Acta Cryst. (1976). A**32**, 274

Least-Squares Refinement of Phases in Direct and Reciprocal Spaces

By J. Gassmann

Max-Planck-Institut für Biochemie, D-8033 Martinsried, Germany (BRD)

(*Received* 23 *June* 1975; *accepted* 7 *October* 1975)

A comparison between least-squares phase refinement as formulated in direct and reciprocal spaces is given. After analysing some common features of different iterative phase determination procedures, these procedures are seen with respect to the least-squares phase refinement techniques. Some results with a myoglobin model structure using an iterative technique are presented.

1. Introduction

Using the concept of conditional joint probabilities a method for phase determination, the 'maximum determinant rule' has been developed (Tsoucaris, 1970). This method was devised to determine simultaneously a number of phases, assuming that the most probable phase set corresponds to the correct solution. The method has been applied to different structures for phase determination (de Rango, Mauguen & Tsoucaris, 1975).

Based on the density squaring principle (Sayre, 1952) a least-squares procedure for phase refinement has been developed (Sayre, 1972; Allegea & Colombo, 1974). The value to be minimized was

$$R=\sum_h |\mathbf{F}_h-\mathbf{F}_h^q|^2=\text{minimum} \qquad (1)$$

$$R'=\sum_h \left|\frac{1}{s_h}(\mathbf{F}_h-\mathbf{F}_h^q)\right|^2=\text{minimum} \qquad (1a)$$

respectively where $\mathbf{F}_h^q$ is the structure factor calculated from

$$\mathbf{F}_h^q=s_h \sum_{h'} \mathbf{F}_{h'}\mathbf{F}_{h-h'}$$

and s_h is to account for the change in atomic form factor. This method has been applied to phase determination in a real structure (Sayre, 1974).

A related method had been given earlier (Hoppe, 1963) but was never applied because of computational labour. For actual structure determinations a stepwise minimization of equation (1) by incremental change of individual phases has also been given (Krabbendam & Kroon, 1971).

These methods in reciprocal space use the squaring principle to obtain phase indications:

$$\varrho^*=\varrho^2 \qquad (2)$$

(ϱ=initial density, ϱ^*=squared density from which phase information is deduced).

Some time ago a general modification scheme for obtaining phase information was proposed (Hoppe & Gassmann, 1968)

$$\varrho^*=f(\varrho)\,. \qquad (3)$$

In subsequent papers (Hoppe, Gassmann & Zechmeister, 1970; Gassmann & Zechmeister, 1972), different forms of density modifications $f(\varrho)$ have been shown to be successful in application to real structure determinations. For analytical forms of local density modification equation (3) can be written as

$$\varrho^*=g(\varrho)\,.\,\varrho \qquad (3a)$$

where $g(\varrho)$ is the local weighting function applied to the density ϱ. Obviously in structure determination the density ϱ is known only approximately and should be more accurately replaced by the approximate density ϱ_p. Equation (3) then reads:

$$\varrho^*=f(\varrho_p)=g(\varrho_p)\,.\,\varrho_p\,. \qquad (3b)$$

This density ϱ_p may be written as a Fourier series over a restricted set of reflexion indices $\mathbf{h}'$:

$$\varrho_p(\mathbf{r})\simeq\sum_{\mathbf{h}'} \mathbf{F}_{\mathbf{h}'} \exp 2\pi i\mathbf{h}'\mathbf{r} \qquad (4)$$

depending on the structure factors $\mathbf{F}_{\mathbf{h}'}$ contributing to the density ϱ_p.

In the following, for the sake of generality, the special case of squaring [equation (2)] will be replaced by the general modification [equation (3)].

2. Least-squares phase refinement in direct space formulation

If one considers the final density ϱ against a modified approximate density ϱ^*, it is possible to define a reliability index

$$R_1(\varphi)=\int(\varrho-\varrho^*)^2\mathrm{d}V=\text{minimum}\,. \qquad (5)$$

This R index should be minimized through choice of phases φ, assuming known observed structure factor amplitudes F_h. The different contributions to R_1 are:

$$R_1(\varphi)=\int\varrho^2\mathrm{d}V+\int\varrho^{*2}\mathrm{d}V-2\int\varrho\varrho^*\mathrm{d}V\,. \qquad (5a)$$

The first two terms are independent of phase:[†]

$$\int \varrho^2 \mathrm{d}V \sim \sum_{h} |F_h|^2,$$

and

$$\int \varrho^{*2} \mathrm{d}V \sim \sum_{h''} |F_{h''}|^2 .$$

The summation in the second expression runs only through the index set h'' resulting from the approximate density index set h' [equation (4)]. From equation (5) we get

$$R_2(\varphi) = \int \varrho\varrho^* \mathrm{d}V = \text{maximum} . \tag{6}$$

This maximum should be reached through choice of phases.

