*Acta Cryst.* (1979). A35, 757–764

# A Theoretical Weighting Scheme for Tangent-Formula Development and Refinement and Fourier Synthesis

BY CARMELO GIACOVAZZO

*Istituto di Mineralogia e Petrografia, Università degli Studi, 70121, Bari, Italy*

(*Received 1 December 1978; accepted 20 March 1979*)


## Abstract

A weighting scheme for use in tangent-formula phase development and refinement is derived by application of joint probability distribution functions. It may easily be incorporated in existing computer programs. A weighting scheme for Fourier synthesis is also described; it takes into account the uncertainty of the phases assigned by a direct procedure.


## 1. Introduction

When $G = 2N^{-1/2}|E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h-k}}|$ is *a priori* known, the triple-phase invariant $\varphi = \varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h-k}}$ has a probability distribution about 0 (modulo $2\pi$) of the form (Cochran, 1955)

$$P(\varphi|G) \simeq [2\pi I_0(G)]^{-1} \exp(G\cos\varphi), \tag{1}$$

with variance (Karle & Karle, 1966)

$$V(G) = \frac{\pi^2}{3} + [I_0(G)]^{-1} \sum_{r=1}^{\infty} \frac{I_{2r}(G)}{r^2} - 4[I_o(G)]^{-1} \sum_{r=0}^{\infty} \frac{I_{2r+1}(G)}{(2r+1)^2}. \tag{2}$$

$I_r$ are the modified Bessel functions of the first kind of order $r$. If $\varphi_{\mathbf{k}}$ and $\varphi_{\mathbf{h-k}}$ are known, the conditional distribution of $\varphi_{\mathbf{h}}$ is given by

$$P(\varphi_{\mathbf{h}}|\varphi_{\mathbf{k}}, \varphi_{\mathbf{h-k}}, G) \simeq [2\pi I_o(G)]^{-1} \exp[G\cos(\varphi_{\mathbf{h}} - \theta_{\mathbf{h}})], \tag{3}$$

where $\theta_{\mathbf{h}} = \varphi_{\mathbf{k}} + \varphi_{\mathbf{h-k}}$, with variance given again by (2). When several pairs of known phases are known (3) is replaced by

$$P(\varphi_{\mathbf{h}}|\{\varphi_{\mathbf{k}_j}, \varphi_{\mathbf{h-k}_j}, G_j\}) \simeq [2\pi I_o(\alpha)]^{-1} \exp[\alpha\cos(\varphi_{\mathbf{h}} - \theta_{\mathbf{h}})], \tag{4}$$

where $\{\varphi_{\mathbf{k}_j}, \varphi_{\mathbf{h-k}_j}, G_j\}$ denotes the set of phases and magnitudes *a priori* known. Furthermore, $\theta_{\mathbf{h}}$, the most efficient value for $\varphi_{\mathbf{h}}$, is given by (Karle & Hauptman, 1956)

$$\tan\theta_{\mathbf{h}} = \frac{\sum_j |E_{\mathbf{k}_j} E_{\mathbf{h-k}_j}| \sin(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h-k}_j})}{\sum_j |E_{\mathbf{k}_j} E_{\mathbf{h-k}_j}| \cos(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h-k}_j})} = \frac{T_{\mathbf{h}}}{B_{\mathbf{h}}}; \tag{5}$$

and the corresponding variance for $\varphi_{\mathbf{h}}$ is given again by (2), but

$$\alpha_{\mathbf{h}} = 2N^{-1/2}|E_{\mathbf{h}}|(T_{\mathbf{h}}^2 + B_{\mathbf{h}}^2)^{1/2} \tag{6}$$

replaces $G$.

Unfortunately, in the practical procedures for phase determination the premise on which (4), (5) and (6) is based is not fully satisfied. In fact the phases $\varphi_{\mathbf{k}_j}$, $\varphi_{\mathbf{h-k}_j}$ are themselves uncertain and have an associated variance. So weighted tangent formulae such as (Germain, Main & Woolfson, 1971)

$$\tan\varphi_{\mathbf{h}} = \frac{\sum_j W_{\mathbf{k}_j} W_{\mathbf{h-k}_j} |E_{\mathbf{k}_j} E_{\mathbf{h-k}_j}| \sin(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h-k}_j})}{\sum_j W_{\mathbf{k}_j} W_{\mathbf{h-k}_j} |E_{\mathbf{k}_j} E_{\mathbf{h-k}_j}| \cos(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h-k}_j})} = \frac{T'_{\mathbf{h}}}{B'_{\mathbf{h}}} \tag{7}$$

can usefully replace (5). The form of weighting should ensure that poorly determined phases have little effect in the determination of other phases. Furthermore, since all determined phases are included in the right hand side of (7), the phase determination process may be very efficient. Of course, a necessary condition is that the weight $W_{\mathbf{k}}$ is proportional to the accuracy of the phase $\varphi_{\mathbf{k}}$.

Several weighting schemes have been proposed. Germain, Main & Woolfson (1971) suggested

$$W_{\mathbf{h}} = \tanh\{N^{-1/2}|E_{\mathbf{h}}|(T_{\mathbf{h}}'^2 + B_{\mathbf{h}}'^2)\}. \tag{8}$$

Weighting criterion (8) was criticized by Schenk (1972) and was revealed as unrealistic in several cases. A procedure afterwards adopted by Germain, Main & Woolfson was

$$W_{\mathbf{h}} = \min(0{\cdot}2\alpha, 1{\cdot}0), \tag{9}$$

which proved a more efficient criterion than (8). Its use however is not able to avoid the problem that in multi-

0567-7394/79/050757-08$01.00 

solution procedures the wrong solutions are often characterized by phases, with which final values of $\alpha$ are associated, which are larger than would be expected from theory. Furthermore, weights easily become and stay at unity during phasing procedure.

The weighting scheme in tangent refinement is not a trivial topic. It has been noted in fact (Busetta, 1976; Lessinger, 1976) that the true phases are not stable under some weighted tangent refinement schemes whereas they remain stable if more suitable weighting criteria are used. Recently Hull & Irwin (1978) proposed a more elaborate weighting scheme. As shown by Germain, Main & Woolfson (1970), when the values $G_{\mathbf{h},\mathbf{k}_j}$ are *a priori* known and the phases $\varphi_{\mathbf{k}_j}$, $\varphi_{\mathbf{h}-\mathbf{k}_j}$ are unknown but are supposed to satisfy Cochran (1955) distribution, then the expected value of $\alpha_{\mathbf{h}}^2$ is given by

$$\alpha_E = \sum_j G^2_{\mathbf{h},\mathbf{k}_j} + \sum_{i\neq j} G_{\mathbf{h},\mathbf{k}_i} G_{\mathbf{h},\mathbf{k}_j} \frac{I_1(G_{\mathbf{h},\mathbf{k}_i})\, I_1(G_{\mathbf{h},\mathbf{k}_j})}{I_0(G_{\mathbf{h},\mathbf{k}_i})\, I_0(G_{\mathbf{h},\mathbf{k}_j})}. \quad (10)$$

