

## Application of maximum likelihood to direct methods: the probability density function of the triple-phase sums. XI

Jordi Rius

Institut de Ciència de Materials de Barcelona (CSIC), Campus de la UAB, 08193-Bellaterra, Catalunya, Spain. Correspondence e-mail: jordi.rius@icmab.es

Received 25 January 2006  
Accepted 27 June 2006

The maximum-likelihood method is applied to direct methods to derive a more general probability density function of the triple-phase sums which is capable of predicting negative values. This study also proves that maximization of the origin-free modulus sum function  $S$  yields, within the limitations imposed by the assumed approximations, the maximum-likelihood estimates of the phases. It thus represents the formal theoretical justification of the  $S$  function that was initially derived from Patterson-function arguments [Rius (1993). *Acta Cryst.* **A49**, 406–409].

© 2006 International Union of Crystallography  
Printed in Great Britain – all rights reserved

## 1. Introduction

Although most recent efforts in *ab initio* phasing of medium-sized crystal structures concentrate on dual-space direct methods (Weeks *et al.*, 1994), direct methods working exclusively in reciprocal space have also been improved considerably in the last few years (Rius, 2004). Reciprocal-space procedures are simpler to implement and faster to calculate but, as a counterpart, require previous knowledge of the small  $E$  magnitudes.

Historically, the evolution of direct methods can be divided into two periods. An initial one making use of the probability associated with the phase sum of each structure invariant and for which Cochran's probability density function (p.d.f.) was crucial (Cochran, 1955), and a second one dominated by the computing facilities allowing the simultaneous handling of the information contained in multiple structure invariants. Decisive for this second period was not only the development of suitable phase refinement functions but also the derivation of the four-phase sum p.d.f. that allowed negative quartets to be predicted from the known weak  $E$ 's (Schenk, 1973; Hauptman, 1975; Giacovazzo, 1976). In the present contribution, a modified p.d.f. of the triple-phase sum is introduced that can also predict negative triplets from weak  $E$ 's, in analogy to the quartets case. This p.d.f. follows from the application of the maximum-likelihood method (Fisher, 1922) to the phases using the observed strong and weak  $E$  values as observations. It is found that the likelihood function corresponds to the 'origin-free' modulus sum function (Rius, 1993), thus providing the latter with a strong theoretical basis and justifying its observed robustness and efficiency for reciprocal-space phase refinements. This also confirms the results published in Rius (1997), where the close relationship between negative triplets and quartets was already deduced from physical considerations. To facilitate the maximum-likelihood

derivation in §6, all relevant quantities involved in its definition are first introduced in §2, 3, 4 and 5.

2. The structure factor  $G$  of the squared point-atom structure

For simplicity, a crystal structure belonging to space group  $P1$  and containing  $N$  equal atoms in the unit cell will be assumed throughout. The normalized structure factor of reflection  $\mathbf{H}$  is then given after Hauptman & Karle (1953) by

$$\mathbf{E}_{\mathbf{H}} = E_{\mathbf{H}} \exp i\varphi_{\mathbf{H}} = (1/N^{1/2}) \sum_j \exp(i2\pi\mathbf{H}\mathbf{r}_j), \quad (1)$$

which corresponds to an electron-density distribution of point atoms with  $1/N^{1/2}$  as scattering power. The modulus  $E_{\mathbf{H}}$  can be obtained from the measured intensity by correcting it for the atomic form factor and thermal vibration decays. By squaring the scattering power  $1/N^{1/2}$ , the structure factor of the squared point-atom structure results:

$$\mathbf{G}_{\mathbf{H}} = G_{\mathbf{H}} \exp i\psi_{\mathbf{H}} = (1/N) \sum_j \exp(i2\pi\mathbf{H}\mathbf{r}_j) \quad (2)$$

so that the experimental  $G_{\mathbf{H}}$  and  $E_{\mathbf{H}}$  are related by the simple expression

$$G_{\mathbf{H}} = E_{\mathbf{H}}/N^{1/2}. \quad (3)$$

In what follows, only those reflections with  $s$  values smaller than  $s_{\max} = (2 \sin \theta_{\max})/\lambda$  will be considered.

