Journal of
Applied Crystallography

ISSN 0021-8898

Received 24 January 2008
Accepted 17 March 2008

Structure refinement: some background theory and practical strategies

David Watkin

Chemical Crystallography Laboratory, University of Oxford, UK. Correspondence e-mail: david.watkin@chem.ox.ac.uk

Most modern small-molecule refinement programs are based on similar algorithms. Details of these methods are scattered through the literature, sometimes in books that are no longer in print and usually in mathematical detail that makes them unattractive to nonprogrammers. This paper aims to discuss these well established algorithms in nonmathematical language, with the intention of enabling crystallographers to use their favourite programs effectively.

1. Introduction

Refinement is a general term that refers to almost all the operations needed to develop a trial model into one that best represents the observed data. Just as there is not a well defined mathematical technique for extracting valid phases from the observed intensities, so also there is no single well defined path from the trial model to the completed structure – if there were, it would have been programmed long ago. There are, however, some well trodden trails to guide a structure analyst, together with a growing number of validation tools. Refinement is a step-wise procedure, with increasingly subtle features being introduced in order to develop the model. Physical and chemical validation is a key feature of every stage of a refinement. In very difficult cases it is rare that mathematics alone will lead to an acceptable structure. In these cases knowledge of the chemistry or physical properties of the material may help to resolve uncertainties. Difficulties in programming 'scientific experience' have prevented the development of fully automatic structure analysis systems, though modern programs are beginning to bring 'do-it-yourself' structure analyses into the hands of nonspecialists. This paper provides an overview of some of the extensive literature on refinement techniques with the aim of helping the analyst to understand how modern programs work. It contains no recipes that will work in every case, though the general strategy is to develop a trial model stepwise into an acceptable representation of the crystalline physical specimen.

The trial model consists of parameters describing the mean positions of the atoms in the unit cell (the atomic coordinates), the amplitudes of the average displacements from these positions (the atomic displacement parameters, ADPs) and other parameters related to the sample itself or the experimental technique [the overall scale factor, extinction coefficient, the Flack (1983) parameter, twin fractions *etc.*]. Approximate atomic coordinates are usually obtained from Fourier syntheses or sometimes, in the case of inorganic structures, by analogy with a related material.

The ADP can represent thermal agitation of an atom (hence the older term temperature factor) or the locus of atoms in slightly different positions in different unit cells (static disorder) or, most generally, a mixture of these two effects. Thermal agitation (dynamic disorder) can sometimes be distinguished from static disorder by performing structure determinations at two different temperatures. Thermal vibrations reduce as the temperature drops. The simplest ADP model consists of isotropic displacements from the mean position ($U_{\rm iso}$). A more complex model, generally used for all non-H atoms in a modern structure determination, is a displacement ellipsoid. This is a symmetric 3×3 tensor, which means that it contains six independent terms, and is generally represented by $\mathbf{U}_{\rm aniso}$, with individual terms represented by U^{ij}. The interpretation of a two-dimensional tensor is shown in Fig. 1.

Since the ellipsoid represents a volume in which the atom could be found, this volume must be positive. An anisotropic ADP corresponding to an ellipsoid with a zero or negative volume is said to be nonpositive definite, and most programs will warn the user if this situation occurs. Either initial values for the ADPs can be estimated (based on experience) or an average value can be found from the Wilson plot (Giacovazzo *et al.*, 2002).

The Wilson plot (Fig. 2) is a graph of the log of the normalized observed intensities averaged over a short resolution interval $\log[\langle I\rangle/\sum(\text{scattering factor})^2]$ against $[\sin(\theta)/\lambda]^2$. If the electron density were distributed uniformly throughout the cell, the resulting plot would be a straight line whose gradient gives the overall isotropic atomic displacement parameter. The electron density is not distributed uniformly (otherwise it would not be a structure), so the observed plot generally shows ripples corresponding to recurrent interatomic distances. A mean gradient can still be computed. Wilson plots showing a positive slope (giving a negative ADP) indicate that there is something very unusual about the data. For heavily absorbing materials, a positive slope could imply that a multi-scan 'absorption correction' has accounted for

variations in the shape of the crystal but not for the absorption of the residual sphere, with a diameter about equal to the mean size of the crystal. For very weakly diffracting materials, the curve may show an upturn at high angles, where the diffractometer was mainly measuring noise.

A class of materials exist in which the molecules in adjacent unit cells differ slightly from each other in a systematic, incremental, way. Such materials are called modulated structures. They are characterized by a diffraction pattern consisting of a reciprocal lattice of strong reflections corresponding to the basic unit cell, with weaker reflections falling as satellites of the strong ones. If the pattern of small changes in a series of adjacent cells is itself repeated [so that the $(n+1)$th cell is the same as the first, the $(n+2)$th is the same as the second *etc*.], the modulation is said to be commensurate. If the pattern of the modulation never gets back in step with the main lattice, the modulation is called incommensurate. If an analyst fails to notice the satellite reflections and only works with the strong, principal, lattice, the resulting structure is the average of the n adjacent cells forming the modulation. If the incremental differences are in the positions of some of the atoms (*e.g.* a small rotation of a group in one cell compared with the position of the group in the previous cell) then the average structure will appear to be disordered, with large ADPs. For many analytical purposes this average structure is quite adequate. However, if the structure is to be described as fully as possible, the weak satellite reflections must also be measured. Since they do not fall on the same reciprocal lattice as the main reflections, they cannot be indexed in a conventional way with three indices h, k, l. One or more additional indices are required, corresponding to a four- or higher-dimensional space, and hence a four- or higher-dimensional space group. Software is becoming available to index area-detector images from these kinds of materials, and programs such *JANA2000* (Petricek *et al.*, 2000) can carry out refinements in this high-dimensionality space.

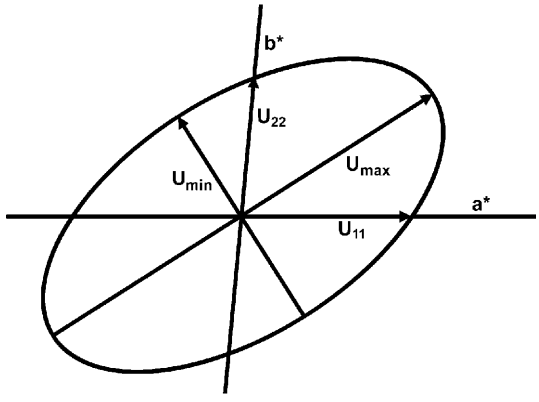

Figure 1
Anisotropic ADP. U^{11} and U^{22} are the intercepts on the a^* and b^* axes; U^{12} is related to the inclination of the ellipsoid with respect to the unit cell.

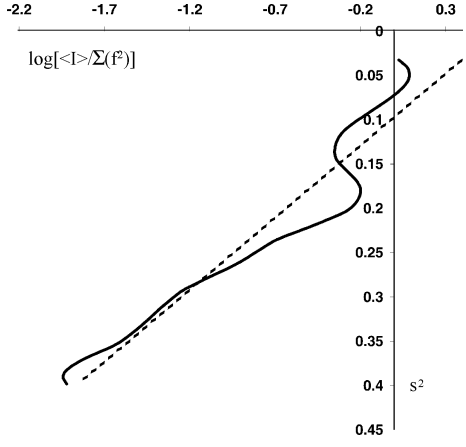

Figure 2
Wilson plot. The abscissa is $\log[\langle I\rangle/\sum(\text{scattering factor})^2]$, where the average is taken over all data in a small resolution range and the scattering factors are computed at the mean resolution of the range. The ordinate is $[\sin(\theta)/\lambda]^2$. The slope of the graph gives the overall isotropic atomic displacement parameter. The ripples are due to recurrent interatomic vectors in the material.

2. Fourier refinement

The reciprocal lattice is determined from the setting angles of a serial diffractometer or from the setting angles and image coordinates on area-detector machines. The locations of diffracted rays in the reciprocal lattice enable one to calculate the size and shape of the crystallographic unit cell and perhaps obtain a strong indication of the crystal class (or system). The positions of the diffracted intensities tell us nothing about the distribution of the atoms inside the cell.

The intensities of the diffracted beams tell us about the electron distribution in the cell, and the diffraction phenomenon is summarized in equation (1):

$$F_{hkl} = \iiint V \rho_{xyz} \exp[2\pi i(hx + ky + lz)]\, \partial x\, \partial y\, \partial z, \qquad (1)$$

where ρ_{xyz} is the electron density at the point (x, y, z). Note that the integration is over the total volume, V, of the unit cell so that the whole structure contributes to each diffracted beam. It is not possible to say that the intensity at a particular reciprocal lattice point is a result of a particular atomic feature. Note also the $\exp(2\pi i)$ term. This indicates that the diffracted wave has both magnitude and phase. In data collection, normally only the intensity I can be measured, and the structure factor (F) is obtained from the transformations $F^2 = (Lp)^{-1}I$ and $F = (F^2)^{1/2}$, where L and p are the Lorentz and polarization corrections.

When both the amplitude and the phase are known, the reverse transformation enables us to compute the electron density at any point:

$$\rho_{xyz} = V^{-1} \sum \sum \sum \left|F_{hkl}\right| \exp\left[-2\pi i(hx + ky + lz + \alpha_{hkl})\right]. \qquad (2)$$

Note here that the triple summation is over the whole of the reciprocal lattice – we cannot use just a part of the diffraction data to compute the electron density at some chosen point in the cell. Note also the modulus sign (||) around F, which means that F is a phaseless quantity. In the rest of this paper, F

will be used to represent an unphased structure amplitude and F_α one that is phased. The best value of F to use in maps is rather different from that to be used in least squares (§3.4). It certainly should never be negative, since in this calculation this corresponds to a phase change of 180°. Nor should it simply be the unsigned square root of F^2, because this introduces bias. Sivia (1996) gives a Bayesian method for estimating the 'best' positive value for weak reflections to use in Fourier syntheses. This value should not be used in least-squares refinement.

The phase of the diffracted beam occurs as α in the exponential. In general, the phases cannot easily be measured [but see Pringle & Shen (2003) and http://staff.chess.cornell.edu/~shen/Research.html#phase_problem for interesting developments], and we generally depend upon direct methods to find us some approximate values. A newly emerging method for solving the phase problem is charge-flipping (see §2.1). Less commonly these days, trial atomic positions can be extracted from a Patterson map and used to compute starting phases. Using these phases and E values or the observed amplitudes, we compute an E or electron density map. This is examined and the regions of high density are generally interpreted as atom sites. It is important to remember that the contoured electron density map is the closest representation we have to the continuous electron density in the crystal and is the best place to start searching for explanations if a parametric refinement fails to converge as expected.

This basic model can be the starting point for a number of refinement strategies.

2.1. Map modification

The actual values at every point in the computed electron density map can be modified so that they conform more closely with our knowledge of density distributions. Hoppe *et al.* (1970) suggest some possible modifications:

(1) Reduction of the enhanced maxima at known atomic sites.

(2) Enhancement of the reduced maxima at new atomic sites.

(3) Further reduction of the reduced maxima at wrongly placed atomic sites.

(4) Reduction of background and elimination of negative regions.

If the modifications are essentially valid, numerical integration of the modified map using equation (1) should yield improved phases, which can be reused with the original observed amplitudes to give an improved map. Strategies based on this technique continue to be widely used in macromolecular crystallography and in a much modified form have recently been reintroduced to small-molecule analyses under the name of charge-flipping (Oszlányi & Sütő, 2004). In this procedure, instead of negative regions of the map being simply set to zero, their sign is reversed (*i.e.* made positive) before new phases are computed. This procedure is often able to recover a complete structure from an initial random set of phases.

2.2. Map interpretation

This is the key stage in almost all small-molecule structure determinations. The map produced from the observed amplitudes and estimated phases is examined, usually by computer, for local maxima. The interpolated coordinates of these maxima (frequently called 'Q peaks'), together with their density, are tabulated for possible interpretation as atomic sites. Generally, the software applies space-group symmetry operators to the coordinates in order to try to bring the peaks close to each other – to assemble molecular fragments. In the simplest cases, these peaks are displayed as two-dimensional projections in text files [*e.g.* in *MULTAN87* (Debaerdemaeker *et al.*, 1987) or *SHELXS* (Sheldrick, 2008)] or three-dimensional models in suitable graphics programs, and the crystallographer is invited to assign an atomic type to each peak or to reject it. Other programs attempt to use the interpolated density and rules about chemical bonding to assign atomic types [*SIR92* (Altomare *et al.*, 1994) and *MULTAN87*].

This list of atomic types and three-dimensional coordinates is the basis of the structural model and is what is generally meant when people refer to 'a crystal structure'. There are a great wealth of programs available that can plot out these coordinates to generate diagrams of the structure. The replacement of the continuous electron density by a list of discrete points is a leap of faith which fortunately is generally justified. Electron density maps are usually computed at about $\frac{1}{4}$ Å resolution, so that the contours in the final two-dimensional plot shown in Fig. 3 will have been drawn through about 1400 data points. The corresponding 12 atoms can be represented by 48 coordinates (12 values of x, y, z and $U_{\rm iso}$), which in this case provide a good approximation to the underlying density.

Structure factors can be computed from the atomic model using equation (3) as an approximation to equation (1):

$$F_{\alpha,hkl} \simeq \sum_j^{\rm atoms} f_j \exp\left(-8\pi^2 s^2 U\right) \exp\left[2\pi i\left(hx_j + ky_j + lz_j\right)\right]. \quad (3)$$

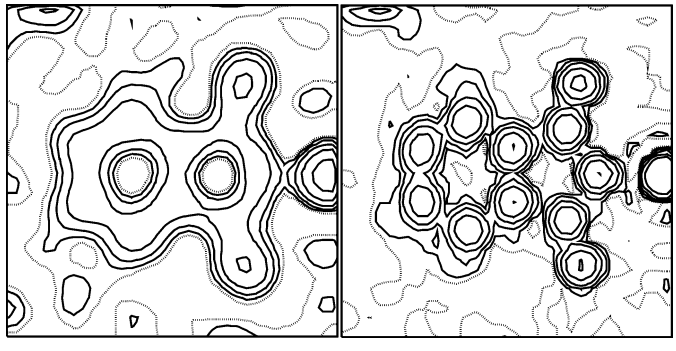

Figure 3
9 × 9 Å section of a Fourier map computed in the plane of a molecule of 2-dimethylsulfuranylideneindan-1,3-dione. Contours at 0, 1, 2.5, 5, 10 and 20 electrons Å^{-1}, computed at $\frac{1}{4}$ Å intervals. The left-hand figure uses all data out to 14° in θ; the right-hand figure uses all data out to 29° (Mo radiation).

The triple integration over the continuous electron density has been replaced by a summation over the atoms in the model. x, y and z are the atomic coordinates, f is the scattering factor, a term expressing the interaction of X-rays with electrons, s is $\sin(\theta)/\lambda$, and U is the ADP, a smearing function to represent time- and space-averaged displacements of the atoms from their equilibrium positions.

Phases computed from this atomic interpretation of the map can be used with the observed amplitudes to compute an improved map. This map may reveal new atoms that were missing from the original map. In addition, the peak interpolation of this new map may give atomic coordinates slightly different from those in the original model. These differences can be used to compute improved coordinates. For a centrosymmetric structure, the original model coordinates are simply replaced by new ones interpolated from the map [c = 1 in equation (4)]:

$$x_{\rm new} = x_{\rm old} + c\left(x_{\rm peak} - x_{\rm old}\right). \quad (4)$$

For noncentrosymmetric structures, experience shows that values of c up to 2 lead to improved coordinates (Cruickshank, 1950).

Automated Fourier refinement, sometimes called Fourier recycling, is an integral part of direct methods programs and is generally interspersed with cycles of least-squares coordinate refinement (see below) and tangent formula phase refinement. Each of these different techniques has a different sensitivity to errors in the model, and it is the skilful alternation between them that accounts for much of the success of modern structure determination programs.

3. Numerical optimization

The two previous methods cycle between computing phases and computing electron density maps and are highly successful in helping to establish the atomic skeleton of a structure. Once a model has been fully characterized, *i.e.* all the atom sites have been approximately located, the parameters describing these sites can be optimized directly.

The underlying aim of optimization is to propose a model that best predicts the observations. The X-ray data are clearly observations, but it is important to remember that other information, such as expected geometrical parameters (distances, angles), can also be regarded as observations. For organic and organometallic materials, we now have good estimates of these parameters. 'There is very little that can be added to the average intramolecular geometrical data collected by use of the Cambridge Structural Database; anything at variance with these well established averages is most probably wrong.' (Gavezzotti & Flack, 2005.)

A common-sense measure of the goodness of a prediction is one in which the residual ($Y_{\rm obs} - Y_{\rm calc}$) is small. $Y_{\rm obs}$ is the observation and $Y_{\rm calc}$ is the corresponding value computed from the model. If there are lots of observations, an overall goodness of prediction could be computed by summing the residuals. If the $Y_{\rm obs}$ values are evenly spread both greater and lower than $Y_{\rm calc}$, the sum should tend to zero. To obtain a measure of the spread of the residuals, G, we need to remove their signs before performing the summation. This can be achieved in two ways:

$$G = \sum^{\text{All observations}} \left|Y_{\rm obs} - Y_{\rm calc}\right| \quad (5)$$

or

$$G = \sum^{\text{All observations}} \left(Y_{\rm obs} - Y_{\rm calc}\right)^2. \quad (6)$$

These numbers will increase with the number of items summed, but can be normalized by dividing either by the number of observations (see §3.5.2) or by the sum of the observed values to give 'reliability indices', R. Applying the latter to equation (5) gives us

$$G1 = \sum^{\text{All observations}} \left|Y_{\rm obs} - Y_{\rm calc}\right| \Big/ \sum^{\text{All observations}} \left|Y_{\rm obs}\right|, \quad (7)$$

where $|Y_{\rm obs}|$ is the absolute magnitude of $Y_{\rm obs}$. If Y in equation (7) is replaced by the unphased structure magnitude F, we obtain the conventional crystallographic R factor (sometimes called $R1$ and sometimes multiplied by 100 to create a percentage). Note that $F_{\rm calc}$ is always positive; $F_{\rm obs}$ may be negative if the original net intensity was negative [see §3.2.2 point (1)]. $R1$, computed with weak data excluded [typically rejecting reflections for which $I < 2\sigma(I)$] is a rough and ready indication of the 'quality' of an analysis.

Dividing equation (6) by the sum of the squares of the observations and taking the square root gives us

$$G2 = \left[\sum^{\text{All observations}} \left(Y_{\rm obs} - Y_{\rm calc}\right)^2 \Big/ \sum \left(Y_{\rm obs}\right)^2\right]^{1/2}. \quad (8)$$

If Y in equation (8) is replaced by F^2, we obtain the crystallographic $R2$. Multiplying each term in the summations by a weighting function (see below) gives $wR2$. This is generally computed using all data, is generally much greater than $R1$ and, although more statistically sound than $R1$, is less commonly used as an informal indicator. Since the values of the weights can be modified by the program or the user, $wR2$ is susceptible to massaging.

In general, minimal values of $G1$ and $G2$ do not correspond to the same model – they correspond to different 'bests' unless appropriate weighting functions are used. Similar reliability indices can be defined for geometric observations, such as bond lengths, discussed below (§5). In this case $Y_{\rm obs}$ is a value obtained from the literature or a database, and $Y_{\rm calc}$ is from the current model.

3.1. Optimization criteria

The task of optimization procedures is to alter the values of the parameters in the model to minimize the discrepancies between $Y_{\rm obs}$ and $Y_{\rm calc}$. Possible minimization functions are

$$M1 = \sum^{\text{All observations}} \left|Y_{\text{obs}} - Y_{\text{calc}}\right|, \quad (9)$$

$$M2 = \sum^{\text{All observations}} (Y_{\text{obs}} - Y_{\text{calc}})^2, \quad (10)$$

$$M3 = -\log P(Y_{\text{obs}}, Y_{\text{calc}}). \quad (11)$$

The least modulus, $M1$ [equation (9)], has been unpopular for many years because of the difficulty in minimizing a function with a discontinuous first derivative, but its use seems to be gaining favour in other fields (Rousseeuw & Leroy, 2003). It is reported to be resistant to the influence of outliers in the data. Outliers (§3.5.3) are observations whose values are so far from those expected (assuming a usual error distribution) that they are probably subject to some gross experimental error.

In equation (11), P is the probability of making the observations Y_{obs} given the current model. This is the maximum likelihood function, now very popular in macromolecular crystallography, and can be formulated to take into account expected errors in the phases of reflections or unusual error distributions. If the error distribution of the observations is normal and uncorrelated and the model is fully parameterized, equation (11) reduces to equation (10). For most small-molecule structure analyses the data are copious, with errors approximating to normal, and the models do represent the actual structure. For this reason, small-molecule programs minimize a weighted variation of the function $M2$, the least-squares solution. It is important to understand that $M2$ only tells us what we are going to minimize, not how we are going to do it. Small-molecule crystallographers generally choose to find the solution by the use of the normal equations (see §3.6). Macromolecular crystallographers often choose the conjugate gradient method (which avoids the computationally costly expense of forming the normal equations). There is, however, a much wider range of optimization techniques available, and we can expect to see these being increasingly used in crystallography in the next few years. Already Monte Carlo and simulated annealing methods are being used in the early stages of refinement and the processing of powder data (Markvardsen *et al.*, 2005) because of their increased range of convergence and ability to escape from shallow false minima (false solutions).

3.2. The function to be minimized

So far we have not closely defined Y_{obs}, the X-ray or neutron observation. An IUCr committee some years ago failed to come to a unanimous decision as to what constituted the best definition (Schwarzenbach *et al.*, 1989). The contenders were

$$Y_{\text{obs}} = I_{\text{obs}}, \quad Y_{\text{obs}} = F_{\text{obs}} \quad \text{and} \quad Y_{\text{obs}} = F^2_{\text{obs}}.$$

3.2.1. $Y_{\text{obs}} = I_{\text{obs}}$, the actual observed intensity. The supposed merit in this definition is that it involves the minimal tinkering with the data. Because of this, the inverse of terms like the Lorentz, polarization and absorption corrections have to be applied to obtain I_{calc} from F_{calc}. Unless these corrections contain parameters that can be refined along with the structural parameters, refinement against I_{obs} is equivalent to refinement against F^2_{obs}. If the model is extended to include what are usually regarded as data processing parameters then refinement should be against I. One example of this is the inclusion of the crystal shape in the main structural refinement (Blanc *et al.*, 1991) – not to be confused with refining the crystal shape against the variation of the intensities of equivalent reflections (Herrendorf, 1993).

