Acta Cryst. (1989). A**45**, 200–203

The Maximum Entropy Distribution Consistent with Observed Structure Amplitudes

BY E. PRINCE

Institute for Materials Science and Engineering, National Bureau of Standards, Gaithersburg, Maryland 20899, *USA*

(*Received* 23 *March* 1988; *accepted* 27 *September* 1988)

Abstract

It is shown that an electron density distribution of the form $\rho_k = \exp[\sum f_j(r_k)x_j]$ has maximum entropy under the constraint that the expected values of a set of functions, $f_j(\mathbf{r})$, are constant. For a Fourier map the functions $f_j(\mathbf{r})$ are the magnitudes of the structure factors for a set of reflections $\mathbf{h}_j$ including $F(000)$. The values of the parameters x_j for which $|\langle\exp(2\pi i\mathbf{h}_j.\mathbf{r})\rangle| = |F_{obs}(\mathbf{h}_j)|$ for an arbitrarily large set of reflections may be found by an iterative algorithm in which $\mathbf{x}_{i+1} = \mathbf{x}_i + \mathbf{H}_i^{-1}\boldsymbol{\Delta}_i$, where the matrix $\mathbf{H}$ is positive definite. Because the distribution $\rho(\mathbf{r})$ is everywhere positive, if non-negativity of electron density is sufficient information to determine a unique structure by direct methods, it follows that the maximum entropy procedure must lead to the same unique structure. Maximum entropy is thus an efficient way of expressing the phase implications of a large set of structure amplitudes.

Introduction

Various workers (Collins, 1982; Wilkins, Varghese & Lehmann, 1983; Bricogne, 1984) have shown that the principle of maximum entropy (Shannon, 1948; Jaynes, 1979) can provide a powerful formalism for the optimum use of the structure information contained in the measured values of X-ray diffraction amplitudes. These workers have proposed procedures for finding maximum entropy distributions that are consistent with the observed data, procedures that generally involve a many-parameter optimization that incorporates the data by use of Lagrangian multipliers. This technique becomes cumbersome when applied to large-molecule structures, leading Woolfson (1987) to observe that 'entropy maximization is adding nothing new to the crystallographic scene, and, since it involves a great deal of effort, perhaps nothing useful'. Prince, Sjölin & Alenljung (1988; hereinafter referred to as PSA), however, have shown that a maximum entropy distribution can be found by expressing the maximum entropy condition in terms of a small number of parameters and varying those parameters to find a distribution that satisfies the constraints. They applied this method to a model in which the only constraints were on the total number of electrons in the unit cell and the expected value of a truncated Fourier series using a set of phases previously determined by other means. This paper extends that method to determine a distribution that has maximum entropy and simultaneously satisfies the amplitudes of an arbitrarily large set of reflections.

Mathematical analysis

In the entropy maximization procedure we divide a unit cell into a large number of subunits, commonly called 'pixels', and seek to maximize the quantity

$$S = -\sum_{k=1}^{N} \rho_k \ln(\rho_k/\rho'_k), \qquad (1)$$

subject to certain constraints that require that the Fourier transform of the density distribution be in agreement with the observed diffraction data. ρ_k represents the mean density in the kth pixel, and ρ'_k is the density in that pixel in a distribution inferred from prior information. We shall discuss below the prior distribution in terms of a sequential approach to a complete model, but we shall assume initially that it is uniform, so that the terms in (1) reduce to $\rho_k \ln \rho_k$. The constraints are that the sum of $(V/N)\rho_k$ (where V is the volume of the unit cell) over the unit cell be equal to $F(000)$, and that the expected values of one or more other Fourier coefficients, or sums of terms in a truncated series, be constant. The case considered in PSA is one in which there is a single constraint on the expected value of a truncated series.

