

3.2. THE LEAST-SQUARES PLANE

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 nondiagonal 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),¹ not even if there is no interatomic correlation and 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)² 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.

¹ 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 nonsingular matrix, or by any matrix. Such are of course excluded here.

² 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.

Our first example of this has already been given.³ 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.⁴

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 A3.2.1, 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 , $k = 1, \dots, n$, 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

³ See first footnote¹.

⁴ 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.

3. DUAL BASES IN CRYSTALLOGRAPHIC COMPUTING

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 $\mathbf{P}_{kij} = \mathbf{P}_{ljk}$, but \mathbf{P}_{kij} will not in general be equal to \mathbf{P}_{kji} . The array $\mathbf{\Lambda}$ 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 ($\mathbf{P} = \mathbf{M}^{-1}$),⁵ note that ' \mathbf{M} ' is unrelated to the ' \mathbf{M} ' of Section 3.2.2. The least-squares sum is now

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

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

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

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

$$\begin{aligned} 0 &= \frac{\partial \Phi}{\partial \mathbf{R}_a} = \mathbf{P}(\mathbf{R} - \mathbf{R}_a) + \mathbf{\Lambda}\mathbf{b}, \\ 0 &= \frac{\partial \Phi}{\partial \mathbf{b}} = \mathbf{\Lambda}\mathbf{R}_a, \\ \mathbf{1} &= \mathbf{b}\mathbf{R}_a. \end{aligned}$$

That the partial derivatives of $S/2$ should be represented by $\mathbf{P}(\mathbf{R} - \mathbf{R}_a)$ depends upon the above-mentioned symmetry of \mathbf{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 $\mathbf{\Lambda}\mathbf{b}$ is just the same as $\mathbf{b}\mathbf{\Lambda}$. 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 \mathbf{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 A3.2.1) that have to be solved for \mathbf{b} and $\mathbf{\Lambda}$:

$$\begin{aligned} \mathbf{R}_a &= \mathbf{R} + \mathbf{M}\mathbf{\Lambda}\mathbf{b} = \mathbf{R} + \mathbf{M}\mathbf{b}\mathbf{\Lambda} \\ 0 &= \mathbf{F} \equiv (\mathbf{\Lambda}\mathbf{M}\mathbf{\Lambda})\mathbf{b} + \mathbf{\Lambda}\mathbf{R} \\ 0 &= \mathbf{G} \equiv (\mathbf{b}\mathbf{M}\mathbf{b})\mathbf{\Lambda} - (\mathbf{1} - \mathbf{b}\mathbf{R}). \end{aligned}$$

The best way of handling these equations, at least if it is desired to find their solutions both for \mathbf{R}_a , $\mathbf{\Lambda}$, 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 $\delta\mathbf{\Lambda}$. 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 $\delta\mathbf{\Lambda}$, and \mathbf{Y} is the vector either of the first terms or of the second terms on the right-hand side of the equations.

$$\begin{aligned} (\mathbf{\Lambda}\mathbf{M}\mathbf{\Lambda})\delta\mathbf{b} + (\mathbf{M}\mathbf{\Lambda}\mathbf{b} + \mathbf{\Lambda}\mathbf{M}\mathbf{b} + \mathbf{R})\delta\mathbf{\Lambda} &= -\mathbf{F}_0 - \mathbf{\Lambda}\delta\mathbf{R} \\ (\mathbf{M}\mathbf{b}\mathbf{\Lambda} + \mathbf{b}\mathbf{M}\mathbf{\Lambda} + \mathbf{R})\delta\mathbf{b} + (\mathbf{b}\mathbf{M}\mathbf{b})\delta\mathbf{\Lambda} &= -\mathbf{G}_0 - \mathbf{b}\delta\mathbf{R}. \end{aligned}$$

When \mathbf{X} becomes the vector ε of errors in \mathbf{b} and $\mathbf{\Lambda}$ 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 $\mathbf{\Lambda}$. 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 $\mathbf{Z}\overline{\eta\eta^T}\mathbf{Z}^T$ is no longer equal to \mathbf{B}^{-1} .

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

$$\begin{aligned} \mathbf{b}\mathbf{R}_a &= \mathbf{b}\mathbf{R} + \mathbf{b}\mathbf{M}\mathbf{b}(\mathbf{b}\mathbf{M}\mathbf{b})^{-1}(\mathbf{1} - \mathbf{b}\mathbf{R}) \\ &= \mathbf{b}\mathbf{R} + \mathbf{1} - \mathbf{b}\mathbf{R} = \mathbf{1}. \end{aligned}$$

At the same time one can see how the lateral shifts occur in response to the anisotropy of \mathbf{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

⁵ 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.

3.2. THE LEAST-SQUARES PLANE

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 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, $m\mathbf{r} = d$, 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

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

the term in the new Lagrange multiplier, μ , and the term in $d\mathbf{I}$ having been added to the previous expression. The \mathbf{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 m components, μ , and $d - 4n + 5$ unknowns in all, as required. These equations are as follows:

$$\begin{aligned} m\mathbf{R}_a &= d\mathbf{I} \\ mm &= 1 \\ 0 &= -\left(\frac{\partial\Phi}{\partial\mathbf{R}_a}\right) = \mathbf{P}(\mathbf{R} - \mathbf{R}_a) + \Lambda m \\ 0 &= \frac{\partial\Phi}{\partial m} = -\Lambda\mathbf{R}_a + \mu m \\ 0 &= \frac{\partial\Phi}{\partial d} = \Lambda\mathbf{I}. \end{aligned}$$

