Maximum entropy methods in neutron scattering: application to the structure factor problem in disordered materials

A. K. Soper
Neutron Science Division
Rutherford Appleton Laboratory
Chilton, Didcot, Oxon, OX11 OQX
UNITED KINGDOM

Abstract: Maximum entropy methods are becoming increasingly important in the analysis of neutron scattering data, principally for the deconvolution of the instrumental resolution function from measured data, and for the inversion of structure factor data to pair correlation function. It is demonstrated that maximizing the entropy of the estimated distribution does not necessarily guarantee that the inverted distribution is free from artifacts associated with the truncation, noise and systematic effects in the data: the result can depend quite markedly on the assumed prior distribution used to calculate the entropy. For the structure problem a novel exponential weight on the Fourier coefficients is introduced which serves to ensure that the structure factor and its derivatives are continuous. The rate of exponential decay, which is related to the width of the narrowest peak in the structure factor is determined by an inverse correlation length that can be obtained from the data. In this way the results are markedly less dependent on the assumed prior distribution.

1. Introduction

A broad class of problems in neutron scattering involve the inversion of a set of measurements, the data D_i, to a desired distribution function, N_j, the N_j being related to the D_i via a transform of some kind:-

$$D_i = \mathrm{Tr}\left\{ N_j \right\}, \qquad i,j= 0,1,\ldots,\infty \tag{1}$$

This inversion is often impossible or ill-conditioned for several reasons:-

(a) the transform may not be linear;
(b) the data may be incomplete, i=i1 ... i2;
(c) the data are measured at discrete points;
(d) the data may be noisy;
(e) the data may have systematic errors.

Because of the ill-conditioning it is likely that several or perhaps a large set of distributions N_j can be regarded as consistent with the measured data.

Over the years a large number of methods have evolved to cope with the variety of difficulties which arise in the inversion of incomplete data. Of these it is claimed, Jaynes (1982), that the Maximum Entropy (ME) approach, which attempts to avoid producing any information which is not justified by the data, provides an independent assessment of all the possible solutions to a particular problem, and so leads to that solution which is "maximally non-committal" with respect to the unmeasured data. This is achieved via an entropy metric which is usually defined as

$$H = - \sum_j N_j \ln \{N_j/P_j\} + \sum_j N_j - \sum_j P_j \quad (2)$$

where P represents a "prior" distribution which incorporates previous knowledge about the distribution not contained in the data. The second and third terms are introduced in (2) in the event that N and P are not normalizable distributions. In the absence of any other information the ME solution, which attempts to maximize H, is simply N = P for all j. When constrained by additional information, i.e. the measured data, the entropy falls below its maximum value. The object of the ME method therefore is to find that solution for N which satisfies the data but which also keeps H as near as possible to its maximum value. It will be seen that this definition of H only exists if $N_j, P_j > 0$.

The goodness of fit to the data is usually measured by a χ-squared statistic or R-factor:

$$R_f^2 = \sum_i (D_i - M_i)^2 / \sum_i D_i^2 \quad (3)$$

where M_i is an estimate (or "model") of the i'th data point obtained from the estimated trial N distribution via (1). With this definition a "quality factor", or Q-factor, which represents how well a particular solution satisfies the dual constraints of entropy maximization and fit to the supplied data, is defined as

$$Q_f^2 = - H + xR_f^2, \quad (4)$$

where x is an undetermined positive multiplier which controls how closely the model fits the data. Therefore it is the Q-factor which is to be minimized, with x determined by constraining the R-factor to a predefined value.

In setting up the ME solution for a given experimental situation, there are two questions that need to be confronted. Firstly what is the most appropriate distribution space, the N distribution, in which entropy is to be calculated? Secondly, what is the most reasonable choice for the prior distribution, or P distribution. It is frequently assumed that the prior distribution should be uniform even though the existence of the data implies that the real distribution is anything but uniform. Unfortunately both of these questions are often ignored in the literature, there being an implicit assumption that somehow maximizing entropy will cover up all the difficulties. In the sections that follow I will apply the ME method to the problem of calculating the pair correlation function for a liquid or amorphous material from structure factor data. I will demonstrate that the obvious choice for the distribution N is in fact quite inappropriate in this case, and that the result of ME analysis can depend markedly on the choice of prior distribution. Full details on the correct choice for the N and P distributions for the entropy estimation for this problem are given elsewhere, Soper (1988), but some examples of the results are shown here. The conclusions to be drawn are applicable to many other applications where ME techniques are used.

