Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 17 April 2000
Accepted 29 August 2000

The method of joint probability distribution functions applied to the one-wavelength anomalous-scattering (OAS) case

Carmelo Giacovazzo[a,b]* and Dritan Siliqi[b,c]

[a]IRMEC c/o Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy, [b]Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy, and [c]Laboratory of X-ray Diffraction, Department of Inorganic Chemistry, Faculty of Natural Sciences, Tirana, Albania. Correspondence e-mail: c.giacovazzo@area.ba.cnr.it

The method of the joint probability distribution function is applied to the case in which the positions of the anomalous scatterers are fully or partially known. The mathematical technique is able to handle errors both in the model structure of the located anomalous scatterers and in measurements. A criterion for ranking the more accurate phase estimates is given.

1. Notation

N: number of atoms in the unit cell.
a: number of anomalous scatterers in the unit cell.
la: number of located anomalous scatterers in the unit cell.
$nla = N - la$: nla includes non-located anomalous scatterers and non-anomalous scatterers.
$f_j = f_j^0 + \Delta f_j + if_j'' = f_j' + if_j''$: scattering factor of the jth atom. f' is its real, f'' is its imaginary part. The thermal factor is included.

$$F^+ = |F^+|\exp(i\varphi^+) = F_{\mathbf{h}} = \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{h}\mathbf{r}_j).$$

$$F_{la}^+ = |F_{la}^+|\exp(i\varphi_{la}^+) = \sum_{la} f_j \exp(2\pi i \mathbf{h}\mathbf{r}_j).$$

$$F^- = |F^-|\exp(i\varphi^-) = F_{-\mathbf{h}} = \sum_{j=1}^{N} f_j \exp(-2\pi i \mathbf{h}\mathbf{r}_j).$$

$$F_{la}^- = |F_{la}^-|\exp(i\varphi_{la}^-) = \sum_{la} f_j \exp(-2\pi i \mathbf{h}\mathbf{r}_j).$$

$|F|\exp(i\varphi)$: structure factor calculated by taking into account non-anomalous scattering (all the atoms in the unit cell included).
$\Sigma_{la}, \Sigma_{nla}, \Sigma_N = \sum (f_j'^2 + f_j''^2)$, where the summation is extended to la, nla and N atoms.
$\Delta_{\rm ano} = |F^+| - |F^-|$.

2. Introduction

The method of joint probability distribution functions has been already applied to the OAS (one-wavelength anomalous-scattering) case. The joint probability distributions

$$P(|F^+|, |F^-|, \varphi^+, \varphi^-) \qquad (1a)$$

and

$$P(|F_{\mathbf{h}_1}^+|, |F_{\mathbf{h}_2}^+|, |F_{\mathbf{h}_3}^+|, |F_{\mathbf{h}_1}^-|, |F_{\mathbf{h}_2}^-|, |F_{\mathbf{h}_3}^-|, \varphi_{\mathbf{h}_1}^+, \ldots, \varphi_{\mathbf{h}_3}^-) \text{ with } \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 = 0 \qquad (1b)$$

were independently obtained by Hauptman (1982) and by Giacovazzo (1983) in the case in which no prior information is available on the anomalous-scatterer positions. The approaches aimed at identifying: (i) from (1a) the conditional probability distribution $P(\Delta\varphi||F^+|, |F^-|)$ of $\Delta\varphi = \varphi^+ + \varphi^-$ given $|F^+|$ and $|F^-|$; (ii) from (1b) the conditional triplet phase distribution

$$P(\Phi||F_{\mathbf{h}_1}^+|, |F_{\mathbf{h}_2}^+|, |F_{\mathbf{h}_3}^+|, |F_{\mathbf{h}_1}^-|, |F_{\mathbf{h}_2}^-|, |F_{\mathbf{h}_3}^-|).$$

This last distribution should help to estimate phase values without any prior knowledge of the atomic positions of the anomalous scatterers.

In this paper, we intend to apply the joint probability distribution technique to the OAS case when the positions of all or a part of the anomalous scatterers have been found *via* one of the current methods (see Blow & Rossmann, 1961; North, 1965; Mathews, 1966; see also Giacovazzo, 1998, for a general description of them). Then the joint distributions

$$P(F^+, F^-|F_{la}^+, F_{la}^-) \qquad (2)$$

will be calculated, from which

$$P(\varphi^+||F^+|, |F^-|, |F_{la}^+|, |F_{la}^-|)$$

and

$$P(\varphi^-||F^+|, |F^-|, |F_{la}^+|, |F_{la}^-|)$$

will be derived. From them, the most probable value of φ is easily derived by geometrical considerations.

While (1a) and (1b) were derived without taking into account errors in measurements, the study of (2) cannot be made without using them. Accordingly, the mathematical approach we describe here involves errors both in the model structure of located anomalous scatterers and in measurements.

To evaluate the potential of the present approach, a comparison will be made between our conclusive formulas and the corresponding expressions obtained by previous authors. We notice that previous probabilistic approaches consider OAS as a special SIR (single isomorphous replacement) case. *I.e.* the classical Blow & Crick (1959) expression, integrated by Terwilliger & Eisenberg (1987) contributions and originally derived for SIR–MIR cases, has been extended by analogy to the OAS case. The result is

$$P(\varphi) \approx \exp[-\varepsilon(\varphi)/(2E^2)], \qquad (3)$$

where

$$\varepsilon = |\Delta_{\rm ano}^{\rm obs} - \Delta_{\rm ano}^{\rm calc}|,$$
$$E = \langle \varepsilon^2 \rangle + 4\sigma^2(\Delta_{\rm ano}),$$

$\sigma^2(\Delta_{\rm ano})$ takes measurement errors into account.