If one inserts the Fourier series for ϱ and ϱ^* equation (6) reads

$$R_2(\varphi) = \sum_{h_1,h_2}^{h_1+h_2=0} \mathbf{F}_{h_1}\mathbf{F}^*_{h_2} = \sum_{h_2} \mathbf{F}_{-h_2}\mathbf{F}^*_{h_2} . \tag{6a}$$

$\mathbf{F}^*_h$ is the Fourier coefficient for the modified density. The summation index h_2 runs over the index set pertaining to the modified approximate density ϱ^*. Inserting equation (3a) and taking the Fourier series for the values ϱ, ϱ_p and $g(\varrho_p)$ results in:

$$\begin{aligned} R_2(\varphi) &= \int \varrho \,.\, g(\varrho_p) \,.\, \varrho_p \mathrm{d}V \\ &= \sum_{h_1,h_2,h_3}^{h_1+h_2+h_3=0} \mathbf{F}_{h_1}\mathbf{F}_{h_2}\mathbf{G}_{h_3} = \text{maximum} \end{aligned} \tag{7}$$

where $\mathbf{G}_h$ is the Fourier coefficient of the weighting function $g(\varrho_p)$. Because of the condition $\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=0$ this may be written as:

$$\begin{aligned} R_2(\varphi) &= \sum_{h_1,h_2} \mathbf{G}_{h_2-h_1}\mathbf{F}_{h_1}\mathbf{F}_{-h_2} \\ &= \sum_{h_2,h_3} \mathbf{G}_{h_3}\mathbf{F}_{h_2-h_3}\mathbf{F}_{-h_2} = \text{maximum} , \end{aligned} \tag{7a}$$

$h_1 \mathrel{\hat{=}}$ total index set (ϱ)

$h_2 \mathrel{\hat{=}}$ restricted index set (ϱ_p)

$h_3 \mathrel{\hat{=}}$ restricted index set [$g(\varrho_p)$] .

Equation (6) has been given earlier[‡] for the special case of $g(\varrho_p)=\varrho_p$ (Cochran, 1952):

$$R_2(\varphi) = \int \varrho \,.\, \varrho_p^2 \mathrm{d}V = \text{maximum} . \tag{7b}$$

Using the weaker condition

$$R_2(\varphi) = \int \varrho \,.\, \varrho_p^2 \mathrm{d}V \geq 0$$

and transforming it into a form equivalent to equation (7a), inequalities among structure factors can be derived (Karle & Hauptman, 1950):

$$R_2(\varphi) = \int \varrho \,.\, \varrho_p^2 \mathrm{d}V = \sum_{h_1,h_2} \mathbf{F}_{h_1}\mathbf{F}_{-h_2}\mathbf{F}_{h_2-h_1} \geq 0 . \tag{7c}$$

For the original derivation of the inequalities it has only been assumed that

$$\varrho^* = f(\varrho_p) = g(\varrho_p) \,.\, \varrho_p \geq 0 \tag{8}$$

for all values of $\varrho_p(\mathbf{r})$, which is certainly true for

$$\varrho^* = f(\varrho_p) = \varrho_p^2 . \tag{8a}$$

The much stronger relation (6), when transformed into reciprocal space, for condition (8a) results in:

$$R_2(\varphi) = \sum_{h_1,h_2} \mathbf{F}_{h_2-h_1}\mathbf{F}_{h_1}\mathbf{F}_{-h_2} = \text{maximum} . \tag{9}$$

The assumption of (8) means a possible separation of $\varrho^* = \sqrt{\varrho^*} \,.\, \sqrt{\varrho^*}$, where $\sqrt{\varrho^*}$ is a real-valued function of space, with a Fourier transform

$$\mathbf{V}_h = \mathscr{F}(\sqrt{\varrho^*}) \tag{10}$$

where the coefficients $\mathbf{V}_h$ obey Friedel's relation. Equation (6) may then be written as:

$$\begin{aligned} R_2(\varphi) = \text{maximum} &= \sum_{h_1,h_2,h_3}^{h_1+h_2+h_3=0} \mathbf{F}_{h_1}\mathbf{V}_{h_2}\mathbf{V}_{h_3} \\ &= \sum_{h_2,h_3} \mathbf{F}_{h_2-h_3}\mathbf{V}_{-h_2}\mathbf{V}_{h_3} , \end{aligned} \tag{11}$$

where the summations h_2 and h_3 run over a restricted index set corresponding to equation (10).

If one considers (11) as a Hermitian form of the variables $\mathbf{V}_h$ the pertaining mathematical theorems can be applied. This is especially useful for the decomposition in Gram determinants or, as they are called in crystallography, Karle–Hauptman determinants. Replacement of indices and application of the Jacobian form results in: $i \mathrel{\hat{=}} -\mathbf{h}_2$; $j \mathrel{\hat{=}} \mathbf{h}_3$; $ij \mathrel{\hat{=}} \mathbf{h}_2-\mathbf{h}_3$

$$R_2(\varphi) = \sum_{i,j} F_{ij}V_iV_j = \sum_{k=1}^{N} |y_k|^2/c_{kk} = \text{maximum}$$

[†] This is only correct if the modified density ϱ^* is formed by using the phases from equation (3) together with the observed structure factor amplitudes. Otherwise equation (6) would read $R_2(\varphi)=2\int\varrho\varrho^*\mathrm{d}V-\int\varrho^{*2}\mathrm{d}V=\text{maximum}$.

