SOLYMOSI, F. (1976). In *Contact Catalysis*, Vol. 2, edited by Z. G. SZABO. Amsterdam: Elsevier Scientific.
SOMORJAI, G. A., POWELL, R. E., MONTGOMERY, P. W. & JURA, G. (1967). In *Small-Angle X-ray Scattering*, edited by H. BRUMBERGER. New York: Gordon & Breach.
STOKES, A. R. & WILSON, A. J. C. (1942). *Proc. Cambridge Philos. Soc.* **38**, 313–322.
THOMAS, J. M. & THOMAS, W. J. (1967). *Introduction to the Principles of Heterogeneous Catalysis*. New York: Academic Press.

Acta Cryst. (1981). A**37**, 180–183

A Fast Algorithm for Determining Phase Values for Symbols from Weighted Symbol Relations in Direct Methods

BY PAUL T. BEURSKENS AND PETER A. J. PRICK

Crystallography Laboratory, Toernooiveld, 6525 ED Nijmegen, The Netherlands

(*Received* 30 *January* 1980; *accepted* 25 *September* 1980)

Abstract

The use of symbols to express phases in direct methods leads to relations between symbols (symbol relations). The relative probability of a symbol relation is defined in such a way that it has additive properties. The algorithm given in this paper is developed for the deconvolution of the symbols and is applicable to symmetric phases (0 or π) and to anti-symmetric phases ($+\frac{1}{2}\pi$ or $-\frac{1}{2}\pi$); weighted symbol relations are used as input and figures of merit are calculated for all permutations of phases for the symbols. The algorithm is especially useful when a large number of symbols is used.

Introduction

Nowadays, the majority of structures solved by direct methods are probably solved by using the multisolution program *MULTAN* (Main, Lessinger, Woolfson, Germain & Declercq, 1977). However, the use of symbols to express unknown phases remains a powerful tool as well. Several methods to process and decode symbolic phases have been developed (Karle & Karle, 1966; Beurskens, 1964; Schenk, 1971; and others).

It is the authors' view that by using more than just a few symbols, causing a large number of reflections to take part in the initiation of the calculations, one avoids the use of single or weak phase relationships in the initial – often crucial – stages of a phase generation procedure. This requires a fast and convenient algorithm for disentangling the large bulk of symbol information. In this paper we describe a suitable algorithm (*SYMAN*) for the deconvolution of symbol relations into symmetric phases $(0,\pi)$ or anti-symmetric phases $(+\frac{1}{2}\pi,-\frac{1}{2}\pi)$.

0567-7394/81/020180-04$01.00

Notation

$\varphi_{\mathbf{h}}$	is the (unknown) phase of a reflection **h**
$a, b, c, \ldots$	are the symbols that represent unknown phases
x	is a linear combination of symbols (*e.g.* $a-2b$)
n	is a numerical phase, or the numerical part of a symbolic phase (*e.g.* $\varphi_{\mathbf{h}} = a - 2b + \pi$ can be written as $\varphi_{\mathbf{h}} = x + n$)
α	is the weight associated with the use of the sigma-2 or tangent formula for the determination of the phase $\varphi_{\mathbf{h}}$

$$\boldsymbol{\alpha} = \alpha \exp(i\varphi_{\mathbf{h}}) = 2\sigma_3\,\sigma_2^{-3/2}\,|E_{\mathbf{h}}| \sum_{\mathbf{k}} |E_{\mathbf{h}-\mathbf{k}} E_{\mathbf{k}}| \times \exp i(\varphi_{\mathbf{h}-\mathbf{k}} + \varphi_{\mathbf{k}}), \quad (1)$$

where the summation is restricted to terms in which $(\varphi_{\mathbf{h}-\mathbf{k}} + \varphi_{\mathbf{k}})$ is expressed by the same symbol x, $\sigma_m = \sum_{j=1}^{N} Z_j^m$ for N atoms per unit cell

Symbol relations

Suppose the phase $\varphi_{\mathbf{h}}$ has been calculated as $x_1 + n_1$, with weight α_1, and, independently, as $x_2 + n_2$, with weight α_2. This gives the symbol relation

$$x_1 + n_1 - x_2 - n_2 = 0 \pmod{2\pi}. \quad (2)$$

The variance of this result (acentric), or the probability that this relation is correct (centric), is a function of α_1 and α_2.