3.2.2. $Y_{\text{obs}} = F_{\text{obs}}$. This was the definition most widely used until the mid 1990s, leading to the minimization function equation (12), where w is a weight associated with each observation:

$$N = \sum^{\text{All observations}} w(F_{\text{obs}} - F_{\text{calc}})^2. \quad (12)$$

However, it is mathematically unsound for two reasons:

(1) Obtaining F_{obs} from I_{obs} involves taking the square root of the observed value. This is a nonlinear transformation, so that the error distribution becomes skewed, making it difficult to assign valid standard uncertainties to small, zero or negative F_{obs} values. Note that, when used for computing minimization functions, the sign of F_{obs} is the same[1] as the sign of I_{obs}: $F_{\text{obs}} = \text{sign}(I_{\text{obs}}) \times (F^2_{\text{obs}})^{1/2}$. The validity of standard uncertainties becomes important if they are used in the computation of weighting functions, w, for the refinement. Atkinson (1987) shows that nonlinear transformations of the data may improve the stability of a refinement and make it less susceptible to the effect of outliers, at the cost perhaps of compromising the precision of the model parameters. Edwards (1992) has shown that maximum likelihood optimization is unaffected by the F^2 to F transformation

(2) The derivative of the residual with respect to the structural parameters is not continuous if, for some reflection, F_{calc} has to pass through zero in order to reverse its phase:

$$\frac{\partial(F_{\text{obs}} - F_{\text{calc}})^2}{\partial x} = -2\sum(F_{\text{obs}} - F_{\text{calc}})\frac{\partial F_{\text{calc}}}{\partial x}. \quad (13)$$

If the observation is relatively intense, this means that a poor trial structure can become stuck in a 'false minimum' (Rollett *et al.*, 1976). Except for Rollett's synthetic data, there does not appear to be any published account of a refinement being trapped in a false minimum simply because the refinement was based on F. Some additional justification for refining on F, at least in the initial stages, is given by Atkinson (1987, ch. 6).

3.2.3. $Y_{\text{obs}} = F^2_{\text{obs}}$ This is the method [equation (14)] recommended by Rollett (1988) and is the only definition used in *SHELXL* (Sheldrick, 2008). Most other programs give the user the choice of definition 3.2.2 or 3.2.3.

$$M = \sum^{\text{All observations}} w\left(F^2_{\text{obs}} - F^2_{\text{calc}}\right)^2. \quad (14)$$

[1] This is a correct definition because the intensities are 'ratio' observations, that is, have a physically meaningful zero origin to their values.

The rate of change of this minimization function is given in equation (15). Note the appearance of the phaseless F_{calc} on the right-hand side, which ensures that the derivative is zero when F_{calc} is zero, thus reducing the risk of the refinement being trapped in false minimum.

$$\frac{\partial(F_{obs}^2 - F_{calc}^2)^2}{\partial x} = -4\sum\left(F_{obs}^2 - F_{calc}^2\right) F_{calc}\frac{\partial F_{calc}}{\partial x}. \qquad (15)$$

In most practical cases there is little to choose between the three definitions. One can speculate that the early popularity of definition (12) was a result of the difficulties with obtaining reliable estimates of the standard uncertainties (and hence relative weights) of photographically measured data. Taking the square root of the observations reduces the dynamic range of the data, so that the minimization function N [equation (12)] behaves quite well if all reflections are given the same weight, usually unit weights.[2] By contrast, M is very unstable with unit weights. Computing w in equation (14) by a suitable function, (16), gives quasi-unit weights:

$$w = 1/\left(4F_{obs}^2\right). \qquad (16)$$

This ensures that both M and N converge to the same model (Cruickshank, 1969), though there is a latent problem as F_{obs} tends to zero, which is usually dealt with by using *ad-hoc* rules (see §3.5.1). Use of equation (16) to convert M minimization to N minimization shows that, without proper weighting, M minimization puts most emphasis on the strong reflections.

3.3. Choice of function to be minimized

The choice of refinement against F^2 or F has generated more discussions than it probably warrants. In modern work, the large excess of observations over parameters makes false minima relatively uncommon, and when they occur they are usually evident as unexpected structural features (Murphy *et al.*, 1998). The definition of §3.2.3 (F^2), which minimizes the squares of squared values, is very susceptible to the influence of badly underestimated strong observations, for example, low-order reflections partially occluded by the beam trap. However, these should be detectable by other methods (*e.g.* a plot of F_{obs} *versus* F_{calc}) and should be eliminated from the data.

Robust–resistant weighting, which uses $\sigma(F^2)$ to detect outliers (Prince, 1994*a*), will mitigate the effect of more marginal data, though if F_{obs}^2 is seriously underestimated, it is very likely that its standard uncertainty will be seriously wrong as well. Prince has also shown that the standard uncertainties of the final parameters in an F refinement can be sensitive to large errors in the weak reflections. Attention to eliminating evident outliers and applying appropriate weights to the refinements leads to essentially indistinguishable final parameters from the two methods. Kassner *et al.* (1993), in a careful study, report 'we are not aware that anybody has yet shown that a refinement based on $|F|$ resulted in an incorrect crystal structure determination because it was based on $|F|$'. This is probably still the case. If the two processes lead to different results, the cause is almost certainly serious undetected errors in the data or totally inappropriate weighting schemes. Seiler *et al.* (1984) give a detailed description of the outcomes with real experimental data, and Harris & Moss (1992) revisit the issue from the point of view of macromolecular crystallography.

[2] If the reflection data follow Poisson statistics ($\sigma^2 = I$), then $\sigma^2(F)$ only depends upon the *Lp* correction. See §3.5.1.

3.4. Minimizing the weighted sum of the squares of the residuals – least squares

The expressions for least squares can be based on F or F^2. The equations are slightly simplified for F, and so these will be used in the rest of this chapter. Equivalent expressions for F^2 refinement are given by Cruickshank (1969). The expression for F_{calc} contains exponential and trigonometric terms – that is, it is a nonlinear function of the model parameters, x and U, as shown in equation (17):

$$F_{calc} = \sum_{i}^{atoms} f_i \exp(-2\pi^2 s^2 U^2)\exp(2\pi i\mathbf{h}\cdot\mathbf{x}). \qquad (17)$$

One technique for dealing with this expression is to use the Taylor expansion to linearize the function:

$$F'_{calc} = F_{calc} + \sum^{All\ parameters} \frac{\partial F_{calc}}{\partial x_i}\delta x_i + \cdots, \qquad (18)$$

where x_i are the refinable parameters.

F'_{calc} is the improved calculated structure factor obtained by applying the shifts δx. Setting F'_{calc} equal to F_{obs}, dropping the higher derivative terms and rearranging gives the observational equations

$$F_{obs} - F_{calc} = \sum^{All\ parameters} \frac{\partial F_{calc}}{\partial x}\delta x, \qquad (19)$$

where the residual $F_{obs} - F_{calc}$ plays the role of the simple observation in linear least squares and the derivatives play the roles of the independent parameters. To be able to use this expression, one must have a starting model in order to compute the residual and the derivative. If the Taylor expansion is applied to a linear function, the derivative is a constant no matter what values are used in the model (including setting them all to zero) so that correct shifts are computed directly. For a nonlinear function, the derivatives only become reliable as the parameters approach their 'true' (but initially unknown) values. This means that, for a very poor starting model, the technique may not converge onto the 'correct' solution. Both true and correct are in quotation marks because we can never know that we really have the true, correct solution. The normal equations derived from equation (19) contain only products of first derivatives (see §3.6). Alternative developments include higher derivatives, though these do not seem to be used in crystallographic applications. Press *et al.* (2002) discuss this issue and justify the exclusion of higher derivatives.

3.5. Relative uncertainties in the data, weights and data filters

In their raw form as given in equations (12) and (14), all reflections are added into the minimization process with equal importance if w is set to unity. As already explained, this is not always appropriate, and different weighting schemes have been developed for different purposes. In practical work, it is common to change the weighting scheme at different stages of an analysis in order to achieve different ends.

3.5.1. Statistical weights. Unit weights and quasi-unit weights were introduced above. With these weights, the least squares processes illustrated by equations (12) or (14) will give the same structure, though in fact this structure will not be the 'best'. This is because no allowance has been made for the different uncertainties associated with the observed intensities. To account for this, the functions represented by equations (9) and (10) must be multiplied by a weight (Hughes, 1941):

$$w = 1/\sigma^2(Y_{\rm obs}). \tag{20}$$

To a first approximation, the errors associated with a single intensity observation follow Poisson statistics, so that if

$$I_{\rm net} = {\rm Peak}_{\rm count} - {\rm Background}_{\rm count} \tag{21}$$

then

$$\sigma^2(I_{\rm net}) = {\rm Peak}_{\rm count} + {\rm Background}_{\rm count}. \tag{22}$$

This expression was used to obtain σ for early serial diffractometer data, but the extensive mathematical calculations used in processing image detector data require the use of more complex, proprietary, expressions. Since F^2 is obtained from I simply by multiplying by a scalar, there is no difficulty in obtaining $\sigma(F^2)$ from $\sigma(I)$. It might appear to be possible to obtain $\sigma(F)$ from $\sigma(F^2)$ using

$$\sigma^2(F) = \sigma^2\big(F^2\big)/4F^2, \tag{23}$$

but this is clearly inappropriate as F approaches zero, and other assumptions must be made about the tails of the probability density distributions. Possible assumptions are that $\sigma^2(F) = \sigma^2(F^2)$ if $F^2 < \sigma^2(F^2)$ (*CRYSTALS*; Betteridge *et al.*, 2003) or $\sigma^2(F) = [(F^2)^2 + \sigma^2(F^2)/2]^{1/2} - F^2$ [*XRAY*; quoted by Seiler *et al.* (1984)]. In the early stages of a refinement based on good data and a well parameterized model, these simple statistical weights lead to a steady convergence to a reliable minimum. However, they are not appropriate for the final stages.

3.5.2. Empirical weights and the goodness of fit. Although the correct procedure to be used for converting $\sigma(F^2)$ into $\sigma(F)$ is of some theoretical interest, in practice it turns out not to be so important. Counting statistics may be the major contribution to the variance, but there are other, usually unidentified, sources of error. In addition, the argument of the minimization is $(Y_{\rm obs} - Y_{\rm calc})^2$, which depends upon the current model. Since there are likely to be unidentified shortcomings in the model, a variance needs to be associated with $Y_{\rm calc}$. Rollett (1988) gives an extensive list of sources of errors in both $F_{\rm obs}$ and $F_{\rm calc}$.

Weights should be chosen to give reasonable estimates of the uncertainties in the parameter values and to reduce the effect, as far as possible, of unidentified, omitted parameters on those parameters included in the current model. For example, the weights needed to refine an isotropic model against a particular data set may be different from the weights needed for an anisotropic model. The reduced χ^2, also called the goodness of fit (GoF, S), gives some insight into potential problems with a refinement.

$$\chi^2 = \sum^{\rm All\ reflections} w(Y_{\rm obs} - Y_{\rm calc})^2, \tag{24}$$

$$S^2 = \sum^{\rm All\ reflections} w(Y_{\rm obs} - Y_{\rm calc})^2/(n-m). \tag{25}$$

We can define S_σ^2 as the GoF obtained by using weights computed only from the standard uncertainties assigned to the reflection data, equation (20) (Press *et al.*, 2002).

A low value of S_σ may indicate an over-pessimistic estimation of the variances of the observations, or that there are errors in the data or model which are compensating for each other. A large value of S_σ (greater than 2 or 3) may mean that the proposed solution is in fact a false solution or that there is some additional source of error in the data that has not been factored into the estimated uncertainties. As pointed out by Rollett (1988), fiddling with w to obtain a value of S close to unity is a risky business, and should only be attempted after the analyst is confident that the model is fundamentally correct (and in particular fully parameterized and not in a false minimum) and has assessed the X-ray data for additional sources of error. At this stage, simply multiplying all w by a constant to achieve S = 1.0 is nothing more than cosmetic (Huml, 1980). Rescaling the weights in this way does not affect the parameter values, so making published values of S (which are almost invariably close to unity) totally uninformative. Rescaling the weights does have an effect on the parameter s.u. values and can be thought of as a normalizing factor to compensate for deficiencies in the estimation of the errors on the X-ray observations.

More useful than looking at S itself is an examination of the change in the value of $\langle\chi^2\rangle$ [*i.e.* $\langle\sum w(Y_{\rm obs} - Y_{\rm calc})^2\rangle$] for groups of reflections selected in some systematic way. Common groupings are by $Y_{\rm obs}$, $Y_{\rm calc}$, resolution, reflection index or parity group. If the weights are appropriate (except for a scale factor) these local averages should have a constant value. Most programs provide a method for using a low-order polynomial to modify w to achieve roughly constant local values of χ^2, at the same time generally ensuring that S is approximately unity. The polynomial could be in terms of $Y_{\rm obs}$, $Y_{\rm calc}$ (which is less affected by outliers) or $[kY_{\rm obs} + (1-k)Y_{\rm calc}]$ (Wilson, 1976). If the model is fundamentally correct, local rescaling of the weights in this fashion reduces the effect of unidentified errors in the model parameters, though there is always the risk that selecting weights on the basis of the residual may give false credibility to an incorrect model. Additional global rescaling to make S close to unity produces

the best estimates of the parameter uncertainties in the absence of a detailed analysis of the sources of all the errors. In the absence of an overall scaling factor, weighting schemes based on

$$w = 1/\left\{[\sigma^2(F^2) + \text{function}(F)]\right\} \tag{26}$$

cannot compensate for pessimistic estimates of the reflection standard uncertainties if the function of F can only take positive values. Walker & Stuart (1983) and Parkin (2000) have used the variation of $F_{\rm obs}$ and $F_{\rm calc}$ as a function of diffractometer setting angles as a way of locating diffraction-geometry-dependant errors.

3.5.3. Robust–resistant weights. If all is well with the data, a large residual $(Y_{\rm obs} - Y_{\rm calc})^2$ means that $Y_{\rm calc}$ is in error, and the residual will drive the model parameters towards better values. However, if there is something wrong with the data and it is $Y_{\rm obs}$ that is in serious error, the large residual will adversely affect the model as the refinement tries to move $Y_{\rm calc}$ towards the bad observation.

The advent of area-detector diffractometers, which are generally programmed to measure each independent reflection many times at different instrumental settings [called high redundancy, or sometimes high multiplicity of observation (Müller *et al.*, 2006)], has greatly reduced the incidence of outliers in the data. Even so, badly measured reflections do sometimes occur in the final reflection listing. The most common occurrence is severe underestimation of a reflection because it is partially occluded by the beam stop or some other piece of the hardware. When this happens to a strong low-order reflection, direct methods yield a structure that may then disappear during either automatic or manual refinement. A low-order reflection that is underestimated will give rise to a very small E value. This will normally not be strongly involved in the phasing or tangent refinement of the data [unless the reflection was unfortunate enough to have been chosen for use in calculating a figure of merit involving negative quartets (De Titta *et al.*, 1975)]. The first E map may show a strong likeness to the expected structure. However, during refinement a seriously underestimated strong reflection will lead to a large (but erroneous) residual, which can cause the refinement to diverge. Such reflections are generally very obvious in a list of large residuals and can be eliminated manually. For more marginal cases, it is useful to employ a modification that will down-weight outliers to the user's favoured weighting scheme in a smoothly progressive way (Nicholson *et al.*, 1982). One weight modifier (w'), illustrated in Fig. 4, is

$$w' = \left\{1 - \left[w(Y_{\rm obs} - Y_{\rm calc})\right]^2/a^2\right\}^2, \tag{27}$$

when $|[w^{1/2}(Y_{\rm obs} - Y_{\rm calc})]| \leq a$ (where a is a user-defined parameter), otherwise

$$w' = 0.$$

Reflections with a weighted residual much less than a will have their normal weight, with the modified weight tending to zero as the residual approaches or exceeds a. The difficulty is to provide a value for a. One method is to use $a = k\sigma(Y_{\rm obs})$, but this suffers from the difficulty that a rogue reflection will probably have a rogue standard uncertainty. A more reliable estimate is to use k (a user-defined parameter) times the local average modulus of the weighted residual, $a = k\sum^n |w^{1/2}(Y_{\rm obs} - Y_{\rm calc})|/n$ (Rollett, 1988).

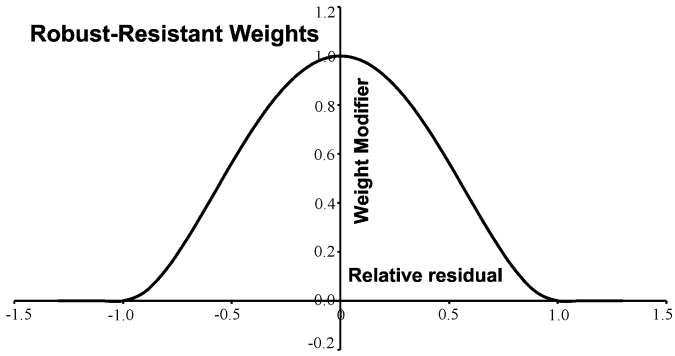

Figure 4
Robust–resistant weight modifier. Reflections with large residuals are down-weighted. In the Nicholson *et al.* (1982) example, a is set to $6\sigma(Y)$. The likelihood of a residual exceeding this value by chance is vanishingly small, so that a discrepancy of this magnitude leads to a weight of zero.

3.5.4. Weighting as a function of resolution. The semi-empirical weights just described are appropriate during the final cycles of refinement of a completed model. At earlier stages in a refinement, other schemes may be useful for other purposes, some of which are described below.

Qurashi & Vand (1953) suggested using a weighting scheme that *decreases* the influence of high-angle reflections, of the form

$$w = 1.0/s^3 \tag{28}$$

[where $s = \sin(\theta)/\lambda$ is the resolution of the reflection], as a means of increasing the radius of convergence of the least-squares method. This is quite strongly related to the procedure in which only low-angle data are used in the initial refinements, and data of increasing resolution are incrementally added to the refinement as the model improves [and which, since high-angle data tend to be weak, is not totally unrelated to the use of an $I/\sigma(I)$ threshold for rejecting data].

Dunitz & Seiler (1973) suggest using a scheme that increases the influence of the high-angle data:

$$w = \exp\left\{c\left[\sin(\theta)/\lambda\right]^2\right\}, \tag{29}$$

where c is a user-defined variable parameter. Since the influence of scattering from H atoms is small at high angles, this weighting scheme reduces the tendency for the heavy-atom parameters to be adjusted by the least squares in a way that partially simulates the scattering from H atoms missing from the model. The resulting atomic parameters are similar to those obtained from a neutron diffraction experiment. Phases for the full data set refined in this way have an increased chance of revealing missing H atoms in a difference Fourier map.

3.6. Solving the observational equations

The unweighted observational equation (19) for each reflection can be written

$$\frac{\partial F_{\rm calc}}{\partial x_1}\delta x_1 + \frac{\partial F_{\rm calc}}{\partial x_2}\delta x_2 + \cdots + \frac{\partial F_{\rm calc}}{\partial x_m}\delta x_m + \cdots = F_{\rm obs} - F_{\rm calc}, \tag{30}$$

where x_1, x_2 *etc.* refer to the refinable parameters, and higher-order terms have been omitted from the expansion. There may be several hundred parameters and typically up to ten times as many reflections. It is this massive over-determination of the parameters by the data that makes X-ray structure analysis so reliable in the absence of serious systematic errors. Given a reasonable starting model, the large number of simultaneous equations can be solved by least squares. Rewriting equation (30) in matrix form for m parameters and n observations gives

$$\begin{bmatrix} \frac{\partial F_{1_{\rm calc}}}{\partial x_1} & \cdot & \frac{\partial F_{1_{\rm calc}}}{\partial x_m} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \frac{\partial F_{n_{\rm calc}}}{\partial x_1} & \cdot & \frac{\partial F_{n_{\rm calc}}}{\partial x_m} \end{bmatrix} \cdot \begin{bmatrix} \delta x_1 \\ \cdot \\ \cdot \\ \cdot \\ \delta x_m \end{bmatrix} = \begin{bmatrix} F_{1_{\rm obs}} - F_{1_{\rm calc}} \\ \cdot \\ \cdot \\ \cdot \\ F_{n_{\rm obs}} - F_{n_{\rm calc}} \end{bmatrix}. \tag{31}$$

The large matrix of derivatives is called the 'design' matrix, because in linear least squares the researcher usually has the opportunity to design an experimental procedure that will achieve optimal coverage of the observational space. In crystallography, the elements of the design matrix are computed from the current model.

As explained in §3.5, each equation in (31) must be weighted by our confidence both in the quality of the observation and in the model. In many experiments it seems likely that errors associated with one observation will be correlated with errors in other observations – this is very likely to be true of adjacent observations in powder data and is probably true of the data from a single frame of an area detector. These correlated weights, if they were known, would be put into a weight matrix, **W**:

$$\begin{bmatrix} w_{11} & w_{12} & \cdot & w_{1n} \\ w_{21} & w_{22} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ w_{n1} & \cdot & \cdot & w_{nn} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial F_{1_{\rm calc}}}{\partial x_1} & \cdot & \frac{\partial F_{1_{\rm calc}}}{\partial x_m} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \frac{\partial F_{n_{\rm calc}}}{\partial x_1} & \cdot & \frac{\partial F_{n_{\rm calc}}}{\partial x_m} \end{bmatrix} \cdot \begin{bmatrix} \delta x_1 \\ \cdot \\ \cdot \\ \cdot \\ \delta x_m \end{bmatrix}$$
$$= \begin{bmatrix} w_{11} & w_{12} & \cdot & w_{1n} \\ w_{21} & w_{22} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ w_{n1} & \cdot & \cdot & w_{nn} \end{bmatrix} \cdot \begin{bmatrix} F_{1_{\rm obs}} - F_{1_{\rm calc}} \\ \cdot \\ \cdot \\ \cdot \\ F_{n_{\rm obs}} - F_{n_{\rm calc}} \end{bmatrix}. \tag{32}$$

For most routine work only the diagonal elements of **W** are used, having values as described in §3.5, though for interesting developments using correlated errors, see McCusker *et al.* (2001).

