Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 11 December 2000
Accepted 8 February 2001

The method of joint probability distribution functions applied to MAD techniques. The centric case

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

Traditional probabilistic approaches consider MAD (multiple-wavelength anomalous-dispersion) data as a special MIR (multiple isomorphous replacement) case. The rigorous method of the joint probability distribution functions has been applied to solve the phase problem, with the assumption that the anomalous scatterers' substructure is *a priori* known. The probabilistic approach is able to handle measurement errors: it has been applied to symmetry-restricted phases and provides simple and efficient formulas.

1. Notation

N: number of atoms in the unit cell.
a: number of anomalous scatterers in the unit cell.
$na = N - a$: number of 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.

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

$$F_a^+ = |F_a^+| \exp(i\varphi_a^+) = \sum_{a} f_j \exp(2\pi i \mathbf{hr}_j).$$

Σ_a, Σ_{na}, $\Sigma_N = \sum (f_j'^2 + f_j''^2)$, where the summation is extended to a, na and N atoms.

2. Introduction

MAD (multiple-wavelength anomalous dispersion) techniques (Hendrickson & Ogata, 1997; Smith, 1997) exploit the structure-factor differences due to the variation of the anomalous-scattering factors at wavelengths around the absorption edges of some atoms in a protein crystal. Owing to the power and tunability of modern synchrotron beamlines, the MAD method has had a profound impact on modern techniques for solving the phase problem in protein crystallography. The procedure usually involves two steps:

(a) the anomalously scattering atoms are first located, which may be difficult because the partial substructure may be very complicated (Terwilliger *et al.*, 1987; Miller *et al.*, 1994; Sheldrick & Gould, 1995);

(b) the phase values are determined assuming the partial structure of the anomalously scattering atoms as prior information.

Previous probabilistic approaches consider: (i) SAD (single-wavelength anomalous dispersion) and MAD data as special SIR (single isomorphous replacement) and MIR (multiple isomorphous replacement) cases, respectively. In particular, the classical Blow & Crick (1959) expressions, integrated by Terwilliger & Eisenberg (1987) contributions, originally derived for SIR–MIR techniques, are extended by analogy to SAD–MAD cases. (ii) The algebraic analysis of the MAD data by Karle (1980) and Hendrickson (1985) has been adapted to a probabilistic description (Pähler *et al.*, 1990; Chiadmi *et al.*, 1993).

In this paper, we are interested in the second stage only. In particular, we will develop below a technique described in a previous paper (Giacovazzo & Siliqi, 2001; from now on paper I), where the joint probability distribution method has been applied to the SAD case on the assumption that the positions of all or 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). In paper I, the joint probability distribution

$$P(F^+, F^- | F_{la}^+, F_{la}^-) \tag{1}$$

has been calculated, from which the phase estimates

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

and

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

were derived. From them, the most probable value of φ was derived by geometrical considerations. In (1), the prior information on F_{la}^+ and F_{la}^- arises from the prior knowledge of the located anomalous scatterers.

In this paper, we will extend the method to the MAD case, with two limiting assumptions: (*a*) all the anomalously scattering atoms have been perfectly localized, thus F_a^+ and F_a^- (the subscript *a* stands for 'anomalous substructure') arise from the full anomalous scatterer substructure; (*b*) only the symmetry-restricted phases will be considered. The two limitations will help in the reading of the paper and an easier understanding of the results, which will be expressed in a rather attractive and simple form.

By analogy with the probabilistic approach described in paper I, the positions of the non-anomalous scatterers will be the primitive random variables. Equation (I.4) now becomes

$$F^+ = F_a^+ + F_{na}^+ + \mu^+ = F_a^+ + F_q^+, \tag{2}$$

where F_{na}^+ is the structure factor corresponding to the non-anomalous scatterers (the subscript *na* stands for 'non-anomalous atoms', all supposed non-located). Furthermore, $\mu^+ = |\mu|^+ \exp(i\theta^+)$ represents the cumulative error arising from errors in measurements: it is inglobated into $F_q^+ = F_{na}^+ + \mu^+$. Equivalently, we should assume

