3.2. The least-squares plane

By R. E. Marsh and V. Schomaker

3.2.1. Introduction

By way of introduction, we remark that in earlier days of crystal structure analysis, before the advent of high-speed computers and routine three-dimensional analyses, molecular planarity was often assumed so that atom coordinates along the direction of projection could be estimated from two-dimensional data [see, *e.g.*, Robertson (1948)]. Today, the usual aim in deriving the coefficients of a plane is to investigate the degree of planarity of a group of atoms as found in a full, three-dimensional structure determination. We further note that, for such purposes, a crystallographer will often be served just as well by establishing the plane in an almost arbitrary fashion as by resorting to the most elaborate, nit-picking and pretentious least-squares treatment. The approximate plane and the associated perpendicular distances of the atoms from it will be all he needs as scaffolding for his geometrical and structural imagination; reasonable common sense will take the place of explicit attention to error estimates.

Nevertheless, we think it appropriate to lay out in some detail the derivation of the 'best' plane, in a least-squares sense, through a group of atoms and of the standard uncertainties associated with this plane. We see two cases: (1) The weights of the atoms in question are considered to be isotropic and uncorrelated (*i.e.* the weight matrix for the positions of all the atoms is diagonal, when written in terms of Cartesian axes, and for each atom the three diagonal elements are equal). In such cases the weights may have little or nothing to do with estimates of random error in the atom positions (they may have been assigned merely for convenience or convention), and, therefore, no one should feel that the treatment is proper in respect to the theory of errors. Nevertheless, it may be desired to incorporate the error estimates (variances) of the atom positions into the *results* of such calculations, whereupon these variances (which may be anisotropic, with correlation between atoms) need to be propagated. In this case the distinction between *weights* (or their inverses) and *variances* must be kept very clear. (2) The weights are anisotropic and are presumably derived from a variance–covariance matrix, which may include correlation terms between different atoms; the objective is to achieve a truly proper Gaussian least-squares result.

3.2.2. Least-squares plane based on uncorrelated, isotropic weights

This is surely the most common situation; it is not often that one will wish to take the trouble, or be presumptive enough, to assign anisotropic or correlated weights to the various atoms. And one will sometimes, perhaps even often, not be genuinely interested in the hypothesis that the atoms actually are rigorously coplanar; for instance, one might be interested in examining the best plane through such a patently non-planar molecule as cyclohexane. Moreover, the calculation is simple enough, given the availability of computers and programs, as to be a practical realization of the off-the-cuff treatment suggested in our opening paragraph. The problem of deriving the plane's coefficients is intrinsically nonlinear in the way first discussed by Schomaker *et al.* (1959; SWMB). Any formulation other than as an eigenvalue–eigenvector problem (SWMB), as far as we can tell, will sometimes go astray. As to the propagation of errors, numerous treatments have been given, but none that we have seen is altogether satisfactory.

We refer all vectors and matrices to Cartesian axes, because that is the most convenient in calculation. However, a more elegant formulation can be written in terms of general axes [*e.g.*, as in Shmueli (1981)].

The notation is troublesome. Indices are needed for atom number and Cartesian direction, and the exponent 2 is needed as well, which is difficult if there are superscript indices. The best way seems to be to write all the indices as subscripts and distinguish among them by context – $i, j, 1, 2, 3$ for directions; k, l, p (and sometimes $K, \ldots$) for atoms. In any case, *atom* first then *direction* if there are two subscripts; *direction,* if only one index for a vector component, but *atom* (in this section at least) if for a weight or a vector. And σ_{d_1}, *e.g.*, for the standard uncertainty of the distance of atom 1 from a plane. For simplicity in practice, we use Cartesian coordinates throughout.

The first task is to find the plane, which we write as

$$0 = \mathbf{m} \cdot \mathbf{r} - d \equiv \mathsf{m}^T \mathsf{r} - d,$$

where $\mathbf{r}$ is here the vector from the origin to any point on the plane (but usually represents the measured position of an atom), $\mathbf{m}$ is a unit vector parallel to the normal from the origin to the plane, d is the length of the normal, and m and r are the column representations of $\mathbf{m}$ and $\mathbf{r}$. The least-squares condition is to find the stationary values of $S \equiv [w_k(\mathsf{m}^T \mathsf{r}_k - d)^2]$ subject to $\mathsf{m}^T \mathsf{m} = 1$, with $\mathsf{r}_k, k = 1, \ldots, n$, the vector from the origin to atom k and with weights, w_k, isotropic and without interatomic correlations for the n atoms of the plane. We also write S as $S \equiv [w(\mathsf{m}^T \mathsf{r} - d)^2]$, the subscript for atom number being implicit in the Gaussian summations ($[\ldots]$) over all atoms, as it is also in the angle-bracket notation for the weighted average over all atoms, for example in $\langle \mathsf{r} \rangle$ – the weighted centroid of the groups of atoms – just below.

First solve for d, the origin-to-plane distance.

$$0 = -\frac{1}{2}\frac{\partial S}{\partial d} = [w(\mathsf{m}^T \mathsf{r} - d)] = 0,$$

$$d = [w\mathsf{m}^T \mathsf{r}]/[w] \equiv \mathsf{m}^T \langle \mathsf{r} \rangle.$$

Then

$$S \equiv [w(\mathsf{m}^T \mathsf{r} - d)^2] = [w\{\mathsf{m}^T(\mathsf{r} - \langle \mathsf{r} \rangle)\}^2]$$

$$\equiv [w(\mathsf{m}^T \mathsf{s})^2] \equiv \mathsf{m}^T[w\mathsf{s}\mathsf{s}^T]\mathsf{m} \equiv \mathsf{m}^T \mathsf{A}\mathsf{m}.$$

Here $\mathsf{s}_k \equiv \mathsf{r}_k - \langle \mathsf{r} \rangle$ is the vector from the centroid to atom k. Then solve for $\mathbf{m}$. This is the eigenvalue problem – to diagonalize A (bear in mind that A_{ij} is just $[ws_i s_j]$) by rotating the coordinate axes, *i.e.*, to find the 3×3 arrays M and L, L diagonal, to satisfy

$$\mathsf{M}^T \mathsf{A}\mathsf{M} = \mathsf{L}, \qquad \mathsf{M}^T \mathsf{M} = \mathsf{I}.$$

A and M are symmetric; the columns m of M are the direction cosines of, and the diagonal elements of L are the sums of weighted squares of residuals from, the best, worst and intermediate planes, as discussed by SWMB.

3.2.2.1. *Error propagation*

Waser *et al.* (1973; WMC) carefully discussed how the random errors of measurement of the atom positions propagate into the derived quantities in the foregoing determination of a least-squares plane. This section presents an extension of their discussion. To begin, however, we first show how standard first-order perturbation theory conveniently describes the propagation of error into M and L when the positions r_k of the atoms are incremented by the amounts $\delta \mathsf{r}_k \equiv \xi_k$ and the corresponding quantities $\mathsf{s}_k \equiv \mathsf{r}_k - \langle \mathsf{r} \rangle$ (the vectors from the centroid to the atoms) by the amounts $\eta_k, (\mathsf{s} \to \mathsf{s} + \eta)$, $\eta_k \equiv \xi_k - \langle \xi \rangle$. (The need to account for the variation in position of the centroid, *i.e.* to distinguish between η_k and ξ_k, was overlooked by WMC.) The consequent increments in A, M and L are

$$\delta \mathbf{A} = [w\eta \mathbf{s}^T] + [w\mathbf{s}\eta^T] \equiv \alpha,$$

$$\delta \mathbf{M} = \mathbf{M}\mu,$$

$$\delta \mathbf{L} \equiv \lambda.$$

Here the columns of μ are expressed as linear combinations of the columns of $\mathbf{M}$. Note also that both perturbations, μ and λ, which are the adjustments to the orientations and associated eigenvalues of the principal planes, will depend on the reduced coordinates $\mathbf{s}$ and the perturbing influences ξ by way of α, which in turn depends only on the reduced coordinates and the reduced shifts η_k. In contrast,

$$\delta \mathbf{d} = \delta(\mathbf{m}^T \langle \mathbf{r} \rangle) = (\delta \mathbf{m}^T)\langle \mathbf{r} \rangle + \mathbf{m}^T \langle \xi \rangle;$$

the change in the origin-to-plane distance for the plane defined by the column vectors $\mathbf{m}$ of $\mathbf{M}$, depends on the $\langle \mathbf{r} \rangle$ and $\langle \xi \rangle$ directly as well as on the $\mathbf{s}$ and η by way of the $\delta \mathbf{m}$.

