

Maximum entropy—A new approach to the crystallographic phase problem

RAMESH NARAYAN*, RAJARAM NITYANANDA and G V VANI

Raman Research Institute, Bangalore 560080, India

*Present address: Theoretical Astrophysics, California Institute of Technology, Pasadena, California 91125, USA

Abstract. We have recently drawn the attention of crystallographers to the maximum entropy method (MEM) of image reconstruction which has found application in radio astronomy and geophysics. In this article, we review the theory of the MEM as applied to the crystallographic phase problem and discuss points of similarity, and differences, with conventional techniques. We then present a number of illustrative calculations in two and three dimensions on simulated and real structures. These examples show that the MEM can be used for phase refinement from partially known phases, say in macromolecular crystallography, and also for solving crystal structures *ab initio* using only measured Fourier amplitudes. With improved algorithms the method promises to become competitive with current techniques in crystal structure analysis.

Keywords. Crystallography; crystal structure analysis; phase problem; maximum entropy method; direct methods.

1. Introduction

In conventional x-ray crystal structure analysis one measures the intensities of Bragg reflections of the crystal under study. The intensity $I_{\mathbf{H}_j}$ corresponding to the reciprocal vector \mathbf{H}_j is related to the complex Fourier structure factor $F_{\mathbf{H}_j}$ by

$$I_{\mathbf{H}_j} = F_{\mathbf{H}_j} F_{\mathbf{H}_j}^* = |F_{\mathbf{H}_j}|^2, \quad (1)$$

where the structure factors $F_{\mathbf{H}_j}$ and the electron density $\rho(\mathbf{r})$ in the crystal unit cell are related by a Fourier transform:

$$F_{\mathbf{H}_j} = |F_{\mathbf{H}_j}| \exp(i\varphi_{\mathbf{H}_j}) = \int \rho(\mathbf{r}) \exp(2\pi i \mathbf{H}_j \cdot \mathbf{r}) dV_r \quad (2)$$

$$\rho(\mathbf{r}) = \frac{1}{V_r} \sum_{\mathbf{H}_j} F_{\mathbf{H}_j} \exp(-2\pi i \mathbf{H}_j \cdot \mathbf{r}) \quad (3)$$

V_r is the volume of the crystal unit cell. Thus, one experimentally measures only the structure amplitudes $|F_{\mathbf{H}_j}|$ whereas according to (2) and (3) one requires both amplitudes as well as phases $\varphi_{\mathbf{H}_j}$ to compute the electron density. This is the famous phase problem of crystallography—one cannot solve the structure unless one has some estimates of the phases. The magnitude of the problem has been brought out strikingly by Ramachandran and Srinivasan (1970) who have given examples of how true phases, coupled even with random amplitudes, can bring out the true crystal structure; on the other hand, true amplitudes with random phases give no information at all.

Obviously one needs to put in some *a priori* information in addition to the measured amplitudes if one is to obtain a solution. The most powerful input turns out to be the

prior knowledge that the crystal consists of atoms so that the (positive) electron density is strongly clumped in well separated peaks. Further, one generally knows in advance the number and types of atoms present and also has limited stereochemical information. Although a rigorous proof is lacking, all these extra constraints in practice seem to ensure a unique solution to the phase problem, apart from trivial ambiguities related to the choice of origin or enantiomorph. This solution is however not easy to determine!

For small structures involving a few atoms, Patterson techniques work quite well. Here one computes the Fourier transform of the intensities $I_{\mathbf{H}_j}$ ($= |F_{\mathbf{H}_j}|^2$). This map can be shown to be the self-convolution of the electron density (*e.g.* Ramachandran and Srinivasan 1970) and so each peak in it corresponds to an interatomic vector in the original structure. If the peaks are well resolved, which can happen in small structures, one "inverts" this information on interatomic bonds to solve the structure.

Even larger structures can be solved using the Patterson technique provided a few of the atoms are significantly heavier (*i.e.* have more electrons) than the rest; one can obtain the heavy atom positions from the Patterson map and solve the rest of the structure by using the heavy atom derived phases for the structure factors. By far the most challenging aspect of the phase problem in crystallography has been the solution of large structures of nearly equal atoms. Here some very powerful techniques have been developed over the years which go collectively under the name of direct methods (see for example Ladd and Palmer 1980). In a pioneering paper Sayre (1952) mathematically expressed the equality of atoms by the statement that the squared structure $\rho^2(\mathbf{r})$ is very similar to $\rho(\mathbf{r})$ except for a "normalisation". He showed that this immediately leads to relations among the phases of certain sets of reflections—the famous triplet phase relation. Later work has extended this approach to higher order relations and probability distributions have been worked out for the values of sums of specific groups of phases (the so-called structure invariants and seminvariants). Coupled with multisolution (Karle and Karle 1963, 1966; Germain *et al* 1971) methods (where one selects a small set of primary reflections, systematically assigns various phase values to these and in each case solves the phases of the other reflections using triplet and other relations), it can be said that the direct methods, exemplified by the computer program MULTAN and its descendants, are the most powerful techniques available today to solve crystal structures. However, despite the enviable success that these techniques have had, it is probably fair to say that the last word has by no means been said on the phase problem. The search for newer approaches must go on as crystallographers attempt to solve progressively larger structures.

A technique called the maximum entropy method (MEM), which has been studied for some time in other fields, has in the last year or two attracted some interest in crystallography. Two of us (RN and RN) studied the MEM in the context of its application in radio astronomical image processing and drew attention to its potential importance in crystallography (Narayan and Nityananda 1981, 1982). In this article we briefly review some of the work already reported in the field and then describe our more recent results on simulated and real crystal structures in two and three dimensions. On the basis of our present studies we believe that the MEM has a promising role to play in crystallography and merits further investigation. Readers interested in the actual results of this method applied to crystallography could proceed straightaway to §§ 5 and 6, although we hope that this will generate enough interest in the MEM to return to the more theoretical sections.

2. Maximum entropy method

Let us suppose, as is the case in many geophysical and radio astronomical applications, that F_{H_j} is available both in amplitude and phase at some of the H_j but is unknown at the rest of the points. Setting the unknown F_{H_j} to zero leads to a map with large termination ripple as well as loss of resolution. In the MEM, the unknown F_{H_j} are chosen so as to maximise the "entropy" S given by

$$S = \int f[\rho(\mathbf{r})] dV_r, \quad (4)$$

where f is some suitable real function of $\rho(\mathbf{r})$. There is an unresolved debate going on in the literature as to the most suitable form for f . One school of thought including the originator of the MEM (Burg 1967; Ables 1972) favours the entropy

$$S_1 = \int \ln[\rho(\mathbf{r})] dV_r, \quad (5)$$

while others (Frieden 1972; Gull and Daneill 1978) prefer

$$S_2 = - \int \rho(\mathbf{r}) \ln[\rho(\mathbf{r})] dV_r. \quad (6)$$

The arguments in favour of one or the other form are usually based on information theory or combinatorial probabilistic arguments, verging sometimes on the metaphysical. It has been claimed that the MEM produces as featureless a map as possible consistent with the data (Ponsonby 1973) and that it is maximally non-committal with regard to the unmeasured data (Ables 1972). In an interesting paper Komesaroff and Lerche (1978) (see also Komesaroff *et al* 1981) showed that there is an intimate connection in one-dimensional problems between the entropy S_1 and the positivity constraint on $\rho(\mathbf{r})$; however, this does not generalize very well to higher dimensions.

