

11.4. DENZO AND SCALEPACK

values of scaling parameters due to insufficient information to determine the values of parameters. In *SCALEPACK*, the method to stabilize such ill-conditioned calculations is closely related to Tikhonov stabilization (Tikhonov & Arsenin, 1977), where additional, *a priori* knowledge about the expected magnitude of the physical effect modelled is used to restrain the solutions, based on the same argument as in the case of restraints in the atomic refinement.

For example, logarithms of scale factors typically do not fluctuate by more than w_s between frames, where expectation about w_s is a function of the data-collection stability (beam stability, goniostat and/or crystal vibrations). This knowledge is described by adding a penalty term (*scale restrain*) to the functions being optimized:

$$(1/w_s^2)(p_{i_s} - p_{(i+1)_s})^2. \quad (11.4.8.9)$$

A similar approach can be used in calculations of absorption coefficients. For smooth absorption with the expectation of decreasing magnitude of parameters for high orders of spherical harmonics [equation (11.4.8.7)], a reasonable restraint term parameterized by w_a results in

$$\frac{I^2(p_{as,lm}^2 + p_{ac,lm}^2)}{w_a^2}. \quad (11.4.8.10)$$

If we do not want to penalize high-order terms more than the low-order ones, the following restraint can be used:

$$\frac{(p_{as,lm}^2 + p_{ac,lm}^2)}{w_a^2}. \quad (11.4.8.11)$$

11.4.9. Global refinement or post refinement

The process of refining crystal parameters using the combined reflection intensity measurements is known as global refinement or post refinement (Rossmann, 1979; Evans, 1993). The implementation of this method in *SCALEPACK* allows for separate refinement of the orientation of each image, but with the same unit-cell value for the whole data set. In each batch of data (a batch is typically one image), different unit-cell parameters may be poorly determined. However, in a typical data set there are enough orientations to determine all unit-cell lengths and angles precisely. Global refinement is also more precise than the processing of a single image in the determination of crystal mosaicity and the orientation of each image.

11.4.10. Merging – assessment of the error model and signal magnitudes in the data

Proper error estimation requires the use of Bayesian reasoning and a multi-component error model (Schwarzenbach *et al.*, 1989; Evans, 1993). In principle, the error estimates may be derived solely from a theoretical understanding of the measurement process. However, the complexity of error propagation and correlations between various sources of effects have led crystallographers to rely on hybrid approaches also involving self-consistency analysis of symmetry-equivalent reflections.

11.4.10.1. Estimation of random errors

The random errors in *DENZO* are estimated by a heuristic procedure that also accounts for small components of systematic errors (Borek *et al.*, 2003). Initially, *DENZO* estimated errors of integrated diffraction peaks from X-ray film. After introducing

detectors with larger dynamic range, the procedure was adjusted accordingly.

The initial estimates of errors are obtained by

$$\sigma_0 = \frac{1}{\sum_i P_i^2/(B_i + P_i I)} \times \left\{ e_d \left[\sum_i P_i^2 (B_i + P_i I) + \frac{e_d}{n_b} \sum_i \frac{P_i^2 B_i}{(B_i + P_i I)^2} \right] \right\}^{1/2}, \quad (11.4.10.1)$$

where n_b is the number of pixels used in background estimation and e_d is the error-density parameter defined for each instrument, which can also be overridden by the user (Gewirth, 2003) with other variables defined in equation (11.4.7.1). The sums are calculated over all the pixels in a reflection profile. The expression within the braces { } describes two components of uncertainty: the left sum accounts for contributions resulting from pixels in the peak area, whereas the right sum adds an adjustment resulting from uncertainty of the background estimate. The denominator in the front of the expression in braces is derived from error propagation for the profile-fitted intensity [equation (11.4.7.3)].

Next, the goodness-of-profile-fitting factor g is calculated:

$$g = \left[\frac{1}{(n_i - 1)} \sum_i \frac{(M_i - B_i - P_i I)^2}{e_d (B_i + P_i I)} \right]^{1/2}, \quad (11.4.10.2)$$

where n_i is the number of pixels in a reflection profile. For weak reflections the parameter g should be relatively close to 1. If it is systematically off by a large factor, the error-density parameter e_d should be adjusted (Borek *et al.*, 2003). *SCALEPACK* applies an additional level of adjustment to the estimates produced by *DENZO* (Borek *et al.*, 2003):

$$\sigma_s = 1.2\sigma_0 g^{1/2}, \quad (11.4.10.3)$$

which is scaled either by the user or by an automatically adjustable factor E_s (called the error scale factor) to make disagreements among symmetry-related measurements consistent:

$$\sigma_I = E_s \sigma_s. \quad (11.4.10.4)$$

Even this scaled estimate of random error σ_I does not account for all types of errors and additional adjustments for systematic effects are needed.

11.4.10.2. Estimation of multiplicative errors

The multiplicative scale factor has its own uncertainty independent of random errors with typical values in the range of a few per cent. However, even such small errors are important in calculations of the phase signal. Errors in the scale factors have a correlated component that equally affects measurements of intensities in phasing differences, so it does not impact on the differences themselves. The important part is estimating the magnitude of the remaining component of scaling errors, described by σ_K . Comparing symmetry-related reflections estimates only the relevant component of multiplicative errors. The total scaling error would have to be estimated differently, but typically it has little relevance to macromolecular crystallography and can be ignored.

The σ_I [equation (11.4.10.4)] can be combined with σ_K to obtain the final estimated error of the scaled measurement:

$$\sigma_E = (1/K)(\sigma_I^2 + I^2 \sigma_K^2)^{1/2}. \quad (11.4.10.5)$$