While the information content of the data appears to depend quantitatively on the nominal resolution, in fact it is dependent on the data quality throughout the resolution range, and both high- and low-resolution completeness and their statistical significance affect the information content of the data and derived model. High-intensity low-resolution terms remain important for refinement at atomic resolution, as they define the contrast in the density maps between solvent and protein, and because their omission biases the refinement, especially that of parameters such as the ADPs. To judge the effective resolution of the diffraction data set, the concept of 'optical resolution', which can be estimated from the shape of the origin peak in the Patterson synthesis, may be very useful (Vaguine *et al.*, 1999).

The rejection of low-intensity observations will also introduce bias. In particular, all the maps calculated for visual or computer inspection by Fourier transformation are diminished in quality by omission of any terms, but are especially affected by omission of strong low-resolution data. This is particularly true in the early stages of structure solution, where low-resolution data can be vital. Although most phase-improvement algorithms rely on relations between all reflections, terms involving low-resolution reflections will be large, will be involved in many relations and will play a dominant role. Hence, omission of these terms will severely degrade the power of these methods, which may indeed converge to solutions that have nothing whatsoever to do with the real structure.

18.4.2.2. Anisotropic scaling

The intensity data from a crystal may display anisotropy, *i.e.*, the intensity fall-off with resolution will vary with direction, and may be much higher along one crystal axis than along another. If the structure is to be refined with an isotropic atomic model (either because there are insufficient data or the programs used cannot handle anisotropic parameters), then the fall-off of the calculated F^2 values will, of necessity, also be isotropic. In this situation, an improved agreement between observed and calculated F^2 values can be obtained either by using anisotropic scaling during data reduction to the expected Wilson distribution of intensities, or by including a maximum of six overall anisotropic parameters during refinement. This will result in an isotropic set of F^2 values. For crystals with a high degree of anisotropy in the experimental data, this can lead to a substantial drop of several per cent in R and $R_{\rm free}$ (Sheriff & Hendrickson, 1987; Murshudov *et al.*, 1998).

This ambiguity effectively disappears with use of an anisotropic atomic model. The individual ADPs, including contributions from both static and thermal disorder, take up relative individual displacements, but also the overall anisotropy of the experimental F^2 values. The significance of the overall anisotropy is a point of some contention, and its physical meaning is not clear. It may represent asymmetric crystal imperfection or anisotropic overall displacement of molecules in the lattice related to TLS parameters. Refinement of TLS parameters, which can be performed using, for example, *RESTRAIN* (Driessen *et al.*, 1989) *REFMAC* (Winn *et al.*, 2001) or *PHENIX.REFINE* (Terwilliger *et al.*, 2008), removes the overall crystal contribution to the ADPs.

It is important that at least the intensity or amplitude data be deposited in the PDB as measured, without any anisotropic correction being applied. For the refined model, complete information on any overall anisotropic and TLS modelling must be explicitly included, as well as the individual atomic ADPs.

18.4.3. Computational algorithms and strategies

18.4.3.1. Classical least-squares refinement of small molecules

The principles of the least-squares method of minimization are described in *International Tables for Crystallography* Volume C (2004). Least squares involves the construction of an order $N \times N$ normal matrix, where N is the number of parameters, representing a system of least-squares equations, whose solution provides estimates of adjustments to the current atomic parameters. The problem is nonlinear and the matrix construction and solution must be iterated until convergence is achieved. In addition, inversion of the matrix at convergence provides an approximation to the standard uncertainties for each individual parameter refined according to the Cramer–Rao inequality (Stuart *et al.*, 1999). Indeed, this is the only method available so far that gives such estimates properly.

However, even for small molecules there may be some disordered regions which will require the imposition of restraints, as is the case for macromolecules (see below), and the presence of such restraints means that the error estimates no longer reflect the information from the X-ray data alone. If the problem of how restraints affect the error estimates could be resolved, then inversion of the matrix corresponding to the second derivative of the posterior distribution would provide standard uncertainties incorporating both the prior knowledge, such as the restraints, and the experimental data. Equation (18.4.1.2) for information measure could then be applied, but this requires further development. For small structures, the speed and memory of modern computers have reduced the requirements for such calculations to the level of seconds, and the computational requirements form a trivial part of the structure analysis. Recent developments in the application of fast Fourier transform methods to normal matrix–vector multiplication (Strokopytov, 2008) and fast information matrix evaluation (Steiner *et al.*, 2003) suggest that fast calculations for macromolecular structures may be available in the foreseeable future.

18.4.3.2. Least-squares refinement of large structures

The size of the computational problem increases dramatically with the size of the unit cell, as the number of terms in the matrix increases with the square of the number of parameters. Furthermore, construction of each element depends on the number of reflections. For macromolecular structures, computation of a full matrix is at present prohibitively expensive in terms of CPU time and memory. A variety of simplifying approaches have been developed, but all suffer from a poorer estimate of the standard uncertainties and from a more limited range and speed of convergence.

The first is the block-matrix approach, where instead of the full matrix, only square blocks along the matrix diagonal are constructed, involving groups of parameters that are expected to be correlated. The correlation between parameters belonging to different blocks is therefore neglected completely. In this way, the whole least-squares minimization is split into a set of smaller independent units. In principle this leads to the same solution, but more slowly and with less precise error estimates. Nevertheless, block-matrix approaches remain essential for tractable matrix inversion for macromolecular structures.

A further simplification involves the conjugate-gradient method or the diagonal approximation to the normal matrix (the second derivative of minus the log of the likelihood function in the case of maximum likelihood), which essentially ignores all off-diagonal terms of the least-squares matrix. For the conjugate-