Acta Cryst. (1976). A32, 17

Phase Extension by Means of the Square Root of the Electron Density

BY A. R. DAVIES* AND J. S. ROLLETT

University Computing Laboratory, 19 Parks Road, Oxford, OX1 3PL, England

(*Received* 4 *June* 1975; *accepted* 26 *June* 1975)

A method is described which provides phase estimates for structure factors. The data required are the moduli of some or all of the structure factors and phases for a sub-set of them. The Fourier components corresponding to a real square root of the electron density are estimated by fitting their self convolution to the available data by least squares. The structure factors are modified to improve resolution and maintain non-negativity of the electron density. Suitable weighting of the observations assists convergence and estimates of error for the electron density map are provided. The results of tests of the method are described.

1. Introduction

Most crystal structures can be made to yield diffracted beams of observable intensity for reflexions from sets of planes with spacings comparable with interatomic distances. It is much less often true that the application of direct methods, or of methods related to that of multiple isomorphous replacement, provides phase estimates for the majority of the structure factors which can be observed. Commonly the phase estimates which can be obtained are relatively few. For protein structures they often correspond to reflexions of large spacing only. Hence there is a need for methods of analysis which can extend phase sets to improve resolution. For protein structures in particular there is a need to do this in such a way that the accuracy of the resulting electron density map can be estimated reliably.

Biraud (1969) has described a method of improving resolution which is applicable to a real non-negative function. This possesses a real (but possibly negative) square root. The Fourier transform of the square root is Hermitian and can be convoluted with itself to fit the Fourier transform of the original function. We have used least squares to fit the self-convolute of the Fourier components of a square-root function to the amplitudes and phase estimates for the structure factors. The transform of the self-convolute is an electron-density estimate which agrees with the experimental data and is non-negative. It can, in general, be made to enhance resolution.

The starting approximation is obtained by assuming that the electron density and its square root resemble each other. Hence the initial Fourier component estimates for the square root are taken to be the same as those for the structure itself, apart from a Gaussian correction for change of shape of the peaks. This starting approximation is based on the relationship of Sayre (1952). The least-squares fitting process does not require that the electron density resemble its square root, however, and there is no reason why it should not yield an accurate fit to a density function with peaks of varying heights. This makes our method attractive for application to protein structures with heavy atoms, variations of vibrational amplitude from one atom to the next, and regions of disordered solvent. So far as our present work has gone, the less restrictive condition of non-negativity (rather than resemblance to the square) has been adequate to define the electron density.

The number of structure factors is always finite. We must therefore use a finite number of terms in the series for the square-root function. Boas & Kac (1945) have found limits on the amplitudes of a one-dimensional Fourier transform which can represent a non-negative function that tends to zero as the argument tends to infinity. Davies (1974) has extended this theory to cover periodic functions in three dimensions. A necessary (but not sufficient) condition on the ratio $|F(\mathbf{h})|/|F(\mathbf{0})|$ is obtained. The condition is $|F(\mathbf{h})| \leq \frac{1}{2}|F(\mathbf{0})|\{2\cos[\pi/([r_0/|\mathbf{h}|]+2)]\}^{n(\mathbf{h})}$ where r_0 is the limiting value of $|\mathbf{h}|$ for non-zero $|F(\mathbf{h})|$, $[r_0/|\mathbf{h}|]$ is the largest integer not exceeding $r_0/|\mathbf{h}|$, and $n(\mathbf{h})$ is the number of non-zero indices in $\mathbf{h}$. This condition turns out to be less stringent than the condition we have applied to ensure that the truncation error of our approximation shall be small. The truncation condition is that $|F_0(\mathbf{h})|$ does not exceed one per cent of $|F_0(\mathbf{0})|$ when $|\mathbf{h}|$ is so large that $F_c(\mathbf{h})$ must be zero, and this condition has been established empirically by some of the test calculations which we describe in § 4. Notation is described in the Appendix.

2. Method

We define P as the set of $\mathbf{h} \equiv (h,k,l)$ for which both modulus and phase are given. $F_0(\mathbf{h})$ is the observed (generally complex) structure factor, which is modified by a Gaussian factor to ensure that the truncation condition is satisfied. $\varphi_0(\mathbf{h})$ is its phase angle. $C(\mathbf{h}')$ is the component of order $\mathbf{h}'$ in the Fourier series for the square root of the electron density, and we calculate

* Present address: Atlas Computing Laboratory, Chilton, Didcot, Oxfordshire OX11 OQY, England.

$F_c(\mathbf{h})$ as

$$F_c(\mathbf{h}) = \sum_{C\text{-set}} C(\mathbf{h}') \, . \, C(\mathbf{h}-\mathbf{h}')$$

where the C-set is the set of $\mathbf{h}'$ for which $C(\mathbf{h}') \neq 0$. If the phase angle of $F_c(\mathbf{h})$ is $\varphi_c(\mathbf{h})$ then we can calculate the deviation of $F_c(\mathbf{h})$ from $F_o(\mathbf{h})$ in the direction parallel to $F_o(\mathbf{h})$ as

$$\Delta_{\parallel}(\mathbf{h}) = |F_o(\mathbf{h})| - |F_c(\mathbf{h})| \cos [\varphi_o(\mathbf{h}) - \varphi_c(\mathbf{h})] \, .$$

Similarly the deviation of $F_c(\mathbf{h})$ from $F_o(\mathbf{h})$ in the direction perpendicular to this is

$$\Delta_{\perp}(\mathbf{h}) = |F_c(\mathbf{h})| \sin [\varphi_o(\mathbf{h}) - \varphi_c(\mathbf{h})] \, .$$

