International Tables for Crystallography (2006). Vol. C, Chapter 8.4, pp. 702–706.

8.4. Statistical significance tests

By E. Prince and C. H. Spiegelman

In Chapter 8.1, we discussed the method of least squares and procedures for estimating the values of the adjustable parameters of a model that predicts the mean of a population from which experimental observations are drawn at random. Any model, however, will have some set of parameter values that gives the best least-squares fit. We must now address the question of whether that best fit is adequate, that is, whether it is plausible, given the precision of the data, to accept the hypothesis that the model really is a correct representation of the phenomena that have been measured in the collection of the data. In this chapter, we discuss the probability density function for the sum of squared residuals if the individual residuals are drawn from a normal distribution, the χ^2 distribution, and the conditions under which this p.d.f. may be assumed to approximate a practical case. Next, we discuss the F distribution, which is the distribution of the ratio of two independent, random variables, each of which has a χ^2 distribution, and its use in comparing the fits of constrained and unconstrained versions of a model. We also discuss a test that is useful for a more general comparison of models. Finally, we discuss the variation among data points of their effectiveness in improving the precision of parameter estimates and the application of this analysis to the optimum design of experiments.

8.4.1. The χ^2 distribution

We have seen [equation (8.1.2.1)] that the least-squares estimate is derived by finding the minimum value of a sum of terms of the form

$$R_i = w_i[y_i - M_i(\mathbf{x})]^2, \tag{8.4.1.1}$$

and, further, that the precision of the estimate is optimized if the weight, w_i, is the reciprocal of the variance of the population from which the observation is drawn, $w_i = 1/\sigma_i^2$. Using this relation, (8.4.1.1) can be written

$$R_i = \left\{\left[y_i - M_i(\mathbf{x})\right]/\sigma_i\right\}^2. \tag{8.4.1.2}$$

Each term is the square of a difference between observed and calculated values, expressed as a fraction of the standard uncertainty of the observed value. But, by definition,

$$\sigma_i^2 = \left\langle [y_i - M_i(\mathbf{x})]^2 \right\rangle, \tag{8.4.1.3}$$

where $\mathbf{x}$ has its unknown 'correct' value, so that $\langle R\rangle = 1$, and the expected value of the sum of n such terms is n. It can be shown (Draper & Smith, 1981) that each parameter estimated reduces this expected sum by one, so that, for p estimated parameters,

$$\langle S\rangle = \left\langle \sum_{i=1}^{n} \left\{\left[y_i - M_i(\widehat{\mathbf{x}})\right]/\sigma_i\right\}^2 \right\rangle = n - p, \tag{8.4.1.4}$$

where $\widehat{\mathbf{x}}$ is the least-squares estimate. The *standard uncertainty of an observation of unit weight*, also referred to as the *goodness-of-fit parameter*, is defined by

$$G = \left[\frac{S}{n-p}\right]^{1/2} = \left[\frac{\left\{\sum_{i=1}^{n} w_i\left[y_i - M_i(\widehat{\mathbf{x}})\right]^2\right\}}{n-p}\right]^{1/2}. \tag{8.4.1.5}$$

From (8.4.1.4), it follows that $\langle G\rangle = 1$ for a correct model with weights assigned in accordance with (8.4.1.2).

A value of G that is close to one, if the weights have been assigned by $w_i = 1/\sigma_i^2$, is an indicator that the model is consistent with the data. It should be noted that it is not necessarily an indicator that the model is 'correct', because it does not rule out the existence of an alternative model that fits the data as well or better. An assessment of the adequacy of the fit of a given model depends, however, on what is meant by 'close to one', which depends in turn on the spread of a probability density function for G. We saw in Chapter 8.1 that least squares with this weighting scheme would give the best, linear, unbiased estimate of the model parameters, with no restrictions on the p.d.f.s of the populations from which the observations are drawn except for the implicit assumption that the variances of these p.d.f.s are finite. To construct a p.d.f. for G, however, it is necessary to make an assumption about the shapes of the p.d.f.s for the observations. The usual assumption is that these p.d.f.s can be described by the *normal p.d.f.*,

$$\Phi_N(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[\frac{(x-\mu)^2}{2\sigma^2}\right]. \tag{8.4.1.6}$$

The justification for this assumption comes from the *central-limit theorem*, which states that, under rather broad conditions, the p.d.f. of the arithmetic mean of n observations drawn from a population with mean μ and variance σ^2 tends, for large n, to a normal distribution with mean μ and variance σ^2/n. [For a discussion of the central limit theorem, see Cramér (1951).]