On the other hand, if the phases $\varphi_{\mathbf{k}_j}$, $\varphi_{\mathbf{h}-\mathbf{k}_j}$ are supposed to be a random set, the expected value of $\alpha_{\mathbf{h}}^2$ is given by

$$\alpha_r^2 = \sum_j G^2_{\mathbf{h},\mathbf{k}_j}. \quad (11)$$

Hull & Irwin's (1978) weighting scheme is the minimum of that given by (9) and

$$W_{\mathbf{h}}' = \psi e^{-x^2} \int_0^x \exp t^2 \, \mathrm{d}t, \quad (12)$$

where $x = \alpha_{\mathbf{h}}/\alpha_E$ and $\psi$ is a normalizing factor chosen so that $W' = 1$ when $x = 1$. $W'$ has its maximum at $x = 1$: when $\alpha_{\mathbf{h}} > \alpha_E$ or $\alpha_{\mathbf{h}} < \alpha_E$, $w < 1$, supporting a more realistic agreement between the calculated and the expected values of $\alpha_{\mathbf{h}}$.

Hull & Irwin's (1978) scheme can be considered more useful than preceding ones; the authors showed in fact that a method which constrains $\alpha_{\mathbf{h}}^2$ to equal $\alpha_E^2$ during tangent-formula refinement is equivalent to forcing the Sayre equation to be obeyed. The scheme however appears criticizable in two aspects: (i) the form of (12) appears to be rather arbitrary. Many other functions can be found which take their maximum at $\alpha = \alpha_E$ where they can assume unitary value; (ii) the scheme assigns $W$ after an *a posteriori* comparison of the experimental value of $\alpha$ with $\alpha_E$. It should therefore be useful to introduce a scheme which leads to more realistic values of $\alpha$ just by taking into account the 'uncertainty' of the various 'known' phases. This is exactly the first aim of this paper. In order to make clear the mathematical approach, we describe in § 2 some important results in the statistics of directional data which are useful for our purposes. In the following we will often use the notation $R = |E|$.

## 2. Some results in the theory of distributions

### 2.1. *The wrapped normal distribution*

A useful representation of this density is

$$W_N(\rho,\theta) = \left[1 + 2 \sum_{p=1}^{\infty} \rho^{p^2} \cos p(\varphi - \theta)\right] / 2\pi, \quad (13)$$

with $0 < \varphi \leq 2\pi$, $0 \leq \rho \leq 1$, $\rho = \exp(-\sigma^2/2)$. The distribution is unimodal and is symmetric about $\theta$. It tends to uniform distribution as $\rho \to 0$ (or $\sigma \to \infty$) while it tends to concentrate at one single point when $\rho \to 1$. From (13),

$$\langle \cos \varphi \rangle = \rho \cos \theta, \quad \langle \sin \varphi \rangle = \rho \sin \theta, \quad (14)$$

$$\mathrm{var}[\cos \varphi] = \tfrac{1}{2}(1 + \rho^4 \cos 2\theta) - \rho^2 \cos^2 \theta, \quad (15)$$

$$\mathrm{var}[\sin \varphi] = \tfrac{1}{2}(1 - \rho^4 \cos 2\theta) - \rho^2 \sin^2 \theta. \quad (16)$$

### 2.2. *Von Mises distribution*

The Von Mises (1918) distribution

$$M(\varphi;\theta,G) = [2\pi I_0(G)]^{-1} \exp[G \cos(\varphi - \theta)] \quad (17)$$

is unimodal and symmetric about $\theta$. The mode is at $\varphi = \theta$ and the antimode is at $\varphi = \theta + \pi$. The ratio of the density at the mode to the density at the antimode is given by $\exp(2G)$ so that the larger the value of $G$ the greater the clustering around the mode. For $G = 0$, $M(\varphi;\theta,G)$ reduces to the uniform distribution. From (17),

$$\langle \cos \varphi \rangle = \frac{I_1(G)}{I_0(G)} \cos \theta, \quad \langle \sin \varphi \rangle = \frac{I_1(G)}{I_0(G)} \sin \theta, \quad (18)$$

$$\mathrm{var}[\cos \varphi] = \frac{1}{2}\left[1 + \frac{I_2(G)}{I_0(G)} \cos 2\theta\right] - \left[\frac{I_1(G)}{I_0(G)}\right]^2 \cos^2 \theta, \quad (19)$$

$$\mathrm{var}[\sin \varphi] = \frac{1}{2}\left[1 - \frac{I_2(G)}{I_0(G)} \cos 2\theta\right] - \left[\frac{I_1(G)}{I_0(G)}\right]^2 \sin^2 \theta. \quad (20)$$

The Fourier expansion of (17) gives

$$M(\varphi;\theta,G) = \left\{1 + \frac{2}{I_0(G)} \sum_{p=1}^{\infty} I_p(G) \cos[p(\varphi - \theta)]\right\} / 2\pi. \quad (21)$$

### 2.3. *Relation between Von Mises and wrapped normal distributions*

Von Mises and wrapped normal distributions can be made to approximate each other closely. To this end (Stephens, 1963), (21) can be compared with (13). When $p = 1$ the coefficients in (13) and (21) agree if

$$\rho = I_1(G)/I_0(G). \tag{22}$$

The coefficients for $p > 1$ will agree approximately if

$$[I_1(G)/I_0(G)]^{p^2} \simeq I_p(G)/I_0(G). \tag{23}$$

Exact agreement occurs between the two distributions when $G = 0$ (both become the uniform distribution). When $G$ is large the left hand side of (23) may be approximated by

$$\left[1 - \frac{1}{2G} - \frac{1}{8G^2} - \frac{1}{8G^3} - \ldots\right]^{p^2},$$

so that the asymptotic result is $(1 - 1/2G)^{p^2} \simeq 1 - p^2/2G$. This is also the asymptotic value of the right hand side if $I_p$ is expanded according to (Abramowitz & Stegun, 1970)

$$I_p(G) \simeq (2\pi G)^{-1/2} \exp(G) \left[1 - \frac{(4p^2-1)}{8G} + \frac{(4p^2-1)(4p^2-9)}{2!(8G)^2} - \ldots\right].$$

It appears therefore that the agreement should be good for all $G$'s. Stephen's (1963) calculations are quite satisfactory; the 'best fit' between the two distributions occurs when (22) is closely satisfied. In order to translate those results in terms of the parameters usually involved in direct procedures for phase estimation we note: (*a*) the expected values of $\varphi$, $\cos\varphi$, $\sin\varphi$ coincide when calculated according to $M(\varphi;\theta,G)$ and $W_N[I_1(G)/I_0(G),\theta]$; (*b*) in the same condition the calculated variances are in good agreement. In Table 1 are shown, for given values of $G$, the cosine and sine variances calculated according to $M(\varphi;0,G)$ (column labelled $V_M$) and to $W_N[I_1(G)/I_0(G),0]$ (column labelled $V_W$).

