Acta Cryst. (1992). A**48**, 28-30

Increasing the Power of Classical Direct Methods of Solving Crystal Structures. The χ^2 Tangent Formula*

BY JORDI RIUS AND CARLES MIRAVITLLES

Institut de Ciencia de Materials de Barcelona, CSIC, Campus Universitari de Bellaterra, 08193 Cerdanyola, Barcelona, Spain

(*Received* 17 *April* 1991; *accepted* 2 *August* 1991)

Abstract

The conventional tangent formula can be derived from the fact that $Z(\Phi)$, *i.e.* the integral $\int_v \rho^3 \, dV$ expressed as a function of the collectivity Φ of phases of the largest E's, is a positive maximum for the correct Φ. In practice, however, refinement of phases with the tangent formula can also lead, for certain space groups and atomic arrangements, to false maxima of $Z(\Phi)$. To reduce the number of such false maxima, $Z(\Phi)$ is multiplied by the penalty function $P(\Phi)=[1+\lambda(1-\chi^2(\Phi)/n)]$, with λ and n being, respectively, a suitable weighting factor and the number of terms in the sum of a χ^2 statistic that takes into account the conditional probability distributions of all the triplets involving two largest E's plus a largest, a medium-large or a weakest one. It will be shown how the combined function $Z(\Phi)\times P(\Phi)$ can be maximized by a tangent formula. The application of this new restrained tangent formula will be illustrated on the basis of a representative example.

1. Introduction

According to Cochran (1952) the integral $\int_v \rho^3 \, dV$ must be a large magnitude. This integral may be approximated in reciprocal space by means of the expression

$$Z(\Phi)=2\sum_{\mathbf{h}}\sum_{\mathbf{h}'} E_{-\mathbf{h}}E_{\mathbf{h}'}E_{\mathbf{h}-\mathbf{h}'}\cos(\varphi_{-\mathbf{h}}+\varphi_{\mathbf{h}'}+\varphi_{\mathbf{h}-\mathbf{h}'}) \quad (1)$$

where the $\varphi_{\mathbf{h}}$ denote the phases of the largest normalized structure factors $\mathbf{E}_{\mathbf{h}}$, hereafter collectively called Φ, and where the phase sums in the cosine terms are triple-phase structure invariants, *i.e.* their values are determined by the structure alone and are independent of the choice of the origin of the unit cell. The conventional tangent formula (Karle & Hauptman, 1956),

$$\varphi_{\mathbf{h}} = \text{phase of } \left\{\sum_{\mathbf{h}'} \mathbf{E}_{\mathbf{h}'}\mathbf{E}_{\mathbf{h}-\mathbf{h}'}\right\},$$

results from the fact that the function $Z(\Phi)$ must be a maximum for the correct Φ, *i.e.* from making

$$\partial Z(\Phi)/\partial\varphi_{\mathbf{h}} = 2E_{\mathbf{h}}\sum_{\mathbf{h}'} E_{\mathbf{h}'}E_{\mathbf{h}-\mathbf{h}'}\sin(-\varphi_{\mathbf{h}}+\varphi_{\mathbf{h}'}+\varphi_{\mathbf{h}-\mathbf{h}'}) = 0 \quad (2)$$

for each $\varphi_{\mathbf{h}}\in\Phi$ (Debaerdemaeker, Tate & Woolfson, 1985). The tangent formula allows the value of a given phase to be refined when the values of the remaining phases are known. This process can be sequentially performed for each $\varphi_{\mathbf{h}}$, thus leading to the improvement of the whole set of phases Φ involved in the refinement. Unfortunately, refinement of phases with the tangent formula can also lead to false maxima of $Z(\Phi)$ for certain space groups and atomic arrangements especially when the initial phases are assigned random values.

2. The χ^2 restraint

To reduce the number of false maxima of $Z(\Phi)$, a modified tangent formula (χ^2 tangent formula) is proposed for space group $P1$ that maximizes $Z(\Phi)$ under the restraint that the reduced χ^2 statistic (Cramér, 1946)

$$\chi^2(\Phi)=\sum_{\mathbf{H}\in K}(\hat{G}_{\mathbf{H}}-G_{\mathbf{H}})^2/\sigma^2_{\mathbf{H}} \quad (3)$$

is low enough ($\leq n_K$). K denotes the set of lattice points in one asymmetric unit of the reciprocal space for which there are useful observations, n_K is the number of terms in the sum and