In contrast to the above approaches which seek to attribute a deep foundation to the MEM, there has been another movement initiated by Högbom (1978) and Subrahmanya (1978, 1980) who suggested that the successes of the method are primarily due to the *a priori* information built in, in particular the penalty against negative values. We have extended this view-point (Nityananda and Narayan 1982; Narayan and Nityananda 1983) and carried out a detailed study of the MEM focussing primarily on the maximisation conditions that the reconstruction satisfies. By a simple geometrical argument it was shown that any "entropy" function f which satisfies $f'' < 0$ and $f''' > 0$ implicitly makes the *a priori* assumption that the map consists of isolated sharp features separated by flat extended regions. The forms (5) and (6) as well as several others, *e.g.* $\rho^{1/2}$, $-1/\rho$, etc., satisfy these conditions, and hence all these are suitable "entropies" to generate satisfactory maps. In our view, the MEM is just a variational way of incorporating our *a priori* information that the true map consists of peaks on a flat background. The relevance of the MEM to crystallography then becomes obvious.

Regardless of the controversy over the foundations of the method, it is well known that the MEM produces excellent reconstructions. Figure 1 shows a two-dimensional simulation of a radio astronomical aperture synthesis reconstruction where the MEM map is clearly far superior to the naive approach of setting unmeasured F_{H_j} to zero. Results such as this, coupled with our conviction that the *a priori* assumptions built into

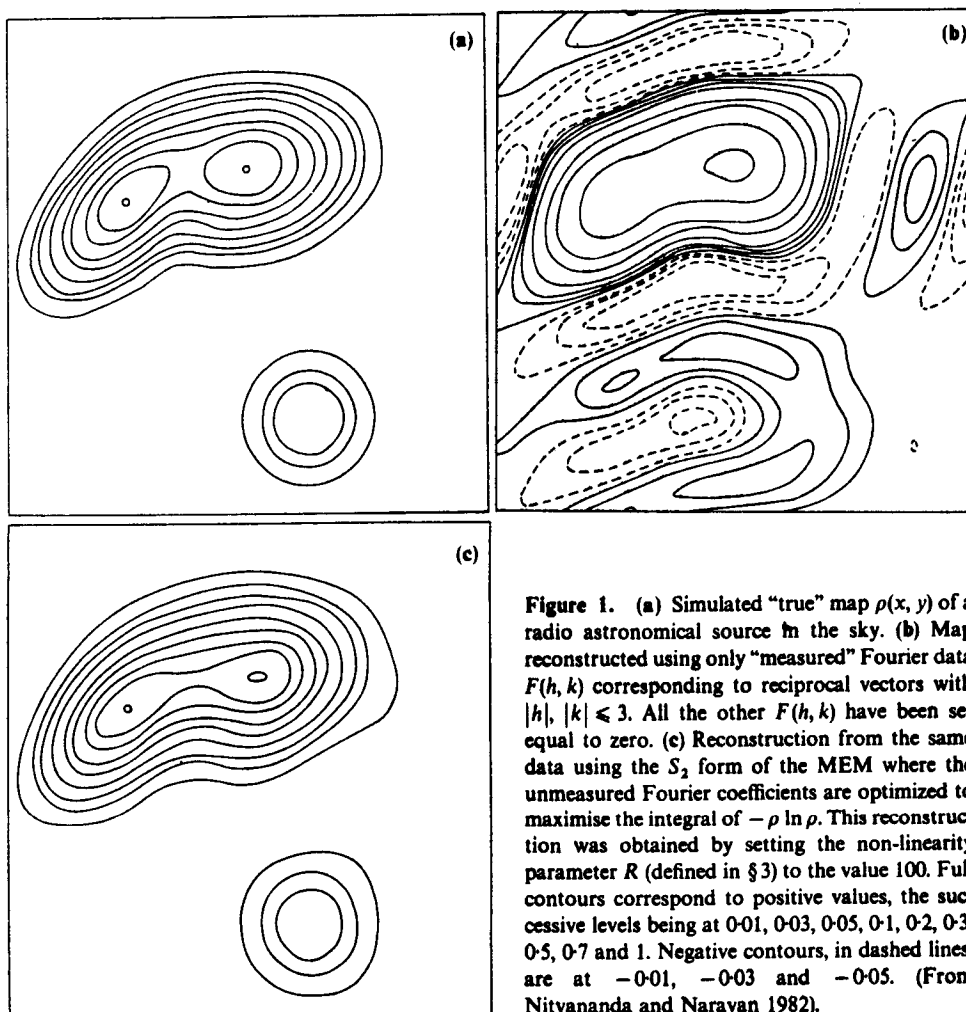


Figure 1. (a) Simulated "true" map $\rho(x, y)$ of a radio astronomical source in the sky. (b) Map reconstructed using only "measured" Fourier data $F(h, k)$ corresponding to reciprocal vectors with $|h|, |k| \leq 3$. All the other $F(h, k)$ have been set equal to zero. (c) Reconstruction from the same data using the S_2 form of the MEM where the unmeasured Fourier coefficients are optimized to maximise the integral of $-\rho \ln \rho$. This reconstruction was obtained by setting the non-linearity parameter R (defined in §3) to the value 100. Full contours correspond to positive values, the successive levels being at 0.01, 0.03, 0.05, 0.1, 0.2, 0.3, 0.5, 0.7 and 1. Negative contours, in dashed lines, are at -0.01, -0.03 and -0.05. (From Nityananda and Narayan 1982).

the MEM are perfectly suited to crystallography, originally led us to propose that the method could help solve the phase problem (Narayan and Nityananda 1981, 1982).

3. Maximum entropy for the phase problem

For the phase problem we write (4) in the form

$$S = \int_V \left[\frac{1}{V_r} \sum_{\mathbf{H}_j} |F_{\mathbf{H}_j}| \exp \{i(\varphi_{\mathbf{H}_j} - 2\pi \mathbf{H}_j \cdot \mathbf{r})\} \right] dV_r, \quad (7)$$

where, because $\rho(\mathbf{r})$ is real, we have

$$\varphi_{\mathbf{H}_j} = -\varphi_{-\mathbf{H}_j}. \quad (8)$$

Let us define the Fourier coefficients $G_{\mathbf{H}_j}$ as follows

$$G_{\mathbf{H}_j} = |G_{\mathbf{H}_j}| \exp(i\theta_{\mathbf{H}_j}) = \int f'[\rho(\mathbf{r})] \exp(2\pi i \mathbf{H}_j \cdot \mathbf{r}) dV_r. \quad (9)$$

Then, maximising the entropy S in (7) with respect to the unknown phases we obtain (Narayan and Nityananda 1981, 1982)

$$\frac{\partial S}{\partial \varphi_{\mathbf{H}_j}} = \frac{2}{V_r} |F_{\mathbf{H}_j}| |G_{\mathbf{H}_j}| \sin(\varphi_{\mathbf{H}_j} - \theta_{\mathbf{H}_j}) = 0. \quad (10)$$

Equations (10) (there are as many equations as the number of unknown phases) have to be satisfied in order to maximise the "entropy". Since $|G_{\mathbf{H}_j}|$ and $\theta_{\mathbf{H}_j}$ are implicitly functions of all the $\varphi_{\mathbf{H}_j}$, these are a highly coupled system of nonlinear equations which will have to be solved numerically by an iterative procedure. However, we note that $\partial S / \partial \varphi_{\mathbf{H}_j}$ are the components of the gradient of S with respect to the unknown $\varphi_{\mathbf{H}_j}$, and the calculation of $|G_{\mathbf{H}_j}|$ and $\theta_{\mathbf{H}_j}$ involves only Fourier transforms which can be very efficiently carried out by means of the fast Fourier transform algorithm. Thus gradient-type algorithms to maximise S (see § 4) are easily developed and these are not likely to be very expensive in terms of computing time.