Equation (32) can be written more succinctly as

$$\mathbf{W}\cdot\mathbf{A}\cdot\delta\mathbf{x} = \mathbf{W}\cdot\Delta\mathbf{F}. \tag{33}$$

If there are more observations than unknowns, the method of least squares provides values for both the unknowns and their standard uncertainties. Note that the matrix of derivatives, **A**, does not appear anywhere in the minimization functions N or M. This has the curious implication that the structure obtained at the minimum is independent of the method used to obtain **A**, leaving the crystallographer with a choice of treatments of **A**. A solution *via* the use of the normal equations is

$$\mathbf{A}^{\rm t}\cdot\mathbf{W}\cdot\mathbf{A}\cdot\delta\mathbf{x} = \mathbf{A}^{\rm t}\cdot\mathbf{W}\cdot\Delta\mathbf{F} \tag{34}$$

and

$$\delta\mathbf{x} = (\mathbf{A}^{\rm t}\cdot\mathbf{W}\cdot\mathbf{A})^{-1}\cdot\mathbf{A}^{\rm t}\cdot\mathbf{W}\cdot\Delta\mathbf{F}, \tag{35}$$

where the superscript t denotes the transpose. The normal matrix, $\mathbf{A}^{\rm t}\cdot\mathbf{W}\cdot\mathbf{A}$ is symmetric and usually inverted by the Cholesky method (Rollett, 1964). This is a very robust method. If the matrix is singular or near-singular, indicating that a parameter or parameter combination cannot be determined by the data, the appropriate elements on the diagonal of the inverse matrix are set to zero (not infinity) so that the shifts $\delta\mathbf{x}$ of the singular parameters from equation (35) are also zero, *i.e.* that parameter is unchanged. The user of the program must be informed of this action so that the singularity can be eliminated by changing the model, adding constraints or adding restraints (see below).

Equation (33) can be solved by other processes, for example by conjugate gradient methods, but these are not frequently used in small-molecule refinements. An old but excellent discussion of possible methods is presented by Sparks (1961), and a good modern survey of methods for forming and solving these equations is given by Tronrud (2004), who shows that different approximations to **A** lead to differing speeds or radii of convergence.

3.7. Problems with the normal equations

If the normal matrix is computed from the first derivatives only, the diagonal is formed from the sum-of-squares and so is always nonnegative. The off-diagonal terms are the sums-of-products of terms, generally with differing signs, so that the totals are usually small. Solution of the equations with this diagonal-dominated matrix is generally straightforward. The off-diagonal elements can become larger for a variety of reasons (which include missed or pseudo-symmetry, or over-harsh geometric restraints). When this happens, parameters linked by these large elements are said to be highly correlated, which may lead to difficulties in solving the equations.

3.7.1. Space-group-related issues. Failure to take proper account of atoms on special positions, or to fix the origin in polar directions (see §5.1.8), will lead to a singular normal matrix (*i.e.* the determinant is very close to zero, so that the matrix cannot be inverted safely). Both of these situations are predictable from knowledge of the space group and the approximate atomic coordinates and so are now generally treated automatically by the software.

3.7.2. Atomic positional disorder. Hughes (1941) showed that for well resolved atoms the possibility of correlation between atom parameters is decreased as the distance between the atoms increases. This means that there is always correlation between bonded atoms, but it is generally not excessive. Correlation leading to large off-diagonal terms and hence instability in the refinement becomes an issue for structures containing disorder. As a rule of thumb, if the disordered atoms can be seen as discrete local peaks in a Fourier map phased by the un-disordered part of the structure, refinement should proceed smoothly (Fig. 5).

3.7.3. Pseudo-symmetry. Curiously, it is not uncommon for structures to be found in which there is extra approximate symmetry in addition to the true space-group symmetry. Sometimes this pseudo-symmetry is very local, such as a pseudo-glide plane perpendicular to a non-unique axis in a monoclinic cell [Fig. 6; Cambridge Structural Database (CSD; Allen, 2002) refcode SAPVAG; Zimmerman & St Claire, 1989]. This kind of problem rarely causes difficulty with refinements.

When the pseudo-symmetry affects the whole structure, such as a pseudo-centre of symmetry in a chirally pure compound, pseudo-centring due to a heavy atom lying near to a special position of a higher-symmetry space group or pseudo-doubling of a cell axis, the effect on the refinement can be much more serious. Older programs would often complain that the normal matrix was singular and then abort. Modern programs, using techniques like the Marquardt modifier (§5.1.9), tend to execute to completion, but the results are unreliable (Harlow, 1996).

3.7.4. Very poor starting model. If the initial estimates of the positional parameters are far from the 'correct' positions (> 0.4 Å), refinement including the ADPs can be very unstable, leading to massive false shifts in some ADPs or the overall scale factor. Programs generally detect this situation and just apply some small fraction of the predicted shift. If the starting model has errors of the order of 0.7 Å, the trial atom could be roughly equidistant from a number of potential atom sites, unless it has fallen into the void, created by van der Waals repulsions, surrounding molecules.

The centrosymmetric/pseudo-centrosymmetric case is a trap for the unwary (Marsh & Spek, 2001). There are sets of space groups that cannot be distinguished from each other by the systematic absences alone (*e.g.* *P*2, *Pm*, *P*2/*m*). It is not unknown for direct methods of structure solution to fail with the centrosymmetric space group but then yield a structure if the space-group symmetry is changed to remove the centre.[3] These structures generally appear to refine satisfactorily in the low-symmetry space group to a low *R* factor, but close examination may show unusual geometric features. Bond lengths and angles deviate more than expected from conventional values, and anisotropic ADPs are exceedingly anisotropic. Closer examination may show that the means of geometric features related by the pseudo-centre are close to conventional values, and the averages of ADPs are much less anisotropic. These are clear signs that the analyst should investigate translating the structure so that the local centre of symmetry becomes a true centre, changing the space group and continuing the refinement. If the molecular geometry becomes decidedly more conventional after further refinement then this is probably the true space group. Flack & Bernardinelli (2006) discuss the opposite situation, in which structures are reported to be in centrosymmetric space groups when there should be no centre of symmetry. In difficult cases the role of the weak reflections becomes important (see §3.9).

Fig. 7 (CSD refcode QEQRUZ; Vigante *et al.*, 2000) is an example in which the space group must be acentric because the material is chiral, but which has a strong pseudo-centre of inversion relating two independent molecules in the asymmetric unit. The refinement does not proceed smoothly, as revealed by the published bond lengths and angles. The large deviations from expected values are probably a consequence

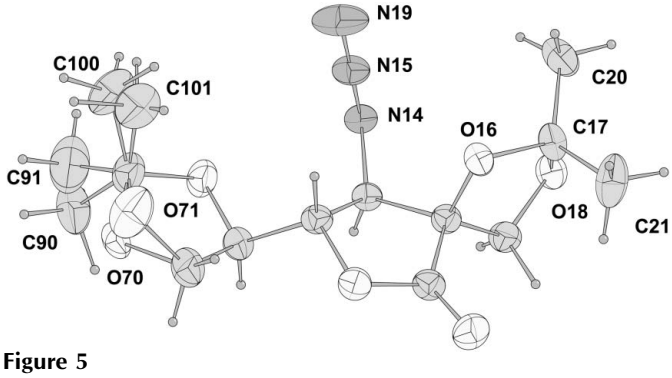

Figure 5
Disorder. The right-hand isopropylidene protecting group (containing O16 and C21) is reasonably well defined. The left-hand moiety refined to very elongated ADPs. These were replaced by partial atoms corresponding to ring-flipping. It was unclear from the actual electron density distribution which of these models was really most appropriate (Harding *et al.*, 2005).

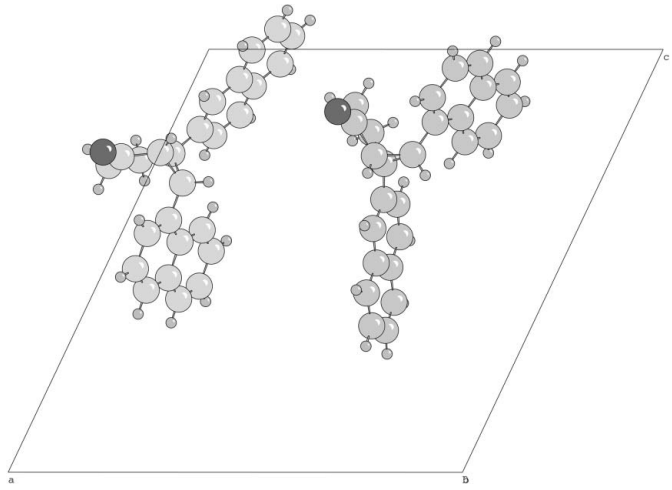

Figure 6
This chiral molecule is in space group $P2_1$ but has an exceedingly good pseudo-glide plane.

[3] This technique may be useful even when there is no ambiguity about the space group. A structure failing to solve in *P*21/*c* or *C*2/*c* may solve easily in *P*21 or *C*2. In this case the missing centre should be added to the structure and the redundant atoms removed.

Table 1
Bond lengths (Å) for each of the two phenyl groups in a pseudo-centrosymmetric material with $Z' = 2$.

Although the spread of values is very large, the overall mean is quite normal. This is a characteristic feature of high correlation in the normal matrix.

Ring 1	Bond length		Ring 2	Bond length	
Atoms	Molecule 1	Molecule 2	Atoms	Molecule 1	Molecule 2
9–10	1.4034	1.3993	15–16	1.2912	1.6016
10–11	1.7301	1.1034	16–17	1.3887	1.4470
11–12	1.3216	1.2308	17–18	1.5639	1.0727
12–13	1.3242	1.4158	18–19	1.3814	1.3423
13–14	1.3917	1.3783	19–20	1.4245	1.4201
14–9	1.6112	1.1427	20–15	1.5164	1.2996
Average	1.4637	1.2784	Average	1.4227	1.3639
Overall Average					1.3834

of high correlation in the normal matrix owing to the pseudo-centre of symmetry (Table 1).

3.8. Eigenvalue filtering

The normal matrix is generally inverted by the Cholesky process, which will correctly handle some kinds of singularities, such as an attempt to refine an unrefinable parameter for an atom on a special position. It cannot handle more complex situations leading to singularities or near-singularities, such as refining a centrosymmetric structure in a noncentrosymmetric space group. In this situation, pairs of parameters should have symmetry-related values, and the shifts in refinement should be equal and opposite. The Cholesky invertor will apply some kind of shift to one member of the pair and no shift to the other. Thus, although the program will execute to completion, the refined structure will be wrong. Eigenvalue filtering is a technique that analyses the full normal matrix for latent singularities and generates parameter shifts that have the correct symmetry relationships between them. The centro/noncentrosymmetric case should be dealt with correctly using this filtering, though it is best resolved by using the correct space group, when this is known. In practice, eigenvalue filtering is most useful as a tool for trying to understand the source of singularities and near-singularities (Watkin, 1994). In macromolecular crystallography, it is used as a method for detecting noncrystallographic symmetry and also allows the detection of singular parameters or parameter combinations that cannot be determined by direct examination of the X-ray data (Cowtan & Ten Eyck, 2000).

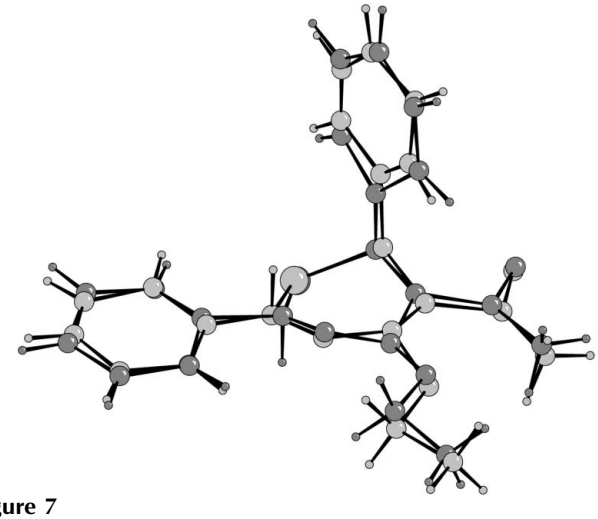

Figure 7
The two independent molecules in *QEQRUZ* (Vigante *et al.*, 2000), overlaid with the best match using *CRYSTALS*. Although the molecules are chiral in $P2_1$, the operator giving the best match between the molecules is a pseudo-centre of symmetry.

3.9. Weak reflections

There is no doubt that, with modern area-detector diffractometers and effectively infinite data storage facilities, all reflections up to some predetermined resolution limit should be estimated and recorded in the primary data file. What weak data should be used for is more problematic. There is now copious evidence that they should be included in data sets used for structure solution by direct methods, since they play an important role in the determination of the scale of the data from the Wilson plot and in the estimation of figures of merit involving negative quartets (Altomare *et al.*, 1995).

A debate about the more general treatment of weak reflections in refinement has been continuing since Hirshfeld & Rabinovich (1973) demonstrated that, for a synthetic data set, the omission of reflections whose observed intensity is less than some threshold standard deviation had a detectable effect upon the ADPs and the overall scale factor. Their findings were supported using real data by Arnberg *et al.* (1979). Seiler *et al.* (1984) reinvestigated the problem and showed that, for data measured with a serial diffractometer, the effect of the weak reflections upon the positional parameters is negligible, and that on the ADPs is only marginal. More recently, Harris & Moss (1992) could see no strong evidence for an improvement in macromolecular refinements by the inclusion of the very weak data in the refinement. Hirshfeld & Rabinovich (1973) concluded by commenting 'our limited experience indicates that in real situations the effect of biased data on the structurally interesting parameters is rarely large enough to matter', but suggesting that 'for safety's sake' one should include all reflections. The problem is to decide what is meant by 'all'.

There are broadly two kinds of weak reflections – those that contain useful information and those that do not. Imagine a data set for a simple organic material measured with Mo $K\alpha$ radiation from a monoclinic crystal with a 10 Å unique b axis. If the $0k0$ reflections with $k = 21, 23, 25$ are weak, these tell us almost nothing, because we can expect the $k = 20, 22, 24$ reflections to be weak also. In contrast, if the $k = 1, 3, 5, 7$ reflections are all weak, we have strong evidence for a 2_1 axis (Glusker *et al.*, 1994). There is clearly no special advantage in including the weak 0 22 0 reflection in a refinement, but if in a particular case the 020 reflection, normally expected to be very strong, is instead very weak, its weakness will be highly

significant. Prince & Nicholson (1985) have shown that, provided reflections are not excluded on the basis of their residual, exclusion of individual reflections should not bias the outcome. That is not the same as saying that omitting a reflection will have no effect. The same authors have shown that some parameters can be strongly influenced by some reflections, that is to say, inclusion of those reflections will have an influence on the parameter standard uncertainties. These significant reflections may be either strong or weak. Omitting reflections on the basis of $I/\sigma(I)$ is insecure because this ratio is indirectly correlated to the residual, since slightly overestimated reflections are selected in preference to underestimated ones (Seiler *et al.*, 1984). This can often be observed for batches of weak reflections, where $\sum I_{\rm obs}$ is generally larger than $\sum I_{\rm calc}$ if an $I/\sigma(I)$ threshold has been used.

It seems that the best strategy for filtering out low-information reflections is to set a threshold in terms of resolution. Schwarzenbach *et al.* (1989) assert that this strategy will effectively discriminate against weak data without introducing bias. As Weiss (2001) points out, for macromolecular crystallographers setting a resolution threshold is largely subjective. Some small-molecule analysts simply accept the arbitrary recommendations in the IUCr guide for authors (*i.e.* $\theta_{\rm max} > 25^\circ$ for Mo $K\alpha$, $\theta_{\rm max} > 67^\circ$ for Cu $K\alpha$). This may lead them to include large numbers of worthless data from a weakly diffracting crystal or omit useful data from a strong diffracter. It is worth recalling the *Parable of the Emperor of China* (Herbstein, 2000), which can be paraphrased as saying that a lot of bad measurements cannot yield a good result.

The advantageous refinement apparently offered by the use of resolution thresholding in order to retain the accidentally weak low-angle data may be illusory. Analysis of the systematic absences from a structure with a centred lattice generally shows a significant number to be observed well above background. This may be for physical reasons [*e.g.* thermal diffuse scattering (Cooper & Rouse, 1968), $\lambda/2$ contributions (Kirschbaum *et al.*, 1997)] or because the peak-search methods being used to integrate reflections are over-enthusiastic. In contrast, Lenstra & Kataeva (2001) describe the opposite situation, in which high-angle data are underestimated because of the method of integration. Fig. 8 illustrates the distribution of the intensity data for the systematic absences from a *C*-centred cell measured from the same crystal on two different diffractometers. The instrument in the lower diagram found many more negative intensities than positive ones – a systematic bias in the weak data – possibly because of inappropriate data collection or processing procedures.

As with the F/F^2 debate, the effects of thresholding only have an impact in marginal cases, in which situation the diffraction data will need careful scrutiny anyway. Perhaps the most important of these marginal cases is the one where there are difficulties in choosing between a centrosymmetric and a noncentrosymmetric space group. Dunitz (1995) and others (*e.g.* Schomaker & Marsh, 1979; Kassner *et al.*, 1993) have shown that it is the weak, low-order, reflections that offer the best chance of resolving the situation. Fig. 9 illustrates that it is the weak reflections (bottom diagram) that show the largest relative changes in structure amplitude when symmetry is not exact.

Crystals suffering from pseudo-translational symmetry (including pseudo-centring due to heavy atoms lying on special positions) can have whole classes of reflections systematically weak (*e.g.* the $h + k$ odd reflections for pseudo-*C*-centring). In his careful analysis of the diffraction data from quartz, Zachariasen (1965) noted that the largest fractional Friedel differences $[2(I_h - I_{\bar{h}})/(I_h + I_{\bar{h}})]$ occurred amongst the weak reflections. This is generally the case when there are no strong anomalous scatterers. Bernardinelli & Flack (1985) describe a weighting scheme that will enhance the influence of reflections with the largest anomalous differences. However, before using this scheme the analyst should be convinced that the weak reflections have not been systematically overestimated.

3.10. Shift multipliers and partial-shift damping factors

For a well behaved refinement in which the normal matrix contains contributions from the second derivative, the shift in each cycle will be the square of the shift in the previous cycle (called quadratic convergence) (Rollett, 1984). In most current crystallographic programs, the second derivatives are neglected because of their destabilizing effect (reduced radius of convergence), which has the effect of making the shifts in

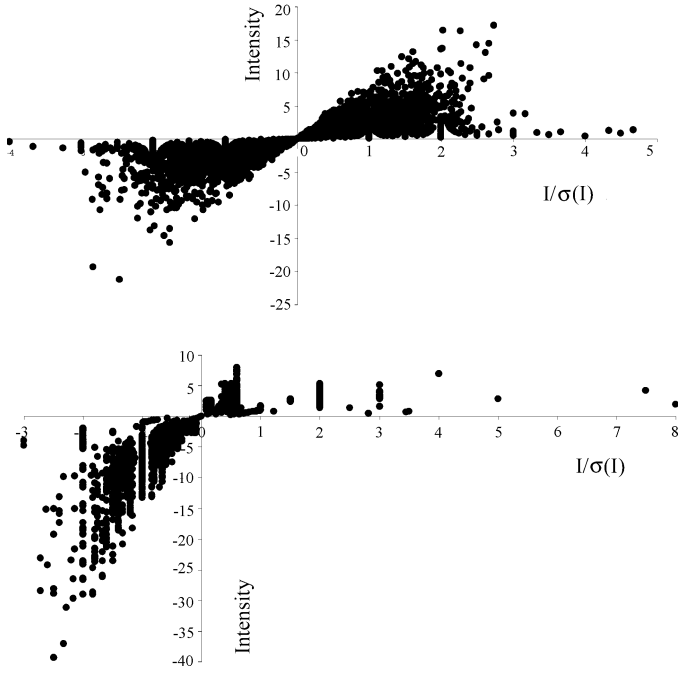

Figure 8
Distribution of intensity data for the systematic absences from a *C*-centred cell measured from the same crystal on two instruments. Ordinate: *I*; abscissa: $I/\sigma(I)$. Top illustration: A well adjusted instrument. Lower illustration: Underestimated weak reflections. There are very many more negative reflections than positive ones; the data-collection and processing parameters should be checked.

each cycle an approximately constant fraction of the shift in the previous cycle, *i.e.* more cycles are needed to reach convergence. Sparks (1961) and later Hodgson & Rollett (1963) showed that, in suitable cases, these shifts can be inflated to accelerate convergence. This idea does not seem to have been automated in any widely used programs. However, the converse operation, deflating the shifts if they seem improbably large, is common to most programs [equation (36)]. Note that these factors are applied after the solution of the normal equations and so can take no account of parameter correlations.

$$x_{\rm new} = x_{\rm old} + p\delta x, \tag{36}$$

where p is the damping factor, usually less than or equal to unity (Lipson & Cochran, 1966). In the event that a predicted shift/s.u. exceeds some threshold, some or all of the shifts are rescaled so that the maximum lies at this threshold. Because a near-singular matrix can lead to massive predicted shifts and s.u. values, most programs also set a limit on the maximum shifts that can be applied to parameters. Shift-limiting restraints (§5.1.8) modify the matrix before inversion, generally leading to better conditioning and hence stability in the refinement.

4. Constraints

Constraints are rules about the permitted values of parameters or of the relationships between them that, in a mathematical sense, must not be disobeyed. Loosely, they are of two types – those implicit in the overall formulation of the model and those explicitly imposed. For example, if a model is refined with isotropic ADPs, a constraint that the ADP is spherical has been imposed. Not refining an extinction correction implies the constraint that its value is zero, exactly. The use of spherical form factors is another implied constraint. One of the roles of the analyst is to judge what level of complexity can be introduced into a model given the available data. Amongst the explicitly imposed constraints are those required to conserve the space-group symmetry requirements. An atom on the special position $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ in $P\bar{1}$ should not be refined at all – its parameters are *constrained* to these special values. Alternatively, an atom on the special position (x, x, x) in Wyckoff position e in space-group $P23$ can be refined but must be constrained to have equal shifts applied in the **x**, **y** and **z** directions. In general, constraints are rules about the relationships between parameters that must be obeyed. Constraints arising from space-group symmetry relationships are usually generated automatically by modern programs (Burzlaff *et al.*, 1978), but other constraints may be imposed by the analyst. Many textbooks advocate applying constraints *via* the method of Lagrange undetermined multipliers (Rollett, 1970), but this method seems to be largely ignored in general-purpose crystallographic programs. Larson (1980) worked through an alternative strategy, which reparameterizes the problem *via* a matrix of constraint.

