

- JAYNES, E. T. (1957). *Phys. Rev.* **106**, 620-630.
 KARLE, J. (1972). *Acta Cryst.* **B28**, 3362-3369.
 KARLE, J. & GILARDI, R. D. (1973). *Acta Cryst.* **A29**, 401-407.
 KARLE, J. & HAUPTMAN, H. (1950). *Acta Cryst.* **3**, 181-187.
 KARLE, J. & HAUPTMAN, H. (1953). *Acta Cryst.* **6**, 131-135.
 KIRKWOOD, J. G. (1935). *J. Chem. Phys.* **3**, 300-313.
 KLUG, A. (1958). *Acta Cryst.* **11**, 515-543.
 KRONENBURG, M. J., PESCHAR, R. & SCHENK, H. (1990). *Proc. K. Ned. Akad. Wet.* **93**, 29-32.
 MAYER, J. E. & MAYER, M. G. (1940). *Statistical Mechanics*. New York: Wiley.
- NAYA, S., NITTA, I. & ODA, T. (1965). *Acta Cryst.* **19**, 734-747.
 NIJBOER, B. R. A. & VAN HOVE, L. (1952). *Phys. Rev.* **85**, 777-783.
 PESCHAR, R. (1987). Thesis, Univ. of Amsterdam, The Netherlands.
 PESCHAR, R. & SCHENK, H. (1987). *Acta Cryst.* **A43**, 513-522.
 SUBRAMANIAN, V. & HALL, S. R. (1982). *Acta Cryst.* **A38**, 577-590.
 VAN KAMPEN, N. G. (1961). *Physica (Utrecht)*, **27**, 783-792.
 WATSON, G. N. (1952). *A Treatise on the Theory of Bessel Functions*. Cambridge Univ. Press.
 WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318-321.

Acta Cryst. (1991). **A47**, 480-484

On the Estimation of Non-Measured Diffraction Magnitudes

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

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

AND N. STEADMAN

Chemical Crystallography Laboratory, 9 Parks Road, Oxford OX1 3PD, England

(Received 29 June 1990; accepted 26 February 1991)

Abstract

Shortage or ambiguity in diffraction data may limit the efficiency of structure analysis techniques. The joint probability distribution method has been used in order to estimate the values of non-measured diffraction magnitudes. Some experimental tests show the formulae have an efficiency which is promising.

1. Introduction

Sometimes the number of measured diffraction magnitudes is not sufficient for the satisfactory attainment of a crystal structure solution and refinement. Shortage of data occurs regularly in macromolecular crystallography (too many structural parameters to determine compared with the number of available independent observations), but occasionally it also occurs in single-crystal small-molecule crystallography when the diffracting crystal is of poor quality or is unstable under experimental conditions. Shortage of data is very critical in powder crystallography, where occasional or systematic overlapping of diffraction effects does not allow experimental measurement of tens or hundreds of single diffraction amplitudes.

The most effective way for overcoming the problem is to optimize and/or simultaneously to use some complementary experimental techniques (e.g. low-temperature apparatus or synchrotron radiation or a combination of X-ray and/or neutron and/or electron diffraction *etc.*). Often such techniques are not immediately available and the practical problem may

be that of predicting the values of some non-measured diffraction amplitudes by exploiting the measured data as prior information. For example, if data up to $(\sin \theta)/\lambda = \rho_1$ are available, one may try to predict the amplitudes in the range $\rho_2 - \rho_1$ with $\rho_2 > \rho_1$. The problem may be of great importance for crystal-structure-solution methods. In Patterson techniques such supplementary information can make the deconvolution of the Patterson map easier. In direct methods it will make the phasing process and the identification of the correct solution more efficient.