We minimize, by least-squares, a function which contains one term for each member of the P-set, of the form

$$w_{\parallel}(\mathbf{h}) \Delta^2_{\parallel}(\mathbf{h}) + w_{\perp}(\mathbf{h}) \Delta^2_{\perp}(\mathbf{h}) \, .$$

In general, the weight $w_{\parallel}$ will be larger than $w_{\perp}$, because of the precision of the radial component of $F_o(\mathbf{h})$ and the relative uncertainty of the cross-radial component. This choice of observations ensures that the covariance between $\Delta_{\parallel}$ and $\Delta_{\perp}$ is zero, which is necessary for the validity of the least-squares method with simple weights.

We define M as the set of $\mathbf{h}$ for which the modulus, but not the phase, is given. For each member of the M-set we define

$$\Delta_M(\mathbf{h}) = |F_b(\mathbf{h})| - |F_c(\mathbf{h})| \, .$$

Each member of the M-set makes a contribution to the minimization function equal to $w_M(\mathbf{h}) \Delta^2_M(\mathbf{h})$. $F_b(\mathbf{h})$ is the centroid of the probability distribution of the unknown true value of the structure factor, $F_t(\mathbf{h})$, based on knowledge of $|F_o(\mathbf{h})|$ and $F_c(\mathbf{h})$.

The contributions to the variances of $\Delta_{\parallel}$, $\Delta_{\perp}$ and Δ_M, and hence the appropriate $w_{\parallel}$, $w_{\perp}$ and w_M, are discussed in § 3.

The parameters in the calculation define the $C(\mathbf{h}')$. It would be possible to use the real and imaginary parts of each $C(\mathbf{h}')$ as parameters, but we have chosen to use $|C(\mathbf{h}')|$ and $\alpha(\mathbf{h}')$, the phase angle of $C(\mathbf{h}')$. This has the disadvantage that $\alpha(\mathbf{h}')$ is indeterminate when $|C(\mathbf{h}')|$ is zero, but that has not presented any problems except the need for caution in setting up a starting model. The $F_c(\mathbf{h})$ are non-linear functions of $|C(\mathbf{h}')|$ and $\alpha(\mathbf{h}')$, and the least-squares equations are linearized in the usual way (see for example Rollett, 1965, pp. 32–35). The derivatives required are:

For $\mathbf{h}$ in the P-set

$$\frac{\partial \Delta_{\parallel}(\mathbf{h})}{\partial |C(\mathbf{h}')|} = -2 \sum_{\mathbf{h}^* \in S\mathbf{h}'} |C(\mathbf{h}-\mathbf{h}^*)| \cos A(\mathbf{h}\mathbf{h}^*) \, ,$$

$$\frac{\partial \Delta_{\parallel}(\mathbf{h})}{\partial \alpha(\mathbf{h}')} = +2|C(\mathbf{h}')| \times \sum_{\mathbf{h}^* \in S\mathbf{h}'} \frac{\partial \alpha(\mathbf{h}^*)}{\partial \alpha(\mathbf{h}')} |C(\mathbf{h}-\mathbf{h}^*)| \sin A(\mathbf{h}\mathbf{h}^*) \, ,$$

$$\frac{\partial \Delta_{\perp}(\mathbf{h})}{\partial |C(\mathbf{h}')|} = -2 \sum_{\mathbf{h}^* \in S\mathbf{h}'} |C(\mathbf{h}-\mathbf{h}^*)| \sin A(\mathbf{h}\mathbf{h}^*) \, ,$$

$$\frac{\partial \Delta_{\perp}(\mathbf{h})}{\partial \alpha(\mathbf{h}')} = -2|C(\mathbf{h}')| \times \sum_{\mathbf{h}^* \in S\mathbf{h}'} \frac{\partial \alpha(\mathbf{h}^*)}{\partial \alpha(\mathbf{h}')} |C(\mathbf{h}-\mathbf{h}^*)| \cos A(\mathbf{h}\mathbf{h}^*) ;$$

and for $\mathbf{h}$ in the M-set

$$\frac{\partial \Delta_M(\mathbf{h})}{\partial |C(\mathbf{h}')|} = -2 \sum_{\mathbf{h}^* \in S\mathbf{h}'} |C(\mathbf{h}-\mathbf{h}^*)| \cos B(\mathbf{h}\mathbf{h}^*) \, ,$$

$$\frac{\partial \Delta_M(\mathbf{h})}{\partial \alpha(\mathbf{h}')} = +2|C(\mathbf{h}')| \sum_{\mathbf{h}^* \in S\mathbf{h}'} |C(\mathbf{h}-\mathbf{h}^*)| \sin B(\mathbf{h}\mathbf{h}^*) ;$$

where

$$A(\mathbf{h}\mathbf{h}^*) = \alpha(\mathbf{h}^*) + \alpha(\mathbf{h}-\mathbf{h}^*) - \varphi_o(\mathbf{h}) \, ,$$

$$B(\mathbf{h}\mathbf{h}^*) = \alpha(\mathbf{h}^*) + \alpha(\mathbf{h}-\mathbf{h}^*) - \varphi_c(\mathbf{h}) \, ,$$

and $\mathbf{h}^* \in S\mathbf{h}'$ means that $\mathbf{h}^*$ is one of the reciprocal vectors related to $\mathbf{h}'$ by symmetry.

