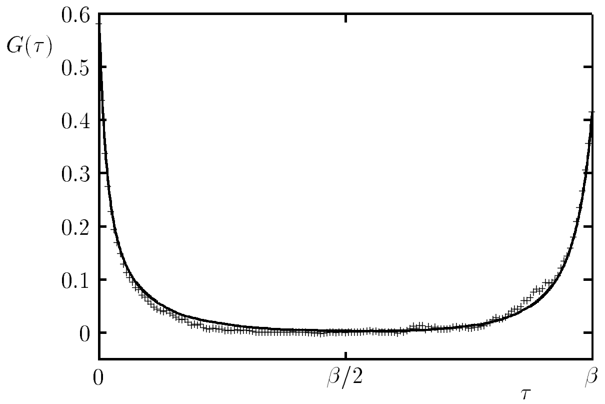

Figure 1. The data from one bin (points) compared to the final average over 800 bins (line).

In the following we will discuss the results of the MaxEnt-procedure considering three cases: (1) Neglecting any information of the covariance matrix, (2) using the covariance matrix only, and (3) taking into account both the covariance and its errors. To determine the dependence on the number of bins and to give an quantitative argument for the amount of the computational effort, which has to be taken, we examined spectra resulting from QMC-data for 200, 400 and 800 bins (160000, 320000 and 640000 sweeps, respectively). The number of time-slices, which corresponds to the dimension of the covariance matrix, is 160 in the present study.

Starting with 200 bins, which is slightly larger than the number of time slices, one can see in Fig. 2 that neither the MaxEnt reconstruction of the plain data (Fig. 2a) nor the additional use of the covariance matrix (Fig. 2b) gives a reliable result. In both spectra the structures are too pronounced and at the wrong position. It appears that the results are generally better if the covariance matrix is not included, since the additional information is treated as exact data-constraints although it suffers from pronounced statistical noise. If, however, the statistical errors of the covariance matrix are taken into account (Fig. 2c) the result reproduces the ED-result very well. There is a small overestimation of the spectral weight at $\omega \approx -3$ only.

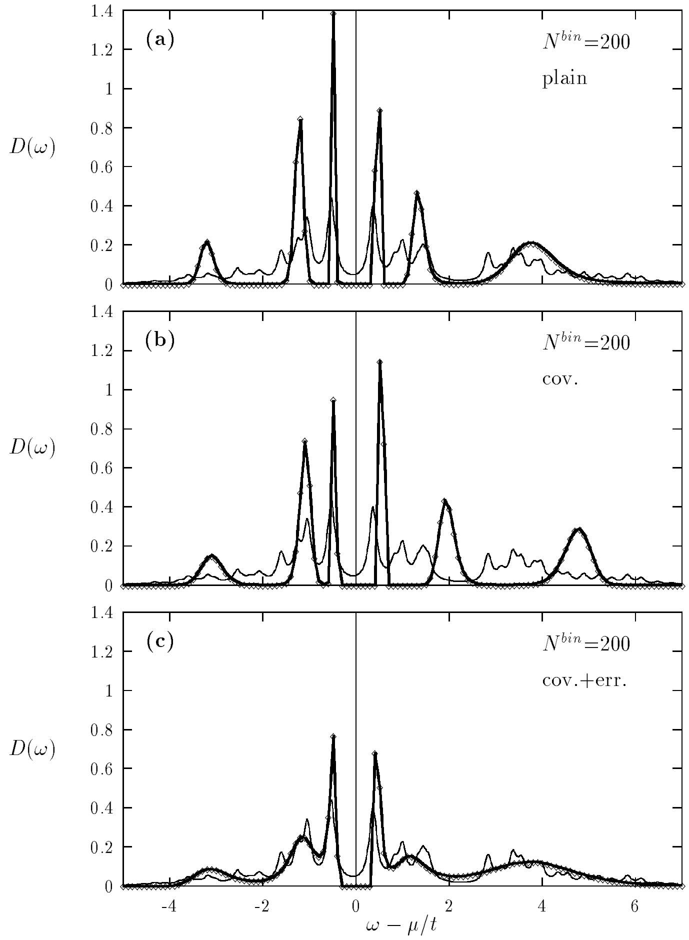

Figure 2. Comparison of the MaxEnt-spectra (thick line) with the ED-result (thin line) for 200 bins: (a) without use of the covariance (first case, see text), (b) with covariance, (c) with the error of the covariance matrix

