International Tables for Crystallography (2006). Vol. F, Section 11.3.3, pp. 221–222.

$$E = w_X \sum_{i=1}^{n} (\Delta_X^i)^2 + w_Y \sum_{i=1}^{n} (\Delta_Y^i)^2 + w_Z \sum_{i=1}^{n} (\Delta_Z^i)^2.$$

The residuals between the calculated (X_i, Y_i, Z_i) and observed spot centroids are

$$\begin{aligned}
\Delta_X^i &= X_i - X_i' = X_0 + F\mathbf{S}_i \cdot \mathbf{d}_1 / \mathbf{S}_i \cdot \mathbf{d}_3 - X_i' \\
\Delta_Y^i &= Y_i - Y_i' = Y_0 + F\mathbf{S}_i \cdot \mathbf{d}_2 / \mathbf{S}_i \cdot \mathbf{d}_3 - Y_i' \\
\Delta_Z^i &= Z_i - Z_i' = \varphi_0 + \Delta_\varphi \sum_{j=-\infty}^{\infty} (j - 1/2) R_j^i - Z_i'.
\end{aligned}$$

Let s_μ $(\mu = 1, \ldots, k)$ denote the k independent parameters for which initial estimates are available. Expanding the residuals to first order in the parameter changes δs_μ gives

$$\Delta(s_\mu + \delta s_\mu) \approx \Delta(s_\mu) + \sum_{\mu=1}^{k} \frac{\partial \Delta}{\partial s_\mu} \delta s_\mu.$$

The parameters should be changed in such a way as to minimize $E(\delta s_\mu)$, which implies $\partial E / \partial \delta s_\mu = 0$ for $\mu = 1, \ldots, k$. The δs_μ are found as the solution of the k normal equations

$$\begin{aligned}
&\sum_{\mu'=1}^{k} \left(w_X \sum_{i=1}^{n} \frac{\partial \Delta_X^i}{\partial s_\mu} \frac{\partial \Delta_X^i}{\partial s_{\mu'}} + w_Y \sum_{i=1}^{n} \frac{\partial \Delta_Y^i}{\partial s_\mu} \frac{\partial \Delta_Y^i}{\partial s_{\mu'}} + w_Z \sum_{i=1}^{n} \frac{\partial \Delta_Z^i}{\partial s_\mu} \frac{\partial \Delta_Z^i}{\partial s_{\mu'}} \right) \delta s_{\mu'} \\
&\quad = - \left(w_X \sum_{i=1}^{n} \Delta_X^i \frac{\partial \Delta_X^i}{\partial s_\mu} + w_Y \sum_{i=1}^{n} \Delta_Y^i \frac{\partial \Delta_Y^i}{\partial s_\mu} + w_Z \sum_{i=1}^{n} \Delta_Z^i \frac{\partial \Delta_Z^i}{\partial s_\mu} \right).
\end{aligned}$$

The parameters are corrected by δs_μ and a new cycle of refinement is started until a minimum of E is reached. The weights

$$w_X = 1 / \sum_{i=1}^{n} (\Delta_X^i)^2, \quad w_Y = 1 / \sum_{i=1}^{n} (\Delta_Y^i)^2, \quad w_Z = 1 / \sum_{i=1}^{n} (\Delta_Z^i)^2$$

are calculated with the current guess for s_μ at the beginning of each cycle.

The derivatives appearing in the normal equations can be worked out from the definitions given in Sections 11.3.2.2 and 11.3.2.4, and only the form of the gradient of the Z residuals is shown. Assuming $\sigma_i = \sigma_M / |\zeta_i|$ $(i = 1, \ldots, n)$ is constant for each reflection, the gradients of the Z residuals are obtained from the chain rule and the relation $\mathrm{d}\,\mathrm{erf}(z)/\mathrm{d}z = [2/(\pi)^{1/2}] \exp(-z^2)$.

$$\begin{aligned}
\frac{\partial \Delta_Z^i}{\partial s_\mu} &= \frac{\partial \Delta_Z^i}{\partial \varphi_i} \frac{\partial \varphi_i}{\partial s_\mu} \\
\frac{\partial \Delta_Z^i}{\partial \varphi_i} &= \frac{\Delta_\varphi}{(2\pi)^{1/2} \sigma_i} \sum_{j=-\infty}^{\infty} \exp[-(\varphi_0 + j\Delta_\varphi - \varphi_i)^2 / 2\sigma_i^2] \\
\frac{\partial \varphi_i}{\partial s_\mu} &= \cos \varphi_i \frac{\partial \sin \varphi_i}{\partial s_\mu} - \sin \varphi_i \frac{\partial \cos \varphi_i}{\partial s_\mu}.
\end{aligned}$$

Obviously, $\partial \Delta_Z^i / \partial s_\mu$ is small for a fully recorded reflection because of the small values of all exponentials appearing in $\partial \Delta_Z^i / \partial \varphi_i$. In contrast, the gradient for a partial reflection, equally recorded on two adjacent images, is most sensitive to parameter variations because one of the exponentials assumes its maximum value. In the limiting case of infinitely fine-sliced data, it can be shown that $\lim_{\Delta_\varphi \to 0} \partial \Delta_Z^i / \partial \varphi_i = 1$. Thus, the refinement scheme based on observed Z centroids, as described here and implemented in *XDS*, is applicable to fine-sliced data – and to data recorded with a large oscillation range as well.