$$F^- = F_a^- + F_{na}^- + \mu^- = F_a^- + F_q^-,$$

where $F_q^- = F_{na}^- + \mu^-$. Since $F^+ = F^-$ in the centric case, from now on we will omit the superscript sign ($F^+ = F^- \equiv F$), we will assume that $\mu^+ \approx \mu^- \equiv \mu$ is a real variable and that F_a, F_{na}, μ are uncorrelated with each other. Then,

$$\langle |F|^2 \rangle = |F_a|^2 + \Sigma_{na} + \langle |\mu|^2 \rangle.$$

As in paper I, we will normalize the structure factors with respect to the unknown part of the structure: accordingly, for centric reflections,

$$A^+ = A^- \equiv A = \left[\sum_{j=1}^{N} \left(f_j' \cos 2\pi \mathbf{hr}_j \right) + \mu \right] \Big/ \Sigma_{na}^{1/2}, \tag{3a}$$

$$B^+ = B^- \equiv B = \left[\sum_{j=1}^{N} \left(f_j'' \cos 2\pi \mathbf{hr}_j \right) \right] \Big/ \Sigma_{na}^{1/2}, \tag{3b}$$

$$E = R \exp(i\varphi) = (A + iB) = F/\Sigma_{na}^{1/2}.$$

Equivalently,

$$E_a = R_a \exp(i\varphi_a) = (A_a + iB_a) = F_a/\Sigma_{na}^{1/2},$$
$$E_q = R_q \exp(i\varphi_q) = (A_q + iB_q) = F_q/\Sigma_{na}^{1/2},$$
$$\sigma^2 = \langle \mu^2 \rangle / \Sigma_{na}.$$

The single-wavelength case will first be considered, to show how the approach may be simplified. Then the two-wavelength case will be analysed: the resulting formulas have a very simple form and are quite instructive. Then the general MAD case will be described.

3. The one-wavelength case: the distribution $P(A \mid E_a)$

Since $E^+ = E^- \equiv E$, $E_a^+ = E_a^- \equiv E_a$, we should study the conditional distribution $P(R, \varphi|E_a)$. On assuming that the anomalous-scatterer substructure is perfectly known, the characteristic function of the distribution $P(R, \varphi|E_a)$ is

$$C(u) = \exp[i(uA_a + vB_a)] \exp[-e(u^2/2)], \tag{4}$$

where $e = (1 + \sigma_\mu^2)$. Then,

$$P(R, \varphi|E_a) = \delta(B - B_a) L(A, A_a, e),$$

where δ is the Dirac delta function and

$$L(A, A_a, e) = (2\pi e)^{-1/2} \exp[-(A - A_a)^2/2e] \tag{5}$$

is the Gausssian distribution of the variable A, centred at A_a, with variance e.

The distribution (5) suggests that we can emphasize the one-dimensionality of our statistical problem by replacing the pair (R, φ) by A and assuming as 'observed' value of $|A|$ the value $|A| = (|E|^2 - B_a^2)^{1/2}$, where $|E|$ is the observed pseudo-normalized structure factor and B_a is calculated *via* the prior information on the anomalous-scatterer substructure. Then the characteristic function of the distribution $P(A|E_a)$ is simply

$$C(u) = \exp i[uA_a - eu^2/2]$$

and

$$P(A|E_a) = L(A, A_a, e). \tag{6}$$

The probability that the sign s of A is equal to the sign s_a of A_a is then:

$$P(s = s_a) = 0.5 + 0.5 \tanh(|AA_a|/e). \tag{7}$$

The probabilistic formula (7) suggests that: (*a*) the larger e, the less reliable the sign indications are; when there is no error in the measurements, the limiting case $e = 1$ occurs (note that e is never smaller than unity); (*b*) large values of A_a are necessary to obtain reliable phase indications: this seldom occurs in protein crystallography because the scattering power of the anomalous-scatterer substructure is usually a very small percentage of the scattering power of the unit cell.

Multiple-wavelength experiments are then necessary to provide additional information for a more fruitful statistical phase assignment.