In view of computer time and programming convenience it is generally desirable to replace variances and probabilities by weights which have additive

properties; this is essential in the symbol analysis algorithm *SYMAN*.

The weight, α_{rel}, associated with (2), will be defined in Appendix 1.

We use the following properties:

α_{rel} is a function of α_1 and α_2;

if for two or more reflections the symbolic phases lead to identical symbol relations, their weights are added;

a negative α_{rel} value implies a contradiction, or a phase shift π (for instance: $a + 2b + \pi = 0$ with weight α_{rel} can be written as $a + 2b = 0$ with weight $-\alpha_{rel}$);

$\alpha_{rel} = 0$ means that the symbol relation has not been observed, or that the contributions to this relation cancel out.

The procedure *SYMAN*

The fast symbol analysis procedure transforms an ordered list of weights for symbol relations into a list of figures of merit for the possible solutions (phase values for symbols). The algorithm is given in Table 1, using the algorithmic language Algol60 (Naur, 1963). The procedure is explained by the numerical example given in Table 2. In the example we have three symbols, a, b and c; each of these symbols represents a phase 0 or π. With the trivial identities $-a = +a$ and $2a = 0$, there are seven $(2^3 - 1)$ possible symbol relations with weights α_{rel}.

A symbol relation is represented and identified by a binary number (decimal value = i) such that a digit 0 or 1 denotes the absence or presence of the corresponding symbol in this relation. For each of the seven symbol relations the weight is stored in a table A at position i; notation: $A[i]$. The trivial identity $0 = 0$ with $\alpha = 0$ is stored at $i = 0$.

In the successive stages of the procedure the symbols are eliminated one by one. For the elimination of the symbol c (stage 1) the contents of table A are diverted into two subtables: one for the substitution of $c = 0$ and one for the substitution of $c = \pi$.

For instance (see Table 2, stage 1):

For $c = 0$ the relation $c + a = 0$ ($i = 5$) becomes $a = 0$, and its weight is added to the weight of the relation $a = 0$ ($i = 1$): $A[1] = +1 - 4 = -3$.

The symbol relation $c = 0$ ($i = 4$) now changes into the identity relation $0 = 0$ ($i = 0$): $A[0] = -3$. The result is a figure of merit for the substitution $c = 0$.

The complete information for $c = 0$ is stored in the subtable $i = 0$–3.

Substitution of $c = \pi$ in the symbol relation $c + a = 0$ ($i = 5$) gives $a = \pi$ with $\alpha = -4$, or $a = 0$ with $\alpha = +4$, which is to be added to the weight of $a = 0$ ($i = 1$). The result, however, is stored in the subtable $i = 4$–7: $A[5] = +1 + 4 = +5$.

Table 1. *The procedure SYMAN in the algorithmic language Algol60*

```
procedure SYMAN(NS,A); value NS; integer NS; integer array A;
comment NS = number of symbols, A is defined as A [0: 2**NS -1];
begin integer MAX, KSTEP, I, LMAX, K, N, L, TEMP;
      MAX: = 2**NS; KSTEP: = MAX*2;
      for I: = 1 step 1 until NS do
      begin KSTEP: = KSTEP/2; LMAX: = KSTEP/2;
            for K: = 0 step KSTEP until MAX - 1 do
            begin N: = K + LMAX;
                  for L: = 0 step 1 until LMAX - 1 do
                  begin TEMP: = A[K + L];
                        A[K + L]: = TEMP + A[N + L];
                        A[N + L]: = TEMP - A[N + L]
end end end end;
```

