A New Approach to the Least-Squares Refinement of Highly Covarying Parameters in Crystal Structure Determinations

BY A. D. RAE

Department of Physical Chemistry, The University of New South Wales, Kensington, N.S.W. 2033, Australia

When two parameters covary greatly, changes in these parameters should be considered together. By selecting new parameters which are the sums and the differences of the parameter shifts of highly covarying parameters and by considering the transformed least-squares equations, it is possible to discuss ways of optimizing the least-squares method.

Introduction

The X-ray structure factor for a particular crystal may be expressed as $F(S) = \sum_{n=1}^{N} f_n T_n(S) \exp(2\pi i \mathbf{r}_n . S)$ where

N is the number of atoms in the unit cell, S is a reciprocal lattice vector, $|S| = 2 \sin \theta / \lambda$, and $\mathbf{r}_n$ is the position of the nth atom of scattering power $f_n T_n(S)$, where $T_n(S)$ is the anisotropic temperature factor.

Let us now transform all points $\mathbf{r}$ in the unit cell to $(\theta, \mathbf{t})\mathbf{r} = \theta\mathbf{r} + \mathbf{t}$ where θ is a point group operation allowed by the lattice and $\mathbf{t}$ is a translation. The X-ray structure factor for this new set of points may be expressed as

$$F_{\theta, t}(S) = \exp(2\pi i \mathbf{t} . S) F(\theta^{-1} S).$$

Thus we have the identity

$$\tfrac{1}{2}(|F(S)|^2 + |F(\theta^{-1}S)|^2)$$
$$= \tfrac{1}{4}(|F(S) + F_{\theta, t}(S)|^2 + |F(S) - F_{\theta, t}(S)|^2)$$

so that for any particular class of reflexions ($h + k = 2n$, etc.)

$$\langle |F(S)|^2 \rangle = \tfrac{1}{4} \langle |F(S) + F_{\theta, t}(S)|^2 \rangle$$
$$+ \tfrac{1}{4} \langle |F(S) - F_{\theta, t}(S)|^2 \rangle \quad (1)$$

where $\langle A \rangle$ means the expectation value of A.

Now $\tfrac{1}{4}|F(S) + F_{\theta, t}(S)|^2$ is the value of $|F(S)|^2$ we would obtain if for every atom at $\mathbf{r}_n$ we create an identical atom at $\theta\mathbf{r}_n + \mathbf{t}$ and give each atom half weight. We shall call this arrangement of atoms the 'mean' structure.

Also $\tfrac{1}{4}|F(S) - F_{\theta, t}(S)|^2$ is the value of $|F(S)|^2$ we would obtain if for every atom at $\mathbf{r}_n$ we create an identical atom at $\theta\mathbf{r}_n + \mathbf{t}$ but of negative scattering power and give each atom half weight. We shall call this arrangement of atoms the 'difference' structure.

If $(\theta, \mathbf{t})$ is a symmetry element of the space group then $F_{\theta, t}(S) = F(S)$ and the final term of expression (1) is zero. If (θ, t) is not a symmetry element of the space group then the final term of expression (1) makes a contribution of between 0 and $\tfrac{1}{2} \langle |F(S)|^2 \rangle$. The contri-

bution is only $\tfrac{1}{2} \langle |F(S)|^2 \rangle$ if atoms at $\mathbf{r}_n$ and atoms created at $\theta\mathbf{r}_n + \mathbf{t}$ can be resolved at the particular value of $|S|$. If the contribution is less than $0.1 \langle |F(S)|^2 \rangle$ then we can say that $(\theta, \mathbf{t})$ is a pseudo symmetry element, and there must be significant overlap of the atoms at $\mathbf{r}_n$ and those created at $\theta\mathbf{r}_n + \mathbf{t}$.

If there is an index condition on S such that $F_{\theta, t}(S) = -F(S)$ for some particular S values, then the only contribution to such reflexions is from the 'difference' structure.

If there is an index condition on S such that $F_{\theta, t}(S) = F(S)$ for some particular S values then the only contribution to such reflexions is from the mean structure. Symmetry elements $(\theta, \mathbf{t})$ where $\mathbf{t}$ is not a unit cell repeat allow this separation on certain reflexions.

Another common pseudo symmetry element is the inversion operation, $\bar{1}$. The phase for the 'difference' structure is $\pm \pi/2$ from the phase for the 'mean' structure for any particular value of S and so no information about the phase of the structure factors for the 'difference' structure can be obtained from refinement of the 'mean' structure.