4.1. The matrix of constraint

In the application of constraints, the physical parameters used in the normal representation of the model (**x**) are related by an expression to a smaller set of parameters, which will actually be refined (**x**′):

$$\mathbf{x} = \mathbf{M}'\mathbf{x}' + \mathbf{c}, \tag{37}$$

where **c** is a vector of constants, so that

$$\delta\mathbf{x} = \mathbf{M} \cdot \delta\mathbf{x}', \tag{38}$$

where **M** is the matrix of derivatives of **M**′ with respect to the parameters **x**′. Substituting (37) into (32) gives

$$\begin{bmatrix} w_{11} & w_{12} & \cdot & w_{1n} \\ w_{21} & w_{22} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ w_{n1} & \cdot & \cdot & w_{nn} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial F_{1_{\rm calc}}}{\partial x_1} & \cdot & \frac{\partial F_{1_{\rm calc}}}{\partial x_m} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \frac{\partial F_{n_{\rm calc}}}{\partial x_1} & \cdot & \frac{\partial F_{n_{\rm calc}}}{\partial x_m} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial x_1}{\partial x'_1} & \cdot & \frac{\partial x_1}{\partial x'_q} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \frac{\partial x_m}{\partial x'_1} & \cdot & \frac{\partial x_m}{\partial x'_q} \end{bmatrix}$$
$$\cdot \begin{bmatrix} \delta x'_1 \\ \cdot \\ \cdot \\ \cdot \\ \delta x'_q \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdot & w_{1n} \\ w_{21} & w_{22} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ w_{n1} & \cdot & \cdot & w_{nn} \end{bmatrix} \cdot \begin{bmatrix} F_{1_{\rm obs}} - F_{1_{\rm calc}} \\ \cdot \\ \cdot \\ \cdot \\ F_{n_{\rm obs}} - F_{n_{\rm calc}} \end{bmatrix} \tag{39}$$

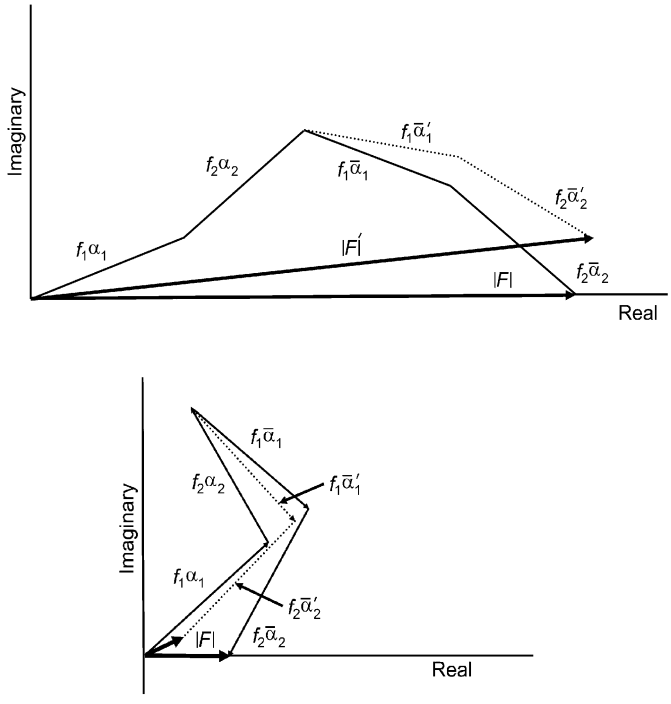

Figure 9
A strong (top) and weak (bottom) reflection at the same Bragg angle, with four equal atoms in the cell. The solid lines are the contributions to $|F|$ in the centrosymmetric case. In the pseudo-centric case, the phase angles for the contributions from the second two atoms are slightly shifted, so that the phase angle of the final structure factor is not zero. The resulting amplitude, $|F'|$, has a larger relative change for the weak reflection.

and the corresponding normal equations

$$\delta\mathbf{x}' = (\mathbf{M}^{\rm t}\mathbf{A}^{\rm t}\cdot\mathbf{W}\mathbf{A}\cdot\mathbf{M})^{-1}\cdot(\mathbf{M}^{\rm t}\mathbf{A}^{\rm t})\cdot\mathbf{W}\Delta\mathbf{F}. \quad (40)$$

The important thing to notice about equation (39) is that the number of least-squares variables ($\mathbf{x}'$) is less than the number of real physical variables ($\mathbf{x}$). The extra knowledge needed to expand $\mathbf{x}'$ to $\mathbf{x}$ is contained in the matrix of constraint, $\mathbf{M}$. This matrix is also needed for the computation of the variance–covariance matrix of the parameters $\mathbf{x}$ from the variance–covariance matrix of the parameters $\mathbf{x}'$. Without a knowledge of $\mathbf{M}$, it is not possible to compute proper standard uncertainties of quantities computed from $\mathbf{x}$, for example, bond lengths (which is why *checkCIF* sometimes returns s.u. values rather different from those computed by the refinement program).

Some examples of matrices of constraint are listed below.

4.1.1. Special positions. If $\mathbf{x}'$ is the least-squares parameter, the matrix needed for Wyckoff position *e* in *P*23 is

$$\begin{bmatrix}\partial x\\ \partial y\\ \partial z\end{bmatrix} = \begin{bmatrix}1\\ 1\\ 1\end{bmatrix}\cdot\partial\mathbf{x}'. \quad (41)$$

4.1.2. Site occupation factors (SOFs) and disorder. It is not uncommon in minerals to find an atomic site that can be occupied by either of two element types. Different asymmetric units can contain one or other of the elements, but the average occupancy of the site, totalled over the whole sample, should come to one (unless there are some asymmetric units in which the site is vacant). If the refinement starts with the two elements distributed evenly on the site then the SOFs will both be 0.5. During refinement, if the occupancy of one element (occ_1) rises, that of the other (occ_2) must decrease by the same amount to keep the total unity. If occ′ is the least-squares parameter, the matrix needed is

$$\begin{bmatrix}\partial\mathrm{occ}_1\\ \partial\mathrm{occ}_2\end{bmatrix} = \begin{bmatrix}1\\ -1\end{bmatrix}\cdot\partial\mathrm{occ}'. \quad (42)$$

If more than two elements are involved, the matrix of constraint takes the form

$$\begin{bmatrix}\partial\mathrm{occ}_1\\ \partial\mathrm{occ}_2\\ \partial\mathrm{occ}_3\\ \partial\mathrm{occ}_4\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\\ -1 & -1 & -1\end{bmatrix}\cdot\begin{bmatrix}\partial\mathrm{occ}'_1\\ \partial\mathrm{occ}'_2\\ \partial\mathrm{occ}'_3\end{bmatrix}. \quad (43)$$

In this example, four physical occupation factors are replaced by three least-squares parameters such that the sum of the shifts of the physical parameters is zero.

A similar situation can occur in organic materials where an atom or group of atoms can occupy either of two sites in the asymmetric unit. For a group of (say) four atoms occupying either a site *a* or a site *b*, all the atoms on site *a* will have the same SOF (occ_a), and all the atoms on site *b* will have $\mathrm{occ}_b = (1 - \mathrm{occ}_a)$. There are thus eight physical SOFs but only one least-squares parameter.

4.1.3. Riding models. The term 'riding model' is used rather loosely in the literature, but one reasonably clear-cut definition is that the analyst wants one set of physical parameters to move synchronously with another set. Most commonly, this technique is used to move H atoms synchronously with the C atoms to which they are bonded, thereby preserving the bond length and direction. The crude way to achieve this is to compute shifts for the C atoms and then apply the same shifts to the H atoms. However, if the matrix of constraint has been properly encoded into the program, this provides a more rigorous mechanism. Consider a CH group (Fig. 10).

For atoms C1 and H1 there are six positional parameters, but if the H and C atoms ride together, there are only three least-squares parameters – the shifts *x*, *y* and *z* of both atoms. The matrix of constraint is

$$\begin{bmatrix}\partial C_x\\ \partial C_y\\ \partial C_z\\ \partial H_x\\ \partial H_y\\ \partial H_z\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\cdot\begin{bmatrix}\partial x'\\ \partial y'\\ \partial z'\end{bmatrix}. \quad (44)$$

The advantage of this method over the crude one is that the terms in the normal matrix contain contributions arising from the scattering from both the C and the H atoms. The matrix of constraint has implemented the 'chain rule' for differentiation. While a riding constraint preserves the relative dispositions of the 'ridden' atoms, the geometry with respect to other atoms becomes distorted – in this example particularly the angles made from H1 to other atoms joined to C1. This may require the geometry to be re-regularized after a cycle of refinement.

In a general least-squares program the user should be able to make any parameter 'ride' on any other, so that, for example, whole groups of positional parameters or ADPs can be made to shift synchronously; such an approach might be useful during the early stages of a difficult refinement. If groups of parameters can be made to ride, but some of the shifts are inverted before being applied, one has a mechanism for applying noncrystallographic symmetry constraints. This provides a mechanism for refining part of a structure in a low-symmetry space group and the rest with a higher symmetry. This situation was foreseen 40 years ago by Cruickshank *et al.* (1964), who proposed that a program should be able to handle more than one space group at a time.

4.1.4. Rigid groups. The mechanism described in §4.1.3 enables translational shifts to be applied uniformly to groups

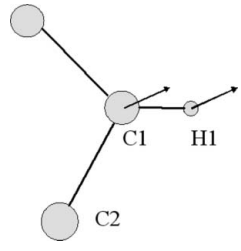

Figure 10
Riding refinement. Atoms C1 and H1 are riding together, preserving both the length and the direction of the C—H bond. Since C2 can move independently, the C2—C1—H1 angle may change from its original value.

of atoms – the whole group moves parallel to the cell edges. More commonly, the analyst may want a group of atoms to shift and rotate as a rigid body (*e.g.* each of the phenyl groups in a triphenylphosphine ligand). If the group contains n atoms, there are $3n$ physical parameters but only six least-squares parameters (three rotational and three translational shifts of the rigid group).

$$\begin{bmatrix} \partial x_1 \\ \partial y_1 \\ \partial z_1 \\ \partial x_2 \\ \partial y_2 \\ \partial z_2 \\ \partial x_3 \\ \partial y_3 \\ \partial z_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \frac{\partial x_1}{\partial \theta} & \frac{\partial x_1}{\partial \varphi} & \frac{\partial x_1}{\partial \chi} \\ 0 & 1 & 0 & \frac{\partial y_1}{\partial \theta} & . & . \\ 0 & 0 & 1 & . & . & . \\ 1 & 0 & 0 & . & . & . \\ 0 & 1 & 0 & . & . & . \\ 0 & 0 & 1 & . & . & . \\ 1 & 0 & 0 & . & . & . \\ 0 & 1 & 0 & . & . & . \\ 0 & 0 & 1 & \frac{\partial z_3}{\partial \theta} & \frac{\partial z_3}{\partial \varphi} & \frac{\partial z_3}{\partial \chi} \end{bmatrix} \cdot \begin{bmatrix} \partial u \\ \partial v \\ \partial w \\ \partial \theta \\ \partial \varphi \\ \partial \chi \end{bmatrix}. \quad (45)$$

Because nonlinear least squares only computes shifts to parameters, the rigid group must have the correct relative geometry before refinement begins. The terms in the matrix of constraint depend upon the current model and so must be recomputed before each refinement cycle. More sophisticated implementations reduce the number of degrees of freedom further, for example, only allowing a *tert*-butyl group to rotate about the terminal linking bond. Few implementations permit rigid groups containing more than one atom on special positions or groups sharing atoms. For these reasons, this constraint has largely been superseded by equivalent restraints (see §5).

4.1.5. ADP constraints and TLS groups. Changing from isotropic to anisotropic atomic displacement parameters replaces the single parameter $U_{\rm iso}$ with the six parameters of the symmetric tensor $\mathbf{U}_{\rm aniso}$, *i.e.* more than doubles the number of parameters to be refined for each atom. Modern data from good crystals will generally permit the anisotropic refinement of non-H atoms. However, if the data are poor (for example, have large standard uncertainties or were only observable to a low Bragg angle), the model may need simplifying or the X-ray observations may need supplementing by observations of restraint (§5).

For example, if the data only extend to low resolution, there may be a temptation to refine some of the 'less important' atoms (*e.g.* peripheral phenyl or *tert*-butyl groups) with isotropic ADPs in order to reduce the number of variables. This is almost certainly a poor model to use if the poor diffracting power of the sample is due to large thermal motion of the atoms. The peripheral atoms will almost certainly be vibrating anisotropically, leading to the paradox that poor data can be best represented by a complex model. The best solution is to use anisotropic ADPs, together with TLS (translation, libration, screw) constraints or copious restraints to assure a physically reasonable model.

If a group of atoms are bound more or less inflexibly together, their individual ADPs will be highly correlated. The individual ADPs can be replaced by a group **TLS** tensor (Schomaker & Trueblood, 1968). This is a six-by-six tensor,

$$\begin{bmatrix} \mathbf{T} & \mathbf{S}^{\rm t} \\ \mathbf{S} & \mathbf{L} \end{bmatrix}.$$

The 3×3 subtensor **T** corresponds to rectilinear displacements of the whole group – it is rather similar to the conventional atomic U^{ij}. The subtensor **L** (degrees or radians squared) represents torsional oscillation (libration) around three perpendicular axes. The screw tensor **S** and its transpose $\mathbf{S}^{\rm t}$ link the libration and the translation tensors. **TLS** contains 20 independent parameters, which replace the six individual U^{ij} terms for each atom in the group, leading to a substantial reduction in the number of parameters to be refined for a rigid group of four or more atoms (Pawley, 1966; Winn *et al.*, 2001). As with rigid-body refinement (§4.1.4), there are implementational problems if different TLS groups share atoms or include symmetry operators. From a physical point of view it is unlikely that large groups of atoms (*i.e.* those that would most benefit from a TLS approach) will be strictly rigid. For this reason, TLS-constrained refinement has largely been superseded by ADP similarity or TLS restraints.

4.1.6. Reparameterization. If pairs of parameters are highly correlated in their natural coordinate system (for example, because of pseudo-translational symmetry), they can be transformed to a coordinate system that reduces the correlation (Prince, 1994*b*). In the new system, one coordinate direction corresponds to the sum of the parameters, the other to the difference. This is similar to eigenvalue filtering, except that the rotation needed to diagonalize the 2×2 matrix is prescribed in the matrix of constraint (Fig. 11).

This technique is commonly used in biological and social sciences but is not widely implemented in crystallography (Watkin, 1994). The matrix of constraint is

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} x'_1 \\ x'_2 \end{bmatrix}. \quad (46)$$

4.1.7. Miscellaneous. With the widely used refinement programs, the user is generally restricted to a set of predefined constraints aimed at commonly occurring situations. This is largely because of the difficulty in providing a general method for converting an arbitrary function defined by the user into a properly integrated matrix of constraint. Some programs will let the user provide simple linear functions of

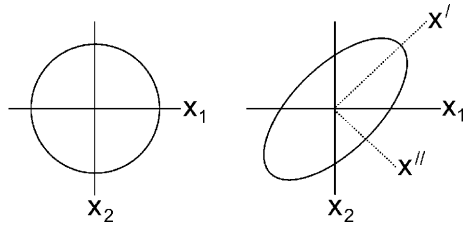

Figure 11
Uncorrelated (left) and correlated (right) parameters. The circle and ellipse are the locus of points of the constant minimization function as x_1 and x_2 are varied. For uncorrelated parameters, a small change in x_1 has almost no effect on x_2. For the correlated parameters, a small shift in either parameter implies a shift in the other. Re-parameterization to the x'–x'' coordinate system removes (or reduces) the correlation.

the usual crystallographic parameters. *TOPAS*, a very modern computer program, permits the user to define new variables and provide completely general functions as constraints (Coehlo, 2004).

4.1.8. Standard uncertainties of constrained parameters. The scaled inverse of the weighted normal matrix is the source of all information about parameter uncertainties and their correlations. If constraints have been applied, the normal matrix only contains information about least-squares parameters, and the matrix of constraint must be used to propagate this information into the physical parameters. Without these two matrices, it is not possible to compute proper uncertainties on derived parameters (such as interatomic distances). In the example of the riding CH group given above, the standard uncertainties on the H atom will be exactly the same as those for the C atom because both come from the same least-squares parameters. However, because the physical parameters are 100% correlated (through the constraint matrix), the bond length between them will have an s.u. of exactly zero. The same argument is true for a rigid group. The atoms in the group will all have associated uncertainties, but the s.u. values of interatomic distances within the group will be zero. Thus while the s.u. of a distance between two H atoms riding on the same C atom will be exactly zero, the s.u. of the distance between two H atoms riding on different C atoms will be the same as the s.u. of the distance between the C atoms. It is for this reason that some programs emphasize that they use the full correlation matrix and yield correct derived parameter s.u. values, which may be different from those computed by programs that only have access to atomic s.u. values.

5. Restraints

When an analyst has no doubt about a functional relationship between parameters in the physical model, constraints are the correct tool to use to feed this information into the refinement. A more flexible tool for influencing the outcome of a refinement under less clear-cut conditions is the use of the 'observations of restraint', usually just called restraints (Waser, 1963). With this technique, instead of using the functional relationship to reduce the number of least-squares parameters, the relationship is used to generate new observations to be used in addition to the usual X-ray observations. Equation (12) was an example of a function of the X-ray observations that was to be minimized during refinement. An equivalent function including restraints is

$$N = \sum^{\text{All observations}} w\left(F^2_{\text{obs}} - F^2_{\text{calc}}\right)^2 + \sum^{\text{All restraints}} w(T_{\text{obs}} - T_{\text{calc}})^2, \tag{47}$$

where T_{calc} is the value of some target function computed from the structural parameters, and T_{obs} is the corresponding value taken from some other source. The T_{obs} values become a few additional observations amongst many X-ray observations, and so their influence is not absolute. The linearized equation of restraint [*cf.* equation (19)] used in least squares is

$$T_{\text{obs}} - T_{\text{calc}} = \sum^{\text{parameters}} \frac{\partial T_{\text{calc}}}{\partial x}\delta x. \tag{48}$$

The strength of the restraint can usually be adjusted by giving it a standard uncertainty (reflecting one's confidence in the assertion), which is converted into a least-squares weight in the usual way (Rollett, 1970). With a sufficiently small s.u., a restraint will influence the model almost as strongly as a constraint. As will be evident from equation (49),

$$\begin{bmatrix} w_{11} & w_{12} & \cdot & w_{1n} & & & & \\ w_{21} & w_{22} & \cdot & \cdot & & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & & & & \\ w_{n1} & \cdot & \cdot & w_{nn} & & & & \\ & \cdot & & & w_1 & \cdot & \cdot & \\ & \cdot & & & \cdot & w_2 & \cdot & \\ & \cdot & & & \cdot & \cdot & w_3 & \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial F_{1_{\text{calc}}}}{\partial x_1} & \cdot & \frac{\partial F_{1_{\text{calc}}}}{\partial x_m} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \frac{\partial F_{n_{\text{calc}}}}{\partial x_1} & \cdot & \frac{\partial F_{n_{\text{calc}}}}{\partial x_m} \\ \frac{\partial T_{1_{\text{calc}}}}{\partial x_1} & \cdot & \frac{\partial T_{1_{\text{calc}}}}{\partial x_m} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \frac{\partial T_{n_{\text{calc}}}}{\partial x_1} & \cdot & \frac{\partial T_{n_{\text{calc}}}}{\partial x_m} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial x_1}{\partial x'_1} & \cdot & \frac{\partial x_1}{\partial x'_q} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \frac{\partial x_m}{\partial x'_1} & \cdot & \frac{\partial x_m}{\partial x'_q} \end{bmatrix} \cdot \begin{bmatrix} \delta x'_1 \\ \cdot \\ \cdot \\ \delta x'_q \end{bmatrix}$$

$$= \begin{bmatrix} w_{11} & w_{12} & \cdot & w_{1n} & & & & \\ w_{21} & w_{22} & \cdot & \cdot & & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & & & & \\ w_{n1} & \cdot & \cdot & w_{nn} & & & & \\ & \cdot & & & w_1 & \cdot & \cdot & \\ & \cdot & & & \cdot & w_2 & \cdot & \\ & \cdot & & & \cdot & \cdot & w_3 & \end{bmatrix} \begin{bmatrix} F_{1_{\text{obs}}} - F_{1_{\text{calc}}} \\ \cdot \\ \cdot \\ \cdot \\ F_{n_{\text{obs}}} - F_{n_{\text{calc}}} \\ T_{\text{obs}} - T_{\text{calc}} \\ \cdot \\ \cdot \end{bmatrix}, \tag{49}$$

the equations of restraint are just additional observations, so they must be filtered through the matrix of constraint. This means that, if a parameter is the subject of both a restraint and a constraint, the constraint will be obeyed even if there is a conflict. This conflict will be evident as a large value for the restraint residual ($T_{\text{obs}} - T_{\text{calc}}$). Residuals larger than about three times the requested standard uncertainty should always be investigated.

In crystallography, restraints are generally applied under one (or both) of two situations:

(1) The starting model is very poor, and the user suspects that the minimization space is full of false minima. Suitable soft restraints may help the minimization move towards an acceptable (and hopefully correct) minimum – they provide a guide through the minimization space. Once the solution is 'correct', the restraints can be slackened or removed so that the structure becomes the one 'seen' by the X-ray data only. This method is commonly used in macromolecular analyses, where phasing and map interpretation are uncertain, and in extended lattice work (*e.g.* zeolites), where the starting model

may be a related compound. In an extreme case, the initial model may be derived from a related compound by refinement without any X-ray data at all, but with an extensive list of restraining interatomic distances and angles. Such a procedure is sometimes called DLS [distance least squares, after a program of the same name (Meier & Villiger, 1969)]. To be effective, there must be sufficient restraints to define the position of every atom relative to its neighbours and sufficient restraints involving symmetry-related atoms to fix the structure with respect to the space-group symmetry operators. Without these latter restraints, the position of the structure in the cell would be undefined.

(2) Normal refinement has produced an unacceptable structure, *i.e.* one that does not conform in detail to the accepted rules of chemical bonding. The structure could quite simply be 'wrong' – a false solution to the phase problem (see Murphy *et al.*, 1998). Alternatively, the result could imply that the data are inadequate in some way or that the model is under-parameterized. If the data cannot be improved and the user cannot think of additional valid parameters to add to the model, restraints should be investigated as a way to achieve a preconceived end result. Ideally, the standard uncertainty of the unrestrained anomalous feature should be sufficiently high that the feature lies within 3σ of the expected value, implying that the X-ray data lack the information content needed to define the feature. A well chosen restraint will not be in conflict with the X-ray data. Fig. 12 shows how a distance restraint will fix the location of an ill-determined parameter in a broad minimization well. There are two complications to watch out for in restrained refinement.

