18.5.2. The least-squares method

18.5.2.1. *The normal equations*

In the unrestrained least-squares method, the residual

$$R = \sum_3 w(hkl)\Delta^2(hkl) \qquad (18.5.2.1)$$

is minimized, where Δ is either $|F_o| - |F_c|$ for R_1 or $|F_o|^2 - |F_c|^2$ for R_2, and $w(hkl)$ is chosen appropriately. The summation is over crystallographically independent planes.

When R is a minimum with respect to the parameter u_j, $\partial R/\partial u_j = 0$, *i.e.*,

$$\sum_3 w\Delta(\partial\Delta/\partial u_j) = 0. \qquad (18.5.2.2)$$

For R_1, $\partial\Delta/\partial u_j = -\partial|F_c|/\partial u_j$; for R_2, $\partial\Delta/\partial u_j = -2|F_c|\partial|F_c|/\partial u_j$. The n parameters have to be varied until the n conditions (18.5.2.2) are satisfied. For a trial set of the u_j close to the correct values, we may expand Δ as a function of the parameters by a Taylor series to the first order. Thus for R_1,

$$\Delta(\mathbf{u}+\mathbf{e}) = \Delta(\mathbf{u}) - \sum_i \varepsilon_i(\partial|F_c|/\partial u_i), \qquad (18.5.2.3)$$

where ε_i is a small change in the parameter u_i, and $\mathbf{u}$ and $\mathbf{e}$ represent the whole sets of parameters and changes. The minus sign occurs before the summation, since $\Delta = |F_o| - |F_c|$, and the changes in $|F_c|$ are being considered.

Substituting (18.5.2.3) in (18.5.2.2), we get the *normal equations* for R_1,

$$\sum_i \varepsilon_i \left[\sum_3 w(\partial|F_c|/\partial u_i)(\partial|F_c|/\partial u_j)\right] = \sum_3 w\Delta(\partial|F_c|/\partial u_j). \qquad (18.5.2.4)$$

There are n of these equations for $j = 1, \ldots, n$ to determine the n unknown ε_j.

For R_2 the normal equations are

$$\sum_i \varepsilon_i \left[\sum_3 w(\partial|F_c|^2/\partial u_i)(\partial|F_c|^2/\partial u_j)\right] = \sum_3 w\Delta(\partial|F_c|^2/\partial u_j). \qquad (18.5.2.5)$$

Both forms of the normal equations can be abbreviated to

$$\sum_i \varepsilon_i a_{ij} = b_j. \qquad (18.5.2.6)$$

For the values of $\partial|F_c|/\partial u_j$ for common parameters see, *e.g.*, Cruickshank (1970).

Some important points in the derivation of the standard uncertainties of the refined parameters can be most easily understood if we suppose that the matrix a_{ij} can be approximated by its diagonal elements. Each parameter is then determined by a single equation of the form

$$\varepsilon_i \sum_3 wg^2 = \sum_3 wg\Delta, \qquad (18.5.2.7)$$

where $g = \partial|F_c|/\partial u_i$ or $\partial|F_c|^2/\partial u_i$. Hence

$$\varepsilon_i = \left(\sum_3 wg\Delta\right) \Big/ \left(\sum_3 wg^2\right). \qquad (18.5.2.8)$$

At the conclusion of the refinement, when R is a minimum, the variance (square of the s.u.) of the parameter u_i due to uncertainties in the Δ's is

$$\sigma_i^2 = \left[\sum_3 w^2 g^2 \sigma^2(F)\right] \Big/ \left(\sum_3 wg^2\right)^2. \qquad (18.5.2.9)$$

If the weights have been chosen as $w(hkl) = 1/\sigma^2(|F_{hkl}|)$ or $1/\sigma^2(|F_{hkl}|^2)$, this simplifies to

$$\sigma_i^2 = 1 \Big/ \left(\sum_3 wg^2\right) = 1/a_{ii}, \qquad (18.5.2.10)$$

which is appropriate for absolute weights. Equation (18.5.2.10) provides an s.u. for a parameter relative to the s.u.'s $\sigma(|F|)$ or $\sigma(|F|^2)$ of the observations.

In general, with the full matrix a_{ij} in the normal equations,

$$\sigma_i^2 = (a^{-1})_{ii}, \qquad (18.5.2.11)$$

where $(a^{-1})_{ii}$ is an element of the matrix inverse to a_{ij}. The covariance of the parameters u_i and u_j is $\mathrm{cov}(i,j) \equiv \sigma_i\sigma_j\mathrm{correl}(i,j) = (a^{-1})_{ij}$.

