

Least-Squares Refinement with Subsidiary Conditions

BY JÜRIG WASER

Gates and Crellin Laboratories of Chemistry, California,
Institute of Technology, Pasadena, California, U.S.A.*

(Received 18 January 1963)

A method of least-squares refinement is described in which the subsidiary conditions are treated like observational equations. The advantages of the method are its generality, its adaptability to machine computing, and the possibility of relaxing the subsidiary conditions to any desired degree by appropriate changes in the weighting. In suitable cases the method extends the range for which least-squares refinements converge to the correct solution.

It is useful at times to refine positional parameters of atoms in such a way that these atoms by necessity represent a molecule of known and specified geometry. In the special case that this molecule is rigid, one may define the positions of all atoms in a molecular coordinate system and consider the parameters describing the orientation and translation of the molecular coordinate system relative to the crystallographic system as parameters to be refined (*e.g.* Sparks, 1958). A molecule that consists of several linked rigid portions may be treated by an extension of the same method. However, with increasing flexibility of the molecules considered the procedure becomes cumbersome.

A more general method of imposing geometrical conditions on the positional parameters of the atoms is to subject them to subsidiary conditions in the least-squares treatment by the classical and powerful method of Lagrange multipliers (*e.g.* Hughes, 1941). The equations that express the desired conditions (*e.g.* specifying that certain interatomic distances and bonding angles have preassigned values) are multiplied by undetermined factors — the Lagrange multipliers — and the results added to the weighted sum of the squared residuals to be minimized. Differentiation with respect to the parameters results in equations for the parameter shifts as functions of the Lagrange multipliers. These equations and the subsidiary conditions may then be solved for the parameter shifts and the Lagrange multipliers. While elegant, this method is often cumbersome in numerical applications and in particular proved to be unsuitable to machine computations for the particular application of interest here.

The idea of treating the desired subsidiary conditions exactly like observational equations presented itself. The squares of residuals resulting from the subsidiary conditions *suitably weighted* are simply added to the sum of the weighted squares of the

residuals coming from the usual observational equations.

This appears to be an entirely general method of introducing conditions into the least-squares procedure, and in particular when refining positional parameters. In the limit that the residuals originating from the subsidiary conditions are supplied with infinite weights these conditions are satisfied precisely. It is, however, possible to relax the conditions to any desired degree by suitably decreasing the weights of the corresponding residuals. In this respect the present method is more general than the method of Lagrange multipliers. It has the same advantage over the method involving molecular coordinates described earlier, as well as the advantage of being a general method, not requiring the setting up of new equations for each new kind of molecule. It works as well for rigid as for highly flexible molecules.

The usual metric conditions imposed on atomic parameters to insure that they represent a given molecule are (1) that certain bonded atoms are at preassigned distances from each other, (2) that certain atom triplets define preassigned bonding angles, and (3) that certain atoms are coplanar. It turns out that the mathematics is most straightforward if conditions of the type (2) and (3) are expressed as distance requirements also. Thus, the condition for a given bonding angle is best restated in terms of the three distances involved. The coplanarity of a set of atoms may be enforced by specifying non-bonded as well as bonded distances between the atoms in this plane. This is not the most effective way to achieve coplanarity because small deviations from coplanarity affect interatomic distances only in the second order. However, this does not appear to be a disadvantage, because the imposition of too stringent requirements may generally be unwise. For some molecules it may be convenient to impose conditions of coplanarity by specifying the distances of these atoms from an atom outside the plane (*e.g.* in ferrocene).