The simplest way to predict the value of a non-measured intensity with vectorial index \mathbf{h} is to use Wilson's statistics. Expectations are

$$\langle |E_{\mathbf{h}}| \rangle = (2/\pi)^{1/2} \quad \text{for centric structures} \quad (1a)$$

$$\langle |E_{\mathbf{h}}| \rangle = (\pi/2)^{1/2} \quad \text{for non-centric structures.} \quad (1b)$$

Relations (1a, b) are too poor to be used for most practical purposes. Probabilistic expressions for estimating $|E_{\mathbf{h}}|$ from all the most reliable quartets in which \mathbf{h} is a cross term were presented by Van der Putten, Schenk & Tsoucaris (1982). More recently, David (1987) suggested a formula which relies on the fact that a Patterson function $P(\mathbf{u})$ is a positive function as well as $P^2(\mathbf{u})$: Sayre's (1952) squaring method was then applied. David's conclusive formula is

$$\langle |F_{\mathbf{h}}|^2 \rangle = \sum_{\mathbf{k}} |F_{\mathbf{k}}|^2 |F_{\mathbf{h}-\mathbf{k}}|^2. \quad (2)$$

Even if of large interest, (2) suffers from two limitations:

(1) Sayre's method is strictly applicable to equal-atom structures. Therefore, a Patterson map, with its very large origin peak, does not comply with Sayre's hypotheses. The use of origin-removed Patterson functions could improve the situation, but in this case equation (2) is no longer valid;

(2) Sayre's equation has asymptotical validity; *i.e.* it requires a rather large number of triplets.

A probabilistic rather than an algebraic relationship among structure-factor magnitudes gives a better fit in practical situations where only a limited number of triplets is available. In accordance with this consideration the first aim of this paper is to provide a probabilistic basis for estimating diffraction magnitudes given prior information on other magnitudes. Centric and non-centric crystals will be taken into account and some practical tests will be presented.

2. The probabilistic estimation of $|E_h|$ in non-centric crystals

Let us assume that the atomic coordinates are the primitive random variables, uniformly distributed in the unit cell. Then the joint probability distribution function for three complex-valued normalized structure factors $E_1 = E_{h_1}$, $E_2 = E_{h_2}$, $E_3 = E_{h_3}$ ($h_1 + h_2 + h_3 = 0$), up to and including terms of order $N^{-1/2}$, is given (Hauptman & Karle, 1953; Cochran, 1955) by

$$P(R_1, R_2, R_3, \varphi_1, \varphi_2, \varphi_3) \approx \pi^{-3} R_1 R_2 R_3 \exp\{-(R_1^2 + R_2^2 + R_3^2) + 2N^{-1/2} R_1 R_2 R_3 \cos(\varphi_1 + \varphi_2 + \varphi_3)\}$$

where $E_i = R_i \exp(i\varphi_i)$. Then

$$P(R_1, R_2, R_3) \approx 8R_1 R_2 R_3 \exp\{-(R_1^2 + R_2^2 + R_3^2)\} \times I_0(2N^{-1/2} R_1 R_2 R_3)$$

and

$$P(R_1|R_2, R_3) \approx 2R_1 \exp(-R_1^2 - R_2^2 R_3^2/N) \times I_0(2N^{-1/2} R_1 R_2 R_3) \quad (3)$$

are obtained.

The first moments of (3) may be calculated *via* the formula

$$\int_0^\infty x^\mu \exp(-ax^2) I_0(qx) dx = \Gamma[(\mu+1)/2] / (2a^{(\mu+1)/2}) \times {}_1F_1[(-\mu+1)/2; 1; -q^2/4a] \exp(q^2/4a)$$

where Γ is the gamma function and ${}_1F_1$ is the confluent hypergeometric function. We obtain

$$\langle R_1|R_2, R_3 \rangle \approx \frac{1}{2} \pi^{1/2} {}_1F_1(-\frac{1}{2}; 1; -R_2^2 R_3^2/N).$$

If $R_2^2 R_3^2/N$ is sufficiently small then ${}_1F_1$ may be expanded as

$$1 + R_2^2 R_3^2/2N - R_2^4 R_3^4/16N^2 + R_2^6 R_3^6/(96N) - \dots$$

The convergence is not fast: a rough approximation gives

$$\langle R_1|R_2, R_3 \rangle \approx \frac{1}{2} \pi^{1/2} [1 + (2N)^{-1} R_2^2 R_3^2]. \quad (4a)$$

More immediate is the estimation of R_1^2 :

$$\langle R_1^2|R_2, R_3 \rangle \approx {}_1F_1(-1; 1; R_2^2 R_3^2/N) = 1 + N^{-1} R_2^2 R_3^2. \quad (4b)$$

If prior information on R_2 and R_3 is not available R_1 and R_1^2 are expected to coincide with values provided by Wilson's theory. However, when this information is available $\langle R_1|R_2, R_3 \rangle$ and $\langle R_1^2|R_2, R_3 \rangle$ are always larger than the values suggested by Wilson. In practice, small R values cannot be identified by (4a) and (4b).