3. The joint probability distribution $P(F^+, F^- | F^+_{la}, F^-_{la})$

In our probabilistic approach, the positions of the non-located anomalous scatterers and the positions of the non-anomalous scatterers will be the primitive random variables. We will assume that

$$F^+ = F^+_{la} + F^+_{nla} + \mu^+ = F^+_{la} + F^+_q, \qquad (4)$$

where: (*a*) F^+_{nla} is the structure factor corresponding to the non-located anomalous scatterers and to the non-anomalous scatterers; (*b*) $\mu^+ = |\mu^+|\exp(i\theta^+)$ represents the cumulative errors arising from different sources (*i.e.* the structural model constituted by the located anomalous scatterers and errors in measurements); (*c*) $F^+_q = F^+_{nla} + \mu^+$.

Equivalently,

$$F^- = F^-_{la} + F^-_{nla} + \mu^- = F^-_{la} + F^-_q, \qquad (5)$$

where $F^-_q = F^-_{nla} + \mu^-$.

If all the anomalous scatterers are located, $F^+_{nla} = (F^-_{nla})^*$, where * indicates the complex conjugate. We make some reasonable assumptions:

(*a*) F^+_{la}, F^+_{nla}, μ^+ are uncorrelated with each other;

(*b*) the same assumption holds for F^-_{la}, F^-_{nla}, μ^-;

(*c*) $\langle \mu^+ \rangle = \langle \mu^- \rangle = 0$;

(*d*) $\langle \mu^+ \mu^- \rangle = 0$. This implies that errors on F^+ and F^- are uncorrelated.

Then,

$$\langle |F^+|^2 \rangle = |F^+_{la}|^2 + \Sigma_{nla} + \langle |\mu^+_\mu|^2 \rangle$$
$$\langle |F^-|^2 \rangle = |F^-_{la}|^2 + \Sigma_{nla} + \langle |\mu^-_\mu|^2 \rangle.$$

When part of the structure is known, numerical reasons (Camalli *et al.*, 1985) suggest that it is more useful to pseudo-normalize the structure factor with respect to the unknown part of the crystal structure. This is in agreement with the fact that atomic coordinates of the located atoms no longer belong to the set of primitive random variables. If structure factors are normalized with respect to the full chemical content of the crystal structure, the final formulas will not change but they will assume a more complicated mathematical form.

Accordingly, on supposing that part or all of the anomalous scatterers are located, we find that

$$R\exp(i\varphi^+) = (A^+ + iB^+) = F^+/\Sigma_{nla}^{1/2},$$
$$G\exp(i\varphi^-) = (A^- + iB^-) = F^-/\Sigma_{nla}^{1/2},$$

where R and G are the pseudo-normalized moduli of $|F^+|$ and $|F^-|$, respectively, and

$$A^+ = \left[\sum_{j=1}^{N}(f'_j\cos 2\pi\mathbf{hr}_j - f''_j\sin 2\pi\mathbf{hr}_j) + |\mu^+|\cos\theta^+\right]\Big/\Sigma_{nla}^{1/2}$$
$$B^+ = \left[\sum_{j=1}^{N}(f'_j\sin 2\pi\mathbf{hr}_j + f''_j\cos 2\pi\mathbf{hr}_j) + |\mu^+|\sin\theta^+\right]\Big/\Sigma_{nla}^{1/2}$$
$$A^- = \left[\sum_{j=1}^{N}(f'_j\cos 2\pi\mathbf{hr}_j + f''_j\sin 2\pi\mathbf{hr}_j) + |\mu^-|\cos\theta^-\right]\Big/\Sigma_{nla}^{1/2}$$
$$B^- = \left[\sum_{j=1}^{N}(-f'_j\sin 2\pi\mathbf{hr}_j + f''_j\cos 2\pi\mathbf{hr}_j) + |\mu^-|\sin\theta^-\right]\Big/\Sigma_{nla}^{1/2}.$$

Equivalently,

$$R_{la}\exp(i\varphi^+_{la}) = (A^+_{la} + iB^+_{la}) = F^+_{la}/\Sigma_{nla}^{1/2}$$
$$G_{la}\exp(i\varphi^-_{la}) = (A^-_{la} + iB^-_{la}) = F^-_{la}/\Sigma_{nla}^{1/2}$$
$$R_q\exp(i\varphi^+_q) = (A^+_q + iB^+_q) = F^+_q/\Sigma_{nla}^{1/2}$$
$$G_q\exp(i\varphi^-_q) = (A^-_q + iB^-_q) = F^-_q/\Sigma_{nla}^{1/2},$$

where

$$A^+_q = [\Re(F^+_{nla}) + |\mu^+|\cos\theta^+]/\Sigma_{nla}^{1/2}$$
$$B^+_q = [\Im(F^+_{nla}) + |\mu^+|\sin\theta^+]/\Sigma_{nla}^{1/2}$$
$$A^-_q = [\Re(F^-_{nla}) + |\mu^-|\cos\theta^-]/\Sigma_{nla}^{1/2}$$
$$B^-_q = [\Im(F^-_{nla}) + |\mu^-|\sin\theta^-]/\Sigma_{nla}^{1/2}.$$

$\Re(\ldots)$ and $\Im(\ldots)$ stand for real and imaginary parts, respectively.

