The main difference from the method of Fox & Holmes (1966) is the introduction of the weights $w_{hl\alpha}$. These weights depend upon the distance between each reflection hl and the positions α. They are monotonically decreasing functions of this distance, implemented as Gaussians in *XDS* and *XSCALE*. This results in a smoothing of the scaling factors since each reflection contributes to the observational equations in proportion to the weights $w_{hl\alpha}$.

Minimization of Ψ is done iteratively. After each step, the g_α are replaced by $\Delta g_\alpha + g_\alpha$ and rescaled to a mean value of 1. The corrections Δg_α are determined from the normal equations

$$\sum_\beta A_{\alpha\beta}\Delta g_\beta = b_\alpha,$$

where

$$\begin{aligned}
A_{\alpha\beta} &= \sum_h [\delta_{\alpha\beta} I_h^2 u_{h\alpha} + (r_{h\alpha}v_{h\beta} + v_{h\alpha}r_{h\beta} - v_{h\alpha}v_{h\beta})/u_h] \\
b_\alpha &= \sum_h I_h r_{h\alpha} \\
I_h &= v_h/u_h \\
r_{h\alpha} &= v_{h\alpha} - g_\alpha u_{h\alpha} I_h \\
u_{h\alpha} &= \sum_l w_{hl\alpha}/\sigma_{hl}^2 \\
v_{h\alpha} &= \sum_l w_{hl\alpha} I_{hl}/\sigma_{hl}^2 \\
u_h &= \sum_\alpha g_\alpha^2 u_{h\alpha} \\
v_h &= \sum_\alpha g_\alpha v_{h\alpha}.
\end{aligned}$$

In case a 'true' intensity I_h is available from a reference data set, the non-diagonal elements are omitted from the sum over h in the normal matrix $A_{\alpha\beta}$. The corrections Δg_α are expanded in terms of the eigenvectors of the normal matrix, thereby avoiding shifts along eigenvectors with very small eigenvalues (Diamond, 1966). This filtering method is essential since the normal matrix has zero determinant if no reference data set is available.

11.3.5. Post refinement

The number of fully recorded reflections on each single image rapidly declines for small oscillation ranges and the complete intensities of the partially recorded reflections have to be estimated. This presented a serious obstacle in early structural work on virus crystals, as the crystal had to be replaced after each exposure on account of radiation damage. A solution of this problem, the 'post refinement' technique, was found by Schutt, Winkler and Harrison, and variants of this powerful method have been incorporated into most data-reduction programs [for a detailed discussion, see Harrison *et al.* (1985); Rossmann (1985)]. The method derives complete intensities of reflections only partially recorded on an image from accurate estimates for the fractions of observed intensity, the 'partiality'. The partiality of each reflection can always be calculated as a function of orientation, unit-cell metric, mosaic spread of the crystal and model intensity distributions. Obviously, the accuracy of the estimated full reflection intensity then strongly depends on a precise knowledge of the parameters describing the diffraction experiment. Usually, for many of the partial reflections, symmetry-related fully recorded ones can be found, and the list of such pairs of intensity observations can be used to refine the required parameters by a least-squares procedure. Clearly, this refinement is carried out after all images have been processed, which explains why the procedure is called 'post refinement'.

Adjustments of the diffraction parameters s_μ $(\mu = 1, \ldots, k)$ are determined by minimization of the function E, which is defined as the weighted sum of squared residuals between calculated and observed partial intensites.

$$\begin{aligned}
E &= \sum_{hj} w_{hj}(\Delta_{hj})^2 \\
\Delta_{hj} &= R_j(\varphi_{hj}) g_j I_h - I_{hj} \\
w_{hj} &= 1/\{\sigma^2(I_{hj}) + [R_j(\varphi_{hj})g_j]^2\sigma^2(I_h)\}.
\end{aligned}$$

Here, I_{hj} is the intensity recorded on image j of a partial reflection with indices summarized as hj, I_h is the mean of the observed intensities of all fully recorded reflections symmetry-equivalent to hj, g_j is the inverse scaling factor of image j, φ_{hj} is the calculated spindle angle of reflection hj at diffraction and R_j is the computed fraction of total intensity recorded on image j.

Expansion of the residuals Δ_{hj} to first order in the parameter changes δs_μ and minimization of $E(\delta s_\mu)$ leads to the k normal equations

$$\sum_{\mu'=1}^{k}\left(\sum_{hj} w_{hj}\frac{\partial\Delta_{hj}}{\partial s_\mu}\frac{\partial\Delta_{hj}}{\partial s_{\mu'}}\right)\delta s_{\mu'} = -\sum_{hj} w_{hj}\Delta_{hj}\frac{\partial\Delta_{hj}}{\partial s_\mu}.$$

Often, the normal matrix is ill-conditioned, since changes in some unit-cell parameters or small rotations of the crystal about the incident X-ray beam do not significantly affect the calculated partiality R_j. To take care of these difficulties, the system of equations is rescaled to yield unit diagonal elements for the normal matrix and the correction vector δs_μ is filtered by projection into a subspace defined by the eigenvectors of the normal matrix with sufficiently large eigenvalues (Diamond, 1966).