The conditional expressions resulting from distance requirements are as follows: Let the starting para-

* Contribution No. 2934 from the Gates and Crellin Laboratories of Chemistry.

meters of the N atoms considered be x_{ij}^0 , $i=1\dots N$, $j=1, 2, 3$. Here and later the indices i and k will always be atom numbers, while j and t refer to the crystallographic axes (\mathbf{a}_j) and run from 1 to 3. The least-squares iterations to be discussed lead to shifts Δ_{ij} and new parameters x_{ij}

$$x_{ij} = x_{ij}^0 + \Delta_{ij} . \quad (1)$$

Certain distances r_{ik} are to be prescribed for select pairs of atoms i and k , and the parameters x_{ij} to be adjusted so that the actual distances between the atoms give a 'best' fit with weights w_{ik} to the prescribed distances. The sum of weighted squares of residuals related to this requirement is then

$$R_1 = \sum_{i,k} w_{ik} \left[\left(\sum_{j,t} g_{jt} (x_{ij} - x_{kj}) (x_{it} - x_{kt}) \right)^{\frac{1}{2}} - r_{ik} \right]^2 \quad (2)$$

where

$$g_{jt} = \mathbf{a}_j \cdot \mathbf{a}_t$$

represents the metric tensor in the crystallographic system used, and the sum over the index pair (i, k) has to be taken over all pairs for which distances r_{ik} have been specified. Using (1), differentiating R_1 with respect to the parameters x_{ij} , and neglecting all powers of the shifts Δ_{ij} higher than the first the following expressions result:

$$\begin{aligned} \left(\frac{1}{2}\right) (\partial R_1 / \partial x_{ij}) &= \sum_{k(i)} w_{ik} \left[(r_{ik}/r_{ik}^0) g_{ikj} \sum_t g_{ikt} (\Delta_{it} - \Delta_{kt}) \right. \\ &\quad \left. + [1 - (r_{ik}/r_{ik}^0)] \sum_t g_{jt} (\Delta_{it} - \Delta_{kt}) + g_{ikj} (r_{ik}^0 - r_{ik}) \right] . \quad (3) \end{aligned}$$

For any given value of i the sum over $k(i)$ is to be taken over all k 's for which a distance has been specified that relates the two atoms i and k . The quantity r_{ik}^0 is the distance between the initial locations x_{ij}^0 and x_{kj}^0 ,

$$r_{ik}^0 = \left[\sum_{j,t} g_{jt} (x_{ij}^0 - x_{kj}^0) (x_{it}^0 - x_{kt}^0) \right]^{\frac{1}{2}}$$

while g_{ikj} is defined by

$$g_{ikj} = \left[\sum_t g_{jt} (x_{it}^0 - x_{kt}^0) \right] / r_{ik}^0 = a_j \cos(\mathbf{a}_j, \mathbf{r}_i - \mathbf{r}_k) .$$

The weighted sum of squared residuals that is usually minimized in structure factor least-squares (*SFLS*) refinements is

$$R_2 = \sum w_F (F_o - F_c)^2$$

where the sum is taken over all reflections. Expansion in terms of the parameter shifts yields the following result* for the partial derivatives:

$$\begin{aligned} \left(\frac{1}{2}\right) (\partial R_2 / \partial x_{ij}) &= \sum w_F \left[\sum_{k,t} \frac{\partial F_c}{\partial x_{kt}} \frac{\partial F_c}{\partial x_{ij}} \Delta_{kt} - \sum_i (F_o - F_c) \frac{\partial^2 F_c}{\partial x_{ij} \partial x_{it}} \Delta_{it} \right. \\ &\quad \left. - (F_o - F_c) \frac{\partial F_c}{\partial x_{ij}} \right] . \quad (4) \end{aligned}$$

In a least-squares refinement with specified distance requirements the linear equations to be solved for the shifts are obtained by setting the sums of (3) and (4) equal to zero,

$$\partial R_1 / \partial x_{ij} + \partial R_2 / \partial x_{ij} = 0 . \quad (5)$$

This procedure will be referred to as the *conditional SFLS* method. As implied earlier the degree to which the distance conditions are satisfied depends on the ratio w_{ik}/w_F of the weighting functions used. If the relative emphasis on the observational equations and the subsidiary conditions is to be commensurate with approximating the observed structure factors to within about 1 electron and the specified distances to within about 0.1 Å, the w_{ik} should be of the order of $(1/0.1)^2 = 100$ times as large as the w_F .