If we make the assumption of a normal distribution of errors and make the substitution $z = (x-\mu)/\sigma$, (8.4.1.6) becomes

$$\Phi_N(z, 0, 1) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right). \tag{8.4.1.7}$$

The probability that z^2 will be less than χ^2 is equal to the probability that z will lie in the interval $-\chi \le z \le \chi$, or

$$\Psi(\chi^2) = \int_0^{\chi^2} \Phi(z^2)\,\mathrm{d}z^2 = \int_{-\chi}^{+\chi} \Phi(z)\,\mathrm{d}z. \tag{8.4.1.8}$$

Letting $t = z^2/2$ and substituting in (8.4.1.7), this becomes

$$\Psi(\chi^2) = \frac{1}{\sqrt{\pi}}\int_0^{\chi^2/2} t^{-1/2}\exp(-t)\,\mathrm{d}t. \tag{8.4.1.9}$$

$\Phi(\chi^2) = \mathrm{d}\Psi(\chi^2)/\mathrm{d}\chi^2$, so that

$$\begin{array}{ll} \Phi\left(\chi^2\right) = \left(2\pi\chi^2\right)^{-1/2}\exp\left(-\chi^2/2\right), & \chi^2 > 0, \\ \Phi\left(\chi^2\right) = 0, & \chi^2 \le 0. \end{array} \tag{8.4.1.10}$$

The joint p.d.f. of the squares of two random variables, z_1 and z_2, drawn independently from the same population with a normal p.d.f. is

$$\Phi_J\left(z_1^2, z_2^2\right) = \frac{1}{2\pi z_1 z_2}\exp\left[-\frac{z_1^2 + z_2^2}{2}\right], \tag{8.4.1.11}$$

and the p.d.f. of the sum, s^2, of these two terms is the integral over the joint p.d.f. of all pairs of z_1^2 and z_2^2 that add up to s^2.

$$\Phi(s^2) = \frac{1}{2\pi}\exp\left(-\frac{s^2}{2}\right)\left[z_1^2(s^2 - z_1^2)\right]\mathrm{d}z_1^2. \tag{8.4.1.12}$$

This integral can be evaluated by use of the gamma and beta functions. The *gamma function* is defined for positive real x by

Table 8.4.1.1. *Values of χ^2/ν for which the c.d.f. $\Psi(\chi^2,\nu)$ has the values given in the column headings, for various values of ν*

ν	0.5	0.9	0.95	0.99	0.995
1	0.4549	2.7055	3.8415	6.6349	7.8795
2	0.6931	2.3026	2.9957	4.6052	5.2983
3	0.7887	2.0838	2.6049	3.7816	4.2794
4	0.8392	1.9449	2.3719	3.3192	3.7151
6	0.8914	1.7741	2.0986	2.8020	3.0913
8	0.9180	1.6702	1.9384	2.5113	2.7444
10	0.9342	1.5987	1.8307	2.3209	2.5188
15	0.9559	1.4871	1.6664	2.0385	2.1868
20	0.9669	1.4206	1.5705	1.8783	1.9999
25	0.9735	1.3753	1.5061	1.7726	1.8771
30	0.9779	1.3419	1.4591	1.6964	1.7891
40	0.9834	1.2951	1.3940	1.5923	1.6692
50	0.9867	1.2633	1.3501	1.5231	1.5898
60	0.9889	1.2400	1.3180	1.4730	1.5325
80	0.9917	1.2072	1.2735	1.4041	1.4540
100	0.9933	1.1850	1.2434	1.3581	1.4017
120	0.9945	1.1686	1.2214	1.3246	1.3638
140	0.9952	1.1559	1.2044	1.2989	1.3346
160	0.9958	1.1457	1.1907	1.2783	1.3114
200	0.9967	1.1301	1.1700	1.2472	1.2763

$$\Gamma(x) = \int_0^\infty t^{x-1}\exp(-t)\,\mathrm{d}t. \qquad (8.4.1.13)$$

Although this function is continuous for all $x > 0$, its value is of interest in the context of this analysis only for x equal to positive, integral multiples of 1/2. It can be shown that $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$, and $\Gamma(x+1) = x\Gamma(x)$. It follows that, for a positive integer, n, $\Gamma(n) = (n-1)!$, and that $\Gamma(3/2) = \sqrt{\pi}/2$, $\Gamma(5/2) = 3\sqrt{\pi}/4$, *etc*. The *beta function* is defined by

