The presence of more planes in this derived equilibrium form than the growth form described by Groth (1919), is quite normal. This difference is probably due to the velocity of growth of crystals in solutions.

The authors wish to express their thanks to Dr K. V. Mirskaya for her kind assistance and to Dr Yu. T. Struchkov for his interest in the work undertaken.

References

GIBBS, J. W. (1928). Collected works. New York: Longman Green & Co.
GROTH, P. (1919). *Chemische Kristallographie*, Teil 5, p.438 Leipzig: Engelman.
HARTMAN, P. (1958). *Acta Cryst.* **11**, 459.
HARTMAN, P. & PERDOK, W. G. (1955). *Acta Cryst.* **8**, 49, 521.
HERRING, G. (1950). *Phys. Rev.* **75**, 344.
KITAIGORODSKY, A. I. (1966). *Extr. J. Chim. Phys.* **1**, 9.
KITAIGORODSKY, A. I. (1970). *Advanc. Struct. Res. Diffr. Methods*, **3**, 173.
KOSSEL, W. (1928). *Leipz. Vortr.* p. 1. Leipzig: Hirzel.
MASON, R. (1964). *Acta Cryst.* **17**, 547.
STRANSKI, I. N. (1928). *Z. Phys. Chem.* (*A*) **136**, 259.
STRANSKI, I. N. (1956). *Bull. Soc. franç. Minér.* **79**, 359.
STRANSKI, I. N. & KAISHEV, R. (1935). *Phys. Z.* **36**, 393.
WOLFF, G. A. & GUALTIERI, J. G. (1962). *Amer. Miner.* **47**, 562.
WULFF, G. (1901). *Z. Kristallogr.* **34**, 449.
YAMADA, M. (1924). *Z. Phys.* **25**, 289.

Acta Cryst. (1972). A**28**, 210

On Least-squares Refinement of the Phases of Crystallographic Structure Factors

BY D. SAYRE

Mathematical Sciences Department, IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598, *U.S.A.*

(*Received* 6 *July* 1971)

A least-squares technique for direct refinement of the phases of structure-factors of crystals which obey the equations $a_hF_h=\sum_k F_kF_{h-k}$ is described. It is shown by tests carried out under somewhat idealized conditions that the initial phases need not be of a completeness or accuracy sufficient to resolve the atoms of the structure. The technique may be able to provide a bridge in protein crystallography and structure determination generally between preliminary phasing techniques and final refinement of atom parameters.

The refinement techniques in current use in crystallography (least-squares refinement of atomic parameters, Fourier refinement of atomic parameters and phases simultaneously) have the property that they cannot commence until the structure determination has progressed to the point where an approximate set of atom parameters, or equivalently a set of phases sufficiently complete and accurate to yield a Fourier map from which such parameters can be read off, is available. Frequently, however, a less complete and accurate set of phases will be available at an earlier point in the determination. In protein crystallography, for example, phases to 3·0–2·5 Å resolution can be obtained by the method of multiple isomorphous replacement (m.i.r.). It may also be noted that Hauptman (1969) and Karle (1970) have given procedures which from the observed magnitudes alone produce rough approximations to the phases for any structure. Thus a refinement technique which operated directly on phases rather than on atom parameters, and which was able to commence from an initial set of phases of less than atom-resolution quality, could be an important aid in completing protein structures and perhaps in determining structures generally.