4. The two-wavelength case: the distribution $P(A_1, A_2 \mid E_{a1}, E_{a2})$

According to §3, we will derive the joint probability distribution $P(A_1, A_2|A_{a1}, A_{a2})$, where the subscripts 1 and 2 indicate that the variables are referred to the wavelengths λ_1 and λ_2, respectively. The characteristic function is

$$C(u_1, u_2) = \exp[i(u_1 A_{a1} + u_2 A_{a2})] \times \exp[-(e_1 u_1^2 + e_2 u_2^2 + 2u_1 u_2)/2], \tag{8}$$

where u_1 and u_2 are carrying variables associated with A_1 and A_2, respectively.

Equation (8) has been obtained under the reasonable assumption that measurement errors at the two wavelengths are uncorrelated. The distribution $P(A_1, A_2|A_{a1}, A_{a2})$ is readily obtained as the Fourier transform of (8); we have

$$P(A_1, A_2|A_{a1}, A_{a2}) = L^{-1}\exp\{-1/2k^{-1}[e_2(A_1 - A_{a1})^2 + e_1(A_2 - A_{a2})^2 - 2(A_1 - A_{a1})(A_2 - A_{a2})]\}, \quad (9)$$

where L is a normalizing factor and $k = (e_1e_2 - 1)$.

A more instructive form of (9) is obtained by introducing the relations

$$e_1 = (1 + \sigma_1^2), \quad e_2 = (1 + \sigma_2^2).$$

We have

$$P(A_1, A_2|A_{a1}, A_{a2}) = L^{-1}\exp\{-1/2k^{-1}[(A_1 - A_{a1}) - (A_2 - A_{a2})]^2 - 1/2k^{-1}[\sigma_2^2(A_1 - A_{a1})^2 + \sigma_1^2(A_2 - A_{a2})^2]\}, \quad (10)$$

where $k = \sigma_1^2 + \sigma_2^2 + \sigma_1^2\sigma_2^2$.

Equation (10) suggests the following:

(*a*) The first term in the exponential is maximized when $A_{q1} = A_1 - A_{a1}$ has the same sign and modulus as $A_{q2} = A_2 - A_{a2}$. The reader will find this obvious if he or she considers that A_{q1} and A_{q2} do not depend on the anomalous scattering and therefore they are expected to be equal. The formula confirms this expectation and, at the same time, uses the above condition as a lack-of-closure criterion.

(*b*) The second term in the exponential supports the tendency of A_1 and A_2 to have the same signs as A_{a1} and A_{a2}, respectively (see §3): the tendency is regulated by the error parameters σ_2^2 and σ_1^2, respectively. Since σ_2^2 and σ_1^2 are usually smaller than unity, the second term in the exponential is usually negligible with respect to the first one. Accordingly, the additional wavelength dispersion data constitute a valuable source of information with respect to the SAS case.

(*c*) Formula (10) supports the expectation that the difference $(A_1 - A_{a1})^2$ plays a more important role if the error on $(A_2 - A_{a2})^2$ (say σ_2^2) is larger than σ_1^2. The reciprocal is also true.

(*d*) Since

$$[(A_1 - A_{a1}) - (A_2 - A_{a2})]^2 = [(A_1 - A_2) - (A_{a1} - A_{a2})]^2, \quad (11)$$

the distribution (10) clearly indicates the possible experimental reasons for MAD failures. Indeed, the size of the term (11) critically depends on the experimental errors related to the estimates of the differences $(|A_1| - |A_2|)$ and $(|A_{a1}| - |A_{a2}|)$, both of which are quite small quantities.