A crucial issue is the question of the uniqueness of the solution. Of course, for every solution there is a whole family related to it by an origin shift and/or enantiomorph reversal. However, apart from this trivial ambiguity, it is possible that there are several distinct maxima representing physically different solutions. To see this let us write the conditions for the vanishing of the gradient in (10) as

$$\theta_{\mathbf{H}_j} = \varphi_{\mathbf{H}_j} + n_{\mathbf{H}_j} \pi, \quad (11)$$

where the $n_{\mathbf{H}_j}$ are integers which have either of the two values 0 or 1. One can attempt to solve (10) for each choice of the $n_{\mathbf{H}_j}$ (though it is possible that for certain choices there will be no solution at all). In addition, for each choice of the $n_{\mathbf{H}_j}$, there may be more than one distinct solution. Presumably, the majority of such solutions correspond to saddle points in the entropy and only a small fraction constitute true maxima. A function of n phase-like variables can be easily seen to have at least one maximum, one minimum, and n saddle points, using the arguments illustrated for $n = 2$ in the context of solid state physics by Ziman (1972). In any case, it is clear that the phase problem introduces a rich new dimension to maximum entropy. The existence of several local maxima (about which we are convinced on the basis of our computer simulations) means that any numerical technique is liable to be trapped in the wrong maximum. The saving grace is that these subsidiary maxima generally display elements of the true structure and are therefore still useful for the elucidation of the structure.

Even if there are multiple MEM solutions to the phase problem, it is crucial that the true structure itself should be close to a solution—only then can we hope to get something out of the method. At a general level, since the electron density distribution in a crystal is always peaky with a flat baseline, it satisfies the *a priori* input built into the MEM (§ 2) and hence is likely to be a solution. However, one can more rigorously establish this for equal atom structures. Figure 2a schematically shows the electron density $\rho(x)$ (in one dimension) of a "structure" of equal symmetric well-resolved atoms. Let us for convenience focus on the entropy S_1 where $f(\rho) \equiv \ln(\rho)$. Then

$$f'(\rho) \equiv 1/\rho, \quad (12)$$

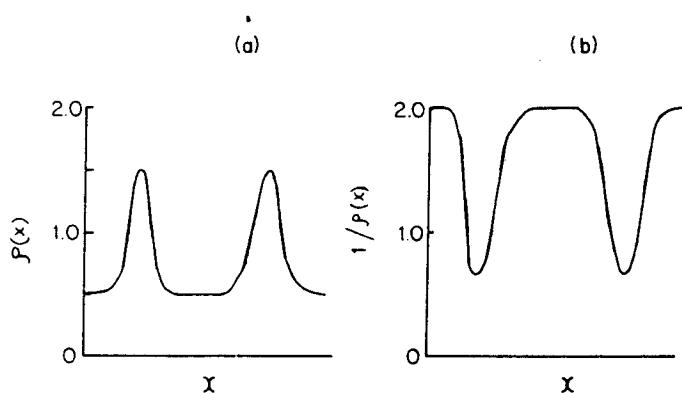


Figure 2. (a) Schematic electron density $\rho(x)$ of a one-dimensional structure with two equal "atoms"; a constant "DC value" of 0.5 has been added. (b) Variation of $1/\rho(x)$, showing that it is also a structure with two equal atoms; the atoms here are negative, have a different shape from those in (a) and hang from a DC level of 2.

and its shape as a function of x for the true structure is shown in figure 2b. It is clear that

$$f'[\rho_{\text{true}}(x)] = K + \rho_{\text{true}}(x) * S(x), \quad (13)$$

where K is a constant, $*$ represents a convolution and $S(x)$ is a symmetric "shape" function. Taking the Fourier transform of (13) we have for any non-origin H_j

$$G_{H_j} = F_{H_j} S_{H_j}, \quad (14)$$

where S_{H_j} , the H_j th Fourier coefficient of $S(x)$, is a real number (positive or negative). Thus we automatically have

$$\theta_{H_j} = \varphi_{H_j} \quad \text{or} \quad \varphi_{H_j} + \pi, \quad (15)$$

depending on the sign of S_{H_j} . This is identical to the conditions written in (10) or (11) and shows that equal atom structures are automatically solutions of the MEM. As a further bonus, since the shape function $S(x)$ is usually known, one has full information on S_{H_j} and hence one can, in principle, determine all the n_{H_j} in (11). This greatly reduces the number of solutions to the problem. It is clear that the above considerations work for all forms of the entropy and are dimension-independent.

In the case of non-equal atom (*i.e.* heavy atom) structures, there is a rather subtle point involved. It will be noticed that both the entropies S_1 and S_2 ((5) and (6)) are defined only for positive values of the argument. Hence, at any intermediate stage of refinement, when one has a lot of spurious features in the map, including negative excursions of the electron density, it is necessary to add a suitable constant C to $\rho(\mathbf{r})$. Any value of C larger than the greatest negative value of $\rho(\mathbf{r})$ is enough to make the problem well-defined. However, it is clear that the reconstructed solutions now have a non-trivial dependence on the value of C . Rather than look at C itself, in our work we have found it useful to employ another parameter R given by

$$R = f''(\rho_{\min} + C) / f''(\rho_{\max} + C). \quad (16)$$

We have borrowed this concept from our earlier work in radio astronomy (Nityananda and Narayan 1982) where we showed that R is a measure of non-linearity. For the

present purposes it is sufficient to note that large values of C imply small values of R (however, R is always > 1) while the smallest permitted C (*viz.* $-\rho_{\min}$) gives $R \rightarrow \infty$. It is also clear that as $R \rightarrow 1$, f'' is constant and f' a linear function of ρ . The phases θ_{H_j} of G_{H_j} in (9) are then just φ_{H_j} or $\varphi_{H_j} + \pi$ and all maps satisfy the MEM conditions (11). Now, in our experience, low values of R lead to a stable centrosymmetric solution even when the true structure is non-centric. When R is increased, the symmetry is broken and one of the two enantiomorphs grows at the expense of the other. This is most clearly seen in the case of a structure with a single heavy atom, where the enantiomorphs are related by a centre of symmetry at the heavy atom location (see figure 3a). As R is increased in magnitude, the ghost peaks in the "wrong" enantiomorph become progressively weaker (figure 3b). This suggests that MEM refinement on non-centric structures, particularly if they have heavy atoms, should be carried out at high values of R , *i.e.* low values of C . The role of the constant C *vis a vis* centric and non-centric structures is discussed in terms of a simple illustrative example in the Appendix.