11.3.3. Integration

A fundamental requirement for a general integration method is that it should distinguish carefully between signal and background points within its integration domain. For weak reflections, this distinction cannot be made reliably because of the errors superimposed on the signal. The problem can be solved, however, provided that both weak and strong reflections share the same profile shape – an assumption that has been adopted by most data-processing packages.

The intensity distribution of a reflection can be modelled analytically or derived from the observed profiles of neighbouring strong spots. For the rotation method, the profile shape depends strongly on the specific path of the reflection through the Ewald sphere and on variations in the angle of incidence of the diffracted beam on a flat detector. These geometrical distortions can be eliminated by mapping the reflections onto the coordinate system defined in Section 11.3.2.3, which simplifies the task of modelling the expected intensity distribution as all reflection profiles become similar.

11.3.3.1. *Spot extraction*

The region around a spot is defined by the two parameters δ_D and δ_M, which represent spot diameter and reflecting range, respectively. It is assumed that the coordinates of all image pixels contributing to the intensity of a spot satisfy $|\varepsilon_1| \le \delta_D/2, |\varepsilon_2| \le \delta_D/2$ and $|\varepsilon_3| \le \delta_M/2$ when mapped to the profile coordinate system $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ defined in Section 11.3.2.3. Regions of neighbouring reflections may overlap. As implemented in *XDS*, potential overlap is dealt with by a simple strategy: pixels within the overlap region are assigned to the nearest spot. This is carried out in two steps. First, reflections predicted to occur on a given rotation image are found by generating and testing all possible indices h, k, l up to the highest resolution recorded by the detector. Reflection indices, coordinates of the diffracted beam wave vector and the expected fraction of spot intensity recorded on the image are saved in a table. In the second step, each reflection boundary is traced in the image and corrected to exclude pixels belonging to overlapping reflections, which are rapidly located in the table by the hash technique. The image scaling factor obtained from the mean image background and the neighbourhood pixel values belonging to the reflections recorded in the image are saved on a scratch file dedicated to the currently processed data image.

At regular intervals, these files are merged such that all pixel values belonging to a spot found in the contributing images follow each other. Reflections for which contributing pixels are expected further ahead in data processing are just copied to a scratch output file. The other reflections are mapped to the Ewald sphere, as described below, and their three-dimensional profiles and accompanying information are routed to the main output file of the spot-extraction step. After the file-merging procedure, spot extraction continues.

11.3.3.2. *Background*

The region around a spot is assumed to have been chosen to be large enough to include a sufficient number of pixels which can be used for determination of the background. Background determination, as implemented in *XDS*, begins by sorting all pixels belonging to a reflection by increasing intensity. For weak or absent reflections, these values should represent a random sample drawn from a normal distribution. If this is not the case, the pixel with the largest intensity is removed until the sampling distribution of the remaining smaller items satisfies the expected distribution. This method will also exclude pixels with unexpectedly high values, such as ice reflections. The background, determined as the mean value of the accepted pixels, is systematically overestimated for strong spots because of some residual intensity extending into the accepted background pixels. This residual intensity is estimated from the expected distribution $\omega(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ defined in Section 11.3.2.3 and removed from the final background value.

11.3.3.3. *Standard profiles*

Reflection profiles are represented on the Ewald sphere within a domain D_0 comprising $2n_1+1, 2n_2+1, 2n_3+1$ equidistant grid-points along $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$, respectively. The sampling distances between adjacent grid points are then $\Delta_1 = \delta_D/(2n_1+1)$, $\Delta_2 = \delta_D/(2n_2+1), \Delta_3 = \delta_M/(2n_3+1)$. Thus, grid coordinate ν_3 $(\nu_3 = -n_3, \ldots, n_3)$ covers the set of rotation angles

$$\Gamma_{\nu_3} = \{\varphi' | (\nu_3 - 1/2)\Delta_3 \le (\varphi' - \varphi)\cdot\zeta \le (\nu_3 + 1/2)\Delta_3\}.$$

Contributions to the spot intensity come from one or several adjacent data images $(j = j_1, \ldots, j_2)$, each covering the set of rotation angles

$$\Gamma_j = \{\varphi' | \varphi_0 + (j-1)\Delta_\varphi \le \varphi' \le \varphi_0 + j\Delta_\varphi\}.$$

Assuming Gaussian profiles along $\mathbf{e}_3$ for all reflections (see Section 11.3.2.3), the fraction of counts (after subtraction of the background) contributed by data frame j to grid coordinate ν_3 is

$$f_{\nu_3 j} \approx \int\limits_{\Gamma_j \cap \Gamma_{\nu_3}} \exp[-(\varphi' - \varphi)^2/2\sigma^2]\,\mathrm{d}\varphi' \times \left\{ \int\limits_{\Gamma_j} \exp[-(\varphi' - \varphi)^2/2\sigma^2]\,\mathrm{d}\varphi' \right\}^{-1},$$

where $\sigma = \sigma_M/|\zeta|$. The integrals can be expressed in terms of the error function, for which efficient numerical approximations are available (Abramowitz & Stegun, 1972). Finally, each pixel on data image j belonging to the reflection is subdivided into 5 × 5 areas of equal size, and $f_{\nu_3 j}/25$ of the pixel signal is added to the profile value at grid coordinates ν_1, ν_2, ν_3 corresponding to each subdivision.