$$G_{\mathbf{H}}=\left|\sum_{\mathbf{h}''}\mathbf{E}_{\mathbf{h}''}\mathbf{E}_{\mathbf{H}-\mathbf{h}''}\right| = \sum_{\mathbf{h}''}E_{\mathbf{h}''}E_{\mathbf{H}-\mathbf{h}''}\cos(\phi_{-\mathbf{H}}+\varphi_{\mathbf{h}''}+\varphi_{\mathbf{H}-\mathbf{h}''}) \quad (4)$$

with

$$\phi_{\mathbf{H}} = \text{phase of } \left\{\sum_{\mathbf{h}''}\mathbf{E}_{\mathbf{h}''}\mathbf{E}_{\mathbf{H}-\mathbf{h}''}\right\}.$$

The expected value of $G_{\mathbf{H}}$ can be found from

$$\hat{G}_{\mathbf{H}}=\sum_{\mathbf{h}''}E_{\mathbf{h}''}E_{\mathbf{H}-\mathbf{h}''}\langle\cos(\phi_{-\mathbf{H}}+\varphi_{\mathbf{h}''}+\varphi_{\mathbf{H}-\mathbf{h}''})\rangle.$$

Evidently, in order to derive $\hat{G}_{\mathbf{H}}$, the expectation values of the cosine terms should be known. For the

* Presented at the Summer Course on Materials Characterization organized by the University Menendez Pelayo, Barcelona, Spain, 15-19 July 1991.

0108-7673/92/010028-03$03.00

special case of **H** corresponding to the largest, a medium-large or the weakest E, these values can be obtained from the approximation

$$\langle \cos(\phi_{-\mathbf{H}} + \varphi_{\mathbf{h}''} + \varphi_{\mathbf{H}-\mathbf{h}''}) \rangle \cong \xi_{\mathbf{H},\mathbf{h}''}$$

where $\xi_{\mathbf{H},\mathbf{h}''}$ is the conditional expectation value of the cosine of the triple-phase invariant estimated from the known structure-factor magnitudes $E_{\mathbf{H}}$, $E_{\mathbf{h}''}$, $E_{\mathbf{H}-\mathbf{h}''}$ and the atomic content of the unit cell (Germain, Main & Woolfson, 1970). This approximation assumes that $\phi_{\mathbf{H}} \cong \varphi_{\mathbf{H}}$ for the largest and medium-large E's and $\xi_{\mathbf{H},\mathbf{h}''} \simeq \langle \cos(\phi_{-\mathbf{H}} + \varphi_{\mathbf{h}''} + \varphi_{\mathbf{H}-\mathbf{h}''}) \rangle \simeq 0$ for the weakest ones. Finally, the variance $\sigma_{\mathbf{H}}^2$ related to $\hat{G}_{\mathbf{H}}$ can be found from the conditional variances $\alpha_{\mathbf{H},\mathbf{h}''}^2$ of the individual cosine terms (Hauptman, 1972) by applying the central limit theorem, so that

$$\sigma_{\mathbf{H}}^2 = \sum_{\mathbf{h}''} (E_{\mathbf{h}''} E_{\mathbf{H}-\mathbf{h}''})^2 \alpha_{\mathbf{H},\mathbf{h}''}^2.$$

3. The χ^2 tangent formula

The derivation of the χ^2 tangent formula requires the prior definition of the penalty function $P(\Phi)$ based on $\chi^2(\Phi)$

$$P(\Phi) = 1 + \lambda[1 - \chi^2(\Phi)/n_K]$$

with λ a suitable weighting factor. The function $P(\Phi)$ is close to one for a correct Φ. For incorrect solutions, however, $P(\Phi)$ tends to be smaller than one. Combination of $Z(\Phi)$ with the penalty function $P(\Phi)$ leads to the more reliable function

$$Q(\Phi) = Z(\Phi) \times P(\Phi) \qquad (5)$$

that can be maximal if and only if $Z(\Phi)$ and $P(\Phi)$ are both maximal. Notice that, for a correct Φ, $Q(\Phi) \simeq Z(\Phi)$. Function $Q(\Phi)$ can be maximized by setting

$$\partial Q/\partial \varphi_{\mathbf{h}} = 0 \quad \forall \varphi_{\mathbf{h}} \in \Phi$$

and, therefore,

$$\partial Z(\Phi)/\partial \varphi_{\mathbf{h}} - k\, \partial \chi^2(\Phi)/\partial \varphi_{\mathbf{h}} = 0 \qquad (6)$$

