Thus, a map calculated with the weighted coefficients $F_w = 2w|F_o| - |F_c|$ will provide the full scattering signal with the correct phase, plus additional random and systematic noise that has been minimized if the true phase is unknown,

$$(2wF_o - F_c)\exp i\alpha_c = F_m \exp i\alpha_m + F_c \exp i\alpha_c$$
(true signal at full strength)
$$+ (F_m^2 - \langle F_m^2 \rangle)/F_c \exp i\alpha_c$$
(minimal systematic noise)
$$+ F_m \exp i(-\alpha_m + 2\alpha_c)$$
(random noise).

References

BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography*. New York: Academic Press.

BRANDEN, C.-I. & JONES, T. A. (1990). *Nature (London)*, **343**, 687-689.

CHEN, L. Q., ROSE, J. P., BRESLOV, E., YANG, D., CHANG, W. R., FUREY, W. F., SAX, M. & WANG, B. C. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 4240-4244.

GREER, J. (1985). *Methods Enzymol.* **115**, 206-224.

MESSERSCHMIDT, A., ROSSI, A., LADENSTEIN, R., HUBER, R., BOLOGNESI, M., GATTI, G., MARCHESINI, A., PETRUZZELLI, R. & FINAZZI-AGRO, A. (1989). *J. Mol. Biol.* **206**, 513-529.

PHILLIPS, S. E. V. (1978). *Nature (London)*, **273**, 247-248.

READ, R. J. (1986). *Acta Cryst.* A**42**, 140-149.

SIM, G. A. (1960). *Acta Cryst.* **13**, 511-512.

TULINSKY, A., PARK, C. H. & SKRZYPCZAK-JANKUN, E. (1988). *J. Mol. Biol.* **202**, 885-901.

WANG, B.-C. (1985). *Methods Enzymol.* **115**, 90-112.

WILLIAMS, T. (1982). PhD thesis, Univ. of North Carolina at Chapel Hill, USA.

WILSON, C., WARDELL, M., WEISGRABER, K. H., MAHLEY, R. W. & AGARD, D. A. (1991). *Science*, **252**, 495-506.

Acta Cryst. (1993). A**49**, 104-106

The Decomposition Scheme for Direct Methods

BY VÁCLAV KŘÍŽ

Fyzikální ústav ČSAV, Cukrovarnická 10, 162 00 *Praha* 6, *Czechoslovakia*

(*Received* 18 *November* 1990; *accepted* 19 *June* 1992)

Abstract

A new method for the decomposition of a set of start phases in two subsets is described. The decomposition enables the derivation of the subsets that have good values of some figures of merit to different ones with nearly the same value. By this operation a new set of start phases is obtained for the next refinement process. The method presented can be used as a simple but useful extension of the advanced program systems for the solution of the phase problem by direct methods.

1. Introduction

Each direct-method routine consists in forming some real function $G(\varphi)$ of the phase set φ and in generating a limited number of sets $\varphi_1, \ldots, \varphi_r$ of phases for which the values $G(\varphi_i)$ are close to the value expected for the correct solution estimated by the statistical theory.

The function $G(\varphi)$ can be, for example, some combination of good figures of merit. We denote by V the range of the function G. V is a subset of an n-dimensional hypercube,

$$V = I_1 \times \ldots \times I_n,$$

where n is the number of unknown phases, I_j is an interval $(-\pi, \pi)$ in the case of general phases or two values (that differ by π) for special phases.

Let the function G be chosen such that, for the correct phase set φ^*, the value $G(\varphi^*)$ can be expected to be very small, *i.e.* the routine has to generate 'the best' minima of G.

The most important component of a direct-method routine is a transformation $P: V \to V$. For some starting set $\varphi \in V$ this transformation derives a set of phases $P(\varphi) \in V$ for which a small value of $G[P(\varphi)]$ can be expected. As an example of P we can consider the traditional tangent procedure. Let $D \subset V$ be a subset of all the sets of starting phases used as the arguments of function P. D may be, for example, a result of the magic-integer procedure (White & Woolfson, 1975) or it may be identical to V [see, for example, the 'random approach' of Yao (1981)]. If P is based on developing phase values directed by a 'convergence map' (Germain, Main & Woolfson, 1970) from a small subset of fixed (origin definition) and trial phases then D is constructed by setting arbitrary acceptable values to the remaining phases. The direct-methods routine is completely described by the triad $\{D, P, G\}$; the pair $\{D, P\}$ determines which phase sets may be produced by means of this routine.

