Consistent application of maximum entropy to quantum Monte Carlo data

1996 J. Phys.: Condens. Matter 8 3881

(http://iopscience.iop.org/0953-8984/8/21/013)

J. Phys.: Condens. Matter **8** (1996) 3881–3888. Printed in the UK

Consistent application of maximum entropy to quantum Monte Carlo data

W von der Linden†, R Preuss‡ and W Hanke‡

† Max-Planck-Institut für Plasmaphysik, EURATOM Association, D-85740 Garching b. München, Germany

‡ Institut für Theoretische Physik, Universität Würzburg, Am Hubland
D-97074 Würzburg, Germany

Received 18 December 1995, in final form 25 March 1996

Abstract. Bayesian probability theory along with the maximum entropy concept is widely used for inferential problems, particularly to infer dynamic properties of strongly correlated fermion systems from quantum Monte Carlo (QMC) imaginary time data. In current QMC applications, however, the error covariance of the QMC data is not treated consistently. Here we present a consistent Bayesian analysis of all the information provided by the QMC simulation. This approach allows us to infer reliable results with the least amount of computer time.

1. Introduction

Bayesian probability theory [1, 2] provides a general and consistent frame for logical inference if there is uncertainty. An important class of problems comprises the determination of distribution functions, say $\rho(x)$, based on incomplete information not providing a unique solution of $\rho(x)$. Bayesian probability theory allows us to exploit any type of testable information, like expectation values $\int \rho(x) f(x) \mathrm{d}x$ of noisy (computer-) experimental data and other forms of prior knowledge. Combined with the entropic prior it is referred to as quantified maximum entropy (MaxEnt) [3, 4] and yields the most uncommittal and unbiased result, given the data constraints and prior knowledge. MaxEnt has originally been introduced to infer celestial images from incomplete and noisy radio-astronomic data. In the sequel, it has been applied successfully to various other data-analysis problems [5, 6, 7, 8].

Here we will focus on the ill-posed inversion problem encountered in quantum Monte Carlo (QMC) simulations [6, 9, 10]. In this field, MaxEnt has become a standard and successful technique to infer dynamic properties of strongly correlated fermion systems from imaginary-time QMC data. QMC simulations yield values for dynamic quantities along the imaginary-time axis for a finite number of times. The inversion is not unique due to the limited number of data and the presence of statistical errors. A direct inversion of the respective Laplace transform would tremendously overfit the data and the desired signal would be lost. Bayesian probability theory provides the consistent frame to separate the signal from the noise. A complication arises, since the errors of the QMC data are correlated. It has been proposed [6] to use the QMC estimate of the error-covariance matrix in a multivariate normal distribution, which assumes that the matrix has no statistical errors. This approach has been heavily debated as it leads in general to less useful results than disregarding the off-diagonal elements in the first place. The reason is that the Bayesian

0953-8984/96/213881+08$19.50

analysis is not performed consistently, by using an *ad hoc* expression for the likelihood which ignores the statistical errors of the QMC estimate of the covariance matrix can, which is usually large, typically of the order 1–10% for diagonal elements and 10–100% for off-diagonal elements. The problem can be avoided by drastically increasing the sample size (N) which is, however, computationally very expensive and rarely feasible. Bayesian probability theory on the other hand provides a perfect tool to exploit whatever information is provided by QMC simulations and irrespective of the sample size.

2. Bayesian formalism

Bayes' theorem allows us to determine the posterior probability (posterior) $P(A|DCH)$ of the sought-for quantity A, given the QMC data D and further information H. The latter summarizes the definition and all hypotheses of the problem. The exact values D_l^{ex} of the quantities measured by QMC simulations are related to A via