Although the MEM is a variational method, having its origins in information theory, it turns out surprisingly to have certain features in common with more conventional techniques. If we look at the MEM conditions (10) or equivalently (11), we see that the phases θ_{H_j} and φ_{H_j} are related. Since θ_{H_j} , by (9), is a function of all the φ_{H_k} through $\rho(\mathbf{r})$, we thus see that the phase relations (11) implicitly involve all the unknown phases. In conventional crystallography Sayre (1952) introduced the concept of triplet phase relations, where the phases of three suitably selected reflections are tied together. This

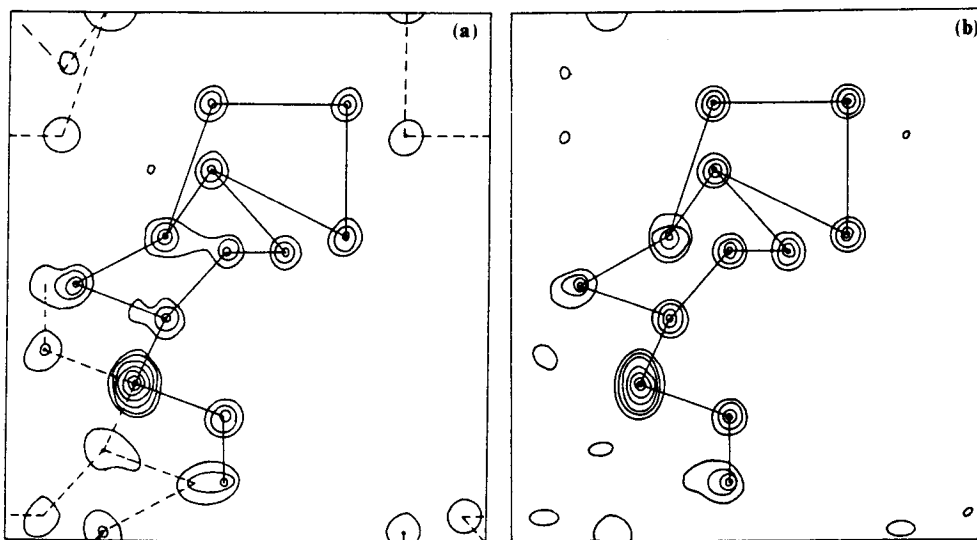


Figure 3. (a) The true structure consists of 12 atoms shown by solid lines, with a heavy atom located at the strongest peak. When the true phases are refined by the MEM using the entropy form S_1 at $R = 9$, the final map has reasonably strong ghost peaks as shown by the dashed lines. These peaks correspond to a spurious enantiomorph molecule sharing the same heavy atom as the true structure. (b) When the refinement is carried out at $R = 100$, the ghost peaks are very much weaker showing that the MEM breaks the symmetry more effectively. On the other hand, if one used $R \ll 9$, one would obtain a completely centrosymmetric map even though the structure amplitudes correspond to a non-centric structure.

has since been extended to quartets, quintets etc., which are believed to be more powerful, though also more difficult to implement (Ladd and Palmer 1980). Since the MEM involves very high order phase relations which are relatively easy to implement through Fourier transforms and the gradient method, there is a case for believing that the method may be an improvement over current direct methods. One can look upon the "non-linearity" parameter R as determining the extent to which different phases are coupled by the MEM condition, since they become decoupled in the $R \rightarrow 1$ limit. Since R can be controlled through the constant C , this may be considered an added advantage.

There has been interest in recent times in a technique called the maximum determinant method (Lajzerowicz and Lajzerowicz 1966; Tsoucaris 1970, 1980) which is based on the Karle-Hauptmann (1950) determinant and involves high order phase relations. We have proved (Narayan and Nityananda 1982) that the maximum determinant method is closely related to maximising the entropy S_1 (Britten and Collins 1982 have also independently reached the same conclusion); this is therefore another contact between conventional crystallography and the MEM.

4. Numerical algorithm

We have carried out a series of calculations on model "structures" in two and three dimensions as well as on a few real structures in three dimensions. We have employed the gradient approach to maximisation of S , using the expression (10) for the components of the gradient. The alternative fixed-point approach which seeks to directly achieve the conditions (11) could be faster when properly implemented. However, fixed point schemes are often notoriously unstable; therefore this approach seemed unsafe to us at this early stage of investigation when speed of convergence is of secondary importance compared to the need to understand the properties of the method. The gradient method on the other hand is very stable and guarantees that S increases with each iteration.

All our studies have been in the lowest symmetry *viz* $p1$ (in 2D) and $P1$ (in 3D). This was dictated by two considerations. Firstly, conventional methods work very well for higher symmetry crystals, particularly when they are centric. It seemed that the MEM could hope to make the greatest contribution in low symmetry non-centric structures. Secondly, the gradient algorithm requires continuous variation of phases whereas centric structures have only two values for each phase *viz* 0 or π .

In our work with the gradient method, we often noticed that a single peak in the map grew at the expense of the others, thus making it virtually impossible to interpret the map. We have been able to suppress this tendency by maximising a modified form of "entropy" *viz*

$$S' = \int \{f[C_1 + \rho(\mathbf{r})] + f[C_2 - \rho(\mathbf{r})]\} dV, \quad (17)$$

where C_1 is a constant ($> -\rho_{\min}$) which is adjusted to give the required value of R (equation (16)) and C_2 is a second constant set to have a value slightly greater than the theoretical height of the heaviest atom (see figure 4). The constant C_2 ensures that any peak which attains the expected height of an atom does not grow any further, thus giving a chance for the other atoms to develop. The constant C_1 is regularly updated at each cycle in order to maintain the value of the nonlinearity parameter R . In the early

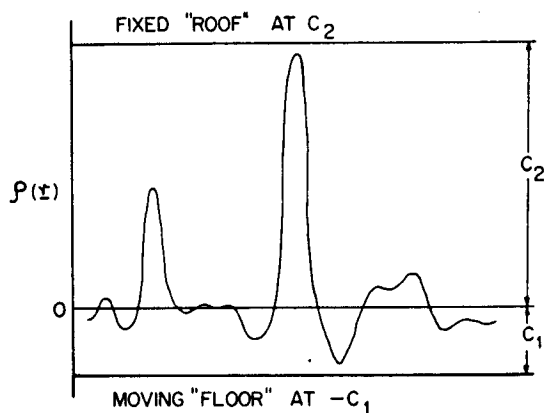


Figure 4. Schematic diagram showing the essentials of our modified entropy (17), which has a variable "floor", adjusted to maintain a preset value of R , and a fixed "roof", set so as to discourage any peak from growing beyond the theoretically expected height of an atom.

stages of *ab initio* structure determination, we usually start with totally random phases for the reflections. The corresponding $\rho(\mathbf{r})$ would have no distinct peaks and would have relatively large negative excursions. At this stage C_1 will have a large value. As the refinement proceeds, peaks will start developing and at the same time the ripple at the bottom would get ironed out, resulting in smaller values for C_1 .