It is common in pseudosymmetric structures to refine the structure initially assuming that the pseudosymmetry is real. Artificial peaks in the electron density map are created at points $\theta\mathbf{r}_n + \mathbf{t}$ and chemical intuition can be used to select the real structure, thus giving some initial information for the refinement of the 'difference' structure. However, refinement at this stage often gives poorer values of bond lengths and bond angles than can be obtained by selecting the real structure from the 'mean' structure. The reason for this can be attributed to the high covariance between parameters that were previously regarded as equal, and the fact that little effort has been made to eliminate the noise resulting from incorrect evaluation of the magnitude and phase of $F_o - F_c$.

Theory

Let us first consider a 2×2 matrix with high covariance. The least-squares equations may be written most generally as

$$\begin{vmatrix} a_{11} & \pm\sqrt{a_{11}a_{22}}(1-\delta) \\ \pm\sqrt{a_{11}a_{22}}(1-\delta) & a_{22} \end{vmatrix} \begin{vmatrix} \Delta u_1 \\ \Delta u_2 \end{vmatrix}$$

$$= \begin{vmatrix} \sqrt{a_{11}}(\beta+\gamma) \\ \pm\sqrt{a_{22}}(\beta-\gamma) \end{vmatrix} = \begin{vmatrix} b_1 \\ b_2 \end{vmatrix}.$$

If $\delta=0$ then $\gamma=0$ and the second equation is a repeat of the first, namely $\sqrt{a_{11}}\Delta u_1 \pm \sqrt{a_{22}}\Delta u_2 = \beta$ and so we can arbitrarily say $\sqrt{a_{11}}\Delta u_1 \mp \sqrt{a_{22}}\Delta u_2 = 0$. The diagonal terms of the 2×2 matrix are positive but the off-diagonal terms can be positive or negative, and we shall consider both cases simultaneously.

$$\text{If } \delta \neq 0 \text{ then } \sqrt{a_{11}}\Delta u_1 = \frac{\beta}{2-\delta} + \frac{\gamma}{\delta}$$

and

$$\sqrt{a_{22}}\Delta u_2 = \pm\left(\frac{\beta}{2-\delta} - \frac{\gamma}{\delta}\right)$$

so that

$$\sqrt{a_{11}}\Delta u_1 \pm \sqrt{a_{22}}\Delta u_2 = \frac{2\beta}{2-\delta} = \frac{1}{2-\delta}\left(\frac{b_1}{\sqrt{a_{11}}} \pm \frac{b_2}{\sqrt{a_{22}}}\right)$$

and

$$\mp\sqrt{a_{11}}\Delta u_1 + \sqrt{a_{22}}\Delta u_2 = \mp\frac{2\gamma}{\delta}$$

$$= \frac{1}{\delta}\left(\mp\frac{1}{\sqrt{a_{11}}}b_1 + \frac{1}{\sqrt{a_{22}}}b_2\right).$$

However, we could regard the least-squares equations as a combination of two independent least-squares equations.

$$(a) \quad (\sqrt{a_{11}}\Delta u_{1a} \pm \sqrt{a_{22}}\Delta u_{2a})\frac{2-\delta}{2} = \beta$$

where

$$\mp\sqrt{a_{11}}\Delta u_{1a} + \sqrt{a_{22}}\Delta u_{2a} = 0,$$

and

$$(b) \quad (\mp\sqrt{a_{11}}\Delta u_{1b} + \sqrt{a_{22}}\Delta u_{2b})\left(\frac{\delta}{2}\right) = \mp\gamma$$

where

$$\sqrt{a_{11}}\Delta u_{1b} \pm \sqrt{a_{22}}\Delta u_{2b} = 0.$$

Equations (a) refine the 'mean' structure and equations (b) refine the 'difference' structure. $\Delta u_1 = \Delta u_{1a} + \Delta u_{1b}$ and $\Delta u_2 = \Delta u_{2a} + \Delta u_{2b}$.

