Acta Cryst. (1977). A33, 1009–1010

A method for optimizing relative weights in least-squares analysis.

By KURT NIELSEN, *Chemistry Department B, The Technical University of Denmark*, DK-2800 *Lyngby, Denmark*

(*Received* 7 *May* 1977; *accepted* 17 *June* 1977)

An automatic method for optimizing a weight function of a specific form for use in least-squares analysis is described. The function to be minimized is $\sum_i (w_i|\Delta F_i|^2) \log (w_i|\Delta F_i|^2)$, subject to the condition that $\sum_i w_i|\Delta F_i|^2$ is constant. Minimization is accomplished *via* a simplex method.

In least-squares analysis a set of observed quantities, $\{F_i\}$, is related to a model described by a set of parameters, $\{x_i\}$. In order to obtain maximum accuracy with the least-squares method, weights should be used, which are related to the accuracy of the observations. Relative weights may reflect the trend in the $|\Delta F|$, the absolute difference between the observed and calculated quantity. A weight function, w, depending on a very small number of parameters, is chosen so that the averages of $w|\Delta F|^2$ are constant when the set of $w|\Delta F|^2$ values is analysed in any systematic fashion. Weights chosen in this way lead to estimates of the variances, $\sigma^2(x_i)$, which allow for all random experimental errors, such systematic experimental errors as cannot be paralleled in the model, and such defects of the model as are not paralleled in the experimental data (Cruickshank, 1970).

A reasonable measure, S_N, of the goodness of the weight function may be defined as

$$S_N = \sum_{i=1}^{N} p_i \log p_i ,$$

where $p_i = kw_i|\Delta F_i|^2$, and the constant, k, is chosen so that

$$\sum_{i=1}^{N} p_i = 1 .$$

The function S_N is the negative entropy function (Brillouin, 1962), and some of its properties are of particular interest:

(*a*) S_N is a minimum and equal to $-\log N$ if all p_i are equal to N^{-1}. This corresponds to $w_i = |\Delta F_i|^{-2}$.

(*b*) S_N is a maximum and equal to zero if all but one of the p_i are zero.

(*c*) Any kind of averaging procedure decreases S_N. Thus if **A** is an N-dimensional square matrix with elements $A_{ij} \geq 0$, and with the property that

$$\sum_i A_{ij} = \sum_j A_{ij} = 1$$

and if

$$p_i' = \sum_j A_{ij} p_j ,$$

then $S_N(p_i') \leq S_N(p_i)$. The sign of equality holds only if **A** is the unit matrix.

Suppose the set p_i is divided into disjoint subsets, and every p_i is replaced by the average value within the subset to which it belongs, then the resulting value, S_N', is smaller than S_N. Thus S_N is the upper bound of all values S_N' obtained in any method of analysis, and an optimal weight function of a chosen form may be found by minimizing S_N with respect to the parameters in the weight function. However, the optimization of the weight function breaks down if a few of the p_i are much larger than the others, for example as a result of extinction, but this can be overcome by performing a normal probability analysis of $w^{1/2}\Delta F$ (Abrahams & Keve, 1971), and then removing the outstanding reflexions from the optimization.

Rather than working with S_N, whose minimum value is dependent on the number of observations, one may define a weight index, S, as $S = 1 + S_N/\log N$. The minimum and maximum of S are 0 and 1 respectively.

Minimization of S is accomplished *via* the simplex method (Nelder & Mead, 1964–65). This method is chosen because S is a fairly complex function of the parameters in the weight function, and because these parameters may be heavily correlated. It is also a convenient method for inclusion of different kinds of penalty functions, for example to secure positive weights or to limit the ratio between the maximum and minimum weight. Furthermore, the method is rather insensitive to the choice of starting point.

As an example of the use of the method, the results from an optimization of the weight function for 1-phenyl-3-methyl-5-(1,2,3-triazolio)sulfide (Søtofte & Nielsen, 1977) is shown. There are 882 observed reflexions, and the form of the weight function was chosen to be

$$w_i = [a + bF_i + cF_i^2 + d(\sin\theta)/\lambda]^{-1} .$$

The parameters for the optimized weight function were determined iteratively, first from the least-squares minimum obtained with unit weights, and second from the least-squares minimum of the weighted equations. In a final least-squares analysis convergence was obtained within one cycle giving an average shift less than one tenth of the estimated error, and the iterative process was considered to be completed. Furthermore the final refinement yielded R and R_w values (0·039 and 0·038 respectively) identical to those obtained in the last cycle of the previous refinement. The parameters in the weight function were $a = 0{\cdot}6820$, $b = -0{\cdot}1089$, $c = 0{\cdot}003619$, and $d = 0{\cdot}8169$. The values of S before and after the final least-squares cycle were 0·107 and 0·109 respectively. The S obtained with unit weights was 0·215.

A subsequent normal probability plot, with weights brought to a scale such that the standard deviation of an observation of unit weight is unity, gave a slope of 0·901 and a

Table 1. *Analysis of* $\langle w|\Delta F|^2 \rangle$ *as a function of* F

Interval	Number of reflexions	$\langle \lvert\Delta F\rvert^2 \rangle$	$\langle w\lvert\Delta F\rvert^2 \rangle$
5·04	110	1·086	1·034
6·43	110	0·915	0·985
7·85	110	0·776	0·980
9·64	110	0·447	0·780
12·26	110	0·516	1·054
15·93	110	0·284	0·835
22·58	110	0·489	1·315
142·98	112	3·447	1·018

correlation coefficient, ϱ, of 0·999, indicating an almost normal distribution of the weighted residuals. The corresponding values before the last least-squares refinement were 0·904 and 0·999 respectively. The averages of $w|\Delta F|^2$ analysed with respect to F and $(\sin\theta)/\lambda$ are shown in Tables 1 and 2 respectively. The values in the tables are scaled such that the total average is unity and refer to the second determination of the weight function.

