

## Construction of Maximum-Entropy Density Maps, and their Use in Phase Determination and Extension

BY E. PRINCE

*Materials Science and Engineering Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA*

(Received 2 June 1992; accepted 18 September 1992)

### Abstract

Methods for constructing everywhere-positive electron-density maps with Fourier amplitudes matching those for arbitrarily large sets of observed data, utilizing dual-function methods for maximization of entropy, are described. Possible strategies for utilizing these maps for the determination and extension of phases in macromolecular structure determination are suggested, and problems are discussed.

### Introduction

From the earliest work on direct methods (Harker & Kasper, 1948) it was recognized that the fact that electron density could never be negative implied restrictions on the possible phases of the structure factors for X-ray diffraction from real crystals. Karle & Hauptman (1950) put this principle on a mathematically rigorous basis by showing that certain matrices constructed from sets of structure factors (Toeplitz matrices) must be positive semi-definite, implying the nonnegativity of the determinants of these matrices and giving inequality relations among the structure factors. Karle & Hauptman expressed these inequalities in the form of determinants, and, because of the rapid increase in the complexity of the expansion of a determinant as its size increases, they did not receive practical use except in rather small problems. The inequalities can also be expressed in terms of the Cholesky factorization of a positive-definite matrix (Prince, 1989a), and it is apparent from the form of the resultant expressions that the restrictions must get tighter as the size of the matrix increases. Furthermore, because the crystal is composed of atoms, all structure factors must be functions of a finite number of parameters, so that only a finite number of them can be independent, and the relations among sufficiently large numbers of them must be exact. If the number of atoms in the asymmetric unit is large, a 'sufficiently large number' of structure factors may be a very large number indeed! If, however, a nonnegative distribution can be found that gives a Fourier transform with amplitudes corresponding to a set of observed reflections, all relationships involving those reflections, including relationships with reflections not in the set, must be satisfied. Therefore phases can be extended from a sufficiently large starting set.

Crystallography is only one of a number of fields in which the data are measures of a Fourier transform of a density distribution. In some of these fields, radio astronomy for example, relative phases can be measured, but the data never extend to infinity in transform space, so they are necessarily incomplete. In recent years the principle of maximum entropy (Jaynes, 1979) has been used to define a unique solution to this underdetermined problem. The maximum-entropy solution can be shown to have a number of interesting properties. The property that has received the most attention is freedom from bias. In Jaynes's terms it is 'maximally noncommittal' with respect to all data that are not used in its construction. However, the property that is most significant in the crystallographic phase problem is nonnegativity; maximum entropy is a sufficient (but not necessary) condition to satisfy the requirement of an everywhere-positive electron-density distribution whose Fourier transform agrees with the observed data. Therefore a procedure that finds a maximum-entropy distribution finds a distribution that satisfies all Karle-Hauptman inequality relationships.

Maximum entropy has been applied to crystallography by Collins (1982), Wilkins (1983), Navaza (1985) and, notably, by Bricogne (1984). Except for Wilkins, who described a 'single-pixel approximation' procedure, these papers discuss the principles of maximum entropy, but do not give practical numerical solutions to the problem. Collins & Mahar (1983) showed that maximum entropy implies that the density can be expressed by an exponential function, but they also did not give a detailed solution. Livesey & Skilling (1985) discussed a procedure for finding the solution, but they did not describe their procedure in sufficient detail for a program to be written by someone else. Prince, Sjölin & Alenljung (1988) described a procedure for finding a maximum-entropy distribution under conditions of an overall 'soft' constraint, and Sjölin, Svensson, Prince & Sundell (1990) applied this procedure to phase extension in a protein. More recently Prince (1989b) described a more general procedure that can be used to construct a density map whose Fourier coefficients match the individual structure amplitudes of an arbitrarily large set of observed reflections. Sjölin, Prince, Svensson & Gilliland (1991) used this procedure to generate phases *ab initio* to produce a map of a protein that was in remarkable agreement with one computed using phases from the refined

structure. In this paper details of the implementation of this procedure are described, and ways of applying it as a tool for phase determination and extension are discussed.