In the present paper we consider, as a possible technique of this kind, minimization of the expression $\sum_h |a_hF_h-\sum_k F_kF_{h-k}|^2$ as a function of the phases. This minimization may be regarded as a reciprocal-space analog of the ordinary least-squares refinement of atom parameters. Starting with an initial set $\{\varphi_j^0\}$ of phases the process generates a sequence of approximations $\{\varphi_j^1\}$, $\{\varphi_j^2\}$, ... tending toward the set $\{\varphi_j\}$ at which the function is a minimum. Preliminary tests of the method conducted under partially realistic conditions are reported to show that the final set $\{\varphi_j\}$ can indeed be adequately close to the set $\{\Phi_j\}$ of true phases and that convergence to $\{\varphi_j\}$ can occur even when the initial set $\{\varphi_j^0\}$ is considerably below atom-resolution quality. $\{\varphi_j\}$ may alternatively be thought of as a least-squares solution of the equations

$$a_hF_h=\sum_k F_kF_{h-k} \qquad (1)$$

(Sayre, 1952) without the usual approximation involved in replacing equation (1) by its corresponding set of triple-product relationships.

We may pause briefly to put the present technique into perspective. Hoppe (1963) proposed, although he apparently did not carry out, minimization of the expression $\sum_h w_h(|F_h|-|a_h^{-1}\sum_k F_kF_{h-k}|)^2$ as a means of phase refinement. This process would have been tantamount, however, to finding a least-squares solution only of the radial equations implied by (1). (It should be noted that each equation of form (1) really expresses two equations, relating the real and imaginary parts of the two sides, or alternatively the radial and angular parts. This fact is useful in producing a considerable degree of overdetermination of $\{\varphi_j\}$.) Subsequently Karle & Karle (1966) have elaborated the method known as tangent-formula refinement; their method is essentially the converse of the method suggested by Hoppe, since only the angular equations are allowed to play a full role. In addition their treatment uses E's throughout instead of F's and replaces the minimization process by a form of fixed-point iteration (see *e.g.* Collatz, 1966). All these features probably affect the convergence and accuracy of the method somewhat adversely. Perhaps most closely related to the present technique is the work of Krabbendam & Kroon (1971), who describe the minimization of a function similar to that employed here, but by a discrete search algorithm.

Tests of the method to date have used error-free data calculated from artificial equi-atom structures in space group $P1$. The structures were computer-generated and obeyed simple requirements on bond-lengths and distances between non-bonded atoms. With data of this type, when all reflections within the Cu $K\alpha$ sphere (*i.e.* out to 0·77 Å resolution) are included, the agreement between $\{\varphi_j\}$ and $\{\Phi_j\}$ is so close as to make it difficult to determine the relationship between them precisely. In the one case of this type that was studied the mean discrepancy between the two sets was 1·7°, and it is likely that much of this was due to round-off error in the computations. (In this and all other tests reported the F's were adjusted to correspond to spherically symmetrical Gaussian atoms of shape $\exp(-4r^2)$ before submission to the phase-refinement procedure.) For this reason the principal tests were devoted to more realistic cases, in which data-incompleteness could be expected to play an important role.

In the principal tests of this type an attempt was made to reproduce the data-incompleteness typically encountered in protein crystallography. It was assumed that only 8 independent reflections/atom could be observed, and that these would coincide with the reflections which were strongest when the atoms were given a temperature-factor B of 20. The size of the structures was set at 100 atoms. Under these conditions 800 reflections were available for a structure, of which approximately 150 lay within the sphere of 3 Å resolution and approximately 650 lay outside that sphere. An occasional reflection to 1·5 Å or 1·4 Å was observed. The initial phases φ_j^0 consisted of the correct phases Φ_j for the structure-factors within the 3 Å sphere, which were assumed to be known from m.i.r.; for $j>\sim150$, on the other hand, φ_j^0 was undefined. $\{\varphi_j^1\}$ was obtained for $1\le j\le\sim150$ by setting $\varphi_j^1=\varphi_j^0$, and for $\sim150\le j\le 800$ by setting φ_j^1 equal to the phase assumed by $\sum F_kF_{j-k}$ when only the terms involving F's with φ^1 already assigned were included; $\{\varphi_j^1\}$ was thus obtained by simple tangent formula extension. The minimization process was begun at this point. Only the phases of structure-factors lying outside the 3 Å sphere were allowed to change in this process. Table 1 summarizes the course of the refinement from this point for each of the five structures studied. It is seen that initially the mean error of the extended phases varied from 54·0 to 69·3°. At the end of the refinement this error was reduced to 8·2–9·6°. In the final sets $\{\varphi_j\}$ the proportion of refined phases correct to within 30° was 96–98%. 3-dimensional Fourier syntheses made with refined phases are essentially indistinguishable from those made with all correct phases, and in them most of the individual atoms can be seen as separate objects.