0108-7673/93/010104-03$06.00

Let $C \subset V$ be a subset of all 'correct' phase sets in the sense that in the interpretation of correct sets a reasonable fragment or all non-hydrogen atoms can be recognized. S denotes all the start sets for all the acceptable generated solutions, *i.e.* $\varphi \in S$ if and only if $P(\varphi) \in C$. U denotes all the uninterpretable generated results with values of figures of merit better than those of all correct solutions, *i.e.* $\varphi \in U$ if and only if $G[P(\varphi)] < g^*$, where g^* is the minimal value of $G(\varphi)$ for all $\varphi \in P(S)$, where $P(S)$ denotes the set of all results of the application of the transformation function P to all the phase sets from S. Note that we have to interpret S and U not only to reflect the quality of the direct-methods routine but also the quality of the experimental data.

The purpose of this contribution is to improve the advanced direct-method routines in the case when S and U are small compared with D and S is not empty. This means that the quality of experimental data and the use of the statistical theory is adequate for this case, but the minimization process based on the arbitrary choice of the starting sets from D ('brute-force approach') will probably be unsuccesful for limited computer time.

2. Decomposition

We will suppose the following.

(i) V is equal to D (the transformation function is defined for the same sets as function G).

(ii) The transformation P is stable, *i.e.* $P[P(\varphi)] = P(\varphi)$ for all $\varphi \in D$.

(iii) The transformation P is realized as a searching process for local minima of the function G.

(iv) $G(\varphi)$ can be expressed as a function of a vector Y, where for each j the value $Y_j(\varphi)$ is given by the modulo 2π of a linear combination of a small number of phases with coefficients equal to 1 or -1. Hence there is a function $\bar{G}$ with the property $\bar{G}[Y(\varphi)] = G(\varphi)$ for each φ. Y_j may be, for example, a triplet or quartet relationship.

All these suppositions are not necessary; they are only aids to simplify later considerations and all are acceptable for current direct-methods routines. The main problem of the application of direct methods results from too many local minima of the functions G.

The random generation of starting sets for minimization processes ('the refinement of random phases') may be successful only if the region S of sets from which the minimization procedure converges to the correct solution φ^* is sufficiently wide (*i.e.* the radius of convergence is great). In the opposite case, we can use the strategy of improvement upon the minima already found.

The problem is how to change the values of phases and keep the value of the function G small. A possible solution (Kříž, 1989) is based on the change of a sufficient number of phases and the simultaneous change of only a small number of values Y_j. This decomposition enables setting arbitrary values to 30–35% of the phases and keeping the values of about 50–70% of triplets.

We can expect better results if we restrict the values for the changed phases as follows. Let N be an n-dimensional vector of components 0 or π. Let $\bar{\varphi}$ be a phase set that has good values of figures of merit and let $\varphi^{(N)} = (\bar{\varphi} + N) \bmod 2\pi$. The values $Y_j(\varphi^{(N)})$ are equal to either $Y_j(\bar{\varphi})$ or $Y_j(\bar{\varphi}) \pm \pi$. They remain unchanged when Y_j depend on an even number of changed phases. This is in contrast to the decomposition scheme given by Kříž (1989). In this scheme the value Y_j remains unchanged only when Y_j does not depend on the changed phases.

3. A simple algorithm for the decomposition of the sets of phases

The problem is how to divide suboptimally all the phases into two subsets for the determination of the vector N. At first we define the weight as

$$w(N) = \sum_{j=1}^{n} w_j |Y_j(N) - Y_j(Z)|,$$

where Z is the vector of all the components that are equal to zero and w_j is an estimate of how a change of the value Y_j affects the value of $\bar{G}[Y(\varphi)]$. The quotient $w(N)/(\pi \sum_{j=1}^{n} w_j)$ expresses the (weighted) number of changed components of the vector $Y(\varphi + N)$ in comparison with $Y(\varphi)$.

Let $e^{(j)}$ be the vector defined by $e_j^{(j)} = \pi$ and $e_i^{(j)} = 0$ for $j \neq i$. The algorithm for the generation of the vectors N with a small weight may be designed quite simply with the following five steps.

Step 1. Generate the vector N randomly.

Step 2. Determine the index k with the property

$$w(N + e^{(k)}) \leq w(N + e^{(i)}) \text{ for } i = 1, \ldots, n.$$

Step 3. If $w(N) > w(N + e^{(k)})$ and the number of nonzero phases of vector $N + e^{(k)}$ is greater than $n/3$, replace N by $(N + e^{(k)}) \bmod 2\pi$ and return to step 2.

Step 4. Store vector N and repeat steps 1–4 a large number of times (*e.g.* 300).

Step 5. Select a subset of different vectors with minimal weights from the stored vectors.