An alternative mathematical approach may be used. The characteristic function of the distribution may be expanded in Taylor series up to and including N^{-1} -order terms and its Fourier transform may be calculated. Then [Naya, Nitta & Oda (1965); but see also Giacobozzo (1976), where the three-variate distribution may be obtained as a marginal of a seven-variate one],

$$P(R_1, R_2, R_3, \varphi_1, \varphi_2, \varphi_3) \approx \pi^{-3} R_1 R_2 R_3 \exp[-(R_1^2 + R_2^2 + R_3^2)] \times \left\{ 1 + 2N^{-1/2} R_1 R_2 R_3 \cos(\varphi_1 + \varphi_2 + \varphi_3) + N^{-1} \left[R_1^2 R_2^2 R_3^2 \cos 2(\varphi_1 + \varphi_2 + \varphi_3) + (R_1^2 - 1)(R_2^2 - 1)(R_3^2 - 1) - \frac{1}{4} \sum_{i=1}^3 L_4(R_i) \right] \right\} \quad (5)$$

where $L_4(R_i) = R_i^4 - 4R_i^2 + 2$. In (5) terms of order larger than N^{-1} have been omitted.

The marginal distribution $P(R_1, R_2, R_3)$ may be obtained by integrating (5) over $\varphi_1, \varphi_2, \varphi_3$: then the conditional distribution

$$P(R_1|R_2, R_3) \approx 2R_1 \exp(-R_1^2) \times \{1 + Q_L^{-1}[-(4N)^{-1} L_4(R_1) + N^{-1}(R_2^2 - 1)(R_3^2 - 1)]\} \quad (6)$$

is obtained, where

$$Q_L = 1 - 4N^{-1} [L_4(R_2) + L_4(R_3)].$$

Distribution (6) reduces to Wilson's distribution (extended to N^{-1} -order terms)

$$P(R_1) \approx 2R_1 \exp(-R_1^2) \{1 - (4N)^{-1} L_4(R_1)\}$$

when prior information on R_2 and on R_3 is not available.

From (6) the following conditional expected values are obtained:

$$\begin{aligned} \langle R_1 | R_2, R_3 \rangle &= \frac{1}{2} \pi^{1/2} \{1 + Q_L^{-1} [(16N)^{-1} \\ &\quad + (2N)^{-1} (R_2^2 - 1)(R_3^2 - 1)]\} \\ &= \frac{1}{2} \pi^{1/2} \{1 + (2Q_L N)^{-1} (R_2^2 - 1)(R_3^2 - 1)\} \end{aligned} \quad (7a)$$

$$\langle R_1^2 | R_2, R_3 \rangle = 1 + (Q_L N)^{-1} (R_2^2 - 1)(R_3^2 - 1) \quad (7b)$$

which may be compared with classical Wilson's values

$$\langle R_1 \rangle = \pi^{1/2}/2, \quad \langle R_1^2 \rangle = 1.$$

According to (7a) and (7b), if R_2 and R_3 are both larger than unity then R_1 and R_1^2 are expected to be larger than $\pi^{1/2}/2$ and 1 respectively; they are expected to be smaller than $\pi^{1/2}/2$ and 1 respectively if R_2 is large and R_3 is small or *vice versa*. The triplet contribution is however of order N^{-1} , too small to be useful in practice. The obvious suggestion is to consider the more complex distribution

