Estimating Errors in Least-Squares Fitting

P. H. Richter

Communications Systems and Research Section

While least-squares fitting procedures are commonly used in data analysis and are extensively discussed in the literature devoted to this subject, the proper assessment of errors resulting from such fits has received relatively little attention. The present work considers statistical errors in the fitted parameters, as well as in the values of the fitted function itself, resulting from random errors in the data. Expressions are derived for the standard error of the fit, as a function of the independent variable, for the general nonlinear and linear fitting problems. Additionally, closed-form expressions are derived for some examples commonly encountered in the scientific and engineering fields, namely, ordinary polynomial and Gaussian fitting functions. These results have direct application to the assessment of antenna gain and system temperature characteristics, in addition to a broad range of problems in data analysis. The effects of the nature of the data and the choice of fitting function on the ability to accurately model the system under study are discussed, and some general rules are deduced to assist workers intent on maximizing the amount of information obtained from a given set of measurements.

I. Summary

The fitting of data of the form $(x_i, y_i), i = 1, 2, \cdots, N$ by a function $y(x; a_1, \cdots, a_M) \equiv y(x; \mathbf{a})$, depending on M coefficients, a_j, and the independent variable x, is common in scientific and engineering work. The procedure most often used for optimizing the coefficients in order to obtain the best fit is the least-squares method, in which the quantity

$$\chi^2(\mathbf{a}) = \sum_{i=1}^{N} \frac{[y_i - y(x_i; \mathbf{a})]^2}{\sigma_i^2}$$

is minimized, where σ_i is the standard deviation of the random errors of y_i, which we assume to be normally distributed.

The result of such a fitting procedure is the function $y(x; \mathbf{a}_0)$, where $\mathbf{a}_0$ is the coefficient vector that minimizes $\chi^2(\mathbf{a})$, and the question arises as to what standard error to associate with the values of this resulting fit. Standard references on statistics and data analysis give the well-known result that the variances of the coefficients, a_j, are given by the diagonal elements of the covariance matrix, $\mathbf{C}$, i.e., $\sigma_{a_j}^2 = C_{jj}$, where $\mathbf{C}$ is the inverse of the matrix $\mathbf{H}$, variously referred to as the curvature or Hessian matrix. While it is often useful to know what the parameter errors are, especially if the parameters

themselves are related to some underlying physical model of the process under study, this does not tell one directly what the error is in the values of the fitting function itself, and a knowledge of this error, which is a function of the independent variable, is frequently of value in characterizing the system performance.

Lacking a general discussion of this in the literature, it seems that various workers assume a mean error equal to either the rms value of the data errors or $1/\sqrt{N}$ times this. It is shown in the present work, however, that for the general least-squares fit, the weighted mean value of the variance of the fit, averaged over the data points $x = x_i$, is given by

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\sigma_y^2(x_i)}{\sigma_i^2} = \frac{M}{N}$$

so that for constant data errors,

$$\overline{\sigma_y^2} = \frac{1}{N}\sum_{i=1}^{N}\sigma_y^2(x_i) = \frac{M}{N}\sigma^2$$

Thus, the mean standard error depends on the order of the fit, increasing as the square root of this value.

The error in the value of the fitted function, however, always depends on x, even when the standard deviations of the data errors, σ_i, are all the same, independent of x. An analysis of these errors leads to the general result that the variance of the value of the fitted function, resulting from the random data errors, is given by

$$\sigma_y^2(x) = \sum_{j=1}^{M}\sum_{k=1}^{M}C_{jk}d_j(x)d_k(x) = \mathbf{d}(x)^T\mathbf{C}\,\mathbf{d}(x)$$

where $[\mathbf{d}(x)]_j \equiv d_j(x) = [\partial y(x;\mathbf{a})/\partial a_j]|_{\mathbf{a}_0}$ and T implies matrix transpose. For the special case of linear fitting, where $y(x;\mathbf{a}) = \sum_{j=1}^{M}a_j X_j(x)$, this becomes

$$\sigma_y^2(x) = \sum_{j=1}^{M}\sum_{k=1}^{M}C_{jk}X_j(x)X_k(x) = \mathbf{x}(x)^T\mathbf{C}\,\mathbf{x}(x)$$

where $\mathbf{x}(x)$ is a column vector whose elements are $X_j(x)$. An example of the application of this result to a set of antenna aperture efficiency versus elevation data is shown in Figs. 1 through 4.