Table 1. *Cosine and sine variances for given values of G*

$V_M$; calculated according to $M(\varphi; 0, G)$. $V_W$; calculated according to $W_N[I_1(G)/I_0(G)]$.

| $G$ | $V_M(\cos)$ | $V_W(\cos)$ | $V_M(\sin)$ | $V_W(\sin)$ |
|---|---|---|---|---|
| 0·0 | 0·500 | 0·500 | 0·500 | 0·500 |
| 1·0 | 0·354 | 0·321 | 0·446 | 0·480 |
| 2·0 | 0·164 | 0·132 | 0·349 | 0·381 |
| 3·0 | 0·074 | 0·059 | 0·270 | 0·285 |
| 4·0 | 0·039 | 0·033 | 0·216 | 0·222 |
| 5·0 | 0·023 | 0·020 | 0·179 | 0·181 |
| 6·0 | 0·016 | 0·014 | 0·152 | 0·154 |
| 7·0 | 0·011 | 0·010 | 0·132 | 0·133 |
| 8·0 | 0·009 | 0·008 | 0·117 | 0·118 |

### 2.4. *Convolutions of wrapped normal distributions*

Let $\varphi_1, \varphi_2, \ldots, \varphi_n$ be $n$ mutually independent variables and let $\theta_i$ have probability distribution $W_N(\rho_i,\theta_i)$. Then the variable sum $\varphi_1 + \varphi_2 + \cdots + \varphi_n$ has probability distribution function $W_N(\rho_1\rho_2\ldots\rho_n, \theta_1 + \theta_2 + \cdots + \theta_n)$.

### 2.5. *Convolutions of Von Mises distributions*

If $\varphi_1$ and $\varphi_2$ are two mutually independent variables distributed according to $M(\varphi;\theta_1,G_1)$ and $M(\varphi;\theta_2,G_2)$ respectively, the probability distribution function of $\varphi = \varphi_1 + \varphi_2$ is given by

$$P(\varphi_1 + \varphi_2) = [2\pi I_0(G_1)\, I_0(G_2)]^{-1} I_0[G_1^2 + G_2^2 + 2G_1 G_2 \times \cos(\varphi - \theta_1 - \theta_2)], \tag{24}$$

which is not a Von Mises distribution. Equation (24), however, can be approximated by a Von Mises distribution by applying the results described in §§ 2.3 and 2.4. We first approximate $M(\varphi;\theta_1,G_1)$ and $M(\varphi;\theta_2,G_2)$ by $W_N[I_1(G_1)/I_0(G_1),\theta_1]$ and $W_N[I_1(G_2)/I_0(G_2),\theta_2]$, respectively. Because of § 2.4, the convolution of these two distributions is the wrapped normal distribution

$$W_N\left(\frac{I_1(G_1)}{I_0(G_1)}\frac{I_1(G_2)}{I_0(G_2)}, \theta_1 + \theta_2\right),$$

which in turn may be approximated by the Von Mises distribution $M(\varphi; \theta_1 + \theta_2, G)$, where $G$ satisfies

$$D_1(G) = D_1(G_1)\, D_1(G_2). \tag{25}$$

In (25) and in the following formulae, $D_p(G_i)$ stands for $I_p(G_i)/I_0(G_i)$. This result may be generalized; let $\varphi_1, \varphi_2, \ldots, \varphi_n$ be $n$ mutually independent variables and let any $\varphi_i$ be distributed according to $M(\varphi_i;\theta_i,G_i)$. Then the variable sum $\varphi = \varphi_1 + \varphi_2 + \cdots + \varphi_n$ can be approximated by the Von Mises distribution $M(\varphi; \theta_1 + \theta_2 + \cdots + \theta_n, G)$ where $G$ satisfies

$$D_1(G) = D_1(G_1)\, D_1(G_2) \ldots D_1(G_n). \tag{26}$$

## 3. Some formulae

In the main text, evaluation of the integral

$$\frac{1}{2\pi}\int_0^{2\pi} \exp\left[\sum_k C_k \cos(\varphi + \alpha_k)\right] d\varphi \tag{27}$$

is of frequent occurrence. For real $C_k$ in (27) we define $X$ and $\xi$ by

$$X = \left[\sum_k \sum_l C_k C_l \cos(\alpha_k - \alpha_l)\right]^{1/2}, \tag{28}$$

$$\xi = \arc\cos\left[\frac{\sum_k C_k \cos \alpha_k}{X}\right]. \tag{29}$$

From (28) and (29) we have

$$\sum_k C_k \cos(\varphi + \alpha_k) = X \cos(\varphi + \xi),$$

so that

$$\frac{1}{2\pi}\int_0^{2\pi} \exp\left[\sum_k C_k \cos(\varphi + \alpha_k)\right] d\varphi = I_0(X). \tag{30}$$

The following formulae from the theory of Bessel functions are also collected for convenient reference:

$$\frac{1}{2\pi}\int_0^{2\pi} \cos m\varphi \exp(z \cos\varphi)\, d\varphi = I_m(z), \tag{31}$$

$$\int_0^{2\pi} \sin m\varphi \exp(-z \cos\varphi)\, d\varphi = 0, \tag{32}$$

$$I_o[(z_1^2 + z_2^2 + 2z_1 z_2 \cos\varphi)^{1/2}] = I_0(z_1) I_0(z_2) + 2\sum_{p=1}^{\infty} I_p(z_1) I_p(z_2) \cos p\varphi. \tag{33}$$

### 4. The distribution of $\varphi_h$ given the distributions of $\varphi_k$ and $\varphi_{h-k}$

In the following we will suppose that the moduli of the normalized structure factors are known without uncertainty. In tangent schemes, however, $\varphi_k$ and $\varphi_{h-k}$ are usually uncertain and have an associated variance; thus the crude application of (3) overestimates the reliability of $\varphi_h$. However, if $P(\varphi_k)$ and $P(\varphi_{h-k})$ are known, then $P(\varphi_h|\varphi_k,\varphi_{h-k},G)$ can be replaced by

$$\int_0^{2\pi}\int_0^{2\pi} P(\varphi_h|\varphi_k,\varphi_{h-k},G)\, P(\varphi_k)\, P(\varphi_{h-k})\, d\varphi_k\, d\varphi_{h-k}. \tag{34}$$