2. Solution of the Maximum Entropy Problem

The general solution of the ME problem is a highly non-linear problem and several solutions exist, mostly using sophisticated search procedures, Bryan and Skilling (1984). I have a developed a Monte Carlo (MC) solution to this problem which has several attractive features. In particular it is simple to execute, can allow error bars on the calculated distributions to be estimated if needed, and by virtue of the stochastic process intrinsic to MC calculations is unlikely to get stuck in local phase space minima. The object of the MC calculation is to set up an ensemble of distributions such that each member occurs with probability

$$p(Q_f^2) = \exp\{-\lambda Q_f^2\} \tag{5}$$

with λ a positive multiplier which determines the size of the Q-factor and its fluctuations: as λ is made larger so the Q-factor is driven smaller and the fluctuations become smaller. Full details of this algorithm are given elsewhere, Soper (1988), and will not be elaborated further here. However it will be noted that in all the examples given below λ is kept as large as possible, so that fluctuations are held to a minimum and the individual trial distributions lie indistinguishably close to the ME solution. Typical run times for this algorithm, which might involve 500,000 individual moves, are ~10 minutes of cpu on a VAX 8650, assuming 200 data points and 500 points in the N distribution.

3. The Structure Problem in Disordered Systems

The underlying transform in the structure of liquids and amorphous materials is in principle a straightforward Fourier transform:

$$S(Q) = 4\pi\rho \int_0^\infty r\{ g(r)-1 \} \sin(Qr)\, dr \tag{6}$$

where S(Q) is the measured structure factor, as a function of wave vector transfer, Q, and g(r) is the underlying pair distribution function as a function of radial distance r from an atom at the origin. The atomic number density is ρ. Inverting this transform directly can lead to significant transform errors because the data can never be measured over a complete range of Q values and in any case invariably contain some form of error, statistical or systematic. Typically one introduces the constraint of only calculating g(r) at certain values of r, according to the Lado (1971) rules for Fourier transforms, i.e. $\Delta r = \pi/Q_{max}$. Furthermore a "window" function is often invoked to further reduce the effects of noise in the calculated distribution function.

This problem is readily amenable to ME analysis. In particular there is apparently an obvious choice for the N distribution by virtue of the normalization

$$4\pi\rho \int_0^\infty r^2\{\, g(r)-1 \,\}\, dr = -1 + \rho\chi k_B T \tag{7}$$

where χ is the isothermal compressibility and T is the absolute temperature. Hence the "obvious" choice for the N distribution is simply

$$N_j = N(r_j) = 4\pi\rho\, r_j^2 g(r_j)\Delta r \tag{8}$$

where Δr is the bin width of the discretized distribution. Figures 1 and 2 show the effects of applying ME analysis to the problem of transforming the hard sphere structure factor to pair correlation function. In this case the input S(Q) is known exactly within the Percus-Yevick approximation but is a particularly severe test of any transform method because g(r) for hard spheres is discontinuous at $r=\sigma$, the hard core diameter. In this case the density was chosen such that $\rho\sigma^3 = 0.5$, and a large Q limit of $Qmax = 15/\sigma$ was imposed on the S(Q) data, with $\sigma = 1$Å. In figure 1 the prior distribution is

$$\begin{aligned} P_j &= 0 && \text{for } r_j < \sigma \\ p_j &= 4\pi\rho r_j^2 \Delta r && \text{for } r_j > \sigma, \end{aligned} \tag{9}$$

while for figure 2 the prior distribution is set at

$$\begin{aligned} P_j &= 0 && \text{for } r_j < 0.75\sigma \\ p_j &= 4\pi\rho r_j^2 \Delta r && \text{for } r_j > 0.75\sigma. \end{aligned} \tag{10}$$

