

3. DUAL BASES IN CRYSTALLOGRAPHIC COMPUTING

$$\sigma^2(l_1) = l_1^T \mathbf{g}(\mathbf{M}_{AA} - \mathbf{S}\mathbf{M}_{AA} - \mathbf{M}_{AA}\mathbf{S}^T + \mathbf{S}\mathbf{M}_{AA}\mathbf{S}^T) \mathbf{g}l_1 / l_1^2 \quad (3.1.10.16)$$

$$\sigma^2(l_2) = l_2^T \mathbf{g}(\mathbf{S}\mathbf{M}_{AA}\mathbf{S}^T - \mathbf{M}_{AC}\mathbf{S}^T - \mathbf{S}\mathbf{M}_{AC} + \mathbf{M}_{CC}) \mathbf{g}l_2 / l_2^2 \quad (3.1.10.17)$$

$$\sigma^2(l_3) = l_3^T \mathbf{g}(\mathbf{M}_{AA} - \mathbf{M}_{AC} - \mathbf{M}_{CA} + \mathbf{M}_{CC}) \mathbf{g}l_3 / l_3^2 \quad (3.1.10.18)$$

$$\text{cov}(l_1, l_2) = l_1^T \mathbf{g}(\mathbf{M}_{AA}\mathbf{S}^T - \mathbf{S}\mathbf{M}_{AA}\mathbf{S}^T - \mathbf{M}_{AC} + \mathbf{S}\mathbf{M}_{AC}) \mathbf{g}l_2 / l_1 l_2 \quad (3.1.10.19)$$

$$\text{cov}(l_1, l_3) = l_1^T \mathbf{g}(-\mathbf{M}_{AA} + \mathbf{S}\mathbf{M}_{AA} + \mathbf{M}_{AC} - \mathbf{S}\mathbf{M}_{AC}) \mathbf{g}l_3 / l_1 l_3 \quad (3.1.10.20)$$

$$\text{cov}(l_2, l_3) = l_2^T \mathbf{g}(-\mathbf{S}\mathbf{M}_{AA} + \mathbf{M}_{CA} + \mathbf{S}\mathbf{M}_{AC} - \mathbf{M}_{CC}) \mathbf{g}l_3 / l_2 l_3. \quad (3.1.10.21)$$

In equations (3.1.10.14)–(3.1.10.21), l_i is a column vector with components the differences of the coordinates of the atoms connected by the vector. Representative formulae involving the angles $\alpha_1, \alpha_2, \alpha_3$ are

$$\begin{aligned} \sigma^2(\alpha_1) = & [\cos^2 \alpha_2 \sigma^2(l_1) - 2 \cos \alpha_2 \text{cov}(l_1, l_2) \\ & + 2 \cos \alpha_2 \cos \alpha_3 \text{cov}(l_1, l_3) + \sigma^2(l_2) \\ & - 2 \cos \alpha_3 \text{cov}(l_2, l_3) \\ & + \cos^2 \alpha_3 \sigma^2(l_3)] / (l_2 / l_1 l_3 \sin \alpha_1)^2 \end{aligned} \quad (3.1.10.22)$$

$$\begin{aligned} \text{cov}(\alpha_1, \alpha_2) = & [\cos \alpha_1 \cos \alpha_2 \sigma^2(l_1) \\ & + (\cos \alpha_2 \cos \alpha_3 - \cos \alpha_1) \text{cov}(l_1, l_2) \\ & + (\cos \alpha_1 \cos \alpha_3 - \cos \alpha_2) \text{cov}(l_1, l_3) \\ & - \cos \alpha_3 \sigma^2(l_2) + (1 + \cos^2 \alpha_3) \text{cov}(l_2, l_3) \\ & - \cos \alpha_3 \sigma^2(l_3)] / (l_1^2 \sin \alpha_1 \sin \alpha_2) \end{aligned} \quad (3.1.10.23)$$

$$\text{cov}(\alpha_1, l_1) = [-\cos \alpha_2 \sigma^2(l_1) + \text{cov}(l_1, l_2) - \cos \alpha_3 \text{cov}(l_1, l_3)] / (l_2 / l_1 l_3 \sin \alpha_1) \quad (3.1.10.24)$$

$$\text{cov}(\alpha_1, l_2) = [-\cos \alpha_2 \text{cov}(l_1, l_2) + \sigma^2(l_2) - \cos \alpha_3 \text{cov}(l_2, l_3)] / (l_2 / l_1 l_3 \sin \alpha_1). \quad (3.1.10.25)$$