In (34), $R_h$, $R_k$, $R_{h-k}$ and the conditional probability distributions of $\varphi_k$ and $\varphi_{h-k}$ are the *a priori* information available. It is no matter in our approach what kind of *a priori* information has been exploited in order to define $P(\varphi_k)$ and $P(\varphi_{h-k})$. We only suppose that $P(\varphi_k)$ and $P(\varphi_{h-k})$ are distributed according to $M(\varphi_k;\theta_k,\alpha_k)$ and $M(\varphi_{h-k};\theta_{h-k},\alpha_{h-k})$, respectively. Furthermore, in accordance with the usual tangent procedure, (34) requires that the distributions of $\varphi_k$, $\varphi_{h-k}$ around their expected values $\theta_k$, $\theta_{h-k}$ are mutually independent. From now on we will refer to (34) as $P(\varphi_h\|\varphi_k,\varphi_{h-k})$. Then

$$P(\varphi_h\|\varphi_k, \varphi_{h-k}) \simeq [2\pi I_0(G)\, 2\pi I_0(\alpha_k)\, 2\pi I_0(\alpha_{h-k})]^{-1} \times \int_0^{2\pi}\int_0^{2\pi} \exp(Q)\, d\varphi_k\, d\varphi_{h-k}, \tag{35}$$

where

$$Q = G\cos(\varphi_h - \varphi_k - \varphi_{h-k}) + \alpha_k \cos(\varphi_k - \theta_k) + \alpha_{h-k}\cos(\varphi_{h-k} - \theta_{h-k}). \tag{36}$$

If terms in (36) which depend on $\varphi_{h-k}$ are combined, the integration in (35) with respect to $\varphi_{h-k}$ may be performed according to formulae (27)–(30). Then

$$P(\varphi_h\|\varphi_k, \varphi_{h-k}) \simeq (2\pi)^{-2}[I_0(G)\, I_0(\alpha_k)\, I_0(\alpha_{h-k})]^{-1} \times \int_0^{2\pi} I_0(X_1) \exp \alpha_k \cos(\varphi_k - \theta_k)\, d\varphi_k, \tag{37}$$

where

$$X_1 = [G^2 + \alpha_{h-k}^2 + 2G\alpha_{h-k}\cos(\varphi_h - \varphi_k - \theta_{h-k})]^{1/2}.$$

Expanding $I_0(X_1)$ according to (33) and integrating (37) with respect to $\varphi_k$ by (31) give

$$P(\varphi_h\|\varphi_k,\varphi_{h-k}) \simeq \frac{1}{2\pi} + \frac{1}{\pi}\sum_{p=1}^{\infty} D_p(G)\, D_p(\alpha_k)\, D_p(\alpha_{h-k}) \times \cos p(\varphi_h - \theta_k - \theta_{h-k}). \tag{38}$$

Equation (38) is the result of a twofold convolution of Von Mises distributions. Because of § 2.5 we can approximate (38) by the Von Mises distribution

$$P(\varphi_h) \simeq [2\pi I_0(\alpha_h)]^{-1} \exp[\alpha_h \cos(\varphi_h - \theta_k - \theta_{h-k})], \tag{39}$$

where $\alpha_h$ is the solution of the equation

$$D_1(\alpha_h) = D_1(G)\, D_1(\alpha_k)\, D_1(\alpha_{h-k}). \tag{40}$$

Equation (39), as (3), has its maximum when $\varphi_h = \theta_h = \theta_k + \theta_{h-k}$, where $\theta_k$ and $\theta_{h-k}$ are the 'known' phase values of $\varphi_k$ and $\varphi_{h-k}$ in standard tangent procedures. The asymptotic behaviour of (39) may be described by: (i) let $\varphi_k$ and $\varphi_{h-k}$ be *a priori* known without uncertainty. Then $D_1(\alpha_k)$ and $D_1(\alpha_{h-k})$ equal unity so that (39) reduces to (3). From a mathematical point of view this case involves in (34) the conditions $P(\varphi_k) = \delta(\varphi_k - \theta_k)$ and $P(\varphi_{h-k}) = \delta(\varphi_{h-k} - \theta_{h-k})$ where $\delta(x)$ is the Dirac function; (ii) if $G_k$ or (and) $G_{h-k}$ is (are) equal to zero, the $\alpha_h$ vanishes too. From a mathematical point of view the twofold convolution of the probability functions involves one (two) uniform distribution(s) and gives a uniform distribution for $\varphi_h$.

Table 2. *Values of* $\alpha_h$ *calculated from equation* (40) *for given values of* $G$, $\alpha_k$ *and* $\alpha_{h-k}$

| $G$ | $\alpha_k$ | $\alpha_{h-k}$ | $\alpha_h$ |
|---|---|---|---|
| 3·0 | 6·0 | 8·0 | 2·0 |
| | 3·0 | 4·0 | 1·4 |
| | 3·0 | 3·0 | 1·3 |
| | 2·0 | 3·0 | 1·0 |
| | 2·0 | 2·0 | 0·9 |
| | 2·0 | 1·0 | 0·6 |
| | 1·0 | 1·0 | 0·4 |

In order to give a practical idea of the effects of (40), the values of $\alpha_h$ for appropriate values of $G$, $\alpha_k$ and $\alpha_{h-k}$ are given in Table 2.

## 5. The distribution of $\varphi_h$ given the distribution of more pairs $\varphi_k, \varphi_{h-k}$

Let us suppose that a phase $\varphi_h$ may be assigned by tangent methods *via* two triplet relationships involving the pairs $(\varphi_{k_1}, \varphi_{h-k_1})$, $(\varphi_{k_2}, \varphi_{h-k_2})$. Provided the distributions of $\varphi_{k_1}, \varphi_{h-k_1}, \varphi_{k_2}, \varphi_{h-k_2}$ (*a*) are mutually independent, (*b*) are known to be the Von Mises distributions $M(\varphi_{k_1};\theta_1,\alpha_1)$, $M(\varphi_{h-k_1};\theta'_1,\alpha'_1)$, $M(\varphi_{k_2};\theta_2,\alpha_2)$, $M(\varphi_{h-k_2};\theta'_2,\alpha'_2)$, respectively (in other words $\theta_1$, $\theta'_1$, $\theta_2$, $\theta'_2$ are the 'known' values of $\varphi_{k_1}, \varphi_{h-k_1}, \varphi_{k_2}, \varphi_{h-k_2}$); then the distribution of $\varphi_h$ may be written, in accordance with § 4, as

$$\begin{aligned}P(\varphi_1 \| \varphi_{k_1}, \varphi_{h-k_1}, \varphi_{k_2}, \varphi_{h-k_2}) &\simeq \int_0^{2\pi}\cdots\int_0^{2\pi} P(\varphi_{k_1})\,P(\varphi_{h-k_1})\,P(\varphi_{k_2})\,P(\varphi_{h-k_2})\,P(\varphi_h|\varphi_{k_1}, \ldots, \varphi_{h-k_2}, G_1, G_2)\,\mathrm{d}\varphi_{k_1}\ldots\mathrm{d}\varphi_{h-k_2}\\ &= (2\pi)^{-5}[I_0(\gamma_h)\,I_0(\alpha_1)\,I_0(\alpha'_1)\,I_0(\alpha_2)\,I_0(\alpha'_2)]^{-1}\\ &\quad\times\int_0^{2\pi}\cdots\int_0^{2\pi}\exp(Q)\,\mathrm{d}\varphi_{k_1}\,\mathrm{d}\varphi_{h-k_1}\,\mathrm{d}\varphi_{k_2}\,\mathrm{d}\varphi_{h-k_2},\end{aligned} \qquad (41)$$