The first-order results arising from the standard conditions, $\mathbf{M}^T \mathbf{M} = \mathbf{I}, \mathbf{L}$ diagonal, and $\mathbf{M}^T \mathbf{AM} = \mathbf{L}$, are

$$\mu^T + \mu = 0, \lambda \text{ diagonal,}$$

and

$$\mu^T \mathbf{M}^T \mathbf{AM} + \mathbf{M}^T \alpha \mathbf{M} + \mathbf{M}^T \mathbf{AM}\mu = \mu^T \mathbf{L} + \mathbf{L}\mu + \mathbf{M}^T \alpha \mathbf{M} = \lambda.$$

Stated in terms of the matrix components λ_{ij} and μ_{ij}, the first condition is $\mu_{ij} = -\mu_{ji}$, hence $\mu_{ii} = 0$, $i,j = 1,2,3$, and the second is $\lambda_{ij} = 0$, $i \neq j$. With these results the third condition then reads

$$\lambda_{jj} = (\mathbf{M}^T \alpha \mathbf{M})_{jj}, \qquad j = 1,2,3$$

$$\mu_{ij} = (\mathbf{M}^T \alpha \mathbf{M})_{ij}/(L_{jj} - L_{ii}), \quad i \neq j, \quad i,j = 1,2,3.$$

All this is analogous to the usual first-order perturbation theory, as, for example, in elementary quantum mechanics.

Now rotate to the coordinates defined by WMC, with axes parallel to the original eigenvectors [$\mathbf{M} = \mathbf{I}$, $A_{ij} = L_{ij}\delta_{ij}$, $(\mathbf{M}^T \alpha \mathbf{M})_{ij} = \alpha_{ij}$], restrict attention to the best plane ($M_{13} \equiv m_1 = 0, M_{23} \equiv m_2 = 0, M_{33} \equiv m_3 = 1$), and define ε^T as ($\delta m_1, \delta m_2, \delta d_c$), keeping in mind $\delta m_3 = \mu_{33} = 0$; d_c itself, the original plane-to-centroid distance, of course vanishes. One then finds

$$\delta m_i \equiv \varepsilon_i = \alpha_{i3}/(L_{33} - L_{ii})$$

$$= [w(s_i \eta_3 + s_3 \eta_i)]/([ws_3^2] - [ws_i^2]), \quad i = 1,2,$$

$$\delta d_c \equiv \varepsilon_3 = [w\xi_3]/[w] \equiv \langle \xi_3 \rangle, \qquad \xi_k \equiv \delta r_k,$$

and also

$$\delta d = \varepsilon_1 \langle r_1 \rangle + \varepsilon_2 \langle r_2 \rangle + \varepsilon_3.$$

These results have simple interpretations. The changes in direction of the plane normal (the δm_i) are rotations, described by ε_1 and ε_2, in response to changes in moments acting against effective torsion force constants. For ε_2, for example, the contribution of atom k to the total relevant moment, about direction 1, is $-w_k s_{k3} s_{k2}$ ($w_k s_{k3}$ the 'force' and s_{k2} the lever arm), and its nominally first-order change has two parts, $-w_k s_{k2} \eta_3$ from the change in force and $-w_k s_{k3} \eta_2$ from the change in lever arm; the resisting torsion constant is $[ws_2^2] - [ws_3^2]$, which, reflection will show, is qualitatively reasonable if not quantitatively obvious. The perpendicular displacement of the plane from the original centroid $\langle r \rangle$ is ε_3, but there are two further contributions to δd, the change in distance from origin to plane along the plane normal, that arise from the two components of out-of-plane rotation of the plane about its centroid. Note that ε_3 is not given by $[w\eta_3]/[w] = [w(\xi_3 - \langle \xi_3 \rangle)]/[w]$, which vanishes identically.

There is a further, somewhat delicate point: If the group of atoms is indeed essentially coplanar, the s_{k3} are of the same order of

magnitude as the η_{ki}, unlike the $s_{ki}, i \neq 3$, which are in general about as big as the lateral extent of the group. It is then appropriate to drop all terms in η_i or ξ_i, $i \neq 3$, and, in the denominators, the terms in s_{k3}^2.

The covariances of the derived quantities (by *covariances* we mean here both variances and covariances) can now be written out rather compactly by extending the implicit designation of atom numbers to double sums, the first of each of two similar factors referring to the first atom index and the second to the second, *e.g.*, $\sum_{kl} ww(s_i s_j) \ldots \equiv \sum_{kl} w_k w_l (s_{ki} s_{lj}) \ldots$. Note that the various covariances, *i.e.* the averages over the presumed population of random errors of replicated measurements, are indicated by overlines, angle brackets having been pre-empted for averages over sets of atoms.

$$\text{cov}(m_i, m_j) \equiv \overline{\varepsilon_i \varepsilon_j}$$

$$= \frac{\sum_{kl} ww(s_i s_j \overline{\eta_3 \eta_3} + s_3 s_3 \overline{\eta_i \eta_j} + s_i s_3 \overline{\eta_3 \eta_j} + s_3 s_j \overline{\eta_i \eta_3})}{\{[w(s_3^2 - s_i^2)]\}\{[w(s_3^2 - s_j^2)]\}}, i,j = 1,2$$

$$\text{cov}(m_i, d_c) \equiv \overline{\varepsilon_i \varepsilon_3} = \frac{\sum_{kl} ww(s_{ki} \overline{\eta_3 \xi_3} + s_{k3} \overline{\eta_i \xi_3})}{\{[w(s_3^2 - s_i^2)]\}[w]}, \quad i,j = 1,2$$

$$\sigma^2(d_c) \equiv \overline{\varepsilon_3^2} = \frac{\sum_{kl} ww \overline{\xi_3 \xi_3}}{[w]^2}$$

$$\sigma^2(d) \equiv \langle (\delta d)^2 \rangle = \langle r_1 \rangle^2 \overline{\varepsilon_1^2} + \langle r_2 \rangle^2 \overline{\varepsilon_2^2} + \overline{\varepsilon_3^2} + 2\langle r_1 \rangle \langle r_2 \rangle \overline{\varepsilon_1 \varepsilon_2}$$

$$+ 2\langle r_1 \rangle \overline{\varepsilon_1 \varepsilon_3} + 2\langle r_2 \rangle \overline{\varepsilon_2 \varepsilon_3}.$$

Interatomic covariance (*e.g.*, $\overline{\eta_{k3} \eta_{l3}}$, $k \neq l$) thus presents no formal difficulty, although actual computation may be tedious. Nonzero covariance for the η's *may* arise explicitly from interatomic covariance (*e.g.*, $\overline{\xi_{ki} \xi_{lj}}$, $k \neq l$) of the errors in the atomic positions r_k, and it will *always* arise implicitly because $\langle \xi \rangle$ in $\eta_k = \xi_k - \langle \xi \rangle$ includes all the ξ_k and therefore has nonzero covariance with all of them and with itself, even if there is no interatomic covariance among the ξ_i's.