Let s_1, s_2, s_{a1}, s_{a2} be the signs of A_1, A_2, A_{a1}, A_{a2}, respectively. The sign probabilities can be derived by first omitting from (10) the terms that are insensitive to the signs:

$$P(A_1, A_2|A_{a1}, A_{a2}) \approx L^{-1}\exp\{k^{-1}[A_1A_2 + (A_1 - A_2) \times (A_{a1} - A_{a2}) + \sigma_2^2A_1A_{a1} + \sigma_1^2A_2A_{a2}]\}, \quad (12)$$

then the marginal probability

$$P(s_1|A_{a1}, A_{a2}) = \sum_{s_2=\pm1} P(s_1, s_2|A_{a1}, A_{a2})$$

is derived, and finally the conclusive expression is obtained:

$$P(s_1 = s_{a1}|A_{a1}, A_{a2}) = [1 + P(s_1 = -s_{a1}|A_{a1}, A_{a2})/P(s_1 = s_{a1}|A_{a1}, A_{a2})]^{-1}. \quad (13)$$

A simpler but robust expression may be obtained by observing that A_1 and A_2 will mostly have the same sign s_1. If the probability that they have opposite sign is considered vanishing, the product (A_1A_2) can be omitted from (12).

Then,

$$(A_1 - A_2)(A_{a1} - A_{a2}) = s_1(|A_1| - |A_2|)(A_{a1} - A_{a2})$$

and

$$P(s = s_{a1}|A_{a1}, A_{a2}) \approx 0.5 + 0.5\tanh\{k^{-1}[s_1\Delta_{12}(A_{a1} - A_{a2}) + \sigma_2^2|A_1A_{a1}|]\}, \quad (14a)$$

where

$$\Delta_{12} = |A_1| - |A_2|.$$

Equations (13) and (14*a*) are the conclusive formulas for the two-wavelength case.

As before underlined, the first term in the tanh argument is more important than the second, owing to the usually quite small value of σ_1^2 (and σ_2^2). If $\Delta f_1'$ and $\Delta f_2'$ have the same sign then

$$A_{a1} - A_{a2} = s_{a1}(|A_{a1}| - |A_{a2}|) = s_{a1}\Delta_{a12},$$

and (14*a*) reduces to the simpler expression

$$P(s_1 = s_{a1}|A_{a1}, A_{a2}) \approx 0.5 + 0.5\tanh\{k^{-1}[\Delta_{12}\Delta_{a12} + \sigma_2^2(|A_1A_{a1}|)]\}, \quad (14b)$$

which mostly relies on the sign of the product $\Delta_{12}\Delta_{a12}$.

Usually, $\sigma_i^2 \ll 1$: accordingly, we can introduce the approximation

$$K = \sigma_1^2 + \sigma_2^2 + \sigma_1^2\sigma_2^2 \approx \sigma_1^2 + \sigma_2^2$$

and simply express (14*b*) in terms of structure factors:

$$P(s_1 = s_{a1}|A_{a1}^F, A_{a2}^F) \approx 0.5 + 0.5\tanh\{[\langle\mu_1^2\rangle + \langle\mu_2^2\rangle]^{-1} \times [\Delta_{12}^F\Delta_{a12}^F + \langle\mu_2^2\rangle|A_{a1}^FA_{a2}^F|/\Sigma_N]\},$$

where $\Delta_{12}^F = |A_1^F| - |A_2^F|$, $\Delta_{a12}^F = |A_{a1}^F| - |A_{a2}^F|$ and A_i^F, A_{ai}^F are the real components of F_i and F_{ai}.

Since the second term of the tanh arguments is often negligible with respect to the first one, we obtain

$$P(s_1 \approx s_{a1}|A_{a1}^F, A_{a2}^F) \approx 0.5 + 0.5\tanh\{[\langle\mu_1^2\rangle + \langle\mu_2^2\rangle]^{-1}\Delta_{12}^F\Delta_{a12}^F\}.$$

Some didactical cases are illustrated in Appendix *A*.

5. The *n*-wavelength case: the distribution $P(A_1, \ldots, A_n|E_{a1}, \ldots, E_{an})$

The procedure described in §4 is now extended to the *n*-wavelength case. The characteristic function of $P(A_1, \ldots, A_n|E_{a1}, \ldots, E_{an})$ is:

$$C(u_1,\ldots,u_n)=\exp[i(u_1A_{a1}+\ldots+u_nA_n)]\times\exp\left[-\left(\sum_i e_iu_i^2+2\sum_{i<j}u_iu_j\right)\Big/2\right].$$