3. The structure factor  $G$  as a function of the phases

Consider the sum  $\sum_{\mathbf{H}} \mathbf{E}_{\mathbf{H}} \mathbf{E}_{-\mathbf{H}}$  in which  $\mathbf{H}$  represents the experimentally accessible reflections. As indicated in Fig. 1, these reflections correspond to vectors going from the reciprocal-lattice node  $-\mathbf{H}$  (point  $O'$ ) to any node inside the

limiting sphere and with lengths  $s_{\mathbf{H}'} < s_{\max}$ . Next, this sum will be decomposed in terms of the atomic position vectors  $\mathbf{r}_j$ :

$$N \sum_{\mathbf{H}'} \mathbf{E}_{\mathbf{H}'} \mathbf{E}_{\mathbf{H}-\mathbf{H}'} = \sum_k \exp(i2\pi\mathbf{H}\mathbf{r}_k) \sum_{\mathbf{H}'} \sum_j \exp[-i2\pi\mathbf{H}'(\mathbf{r}_k - \mathbf{r}_j)] \quad (4)$$

and, if  $\mathbf{r}_{kj} = \mathbf{r}_k - \mathbf{r}_j$ , then

$$N^{1/2} \sum_{\mathbf{H}'} \mathbf{E}_{\mathbf{H}'} \mathbf{E}_{\mathbf{H}-\mathbf{H}'} = \sum_k \exp(i2\pi\mathbf{H}\mathbf{r}_k) \sum_j \sum_{\mathbf{H}'} (1/N^{1/2}) \times \exp(-i2\pi\mathbf{H}'\mathbf{r}_{kj}). \quad (5)$$

The inner sum  $\sum_{\mathbf{H}'} (1/N^{1/2}) \exp(-i2\pi\mathbf{H}'\mathbf{r}_{kj})$  yields the electron density of the origin peak of strength  $(1/N^{1/2})$  at position  $\mathbf{r}_{kj}$ . For  $j = k$ ,  $\mathbf{r}_{kk} = 0$  and the value of the inner sum reduces to  $n_{\mathbf{H}'}/N^{1/2}$  with  $n_{\mathbf{H}'}$  being the number of  $\mathbf{H}'$  reflections. For  $j \neq k$ , this value will also depend on the sampling volume of  $\mathbf{H}'$ . If a given  $-\mathbf{H}$  is close to the centre of the limiting sphere (*i.e.* for short  $O-O'$  distances in Fig. 1), the sampling volume of  $\mathbf{H}'$  can be approximated to a sphere of effective radius  $s_{\text{eff}} \simeq s_{\max}$ . The summation  $\sum_{\mathbf{H}'} \exp(-i2\pi\mathbf{H}'\mathbf{r})$  can be expressed as the integral over the whole of the reciprocal space  $V^*$ ,

$$\sum_{\mathbf{H}'} \exp(-i2\pi\mathbf{H}'\mathbf{r}) = K \int_{V^*} t(s) \exp(i2\pi\mathbf{r}\mathbf{s}) \, ds, \quad (6)$$

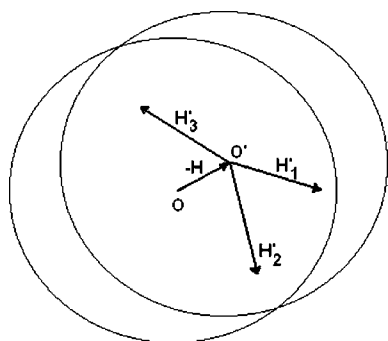
where function  $t(s)$  is 1 for  $s < s_{\text{eff}}$  and 0 otherwise (Lipson & Cochran, 1966). This integral is the Fourier transform of  $t(s)$  and is a function of  $r$  only. It is given by

$$T_3(r) = 4\pi s_{\text{eff}}^3 [\sin(2\pi s_{\text{eff}}r) - 2\pi s_{\text{eff}}r \cos(2\pi s_{\text{eff}}r)] / (2\pi s_{\text{eff}}r)^3.$$

The value of the scaling constant  $K$  can be estimated by considering that the value of the sum  $\sum_{\mathbf{H}'} \exp(-i2\pi\mathbf{H}'\mathbf{r})$  at  $\mathbf{r} = 0$  is  $n_{\mathbf{H}'}$ , *i.e.*  $K = n_{\mathbf{H}'}/T_3(0)$ . Consequently, the value of the sum for an arbitrary  $\mathbf{r}_{kj}$  is

$$\sum_{\mathbf{H}'} \exp(-i2\pi\mathbf{H}'\mathbf{r}_{kj}) \simeq n_{\mathbf{H}'} [T_3(r_{kj})/T_3(0)]. \quad (7)$$

For data reaching atomic resolution,  $T_3$  falls off very fast. Therefore it will only be significant for  $\mathbf{r}_{kk} = 0$  and, since the quotient  $T_3(r_{kk})/T_3(0)$  is unity, the approximation