In either case the fit to the data was the same, (R-factor = 1%); however it is readily apparent that the two results for g(r) are not the same. In particular the distribution in figure 2 has greater entropy than figure 1 (-2.064 for figure 2 compared to -2.534 for figure 1) when measured against the uniform prior, thus confirming that the algorithm has found the true maximum entropy solution for figure 2. This result is apparently at odds with our intuition which might tend to favour the one in figure 1 as being less "structured", if the word "structure" in this case is taken to indicate the number and size of peaks and valleys in the calculated distribution. In fact the distribution in figure 1 is very close to the known exact solution, Throop and Bearman (1965).

The difference between the two solutions is manifested in Q space not in the region of the input data, where the two solutions give equally good fits, but beyond the input region. Figure 3 shows the calculated structure factor for the distribution in figure 2, and also the difference between model and data. It is seen that immediately beyond the data (Q ~ 15-16 Å^{-1}) there is a strange cusp in the S(Q) from figure 2, a phenomenon which has been seen before, Root, Egelstaff and Nickel (1986). The ripples seen in figure 2 become suspect when it is realized they have a period of ~ $2\pi/Qmax$, where Qmax is the largest Q value for the input data. Hence it is concluded that maximizing the entropy has not avoided the truncation ripples associated with the discontinuity in the input data at Q=Qmax. For measured datasets which usually have a noise component, the discontinuities become important since there is effectively a discontinuity at every data point.

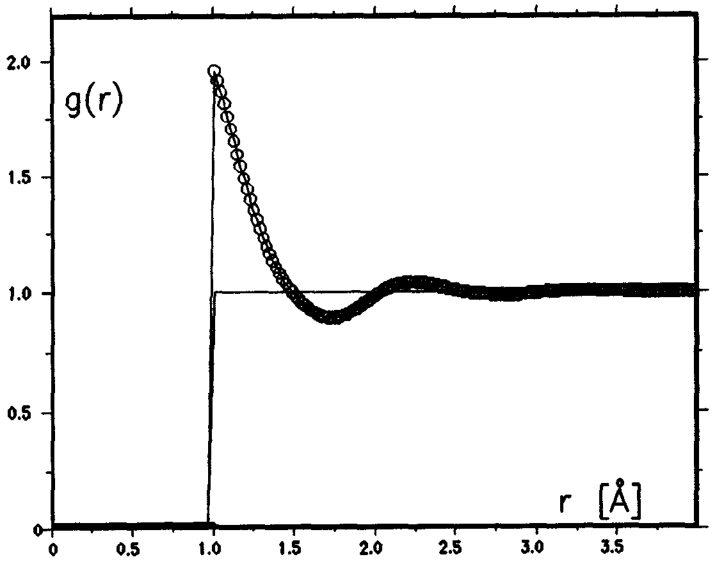

Figure 1 Maximum entropy pair correlation function derived from Percus Yevick hard sphere structure factor. The S(Q) data were truncated at Q = 15Å^{-1}, and the prior distribution used is zero in the region r = 0 to r = 1Å. The circles show the calculated distribution and the line shows the prior.

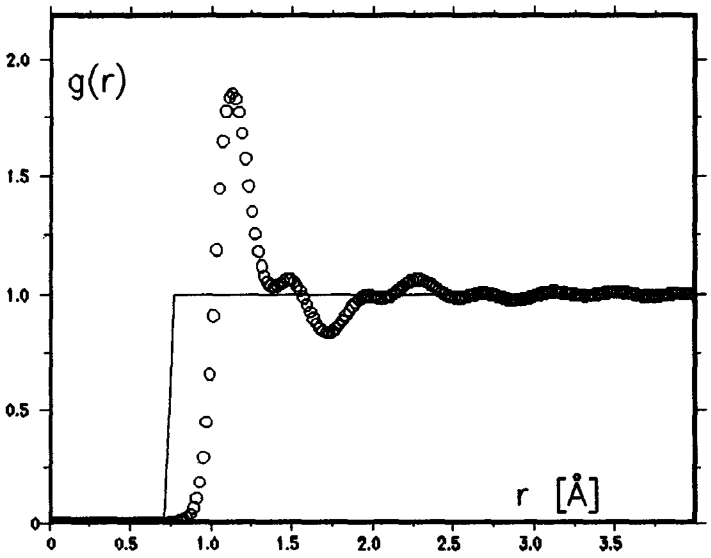

Figure 2 Same calculation and notation as for figure 1 except that the prior is zero in the region r = 0 to r = 0.75Å.