For the important class of basis functions corresponding to ordinary polynomials, $X_j(x) = x^{j-1}$, it is shown that if the data are uniformly distributed along the x-axis and the data standard errors are constant, $\sigma_i = \sigma$, then simple, closed-form expressions can be derived for $\sigma_y^2(x)$. Thus, we find

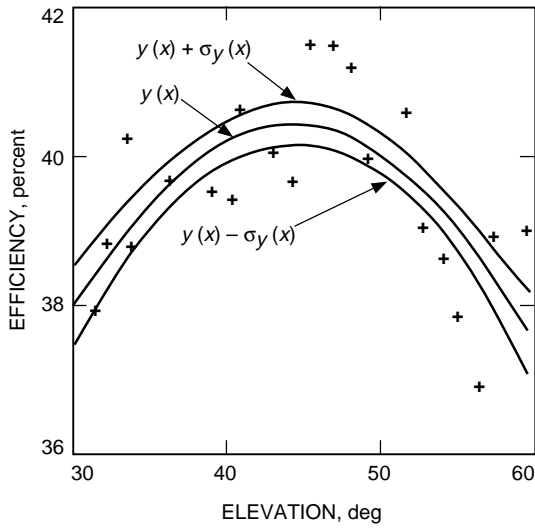$$\eta^2 = \sum_{j=0}^{M-1}A_{2j}^{(M-1)}(N)\xi^{2j}$$

Fig. 1. Quadratic fit to antenna aperture efficiency versus elevation data showing the confidence limits corresponding to 68.3 percent ($\pm\sigma_y(x)$). The data standard errors were constant and equal to 0.91 efficiency percent, and the computed reduced χ^2 was 1.06.
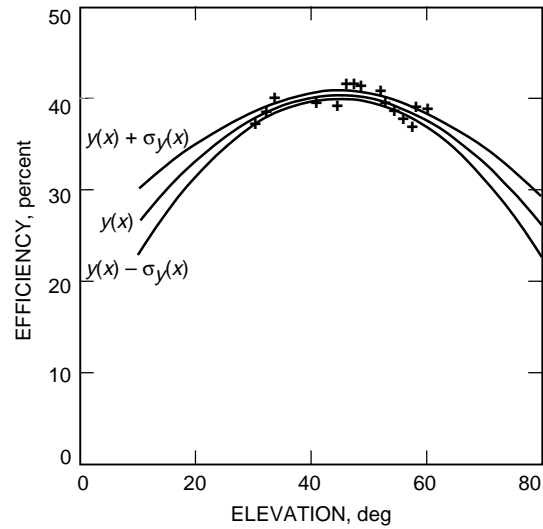

Fig. 2. The same as Fig. 1 except the fit and limits are extended beyond the existing data to illustrate the effect of the rapid increase in the error of the fit.
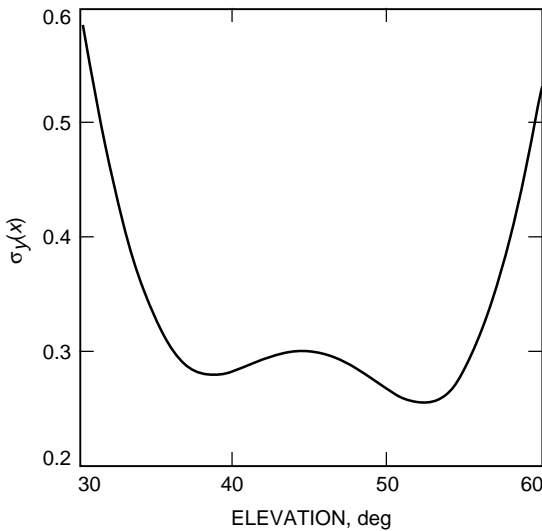

Fig. 3. The standard error of the fit corresponding to Fig. 1 for the range of elevation over which the data exist.
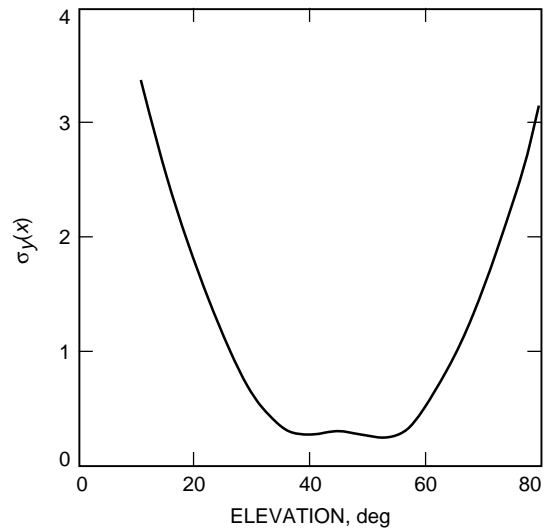

Fig. 4. The same as Fig. 3 except the range of elevation is extended beyond the existing data.

