3. Concluding remarks

The probabilistic approaches for the estimation of quartet invariants in $P1$ and $P\bar{1}$ by Hauptman and by Giacovazzo have been further developed to derive the joint probability distribution function (including terms up to order N^{-1}) of n phases given $p \geq n$ moduli. The formulae (14) and (21) are obtained.

In (14) the weights w and w' suggest that a cross term of a quartet with large known modulus but unknown phase provides a positive contribution to Q [see (15)], while a negative contribution is provided by a cross term when both its modulus and its phase are known. According to (21), a cross term with known large modulus will provide a positive contribution to Q'' no matter whether the corresponding phase is known or not. Such behaviour should have striking consequences in practical applications. Indeed, as long as the largest $|E|$ values are phased during the phasing process, (14) and (21) will use such information in different ways. In particular, the positivity of the quartet term is expected to decrease in (14) and increase in (21).

In conclusion, while the approaches of Hauptman and Giacovazzo produce nearly equivalent quartet estimates (Giacovazzo, Camalli & Spagna, 1989) when only moduli are *a priori* known, the two formalisms lead to quite different estimates when applied to a situation in which a large number of phases are also known. This unexpected result will prove of large practical interest, as shown in the paper by Burla, Cascarano & Giacovazzo (1992), and suggests that the probabilistic quartet theory, as formulated so far, is not completely satisfactory. Indeed, the different mathematical approximations involved in the approaches of Hauptman and Giacovazzo are far from being insignificant if they cause such striking differences.

References

BURLA, M. C., CASCARANO, G. & GIACOVAZZO, C. (1992). *Acta Cryst.* A**48**, 906-912.

COCHRAN, W. (1955). *Acta Cryst.* **8**, 473-478.

GIACOVAZZO, C. (1975). *Acta Cryst.* A**31**, 252-259.

GIACOVAZZO, C. (1976). *Acta Cryst.* A**32**, 91-99.

GIACOVAZZO, C. (1977*a*). *Acta Cryst.* A**33**, 933-944.

GIACOVAZZO, C. (1977*b*). *Acta Cryst.* A**33**, 50-54.

GIACOVAZZO, C. (1980*a*). *Acta Cryst.* A**36**, 362-372.

GIACOVAZZO, C. (1980*b*). *Direct Methods in Crystallography.* London: Academic Press.

GIACOVAZZO, C., CAMALLI, M. & SPAGNA, R. (1989). *Acta Cryst.* A**45**, 141-143.

GREEN, E. A. & HAUPTMAN, H. (1976). *Acta Cryst.* A**32**, 43-45.

HAUPTMAN, H. (1975). *Acta Cryst.* A**31**, 680-687.

HAUPTMAN, H. (1978). *Acta Cryst.* A**34**, 525-528.

HAUPTMAN, H. & KARLE, J. (1953). *The Solution of the Phase Problem* I. *The Centrosymmetric Crystal*, ACA Monograph No. 3. New York: Polycrystal Book Service.

HEINERMAN, J. J. L. (1977). Thesis, Univ. of Utrecht, The Netherlands.

KLUG, A. (1958). *Acta Cryst.* **11**, 515-543.

Acta Cryst. (1992). A**48**, 906-912

The Joint Probability Distribution of Any Set of Phases Given Any Set of Diffraction Magnitudes. II. Practical Applications

BY M. C. BURLA

Dipartimento di Scienze della Terra, Università, 06100 *Perugia, Italy*

AND G. CASCARANO AND C. GIACOVAZZO

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

(*Received* 13 *December* 1992; *accepted* 6 *May* 1992)

Abstract

In the first paper in this series [Giacovazzo, Burla & Cascarano (1992). *Acta Cryst.* A**48**, 901-906], the conditional joint probability distribution function of n phases given $p \geq n$ moduli was derived. The properties of the concluding formulae are checked here. It is found that the distribution is not maximized by the correct phases, mostly because of bias in the formulae. If the triplets are estimated *via* the $P10$ formula [Cascarano, Giacovazzo, Camalli, Spagna, Burla, Nunzi & Polidori (1984). *Acta Cryst.* A**40**, 278-283] instead of being estimated by the Cochran relationship [Cochran (1952). *Acta Cryst.* **5**, 65-67], the situation is remarkably improved but further improvements are needed. A practical procedure is also described that successfully uses phase relationships to solve difficult structures.

0108-7673/92/060906-07$06.00

Symbols and notation

We will adopt the same symbols and notation used in the first paper in this series (Giacovazzo, Burla & Cascarano, 1992) (hereafter referred to as paper I).