$$P(R_h, \varphi_h, \{R_k, R_{h-k}, \varphi_k, \varphi_{h-k}\})$$

where \mathbf{k} varies over reciprocal space. Such a distribution involves quartet phase relationships. If they are neglected and if each triplet is assumed to give a contribution to the estimation of R_h independent of the others then

$$\begin{aligned} P(R_h, \{R_k, R_{h-k}\}) \\ \approx \prod_n [\pi^{-1} R_n \exp(-R_n^2)] \{1 + (R_h^2 - 1)N^{-1} \\ \times \sum_k (R_k^2 - 1)(R_{h-k}^2 - 1) - (4N) \sum_n L_4(R_n)\} \end{aligned}$$

from which

$$\langle R_h | \dots \rangle = \frac{1}{2} \pi^{1/2} \left\{ 1 + (2N)^{-1} \times \sum_k \frac{(R_k^2 - 1)(R_{h-k}^2 - 1)}{1 - (4N)^{-1} [L_4(R_k) + L_4(R_{h-k})]} \right\} \quad (8a)$$

$$\langle R_h^2 | \dots \rangle = 1 + N^{-1} \sum_k \frac{(R_k^2 - 1)(R_{h-k}^2 - 1)}{1 - (4N)^{-1} [L_4(R_k) + L_4(R_{h-k})]} \quad (8b)$$

3. The probabilistic estimation of $|E_h|$ in centric crystals

If the exponential form of a trivariate distribution (up to $N^{-1/2}$ -order terms) is used in $P\bar{1}$ then

$$\begin{aligned} P(R_1, R_2, R_3) &= (2/\pi)^{3/2} \exp[-(R_1^2 + R_2^2 + R_3^2)] \\ &\quad \times \cosh(R_1 R_2 R_3 N^{-1/2}) \end{aligned}$$

and

$$\begin{aligned} P(R_1 | R_2, R_3) &= (2/\pi)^{1/2} \exp(-R_1^2/2 - R_2^2 R_3^2/2N) \\ &\quad \times \cosh(R_1 R_2 R_3 N^{-1/2}). \end{aligned}$$

Further calculations bring results equivalent to (4a) and (4b). Therefore, in centric crystals we also prefer to use the distribution arising from the Fourier transform of the expansion in Taylor series of the characteristic function. The result is

$$\begin{aligned} P(R_1 | R_2, R_3) &= (2/\pi)^{1/2} \exp(-R_1^2/2) \\ &\quad \times \{1 + Q_H^{-1} [-(8N)^{-1} H_4(R_1) \\ &\quad + (2N)^{-1} (R_1^2 - 1)(R_2^2 - 1)(R_3^2 - 1)]\} \end{aligned}$$

where $H_4(R_i) = R_i^4 - 6R_i^2 + 3$ is the Hermite polynomial of order four and $Q_H = 1 - (8N)^{-1} [H_4(R_2) + H_4(R_3)]$. Accordingly,

$$\begin{aligned} \langle R_h | \{R_k, R_{h-k}\} \rangle \\ = (2/\pi)^{1/2} \left\{ 1 + 2N^{-1} \times \sum_k \frac{(R_k^2 - 1)(R_{h-k}^2 - 1)}{1 - (8N)^{-1} [H_4(R_k) + H_4(R_{h-k})]} \right\} \end{aligned} \quad (9a)$$

$$\begin{aligned} \langle R_h^2 | \{R_k, R_{h-k}\} \rangle \\ = 1 + N^{-1} \sum_k \frac{(R_k^2 - 1)(R_{h-k}^2 - 1)}{1 - (8N)^{-1} [H_4(R_k) + H_4(R_{h-k})]} \end{aligned} \quad (9b)$$

4. Experimental and conclusions

In order to check the efficiency of (8a, b) and (9a, b) we have used nine structures, the code names of which, together with relevant crystallochemical data, are shown in Table 1. GEN1 and GEN2 are model structures: their diffraction magnitudes were generated in order to check data without errors up to the desired $(\sin \theta)/\lambda$ limit. PGE2, SKN1, LOGANIN and FEGAS are non-centrosymmetrical real structures whose data were collected by single-crystal techniques. SALEX, PIC and TIPORF are centrosymmetrical structures: data for the last two were collected by powder techniques and synchrotron radiation.