The weights vary from cycle to cycle. They are regarded as a means of choosing the most appropriate minimization function for the cycle, rather than as variables in this function. Thus we do not differentiate the weights in forming the normal equations. The number of $\mathbf{h}$ in the C-set will usually be comparable with the number in the P-set plus the M-set. For a typical three-dimensional calculation on a large structure the number of parameters is nearly twice the number of $\mathbf{h}'$ in the C set [some of the $C(\mathbf{h}')$ may have phases restricted by symmetry] and is inconveniently large. We have adopted the following procedure to avoid severe data handling difficulties:

(*a*) Refine $|C(\mathbf{h}')|$ and $\alpha(\mathbf{h}')$ for all k' and l' corresponding to a particular value of h', keeping the parameters for other h' fixed.

(*b*) Advance h' and repeat (*a*) until all h' have been dealt with, but base the calculation for each value of h' on the same starting parameters.

(*c*) Create a vector of shifts by adding together all the shifts produced by operations (*a*) and (*b*), and search along the direction of this vector in parameter space for a minimum of the minimization function. Use this minimum as the set of starting parameters for operation (*d*)

(*d*) Repeat operations (*a*) to (*c*), refining parameters for particular values of k' at each stage.

(*e*) Repeat operations (*a*) to (*c*), refining parameters for particular values of l' at each stage.

It might be supposed that the linear search procedure of operation (*c*) would be time consuming, but this is not in fact a dominant part of the time required for the program because no derivatives of the minimization function are needed except that at the start point, and this can be calculated easily from the derivatives in the normal equations. A search procedure of

Bard (1970) was used. In this a parabola is fitted to the value and gradient of the function at the start of the step and the value only at one other point. It was found to require no more than three function evalutions per step in most cases.

It is not easy to eastablish the effectiveness of this 'alternating-directions' algorithm in a rigorous way, but it has been found to converge quite well, and faster than an algorithm in which steps (*d*) and (*e*) are omitted.

The program we have used was written in Fortran for the ICL 1906A computer, and was designed for orthorhombic symmetry only. Care was taken to minimize the time required to run it by removing all possible operations from the innermost loops. Details of the programming techniques are given by Davies (1974), Chap. 5.

3. Error analysis

We have wished to carry out the calculations in such a way as to produce convergence as rapidly as possible towards a minimum of the minimization function which corresponds to physical reality. We have therefore taken into account in our weighting system not only errors of the data, but also the contributions to $\Delta_{\parallel}$, $\Delta_{\perp}$ and Δ_M from the errors of the model used to fit the data.

Our guiding principle has been to calculate the sum of the contributions to the variance for each Δ value and to use its reciprocal as a weight. For the *P*-set the contributions are, for $\Delta_{\parallel}$:

(*a*) The variance of $F_o(\mathbf{h})$ in the radial direction due to experimental error; we have assumed that this can be provided with the data.

(*b*) The truncation error of $F_c(\mathbf{h})$ due to the finite size of the *C*-set. This is discussed below.

For $\Delta_{\perp}$, the contributions are the same, except that the variance of $F_o(\mathbf{h})$ is that in the cross-radial direction due to the error in determining the phase angle $\varphi_o(\mathbf{h})$. We have assumed that the cross-radial variance of $F_o(\mathbf{h})$ can be provided with the data.

For the *M*-set, the contributions are:

(*a*) The variance of $|F_o(\mathbf{h})|$ due to experimental error.

(*b*) The truncation error of $|F_c(\mathbf{h})|$.

(*c*) The variance of $|F_b(\mathbf{h})|$ as an estimate of the component of $F_t(\mathbf{h})$ in the direction defined by $\varphi_c(\mathbf{h})$, due to the uncertainty of $\varphi_c(\mathbf{h})$.

At any stage we have taken the *C*-set to consist of all reciprocal points in a sphere bounded by a value of $|\mathbf{h}|$ which is small enough to ensure that the number of parameters is not greater than the number of observations. Outside a radius twice that of the *C*-set the value of $|F_c(\mathbf{h})|$ must be zero and the magnitude of the truncation error is therefore equal to $|F_o(\mathbf{h})|$ which can be estimated from the Gaussian factor applied to the observations and the requirement that $\langle |E|^2 \rangle = 1$ at all $|\mathbf{h}|$. In order to estimate the truncation error at a radius at which F_c is not zero we use a technique in which we assume that the truncation error is a complex random variable isotropically distributed about $F_c(\mathbf{h})$ and which is not correlated with the experimental errors in $F_o(\mathbf{h})$. An estimate of the truncation error is then given by its variance, which is found as follows. When a *C*-set is fitted to a *P*-set alone, and the moduli and phases for the *P*-set are exactly those for a non-negative electron density, the errors of fit are entirely due to truncation. We have found, in such cases, that it is satisfactory to take the natural logarithm of the truncation error to be a linear function of $|\mathbf{h}|^2$. We have estimated this function by fitting a straight line to the natural logarithm of the r.m.s. error for groups of *P*-set data points of various $|\mathbf{h}|$ values and also for the point at the limit of the non-zero F_c. It has also been satisfactory to estimate the truncation error for $|\mathbf{h}|$ values corresponding to the *M*-set by interpolation from the straight line. Our normal procedure has been to begin by fitting the *P*-set alone. At each cycle the truncation variance has been estimated from a straight-line plot. When convergence has been attained we have added the *M*-set to the data, extended the *C*-set to reduce the truncation error, and continued to estimate its variance from the fit to the *P*-set and from the point at which $|F_c(\mathbf{h})|$ falls to zero.