Under the above assumptions, the characteristic function $C(u^+, u^-, v^+, v^-)$ of the distribution

$$P(A^+, A^-, B^+, B^- | A^+_{la}, A^-_{la}, B^+_{la}, B^-_{la})$$

[in short $P(A^+, A^-, B^+, B^-)$] may be calculated.

We have

$$\begin{aligned} C(u^+, u^-, v^+, v^-) &= \langle \exp i(u^+A^+ + u^-A^- + v^+B^+ + v^-B^-) \rangle \\ &= \exp i(u^+A^+_{la} + u^-A^-_{la} + v^+B^+_{la} + v^-B^-_{la}) \\ &\quad \times \langle \exp i(u^+A^+_q + u^-A^-_q + v^+B^+_q + v^-B^-_q) \rangle, \end{aligned} \qquad (6)$$

where u^+, u^-, v^+, v^- are carrying variables associated with A^+, A^-, B^+, B^-, respectively. Expanding the right-hand side of (6) in cumulants gives

$$\begin{aligned}
&C(u^+, u^-, v^+, v^-)\\
&\quad = \langle \exp i(u^+A^+_{la} + u^-A^-_{la} + v^+B^+_{la} + v^-B^-_{la})\rangle\\
&\qquad \times \exp\{-[e^+(u^{+2} + v^{+2}) + e^-(u^{-2} + v^{-2})]/4\\
&\qquad - \tfrac{1}{2}c'_1(u^+u^- - v^+v^-) - \tfrac{1}{2}c'_2(u^+v^- + u^-v^+)\},
\end{aligned}$$

where

$$\begin{gathered}
e^+ = (1 + \sigma_\mu^{+2}), \quad e^- = (1 + \sigma_\mu^{-2})\\
\sigma_\mu^{+^2} = \langle|\mu_\mu^+|^2\rangle/\Sigma_{nla}, \quad \sigma_\mu^{-^2} = \langle|\mu_\mu^-|^2\rangle/\Sigma_{nla}\\
c'_1 = \sum_{nla}(f_j'^2 - f_j''^2)\Big/\sum_{nla}(f_j'^2 + f_j''^2)\\
c'_2 = \left(2\sum_{nla} f'_j f''_j\right)\Big/\sum_{nla}(f_j'^2 + f_j''^2).
\end{gathered}$$

Then,

$$\begin{aligned}
&P(A^+, A^-, B^+, B^-)\\
&\quad \approx (2\pi)^{-4}\int\limits_{-\infty}^{+\infty}\ldots\int\limits_{-\infty}^{+\infty}\exp\{-i[u_1(A^+ - A^+_{la})\\
&\qquad + u_2(A^- - A^-_{la}) + v_1(B^+ - B^+_{la})\\
&\qquad + v_2(B^- - B^-_{la})] - [e^+(u^{+2} + v^{+2})\\
&\qquad + e^-(u^{-2} + v^{-2})]/4 - \tfrac{1}{2}c'_1(u^+u^- - v^+v^-)\\
&\qquad - \tfrac{1}{2}c'_2(u^+v^- + u^-v^+)\}\,\mathrm{d}u^+\ldots\mathrm{d}v^-.
\end{aligned}$$

Define

$$\begin{aligned}
u^+ &= (2/e^+)^{1/2}u^{+\prime}, \quad u^- = (2/e^-)^{1/2}u^{-\prime}\\
v^+ &= (2/e^+)^{1/2}v^{+\prime}, \quad v^- = (2/e^-)^{1/2}v^{-\prime}
\end{aligned}$$

and replace c'_1 and c'_2 by

$$c_1 = c'_1(e^+e^-)^{-1/2}, \quad c_2 = c'_2(e^+e^-)^{-1/2},$$

respectively. Then,

$$\begin{aligned}
P(A^+, A^-, B^+, B^-) &\approx (2\pi)^{-4}2^2(e^+e^-)^{-1}\\
&\quad \times \int\limits_{-\infty}^{+\infty}\ldots\int\limits_{-\infty}^{+\infty}\exp(-i\overline{\mathbf{T}}\mathbf{U}' - \tfrac{1}{2}\overline{\mathbf{U}}'\boldsymbol{\lambda}\mathbf{U}')\,\mathrm{d}\overline{\mathbf{U}}'\\
&= \pi^{-2}(e^+e^-)^{-1}\lambda^{-1/2}\exp(-\tfrac{1}{2}\overline{\mathbf{T}}\boldsymbol{\lambda}^{-1}\overline{\mathbf{T}}),
\end{aligned}$$

where

$$\begin{aligned}
\overline{\mathbf{U}}' &= (u^{+\prime}, u^{-\prime}, v^{+\prime}, v^{-\prime})\\
\overline{\mathbf{T}} &= \{(A^+ - A^+_{la})(2/e^+)^{1/2}, (A^- - A^-_{la})(2/e^-)^{1/2},\\
&\quad (B^+ - B^+_{la})(2/e^+)^{1/2}, (B^- - B^-_{la})(2/e^-)^{1/2}\}\\
\boldsymbol{\lambda} &= \begin{vmatrix} 1 & c_1 & 0 & c_2\\ c_1 & 1 & c_2 & 0\\ 0 & c_2 & 1 & -c_1\\ c_2 & 0 & -c_1 & 1\end{vmatrix}.
\end{aligned}$$