where

$$Q = Q_1 + Q_2,$$

$$Q_1 = G_1\cos(\varphi_h - \varphi_{k_1} - \varphi_{h-k_1}) + \alpha_1\cos(\varphi_{k_1} - \theta_1) + \alpha'_1\cos(\varphi_{h-k_1} - \theta'_1),$$

$$Q_2 = G_2\cos(\varphi_h - \varphi_{k_2} - \varphi_{h-k_2}) + \alpha_2\cos(\varphi_{k_2} - \theta_2) + \alpha'_2\cos(\varphi_{h-k_2} - \theta'_2),$$

$$G_1 = 2R_h R_{k_1} R_{h-k_1}/\sqrt{N},$$

$$G_2 = 2R_h R_{k_2} R_{h-k_2}/\sqrt{N},$$

$$\gamma_h = [G_1^2 + G_2^2 + 2G_1G_2\cos(\varphi_{k_1} + \varphi_{h-k_1} - \varphi_{k_2} - \varphi_{h-k_2})]^{1/2}.$$

The integration of (41) is not straightforward, mostly because $I_0(\gamma_h)$ depends on the variables $\varphi_{k_1}, \varphi_{h-k_1}, \varphi_{k_2}, \varphi_{h-k_2}$. We introduce therefore the following approach. We first make the approximation

$$P(\varphi|\varphi_{k_1},\ldots,\varphi_{h-k_2}, G_1, G_2) \simeq SP(\varphi_h|\varphi_{k_1}, \varphi_{h-k_1}, G_1)\,P(\varphi_h|\varphi_{k_2}, \varphi_{h-k_2}, G_2),$$

where $S$ is a suitable normalizing factor. Now the right hand side of (41) becomes

$$\begin{aligned}&S\frac{1}{(2\pi)^3}\frac{1}{I_0(G_1)\,I_0(\alpha_1)\,I_0(\alpha'_1)}\int_0^{2\pi}\int_0^{2\pi}\exp Q_1\,\mathrm{d}\varphi_{k_1}\,\mathrm{d}\varphi_{h-k_1}\\ &\times\frac{1}{(2\pi)^3}\frac{1}{I_0(G_2)\,I_0(\alpha_2)\,I_0(\alpha'_2)}\int_0^{2\pi}\int_0^{2\pi}\exp Q_2\,\mathrm{d}\varphi_{k_2}\,\mathrm{d}\varphi_{h-k_2},\end{aligned} \qquad (42)$$

which, because of the results described in § 4, reduces to

$$S[2\pi I_0(\beta_1)]^{-1}\exp[\beta_1\cos(\varphi_h - \theta_1 - \theta'_1)]\,[2\pi I_0(\beta_2)]^{-1} \times\exp[\beta_2\cos(\varphi_h - \theta_2 - \theta'_2)]. \qquad (43)$$

In (43), $\beta_1$ and $\beta_2$ satisfy

$$D_1(\beta_1) = D_1(G_1)\,D_1(\alpha_1)\,D_1(\alpha'_1),$$

$$D_1(\beta_2) = D_1(G_2)\,D_1(\alpha_2)\,D_1(\alpha'_2),$$

respectively. The normalization of (43) leads then to

$$P(\varphi_h \| \varphi_{k_1}, \ldots, \varphi_{h-k_m}) \simeq [2\pi I_0(\alpha_h)]^{-1}\exp[\alpha_h\cos(\varphi_h - \theta_h)], \qquad (44)$$

where

$$\alpha_h = \left\{\left[\sum_{j=1}^m \beta_j\cos(\theta_j + \theta'_j)\right]^2 + \left[\sum_{j=1}^m \beta_j\sin(\theta_j + \theta'_j)\right]^2\right\}^{1/2}, \qquad (45)$$

$$\tan\theta_h = \frac{\sum_{j=1}^m \beta_j\sin(\theta_j + \theta'_j)}{\sum_{j=1}^m \beta_j\cos(\theta_j + \theta'_j)}, \qquad (46)$$

and $m = 2$.

If $\varphi_h$ may be estimated by $m > 2$ pairs $(\varphi_{k_j}, \varphi_{h-k_j})$, (44), (45) and (46) are still valid whatever $m$ may be. (46) is our weighted tangent formula. We note that the weights $W$ which appear in (7) do not appear explicitly in (46); furthermore, $\theta_j$ and $\theta'_j$ are the same quantities as $\varphi_{k_j}$ and $\varphi_{h-k_j}$. $\alpha_h$ as given by (45) measures the accuracy with which the value $\theta_h$ has been assigned to $\varphi_h$.

## 6. Weighting criteria for phases not evaluated by Von Mises distributions

One-, two-, three-phase structure seminvariants and four-, five-phase structure invariants are not distributed according to Von Mises distributions, even if in several cases useful Von Mises distributions have been devised which reliably estimate them (Giacovazzo, 1976, 1977, 1978). When a phase $\varphi$ is estimated by probabilistic formulae, both the expected value $\theta$ and the variance value $v$ are usually evaluated. It may be expected that without large error, $\varphi$ can be used as a known phase in tangent procedures as it was distributed according to $M(\varphi;\theta,\alpha)$, where $\alpha$ is the value of the concentration parameter which corresponds to the variance $v$.

### 7. The sign probability for $E_{\mathbf{h}}$ given the sign probabilities for $E_{\mathbf{k}_1}, E_{\mathbf{h}-\mathbf{k}_1}, E_{\mathbf{k}_2}, E_{\mathbf{h}-\mathbf{k}_2},\ldots$

For the centrosymmetric space groups we will denote by $S_{\mathbf{h}}, S_{\mathbf{k}},\ldots$, the signs of $E_{\mathbf{h}}, E_{\mathbf{k}},\ldots$, and by $S_{\mathbf{h},\mathbf{k}}$ the sign of any triplet $E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}$. As is well known (Cochran & Woolfson, 1955), the probability density of $S_{\mathbf{h},\mathbf{k}}$ given $G = R_{\mathbf{h}}R_{\mathbf{k}}R_{\mathbf{h}-\mathbf{k}}/\sqrt{N}$ is

$$P(S_{\mathbf{h},\mathbf{k}}|G) \simeq 0\cdot5 + 0\cdot5 \tanh G. \qquad (47)$$

Correspondingly, the probability density for $S_{\mathbf{h}}$ given $S_{\mathbf{k}}$ and $S_{\mathbf{h}-\mathbf{k}}$ is given by

$$P(S_{\mathbf{h}}|S_{\mathbf{k}}, S_{\mathbf{h}-\mathbf{k}}, G) \simeq 0\cdot5 + 0\cdot5 \tanh(S_{\mathbf{h}}S_{\mathbf{k}}S_{\mathbf{h}-\mathbf{k}}G). \qquad (48)$$