**Figure 1**  
Schematic representation of the experimentally accessible  $\mathbf{H}'$  reflections in expression (4). The sphere of radius  $s_{\max} = (2\sin\theta_{\max})/\lambda$  centred at the origin  $O$  is the so-called limiting sphere. The wanted  $\mathbf{H}'$  reflections can be found by placing an additional sphere of same radius  $s_{\max}$  at  $O'$ , *i.e.* the reciprocal-lattice node  $-\mathbf{H}$ , and by looking for all nodes  $\mathbf{H}'$  inside the intersection area.

$$\sum_{\mathbf{H}'} (1/N^{1/2}) \exp(-i2\pi\mathbf{H}'\mathbf{r}_{kj}) \simeq n_{\mathbf{H}'}/N^{1/2} \quad (8)$$

holds. In contrast, for  $-\mathbf{H}$  close to the border of the limiting sphere, the sampling volume of the corresponding  $\mathbf{H}'$  cannot be approximated by a complete sphere, *i.e.* either the term  $\mathbf{H}'$  or  $-\mathbf{H}'$  is sometimes missing in  $\sum_{\mathbf{H}'} \exp(-i2\pi\mathbf{H}'\mathbf{r}_{kj})$ , so that the resulting imaginary contribution will not vanish. Fortunately, for data at atomic resolution, the broadening of the origin peak caused by the Fourier truncation is small in comparison to typical next-neighbour distances (1.4–1.5 Å), so that approximation (8) is still valid. This situation is illustrated in detail in Fig. 2 for the one-dimensional case. Introduction of (8) in (5) under consideration of (2) leads to

$$\mathbf{G}_{\mathbf{H}} \simeq \langle \mathbf{E}_{\mathbf{H}'} \mathbf{E}_{\mathbf{H}-\mathbf{H}'} \rangle_{\mathbf{H}'} \quad (9)$$

(Hughes, 1953). The modulus  $G_{\mathbf{H}}$  can be obtained by multiplying  $\mathbf{G}_{\mathbf{H}}$  with the phase term  $\exp[i\psi_{-\mathbf{H}}]$  of its complex conjugate, *i.e.* by making

$$G_{\mathbf{H}} = \Re\{\exp[i\psi_{-\mathbf{H}}]\mathbf{G}_{\mathbf{H}}\} \quad (10)$$

with

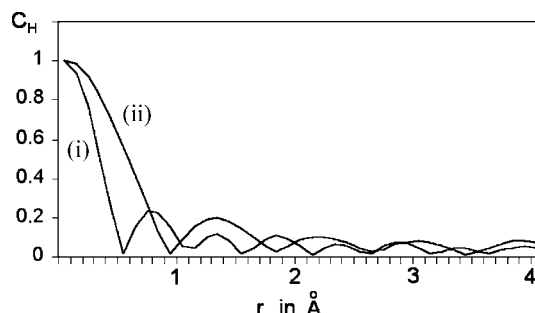
$$\psi_{\mathbf{H}} = \text{phase of } \left\{ \sum_{\mathbf{H}'} \mathbf{E}_{\mathbf{H}'} \mathbf{E}_{\mathbf{H}-\mathbf{H}'} \right\}. \quad (11)$$

Finally, from (9) and (10), the desired modulus  $G_{\mathbf{H}}$  in terms of the phases follows:

$$G_{\mathbf{H}}(\Phi) \simeq \langle E_{\mathbf{H}'} E_{\mathbf{H}-\mathbf{H}'} \cos \Phi_{3_{\mathbf{H},\mathbf{H}'}} \rangle_{\mathbf{H}'}, \quad (12)$$

where  $\Phi$  denotes here the collectivity of phases  $\varphi$  of the true structure and where the triple-phase sum is

$$\Phi_{3_{\mathbf{H},\mathbf{H}'}} = \psi_{-\mathbf{H}} + \varphi_{\mathbf{H}'} + \varphi_{\mathbf{H}-\mathbf{H}'}. \quad (13)$$



**Figure 2**  
Evolution of  $\langle \exp(-i2\pi\mathbf{H}'\mathbf{r}) \rangle_{\mathbf{H}'(\mathbf{H})} = C_{\mathbf{H}}(r) \exp i\alpha_{\mathbf{H}}(r)$  for a one-dimensional crystal structure of period  $a = 30$  Å at atomic resolution ( $s_{\max} = 1 \text{ \AA}^{-1}$ ): (i) for a reflection  $\mathbf{H}$  near the centre of the limiting 'sphere' with  $s_{\mathbf{H}} = 0.1 \text{ \AA}^{-1}$  and  $\mathbf{H}'$  ranging from  $-30$  to  $27$ ; (ii) for a reflection  $\mathbf{H}$  close to the border of the limiting 'sphere' with  $s_{\mathbf{H}} = 0.9 \text{ \AA}^{-1}$  and  $\mathbf{H}'$  ranging from  $-30$  to  $3$ . The sharpest  $C_{\mathbf{H}}$  function corresponds to  $s_{\mathbf{H}} = 0.1 \text{ \AA}^{-1}$ . In both cases, however,  $C_{\mathbf{H}}$  is small enough at approximately  $1.5 \text{ \AA}$ , *i.e.* a typical next-neighbour distance in organic compounds. This justifies approximation (8).