Each cycle of gradient refinement involves the following calculations. Using the current phases $\varphi_{\mathbf{H}}$, we compute $\rho(\mathbf{r})$ by means of a Fourier transform, compute the function $\{f'[C_1 + \rho(\mathbf{r})] - f'[C_2 - \rho(\mathbf{r})]\}$ and calculate $G_{\mathbf{H}}$, its Fourier transform. Equations (10) then give the components of the gradient of S' . Now we shift the $\varphi_{\mathbf{H}}$ by small amounts proportional to the gradient components and compute the new gradient at the shifted phases. We then calculate and apply the shift necessary in order to minimise the magnitude of the gradient at the final point (one could also use the more standard scheme of requiring the final gradient to be orthogonal to the starting one). Thus, one cycle involves four Fourier transforms.

5. Two-dimensional simulations*

We have carried out a number of tests of the MEM on model two-dimensional "structures" with typically 10–15 equal "atoms" in the unit cell. The electron density was sampled on a 32×32 grid and correspondingly the structure factors ranged over reciprocal indices going from -15 to $+16$ (-16 being identified with $+16$ because of the translation symmetry in reciprocal space introduced by the discrete sampling in real space).

Figure 5 shows the result of a typical refinement from partial phase information. The structure factors of the model structure with 15 equal atoms were calculated and their phases were given random errors, with a rectangular distribution, of rms 70° . The

* The results described here were presented at the XIII National Seminar on Crystallography, Nagpur, India, March 1982.

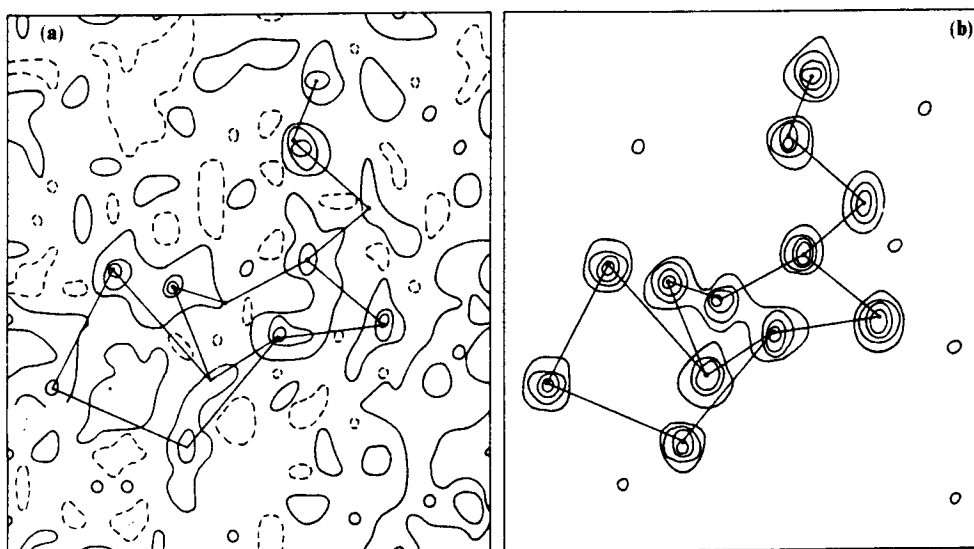


Figure 5. (a) The true structure consists of 12 equal atoms at the vertices joined by the straight lines. The map shown here is obtained by adding phase errors of rms 70° to all the reflections. (b) Map obtained after refining the phases using the MEM.

corresponding electron density map is shown in figure 5a. The structure is still clearly seen but there are many spurious low-level features in the map. Figure 5b shows the remarkable improvement that can be achieved by means of MEM refinement. Here we did 20 cycles of gradient refinement with the entropy S_1 modified as in (17), using $R = 10$. The final rms phase error is only 21.7° . Actually, most of the improvement takes place in the first few cycles. For instance, after 10 cycles, the rms is as low as 30.8° . From a number of similar studies we find that the structure refines well from initial phase errors of up to $70\text{--}75^\circ$. There are many situations, particularly in macromolecular crystallography, where one can obtain limited phase information using techniques such as isomorphous replacement and anomalous dispersion. There appears to be good hope that the MEM can greatly help in phase refinement in such cases. Collins (1982) has published some results on phase refinement on real protein data using the MEM. While there is some improvement, his results are not as impressive as those in figure 5. This could be because of experimental errors in the structure amplitudes or, more probably, because the atoms in his structure are not well resolved (the MEM, being peak-seeking, is likely to work best with high resolution data).

The more difficult, and therefore more interesting problem is to solve a crystal structure *ab initio*, starting from pure amplitudes and no phase information. We have tried this in a number of two-dimensional simulations and figures 6–8 show some typical results. In each case we started the refinement with totally random phases and refined for 20 cycles with entropy S_1 at $R = 10$. Although the final map usually had a few spurious peaks, invariably the structure could be identified, with at most one or two atoms not seen. As a more objective test of the power of the method, we carried out a few “blind tests” of the following kind. One of us (Narayan) generated a model and produced a list of structure amplitudes along with stereochemical information on the

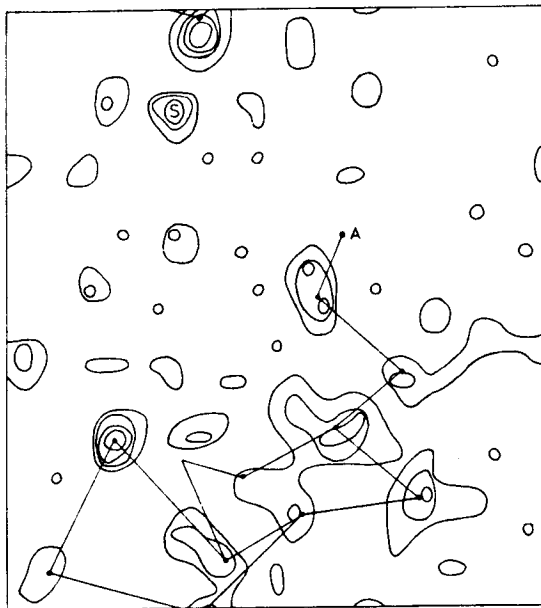


Figure 6. The true structure is the same as in figure 5. Random phases were given to the true structure amplitudes and refined using the MEM. Eleven out of 12 atoms can be fitted on peaks in the refined map. The atom marked *A* is absent with no nearby peak, while the peak *S* is a strong spurious feature. This is a typical result obtained with the MEM in two-dimensional *ab initio* structure analysis. Note the origin shift.



Figure 7. The true structure consists of 2 molecules, of 8 equal atoms each, with the orientations and relative spacings shown. The MEM map obtained by refining random starting phases reveals most of the structure except the absent atom *A*. There is one strong spurious peak *S*.

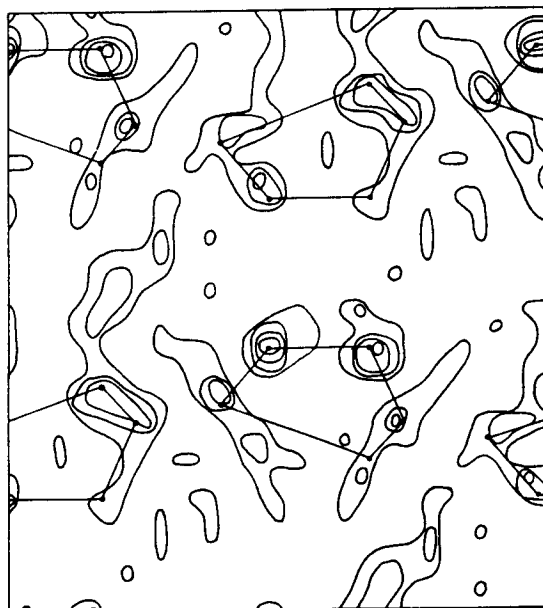


Figure 8. The true structure consists of 4 molecules, of 5 equal atoms each, in cm symmetry. The MEM map, refined from random starting phases, without including the information on the mirror, shows all the 10 independent atoms and has no strong spurious peaks.