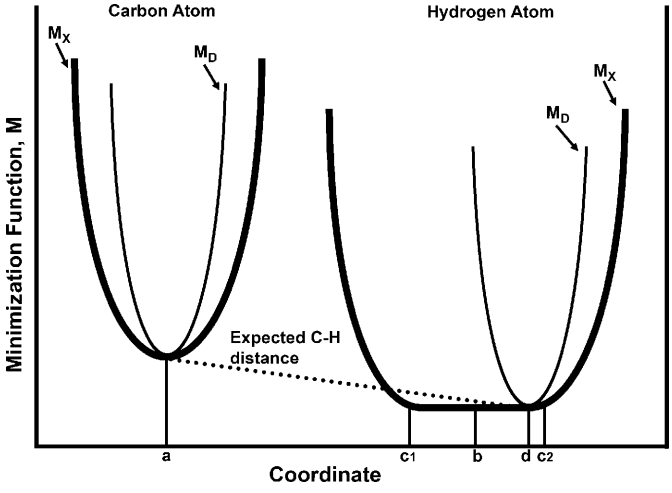

Figure 12
Schematic representation of a C—H distance restraint. The bond is assumed to be parallel to the x axis so that only one coordinate need be visualized. The heavy lines are the X-ray minimization function (M_X) for small changes in the C- or H-atom position. Note the broad minimum for hydrogen ($c1$–$c2$), implying that the X-rays only poorly determine its position. The light lines are the minimization function (M_D) for a C—H distance restraint and are the same for both atoms. The C atom is well defined by the X-rays (a), and the restraint places the H atom at an optimal position (d) on the broad plateau. If the C—H distance had been restrained to a much longer value, this would have driven the H-atom position up the side of the X-ray minimization function, revealing a conflict with the X-ray data.

One is that the values of the restrained parameters deviate from the target values by more than 3σ. This implies a conflict between the X-ray data and the requested restraint.

The other is applying restraints with very small s.u. values. The crystallographer should be aware that many restraints (*e.g.* distances, angles, geometric similarity, ADP similarity, Hirshfeld restraints) involve the differences between parameters, rather than the parameter values themselves. This means that the restraints will make significant contributions to the off-diagonal elements of the normal matrix. This is inevitable since the restraint is simply encoding the fact that the parameters are correlated. Excessive weighting of the equations of restraint can reduce the diagonal-dominance of the matrix and have a detrimental effect upon the conditioning, leading to numerical instability.

Restraints are commonly used in cases of disorder to ensure normal bond lengths and ADPs. The standard uncertainties of the restraints are adjusted manually to achieve a desired conformity with the target values. Nocedal & Wright (1999) put this concept on a formal basis, varying the restraint weights on a cycle-by-cycle basis. Note that if, at the end of the refinement, $|T_{\rm obs} - T_{\rm calc}|$ is much greater that the requested uncertainty there is almost certainly an incompatibility between the model and the X-ray data.

5.1. Restraints commonly used in small-molecule crystallography

5.1.1. Geometric restraints. The CSD contains the atomic coordinates of 25 000 000 atoms, and one can estimate that over 50 000 000 bond lengths can be computed from them. From these, reliable mean values and dispersions can be computed for the more common molecular geometries. The most common of these are tabulated in *International Tables for Crystallogaphy*, Vol. C. Because bond lengths between atoms are influenced by the nature of the environment (*e.g.* other atoms connected to the bonded atoms), the tables have entries for common geometries. For less common atomic arrangements, programs such as *MOGUL* (Bruno *et al.*, 2004) can directly search an auxiliary database compiled from the CSD for the required bonds and standard uncertainties. Providing that the geometry of interest is well represented, mean bond lengths extracted from the CSD make good target values for restraints, and the observed dispersion about the mean can be used to provide a measure of confidence in the mean. The observational equation is

$$D_{\rm obs} - D_{\rm calc} = \sum^{\text{All parameters for two atoms}} \frac{\partial D_{\rm calc}}{\partial x}\delta x, \qquad (50)$$

where D are the interatomic distances and x are the coordinates of the two atoms (A and B) forming the bond. The derivatives are given in equation (51), where g_{ij} are elements of the metric tensor:

$$\begin{aligned} \partial D/\partial x_A &= (g_{11}\Delta x + g_{12}\Delta y + g_{13}\Delta z)/D, \\ \partial D/\partial x_B &= -(g_{11}\Delta x + g_{12}\Delta y + g_{13}\Delta z)/D. \end{aligned} \qquad (51)$$

Note that only differences in atomic coordinates occur in these expressions, so that unless atom A is related to B by a symmetry operator, absolute values cannot be found for x, y and z (Watkin, 1988).

Angles can be restrained either directly to some target value or by restraints on the two 1–2 and one 1–3 distances. In the latter case a nondiagonal weight matrix should be used, but this is rarely done in practice. Because inter-bond angles are more flexible than bond lengths, the target values are more difficult to assign, and the standard uncertainties will be much greater. Similarly, torsion angles can be represented either as a function of the coordinates of the four atoms or by six distance restraints. The standard uncertainties will be even greater than for angles.

5.1.2. Rigid-group restraints. Distance restraints do not necessarily have to be between directly bonded atoms (*e.g.* the torsion angle restraint above involves a 1–4 distance). If sufficient bonded and nonbonded distances can be specified, the atoms involved will behave like a rigid group. However, if the group is approximately planar, interatom distances will not be effective in controlling out-of-plane displacements, since quite large displacements have only a small effect on the distances.[4] There are several better algorithms for restraining groups of atoms to be coplanar (Urzhumtsev, 1991) or to have similar geometry, that is, one group is related to the other by a translation vector and a 3×3 matrix (Blanc & Paciorek, 2001). This matrix can represent a pure rotation, a rotation–inversion or a rotation–dilation (Watkin, 1980).

5.1.3. Anti-bumping restraints. These are used to control nonbonded relationships between atoms, usually to prevent solvent molecules or disordered groups 'bumping' into other parts of the structure. The restraint can be implemented either as a normal distance restraint, which is only applied if the interatomic distance becomes less than some threshold, or as a distance restraint where the weight is itself a function of the interatomic distance. An expression of the form

$$w \propto \left(\frac{\text{Sum of van der Waals radii}}{\text{Interatomic separation}}\right)^{k} \qquad (52)$$

is similar to the repulsive part of the Lennard–Jones potential and avoids a discontinuity when the observed separation equals the target separation (*CRYSTALS*; Betteridge *et al.*, 2003). Note that, since the weight is a function of the residual, it must be recomputed for each cycle of refinement and that the refinement is thus no longer strictly least-squares.

5.1.4. Linked-parameter restraints. In §4.1.3 we discussed the riding model, in which pairs or whole groups of parameters are constrained to move synchronously. Sometimes, something less rigorous may be appropriate. For example, in a mineral, a particular site might be occupied by either of two elements, A and B. If we know that the site must be fully occupied, then the constraint of §4.1.2 is appropriate. However, if there is the possibility that the site may be only partially occupied, or may be contaminated by a small amount of an unidentified heavy impurity, a restraint should be used. An appropriate equation of restraint is $\mathrm{SOF}_A + \mathrm{SOF}_B \simeq 1.0\,(1)$, where 1.0 is the expected total site occupancy, and the s.u. in parentheses expresses the type of deviation from this value that we think possible.

5.1.5. Similarity restraints. In cases where exact target values for restraints are not available, targets can be inferred from the symmetry of the material. For example, in a CF_3 fragment, there is a high probability that all three C—F bonds will have the same length. Their average value, recomputed before each cycle of refinement, can be used as the target. Similar symmetry arguments can be applied to a phenyl group. If the data are of poor quality, all six C—C bonds can be made to be similar. A less stringent model might simply aim at achieving twofold similarity. If there are repeated fragments in a structure, equivalent bonds (and angles) in each fragment can be restrained to be similar. A similar approach can be applied to the independent molecules in a $Z' > 1$ structure. In general, geometrical similarity restraints are limited to 1–2 and 1–3 distances. This permits the groups to have independent torsional flexibility. Longer nonbonded distances can be restrained if total similarity is required.

5.1.6. Atomic displacement parameter restraints. Our knowledge about the behaviour of atomic displacement parameters is much less secure than our knowledge of molecular geometry, so ADP restraints are generally applied rather slackly (Irmer, 1990). To set up a restraint one needs to have some physical model for the atomic displacements.

(1) Approximate sphericity. The component U^{ij} is restrained so that the ellipse is only slightly distorted from a sphere. The required expressions are

$$U^{11} \simeq U_{\text{equiv}}, \quad U^{12} \simeq U_{\text{equiv}} \cos\gamma^* \quad \textit{etc.} \qquad (53)$$

There are two commonly used definitions of U_{equiv}. If U_i *etc.* are the principal axes of the ellipse $U^{[ij]}$ then

$$\begin{aligned} U_{\text{arithmetic}} &= (U_1 + U_2 + U_3)/3, \\ U_{\text{geometric}} &= (U_1\, U_2\, U_3)^{1/3}. \end{aligned} \qquad (54)$$

The arithmetic mean is normally quoted in CIF files; the geometric mean corresponds to a sphere with the same volume as the ellipsoid and is preferable in some calculations since it is sensitive to extreme values of the principal axes (Watkin, 2000).

(2) Bond stretching. The magnitude of the mean displacements of pairs of atoms, u_r, is restrained to be equal along the direction linking the atoms A and B, *i.e.* the differences in the components of the ADPs are reduced. This restraint implements Hirshfeld's 'rigid-bond' condition, which postulates that X-rays cannot detect bond stretching vibrations (Hirshfeld, 1976).

$$\langle u_r^2 \rangle = \mathbf{r}'\mathbf{U}_A\mathbf{r} \simeq \mathbf{r}'\mathbf{U}_B\mathbf{r}, \qquad (55)$$

where $\mathbf{r}$ is the normalized vector between the two atoms, and $\mathbf{r}'$ is its transpose.

The atoms A and B are often next neighbours (linked atoms) but in fact can be any or all pairs in a structural motif

[4] If D is the bond length, a the small change in this length and d the out-of-plane displacement, then $d^2 = a^2 + 2aD$.

behaving as a rigid body. This scheme provides restraints equivalent to a TLS constraint, with the advantage that it permits slight deviations from absolute rigidity. It ensures that a refinement will conform to the Hirshfeld criteria, that the components of the ADPs will be equal along the directions of the interatomic vectors for a rigid fragment of a well determined structure (Hirshfeld, 1976). Its main weakness is that it provides little control over out-of-plane components of U^{ij} for approximately planar groups (see §5.1.2).

(3) Similarity. The individual components U^{ij} of the ADPs of adjacent atoms (A and B) are restrained to be similar. This restraint implements the idea that, even in flexible systems, adjacent atoms generally behave similarly.

$$U_A^{ij} \simeq g U_B^{ij}, \tag{56}$$

where g is a user-defined scale factor, generally unity. This model is likely to be more uncertain than the previous and so is generally applied with a smaller weight (larger s.u.). Its uses include controlling out-of-plane components of the ADPs and ensuring that equivalent atoms in a disordered fragment have similar ADPs. This similarity restraint can be applied simultaneously with the Hirshfeld restraint [§5.1.6 point (2)].

(4) Rigid body. The individual ADPs are restrained to conform to some predetermined rigid-body atomic displacement model (TLS model; see §4.1.5). The equation of restraint is

$$\mathbf{U}_{\rm obs} \simeq \mathbf{U}_{\rm calc} = \begin{bmatrix} \mathbf{I} & \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{T} & \mathbf{S}^{\rm t} \\ \mathbf{S} & \mathbf{L} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{A}^{\rm t} \end{bmatrix}, \tag{57}$$

where **A** is a matrix containing functions of the atomic positions [not to be confused with the design matrix, equation (32)]. Such a restraint can be used to 'regularize' the individual ADPs of a group of atoms. **T**, **L** and **S** are determined from the current model and used to guide refinement in a subsequent cycle of refinement, after which they are recomputed for the following cycle. Madsen *et al.* (2004) extend the regularization concept to H atoms, where the ADP consists of an ellipsoid computed from the local rigid-group **TLS** tensor augmented by additional components parallel and perpendicular to the bond from the adjacent heavier atom.

5.1.7. Sum and average restraints. Groups of site occupation factors can be constrained so that their sum is unity exactly [equation (43)]. However, with some materials, for example minerals, it is possible to have more than two sites or species involved in a relationship, and for the sum of the occupation factors to be less than unity if vacancies can occur or greater if an unidentified heavier element is present in small quantities. Restraining the sum of the occupation factor shifts to be about zero, with a standard uncertainty, will deal with this situation (see §4.1.2):

$$0.00\,(1) = \delta{\rm occ}_1 + \delta{\rm occ}_2 + \delta{\rm occ}_3 + \delta{\rm occ}_4. \tag{58}$$

A related situation is where a number of sites are expected to have approximately the same occupation factor. This can be achieved by restraining the individual sites to their average value.

5.1.8. Origin fixing. There are a number of space groups where the origin of one or more directions is not fixed by symmetry operators. These space groups are often called polar, and axes whose origins are not fixed are called the polar or origin-free axes. The origins of these axes are arbitrary and are said to be floating origins. Examples are $P12_11$ (one floating origin parallel to b), $P1c1$ (floating origins parallel to a and c) and $P1$ (three floating origins). In these polar directions, the relative positions of the atoms can be determined (so that their separations can be computed), but their absolute positions are undefined. Simple attempts to refine the coordinates of all atoms in the structure in the polar direction will lead to a singular normal matrix. The ill-conditioning of the matrix is generally controlled automatically in modern programs by applying restraints to keep the centre of gravity of the structure invariant during the refinement (Flack & Schwarzenbach, 1988). The origin can also be fixed by a constraint of the form given in equation (42), but with the occupation parameter shifts replaced by the positional parameter shifts, weighted by the atomic scattering powers, along the free axis. In the older literature there are many structures reported where the origin was fixed by not refining the appropriate coordinate of one atom. Under this constraint the selected coordinate, being unrefined, had a standard uncertainty of zero exactly, and the s.u. that should be associated with it was added to the corresponding coordinate uncertainties of all other atoms. This is not a serious problem if the selected atom is a very heavy atom. If a light atom (C, N, O) is used, bond length s.u. values computed from the positional s.u. values only (*e.g.* in *PARST* or *PLATON*) will be seriously overestimated. Those computed from the full variance–covariance matrix will be correct.

5.1.9. Shift-limiting restraints. A technique for stabilizing refinements widely used in noncrystallographic domains is the Levenberg–Marquardt method (Gill *et al.*, 1981). In this method, all the terms on the diagonal of the normal matrix are multiplied by a factor $(1 + \lambda)$, thus increasing the dominance of the diagonal elements and improving the conditioning of the matrix, but at a cost of predicting smaller shifts, *i.e.* providing a kind of 'damping' for the refinement. In some applications the value of λ is obtained in an iterative process by the program itself. In crystallography, this is impractical because, for each iteration to determine λ, shifts have to be applied and structure factors recomputed. In *SHELXL*, a simplified strategy is used in which λ is set to a small default value suitable for dealing with the effects of rounding errors in the mathematics or moderate parameter correlation. The user should increase λ in cases of severe correlation and set it to zero for a final cycle of refinement in order to obtain the best estimates of parameter correlations and standard uncertainties.

A different implementation of the same idea is *via* shift-limiting restraints (Watkin, 1994). Here, only terms affecting volatile parameters are modified. The restraining equation is

$$0.00\,(1) = x_{\rm new} - x_{\rm old},$$

leading to

$$x_{\rm new} = x_{\rm old} + \frac{\partial x_{\rm old}}{\partial x}\delta x, \qquad (59)$$

where 0.00 (1) is the expected shift (zero) and tolerance (0.01) in the natural units of x. This equation simply augments the diagonal element for the parameter x by a factor depending on the tolerance. Unlike proper Marquardt damping, the restraint can be different for different parameters and does not require the normal matrix to be normalized. All Marquardt-type methods for damping the refinement have the advantage that they modify the normal matrix before it is inverted, and this approach should not be confused with the application of partial shift factors after inversion (§3.10).

One of the criteria for the satisfactory convergence of a refinement is that the largest shift in every parameter is only a small fraction of its s.u. Occasionally this criterion cannot be achieved for one or more parameters, with the shift changing sign on successive cycles but maintaining about the same magnitude. This might happen for the cross terms of the ADP of an almost cylindrical or spherical ellipsoid, where the direction of the axes is almost undefined. A few cycles of refinement with heavier damping may stabilize the parameter, after which the damping can be relaxed.

5.1.10. Miscellaneous restraints. Because restraints are just additional observational equations, they are easy for programmers to implement in a general way in order to provide the user with flexible choices. In *CRYSTALS* (Betteridge *et al.*, 2003), the user can include in the data file a Fortran-like expression involving the refinable parameters and other crystallographic data (unit cell, atomic weight, ionic charge *etc.*) to be used as a restraint. For example, the mineral garnet has the general formula $A_3B_2(SiO_4)_3$, where A could be Ca^{2+}, Fe^{2+}, Mg^{2+} or Mn^{2+} and B could be Al^{3+}, Fe^{3+}, Cr^{3+} *etc.* In melanitic garnet (titanium andradite), Ti^{4+} can partially substitute for B and for Si in (SiO_4). To preserve the charge balance, some of the Si and O will be chargeless, leading to the relationship

$$\begin{aligned}3(2Ca^{2+}_{\rm occ} + 2Fe^{2+}_{\rm occ}) + 2(3Fe^{3+}_{\rm occ} + 3Al^{3+}_{\rm occ} + 4Ti^{4+}_{\rm occ}) \\ + 3(Si_{\rm occ} + 4Si^{4+}_{\rm occ} + 4Ti^{4+}_{\rm occ}) = 12(O_{\rm occ} + 2O^{2-}_{\rm occ}),\end{aligned}$$

derived from $3Ca^{2+} + 2Al^{3+} + 3Si^{4+} = 12O^{2-}$, where the occupancies of these atomic sites are refinable parameters (Agrosi *et al.*, 2002). Even with this equation of restraint, the site occupation factors (which provide information about the origins of the mineral) can only be determined by the use of very high quality data. The most common problem occurring when site occupancies are being refined is instability of the ADPs. This is because the shape of an atomic form factor and the shape of the ADP smearing function are similar over a small resolution range. As a consequence, a slight change in an ADP has almost the same effect on a computed structure factor as a slight change in occupancy (*i.e.* the two parameters are highly correlated). This correlation is best reduced by collecting diffraction data to a high resolution and applying a very reliable geometrical absorption correction (which is also resolution-dependent).

6. Some numbers

6.1. Why crystallographic refinement works so well

So far, we have been concerned with outlining the mathematics and physics encoded in most modern programs on the assumption that the methods will work in practice – and of course experience shows that they do. Nonetheless, it is illuminating to carry out an order-of-magnitude estimate of the calculations involved.

Imagine an organic structure in $P1$ with a cell of $10 \times 10 \times 10$ Å. For an averagely well diffracting crystal on an ordinary diffractometer, it turns out that the index of the highest-order reflection ($h_{\rm max}$) observable along an axis is at least equal to the length of the axis in ångstroms. Since in $P1$ the asymmetric part of the reciprocal lattice only needs any two of h, k or l to take both positive and negative values, the number of unique reflections that can be observed is

$$n \simeq 20 \times 20 \times 10 = 4000.$$

For organic and organometallic materials, the approximate volume of the molecule can be estimated as about 20 Å^3 per non-H atom. The number of atoms in the cell is thus

$$m \simeq 1000/20 = 50.$$

For an isotropic model (parameters x, y, z, $U_{\rm iso}$ per atom), this gives $4000/(50 \times 4) = 20$ observations per refinable parameter. For the anisotropic model (x, y, z and six U^{ij}), this gives nine observations per parameter. This calculation would be equivalent to determining the gradient and intercept of a straight line through 18 experimental observations. In the simplified analogy of fitting a straight line to experimental observations, one only needs two error-free observations to determine the gradient and intercept. In the presence of experimental errors, more observations enable one to determine their effects on the parameter values and s.u. values, and to ensure that the model is functionally valid. For example, a straight line, an arc of a circle and a sine curve can all be drawn through two points. It is unlikely that 18 observations from a linear function would also fit an arc or sine curve unless the observations were bunched close together. The IUCr guidelines of ten observations per parameter for centric or heavy-atom structures reduces the possibility of a fundamentally wrong structure being published. For noncentric structures, there are only one-half of the number of theoretically independent reflections per parameter, so that the reduced guidelines are eight observations per variable in the absence of any atoms heavier than chlorine. Since it is often possible to infer that the model is essentially correct by reference to the chemistry or physics of the material, these guides represent the most pessimistic case, and convincing refinements can be completed with many fewer observations [see, for example, Lozano-Casal *et al.* (2005), where a sample contained in a diamond anvil cell yields only 5.7 observations per variable].

Crystallographers are in the privileged position of apparently having an experimental technique that yields copious independent data. However, not all data are of equal value for determining the structural parameters. Some reflections,

unidentified at the start of an analysis, may have a special influence on some parameters. The effect is called the 'leverage' of an observation (Prince, 1994*c*), and the matrix that shows the leverage of each reflection on each parameter is called the 'projection' or 'hat' matrix. While for the computation of a Fourier synthesis it is important to include all possible observations, it is quite permissible to drop observations from a least-squares refinement at will, provided the rejection is not based on the residual. The refined parameters should not be affected by the omissions, but their standard uncertainties will be. The hat matrix shows which reflections should be included in order to obtain the most reliable parameters and standard uncertainties (Merli *et al.*, 2000). This is an expensive computation that can only be completed once the structure has been solved, so that the IUCr guidelines represent generous margins set in the hope of including all important data for day-to-day working.

6.2. Computing requirements

The design matrix [equation (31)] has one row for each reflection and one column for each parameter. For a structure with n = 100 atoms, there will be approximately 1000 parameters and 10 000 reflections, giving a matrix of 10 000 000 elements. The matrix is massive in terms of personal computing and so is almost never computed and stored as such. Instead, if solution of the least-squares equations is to be *via* the normal equations [equation (34)], the normal matrix $\mathbf{A}^{\mathrm{t}}\mathbf{WA}$ is computed directly. This is a square matrix, with one column and row for each parameter – in this case about 1 000 000/2 elements since the matrix is symmetric. It is usually possible to store a matrix of this size in a modern personal computer, but its formation takes considerable computing effort – in this example 1000 × 1000 × 10 000/2 = 5 000 000 000 multiplications and additions. As memory becomes ever cheaper and processors ever faster, there is a temptation to simply increase the amount of storage allocated in a program for the normal matrix. Because the matrix is of the order $O(n^2)$, this rapidly becomes ineffective. The best solution when dealing with very large structures is to use macromolecular refinement programs, which do not involve the formation of this matrix, and use fast Fourier techniques for computing both the structure factors and the derivatives (Tronrud, 2003).