#### 4. Observational variance of the structure factor $G$ of the squared point-atom structure

Let the  $G_{\mathbf{H}}(\Phi)$  in (12) be decomposed into the  $n_{\mathbf{H}}$  individual terms

$$g_{\mathbf{H}'} = E_{\mathbf{H}'} E_{\mathbf{H}-\mathbf{H}'} \cos \Phi 3_{\mathbf{H},\mathbf{H}'} \quad (14)$$

and let the  $g_{\mathbf{H}'}$  be assumed to be independent observations from the same population following a p.d.f. centred at  $G$  (the 'measured' value) with variance

$$\sigma_{g_{\mathbf{H}}}^2 = n_{\mathbf{H}}^{-1} \sum_{\mathbf{H}'} (g_{\mathbf{H}'} - G_{\text{Hav}})^2 \quad (15)$$

and with estimated mean of the population,  $G_{\text{Hav}} = \langle g_{\mathbf{H}'} \rangle_{\mathbf{H}'}$ . For its computation, it is more convenient to write  $\sigma_{g_{\mathbf{H}}}^2$  in the form

$$\sigma_{g_{\mathbf{H}}}^2 = (1/n_{\mathbf{H}}) \left[ \sum_{\mathbf{H}'} g_{\mathbf{H}'}^2 - \left( \sum_{\mathbf{H}'} g_{\mathbf{H}'} \right)^2 / n_{\mathbf{H}} \right] \quad (16)$$

so that by introducing expression (14) for  $g_{\mathbf{H}'}$  in (16),

$$\begin{aligned} \sigma_{g_{\mathbf{H}}}^2 = & (1/n_{\mathbf{H}}) \sum_{\mathbf{H}'} E_{\mathbf{H}'}^2 E_{\mathbf{H}-\mathbf{H}'}^2 \cos^2 \Phi 3_{\mathbf{H},\mathbf{H}'} \\ & - (1/n_{\mathbf{H}}^2) \left[ \sum_{\mathbf{H}'} E_{\mathbf{H}'}^2 E_{\mathbf{H}-\mathbf{H}'}^2 \cos^2 \Phi 3_{\mathbf{H},\mathbf{H}'} + \sum_{\mathbf{H}'} \sum_{\mathbf{H}'' \text{ for } \mathbf{H}' \neq \mathbf{H}''} \dots \right] \end{aligned} \quad (17)$$

the approximate value is found:

$$\sigma_{g_{\mathbf{H}}}^2 \cong \langle E_{\mathbf{H}'}^2 E_{\mathbf{H}-\mathbf{H}'}^2 \cos^2 \Phi 3_{\mathbf{H},\mathbf{H}'} \rangle_{\mathbf{H}'}. \quad (18)$$

The desired variance of the estimated mean of the population,  $G_{\text{Hav}}$  is then given by

$$\sigma_{\mathbf{H}}^2 = \langle (G_{\text{Hav}} - G)^2 \rangle_{\mathbf{H}'} = (1/n_{\mathbf{H}}^2) \sum_{\mathbf{H}'} \sigma_{g_{\mathbf{H}}}^2 = \sigma_{g_{\mathbf{H}}}^2 / n_{\mathbf{H}}. \quad (19)$$

The numerical value of  $\sigma_{g_{\mathbf{H}}}^2$  can be roughly estimated by supposing in (18) that  $E_{\mathbf{H}'}$ ,  $E_{\mathbf{H}-\mathbf{H}'}$  and  $\cos^2 \Phi 3_{\mathbf{H},\mathbf{H}'}$  are uncorrelated. According to the  $E$  statistics, the theoretical values of  $\langle E_{\mathbf{H}'}^2 \rangle_{\mathbf{H}'}$  and  $\langle E_{\mathbf{H}-\mathbf{H}'}^2 \rangle_{\mathbf{H}'}$  are both one and, on the hypothesis of a uniform distribution of  $\Phi 3$  on the trigonometric circle, the value of  $\langle \cos^2 \Phi 3_{\mathbf{H},\mathbf{H}'} \rangle_{\mathbf{H}'}$  is  $1/2$ , so that

$$\sigma_{g_{\mathbf{H}}}^2 \cong \langle E_{\mathbf{H}'}^2 \rangle_{\mathbf{H}'} \langle E_{\mathbf{H}-\mathbf{H}'}^2 \rangle_{\mathbf{H}'} \langle \cos^2 \Phi 3_{\mathbf{H},\mathbf{H}'} \rangle_{\mathbf{H}'} = 1/2. \quad (20)$$