4. The shift from good local minima

Let $N^{(k)}$ for $k = 1, \ldots, n$ be a set of shifting vectors with a small weight $w(N^{(k)})$ and let the value $G(\bar{\varphi})$ be small. We can expect the values $G(\bar{\varphi} + N^{(k)})$ to be small too and we can expect that the phase sets $\varphi^{(k)} = P(\bar{\varphi} + N^{(k)})$ will probably be better local minima than those obtained by the application of function P to the randomly generated starting sets.

Note that since $(N^{(k)}+N^{(k)})$ mod 2π is a zero vector, the phase set $P[P(\bar{\varphi}+N^{(k)})+N^{(k)}]$ is expected to be nearly identical to $\bar{\varphi}$. Analogously for $N^{(k)}=(N^{(i)}+N^{(j)})$ mod 2π we can expect $P(\bar{\varphi}+N^{(k)})$ to be very close to $P[P(\bar{\varphi}+N^{(i)})+N^{(j)}]$ or to $P[P(\bar{\varphi}+N^{(j)})+N^{(i)}]$. It is useful to take heed of this fact in the implementation of step 5 of the algorithm and to generate sets of shifting vectors that are not only different but also 'independent'.

5. Application

The decomposition scheme presented may be simply incorporated into the direct-methods routines. It was verified with real structures by the program system *TRYMIN*90. This system and a detailed definition of the form of the function $\bar{G}$ used here (the so-called centrosymmetric approximation) is presented by Kříž (1992). The form of centrosymmetric approximation is derived from Cochran's (1955) distribution. Examples (of real structures) show that there are a number of independent shifting vectors that change about 50% of phases and keep about 90% of triplet invariants unchanged. Further numerical experiments show that we can expect similar results for other combinations of phases, for example, the traditional tangent formula or quartets.

References

COCHRAN, W. (1955). *Acta Cryst.* **8**, 473-478.
GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* B**26**, 274-285.
KŘÍŽ, V. (1989). *Acta Cryst.* A**45**, 456-463.
KŘÍŽ, V. (1992). In preparation.
WHITE, P. S. & WOOLFSON, M. M. (1975). *Acta Cryst.* A**31**, 53-56.
YAO, J. (1981). *Acta Cryst.* A**37**, 642-644.

Acta Cryst. (1993). A**49**, 106-115

Description and Peak-Position Determination of a Single X-ray Diffraction Profile for High-Accuracy Lattice-Parameter Measurements by the Bond Method. I. An Analysis of Descriptions Available

BY EWA GAŁDECKA

Institute of Low-Temperature and Structure Research, Polish Academy of Sciences, ul. Okólna 2, 50-950 *Wrocław, Poland*

(*Received* 20 *October* 1991; *accepted* 26 *June* 1992)

Abstract

The accuracy of lattice-parameter measurements is defined by the actual error of Bragg-angle determination. The total error in the angular position determined (peak position) - corresponding to the Bragg angle - depends on the experiment itself (physical and geometrical aberrations) and on the method of calculation used. The aim of this paper is to find the best method of approximation of the measurement data to ensure a given accuracy - here, 1 part in 10^6 - with the assumption that the technique and the method used for correcting the aberrations allow this accuracy. Considering some disadvantages of interpolation and approximation with polynomials, commonly used in practice, it is suggested the calculations are based on a model of the measured diffraction profile. In the present paper (paper I), desired properties of such a model are discussed. Various possible descriptions of the diffraction profile - including popular 'shape functions' widely used in practice - are collected in a unified and standardized form and classified and analysed, with account taken of (i) physical aspects, (ii) mathematical aspects, (iii) statistical aspects and (iv) practical aspects (applications), with premises resulting from the measurement method, the Bond method [Bond (1960). *Acta Cryst.* **13**, 814-818]. A special emphasis has been put on the best description of the moderate asymmetry characterizing the reflections from nearly perfect single crystals and on statistical properties of the model.

1. The problem

1.1. *Precision and accuracy of the lattice-spacing determination*

To ensure a given accuracy $|\Delta d|/d$ of the lattice-parameter determination, the error, $\mathrm{d}\theta$, in the Bragg angle, θ, cannot exceed that resulting from differentiation of the Bragg law, *i.e.*

$$|\mathrm{d}\theta| \le (|\Delta d|/d)\tan\theta. \qquad (1)$$

For example, to achieve the accuracy of 1 part in 10^6 of the d_{111}-spacing measurement of a silicon single crystal when Cu $K\alpha$ radiation is used, the error in θ

0108-7673/93/010106-10$06.00