In order to maximize the entropy expression in (1) subject to n constraints, it is necessary to partition the N-dimensional parameter space into an n-dimensional subspace reachable by a linear combination of the constraint functions and an $(N-n)$-dimensional 'null' space orthogonal to it. This partition may be made efficiently by use of the variable reduction method (Gill, Murray & Wright, 1981) as follows. Define a constraint matrix $\mathbf{C}$ by $C_{jk} = \partial f_j(\boldsymbol{\rho})/\partial\rho_k$, where $f_j(\boldsymbol{\rho})$ is the jth constraint function. If no two constraint functions are redundant, $\mathbf{C}$ has full-row rank (Stewart, 1973) and can be partitioned into a nonsingular $(n\times n)$ matrix $\mathbf{V}$ and an $n\times(N-n)$ matrix $\mathbf{U}$. The rows of the matrix $[-(\mathbf{V}^{-1}\mathbf{U})^T, \mathbf{I}]$, where $\mathbf{I}$ is the identity matrix of order $(N-n)$, are then

0108-7673/89/020200-04$03.00

orthogonal to the rows of **C** and form a sufficient basis set for the null space. The condition for maximum entropy with constant $\langle f_j(\mathbf{\rho})\rangle$ for all j is that the partial derivatives of S with respect to scalar multiples of each of the null-space basis vectors vanish individually. This matrix has a number of useful properties. (A simple example applying this procedure is worked out in PSA.) Each row has no more than $n+1$ nonzero elements, all except one of which fall in the same set of n columns, while that one always has the value +1. Furthermore, one row of **C** serves to keep the sum of the numbers of electrons in all pixels equal to $F(000)$, which means that all of its elements also have the value +1. Because each row of $[-(\mathbf{V}^{-1}\mathbf{U})^T, \mathbf{I}]$ must be orthogonal to this, the sum of all elements in a row must be zero, and the sum of n elements must therefore be −1. Given that $\partial S/\partial \rho_k = -(\ln \rho_k + 1)$, it follows that the condition for maximum entropy is satisfied when

$$\ln \rho_k = \sum_{l=1}^{n} Z_{kl} \ln \rho_l, \tag{2}$$

where $Z = (\mathbf{V}^{-1}\mathbf{C})^T$, and ρ_l are the densities in the n pixels corresponding to the columns of **V**. Thus the maximum entropy condition provides a unique specification of the densities in all N pixels in terms of the densities in only n of them. Note also that (2) is unaffected by the addition of a constant to each side, so that the maximum entropy condition is independent of any linear rescaling of the entire map.

Most procedures previously proposed for the determination of a maximum entropy distribution have involved a search within the null space starting at a point where the constraints are satisfied, a *feasible point*, for a point at which the gradient of the entropy function vanishes. This requires the use of a fitting algorithm in a space that may be of extremely high dimension, with the additional problem that safeguards must be used to prevent the density in any pixel from being negative at any point in the search. PSA showed that it is much more efficient computationally to reverse the process and to search among points satisfying (2) for a feasible point. They showed that the function

$$E(\mathbf{\rho}) = \left\langle \sum_{j=1}^{m} |F(\mathbf{h}_j)| \cos(2\pi \mathbf{h}_j.\mathbf{r} - \varphi_j) \right\rangle \tag{3}$$

is a well behaved function of the single variable $x = \ln(\rho_{\max}/\rho_{\min})$, where $\rho_{\max}$ and $\rho_{\min}$ are the maximum and minimum densities in the map, respectively, and that the equation

$$E(\mathbf{\rho}) = \sum_{j=1}^{m} |F(\mathbf{h}_j)|^2 \tag{4}$$

can be easily solved using standard numerical methods. This procedure is known in the optimization literature as the dual method (Gill, Murray & Wright, 1981; Luenberger, 1984). It is particularly useful when the conditions for a stationary point in the null space can be written in closed form, as in (2).

A maximum entropy density distribution with a constrained expected value of a truncated Fourier series like that in (3) reproduces many features of the truncated series. For example, it has maxima, minima and saddle points in the same places as the truncated series, but the peaks are sharper and the valleys are flatter. Its Fourier transform, however, has different values for the amplitudes of the individual terms in the series and nonzero values for the amplitudes and values for the phases of the terms that were not included in the series. The latter feature may be used to extend phases to higher resolution, the calculated phases being used with the observed amplitudes to compute a higher-resolution map. It is of interest, however, to find a density map that is in *exact* agreement with the amplitudes of all observed reflections, or at least a large set of the strongest ones. To accomplish this, we define an additional set of constraint functions of the form $f_j(\mathbf{\rho}) = \langle \cos(2\pi \mathbf{h}_j.\mathbf{r} - \varphi_j)\rangle$, and solve the system of nonlinear equations $f_j(\mathbf{\rho}) = |F_{\text{obs}}(\mathbf{h}_j)|$. If one sets the scale to give the correct value of $F(000)$, (2) is equivalent to

$$\rho_k = [NF(000)/V] \frac{\exp\left[\sum_{l=1}^{n} Z_{kl} \ln \rho_l\right]}{\sum_{m=1}^{N} \exp\left[\sum_{l=1}^{n} Z_{ml} \ln \rho_l\right]}, \tag{5}$$

and $f_j(\mathbf{\rho})$ is then

$$f_j(\mathbf{\rho}) = \sum_{k=1}^{N} \rho_k \cos(2\pi \mathbf{h}_j.\mathbf{r}_k - \varphi_j). \tag{6}$$