The first result of the tests is that (8) and (9) are rather insensitive to the presence of the scaling functions L_4 and H_4 respectively. Therefore, to a good approximation it may be stated that, for both centric and non-centric crystals,

$$\langle R_h^2 | \dots \rangle = 1 + N^{-1} \sum_k (R_k^2 - 1)(R_{h-k}^2 - 1), \quad (10)$$

while it is

$$\langle R_h | \dots \rangle = \frac{1}{2} \pi^{1/2} [1 + (2N)^{-1} \sum_k (R_k^2 - 1)(R_{h-k}^2 - 1)] \quad (11a)$$

Table 1. Code name, space group and crystallochemical data for the test structures

Code	Space group	Chemical formula	Z
GEN1	<i>P</i> 1	C ₇	1
GEN2	<i>P</i> 2 ₁	Pb ₂ O ₄	4
PGE2 ⁽¹⁾	<i>P</i> 1	C ₂₀ H ₃₂ O ₅	1
SKN1 ⁽²⁾	<i>P</i> 3 ₁	C ₇ H ₁₆ ClNO ₄	3
LOGANIN ⁽³⁾	<i>P</i> 2 ₁ 2 ₁ 2 ₁	C ₁₇ H ₂₆ O ₁₀	4
FEGAS ⁽⁴⁾	<i>P</i> 6 ₃ / <i>mmc</i>	Fe ₂ Ga ₂ S ₅	2
SALEX ⁽⁵⁾	<i>P</i> 3̄	K _{3.86} Na _{5.30} (H ₃ O ⁺) _{0.84} · · · Fe ₆ ³⁺ O ₂ (SO ₄) ₁₂ · 17·08 H ₂ O	1
PIC ⁽⁶⁾	<i>P</i> 1̄	PbC ₂ O ₄	2
TIPORF ⁽⁷⁾	<i>I</i> 4/ <i>m</i>	C ₄₄ H ₂₈ N ₄ TiCl ₂	2

References: (1) DeTitta, Langs, Edmonds & Duax (1980); (2) unpublished; (3) Jones, Sheldrick, Glösenkamp & Tietze (1980); (4) Cascarano, Dogguy-Smiri & Nguyen-Huy Dung (1987); (5) Scordari & Stasi (1990); (6) Christensen & Lehmann (1989); (7) Christensen, Grand, Lehmann & Cox (1990).

for non-centric and

$$\langle R_h | \dots \rangle = (2/\pi)^{1/2} [1 + (2N)^{-1} \sum_k (R_k^2 - 1)(R_{h-k}^2 - 1)] \quad (11b)$$

for centric crystals. Accordingly, experimental results will refer to the use of (10) and (11) rather than to (8) and (9).

Since the number of observations is finite the number of triplets involved in (11) varies according to *h*. We have always used in our calculations the full set of the triplets experimentally available. Due to the mathematical approximations introduced in the probabilistic approach described in § 3, negative estimates for *R_h* and for the variance can be obtained. We have therefore decided to avoid any further use of the variance in the calculations. Furthermore, we also decided to normalize calculated *R_h* values with respect to the number of involved triplets [*L*(2*N*)⁻¹⟨(R_k² - 1)(R_{h-k}² - 1)⟩ has been used in the formulas rather than (2*N*)⁻¹∑_k (R_k² - 1)(R_{h-k}² - 1), where *L* is a scale constant] and to rescale them under the constraint that the average of the *R_h*²'s must be unity.

As a first application we have estimated for SKN1 the values of *R* in the (sin θ)/λ range 0.62-0.63 Å⁻¹ by exploiting on the right-hand side of (11a) data measured in the range 0-0.62 Å⁻¹. Predicted and true values are compared in Table 2. Estimated magnitudes show a remarkable fit to true ones.