“molecule”. Another (Vani), given a map with random phases, would refine them by means of the MEM and try to interpret the resulting map. In all the three cases where this was tried the structure was solved correctly—usually, one or two atoms did not appear in the MEM map, but these could be easily located by standard Fourier techniques (Ramachandran and Srinivasan 1970). Because we start from random phases in our calculations, there is invariably an origin shift in the solution and in 50% of the cases there would also be an enantiomorph reversal.

A point to be emphasized is that all our calculations were carried out in $p1$ symmetry which is considered the most difficult. Figure 8 shows an interesting case where the true structure has cm symmetry but we did not use the information on the mirror during the refinement. Thus, although this structure has only five independent atoms, we solved it as if it was a 10-atom structure in $c1$ symmetry. It is interesting that the mirror symmetry comes to light during the structure solution, though not perfectly.

Another interesting feature is that in all our calculations we directly used the structure amplitudes $|F_H|$ without any modification. In the direct methods a crucial prerequisite is “sharpening” of the data to obtain so-called E 's, corresponding to a point atom structure. In our experience, the MEM works very well with F 's and, in fact, does not refine so well when E 's are used.

6. Three-dimensional refinement—real structures

Encouraged by the results in two dimensions we have very recently embarked on three-dimensional structure determination using the MEM. Due to memory limitations in the

mini-computer (PDP 11/34) with which we worked, we used a $16 \times 16 \times 16$ unit cell for the electron density, corresponding to (h, k, l) ranging from -7 to $+8$. Once again all refinement was carried out in $P1$ symmetry using the gradient method. Whereas in two dimensions the entropy S_1 appears to be marginally better than S_2 , we find the opposite to be true in three dimensions. Hence all the results described here correspond to the entropy

$$S'_2 = - \int \{ [C_1 + \rho(\mathbf{r})] \ln [C_1 + \rho(\mathbf{r})] + [C_2 - \rho(\mathbf{r})] \ln [C_2 - \rho(\mathbf{r})] \} dV_r. \quad (18)$$

The constant C_1 was adjusted to maintain $R = 10$ while C_2 was taken to be slightly greater than the theoretical height of the strongest atom. The refinement is typically slower than in two dimensions, needing of the order of 50–100 cycles, though once again most of the gain is achieved early on.

To begin with we generated a model structure with 15 equal atoms in $P1$ symmetry and refined the phases using the MEM, starting from random values. The electron density after 60 cycles of refinement was scanned for peaks and these were listed in decreasing order of magnitude. The structure was very clearly identified from the peaks. Among the first 15 peaks, 10 belonged to atoms in the structure and 5 were spurious. Among the 10 correct peaks, 3 (*viz* peaks 2, 3, 4 in the order of ranking) were elongated and had a second atom in the tail region. This is a problem we faced in all our three-dimensional calculations—because of the lack of resolution (since we have only 16 grid points along each axis), peaks sometimes merged and two atoms would be represented by a single elongated peak. In such cases, usually one atom is close to the strongest region of the peak and the other weakly present in the tail. Thus, in our simulated 15-atom structure, the first 15 peaks accounted for 13 atoms and in fact the first 8 peaks were all correct and contained 11 atoms. The final 2 atoms in the structure were at peaks 23 and 26. We feel that the performance of the MEM is very encouraging.

Emboldened by this we have tried to “solve” a few real structures using real experimental data. Unfortunately, we could not find in the literature any suitable equal-atom $P1$ structure whose data would fit within our 16^3 grid. As the next best thing, we selected a few structures in $P\bar{1}$ symmetry and refined them in $P1$ symmetry, ignoring the centre of symmetry altogether.

Martuscelli (1969) has reported the crystal structure of octa-trans-2, trans-6-diene-1, 8-dioic acid which crystallizes in $P\bar{1}$ symmetry with unit cell parameters $a = 12.00 \text{ \AA}$, $b = 3.94 \text{ \AA}$, $c = 7.72 \text{ \AA}$, $\alpha = 121.35^\circ$, $\beta = 139.4^\circ$, $\gamma = 70.05^\circ$. He solved the structure from about 500 observed reflections using model-fitting procedures based on the Patterson maps. Because the cell parameter a is rather large, our calculations on this structure were hampered by the extremely poor resolution (0.75 \AA) in this direction. Also, we could include only reflections with h in the range -7 to $+8$ which reduced the number of structure amplitudes we could use to 326. In spite of these limitations, the structure refined quite easily from random phases. After 100 cycles of refinement, the first 12 peaks in the map included all the 12 atoms of the structure. Peaks 7 and 9 were spurious while peaks 1 and 2 included 2 atoms each. This is very gratifying since we had very limited data and severe resolution problems.

Our next attempt on a much larger equal atom structure was not so spectacular though still rather encouraging. Karle (1969) has reported the structure of a thymine-thymine adduct (obtained from irradiated thymine). This is again in $P\bar{1}$ symmetry with 38 atoms in the unit cell (2×19 independent atoms); $a = 9.44 \text{ \AA}$, $b = 8.29 \text{ \AA}$, c

= 7.57 Å, $\alpha = 99.0^\circ$, $\beta = 91.5^\circ$, $\gamma = 89.8^\circ$. The data consisted of 1844 independent reflections and the centrosymmetric structure was originally solved by obtaining the structure factor signs using the symbolic addition procedure (Karle and Karle 1966). We attempted to solve the 38 atoms of the structure in $P1$ symmetry with the MEM, using 1370 reflections which fitted within our grid. After 100 cycles of refinement starting from random phases, we found that, of the first 38 peaks, 18 were correct, including 22 atoms (peaks 1, 6, 11 and 14 had two atoms each), while the first 80 peaks had 25 correct peaks with 31 atoms; 7 atoms had no peaks at all in the map. Of the two molecules, one was seen quite clearly while the other was rather weak. The refinement has clearly not been too successful; on the other hand, we believe we could, if necessary, have solved the structure from our map with a little patience if the structure were not already known. Karle (1969) used 6 symbols for which a unique assignment of signs was indicated by the symbolic addition method and the first 19 peaks of the resulting E -map showed the 19 independent atoms in the structure—a truly remarkable performance. However, considering that (i) we had to use less data, (ii) we had poor resolution in the electron density map, (iii) we refined 38 independent peaks, (iv) we worked in a non-centric framework and (v) we have not really optimised our refinement procedure, we feel our results using the MEM are quite good. We may point out that a structure with 38 atoms in $P1$ is not trivial even today and would probably have been considered quite difficult a decade back.