Table 2. *Analysis of* $\langle w|\Delta F|^2\rangle$ *as a function of* $(\sin\theta)/\lambda$

Interval	Number of reflexions	$\langle\|\Delta F\|^2\rangle$	$\langle w\|\Delta F\|^2\rangle$
0·3084	110	3·093	1·106
0·3861	110	0·667	1·131
0·4418	110	0·520	0·878
0·4885	110	0·488	0·807
0·5302	110	0·706	1·103
0·5665	110	0·580	0·823
0·6049	110	0·932	1·116
0·6491	112	1·015	1·036

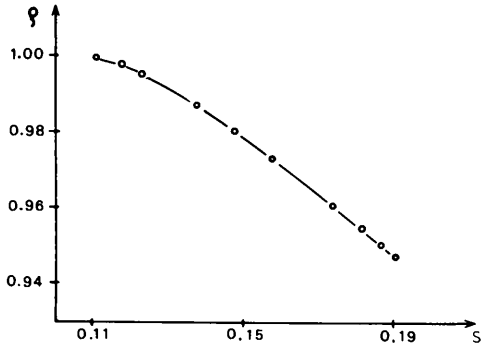

Fig. 1. Plot of the correlation coefficient, ϱ, *versus* weight index, S.

Fig. 1 shows the correlation coefficient as a function of the weight index S for a fixed set ΔF. The different values of S and ϱ are obtained by varying the parameters in the weight function. The correlation coefficient, ϱ, tends to 1 with decreasing value of S, in which case the $w^{1/2}\Delta F$ are drawn from a normal distribution, and the least-squares method and the maximum-likelihood method become equivalent (Kendall, 1946). Although, as a general feature, this remains to be proved, the present method, as demonstrated by the example, is nevertheless, together with a normal probability plot, a useful tool in the evaluation of crystal structure data.

References

ABRAHAMS, S. C. & KEVE, E. T. (1971). *Acta Cryst.* A**27**, 157–165.

BRILLOUIN, L. (1962). *Science and Information Theory*, 2nd ed. New York: Academic Press.

CRUICKSHANK, D. W. J. (1970). In *Crystallographic Computing*, edited by F. R. AHMED, p. 195. Copenhagen: Munksgaard.

KENDALL, M. G. (1946). *The Advanced Theory of Statistics*, Vol. II, p. 59. London: Charles Griffin.

NELDER, J. A. & MEAD, R. (1964–65). *Comput. J.* **7**, 308–313.

SØTOFTE, I. & NIELSEN, K. (1977). In preparation.

Acta Cryst. (1977). A**33**, 1010–1011

Further notes on the generalized inverse.

By D. N. J. WHITE, *Chemistry Department, The University, Glasgow G12 8QQ, Scotland*

(*Received* 22 *April* 1977; *accepted* 6 *May* 1977)

Fast algorithms for 'inverting' singular matrices by modified Cholesky procedures produce a non-unique A^{r2} generalized reciprocal matrix rather than the generalized inverse A^+.

Some crystallographic applications of generalized inverse matrices have been described (Mackay, 1977) and further food for thought may be gleaned from standard texts (Pringle & Rayner, 1971; Boullion & Odell, 1971). In addition to these applications the use of generalized inverse matrices is currently undergoing something of a renaissance in a related area of physical science, namely empirical valence-force-field or molecular-mechanics calculations of crystal (Warshel & Lifson, 1970) and molecular structure (White, 1977). Unfortunately, the search for efficient algorithms to generate generalized inverses has precipitated several difficulties which have been discussed by Ermer (1975) although no explanation was offered. These problems have an almost exact parallel in the refinement of structure factors by the methods described in Mackay's paper and we offer the following clarification.

Some definitions are required because the nomenclature of generalized inverses is something of a semantic minefield. For any matrix P, square or rectangular, there exists a *unique* matrix Q satisfying the conditions

$$PQP = P \tag{1}$$

$$QPQ = Q \tag{2}$$

$$(PQ)^* = PQ \tag{3}$$

$$(QP)^* = QP\,. \tag{4}$$

Furthermore, (*a*) a one-condition, generalized reciprocal matrix of P is a matrix $Q = P^{r1}$ satisfying (1); (*b*) a two-condition, generalized reciprocal matrix of P is a matrix $Q = P^{r2}$ satisfying (1) and (2); (*c*) a left-weak, generalized reciprocal matrix of P is a matrix $Q = P^{r3}$ satisfying (1), (2) and (3); (*d*) a right-weak, generalized reciprocal matrix of P is a matrix $Q = P^{r3'}$ satisfying (1), (2) and (4); (*e*) the generalized inverse matrix of P is a matrix $Q = P^+$ satisfying (1)–(4).

The following inclusion relations are true

$$P^+ \subseteq P^{r3} \subseteq P^{r2} \subseteq P^{r1}$$

$$P^+ \subseteq P^{r3'} \subseteq P^{r2}$$