If any of the angles approach 0 or 180°, the denominators in (3.1.10.22)–(3.1.10.25) will become very small, necessitating high-precision arithmetic. Indeterminacies resulting from special relationships between atomic positions may require rederivation of the equations for variances and covariances, to take the relationships into account explicitly and avoid the indeterminacies. A true symmetry condition requiring, for example, a linear bond should cause little problem, and the corresponding variance will be zero. It is the indeterminacies not originating from crystal symmetry that demand caution, in recognizing them and in coping with them correctly.

A general expression for the variance of a dihedral angle, in terms of the variances and covariances of the coordinates of the four atoms, is (Shmueli, 1974)

$$\sigma^2(\tau) = \sum_k \sum_n \frac{\partial \tau}{\partial x_{(k)}^i} \frac{\partial \tau}{\partial x_{(n)}^j} \text{cov}[x_{(k)}^i, x_{(n)}^j], \quad (3.1.10.26)$$

where, in addition to the usual tensor summation over i and j from 1 to 3, summation must be carried out over the four atoms (*i.e.*, k and n vary from 1 to 4). Special cases of (3.1.10.26), corresponding to various levels of approximation of diagonal matrices and isotropic errors, are given in Shmueli (1974). Formulae in dyadic notation are given in Waser (1973) for the

variances and covariances of dihedral angles, of best planes, of torsion angles, and of other molecular parameters.

3.1.11. Mean values

The weighted mean of a set of quantities X_i is

$$\langle X \rangle = \sum w_i X_i / \sum w_i, \quad (3.1.11.1)$$

where the weights are typically chosen to minimize the variance of $\langle X \rangle$. The variance may be computed from the variance–covariance matrix \mathbf{M} of the X_i by

$$\sigma^2(\langle X \rangle) = \mathbf{w}^T \mathbf{M} \mathbf{w} / (\sum w_i)^2. \quad (3.1.11.2)$$

Minimization of $\sigma^2(\langle X \rangle)$ leads to weights given by

$$\mathbf{w} = \mathbf{M}^{-1} \mathbf{v}, \quad (3.1.11.3)$$

where the components of vector \mathbf{v} are all equal ($v_i = v_j$ for all i and j); since (3.1.11.1) and (3.1.11.2) require only relative weights, we can assign $v_i = 1$ for all i . Placing these weights in (3.1.11.2) yields

$$\sigma^2(\langle X \rangle) = 1 / \sum w_i. \quad (3.1.11.4)$$

For the case of uncorrelated X_i , the weights are inversely proportional to the corresponding variances

$$w_i = 1 / \sigma^2(X_i). \quad (3.1.11.5)$$

For the case of two correlated variables,

$$w_i = 1 / [\sigma^2(X_i) - \text{cov}(X_1, X_2)]. \quad (3.1.11.6)$$

Derivation and discussion of these equations may be found in Sands (1966, 1982*b*).

The presence of systematic errors in the experimental data often results in these formulae producing estimates of the standard uncertainties of molecular dimensions that are too small; it has been suggested that such error estimates should be multiplied by 1.5 to make them more realistic (Taylor & Kennard, 1983). It is essential also that averages be computed only of similar quantities, and interatomic distances corresponding to different bond orders or different environments may not represent the same physical quantities; that is, there are reasons for the discrepancies, and averaging may obscure important information. Another source of error in molecular geometry parameters determined from crystallographic measurements is thermal motion, and distances should be corrected for such effects before making comparisons (Busing & Levy, 1964; Johnson, 1970, 1980).

A discussion of the appropriateness of weighted and unweighted means may be found in Taylor & Kennard (1985), which suggests that the unweighted mean might even be preferable if environmental effects are large.

3.1.12. Computation

It has been mentioned that the tensor formulation used in this chapter is particularly amenable to machine computation. As a simple illustration of this point, the following Fortran program will compute the lengths of vectors \mathbf{X} and \mathbf{Y} and the angle between them.