1. Introduction

In paper I, the joint probability distribution of n phases given $p \geq n$ moduli was derived. Two formulae were obtained, both including terms of order up to N^{-1}. The first [(I.14) *i.e.* equation (14) of paper I] was derived by following the Hauptman approach. For ease of reference it is quoted here:

$$P(\varphi_1,\ldots,\varphi_n|R_1,\ldots,R_p) \simeq (1/L)\exp\left(\sum_{\text{triplets}} T_{ijl}\cos t_{ijl}+\sum_{\text{quartets}} Q_{ijlm}\cos q_{ijlm}\right), \quad (1)$$

where

$$Q_{ijlm}=B_{ijlm}[w+(w'_5R_5^2-w_5)+(w'_6R_6^2-w_6)+(w'_7R_7^2-w_7)].$$

$w=1$, $w_\nu=0$ always except when the cross magnitude R_ν is *a priori* known (in this case, $w_\nu=1$) and $w'_\nu=0$ always except when R_ν is *a priori* known and φ_ν is unknown (in this case, $w'_\nu=1$).

The second formula [(I.21)] was derived by following Giacovazzo's approach and is

$$P(\varphi_1,\ldots,\varphi_n|R_1,\ldots,R_p) \simeq (1/L)\exp\left(\sum_{\text{triplets}} T_{ijl}\cos t_{ijl}+\sum_{\text{quartets}} Q''_{ijlm}\cos q_{ijlm}\right), \quad (2)$$

where

$$Q''_{ijlm}=B_{ijlm}(w+w_5\varepsilon_5+w_6\varepsilon_6+w_7\varepsilon_7)/(1+Z_{ijlm})^{-1},$$
$$Z_{ijlm}=[(\varepsilon_1\varepsilon_2+\varepsilon_3\varepsilon_4)w_5\varepsilon_5+(\varepsilon_1\varepsilon_3+\varepsilon_2\varepsilon_4)w_6\varepsilon_6+(\varepsilon_1\varepsilon_4+\varepsilon_2\varepsilon_3)w_7\varepsilon_7]/2N.$$

In paper I it was emphasized that the two formulae diverge when a sufficiently large number of phases are involved in the distributions. The first aim of this paper is to study the main features of (1) and (2) and to check their potential usefulness in practical applications.

2. A practical procedure

To check the features of (1) and (2) we should first extend the preceding theoretical results to space groups with symmetry higher than triclinic.

A rigorous calculation of the triplet contribution to the formulae should take into account the complete first representation (Giacovazzo, 1977) of each triplet to discover and treat suitably the so-called symmetry-inconsistent three-phase invariants (Giacovazzo, 1974; Han & Langs, 1988; Giacovazzo, 1989; Burla & Giacovazzo, 1991). In the space groups handled in this paper, the percentage of symmetry-inconsistent triplets is negligible. Therefore, in our tests no special computing effort is made for selecting and properly evaluating the distribution of such triplets.

Quartet estimates are markedly affected by the symmetry. It was shown by Giacovazzo (1976) that a quartet can have more than three cross reflections. For example,

$$\Phi=\varphi_{\mathbf{h}_1}+\varphi_{\mathbf{h}_2}+\varphi_{\mathbf{h}_3}+\varphi_{\mathbf{h}_4} \qquad \mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3+\mathbf{h}_4=0$$

and

$$\Phi'=\varphi_{\mathbf{h}_1}+\varphi_{\mathbf{h}_2\mathbf{R}_s}+\varphi_{\mathbf{h}_3\mathbf{R}_s}+\varphi_{\mathbf{h}_4} \quad \mathbf{h}_1+\mathbf{h}_2\mathbf{R}_s+\mathbf{h}_3\mathbf{R}_s+\mathbf{h}_4=0$$

are symmetry equivalent [$\mathbf{R}_s$ is the rotation component of the sth symmetry operator $\mathbf{C}_s\equiv(\mathbf{R}_s,\mathbf{T}_s)$]. Accordingly, there are five cross vectors of Φ:

$$\mathbf{h}_1+\mathbf{h}_2,\quad \mathbf{h}_1+\mathbf{h}_3,\quad \mathbf{h}_2+\mathbf{h}_3,$$
$$\mathbf{h}_1+\mathbf{h}_2\mathbf{R}_s,\quad \mathbf{h}_1+\mathbf{h}_3\mathbf{R}_s.$$

The theory of representations is able to treat all such symmetry aspects. Accordingly, in our calculations the complete contribution arising from all the cross terms is considered for each quartet. Thus, each quartet contribution in (1) or (2) takes into account the complete first representation of the quartet (Giacovazzo, 1980*a*). Again, as in the triplet case, symmetry-inconsistent quartets are not considered.

The search for triplet and quartet invariants was made by means of a modified version of *SIR*88 (Burla, Camalli, Cascarano, Giacovazzo, Polidori, Spagna & Viterbo, 1989). To save CPU time all the reflections were stored in the central memory of the computer. Quartets were calculated not as the difference of two triplets but directly by seeking four vectorial indices that summed to zero. Therefore, the set of quartets contained both positive and negative estimated quartets.