### Construction of a maximum-entropy map

Consider a unit cell divided into  $n$  subcells, commonly called *pixels*, and let  $\rho_k$  be the number of electrons in the pixel located at position  $\mathbf{r}_k$  within the unit cell. We wish to maximize

$$S = - \sum_{k=1}^n \rho_k \ln \rho_k \quad (1)$$

subject to

$$\sum_{k=1}^n \rho_k = F(000), \quad (2)$$

and

$$\left| \sum_{k=1}^n \rho_k \exp(2\pi i \mathbf{h}_j \cdot \mathbf{r}_k) \right| = |F(\mathbf{h}_j)| \quad (3)$$

for  $j = 1, 2, \dots, m$ . Because the phase of  $F(\mathbf{h}_j)$  is in principle unknown, the actual entropy maximum is a function of the assumed phases (Sjölin, Prince, Svensson & Gilliland, 1991). However, because maximum entropy is not a necessary condition for a nonnegative distribution, any distribution that maximizes  $S$  for some set of phases is a possible density map.

Of the various procedures for constrained optimization two are based on a partition of  $n$ -dimensional space into two mutually orthogonal subspaces of  $m$  (the constraint space) and  $n-m$  (the null space) dimensions, respectively. The algorithm described by Livesey & Skilling (1985) was (apparently) the *primal* method (Luenberger, 1984), in which a point is found in the null space that satisfies the constraints, a *feasible point*, and then a search is made entirely within the null space to find the maximum. It was shown previously (Prince, 1989b) that the conditions for a maximum in the null space for constant values of the constraint functions can be solved analytically, leading to a straightforward application of the *dual* method, in which a search is made in the constraint space for a feasible point. The problem of finding the constrained maximum of  $S$  is in fact equivalent to the problem of finding the unconstrained minimum of its *dual function*, which is defined by

$$\Phi(\mathbf{x}) = - \sum_{k=1}^n \rho_k \ln \rho_k + \sum_{j=1}^m x_j [|F_{\text{map}}(\mathbf{h}_j, \mathbf{x})| - |F_{\text{obs}}(\mathbf{h}_j)|] \cos \Delta\varphi_j, \quad (4)$$

where

$$F_{\text{map}} = \sum_{k=1}^n \rho_k(\mathbf{x}) \exp(2\pi i \mathbf{h}_j \cdot \mathbf{r}_k), \quad (5)$$

and

$$\rho_k(\mathbf{x}) = \exp\left[\sum_{j=1}^m x_j \cos(2\pi \mathbf{h}_j \cdot \mathbf{r}_k - \varphi_j)\right]. \quad (6)$$

$\mathbf{x}$  is a vector of parameters that is related to the Lagrange multipliers of the problem, although, as has been pointed out by Lemaréchal & Navaza (1991), the true Lagrange multipliers are complex, whereas the elements of  $\mathbf{x}$  are real.  $\varphi_j$  is a trial phase for  $F(\mathbf{h}_j)$ , while the factor  $\cos \Delta\varphi_j$  reflects the fact that the phase of  $F_{\text{map}}(\mathbf{h}_j)$ , although it is in practice close to  $\varphi_j$ , is not necessarily equal to it. The gradient of the dual function is given by

$$\nabla \Phi(\mathbf{x})_j = [|F_{\text{map}}(\mathbf{h}_j, \mathbf{x})| - |F_{\text{obs}}(\mathbf{h}_j)|] \cos \Delta\varphi_j, \quad (7)$$

so that it vanishes when the constraints are satisfied. Its Hessian matrix is

$$\mathbf{H}(\mathbf{x}) = \mathbf{C}\mathbf{P}(\mathbf{x})\mathbf{C}^T, \quad (8)$$

where

$$C_{jk} = \cos(2\pi \mathbf{h}_j \cdot \mathbf{r}_k - \varphi_j), \quad (9)$$

and  $\mathbf{P}$  is a diagonal matrix in which  $P_{kk} = \rho_k(\mathbf{x})$ . If the constraints are independent,  $\mathbf{C}$  has full-row rank, and, because, from (6),  $\rho_k$  is positive for all  $k$ ,  $\mathbf{H}$  is everywhere positive definite (Stewart, 1973).  $\Phi(\mathbf{x})$  therefore has a unique minimum unless some  $\varphi_j$  is impossible because of a Karle-Hauptman inequality, in which case the entropy is negatively infinite.