A similar expression could also be constructed as:

$$\lambda(\varphi) = \int \varrho\varrho^*\mathrm{d}V / \int \varrho^{*2}\mathrm{d}V = \text{maximum} .$$

Such a value $\lambda(\varphi)$ would be related to the 'Raleigh quotient' for the approximate determination of eigenvalues of a matrix. In our case this value $\lambda(\varphi)$ is equivalent to the correlation coefficient

$$r(\varphi) = \int \varrho\varrho^*\mathrm{d}V / (\int \varrho^2\mathrm{d}V \,.\, \int \varrho^{*2}\mathrm{d}V)^{1/2}$$

because $\int\varrho^2\mathrm{d}V$ is independent of the phases.

[‡] It is interesting to note, that an equivalent derivation of equations (5)–(7) has been used in another phase determination method assuming quite different conditions (Rossmann & Blow, 1963), where the density improvement or 'modification process' is the averaging of density related by non-crystallographic symmetry (Colman, 1974; Bricogne, 1974). Because of this analogous behaviour, results of the present paragraph might be transferable to the molecular replacement method or *vice versa.*

where

$$c_{kk}=\frac{D_k}{D_{k-1}}\,; \quad D_k=\begin{vmatrix} F_{11}\ldots F_{1k} \\ \cdot \quad\quad \cdot \\ \cdot \quad\quad \cdot \\ \cdot \quad\quad \cdot \\ F_{k1}\ldots F_{kk} \end{vmatrix}; \quad D_0=1$$

and

$$y_k=\sum_{l=k}^{\max} c_{kl}V_l$$

with

$$c_{kl}=\frac{1}{D_{k-1}}\begin{vmatrix} F_{11}\ldots F_{1,k-1}F_{1l} \\ \cdot \quad\quad \cdot \quad\quad \cdot \\ \cdot \quad\quad \cdot \quad\quad \cdot \\ \cdot \quad\quad \cdot \quad\quad \cdot \\ F_{k1}\ldots F_{k,k-1}F_{kl} \end{vmatrix}.$$

The summation value N is the rank of the matrix $\|F_{ij}\|$. This may be rewritten as:

$$R_2(\varphi)=\sum_{k=1}^{N} c_{kk}|V_k|^2 + \sum_{k=1}^{N}\sum_{l_1\geq k}^{\max}\sum_{l_2>l_1}^{\max} V_{l_1}V_{-l_2}\,\frac{c_{kl_1}c^{+}_{kl_2}}{c_{kk}} = \text{maximum}\,. \quad (12)$$

The free choice of $\sqrt{\varrho^*}$ (apart from $\varrho^*\geq 0$), *i.e.* the arbitrariness of $\mathbf{V}_h$ (apart from Friedel's relation), implies maximization of the coefficients in (12):

$$c_{kk}=D_k/D_{k-1}=\text{maximum} \quad (13)$$

which is essentially the 'maximum determinant rule' (Tsoucaris, 1970; Lajzérowicz & Lajzérowicz, 1966).

Instead of separating the unspecified modification ϱ^* symmetrically $\varrho^*=\sqrt{\varrho^{*2}}$ it is possible to partition as in equation (8):

$$\varrho^*=g(\varrho)\cdot\varrho$$

which results in equation (7*a*).

Decomposition of (7*a*) into Gram determinants with the Jacobian form gives: $i \triangleq -\mathbf{h}_2$; $j \triangleq \mathbf{h}_1$; $ij \triangleq \mathbf{h}_2-\mathbf{h}_1$

$$R_2(\varphi)\cdot\sum_{h_1,h_2}\mathbf{G}_{h_2-h_1}\mathbf{F}_{h_1}\mathbf{F}_{-h_2} = \sum_{i,j} G_{ij}F_iF_j=\sum_{k=1}^{N}\frac{|X_k|^2}{D_kD_{k-1}}=\text{maximum} \quad (14)$$

where

$$D_k=\begin{vmatrix} G_{11}\ldots G_{1k} \\ \vdots \quad\quad \vdots \\ G_{k1}\ldots G_{kk} \end{vmatrix}; \qquad D_0=1$$

and

$$X_k=\begin{vmatrix} G_{11}\ldots G_{1\ k-1}H_1 \\ \vdots \quad\quad \vdots \quad\quad \vdots \\ G_{k1}\ldots G_{k\ k-1}H_k \end{vmatrix}; \quad X_1=H_1$$

and

$$H_l=\sum_m G_{lm}\,,\,F_m\,. \quad (14a)$$

The values H_l written with crystallographic indices $\mathbf{H}_{-h}=\sum_{h'}\mathbf{G}_{-h-h'}\mathbf{F}_{h'}$ are precisely the Fourier coefficients of the modified electron density ϱ^*. The summation value N is the rank of the coefficient matrix $\mathbf{G}_{ij}$.