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t. \qquad (8.4.1.14)$$

It can be shown (Prince, 1994) that $B(x,y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$. Making the substitution $t = z_1^2/s^2$, (8.4.1.12) becomes

$$\begin{aligned}\Phi(s^2) &= \frac{1}{2\pi}\exp\left(-\frac{s^2}{2}\right)\int_0^1 [t(1-t)]^{-1/2}\,\mathrm{d}t \\ &= \frac{1}{2\pi}\exp\left(-\frac{s^2}{2}\right)B(1/2,1/2) \\ &= \tfrac{1}{2}\exp\left(-\frac{s^2}{2}\right), \qquad s^2 \geq 0. \end{aligned} \qquad (8.4.1.15)$$

By a similar procedure, it can be shown that, if χ^2 is the sum of ν terms, $z_1^2, z_2^2, \ldots, z_\nu^2$, where all are drawn independently from a population with the p.d.f. given in (8.4.1.10), χ^2 has the p.d.f.

$$\begin{aligned}\Phi(\chi^2,\nu) &= \frac{(\chi^2)^{\nu/2-1}}{2^{\nu/2}\Gamma(\nu/2)}\exp\left(-\frac{\chi^2}{2}\right), \qquad \chi^2 > 0, \\ \Phi(\chi^2,\nu) &= 0, \qquad \chi^2 \leq 0. \end{aligned} \qquad (8.4.1.16)$$

The parameter ν is known as the number of *degrees of freedom*, but this use of that term must not be confused with the conventional use in physics and chemistry. The p.d.f. in (8.4.1.16) is the *chi-squared distribution with ν degrees of freedom*. Table 8.4.1.1 gives the values of χ^2/ν for which the *cumulative distribution function* (*c.d.f.*) $\Psi(\chi^2,\nu)$ has various values for various choices of ν. This table is provided to enable verification of computer codes that may be used to generate more extensive tables. It was generated using a program included in the statistical library DATAPAC (Filliben, unpublished). Fortran code for this program appears in Prince (1994).

The quantity $(n-p)G$ is the sum of n terms that have mean value $(n-p)/n$. Because the process of determining the least-squares fit establishes p relations among them, however, only $(n-p)$ of the terms are independent. The number of degrees of freedom is therefore $\nu = (n-p)$, and, if the model is correct, and the terms have been properly weighted, $\chi^2 = (n-p)G^2$ has the chi-squared distribution with $(n-p)$ degrees of freedom. In crystallography, the number of degrees of freedom tends to be large, and the p.d.f. for G correspondingly sharp, so that even rather small deviations from $G^2 = 1$ should cause one or both of the hypotheses of a correct model and appropriate weights to be rejected. It is common practice to assume that the model is correct, and that the weights have correct *relative* values, that is that they have been assigned by $w_i = k/\sigma_i^2$, where k is a number different from, usually greater than, one. G is then taken to be an estimate of k, and all elements of $(\boldsymbol{A}^T\boldsymbol{W}\boldsymbol{A})^{-1}$ (Section 8.1.2) are multiplied by G^2 to get an estimated variance–covariance matrix. The range of validity of this procedure is limited at best. It is discussed further in Chapter 8.5.

8.4.2. The *F* distribution

Consider an unconstrained model with p parameters and a constrained one with q parameters, where $q < p$. We wish to decide whether the constrained model represents an adequate fit to the data, or if the additional parameters in the unconstrained model provide, in some important sense, a better fit to the data. Provided the $(p-q)$ additional columns of the design matrix, $\boldsymbol{A}$, are linearly independent of the previous q columns, the sum of squared residuals must be reduced by some finite amount by adjusting the additional parameters, but we must decide whether this improved fit would have occurred purely by chance, or whether it represents additional information.