3. Maximization of $P(\varphi_1,\ldots,\varphi_n|R_1,\ldots,R_p)$

An important feature of a well behaved conditional probability distribution function $P(\varphi_1,\ldots,\varphi_n|R_1,\ldots,R_p)$ should be

P = maximum for the correct set of phases.

In other words, the value of P calculated for the correct set should be larger than for any of the infinite sets of phases generated by an unbiased source of random phase values. In this case, two important results ensue: (*a*) P could be used as a powerful figure of merit in a multisolution process; (*b*) the maximization of P could be the strategy for the phase-expansion process. In other words, as a phasing criterion we could associate with the reflection $\mathbf{h}$ the phase $\varphi_{\mathbf{h}}$ that maximizes P.

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

Structure code	Space group	Molecular formula	Z
APAPA[1]	$P4_12_12$	$C_{30}H_{37}N_{15}O_{16}P_2.6H_2O$	8
AZET[2]	$Pca2_1$	$C_{21}H_{16}ClNO$	8
CEPHAL[3]	$C2$	$C_{18}H_{21}NO_4$	8
ERGO[4]	$P2_12_12_1$	$C_{28}H_{44}O$	8
GRA4[5]	$P\bar{1}$	$C_{30}H_{22}N_2O_4$	2
INOS[6]	$P2_1/n$	$C_6H_{12}O_6.H_2O$	8
LOGANIN[7]	$P2_12_12_1$	$C_{17}H_{26}O_{10}$	4
MGHEX[8]	$P3_1$	$C_{48}H_{68}MgN_{12}O_{12}.2ClO_4.4CH_3CN$	3
MUNICH[9]	$C2$	$C_{20}H_{16}$	8
NEWQB[10]	$P\bar{1}$	$C_{24}H_{20}N_2O_5$	4
POCRO[11]	$B112/m$	Cr_5KSe_8	2
SCHWARZ[12]	$P1$	$C_{46}H_{70}O_{27}$	1
TPALA[13]	$P2_1$	$C_{28}H_{42}N_4O_7$	2
TPH[14]	$B22_12$	$C_{24}H_{20}N_2$	12
WINTER[15]	$P2_1$	$C_{52}H_{83}N_{11}O_{16}.3CH_2Cl_2$	2

References: (1) Suck, Manor & Saenger (1976); (2) Colens, Declercq, Germain, Putzeys & Van Meerssche (1974); (3) Arora, Bates, Grady, Germain, Declercq & Powell (1976); (4) Hull, Leban, Main, White & Woolfson (1976); (5) unpublished; (6) D. A. Langs, H. C. Freeman, C. E. Nockolds & Y. L. Oh, unpublished; (7) Jones, Sheldrick, Glüsenkamp & Tietze (1980); (8) Karle & Karle (1981); (9) Szeimies-Seebach, Harnisch, Szeimies, Van Meerssche, Germain & Declercq (1978); (10) Grigg, Kemp, Sheldrick & Trotter (1978); (11) Nguyen-Huy Dung, Vo-Van Tien, Behm & Beurskens (1987); (12) B. Schweizer, unpublished; (13) G. D. Smith (Medical Foundation of Buffalo), unpublished; (14) Hoekstra, Vos, Braun & Hornstra (1975); (15) Butters, Hütter, Jung, Pauls, Schmitt, Sheldrick & Winter (1981).

To check if (1) and (2) can be considered to be well behaved distribution functions we selected the test structures given in Table 1. We found the following results.

(*a*) The maximum of $P(\varphi_1, \ldots, \varphi_n | R_1, \ldots, R_n)$, when this is expanded only up to order $N^{1/2}$, is obtained when

$$S = \sum_{\text{triplets}} T_{ijl} \cos t_{ijl} = \max. \qquad (3)$$

Equation (3) coincides with Cochran's (1952) relationship

$$S = \int_V \rho^3(\mathbf{r})\, d\mathbf{r} = \max.$$

Stanley (1979, 1986) suggested that the maximization of S could be used for expanding and refining phase information. However, Altomare, Cascarano & Giacovazzo (1992) showed that such a criterion cannot work for complex structures.

(*b*) The functions (1) and (2) will be maximum when

$$S' = \sum_{\text{triplets}} T_{ijl} \cos t_{ijl} + \sum_{\text{quartets}} Q_{ijlm} \cos q_{ijlm} = \max \qquad (4)$$

and

$$S'' = \sum_{\text{triplets}} T_{ijl} \cos t_{ijl} + \sum_{\text{quartets}} Q''_{ijlm} \cos q_{ijlm} = \max, \qquad (5)$$

respectively. We must check if the quartet contribution makes S' or S'' a more useful figure than S.