In a more explicit form,

$$\begin{aligned}
&P(A^+, A^-, B^+, B^-)\\
&\quad = \pi^{-2}(e^+e^-)^{-1}c^{-1}\exp\left\{-\frac{1}{c}\left[\frac{(A^+ - A^+_{la})^2 + (B^+ - B^+_{la})^2}{e^+}\right.\right.\\
&\qquad \left. + \frac{(A^- - A^-_{la})^2 + (B^- - B^-_{la})^2}{e^-}\right]\\
&\qquad + 2\frac{c_1}{c}\frac{[(A^+ - A^+_{la})(A^- - A^-_{la}) - (B^+ - B^+_{la})(B^- - B^-_{la})]}{(e^+e^-)^{1/2}}\\
&\qquad \left. + 2\frac{c_2}{c}\frac{[(A^+ - A^+_{la})(B^- - B^-_{la}) - (A^- - A^-_{la})(B^+ - B^+_{la})]}{(e^+e^-)^{1/2}}\right\},
\end{aligned} \tag{7}$$

where $c^2 = \det\boldsymbol{\lambda} = [1 - (c_1^2 + c_2^2)]^2$.

The change of variables

$$\begin{aligned}
A^+ &= R\cos\varphi^+, & B^+ &= R\sin\varphi^+\\
A^+_{la} &= R_{la}\cos\varphi^+_{la}, & B^+_{la} &= R_{la}\sin\varphi^+_{la}\\
A^- &= G\cos\varphi^-_{la}, & B^- &= G\sin\varphi^-\\
A^-_{la} &= G_{la}\cos\varphi^-_{la}, & B^-_{la} &= G_{la}\sin\varphi^-_{la}
\end{aligned}$$

changes (7) into

$$\begin{aligned}
&P(R, G, \varphi^+, \varphi^-)\\
&\quad = \frac{RG}{\pi^2e^+e^-c}\exp\left\{-\frac{1}{c}\left[\frac{R^2 + R^2_{la} - 2RR_{la}\cos(\varphi^+ - \varphi^+_{la})}{e^+}\right.\right.\\
&\qquad \left. + \frac{G^2 + G^2_{la} - 2GG_{la}\cos(\varphi^- - \varphi^-_{la})}{e^-}\right] + \frac{2c_3}{c}\frac{1}{(e^+e^-)^{1/2}}\\
&\qquad \times [RG\cos(\varphi^+ + \varphi^- - \gamma) + R_{la}G_{la}\cos(\varphi^+_{la} + \varphi^-_{la} - \gamma)\\
&\qquad \left. - RG_{la}\cos(\varphi^+ + \varphi^-_{la} - \gamma) - R_{la}G\cos(\varphi^- + \varphi^+_{la} - \gamma)]\right\},
\end{aligned}$$

where

$$c_3^2 = c_1^2 + c_2^2, \quad \gamma = \tan^{-1}(c_2/c_1).$$

Again, for shortness, $P(R, G, \varphi^+, \varphi^-)$ stands for $P(R, G, \varphi^+, \varphi^-|R_{la}, G_{la}, \varphi^+_{la}, \varphi^-_{la})$.

Let us define

$$(E^+_{nla})_{\rm calc} = (E^+ - E^+_{la}), \quad (E^-_{nla})_{\rm calc} = (E^- - E^-_{la}).$$

Then,

$$\begin{aligned}
(R_{nla})^2_{\rm calc} &= [R^2 + R^2_{la} - 2RR_{la}\cos(\varphi^+ - \varphi^+_{la})]\\
(G_{nla})^2_{\rm calc} &= [G^2 + G^2_{la} - 2GG_{la}\cos(\varphi^- - \varphi^-_{la})]
\end{aligned}$$

and (6) may be written in the simpler form

$$\begin{aligned}
&P(R, G, \varphi^+, \varphi^-)\\
&\quad = \frac{RG}{\pi^2e^+e^-c}\exp\left\{-\frac{1}{c}\left[\frac{(R^2_{nla})_{\rm calc}}{e^+} + \frac{(G^2_{nla})_{\rm calc}}{e^-}\right]\right.\\
&\qquad + \frac{2c_3}{c}\frac{1}{(e^+e^-)^{1/2}}[RG\cos(\varphi^+ + \varphi^- - \gamma)\\
&\qquad + R_{la}G_{la}\cos(\varphi^+_{la} + \varphi^-_{la} - \gamma) - RG_{la}\cos(\varphi^+ + \varphi^-_{la} - \gamma)\\
&\qquad \left. - R_{la}G\cos(\varphi^- + \varphi^+_{la} - \gamma)]\right\}.
\end{aligned} \tag{8}$$

The distribution (8) is the main result of this paper, from which marginal and conditional distributions will be derived.

4. The conditional distribution of the phases

We first calculate the marginal probability distribution

$$P(\varphi^+, R, G) \approx \frac{2RG}{\pi e^+ e^- c} I_0(Z_{\varphi^+}) \exp\left[-\frac{1}{c}\left(\frac{(R^2_{nla})_{\rm calc}}{e^+} + \frac{G^2 + G^2_{la}}{e^-} - \{[2c_3 R_{la} G_{la} \cos(\varphi^+_{la} + \varphi^-_{la} - \gamma) - 2c_3 R G_{la} \cos(\varphi^+ + \varphi^-_{la} - \gamma)](e^+e^-)^{-1/2}\}\right)\right], \quad (9)$$

where I_0 is the modified Bessel function of order zero and

$$Z_{\varphi^+} = \frac{2G}{c(e^-)^{1/2}}\left\{c_3^2 \frac{(R^2_{nla})_{\rm calc}}{e^+} + \frac{G^2_{la}}{e^-} + 2c_3 G_{la} \times \frac{[R\cos(\varphi^+ + \varphi^-_{la} - \gamma) - R_{la}\cos(\varphi^+_{la} + \varphi^-_{la} - \gamma)]}{(e^+e^-)^{1/2}}\right\}^{1/2}.$$