Let s_c^2 and s_u^2 be the weighted sums of squared residuals for the constrained and unconstrained models, respectively. If the constrained and unconstrained models are equally good representations of the data, and the weights have been assigned by $w_i = 1/\sigma_i^2$, the expected values of the sums of squares are $\langle s_c^2\rangle = (n-q)$ and $\langle s_u^2\rangle = (n-p)$, and, further, they should be distributed as χ^2 with $(n-q)$ and $(n-p)$ degrees of freedom, respectively. Also, $\langle s_c^2 - s_u^2\rangle = (p-q)$, and $(s_c^2 - s_u^2)$ is distributed as χ^2 with $(p-q)$ degrees of freedom. s_c^2 and s_u^2 are not independent, but $(s_c^2 - s_u^2)$ is the squared magnitude of a vector in a $(p-q)$-dimensional subspace that is orthogonal to the $(n-p)$-dimensional space of s_u^2. Therefore, s_u^2 and $(s_c^2 - s_u^2)$ are independent, random variables, each with a χ^2 distribution. Let $\chi_1^2 = (s_c^2 - s_u^2)$, $\chi_2^2 = s_u^2$, $\nu_1 = p-q$, and $\nu_2 = n-p$. The ratio $F = (\chi_1^2/\nu_1)/(\chi_2^2/\nu_2)$ should have a value close to one, even if the weights have relative rather than absolute values, but we need a measure of how far away from one this ratio can be before we must reject the hypothesis that the two models are equally good representations of the data. The conditional p.d.f. for F, given a value of χ_2^2 , is

$$\Phi_C(F|\chi_2^2) = \frac{\left[(\nu_1/\nu_2)\chi_2^2\right]^{\nu_1/2}F^{\nu_1/2-1}}{2^{\nu_1/2}\Gamma(\nu_1/2)}\exp\left[-(\nu_1/\nu_2)\chi_2^2F/2\right], \qquad (8.4.2.1)$$

and the marginal p.d.f. for χ_2^2 is

$$\Phi_M(\chi_2^2) = \frac{(\chi_2^2)^{\nu_2/2-1}}{2^{\nu_2/2}\Gamma(\nu_2/2)}\exp(-\chi_2^2/2). \qquad (8.4.2.2)$$

Table 8.4.2.1. *Values of the F ratio for which the c.d.f. $\Psi(F,\nu_1,\nu_2)$ has the value 0.95, for various choices of ν_1 and ν_2*

ν_2 \ ν_1	1	2	4	8	15
10	4.9646	4.1028	3.4781	3.0717	2.8450
20	4.3512	3.4928	2.8661	2.4471	2.2033
30	4.1709	3.3158	2.6896	2.2662	2.0148
40	4.0847	3.2317	2.6060	2.1802	1.9245
50	4.0343	3.1826	2.5572	2.1299	1.8714
60	4.0012	3.1504	2.5252	2.0970	1.8364
80	3.9604	3.1108	2.4859	2.0564	1.7932
100	3.9361	3.0873	2.4626	2.0323	1.7675
120	3.9201	3.0718	2.4472	2.0164	1.7505
150	3.9042	3.0564	2.4320	2.0006	1.7335
200	3.8884	3.0411	2.4168	1.9849	1.7167
300	3.8726	3.0259	2.4017	1.9693	1.6998
400	3.8648	3.0183	2.3943	1.9616	1.6914
600	3.8570	3.0107	2.3868	1.9538	1.6831
1000	3.8508	3.0047	2.3808	1.9477	1.6764

The marginal p.d.f. for F is obtained by integration of the joint p.d.f.,

$$\Phi(F) = \int_0^\infty \Phi_C\left(F|\chi_2^2\right)\Phi_M\left(\chi_2^2\right)\,\mathrm{d}\chi_2^2, \tag{8.4.2.3}$$

yielding the result

$$\Phi(F,\nu_1,\nu_2) = \frac{(\nu_1/\nu_2)F^{\nu_1/2-1}}{B(\nu_1/2,\nu_2/2)\left[1+(\nu_1/\nu_2)F\right]^{(\nu_1+\nu_2)/2}}. \tag{8.4.2.4}$$

This p.d.f. is known as the *F distribution with ν_1 and ν_2 degrees of freedom*. Table 8.4.2.1 gives the values of F for which the c.d.f. $\Psi(F,\nu_1,\nu_2)$ is equal to 0.95 for various choices of ν_1 and ν_2. Fortran code for the program from which the table was generated appears in Prince (1994).