#### 5. The linear approximation to $G_{\mathbf{H}}(\Phi)^2$

The treatment of the phases  $\varphi$  is much simpler in  $G_{\mathbf{H}}(\Phi)$  than in the corresponding squared function. For simplicity, it will therefore be convenient to replace in the likelihood expression the squared function  $G_{\mathbf{H}}(\Phi)^2$  by the linear approximation

$$G_{\mathbf{H}}(\Phi)^2 \cong k G_{\mathbf{H}}(\Phi). \quad (21)$$

This approximation requires  $G_{\mathbf{H}}(\Phi)$  to be positive, which is ensured through the condition  $\psi = f(\Phi)$  imposed in (11) (Rius *et al.*, 2002). The value of  $k$  is found by averaging the individual slopes obtained from the different experimental  $G_{\mathbf{H}}$ , *i.e.*

$$k = \langle dG^2/dG |_{G=G_{\mathbf{H}}} \rangle_{\mathbf{H}} = 2 \langle G \rangle_{\mathbf{H}}. \quad (22)$$

#### 6. The likelihood expressed as a function of the phases

Let us assume that the  $G_{\mathbf{H}}$  are independent observations and that the errors or discrepancies  $G_{\mathbf{H}} - G_{\mathbf{H}}(\Phi)$  are random and distributed according to a Gaussian density function with variance  $\sigma_{\mathbf{H}}^2$ . According to the maximum-likelihood method (Fisher, 1922), the maximum-likelihood estimate of the parameters, the phases, corresponds to the maximum of the likelihood function  $L$ :

$$L(\Phi) = \prod_{\mathbf{H}} [\sigma_{\mathbf{H}}(2\pi)^{1/2}]^{-1} \exp\{-[G_{\mathbf{H}} - G_{\mathbf{H}}(\Phi)]^2 / 2\sigma_{\mathbf{H}}^2\}, \quad (23)$$

in which the joint density of the data is the product of their individual p.d.f.'s. This expression can be simplified by squaring the binomial, by considering (21) and (22) and collecting together those factors that are independent of the phases in a single constant  $c$ , *i.e.*

$$L(\Phi) = c \prod_{\mathbf{H}} \exp\{[2G_{\mathbf{H}}G_{\mathbf{H}}(\Phi) - 2\langle G \rangle G_{\mathbf{H}}(\Phi)] / 2\sigma_{\mathbf{H}}^2\}. \quad (24)$$

In view of (3), (24) can be modified to

$$L(\Phi) \cong c \prod_{\mathbf{H}} \exp[(1/N^{1/2})(E_{\mathbf{H}} - \langle E \rangle)G_{\mathbf{H}}(\Phi) / \sigma_{\mathbf{H}}^2] \quad (25)$$

and, by taking into account (12), (19) and (20), it further reduces to

$$L(\Phi) \cong c \prod_{\mathbf{H},\mathbf{H}'} \exp[(2/N^{1/2})(E_{\mathbf{H}} - \langle E \rangle)E_{\mathbf{H}'}E_{\mathbf{H}-\mathbf{H}'} \cos \Phi 3_{\mathbf{H},\mathbf{H}'}]. \quad (26)$$

Next, by grouping all phase-independent factors present in the exponential argument into a single parameter

$$\kappa_{\mathbf{H},\mathbf{H}'} = (2/N^{1/2})(E_{\mathbf{H}} - \langle E \rangle)E_{\mathbf{H}'}E_{\mathbf{H}-\mathbf{H}'} \quad (27)$$

the definitive form for  $L(\Phi)$  is reached:

$$L(\Phi) \cong c \prod_{\mathbf{H},\mathbf{H}'} \exp(\kappa_{\mathbf{H},\mathbf{H}'} \cos \Phi 3_{\mathbf{H},\mathbf{H}'}). \quad (28)$$

Each exponential term in (28) involves a triplet of reflections  $-\mathbf{H}$ ,  $\mathbf{H}'$ ,  $\mathbf{H} - \mathbf{H}'$  and represents its individual contribution to the likelihood. It is obvious that only the exponential terms with large concentration parameters  $|\kappa|$  can contribute significantly to the likelihood and that those with small  $|\kappa|$  could be disregarded. Inspection of (27) readily indicates that all  $\kappa$  with at least two weak  $E$  values must have small  $|\kappa|$  values. In contrast, if  $E_{\mathbf{H}}$  and  $E_{\mathbf{H}-\mathbf{H}'}$  are both large, then the contribution of the corresponding exponential terms can be significant. Consequently, only this second case will be considered. To simplify the analysis, the strong (weak)  $\mathbf{H}$  reflections are denoted by the symbols  $\mathbf{h}$  ( $\mathbf{k}$ ), and the two other reflections of the triplet are referred to as  $\mathbf{h}'$  and  $\mathbf{h} - \mathbf{h}'$  ( $\mathbf{h}''$  and  $\mathbf{k} - \mathbf{h}''$ ). The two types of exponential terms that are most important for the likelihood maximization are the following.