where $\eta = \sqrt{N}([\sigma_y(x)]/\sigma)$, $\xi = [(x-\overline{x})/\sigma_x]$, $\overline{x} = (1/N)\sum_{i=1}^{N} x_i$, $\sigma_x^2 = (1/N)\sum_{i=1}^{N} x_i^2 - (1/N^2)(\sum_{i=1}^{N} x_i)^2$, and the coefficients $A_{2j}^{(M-1)}N$ are listed in Table 1 for the cases $M = 2, 3, 4$, corresponding to straight-line, quadratic, and cubic fits. For the straight-line fit, the coefficients appearing in the above expression are independent of the number of data points, N, while for the quadratic and cubic cases they become independent for reasonable values of N, say, $N > 10$. These results are summarized graphically with the set of universal error curves shown in Fig. 5.

Table 1. The coefficients in the equation for the squared normalized standard error of the fit for straight-line, quadratic, and cubic fits.[a]

$M-1$	A_0	A_2	A_4	A_6
1	1	1	0	0
2	$\dfrac{3(3N^2-7)}{4(N^2-4)}$ $\left[\dfrac{9}{4}\right]$	$-\dfrac{6(N^2+1)}{4(N^2-4)}$ $\left[-\dfrac{3}{2}\right]$	$\dfrac{5(N^2-1)}{4(N^2-4)}$ $\left[\dfrac{5}{4}\right]$	0
3	$\dfrac{3(3N^2-7)}{4(N^2-4)}$ $\left[\dfrac{9}{4}\right]$	$\dfrac{5}{12}\dfrac{9N^4-12N^2-61}{(N^2-4)(N^2-9)}$ $\left[\dfrac{15}{4}\right]$	$-\dfrac{5}{36}\dfrac{33N^4-23N^2-226}{(N^2-4)(N^2-9)}$ $\left[-\dfrac{55}{12}\right]$	$\dfrac{175}{108}\dfrac{(N^2-1)^2}{(N^2-4)(N^2-9)}$ $\left[\dfrac{175}{108}\right]$

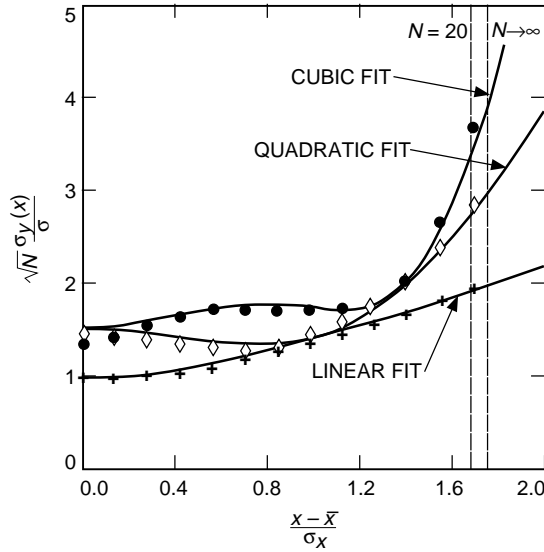[a] The values for $N \to \infty$ are shown in brackets.

Fig. 5. Universal, normalized error curves for straight-line, quadratic, and cubic fits for constant, normally distributed data errors, $\sigma_i = \sigma$, with N uniformly distributed data points. The symbols associated with each curve correspond to the results of Monte Carlo calculations carried out as a check (see Appendix for details).

As an example of a similar development for nonlinear fitting, the case of a Gaussian function given by

$$y(x; \mathbf{a}) = a_1 \exp\left[\frac{-(x-a_2)^2}{2a_3^2}\right]$$

is treated exactly, and it is shown that for uniformly distributed data points located symmetrically relative to the peak, and constant data errors,

$$\eta^2 = \sqrt{\frac{3}{\pi}}\sigma_t e^{-(\xi\sigma_t)^2}\left(3 + 4\sigma_t^4\xi^4\right)$$

where $\sigma_t = (\sigma_x/a_3)$, while if the data errors are proportional to the value of the function, $\sigma(x) \propto y(x; \mathbf{a})$, one finds

$$\eta^2 = N \left[\frac{\sigma_y(x)}{\sigma(x)} \right]^2 = \frac{9}{4} - \frac{3}{2}\xi^2 + \frac{5}{4}\xi^4$$