If both types of interatomic covariance (explicit and implicit) are negligible, the ε covariances simplify a great deal, the double summations reducing to single summations. [The formal expression for $\sigma^2(d)$ does not change, so it will not be repeated.]

$$\text{cov}(m_i, m_j) \equiv \overline{\varepsilon_i \varepsilon_j}$$

$$= \frac{[w^2(s_i s_j \overline{\eta_3^2} + s_3^2 \overline{\eta_i \eta_j} + s_i s_3 \overline{\eta_3 \eta_j} + s_3 s_j \overline{\eta_i \eta_3})]}{\{[w(s_3^2 - s_i^2)]\}\{[w(s_3^2 - s_j^2)]\}}, i,j = 1,2$$

$$\text{cov}(m_i, d_c) \equiv \overline{\varepsilon_i \varepsilon_3} = \frac{[w^2(s_i \overline{\eta_3 \xi_3} + s_3 \overline{\eta_i \xi_3})]}{\{[w(s_3^2 - s_i^2)]\}[w]}, \quad i,j = 1,2$$

$$\sigma^2(d_c) \equiv \overline{\varepsilon_3^2} = \frac{[w^2 \overline{\xi_3^2}]}{[w]^2}.$$

When the variances are the same for ξ as for η (i.e. $\overline{\xi_i \xi_j} = \overline{\eta_i \eta_j}$, all i,j) and the covariances all vanish ($\overline{\xi_i \xi_j} = 0$, $i \neq j$), the $\overline{\varepsilon_i \varepsilon_j}$ simplify further. If the variances are also isotropic ($\overline{\xi_i^2} = \overline{\xi_j^2} = \sigma^2$, all i, j), there is still further simplification to

$$\sigma^2(m_i) \equiv \overline{\varepsilon_i^2} = \frac{[w^2 \sigma^2(s_i^2 + s_3^2)]}{\{[w(s_3^2 - s_i^2)]\}^2}, \quad i = 1,2$$

$$\sigma^2(d_c) \equiv \overline{\varepsilon_3^2} = [w^2 \sigma^2]/[w]^2$$

$$\text{cov}(m_1, m_2) \equiv \overline{\varepsilon_1 \varepsilon_2} = \frac{[w^2 \sigma^2 s_1 s_2]}{\{[w(s_3^2 - s_1^2)]\}\{[w(s_3^2 - s_2^2)]\}}$$

$$\text{cov}(m_i, d_c) \equiv \overline{\varepsilon_i \varepsilon_3} = \frac{[w^2 \sigma^2 s_i]}{\{[w(s_3^2 - s_i^2)]\}[w]}, \quad i = 1,2.$$

If allowance is made for the difference in definition between ε_3 and

δd, these expressions are equivalent to the ones (equations 7–9) given by WMC, who, however, do not appear to have been aware of the distinction between η and ξ and the possible consequences thereof.

If, finally, w^{-1} for each atom is taken equal to its $\overline{\eta_{lj}^2} = \sigma^2$, all j, there is still further simplification.

$$\sigma^2(m_i) \equiv \overline{\varepsilon_i^2} = \frac{[w(s_3^2 + s_i^2)]}{\{[w(s_3^2 - s_i^2)]\}^2}, \quad i = 1, 2$$

$$\sigma^2(d_c) \equiv \overline{\varepsilon_3^2} = [w]/[w]^2 = 1/[w]$$

$$\mathrm{cov}(m_1, m_2) \equiv \overline{\varepsilon_1 \varepsilon_2} = \frac{[ws_1 s_2]}{\{[w(s_3^2 - s_1^2)]\}\{[w(s_3^2 - s_2^2)]\}}$$

$$\mathrm{cov}(m_i, d_c) \equiv \overline{\varepsilon_i \varepsilon_3} = \frac{[ws_i]}{\{[w(s_3^2 - s_i^2)]\}[w]}, \quad i = 1, 2.$$

For the earlier, more general expressions for the components of $\overline{\varepsilon \varepsilon^T}$ it is still necessary to find $\overline{\eta_{ki} \eta_{lj}}$ and $\eta_{ki} \xi_{l3}$ in terms of $\xi_{ki} \xi_{lj}$, with $\eta_{ki} \equiv \delta s_{ki} = \delta(r_{ki} - \langle r_{ki}\rangle) = \xi_{ki} - \langle \xi_i\rangle = \sum_l w_l (\xi_{ki} - \xi_{li})/[w]$.

$$\overline{\eta_{ki} \eta_{pj}} = \sum_{l, q} w_l w_q \overline{(\xi_{ki} - \xi_{li})(\xi_{pj} - \xi_{qj})}/[w]^2$$

$$= \overline{(\xi_{ki} - \langle \xi_i\rangle)(\xi_{pj} - \langle \xi_j\rangle)}$$

$$\overline{\eta_{ki} \xi_{p3}} = \sum_l w_l \overline{(\xi_{ki} - \xi_{li})\xi_{p3}}/[w] = \overline{\xi_{ki} \xi_{p3}} - \overline{\langle \xi_i\rangle \xi_{p3}}.$$

In the isotropic, 'no-correlation' case, for example, these reduce to

$$\overline{\eta_{ki} \eta_{pi}} = -w_k \overline{\xi_{ki}^2}/[w] - w_p \overline{\xi_{pi}^2}/[w]$$
$$+ [w^2 \overline{\xi_i^2}]/[w]^2, \quad k \neq p; i = 1, 2$$

$$\overline{\eta_{ki}^2} = (1 - 2w_k/[w])\overline{\xi_{ki}^2} + [w^2 \overline{\xi_i^2}]/[w]^2, \quad i = 1, 2$$

$$\overline{\eta_{k3} \eta_{p3}} = -w_p \overline{\xi_{p3}^2}/[w],$$

and

$$\overline{\eta_{k3}^2} = \overline{\xi_{k3}^2} - w_k \overline{\xi_{k3}^2}/[w] = \overline{\xi_{k3}^2}(1 - w_k/[w]).$$

Here the difference between the correct covariance values and the values obtained on ignoring the variation in $\langle r\rangle$ may be important if the number of defining atoms is small, say, 5 or 4 or, in the extreme, 3.

3.2.2.2. *The standard uncertainty of the distance from an atom to the plane*

There are two cases, as has been pointed out, *e.g.*, by Ito (1982).

(1) The atom (atom K) was *not* included in the specification of the plane.

$$d_K = \mathsf{m}^T(\mathsf{r}_K - \langle \mathsf{r}\rangle) = r_{k3} - \langle r_3\rangle$$
$$\delta d_K = \xi_{K3} + s_{K1} \varepsilon_1 + s_{K2} \varepsilon_2 - \varepsilon_3$$
$$\sigma_{d_K}^2 = \overline{\xi_{K3}^2} + s_{K1}^2 \overline{\varepsilon_1^2} + s_{K2}^2 \overline{\varepsilon_2^2} + \overline{\varepsilon_3^2}$$
$$+ 2s_{K1} s_{K2} \overline{\varepsilon_1 \varepsilon_2} - 2s_{K1} \overline{\varepsilon_1 \varepsilon_3} - 2s_{K2} \overline{\varepsilon_2 \varepsilon_3}$$
$$+ 2s_{K1} \overline{\xi_{K3} \varepsilon_1} + 2s_{K2} \overline{\xi_{K3} \varepsilon_2} - 2\overline{\xi_{K3} \varepsilon_3}.$$

In the isotropic, 'no-correlation' case the last three terms, *i.e.* the terms in $\overline{\varepsilon_i \xi_{K3}}$, are all negligible or zero.

In either case the value for $\overline{\xi_{K3}^2}$ and the appropriate $\overline{\varepsilon_i \varepsilon_j}$ values from the least-squares-plane calculation need to be inserted.

(2) Atom K *was* included in the specification of the plane. The expression for $\sigma_{d_K}^2$ remains the same, but the averages in it may be importantly different.

For example, consider a plane defined by only three atoms, one of overwhelmingly great w at $(0, 0, 0)$, one at $(1, 0, 0)$ and one at $(0, 1, 0)$. The centroid is at $(0, 0, 0)$ and we take $K = 2$, *i.e.* σ_{d_2} is the item of interest. Of course, it is obvious without calculation that the standard uncertainties vanish for the distances of the three atoms from the plane they alone define; the purpose here is only to show, in one case for one of the atoms, that the calculation gives the same result, partly, it will be seen, because the change in orientation of the plane is taken into account. If the only variation in the atom positions is described by $\overline{\xi_{23}^2} = \sigma^2$, one has $s_{21} = 1, \varepsilon_3 = \varepsilon_2 = 0$, $\varepsilon_1 = -\xi_{23}$, and $\overline{\xi_{K3} \varepsilon_1} = \sigma^2$. The non-vanishing terms in the desired variance are then

$$\sigma_{d_2}^2 = \overline{\xi_{23}^2} + s_{21}^2 \overline{\varepsilon_1^2} + 2s_{21} \overline{\xi_{23} \varepsilon_1}$$
$$= (1 + 1 - 2)\sigma^2 = 0.$$

If, however, the problem concerns the same plane and a fourth atom at position $(1, 0, r_{43})$, *not* included in the specification of the plane and uncertain only in respect to r_{43} (which is arbitrary) with $\overline{\xi_{43}^2} = \sigma^2$ (the same mean-square variation in direction 3 as for atom 2) and $\overline{\xi_{43} \xi_{23}} = 0$, the calculation for $\sigma_{d_4}^2$ runs the same as before, except for the third term:

$$\sigma_{d_4}^2 = (1 + 1 - 0)\sigma^2 = 2\sigma^2.$$

Extreme examples of this kind show clearly enough that variation in the direction of the plane normal or in the normal component of the centroid position will sometimes be important, the remarks to the contrary by Shmueli (1981) and, for the centroid, the omission by WMC notwithstanding. If only a few atoms are used to define the plane (*e.g.*, three or, as is often the case, a very few more), both the covariance with the centroid position and uncertainty in the direction of the normal are likely to be important. The uncertainty in the normal may still be important, even if a goodly number of atoms are used to define the plane, whenever the test atom lies near or beyond the edge of the lateral domain defined by the other atoms.