In the practice of direct procedures, $S_{\mathbf{k}}$ and $S_{\mathbf{h}-\mathbf{k}}$ are known with some uncertainty due to the fact that they are usually assigned by a probabilistic approach. Let $S_{\mathbf{k}}$ and $S_{\mathbf{h}-\mathbf{k}}$ have been assigned with probability values given by

$$P(S_{\mathbf{k}}) \simeq 0\cdot5 + 0\cdot5 \tanh(S_{\mathbf{k}}\alpha_{\mathbf{k}}), \qquad (49)$$

$$P(S_{\mathbf{h}-\mathbf{k}}) \simeq 0\cdot5 + 0\cdot5 \tanh(S_{\mathbf{h}-\mathbf{k}}\alpha_{\mathbf{h}-\mathbf{k}}), \qquad (50)$$

respectively, no matter what kind of *a priori* information has been exploited in order to obtain (49) and (50). According to § 4, the probability density (48) may be replaced by

$$P(S_{\mathbf{h}}\|S_{\mathbf{k}}, S_{\mathbf{h}-\mathbf{k}}) \simeq \sum_{S_{\mathbf{k}},S_{\mathbf{h}-\mathbf{k}}=\pm1} P(S_{\mathbf{h}}|S_{\mathbf{k}}, S_{\mathbf{h}-\mathbf{k}}, G) \times P(S_{\mathbf{k}})P(S_{\mathbf{h}-\mathbf{k}}). \qquad (51)$$

The asymptotic behaviour of (51) may be described by: (i) if $S_{\mathbf{k}}$ and $S_{\mathbf{h}-\mathbf{k}}$ are known to be equal to $\theta_{\mathbf{k}}$ and $\theta_{\mathbf{h}-\mathbf{k}}$ without uncertainty, then

$$P(S_{\mathbf{k}}=\theta_{\mathbf{k}}) = P(S_{\mathbf{h}-\mathbf{k}}=\theta_{\mathbf{h}-\mathbf{k}}) = 1,$$

$$P(S_{\mathbf{k}}=-\theta_{\mathbf{k}}) = P(S_{\mathbf{h}-\mathbf{k}}=-\theta_{\mathbf{h}-\mathbf{k}}) = 0.$$

Therefore (51) reduces to $P(S_{\mathbf{h}}|S_{\mathbf{k}}, S_{\mathbf{h}-\mathbf{k}}, G)$; (ii) if at least one of $P(S_{\mathbf{k}})$ and $P(S_{\mathbf{h}-\mathbf{k}})$ equals 0·5 then also $P(S_{\mathbf{h}}\|S_{\mathbf{k}}, S_{\mathbf{h}-\mathbf{k}}, G)$ equals 0·5.

In order to show the differences between the probability values for $S_{\mathbf{h}}$ calculated according to (48) and (51), we give in Table 3, for a fixed sign probability calculated according to (48), the corresponding values of (51) when (49) and (50) assume some fixed values.

Table 3. *Values for $S_{\mathbf{h}}$ calculated from equation* (51) *for a fixed sign probability calculated by equation* (48)

Equations (49) and (50) assume some fixed values.

| (48) | (49) | (50) | (51) |
|---|---|---|---|
| 0·95 | 0·95 | 0·95 | 0·86 |
| | 0·90 | 0·95 | 0·82 |
| | 0·90 | 0·90 | 0·79 |
| | 0·80 | 0·85 | 0·69 |
| | 0·80 | 0·80 | 0·66 |

If the sign of $E_{\mathbf{h}}$ is estimated *via r* pairs $E_{\mathbf{k}_j}, E_{\mathbf{h}-\mathbf{k}_j}$, then the sign probability for $E_{\mathbf{h}}$ is given by

$$\sum_{S_{\mathbf{k}_1},S_{\mathbf{k}_2},\ldots,S_{\mathbf{h}-\mathbf{k}_r}=\pm1} P(S_{\mathbf{h}}|S_{\mathbf{k}_1}, S_{\mathbf{h}-\mathbf{k}_1},\ldots, S_{\mathbf{h}-\mathbf{k}_r}) \times P(S_{\mathbf{k}_1})P(S_{\mathbf{h}-\mathbf{k}_1})\ldots P(S_{\mathbf{h}-\mathbf{k}_r}). \qquad (52)$$

The calculation of (52) may be time consuming when $r$ is large. A useful approximation of (52) which has the advantage of being algebraically similar to the classical Woolfson formulation is

$$P(S_{\mathbf{h}}) = 0\cdot5 + 0\cdot5 \tanh \sum_{j=1}^{r} \beta_j,$$

where the $\beta_j$'s are obtained by the following two steps: (i) (51) is calculated for each pair $E_{\mathbf{k}_j}, E_{\mathbf{h}-\mathbf{k}_j}$, let $P_j(S_{\mathbf{h}})$ be its value; (ii) putting $P_j = 0\cdot5 + 0\cdot5 \tanh\beta_j$ and inverting with respect to $\beta_j$ to give $\beta_j = \operatorname{arctanh}(2P_j - 1)$.

### 8. Weighting criteria for the starting set of phases

To start the phase determination with tangent procedure a number of phases are considered to be *a priori* known: (*a*) origin-defining phases; (*b*) symbolic phases to which all combinations of the values $\pm\pi/4$, $\pm3\pi/4$ are given if the phases are non-centrosymmetrical, the value 0 or $\pi$ if they are centrosymmetrical; (*c*) phases anyhow determined by probabilistic formulae.

Phases in categories (*a*), (*b*), (*c*) play a different role in the phase determination process. Furthermore they can be considered 'known' with different accuracy so that distributions proper to the category are needed.

For the phases in (*a*), the problem is quite trivial. Let us suppose that the value of any origin defining phase $\varphi$ can be arbitrarily fixed without error. If $\theta$ is its chosen value, the theoretical distribution of $\varphi$ can be assumed to be the Dirac function $\delta(\varphi-\theta)$. From our point of view this is equivalent to associating with $\varphi$ the distribution $M(\varphi;\theta,\alpha)$, where $\alpha$ must be large enough to simulate (for the practice of tangent procedures) the Dirac function: $\alpha > 20$ may be a reasonable choice. If $\varphi$ is a centrosymmetrical phase, its sign can be considered known with unitary probability.

