*Acta Cryst.* (1956). **9**, 616

## Note on the statistical approach to the phase problem.

By J. Gillis, *Department of Applied Mathematics, Weizmann Institute of Science, Rehovoth, Israel*

(*Received* 3 *April* 1956)

1. Consider a centrosymmetric structure and let $H$ denote the general reciprocal-lattice point $(hkl)$, $U_H$ the corresponding unitary structure factor, and $S_H$ its sign. Then the method of Cochran & Douglas (1955) for determining phases is to maximize the expression

$$F_1 = \sum_{H,H'} P(H, H')S_H S_{H'} S_{H+H'} , \qquad (1)$$

where the $P(H, H')$ are suitable non-negative weighting factors. This is based on the fact that, for sufficiently large values of $|U_H U_{H'} U_{H+H'}|$, the expression $S_H S_{H'} S_{H+H'}$ has probability greater than $\frac{1}{2}$ of being positive. However, the expression $F_1$ does not utilize all the information contained in the list of absolute values of $U_H$. In particular, it follows from equation (7) of Zachariasen (1952) that, provided that

$$|U_H|^2 + |U_{H'}|^2 \leqslant |U_{H+H'} U_{H-H'}| , \qquad (2)$$

we shall, on the average, have

$$S_H S_{H'}(U_{H+H'} + U_{H-H'}) \approx -U_{H+H'} U_{H-H'} . \qquad (3)$$

Hence we have

$$(1 + S_H S_{H'} U_{H+H'})(1 + S_H S_{H'} U_{H-H'}) \approx 1 , \qquad (4)$$

and so, unless $|U_{H+H'}|$, $|U_{H-H'}|$ are both small,

$$S_{H+H'} = -S_{H-H'} . \qquad (5)$$

The relation (5) has been checked for a number of solved structures and has been found to hold with high probability, subject to the conditions stated. This suggests that $F_1$ be replaced by

$$F_2 = \sum_{H,H'} P(H, H')S_H S_{H'} S_{H+H'} - \sum_{H,H'} Q(H, H')S_{H+H'} S_{H-H'} . \qquad (6)$$

It is known (Woolfson, 1954; Cochran & Douglas, 1955) that $P(H, H')$ may be taken to be $|U_H U_{H'} U_{H+H'}|$. By a similar argument it can be shown that $Q(H, H')$ may be defined as follows:

$$\left.\begin{array}{ll}\text{if (2) holds,} & Q(H, H') = |U_{H+H'} U_{H-H'}| , \\ \text{otherwise,} & Q(H, H') = 0 .\end{array}\right\} \qquad (7)$$

The techniques developed by Cochran & Douglas for the application of high-speed digital computing machinery to the exploitation of (1) can be used, with some modification, on (6). Further details, along with an analytical justification, will be published later. Meanwhile it may be of interest to report on a practical combination of these concepts with the method of inequalities in the study of a known structure.

2. The data obtained by Abrahams & Robertson (1948) for $p$-nitroaniline were investigated by the inequalities method and, as a result, it was possible to express the signs of a number of the $h0l$ structure factors in terms of five unknowns, $a$, $b$, $c$, $d$, $e$, (cf. Gillis, 1948). It was then decided to try to evaluate these unknowns by statistical methods. However, the relations obtained from the inequalities had all been of the form $S_{H+H'} = S_H S_{H'}$ and so, to reduce the overlap of information, the first sum was omitted from $F_2$ and only the second sum was used. The problem was thus to minimize

$$F_3 = \sum_{H,H'} Q(H, H')S_{H+H'} S_{H-H'} . \qquad (8)$$

It turned out that

$$\begin{aligned}F_3 = {} & 2{\cdot}31a + 2{\cdot}30e + 2{\cdot}43ae + b(1{\cdot}16 + 0{\cdot}62a + 0{\cdot}95ae + 0{\cdot}34e) \\ & + d(0{\cdot}13 + 0{\cdot}28a + 0{\cdot}19e + 0{\cdot}10ae + 0{\cdot}20b + 0{\cdot}12ab) \\ & + c(0{\cdot}19 + 0{\cdot}09 + 0{\cdot}13ae) . \qquad (9)\end{aligned}$$

The actual values of $F_3$ for the 32 possible sets of signs range from $-3{\cdot}88$ to $+11{\cdot}54$. The lowest-scoring sets were

| $a$ | $b$ | $c$ | $d$ | $e$ | $F_3$ |
|---|---|---|---|---|---|
| − | − | − | + | − | −3·88 |
| − | − | − | + | + | −3·42 |
| − | − | − | − | − | −3·24 |
| + | − | − | + | + | −3·14 |
| + | − | − | + | − | −3·08 |
| − | − | − | − | + | −2·78 |

The top line of the table was in fact the correct one. Even without this knowledge, however, it would have been tempting to try $a = b = c = -1$, and these would have correctly determined the signs of 21 among the larger structure factors.

3. In some examples, constructed artificially, it was found that $F_2$ could be maximized by repeated adjustment of the $S_H$'s. This process was rejected by Cochran & Douglas as a method for maximizing $F_1$, for reasons which seem to be very substantial. It has not so far been possible to determine sufficient conditions under which the process can be expected to give the correct result for $F_2$. However, it is clear that the presence of the second sum in $F_2$ removes a simple and fundamental weakness of $F_1$, namely that the latter is always maximized by the wrong set $S_H = +1$ for all $H$.

### References

Abrahams, S. C. & Robertson, J. M. (1948). *Acta Cryst.* **1**, 252.
Cochran, W. & Douglas, A. M. B. (1955). *Proc. Roy. Soc.* A, **227**, 486.
Gillis, J. (1948). *Acta Cryst.* **1**, 174.
Woolfson, M. M. (1954). *Acta Cryst.* **7**, 61.
Zachariasen, W. H. (1952). *Acta Cryst.* **5**, 68.