There exists the following problem concerning equation (5). In most *SFLS* refinements of recent years it is not the simultaneous linear equations $\partial R_2 / \partial x_{ij} = 0$ (and the analogous equations resulting from setting equal to zero the derivatives of R_2 with respect to temperature parameters and scale factor) that are solved, but rather an approximation of these equations called the block diagonal approximation. In it all elements of the coefficient matrix of the shifts of positional and temperature parameters relating to different atoms are set equal to zero, because they usually are expected to be negligibly small. (The approximation is sometimes carried further, e.g. by setting to zero all but diagonal elements and those non-diagonal elements that are not negligible because the crystallographic axes they refer to do not enclose a right angle.) In the conditional *SFLS* procedure the contributions from (3) result in large matrix elements outside the diagonal blocks for the positional parameters of all atom pairs i, k for which distances r_{ik} have been specified. Equation (5) is thus appropriate only for the so-called full matrix approach to least-squares refinements.

When the block diagonal approach is used the following variant has proved of advantage. (As discussed later it may be of advantage even for the full matrix approach.) Each unconditional *SFLS* cycle is followed by one or several cycles that readjust the positional parameters so as to give a 'best' fit to the parameters resulting from the *SFLS* cycle as well as to the prescribed distances. The sum R_2 of weighted squares of structure factor residuals is thus replaced by a sum R_3 of weighted squares of distances to the positions resulting from minimizing R_2 unconditionally.

In more detail, let p_{ij} be the j th component of the positional parameter of the i th atom, as determined

* The second term on the right side of (4) and similar terms involving mixed derivatives with respect to the temperature parameters and the scale factor are usually omitted from least-squares refinements, probably because they are lost in the customary derivation of the normal equations (e.g. Lipson & Cochran, 1953), which strictly applies to linear observational equations only. The existence of these terms has been pointed out by Cruickshank (1952). There appears to be no reason why they should not be included in least-squares refinements, particularly if electronic computers are used.

from a *SFLS* cycle, and let w_i be the weight attached to the squared distance between the position of atom i described by p_{ij} and that described by the parameter x_{ij} to be adjusted. The weighted sum of squared residuals to be added to R_1 is thus

$$R_3 = \sum_i w_i \sum_{j,t} g_{jt} (x_{ij} - p_{ij}) (x_{it} - p_{it}). \quad (6)$$

The derivative of R_3 is given by

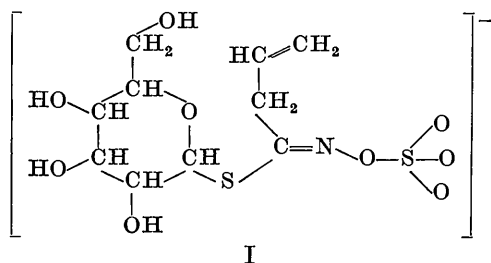
$$\left(\frac{1}{2}\right) \partial R_3 / \partial x_{ij} = w_i \sum_t g_{jt} (x_{it} - p_{it}) = w_i \sum_i g_{jt} (\Delta_{it} + x_{it}^0 - p_{it}) \quad (7)$$

and the simultaneous linear equations to be solved for the shifts are

$$\partial R_1 / \partial x_{ij} + \partial R_3 / \partial x_{ij} = 0. \quad (8)$$

This procedure will be called the *fitting of a molecule to given positions* p_{ij} .