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

$$\begin{aligned} m\mathbf{R} + m\mathbf{M}\mathbf{m}\Lambda &= d\mathbf{I}, \\ mm &= 1, \\ \Lambda\mathbf{R} + \Lambda\mathbf{M}\Lambda m &= \mu m, \\ \Lambda\mathbf{I} &= 0 \end{aligned}$$

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

$$\begin{aligned} \Lambda &= (m\mathbf{M}m)^{-1}(d\mathbf{I} - m\mathbf{R}) \\ \mu &= m\Lambda\mathbf{R} + m\Lambda\mathbf{M}\Lambda m, \end{aligned}$$

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

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

and

$$\begin{aligned} \nu &= n_{\text{observations}} - n_{\text{adjusted parameters}} - n_{\text{constraints}} \\ &= 3n - (n + 3) - n = n - 3, \end{aligned}$$

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

3. DUAL BASES IN CRYSTALLOGRAPHIC COMPUTING

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 A3.2.1

Consider n atoms at observed vector positions r (expressed in Cartesians), n constraints (each adjusted position r_a – ‘a’ for ‘adjusted’ – must be on the plane) and $3n + 3$ adjustable parameters ($3n r_a$ components and the 3 components of the vector 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, $\mathbf{P} \equiv \mathbf{M}^{-1}$) but ‘ \mathbf{M} ’ is no longer the ‘ \mathbf{M} ’ of Section 3.2.2. The appropriate least-squares sum is

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

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

$$\Phi = S/2 - [\lambda a^T r_a].$$

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

$$\begin{aligned} 0 &= \frac{-\partial\phi}{\partial r_a^T} = \mathbf{P}(r - r_a) + \lambda \mathbf{b} \text{ and} \\ 0 &= \frac{-\partial\Phi}{\partial \mathbf{b}^T} = [\lambda r_a], \end{aligned}$$

subject to the plane conditions $\mathbf{b}^T r_a = 1$, each atom, with $d^2 = 1/(\mathbf{b}^T \mathbf{b})$, $m = \mathbf{b}/\sqrt{\mathbf{b}^T \mathbf{b}}$. 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

$$\begin{aligned} r_a &= r + \lambda \mathbf{M} \mathbf{b}, \quad \mathbf{M} \equiv \mathbf{P}^{-1} \\ 1 &= \mathbf{b}^T r + \lambda \mathbf{b}^T \mathbf{M} \mathbf{b}, \\ \lambda &= (1 - \mathbf{b}^T r)/(\mathbf{b}^T \mathbf{M} \mathbf{b}). \end{aligned}$$

Next insert these values for the λ 's and r_a 's into the second equation:

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

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 $0 = [(1 - \mathbf{b}^T r)r]$; *i.e.*, $[\mathbf{r}\mathbf{r}] \cdot \mathbf{b} = [\mathbf{r}]$, in dyadic notation. $[\mathbf{M} = \mathbf{I}$, all atoms; $1 - \mathbf{b}^T r = 0$ in the multiplier of $\mathbf{M}\mathbf{b}/(\mathbf{b}^T \mathbf{M}\mathbf{b})$.] A linear equation in \mathbf{b} , this approximation usually works well.⁶ 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

$$\begin{aligned} 0 &= [\lambda r + \lambda^2 \mathbf{M} \mathbf{b}] \\ &\approx \left[r \frac{\partial\lambda}{\partial \mathbf{b}} + 2\mathbf{M} \mathbf{b} \lambda \frac{\partial\lambda}{\partial \mathbf{b}} + \lambda^2 \mathbf{M} \right] \delta \mathbf{b} + [\lambda(r + \lambda \mathbf{M} \mathbf{b})]_0, \end{aligned}$$

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

$$\begin{aligned} \mathbf{B} \delta \mathbf{b} &\equiv \left[\lambda^2 \mathbf{M} + (r + 2\lambda \mathbf{M} \mathbf{b}) \frac{\partial\lambda}{\partial \mathbf{b}} \right] \delta \mathbf{b} \\ &\equiv \mathbf{Y} = -[\lambda(r + \lambda \mathbf{M} \mathbf{b})]_0, \end{aligned}$$

with

$$\begin{aligned} \frac{\partial\lambda}{\partial \mathbf{b}} &= r^T / \mathbf{b}^T \mathbf{M} \mathbf{b} - [2(1 - \mathbf{b}^T r)/(\mathbf{b}^T \mathbf{M} \mathbf{b})^2] \\ &= -(\mathbf{r}^T + 2\lambda \mathbf{b}^T \mathbf{M}) / \mathbf{b}^T \mathbf{M} \mathbf{b}. \end{aligned}$$

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

$$\begin{aligned} \text{GOF2} &= \left(\frac{S_{\min}}{n-3} \right)^{1/2} = \left(\frac{1}{n-3} [\lambda^2 \mathbf{b}^T \mathbf{M} \mathbf{P} \mathbf{M} \mathbf{b}] \right)^{1/2} \\ &= \left(\frac{1}{n-3} [\lambda^2 \mathbf{b}^T \mathbf{M} \mathbf{b}] \right)^{1/2} \\ &= \left(\frac{1}{n-3} \left[\frac{(1 - \mathbf{b}^T r)^2}{\mathbf{b}^T \mathbf{M} \mathbf{b}} \right] \right)^{1/2}. \end{aligned}$$

⁶ We do not fully understand the curious situation of this equation. It arises immediately if the isotropic problem is formulated as one of minimizing $[(1 - \mathbf{b}^T r)^2]$ 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 $S_{\min} = [(1 - \mathbf{b}^T r)^2]$ 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 $S_{\min} = [w(d - m^T r)^2]$, 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 $1 = c x_a$ and weights w reduces immediately to the standard result $1/c = [w x]/[w]$.