The distribution (9) may be approximated as follows. Since $c \ll 1$, Z_{φ^+} is large for the cases of interest. Then, $I_0(Z_{\varphi^+})$ may be expanded (Abramowitz & Stegun, 1972) according to

$$I_0(Z_{\varphi^+}) = \exp(|Z_{\varphi^+}|)/(2\pi|Z_{\varphi^+}|)^{1/2}. \quad (10)$$

Standard techniques will then lead to the conditional distribution

$$P(\varphi^+|R, G) \approx (G/Z_{\varphi^+})^{1/2} \exp[-(1/c)(G^2/e^- + G^2_{\rm calc} - cZ_{\varphi^+})], \quad (11)$$

where

$$G^2_{\rm calc} = (e^+)^{-1}(R^2_{nla})_{\rm calc} + (e^-)^{-1}G^2_{la} + 2c_3(e^+e^-)^{-1/2} \times G_{la}[R\cos(\varphi^+ + \varphi^-_{la} - \gamma) - R_{la}\cos(\varphi^+_{la} + \varphi^-_{la} - \gamma)]. \quad (12)$$

The same procedure leads to the distribution of φ^-:

$$P(\varphi^-|R, G) \approx (R/Z_{\varphi^-})^{1/2} \exp[-(1/c)(R^2/e^+ + R^2_{\rm calc} - cZ_{\varphi^-})], \quad (13)$$

where

$$R^2_{\rm calc} = (e^-)^{-1}(G^2_{nla})_{\rm calc} + (e^+)^{-1}R^2_{la} + 2c_3(e^+e^-)^{-1/2}R_{la} \times [G\cos(\varphi^- + \varphi^+_{la} - \gamma) - G_{la}\cos(\varphi^+_{la} + \varphi^-_{la} - \gamma)], \quad (14)$$

$$Z_{\varphi^-} = \frac{2R}{c(e^+)^{1/2}}\left\{c_3^2 \frac{(G^2_{nla})_{\rm calc}}{(e^-)} + \frac{R^2_{la}}{(e^+)} + 2c_3 R_{la} \times \frac{[G\cos(\varphi^+_{la} + \varphi^- - \gamma) - G_{la}\cos(\varphi^+_{la} + \varphi^-_{la} - \gamma)]}{(e^+e^-)^{1/2}}\right\}^{1/2}. \quad (15)$$

The formulas obtained so far are rather complicated: a simplified case to better understand them is described in Appendix *A*.

Formulas (11) and (13) have no counterpart in Hauptman (1982) and Giacovazzo (1983) approaches (indeed, in the absence of prior information on the anomalous-scatterer positions, only $\varphi^+ + \varphi^-$ may be estimated, not the single values of φ^+ and φ^-). However, (8) exactly coincides with the distribution (1*a*) as obtained by the above authors, when both the number of located anomalous scatterers and the measurement errors tend to vanish.

5. The centrosymmetric case

Since $E^+_{la} = E^-_{la}$ and $E^+ = E^-$, the joint probability to study is $P(R, \varphi^+|E^+_{la})$. Its characteristic function is

$$C(u^+, v^+) = \exp i(u^+A^+_{la} + v^+B^+_{la})\exp[-e^+(u^{+2} + v^{+2})/4],$$

from which

$$P(R, \varphi^+) \approx (\pi e^+)^{-1} R \exp\{-(1/e^+)[R^2 + R^2_{la} - 2RR_{la}\cos(\varphi^+ - \varphi^+_{la})]\}. \quad (16)$$

From (16), the following marginal distributions are obtained:

$$P(R) = \frac{2R}{e^+}\exp\left\{-\frac{1}{e^+}(R^2 + R^2_{la})\right\} I_0\left(\frac{2RR_{la}}{e^+}\right), \quad (17)$$

$$P(\varphi^+|R) = \left[2\pi I_0\left(\frac{2RR_{la}}{e^+}\right)\right]^{-1} \exp\left[\frac{2RR_{la}}{e^+}\cos(\varphi^+ - \varphi^+_{la})\right]. \quad (18)$$

According to (18), the most probable phase for φ^+ is always φ^+_{la}, with reliability parameter equal to $2RR_{la}/e^+$. Since R_{la} is in general a small number, the phase prediction is generally weak. Good predictions will be obtained when the scattering power of located anomalous-scattering atoms is not a negligible fraction of the unit-cell scattering power.

6. Applications

The distributions (11) and (13) have been plotted in Figs. 1–3 for some specific cases by using the diffraction data of cyanase (Walsh *et al.*, 2001), a homodecamer that crystallizes in $P1$ with four selenomethionines per monomer (40 in the unit cell), $a = 76.3$, $b = 81.0$, $c = 82.3$ Å, $\alpha = 70.30$, $\beta = 72.20$, $\gamma = 66.40°$. Multiwavelength data were collected up to 1.65 Å resolution: in our tests, we will only use data at $\lambda = 0.9465$ Å for which $f' = -2.618$, $f'' = 3.578$.