Increasing the number of bins to 400 (not shown, see [14]) gives still an overfitted result for the first case. Taking the covariance matrix into account the spectrum is slightly improved for $\omega > 0$, but for $\omega \approx -2$ the spectral weight is suppressed completely. Again the best spectrum is obtained if the errors of the covariance matrix are properly accounted for. Eventually for the large number of 800 bins all three spectra show satisfactory results.

The convergency of the various approaches is reasonable, since with increasing number of bins, the correlation of the QMC-errors for different imaginary times vanishes and the covariance matrix becomes diagonal and the covariance of the errors can be ignored. At the same time, the errors of the covariance matrix decrease and assuming exact data-constraints becomes also exact.

Further investigations of the effect of the error of the covariance matrix revealed that the procedure can be simplified by assuming a sufficiently large constant relative error (in our case 20%). The results show no significant deviation from the results obtained by taking the correctly generated errors.

6. Conclusion

Imaginary time data, obtained by standard QMC simulations, suffer from strongly correlated statistical errors if only a small sample is used. It appears reasonable to include the additional knowledge provided by the covariance matrix. But it has been generally observed, the QMC data of the covariance matrix are useless if the sample size is smaller than at least two times the number of time-slices. We have shown that treating consistently the covariance matrix and its errors in the Bayesian frame reliable results can be obtained almost regardless of the number bins. At least our approach represents a way of acquiring reliable data if only a small sample is at hand.

7. Acknowledgments

We would like to thank R. Silver, S.R. White and J. Gubernatis for stimulating discussions. R.P. is grateful for support by the DFG-project Ha 1537/12-1. The calculations were performed at HLRZ Jülich and at LRZ München.

References

1. G. Bednorz and K. A. Müller *Z. Phys. B*, **64**, p. 188, 1986.
2. V. J. Emery *Phys. Rev. Lett.*, **58**, p. 2794, 1987.
3. R. N. Silver, D. S. Sivia, and J. E. Gubernatis *Phys. Rev. B*, **41**, p. 2380, 1990. See references therein.
4. R. N. Silver et al. *Phys. Rev. Lett.*, **65**, p. 496, 1990.
5. M. Jarrell and J. E. Gubernatis, "Bayesian inference and the analytic continuation of imaginary-time quantum monte carlo data," *Phys. Rep.*, 1995. To be published.
6. J. Skilling in *Maximum Entropy and Bayesian Methods*, J. Skilling, ed., p. 45, Kluwer Academic, Dordrecht, 1989.
7. S. F. Skilling in *Maximum Entropy and Bayesian Methods*, J. Skilling, ed., p. 53, Kluwer Academic, Dordrecht, 1989.
8. B. Buck and V. A. Macaulay, eds., *Maximum Entropy in Action*, Oxford Science, Oxford, 1990.
9. W. von der Linden, M. Donath, and V. Dose *Phys. Rev. Lett.*, **71**, p. 899, 1993.
10. W. von der Linden *Appl.Phys.*, **A60**, p. 155, 1994.
11. J. Skilling in *Maximum Entropy in Action*, B. Buck and V. A. Macaulay, eds., p. 19, Oxford Science, Oxford, 1990.
12. R. Silver and H. Martz in *Maximum Entropy and Bayesian Methods*, G. Heidbreder, ed., Kluwer Academic, Dordrecht, 1993. To be published.
13. S. R. White *Phys. Rev. B*, **44**, p. 4670, 1991.
14. W. von der Linden, R. Preuss, and W. Hanke, "Consistent application of maximum entropy to quantum monte carlo data," submitted to Phys. Rev. B (1995).