The values X_k, which are linear forms in $\mathbf{F}_h$, are linearly independent. This indicates that a maximum for $R_2(\varphi)$ in equation (14) can be reached by maximizing every $|X_k|^2/D_kD_{k-1}$ through choice of phases. This results in:

$$\frac{|X_k|^2}{D_kD_{k-1}}=\text{maximum}\,. \quad (15)$$

For (15) to be true through choice of phases, it is necessary that equation (14) is a positive definite Hermitian form, which places conditions on the matrix $\mathbf{G}_{ij}$ and therefore requires positivity of the weighting function $g(\varrho_p)$:

$$D_k>0; \quad k=1,2\ldots N\,. \quad (16)$$

Appropriate choice of the independent summation indices i,j in (14) results in real values for the diagonal elements. The results of (15) for $k=1$, 2 and 3 are then:

(*a*) $k=1$

$$D_0=1; \quad D_1=G_{11}; \quad X_1=H_1=\sum_m G_{1m}F_m\,.$$

To maximize $|X_1|^2/G_{11}$ through choice of the phases results in:

$$X_1=\text{maximum};$$

$$\varphi(G_{1m}F_m)=\text{const.}=\varphi(H_1)=\varphi(F_1) \text{ because } G_{11}=\text{real}\,.$$

In more familiar crystallographic terms this reads:

$$\varphi(\mathbf{H}_h)=\varphi(\mathbf{F}_h)=\varphi(\mathbf{G}_{h-h'}\mathbf{F}_{h'}) = \text{const. (for all values of } h')\,.$$

For the special case $\mathbf{G}_h=\mathbf{F}_h$ this leads to the normal triple product relation:

$$\varphi(\mathbf{F}_h)=\varphi(\mathbf{F}_{h-h'}\mathbf{F}_{h'})\,.$$

(*b*) $k=2$

$$D_1=G_{11}; \quad D_2=\begin{vmatrix} G_{11}G_{12} \\ G_{21}G_{22} \end{vmatrix};$$

$$X_2=\begin{vmatrix} G_{11}H_1 \\ G_{21}H_2 \end{vmatrix}=G_{11}H_2-G_{21}H_1\,.$$

To maximize $|X_2|^2/D_1D_2$ through choice of phases results in (D_2 is independent of phases):

$$X_2=\begin{vmatrix} G_{11}H_1 \\ G_{21}H_2 \end{vmatrix}=\text{maximum};$$

$$\varphi(H_2)=\varphi(F_2)=\varphi(G_{2m}F_m) = \text{const. (for all values } m\text{) as for the case } k=1;$$

and

$$\varphi(G_{21}H_1)=\text{const.}=\varphi(H_2)+\pi\,.$$

The combination of both conditions results in:

$$\varphi(F_2)=\varphi(G_{2n}F_n)=\varphi(G_{21}G_{1m}F_m)+\pi\,.$$

For the special case of squaring ($\mathbf{G}_h=\mathbf{F}_h$) this is:

$$\varphi(F_2)=\varphi(F_{2n}F_n)=\varphi(F_{21}F_{1m}F_m)+\pi;$$

and this means that maximization of $|X_2|^2/D_1D_2$ leads to triplet and quartet phase indications, where the quartet term has the opposite sign of the triplet term (Hauptman, 1974; Gassmann, 1975).

(*c*) $k=3$

$$\begin{aligned}D_2&=\begin{vmatrix}G_{11}G_{12}\\G_{21}G_{22}\end{vmatrix};\quad D_3=\begin{vmatrix}G_{11}G_{12}G_{13}\\G_{21}G_{22}G_{23}\\G_{31}G_{32}G_{33}\end{vmatrix}\\&=G_{11}G_{22}G_{33}+G_{21}G_{13}G_{32}+G_{31}G_{12}G_{23}\\&\quad-(G_{31}G_{13}G_{22}+G_{21}G_{12}G_{33}+G_{32}G_{23}G_{11})\,.\end{aligned}$$

If $g(\varrho_p)$ is chosen so that D_3 is a real number, *i.e.*

$$\varphi(G_{21}G_{13}G_{32})=\varphi(G_{31}G_{12}G_{23})=\begin{cases}0\\\pi\end{cases},$$

one obtains:

$$\begin{aligned}X_3=\text{maximum}&=\begin{vmatrix}G_{11}G_{12}H_1\\G_{21}G_{22}H_2\\G_{31}G_{32}H_3\end{vmatrix}\\&=(G_{32}G_{21}-G_{22}G_{31})H_1\\&\quad+(G_{31}G_{12}-G_{11}G_{32})H_2\\&\quad+(G_{11}G_{22}-G_{12}G_{21})H_3\,.\end{aligned}$$

Using equation (14*a*) results in:

$$\varphi(H_3)=\varphi(F_3)=\varphi(G_{32}G_{21}G_{1m}F_m)=\varphi(G_{31}G_{12}G_{2n}F_n)$$

and

$$\varphi(F_3)+\pi=\varphi(G_{22}G_{31}G_{1l}F_l)=\varphi(G_{11}G_{32}G_{2r}F_r)\,.$$

As in the case where $k=2$, there are phase indications from multiplets ($n=3$, $n+1=4$) which differ in phase by π.