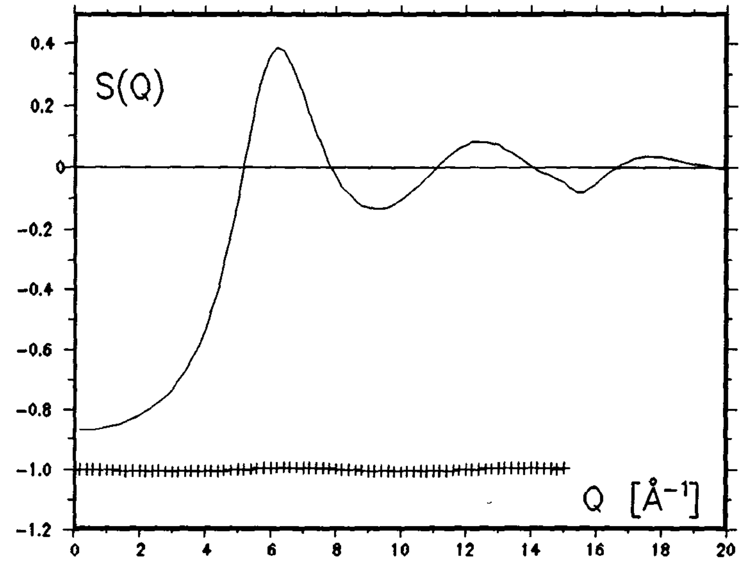

Figure 3 Maximum entropy structure factor corresponding to the correlation function of figure 2. The crosses show the residual between fit and data shifted below zero by unity.

4. Solution of the Discontinuity Problem

I propose two solutions to the problem of discontinuities. The first is to effectively force S(Q) to be everywhere continuous and have continuous derivatives. To see how to do this a well known theorem from Fourier transforms is invoked, Lighthill (1959). From (6) it can be seen that the n'th derivative of S(Q) is given by

$$S^n(Q) = 4\pi\rho(-1)^{n/2} \int_0^\infty r^{n+1}\{ g(r)-1 \} \sin(Qr)dr \qquad \text{for n=even} \quad (11)$$

and

$$S^n(Q) = 4\pi\rho(-1)^{(n-1)/2} \int_0^\infty r^{n+1}\{ g(r)-1 \} \cos(Qr)dr \qquad \text{for n=odd} \quad (12).$$

Therefore if $\{g(r)-1\}$ converges slower than $1/r^{n+2}$ as $r \to \infty$, there will be discontinuities in the n'th derivative. On the other hand if $\exp(\alpha r)\{g(r)-1\}$, where α is a finite positive number, is convergent as $r \to \infty$ then $r^{n+2}\{g(r)-1\}$ is also convergent at large r for all n. For the hard sphere pair correlation function the exponential decay of $\{g(r)-1\}$ with increasing r is an analytic consequence of the theory which describes the hard sphere structure factor, Perry and Throop (1972). For other liquid and amorphous structures the requirement that S(Q) be continuous and have continuous derivatives is a necessary consequence of there being no long range order in the material.