Then,

$$P(A_1,\ldots,A_n|E_{a1},\ldots,E_{an})=(2\pi)^{-n}\int_{-\infty}^{+\infty}\ldots\int_{-\infty}^{+\infty}\exp\left\{-i\left[\sum_i u_i(A_i-A_{ai})\right]-(1/2)\left[\sum_i e_iu_i^2+2\sum_{i<j}u_iu_j\right]\right\}. \quad (15)$$

Equation (15) may be rewritten in a shorter form as

$$P(\mathbf{A}|\mathbf{E}_a)=(2\pi)^{-n}\int_{-\infty}^{+\infty}\ldots\int_{-\infty}^{+\infty}\exp(-i\,\overline{\mathbf{T}}\mathbf{U}-1/2\overline{\mathbf{U}}\mathbf{K}\mathbf{U}),$$

where

$$\overline{\mathbf{T}}=[(A_1-A_{a1}),\ldots,(A_n-A_{an})],\quad \overline{\mathbf{U}}=[u_1,\ldots,u_n],\quad \mathbf{K}=\begin{vmatrix} e_1 & 1 & \ldots & 1\\ 1 & e_2 & \ldots & 1\\ \vdots & \vdots & \vdots & \vdots\\ 1 & 1 & \ldots & e_n\end{vmatrix}.$$

Then,

$$P(\mathbf{A}|\mathbf{E}_a)=(2\pi)^{-n/2}[\det(\mathbf{K})]^{-1/2}\exp(-1/2\overline{\mathbf{T}}\mathbf{K}^{-1}\mathbf{T}) \quad (16)$$

or, in a more explicit form,

$$P(\mathbf{A}|\mathbf{E}_a)=(2\pi)^{-n/2}[\det(\mathbf{K})]^{-1/2}\exp\left\{-1/2\sum_i\lambda_{ii}(A_i-A_{ai})^2+2\sum_{i<j}\lambda_{ij}(A_i-A_{ai})(A_j-A_{aj})\right\}, \quad (17)$$

where the λ_{ij} are the elements of the matrix $\mathbf{K}^{-1}$. The value of $\det(\mathbf{K})$ may be estimated by the relation

$$\det\begin{vmatrix}1+d_1 & 1 & \ldots & 1\\ 1 & 1+d_2 & \ldots & 1\\ \vdots & \vdots & \vdots & \vdots\\ 1 & 1 & \ldots & 1+d_n\end{vmatrix}=d_1d_2\ldots d_n\left[1+\sum_{i=1}^{n}(1/d_i)\right].$$

Accordingly, we have

$$\det(\mathbf{K})=\sigma_1^2\sigma_2^2\ldots\sigma_n^2\left[1+\sum_{i=1}^{n}(1/\sigma_i^2)\right]. \quad (18)$$

It is easily verified that (6) and (9) are special cases of (17). A form generalizing (10) may be obtained after some algebraic rearrangement of the terms in (17). We obtain:

$$P(\mathbf{A}|\mathbf{E}_a)=(2\pi)^{-n/2}[\det(\mathbf{K})]^{-1/2}\exp\left\{-1/2\sum_{i<j}\lambda_{ij}[(A_i-A_j)-(A_{ai}-A_{aj})]^2-1/2\sum_i R_i(A_i-A_{ai})^2\right\}, \quad (19)$$

where $R_i=\sum_{i=1}^{n}\lambda_{ij}$.

The sign probabilities may be derived as in §4. If we assume that all the A_i's have the same sign s_1, then

$$[(A_i-A_j)(A_{ai}-A_{aj})]=[s_1\Delta_{ij}(A_{ai}-A_{aj})],$$

where $\Delta_{ij}=|A_i|-|A_j|$ and

$$P(s_1=s_{a1}|\mathbf{E}_a)=0.5+0.5\tanh\left\{-\sum_{j=2}^{n}\lambda_{1j}[s_1\Delta_{1j}(A_{a1}-A_{aj})]+R_1|A_1A_{a1}|\right\}. \quad (20)$$

If the $\Delta f_j'$ have the same sign, (20) reduces to

$$P(s_1=s_{a1}|\mathbf{E}_a)=0.5+0.5\tanh\left\{-\sum_{j=2}^{n}\lambda_{1j}[\Delta_{1j}\Delta_{a1j}]+R_1|A_1A_{a1}|\right\}, \quad (21)$$

where $\Delta_{a1j}=|A_{a1}|-|A_{aj}|$.