6.3. Blocked matrix strategies

In the past, various modified normal matrix strategies were designed to reduce storage and computation time for slower and smaller computers. They can still be useful today when dealing with moderately large systems (100–1000 atoms) and traditional small-molecule programs. The simplest strategy is to refine different parts of the structure in different least-squares cycles. This method is seductive in programming terms since it is simple to implement, but suffers from the cost of having to compute structure factors for all the atoms in the structure for each cycle – even the atoms that are not being refined. An improved strategy is to compute structure factors and derivatives for all atoms, but only accumulate certain 'blocks' of the normal matrix. For example, if there are n atoms, the positional parameters can be refined in one block ($3n$ parameters) and the anisotropic ADPs in the next ($6n$ parameters), or $n/2$ positions and ADPs can be refined in one block and the remaining $n/2$ in the next. As explained in §3.7, the off-diagonal terms in the matrix can be expected to be small so that the omission of carefully selected parameter combinations has a reduced impact on the rate of convergence of the refinement. Either of the strategies described above can be used to define the parameter blocks – in practice it is best to alternate the two strategies from time to time. For very large structures, it may make sense to pass a window of refined parameters through the structure. The idea can be extended still further and the full matrix reduced to nothing more than a chain of 4 × 4 (positions and U_{iso}) or 9 × 9 (positions and $\mathbf{U}_{\mathrm{aniso}}$) blocks along the leading diagonal. See Rollett (1970) for some suggested blocking schemes and cautions (Fig. 13).

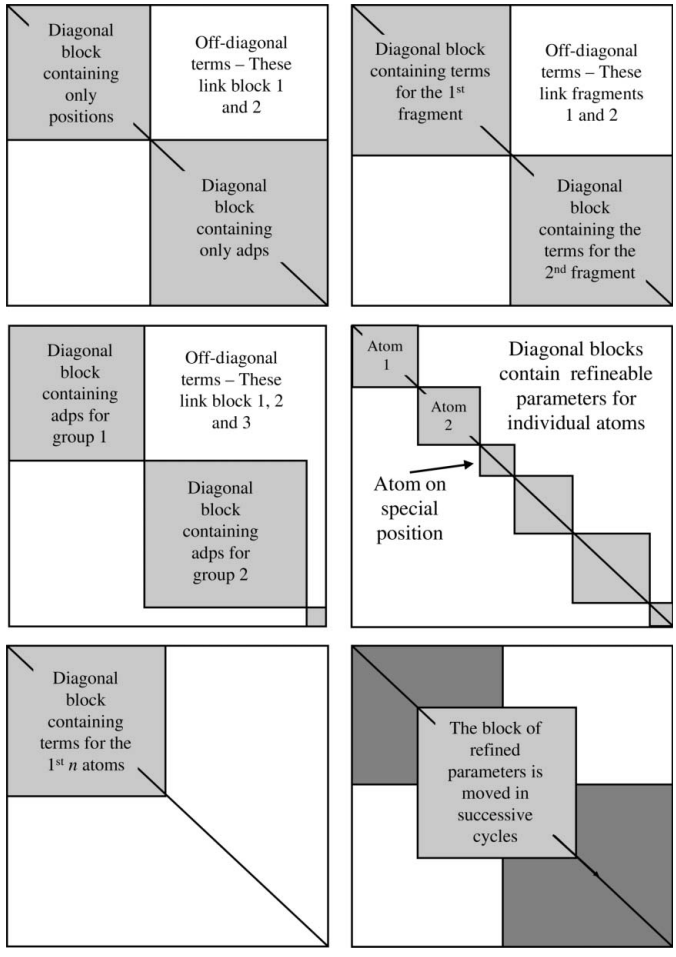

Figure 13
Matrix blocking schemes. Refining different parts of the structure in different matrix blocks (or different cycles) enables relatively large structures to be refined with normal small-molecule programs. Top row: Large-block schemes. Middle row: Schemes that put ADPs into different blocks. Note the small Cruickshank block containing the overall scale factor and dummy isotropic ADP. Bottom row: Sliding window scheme. A large block is moved through the structure with some parameters (atoms) from the previous block included in the current block.

Whenever the full normal matrix is replaced by sub-blocks, there is a risk of over-shifting parameters, so that partial shift factors of about 0.8 are sometimes automatically applied. In the extreme case for the 100-atom example above, if the full matrix of 500 000 elements is replaced by a chain of 100 9×9 atom blocks, only about 4500 elements need computing, with a corresponding massive reduction in the time (1/100) to compute and invert the normal matrices. The cost of this saving is that many more cycles must be computed to approach convergence, that the path to convergence is less certain and that no information is available about the correlation between parameters. In addition, structure factors and derivatives must be recomputed for each cycle. Nonetheless, this strategy has been implemented satisfactorily in the automatic structure development stages of the *SIR2004* program (Burla *et al.*, 2005), where a partial shift factor of 0.5 is applied for each cycle.

There is inevitably a large degree of correlation between the overall scale factor and the ADPs. If the ADPs are split into several blocks, their correlation with the scale factor can be handled *via* a special 2×2 block consisting of the scale and a 'dummy overall isotropic ADP' (Cruickshank, 1961). It must be emphasized that any normal matrix method of refinement that does not accumulate the whole matrix simultaneously is deliberately throwing away information about parameter correlations and so should only be used if experience shows that it is safe to do so. One case where this is definitely not permitted is when there are motifs related by either pseudo-translational symmetry or an approximate centre of symmetry (or worse still, the space group has been incorrectly assigned and the operators should be exact). In these cases ordinary full matrix refinement may be unstable because of the high parameter correlations. Refining different motifs in different matrix blocks (or different refinement cycles) will only conceal this correlation (leading to incorrect parameter values) and not cure it.

Finally, the cosines of the unit-cell angles have an influence on the correlation between parameters, so that the correlation increases for cell angles far from 90°. For monoclinic crystals it is better to choose a unit cell with angles close to 90°, even if this means using a nonstandard space group. If a cell with oblique angles must be used, parameters related by the oblique angle must be refined together (Giacovazzo *et al.*, 2002).

7. Refinement strategies

It is probably true to say that, for a university structure analyst dealing with a mix of organic and organometallic materials, something like 50% of the samples will solve and refine effortlessly, with the analyst only having to verify the atom-type assignments. It is the remaining 50% that provide interest and challenges for both the analyst and software designers. However, even in easy cases the users (or the program) need a strategy and methods for assessing the progress of the work.

7.1. Data assessment

It is important to have realistic expectations about the outcome of a refinement, based upon an assessment of the original X-ray data. In general, high redundancy or high multiplicity of observations is more important than high completeness (*i.e.* the number of independent reflections observed as a fraction of those possible). This is because high redundancy enables one to detect outliers and enables data processing programs to make better compensations for deficiencies in the crystal quality or data-collection procedures. $R_{\rm int}$ (or the merging R) is a measure of the self-consistency of the data and should be looked at as a function of $I/\sigma(I)$ and of resolution. Values of $R_{\rm int}$ in excess of 0.50 (50%) for batches of data indicate that, for the batch in question, the data are not even self-consistent and are therefore unlikely to be consistent with any model. The mean intensity and maximum $I/\sigma(I)$ of the systematic absences should be used as a guide to the quality of the weak data – it is often found that the weak data are systematically over- or underestimated. A Wilson plot deviating from the usual gentle undulation, especially an upturn at high angles, may indicate a problem with the data. A Wilson plot giving an abnormally small (or even negative) overall ADP may point to an inadequate absorption correction or warn of high pseudo-symmetry.

7.2. Getting started

Nonlinear least squares converges to the nearest local minimum of the minimization function. If the starting model is basically wrong, least squares is unlikely to make it come right. Make a careful examination of the model extracted from a Fourier map. If there are peaks (atoms) in really improbable places, eliminate them. If there are peaks that seem to conform to a distorted image of an expected moiety, use a geometrical regularization procedure to improve the geometry. If the model from direct methods or Patterson methods looks as expected, try least squares; otherwise try Fourier refinement. If a peak is eliminated by accident, do not worry (unless it was the heaviest atom in the structure) – it will reappear in a subsequent Fourier or difference map. Occasionally, when a structure is solved with *SIR* (Burla *et al.*, 2005) (which displays the atomic model throughout the development process), the initial E map looks convincing but the structure disintegrates during the refinement stage. If this happens, re-run the program, but stop it at the end of direct methods and then simply refine the overall scale factor. An examination of the reflections with the largest discrepancies between $F_{\rm obs}$ and $F_{\rm calc}$ will almost certainly reveal one or more seriously mismeasured reflections, which should be eliminated. Refinement should then proceed normally.

For extended lattice structures, it may be appropriate to use geometric regularization to improve the initial model. This can be performed with a program designed for this purpose (*DLS*; Meier & Villiger, 1969) or a more general program that permits refinement without X-ray data. An example of the use of this technique might be where an inorganic material undergoes a phase transition with a change of space group and

small changes in cell dimensions. A trial structure for the new phase can be obtained by putting the original atom coordinates (plus some symmetry-equivalent ones if necessary) into the new cell and refining with appropriate distance restraints. The main problem is creating the very large number of distance restraints (which will generally involve symmetry elements) that are needed to define the material.

7.3. Proceed cautiously

If the model is very good, a full anisotropic refinement may be possible immediately. However, if the model is unreliable or data are known to be poor or scarce, a more cautious approach is required. Start with pure statistical weights if you are confident in the estimates of $\sigma(I)$; otherwise use unit weights for F or modified unit weights for F^2 refinement (Rollett, 1984). The overall scale factor (which puts the observed structure factors onto an absolute scale) and average isotropic ADP can be estimated from the Wilson plot – start with these values and only refine positional parameters for a few cycles. During the development of a structure, it is not necessary (or useful) to refine each stage to completion. It is generally advisable to compute a few cycles of isotropic refinement before using an anisotropic model. If there are a few very heavy atoms and lots of light ones, some people recommend extending the anisotropic model first to the heavy atoms and then to all atoms. Careful examination of the ADPs at this stage may reveal misidentified atoms. Thus, if a C atom (six electrons) is erroneously assigned as an N atom (seven electrons), least squares will try to dissipate the extra electron by increasing the ADP (compared with near neighbours). This may be evident in a graphics program (Fig. 14) or *via* the Hirschfeld (1976) test.

If the model contains a totally false atom, in the least squares procedure its ADP will rise to a very large value, indicating that it should be removed from the model. Least squares can reveal spurious atoms, but it cannot introduce new ones. Missing atoms can be inferred from the chemistry [*e.g.* a phenyl group revealing three or more atoms can usually be improved geometrically (Fig. 15)] or found in a Fourier synthesis. There is some difference of opinion as to when a search should be made for H atoms – or even whether a search should be made at all. If one is going to look for the H atoms, one argument proposes that difference Fourier maps should be computed at the end of the isotropic ADP refinement, on the basis that anisotropic ADPs may elongate to try to simulate the missing H atoms. The alternative argument is that, since anisotropic ADPs are probably a better representation of the actual physical state, their inclusion should give more reliable estimates of the phase angles and hence a sharper difference map. The best solution is probably to place geometrically as many H atoms as possible and refine the atoms they are bonded to anisotropically. If the other H atoms cannot be found when their neighbouring atom is anisotropic, restore it to isotropic and try again. For extremely difficult cases, the Dunitz & Seiler (1973) weighting scheme may help (§3.5.4).

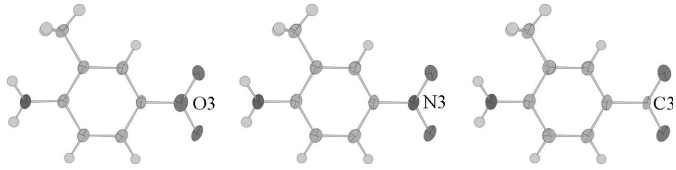

Figure 14
In the left image, the N atom of the nitro group has been misassigned as an O atom, O3. Note that the ADP is relatively large, as least squares tries to dissipate the extra electron. In the right image, the same atom has been misassigned as a C atom, C3. In this case, least squares has compressed the atom to try to increase the electron density to make up for the missing electron.

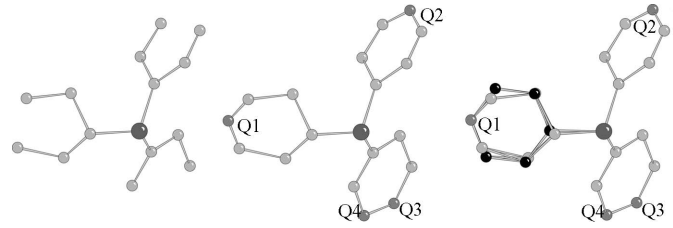

Figure 15
The left image is an incomplete triphenylphosphine group from the E map for a $Z' = 3$ coordination compound containing over 180 non-H atoms. In the middle image local geometry has been used to add four 'Q' peaks at approximately correct places, and in the right image one whole phenyl group has been 'regularized' (Q1 plus the black atoms). In the original work, completion and regularization were performed on all 27 phenyl groups, leading to a satisfactory refinement.

Having placed or found the H atoms, there is also much debate about their treatment during refinement. With the use of modern data from a good crystal, Harlow (1998) remarks 'I have a lot of confidence in structures where the hydrogen atoms were found and refined to reasonable positions (*e.g.* 0.85 < C—H < 1.05 Å) and with reasonable thermal parameters (*e.g.* $2.0 < B_{iso} < 6.0$ Å^2).[5] The hydrogen atoms appear to be very sensitive indicators of a reliable structure and simply don't refine well if there are even modest errors in the data or the model, or if the data is insufficient for the structural analysis.' However, since for the majority of organic and organometallic compounds there are roughly as many H as non-H atoms, this strategy raises the number of refined parameters by about 50%, running the risk that the analysis will fail to meet the IUCr parameter:observation ratio criterion. The most widely used strategies either refine the H atoms with riding constraints to their neighbouring atoms, refine them with copious restraints or simply recompute their positions geometrically after each cycle of refinement. Unless the analysis is specifically aimed at locating H atoms, their exact treatment is not really important. H atoms must be included somewhere near their 'real' positions because of the contribution they make to F_{calc}, but the exact parameter values do not have too much significance.

[5] $B = 8\pi^2 U$.

$$F_{\rm calc} \simeq \sum_{j}^{\rm Non\text{–}H\ atoms} f_j \exp\left[2\pi i\left(hx_j + ky_j + lz_j\right)\right] + \sum_{j}^{\rm H\ atoms} f_j \exp\left[2\pi i\left(hx_j + ky_j + lz_j\right)\right]. \quad (60)$$

Omitting the H atoms altogether leads to a systematic underestimation of $F_{\rm calc}$ [equation (60)] and hence to a systematic bias in $F_{\rm obs} - F_{\rm calc}$, which will have a small effect on the refinement of other atomic parameters.

Minor parameters such as extinction (§7.4.2) or the Flack parameter (§7.4.4) should only be introduced once the atomic model is seemingly resolved. To obtain valid values and standard uncertainties, these parameters must be refined together with the other refinable parameters.

7.4. Out-of-range parameter values

Occasionally during a refinement, parameters may take on values that are outside of their physically sensible range. There are two rather different ways in which a parameter may be out of range. If the discrepancy is of similar magnitude to the parameter standard uncertainty then the deviation is not significant. If the discrepancy is of the order of several standard uncertainties then there is almost certainly either something systematically wrong with the data or a serious problem elsewhere in the model, so that refinement is being performed in the region of a false minimum.

7.4.1. ADPs. Quite reliable values for the ADPs can be estimated by looking at the results for similar materials under the same conditions. For an organic or organometallic material at room temperature, $U_{\rm equiv}$ is generally in the range 0.04–0.06 Å^2, and at low temperature, 0.02–0.04 Å^2. For extended lattice materials, these values might be halved. The ADPs are always correlated with the overall scale factor. Note that, although the refined scale factor k is optimized to minimize $(Y_{\rm obs} - kY_{\rm calc})^2$, in general it should be similar to the Wilson plot scale factor or to k', where $k' = \sum F_{\rm obs}/\sum F_{\rm calc}$. Large differences between these three definitions may indicate an incorrect molecular composition being used for the Wilson plot, incorrect atom assignments in the trial structure or a few very badly measured reflections. As explained above, very large ADPs for a few atoms probably mean that they are spurious and should be removed from the model, or that they are part of an entity that is disordered in some way and which needs more careful modelling. If most of the ADPs are large, verify the quality of the data and begin to think about looking for twinning or the possibility that the structure is modulated. Look carefully at the original diffraction images for split reflections or weak satellite reflections.

If all the ADPs are unusually small, or possibly negative for $U_{\rm iso}$ and nonpositive definite for $\mathbf{U}_{\rm aniso}$, check the variation in absorption correction for the value of μr (where r is half the medial dimension of the crystal and μ is the absorption coefficient) in Table 6.3.3.3 of *International Tables for Crystallography* (1995, Vol. C) at the minimum and maximum observed θ values. For example, anisotropic refinement of inorganic materials containing very heavy atoms can lead to nonpositive definite ADPs if the only absorption correction applied is based on discrepancies between equivalent reflections (multi-scan method; Blessing, 1995). This correction effectively reduces the sample to a small sphere. This sphere may still give rise to a substantial variation in the absorption as a function of θ, which should also be corrected for. Fig. 16 shows the θ-dependent correction to be applied to samples with different values of μr. For a value of μr = 3.0, the data measured at 30° will be twice as strong as they should be, relative to data at 0°, *i.e.* giving the appearance that the ADPs are small.

If just a few atoms go nonpositive definite, this may indicate that there is some unidentified systematic error with the data or that the model is inadequate in some unidentified way. When the source of the problem cannot be located, the correct way forward is to look at the parameter standard uncertainties. If these are large compared with the parameter values themselves then the X-ray data do not contain sufficient information to define the parameter fully, and it is appropriate to apply the 'brutal Bayesian' strategy of setting the parameter to a 'reasonable' value and not refining it further, or to make it subject to restraints. If the s.u. is relatively small, restraints can be tried, but it is important to verify that the observed and target values do not differ by more than three times the requested standard uncertainty. If the restraint residual is large, there is something seriously wrong, and data should be re-collected from a different sample.

7.4.2. Secondary extinction parameter. Secondary extinction affects the most intensely diffracted beams from a high-quality crystal. In effect, so much of the incident radiation is diffracted by the side of the crystal facing the source that the opposite side receives an attenuated beam. This means that the integrated intensity of the diffracted beam is less than it should be (Larson, 1970). For most X-ray work, a single isotropic parameter is refined to compensate for this effect. For neutrons, an anisotropic model is often used.

The decision to refine an extinction correction is usually based on an examination of an $F_{\rm obs}$ *versus* $F_{\rm calc}$ plot. If this shows $F_{\rm obs}$ falling below the expected $F_{\rm calc}$ value for the strong

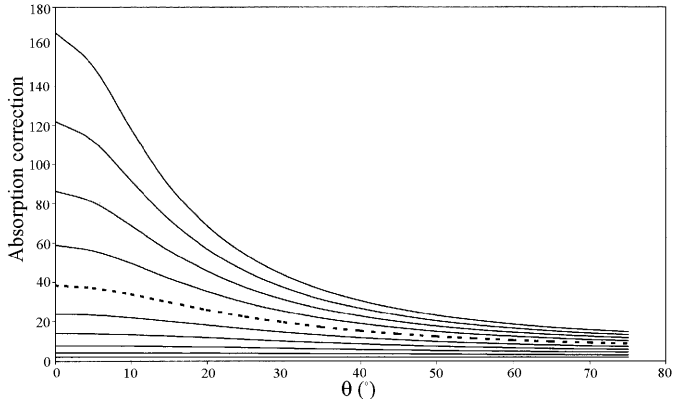

Figure 16
Absorption correction for spheres. Curves at intervals of 0.5 in μr, maximum = 5.0. For μr of 3.0 (dotted curve), the correction at 0° is twice that at 30°.

reflections, or if the ratio $\langle F_{obs}\rangle/\langle F_{calc}\rangle$ is substantially below unity for batches of high-intensity reflections, then an extinction parameter may need refining. If after refinement the value of the extinction correction is small compared with its standard uncertainty, it should be reset to zero and not refined further. A very negative value indicates something seriously wrong with the data or the current model.

7.4.3. Twinning. The phenomenon of twinning was relatively well understood when intensity data were measured photographically (*e.g.* Dunitz, 1964), but much of this knowledge was forgotten during the period dominated by serial diffractometers. Area-detector instruments have once again made working with twinned crystals relatively simple (Cooper *et al.*, 2002; Müller *et al.*, 2006). For the moment, we can consider a sample crystal that consists of two parts for which the reciprocal lattices interpenetrate in a rational way. This means that some (or all) of the diffraction spots will overlap, so that the measured intensities will contain contributions from both parts of the sample. Since the diffraction effects from the two components are not coherent, the observed intensity is simply the sum of the two component intensities,

$$F^2_{obs} = t_1 F^2_{component_1} + t_2 F^2_{component_2}, \quad (61)$$

where t_1 and t_2 are the fractions of the sample corresponding to the two components and are parameters to be included in the refinement. The constraint $t_1 + t_2 = 1.0$ must be applied if the overall scale factor is also being refined. The presence of twinning may not always be evident when the data are being collected – it may become evident during refinement because the analysis fails to converge to a reasonable R factor, because there is apparent disorder that cannot be resolved with any reasonable model, because there are inexplicable values for some of the ADPs or because there are a substantial number of reflections for which F_{obs} is significantly higher than F_{calc}. The twin fraction values must fall in the interval 0–1; values outside of this are physically meaningless. A very small (or negative) value for a twin fraction coupled with a large standard uncertainty indicates that there is no detectable diffraction from that component, which should therefore be removed from the model.