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

The QMC values D_l deviate from D_l^{ex} by statistical errors η_l. Due to the QMC-algorithm, the errors are correlated for different indices l and the information about the error-covariance matrix will be denoted by C. Bayes' theorem relates the posterior to the likelihood function $P(D|ACH)$, which contains the error statistics of the QMC data, and the prior probability $P(A|CH)$

$$P(A|DCH) = \frac{P(D|ACH)P(A|CH)}{P(D|CH)}. \tag{2}$$

The most honest prior should summarize all our prior knowledge—the knowledge we have about A prior to receiving the QMC data A—and nothing else, i.e. it should be as ignorant as possible otherwise [2]. In the case of a positive, additive distribution function (PAD), like the spectral density, the most ignorant prior is the entropic prior [11]

$$P(A|CH) = \frac{1}{Z_S}\exp\Big(\alpha \underbrace{\int [A(\omega) - m(\omega) - A(\omega)\ln(\frac{A(\omega)}{m(\omega)})]\mathrm{d}\omega}_{S}\Big). \tag{3}$$

Z_S is the normalization constant guaranteeing $\int P(A|CH)\mathcal{D}A = 1$. The entropy S is expressed relative to the invariant measure (default-model) $m(\omega)$.

Next we determine the likelihood $P(D|ACH)$—the central topic of this paper. It quantifies the probability for the specific data values D_l measured by the QMC simulation, supposing the exact function A is known. Knowing A implies the knowledge of the exact values D^{ex} due to 1 and the likelihood therefore describes the error statistics of the QMC data

$$P(D|ACH) =: \rho(D^{\mathrm{ex}} - D) = \rho(\eta) \quad \text{given C and H.} \tag{4}$$

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

$$C_{ij} = \frac{1}{N-1}\langle\eta_i\eta_j\rangle = \frac{1}{N-1}\langle(D_i^{\mathrm{ex}} - D_i)(D_j^{\mathrm{ex}} - D_j)\rangle \tag{5}$$

measured by QMC sample of size N. To begin with, we assume that the error-covariance is exact. The resulting problem is to determine the PAD $\rho(\eta)$ given the constraints

$$C_{ij} = \int \rho(\eta)\eta_i\eta_j\mathrm{d}\eta \tag{6}$$

$$1 \quad = \int \rho(\eta)\mathrm{d}\eta. \tag{7}$$

This is a problem falling into the realm of Jaynes' MaxEnt [1, 12], which is analogous to deriving the barometric formula, Maxwell's velocity distribution, or Fermi and Bose statistics. In this framework, $\rho(\eta)$ is obtained upon maximizing the entropy, subject to the data constraints. Treating the constraints with Lagrange parameters we are out for the maximum of

$$\mathcal{L} = S - \sum_{ij} \lambda_{ij} \left(\int \rho(\eta)\eta_i\eta_j \mathrm{d}\eta - C_{ij} \right) - \lambda_0 \sum_i \left(\int \rho(\eta)\mathrm{d}\eta - 1 \right) \tag{8}$$

with λ_0, λ_{ij} being Lagrange parameters. Upon maximizing $\mathcal{L}$ with respect to $\rho(\eta)$ one obtains an analytic expression for the solution

$$\rho(\eta) = \frac{1}{Z} \mathrm{e}^{-\sum_{ij} \lambda_{ij}\eta_i\eta_j}. \tag{9}$$

An ignorant, flat default model ($\int m(\omega)\mathrm{d}\omega = 1$) has been assumed. Z is determined via the normalization constraint

$$Z = \int \mathrm{e}^{-\sum_{ij} \lambda_{ij}\eta_i\eta_j} \mathrm{d}^N\eta = \frac{\pi^{N/2}}{\sqrt{\det(\lambda_{ij})}}. \tag{10}$$

The covariance constraint implies

$$C_{ij} = -\frac{\partial \ln(Z)}{\partial \lambda_{ij}} = \frac{1}{2}(\lambda^{-1})_{ij} \quad \Rightarrow \lambda_{ij} = \frac{1}{2}(C^{-1})_{ij}. \tag{11}$$

Hence the likelihood is the ubiquitous multivariate normal distribution

$$\rho(\eta) = \frac{1}{\sqrt{\det(2\pi C)}} \mathrm{e}^{-\frac{1}{2}\sum_{ij} \eta_i C_{ij}^{-1} \eta_j} \tag{12}$$