As a final example we attempted a simple heavy atom structure. Dutta and Woolfson (1969) solved the structure of tetraethyl diphosphine disulphide, which crystallizes in $P\bar{1}$ symmetry, by direct methods from two projections using two symbols in each. The unit cell has 12 atoms (6 independent atoms if one uses the $P\bar{1}$ symmetry), of which 2 sulphur and 2 phosphorus atoms are heavy, while the other 8 are (light) carbon atoms. The unit cell parameters are $a = 8.98$ Å, $b = 6.45$ Å, $c = 6.15$ Å, $\alpha = 113.0^\circ$, $\beta = 85.2^\circ$, $\gamma = 102.5^\circ$. Of the 864 reflections observed, we could accommodate only 760 within our grid and the phases of these were refined as usual with the MEM. After 70 cycles of refinement the first 12 peaks contained 10 atoms with 2 spurious peaks (8 and 10 in ranking). The first 4 peaks were the 4 heavy atoms and the two atoms that were missing were the weakest in the final solution, having heights only $\sim 18\%$ of the sulphur atoms. The success of the MEM with this structure is interesting as it shows that the method is equally applicable to heavy atom structures (as indeed the direct methods also are, although they were originally developed for equal or near-equal atom structures).

In the various structure determinations described above, there was nothing special about the starting random phases which could have helped in the refinement. We have tried different sets of random phases and usually obtained comparable results. Also, although we refined for ~ 100 cycles in all the cases, most of the refinement is over within the first 20 cycles and the later calculations only help to “sharpen” the peaks and define them better. Finally, we should mention that, as in two dimensions, here again we obtained much better results using the F 's rather than E 's.

7. Discussion and conclusions

The phase problem of crystallography is probably the most difficult and challenging among the variety of Fourier inversion problems known in several fields. The success of the MEM in other fields in different contexts is therefore no guarantee of success when

applied to crystallography. Although we and a few others have been excited about the prospects of the MEM in crystallography, there was until recently only limited reason for optimism. Firstly, we (as well as Britten and Collins 1982) had shown that the MEM is related to the maximum determinant method of crystallography; we had further argued that the MEM phase relations (equations (10)) are higher order phase relations which are likely to be more powerful than the lower order relations (triplets, quartets, etc.) conventionally used and that equal atom structures satisfy the MEM relations. Secondly, Collins (1982) had shown that phase refinement is feasible in macromolecular crystallography, starting from approximately known phases.

We believe that the results presented in this paper greatly improve the case for trying the MEM in crystal structure analysis. We have shown that the method works well in small- and medium-sized *real* crystals and has the ability to locate a high percentage of the atoms starting from random phases for the structure factors. It is worth stressing once again that all our work has been in *P1* symmetry, which is usually considered the most difficult.

A comparison of MEM refinement (as we implement it) with conventional direct methods shows two important differences.

(a) In the direct methods one works with *E*'s and solves for point atoms. On the other hand, we find that the MEM works best with *F*'s. It is possible that this difference arises because the MEM, apart from looking for peaks as do the direct methods, is also concerned about the flatness of the baseline. Since an *E*-map is likely to have a lot of ripple even with correct phases (due to data truncation and experimental errors), this might be a possible explanation.

(b) In the direct methods, the multisolution approach is the rule where a small set of primary reflections are systematically given various phases. One set then produces the structure while the others are usually uninterpretable. In our tests so far we have found that, regardless of the starting set of random phases, the structure usually comes up to about the same level of interpretability. It is possible that we are still working with structures that are too small and that the MEM also may need a multisolution approach with larger structures. On the other hand, there is a distinct possibility that, because of its higher order phase relations, the method is able to attract the solution towards the correct structure from a wider range of starting phases.

Because of the above reasons we feel that the MEM is not just the direct methods in a new disguise but is distinctly different. As such it merits further study in more complicated situations. The ability to work with *F*'s rather than *E*'s might turn out to be a big advantage since the method will be less susceptible to experimental errors. In our view the following further questions now need urgent attention.

(a) Although the gradient method we have employed is useful for the sort of preliminary studies we have made, the future probably belongs to fixed point schemes (*e.g.* Sayre 1980) which directly seek to satisfy the MEM conditions (10). Not only would the method then easily work for centrosymmetric structures, there is also the real advantage that one can use one's *a priori* information on the shapes of atoms through the S_H in (14) to determine the n_H in (11). This information cannot be introduced into the gradient algorithm.

(b) One needs sensitive figures of merit to measure the goodness of the refinement, particularly if the MEM is to be combined with the multisolution approach.

(c) Techniques such as isomorphous replacement and anomalous dispersion are employed to give added information in macromolecular phase refinement and these

could be put in variationally into the MEM. For instance, in isomorphous replacement, one knows that the electron density distribution in a protein and its derivatives are identical except at the locations of the replaced atoms where the differences are known. So too, in the anomalous dispersion case, where the electron density becomes complex, one has the powerful *a priori* information that the peaks in the imaginary part occur only at the centres of the peaks in the real part. One technique to introduce these extra constraints is to maximise the integral of $f(\rho_{\text{derivative}} - \rho_{\text{protein}})$ and $f(\rho_{\text{real}} - \rho_{\text{imaginary}})$, which is equivalent to introducing the *a priori* input that $(\rho_{\text{derivative}} - \rho_{\text{protein}})$ has a flat baseline with sharp well-resolved peaks and that $(\rho_{\text{real}} - \rho_{\text{imaginary}})$ has no negative peaks. These and other similar ideas deserve to be pursued, particularly since Collins (1982) has obtained interesting results even with just the bare MEM.

Acknowledgements

It is both a privilege and a pleasure to contribute this article to a volume commemorating Prof. S Ramaseshan's sixtieth birthday. A whole band of friends, colleagues, students, grandstudents, etc have benefited from his initiative, insight and encouragement. We join with them in wishing him many active and fruitful years of scientific work.

Appendix

We present here an illustrative example in one dimension bringing out the crucial role played by the zero Fourier coefficient F_0 . We consider a periodic real function in one dimension which is sampled at 4 points in the unit cell so that the function is completely described by its values $\rho_{-1}, \rho_0, \rho_1, \rho_2$ (the origin has been taken at the second point). The Fourier coefficients are also periodic. As the ρ 's are real, F_0 and F_2 are real and $F_{-1}^* = F_1 = |F_1| \exp(i\varphi_1)$ are the only structure factors with a continuously variable phase; even here, if we eliminate physically identical solutions (related by an origin shift or enantiomorph reversal), the range of φ_1 from 0 to $\pi/2$ completely describes all distinct solutions compatible with a given set of amplitudes $F_0, |F_1|, F_2$.

Consider the centric "structure" given by

$$\rho_0 = 2.0, \rho_{-1} = \rho_1 = 0.5, \rho_2 = 0, \quad (\text{A1})$$

whose Fourier transform gives

$$F_0 = 1.0, |F_1| = 0.75, \varphi_1 = 0^\circ, F_2 = 0.5. \quad (\text{A2})$$

We investigate the variation of the entropy

$$S_1(C, \varphi_1) = \int \ln [C + \rho(\varphi_1)] dx = \sum_{j=-1}^2 \ln [C + \rho_j(\varphi_1)], \quad (\text{A3})$$

where $\rho_j(\varphi_1)$ are calculated with the true amplitudes and the phase φ_1 . Adding C to all the ρ_j is equivalent to replacing F_0 by $F_0 + C$. Figure 9 shows the variation of S_1 with φ_1 for different values of C . At high values of C , the maximum value of S_1 is seen to occur at the correct phase, $\varphi_1 = 0$. However, at low values of C the maximum shifts away from 0 and at $C = 0.02$, for instance, occurs at $\varphi_1 \sim 10^\circ$. Thus, at low values of C , the MEM solution is non-centric even though the data correspond to a centric structure.