Earlier attempts to extend protein data beyond 3 Å resolution have been reported by Hoppe & Gassmann (1964), Reeke & Lipscomb (1969), and Barrett & Zwick (1971). In the first two of these papers the method employed was essentially that of tangent formula extension. The first paper does not provide an estimate of the errors in the extended phases, but in the second the mean errors of various groups of the extended phases were estimated at 67° to 88°. (Note the approximate agreement with the errors in $\{\varphi_j^1\}$ in the present work.)

Table 1. *Summary of the refinements of five* 100-*atom* '*protein-like*' *structures*

Rows refer to $\{\varphi_j^k\}$. The first five columns trace $R=\sum|a_hF_h-\sum F_kF_{h-k}|^2$, which is the quantity being minimized. The last five columns trace the mean error of the extended phases. The number of extended phases in each structure was: I, 653; II, 657; III, 657; IV, 654; V, 652.

	R					Mean of $\Delta\varphi$				
k	I	II	III	IV	V	I	II	III	IV	V
1	$1{\cdot}01\cdot10^9$	$1{\cdot}03\cdot10^9$	$1{\cdot}74\cdot10^9$	$1{\cdot}11\cdot10^9$	$3{\cdot}68\cdot10^9$	54·0°	60·7°	61·9°	54·8°	69·3°
2	$4{\cdot}26\cdot10^8$	$5{\cdot}23\cdot10^8$	$6{\cdot}28\cdot10^8$	$4{\cdot}56\cdot10^8$	$7{\cdot}74\cdot10^8$	42·6	53·7	54·7	43·8	60·7
3	$2{\cdot}98\cdot10^8$	$3{\cdot}84\cdot10^8$	$4{\cdot}15\cdot10^8$	$2{\cdot}85\cdot10^8$	$5{\cdot}03\cdot10^8$	34·9	47·3	47·2	34·0	54·6
4	$2{\cdot}39\cdot10^8$	$3{\cdot}17\cdot10^8$	$3{\cdot}34\cdot10^8$	$2{\cdot}13\cdot10^8$	$3{\cdot}98\cdot10^8$	30·7	43·2	43·5	27·8	49·3
5	$2{\cdot}03\cdot10^8$	$2{\cdot}67\cdot10^8$	$2{\cdot}82\cdot10^8$	$1{\cdot}76\cdot10^8$	$3{\cdot}14\cdot10^8$	27·4	38·2	39·3	23·1	44·7
10	$8{\cdot}64\cdot10^7$	$1{\cdot}03\cdot10^8$	$1{\cdot}20\cdot10^8$	$7{\cdot}74\cdot10^7$	$1{\cdot}46\cdot10^8$	12·7	16·3	19·5	9·7	25·0
15	$7{\cdot}37\cdot10^7$	$6{\cdot}85\cdot10^7$	$8{\cdot}06\cdot10^7$	$7{\cdot}06\cdot10^7$	$8{\cdot}61\cdot10^7$	9·5	8·6	11·3	8·5	12·8
20	$7{\cdot}37\cdot10^7$	$6{\cdot}79\cdot10^7$	$6{\cdot}88\cdot10^7$	$7{\cdot}06\cdot10^7$	$7{\cdot}14\cdot10^7$	9·4	8·3	9·6	8·4	9·0
25	$7{\cdot}37\cdot10^7$	$6{\cdot}79\cdot10^7$	$6{\cdot}88\cdot10^7$	$7{\cdot}06\cdot10^7$	$7{\cdot}06\cdot10^7$	9·4	8·2	9·6	8·4	8·8

In the third paper a method based upon truncating the negative portions of $\varrho(x)$ was used, with an estimate of the mean error of 78°. These results are in sharp contrast with the mean errors of less than 10° found for the present method. It must be remembered, however, that the above authors used real protein data, and also that not having an accurate knowledge of the correct phases they may have overestimated somewhat the mean error of their extended phases.