A FORTRAN program for an IBM 7090 digital computer has been written by R. A. Sparks and arranged so that refinement cycles with a block diagonal *SFLS* program (7090 program of Gantzel, Sparks & Trueblood, 1961) are automatically followed by one or more cycles that fit a prescribed molecule to the resulting parameters p_{ij} . This procedure proved very useful in the initial refinement stages of the crystal structure of sinigrin ($C_{10}H_{16}O_9NS_2K \cdot H_2O$) (Waser & Watson, 1963). This substance crystallizes in the acentric space group $P2_12_12_1$ with one formula in the asymmetric unit. While *SFLS* refinement without subsidiary conditions failed to converge, convergence was achieved by the above method. The following details are provided to give a general idea of the workings of the method and the choice of weights. Sinigrin is a salt and neither the position of the K^+ ion nor that of the extra H_2O molecule were restricted by any distance requirements. The distance conditions used affected, however, all other atoms (excepting hydrogens) of the anion I of the salt and included specifica-



tions for bonding distances and angles as well as for rigid groups such as the sulfate group and the glucose ring, and for the group of five atoms that are held in a plane by the $C=N$ double bond. Initially the weights for the distance requirements were set at 100, except that the distances holding the five atoms just discussed close to a plane were given weights 10. The weights chosen for the positional requirements

were 50 for the sulfur atoms and 10 (and in early cycles 1) for all but two of the lighter atoms. Because the positions of the two double bonded carbon atoms of the propylene group were felt to be less certain they were given the weights 0.01 thus minimizing any pull on the remainder of the anion. In later stages the weights of all lighter atoms were set uniformly at 10, and the distance requirements were gradually relaxed by decreasing their number and weights, the latter eventually by a factor 10. During the course of refinements the shifts of the two double bonded propylene carbon atoms were larger than 1 Å and many other atoms shifted by several tenths of one Å. This is illustrated by Fig. 1 in which the positions of the 22 atoms of the anion I are shown for an early set of parameters (with the customary R factor about 0.35) and for the cycle after which all distance requirements were relinquished (R about 0.15).

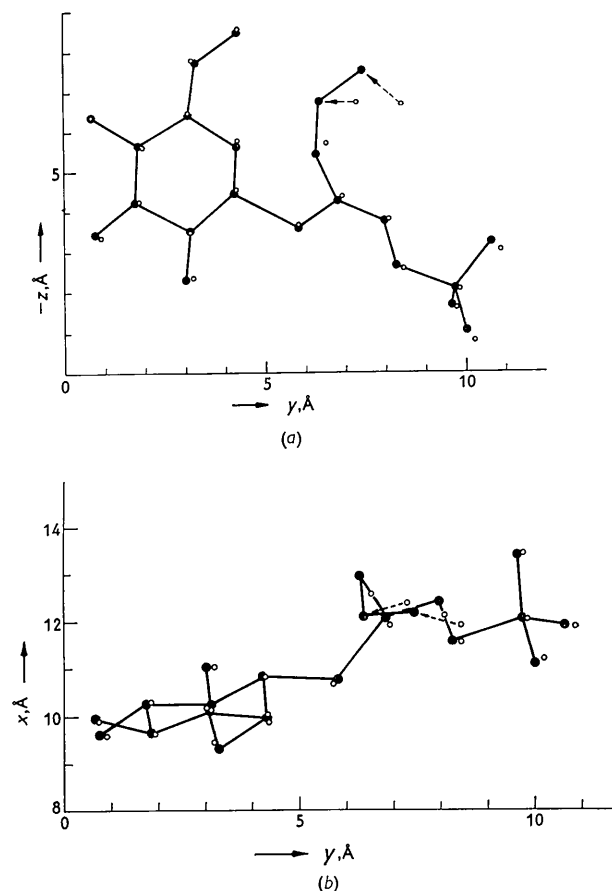


Fig. 1. (a) Top view and (b) side view of anion I at early (empty circles) and late refinement stages (filled and connected circles). The arrangement of the atoms in the top view is similar to that in the chemical formula I.

It was found that after each *SFLS* cycle the adjustment of atomic positions demanded by the

weighted positional and distance requirements usually took several iterations of equation (8) until all shifts had become negligibly small. It thus appears that even when full matrix *SFLS* methods are employed it may be preferable to use the procedure of fitting a molecule to the parameters resulting from unconditional *SFLS* cycles rather than the conditional *SFLS* method of incorporating distance conditions into the *SFLS* matrix. Another reason for such a preference is that the results of unconditional *SFLS* refinement cycles are, of course, always of interest, and to obtain them it is necessary to solve the equations for the shifts twice, once without and once with the terms originating from (3).