3. Least-squares phase refinement in reciprocal space formulation

The equivalence of phase refinement in direct and reciprocal spaces is best shown by converting the expression for the reliability factor [equation (5)]:

$$R_1(\varphi)=\int(\varrho-\varrho^*)^2\mathrm{d}V\simeq\sum_h(\mathbf{F}_h-\mathbf{F}_h^*)\,(\mathbf{F}_{-h}-\mathbf{F}_{-h}^*)=\text{minimum}\qquad(4b)$$

where Parseval's formula has been applied. $\mathbf{F}_h^*$ here denotes the structure factor for the modified density ϱ^*. For the special case $\varrho^*=\varrho^2$, *i.e.* $\mathbf{F}_h^*=\mathbf{F}_h^q\sim\sum_{h'}\mathbf{F}_{h-h'}\mathbf{F}_{h'}$, this amounts to

$$R(\varphi)=\sum_{h'}|\mathbf{F}_h-\mathbf{F}_h^q|^2\qquad(4c)$$

which is essentially the minimization condition given by Sayre (1972).

Rewriting $R(\varphi)$ gives:

$$R(\varphi)=\sum_h|\mathbf{F}_h|^2+\sum_h|\mathbf{F}_h^*|^2-2\sum_h\mathbf{F}_h\mathbf{F}_{-h}^*=\text{minimum}\,.$$

Taking the same assumptions as for the derivation of (6) one obtains the same result:

$$R_2(\varphi)=\sum_{h'}\mathbf{F}_{h'}\mathbf{F}_{-h'}^*=\text{maximum}\qquad(6a)$$

where h' runs over the restricted index set resulting from the Fourier transformation of the approximate density $\varrho^*=f(\varrho_p)$.

Using a procedure equivalent to that used in least-squares refinement of atomic parameters it is possible to minimize $R(\varphi)$ (Sayre, 1972). However, the derivations above for phase refinement formulated in direct space indicate an alternative way to minimize the reliability index R. This can be achieved through maximizing R_2. This is most conveniently done by maximization of the coefficients c_{kk} ('maximum determinant rule') or by maximization of the components $|X_k|^2/D_kD_{k-1}$ belonging to the Jacobian form of equation (14). At the same time these derivations indicate the possible different forms of modifications, leaving the squaring principle and the resulting equation (4*c*) as a special case.

4. Phase determination by iterative procedures in comparison with least-squares procedures

The commonly used methods to determine phases so far have been iterative procedures without reference to the least-squares principle (*e.g.* tangent refinement, phase correction). The effect of such procedures is twofold since they contain the possibility firstly for an extension to determine new phases and secondly for a refinement to improve old phases. This twofold function of extension and refinement is, however, not fully present in all these iterative methods and moreover depends strongly on the available experimental data, *i.e.* on the resolution.

Since tangent refinement and phase correction (in its local weighting form) both use the atomic envelope principle of resolved atoms (Gassmann, 1976), these methods in a strict sense are only applicable at atomic resolution. At lower resolution the use of these methods is only valid if the original and the modified density show a qualitative similar structure (Hoppe, 1963). This restricts especially the range of phase extension, because only weak phase relationships exist outside this range (Cochran, 1955). Other methods like the molecular envelope principle or non-local phase correction may allow extension and refinement at lower resolution with less restrictions.

Another important feature of iterative phase determination is the inclusion of observed structure factor amplitudes. This is a decisive point for the convergence of iterative phase determination. These procedures are not a simple 'form of fixed-point iteration' (Sayre, 1972) and may at most be described as fixed-point iterations with an iteration-dependent, variable operator. The observed structure factor amplitudes restrict the iterated structure factors (as input to the next cycle) to lie on a circle with radius F_{obs} better than any least-squares process can achieve (Fig. 1). In iterative methods the radial part of equation (4*b*), *i.e.* the absolute value of the structure factor is always 'solved' exactly, by using the observed amplitude $F_{\rm obs}$ (see also Bricogne, 1974: p. 405). It is therefore probable that the approximate solution for the absolute value F_h by the least-squares process in normal matrix formulation affects the convergence somewhat adversely.

For tangent refinement and any other modification $\varrho^*=\varrho^n$, $n\geq 2$; as an extension process it is quite obvious that a pure fixed-point iteration could never converge to the correct result, because there is only convergence to one finite fixed point at $\varrho=0$ (Fig. 2).

For phase correction there is convergence to two finite fixed points at $\varrho=0$ and $\varrho=1$ which confines the density to the interval $0\leq\varrho\leq 1$. Also here, in addition to the critical density level T, inclusion of structure factor amplitudes plays the ultimate role for increasing or decreasing the density at a specific point in real space. The properties of refinement relative to the correct phases for these iterative methods have been treated earlier (Gassmann & Zechmeister, 1972).

The phase determination properties for least-squares procedures are dependent upon the formulation as a maximizing rule (Tsoucaris, 1970) or a normal matrix minimizing problem (Sayre, 1972). Whereas the former formulation allows principally the determination of new phases, *i.e.* phase extension, this is not possible for the latter one. The initial starting phases in this case have to be obtained by an iterative-type process. Therefore to demonstrate the effect of normal-matrix least-squares techniques one should show the improvement from this initial, extended phase set (which is the actual least-squares starting point) to the refined set. This has not been done in the past (Sayre, 1974).