3.2.3. **The proper least-squares plane, with Gaussian weights**

If it is desired to weight the points to be fitted by a plane in the sense of Gaussian least squares, then two items different from what we have seen in the crystallographic literature have to be brought into view: (1) the weights may be anisotropic and include interatomic correlations, because the error matrix of the atom coordinates may in general be anisotropic and include interatomic correlations; and (2) it has to be considered that the atoms truly lie on a plane and that their observed positions are to be adjusted to lie precisely on that plane, whatever its precise position may turn out to be and no matter what the direction, in response to the anisotropic weighting, of their approach to the plane.

An important consequence of (1), the non-diagonal character of the weight matrix, even with Cartesian coordinates, is that the problem is no longer an ordinary eigenvalue problem as treated by SWMB (1959),* not even if there is no interatomic correlation and

---

* A simple two-dimensional problem illustrates the point. A regular polygon of n atoms is to define a 'best' line (always a *central* line). If the error matrix (the same for each atom) is isotropic, the weighted sum of squares of deviations from the line is independent of its orientation for $n > 2$, *i.e.* the problem is a degenerate eigenvalue problem, with two equal eigenvalues. However, if the error ellipsoids are not isotropic and are all oriented radially or all tangentially (these are merely the two orientations tried), the sum has $n/2$ equal minima for even n and 2 equal minima for odd n, in the one-π range of possible orientations of the line.

Possibly similar peculiarities might be imagined if the anisotropic weights were more complicated (*e.g.*, 'star' shaped) than can be described by a non-singular matrix, or by any matrix. Such are of course excluded here.

the anisotropy is the same for each atom. On this last case the contrary remark in SWMB at the beginning of the footnote, p. 601, is incorrect, and the treatments in terms of the eigenvector–eigenvalue problem by Hamilton (1961, 1964, pp. 174–177) and Ito (1981a)* evade us. At best the problem is still not a genuine eigenvalue problem if the anisotropies of the atoms are not all alike.

Hypothesis (2), of perfect planarity, may be hard to swallow. It has to be tested, and for any set of atoms the conclusion may be that they probably do not lie on a plane. But if the hypothesis is provisionally adopted (and it has to be decided beforehand which of the following alternatives is to be followed), the adjusted positions are obtained by moving the atoms onto the plane

(a) along paths normal to the plane, or

(b) along the *proper* paths of 'least resistance' – that is, paths with, in general, both normal and lateral components differently directed for each atom so as to minimize the appropriately weighted quadratic form of differences between the observed and adjusted coordinates. The lateral motions (and the anisotropic weights that induce them) may change the relative weights of different atoms in accordance with the orientation of the plane; change the perpendicular distance of origin-to-plane; and change the orientation of the plane in ways that may at first give surprise.

Our first example of this has already been given.† A second, which actually inspired our first, is to be found in Hamilton (1964; example 5-8-1, p. 177), who discusses a rectangle of four points with identical error ellipsoids elongated in the long direction of the rectangle. The unweighted best line bisects the pattern along this direction, but the weighted best line is parallel to the short direction, if the elongation of the ellipsoids is sufficient. A third example (it is severely specialized so that a precise result can be attained without calculation) has three atoms *ABC* arranged like a $C_{2v} - mm$ molecule with bond angle 90°. The central atom, *B*, has overwhelming isotropic weight; *A* and *C* have parallel extremely elongated error ellipsoids, aligned parallel to the *A—B* bond. The unweighted best line passes through *B* parallel to $A \cdots C$; the weighted best line passes through *B* and through *C*. Our last example is of a plane defined by a number of atoms of which one lies a considerable distance above the plane and at a distance from the normal through the centroid, but with the downward semi-axis of its extremely elongated prolate error ellipsoid intersecting that normal before it intersects the plane. If this atom is moved at right angles to the plane and further away from it, the centroid normal tips toward the atom, whereas it would tip away if the atom's weight function were isotropic or if the calculation were the usual one and in effect constrained the adjusted position of the atom to move at right angles to the plane.

The lead notion here – that the observed points are to be adjusted individually to fit a curve (surface) of required type exactly, rather than that the curve should simply be constructed to fit the observed points as well as possible in the sense of minimizing the weighted sum of squares of the distances along some preordained direction (perhaps normal to the plane, but perhaps as in ordinary curve fitting parallel to the *y* axis) – we first learned from the book by Deming (1943), *Statistical Adjustment of Data*, but it is to be found in Whittaker & Robinson (1929), Arley & Buch (1950), Hamilton (1964, our most used reference), and doubtless widely throughout the least-squares literature. It has recently been strongly emphasized by Lybanon (1984), who gives a number of modern references. It is the only prescription that properly satisfies the least-squares

conditions, whereas (a) and other analogous prescriptions are only arbitrary regressions, in (a) a normal regression onto the plane.‡

We have explored the problem of least-squares *adjustment* of observed positions subject to anisotropic weights with the help of three Fortran programs, one for the straight line and two for the plane. In the first plane program an approximate plane is derived, coordinates are rotated as in WMC (1973), and the parameters of the plane are adjusted and the atoms moved onto it, either normally or in full accord with the least-squares condition, but in either case subject to independent anisotropic weight matrices. The other plane program, described in Appendix 3.2.1, proceeds somewhat more directly, with the help of the method of Lagrange multipliers. However, neither program has been brought to a satisfactory state for the calculation of the variances and covariances of the derived quantities.

3.2.3.1. *Formulation and solution of the general Gaussian plane*

We conclude with an outline for a complete feasible solution, including interatomic weight-matrix elements. Consider atoms at observed vector positions $\mathbf{r}_k$, $k = 1, \ldots, n$, designated in the following equations by $\mathbf{R}$, an *n*-by-3 array, with $\mathbf{R}_{ki} = r_{ki}$; the corresponding adjusted positions denoted by the array $\mathbf{R}_a$; *n* constraints (each adjusted position $\mathbf{r}_a - a$ for 'adjusted' – must be on the plane), and $3n + 3$ adjustable parameters ($3n$ $\mathbf{R}_a$ components and the 3 components of the vector $\mathbf{b}$ of reciprocal intercepts of the plane), so that the problem has $n - 3$ degrees of freedom. The $3n$-by-$3n$ weight matrix $\mathbf{P}$ may be anisotropic for the separate atoms, and may include interatomic elements; it is symmetric in the sense $\mathbf{P}_{kilj} = \mathbf{P}_{ljki}$, but $\mathbf{P}_{kilj}$ will not in general be equal to $\mathbf{P}_{kjli}$. The array Λ denotes *n* Lagrange multipliers, one for each atom and unrelated to the λ's of Section 3.2.2; *m* and *d* still represent the direction cosines of the plane normal and the perpendicular origin-to-plane distance.

For a linear least-squares problem we know (see, *e.g.*, Hamilton, 1964, p. 143) that the proper weight matrix is the reciprocal of the atomic error matrix ($\mathbf{P} = \mathbf{M}^{-1}$);§ note that '$\mathbf{M}$' is unrelated to the '$\mathbf{M}$' of Section 3.2.2. The least-squares sum is now

$$S = (\mathbf{R} - \mathbf{R}_a)P(\mathbf{R} - \mathbf{R}_a),$$

and the augmented sum for applying the method of Lagrange multipliers is

$$\Phi = S/2 - \Lambda\mathbf{b}\mathbf{R}_a.$$

---

* Ito observes that his method fails when there are only three points to define the plane, his least-squares normal equations becoming singular. But the situation is worse: his equations are singular for any number of points, if the points fit a plane exactly.

† See first footnote.