The cumulative distribution function $\Psi(F,\nu_1,\nu_2)$ gives the probability that the F ratio will be less than some value by chance if the models are equally consistent with the data. It is therefore a necessary, but not sufficient, condition for concluding that the unconstrained model gives a significantly better fit to the data that $\Psi(F,\nu_1,\nu_2)$ be greater than $1-\alpha$, where α is the desired level of significance. For example, if $\Psi(F,\nu_1,\nu_2)=0.95$, the probability is only 0.05 that a value of F this large or greater would have been observed if the two models were equally good representations of the data.

Hamilton (1964) observed that the F ratio could be expressed in terms of the crystallographic weighted R index, which is defined, for refinement on |F| (and similarly for refinement on $|F|^2$), by

$$R_w = \left[\sum w_i(|F_o|_i - |F_c|_i)^2 \Big/ \sum w_i|F_o|_i^2\right]^{1/2}. \tag{8.4.2.5}$$

Denoting by R_c and R_u the weighted R indices for the constrained and unconstrained models, respectively,

$$F = (\nu_2/\nu_1)[(R_c/R_u)^2 - 1], \tag{8.4.2.6}$$

and a c.d.f. for R_c/R_u can be readily derived from this relation. A significance test based on R_c/R_u is known as *Hamilton's R-ratio test*; it is entirely equivalent to a test on the F ratio.

8.4.3. Comparison of different models

Tests based on F or the R ratio have several limitations. One important one is that they are applicable only when the parameters of one model form a subset of the parameters of the other. Also, the F test makes no distinction between improvement in fit as a result of small improvements throughout the entire data set and a large improvement in a small number of critically sensitive data points. A test that can be used for comparing arbitrary pairs of models, and that focuses attention on those data points that are most sensitive to differences in the models, was introduced by Williams & Kloot (1953; also Himmelblau, 1970; Prince, 1982).

Table 8.4.3.1. *Values of t for which the c.d.f. $\Psi(t,\nu)$ has the values given in the column headings, for various values of ν*

ν	0.75	0.90	0.95	0.99	0.995
1	1.0000	3.0777	6.3138	31.8206	63.6570
2	0.8165	1.8856	2.9200	6.9646	9.9249
3	0.7649	1.6377	2.3534	4.5407	5.8409
4	0.7407	1.5332	2.1319	3.7469	4.6041
6	0.7176	1.4398	1.9432	3.1427	3.7074
8	0.7064	1.3968	1.8596	2.8965	3.3554
10	0.6998	1.3722	1.8125	2.7638	3.1693
12	0.6955	1.3562	1.7823	2.6810	3.0546
14	0.6924	1.3450	1.7613	2.6245	2.9769
16	0.6901	1.3368	1.7459	2.5835	2.9208
20	0.6870	1.3253	1.7247	2.5280	2.8453
25	0.6844	1.3164	1.7081	2.4851	2.7874
30	0.6828	1.3104	1.6973	2.4573	2.7500
35	0.6816	1.3062	1.6896	2.4377	2.7238
40	0.6807	1.3031	1.6839	2.4233	2.7045
50	0.6794	1.2987	1.6759	2.4033	2.6778
60	0.6786	1.2958	1.6707	2.3901	2.6603
80	0.6776	1.2922	1.6641	2.3739	2.6387
100	0.6770	1.2901	1.6602	2.3642	2.6259
120	0.6765	1.2886	1.6577	2.3578	2.6174

Consider a set of observations, y_{0i}, and two models that predict values for these observations, y_{1i} and y_{2i}, respectively. We determine the slope of the regression line $z=\lambda x$, where $z_i = \left[y_{0i} - (1/2)(y_{1i}+y_{2i})\right]/\sigma_i$, and $x_i = (y_{1i}-y_{2i})/\sigma_i$. Suppose model 1 is a perfect fit to the data, which have been measured with great precision, so that $y_{0i}=y_{1i}$ for all i. Under these conditions, $\lambda=+1/2$. Similarly, if model 2 is a perfect fit, $\lambda=-1/2$. Real experimental data, of course, are subject to random error, and $|\lambda|$ in general would be expected to be less than 1/2. A least-squares estimate of λ is

$$\widehat{\lambda} = \frac{\sum_{i=1}^n z_i x_i}{\sum_{i=1}^n x_i^2}, \tag{8.4.3.1}$$

and it has an estimated variance

$$\widehat{\sigma}_\lambda^2 = \frac{\sum_{i=1}^n z_i^2 - \widehat{\lambda}^2\sum_{i=1}^n x_i^2}{(n-1)\sum_{i=1}^n x_i^2}. \tag{8.4.3.2}$$