This complicated procedure leads to more uniform intensity profiles for all reflections than using their untransformed shape. This simplifies the task of modelling the expected intensity distribution needed for integration by profile fitting. As implemented in *XDS*, reference profiles are learnt every 5° of crystal rotation at nine positions on the detector, each covering an equal area of the detector face. In the learning phase, profile boxes of the strong reflections are normalized and added to their nearest reference profile boxes. The contributions are weighted according to the distance from the location of the reference profile. Each grid point within the average profile boxes is classified as signal if it is above 2% of the peak maximum. Finally, each profile is scaled such that the sum of its signal pixels normalizes to one. The analytic expression $\omega(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ defined in Section 11.3.2.3 for the expected intensity distribution is only a rough initial approximation which is now replaced by the empirical reference profiles.

11.3.3.4. *Intensity estimation*

If an expected intensity distribution $\{p_i | i \in D_0\}$ of the observed profile is given in a domain D_0, the reflection intensity I can be estimated as

$$I = \sum_{i\in D} (c_i - b_i) p_i / v_i \bigg/ \sum_{i \in D} p_i^2 / v_i,$$

which minimizes the function

$$\psi(I) = \sum_{i \in D} (c_i - I \cdot p_i - b_i)^2 / v_i, \qquad \sum_{i \in D_0} p_i = 1.$$

b_i, c_i, v_i $(i \in D)$ are background, contents and variance of pixels observed in a subdomain $D \subseteq D_0$ of the expected distribution. The background b_i underneath a diffraction spot is often assumed to be a constant which is estimated from the neighbourhood around the reflection. Determination of reflection intensities by profile fitting has a long tradition (Diamond, 1969; Ford, 1974; Kabsch, 1988*b*; Otwinowski, 1993). Implementations of the method differ mainly in their assumptions about the variances v_i. Ford uses constant variances, which works well for films, which have a high intrinsic background. In *XDS*, which was originally designed for a multiwire detector, $v_i \propto p_i$ was assumed, which results in a straight summation of background-subtracted counts within the expected profile region, $I = \sum_{i \in D}(c_i - b_i)/\sum_{i\in D} p_i$. This particular simple formula is very satisfactory for the low background typical of these detectors. For the general case, however, better results can be obtained by using $v_i = b_i + Ip_i$ for the pixel variances as shown by Otwinowski and implemented in *DENZO* and in the later version of *XDS*. Starting with $v_i = b_i$, the intensity is now found by an iterative process which is terminated if the new intensity estimate becomes negative or does not change within a small tolerance, which is usually reached after three cycles. It can be shown that the solution thus obtained is unique.

11.3.4. Scaling

Usually, many statistically independent observations of symmetry-related reflections are recorded in the rotation images taken from one or several similar crystals of the same compound. The squared structure-factor amplitudes of equivalent reflections should be equal and the idea of scaling is to exploit this *a priori* knowledge to determine a correction factor for each observed intensity. These correction factors compensate to some extent for effects such as radiation damage, absorption, and variations in detector sensitivity and exposure times, as well as variations in size and disorder between different crystals.

The usual methods of scaling split the data into batches of roughly the same size, each covering one or more adjacent rotation images, and then determine a single scaling factor for all reflections in each batch. Neighbouring reflections may then receive quite different corrections if they are assigned to different batches. Since the selection of batch boundaries is to some extent arbitrary, a more continuous correction function would be preferable. This function could be modelled analytically (for example by using spherical harmonics) or empirically, as implemented in *XSCALE* and described below.

For each reflection, observational equations are defined as

$$\psi_{hl\alpha} = (I_{hl} - g_\alpha I_h)/\sigma_{hl}.$$

The subscript h represents the unique reflection indices and l enumerates all symmetry-related reflections to h. By definition, the unique reflection indices have the largest h, then k, then l value occuring in the set of all indices related by symmetry to the original indices, including Friedel mates. Thus, two reflections are symmetry-related if and only if their unique indices are identical. I_h is the unknown 'true' intensity and I_{hl}, σ_{hl} are symmetry-related observed intensities and their standard deviations, respectively. The subscript α denotes the coordinates at which the scaling function g_α should be evaluated. As implemented in *XDS* and *XSCALE*, $\alpha = 1, \ldots, 9$ denotes nine positions uniformly distributed in the detector plane at the beginning of data collection, $\alpha = 10, \ldots, 18$ the same positions on the detector but after the crystal has been rotated by, say, 5°, and so on. The scaling factors g_α and the estimated intensities I_h are found at the minimum of the function

$$\Psi = \sum_{hl\alpha} w_{hl\alpha} \Psi^2_{hl\alpha}.$$

references