In each figure, (11) is the full line, (13) the broken line. The bimodal nature of the distribution is clearly observable in

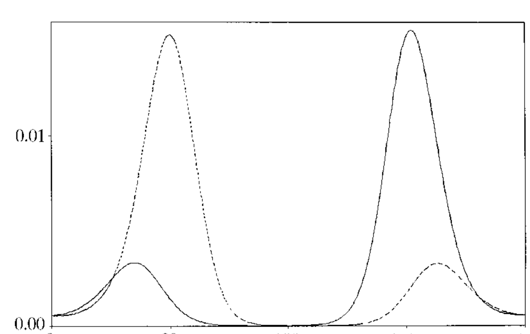

Figure 1
Distributions (11) (full line) and (13) (broken line) for the reflection 21,2,1: $R = 1.26$, $G = 1.24$, $R_{la} = 0.32$, $G_{la} = 0.32$, $\varphi^+_{la} = 267$, $\varphi^-_{la} = 108$, $\varphi_{\rm best} = 285$.

Figs. 1 and 2; occasionally (see Fig. 3) they are unimodal. The best estimates of φ^+ and φ^- (say $\varphi^+_{\rm best}$ and $\varphi^-_{\rm best}$) are obtained by calculating the centroids of the probability distributions (11) and (13), respectively (Blow & Crick, 1959). *I.e.* $\varphi^+_{\rm best}$ is obtained by calculating

$$x^+ = \textstyle\int P(\varphi^+|R, G)\cos\varphi^+\,{\rm d}\varphi^+,$$
$$y^+ = \textstyle\int P(\varphi^+|R, G)\sin\varphi^+\,{\rm d}\varphi^+,$$
$$\varphi^+_{\rm best} = \tan^{-1}(y^+/x^+).$$

The classical figure of merit for the phase estimate, widely used in protein crystallography, is given by

$$m^+ = (x^{+2} + y^{+2}).$$

Once $\varphi^+_{\rm best}$ is available, φ may be estimated by trivial geometrical considerations (the vector $\mathbf{F}$ is obtained by subtracting the anomalous scattering from $F^+_{\rm best}$; see notation in §1). The same procedure may be applied to derive the values of $\varphi^-_{\rm best}$ and m^-, from which another estimate of φ may be obtained. In the absence of measurement errors, the two estimates of φ should be perfectly coincident: in practice, they may differ by a few degrees and we assume their average (say $\varphi_{\rm best}$) as the best estimate of φ.

In order to check the potential of our approach, the procedure has been first applied to calculated (without error) data: we assumed $e^+ = e^- = 1 + (0.01|E_{\rm calc}|)^2$ to avoid singularities in distributions (11) and (13). The results are shown in Table 1, where $\varphi_{\rm best}$ estimates are ranked *versus m*. We notice: (*a*) the figure of merit m is a good criterion to select the most reliable estimates; (*b*) errors in the estimates are unavoidable owing to the bimodal nature of the distributions.

The corresponding results for observed data are shown in Table 2: the values of e^+ and e^- used in the calculations arise now from σ^{+2}_μ and σ^{-2}_μ deriving from measurement counting statistics. We notice: (*a*) m is still a good criterion to select the most reliable phases; (*b*) errors in the phase estimates are now larger owing to errors in measurements.

Results in Table 2 can help to address the question of the significances of our one-wavelength estimates. Lower values of f'' would depress the average value of the signal $|R - G|$ and therefore would reduce the number of reflections with high m value. Larger experimental errors (*i.e.* larger e^+ and e^- values) would depress the reliability of (11) and (13): accordingly, in Table 2, the number of reflections with $m > 0.9$ is lower than the corresponding number in Table 1. Underestimation of e^+ and e^- can influence the reliability of the estimates: *i.e.*, in Table 2, 2113 reflections have $m > 0.9$ but their average phase error is higher than for the corresponding 12869 reflections in Table 1. A more realistic scheme for the error (including the error in the structural model of the anomalous scatterers) should restrict m to a lower range and establish a more realistic correspondence between m and $\Delta\varphi$. Updating the error scheme does not require any change in our mathematical approach.

Table 1
Cyanase calculated data.

Phase estimates are ranked as a function of m. Numb is the number of reflections with figure of merit $m >$ Sog, $\Delta\varphi = \langle|\varphi_{\rm true} - \varphi_{\rm best}|\rangle$ is the corresponding average phase error (the weighted average phase error is in parentheses).

Sog	Numb	$\Delta\varphi$ (°)
0.1	62370	38 (32)
0.2	60369	37 (32)
0.3	56777	35 (31)
0.4	52325	33 (30)
0.5	46965	30 (28)
0.6	41016	27 (26)
0.7	33700	24 (23)
0.8	24644	20 (19)
0.9	12869	15 (15)

7. Conclusions

A new probabilistic approach for handling the OAS case is described, aiming at phasing structure factors under the assumption that anomalous-scatterer positions are fully or partially known. The method uses the technique of the joint probability distribution functions and provides conclusive formulas that, applied to a practical case, provides efficient tools for phasing reflections. Distinctive features of our approach are the following:

(*a*) The formulas are not obtained by analogy with the SIR case, but rigorously derived for the OAS case.

(*b*) The formulas do not have the usual exponential form.

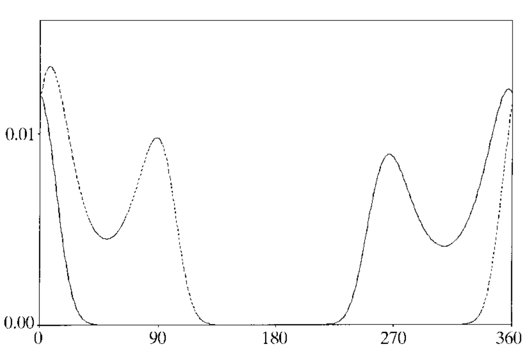

Figure 2
Distributions (11) (full line) and (13) (broken line) for the reflection 10,7,11: $R = 0.47$, $G = 0.52$, $R_{la} = 0.24$, $G_{la} = 0.24$, $\varphi^+_{la} = 48$, $\varphi^-_{la} = 326$, $\varphi_{\rm best} = 318$.