18.5.2.2. *Weights*

In the early stages of refinement, artificial weights may be chosen to accelerate refinement. In the final stages, the weights must be related to the precision of the structure factors if parameter variances are being sought. There are two distinct ways, covering two ranges of error, in which this may be done.

(1) The weights for R_1, say, may reflect the precision of the $|F_o|$, so that $w(hkl) = 1/\sigma^2(|F_{hkl}|)$, where σ^2 is the estimated variance of $|F_o|$ due to a specific class of experimental uncertainties. These absolute weights are derived from an analysis of the experiment. Weights chosen in this way lead to estimated parameter variances $\sigma_i^2 = (a^{-1})_{ii}$, (18.5.2.11), which cover only the specific class of experimental uncertainties.

(2) The weights may reflect the trends in the $|\Delta| \equiv \|F_o| - |F_c\|$. A weighting function with a small number of parameters is chosen so that the averages of $w\Delta^2$ are constant when the set of $w\Delta^2$ values is analysed in any pertinent fashion (*e.g.* in bins of increasing $|F_o|$ and $2\sin\theta/\lambda$). Weights chosen in this way are relative weights, and the expression for the parameter variances needs a scaling factor,

$$S^2 = \left(\sum_3 w\Delta^2\right) \Big/ (n_{\mathrm{obs}} - n_{\mathrm{params}}). \qquad (18.5.2.12)$$

Hence, in the full-matrix case,

$$\sigma_i^2 = \left[\left(\sum_3 w\Delta^2\right) \Big/ (n_{\mathrm{obs}} - n_{\mathrm{params}})\right](a^{-1})_{ii}, \qquad (18.5.2.13)$$

which allows for all random experimental errors, such systematic experimental errors as cannot be simulated in the $|F_c|$ and imperfections in the calculated model.

18.5.2.3. *Statistical descriptors and goodness of fit*

In recent years, there have been developments and changes in statistical nomenclature and usage. Many aspects are summarised in the reports of the IUCr Subcommittee on Statistical Descriptors in Crystallography (Schwarzenbach *et al.*, 1989, 1995). In the second report, *inter alia*, the Subcommittee emphasizes the terms *uncertainty* and *standard uncertainty* (s.u.). The latter is a replacement for the older term *estimated standard deviation* (e.s.d.). The Subcommittee classify uncertainty components in two categories, based on their method of evaluation: type A, estimated by the statistical analysis of a series of observations, and type B, estimated otherwise. As an example of the latter, a type B component could allow for doubts concerning the estimated shape and dimensions of the diffracting crystal and the subsequent corrections made for absorption.

The square root S of the expression S^2, (18.5.2.12) above, is called the *goodness of fit* when the weights are the reciprocals of the absolute variances of the observations.

One recommendation in the second report does call for comment here. While agreeing that formulae like (18.5.2.13) lead to conservative estimates of parameter variances, the report suggests that this practice is based on the questionable assumption that the variances of the observations by which the weights are assigned are relatively correct but uniformly underestimated. When the goodness of fit $S > 1$, then either the weights or the model or both are suspect.

Comment is needed. The account in Section 18.5.2.2 describes two distinct ways of estimating parameter variances, covering two ranges of error. The kind of weights envisaged in the reports (based on variances of type A and/or of type B) are of a class described for method (1). They are not the weights to be used in method (2) (though they may be a component in such weights). Method (2) implicitly assumes from the outset that there are experimental errors, some covered and others not covered by method (1), and that there are imperfections in the calculated model (as is obviously true for proteins). Method (2) avoids exploring the relative proportions and details of these error sources and aims to provide a realistic estimate of parameter uncertainties which can be used in external comparisons. It can be formally objected that method (2) does not conform to the criteria of random-variable theory, since clearly the Δ's are partially correlated through the remaining model errors and some systematic experimental errors. But it is a useful procedure. Method (1) on its own would present an optimistic view of the reliability of the overall investigation, the degree of optimism being indicated by the inverse of the goodness of fit (18.5.2.12). In method (2), if the weights are on an arbitrary scale, then S^2 can have an arbitrary value.

For an advanced-level treatment of many aspects of the refinement of structural parameters, see Part 8 of *International Tables for Crystallography*, Volume C (1999). The detection and treatment of systematic error are discussed in Chapter 8.5 therein.