‡ Ito's second method (Ito, 1981b), of 'substitution', is also a regression, essentially like the regression along *z* at fixed *x* and *y* used long ago by Clews & Cochran (1949, p. 52) and like the regressions of *y* on fixed *x* that – despite the fact that both *x* and *y* are afflicted with random errors – are commonly taught or practised in schools, universities and laboratories nearly 200 years after Gauss, to the extent that Deming, Lybanon and other followers of Gauss have so far had rather little influence. Kalantar's (1987) short note is a welcome but still rare exception.

§ Is this statement firm for a nonlinear problem? We use it, assuming that at convergence the problem has become effectively linear. But in fact this will depend on how great the nonlinearity is, in comparison with the random errors (variances) that eventually have to be considered. Another caveat may be in order in regard to our limited knowledge of Gauss's second derivation of the method of least squares, the one he preferred [see Whittaker & Robinson (1929)] and which establishes for a linear system that the best *linear combination* of a set of observations, afflicted by random errors, for estimating any arbitrary derived quantity – best in the sense of being unbiased and having minimal mean-square error – is given by the method of least squares with the weight matrix set equal to the inverse error matrix of the observations. Hamilton, and Whittaker & Robinson, prove this only for the case that the derived parameters are not constrained, whereas here they are. We believe, however, that the best choice of weights is a question concerning only the observations, and that it cannot be affected by the method used for minimizing *S* subject to any constraints, whether by eliminating some of the parameters by invoking the constraints directly or by the use of Lagrange multipliers.

Here the summation brackets ([...]) have been replaced by matrix products, but the simplified notation is not matrix notation. It has to be regarded as only a useful mnemonic in which all the indices and their clutter have been suppressed in view of their inconveniently large number for some of the arrays. If the dimensions of each are kept in mind, it is easy to recall the indices, and if the indices are resupplied at any point, it is not difficult to discover what is really meant by any of the expressions or whether evaluations have to be executed in a particular order.

The conditions to be satisfied are

$$0 = \frac{-\partial \Phi}{\partial \mathbf{R}_a} = \boldsymbol{P}(\mathbf{R} - \mathbf{R}_a) + \Lambda \mathbf{b},$$

$$0 = \frac{-\partial \Phi}{\partial \mathbf{b}} = \Lambda \mathbf{R}_a,$$

$$\mathbf{1} = \mathbf{b} \mathbf{R}_a.$$

That the partial derivatives of $S/2$ should be represented by $\boldsymbol{P}(\mathbf{R} - \mathbf{R}_a)$ depends upon the above-mentioned symmetry of $\boldsymbol{P}$. Note that each of the n unit elements of $\mathbf{1}$ expresses the condition that its $\mathbf{r}_a$ should indeed lie on the plane, and that $\Lambda \mathbf{b}$ is just the same as $\mathbf{b}\Lambda$. The perpendicular origin-to-plane distance and the direction cosines of the plane normal are $d^2 = 1/\mathbf{b}^T \mathbf{b}$ and $\mathbf{m} = \mathbf{b}/\sqrt{\mathbf{b}^T \mathbf{b}}$.

On multiplication by $\boldsymbol{M}$ the first condition solves for $\mathbf{R}_a$, and that expression combined separately with the second condition and with the third gives highly nonlinear equations (there is further mention of this nonlinearity in Appendix 3.2.1) that have to be solved for $\mathbf{b}$ and Λ:

$$\mathbf{R}_a = \mathbf{R} + \boldsymbol{M} \Lambda \mathbf{b} = \mathbf{R} + \boldsymbol{M} \mathbf{b} \Lambda$$

$$0 = \mathbf{F} \equiv (\Lambda \boldsymbol{M} \Lambda)\mathbf{b} + \Lambda \mathbf{R}$$

$$0 = \mathbf{G} \equiv (\mathbf{b}\boldsymbol{M}\mathbf{b})\Lambda - (\mathbf{1} - \mathbf{b}\mathbf{R}).$$

The best way of handling these equations, at least if it is desired to find their solutions both for $\mathbf{R}_a$, Λ, and $\mathbf{b}$ and for the error matrix of these derived quantities in terms of the error matrix for $\mathbf{R}$, seems to be to find an approximate solution by one means or another, to make sure that it is the desired solution, if (as may be) there is ambiguity, and to shift the origin to a point far enough from the plane and close enough to the centroid normal to avoid the difficulties discussed by SWMB. Then linearize the first and second equations by differentiation and solve the results first iteratively to fit the current residuals $\mathbf{F}_0$ and $\mathbf{G}_0$ and then for nominally infinitesimal increments $\delta \mathbf{b}$ and $\delta \Lambda$. In effect, one deals with equations $\mathbf{Q}\mathbf{X} = \mathbf{Y}$, where $\mathbf{Q}$ is the $(n+3) \times (n+3)$ matrix of coefficients of the following set of equations, $\mathbf{X}$ is the $(n+3)$-dimensional vector of increments $\delta \mathbf{b}$ and $\delta \Lambda$, and $\mathbf{Y}$ is the vector either of the first terms or of the second terms on the right-hand side of the equations.

$$(\Lambda \boldsymbol{M} \Lambda)\delta \mathbf{b} + (\boldsymbol{M}\Lambda \mathbf{b} + \Lambda \boldsymbol{M}\mathbf{b} + \mathbf{R})\delta \Lambda = -\mathbf{F}_0 - \Lambda \delta \mathbf{R}$$

$$(\boldsymbol{M}\mathbf{b}\Lambda + \mathbf{b}\boldsymbol{M}\Lambda + \mathbf{R})\delta \mathbf{b} + (\mathbf{b}\boldsymbol{M}\mathbf{b})\delta \Lambda = -\mathbf{G}_0 - \mathbf{b}\delta \mathbf{R}.$$

When $\mathbf{X}$ becomes the vector ε of errors in $\mathbf{b}$ and Λ as induced by errors $\eta \equiv \delta \mathbf{R}$ in the measured atom positions, these equations become, in proper matrix notation, $\mathbf{Q}\varepsilon = \mathbf{Z}\eta$, with solution $\varepsilon = \mathbf{Q}^{-1}\mathbf{Z}\eta$, where $\mathbf{Z}$ is the $(n+3)$-dimensional vector of components, first of $\mathbf{b}$ then of Λ. The covariance matrix $\overline{\varepsilon \varepsilon^T}$, from which all the covariances of $\mathbf{b}$, $\mathbf{R}_a$, and $\mathbf{R}$ (including for the latter any atoms that were not included for the plane) can be derived, is then given by

$$\overline{\varepsilon \varepsilon^T} = \mathbf{Q}^{-1}\mathbf{Z}\overline{\eta \eta^T}\mathbf{Z}^T(\mathbf{Q}^{-1})^T.$$

This is not as simple as the familiar expression for propagation of errors for a least-squares problem solved without the use of

Lagrange multipliers, i.e. $\overline{\varepsilon \varepsilon^T} = \mathbf{B}^{-1}$, where $\mathbf{B}$ is the matrix of the usual normal equations, both because $\mathbf{B} = \mathbf{B}^T$ is no longer valid and because the middle factor $\mathbf{Z}\eta \eta^T \mathbf{Z}^T$ is no longer equal to $\mathbf{B}^{-1}$.