The hypothesis that the two models give equally good fits to the data can be tested by considering $\widehat{\lambda}$ to be an unconstrained, one-parameter fit that is to be compared with a constrained, zero-parameter fit for which $\lambda=0$. A p.d.f. for making this comparison can be derived from an F distribution with $\nu_1=1$ and $\nu_2=\nu=(n-1)$.

$$\Phi(F,1,\nu) = \frac{\Gamma[(\nu+1)/2]}{\sqrt{\pi\nu F}\,\Gamma(\nu/2)(1+F/\nu)^{(\nu+1)/2}}. \tag{8.4.3.3}$$

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

$$\int_0^{F_0} \Phi(F, 1, \nu)\, \mathrm{d}F = \int_{-t_0}^{+t_0} \Phi(t, \nu)\, \mathrm{d}t, \tag{8.4.3.4}$$

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

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

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)],

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

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

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

[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 $\boldsymbol{Z} = \boldsymbol{R}\boldsymbol{A}$, $\mathbf{y}' = \mathbf{y} - \mathbf{y}_0$, and $\widehat{\mathbf{y}}' = \widehat{\mathbf{y}} - \mathbf{y}_0$. The least-squares estimate may then be written

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

and

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

Thus, the matrix $\boldsymbol{P} = \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})\boldsymbol{Z}^T$, 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*.) $\boldsymbol{P}^2 = \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T = \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T = \boldsymbol{P}$, 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,

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

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,

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

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

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

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 $\mathbf{x} = \hat{\mathbf{x}}_n + \alpha \boldsymbol{V}_{n+1}\mathbf{z}^T y'_{n+1}$, 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

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

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

8.3

Cruickshank, D. W. J. (1961). *Coordinate errors due to rotational oscillations of molecules*. *Acta Cryst*. **14**, 896–897.

Finger, L. W. (1969). *The crystal structure and cation distribution of a grunerite*. *Mineral. Soc. Am. Spec. Pap*. **2**, 95–100.

Gill, P. E., Murray, W. & Wright, M. M. (1981). *Practical optimization*. New York: Academic Press.

Hamilton, W. C. (1964). *Statistics in physical science: estimation, hypothesis testing and least squares*. New York: Ronald Press.

Hendrickson, W. A. (1985). *Stereochemically restrained refinement of macromolecular structures*. *Methods in enzymology*, Vol. 115. *Diffraction methods for biological macromolecules, Part B*, edited by H. W. Wyckoff, C. H. W. Hirs & S. N. Timasheff, pp. 252–270. New York: Academic Press.

Hendrickson, W. A. & Konnert, J. H. (1980). *Incorporation of stereochemical information into crystallographic refinement*. *Computing in crystallography*, edited by R. Diamond, S. Ramaseshan & D. Venkatesan, pp. 13.01–13.26. Bangalore: Indian Academy of Sciences.

Hestenes, M. & Stiefel, E. (1952). *Methods of conjugate gradients for solving linear systems*. *J. Res. Natl Bur. Stand*. **49**, 409–436.

Jack, A. & Levitt, M. (1978). *Refinement of large structures by simultaneous minimization of energy and R factor*. *Acta Cryst*. A**34**, 931–935.

Johnson, C. K. (1970). *Generalized treatments for thermal motion*. *Thermal neutron diffraction*, edited by B. T. M. Willis, pp. 132–160. Oxford University Press.

Konnert, J. H. (1976). *A restrained-parameter structure-factor least-squares refinement procedure for large asymmetric units*. *Acta Cryst*. A**32**, 614–617.

Konnert, J. H. & Hendrickson, W. A. (1980). *A restrained-parameter thermal-factor refinement procedure*. *Acta Cryst*. A**36**, 344–350.

Levy, H. A. (1956). *Symmetry relations among coefficients of anisotropic temperature factors*. *Acta Cryst*. **9**, 679.

Prince, E. (1994). *Mathematical techniques in crystallography and materials science*, 2nd ed. Berlin/Heidelberg/New York/London/Paris/Tokyo/Hong Kong/Barcelona/Budapest: Spring-Verlag.