A typical element of  $\mathbf{H}(\mathbf{x})$  is

$$H_{jk}(\mathbf{x}) = \sum_{l=1}^n \rho_l(\mathbf{x}) \cos(2\pi \mathbf{h}_j \cdot \mathbf{r}_l - \varphi_j) \times \cos(2\pi \mathbf{h}_k \cdot \mathbf{r}_l - \varphi_k), \quad (10)$$

which reduces to

$$H_{jk}(\mathbf{x}) = \frac{1}{2} \sum_{l=1}^n \rho_l(\mathbf{x}) \{ \cos[2\pi(\mathbf{h}_j + \mathbf{h}_k) \cdot \mathbf{r}_l - (\varphi_j + \varphi_k)] + \cos[2\pi(\mathbf{h}_j - \mathbf{h}_k) \cdot \mathbf{r}_l - (\varphi_j - \varphi_k)] \}. \quad (11)$$

Each term in this expression is itself a Fourier summation, so that

$$H_{jk}(\mathbf{x}) = \frac{1}{2} \{ |F_{\text{map}}(\mathbf{h}_j + \mathbf{h}_k)| \times \cos[\varphi_{j+k} - (\varphi_j + \varphi_k)] + |F_{\text{map}}(\mathbf{h}_j - \mathbf{h}_k)| \times \cos[\varphi_{j-k} - (\varphi_j - \varphi_k)] \}. \quad (12)$$

Thus each element is a simple function of two amplitudes and four phases, all of which are available without further calculation if the Fourier transform of the map is computed using fast Fourier transform (FFT) routines. [Note that the expressions in (10), (11) and (12) are exact for the actual

starting map. Problems often occurring in discrete Fourier transforms because of an index that is close to or greater than half the number of grid points in that direction do not arise.]

Because there is always a small number of reflections whose phases can be defined arbitrarily, in order to fix the origin and sometimes to choose an enantiomorph, choosing phases can be considered to be 'phase extension from a starting set' even when there is no *a priori* phase information. Therefore, the map constructed always contains prior knowledge, and it is necessary to consider how this prior knowledge can be incorporated into the construction of a maximum-entropy map containing additional information. In this work the prior map is represented by an image of the map in logarithmic space and two additional parameters, the logarithm of a scale factor, which is adjusted so as to hold  $F(000)$  constant, and the logarithm of a contrast factor. Let  $\psi_k$  be the value of the logarithmic map in the  $k$ th pixel, and denote the logarithms of the contrast factor and the scale factor by  $x_0$  and  $x_{m+1}$ , respectively. It is convenient to scale the logarithmic map so that  $0 \leq \psi_k \leq 1$ . Then the number of electrons in the  $k$ th pixel is

$$\rho_k(\mathbf{x}) = \exp(x_0\psi_k + x_{m+1}) \times \exp\left[\sum_{j=1}^m x_j \cos(2\pi\mathbf{h}_j \cdot \mathbf{r}_k - \varphi_j)\right]. \quad (13)$$

As additional reflections are added,  $x_{m+1}$  is adjusted to hold  $F(000)$  constant. The overall constraint used by Prince, Sjölin & Alenljung (1988) also holds the quantity

$$E = \sum_{k=1}^n \rho_k \psi_k, \quad (14)$$

the 'expected value' of  $\psi$ , constant. To see the effect of this constraint, consider the fact that  $\psi(\mathbf{r})$  is the sum of a Fourier series containing a finite number,  $p$ , of terms, which may be represented by

$$\psi_k = \sum_{l=1}^p a_l \cos(2\pi\mathbf{h}_l \cdot \mathbf{r}_k - \varphi_l), \quad (15)$$

whereas  $\rho(\mathbf{r})$  is represented by a general Fourier series. Because of the orthogonality of the trigonometric functions, however, all terms of the sum in (14) vanish except for those that also appear in the sum in (15), so that

$$E = \sum_{k=1}^n \sum_{l=1}^p a_l |F(\mathbf{h}_l)| \cos^2(2\pi\mathbf{h}_l \cdot \mathbf{r}_k - \varphi_l), \quad (16)$$

which reduces to

$$E = \frac{1}{2} \sum_{l=1}^p a_l |F(\mathbf{h}_l)|. \quad (17)$$

If the phase of  $F(\mathbf{h}_l)$  was equal to  $\varphi_l$  and  $d|F(\mathbf{h}_l)|/da_l$  was the same for all values of  $l$ , this would place a rigid constraint on the height of the origin peak of the Patterson function. However, neither of these conditions is true in general, so that, although the constraint does require the updated map to retain most of the features of the prior map, the phases of acentric reflections are allowed to drift, and some readjustment of the ratios of the  $|F|$  values also occurs. This constraint adds an additional row and column to the Hessian matrix of the dual function, which takes the form

$$H_{0j}(\mathbf{x}) = \sum_{k=1}^n \psi_k \rho_k \cos(2\pi\mathbf{h}_j \cdot \mathbf{r}_k - \varphi_j). \quad (18)$$