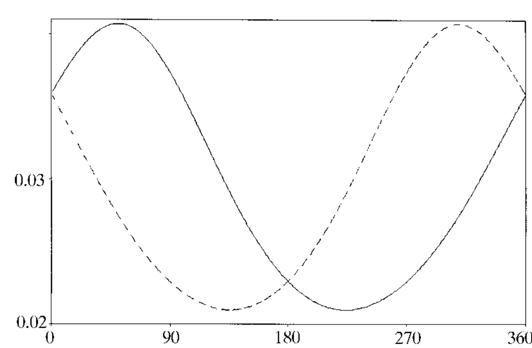

Figure 3
Distributions (11) (full line) and (13) (broken line) for the reflection 19,1,6: $R = 2.96$, $G = 2.99$, $R_{la} = 0.05$, $G_{la} = 0.05$, $\varphi^+_{la} = 81$, $\varphi^-_{la} = 294$, $\varphi_{\rm best} = 48$.

Table 2
Cyanase observed data.

Phase estimates are ranked as a function of m. Numb is the number of reflections with figure of merit $m > \mathrm{Sog}$, $\Delta\varphi = \langle|\varphi_{\rm true} - \varphi_{\rm best}|\rangle$ is the corresponding average phase error (the weighted average phase error is in parentheses).

Sog	Numb	$\Delta\varphi$ (°)
0.1	57211	66 (62)
0.2	51546	65 (62)
0.3	45073	63 (61)
0.4	38471	61 (60)
0.5	32138	60 (59)
0.6	25284	58 (57)
0.7	17595	56 (56)
0.8	9402	53 (53)
0.9	2113	52 (52)

(*c*) A change in the error magnitude both influences the phase reliability and modifies the phase estimates. This property is not shared by distributions like (3), which assigns different reliabilities for different error rates, but always provides the same estimate. The evaluation of the practical consequences of such a feature requires additional study.

(*d*) The formulas are still valid when some non-anomalous in addition to the anomalous atoms are localized. It is sufficient to understand the symbol '*la*' specified in §1 as 'located atoms', instead of 'located anomalous scatterers'. As a consequence, '*nla*' will represent the 'non-located atoms' instead of the original 'non-located anomalous scatterers'. This feature suggests the usefulness of a cyclic procedure capable of using as prior the structural information available at a given step of the phasing process.

(*e*) The final formulas may be easily generalized for application to the MAD case. This is the next important step of our work.

APPENDIX *A*

Let us suppose that all the anomalous scatterers have been correctly located: equivalently, assume that the errors come out mainly from measurements. Then,

$$c_2 \approx 0, \quad c_1 \approx c_3 = (e^+e^-)^{-1/2}$$
$$c \approx 1 - (e^+e^-)^{-1} = e/(e^+e^-), \quad \gamma \approx 0$$
$$e = (e^+e^- - 1).$$

Since ($\Re$ stands for 'real part of'),

$$\begin{aligned}\Re[(E^+_{nla})_{\rm calc}(E^-_{nla})_{\rm calc}] &= [RG\cos(\varphi^+ + \varphi^-) + R_{la}G_{la}\cos(\varphi^+_{la} + \varphi^-_{la}) \\ &\quad - RG_{la}\cos(\varphi^+ + \varphi^-_{la}) - R_{la}G\cos(\varphi^+_{la} + \varphi^-)],\end{aligned}$$

the joint probability distribution (8) may be written in the simpler form

$$\begin{aligned}P(R, G, \varphi^+, \varphi^-) &= (RG/\pi^2 e)\exp\big(-(1/e)\{e^-(R_{nla})^2_{\rm calc} + e^+(G_{nla})^2_{\rm calc} \\ &\quad - 2\Re[(E^+_{nla})_{\rm calc}(E^-_{nla})_{\rm calc}]\}\big). \qquad (19)\end{aligned}$$

Then equations (11)–(15) may be replaced by

$$\begin{aligned}P(\varphi^+|R, G) \approx (G/Z_{\varphi^+})^{1/2}\exp\{-[(e^+e^-)/e][G^2/e^- + G^2_{\rm calc} \\ - (e/e^+e^-)Z_{\varphi^+}]\} \qquad (20)\end{aligned}$$

$$\begin{aligned}P(\varphi^-|R, G) \approx (R/Z_{\varphi^-})^{1/2}\exp\{-[(e^+e^-)/e][R^2/e^+ \\ + R^2_{\rm calc} - (e/e^+e^-)Z_{\varphi^-}]\} \qquad (21)\end{aligned}$$

$$\begin{aligned}G^2_{\rm calc} &= (e^+)^{-1}(R^2_{nla})_{\rm calc} + (e^-)^{-1}G^2_{la} + 2(e^+e^-)^{-1}G_{la} \\ &\quad \times [R\cos(\varphi^+ + \varphi^-_{la}) - R_{la}\cos(\varphi^+_{la} + \varphi^-_{la})] \qquad (22)\end{aligned}$$

$$\begin{aligned}R^2_{\rm calc} &= (e^-)^{-1}(G^2_{nla})_{\rm calc} + (e^+)^{-1}R^2_{la} + 2(e^+e^-)^{-1}R_{la} \\ &\quad \times [R\cos(\varphi^- + \varphi^+_{la}) - G_{la}\cos(\varphi^+_{la} + \varphi^-_{la})] \qquad (23)\end{aligned}$$