We can express the least-squares equations in matrix notation as $AU=B$. These equations can be re-expressed as $A_1U_1=B_1$ where $U=CU_1$, $B_1=C^+B$ and $A_1=C^+AC$. If we change from variables Δu_1, Δu_2 to variables

$$(\sqrt{a_{11}}\Delta u_1 \pm \sqrt{a_{22}}\Delta u_2) \text{ and } (\mp\sqrt{a_{11}}\Delta u_1 + \sqrt{a_{22}}\Delta u_2)$$

$$\text{then } C = \tfrac{1}{2}\begin{vmatrix} \dfrac{1}{\sqrt{a_{11}}} & \mp\dfrac{1}{\sqrt{a_{11}}} \\ \pm\dfrac{1}{\sqrt{a_{22}}} & \dfrac{1}{\sqrt{a_{22}}} \end{vmatrix}$$

$$\text{giving } B_1 = \tfrac{1}{2}\begin{vmatrix} \dfrac{b_1}{\sqrt{a_{11}}} \pm \dfrac{b_2}{\sqrt{a_{22}}} \\ \mp\dfrac{b_1}{\sqrt{a_{11}}} + \dfrac{b_2}{\sqrt{a_{22}}} \end{vmatrix} = \begin{vmatrix} \beta \\ \mp\gamma \end{vmatrix}$$

and

$$A_1 = \begin{vmatrix} \tfrac{1}{2}(2-\delta) & 0 \\ 0 & \dfrac{\delta}{2} \end{vmatrix} \text{ thus giving } U = (CA_1^{-1}C^+)B.$$

In the light of these observations we are now in a position to discuss the least-squares refinement for structures containing a twofold pseudo symmetry element.

Had we simply ignored off-diagonal terms in the least-squares equations, we would have obtained

$$\sqrt{a_{11}}\Delta u_1 = \beta+\gamma.$$

This overweights the change in the 'mean' structure and underweights the change in the 'difference' structure. To get the 'mean' structure to refine, a fractional shift is commonly applied to all parameters, meaning that the refinement of the difference structure is weighted out of existence. Oscillation of the parameters of the mean structure is commonly regarded as indicative of convergence.

If a full-matrix approach is employed, it is found that excessive shifts in the difference structure are obtained.

In our simple 2×2 problem, the difference structure refinement depended on γ/δ and $\gamma \to 0$ as $\delta \to 0$ and it appears that γ/δ is not usually well behaved in actual practice, even when δ is reasonably large, say 0·2.

There are obvious advantages to be obtained by regarding the refinement of the 'mean' and 'difference' structures independently of each other. This can be done by pairing pseudo-equivalent atoms. We invent new parameters as linear combinations of old parameters to give a new set of least-squares equations $A_1U_1 = B_1$ from the old set of least-squares equations $AU=B$. As before, $U=CU_1$, $B_1=C^+B$, $A_1=C^+AC$ and $U=(CA_1^{-1}C^+)B$.

If consider the highly covarying pair of variables $(\Delta u_i, \Delta u_j)$ then

then

$$\Delta u_i = \frac{1}{2\sqrt{a_{ii}}}(\Delta u_{1i'} \mp \Delta u_{1j'})$$

$$\Delta u_j = \frac{1}{2\sqrt{a_{jj}}}(\pm \Delta u_{1i'} + \Delta u_{1j'})$$

i.e.

$$\Delta u_{1i'} = \sqrt{a_{ii}}\Delta u_i \pm \sqrt{a_{jj}}\Delta u_j$$

$$\Delta u_{1j'} = \mp\sqrt{a_{ii}}\Delta u_i + \sqrt{a_{jj}}\Delta u_j. \tag{1}$$

$\Delta u_{1i'}$ refines the mean structure and $\Delta u_{1j'}$ refines the difference structure. The sign of the off-diagonal term

a_{ij} of A determines the signs used. The upper sign is chosen if a_{ij} is positive, the lower sign if a_{ij} is negative. For example, if the pseudo-symmetry operation relating the variable Δu_i to the variable Δu_j is the inversion operator, $\bar{1}$, then a_{ij} is positive for temperature parameters but negative for positional parameters.

If a variable does not have a variable with high covariance associated with it (e.g. scale constant, extinction coefficient, a variable for an atom without a pseudo symmetric equivalent) then we can still use equations (1) by inventing a parameter $\Delta u_j \equiv 0$. This can be done by expanding the size of A and making $a_{jj} = 1$, $a_{ij} = 0$ ($i \neq j$) and $b_j = 0$. Thus the matrix C^{-1} is known ($U_1 = C^{-1} U$) so that the inverse transformation $A = (C^+)^{-1} A_1 C^{-1}$ can be evaluated.

The first $n/2$ equations $A_1 U_1 = B_1$ refine the 'mean' structure and the final $n/2$ equations $A_1 U_1 = B_1$ refine the 'difference' structure (dimension of A_1 is $n \times n$) if we label our new variables $\Delta u_{1i'}$ ($i' = 1$ to n) in a convenient way and eliminate all covariance in the A_1 matrix by setting all elements not in one of the two ($n/2 \times n/2$) diagonal blocks to zero. This prevents the refinement of the 'mean' structure and any noise associated with it from upsetting the refinement of the 'difference' structure. Fractional shifts, ω, can be applied to either block, by multiplying the elements of A_1 by ω^{-1}. The modified matrix A_1' can then be transformed back to a modified matrix A' for the variables Δu, $[A' = (C^{-1})^+ A_1' C^{-1}]$, and the extra dummy variables eliminated.