Table 2. *Numerical example for three symbols*

i	0	1	2	3	4	5	6	7
Binary representation	000	001	010	011	100	101	110	111
Symbol relation	*	$a = 0$	$b = 0$	$b + a = 0$	$c = 0$	$c + a = 0$	$c + b = 0$	$c + b + a = 0$
$\alpha_{rel} = A[i]$	0	+1	0	−5	−3	−4	−3	+1
Stage 1:								
substitute	$c = 0$	$c = 0$	$c = 0$	$c = 0$	$c = \pi$	$c = \pi$	$c = \pi$	$c = \pi$
add pairs (i,i')	0,4	1,5	2,6	3,7	0,4	1,5	2,6	3,7
for symbol relation	*	$a = 0$	$b = 0$	$b + a = 0$	*	$a = 0$	$b = 0$	$b + a = 0$
$A[i]$	−3	−3	−3	−4	+3	+5	+3	−6
Stage 2:								
eliminate b	$b = 0$	$b = 0$	$b = \pi$	$b = \pi$	$b = 0$	$b = 0$	$b = \pi$	$b = \pi$
add pairs (i,i')	0,2	1,3	0,2	1,3	4,6	5,7	4,6	5,7
for symbol relation	*	$a = 0$	*	$a = 0$	*	$a = 0$	*	$a = 0$
$A[i]$	−6	−7	0	+1	+6	−1	0	+11
Stage 3:								
eliminate a	$a = 0$	$a = \pi$	$a = 0$	$a = \pi$	$a = 0$	$a = \pi$	$a = 0$	$a = \pi$
add pairs (i,i')	0,1	0,1	2,3	2,3	4,5	4,5	6,7	6,7
$A[i]$	−13	+1	+1	−1	+5	+7	+11	−11
FOM	<0	0·06	0·06	<0	0·29	0·41	0·65	<0
for phases c, b, a	0,0,0	0,0,π	0,π,0	0,π,π	π0,0	π,0,π	π,π,0	π,π,π

* Identity: 0 = 0, see text.

Now the position 1 is associated with the phase of c and with the symbol relations involving b and a.

In stage 2 each of the two subtables is diverted for the elimination of the symbol b. In stage 3 each of the four subtables is diverted for the elimination of the symbol a.

After elimination of all symbols the binary representation of i corresponds to a combination of phases for the symbols in such a way that a digit 0 or 1 denotes a phase 0 or π. The final weight, stored at position i, is a measure of the consistency of the symbol relations for the corresponding phase values. The $A[i]$ values may be divided by the sum of the absolute values of the input weights, to obtain relative figures of merit (FOM). One or more of the most probable solutions may be selected by the user. (For instance: Table 2, stage 3. $A[6] = +11$, FOM = 0·65; the most probable solution is $c = \pi$, $b = \pi$, $a = 0$.)

The deconvolution of anti-symmetric phases requires a redefinition of the binary codes for symbol relations and for the final phase values. The symbols are chosen to represent phases $+\frac{1}{2}\pi$ and $-\frac{1}{2}\pi$, with trivial identities $2a = \pi$ and $-a = a + \pi$. Because of the binary nature of the transformation we can only use symbol relations of the type $a = b$ and $a = b + \pi$, *i.e.* the phases of a and b are equal or not equal. The binary code defined earlier for $a + b = n$ now represents the relation $a = b + n$; the binary codes for the phases 0 and π now refer to the phases $+\frac{1}{2}\pi$ and $-\frac{1}{2}\pi$, respectively. Apart from these redefinitions the procedure is the same.

Discussion and applications

The usual consistency criteria or figures of merit are based upon phase permutations performed for every individual reflection and, consequently, the computer time depends on the number of symbols, the number of reflections and the number of different phase indications per reflection. In *SYMAN* all symbol information is deconvoluted simultaneously and the computer time depends on the number of symbols only.

The computer time involved is negligible as long as no more than ten symbols (2^{10} possible solutions) are analyzed. The time is doubled for each additional symbol. For an IBM 370/158 the computer time for the analysis of 15 symbols is 5 s.

The gain of time is most considerable when a large number of symbols is used and many reflections take part in the phase generating procedure.