The general efficiency of the formulas may be deduced from the values of the residual

$$RES = \frac{\sum_h |R_{obs} - R_{calc}|}{\sum_h R_{obs}}$$

In Table 3 for each test structure the upper limit of (sin θ)/λ and two RES values are given: for RES_w and RES_n, *R_{calc}* is calculated according to (1) and (11a, b) respectively. Unlike in Table 2, in Table 3 the RES values refer to all the measured data.

Table 2. SKN1: predicted (*R_{calc}*) and true values (*R_{obs}*) of normalized amplitudes in the sin θ/λ range 0.62-0.63 Å⁻¹

<i>R_{calc}</i>	<i>R_{obs}</i>	<i>R_{calc}</i>	<i>R_{obs}</i>	<i>R_{calc}</i>	<i>R_{obs}</i>
2.13	1.75	0.99	1.00	0.55	0.50
2.12	1.57	0.95	0.87	0.53	0.33
1.96	1.25	0.93	0.50	0.52	0.70
1.91	1.57	0.89	0.16	0.52	0.79
1.73	1.58	0.85	0.68	0.49	0.40
1.64	1.17	0.78	1.02	0.44	0.11
1.56	1.05	0.78	1.25	0.44	0.63
1.52	1.05	0.78	0.18	0.43	0.28
1.41	1.11	0.76	0.65	0.40	0.67
1.33	1.10	0.75	0.60	0.37	0.67
1.26	0.67	0.71	0.56	0.37	0.51
1.21	0.55	0.70	0.63	0.34	0.80
1.16	0.93	0.69	1.09	0.34	0.71
1.15	0.59	0.67	1.07	0.30	0.21
1.15	0.22	0.67	0.45	0.29	0.33
1.14	0.46	0.65	0.64	0.27	0.41
1.14	1.14	0.62	0.38	0.26	0.25
1.09	0.83	0.57	0.25	0.00	0.10
1.08	0.56	0.57	0.48		
1.05	1.19	0.55	0.56		

Table 3. For each test structure the (sin θ)/λ upper limit and the residual values according to Wilson's estimates (RES_w) and to equations (11) (RES_n) are given

Code	[(sin θ)/λ] _{max} (Å ⁻¹)	RES _w	RES _n
GEN1	0.70	0.40	0.14
GEN2	0.70	0.43	0.18
PGE2	0.57	0.46	0.43
SKN1	0.63	0.41	0.30
LOGANIN	0.67	0.48	0.42
FEGAS	0.81	0.74	0.54
SALEX	0.59	0.45	0.34
PIC	0.43	0.49	0.32
TIPORF	0.44	0.60	0.46

Analysis of Table 3 shows that:

- (11) is always better than Wilson's estimate;
- The efficiency of (11) is quite good for model structures, where data are without errors and resolution is high. The presence of heavy atoms does not remarkably disturb the efficiency of the formula.
- The efficiency declines with real data. Structural complexity and low resolution of data also weaken the effectiveness of the formula.
- The presence of structural regularities disturbs the efficiency of the formulas. FEGAS is affected by a strong pseudotranslational symmetry (the pseudo-translation vector is *u* = 1/3*c*; the estimated percentage of electrons suffering from pseudotranslational symmetry is 43%). The high value of RES_w is a consequence of the large dispersion of the |*E*|'s around the average value (hypercentric distribution). The application of (11) produces a great but not completely satisfactory improvement.

In order to obtain further insight into the behaviour of the residual RES_n, for each test structure we have arranged the reflections in decreasing order of *R_{calc}*. The values of RES_n are then calculated as a function

of N_r , where N_r is the number of reflections with R_{calc} larger than a given threshold. In Fig. 1 the trend of RES_n is shown for the nine test structures: in accordance with our observations in Table 2, RES_n generally increases with N_r .