Table 2. *CEPHAL results*

For each value of n we give the number of triplets (NTRIP) and of quartets (NQUAR) found among the n phased reflections and the values of S, S' and S'' relative to the correct structure, to the Patterson solution (in parentheses) and to a random solution (in square brackets).

n	NTRIP	NQUAR	S	S'	S''
50	6	298	14.4	212.8	136.4
			(16.7)	(358.1)	(231.2)
			[5.4]	[5.3]	[2.5]
100	113	4631	177.1	1276	1253
			(249.8)	(2684)	(2637)
			[2.7]	[−15]	[19]
200	836	76092	927	5641	10844
			(1393)	(15939)	(27063)
			[20]	[136]	[173]
334	3751	645213	2768	10742	49344
			(4754)	(37945)	(148354)
			[11]	[−10]	[313]

(*c*) The property that S and/or S' and/or S'' are expected to be maxima for the correct solution holds asymptotically: it is probably satisfied when large numbers of phases and moduli are involved in the distributions. While p can always be assumed to coincide with the total number of measured reflections (*i.e.* the total prior information), the choice of n is not univocal (n is continuously varying during the phasing process). Very efficient probability distributions will be those for which (3), (4) or (5) are satisfied by the correct solution and for relatively small values of n.

(*d*) Symmorphic space groups are of particular interest for our tests. Indeed, in such space groups the so called 'Patterson solution' (where all the t_{ijl} and q_{ijlm} are zero) is frequently encountered among the trials produced in a multisolution process. Even if the Patterson solution is not the required solution it always corresponds, by definition, to the maximum value of S. Thus the effectiveness of the criteria (3), (4) and (5) can easily be estimated by comparing in symmorphic space groups the values S, S' and S'' of the true solution with those of the Patterson solution. A supplementary test on a random set of phases (random solution) will show how the figures corresponding to a random set differ from those of the organized (true and Patterson) solutions.

In accordance with points (*a*)-(*d*) we selected from Table 1 the structures CEPHAL, MUNICH and SCHWARZ, which all crystallize in symmorphic space groups. In Tables 2 to 4, the values of S, S' and S'' corresponding to the correct structure, to the Patterson solution and to a random solution are shown for each structure and for different values of n. To study the asymptotic behaviour of (1) and (2) we used different values of n. The maximum value of n in Tables 2 to 4 is NREF, the number of active reflections chosen by *SIR*88 in the phasing process. For each value of n, the numbers of triplets (NTRIP) and quartets (NQUAR) are given. Values of S, S' and

Table 3. *MUNICH results*

See caption of Table 2 for explanation.

n	NTRIP	NQUAR	S	S′	S″
50	34	317	85.3	294	273
			(100.2)	(494)	(434)
			[−15.8]	[−26]	[−18]
150	535	27964	723	3159	7076
			(1053)	(6831)	(15767)
			[−63]	[28]	[−72]
310	3704	509781	2864	2860	50449
			(4810)	(6486)	(140937)
			[−45]	[218]	[151]

Table 4. *SCHWARZ results*

See caption of Table 2 for explanation.

n	NTRIP	NQUAR	S	S′	S″
50	22	192	110	559	355
			(118)	(710)	(433)
			[−27]	[9.7]	[21]
250	917	57414	1920	16446	27333
			(2417)	(33327)	(48734)
			[−32]	[123]	[367]
470	4556	606573	6002	24615	148620
			(8551)	(74364)	(312052)
			[−105]	[208]	[−92]

Table 5. *The values of the quartet term in* (4) *and* (5) *for CEPHAL, MUNICH and SCHWARZ at different values of n*

	CEPHAL		MUNICH		SCHWARZ	
n	SQ	SQ″	SQ	SQ″	SQ	SQ″
50	198	122	208	188	449	245
150	2964	4024	2435	6352	5715	5609
250	6630	19467	2300	27614	14526	25413
300	8323	33958	782	43519	17107	42100
500	−10425	140568	−29555	155110	16874	165631

S'' are given for the correct solution, for the Patterson solution (in parentheses) and for a random solution (in square brackets). The following points are easily seen.

(*a*) Equation (3) is in general not satisfied for the correct structure. For any value of n, S is always maximum for the Patterson solution.

(*b*) Even if the average reliability parameter of the quartets is small, their large number makes their contribution non-negligible and often dominant. Thus, their unbiased estimate is of basic importance for the success of criteria (4) and (5).

(*c*) When n is large, the contribution of the quartets is greatly underestimated in (1) and overestimated in (2). We have listed in Table 5 the quartet contributions to (1) and to (2) [as SQ and SQ″ respectively] for the correct structures and for different values of n. SQ decreases when n is above a certain threshold: then all the large cross reflections are phased and they give no contribution to SQ. In some cases, SQ becomes negative, which is unacceptable for a well behaved distribution.

