International Tables for Crystallography (2006). Vol. C, Section 8.4.4, pp. 705–706 .

If we let $|t| = \sqrt{F}$, and use

$$\int\limits_0^{F_0} \Phi(F, 1, \nu)\,\mathrm{d}F = \int\limits_{-t_0}^{+t_0} \Phi(t, \nu)\,\mathrm{d}t, \qquad (8.4.3.4)$$

we can derive a p.d.f. for t, which is

$$\Phi(t, \nu) = \frac{\Gamma[(\nu+1)/2]}{\sqrt{\pi\nu}\,\Gamma(\nu/2)[1+t^2/\nu]^{(\nu+1)/2}}. \qquad (8.4.3.5)$$

This p.d.f. is known as *Student's t distribution with ν degrees of freedom*. Setting $t = \widehat{\lambda}/\widehat{\sigma}_\lambda$, the c.d.f. $\Psi(t, \nu)$ can be used to test the alternative hypotheses $\lambda = 0$ and $\lambda = \pm 1/2$. Table 8.4.3.1 gives the values of t for which the c.d.f. $\Psi(t, \nu)$ has various values for various values of ν. Fortran code for the program from which this table was generated appears in Prince (1994).

Again, it must be understood that the results of these statistical comparisons do not imply that either model is a correct one. A statistical indication of a good fit says only that, given the model, the experimenter should not be surprised at having observed the data values that were observed. It says nothing about whether the model is plausible in terms of compatibility with the laws of physics and chemistry. Nor does it rule out the existence of other models that describe the data as well as or better than any of the models tested.

8.4.4. Influence of individual data points

When the method of least squares, or any variant of it, is used to refine a crystal structure, it is implicitly assumed that a model with adjustable parameters makes an unbiased prediction of the experimental observations for some (*a priori* unknown) set of values of those parameters. The existence of any reflection whose observed intensity is inconsistent with this assumption, that is that it differs from the predicted value by an amount that cannot be reconciled with the precision of the measurement, must cause the model to be rejected, or at least modified. In making precise estimates of the values of the unknown parameters, however, different reflections do not all carry the same amount of information (Shoemaker, 1968; Prince & Nicholson, 1985). For an obvious example, consider a space-group systematic absence. Except for possible effects of multiple diffraction or twinning, any observed intensity at a position corresponding to a systematic absence is proof that the screw axis or glide plane is not present. If no intensity is observed for any such reflection, however, any parameter values that conform to the space group are equally acceptable. It is to be expected, on the other hand, that some intensities will be extremely sensitive to small changes in some parameter, and that careful measurement of those intensities will lead to correspondingly precise estimates of the parameter values. For the purpose of precise structure refinement, it is useful to be able to identify the influential reflections.

Consider a vector of observations, $\mathbf{y}$, and a model $\boldsymbol{M}(\mathbf{x})$. The elements of $\mathbf{y}$ define an n-dimension space, and the model values, $M_i(\mathbf{x})$, define a p-dimensional subspace within it. The least-squares solution [equation (8.1.2.7)],

$$\widehat{\mathbf{x}} = (\boldsymbol{A}^T\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{W}(\mathbf{y} - \mathbf{y}_0), \qquad (8.4.4.1)$$

is such that $\widehat{\mathbf{y}} = \boldsymbol{M}(\widehat{\mathbf{x}})$ is the closest point to $\mathbf{y}$ that corresponds to some possible value of $\mathbf{x}$. In (8.4.4.1), $\boldsymbol{W} = \boldsymbol{V}^{-1}$ is the inverse of the variance–covariance matrix for the joint p.d.f. of the elements of $\mathbf{y}$, and $\mathbf{y}_0 = \boldsymbol{M}(\mathbf{x}_0)$ is a point in the p-dimensional subspace close enough to $\boldsymbol{M}(\widehat{\mathbf{x}})$ so that the linear approximation

$$\boldsymbol{M}(\mathbf{x}) = \mathbf{y}_0 + \boldsymbol{A}(\mathbf{x} - \mathbf{x}_0) \qquad (8.4.4.2)$$

[where $A_{ij} = \partial M_i(\mathbf{x})/\partial x_j$] is a good one. Let $\boldsymbol{R}$ be the Cholesky factor of $\boldsymbol{W}$, so that $\boldsymbol{W} = \boldsymbol{R}^T\boldsymbol{R}$, and let $\mathbf{Z} = \boldsymbol{R}\boldsymbol{A}$, $\mathbf{y}' = \mathbf{y} - \mathbf{y}_0$, and $\widehat{\mathbf{y}}' = \widehat{\mathbf{y}} - \mathbf{y}_0$. The least-squares estimate may then be written

$$\widehat{\mathbf{x}} = \mathbf{x}_0 + (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{y}' \qquad (8.4.4.3)$$

and

$$\widehat{\mathbf{y}}' = \mathbf{Z}(\widehat{\mathbf{x}} - \mathbf{x}_0) = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{y}'. \qquad (8.4.4.4)$$