This exponential constraint leads to a simple revision of the definition of the N and P distributions, namely

$$N_j' = \exp(\alpha r_j) * N(r_j) = 4\pi\rho \exp(\alpha r_j) r_j^2 g(r_j)\Delta r \qquad (13)$$

and

$$P_j' = 0 \qquad \text{for } r_j < \sigma$$
$$P_j' = 4\pi\rho \exp(\alpha r_j) r_j^2 \Delta r \qquad \text{for } r_j > \sigma. \qquad (14)$$

The primed distributions are used instead of the unprimed distributions in the definition of entropy, equation (2). Otherwise the calculation proceeds as before. The inverse correlation length, α, is determined from the width of the narrowest peak in S(Q), or by inspection of the large r behaviour of g(r). Hence for a previously unknown dataset it may be necessary to redetermine its value once an initial solution has been achieved.

Figures 4 and 5 show the results of applying this exponential constraint in the definition of entropy, with $\alpha = 1.8/\text{Å}$. It can be seen that the ripples in figure 2 have been largely eliminated in figure 4, and that the problem of the cusp in S(Q) at Q=Qmax has now been erradicated (figure 5). The fit is a good as before (R-factor = 1%), and the entropy is only marginally lower than for figure 2, being equal to -2.096.

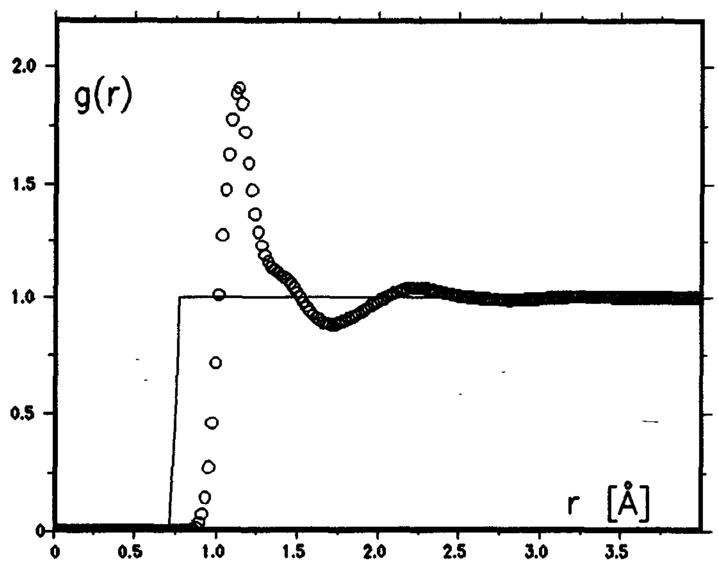

Figure 4 Maximum entropy pair correlation function as for figure 2, but this time derived using the exponential weighting on the distributions used to calculate the entropy, as described in the text.

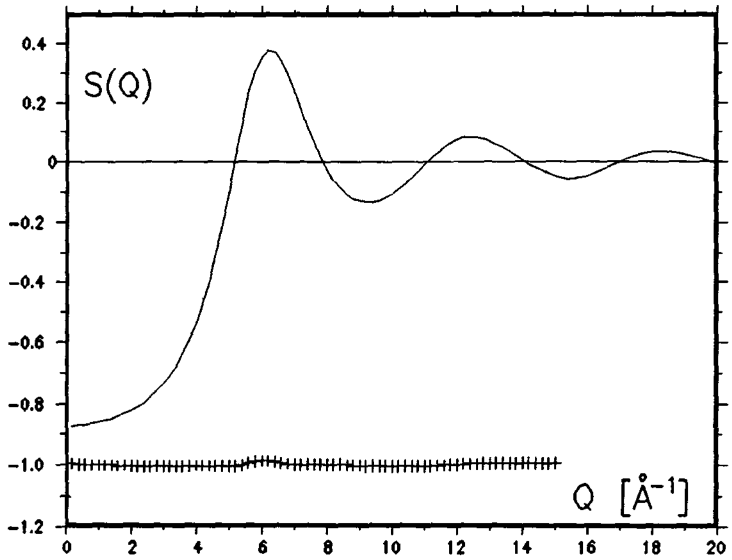

Figure 5 Maximum entropy structure factor corresponding to figure 4. Note that the cusp near Q = 15/Å is absent in this case.