*Type 1: one weak and two strong  $E$  values*

The corresponding concentration parameter and triple-phase sum are

$$\kappa_{\mathbf{k},\mathbf{h}''} = (2/N^{1/2})(E_{\mathbf{k}} - \langle E \rangle)E_{\mathbf{h}'}E_{\mathbf{k}-\mathbf{h}''} \quad (29)$$

$$\Phi 3_{\mathbf{k},\mathbf{h}''} = \psi_{-\mathbf{k}} + \varphi_{\mathbf{h}''} + \varphi_{\mathbf{k}-\mathbf{h}''}. \quad (30)$$

The absolute value of  $\kappa_{k,h''}$  is large due to the presence of the factor  $(E_k - \langle E \rangle)$ . Since  $\kappa_{k,h''}$  is here negative, the contribution to the likelihood will be largest for  $\cos \Phi 3_{k,h''} = -1$ , that is for  $\Phi 3_{k,h''} = \pi$ . Notice that the triple-phase sum contains the phase  $\psi_{-k} = f(\Phi)$  and not  $\phi_{-k}$ .

Type 2: three strong  $E$  values

When all three  $E$  values are large, the equalities  $\psi_{-\mathbf{h}} = \varphi_{-\mathbf{h}}$ ,  $\psi_{\mathbf{h}'} = \varphi_{\mathbf{h}'}$ ,  $\psi_{\mathbf{h}-\mathbf{h}'} = \varphi_{\mathbf{h}-\mathbf{h}'}$  can be assumed to be valid, and therefore there are three exponential terms with identical  $\Phi 3$  values. These three terms can be collected together and the cosine term factorized out to give

$$\begin{aligned} \kappa_{\mathbf{h},\mathbf{h}'} \cos \Phi 3_{\mathbf{h},\mathbf{h}'} + \kappa_{\mathbf{h},\mathbf{h}-\mathbf{h}'} \cos \Phi 3_{\mathbf{h},\mathbf{h}-\mathbf{h}'} + \kappa_{\mathbf{h}',\mathbf{h}-\mathbf{h}'} \cos \Phi 3_{\mathbf{h}',\mathbf{h}-\mathbf{h}'} \\ = 3\kappa_{\mathbf{H},\mathbf{H}'} \cos \Phi 3_{\mathbf{H},\mathbf{H}'} \end{aligned} \quad (31)$$

with

$$\kappa_{\mathbf{h},\mathbf{h}'} = (2/N^{1/2})\{1 - (\langle E \rangle/3)[E_{\mathbf{h}}^{-1} + E_{\mathbf{h}'}^{-1} + E_{\mathbf{h}-\mathbf{h}'}^{-1}]\}E_{\mathbf{h}}E_{\mathbf{h}'}E_{\mathbf{h}-\mathbf{h}'} \quad (32)$$

$$\Phi 3_{\mathbf{h},\mathbf{h}'} = \varphi_{-\mathbf{h}} + \varphi_{\mathbf{h}'} + \varphi_{\mathbf{h}-\mathbf{h}'}. \quad (33)$$

By applying the above conclusions to (28), the likelihood in terms of the set  $\Phi$  of phases  $\varphi$  of the strong reflections may be approximated with

$$L(\Phi) \cong c \prod_{\mathbf{h},\mathbf{h}'} \exp(\kappa_{\mathbf{h},\mathbf{h}'} \cos \Phi 3_{\mathbf{h},\mathbf{h}'}) \prod_{\mathbf{k},\mathbf{h}''} \exp(\kappa_{\mathbf{k},\mathbf{h}''} \cos \Phi 3_{\mathbf{k},\mathbf{h}''}). \quad (34)$$