The variance of $|F_b(\mathbf{h})|$ due to lack of knowledge of $\varphi_t(\mathbf{h})$ in the *M*-set is derived by a Bayesian argument. The distribution of $F_c(\mathbf{h})$ about the true $F_t(\mathbf{h})$ is taken to be normal, bivariate for $F(\mathbf{h})$ with general phase, and univariate for $F(\mathbf{h})$ with phase restricted by symmetry. Bayes's theorem [see for example Lindley (1965)] is then used to combine this with the information available from knowledge of $|F_o(\mathbf{h})|$. If we ignore for this purpose the uncertainty in $|F_o(\mathbf{h})|$ then the true $F_t(\mathbf{h})$ has an *a priori* distribution [*i.e.* a distribution before we take into account knowledge of $F_c(\mathbf{h})$] which is either uniform on a circle centred at the origin, or concentrated at two points in the case of restricted phase. By Bayes's theorem

$$p[F_t(\mathbf{h})|F_c(\mathbf{h})] = \frac{p[F_c(\mathbf{h})|F_t(\mathbf{h})] \,.\, p[F_t(\mathbf{h})]}{\iint p[F_c(\mathbf{h})|F_t(\mathbf{h})] \,.\, p[F_t(\mathbf{h})] \mathrm{d}A},$$

where the integration is over the complex plane. $p[F_t(\mathbf{h})|F_c(\mathbf{h})]$ means the probability density of $F_t(\mathbf{h})$ given the value of $F_c(\mathbf{h})$, and so on. The centroid of the distribution and the variance in the direction defined by the phase of $F_c(\mathbf{h})$ can be found analytically. We write

$$\gamma_1(\mathbf{h}) = |F_c(\mathbf{h})| \, |F_o(\mathbf{h})|/\sigma_A^2(\mathbf{h})$$

where $\sigma_A^2(\mathbf{h})$ is the estimated variance of the distribution of $p[F_c(\mathbf{h})|F_t(\mathbf{h})]$, and

$$\mu_1(\mathbf{h}) = I_1[\gamma_1(\mathbf{h})]/I_0[\gamma_1(\mathbf{h})] \,,$$

where I_0 and I_1 are the zeroth and first-order Bessel functions of imaginary argument. For reflexions of

general phase we then get the centroid

$$|F_b(\mathbf{h})| = \mu_1(\mathbf{h})|F_o(\mathbf{h})|$$

and the variance

$$\sigma_b^2(\mathbf{h}) = |F_o(\mathbf{h})|^2\{1-\mu_1(\mathbf{h})/\gamma_1(\mathbf{h})-[\mu_1(\mathbf{h})]^2\} .$$

For reflexions of restricted phase we write

$$\mu_2(\mathbf{h}) = \tanh [\gamma_1(\mathbf{h})] .$$

The Bayes theorem integration reduces to a summation over the two permitted values of $F_t(\mathbf{h})$ and we get

$$F_b(\mathbf{h}) = \mu_2(\mathbf{h})|F_o(\mathbf{h})| \exp [i\varphi_c(\mathbf{h})]$$

$$\sigma_b^2(\mathbf{h}) = |F_o(\mathbf{h})|^2\{1-[\mu_2(\mathbf{h})]^2\} .$$

A problem is to estimate $\sigma_A^2(\mathbf{h})$ correctly. If the distribution of $F_t(\mathbf{h})$ about $F_c(\mathbf{h})$ could be assumed to be isotropic we could take this as

$$\sigma_A^2(\mathbf{h}) = \langle[|F_c(\mathbf{h})|-|F_o(\mathbf{h})|]^2\rangle_M .$$

Refinement of the parameters to fit the moduli for the M-set, but not the phases, destroys this isotropy however, and we have found empirically that it is satisfactory to use, for both types of reflexion:

$$\sigma_A^2(\mathbf{h}) = 4\langle[|F_c(\mathbf{h})|-|F_o(\mathbf{h})|]^2\rangle_M .$$

At early stages $\sigma_b^2(\mathbf{h})$ is large and cuts down the effect of terms with uncertain phases. Later $\sigma_b^2(\mathbf{h})$ is small and we use essentially minimum-variance weights. Small errors in the estimate of $\sigma_A^2(\mathbf{h})$ are unlikely to affect the weights seriously because μ_1 and μ_2 become slowly varying functions for reasonably small σ_A values.

Provided that the least-squares refinement is carried out with proper weighting, we can use the knowledge of the errors in the $F_c(\mathbf{h})$ that it provides to estimate the variance of the electron density obtained from them. An estimate can be made of the mean-square difference between the electron density based on $F_t(\mathbf{h})$ and that based on $F_c(\mathbf{h})$ [rather than the modulus of $F_o(\mathbf{h})$ and the phase of $F_c(\mathbf{h})$]. This is:

$$\langle(\Delta\varrho)^2\rangle = \frac{1}{V^2}\{[F_c(0)-F_o(0)]^2+2\sum_{\mathbf{h}'\neq 0}[|F_c(\mathbf{h})|^2 - 2K(\mathbf{h})|F_c(\mathbf{h})|\,|F_o(\mathbf{h})|+|F_o(\mathbf{h})|^2]\} ,$$

where

$$K(\mathbf{h}) = \cos(\varphi_c(\mathbf{h})-\varphi_o(\mathbf{h})) , \quad \mathbf{h}\in P ,$$
$$= \mu_1(\mathbf{h}) , \quad \mathbf{h}\in M \text{ with general phase,}$$
$$= \mu_2(\mathbf{h}) , \quad \mathbf{h}\in M \text{ with restricted phase}$$

and the F_c quantities are those obtained at convergence of the least-squares fitting process. The derivations of the $K(\mathbf{h})$ expressions depend on the assumption of isotropy of the distribution of $F_c(\mathbf{h})$ about $F_t(\mathbf{h})$, but it is unlikely that failure of this assumption would alter the estimate of $\langle(\Delta\varrho)^2\rangle$ by much. At the end of the refinement the $K(\mathbf{h})$ values are likely to be near unity and slowly varying with respect to $\gamma_1(\mathbf{h})$.