According to equation (5) least-squares phase refinement can be considered as the minimization of the real-space average of a density modification with respect to the correct density. This averaging certainly implies stronger restrictions than purely iterative methods which depend only on local modification effects. The additional restrictions on iterative methods must be applied throughout the modification process itself (confining densities *etc.*) and through the use of observed amplitudes. Assuming these restrictions to be applied, it is hard to see why least-squares techniques should be superior to iterative methods.

With atomic resolution there are many examples to show the range of convergence for iterative, direct methods. Apart from the phase-extension problem in a least-squares procedure, it remains to be shown what range of convergence may be reached, using phase refinement with a least-squares process.

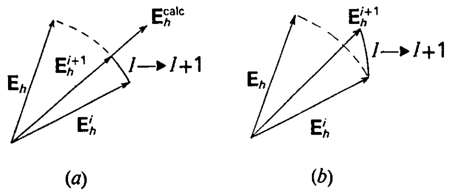

Fig. 1. (*a*) Iterative and (*b*) least-squares structure-factor improvement by one cycle of phase refinement. For iterative procedures the observed structure factor amplitude is used, whereas the normal least-squares process solves only approximately for the radial part of equation (4*b*).

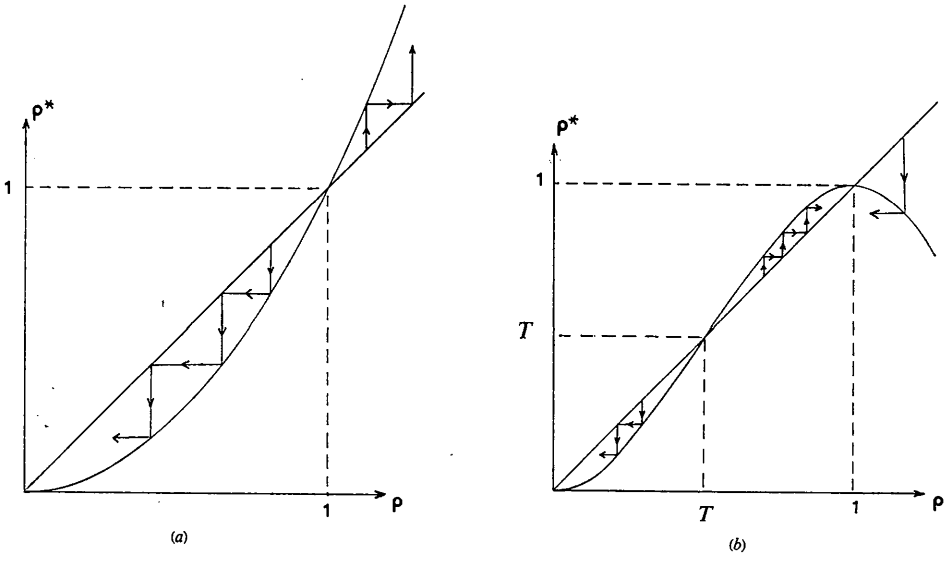

Fig. 2. (*a*) Convergence properties of tangent-refinement and (*b*) phase correction as if they were used as simple fixed-point iterations. Tangent refinement would converge to $\varrho=0$ whereas for phase correction convergence occurs to one of the fixed points $\varrho=0$ or $\varrho=1$ depending on the critical level T. The inclusion of observed amplitudes $F_h^{\rm obs}$ changes the modification drastically from cycle to cycle.

The different effects on phase determination of lower resolution and large molecules are more difficult to demonstrate for both iterative and least-squares methods.

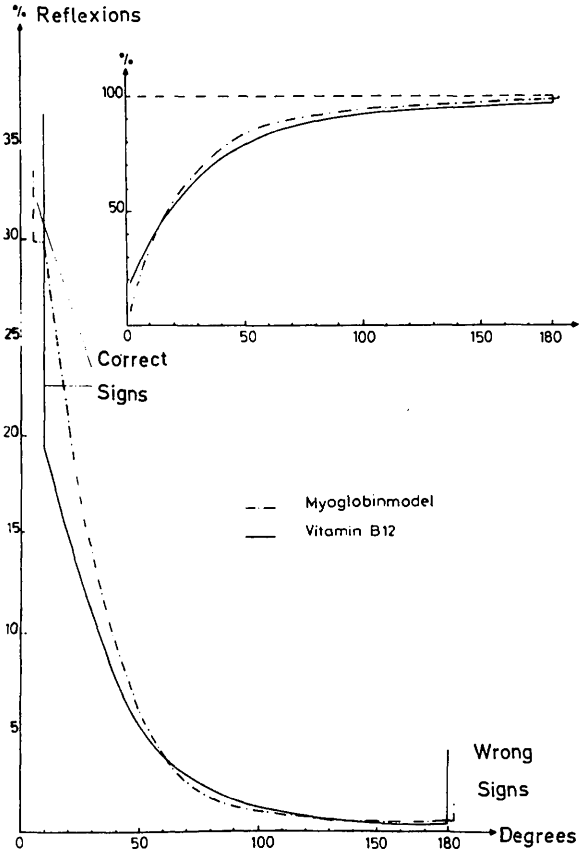

Fig. 3. Phase deviation of extended phases in the range between 2 and 1·4 Å resolution for the myoglobin model structure (dashed lines). The small diagram shows the proportion of phases with a deviation less or equal to the value on the abscissa. The comparison with the same calculations for vitamin $B12$ shows little effect of the molecular size.