Table 6. *The contribution to SQ″ arising from the negative estimated quartets (SQN″) only*

The contributions are given for the correct structure and, in parentheses, for the Patterson solution.

	CEPHAL		MUNICH		SCHWARZ	
n	SQ″	SQN″	SQ″	SQN″	SQ″	SQN″
500	140568	4021	155110	4666	165631	1946
	(517477)	(−56536)	(590584)	(−67642)	(359618)	(−19029)

(*d*) In addition to (3), the criteria (4) and (5) are not satisfied by the correct structure (things get worse!). The asymptotic properties of (1) are worse than those of (2) but both seem to have some bias.

From the points (*a*)-(*d*), the following questions arise: is the Patterson solution the 'natural' set of phases and moduli to maximize S, S' and S''? Does some information exist that is not well exploited by (1) and (2) and that can make S' and/or S'' maxima for the true structures? A possible answer to the above questions is not too difficult: such information is contained in the negative triplet and quartet invariants. Distributions able to recognize and estimate them closely should also make (4) and (5) true for the correct structure. In Table 6, we emphasize, for the correct and for the Patterson structures, the contribution to SQ″ arising from the negative estimated quartets (SQN″) only. It is immediately seen that SQN″ tends to maximize S'' for the correct structure and simultaneously to reduce the value of S'' for the Patterson solution. However, its contribution is not relevant owing to the fact that the reliability parameter (*i.e.* Q''_{ijlm}) for negative quartets is rather small in modulus. More accurate expressions for Q''_{ijlm} will produce more accurate distributions.

The same observation may be applied to the triplet invariants. We could replace in (2) the Cochran reliability parameter $T_{ijl} = 2|E_iE_jE_l|/N^{1/2}$ by the reliability parameter offered by the $P10$ formula (Cascarano, Giacovazzo, Camalli, Spagna, Burla, Nunzi & Polidori, 1984). This parameter is able to recognize in favourable cases most of the negative triplets and is thus able in principle to improve the efficiency of the criteria (3), (4) and (5). In Table 7 the values of S obtained by using the $P10$ formula are compared with those given in Tables 2 to 4. Clearly, the use of the $P10$ formula remarkably improves the efficiency of the distributions (1) and (2) but again the contribution arising from the negative triplets is not sufficient, for CEPHAL, MUNICH and SCHWARZ, to make S maximum for the correct structure. The above results, however, strongly recommend the use of the $P10$ formula.

A supplementary test has been made for POCRO (space group $B112/m$): owing to the small structural complexity, invariant estimates are more reliable. S, SQ″ and S'' were calculated for $n = \text{NREF} = 96$ by estimating triplets *via* the $P10$ formula. The results

Table 7. *A comparison of the values of S obtained by using the P10 formula with those given in Tables 2 to 4*

The *P*10 formula values were calculated with n = NREF. The corresponding values obtained by using Cochran's parameter (see Tables 2 to 4) are given in parentheses.

	n	Patterson solution	True structure
CEPHAL	334	4534 (4754)	3034 (2768)
MUNICH	310	4518 (4810)	3178 (2864)
SCHWARZ	470	8296 (8551)	6495 (6002)

Table 8. *Values of S, SQ″ and S″ for POCRO, CEPHAL, MUNICH and SCHWARZ*

Values estimated using the *P*10 formula with n = NREF = 96 are given together with the corresponding values for the Patterson solution (in parentheses).

	NTRIP	NQUAR	S	SQ″	S''
POCRO	1163	30001	3862 (3862)	37417 (35486)	41279 (39348)
CEPHAL	3121	652539	3034 (4534)	46577 (143600)	49611 (148134)
MUNICH	2839	517215	3178 (4518)	47584 (136128)	50763 (140646)
SCHWARZ	3825	606573	6495 (8296)	142618 (303501)	149113 (311798)

are summarized in Table 8; the data relating to the Patterson solution are given in parentheses. Now the higher accuracy of the probabilistic formulae makes S'' maximum for the correct solution. For the more complicated structures CEPHAL, MUNICH and SCHWARZ, S'' is not maximum for the correct solution (see Table 8).