If we let the F_o data correspond to point atoms each with a vibration parameter β, then the truncation condition mentioned at the end of § 2 gives $\exp(-\beta|\mathbf{h}|^2) \leq 0{\cdot}01$ at the limit of non-zero $F_c(\mathbf{h})$. The error due to the finiteness of the series should then be small. We can therefore regard this estimate as a realistic error estimate for the electron density map consisting of point atoms with vibration β. If however the $F_c(\mathbf{h})$ are sharpened (*i.e.* the effective value of β is reduced) after the fitting process is complete and before the map is calculated, then there may be an appreciable finite series error. We comment further on this in § 4.

4. Tests of the method

We have applied the method described in §§ 2 and 3 to sets of F_o data for structures of symmetry $P2_12_12_1$. These sets correspond to hypothetical structures, generated by a program written by D. Sayre. That program selected coordinates for atoms of equal scattering power as the values of pseudo-random numbers uniformly distributed over ranges corresponding to the asymmetric unit. Atomic positions were rejected if they were (*a*) less than 1·3 Å from any previous position or (*b*) between 1·6 and 2·5 Å from a previous position. Three structures, one with two atoms, one with six atoms, and one with 16 atoms in the asymmetric unit, were generated and E_o values were produced for all $\mathbf{h}$ of spacing greater than 1 Å in each case.

The P-set was chosen in each case as the set of $\mathbf{h}$ of spacing $d(\mathbf{h})$ greater than 2·0 Å. The M-set was chosen so that $2{\cdot}0 \text{ Å} > d(\mathbf{h}) \geq d_M$ and the C-set with $d(\mathbf{h}) \geq d_C$. Values of d_M and d_C are given in Tables 1 to 3 and vary with the cycle number in most calculations. In order to investigate the effects of truncation errors, we have worked with

$$F_o(\mathbf{h}) = \text{const. } \exp(-\beta|\mathbf{h}|^2) \, . \, E_o(\mathbf{h}) .$$

The values of β are also given in the tables. For such a set of $F_o(\mathbf{h})$, the analysis of Sayre (1952) can be modified to show that $C(\mathbf{h}) = \theta(\mathbf{h}) \, . \, F_o(\mathbf{h})$ where $\theta(\mathbf{h}) =$ const. $\exp(-\beta|\mathbf{h}|^2)$. We have generated our starting approximations for $C(\mathbf{h})$ $(\mathbf{h}\in P)$ from this rule, with the constant chosen so that $F_c(\mathbf{0}) = F_o(\mathbf{0})$. When necessary we have generated starting approximations for $C(\mathbf{h})$ $(\mathbf{h}\in M)$ by

$$C(\mathbf{h}) = \text{const. } \theta(\mathbf{h})|F_o(\mathbf{h})| \, . \, \exp[i\varphi_c(\mathbf{h})] ,$$

where $\varphi_c(\mathbf{h})$ is the phase of the M-set term taken from the previous cycle. We define

$$j = \text{the number of cycles since the start of the run}$$
$$\langle|\Delta\varphi|\rangle = \langle|\varphi_t(\mathbf{h})-\varphi_c(\mathbf{h})|\rangle \text{ in degrees}$$
$$\sigma(F) = \langle|F_t(\mathbf{h})-F_c(\mathbf{h})|^2\rangle^{1/2}$$
$$\sigma(|F|) = \langle||F_t(\mathbf{h})|-|F_c(\mathbf{h})||^2\rangle^{1/2}$$
$$R = \sum ||F_t(\mathbf{h})|-|F_c(\mathbf{h})|| / \sum |F_t(\mathbf{h})| .$$

The summations are carried out separately for the *P*- and *M*-sets.

Table 1 shows the results of two runs on the two-atom structure. This had $a=5{\cdot}0$, $b=6{\cdot}0$, $c=7{\cdot}0$ Å. The *P*-set consisted of 22 independent reflexions and the *M*-set of a further 124. Unit weights $w_{\parallel}(\mathbf{h})=w_{\perp}(\mathbf{h})=w_M(\mathbf{h})=1$ were used throughout. In run 1 the value of β is just sufficient to satisfy the necessary condition for non-negativity but not sufficient to ensure a small truncation error. The final value of $\sigma_P(|F|)=0{\cdot}2351$ compares poorly with $\langle|F_o|^2\rangle_P^{1/2}\approx 0{\cdot}8$, and for the *M*-set $\sigma=0{\cdot}1713$ compares poorly with $\langle|F_o|^2\rangle_M^{1/2}\approx 0{\cdot}4$. For run 2 each β value was chosen so that $\exp(-4\beta|\mathbf{h}|^2)\leq 0{\cdot}01$ for the maximum $|\mathbf{h}|$ of the *C*-set. The value of $\sigma_P(|F|)=0{\cdot}0117$ compares well with $\langle|F_o|^2\rangle_P^{1/2}\approx 0{\cdot}7$ and $\sigma_M(|F|)=0{\cdot}0118$ also compares well with $\langle|F_o|^2\rangle_M^{1/2}\approx 0{\cdot}1$. These two runs demonstrate the necessity of ensuring a small truncation error.