Thus, the matrix $\boldsymbol{P} = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})\mathbf{Z}^T$, the *projection matrix*, is a linear relation between the observed data values and the corresponding calculated values. (Because $\widehat{\mathbf{y}}' = \boldsymbol{P}\mathbf{y}'$, the matrix $\boldsymbol{P}$ is frequently referred to in the statistical literature as the *hat matrix*.) $\boldsymbol{P}^2 = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T = \boldsymbol{P}$, so that $\boldsymbol{P}$ is *idempotent*. $\boldsymbol{P}$ is an $n \times n$ positive semidefinite matrix with rank p, and its eigenvalues are either 1 (p times) or 0 ($n - p$ times). Its diagonal elements lie in the range $0 \le P_{ii} \le 1$, and the trace of $\boldsymbol{P}$ is p, so that the average value of P_{ii} is p/n. Furthermore,

$$P_{ii} = \sum_{j=1}^{n} P_{ij}^2. \qquad (8.4.4.5)$$

A diagonal element of $\boldsymbol{P}$ is a measure of the influence that an observation has on its own calculated value. If P_{ii} is close to one, the model is forced to fit the ith data point, which puts a constraint on the value of the corresponding function of the parameters. A very small value of P_{ii}, because of (8.4.4.5), implies that all elements of the row must be small, and that observation has little influence on its own or any other calculated value. Because it is a measure of influence on the fit, P_{ii} is sometimes referred to as the *leverage* of the ith observation. Note that, because $(\boldsymbol{Z}^T\boldsymbol{Z})^{-1} = \boldsymbol{V}_{\mathbf{x}}$, the variance–covariance matrix for the elements of $\widehat{\mathbf{x}}$, $\boldsymbol{P}$ is the variance–covariance matrix for $\widehat{\mathbf{y}}$, whose elements are functions of the elements of $\widehat{\mathbf{x}}$. A large value of P_{ii} means that y_i is poorly defined by the elements of $\widehat{\mathbf{x}}$, which implies in turn that some elements of $\widehat{\mathbf{x}}$ must be precisely defined by a precise measurement of y_i'.

It is apparent that, in a real experiment, there will be appreciable variation among observations in their leverage. It can be shown (Fedorov, 1972; Prince & Nicholson, 1985) that the observations with the greatest leverage also have the largest effect on the volume of the p-dimensional confidence region for the parameter estimates. Because this volume is a rather gross measure, however, it is useful to have a measure of the influence of individual observations on individual parameters. Let $\boldsymbol{V}_n$ be the variance–covariance matrix for a refinement including n observations, and let $\mathbf{z}$ be a row vector whose elements are $z_j = [\partial M(\mathbf{x})/\partial x_j]/\sigma$ for an additional observation. $\boldsymbol{V}_{n+1}$, the variance–covariance matrix with the additional observation included, is, by definition,

$$\boldsymbol{V}_{n+1} = (\mathbf{Z}^T\boldsymbol{Z} + \mathbf{z}^T\mathbf{z})^{-1}, \qquad (8.4.4.6)$$

which, in the linear approximation, can be shown to be

$$\boldsymbol{V}_{n+1} = \boldsymbol{V}_n - \boldsymbol{V}_n\mathbf{z}^T\mathbf{z}\boldsymbol{V}_n/(1 + \mathbf{z}\boldsymbol{V}_n\mathbf{z}^T). \qquad (8.4.4.7)$$

The diagonal elements of the rank one matrix $\boldsymbol{D} = \boldsymbol{V}_n\mathbf{z}^T\mathbf{z}\boldsymbol{V}_n/(1 + \mathbf{z}\boldsymbol{V}_n\mathbf{z}^T)$ are therefore the amounts that the variances of the estimates of individual parameters will be reduced by inclusion of the additional observation.

This result depends on the elements of $\boldsymbol{Z}$ and $\mathbf{z}$ not changing significantly in the (presumably small) shift from $\widehat{\mathbf{x}}_n$ to $\widehat{\mathbf{x}}_{n+1}$. That this condition is satisfied may be verified by the following procedure. Find an approximation to $\widehat{\mathbf{x}}_{n+1}$ by a line search along the line $\mathbf{x} = \widehat{\mathbf{x}}_n + \alpha \boldsymbol{V}_{n+1}\mathbf{z}^T y'_{n+1}$, and then evaluate $\boldsymbol{B}$, a quasi-Newton update such as the BFGS update (Subsection 8.1.4.3) at that point. If $\alpha = 1$, and the gradient of the sum of squares vanishes, then the linear approximation is exact, and $\boldsymbol{B}$ is null. If

$$|B_{ij}| \ll \left[(\boldsymbol{Z}^T\boldsymbol{Z} + \mathbf{z}^T\mathbf{z})_{ii} (\boldsymbol{Z}^T\boldsymbol{Z} + \mathbf{z}^T\mathbf{z})_{jj} \right]^{1/2} \qquad (8.4.4.8)$$

for all i and j, then (8.4.4.7) can be expected to be an excellent approximation for a nonlinear model.

references