We now give the explicit expressions for the λ_{1j} elements:

$$\lambda_{1j}=-\left[\prod_{p=2,p\neq j}^{n}(\sigma_p^2)\right]\Big/[\det(\mathbf{K})]=-\left\{\sigma_1^2\sigma_j^2\left[1+\sum_{i=1}^{n}(1/\sigma_i^2)\right]\right\}^{-1}\quad\text{for } j\neq 1, \quad (22)$$

$$\lambda_{11}=\left[\prod_{p=2}^{n}(\sigma_p^2)\right]\Big/[\det(\mathbf{K})]-\sum_{j=2}^{n}\lambda_{1j},\qquad R_1=\left[\prod_{p=2}^{n}(\sigma_p^2)\right]\Big/[\det(\mathbf{K})]=\left\{\sigma_1^2\left[1+\sum_{i=1}^{n}(1/\sigma_i^2)\right]\right\}^{-1}. \quad (23)$$

We note that, according to (23), λ_{1j} will have a small modulus if σ_1^2 and σ_j^2 are large enough (*i.e.* unreliable measurements). Similarly, according to (23), a large value of σ_1^2 would weaken the probability of the sign relationship between A_1 and A_{a1}.

For $n=3, 4$ the λ_{1j}'s are:

$n=3$

$$\det(\mathbf{K})=\sigma_1^2\sigma_2^2\sigma_3^2+\sigma_1^2\sigma_2^2+\sigma_1^2\sigma_3^2+\sigma_2^2\sigma_3^2$$
$$\lambda_{11}=(\sigma_2^2\sigma_3^2+\sigma_2^2+\sigma_3^2)/\det(\mathbf{K})$$
$$\lambda_{22}=(\sigma_1^2\sigma_3^2+\sigma_1^2+\sigma_3^2)/\det(\mathbf{K})$$
$$\lambda_{33}=(\sigma_1^2\sigma_2^2+\sigma_1^2+\sigma_2^2)/\det(\mathbf{K})$$
$$\lambda_{12}=-\sigma_3^2/\det(\mathbf{K})$$
$$\lambda_{13}=-\sigma_2^2/\det(\mathbf{K})$$
$$\lambda_{23}=-\sigma_1^2/\det(\mathbf{K});$$

Table 1
Expected $\Delta f'$ and f'' values for each chosen λ value.

λ (Å)	$\Delta f'$	f''
1.1271	−1.805	0.646
0.9793	−8.582	3.843
0.9791	−7.663	3.841
0.9465	−2.618	3.578

$n = 4$

$$\det(\mathbf{K}) = \sigma_1^2\sigma_2^2\sigma_3^2\sigma_4^2 + \sigma_1^2\sigma_2^2\sigma_3^2 + \sigma_1^2\sigma_2^2\sigma_4^2 + \sigma_1^2\sigma_3^2\sigma_4^2 + \sigma_2^2\sigma_3^2\sigma_4^2$$
$$\lambda_{11} = (\sigma_2^2\sigma_3^2\sigma_4^2 + \sigma_2^2\sigma_3^2 + \sigma_2^2\sigma_4^2 + \sigma_3^2\sigma_4^2)/\det(\mathbf{K})$$
$$\lambda_{22} = (\sigma_1^2\sigma_3^2\sigma_4^2 + \sigma_1^2\sigma_3^2 + \sigma_1^2\sigma_4^2 + \sigma_3^2\sigma_4^2)/\det(\mathbf{K})$$
$$\vdots$$
$$\lambda_{12} = -\sigma_3^2\sigma_4^2/\det(\mathbf{K})$$
$$\lambda_{13} = -\sigma_2^2\sigma_4^2/\det(\mathbf{K})$$
$$\vdots$$

6. The estimate of symmetry restricted phases

The formulas so far derived have been explicitly obtained for the case $P\bar{1}$, but they may handle reflections with symmetry-restricted phase values different from $(0, \pi)$ without any special modification. For example, let us suppose that the allowed phase values are $(+\pi/2, -\pi/2)$. In this case, A is the component along the imaginary direction in the Gaussian plane. We still retain the A and B definitions of §I.3 and we can apply (20) provided the 'observed' value of A is $|A| = (|E|^2 - |B_a|^2)^{1/2}$.