### Implementation

Because the constrained maximization of entropy corresponds to the unconstrained minimization of a dual function whose Hessian matrix is everywhere positive definite, the problem can be solved numerically using well developed procedures for nonlinear optimization. Many existing crystallographic least-squares programs are based on the Gauss-Newton algorithm, which is the iterative application of a local solution of the linear least-squares problem. Denoting the estimate of  $\mathbf{x}$  in the  $n$ th iteration by  $\mathbf{x}_n$  and the gradient of  $\Phi(\mathbf{x})$  at  $\mathbf{x}_n$  by  $\mathbf{g}_n$ ,

$$\mathbf{x}_{n+1} = \mathbf{x}_n - [\mathbf{H}(\mathbf{x}_n)]^{-1} \mathbf{g}_n. \quad (19)$$

While this procedure works satisfactorily for many problems, it has a number of defects, particularly for large problems. The Hessian matrix must be computed and inverted in each iteration, and global convergence is not assured. Both computational efficiency and stability may be improved in various ways. Equation (19) has the form  $\mathbf{x}_{n+1} = \mathbf{x}_n - \mathbf{s}_n$ , where  $\mathbf{s}_n$  is the solution of the matrix equation  $\mathbf{H}(\mathbf{x}_n)\mathbf{s}_n = \mathbf{g}_n$ . Convergence can be assured by including a line search, so that  $\mathbf{x}_{n+1} = \mathbf{x}_n - \alpha\mathbf{s}_n$ , where  $\alpha$  is a positive scalar. Because  $\mathbf{H}(\mathbf{x}_n)$  is positive definite, its Cholesky factor, a lower triangular matrix  $\mathbf{L}$  with positive diagonal elements such that  $\mathbf{L}\mathbf{L}^T = \mathbf{H}(\mathbf{x}_n)$ , exists, and  $\mathbf{s}_n$  can be found by successive solution of the two triangular systems  $\mathbf{L}\mathbf{b} = \mathbf{g}_n$  and  $\mathbf{L}^T\mathbf{s}_n = \mathbf{b}$ . The efficiency of computation is further enhanced by making use of the fact that the elements of  $\mathbf{H}(\mathbf{x})$  are often not rapidly changing functions of  $\mathbf{x}$ . If  $\mathbf{H}$  were the Hessian matrix of a quadratic function, the gradient would vanish at the point  $\mathbf{x}_n - \mathbf{s}_n$ . Since  $\Phi(\mathbf{x})$  is *not* a quadratic function, the gradient will likely not vanish at  $\mathbf{x}_n - \alpha\mathbf{s}_n$  for any value of  $\alpha$ . However, let  $\mathbf{y}_n = \mathbf{g}_{n+1} - \mathbf{g}_n$ . Denoting  $\mathbf{H}(\mathbf{x}_n)$  by  $\mathbf{H}_n$ , it can be shown (Gill, Murray & Wright, 1981) that

$$\begin{aligned} \mathbf{H}_{n+1} &= \mathbf{H}_n + \mathbf{y}_n \mathbf{y}_n^T / \alpha \mathbf{y}_n^T \mathbf{s}_n \\ &\quad - \mathbf{H}_n \mathbf{s}_n \mathbf{s}_n^T \mathbf{H}_n / \mathbf{s}_n^T \mathbf{H}_n \mathbf{s}_n \end{aligned} \quad (20)$$

is the Hessian matrix of a quadratic function that has the same values of the gradient at  $x_n$  and  $x_{n+1}$  as  $\Phi(x)$ . It can be further shown that this correction formula, which is known as the *Broyden-Fletcher-Goldfarb-Shanno* (BFGS) update, has the property of *hereditary positive definiteness*. That is, if  $H(x_n)$  is positive definite, then  $H(x_{n+1})$  is also, provided a line search that is sufficient to assure convergence has been performed.