It is easy to verify that Λ consists of a set of coefficients for combining the n row vectors of $\boldsymbol{M}\mathbf{b}$, in the expression for $\mathbf{R}_a$, into corrections to $\mathbf{R}$ such that each adjusted position lies exactly on the plane:

$$\mathbf{b}\mathbf{R}_a = \mathbf{b}\mathbf{R} + \mathbf{b}\boldsymbol{M}\mathbf{b}(\mathbf{b}\boldsymbol{M}\mathbf{b})^{-1}(\boldsymbol{I} - \mathbf{b}\mathbf{R})$$

$$= \mathbf{b}\mathbf{R} + \boldsymbol{I} - \mathbf{b}\mathbf{R} = \boldsymbol{I}.$$

At the same time one can see how the lateral shifts occur in response to the anisotropy of $\boldsymbol{M}$, especially if, now, only the anisotropic case without interatomic correlations is considered. For atom k write $\mathbf{b}$ in terms of its components along the principal axes of $\mathbf{M}_k$, associated with the eigenvalues α, β and γ; the shifts are then proportional to $\mathbf{M}_{\alpha\alpha}b_\alpha^2$, $\mathbf{M}_{\beta\beta}b_\beta^2$ and $\mathbf{M}_{\gamma\gamma}b_\gamma^2$, each along its principal axis, and the sums of the contributions of these shift components to the total displacement along the plane normal or along either of two orthogonal directions in the plane can readily be visualized. In effect $\mathbf{M}_k$ is the compliance matrix for these shifts of atom k. Similarly, it can be seen that in the isotropic case with interatomic correlations a pair of equally weighted atoms located, for example, at some distance apart and at about the same distance from the plane, will have different shifts (and different influences on d and $\mathbf{m}$) depending on whether the covariance between the errors in the perpendicular components of their observed positions relative to the plane is small, or, if large, whether it is positive or is negative. If the covariance is large and positive, the adjusted positions will both be pulled toward the plane, strongly resisting, however, the apparent necessity that both reach the plane by moving by different amounts; in consequence, there will be a strong tendency for the plane to tilt toward the more distant atom, and possibly even away from the nearer one. But if the covariance is large and negative, the situation is reversed: the more distant atom can readily move more than the nearer one, while it is very difficult to move them together; the upshot is then that the plane will move to meet the original midpoint of the two atoms while they tilt about that midpoint to accommodate the plane.

It is attractive to solve our problem with the 'normal' formulation of the plane, $\mathbf{m}\mathbf{r} = d$, and so directly avoid the problems that arise for $d \approx 0$. The solution in terms of the reciprocal intercepts $\mathbf{b}$ has been given first, however, because reducing by two (d and a Lagrange multiplier) the number of parameters to be solved for may more than make up for the nuisance of having to move the origin. The solution in terms of d follows.

The augmented variation function is now

$$\Phi = (\mathbf{R} - \mathbf{R}_a)\boldsymbol{P}(\mathbf{R} - \mathbf{R}_a)/2 - \Lambda(\mathbf{m}\mathbf{R}_a - d\boldsymbol{I}) + \mu \mathbf{m}\mathbf{m}/2,$$

the term in the new Lagrange multiplier, μ, and the term in $d\boldsymbol{I}$ having been added to the previous expression. The $\boldsymbol{I}$, an n-vector of 1's, is needed to express the presence of n terms in the Λ sum. There are then five equations to be satisfied – actually $n + 1 + 3n + 3 + 1 = 4n + 5$ ordinary equations – in the $3n$ $\mathbf{R}_a$ components, the n Λ's, the 3 $\mathbf{m}$ components, μ, and $d - 4n + 5$ unknowns in all, as required. These equations are as follows:

$$\mathbf{m}\mathbf{R}_a = d\boldsymbol{I}$$

$$\mathbf{m}\mathbf{m} = 1$$

$$0 = -\left(\frac{\partial \Phi}{\partial \mathbf{R}_a}\right) = \boldsymbol{P}(\mathbf{R} - \mathbf{R}_a) + \Lambda \mathbf{m}$$

$$0 = \frac{\partial \Phi}{\partial \mathbf{m}} = -\Lambda \mathbf{R}_a + \mu \mathbf{m}$$

$$0 = \frac{\partial \Phi}{\partial d} = \Lambda I.$$

As before, multiply the third equation by M and solve for $\mathbf{R}_a$. Then substitute the result into the first and fourth equations to obtain

$$\mathsf{m}\mathbf{R} + \mathsf{m}M\mathsf{m}\Lambda = dI,$$

$$\mathsf{m}\mathsf{m} = 1,$$

$$\Lambda\mathbf{R} + \Lambda M \Lambda\mathsf{m} = \mu\mathsf{m},$$

$$\Lambda I = 0$$

as the $n + 5$ mostly nonlinear equations to be solved for m, Λ, d and μ by linearizing (differentiation), solving for increments, and iterating, in the pattern described more fully above. An approximate solution for m and d has first to be obtained somehow, perhaps by the method of SWMB (with isotropic uncorrelated weights), checked for suitability, and extended to a full complement of first approximations by

$$\Lambda = (\mathsf{m}M\mathsf{m})^{-1}(dI - \mathsf{m}\mathbf{R})$$

$$\mu = \mathsf{m}\Lambda\mathbf{R} + \mathsf{m}\Lambda M \Lambda\mathsf{m},$$

which readily follow from the previous equations. As in the 'intercepts' solution the linearized expression for the increments in μ, Λ, d and m can be used together with the equation for $\mathbf{R}_a$ to obtain all the covariances needed in the treatment described in Section 3.2.2.

3.2.3.2. *Concluding remarks*

Proper tests of statistical significance of this or that aspect of a least-squares plane can be made if the plane has been based on a proper weight matrix as discussed in Section 3.2.3; if it can be agreed that the random errors of observation are normally distributed; and if an agreeable test (*null*) hypothesis can be formulated. For example, one may ask for the probability that a degree of fit of the observed positions to the least-squares plane at least as poor as the fit that was found might occur if the atoms in truth lie precisely on a plane. The χ^2 test answers this question: a table of probabilities displayed as a function of χ^2 and ν provides the answer. Here χ^2 is just our minimized

$$S = \mathsf{b}\Lambda M P M \Lambda\mathsf{b} = \mathsf{b}\Lambda M \Lambda\mathsf{b},$$

and

$$\nu = n_{\text{observations}} - n_{\text{adjusted parameters}} - n_{\text{constraints}}$$

$$= 3n - (n + 3) - n = n - 3,$$

is the number of degrees of freedom for the problem of the plane (erroneously cited in at least one widely used crystallographic system of programs as $3n - 3$). There will not usually be any reason to believe that the atoms are exactly coplanar in any case; nevertheless, this test may well give a satisfying indication of whether or not the atoms are, in the investigator's judgment, essentially coplanar. It must be emphasized that χ^2 as calculated in Section 3.2.3 will include proper allowance for uncertainties in the d and orientation of the plane with greater reliability than the estimates of Section 3.2.2, which are based on nominally arbitrary weights. Both, however, will allow for the large variations in d and *tilt* that can arise in either case if n is small. Some of the earlier, less complete discussions of this problem have been mentioned in Section 3.2.2.

Among the problems not considered here are ones of fitting more than one plane to a set of observed positions, *e.g.* of two planes fitted to three sets of atoms associated, respectively, with the first plane, the second plane, and both planes, and of the angle between the two planes. For the atoms common to both planes there will be a fundamental point of difference between existing programs (in which, in effect, the positions of the atoms in common are inconsistently adjusted to one position on the first plane and, in general, a different position on the second) and what we would advocate as the proper procedure of requiring the adjusted positions of such atoms to lie on the line of intersection of the two planes. As to the dihedral angle there is a difficulty, noted by WMC (1973, p. 2705), that the usual formulation of $\sigma^2(\theta_0)$ in terms of the cosine of the dihedral angle reduces to 0/0 at $\theta_0 = 0$. However, this variance is obviously well defined if the plane normals and their covariances are well defined. The essential difficulty lies with the ambiguity in the direction of the line of intersection of the planes in the limit of zero dihedral angle. For the torsion angle about a line defined by two atoms, there should be no such difficulty. It seems likely that for the two-plane problem proposed above, the issue that decides whether the dihedral angle will behave like the standard dihedral angle or, instead, like the torsion angle, will be found to be whether or not two or more atoms are common to both planes.

All that we have tried to bring out about the covariances of derived quantities involving the plane requires that the covariances of the experimental atom positions (reduced in our formulations to Cartesian coordinates) be included. However, such covariances of derived quantities are often not available in practice, and are usually left unused even if they are. The need to use the covariances, not just the variances, has been obvious from the beginning. It has been emphasized in another context by Schomaker & Marsh (1983) and much more strongly and generally by Waser (1973), whose pleading seems to have been generally ignored, by now, for about thirty years.

Appendix 3.2.1.

Consider n atoms at observed vector positions $\mathbf{r}$ (expressed in Cartesians), n constraints (each adjusted position $\mathbf{r}_a - a$ for 'adjusted' – must be on the plane) and $3n + 3$ adjustable parameters ($3n$ $\mathbf{r}_a$ components and the 3 components of the vector $\mathbf{a}$ of reciprocal intercepts of the plane), so that the problem has $n - 3$ degrees of freedom. The weight matrices $\mathbf{P}$ may be differently anisotropic for each atom, but there are no interatomic correlations. As before, square brackets, '[...]', represent the Gaussian sum over all atoms, usually suppressing the atom indices. We also write λ, *not* the λ of Section 3.2.2, for the Lagrange multipliers (one for each atom); m for the direction cosines of the plane normal; and d for the perpendicular origin-to-plane distance.