4. Active use of the probability distribution functions

The distributions (1) and (2) can be modified by standard methods to obtain the conditional distribution $P(\varphi_1, \varphi_2, \ldots, \varphi_n | R_1, \ldots, R_p)$. Two tangent formulae can then be derived that are similar to those already obtained by Giacovazzo (1980*b*). The first derives from (1) and may be written as

$$\tan \varphi_{\mathbf{h}} \simeq \left[\sum_{\mathbf{k}} T_{\mathbf{h},\mathbf{k}} \sin(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}}) + \sum_{\mathbf{k},\mathbf{l}} Q_{\mathbf{h},\mathbf{k},\mathbf{l}} \sin(\varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} + \varphi_{\mathbf{h}-\mathbf{k}-\mathbf{l}})\right] \times \left[\sum_{\mathbf{k}} T_{\mathbf{h},\mathbf{k}} \cos(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}}) + \sum_{\mathbf{k},\mathbf{l}} Q_{\mathbf{h},\mathbf{k},\mathbf{l}} \cos(\varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} + \varphi_{\mathbf{h}-\mathbf{k}-\mathbf{l}})\right]^{-1} = A_{\mathbf{h}}/B_{\mathbf{h}} \quad (6)$$

with the reliability parameter given by $\alpha_{\mathbf{h}} = (A_{\mathbf{h}}^2 + B_{\mathbf{h}}^2)^{1/2}$. The second type of formula derives from (2) and leads to

$$\tan \varphi_{\mathbf{h}} \simeq \left[\sum_{\mathbf{k}} T_{\mathbf{h},\mathbf{k}} \sin(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}}) + \sum_{\mathbf{k},\mathbf{l}} Q''_{\mathbf{h},\mathbf{k},\mathbf{l}} \sin(\varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} + \varphi_{\mathbf{h}-\mathbf{k}-\mathbf{l}})\right] \times \left[\sum_{\mathbf{k}} T_{\mathbf{h},\mathbf{k}} \cos(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}}) + \sum_{\mathbf{k},\mathbf{l}} Q''_{\mathbf{h},\mathbf{k},\mathbf{l}} \cos(\varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} + \varphi_{\mathbf{h}-\mathbf{k}-\mathbf{l}})\right]^{-1} = A''_{\mathbf{h}}/B''_{\mathbf{h}} \quad (7)$$

with the reliability parameter given by $\alpha''_{\mathbf{h}} = (A''^2_{\mathbf{h}} + B''^2_{\mathbf{h}})^{1/2}$.

The following points should be noted.

(*a*) The use of (6) or (7) in a refinement process requires quite different procedures. Q'' in (7) depends on the cross magnitudes only: its value, once estimated, does not change during the phasing process. In contrast, Q in (6) depends on both the cross magnitudes and the cross phases. Its value has to be recalculated during the phasing process as soon as a cross reflection is phased (which is time consuming).

(*b*) The experimental results in the preceding section discourage any attempt to use (6). However, we wrote a special program to check its usefulness: we found (6) much less efficient than the classical tangent formula based on triplets only (much computing time was wasted). For brevity, the relative experimental results are not quoted.

(*c*) In the preceding section, the use of the *P*10 formula has been encouraged. Thus, from now on, when we refer to (7), we will assume that $T_{\mathbf{h},\mathbf{k}}$ corresponds to the Cochran or the *P*10 parameter according to circumstances.

In accordance with the points (*a*)-(*c*), we devised the following direct procedure.

(1) Triplet relationships are found among the NREF reflections. They can be estimated according to the Cochran relationship or the *P*10 formula.

(2) The convergence-divergence program of *SIR*88 is run to select among NRAND ≤ NREF reflections the origin- and enantiomorph-fixing reflections. Random phase values (with the symmetry restricted when necessary) are associated with the NRAND reflections (Baggio, Woolfson, Declercq & Germain, 1978), with unit weight for the origin- and enantiomorph-fixing reflections and with a weight of 0.8 for the others.

(3) Quartet relationships can eventually be found and estimated among the NRAND reflections. In this case, the user has to take care that NRAND does not exceed a certain threshold (see Tables 2–4) otherwise

Table 9. *Success and failure for the phasing process according to various protocols*

Protocol 1: only triplets estimated by the Cochran formula are used; Protocol 2: only triplets estimated by the *P*10 formula are used; Protocol 3: triplets estimated by Cochran and quartets are simultaneously used; Protocol 4: *P*10 triplets and quartets are used. NOT: number of trials. *Y* denotes success of the phasing process, *N* denotes failure.

Structure code	NOT	Protocol 1	Protocol 2	Protocol 3	Protocol 4
APAPA	300	*Y*	*Y*	*Y*	*Y*
AZET	200	*N*	*Y*	*N*	*Y*
CEPHAL	200	*N*	*Y*	*N*	*Y*
ERGO	400	*N*	*Y*	*N*	*Y*
GRA4	50	*N*	*Y*	*N*	*Y*
INOS	50	*Y*	*Y*	*Y*	*Y*
LOGANIN	100	*Y*	*Y*	*Y*	*Y*
MGHEX	400	*N*	*Y*	*N*	*Y*
MUNICH	400	*N*	*Y*	*N*	*N*
NEWQB	200	*N*	*Y*	*Y*	*Y*
POCRO	20	*Y*	*Y*	*Y*	*Y*
SCHWARZ	400	*N*	*Y*	*N*	*Y*
TPALA	400	*N*	*Y*	*N*	*N*
TPH	400	*Y*	*Y*	*Y*	*Y*
WINTER	400	*N*	*Y*	*N*	*N*

a very large set of quartets must be estimated (which is time consuming).