Since the exponential quantities must be everywhere greater than or equal to zero,  $L$  must have a real logarithm. The logarithm is a monotonically increasing function of its argument, so the maximum value of  $L$  corresponds also to the maximum value of  $\ln(L)$  (Prince, 1982). Taking natural logarithms on both sides of (34),

$$\ln[L(\Phi)] = \ln(c) + S(\Phi) \quad (35)$$

with

$$S(\Phi) = \sum_{\mathbf{h}} \sum_{\mathbf{h}'} \kappa_{\mathbf{h},\mathbf{h}'} \cos \Phi 3_{\mathbf{h},\mathbf{h}'} + \sum_{\mathbf{k}} \sum_{\mathbf{h}''} \kappa_{\mathbf{k},\mathbf{h}''} \cos \Phi 3_{\mathbf{k},\mathbf{h}''}, \quad (36)$$

then, since  $\ln(c)$  is constant, the maximum of  $S(\Phi)$  will correspond to the maximum of  $L(\Phi)$  and will give the maximum-likelihood estimate of the phases.  $S(\Phi)$  is the origin-free modulus sum function which was first derived from purely physical considerations (Rius, 1993, 2002). Since  $L$ ,  $\ln(L)$  and  $S$  have their maxima for the same  $\Phi$ , the elementary theory of maxima and minima tells us that there will be one likelihood equality of type

$$\partial \ln(L)/\partial \varphi_{\mathbf{h}} = \partial S/\partial \varphi_{\mathbf{h}} = 0 \quad (37)$$

for each phase  $\varphi_{\mathbf{h}}$  to be determined (Debaerdemaeker *et al.*, 1985). It can be shown that (37) can be solved with the iterative application of the  $S$ -tangent formula [expression (12) in Rius (1993)]. The  $S$  phase-refinement function has been applied successfully to the solution of a number of crystal structures of variable complexity under a variety of situations, *e.g.* single-crystal diffraction, powder diffraction and surface diffraction.

7. The probability density function of the triple-phase sums

According to (28),  $L$  can be approximated by the product of the exponential terms with triplets formed by either one large or one weak  $\mathbf{H}$  reflection and two additional large ones ( $\mathbf{h}'$  and  $\mathbf{H} - \mathbf{h}'$ ). If the sign of  $\kappa$  is incorporated in the argument of the respective cosine, (28) takes the form

$$L(\Phi) \cong \prod_{\mathbf{H},\mathbf{h}'} c \exp[|\kappa_{\mathbf{H},\mathbf{h}'}| \cos(\mu_{\mathbf{H},\mathbf{h}'} - \Phi 3_{\mathbf{H},\mathbf{h}'})], \quad (38)$$

where  $\mu$  is 0 for positive and  $\pi$  for negative  $\kappa$ . Suppose now the normal circular probability density function of an angular variable  $x$  (von Mises, 1918)

$$P(x) = [2\pi I_0(\kappa)]^{-1} \exp[|\kappa| \cos(\mu - x)] \quad (39)$$

in which

- (i)  $\mu$  is the most probable value of  $x$ ,
- (ii)  $|\kappa|$  is the concentration parameter (*i.e.* a measure of the tightness or sharpness of the function around  $\mu$ ),
- (iii) the first term is a normalizing constant to ensure that the probability in the  $x$  interval  $[-\pi, +\pi]$  is unity, and with  $I_0(\kappa)$  being the modified Bessel function of zero order.

Making  $x = \Phi 3$  in (39) and comparing the resulting expression with (38), one can readily see that  $L(\Phi)$  is proportional to the product of the normal circular p.d.f.'s of the  $\Phi 3$ ,

$$L(\Phi) \cong c' \prod_{\mathbf{H},\mathbf{h}'} P[\Phi 3_{\mathbf{H},\mathbf{h}'}], \quad (40)$$

wherein the phase-independent constant  $\prod_{\mathbf{H},\mathbf{h}'} 2\pi I_0(\kappa_{\mathbf{H},\mathbf{h}'})$  is contained in  $c'$ . Notice that, if the  $\Phi 3$  were assumed to be independent variables, the maximum of  $L$  would be reached for the  $\Phi 3$  being equal to the respective  $\mu$  for all triplets  $\mathbf{H},\mathbf{h}'$ . The shape of the p.d.f. of  $\Phi 3$  for different  $\kappa$  values is depicted in Fig. 3.