It is necessary to bear in mind, in considering the present technique, that enough reflections must be included to make a reasonably well resolved structure possible. If, for example, the observations in the preceding experiments had ended at 2 Å resolution the mean errors of the final sets of extended phases would have risen to 17·2–24·7°. This is in keeping with the fact that equations (1) express a condition on the phases of a reasonably well resolved structure. The omission of terms in (1) can to some extent be compensated for by reducing the a_h, which are known real numbers, appropriately in size. This was done, in all the experiments reported, by multiplication of each a_h by an empirical factor of the form $p-qR-rR^2$, where $R=|h|$. It is found that the optimal values of the parameters p, q, r depend upon the amount and type of data incompleteness, but are independent of the structure.

The computer program which was used to accomplish the minimization in these tests is quite straightforward. Working from the current phases it evaluates the difference between the left- and right-hand sides of equations (1), noting the amounts (real part and imaginary part) by which each equation fails and computing the derivatives of these parts with respect to each phase under refinement; from these quantities it computes a set of increments to the phases by producing and solving the normal equations (full set); finally it adds these increments and recycles. Coded in Fortran and run on a 360 model 91 it takes approximately $4\frac{1}{2}$ minutes/cycle for the 650-parameter refinements described. The initial cycle is modified to produce $\{\varphi_j^1\}$ from $\{\varphi_j^0\}$ as noted earlier. The program needs to be generalized to deal with the usual non-centrosymmetrical space-groups, which unlike $P1$ present special classes of structure-factors whose phases cannot be handled by a purely continuous refinement process.

Tests of the method on data from a real protein are planned in the near future. The severest problem is likely to be the computation itself, with its very considerable problems of scale, but errors in the data and m.i.r. phases, and departures of the structure from the idealized structures considered here, may also provide difficulties.

It may be worth mentioning that another, somewhat smaller, group of tests was carried out in a setting intended to mimic the situation which might arise in structure determination generally if approximations to the phases were calculated directly from the magnitudes by a method like those mentioned earlier. Here the problem would be one of extending phases, not from the inner to the outer reflections, but from the strongest to the somewhat less strong. Using 50-atom structures, phases were extended from the 50 strongest F's within the Cu $K\alpha$ sphere to the 450 next-strongest, with mean errors in the final values of the extended phases of 5·8–9·5°.

In conclusion the author wishes to express his thanks to Dr Alexander Tulinsky, of the Department of Chemistry at Michigan State University, whose suggestion that a direct method for phase extension would be of value in protein crystallography was the impetus for this work.

References

BARRETT, A. N. & ZWICK, M. (1971). *Acta Cryst.* A**27**, 6.

COLLATZ, L. (1966). *Functional Analysis and Numerical Mathematics.* (English edition). New York and London: Academic Press.

HAUPTMAN, H. (1969). *Crystallographic Computing* (Proceedings of the International Summer School on Crystallographic Computing). Edited by F. R. AHMED. Copenhagen: Munksgaard.

HOPPE, W. (1963). *Z. Kristallogr.* **118**, 121.

HOPPE, W. & GASSMANN, J. (1964). *Ber. Bunsenges.* **68**, 808.

KARLE, J. (1970). *Acta Cryst.* B**26**, 1614.

KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849.

KRABBENDAM, H. & KROON, J. (1971). *Acta Cryst.* A**27**, 48.

REEKE, G. N. JR & LIPSCOMB, W. N. (1969). *Acta Cryst.* B**25**, 2614.

SAYRE, D. (1952). *Acta Cryst.* **5**, 60.