As before, $\mathbf{P}_k$ is the reciprocal of the atomic error matrix: $\mathbf{P}_k = \mathbf{M}_k^{-1}$ (correspondingly, $P \equiv M^{-1}$) but '$\mathbf{M}$' is no longer the '$\mathbf{M}$' of Section 3.2.2. The appropriate least-squares sum is

$$S = [(\mathbf{r} - \mathbf{r}_a)^T \mathbf{P}(\mathbf{r} - \mathbf{r}_a)]$$

and the augmented sum for applying the method of Lagrange multipliers is

$$\Phi = S/2 - [\lambda \mathbf{a}^T \mathbf{r}_a].$$

Φ is to be minimized with respect to variations of the adjusted atom positions $\mathbf{r}_{ka}$ and plane reciprocal intercepts b_i, leading to the equations

$$0 = \frac{-\partial \phi}{\partial \mathbf{r}_a^T} = \mathbf{P}(\mathbf{r} - \mathbf{r}_a) + \lambda\mathbf{b} \text{ and}$$

$$0 = \frac{-\partial \Phi}{\partial \mathbf{b}^T} = [\lambda \mathbf{r}_a],$$

subject to the plane conditions $\mathbf{b}^T \mathbf{r}_a = 1$, each atom, with $d^2 = 1/(\mathbf{b}^T \mathbf{b})$, $\mathsf{m} = \mathbf{b}/\sqrt{\mathbf{b}^T \mathbf{b}}$. These equations are nonlinear.

A convenient solution runs as follows: first multiply the first equation by $\mathbf{M}$ and solve for the adjusted atom positions in terms of the Lagrange multipliers λ and the reciprocal intercepts $\mathbf{b}$ of the plane; then multiply *that* result by $\mathbf{b}^T$ applying the plane conditions, and solve for the λ's

$$\mathbf{r}_a = \mathbf{r} + \lambda\mathbf{Mb}, \quad \mathbf{M} \equiv \mathbf{P}^{-1}$$

$$1 = \mathbf{b}^T\mathbf{r} + \lambda\mathbf{b}^T\mathbf{Mb},$$

$$\lambda = (1 - \mathbf{b}^T\mathbf{r})/(\mathbf{b}^T\mathbf{Mb}).$$

Next insert these values for the λ's and $\mathbf{r}_a$'s into the second equation:

$$0 = \frac{\partial\Phi}{\partial\mathbf{b}^T} = [\lambda\mathbf{r}_a] = \left[\frac{1 - \mathbf{b}^T\mathbf{r}}{\mathbf{b}^T\mathbf{Mb}}\left(\mathbf{r} + \frac{1 - \mathbf{b}^T\mathbf{r}}{\mathbf{b}^T\mathbf{Mb}}\mathbf{Mb}\right)\right].$$

This last equation, $F(\mathbf{b}) = 0$, is to be solved for $\mathbf{b}$. It is highly nonlinear: $F(\mathbf{b}) = O(\mathbf{b}^3)/O(\mathbf{b}^4)$. One can proceed to a first approximation by writing $0 = [(1 - \mathbf{b}^T\mathbf{r})\mathbf{r}]$; *i.e.*, $[\mathbf{rr}] \cdot \mathbf{b} = [\mathbf{r}]$, in dyadic notation. [$\mathbf{M} = \mathbf{I}$, all atoms; $1 - \mathbf{b}^T\mathbf{r} = 0$ in the multiplier of $\mathbf{Mb}/(\mathbf{b}^T\mathbf{Mb})$.] A linear equation in $\mathbf{b}$, this approximation usually works well.* We have also used the iterative Frazer, Duncan & Collar eigenvalue solution as described by SWMB (1959), which works even when the plane passes exactly through the origin. To continue the solution of the nonlinear equations, by linearizing and iterating, write $F(\mathbf{b}) = 0$ in the form

$$0 = [\lambda\mathbf{r} + \lambda^2\mathbf{Mb}]$$

$$\approx \left[\mathbf{r}\frac{\partial\lambda}{\partial\mathbf{b}} + 2\mathbf{Mb}\lambda\frac{\partial\lambda}{\partial\mathbf{b}} + \lambda^2\mathbf{M}\right]\delta\mathbf{b} + [\lambda(\mathbf{r} + \lambda\mathbf{Mb})]_0,$$

solve for $\delta\mathbf{b}$, reset $\mathbf{b}$ to $\mathbf{b} + \delta\mathbf{b}$, *etc.*, until the desired degree of convergence of $|\delta\mathbf{b}|/|\mathbf{b}|$ toward zero has been attained. By $\partial\lambda/\partial\mathbf{b}$ is meant the partial derivative of the above expression for λ with respect to $\mathbf{b}$, as detailed in the next paragraph.

In the Fortran program *DDLELSP* (double precision Deming Lagrange, with error estimates, least-squares plane, written to explore this solution) the preceding equation is recast as

$$\mathbf{B}\delta\mathbf{b} \equiv \left[\lambda^2\mathbf{M} + (\mathbf{r} + 2\lambda\mathbf{Mb})\frac{\partial\lambda}{\partial\mathbf{b}}\right]\delta\mathbf{b}$$

$$\equiv \mathbf{Y} = -[\lambda(\mathbf{r} + \lambda\mathbf{Mb})]_0,$$

---

* We do not fully understand the curious situation of this equation. It arises immediately if the isotropic problem is formulated as one of minimizing $[(1 - \mathbf{b}^T\mathbf{r})^2]$ by varying $\mathbf{b}$, and it fails then [SWMB (1959) referred to it as 'an incorrect method'], as it obviously must – observe the denominator – if the plane passes too close to the origin. However, it fails in other circumstances also. The main point about it is perhaps that it is linear in $\mathbf{b}$ and is obtained as the supposedly exact and *unique* solution of the isotropic problem, whereas the problem has no unique solution but three solutions instead (SWMB, 1959). From the point of view of Gaussian least squares, the essential fault in minimizing $S_{\text{lin}} = [(1 - \mathbf{b}^T\mathbf{r})^2]$ may be that the apparently simple weighting function in it, *i.e.* the identity, is actually complicated and unreasonable. In terms of distance deviations from the plane, we have $S_{\text{lin}} = [w(d - \mathbf{m}^T\mathbf{r})^2]$, with $w = \mathbf{b}^T\mathbf{b} = d^{-2}$. Prudence requires that the origin be shifted to a point sufficiently far from the plane and close enough to the centroid normal to avoid the difficulties discussed by SWMB. Note that for the one-dimensional problem of fitting a constant to a set of measurements of a single entity the Deming–Lagrange treatment with the condition $1 = cx_a$ and weights w reduces immediately to the standard result $1/c = [wx]/[w]$.

with

$$\frac{\partial\lambda}{\partial\mathbf{b}} = \mathbf{r}^T/\mathbf{b}^T\mathbf{Mb} - [2(1 - \mathbf{b}^T\mathbf{r})/(\mathbf{b}^T\mathbf{Mb})^2]$$

$$= -(\mathbf{r}^T + 2\lambda\mathbf{b}^T\mathbf{M})/\mathbf{b}^T\mathbf{Mb}.$$

The usual goodness of fit, GOF2 in *DDLELSP*, evaluates to

$$\text{GOF2} = \left(\frac{S_{\min}}{n-3}\right)^{1/2} = \left(\frac{1}{n-3}[\lambda^2\mathbf{b}^T\mathbf{MPMb}]\right)^{1/2}$$

$$= \left(\frac{1}{n-3}[\lambda^2\mathbf{b}^T\mathbf{Mb}]\right)^{1/2}$$

$$= \left(\frac{1}{n-3}\left[\frac{(1 - \mathbf{b}^T\mathbf{r})^2}{\mathbf{b}^T\mathbf{Mb}}\right]\right)^{1/2}.$$

This is only an approximation, because the residuals $1 - \mathbf{b}^T\mathbf{r}$ are not the differences between the observations and appropriate linear functions of the parameters, nor are their variances (the $\mathbf{b}^T\mathbf{Mb}$'s) independent of the parameters (or, in turn, the errors in the observations).