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

$$\rho(\eta) = \frac{1}{\sqrt{\prod(2\pi\sigma_i^2)}} \mathrm{e}^{-\frac{1}{2}\sum_i \frac{\eta_i^2}{\sigma_i^2}}. \tag{13}$$

Unfortunately, this handy result is only valid if the covariance matrix is known exactly, which is not the case. Therefore we proceed one step further in the Bayesian hierarchy and treat the errors of the covariance matrix on the same footing as the errors of the QMC data D in the first place—by quantified MaxEnt [3]. The additional information available by QMC simulations is the estimate σ_{ij} of the variance of the individual matrix elements C_{ij}. Further reliable information cannot be taken from present QMC simulations. Hence Bayesian probability theory is invoked to infer $\rho(\eta)$ from the QMC error covariance C_{ij} and the variances σ_{ij} of C_{ij}. The posterior for $\rho(\eta)$ reads

$$P(\rho|C\sigma H) = \frac{P(C|\rho\sigma H)P(\rho|H)}{P(C|H)}. \tag{14}$$

Superfluous conditions are suppressed. Again the entropic prior is used. The likelihood $P(C|\rho\sigma H)$ factorizes for the individual matrix elements C_{ij} since they are *logically* independent

$$P(C|\rho\sigma H) = \prod_{i<j} P(C_{ij}|\rho\sigma_{ij}H). \tag{15}$$

The individual factors describe the probability distribution for a random variable, say x, if its expectation value $\overline{x}$ and an estimate σ of the true variance $\hat{\sigma}$ are given. The proper statistics in this case is formed by the random variable [14]

$$t = \frac{x - \overline{x}}{\sigma/\sqrt{f}} \tag{16}$$

which follows a Student t-distribution for $f = N - 1$ degrees of freedom [14]

$$P(t) \propto \left(1 + t^2/f\right)^{-\frac{f+1}{2}}. \tag{17}$$

The result was derived in the framework of Bayesian probability theory by Bretthorst [15] upon marginalizing over the true variance $\hat{\sigma}$

$$P(x|\overline{x}\sigma) = \int_0^\infty P(x|\overline{x}\hat{\sigma}\,H)\,P(\hat{\sigma}|\sigma)\mathrm{d}\hat{\sigma}. \tag{18}$$

According to Bayes theorem $P(\hat{\sigma}|\sigma) \propto P(\sigma|\hat{\sigma})P(\hat{\sigma})$ with $P(\sigma|\hat{\sigma})$ being the χ^2 distribution. The invariant prior $P(\hat{\sigma})$ for a scale variable is Jeffreys' prior [1]. The marginalization leads to (17). In this univariate case, the t-distribution is extremely well approximated by the Gaussian $\exp(-t^2/2)$ already for small samples ($N > 3$)†. The likelihood under consideration is therefore

$$P(C|\rho\sigma H) \propto \exp\Bigg(-\underbrace{\frac{1}{2}\sum_{ij}\frac{(C_{ij} - \int \rho(\eta)\eta_i\eta_j\mathrm{d}\eta)^2}{\sigma_{ij}^2}}_{\chi^2}\Bigg). \tag{19}$$

The variance σ_{ij} of the individual matrix elements of the covariance matrix C_{ij} is determined from the QMC measurements D_i^ν via the standard expression

$$\sigma_{ij} = \frac{1}{N}\sum_{\nu=1}^{N}\left((D_i^\nu - \overline{D_i})(D_j^\nu - \overline{D_j})\right)^2 - \left(\frac{1}{N}\sum_{\nu=1}^{N}(D_i^\nu - \overline{D_i})(D_j^\nu - \overline{D_j})\right)^2 \tag{20}$$

which justifies the use of the Gaussian likelihood in (19). Here ν enumerates the N measurements and $\overline{D_i}$ stands for the QMC expectation value for D_i.