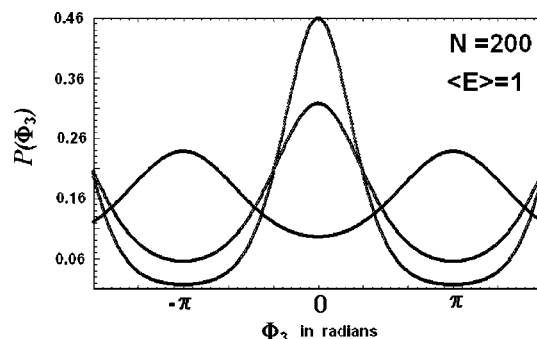


Figure 3 Probability density function of  $\Phi 3 = \psi(-\mathbf{H}) + \varphi(\mathbf{h}') + \varphi(\mathbf{H} - \mathbf{h}')$  for different  $\kappa(\mathbf{H},\mathbf{h}')$  values. The sign of  $\kappa$  determines the position of the maximum of the p.d.f., *i.e.*  $\Phi 3 = 0$  for positive and  $\pi$  for negative  $\kappa$ . The dispersion of the p.d.f. is smaller for higher  $|\kappa|$ : (i)  $\kappa = 1.651$  [ $E(\mathbf{h}) = 3.0$ ,  $E(\mathbf{h}') = 2.5$ ,  $E(\mathbf{h} - \mathbf{h}') = 2.5$ ]; (ii)  $\kappa = 0.875$  [ $E(\mathbf{h}) = 3.0$ ,  $E(\mathbf{h}') = 2.1$ ,  $E(\mathbf{h} - \mathbf{h}') = 1.8$ ]; (iii)  $\kappa = -0.454$  [ $E(\mathbf{k}) = 0.15$ ,  $E(\mathbf{h}') = 2.1$ ,  $E(\mathbf{k} - \mathbf{h}') = 1.8$ ].

## 8. Conclusions

The principal conclusions are: (i) that the phase-refinement procedures based on the maximization of the origin-free modulus sum function yield the maximum-likelihood estimates of the phases; (ii) that the proposed probability density function of the triple-phase sums allows the prediction not only of positive but also of negative triplets. This is achieved through the introduction of the phase  $\psi_{\mathbf{H}}$  of the  $G_{\mathbf{H}}$  which is the principal difference with Cochran's p.d.f. (Cochran, 1955) and is also necessary to preserve the parallelism between triplets and quartets. Effectively, if the likelihood is defined with the  $G_{\mathbf{H}}^2$  as observations, the quartets which result from working out  $G_{\mathbf{H}}(\Phi)^2 = \langle E_{\mathbf{H}'} E_{\mathbf{H}-\mathbf{H}'} \rangle_{\mathbf{H}'} \langle E_{\mathbf{H}''} E_{-\mathbf{H}-\mathbf{H}''} \rangle_{\mathbf{H}''}$  only involve the phases  $\varphi$  and the  $\psi$  are then not necessary (Rius, 1997). This is surely the reason why the p.d.f. predicting the negative quartets was found earlier (Schenk, 1973; Hauptman, 1975; Giacovazzo, 1976).

This work was supported by the Spanish 'Ministerio de Educación y Ciencia' (Project No. MAT2005-01736) and by the 'Generalitat de Catalunya' (SGR2005-00452).

## References

- Cochran, W. (1955). *Acta Cryst.* **8**, 473–478.
- Debaerdemaeker, T., Tate, C. & Woolfson, M. M. (1985). *Acta Cryst.* **A41**, 286–290.
- Fisher, R. A. (1922). *Philos. Trans. R. Soc. London Ser. A*, **222**, 309–368.
- Giacovazzo, C. (1976). *Acta Cryst.* **A32**, 74–82.
- Hauptman, H. (1975). *Acta Cryst.* **A31**, 680–687.
- Hauptman, H. & Karle, H. (1953). *The Solution of the Phase Problem: I. The Centrosymmetric Crystal. Am. Crystallogr. Assoc. Monogr.* No. 3. Wilmington: The Lettershop.
- Hughes, E. W. (1953). *Acta Cryst.* **6**, 871.
- Lipson, H. & Cochran, W. (1966). *The Determination of Crystal Structures*. In *The Crystalline State*, Vol. III, edited by L. Bragg. London: G. Bell and Sons Ltd.
- Mises, R. von (1918). *Phys. Z.* **19**, 490–500.
- Prince, E. (1982). In *Mathematical Techniques in Crystallography*. New York: Springer-Verlag.
- Rius, J. (1993). *Acta Cryst.* **A49**, 406–409.
- Rius, J. (1997). *Acta Cryst.* **D53**, 535–539.
- Rius, J. (2002). *Structure Determination from Powder Diffraction Data*, edited by W. I. F. David, K. Shankland, L. B. McCusker & Ch. Baerlocher, pp. 219–232. *IUCr Monographs on Crystallography* 13. Oxford University Press.
- Rius, J. (2004). *Z. Kristallogr.* **219**, 826–832.
- Rius, J., Torrelles, X., Miravittles, C., Amigó, J. M., Reventós, M. M. (2002). *Acta Cryst.* **A58**, 21–26.
- Schenk, H. (1973). *Acta Cryst.* **A29**, 77–82.
- Weeks, C. M., DeTitta, G. T., Hauptman, H. A., Thuman, P. & Miller, R. (1994). *Acta Cryst.* **A50**, 210–220.