The number of arithmetic operations required to compute  $L$  from  $H$  is proportional to  $m^3$ , where  $m$  is the size of the matrix, and the application of (20) would appear to require that this be done in every iteration. However, an update applied directly to  $L$  that is equivalent to the application of the BFGS update to  $H$  can be computed with a number of operations that is proportional to  $m^2$  (Gill, Murray & Wright, 1981; Dongarra, Bunch, Moler & Stewart, 1979), so that the Cholesky factorization needs to be performed only once. As noted above, all of the information required to compute the initial elements of  $H$  are available when the prior map has been transformed using fast Fourier transform (FFT) routines, and FFT routines that reduce the computation by a factor of the order of the space group can be written using the methods described by Ten Eyck (1973). Using these techniques, a maximum-entropy map with 100 000 pixels in the asymmetric unit and fitting a block of 500 amplitudes can be constructed in less than 1 min of CPU time on a MicroVAX 4000.\* Fig. 1 is a flowchart of a program written to implement this procedure.

### Discussion

Application of the procedures described above will generate an everywhere-positive density map with Fourier amplitudes that agree with the observed data for an arbitrarily large set of reflections. This can in principle lead to extension of phases from a sufficiently large starting set, but the question of how large is sufficient remains. In addition, the analysis assumes implicitly that the data are on an absolute scale, or, equivalently, that the value of  $F(000)$  on the scale of the data is known. If there are values of the unitary structure factor,  $U(h) = F(h)/F(000)$ , that are large enough for Harker-Kasper or Karle-Hauptman inequalities to put restrictions on phases, the maximum-entropy procedure will give the same result, because an impossible phase will lead to a negatively infinite entropy. In an early study of the structure of L-glutamine (Cochran & Penfold, 1952; Woolfson, 1987) it was discovered afterwards that the value of  $F(000)$  had been badly underestimated, and that none of the inequality relationships used was valid! If  $F(000)$  had been correctly estimated,

\*MicroVAX 4000 is a registered trademark of the Digital Equipment Corporation. Trade names are identified in order to specify adequately the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the product is the best available for the purpose.

the inequality relationships would have given no phase indications, because the maps computed using incorrect phases would still have been nonnegative. The entropies of the incorrect phase maps would, however, have been very low, and a maximum-entropy procedure would have led to the correct result.

For macromolecular crystals the largest values of  $|F(h)|$  tend to be at most a few percent of  $F(000)$ . Nevertheless, Sjölin, Prince, Svensson & Gilliland (1991) carried out a study, using the entropy as a figure of merit, that compared maps with various phase combinations and the observed amplitudes from the known (Gilliland, Winborne, Nachman & Wlodawer, 1990) structure of recombinant bovine chymosin, and obtained a map with 1970 of the largest reflections that was strikingly similar to one computed with all data and phases from the refined structure. These authors also studied the effect of scale by testing the early stages of phase determination with assumed values of  $F(000)$  that were a factor of two larger and a factor of two smaller than the correct value. They found that the correct estimation of a block of 16 signs for centric reflections when the prior map contained only origin-defining reflections was rather sensitive to scale, but that the estimation of the phases of acentric reflections when the prior map was made with 67 centric reflections was not seriously affected by scale.

Maximum-entropy methods have been used to extend phases in fragment TR<sub>2</sub>C from bull testis calmodulin (Sjölin, Svensson, Prince & Sundell, 1990), but no macromolecular structure has been solved *ab initio* by these methods. It is not clear that entropy is the best available,

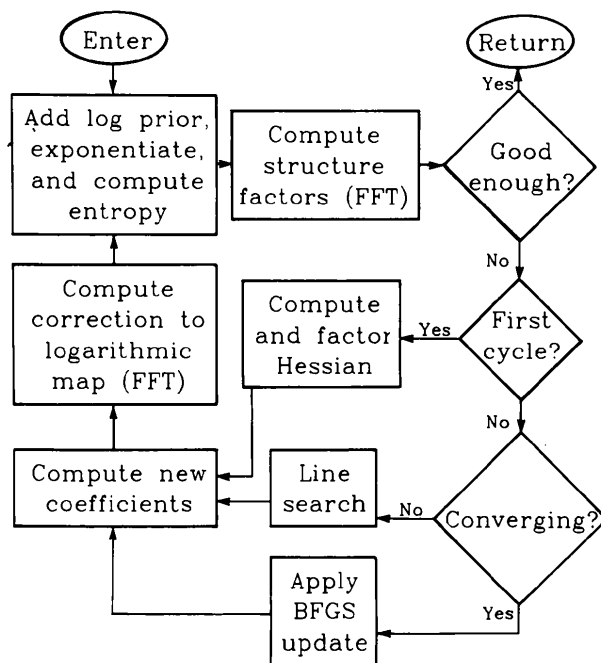


Fig. 1. Flow chart of an algorithm for finding the maximum-entropy map consistent with a set of observed amplitudes.

or even a particularly good, figure of merit. In the early stages of phase determination any set of phases produces an everywhere-positive map, but some figures of merit should be able to select phase sets that are more likely to be correct than others. A characteristic of entropy is that it is biased against maps that have high sharp peaks, which, because proteins are composed mostly of atoms with low atomic numbers, suggests that low-resolution maps should have high entropy.