The MaxEnt result for $\rho(\eta)$ is obtained upon maximizing the posterior (14), i.e. $P(\rho|C\sigma H) \propto \exp\left(\alpha S - \frac{1}{2}\chi^2\right)$ or rather

$$\mathcal{L}(\rho, C) = \alpha S - \tfrac{1}{2}\chi^2. \tag{21}$$

It is expedient to perform a Legendre transform

$$\mathcal{L}(\rho, C) \to \tilde{\mathcal{L}}(\rho, \lambda) = \mathcal{L}(\rho, C) - \sum_{ij}\lambda_{ij}C_{ij}$$

$$\lambda_{ij} := \alpha\frac{\partial\mathcal{L}}{\partial C_{ij}} \tag{22}$$

which reveals the remarkable fact that the solution in the case of noisy data has the same functional form (9) as in the case of exact data constraints [13, 8], merely the determination of the Lagrange parameters is modified due to the presence of noise to

$$C_{ij} + \alpha\lambda_{ij}\sigma_{ij}^2 = \int \rho(\eta)\eta_i\eta_j\mathrm{d}\eta = \frac{1}{2}(\lambda^{-1})_{ij}. \tag{23}$$

This relation coincides with (11) for $\sigma_{ij} = 0$. In practice, however, the errors of the covariance matrix are considerable. In particular if only a moderate number of QMC data

† The convergence is much slower in the multivariate case.

is available the inclusion of the errors of the covariance matrix are essential to obtain a meaningful result at all. The covariance matrix C_{ij} is determined from QMC data D_i^ν via $C_{ij} \propto \sum_{\nu=1}^{N} \Delta D_i^\nu \Delta D_j^\nu$, where ν represents the independent measurements (bins) of D_i, as discussed below. It is obvious that the rank of C is less or equal to the number of bins (N). Hence, if the number of bins is less than the dimension of the covariance matrix, the inverse of C, entering (12), does not exist, and the regularization term in (23) is essential (no matter how small σ_{ij}) to determine $\rho(\eta)$. The regularization parameter α entering (23) can be determined either self-consistently upon maximizing the marginal posterior $P(\alpha|C\sigma H)$ [3, 4], 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 [4]. Independent of the α-determination, the key result is that the likelihood function entering the MaxEnt analysis of QMC imaginary-time data is multinormal with an entropy-regularized covariance matrix. The non-linear equation (23) is solved by the Newton–Raphson scheme. It turned out that $\lambda_{ij} = 0.5(C^{-1})_{ij}$, the solution of the noise-free case, is a reasonable starting point for the iteration scheme.

3. 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

$$H = -t \sum_{\langle i,j\rangle,\sigma} \left(c_{i,\sigma}^\dagger c_{j,\sigma} + \mathrm{HC} \right) + U \sum_i n_{i\uparrow} n_{i\downarrow} \tag{24}$$

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

Unfortunately, dynamic properties cannot be measured directly by QMC simulations. Dynamical information is provided by Matsubara Greens functions $D_l = -\langle T_\tau P(\tau_l) Q(0)\rangle$ 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. P and Q are operators which define the correlation function. Here we will consider the one-particle properties of strongly correlated fermions, and the operators are therefore $P = c$ and $Q = c^\dagger$, respectively. In order to determine the spectral density $A(\omega)$ for real frequencies ω the spectral theorem (1) is applied which is 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 $A(\omega)$,

$$\mu_m = \int \omega^m A(\omega) \mathrm{d}\omega \tag{25}$$

which are given by commutation relations, $\mu_m = [[c, H]_m, c^\dagger]$, and are of simple shape for $m = 1, 2$ [16].

4. 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 behaviour is achieved for this system size and a comparison with the $T = 0$ ED results is possible. After obtaining thermal equilibrium, up to 640 000 sweeps through the space–time Ising-fields were performed.

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 behaviour at least 200 measurements, each separated by four sweeps, have to be accumulated to form one bin. Then the results of a certain number of bins are used for the inversion process.