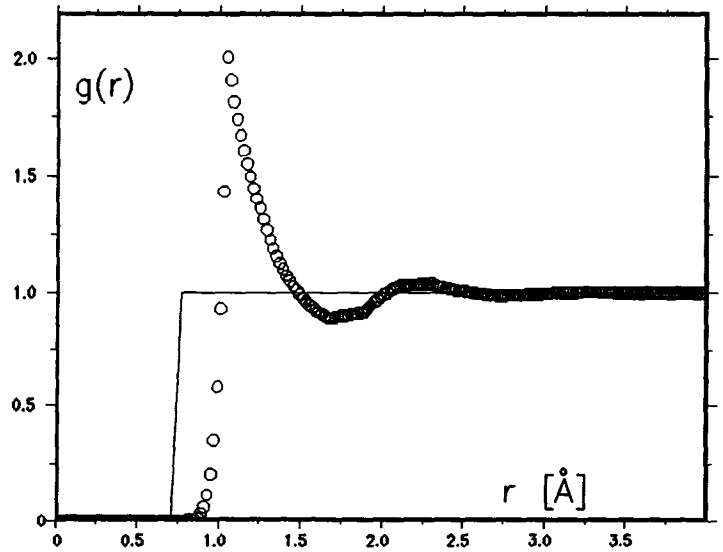

Figure 6 Maximum entropy pair correlation function as before, but this time using the fluctuations in the Q-factor linearly interpolated between the allowed discrete Fourier values.

5. Coping with Truncation Effects

It is apparent in figure 4 that some truncation effects may still be present in the estimated distribution functions. These arise because in estimating the change in Q-factor at each move, there is a sum over the input data which in effect is a Fourier transform of the difference $D_i - M_i$, Soper (1988). Since the change in Q-factor at each move is the driving force behind the calculation this Fourier transform can give rise to exactly the same truncation effects seen in a direct transform of the raw data. My solution to this difficulty is to evaluate the transform only at the allowed r values ($r_j = j\pi/Qmax$), and then interpolate the result onto the required grid of r_i values by linear interpolation.

Figure 6 shows the result of doing this for the same input dataset as before. Now it will be noted that truncation effects are diminished even further: the result is now approaching that of figure 1, but with greater entropy (H = -2.141).

6. Conclusion

The foregoing text has described the application of the ME method to the calculation of the pair correlation function from structure factor data for liquid and amorphous materials. The main conclusion is that ME does not automatically guarantee that the results are free from artifacts associated with noise and truncation in the data. To avoid these artifacts it is necessary to build into the distributions used to calculate entropy known physical constraints which must be satisfied, whatever the detailed form the distributions are to take. For the structure factor problem these constraints include the requirement that the structure factor must everywhere be continuous and have continuous derivatives, and that the fluctuations in the distributions away from the prior distribution are not biassed by the truncation of the input data. Further details and applications of the Monte Carlo algorithm used here are available elsewhere, Soper (1988), as well as a discussion of the present approach in the context of other recent attempts at the structure factor problem.

7. Acknowledgement

I would like to acknowledge invaluable discussions on the subject matter of this paper with S. Baer, T. Gaskell, and N. Rivier.

References

Bryan R K and Skilling J, 1984, Monthly Not. Astr. Soc., 211, 111

Jaynes E T, 1982, Proc. of the IEEE. 70, 939

Lado F, 1971, J. Comput. Phys. 8, 417

Lighthill M J, 1959, "Introduction to Fourier Analysis and Generalised Functions", Cambridge University Press.

Perry G and Throop G J, 1972, J.Chem. Phys. 57, 1827

Root J H, Egelstaff P A, and Nickel B G, 1986, Inst. of Phys. Conf. Ser. No. 81 "Neutron Scattering Data Analysis, 1986", ed. by M W Johnson, IOP Bristol.

Soper A K, 1988, "Proceedings of the International Workshop on Static and Dynamic Properties of Liquids", June 27th-30th, Dubrovnik, Yugosalvia, ed. by M. Davidovich and A.K.Soper, to be published in Springer-Verlag Proceedings in Physics Series; also manuscript in preparation.

Throop G J and Bearman R J, 1965, J.Chem. Phys. 42, 2408