(5) Cycles of phase refinement are made on the NRAND phases. After convergence, the phasing process is extended to the NREF reflections.

We first checked the efficiency of the above-described phasing process by using only triplet relationships. NRAND ≃ NREF/2 proved to be a reasonable choice: Cochran estimates were first used (protocol 1) and then *P*10 estimates (protocol 2). The results are shown in Table 9. The following points should be noted.

(*a*) Large numbers of trials (NOT) were made for most of the structures to emphasize the different efficiencies of the Cochran and P10 estimates. In fact, smaller numbers of trials also allow the structure solution in most of the cases: *e.g.* the first correct solution for APAPA is at 11, for ERGO it is at 109, for MGHEX it is at 85 *etc.*

(*b*) The combination '*P*10 - random approach' is much more efficient than the combination 'Cochran - random approach'. This is probably due to the larger reliability of the *P*10 estimates. Accordingly, it may be concluded that the recovery of the correct set of phases from a random one is easier when reliable phase relationships are used.

To understand the role of the quartet invariants in the above phasing process: (*a*) we chose NRAND = 90 to limit the amount of computing time; (*b*) we constructed quartets by allowing the basis vectors to vary among the NRAND reflections; (*c*) formula (7) was applied for phase determination and refinement. The triplet reliability factor T was derived from the Cochran formula (protocol 3) and from *P*10 (protocol 4).

Table 10. *Results for TPALA showing that triplet (P10) and quartet relationships are arranged according to their reliability parameters*

T: triplets; *Q*″; quartets. NT and NQ are the numbers of triplet and quartet relationships with concentration parameters larger than given values of T and Q'', respectively. The numbers of wrongly estimated cosine signs are given in parentheses.

T, *Q*″	NT	NQ
0.8	2557 (263)	8509 (1518)
1.2	1960 (174)	5372 (790)
1.6	1429 (111)	2394 (331)
2.0	1027 (68)	639 (84)
2.6	579 (25)	
3.8	166 (6)	

The results are shown in Table 9. We observe that the use of quartets allows the additional solution of NEWQB (unsolved in protocol 1). In contrast, the combined use of *P*10-estimated triplets and quartets is not able to solve MUNICH, TPALA and WINTER, which are solved by using P10 triplets only. In conclusion, the additional use of quartet relationships (their number is generally comparable with that of the triplets) proved to be less advantageous than expected. This loss of efficiency is only partially due to the correlation between triplet and quartet information (*i.e.* a quartet is the sum of two triplets). The main reason is probably the lower reliability of the quartet estimates (first representation formula) with respect to the triplet estimates (first and second representation formulae). As an example, in Table 10 we show how triplets and quartets, as calculated for TPALA, are arranged in decreasing order of their concentration parameters T and Q'', respectively.

A double-check of our conclusions may be obtained by eliminating from the set of active quartets those that are wrongly estimated (positive estimated quartets that are actually negative and *vice versa*). Under these conditions, the application of (7) is successful.

5. Concluding remarks

In paper I, two expressions for the conditional joint probability distribution function of n phases given $p > n$ moduli were derived. Both the distributions include triplet and quartet invariant contributions: the first distribution may be considered a development of Hauptman's mathematical approach, the second of Giacovazzo's approach. In this paper, we have checked some properties of the distributions and their usefulness for phase solution. We found that both distributions are not maximized by the correct phases, as one would expect for a well behaved distribution. This may be ascribed to the limited accuracy of the probabilistic formulae estimating triplet and quartet invariants and to some insufficiency in the mathematics used by Hauptman and

Giacovazzo. Replacing in the distributions the Cochran concentration parameter of a triplet by the corresponding $P10$ parameter of the same triplet remarkably improves the behaviour of the distributions although further improvements are needed. Their maximization requires better probabilistic theories; in particular, higher efficiency for the estimation of negative quartet invariants.

The tangent formula (including triplets and quartet contributions) based on the mathematical approach of Giacovazzo proved to be the only one suitable for phase expansion and refinement. The formula was included in a random approach to structure determination. The additional use of quartets was not helpful owing to the limited accuracy of quartet estimates. Replacing in the triplet contribution the Cochran concentration parameter by the corresponding $P10$ parameter remarkably improved the efficiency of the phasing process. But again the combination of $P10$ estimated triplets with quartets proved less efficient. The reason for such a failure is ascribed to the limited accuracy of the probabilistic formulae estimating quartets. A substantial improvement of such formulae is considered a necessary condition for the success of the active use of the quartets in the phasing process.