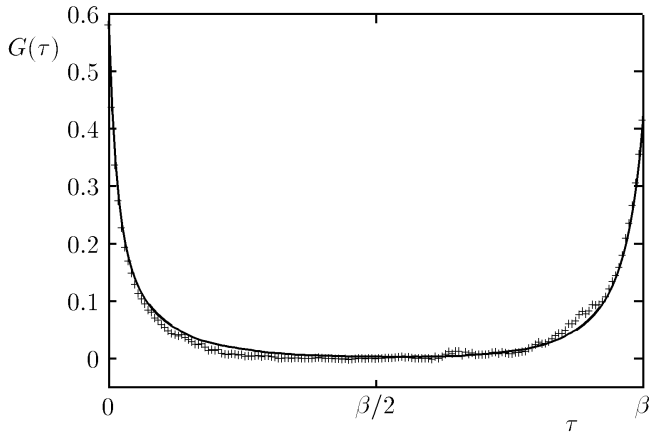

Figure 1. The data from one bin (points) compared to the final average over 800 bins (line).

But binning the data does not suffice to get rid of all the correlations [17]. Still one has to consider the correlations in imaginary time τ (i.e. between D_l and $D_{l'}$). In figure 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 and particularly by accounting for the statistical errors of the covariance matrix.

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 a quantitative argument for the amount of the computational effort, which has to be taken, we show spectra resulting from QMC data for 200, 400 and 800 bins (160 000, 320 000 and 640 000 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 (first column in figure 2) that neither the MaxEnt reconstruction of the plain data (figure 2(a)) nor the additional use of the covariance matrix (figure 2(d)) 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,

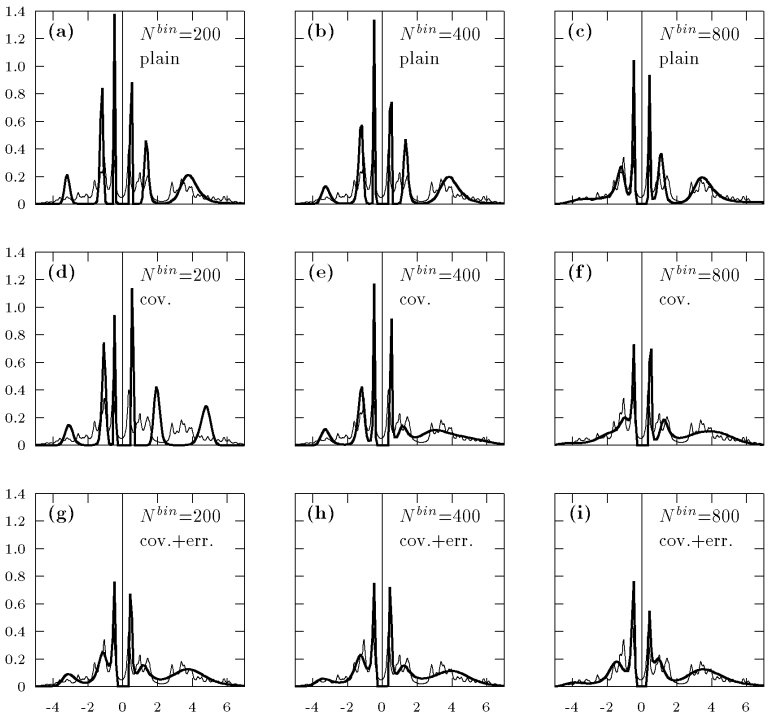

Figure 2. Comparison of the MaxEnt spectra (thick line) with the ED result (thin line) for different parameters. First row (a)–(c), without use of the covariance (first case, see text); second row (d)–(f), with covariance (second case); third row (g)–(i), with the error of the covariance matrix (third case). In dependence of the number of bins: first column (a)–(g), 200 bins; second column (b)–(h), 400 bins; third column (c)–(i), 800 bins.

however, the statistical errors of the covariance matrix are taken into account (figure 2(g)) the result reproduces the ED result very well. There is a small overestimation of the spectral weight at $\omega \approx -3$ only.

Increasing the number of bins to 400 gives still an overfitted result for the first case (figure 2(b)). Taking the covariance matrix into account (figure 2(e)) shows a slightly improved spectrum 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 (figure 2(h)). The maximum at $\omega \approx -3$ is damped to the correct shape and the agreement with the ED result is nearly perfect now.