The least-squares refinement program can then proceed in the usual manner.

The modification of the matrix A to A' can be done in an auxiliary program, specific to the example under consideration by having a read-out/read-in interruption in the least-squares refinement program prior to the actual matrix inversion.

An alternative approach to the problem is to use a block-diagonal solution where only variables associated with a single atom are allowed to covary. One can then evaluate the changes in a pair of highly covarying parameters as

$$\Delta u_i' = \omega_1(\Delta u_i \pm \Delta u_j) + \omega_2(\Delta u_i \mp \Delta u_j)$$
$$\Delta u_j' = \mp[-\omega_2(\Delta u_i \pm \Delta u_j) + \omega_1(\Delta u_i \mp \Delta u_j)]$$

where the upper sign is used if the term a_{ij} relating variables Δu_i and Δu_j is positive. The weight ω_1 is the weight for the refinement of the 'mean' structure and ω_2 is the weight for the refinement of the 'difference' structure. If the term a_{ij} is given by $\pm \sqrt{(a_{ii} a_{jj})} (1 - \delta)$ where δ is small then ideally

$$\omega_1 = \frac{1}{2 - \delta} \text{ and } \omega_2 = \frac{1}{\delta}.$$

We should now consider the errors in refined parameters. The covariance between two highly covarying parameters has the effect of making the value of the standard error for each parameter far worse than if

the pseudo symmetry were considered real. Consider our simple 2×2 problem again.

$$A = \begin{vmatrix} a_{11} & \pm \sqrt{(a_{11} a_{22})} (1 - \delta) \\ \pm \sqrt{(a_{11} a_{22})} (1 - \delta) & a_{22} \end{vmatrix}.$$

The evaluation of the variance and covariance depends upon the elements of

$$A^{-1} = \frac{1}{a_{11} a_{22} \delta(2 - \delta)} \begin{vmatrix} a_{22} & \mp \sqrt{(a_{11} a_{22})} (1 - \delta) \\ \mp \sqrt{(a_{11} a_{22})} (1 - \delta) & a_{11} \end{vmatrix}$$

Then $\text{var}(u_1) = \dfrac{1}{a_{11} \delta(2 - \delta)} \dfrac{\sum \omega \Delta^2}{n - k}$ where n is the number of observations and k is the number of independent variables. However, assuming $a_{11} = a_{22}$, the covariance between the parameters u_1 and u_2 makes
$$\text{cov}(u_1 + u_2, u_1 - u_2) = 0$$

$$\text{var}\left[\tfrac{1}{2}(u_1 \pm u_2)\right] = \frac{1}{2a_{11}(2 - \delta)} \frac{\sum \omega \Delta^2}{n - k}$$

$$\text{var}\left[\tfrac{1}{2}(u_1 \mp u_2)\right] = \frac{1}{2a_{11}\delta} \frac{\sum \omega \Delta^2}{n - k}.$$

We see that the 'mean' structure is still accurately determined if $\delta \simeq 0$ but the 'difference' structure is not, though a reduction in the value of $\sum \omega \Delta^2$ made possible by the extra degrees of freedom allowed to the solution should compensate for this.

Extension of the problem to the case when there is more than one pseudo symmetry operation is feasible We change a set of n variables Δu_i which shows high covariance between members of the set to the variable $\dfrac{1}{n} \sum_{i=1}^{n} \pm \sqrt{a_{ii}} \Delta u_i$ and the n variables $\pm \sqrt{a_{jj}} \Delta u_j -$

$\dfrac{1}{n} \sum_{i=1}^{n} \pm \sqrt{a_{ii}} \Delta u_i$; $j = 1$ to n. One of these final n variables must be eliminated as redundant. The upper sign is used for $\pm \sqrt{a_{ii}}$ if the element a_{ij} is positive for some particular Δu_k in the set.