Table 2 shows the results of two runs with the six-atom structure. This had $a=6{\cdot}30$, $b=7{\cdot}56$, $c=8{\cdot}82$ Å. The *P*-set consisted of 42 independent reflexions and the *M*-set (to 1·2 Å spacing) of a further 120. In run 3 unit weights were used throughout, as for the two-atom structure. In run 4 unit weights were used for the *P*-set but for the *M*-set weights included an allowance for truncation (given by the variance of F_o where F_c became zero) and also the Bayesian contribution to the variance, with $\sigma_A(\mathbf{h})$ estimated by $\frac{1}{2}\langle|F_c(\mathbf{h})-F_t(\mathbf{h})|\rangle_M^2$. Run 3 was repeated with unit weights and $\beta=4{\cdot}6$ throughout. This reduced $\langle|\Delta\varphi|\rangle_M$ to 26°, which still compares poorly with the value of 11·9° for run 4. Thus these runs show clearly the need for proper weighting to cause the refinement to reach a true rather than a false minimum. For the final cycle of run 4, $R_P=0{\cdot}0082$ and $R_M=0{\cdot}075$, indicating a very satisfactory quality of fit. It is also noticeable that the results in run 4 improve considerably when d_M becomes $\leq 1{\cdot}4$ Å.

Table 3 shows the results of two runs with the 16-atom structure. This had $a=9{\cdot}08$, $b=10{\cdot}90$, $c=12{\cdot}72$ Å. The *P*-set consisted of 116 independent reflexions and the *M*-set (to a spacing of 1·4 Å) a further 184. In run 5 the weights took account of truncation and the Bayesian term, and the results show the same features as those of run 4. A repeat of this run (not tabulated) was made in which d_M was reduced to 1·4 and d_C to 1·6 Å in one step after cycle 7. The same minimum was reached in 28 instead of 48 cycles. Hence it is not necessary to add the *M*-set in several stages, and may be inefficient. For run 6, the data were altered by the addition of artificially generated errors. The $|F_o(\mathbf{h})|$ were normally distributed about $|F_t(\mathbf{h})|$ with a standard deviation of 5% of $|F_t(\mathbf{h})|$ for both *P*- and *M*-sets. The $\varphi_o(\mathbf{h})$ for the *P*-set were normally distributed about the $\varphi_t(\mathbf{h})$ with a standard deviation of 15°. It can be seen from the table that the quality of the results was not spoilt by the introduction of these errors. In fact, for R_M there was an improvement as a result of some experiments carried out to improve the estimation of truncation error. For the $F_o(\mathbf{h})$ data without errors, the value of $\ln\langle|F_o(\mathbf{h})-F_c(\mathbf{h})|\rangle^2$ was found for each of a series of annular regions in the *P*-set. Various functions $(a+b|\mathbf{h}|,\ a+b|\mathbf{h}|+c|\mathbf{h}|^2,\ a+b|\mathbf{h}|^2,\ a+b|\mathbf{h}|+c|\mathbf{h}|^2+d|\mathbf{h}|^3,$

Table 1. *Results of two runs with the two-atom structure*

Run	j	β	d_M	d_C	$\sigma_P(\|F\|)$	$\langle\|\Delta\varphi\|\rangle_P$	$\sigma_M(\|F\|)$	$\langle\|\Delta\varphi\|\rangle_M$
1	5	1·57	1·0	2·0	0·3271	41·6	0·2761	46·5
1	12	1·57	1·0	1·6	0·2995	38·4	0·2205	52·3
1	17	1·57	1·0	1·28	0·2351	32·3	0·1713	47·3
2	7	4·71	1·0	2·0	0·0191	0·0	0·0142	24·4
2	14	3·93	1·0	1·6	0·0094	18·7	0·0118	13·4
2	18	3·24	1·0	1·28	0·0117	0·0	0·0118	11·8

Table 2. *Results of two runs with the six-atom structure*

Run	j	β	d_M	d_C	$\sigma_P(F)$	$\langle\|\Delta\varphi\|\rangle_P$	$\sigma_M(F)$	$\langle\|\Delta\varphi\|\rangle_M$
3	9	4·6	1·2	2·0	0·5273	13·9	0·4964	35·5
3	17	2·83	1·2	1·58	1·0176	13·8	1·0385	30·0
4	6	4·6	2·0	2·0	0·2303	0·9	–	–
4	13	4·6	1·8	2·0	0·2255	1·0	0·3336	20·3
4	28	4·6	1·6	1·8	0·1788	0·9	0·4376	24·7
4	39	4·6	1·4	1·6	0·1480	1·0	0·2607	10·7
4	54	4·6	1·2	1·4	0·0606	0·3	0·1519	11·9