Because of quadrant permutation, the maximum error associable with any non-centrosymmetric reflexion $\varphi$ belonging to category (*b*) is 45°. Thus, for any permutation, $\varphi$ is uniformly distributed in one quadrant with expected value $\theta$ equal to the mean direction of the quadrant, and variance equal to $\pi^2/48$. Our procedure suggests for $\varphi$ a Von Mises distribution with the same mean value and the same value of the variance, $M(\varphi;\theta,5\cdot 5)$ can be a reasonable distribution. If $\varphi$ is centrosymmetrical, the sign permutation process assigns to $\varphi$ all the symmetry allowed values. Thus for any permutation the value of $\varphi$ can be considered known without ambiguity.

The procedure can be readily extended to the 'magic integers' representation of the phases. Main (1977) has given, for the most useful sequences of magic integers, the root mean square error involved in the representation of phases.

The use in tangent procedures of the phases in category (*c*) is discussed in § 6.

## 9. Weights for centrosymmetric reflexions in non-centrosymmetric space groups

Let us suppose that $\varphi_{\mathbf{h}}$ is a centrosymmetric reflexion and $\varphi_{\mathbf{k}}$, $\varphi_{\mathbf{h-k}}$ are general. If $\varphi_{\mathbf{k}}$ and $\varphi_{\mathbf{h-k}}$ are known without uncertainty, then

$$P(E_{\mathbf{h}}, \varphi_{\mathbf{k}}, \varphi_{\mathbf{h-k}}, R_{\mathbf{k}}, R_{\mathbf{h-k}}) \simeq \frac{1}{\pi^2}\frac{1}{\sqrt{2\pi}} R_2 R_3 \exp\left[-\frac{E_1^2}{2} - R_2^2 - R_3^2 + \sqrt{2}\frac{E_1 R_2 R_3}{\sqrt{N}}\cos(\varphi_2+\varphi_3)\right],$$

from which

$$P(\varphi_{\mathbf{h}}|\varphi_{\mathbf{k}}, \varphi_{\mathbf{h-k}}, G') \simeq \frac{1}{L}\exp[G'\cos(\varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h-k}})].$$

$L$ is a constant which does not depend on $\varphi_{\mathbf{h}}$ and $G' = G/\sqrt{2}$. Since $\varphi_{\mathbf{h}}$ can assume two values only – *i.e.* $\theta_{\mathbf{h}}$ and $\theta_{\mathbf{h}} + \pi$ – we have

$$P(\theta_{\mathbf{h}}|\varphi_{\mathbf{k}}, \varphi_{\mathbf{h-k}}, G') \simeq \frac{\exp[G'\cos(\theta_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h-k}})]}{2\cosh[G'\cos(\theta_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h-k}})]} = 0\cdot 5 + 0\cdot 5\tanh[G'\cos(\theta_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h-k}})]. \quad (53)$$

If the argument of the hyperbolic tangent is positive, the relation $\varphi_{\mathbf{h}} = \theta_{\mathbf{h}}$ probably holds, if it is negative then the value $\theta_{\mathbf{h}} + \pi$ must be assigned to $\varphi_{\mathbf{h}}$. The larger the absolute value of the cosine, the more reliable is the phase indication. If $\cos(\theta_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h-k}}) = 0$ then $\varphi_{\mathbf{h}}$ is not determined.

This result may be extended to the more general case in which we know the distributions of $\varphi_{\mathbf{k}}$ and $\varphi_{\mathbf{h-k}}$. The procedure described in § 4 suggests that the right hand side of (53) is a useful approximation of $P(\theta_{\mathbf{h}}||\varphi_{\mathbf{k}}, \varphi_{\mathbf{h-k}})$ provided $\beta$ replaces $G'$, where

$$D_1(\beta) = D_1(G')\, D_1(\alpha_{\mathbf{k}})\, D_1(\alpha_{\mathbf{h-k}}).$$

If several pairs $\varphi_{\mathbf{k}_j}$, $\varphi_{\mathbf{h-k}_j}$ contribute to define $\varphi_{\mathbf{h}}$, then

$$P(\theta_{\mathbf{h}}||\{\varphi_{\mathbf{k}_j}, \varphi_{\mathbf{h-k}_j}\}) \simeq 0\cdot 5 + 0\cdot 5\tanh\left[\sum_j \beta_j\cos(\theta_{\mathbf{h}} - \varphi_{\mathbf{k}_j} - \varphi_{\mathbf{h-k}_j})\right].$$

Let us suppose now that the argument of the hyperbolic tangent is negative, then we will assign to $\varphi_{\mathbf{h}}$ the value $\theta_{\mathbf{h}} + \pi$ with a probability larger than 0·5. What is the weight to associate with the relation $\varphi_{\mathbf{h}} = \theta_{\mathbf{h}} + \pi$? In accordance with § 6 we suggest the following procedure:

(*a*) the variance of the relation $\varphi_{\mathbf{h}} = \theta_{\mathbf{h}} + \pi$ is given by

$$V = \pi^2 P(\varphi_{\mathbf{h}} = \theta_{\mathbf{h}} + \pi)\,[1 - P(\varphi_{\mathbf{h}} = \theta_{\mathbf{h}} + \pi)];$$

(*b*) the concentration parameter of the Von Mises distribution with variance $V$ is calculated and is used as $\alpha_{\mathbf{h}}$. However, the procedure can strongly overestimate $\alpha_{\mathbf{h}}$.

## 10. Generalized tangent refinement

So far we have described a tangent refinement procedure which only uses triplet relationships. However, tangent formulae which hold for quartet invariants and for two-phase seminvariants have already been given (Giacovazzo, 1976, 1977). Let us now calculate the reliability of $\varphi_{\mathbf{h}_1}$ when $\varphi_{\mathbf{h}_2}, \ldots, \varphi_{\mathbf{h}_n}$ are 'known' under the hypothesis: (i) $\varphi_{\mathbf{h}_1} + \cdots + \varphi_{\mathbf{h}_n}$ is a structure seminvariant; (ii) the distribution of $\varphi_{\mathbf{h}_1}$, when $\varphi_{\mathbf{h}_2}, \ldots, \varphi_{\mathbf{h}_n}$ and a suitable set of diffraction magnitudes are known, may be approximated by the Von Mises distribution $M(\varphi_{\mathbf{h}_1};\theta_1,G)$; (iii) the distributions of $\varphi_{\mathbf{h}_2}, \ldots, \varphi_{\mathbf{h}_n}$ are $M(\varphi_{\mathbf{h}_2};\theta_2,\alpha_2), \ldots, M(\varphi_{\mathbf{h}_n};\theta_n,\alpha_n)$. Then (40) is replaced by

$$D_1(\alpha_1) = D_1(G)\, D_1(\alpha_2) \ldots D_1(\alpha_n). \quad (54)$$

The generalization of (45) and (46) is now straightforward. An important conclusion arises from (53): the larger $n$ is, the smaller in average will $\alpha_1$ be. In other words, structure seminvariants with a small number of phases are in general more advantageous in tangent refinement than those with a larger number of phases. Therefore, an easy propagation of the errors must be expected for these seminvariants.