5. Test calculations with a myoglobin model structure

Some time ago calculations at non-atomic resolution with iteration type phase determination were carried out (Gassmann, 1966) which demonstrated the following:

(*a*) a considerable part of the new structure information is already included through phase extension without any phase refinement.

(*b*) the accuracy of the determined phases depends on the size of the structure factor amplitudes.

(*c*) the accuracy of the determined phases does not depend on sin θ/λ for a physically sensible range.

Since at the time of these calculations no refined protein structure was available,* atomic coordinates from the Kendrew–Watson myoglobin model were used to calculate structure factors. To be confined to a sensible range of sin θ/λ, a resolution range between 2 and 1·4 Å was selected for phase extension.† The results are shown in Figs. 3–6.

(*a*) Starting with about 10 000 model phases to 2 Å resolution, about 7 000 additional phases to 1·4 Å resolution were determined in one iteration cycle. The phase deviations are shown in Fig. 3. The mean phase deviation is about 45°. Despite these deviations, the resulting Fourier map was quite similar to the Fourier map with calculated phases out to 1·4 Å resolution. This is shown for the example of the heme group in Fig. 4.

(*b*) The similarity of the densities with extended and calculated phases indicates that the main contributions, *i.e.* the large structure factors, have reasonably accurate phases. This is shown in Fig. 5 (for similar results see also Rango, Mauguen & Tsoucaris, 1975).

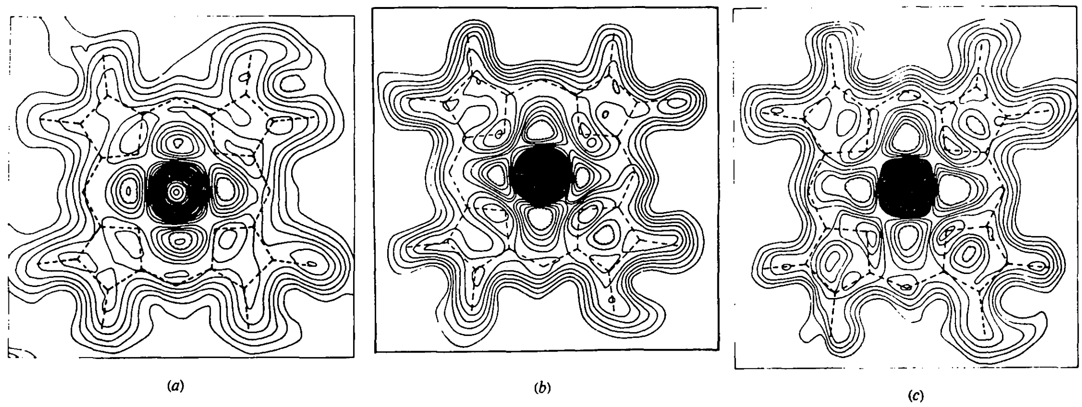

Fig. 4. Improvement of density by phase extension for the heme group of the myoglobin model structure. The phase information used was (*a*) calculated phases to 2 Å resolution, (*b*) calculated phases to 2 Å resolution, extended phases from 2 to 1·4 Å resolution, (*c*) calculated phases to 1·4 Å resolution.

---

* Recently, corresponding calculations have been done on the refined structure of rubredoxin (Collins, Brice, La Cour & Legg, 1976) which can be directly compared with the least-squares phasing results (Sayre, 1974).

† Similar calculations with experimental myoglobin data were done in parallel (Hoppe & Gassmann, 1964).

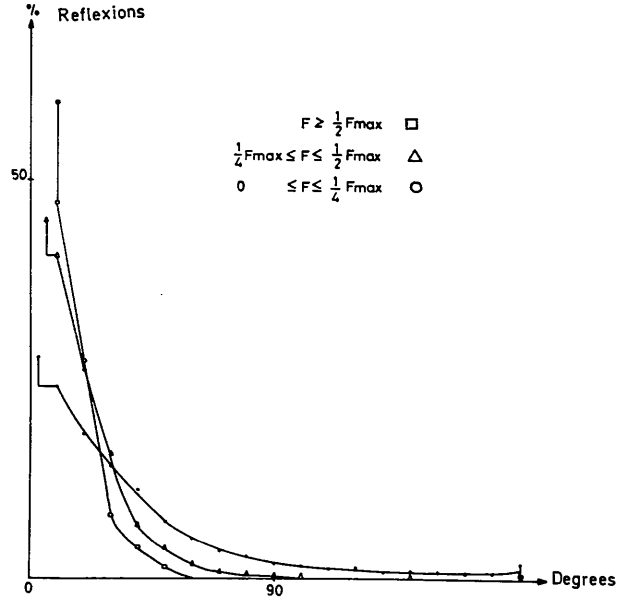

Fig. 5. Distribution of phase deviation in the myoglobin model structure for extension in the 2 to 1·4 Å range in relation to the structure factor amplitude. Large amplitudes have a distinctly smaller phase deviation, as expected from theory.