SYMAN has been implemented in the sign correlation program (Van den Hark, 1976) for the analysis of 26 symbols.

SYMAN is also used in the special *DIRDIF* procedures for origin specification (Beurskens, Prick, Van den Hark & Gould, 1980) and for enantiomorph discrimination (Prick, Beurskens & Gould, 1978). Ten symbols, together with hundreds of phases derived from known heavy-atom positions, are input to the tangent formula; the symbol phases destroy the heavy-atom super-symmetry.

In the general case *SYMAN* can be applied to the symmetrical and antisymmetrical parts of the complex symbol relations, and the results are averaged to obtain phases $\pm\frac{1}{4}\pi$, $\pm\frac{3}{4}\pi$. We have not yet investigated this possible application.

Part of this work was performed at the University of Michigan; one of us (PTB) is most grateful for the hospitality he enjoyed at the Chemistry Department and wishes to thank Dr Chr. Nordman for many pleasant discussions.

APPENDIX 1

The weight for a symbol relation

Formula (1) is equivalent to the sigma-2 formula or the tangent formula (Karle & Karle, 1966). The weight for a particular result is given by α. For centric phase distributions (all reflections in centrosymmetric space groups, or special reflections in non-centrosymmetric space groups), the probability that (1) gives the correct value of φ_h is given by Cochran & Woolfson (1955):

$$P = \tfrac{1}{2} + \tfrac{1}{2}\tanh(\tfrac{1}{2}\alpha). \qquad (3)$$

For acentric phase distributions (general reflections in non-centrosymmetric space groups) the variance σ^2 of the result for φ_h is given by Karle & Karle (1966; see also Karle, 1976); in 'shorthand' notation:

$$\sigma^2 = V(\alpha). \qquad (4)$$

The weight α_{rel} for a symbol relation (2) obtained from two independent results (1) with weights α_1 and α_2 is defined as follows.

Centric

Equation (3) gives the probabilities P_1 and P_2 for the two sign indications, $x_1 + n_1$ and $x_2 + n_2$.

The probability for the symbol relation (2) is given by

$$P_{rel} = P_1 P_2/[P_1 P_2 + (1 - P_1)(1 - P_2)]. \qquad (5)$$

Similarly to the use of the weight α for symbol phases, we introduce α_{rel} as the weight for the symbol relation.

Define α_{rel} in analogy with (3) by

$$P_{rel} = \tfrac{1}{2} + \tfrac{1}{2}\tanh(\tfrac{1}{2}\alpha_{rel}). \qquad (6)$$

From (3), (5) and (6), and using the properties of the hyperbolic function, we derive

$$\tanh \tfrac{1}{2}\alpha_{rel} = \tanh \tfrac{1}{2}\alpha_1 \tanh \tfrac{1}{2}\alpha_2. \qquad (7)$$

Some numerical examples are given in Table 3.

If the same symbol relation is obtained from different reflections, its probability increases according to

$$P_{\rm rel} = \prod_i P_{{\rm rel},i} \Big/ \left[\prod_i P_{{\rm rel},i} + \prod_i (1 - P_{{\rm rel},i}) \right] \qquad (8)$$

and its weight increases according to

$$\alpha_{\rm rel} = \sum_i \alpha_{{\rm rel},i}, \qquad (9)$$

as can be seen by substitution of (6) in (8). Thus we have an expression for the weight (7) that has additive properties (9).

Acentric

The variances $V(\alpha_1)$ and $V(\alpha_2)$ of two phase indications for one reflection can be calculated. In view of the large number of phase relationships, we may assume a normal distribution of deviations.

Table 3. *The weight* $\alpha_{\rm rel}$ *for centric symbol relations, for various* α_1 *and* α_2 *values (equations 6 and 7)*

Accuracy = 0·05.

$\alpha_1 =$	0·5	1·0	1·5	2·0	4·0	10·0
$P_1 =$	0·62	0·73	0·82	0·88	0·98	1·00
α_2						
1·0	0·2	0·4	0·6	0·7	1·0	1·0
2·0	0·4	0·7	1·1	1·3	1·9	2·0
10·0	0·5	1·0	1·5	2·0	4·0	9·3

Table 4. *The weight* $\alpha_{\rm rel}$ *for acentric symbol relations, for various* α_1 *and* α_2 *values (equations 10–12)*

Accuracy = 0·05.