Table 3. *Results of two runs with the* 16*-atom structure*

Run	j	β	d_M	d_C	$\sigma_P(F)$	$\langle\|\Delta\varphi\|\rangle_P$	R_P	$\sigma_M(F)$	$\langle\|\Delta\varphi\|\rangle_M$	R_M
5	7	4·0	2·0	2·0	0·3181	2·3	0·050	–	–	–
5	20	4·0	1·8	2·0	0·3415	2·4	0·052	0·7171	42·5	0·200
5	32	4·0	1·6	1·8	0·2113	3·6	0·033	0·5524	25·9	0·164
5	48	4·0	1·4	1·6	0·1145	1·4	0·019	0·2426	13·0	0·090
6	6	4·0	2·0	2·0	1·0470	4·4	0·055	–	–	–
6	22	4·0	1·4	1·6	0·9491	5·0	0·055	0·264	16·4	0·047

and a cubic spline for which the derivative was fitted to the Wilson plot at the limit of F_c) were fitted to the logarithm. Run 5 was repeated with truncation variances estimated by each of these fitted curves in turn. The most successful run was that in which $a+b|\mathbf{h}|^2$ was used. This method was therefore adopted for run 6.

The quantity $\frac{1}{2}\langle|F_c(\mathbf{h})-F_t(\mathbf{h})|\rangle_M^2$ used to estimate $\sigma_A^2(\mathbf{h})$ for run 6 would not be available in practice. Hence another run was carried out, with $4\sigma_M^2(|F|)$ instead. The factor 4 allowed for anisotropy in the probability distribution of F in the M-set. After 19 cycles, this run was giving results essentially the same as those for run 6, but with $\sigma_M^2(|F|)=0{\cdot}0099$ instead of 0·0079. We conclude that a good estimate of $\sigma_A^2(\mathbf{h})$ is desirable, but that the success of the process is not critically dependent on its exactness.

We show in Fig. 1 a composite of electron density sections computed from the $F_o(\mathbf{h})$ of the P-set of the 16-atom structure. Fig. 2 shows the same composite computed from the output of run 6, using as terms $F_c(\mathbf{h})$. exp $(+2{\cdot}0|\mathbf{h}|^2)$. This sharpening of the final $F_c(\mathbf{h})$ does not appear to introduce a significant finite series error in addition to the truncation error which we can estimate. The excursions from zero in the background of the map do not exceed about 1·0 Å^{-3} while the e.s.d. is 0·5 e Å^{-3}.

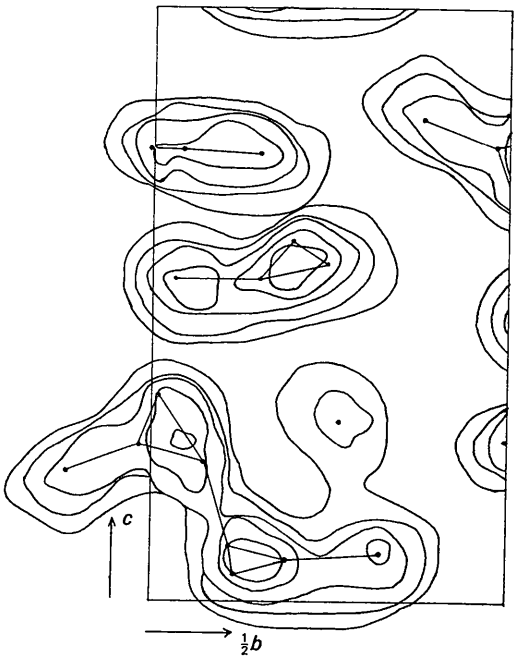

Fig. 1. Composite of electron density sections computed for the 16-atom test structure. The terms used are the F_o for the P-set only. In most cases individual atoms are not resolved by these data, which have spacings not less than 2·0 Å. The contour interval is 1 e Å^{-3}.

5. Conclusions

There are many variations of conditions which can be explored, all of which would affect in some degree the success of this method. So far we have established:

(*a*) It is necessary to ensure that the truncation error is small, and this can be done by ensuring reduction of the E values by a Gaussian factor which falls to 0·01 where the $F_c(\mathbf{h})$ terminate.

(*b*) Correct weights are needed to avoid false minima. If the main sources of error are allowed for, small variations in the method of weight calculation are not critical.

(*c*) The block-diagonal approximation to the normal equations described in § 2 gives reasonably fast convergence and avoids the need for large volumes of stored information. The algorithm is not impossibly expensive, and it is convenient to operate.

(*d*) The estimate of error for the electron density which we give provides a correct indication of the quality of the final map.

We conclude that this method offers sufficient promise of success to be applied to the improvement of resolution for structures with large numbers of atoms, such as proteins. Work to extend the applicability of the program to such structures is in hand.

Much of the analysis which is discussed in this paper is also applicable, with suitable modifications, to the method described by Sayre (1972, 1974).

It is quite probable that the method we describe can be used in other contexts than that of enhancing resolution for protein structures. It may be of value in

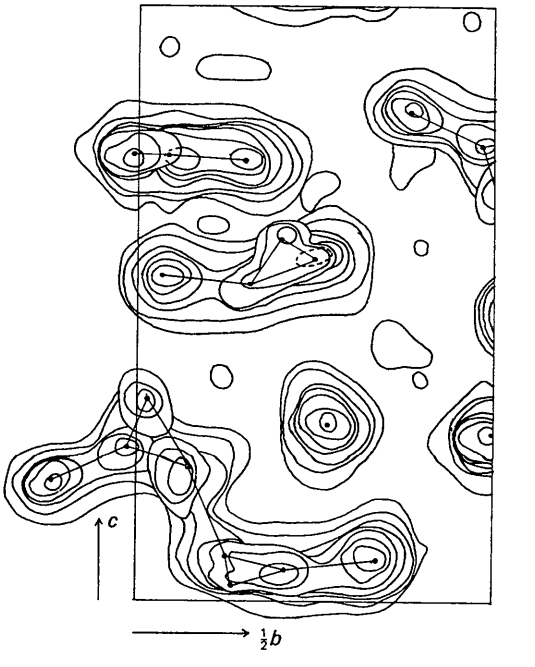

Fig. 2. Composite of electron density sections computed for the 16-atom test structure. The terms used are the F_c corresponding to the P-set and the M-set, multiplied by exp $(+2{\cdot}0|\mathbf{h}|^2)$. All individual atoms are resolved. The contour interval is 1 e Å^{-3}.