7. Experimental tests

To check the correctness of our mathematical approach, we have applied the theory so far developed to the calculated (without error) data of 1SRV (Walsh *et al.*, 1999), space group $C222_1$, $a = 63.470$, $b = 65.960$, $c = 75.030$ Å. Multiwavelength data were collected up to 1.70 Å resolution. The expected $\Delta f'$ and f'' values for each chosen λ are shown in Table 1. Structure factors were calculated for the 958 reflections with restricted phase values. To avoid singularities in (14*b*) and (21), we assumed $e = 1 + (0.05|E_{\rm calc}|^2)$. In Table 2, we use λ_1 and λ_2 wavelengths. For each threshold value, we give the number of reflections (Nr) for which $G >$ ARG and the percentage (%) of the correct sign estimates. G is the modulus of the tanh argument in (14*b*). The two-wavelength (ideal) case completely solves the phase problem. The errors at very small values of G are exclusively rounding errors of the calculations. The reader will note that the percentage of the correct sign estimates does not depend on the value of G. This is due to: (*a*) calculated data are used; (*b*) an unrealistic weight is employed.

The simultaneous use of the four wavelengths has only the effect of increasing the expected reliability of the phase estimates (that is the G value): for shortness, the obvious result is not shown.

Table 2
1SRV calculated data, centric reflections: number of reflections (Nr) for which $G >$ ARG and percentage (%) of the correct sign estimates.

G is the modulus of the tanh argument in equation (14*b*) when the two-wavelength case is checked.

ARG	Nr	%
0.00	958	98.2
0.33	321	99.4
2.00	154	98.7
4.00	103	99.1
7.0	71	98.6
10	57	98.3

8. Conclusions

A new probabilistic approach for handling the MAD case is described, aiming at phasing structure factors under the assumption that the anomalous-scatterer positions are entirely known. The approach uses the technique of the joint probability distribution functions and provides simple and

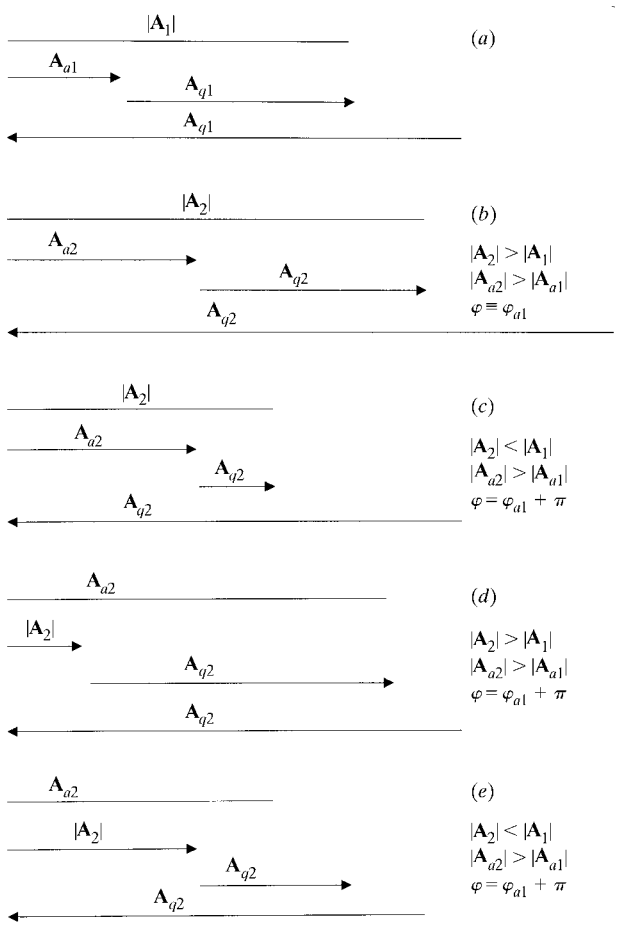

Figure 1
(*a*) One-wavelength centric case. The sign of A_1 is not defined by the prior knowledge of A_{a1}; (*b*), (*c*), (*d*), (*e*) Four specific cases for a second wavelength.