We ask also about the perpendicular distances, e, of atoms to plane and the mean-square deviation $\overline{(\delta e)^2}$ to be expected in e.

$$e = (1 - \mathbf{b}^T\mathbf{r})/\sqrt{\mathbf{b}^T\mathbf{b}} = d(1 - \mathbf{b}^T\mathbf{r})$$

$$\delta e = -d(\mathbf{b}^T\eta + \mathbf{r}^T\varepsilon) + O(\eta^2).$$

Here η and ε are the errors in $\mathbf{r}$ and $\mathbf{b}$. Neglecting '$O(\eta^2)$' then leads to

$$\overline{(\delta e)^2} = d^2(\mathbf{b}^T\overline{\eta\eta^T}\mathbf{b} + 2\mathbf{b}^T\overline{\eta\varepsilon^T}\mathbf{r} + \mathbf{r}^T\overline{\varepsilon\varepsilon^T}\mathbf{r}).$$

We have $\overline{\eta\eta^T} = \mathbf{M} = \mathbf{P}^{-1}$, but $\overline{\eta\varepsilon^T}$ and $\overline{\varepsilon\varepsilon^T}$ perhaps still need to be evaluated.

359

References

3.1

Busing, W. R. & Levy, H. A. (1964). *Effect of thermal motion on the estimation of bond lengths. Acta Cryst.* **17**, 142–146.

Hamilton, W. C. (1964). *Statistics in physical science.* New York: Ronald Press.

Johnson, C. K. (1970). *The effect of thermal motion on interatomic distances and angles.* In *Crystallographic computing*, edited by F. R. Ahmed, pp. 220–226. Copenhagen: Munksgaard.

Johnson, C. K. (1980). *Thermal motion analysis.* In *Computing in crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 14.01–14.19. Bangalore: Indian Academy of Sciences.

Sands, D. E. (1966). *Transformations of variance–covariance tensors. Acta Cryst.* **21**, 868–872.

Sands, D. E. (1977). *Correlation and covariance. J. Chem. Educ.* **54**, 90–94.

Sands, D. E. (1982*a*). *Vectors and tensors in crystallography.* Reading: Addison Wesley. Reprinted (1995) Dover Publications.

Sands, D. E. (1982*b*). *Molecular geometry.* In *Computational crystallography*, edited by D. Sayre, pp. 421–429. Oxford: Clarendon Press.

Shmueli, U. (1974). *On the standard deviation of a dihedral angle. Acta Cryst.* A**30**, 848–849.

Taylor, R. & Kennard, O. (1983). *The estimation of average molecular dimensions from crystallographic data. Acta Cryst.* B**39**, 517–525.

Taylor, R. & Kennard, O. (1985). *The estimation of average molecular dimensions. 2. Hypothesis testing with weighted and unweighted means. Acta Cryst.* A**41**, 85–89.

Waser, J. (1973). *Dyadics and the variances and covariances of molecular parameters, including those of best planes. Acta Cryst.* A**29**, 621–631.

3.2

Arley, N. & Buch, K. R. (1950). *Introduction to the theory of probability and statistics.* New York: John Wiley. London: Chapman & Hall.

Clews, C. J. B. & Cochran, W. (1949). *The structures of pyrimidines and purines. III. An X-ray investigation of hydrogen bonding in aminopyrimidines. Acta Cryst.* **2**, 46–57.

Deming, W. E. (1943). *Statistical adjustment of data.* New York: John Wiley. [First edition 1938.]

Hamilton, W. C. (1961). *On the least-squares plane through a set of points. Acta Cryst.* **14**, 185–189.

Hamilton, W. C. (1964). *Statistics in physical science.* New York: Ronald Press.

Ito, T. (1981*a*). *Least-squares refinement of the best-plane parameters. Acta Cryst.* A**37**, 621–624.

Ito, T. (1981*b*). *On the least-squares plane through a group of atoms. Sci. Pap. Inst. Phys. Chem. Res. Saitama*, **75**, 55–58.

Ito, T. (1982). *On the estimated standard deviation of the atom-to-plane distance. Acta Cryst.* A**38**, 869–870.

Kalantar, A. H. (1987). *Slopes of straight lines when neither axis is error-free. J. Chem. Educ.* **64**, 28–29.

Lybanon, M. (1984). *A better least-squares method when both variables have uncertainties. Am. J. Phys.* **52**, 22–26.

Robertson, J. M. (1948). *Bond-length variations in aromatic systems. Acta Cryst.* **1**, 101–109.

Schomaker, V. & Marsh, R. E. (1983). *On evaluating the standard deviation of U_{eq}. Acta Cryst.* A**39**, 819–820.

Schomaker, V., Waser, J., Marsh, R. E. & Bergman, G. (1959). *To fit a plane or a line to a set of points by least squares. Acta Cryst.* **12**, 600–604.

Shmueli, U. (1981). *On the statistics of atomic deviations from the 'best' molecular plane. Acta Cryst.* A**37**, 249–251.

Waser, J. (1973). *Dyadics and variances and covariances of molecular parameters, including those of best planes. Acta Cryst.* A**29**, 621–631.

Waser, J., Marsh, R. E. & Cordes, A. W. (1973). *Variances and covariances for best-plane parameters including dihedral angles. Acta Cryst.* B**29**, 2703–2708.

Whittaker, E. T. & Robinson, G. (1929). *The calculus of observations.* London: Blackie.

3.3

Abad-Zapatero, C., Abdel-Meguid, S. S., Johnson, J. E., Leslie, A. G. W., Rayment, I., Rossmann, M. G., Suck, D. & Tsukihara, T. (1980). *Structure of southern bean mosaic virus at 2.8 Å resolution. Nature (London)*, **286**, 33–39.

Abi-Ezzi, S. S. & Bunshaft, A. J. (1986). *An implementer's view of PHIGS. IEEE Comput. Graphics Appl.* Vol. 6, Part 2.

Aharonov, Y., Farach, H. A. & Poole, C. P. (1977). *Non-linear vector product to describe rotations. Am. J. Phys.* **45**, 451–454.

Allen, F. H., Bellard, S., Brice, M. D., Cartwright, B. A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B. G., Kennard, O., Motherwell, W. D. S., Rodgers, J. R. & Watson, D. G. (1979). *The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information. Acta Cryst.* B**35**, 2331–2339.

Allinger, N. L. (1976). *Calculation of molecular structure and energy by force field methods. Adv. Phys. Org. Chem.* **13**, 1–82.

Altona, C. & Sundaralingam, M. (1972). *Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. J. Am. Chem. Soc.* **94**(23), 8205–8212.

American National Standards Institute, American National Standard for Information Processing Systems – Computer Graphics – Graphical Kernel System (GKS) Functional Description (1985). ISO 7942, ISO Central Secretariat, Geneva, Switzerland.

American National Standards Institute, American National Standard for Information Processing Systems – Computer Graphics – Programmer's Hierarchical Graphics System (PHIGS) Functional Description, Archive File Format, Clear-Text Encoding of Archive File (1988). ANSI X3.144-1988. ANSI, New York, USA.

Anderson, S. (1984). *Graphical representation of molecules and substructure-search queries in MACCS. J. Mol. Graphics*, **2**, 83–90.

Arnold, D. B. & Bono, P. R. (1988). *CGM and CGI: metafile and interface standards for computer graphics.* Berlin: Springer-Verlag.

Barry, C. D. & North, A. C. T. (1971). *The use of a computer-controlled display system in the study of molecular conformations. Cold Spring Harbour Symp. Quant. Biol.* **36**, 577–584.

Bash, P. A., Pattabiraman, N., Huang, C., Ferrin, T. E. & Langridge, R. (1983). *Van der Waals surfaces in molecular modelling: implementation with real-time computer graphics. Science*, **222**, 1325–1327.

Beddell, C. J. (1970). *An X-ray crystallographic study of the activity of lysozyme.* DPhil thesis, University of Oxford, England.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol.* **112**, 535–542.

Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. (1978). *Protein disk of tobacco mosaic virus at 2.8 Å resolution showing the interactions within and between subunits. Nature (London)*, **276**, 362–368.

Boyd, D. B. & Lipkowitz, K. B. (1982). *Molecular mechanics, the method and its underlying philosophy. J. Chem. Educ.* **59**, 269–274.