Gilmore, Bricogne & Bannister (1990) have proposed a likelihood function, which is based on the effect that the reflections in the set that is being tested have on others outside of that set. This function has been used effectively for determining the structures of molecules that are outside the range of conventional direct methods, but it depends on the existence of reflections with moderately large values of  $|F(h)|/F(000)$ , a condition that is not generally satisfied for crystals of proteins.

Centric reflections, for which phases can have only one of two values, can be tested in blocks, using a fractional factorial design (Box, Hunter & Hunter, 1978), for example. The phases for acentric reflections must be chosen by a much more laborious, one-at-a-time procedure. The tolerance on the agreement between observed and calculated amplitudes must be made tight enough so that the figure of merit changes much less within the acceptance region than it does when phases are changed, even though this implies a tighter tolerance than could be justified on the basis of the statistical precision of the observed values.

Many stimulating discussions with D. M. Collins are gratefully acknowledged. Part of this work was performed with financial assistance from the Swedish Natural Sciences Research Council (NFR).

## References

- BOX, G. E. P., HUNTER, W. G. & HUNTER, J. S. (1978). *Statistics for Experimenters*. New York: John Wiley.
- BRICOGNE, G. (1984). *Acta Cryst.* **A40**, 410-445.
- COCHRAN, W. & PENFOLD, B. R. (1952). *Acta Cryst.* **5**, 644-654.
- COLLINS, D. M. (1982). *Nature (London)*, **298**, 49-51.
- COLLINS, D. M. & MAHAR, M. C. (1983). *Acta Cryst.* **A39**, 252-256.
- DONGARRA, J. J., BUNCH, J. R., MOLER, C. B. & STEWART, G. W. (1979). *LINPACK User's Guide*. Philadelphia: Society for Industrial and Applied Mathematics.
- GILL, P. E., MURRAY, W. & WRIGHT, M. H. (1981). *Practical Optimization*. New York: Academic Press.
- GILLILAND, G. L., WINBORNE, E. L., NACHMAN, J. & WLODAWER, A. (1990). *Proteins Struct. Funct. Genet.* **8**, 82-101.
- GILMORE, C. J., BRICOGNE, G. & BANNISTER, C. (1990). *Acta Cryst.* **A46**, 297-308.
- HARKER, D. & KASPER, J. S. (1948). *Acta Cryst.* **1**, 70-75.
- JAYNES, E. T. (1979). *The Maximum Entropy Formalism*, edited by R. D. LEVINE & M. TRIBUS, pp. 15-118. Cambridge, MA: MIT Press.
- KARLE, J. & HAUPTMAN, H. (1950). *Acta Cryst.* **3**, 181-187.
- LEMARECHAL, C. & NAVAZA, J. (1991). *Acta Cryst.* **A57**, 631-632.
- LIVESKEY, A. K. & SKILLING, J. (1985). *Acta Cryst.* **A41**, 113-122.
- LUENBERGER, D. G. (1984). *Linear and Non-linear Programming*. Cambridge, MA: Addison-Wesley.
- NAVAZA, J. (1985). *Acta Cryst.* **A41**, 232-244.
- PRINCE, E. (1989a). *Acta Cryst.* **A45**, 144.
- PRINCE, E. (1989b). *Acta Cryst.* **A45**, 200-203.
- PRINCE, E., SJÖLIN, L. & ALENLJUNG, R. (1988). *Acta Cryst.* **A44**, 216-222.
- SJÖLIN, L., PRINCE, E., SVENSSON, L. A. & GILLILAND, G. L. (1991). *Acta Cryst.* **A47**, 216-223.
- SJÖLIN, L., SVENSSON, L. A., PRINCE, E. & SUNDELL, S. (1990). *Acta Cryst.* **B46**, 209-215.
- STEWART, G. W. (1973). *Introduction to Matrix Computations*. New York: Academic Press.
- TEN EYCK, L. F. (1973). *Acta Cryst.* **A29**, 183-191.
- WILKINS, S. W. (1983). *Acta Cryst.* **A39**, 892-896.
- WOOLFSON, M. M. (1987). *Acta Cryst.* **A43**, 593-612.