instructive formulas. The approach is quite general: it may also be used to treat the case in which errors at different wavelengths are correlated (Terwillinger & Berendzen, 1997). This only requires suitable coefficients in the Gaussian component of the characteristic function [see equation (8) for the two-wavelength case].

Our next theoretical step will generalize the results for the non-centrosymmetrical case. Since the accuracy of the phase estimates depends on the accuracy of A_a and B_a, the practical applications of our approach would require the integration of our theoretical results with some of the current computer programs for defining and refining the anomalous-scatterer substructure.

APPENDIX *A*

In this *Appendix*, some geometrical considerations supporting the results obtained *via* our probabilistic approach are described.

The one-wavelength centric case may be summarized as in Fig. 1(*a*). A_1 is a non-directional segment of which we want to determine the correct orientation, $\mathbf{A}_{a1}$ is a vector with known modulus and direction. Two possible solutions are available for $\mathbf{A}_{q1}$, with opposite directions (the restraint that $|\mathbf{A}_q + \mathbf{A}_{a1}|$ must be equal to $|\mathbf{A}_1|$ holds). Consequently, both the alternatives are allowed for A_1.

The sections (*b*), (*c*), (*d*), (*e*) of Fig. 1 represent four cases for a second wavelength. We assume that $\Delta f_1' \Delta f_2' > 0$ or, in other words, that $\mathbf{A}_{a1}$ has the same sign as $\mathbf{A}_{a2}$. Each of the four cases is marked by specific relations between the moduli $|\mathbf{A}_j|$ and $|\mathbf{A}_{aj}|$. If the λ_1 and the λ_2 cases are combined, a unique solution is found, owing to the necessity that $\mathbf{A}_{q1} = \mathbf{A}_{q2}$ in modulus and sign. This is in perfect agreement with the main term of the tanh argument in equation (14*b*).

References

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.
Chiadmi, M., Kahn, R., De La Fortelle, E. & Fourme, R. (1993). *Acta Cryst.* D**49**, 522–529.
Giacovazzo, C. (1998). *Direct Phasing in Crystallography*. Oxford University Press.
Giacovazzo, C. & Siliqi, D. (2001). *Acta Cryst.* A**57**, 40–46.
Hendrickson, W. A. (1985). *Trans. Am. Crystallogr. Assoc.* **21**, 11–21.
Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
Karle, J. (1980). *Int. J. Quantum Chem. Symp.* **7**, 357–367.
Mathews, B. W. (1966). *Acta Cryst.* **20**, 82–86.
Miller, R., Gallo, S., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
North, A. C. T. (1965). *Acta Cryst.* **18**, 212–216.
Pähler, A., Smith, J. L. & Hendrickson, W. A. (1990). *Acta Cryst.* A**46**, 537–540.
Sheldrick, G. M. & Gould, R. G. (1995). *Acta Cryst.* B**51**, 423–431.
Smith, J. L. (1997). *Proceedings of the CCP4 Study Weekend. Recent Advances in Phasing*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey, pp. 25–39. Warrington: Daresbury Laboratory.
Terwilliger, T. C. & Berendzen, J. (1997). *Acta Cryst.* D**53**, 571–579.
Terwilliger, T. C. & Eisenberg, D. (1987). *Acta Cryst.* A**43**, 6–13.
Terwilliger, T. C., Kim, S.-H. & Eisenberg, D. (1987). *Acta Cryst.* A**43**, 1–5.
Walsh, M. A., Dementieva, I., Evans, G., Sanishvili, R. & Joachimiak, A. (1999). *Acta Cryst.* D**55**, 1168–1173.