The parameters are corrected by the filtered δs_μ and a new cycle of refinement is started until a minimum of E is reached. The weights, residuals and their gradients are calculated using the current values for s_μ and g_j at the beginning of each cycle. The derivatives

$$\frac{\partial\Delta_{hj}}{\partial s_\mu} = g_j I_h\left(\frac{\partial R_j}{\partial\varphi_{hj}}\frac{\partial\varphi_{hj}}{\partial s_\mu} + \frac{\partial R_j}{\partial\sigma_M}\frac{\partial\sigma_M}{\partial s_\mu} + \frac{\partial R_j}{\partial|\zeta_{hj}|}\frac{\partial|\zeta_{hj}|}{\partial s_\mu}\right)$$

appearing in the normal equations can be worked out from the definitions given in Sections 11.3.2.2 and 11.3.2.4 (to simplify the following equations, the subscript hj is omitted). The fraction R_j of the total intensity can be expressed in terms of the error function (see Section 11.3.2.4) as

$$\begin{aligned}
R_j &= [\mathrm{erf}(z_1) - \mathrm{erf}(z_2)]/2 \\
z_1 &= |\zeta|(\varphi_0 + j\Delta_\varphi - \varphi)/(2)^{1/2}\sigma_M \\
z_2 &= |\zeta|[\varphi_0 + (j-1)\Delta_\varphi - \varphi]/(2)^{1/2}\sigma_M.
\end{aligned}$$

Using the relation $\mathrm{d}\,\mathrm{erf}(z)/\mathrm{d}z = [2/(\pi)^{1/2}]\exp(-z^2)$, the derivatives of R_j are

$$\begin{aligned}
\partial R_j/\partial\varphi &= [\exp(-z_2^2) - \exp(-z_1^2)]|\zeta|/[\sigma_M(2\pi)^{1/2}] \\
\partial R_j/\partial\sigma_M &= [z_2\exp(-z_2^2) - z_1\exp(-z_1^2)]/[\sigma_M(\pi)^{1/2}] \\
\partial R_j/\partial|\zeta| &= [z_1\exp(-z_1^2) - z_2\exp(-z_2^2)]/[|\zeta|(\pi)^{1/2}].
\end{aligned}$$

It remains to work out the derivatives $\partial\varphi/\partial s_\mu$, $\partial\sigma_M/\partial s_\mu$ and $\partial|\zeta|/\partial s_\mu$ (not shown here). As discussed in detail by Greenhough & Helliwell (1982), spectral dispersion and asymmetric beam cross fire lead to some variation of σ_M, which makes it necessary to include additional parameters in the list s_μ. The effect of these parameters on the partiality is dealt with easily by the derivatives $\partial\sigma_M/\partial s_\mu$.

The refinement scheme described above requires initial scaling factors g_j. With the now improved estimates for the partialities R_j, a new set of scaling factors can be obtained by the method outlined in Section 11.3.4. This alternating procedure of scaling and post refinement usually converges within three cycles.

The use of error functions for modelling partiality, as implicated by a Gaussian model for describing spot shape, was chosen here for reasons of conceptual simplicity and coherence. This choice is unlikely to alter significantly the results of post refinement that are based on other functions of similar form [see the discussion by Rossmann (1985)].

11.3.6. Space-group assignment

Identification of the correct space group is not always an easy task and should be postponed for as long as possible. Fortunately, all data processing as implemented in the program *XDS* can be carried out even in the absence of any knowledge of crystal symmetry and cell constants. In this case, a reduced cell is extracted from the observed diffraction pattern and processing of the data images continues to completion as if the crystal were triclinic. Clearly, the reflection indices then refer to the reduced cell and must be reindexed once the space group is known. For all space groups, the required reindexing transformation is linear and involves only whole numbers as shown in Part 9 of IT *A*. The following description and example are taken from Kabsch (1993).

Space-group assignment is carried out in two steps under control of the crystallographer once integrated intensities of all reflections are available. First, the Bravais lattices that are compatible with the observed reduced cell are identified. In the second step, any of the plausible space groups may be tested and rated according to symmetry *R* factors and systematic absences of integrated reflection intensities after reindexing. Additional acceptance criteria are obtained from refinement, now using a reduced set of independent parameters describing the conventional unit cell which should not lead to a significant increase of r.m.s. deviations between observed and calculated reflection positions and angles.