7.4.4. Flack parameter. The Flack parameter (usually given the symbol x) is defined by equation (62):

$$I_{calc}(h,k,l) = (1-x)\left|F_{calc}(h,k,l)\right|^2 + x\left|F_{calc}(\bar{h},\bar{k},\bar{l})\right|^2, \quad (62)$$

where hkl and $\bar{h}\bar{k}\bar{l}$ are a Friedel pair of reflections. Physically, equation (62) represents the diffraction from a crystal twinned by inversion and consisting of two domain states, the first domain state being that of the model crystal structure and the second being that of the inverted crystal structure. The mass fractions of the two domain states are x and $1 - x$. For an enantiopure compound, a value of 0.0 for x indicates that the molecules in the crystal are of the same chirality as the structural model, whereas a value of 1.0 indicates that the chirality of the model and the molecules in the crystal are inverted one with respect to the other. x should fall in the interval 0–1, and values outside of this interval are physically meaningless. Flack & Bernardinelli (2000) show that if the s.u. on the Flack parameter is greater than 0.30, the X-ray data contain no information for obtaining a reliable value for the parameter itself. In this case, the absolute configuration of the molecules can only be assigned, if at all, by non-X-ray means, and the value of x can be set to zero and not refined. It may be appropriate to report x and its s.u. to demonstrate that the calculation was actually performed. If the s.u. is small enough (< 0.04) then the parameter itself can be evaluated. If x falls close to zero or unity then the absolute configuration has been determined. If it falls within this range, but is significantly distant from the bounds, the material is probably twinned by inversion. If it falls significantly outside the range, there is probably something wrong with the data or other aspects of the model. In this case, one strategy would be to reset x to the closest limiting value and accept the rise in the residual $\sum w(|F_{obs}| - |F_{calc}|)$ as an indicator of the problems with the analysis. In any case, the refined value of x and its s.u. should be reported. Fig. 17 illustrates this graphically.

In the event that the Flack parameter refines to a significant value essentially equal to unity, the refined model is the inverted image of the real crystal. To make the model and the crystal correspond, the model structure must be inverted. For most of the noncentrosymmetric space groups in the standard settings given in *International Tables for Crystallography*, Vol. A, this can be achieved simply by replacing all atomic coordinates x, y, z by $-x$, $-y$, $-z$ (the ADPs are centrosymmetric so will not need changing). However, within the 11 pairs of enantiomorphous space groups, the space group will also need to be interchanged with the other member of the pair. Moreover, there are seven space groups where the coordinate inversion must be undertaken at some particular point other than the origin (Table 2).

Table 2
The 11 enantiomorphous space-group pairs and the seven space groups requiring a change of origin when the structure is inverted.

The enantiomorphic space groups					
$P3_1$	$P3_2$	$P6_1$	$P6_5$	$P4_1$	$P4_3$
$P3_112$	$P3_212$	$P6_2$	$P6_4$	$P4_122$	$P4_322$
$P3_121$	$P3_221$	$P6_122$	$P6_522$	$P4_12_12$	$P4_32_12$
		$P6_222$	$P6_422$	$P4_132$	$P4_332$

Space groups requiring a change of origin on inversion		
*Fdd*2	$I4_1$	$I4_122$
	$I4_1md$	$I\bar{4}_1cd$
	$I42d$	$F4_132$

7.5. Disorder

Substitutional disorder in inorganic materials (§§4.1.2 and 5.1.7), in which one element is sometimes replaced by other elements, can be dealt with either by a constraint that fixes the sum of the shifts in the site occupancies to be zero [equations (42) and (43)] or by a similarity restraint [equation (58)]. Such analyses are notoriously difficult and unstable. This is because

the site occupation factors are very highly correlated with the ADPs (Fig. 18). The best chance of resolving the occupation factors and ADPs requires good quality, absorption-corrected data covering an extended resolution range. Even so, it is generally necessary to assign a single ADP to the disordered atoms and possibly assign it a fixed value by analogy with neighbouring atoms.

Positional disorder in organic and organometallic compounds raises different problems, and it is not uncommon for more time to be spent on this part of the structure than on all the rest of the analysis. Symptoms of positional disorder include the following.

7.5.1. Positional parameters. If free refinement leads to unacceptable bond lengths and inter-bond angles, these should be restrained either to known values or to preserve chemically appropriate symmetry.

7.5.2. Site occupation factors. The shifts in these will generally be constrained to conserve a total occupancy of unity.

7.5.3. ADPs. The following problems may arise.

(1) Cigar-shaped ADPs. Anisotropic ADPs that are extremely elongated compared with their neighbours are always suspicious. The least squares procedure may have tried to 'stretch' the ellipsoid so that it includes two or more atomic positions. Delete the troublesome atoms from the model (or set their site occupation factors to zero) and view a difference electron density map in a suitable graphical program. If the electron density at the site of the troublesome atoms is more or less continuous then the cause is likely to be dynamic disorder or structural modulation. If the structure really does consist of two alternative sites for the atom, it may be possible to 'split' the atom into two partially occupied sites located near the extremities of the elongated ADP. When the refinement of the partial atoms is unstable (unacceptable ADPs or bond lengths, or the two atoms running together again), restraints or constraints may be needed. Imagine a CF_3 group (Fig. 19), with the F atoms split over two sites F'_n and F''_n. If the anisotropic ADPs cannot be refined independently for each F atom, constraints can be used to make the corresponding components of each atom ride together ($F'_n U^{ij} = F''_n U^{ij}$) or restraints can be used to ensure they are similar ($F'_n U^{ij} \simeq F''_n U^{ij}$). These subsidiary conditions are not completely appropriate – they relate the pairs of ellipsoids by a centre of symmetry rather than a twofold axis – but are probably adequate in most cases. Bond-stretching restraints can be applied to each C—F bond, and may also be applied in the F···F direction of atoms in the same disordered moiety. Note that $F'_n U^{ij} \simeq F''_m U^{ij}$ similarity restraints (*e.g.* F'_1–F''_2) are inappropriate since the principal

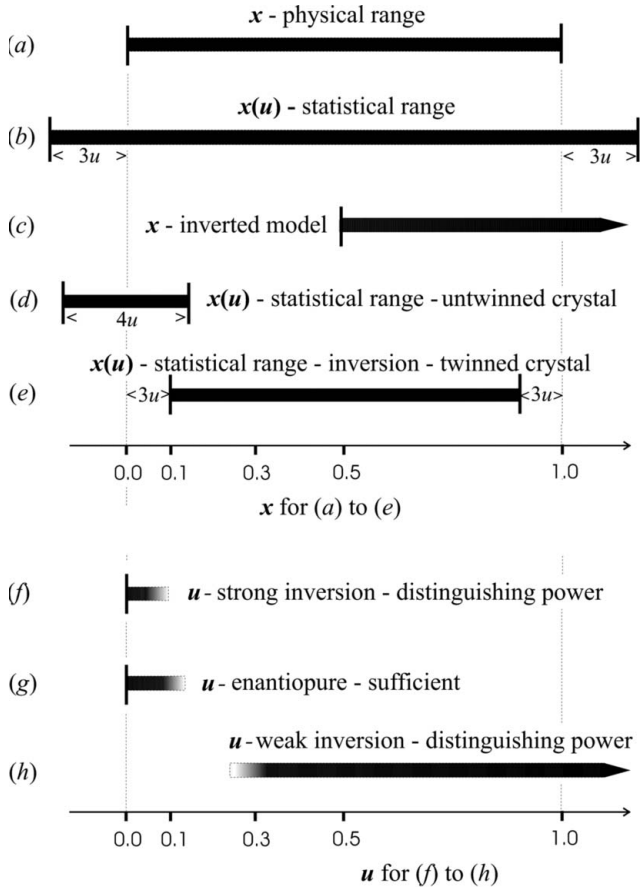

Figure 17
Domains of values of x, its u and the inversion-distinguishing power: (a) the physical domain of x; (b) the statistical domain of x; (c) the domain of x where the crystal and the model are inverted one with respect to the other; (d) the statistical domain of a crystal untwinned by inversion; (e) the statistical domain of a crystal twinned by inversion; (f) the domain of strong inversion-distinguishing power; (g) the domain of enantiopure sufficient inversion-distinguishing power; (h) the domain of weak inversion-distinguishing power. For (f), (g) and (h), the horizontal lines are of varying intensity. In the part of the line that is black, the inversion-distinguishing power may be deduced from the value of u alone. In the part of the line that is grey, the inversion-distinguishing power may not be deduced from the value of u alone. In (b), (d) and (e), arbitrary values of u have been drawn and in practical applications the value of u yielded by the experiment must be used. (Flack & Bernardinelli, 2000.) Reproduced by permission of the authors and IUCr.

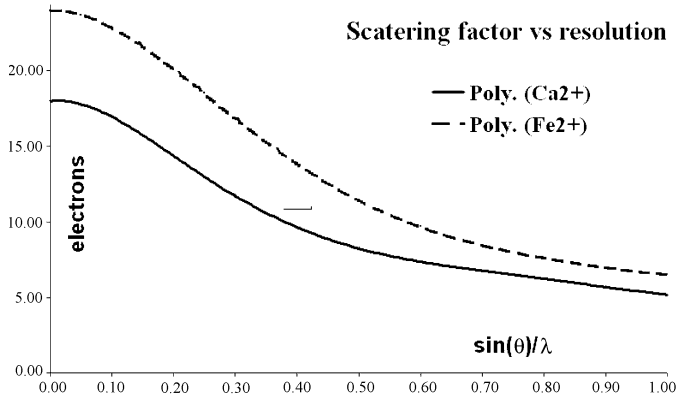

Figure 18
Scattering factors for Ca^{2+} and Fe^{2+} (see §5.1.10). Note that until $\sin(\theta)/\lambda$ is about 0.5, the two curves are almost related by a simple scale factor. An even closer fit could be achieved in this range by applying an ADP factor to one or both of them. This explains why the occupation numbers and ADPs are strongly correlated. Taken over a longer resolution range, the dissimilarity between the curves becomes more evident, so that refined parameters become less correlated.

axes of all of the ellipsoids will be tangential to the circle passing through the F atoms.

(2) Extremely thin disk-like ADPs. These may indicate some kind of two-dimensional disorder, but are more likely to indicate that there is something more seriously wrong with the data or the model, such as twinning.

(3) Physically unrealistic ADPs in a $Z' > 1$ structure. In this case, the Hirshfeld condition is not satisfied for some atom pairs, or the ellipsoid for one or more atoms is unexpectedly small, large or eccentric. As explained above, pseudo-symmetry can have a degrading effect on a refinement, which may need to be controlled by the use of restraints. However, if the pseudo-operator looks exact, it may be that the real and the pseudo-operators have been misassigned. For example, if there is an additional good pseudo-centre of symmetry in $P\bar{1}$, try shifting the origin so as to make the former pseudo-centre into the true centre.

7.5.4. Satellite peaks in the difference Fourier synthesis adjacent to existing atoms. If these occur symmetrically around a heavy atom (*e.g.* a metal) and at unreasonable interatomic distances, they may be due to uncorrected absorption effects. In an F_{obs} or $2F_{obs} - F_{calc}$ map, they could be due to termination-of-series effects if the data in the highest-resolution shell are still appreciably strong. If such peaks are due to disorder, do not put too much reliance on the atomic positions found by an automated peak-search program. These find local maxima, which in the case of a broad and diffuse feature may be unrealistically precise. The best procedure under these circumstances is to look at the contoured three-dimensional electron density maps (see Fig. 20).

Whenever a disordered model is proposed, it is important that it is chemically feasible. It must be possible to construct acceptable patterns of connectivity between the proposed atoms, with acceptable bond lengths and angles. Since disordered atoms are generally only poorly defined by the X-ray data, it is permitted and often desirable to restrain geometric features and ADPs to normal values.

For very highly disordered systems, such as an unsymmetrical solvent molecule disordered over a site with $\bar{3}$ symmetry, solvent molecules disordered in an infinite channel or crystals solvated with a mixture of solvents, modelling with partial atoms can become unrealistically complicated. In this situation, a strong argument can be made for modelling the solvent volume simply with the discrete Fourier transform of the residual electron density in the solvent-accessible volume [*BYPASS* (van der Sluis & Spek, 1990); *PLATON*/*SQUEEZE* (Spek, 2003)]. In equation (1) we saw that the structure factor could be represented as an integral of the electron density, and in equation (3) we represented it as the sum of contributions from resolved atoms. *SQUEEZE* uses a hybrid structure factor expression,

$$A_{hkl} = \sum_j f_j \cos 2\pi(hx + ky + lz) + \int_v \rho_{xyz} \cos 2\pi(hx + ky + lz)\partial v \qquad (63)$$

(where the integration is over the solvent-accessible volume v), with a similar expression for the 'B' part. The first term is a summation over the resolved atoms. The integral in the second term is actually replaced by a summation over small elements from the unresolved solvent-accessible parts of the electron density map. The recovery of the solvent contribution involves an iterative series of difference Fourier maps. The result is a solvent contribution F_{solv} with a phase that differs from the phase of F_{calc}. The *PLATON*/*SQUEEZE* program is independent of any refinement program and integrates smoothly with refinement programs, such as *CRYSTALS*, that can take fixed contributions to the structure factor calculation. Alternatively, a 'solvent-free' F_{obs} (or F^2_{obs}), calculated by subtracting F_{solv} from F_{obs} as complex numbers, can be used for the refinement of the ordered part of the structure with programs that do not have that option. In principle, this

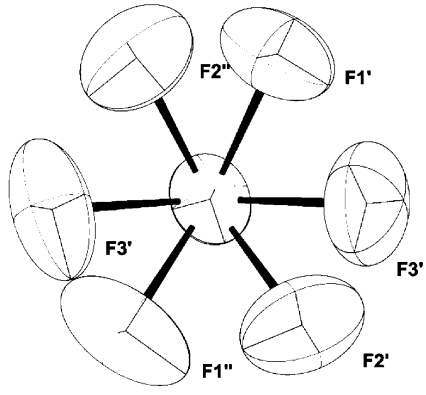

Figure 19
A rotationally disordered CF_3 group. If the data are of good quality and the F atoms have low dynamic disorder, it may be possible to refine them without constraints or restraints. Improbably eccentric ADPs can be controlled by the use of bond-stretching restraints along the C—F directions, and diagonally between the F atoms in each moiety. Weak ADP similarity restraints can be set up between each F atom and its counterpart by disorder.

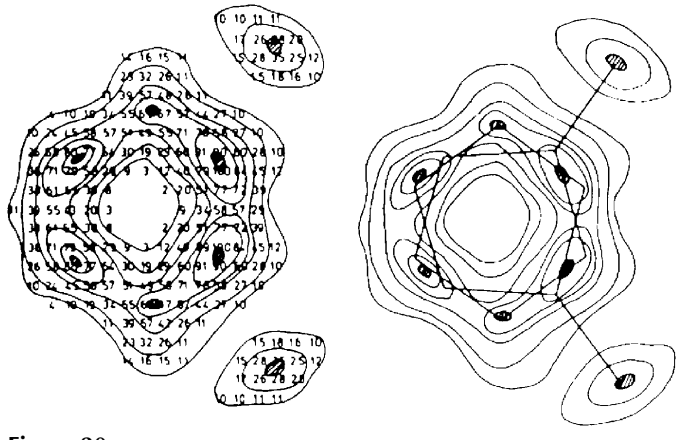

Figure 20
Map Interpretation. The left-hand figure was computer interpreted as a phenyl group with two spurious peaks. The right-hand figure was the manual interpretation as a disordered methylcyclopentadienyl ligand. (Prout & Daran, 1978.) Reproduced by permission of the authors and IUCr.

subtraction process should be iterated. However, in practice this is found to be rarely necessary. The solvent contribution should be added again to F_{calc}, after the convergence of the refinement of the model, for comparison with the original F_{obs}. The result of both approaches is approximately the same.

7.6. Weights and outliers

Initial refinement should always be performed with a simple weighting scheme – unit weights for F, quasi-unit weights for F^2 or 1/variance for either. Once the model is fully parameterized the data should be examined for outliers, by looking at either a table of disagreeable reflections or a plot of F_{obs} *versus* F_{calc}. Fig. 21 shows how a few outliers can spoil the fit of the bulk of the data (top figure).

If there are more than just a few outliers, be suspicious of both the data and the model. If there are several independent estimates of the errant reflections, check that they are in agreement among themselves. If they are in agreement, this may indicate either a systematic error in the data collection procedure or a failure in the model. The output from the data processing software should give an indication of the overall self-consistency of the data (R_{merge} or R_{int}) and may also list groups of reflections having particularly poor self-consistency. If possible, examine the original data. If the data were collected on an area-detector machine, try to look at the actual images.

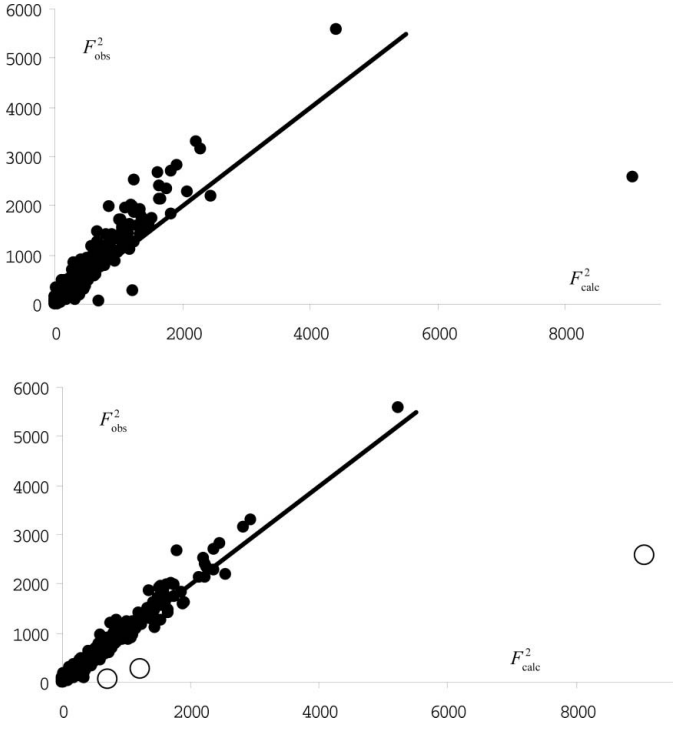

Figure 21
F_{obs}–F_{calc} plots for an organic material. Top image: The structure is fully parameterized, including hydrogen, but the refinement has converged to $R = 15\%$. The model has reasonable bond lengths; the ADPs are a little smaller than expected. There are four reflections where F_{obs} is much lower than F_{calc}. Note that the bulk of the data lie above the line of unit gradient and have a wide spread. Bottom image: Three reflections (circled) have been flagged as outliers because $F_{obs} << F_{calc}$ and omitted from the refinement, which converged to 5% with reasonable ADPs. Note that the bulk of the data now lie close to the line of unit gradient and have less spread. Examination of the data shows that these three reflections occur at $\theta < 4^\circ$ and are probably partially occluded by the beam trap or its support.

Once the model has been finalized and the outliers dealt with, the weights should be optimized. For optimized weights, the overall GoF has virtually no diagnostic use, but the weighted residual, $w\Delta^2$ (where Δ can be in terms of F or F^2), should be examined as a function of F and of resolution. A suitable weighting scheme will have constant $w\Delta^2$ for all systematic rankings of the data (Cruickshank, 1961). If the application of weights causes substantial shifts in the atomic parameters, the weights may need to be reoptimized. Whenever this is done, the results must be carefully scrutinized, since it is possible for repeatedly reoptimized weights to discriminate in favour of a false solution.

7.7. A reasonable model

At the end of an analysis, the model should be assessed against the background of known related structures. If the current structure has totally novel features, consider very seriously the possibility that it is incorrect (Murphy *et al.*, 1998). If possible, impose a more reasonable geometry onto the model and use this to start a new refinement. Alternatively, try solving the structure from scratch using different starting criteria. Failing all else, apply random perturbations to the structure and verify that it refines back to the same final solution.

If the model makes chemical sense but some groups of ADPs seem unexpectedly large, try to ascertain (possibly with the aid of a Dreiding model[6]) if the groups could be in a state of fluxion or libration (rotatory vibration). Libration can be measured by performing a TLS analysis of the U^{ij} parameters of molecular fragments (Schomaker & Trueblood, 1968). **T** is a measure of the overall translational vibration of the group and **L** a measure of the overall libration. Verify that the **T** and **L** subtensors have positive diagonal elements and reasonable magnitudes, and that the U^{ij}s back-computed from **T**, **L** and **S** are substantially the same as the least-squares refined values. If a realistic TLS model cannot be fitted to the anisotropic ADPs, look again for disorder. A large diagonal element in **L** means that there is substantial libration. This can have a large effect on bond lengths, making them appear shorter than they are by a factor much larger than the least-squares standard uncertainty. For a bond lying substantially in the plane of maximum libration, the adjusted bond length is approximately $d_c \simeq d_o(1 + L_{33}/2)$, where d_o is the uncorrected bond length and L_{33} is in radians^2 (Haestier *et al.*, 2008). Fig. 22 shows a ruthenium–Cp* compound in which the Cp* ligand is undergoing substantial libration. Table 3 shows the substantial

[6] Dreiding models are built from modules consisting of an atomic centre with 'bonds' pointing out at angles appropriate for the expected hybridization. The modules can be plugged together to make molecular fragments in which the valence distance and angles are conserved but the dihedral angles are unconstrained.

Table 3
Application of a libration correction to a ruthenium–Cp* compound.

(*a*) The principal axes of the ADPs of a librating Cp* group (Å^2).

C1	0.0393	0.0521	0.1858	
C2	0.0314	0.0616	0.2063	
C3	0.0233	0.0700	0.2641	
C4	0.0266	0.0801	0.2555	
C5	0.0420	0.0842	0.1724	
C11	0.0716	0.0949	0.3856	Might be split
C12	0.0488	0.0962	0.4511	Might be split
C13	0.1059	0.1109	0.6858	Might be split
C14	0.0638	0.1430	0.8415	Might be split
C15	0.0640	0.1104	0.8912	Might be split

(*b*) The **TLS** tensors. Note the value of L_{33}. This large libration results in a significant apparent shortening of the bond lengths.

Centre of gravity, centre of libration	−0.1809	−0.1139	−0.2056
Centre for which **S** is symmetric	−0.0916	−0.0972	−0.1615

L			**T**			**S**		
14.35	0.00	0.00	0.05	0.00	−0.01	0.22	0.00	0.00
0.00	40.14	0.00	0.00	0.09	0.01	0.00	0.19	0.00
0.00	0.00	249.57	−0.01	0.01	0.03	0.00	0.00	−0.03

(*c*) The bond-length adjustments are almost ten times as large as the s.u. computed from the normal matrix.