$$\begin{aligned}Z_{\varphi^+} &= (2G/e)e^+\{(R^2_{nla})_{\rm calc}/(e^+)^2 + G^2_{la} + 2(G_{la}/e^+) \\ &\quad \times [R\cos(\varphi^+ + \varphi^-_{la}) - R_{la}\cos(\varphi^+_{la} + \varphi^-_{la})]\}^{1/2} \qquad (24)\end{aligned}$$

$$\begin{aligned}Z_{\varphi^-} &= (2R/e)e^-\{(G^2_{nla})_{\rm calc}/(e^-)^2 + R^2_{la} + 2(R_{la}/e^-) \\ &\quad \times [G\cos(\varphi^- + \varphi^+_{la})] - G_{la}\cos(\varphi^+_{la} + \varphi^-_{la})]\}^{1/2}. \qquad (25)\end{aligned}$$

Let us now find some simplified forms of (20) and (21). We note that the exponential term in (20) and (21) varies more rapidly than $(G/Z_{\varphi^+})^{1/2}$ or $(R/Z_{\varphi^-})^{1/2}$. Accordingly, the expected value of φ^+ will mostly depend on it and consequently $(G/Z_{\varphi^+})^{1/2}$ and $(R/Z_{\varphi^-})^{1/2}$ may be replaced by unity. Furthermore, if e^+ and e^- are not too large we can replace (20) and (21) by the simpler expressions

$$P(\varphi^+|R, G) \approx \exp[-(e^+/e)(G - G_{\rm calc})^2] \qquad (26)$$

and

$$P(\varphi^-|R, G) \approx \exp[-(e^-/e)(R - R_{\rm calc})^2], \qquad (27)$$

where

$$\begin{aligned}G^2_{\rm calc} &= (R^2_{nla})_{\rm calc} + G^2_{la} + 2G_{la}[R\cos(\varphi^+ + \varphi^-_{la}) \\ &\quad - R_{la}\cos(\varphi^+_{la} + \varphi^-_{la})] \qquad (28)\end{aligned}$$

$$\begin{aligned}R^2_{\rm calc} &= (G^2_{nla})_{\rm calc} + R^2_{la} + 2R_{la}[R\cos(\varphi^- + \varphi^+_{la}) \\ &\quad - G_{la}\cos(\varphi^+_{la} + \varphi^-_{la})]. \qquad (29)\end{aligned}$$

Even if expressed in a concise form, the above expressions are of not immediate understanding.

A further simplification allows a clearer insight into (19). Distribution (19) reduces to (30) if e^+ and e^- are sufficiently small and close to each other:

$$\begin{aligned}P(R, G, \varphi^+, \varphi^-) \approx (RG/\pi e)\exp[-(1/e) \\ \times |(E^+ - E^+_{la}) - (E^- - E^-_{la})|^2]. \qquad (30)\end{aligned}$$

We note:

(*a*) If only E^+_{la} and R are known then (30) reduces to $P(R, \varphi^+)$ and the maximum of the probability distributions is attained when $|E^+ - E^+_{la}|$ is a minimum. This occurs when $\varphi^+ = \varphi^+_{la}$ (in this case, E^+ and E^+_{la} are collinear vectors).

(*b*) If only E^-_{la} and G are known, the maximum value of $P(G, \varphi^-)$ is attained when $\varphi^- = \varphi^-_{la}$.

(*c*) If E^+_{la}, E^-_{la}, R and G are known, the largest value of (30) is no longer attained when $\varphi^+ = \varphi^+_{la}$ and/or $\varphi^- = \varphi^-_{la}$. These values indeed maximize $(|E^+ - E^+_{la}|^2 + |E^- - E_{la}|^2)$ but not $|(E^+ - E^+_{la}) - (E^- - E_{la})|^2$. Thus, (30) takes into account the correlation between E^+ and E^-.

Let us now write (30) as

$$P(R, G, \varphi^+, \varphi^-) \approx (RG/\pi e)\exp[-(1/e)|(E^+ - E^-) - (E^+_{la} - E^-_{la})|^2].$$

Accordingly, the most probable values of φ^+ and φ^- are those for which $(E^+ - E^-)$ is as close as possible (in the complex phase) to $(E^+_{la} - E^-_{la})$. This result is a sensitive way to define the maximization condition.

The simplifying assumptions introduced in this section suggest [by analogy with (27)] that distributions (11) and (13) are expected to be mostly bimodal.

M. Walsh is kindly acknowledged for having provided us with a set of experimental data.

References

Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions.* New York: Dover.
Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
Blow, D. M. & Rossmann, M. G. (1961). *Acta Cryst.* **14**, 1195–1202.
Camalli, M., Giacovazzo, C. & Spagna, R. (1985). *Acta Cryst.* A**41**, 605–613.
Giacovazzo, C. (1983). *Acta Cryst.* A**39**, 585–592.
Giacovazzo, C. (1998). *Direct Phasing in Crystallography.* Oxford University Press.
Hauptman, H. (1982). *Acta Cryst.* A**38**, 632–641.
Mathews, B. W. (1966). *Acta Cryst.* **20**, 82–86.
North, A. C. T. (1965). *Acta Cryst.* **18**, 212–216.
Terwilliger, T. C. & Eisenberg, D. (1987). *Acta Cryst.* A**43**, 6–13.
Walsh, M. A., Otwinowski, Z., Perrakis, A., Anderson, P. M. & Joachimiak, A. (2001). In preparation.