## 11. Weighted Fourier synthesis

At the end of the phasing procedure the final phases have associated $\alpha$ weights; small weights indicate uncertain phases. Germain, Main & Woolfson (1971) showed that a Fourier synthesis with suitably weighted coefficients gives a higher signal to noise ratio than an unweighted $E$ map. By a semi-empirical analysis, these authors suggested in the Fourier map

$$\rho(\mathbf{r}) = \sum_{\mathbf{h}} |E_{\mathbf{h}}| \exp(i\varphi_{\mathbf{h}}) \exp(-2\pi i \mathbf{hr}), \qquad (55)$$

the replacement of the coefficient $|E_{\mathbf{h}}|$ by $[\tanh(\alpha_{\mathbf{h}}/2)]|E_{\mathbf{h}}|$.

Our weighting scheme immediately gives the expression of the weighted Fourier coefficient, provided in (55) $\exp(i\varphi_{\mathbf{h}})$ is replaced by its expected value [see Blow & Crick's (1959) criterion for the 'best Fourier']. If the phase $\varphi_{\mathbf{h}}$ is supposed to be distributed according to $M(\varphi_{\mathbf{h}};\theta_{\mathbf{h}},\alpha_{\mathbf{h}})$ and $\theta_{\mathbf{h}}$ is its assigned value, then

$$\langle \exp i\varphi_{\mathbf{h}} \rangle \simeq \int_0^{2\pi} \exp(i\varphi_{\mathbf{h}})\, M(\varphi_{\mathbf{h}};\theta_{\mathbf{h}},\alpha_{\mathbf{h}})\, \mathrm{d}\varphi_{\mathbf{h}}$$

$$= D_1(\alpha_{\mathbf{h}}) \exp(i\theta_{\mathbf{h}}).$$

Finally, in (55) $D_1(\alpha_{\mathbf{h}})|E_{\mathbf{h}}|$ should replace $|E_{\mathbf{h}}|$.

For centrosymmetric structures, or centrosymmetric reflexions in non-centrosymmetric structures, the distribution of $\varphi_{\mathbf{h}}$ is a two-value function: then $\langle \exp i\varphi_{\mathbf{h}} \rangle$ reduces to $\langle S_{\mathbf{h}} \rangle$, the expected sign of $E_{\mathbf{h}}$. If the direct procedure assigns $S_{\mathbf{h}}$ to $E_{\mathbf{h}}$ with probability $P(S_{\mathbf{h}})$, then

$$\langle S_{\mathbf{h}} \rangle = [2P(S_{\mathbf{h}}) - 1].$$

In conclusion, the coefficient of the Fourier synthesis should be $[2P(S_{\mathbf{h}}) - 1]\, E_{\mathbf{h}}$ instead of $E_{\mathbf{h}}$.

## 12. Concluding remarks

Weighting schemes for tangent refinement are usually semi-empirical and are suggested by triplet conditional distributions giving the value of a phase when some others are exactly known. The present scheme is obtained theoretically, and it is able to predict the reliability of any individual phase when the distributions of other suitable phases are known. We emphasize the fact that our approach does not require that some phase values are known, but only that their distributions are known. Therefore it is able to take into account the uncertainty of the assigned phase values.

The scheme does not force tangent-formula refinement to obey Sayre's equation, as Hull & Irwin's (1978) scheme does. However, probabilistic considerations can easily be found which force our weighting scheme to obey Sayre's equation.

A fundamental hypothesis of the theory above developed is that, for a given $\mathbf{h}$, the quantities $\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j}$, $j = 1,\ldots, n$, are distributed around $\varphi_{\mathbf{h}}$ according to Von Mises distributions $M(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j}; \varphi_{\mathbf{h}}, G_j)$.

If the model is to be practically useful, some kind of general agreement must be found between the theoretical propositions and the observations. For any $\mathbf{h}$ we have at our disposal a sample of $n$ 'observed' values (the $n$ assigned values $\theta_{\mathbf{k}_j} + \theta_{\mathbf{h}-\mathbf{k}_j}$) and we want to know if they can be reasonably regarded as drawn by a simple random sampling from populations having Von Mises distributions $M(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j}; \varphi_{\mathbf{h}}, G_j)$. We can expect, if the hypothesis is true, that the sample values should form a statistical image of the hypothetical distribution.

Let us now suppose that the algebraic form of the hypothetical distribution is known; we are concerned with the mean or some other characteristics of the distribution, and we ask whether the differences between the observed values should be ascribed to random fluctuations or judged to be significant.

According to Hull & Irwin's (1978) scheme, we choose $\alpha$ as the characteristic to be investigated. The observed value of $\alpha$ as given by (6) (from now on $\alpha_{\rm obs}$) may be compared with the expected value $\alpha_E$ given by (10). If $\alpha_{\rm obs} \ll \alpha_E$, then the hypothesis about the distribution of the quantities $\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j}$ is disproved; small values of $\alpha_{\rm obs}$ correctly mark this situation. If $\alpha_{\rm obs} \simeq \alpha_E$, then the agreement between theory and observations is satisfactory and the phase $\varphi_{\mathbf{h}}$ may be considered as well established.

If $\alpha_{\rm obs} \gg \alpha_E$, the observations are again inconsistent with the hypothesis; then $\varphi_{\mathbf{h}}$ is not to be considered well established and a penalty for $\alpha_{\rm obs}$ should be introduced.

In a forthcoming paper, practical aspects and applications of the theory above developed will be shown. We only anticipate here that the first tests will be successful even in cases where the application of the normal tangent formula is not very effective.

### References


ABRAMOWITZ, M. & STEGUN, I. A. (1970). *Handbook of Mathematical Functions*. New York: Dover Publications.

BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.

BUSETTA, B. (1976). *Acta Cryst.* A**32**, 139–143.

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

COCHRAN, W. & WOOLFSON, M. M. (1955). *Acta Cryst.* **8**, 1–12.

GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* B**26**, 274–285.

GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1971). *Acta Cryst.* A**27**, 368–376.

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

GIACOVAZZO, C. (1977). *Acta Cryst.* A**33**, 539–547.

GIACOVAZZO, C. (1978). *Acta Cryst.* A**34**, 562–574.

HULL, S. E. & IRWIN, M. J. (1978). *Acta Cryst.* A**34**, 863–870.

KARLE, I. & KARLE, J. (1966). *Acta Cryst.* **21**, 849–859.

KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635–651.

LESSINGER, L. (1976). *Acta Cryst.* A**32**, 538–550.

MAIN, P. (1977). *Acta Cryst.* A**33**, 750–757.

SCHENK, H. (1972). *Acta Cryst.* A**28**, 412–422.

STEPHENS, H. A. (1963). *Biometrika*, **50**, 385–390.

VON MISES, R. (1918). *Phys. Z.* **19**, 490–500.