$\alpha_1 = 0{\cdot}5$	1·0	1·5	2·0	4·0	10·0
$\sigma_1 = 88°$	72°	60°	50°	31°	19°
α_2					
1·0	0·0	0·0	0·3	0·5	0·8	1·0
2·0	0·1	0·5	0·8	1·1	1·5	1·8
10·0	0·4	0·9	1·4	1·8	3·2	5·3

Table 5. *Comparison of true and approximated* $\alpha_{\rm rel}$ *values for acentric multiple symbol relations*

$\alpha_{{\rm rel},1}$	$\alpha_{{\rm rel},2}$	$\alpha_{{\rm rel},3}$	$\alpha_{\rm rel}$ (true)	$\alpha_{\rm rel}$ (equation 14)
0·5	0·5		1·5	1·0
0·5	0·5	0·5	2·0	1·5
0·5	1·0		1·7	1·5
0·5	2·0		2·5	2·5
1·0	1·0	1·0	2·6	3·0
2·0	2·0	2·0	4·5	6·0
2·0	4·0		5·4	6·0
4·0	4·0		7·3	8·0

Then the variance of the symbol relation is given by

$$\sigma^2_{\rm rel} = \sigma^2_1 + \sigma^2_2 = V(\alpha_1) + V(\alpha_2). \qquad (10)$$

In analogy with the centric case, we define the weight of the symbol relation as

$$\alpha_{\rm rel} = V^{-1}(\sigma^2_{\rm rel}), \qquad (11)$$

where the function V^{-1} is the inverse of the function V (equation 4). Substitution of (10) into (11) gives

$$\alpha_{\rm rel} = V^{-1}[V(\alpha_1) + V(\alpha_2)]. \qquad (12)$$

In this way, $\alpha_{\rm rel}$ is defined as a function of α_1 and α_2. Numerical values are given in Table 4. By comparing Tables 3 and 4, it is clear the acentric symbol relations are less powerful than centric symbol relations. The reliability of a multiple symbol relation increases, or the standard deviation decreases, according to

$$\sigma^{-2}_{\rm rel} = \sum_i \sigma^{-2}_{{\rm rel},i}. \qquad (13)$$

From this result and definition (11), we can calculate $\alpha_{\rm rel}$ of a multiple symbol relation. In contrast with the centric case, $\alpha_{\rm rel}$ has no strictly additive properties and, in analogy with (9), we define the approximation

$$\alpha_{\rm rel} \simeq \sum_i \alpha_{{\rm rel},i}. \qquad (14)$$

True and approximated $\alpha_{\rm rel}$ values are compared in Table 5. The agreement appears to be good enough for practical applications. Fortunately, the approximation is an underestimation for the weaker contributions.

References

BEURSKENS, P. T. (1964). *Acta Cryst.* **17**, 462.

BEURSKENS, P. T., PRICK, P. A. J., VAN DEN HARK, TH. E. M. & GOULD, R. O. (1980). *Acta Cryst.* A**36**, 653–656.

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

KARLE, J. (1976). *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 155–164. Copenhagen: Munksgaard.

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

MAIN, P., LESSINGER, L., WOOLFSON, M. M., GERMAIN, G. & DECLERCQ, J.-P. (1977). *MULTAN. A Computer Program for the Automatic Solution of Crystal Structures*. Univs. of York, England and Louvain, Belgium.

NAUR, P. (1963). *Numer. Math.* **4**, 420–453.

PRICK, P. A. J., BEURSKENS, P. T. & GOULD, R. O. (1978). *Acta Cryst.* A**34**, S42.

SCHENK, H. (1971). *Acta Cryst.* B**27**, 2037–2039.

VAN DEN HARK, TH. E. M. (1976). Thesis, Univ. of Nijmegen.