with

$$k = \lambda Z(\Phi) n_K^{-1} P(\Phi)^{-1}.$$

Obviously, the utility of (6) depends on the possibility of expressing $\partial \chi^2(\Phi)/\partial \varphi_{\mathbf{h}}$ as an explicit function of $\varphi_{\mathbf{h}}$. In view of (3) and (4), and after some algebraic manipulation (see also Rius & Miravitlles, 1992), one obtains

$$\partial \chi^2(\Phi)/\partial \varphi_{\mathbf{h}} = 4E_{\mathbf{h}} \sum_{\mathbf{H}'} [(\hat{G}_{\mathbf{H}'} - G_{\mathbf{H}'})/\sigma_{\mathbf{H}'}^2] E_{\mathbf{h}-\mathbf{H}'} \times \sin(\varphi_{\mathbf{h}} - \phi_{\mathbf{H}'} - \varphi_{\mathbf{h}-\mathbf{H}'}) \qquad (7)$$

where the $\mathbf{H}'$ summation also includes the Friedel pairs. Substitution of (2) and (7) in (6) leads finally to the desired χ^2 tangent formula,

$$\varphi_{\mathbf{h}} = \text{phase of} \left\{ \sum_{\mathbf{h}'} \mathbf{E}_{\mathbf{h}'} \mathbf{E}_{\mathbf{h}-\mathbf{h}'} + 2k \sum_{\mathbf{H}'} [(\hat{G}_{\mathbf{H}'} - G_{\mathbf{H}'})/\sigma_{\mathbf{H}'}^2] \mathbf{E}_{\mathbf{h}-\mathbf{H}'} \exp i\phi_{\mathbf{H}'} \right\}.$$

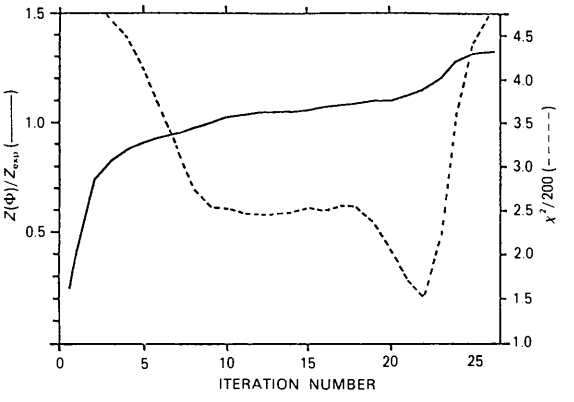
For a correct Φ, $P(\Phi) \simeq 1$ holds and k reduces to $\lambda Z(\Phi)/n_K$. It has been found that the use of the reduced k does not greatly affect the effectiveness of the χ^2 tangent formula in refining random phases. The principal advantage of using the reduced k is that any possible scaling error in the variances $\sigma_{\mathbf{H}}^2$ can be compensated for by slightly changing the value of λ.

Logically, the χ^2 sum need not involve all the measured $E_{\mathbf{H}}$. For the particular case of $\mathbf{H}'$ only including those reflections with weakest E magnitudes, an expression that essentially coincides with the reduced form of the Sayre-equation tangent formula results (Debaerdemaeker, Tate & Woolfson, 1988). Since the effectiveness of using the weakest E's as a source of additional information for solving the structures of even small proteins has been previously discussed (Woolfson & Yao, 1990), this particular case will not be handled here. Instead, the capability of the χ^2 tangent formula to refine initially random phases when only some medium-large E's are considered in the χ^2 sum will be emphasized. This is the first time that this type of information has been actively used in a tangent formula refinement.