Prince, E., Dickens, B. & Rush, J. J. (1974). *A study of one-dimensional hindered rotation in NH_3OHClO_4*. *Acta Cryst*. B**30**, 1167–1172.

Prince, E. & Finger, L. W. (1973). *Use of constraints on thermal motion in structure refinement of molecules with librating side groups*. *Acta Cryst*. B**29**, 179–183.

Rae, A. D. (1978). *An optimized conjugate gradient solution for least-squares equations*. *Acta Cryst*. A**34**, 578–582.

Ralph, R. L. & Finger, L. W. (1982). *A computer program for refinement of crystal orientation matrix and lattice constants from diffractometer data with lattice symmetry constraints*. *J. Appl. Cryst*. **15**, 537–539.

Rietveld, H. M. (1969). *A profile refinement method for nuclear and magnetic structures*. *J. Appl. Cryst*. **2**, 65–71.

Schomaker, V. & Trueblood, K. N. (1968). *On the rigid-body motion of molecules in crystals*. *Acta Cryst*. B**24**, 63–76.

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.

Sygusch, J. (1976). *Constrained thermal motion refinement for a rigid molecule with librating side groups*. *Acta Cryst*. B**32**, 3295–3298.

Waser, J. (1963). *Least-squares refinement with subsidiary conditions*. *Acta Cryst*. **16**, 1091–1094.

8.4

Cramér, H. (1951). *Mathematical methods of statistics*. Princeton, NJ: Princeton University Press.

Draper, N. & Smith, H. (1981). *Applied regression analysis*. New York: John Wiley.

Fedorov, V. V. (1972). *Theory of optimal experiments*, translated by W. J. Studden & E. M. Klimko. New York: Academic Press.

Hamilton, W. C. (1964). *Statistics in physical science: estimation, hypothesis testing and least squares*. New York: Ronald Press.

Himmelblau, D. M. (1970). *Process analysis by statistical methods*. New York: John Wiley.

Prince, E. (1982). *Comparison of the fits of two models to the same data set*. *Acta Cryst*. B**38**, 1099–1100.

Prince, E. (1994). *Mathematical techniques in crystallography and materials science,* 2nd ed. Berlin/Heidelberg/New York/London/Paris/Tokyo/Hong Kong/Barcelona/Budapest: Springer-Verlag.

Prince, E. & Nicholson, W. L. (1985). *Influence of individual reflections on the precision of parameter estimates in least squares refinement*. *Structure and statistics in crystallography*, edited by A. J. C. Wilson, pp. 183–195. Guilderland, NY: Adenine Press.

Shoemaker, D. P. (1968). *Optimization of counting time in computer controlled X-ray and neutron single-crystal diffractometry*. *Acta Cryst*. A**24**, 136–142.

Williams, E. J. & Kloot, N. H. (1953). *Interpolation in a series of correlated observations*. *Aust. J. Appl. Sci*. **4**, 1–17.

8.5

Abrahams, S. C. & Keve, E. T. (1971). *Normal probability plot analysis of error in measured and derived quantities and standard deviations*. *Acta Cryst*. A**27**, 157–165.

Beckman, R. J. & Cook, R. D. (1983). *Outlier…….s*. *Technometrics*, **25**, 119–149.

Belsley, D. A. (1991). *Conditioning diagnostics*. New York: John Wiley & Sons.

Belsley, D. A., Kuh, E. & Welsch, R. E. (1980). *Regression diagnostics*. New York: John Wiley & Sons.

Chatterjee, S. & Hadi, A. S. (1986). *Influential observations, high leverage points, and outliers in linear regression*. *Stat. Sci*. **1**, 379–393.

Fedorov, V. V. (1972). *Theory of optimal experiments*, translated by W. J. Studden & E. M. Klimko. New York: Academic Press.

ISO (1993). *Guide to the expression of uncertainty in measurement*. Geneva: International Organization for Standardization.

Kafadar, K. & Spiegelman, C. H. (1986). *An alternative to ordinary Q–Q plots: conditional Q–Q plots*. *Comput. Stat. Data Anal*. **4**, 167–184.

Prince, E. (1994). *Mathematical techniques in crystallography and materials science*, 2nd ed. Berlin/Heidelberg/New York/London/Paris/Tokyo/Hong Kong/Barcelona/Budapest: Springer-Verlag.