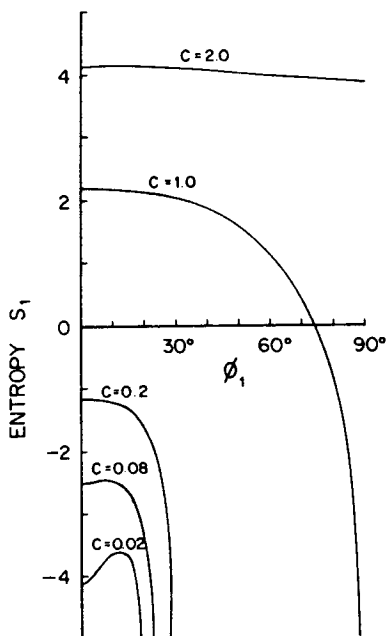


Figure 9. Variation of entropy S_1 as a function of phase ϕ_1 for the centrosymmetric structure (A1), (A2) for various "DC offsets" C . Note that the maximum entropy solution at small C is non-centric ($\phi_1 \neq 0$).

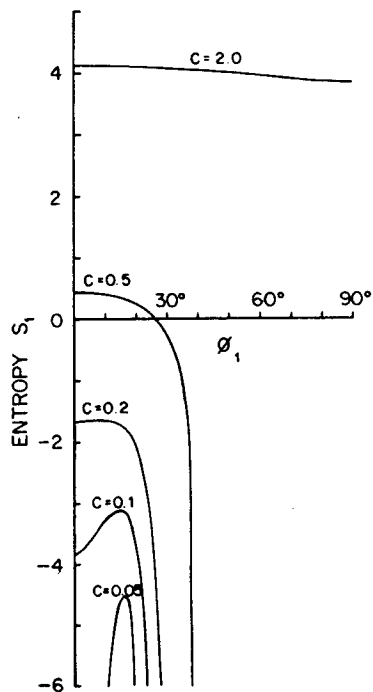


Figure 10. Variation of entropy S_1 as a function of phase ϕ_1 for the non-centrosymmetric structure (A4), (A5) for various "DC offsets" C . Note that the maximum entropy solution at large C is centric ($\phi_1 = 0$).

Consider next the non-centric structure

$$\rho_{-1} = 0, \rho_0 = 3.0, \rho_1 = 1.0, \rho_2 = 0, \quad (\text{A4})$$

$$F_0 = 1.0, |F_1| = 0.7906, \phi_1 = 18.43^\circ, F_2 = 0.5. \quad (\text{A5})$$

The variation of the entropy S_1 for this case is shown in figure 10. At low values of C , the maximum value of S_1 occurs close to the true phase *e.g.* for $C = 0.05$, we obtain $\phi_1 \sim 18^\circ$. However, as C increases, the maximum shifts to lower values of ϕ_1 until above a critical value of C the maximum is at $\phi_1 = 0$, *i.e.* a centric structure. This tendency of a non-centric structure to move towards a centric solution at high C (or low R) is also discussed in §3 with a two-dimensional example.

The above results suggest that the value of C should be set by the user depending on prior knowledge of the presence or absence of a centre of symmetry. We have found by experience that very low values of C lead to slow convergence while very high values of C can cause round-off errors in the calculations. In our work, with these two auxiliary constraints in mind, we constantly update C at each cycle so as to maintain the non-linearity parameter R at a value ~ 10 . This has given good results with both centric and non-centric structures.

References

- Ables J G 1972 in *Symposium on the collection and analysis of astrophysical data*, *Astron. Astrophys. Suppl.* **15** 383 (1974)
- Britten P L and Collins D M 1982 *Acta Crystallogr.* **A38** 129
- Burg J P 1967 Annual Meeting of the International Society of Exploration Geophysicists, Oklahoma City
Reprinted: 1978 *Modern spectral analysis* (ed.) D G Childers (New York: IEEE Press) p. 34
- Collins D M 1982 *Nature (London)* **298** 49
- Dutta S N and Woolfson M M 1969 *Acta Crystallogr.* **14** 178
- Frieden B R 1972 *J. Opt. Soc. Am.* **62** 511
- Germain G, Main P and Woolfson M M 1971 *Acta Crystallogr.* **A27** 368
- Gull S F and Daniell G J 1978 *Nature (London)* **272** 686
- Högbom J A 1978 in *Image formation from coherence functions in astronomy*, (ed.) C van Schooneveld (D. Reidel Publishing Company) p. 237
- Karle I L 1969 *Acta Crystallogr.* **B25** 2119
- Karle I L and Karle J 1963 *Acta Crystallogr.* **16** 969
- Karle J and Karle I L 1966 *Acta Crystallogr.* **21** 849
- Karle J and Hauptman H 1950 *Acta Crystallogr.* **3** 181
- Komesaroff M M and Lerche I 1978 in *Image formation from coherence functions in astronomy* (ed.) C van Schooneveld (D. Reidel Publishing Company) p. 241
- Komesaroff M M, Narayan R and Nityananda R 1981 *Astron. Astrophys.* **93** 269
- Ladd M F C and Palmer R A 1980 *Theory and practice of direct methods in crystallography*. (New York and London: Plenum Press)
- Lajzerowicz J and Lajzerowicz J 1966 *Acta Crystallogr.* **21** 8
- Martuscelli E 1969 *Acta Crystallogr.* **B25** 2540
- Narayan R and Nityananda R 1981 *Curr. Sci* **50** 168
- Narayan R and Nityananda R 1982 *Acta Crystallogr.* **A38** 122
- Narayan R and Nityananda R 1983 to appear in *Proceedings of the symposium on measurement and processing for indirect imaging*, Sydney (Cambridge: University Press)
- Nityananda R and Narayan R 1982 *J. Astrophys. Astron.* **3** 419
- Ponsonby J E B 1973 *Mon. Not. R. Astron. Soc.* **163** 369
- Ramachandran G N and Srinivasan R 1970 *Fourier methods in crystallography* (Wiley-Interscience)
- Sayre D 1952 *Acta Crystallogr.* **5** 60
- Sayre D 1980 in *Theory and practice of direct methods in crystallography*, (eds) M F C Ladd and R A Palmer, (New York and London: Plenum Press) p. 271
- Subrahmanya C R 1978 in *Image reconstruction from coherence functions in astronomy*, (ed.) C van Schooneveld (D. Reidel Publishing Company) p. 287
- Subrahmanya C R 1980 *Bull. Astron. Soc. India* **8** 5
- Tsoucaris G 1970 *Acta Crystallogr.* **26** 492
- Tsoucaris G 1980 in *Theory and practice of direct methods in crystallography*, (eds) M F C Ladd and R A Palmer, (New York and London: Plenum Press) p. 287
- Ziman J M 1972 *Principles of the theory of solids*, 2nd edition (Cambridge: University Press) p. 50