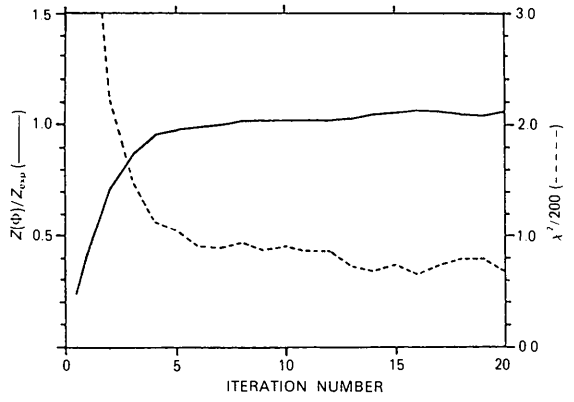
4. Test example

The selected example is a known derivative of the punctatin A antibiotic produced by the dung fungus *Poronia punctata* (molecular formula: $C_{15}H_{24}O_3$, triclinic, space group $P1$, $Z = 3$, cell volume = 1106 Å^3) (Poyser, Edwards, Anderson, Hursthouse, Walker, Sheldrick & Walley, 1986). To prove the difficulty of using the unweighted conventional tangent formula to solve this structure (*i.e.* without introducing the χ^2 restraint), 400 sets of 190 initially random phases of the largest E's (>1.70) have been refined.* According to the figures of merit, no correct solution is present. Most of the calculated Fourier maps show a dominant peak. Fig. 1(*a*) describes the evolution of the normalized $Z(\Phi)$ function as the refinement progresses. It can be seen how the final value of $Z(\Phi)$ is much greater than expected [$Z(\Phi)/Z_{\text{exp}} = 1.33$]. In other words, instead of reaching the correct local maximum of $Z(\Phi)$, the global maximum of $Z(\Phi)$ is found. This

* A section containing the complete mathematical derivations as well as the computer outputs of the test example has been deposited with the British Library Lending Division as Supplementary Publication No. SUP54561 (32 pp.). Copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.



(*a*)



(*b*)

Fig. 1. Evolution of the normalized Cochran integral $Z(\Phi)/Z_{exp}$ and of the $\chi^2/200$ goodness of fit during the refinement of the 190 initially random phases of the largest normalized structure factors using (*a*) the conventional tangent formula and (*b*) the χ^2 tangent formula. To facilitate the comparison between the refinements, the same starting phase values have been used. While the conventional tangent formula maximizes $Z(\Phi)$ to its global maximum (*a*), the addition of the χ^2 restraint succeeds in keeping $Z(\Phi)$ at the correct local maximum (*b*).

is also reflected in the high χ^2 values computed at the end of each iteration.

To avoid this behaviour, 200 medium-large E's in the range 1.70 to 1.41 have been considered in the χ^2 sum. This represents the addition of 3271 new triplets to the 3706 involving only the largest E's. 40 sets of initially random phases were refined using the χ^2 tangent formula with $\lambda = 0.095$, recalculating ϕ_H and G_H every 38 refined phases. From the 40 sets, two showed the images of the three symmetry-independent molecules. The evolution of $Z(\Phi)$ and χ^2 for one of these two sets is represented in Fig. 1(*b*). Comparison with Fig. 1(*a*) shows that $Z(\Phi)/Z_{exp}$ is here closer to 1 and that $\chi^2/200$ gradually decreases from 4.86 at the first iteration to 0.68 at the end of the refinement. This clearly indicates that the addition of the χ^2 restraint hinders $Z(\Phi)$ from reaching the global maximum.

Although derived in space group $P1$, the χ^2 tangent formula is completely general and can be applied to all space groups with only minor modifications.

This work has been supported by the DGICYT (Project PB89-0036).

References

COCHRAN, W. (1952). *Acta Cryst.* **5**, 65-67.
CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.
DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1985). *Acta Cryst.* A**41**, 286-290.
DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1988). *Acta Cryst.* A**44**, 353-357.
GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* B**26**, 274-285.
HAUPTMAN, H. (1972). *Z. Kristallogr.* **135**, 1-17.
KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635-651.
POYSER, J. P., EDWARDS, R., ANDERSON, J., HURSTHOUŞE, M., WALKER, N., SHELDRICK, G. & WALLEY, A. (1986). *J. Antibiot.* **39**, 167-169.
RIUS, J. & MIRAVITLLES, C. (1992). *Acta Cryst.* A**48**, 69-70.
WOOLFSON, M. M. & YAO, J. (1990). *Acta Cryst.* A**46**, 409-413.

Acta Cryst. (1992). A**48**, 30-36

Integration of Patterson Information into Direct Methods. II. The First Applications

BY A. ALTOMARE, G. CASCARANO AND C. GIACOVAZZO

Istituto di Ricerca per lo Sviluppo di Metodologie Cristallografiche, c/o Dipartimento Geomineralogico, Campus Universitario, 70124 *Bari, Italy*

(*Received* 6 *February* 1991; *accepted* 21 *June* 1991)

Abstract

A probabilistic formula [Giacovazzo (1991). *Acta Cryst.* A**47**, 256-263] estimates triplet invariant phases given prior information on a non-Harker Patterson peak **u**. The formula requires prior information both on the coordinates of the peak and on the scattering factors of the atoms with mutual distance **u**. Since

0108-7673/92/010030-07$03.00