18.5.3. Restrained refinement

18.5.3.1. *Residual function*

Protein structures are often refined by a restrained refinement program such as *PROLSQ* (Hendrickson & Konnert, 1980). Here, a function of the type

$$R' = \sum w_h(\Delta F)^2 + \sum w_{\rm geom}(\Delta Q)^2 \qquad (18.5.3.1)$$

is minimized, where Q denotes a geometrical restraint such as a bond length. Formally, all one is doing is extending the list of observations. One is adding to the protein diffraction data geometrical data from a stereochemical dictionary such as that of Engh & Huber (1991). A chain C—N bond length may be known from the dictionary with much greater precision $1/w_{\rm geom}^{1/2}$, say 0.02 Å, than from an unrestrained diffraction-data-only protein refinement.

In a high-resolution unrestrained refinement of a small molecule, the standard uncertainty (s.u.) of a bond length A—B is often well approximated by

$$\sigma(l) = (\sigma_A^2 + \sigma_B^2)^{1/2}. \qquad (18.5.3.2)$$

However, in a protein determination $\sigma(l)$ is often much smaller than either σ_A or σ_B because of the excellent information from the stereochemical dictionary, which correlates the positions of A and B.

Laying aside computational size and complexity, the protein precision problem is straightforward in principle. When a restrained refinement has converged to an acceptable structure and the shifts in successive rounds have become negligible, invert the full matrix. The inverse matrix immediately yields estimates of the variances and covariances of all parameters.

The dimensions of the matrix are the same whether or not the refinement is restrained. The full matrix will be rather sparse, but not nearly as sparse as in a small-molecule refinement. For the purposes of Section 18.5.3, it is irrelevant whether the residual for the diffraction data is based on $|F|$ or $|F|^2$. On the relative weighting of the diffraction and restraint terms, see Section 18.5.3.3.

18.5.3.2. *A very simple protein model*

Some aspects of restrained refinement are easily understood by considering a *one-dimensional protein consisting of two like atoms* in the asymmetric unit, with coordinates x_1 and x_2 relative to a fixed origin and bond length $l = x_2 - x_1$. In the refinement, the normal equations are of the type $\mathbf{N}\Delta\mathbf{x} = \mathbf{e}$. For two non-overlapping like atoms, the *diffraction data* will yield a normal matrix

$$\mathbf{N} = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}, \qquad (18.5.3.3)$$

with inverse

$$\begin{pmatrix} 1/a & 0 \\ 0 & 1/a \end{pmatrix}, \qquad (18.5.3.4)$$

where

$$a = \sum w_h(\partial|F_n|/\partial x_i)^2. \qquad (18.5.3.5)$$

A *geometric restraint* on the length will yield a normal matrix

$$\begin{pmatrix} b & -b \\ -b & b \end{pmatrix} \qquad (18.5.3.6)$$

with no inverse, since its determinant is zero, where

$$b = w_{\rm geom}(\partial l/\partial x_i)^2. \qquad (18.5.3.7)$$

Note $\partial l/\partial x_2 = -\partial l/\partial x_i = 1$, so that

$$b = w_{\rm geom} = 1/\sigma_{\rm geom}^2(l), \qquad (18.5.3.8)$$

where $\sigma_{\rm geom}^2(l)$ is the variance assigned to the length in the stereochemical dictionary.

Combining the diffraction data and the restraint, the normal matrix becomes

$$\begin{pmatrix} a+b & -b \\ -b & a+b \end{pmatrix}, \qquad (18.5.3.9)$$

with inverse

$$\{1/[a(a+2b)]\}\begin{pmatrix} a+b & b \\ b & a+b \end{pmatrix}. \qquad (18.5.3.10)$$

For the diffraction data alone, the variance of x_i is

$$\sigma_{\rm diff}^2(x_i) = 1/a. \qquad (18.5.3.11)$$

For the diffraction data plus restraint, the variance of x_i is

$$\sigma_{\rm res}^2(x_i) = (a+b)/[a(a+2b)] < \sigma_{\rm diff}^2(x_i). \qquad (18.5.3.12)$$

Note that though the restraint says nothing about the position of x_i, the variance of x_i has been reduced because of the coupling to the position of the other atom. In the limit when $a \ll b$, $\sigma_{\rm res}^2(x_i)$ is only half $\sigma_{\rm diff}^2(x_i)$.