The above considerations suggest that estimating non-measured diffraction magnitudes is rather complicated. The statistical relationships which can be used are of order N^{-1} (and therefore rather weak)

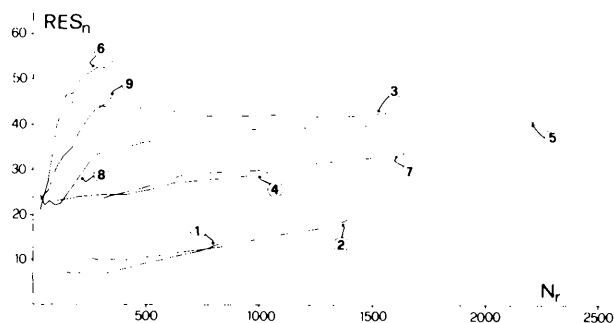


Fig. 1. RES_n is plotted as a function of N_r , where N_r is the number of reflections with R_{calc} larger than a given threshold. The curves correspond to the following test structures: (1) GEN1; (2) GEN2; (3) PGE2; (4) SKN1; (5) LOGANIN; (6) FEGAS; (7) SALEX; (8) PIC; (9) TIPORF.

and do not offer a satisfactory solution of the problem, unless some supplementary structural information is available.

The initial contribution by L. Favia is kindly acknowledged.

References

- CASCARANO, G., DOGGUY-SMIRI, L. & NGUYEN-HUY DUNG (1987). *Acta Cryst.* C43, 2050-2053.
 CHRISTENSEN, A. N., GRAND, A., LEHMANN, M. S. & COX, D. E. (1990). *Acta Chem. Scand.* 44, 103.
 CHRISTENSEN, A. N. & LEHMANN, M. S. (1989). *Acta Chem. Scand.* 43, 19-25.
 COCHRAN, W. (1955). *Acta Cryst.* 8, 473-478.
 DAVID, W. I. F. (1987). *J. Appl. Cryst.* 20, 316-319.
 DETITTA, G. T., LANGS, D. A., EDMONDS, J. W. & DUAX, W. L. (1980). *Acta Cryst.* B36, 638-645.
 GIACOVAZZO, C. (1976). *Acta Cryst.* A32, 91-99.
 HAUPTMAN, H. & KARLE, J. (1953). *The Solution of the Phase Problem. I. The Centrosymmetric Crystal*. Am. Crystallogr. Assoc. Monogr. No 3. New York: Polycrystal Book Service.
 JONES, P. G., SHELDRIK, G. M., GLÜSENKAMP, K.-H. & TIETZE, L. F. (1980). *Acta Cryst.* B36, 481-483.
 NAYA, S., NITTA, I. & ODA, T. (1965). *Acta Cryst.* 19, 734-747.
 SAYRE, W. (1952). *Acta Cryst.* 5, 60-65.
 SCORDARI, F. & STASI, F. (1990). *Z. Kristallogr.* 190, 47-62.
 VAN DER PUTTEN, N., SCHENK, H. & TSOUCARIS, G. (1982). *Acta Cryst.* A38, 98-102.

Acta Cryst. (1991). A47, 484-490

TELS: Least-Squares Solution of the Structure-Invariant Equations

BY FUSEN HAN

Upjohn Laboratories, The Upjohn Company, Kalamazoo, MI 49001, USA

AND GEORGE DETITTA AND HERBERT HAUPTMAN

Medical Foundation of Buffalo, Inc., 73 High Street, Buffalo, NY 14203, USA

(Received 21 June 1990; accepted 6 March 1991)

Abstract

Procedures are described to extract the values of individual phases from estimated structure invariants. The linear-equation and least-squares methods are used as two separate techniques in these procedures. The linear-equation method uses a linearly independent set of equations, with arbitrarily assigned integers, which are sufficient in number to solve for values of an equal number of phases. The least-squares method uses a set of overdetermined equations in which an 'integer problem' has to be considered. The assembly of these two techniques with a novel integer-trial-and-error method shows a remarkable ability to overcome the 'integer problem'. As a test of the whole procedure, phases were extrac-

ted from three-phase structure invariants calculated from the theoretical data for the platinum chloride derivative of cytochrome C_{550} .

Introduction

The structure invariant continues to play a central role in the direct-methods approach to the phase problem. Over the past few years, several investigators have attempted to derive new probabilistic formulae to improve structure-invariant estimations (Hauptman, 1982; Giacovazzo, 1983). Some of these formulae, for example the formula for anomalous-scattering data, yield unique estimates for the structure invariants themselves, as opposed to their cosines.