The general choice of weights depends on the details of the problem on hand. It should also be noted that the degree to which the distance requirements are satisfied does not depend on the ratio of the weights w_{ik}/w_i alone, but also on the degree to which the positional requirements are compatible with the distance requirements. The poorer the positional parameters, the larger are the relative weights w_{ik} needed to achieve a molecule of reasonable geometry. Positions of heavy atoms should be given considerable weight; unless these positions are close to being correct, refinement is liable to fail in any case. Suspect positions should be given little weight lest they pull on the remainder of the molecule rather than being pulled.

As already stated, use of weighted distance requirements strengthens the least-squares refinement method in suitable cases so that the trial parameters for which this refinement begins to converge may be further removed from their correct values than when such requirements are not made. Such distance requirements will in general be relaxed as the refinement proceeds. In special cases, however, it may be useful to keep at least some of the distance requirements even in the last refinement stages. For example, the quantity and/or quality of the data available may not warrant refinement of all individual atomic parameters. In the presence of heavy atoms as in some organometallic compounds it may be appropriate to refine the positions of molecular groupings of atoms rather than of individual atoms. It may also be useful to refine positions of hydrogen atoms in organic molecules subject to suitable weighted distance requirements.

The method of fitting a molecule to given positions is also useful for finding 'best' atomic parameters from the results of preliminary Fourier work. It may be applied when Fourier peaks are not resolved and

even when some Fourier peaks are missing, provided that enough distances can be specified to define all atomic positions sought.

The method may further be modified to apply to two-dimensional Fourier syntheses. For purposes of discussion we assume a projection along the crystallographic axis \mathbf{a}_3 . The positional requirements leading to R_3 must then be modified to include only terms affecting the x_1 and x_2 parameters, while the distance requirements leading to R_1 are, of course, kept three-dimensional. Enough distances must be specified to define all relative x_3 parameters. To keep the molecule to be fitted from shifting arbitrarily along \mathbf{a}_3 , the x_3 parameter of one atom must be given a fixed value, like 0. Convergence to a solution may be slow, unless the starting parameters x_3^0 are given roughly correct relative values. This also serves to remove another source of ill definition that exists when \mathbf{a}_1 and \mathbf{a}_2 are perpendicular to \mathbf{a}_3 (or nearly so), in which case at least two solutions exist, related by a mirror plane (or near mirror plane) perpendicular to \mathbf{a}_3 . (Additional solutions exist if mirroring of parts of the molecule is compatible with the set of preassigned distances.)

We have found that when the fitting of a molecule to prescribed positions requires large shifts it may be necessary for the convergence of the iterations to scale the shifts so that the largest of them does not exceed a certain prescribed value. In our experience parameter shifts that correspond to more than about 0.1 Å are detrimental to rapid convergence.

I wish to thank Dr Richard E. Marsh for valuable advice and criticism, Dr Robert A. Sparks for helpful discussions and for the writing of the FORTRAN program that fits a molecule to given positions, and Mr Robert S. Deverill for his help with the incorporation of this program into the *SFLS* block diagonal program.

References

- CRUICKSHANK, D. W. (1952). *Acta Cryst.* **5**, 511.
GANTZEL, SPARKS & TRUEBLOOD (1961). Modification of full matrix program No. 384 listed in *International Union of Crystallography World List of Computer Programs*, Sept. 1962.
HUGHES, E. W. (1941). *J. Amer. Chem. Soc.* **63**, 1737.
LIPSON, H. & COCHRAN, W. (1953). *The Determination of Crystal Structures*. London: Bell.
SPARKS, R. A. (1958). Ph. D. Thesis, University of California at Los Angeles, California, U.S.A.
WASER, J. & WATSON, W. H. (1963). *Nature, Lond.* **198**, 1297.