Eventually for the large number of 800 bins all three spectra show satisfactory results (third column of figure 2). Only in the first case (figure 2(c)) the MaxEnt curve decreases still too fast for $\omega > 5$ leading to a wrong width.

The convergence 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.

In principle the method we suggested is one further step in a hierarchy of probabilities. One could argue that it would be necessary to include the errors of the variance of the

covariance matrix as well. To illustrate that this is not the case we studied the robustness of the presented scheme against changes in the variances σ_{ij}. It turned out that, as long as the errors are in the right order of magnitude the details do not matter. Even a constant relative error (which is certainly no valid approximation for the errors of a Poisson process) yields no significant deviations from the results obtained by taking the correct data. The statistical errors of the covariance matrix were in the problem under consideration about 20%. In summary, the recovered spectral density depends strongly on the statistical errors of the covariance matrix which can lead to spurious structures. It is therefore necessary to take these errors into account, e.g. as presented in this paper by using the entropy-regularized covariance matrix. The accuracy of the statistical errors, however, has little impact on the sought-for spectral density.

5. Conclusion

Bayesian probability theory is the quantification of inductive logic if there is uncertainty. It allows an unbiased inference from any type of testable information, like e.g. incomplete and noisy data and prior knowledge. In the case of the inversion problem posed by QMC simulations, the information consists of exact sum-rules, estimates for expectation values, the variance-covariance matrix, and the statistical errors of these matrix elements. We have shown that only if the covariance matrix and its errors are treated consistently in the Bayesian frame reliable results are obtained regardless of the number bins.

Acknowledgments

We would like to thank R Silver and S R White for stimulating discussions and we are grateful to the Bavarian FORSUPRA program on high-T_c research for financial support. The calculations were performed at HLRZ Jülich and at LRZ München.

References

[1] Jaynes E T 1983 *Papers on Probability, Statistics and Statistical Physics* ed R Rosenkrantz (Dordrecht: Reidel)
[2] O'Hagan A 1994 *Kendall's Advanced Theory of Statistics*, 2b, *Bayesian Inference* (Halsted)
[3] Skilling J 1989 *Maximum Entropy and Bayesian Methods* ed J Skilling (Dordrecht: Kluwer) p 45; 1990 *Maximum Entropy and Bayesian Methods* ed P F Fougère (Dordrecht: Kluwer) p 341
[4] Gull S F 1989 *Maximum Entropy and Bayesian Methods* ed J Skilling (Dordrecht: Kluwer) p 53
[5] Buck B and Macaulay V A (eds) 1990 *Maximum Entropy in Action* (Oxford: Oxford Science)
[6] Silver R N, Sivia D S and Gubernatis J E 1990 *Phys. Rev.* B **41** 2380 and references therein.
[7] von der Linden W, Donath M and Dose V 1993 *Phys. Rev. Lett.* **71** 899
[8] von der Linden W 1994 *Appl. Phys.* **A60** 155
[9] Silver R N *et al* 1990 *Phys. Rev. Lett.* **65** 496
[10] Gubernatis J E *et al* 1991 *Phys. Rev.* B **44** 6011
[11] Skilling J 1990 *Maximum Entropy in Action* eds B Buck and V A Macaulay (Oxford: Oxford Science) p 19
[12] Kapur J N and Kesavan H K 1992 *Entropy Optimization Principles with Applications* (San Diego: Academic)
[13] Silver R N and Martz H F 1993 *Maximum Entropy and Bayesian Methods* ed G Heidbreder (Dordrecht: Kluwer) to be published
[14] Frieden B R 1991 *Probability, Statistical Optics, and Data Testing* ed T S Huang (Berlin: Springer)
[15] Bretthorst G L 1990 *J. Magn. Res.* **88** 552
[16] White S R 1991 *Phys. Rev.* B **44** 4670
[17] Jarrell M and Gubernatis J 1995 *Phys. Rep.* submitted