To solve, by Newton's method (Prince, 1982), the system of equations $f_j(\mathbf{\rho}) = |F_{\text{obs}}(\mathbf{h}_j)|$ we require the matrix **M** whose elements are $M_{jl} = \partial f_j(\mathbf{\rho})/\partial \lambda_l$, where $\lambda_l = \ln \rho_l$. These are given by

$$\partial f_j(\mathbf{\rho})/\partial \lambda_l = \sum_{k=1}^{N} [\partial f_j(\mathbf{\rho})/\partial \rho_k](\partial \rho_k/\partial \lambda_l), \tag{7}$$

but $\partial f_j(\mathbf{\rho})/\partial \rho_k = C_{jk}$, and $\partial \rho_k/\partial \lambda_l = Z_{kl}\rho_k$, so that **M** is given by

$$\mathbf{M} = \mathbf{CPZ}, \tag{8}$$

where **P** is a diagonal matrix such that $P_{kk} = \rho_k$.

M is a general matrix with no simplifying symmetry properties, and it requires selection of a set of columns of **C** to form a nonsingular matrix **V**. Note, however, that $\mathbf{Z} = \mathbf{C}^T(\mathbf{V}^{-1})^T$, and, without loss of generality, a feasible point satisfying the maximum entropy condition can be expressed, rather than in terms of a subset of $\ln \rho_k$, in terms of the linearly transformed variable $\mathbf{x} = (\mathbf{V}^{-1})^T \boldsymbol{\lambda}_{\mathbf{V}}$. The Newtonian solution of the system of equations then becomes

$$\mathbf{x}_{i+1} - \mathbf{x}_i = \mathbf{H}_i^{-1} \boldsymbol{\Delta}_i, \tag{9}$$

where $\mathbf{H}=\mathbf{CPC}^T$, and $(\Delta_i)_j=|F_{obs}(\mathbf{h}_j)|-f_j[\boldsymbol{\rho}(\mathbf{x}_i)]$. $\mathbf{H}$ is symmetric, and, because $\mathbf{C}$ has full-row rank and $P_{kk}>0$ for all k, is in fact positive definite. Thus the solution of the maximum entropy problem closely resembles the least-squares problem, except that it is $\mathbf{P}$, which plays the role of a weight matrix, that is refined.

An application example

To illustrate how this method may be used in crystallography, consider a simple one-dimensional example based on the projection of the diamond structure on a [111] axis. The dashed curve in Fig. 1 is a plot of the truncated series

$$\rho(y)=F(0)+2F(1)\cos 2\pi y+2F(3)\cos 6\pi y+2F(4)\cos 8\pi y,$$

where $F(0)=1$, $F(1)=0{\cdot}68$, $F(3)=-0{\cdot}65$, and $F(4)=-0{\cdot}90$. It has several regions of negative density, as well as some extra bumps in addition to the principal ones at $y=0{\cdot}125$ and $y=0{\cdot}875$. The map was divided into 200 pixels. The constraint matrix has four rows: the first row is all ones to maintain constant normalization; the second row is an image of $\rho(y)$ with the minimum density set to zero and scaled so that the maximum density is one; the third and fourth rows are the values of $\cos 2\pi y$ and $\cos 8\pi y$ in the middle of each pixel. (Note that the three cosine terms cannot all be included individually, or the second row would be a linear combination of the others, leading to rank deficiency.) Because the first row of $\mathbf{C}$ is all ones, $\exp(x_1)$ is a linear scale factor that can be adjusted to the proper normalization without affecting the shape of the curve. The procedure of (9) was applied, starting with