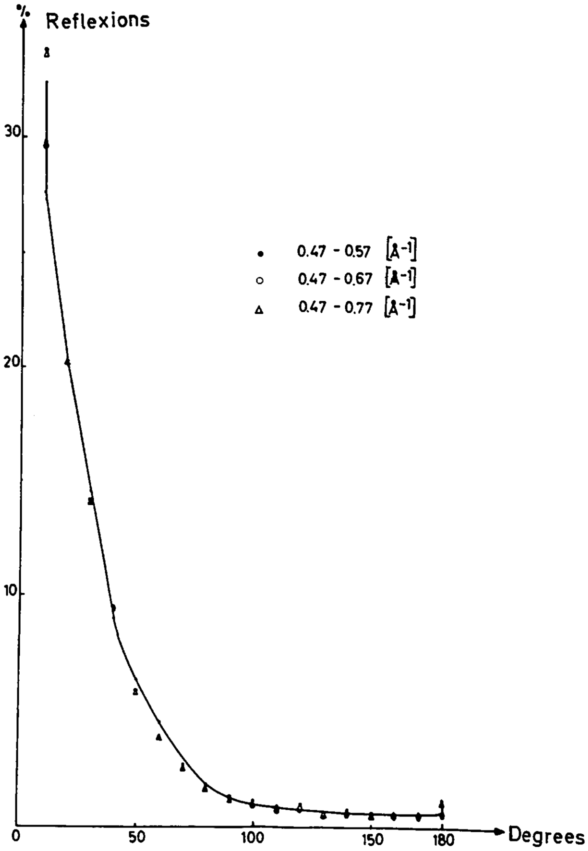

Fig. 6. Distribution of phase deviation in the myoglobin model structure for different extension ranges. There was no effect on the distribution when referred to reciprocal space, *i.e.* $h=2\sin\theta/\lambda$.

(*c*) Considering phase extension in reciprocal space as convolution $\mathbf{F}_h^* \simeq \sum_{h'} \mathbf{G}_{h-h'}\mathbf{F}_{h'}$, one could think of an improvement of the phase accuracy depending on the number of contributions in the convolution, *i.e.* with relation to the index h. This is not true for the considered range (Fig. 6). Also incremental phase extension in shells of resolution ranges did not improve accuracy of phases appreciably.

It would be feasible to refine these phases further. For an unknown structure the progress of refinement must be then judged by an R index, which is a very weak indication. Even for non-sensible refinement this number may show a normal decrease (Sayre, 1974). It is therefore considered to be more cautious at non-atomic resolution to use only one, or a few cycles of phase determination to gain at least the new information added through phase extension. The range of extension should be selected carefully in order to obtain phases of sufficient accuracy.

The author is grateful to Professor W. Hoppe for initiating the calculations on the myoglobin model and for his continuous support for this work.

References

Allegra, G. & Colombo, A. (1974). *Acta Cryst.* A**30**, 727–729.

Bricogne, G. (1974). *Acta Cryst.* A**30**, 395–405.

Cochran, W. (1952). *Acta Cryst.* **5**, 65–67.

Cochran, W. (1955). *Acta Cryst.* **8**, 473–478.

Collins, D. M., Brice, M. D., La Cour, T. F. M. & Legg, M. J. (1976). *Crystallographic Computing* II, edited by F. R. Ahmed. Copenhagen: Munksgaard. To be published.

Colman, P. M. (1974). *Z. Krist.* **140**, 344–349.

Gassmann, J. (1966). Thesis, Technical Univ. Munich.

Gassmann, J. (1975). *Acta Cryst.* A**31**, 825–831.

Gassmann, J. (1976). *Crystallographic Computing* II, edited by F. R. Ahmed. Copenhagen: Munksgaard. To be published.

Gassmann, J. & Zechmeister, K. (1972). *Acta Cryst.* A**28**, 270–280.

Hauptman, H. (1974). *Acta Cryst.* A**30**, 472–476.

Hoppe, W. (1962). *Acta Cryst.* **15**, 13–17.

Hoppe, W. (1963). *Z. Kristallogr.* **118**, 121–126.

Hoppe, W. & Gassmann, J. (1964). *Ber. Bunsenges. Phys. Chem.* **68**, 608–611.

Hoppe, W. & Gassmann, J. (1968). *Acta Cryst.* B**24**, 97–107.

Hoppe, W., Gassmann, J. & Zechmeister, K. (1970). *Crystallographic Computing* I, pp. 26–36. Edited by F. R. Ahmed. Copenhagen: Munksgaard.

Karle, J. & Hauptman, H. (1950). *Acta Cryst.* **3**, 181–187.

Krabbendam, H. & Kroon, J. (1971). *Acta Cryst.* A**27**, 48–53.

Lajzérowicz, J. & Lajzérowicz, J. (1966). *Acta Cryst.* **21**, 8–12.

Rango C. de, Mauguen, Y. & Tsoucaris, G. (1975). *Acta Cryst.* A**31**, 227–233.

Rossmann, M. G. & Blow, D. M. (1963). *Acta Cryst.* **16**, 39–45.

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.

Tsoucaris, G. (1970). *Acta Cryst.* A**26**, 492–499.