extending the sets of phases obtained by the application of direct methods to smaller structures. It may also be possible to enhance the resolution of Patterson series by this method more effectively than by simple sharpening of the data.

We wish to thank Carmarthenshire County Council for a Thomas and Elizabeth Williams Scholarship, and the Science Research Council for an Atlas Computer Laboratory, Pembroke College Oxford, Joint Research Fellowship, which have made it possible for one of us (A.R.D.) to take part in this work. We also wish to thank the Oxford University Computing Service and the Atlas Computer Laboratory for computing facilities. We have benefited from discussions with Dr D. Sayre and wish to thank him for making available to us a program for generating artificial structures.

APPENDIX

Notation

Some quantities are not used except near the point in the text at which they are defined. They have been omitted here.

$\|\mathbf{h}\|$	Radius in reciprocal space ($=2\sin\theta/\lambda$).
$E(\mathbf{h})$	Normalized structure factor.
$F_o(\mathbf{h}), \varphi_o(\mathbf{h})$	The observed structure factor and its phase.
$F_c(\mathbf{h}), \varphi_c(\mathbf{h})$	The calculated structure factor and its phase [calculated by self-convolution of the $C(\mathbf{h})$].
$F_t(\mathbf{h}), \varphi_t(\mathbf{h})$	The true (exact) structure factor and its phase.
$F_b(\mathbf{h})$	The centroid (in the complex plane) of the probability distribution $p[F_t(\mathbf{h})\|F_c(\mathbf{h})]$.
P	The set of reflexions for which both modulus and phase of F_o are given.
M	The set of reflexions for which only the modulus of F_o is given.
C	The set of reflexions for which $C(\mathbf{h}')\neq 0$.
$C(\mathbf{h}')$	The (complex) Fourier component of order $\mathbf{h}'$ in the Fourier series for the square root of the electron density.
$\alpha(\mathbf{h}')$	The phase of $C(\mathbf{h}')$.
$\Delta_{\parallel}(\mathbf{h}), \Delta_{\perp}(\mathbf{h})$	The deviations of $F_c(\mathbf{h})$ from $F_o(\mathbf{h})$ in the directions parallel and perpendicular to $F_o(\mathbf{h})$, for $\mathbf{h}$ in P.
$\Delta_M(\mathbf{h})$	The deviation of $\|F_c(\mathbf{h})\|$ from $\|F_b(\mathbf{h})\|$, for $\mathbf{h}$ in M.
$w_{\parallel}(\mathbf{h}), w_{\perp}(\mathbf{h}), w_M(\mathbf{h})$	Weights for $\Delta_{\parallel}(\mathbf{h}), \Delta_{\perp}(\mathbf{h}), \Delta_M(\mathbf{h})$.
β	The parameter in $\exp(-\beta\|\mathbf{h}\|^2)$, by which the E values are multiplied to generate F_o.
$p(X)$	The probability density function for the random variable X.
$p(X\|Y)$	The probability density function for X, given the value of the random variable Y.
$\sigma_A^2(\mathbf{h})$	The estimated variance (assumed isotropic) of the distribution $p[F_c(\mathbf{h})\|F_t(\mathbf{h})]$.
$\sigma_b^2(\mathbf{h})$	The variance in the radial direction of the distribution $p[F_t(\mathbf{h})\|F_c(\mathbf{h})]$.
$\langle\|\Delta\varphi\|\rangle_{P,M}$	The mean value of $\|\varphi_t(\mathbf{h})-\varphi_c(\mathbf{h})\|$ over P, M.
$\sigma_{P,M}(F)$	The r.m.s. value of $\|F_t(\mathbf{h})-F_c(\mathbf{h})\|$ over P, M.
$\sigma_{P,M}(\|F\|)$	The r.m.s. value of $\|\|F_t(\mathbf{h})\|-\|F_c(\mathbf{h})\|\|$ over P, M.
$R_{P,M}$	The conventional R index evaluated over P, M.

References

BARD, Y. (1970). *S.I.A.M. J. Num. Anal.* **7**, 157–186.
BIRAUD, Y. (1969). *Astr. Astrophys.* **1**, 124–127.
BOAS, R. P. & KAC, M. (1945). *Duke Math. J.* **12**, 189–206.
DAVIES, A. R. (1974). D. Phil. Thesis. Univ. of Oxford.
LINDLEY, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint.* Cambridge Univ. Press.
ROLLETT, J. S. (1965). *Computing Methods in Crystallography*, Edited by J. S. ROLLETT, pp. 32–35. Oxford: Pergamon Press.
SAYRE, D. (1952). *Acta Cryst.* **5**, 60–65.
SAYRE, D. (1972). *Acta Cryst.* A**28**, 210–212.
SAYRE, D. (1974). *Acta Cryst.* A**30**, 180–184.