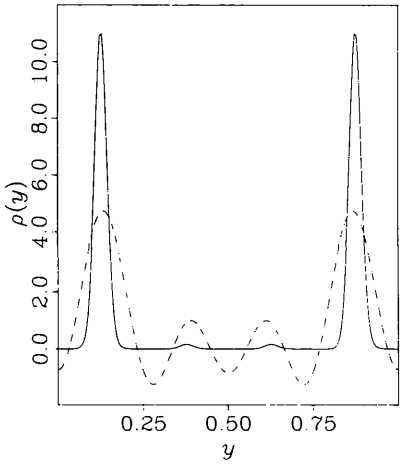

Fig. 1. Plot (dashed curve) of the Fourier series $\rho(y)=1+1{\cdot}36\cos 2\pi y-1{\cdot}30\cos 6\pi y-1{\cdot}80\cos 8\pi y$ and (solid curve) an entropy-maximized distribution having the same coefficients for corresponding terms in its Fourier expansion.

Table 1. *Coefficients of the first* 13 *terms in the Fourier expansion of the entropy-maximized curve in Fig.* 1

h	F(h)	h	F(h)
0	1·000	7	0·504
1	0·680	8	0·657
2	−0·003	9	0·396
3	−0·650	10	−0·012
4	−0·900	11	−0·323
5	−0·576	12	−0·395
6	0·009		

$x_1=\ln(0{\cdot}1\rho_{max})$, where ρ_{max} is the maximum density in the initial map, $x_2=\ln(10)$, $x_3=x_4=0$. It converged in four iterations to $x_1=-6{\cdot}024$, $x_2=3{\cdot}821$, $x_3=1{\cdot}295$, $x_4=3{\cdot}719$, giving the solid curve in Fig. 1. The coefficients of the first few terms in the Fourier expansion of this curve are given in Table 1. The coefficients corresponding to the terms of the initial truncated series match exactly.

Discussion

In this example when the fitting procedure was divided into two stages, with a preliminary step using the soft constraint of PSA followed by adjustment of all parameters, convergence was no more rapid than it was when adjustment of all parameters started with the crude approximation. Because it would be expected that more information included in determining the starting point for the iterative solution would lead to a closer approximation, this may be a feature of this simple model that does not carry over to more complex problems. Note that the sequential process corresponds to

$$\rho_k=\rho'_k\exp\left[\sum_{j=1}^{n}f_j(r_k)\Delta x_j\right], \qquad (10a)$$

where

$$\rho'_k=\left\{[NF(000)/V]\Big/\sum_{l=1}^{N}\exp\left[\sum_{m=1}^{n}f_m(r_l)x'_m\right]\right\}\times\exp\left[\sum_{m=1}^{n}f_m(r_k)x'_m\right], \qquad (10b)$$

and $\mathbf{x}'$ is the set of parameters found in the previous step. ρ' thus plays the role of the 'prior' distribution that appears in (1). It makes no difference, mathematically at least, whether the fitting is performed in one step starting with a uniform prior distribution or is performed in several steps with increasingly tight constraints. This is in accord with the consistency conditions that maximum entropy has been shown to satisfy (Shore & Johnson, 1980; Livesey & Skilling, 1985). It seems likely, however, that the sequential procedure would be better behaved in practice, particularly when phase extension must be included.

By using this procedure it is possible to find a vector of parameters, $\mathbf{x}$, that will make $|F_{obs}(\mathbf{h}_j)|=|f_j(\rho)|$,

where $f_j(\rho)$ is defined in (6), to an arbitrary precision. Real data, however, are subject to a statistical uncertainty, and it is pointless to demand agreement that is more precise than this uncertainty. Previously described procedures (*e.g.* Livesey & Skilling, 1985) for finding a maximum entropy distribution have defined the set of feasible points by a constraint of the form $\sum\{[|F_{obs}(\mathbf{h}_j)| - |f_j(\boldsymbol{\rho})|]/\sigma_j\}^2 = m$, where σ_j is the standard deviation of $|F_{obs}(\mathbf{h}_j)|$ and the sum is over m reflections. This constraint contains nothing that requires the quantities within the braces to have a distribution that in any way resembles a normal distribution, and, further, has resulted in problems in computation. By contrast, the lack of precision in the data is relevant to the procedure described here only in the definition of the stopping rule. Refinement can continue until $[|F_{obs}(\mathbf{h}_j)| - |f_j(\boldsymbol{\rho})|]^2/\sigma_j^2 < 1$ for all j.