If we consider $a_{ij} = \pm \sqrt{a_{ii} a_{jj}} (1 - \delta)$ where δ is a constant for any i, j and then consider the n variables in isolation we find

$$\Delta u_i = \frac{\langle b \rangle}{n - (n-1)\delta} + \left(\frac{b_i}{\sqrt{a_{ii}}} - \langle b \rangle \right) \Big/ \delta$$

where

$$\langle b \rangle = \frac{1}{n} \sum_{j=1}^{n} \pm \frac{1}{\sqrt{a_{jj}}} b_j.$$

b_j is the jth member of the column matrix B and the upper sign is used if a_{ij} is positive. We see the problem still has a great resemblance to the initial two-variable problem. Δu_i is still considered to be a consequence of the refinement of the mean structure and the difference from the mean and a weighting scheme for a block diagonal solution may accordingly be evaluated.

Likewise, the elimination of the noise of the refinement of the mean structure from the A matrix may be performed in a similar manner as before.

The pseudosymmetry operator which gives high covariance between parameters need not always allow a sensible selection of combination of parameters. An example is the pseudosymmetry operators $(\frac{1}{3}+x, y, \frac{1}{2}-z)$, $(\frac{2}{3}+x, y, \frac{1}{2}-z)$. A refinement procedure for a crystal structure with these pseudosymmetry operators is currently being investigated, and will be published elsewhere.

We have seen earlier that a full-matrix solution tends to overweight the changes in the 'difference' structure whereas a block diagonal approach overweights the changes in the 'mean' structure but underweights the 'difference' structure, and so other refinement procedures may be considered.

As stated in the Introduction, the refinement of the 'difference' structure is only possible as $(\sin\theta)/\lambda$ increases. Thus the simplest method to reduce covariance between highly covarying parameters is to use high-angle data or to weight data according to $(\sin\theta)/\lambda$. Similarly, if there is an index condition so that only the 'difference' structure contributes to certain reflexions then the weight of these reflexions can be increased.

A combination of a full-matrix approach and a diagonal approach can be considered. This is simply brought about by multiplying the diagonal elements a_{ii} of the A matrix by a constant, k_1, greater than 1, and the elements of the column matrix B by a constant k_2. The constant k_2 should be less than k_1, otherwise

the refinement of the 'mean' structure will be over-weighted. A down weighting of the refinement of the 'difference' structure is brought about by a reduction of the apparent covariance between parameters. The shifts of highly covarying parameters are reduced more than the shifts in parameters which are less covariant. A read-out/read-in interruption to the least-squares refinement program prior to the inversion of the A matrix allows any number of modifications to be tested without re-evaluating the A matrix each time.

This final method has a lot to recommend it. Consider our simple two-variable example again. The equation

$$\sqrt{a_{11}}\Delta u_1 = \frac{\beta}{2-\delta} + \frac{\gamma}{\delta}$$

becomes

$$\frac{k_1}{k_2}\sqrt{a_{11}}\Delta u_1 = \frac{\beta}{2-\delta'} + \frac{\gamma}{\delta'} = \omega_1\beta + \omega_2\gamma$$

where

$$1-\delta' = k_1^{-1}(1-\delta); \quad k_1 > 1 .$$

δ'/δ varies from ∞ for $\delta=0$ to 1 for $\delta=1$. Also $(2-\delta)/\delta$ varies from ∞ for $\delta=0$ to 1 for $\delta=1$, whereas $\omega_2/\omega_1 = (2-\delta')/\delta'$ varies from $(k_1+1)/(k_1-1)$ for $\delta=0$ to 1 for $\delta=1$.

Thus we have a built-in discrimination against those parameters of the 'difference' structure which are least accurately determined. If we choose k_1 as 1·4 say, then ω_2/ω_1 varies between 6 and 1. This procedure will also work when a larger set of covarying variables is considered.

The Use of Phase Relationships between Quartets of Reflexions

By H. Schenk

Laboratory for Crystallography, University of Amsterdam, Nieuwe Prinsengracht 126, Amsterdam, The Netherlands

Phase relations between quartets of reflexions can be found either directly from a generalized Hughes expression or by elimination of the phase of a reflexion common to two Σ_2 relationships. By combining the information from both types, strengthened quartet relationships (SQR) can be constructed, which are comparable to the Σ_2 relations in reliability and number. Most of the reliable SQR's involve the strong reflexions only. The phases of these reflexions are related by a highly overdetermined system of equations leading in a simple manner to a good starting set.

List of symbols

$F(H)$	Structure factor of reflexion H
E_H	Normalized structure factor
U_H	Unitary structure factor
f	Atomic scattering factor
g	Scattering factor of the squared electron density of an atom
N	Number of atoms in the unit cell
V	Volume of the unit cell
$S(H)$	The sign of reflexion H
φ_H	The phase of reflexion H