11.3.6.1. *Determination of the Bravais lattice*

The determination of possible Bravais lattices is based upon the concept of the reduced cell whose metric parameters characterize 44 lattice types as described in Part 9 of IT *A*. A primitive basis $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ of a given lattice is defined there as a reduced cell if it is right-handed and if the components of its metric tensor

$$A = \mathbf{b}_1 \cdot \mathbf{b}_1, \quad B = \mathbf{b}_2 \cdot \mathbf{b}_2, \quad C = \mathbf{b}_3 \cdot \mathbf{b}_3,$$
$$D = \mathbf{b}_2 \cdot \mathbf{b}_3, \quad E = \mathbf{b}_1 \cdot \mathbf{b}_3, \quad F = \mathbf{b}_1 \cdot \mathbf{b}_2$$

satisfy a number of conditions (inequalities). The main conditions state that the basis vectors are the shortest three linear independent lattice vectors with either all acute or all non-acute angles between them. As specified in IT *A*, each of the 44 lattice types is characterized by additional equality relations among the six components of the reduced-cell metric tensor. As an example, for lattice character 13 (Bravais type *oC*) the components of the metric tensor of the reduced cell must satisfy

$$A = B, \quad B \le C, \quad D = 0, \quad E = 0, \quad 0 \le -F \le A/2.$$

Any primitive triclinic cell describing a given lattice can be converted into a reduced cell. It is well known, however, that the reduced cell thus derived is sensitive to experimental error. Hence, the direct approach of first deriving the correct reduced cell and then finding the lattice type is unstable and may in certain cases even prevent the identification of the correct Bravais lattice.

A suitable solution of the problem has been found that avoids any decision about what the 'true' reduced cell is. The essential requirements of this procedure are: (*a*) a database of possible reduced cells and (*b*) a backward search strategy that finds the best-fitting cell in the database for each lattice type.

The database is derived from a seed cell which strictly satisfies the definitions for a reduced cell. All cells of the same volume as the seed cell whose basis vectors can be linearly expressed in terms of the seed vectors by indices -1, 0, or $+1$ are included in the database. Each unit cell in the database is considered as a potential reduced cell even though some of the defining conditions as given in Part 9 of IT *A* may be violated. These violations are treated as being due to experimental error.

The backward search strategy starts with the hypothesis that the lattice type is already known and identifies the best-fitting cell in the database of possible reduced cells. Contrary to a forward directed search, it is now always possible to decide which conditions have to be satisfied by the components of the metric tensor of the reduced cell. The total amount by which all these equality and inequality conditions are violated is used as a quality index. This measure is defined below for lattice type 13 *oC* testing a potential reduced cell $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ from the database for agreement. Positive values of the quality index p_{13} indicate that some conditions are not satisfied.

$$p_{13}(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3) = |A - B| + \max(0, B - C) + |D| + |E| + \max(0, F) + \max(0, -F - A/2).$$

All potential reduced cells in the database are tested and the smallest value for p_{13} is assigned to lattice type 13. This test is carried out for all 44 possible lattice types using quality indices derived in a similar way from the defining conditions as listed in Part 9 of IT *A*. For each of the 44 lattice types thus tested, the procedure described here returns the quality index, the conventional cell parameters and a transformation matrix relating original indices with respect to the seed cell to the new indices with respect to the conventional cell. These index-transformation matrices are derived from those given in Table 9.3.1 in IT *A*.

The results obtained by this method are shown in Table 11.3.6.1 for the example of a 1.5° oscillation data film containing 1313 strong diffraction spots which were located automatically. The space group of the crystal is $C222_1$ and the cell constants are $a = 72.9, b = 100.1, c = 92.6$ Å. The entry for the correct Bravais lattice *oC* with derived cell constants close to the true ones has a low value for its quality index and thus appears as a possible explanation of the observed diffraction pattern.

11.3.6.2. *Finding possible space groups*

Inspection of the table rating the likelihood of each of the 44 lattice types usually reveals a rather limited set of possible space groups. Furthermore, the absence of parity-changing symmetry operators required for protein crystals restricts the number of possible space groups to 65 instead of 230. Any space group can be tested by repeating only the final steps of data processing. These steps include a comparison of symmetry-related reflection intensities, as well as a refinement of the parameters controlling the diffraction pattern after reindexing the reflections by the appropriate transformation. Low r.m.s. deviations between the observed and refined spot positions, as well as small *R* factors for symmetry-related reflection intensities, indicate that the constraints imposed by the tentatively chosen space group are satisfied. The space group with highest symmetry compatible with the data is almost certainly correct if the data set is sufficiently complete and redundant, which requires that each symmetry element relates a sufficient number of reflections to one another.

For the example of a 1.5° oscillation data film given above, space-group determination consists of the following steps. Inspection of Table 11.3.6.1 indicates that lattice characters 10, 13, 14 and 34, besides the triclinic characters 31 and 44, are approximately