The authors thank C. Chiarella for technical contribution.

References

ALTOMARE, A., CASCARANO, G. & GIACOVAZZO, C. (1992). *Acta Cryst.* **A48**, 30-36.

ARORA, S. A., BATES, R. B., GRADY, R. A., GERMAIN, G., DECLERCQ, J. P. & POWELL, R. G. (1976). *J. Org. Chem.* **41**, 551-554.

BAGGIO, R., WOOLFSON, M. M., DECLERQ, J. P. & GERMAIN, G. (1978). *Acta Cryst.* **A34**, 883-892.

BURLA, M. C., CAMALLI, M., CASCARANO, G., GIACOVAZZO, C., POLIDORI, G., SPAGNA, R. & VITERBO, D. (1989). *J. Appl. Cryst.* **22**, 389-393.

BURLA, M. C. & GIACOVAZZO, C. (1991). *Acta Cryst.* **A47**, 346-352.

BUTTERS, P., HÜTTER, P., JUNG, G., PAULS, N., SCHMITT, H., SHELDRICK, G. M. & WINTER, W. (1981). *Angew. Chem.* **93**, 904-905.

CASCARANO, G., GIACOVAZZO, C., CAMALLI, M., SPAGNA, R., BURLA, M. C., NUNZI, A. & POLIDORI, G. (1984). *Acta Cryst.* **A40**, 278-283.

COCHRAN, W. (1952). *Acta Cryst.* **5**, 65-67.

COLENS, A., DECLERCQ, J. P., GERMAIN, G., PUTZEYS, J. P. & VAN MEERSSCHE, M. (1974). *Cryst. Struct. Commun.* **3**, 119-122.

GIACOVAZZO, C. (1974). *Acta Cryst.* **A30**, 626-630, 631-634.

GIACOVAZZO, C. (1976). *Acta Cryst.* **A32**, 958-966.

GIACOVAZZO, C. (1977). *Acta Cryst.* **A33**, 527-531, 933-944.

GIACOVAZZO, C. (1980*a*). *Direct Methods in Crystallography.* London: Academic Press.

GIACOVAZZO, C. (1980*b*). *Acta Cryst.* **A36**, 74-82.

GIACOVAZZO, C. (1989). *Acta Cryst.* **A45**, 534-538.

GIACOVAZZO, C., BURLA, M. C. & CASCARANO, G. (1992). *Acta Cryst.* **A48**, 901-906.

GRIGG, R., KEMP, J., SHELDRICK, G. M. & TROTTER, J. (1978). *J. Chem. Soc. Chem. Commun.* pp. 109-111.

HAN, F. & LANGS, D. A. (1988). *Acta Cryst.* **A44**, 563-566.

HOEKSTRA, A., VOS, A., BRAUN, P. B. & HORNSTRA, J. (1975). *Acta Cryst.* **B31**, 1708-1715.

HULL, S. E., LEBAN, I., MAIN, P., WHITE, P. S. & WOOLFSON, M. M. (1976). *Acta Cryst.* **B32**, 2374-2381.

JONES, P. G., SHELDRICK, G. M., GLÜSENKAMP, K.-H. & TIETZE, L. F. (1980). *Acta Cryst.* **B36**, 481-483.

KARLE, I. L. & KARLE, J. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 681-685.

NGUYEN-HUY DUNG, VO-VAN TIEN, BEHM, H. J. & BEURSKENS, P. T. (1987). *Acta Cryst.* **C43**, 2258-2260.

STANLEY, E. (1979). *Acta Cryst.* **A35**, 966-970.

STANLEY, E. (1986). *Acta Cryst.* **A42**, 297-299.

SUCK, D., MANOR, P. C. & SAENGER, W. (1976). *Acta Cryst.* **B32**, 1727-1737.

SZEIMIES-SEEBACH, U., HARNISCH, J., SZEIMIES, G., VAN MEERSSCHE, M., GERMAIN, G. & DECLERCQ, J. P. (1978). *Angew. Chem. Int. Ed. Engl.* **17**, 848-850.

Acta Cryst. (1992). A**48**, 912-916

Conjugate-Direction Minimization: an Improved Method for the Refinement of Macromolecules

BY D. E. TRONRUD

Howard Hughes Medical Institute and Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, *USA*

(*Received* 29 *August* 1991; *accepted* 13 *May* 1992)

Abstract

A novel method of function minimization that combines the power of the diagonal approximation to the normal matrix with conjugate directions is described. This method approaches closer to the local minimum than the methods that are commonly used in macromolecular refinement. The weaknesses of the current methods are analyzed to explain the advantage of the conjugate-direction method.

0108-7673/92/060912-05$06.00