	Uncorrected	Corrected	Difference
Tangential bonds			
C5—C4	1.416 (5)	1.474	0.058
C5—C1	1.432 (5)	1.492	0.060
C4—C3	1.416 (5)	1.478	0.062
C3—C2	1.414 (5)	1.470	0.056
C2—C1	1.411 (5)	1.473	0.062
Radial bonds			
C15—C5	1.636 (5)	1.705	0.069
C14—C4	1.616 (5)	1.680	0.064
C13—C3	1.611 (5)	1.682	0.071
C12—C2	1.610 (5)	1.676	0.066
C11—C1	1.608 (5)	1.672	0.064

changes in bond lengths that occur when a libration correction is applied (Busing & Levy, 1964).

7.8. Under- and over-parameterization

The IUCr guidelines for the observation:parameter ratio are only guidelines, and the model refined in every structure analysis must be judged on its own merits. If there are very many high-quality X-ray observations, they may easily support the refinement of a very detailed model – for example, the inclusion of a more complex representation of the ADPs than the usual Gaussian (Trueblood *et al.*, 1996).

If the data are poor or sparse, the choice of parameters to be refined becomes more difficult. If more parameters are refined than the data will support, their values risk becoming meaningless – hence Stewart Pawley's comment 'It is often said that with enough parameters you could fit an elephant.' Let us coin the phrase 'elephant parameter' for any model parameter that has no relevance to reality. Now let us ask ourselves the question: how many elephant parameters did we use in our last model refinement (Pawley, 1972)? In macromolecular crystallography, this over-parameterization is frequently called 'over refinement'. As explained in §4.1.5, it is not uncommon for poor or sparse data to require a complex model in order to predict a reasonable fit to the observations. Examples of complexity include resolvable positional disorder and very aspherical anisotropic ADPs. If, under free refinement, these give models with physically unacceptable parameters or derived parameters (*e.g.* weird bond lengths) accompanied by large standard uncertainties, the recommended procedure is to use restraints to regularize the model. The restraints should be chosen on the basis of sound physical arguments and should be assigned realistic uncertainties. In general, similarity restraints are easier to justify than actual numerical target values. Table 4 shows the consequences of applying reasonable and unreasonable restraints to one of the phenyl groups in tetraphenylene (which, forming good quality stable crystals in a centred space group and having chemical fourfold symmetry,

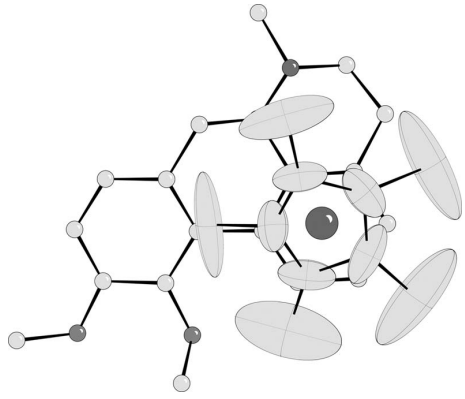

Figure 22
The Cp* ligand, which lies above the Ru atom, is clearly undergoing librational motion. This leads to a shortening of both the radial and the tangential bonds, as determined with X-rays. The correction for the shortening is about ten times the s.u. values of the bonds as determined by the least-squares refinement.

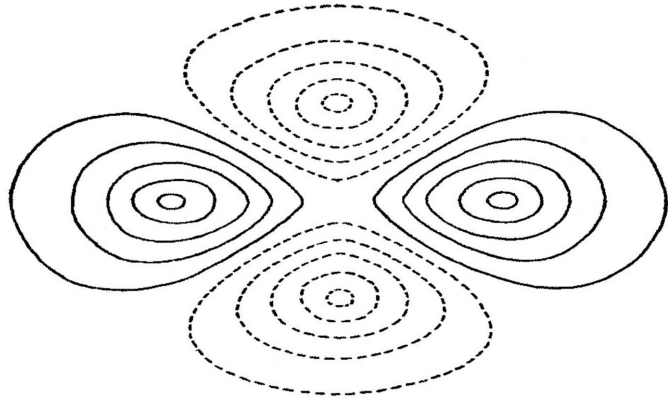

Figure 23
Idealized difference electron density through the site of a heavy atom that was refined with an isotropic ADP when an anisotropic model would have been more appropriate. The negative (dotted) lobes indicate that the model has too much density in these regions.

Table 4
F^2 refinements of tetraphenylene with 4146 observations and 218 parameters (distance and s.u. values in Å).

In the free refinement, a pair of chemically equivalent bonds differ in length by 0.004 Å. Restraining them to be the same with a reasonable standard uncertainty on the difference has no appreciable effect. Requesting an s.u. of 0.001 Å makes the two bonds very similar but is slightly in conflict with the X-ray data – the *R* factor rises. Restraining them to the unrealistic bond length of 1.29 Å with a reasonable s.u. has little effect on the model, and the residual $(1.29 - D_{calc}) = 0.10$ greatly exceeds the requested s.u. of 0.01 Å, showing that the restraint is inappropriate. When the requested s.u. is reduced to 0.001, this has a serious effect on the model. The restraint is now obeyed, with the consequence that the X-ray *R* factor rises substantially.

Type	s.u.	*R*	*Rw*	C1—C2	C1—C6	Mean	Δ	Residual
Free	–	4.16	10.50	1.3928 (19)	1.4011 (18)	1.39695	0.00415	–
Modest Mean	0.01	4.16	10.50	1.3928 (18)	1.4010 (18)	1.39690	0.00410	0.00410
Harsh Mean	0.0001	4.20	10.61	1.39691 (10)	1.39694 (10)	1.39693	0.00002	0.00002
Modest 1.29	0.01	4.16	10.52	1.391 (2)	1.3992 (19)	1.39510	0.00410	0.10092
Harsh 1.29	0.0001	12.17	29.28	1.29005 (10)	1.29006 (10)	1.29006	0.00001	0.00006

is an excellent material for testing data-collection and refinement strategies). There are copious good data, so that a reasonable restraint has no impact. An unreasonable restraint with a tiny uncertainty causes the model to come into conflict with the X-ray data, so that the *R* factor rises.

Under-parameterization is the opposite situation, in which the model is too simple, for example refining a routine organometallic compound using only isotropic ADPs or omission of a solvent molecule. Symptoms of under-parameterization may include unexpectedly high conventional *R* factors, structured noise in a difference map (Fig. 23) or a systematic trend in $F_{obs} - F_{calc}$. In addition, the trial structure may appear distorted or have unrealistic ADPs as the refinement tries to adjust the available parameters to compensate for the missing parameters.

8. Conclusion

In the introduction we said that there is no simple recipe for refining difficult structures. This paper has described some commonly available tools and hopefully provided enough theoretical background to enable them to be applied successfully. At all times, it must be remembered that the atomic model is only a model and that there may be several marginally different models that yield structure amplitudes in fair agreement with the observed values. If we had more reliable estimates of the errors in the data, and an understanding of the consequences of shortcomings in the model, purely statistical tests might select between alternative models. Even if we have good statistical information, final distinctions have to be made on the basis of experience and the general consensus of the crystallographic community. This may sound dangerously close to 'chi-by-eye' (Press *et al.*, 2005), but it is perhaps better than putting blind faith in insecure statistical inference. If a structure 'looks wrong', it probably is wrong. The converse is not necessarily true. Harlow (1996) divides refined structures into four classes – 'quality structures', which are the gold standard analyses obtained by careful work on very good crystals; 'fuzzy structures', which are the normal run-of-the-mill products from routine analytical work; 'incorrect structures', which are ones where a fundamental error has been made; and finally 'junk' structures. This final class could be subdivided. A structure could satisfy Harlow's 'junk' criteria simply as a result of careless work, in which case the result is worthless. However, if the sample preparation and data collection have been carefully performed, failure to refine to a conventionally fuzzy structure (or better) is an indication that something unusual is happening in the diffraction process, which may be worth reporting and reinvestigating. Nature is not always so obliging that she invariably follows the laws that we make. Watkin (1994) points out that the conventional space groups are just mathematical points in what is almost certainly a continuum. Modulated, incommensurate and quasi-crystals stretch the scientific imagination. The problem for the journals is to try to distinguish between good work on bad crystals and bad work on good crystals. If all you have is a CIF, the two cases must look very similar.

This paper is largely a collection of other people's flowers, selected to show the sometimes conflicting ideas about an optimal strategy for structure refinement. I am very grateful to the referees for their helpful comments, and to many colleagues for their suggestions for improvements and additions.

References

Agrosi, G., Schingaro, E., Pedrazzi, G., Scandale, E. & Scordari, F. (2002). *Eur. J. Mineral.* **14**, 785–794.
Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Altomare, A., Burla, M. C., Cascarano, G., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G. & Polidori, G. (1995). *Acta Cryst.* A**51**, 305–309.
Altomare, A., Cascarano, G., Giacovazzo, C., Guagliardi, A., Burla, M. C., Polidori, G. & Camalli, M. (1994). *J. Appl. Cryst.* **27**, 435.
Arnberg, L., Hovmöller, S. & Westman, S. (1979). *Acta Cryst.* A**35**, 497–499.
Atkinson, A. C. (1987). *Plots, Transformations and Regression*. Oxford: Oxford Science Publications.
Bernardinelli, G. & Flack, H. D. (1985). *Acta Cryst.* A**41**, 500–511.
Betteridge, P. W., Carruthers, J. R., Cooper, R. I., Prout, K. & Watkin, D. J. (2003). *J. Appl. Cryst.* **36**, 1487.
Blanc, E. & Paciorek, W. (2001). *J. Appl. Cryst.* **34**, 480–483.
Blanc, E., Schwarzenbach, D. & Flack, H. D. (1991). *J. Appl. Cryst.* **24**, 1035–1041.
Blessing, R. H. (1995). *Acta Cryst.* A**51**, 33–38.
Bruno, I. J., Cole, J. C., Kessler, M., Luo, J., Motherwell, W. D. S., Purkis, L. H., Smith, B. R., Taylor, R., Cooper, R. I., Harris, S. E. & Orpen, A. G. (2004). *J. Chem. Inf. Comput. Sci.* **44**, 2133–2144.
Burla, M. C., Caliandro, R., Camalli, M., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., Polidori, G. & Spagna, R. (2005). *J. Appl. Cryst.* **38**, 381–388.

Burzlaff, H., Bohme, R. & Gomm, M. (1978). *Computing in Crystallography*, edited by H. Schenk, R. Olthof-Hazekamp, H. van Koningsveld & G. C. Bassi, pp. 75–79. Delft University Press.
Busing, W. R. & Levy, H. A. (1964). *Acta Cryst.* **17**, 142–146.
Coehlo, A. (2004). *TOPAS-Academic.* Bruker AXS, Karlsruhe, Germany.
Cooper, M. J. & Rouse, K. D. (1968). *Acta Cryst.* A**24**, 405–410.
Cooper, R. I., Gould, R. O., Parsons, S. & Watkin, D. J. (2002). *J. Appl. Cryst.* **35**, 168–174.
Cowtan, K. & Ten Eyck, L. F. (2000). *Acta Cryst.* D**56**, 842–856.
Cruickshank, D. W. J. (1950). *Acta Cryst.* **3**, 10–13.
Cruickshank, D. W. J. (1961). *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 43–46. London: Pergamon.
Cruickshank, D. W. J. (1969). *Crystallographic Computing*, edited by F. R. Ahmed, pp. 187–197. Copenhagen: Munksgaard.
Cruickshank, D. W. J., Sime, J. G., Smith, J. G. F., Truter, W. R., Wells, M., Rollett, J. S. & Freeman, H. C. (1964). *Edition 1*. Computing Laboratory, University of Oxford, UK.
Debaerdemaeker, T., Germain, G., Main, P., Tate, C. & Woolfson, M. M. (1987). *MULTAN87*. University of York, UK.
De Titta, G. T., Edmonds, J. W., Langs, D. A. & Hauptman, H. (1975). *Acta Cryst.* A**31**, 472–479.
Dunitz, J. D. (1964). *Acta Cryst.* **17**, 1299–1304.
Dunitz, J. D. (1995). *X-ray Analysis and the Structure of Organic Molecules*. Basel: Cornell University Press, Verlag HCA.
Dunitz, J. D. & Seiler, P. (1973). *Acta Cryst.* B**29**, 589–595.
Edwards, A. W. F. (1992). *Likelihood*. Baltimore: The John Hopkins University Press.
Flack, H. D. (1983). *Acta Cryst.* A**39**, 876–881.
Flack, H. D. & Bernardinelli, G. (2000). *J. Appl. Cryst.* **33**, 1143–1148.
Flack, H. D. & Bernardinelli, G. (2006). *Inorg. Chim. Acta*, **359**, 383–387.
Flack, H. D. & Schwarzenbach, D. (1988). *Acta Cryst.* A**44**, 499–506.
Gavezzotti, A. & Flack, H. D. (2005). http://www.iucr.org/iucr-top/comm/cteach/pamphlets/21/index.html.
Giacovazzo, C., Monaco, H. L., Artoli, G., Veterbo, D., Ferraris, G., Gilli, G., Zanotti, G. & Catti, M. (2002). *Fundamentals of Crystallography*, 2nd ed., p. 105. Oxford: Oxford Science Publications.
Gill, P. E., Murray, W. & Wright, M. H. (1981). *Practical Optimisation*. London: Academic Press.
Glusker, J. P., Lewis, M. & Rossi, M. (1994). *Crystal Structure Analysis for Chemists and Biologists*. New York: VCH Publishing Inc.
Haestier, J., Sadki, M., Thompson, A. L. & Watkin, D. J. (2008). *J. Appl. Cryst.* **41**, 531–536.
Harding, C. C., Watkin, D. J., Cowley, A. R., Soengas, R., Skytte, U. P. & Fleet, G. W. J. (2005). *Acta Cryst.* E**61**, o250–o252.
Harlow, R. (1996). *J. Res. Natl Inst. Technol.* **101**, 327–339.
Harlow, R. (1998). *The Hydrogen Challenge*. http://www.pitt.edu/~geib/challenge.html.
Harris, G. W. & Moss, D. S. (1992). *Acta Cryst.* A**48**, 42–45.
Herbstein, F. H. (2000). *Acta Cryst.* B**56**, 547–557.
Herrendorf, W. (1993). *HABITUS*. University of Karlsruhe, Germany.
Hirshfeld, F. L. (1976). *Acta Cryst.* A**32**, 239–244.
Hirshfeld, F. L. & Rabinovich, D. (1973). *Acta Cryst.* A**29**, 510–513.
Hodgson, L. I. & Rollett, J. S. (1963). *Acta Cryst.* **16**, 329–335.
Hoppe, W., Gassman, J. & Zechmeister, K. (1970). *Crystallographic Computing*, edited by F. R. Ahmed, pp. 26–36. Copenhagen: Munksgaard.
Hughes, E. W. (1941). *J. Am. Chem. Soc.* **63**, 1737–1752.
Huml, K. (1980). *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Ventatesan, pp. 12.01–12.22. Bangalore: Indian Academy of Sciences.
Irmer, E. (1990). PhD thesis, University of Göttingen, Germany.
Kassner, D., Baur, W. H., Joswig, W., Eichhorn, K., Wendschuh-Josties, M. & Kupčik, V. (1993). *Acta Cryst.* B**49**, 646–654.
Kirschbaum, K., Martin, A. & Pinkerton, A. A. (1997). *J. Appl. Cryst.* **30**, 514–516.
Larson, A. C. (1970). *Crystallographic Computing*, edited by F. R. Ahmed. Copenhagen: Munksgaard.
Larson, A. C. (1980). *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 11.01–11.07. Bangalore: Indian Academy of Sciences.
Lenstra, A. T. H. & Kataeva, O. N. (2001). *Acta Cryst.* B**57**, 497–506.
Lipson, H. & Cochran, W. (1966). *The Determination of Crystal Structures*, 3rd ed., p. 347. London: G. Bell and Sons Ltd.
Lozano-Casal, P., Allan, D. R. & Parsons, S. (2005). *Acta Cryst.* B**61**, 717–723.
Madsen, A. Ø., Sørensen, H. O., Flensburg, C., Stewart, R. F. & Larsen, S. (2004). *Acta Cryst.* A**60**, 550–561.
Markvardsen, A. J., Shankland, K., David, W. I. F. & Didlick, G. (2005). *J. Appl. Cryst.* **38**, 107–111.
Marsh, R. E. & Spek, A. L. (2001). *Acta Cryst.* B**57**, 800–805.
McCusker, L. B., Baerlocher, C., Grosse-Kunstleve, R., Brenner, S. & Wessels, T. (2001). *Chimia*, **55**, 497–504.
Meier, W. M. & Villiger, H. (1969). *Z. Kristallogr.* **129**, 411–423.
Merli, M., Ungaretti, L. & Oberti, R. (2000). *Am. Mineral.* **85**, 532–542.
Müller, P., Herbst-Irmer, R., Spek, A. L., Schneider, T. R. & Sawaya, M. R. (2006). *Crystal Structure Refinement*, edited by P. Müller, p. 10. Oxford: Oxford Science Publications.
Murphy, V. J., Rabinovich, D., Hascall, T., Klooster, W. T., Koetzle, T. F. & Parkin, G. (1998). *J. Am. Chem. Soc.* **120**, 4372–4387.
Nicholson, W. L., Prince, E., Buchanan, J. & Tucker, P. (1982). *Crystallographic Statistics*, edited by S. Ramaseshan, M. F. Richardson & A. J. C. Wilson, pp. 229–263. Bangalore: Indian Academy of Sciences.
Nocedal, J. & Wright, S. J. (1999). *Numerical Optimization.* Springer Series in Operations Research. New York: Springer.
Oszlányi, G. & Sütő, A. (2004). *Acta Cryst.* A**60**, 134–141.
Parkin, S. (2000). *Acta Cryst.* A**56**, 157–162.
Pawley, G. S. (1966). *Acta Cryst.* **20**, 631–638.
Pawley, S. (1972). NATO Advanced Study Institute: Experimental Aspects of X-ray and Neutron Single Crystal Diffraction Methods, Chemistry Department, Aarhus University, Denmark, 31 July–11 August.
Petricek, V., Dusek, M. & Palatinus, L. (2000). *JANA2000*. Institute of Physics, Prague, Czech Republic.
Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (2002). *Numerical Recipes in C++*, pp. 686–688. Cambridge University Press.
Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (2005). *Numerical Recipes in C++*, p. 662. Cambridge University Press.
Prince, E. (1994*a*). *Mathematical Techniques in Crystallography & Material Science*, 2nd ed., pp. 79–82. Berlin: Springer–Verlag.
Prince, E. (1994*b*). *Mathematical Techniques in Crystallography & Material Science*, 2nd ed., p. 115. Berlin: Springer–Verlag.
Prince, E. (1994*c*). *Mathematical Techniques in Crystallography & Material Science*, 2nd ed., p. 121. Berlin: Springer–Verlag.
Prince, E. & Nicholson, W. L. (1985). *Structure and Statistics in Crystallography*, edited by A. J. C. Wilson, p. 191. New York: Adenine Press.
Pringle, D. & Shen, Q. (2003). *J. Appl. Cryst.* **36**, 29.
Prout, K. & Daran, J.-C. (1978). *Acta Cryst.* B**34**, 3586–3591.
Qurashi, M. M. & Vand, V. (1953). *Acta Cryst.* **6**, 341–349.
Rollett, J. S. (1964). *Computing Methods in Crystallography*, edited by J. S. Rollett, pp. 19–21. Oxford: Pergamon Press.
Rollett, J. S. (1970). *Crystallographic Computing*, edited by F. R. Ahmed, pp. 169–175. Copenhagen: Munksgaard.
Rollett, J. S. (1984). *Methods and Applications in Crystallographic Computing*, edited by S. R. Hall & T. Ashida. Oxford: Clarendon Press.

Rollett, J. S. (1988). *Crystallographic Computing 4*, edited by N. W. Isaacs & M. R. Taylor, pp. 149–166. Oxford University Press.
Rollett, J. S., McKinlay, T. G. & Haigh, P. N. (1976). *Crystallographic Computing Techniques*, edited by F. R. Ahmed, pp. 413–419. Copenhagen: Munksgaard.
Rousseeuw, P. J. & Leroy, A. M. (2003). *Robust Regression and Outlier Detection*. New Jersey: Wiley.
Schomaker, V. & Marsh, R. E. (1979). *Acta Cryst.* B**35**, 1933–1934.
Schomaker, V. & Trueblood, K. N. (1968). *Acta Cryst.* B**24**, 63–76.
Schwarzenbach, D., Abrahams, S. C., Flack, H. D., Gonschorek, W., Hahn, Th., Huml, K., Marsh, R. E., Prince, E., Robertson, B. E., Rollett, J. S. & Wilson, A. J. C. (1989). *Acta Cryst.* A**45**, 63–75.
Seiler, P., Schweizer, W. B. & Dunitz, J. D. (1984). *Acta Cryst.* B**40**, 319–327.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Sivia, D. S. (1996). *Data Analysis – a Bayesian Tutorial*. Oxford University Press.
Sluis, P. van der & Spek, A. L. (1990). *Acta Cryst.* A**46**, 194–201.
Sparks, R. A. (1961). *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 170–187. London: Pergamon Press.
Spek, A. L. (2003). *J. Appl. Cryst.* **36**, 7–13.
Tronrud, D. E. (2004). *Acta Cryst.* D**60**, 2156–2168.
Trueblood, K. N., Bürgi, H.-B., Burzlaff, H., Dunitz, J. D., Gramaccioli, C. M., Schulz, H. H., Shmueli, U. & Abrahams, S. C. (1996). *Acta Cryst.* A**52**, 770–781.
Urzhumtsev, A. G. (1991). *Acta Cryst.* A**47**, 723–727.
Vigante, B., Ozols, Y., Mishnev, A., Duburs, G. & Chekavichus, B. (2000). *Khim. Geterotsikl. Soedin. SSSR (Russ.) Chem. Heterocycl. Compd.* p. 978.
Walker, N. & Stuart, D. (1983). *Acta Cryst.* A**39**, 158–166.
Waser, J. (1963). *Acta Cryst.* **16**, 1091–1094.
Watkin, D. J. (1980). *Acta Cryst.* A**36**, 975–978.
Watkin, D. J. (1988). *Crystallographic Computing 4*, edited by N. W. Isaacs & M. R. Taylor, pp. 111–125. Oxford University Press.
Watkin, D. (1994). *Acta Cryst.* A**50**, 411–437.
Watkin, D. (2000). *Acta Cryst.* B**56**, 747–749.
Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.
Wilson, A. J. C. (1976). *Acta Cryst.* A**32**, 994–996.
Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.
Zachariasen, W. H. (1965). *Acta Cryst.* **18**, 714–716.
Zimmerman, H. E. & St Clair, J. D. (1989). *J. Org. Chem.* **54**, 2125–2137.