It can be shown (Luenberger, 1984) that any set of starting phases will lead to a unique maximum entropy map. As Bricogne (1984) has shown, however, different sets of starting phases may lead to different maps, each of which is equally consistent with the data. In these circumstances the total entropy of the map provides a criterion for choosing among different possible sets of starting phases.

The entropy maximization procedure described here is a means of finding an everywhere-positive electron distribution for which the amplitude of the structure factor, $|\langle \exp(2\pi i \mathbf{h}.\mathbf{r})\rangle|$, is equal to $|F_{obs}(\mathbf{h})|$ for an arbitrarily large set of reflections. It does not, however, make any use of the fact that a crystal is composed of atoms that have definite well known physical and chemical properties. It is thus a tool for obtaining a map into which an atomic model may be fitted, with subsequent refinement by least-squares methods. If non-negativity of electron density is a sufficient condition to determine a unique atomic structure by direct methods (Woolfson, 1987, and references therein), it follows logically that entropy maximization, which is a stronger condition, will find the same atomic structure. A characteristic of the maximum entropy distribution is that it maximizes the minimum density in the unit cell, thereby minimizing the probability that some reflection in the unmeasurable region of reciprocal space will have an amplitude for which there is no phase that will not cause the density at some point to be negative. Maximum entropy is not a necessary condition for an acceptable structure, so structures whose entropies are far from the global maximum cannot be ruled out. Nevertheless, it is at least a plausible conjecture that, in a centrosymmetric structure, the sign combination for the strongest reflections that has the highest entropy is likely to be the correct one, and that, in a non-centrosymmetric structure, the correct set of phases will give a distribution that has an entropy close to the maximum. Maximum entropy *is* an efficient way to express the mutual phase implications of a large set of amplitudes simultaneously.

References

BRICOGNE, G. (1984). *Acta Cryst.* A**40**, 410-445.

COLLINS, D. M. (1982). *Nature (London)*, **298**, 49-51.

GILL, P. E., MURRAY, W. & WRIGHT, M. M. (1981). *Practical Optimization.* New York: Academic Press.

JAYNES, E. T. (1979). *The Maximum Entropy Formalism*, edited by R. D. LEVINE & M. TRIBUS, pp. 15-118. Cambridge, MA: Massachusetts Institute of Technology.

LIVESEY, A. K. & SKILLING, J. (1985). *Acta Cryst.* A**41**, 113-122.

LUENBERGER, D. G. (1984). *Linear and Nonlinear Programming.* Reading, MA: Addison Wesley.

PRINCE, E. (1982). *Mathematical Techniques in Crystallography and Materials Science.* New York: Springer-Verlag.

PRINCE, E., SJÖLIN, L. & ALENLJUNG, R. (1988). *Acta Cryst.* A**44**, 216-222.

SHANNON, C. E. (1948). *Bell Syst. Tech. J.* pp. 379-423, 623-656.

SHORE, J. E. & JOHNSON, R. W. (1980). *IEEE Trans. Inf. Theory*, **IT-26**, 26-37; **IT-29**, 942-943.

STEWART, G. W. (1973). *Introduction to Matrix Computations.* New York: Academic Press.

WILKINS, S. W., VARGHESE, J. N. & LEHMANN, M. S. (1983). *Acta Cryst.* A**39**, 47-60.

WOOLFSON, M. M. (1987). *Acta Cryst.* A**43**, 593-612.

Acta Cryst. (1989). A**45**, 203-208

The Estimation of Triplet Invariants from Multi-Wavelength Data

BY E. A. KLOP, H. KRABBENDAM AND J. KROON

Laboratorium voor Kristal- en Structuurchemie, Rijksuniversiteit, Padualaan 8, 3584 *CH Utrecht, The Netherlands*

(*Received* 16 *May* 1988; *accepted* 28 *September* 1988)

Abstract

A two-step procedure is presented for the estimation of triplet invariants from multi-wavelength data. In the first step wavelength-independent structure-factor magnitudes of both the total structure and the light-atom substructure, together with associated structure-factor phase differences, are calculated explicitly *via* a modified Singh & Ramaseshan [*Acta Cryst.* (1968), B**24**, 35-40] procedure. In the second step these

0108-7673/89/020203-06$03.00