where in both cases it is assumed that the number of data points, N, is reasonably large, of the order of 20 or more, and in the former case, it is also assumed that the spread of the data points, L, is greater than about $\pm 2a_3$. These results are shown graphically in Figs. 6 and 7.
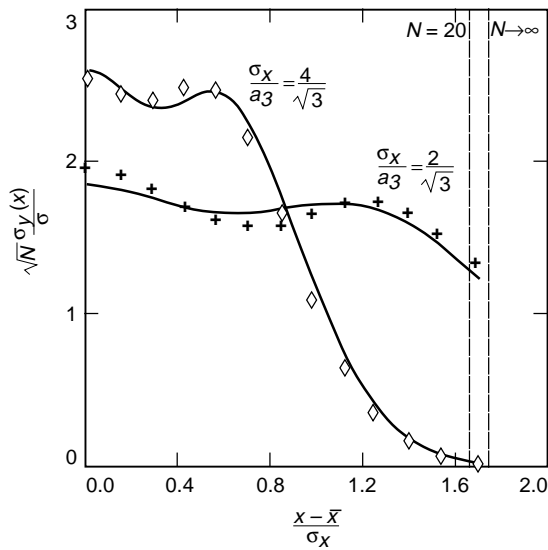

Fig. 6. Universal, normalized error curves for the general Gaussian function $y(x;a) = a_1 \exp[-(x-a_2)^2/2a_3]$ for constant, normally distributed data errors, $\sigma_i = \sigma$, with N uniformly distributed data points centered on the peak of the Gaussian. The symbols associated with each curve correspond to the results of Monte Carlo calculations carried out as a check (see Appendix for details).

Fig. 7. The same as Fig. 6 except that the data errors are now proportional to the function, $\sigma_i = \sigma(x_i) \propto y(x_i;a)$. The symbols associated with each curve correspond to the results of Monte Carlo calculations carried out as a check (see Appendix for details).
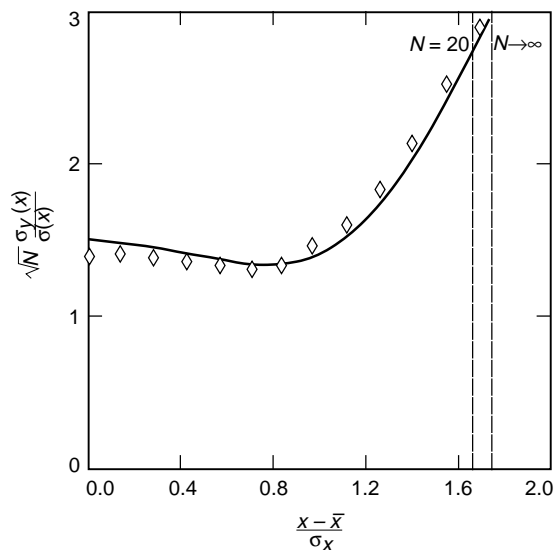
Another important aspect of the general least-squares fitting problem is the optimization of the sampling during the taking of data, e.g., what spacing should one use along the x-axis and how many points should one use in order to reduce the parameter errors to acceptable levels? Since the parameter errors for the case of polynomial fits depend sensitively on the location of the origin of the x-scale, and in any event the coefficients themselves are unlikely to have a fundamental significance in terms of an underlying physical model of the process under study, we restrict ourselves to a consideration of Gaussian fits as an example of some practical importance.

Thus, for uniformly and symmetrically distributed data points, we find the following. For constant data errors $\sigma_i = \sigma$,

$$N\frac{\sigma_{a_1}^2}{\sigma^2} = \frac{NS_4}{(2S_0+1)S_4 - 2S_2^2}$$

$$N\frac{(\sigma_{a_2}/a_3)^2}{(\sigma/a_1)^2} = \frac{1}{2}\frac{N}{\delta_t^2 S_2}$$

$$N\frac{(\sigma_{a_3}/a_3)^2}{(\sigma/a_1)^2} = \frac{1}{2}\frac{N(2S_0+1)}{\delta_t^4[(2S_0+1) - 2S_2^2]}$$

where the sums S_0, S_2, and S_4 are given by Eq. (A-7) of the Appendix and $\delta_t^2 = [12/(N^2-1)](\sigma_x^2/a_3^2) = [12/(N^2-1)]\sigma_t^2$. The normalized errors $\sqrt{N}(\sigma_{a_1}/\sigma)$, $\sqrt{N}[(\sigma_{a_2}/a_3)/(\sigma/a_1)]$, and $\sqrt{N}[(\sigma_{a_3}/a_3)/(\sigma/a_1)]$ are plotted as functions of the variable L/a_3 for selected values of the number of data points, N, in Fig. 8. In the limit $N \to \infty$ and $L/a_3 \geq 4$, the above normalized standard errors become

$$N\frac{\sigma_{a_1}^2}{\sigma^2} = \frac{3}{2\sqrt{\pi}}\frac{L}{a_3}$$

$$N\frac{(\sigma_{a_2}/a_3)^2}{(\sigma/a_1)^2} = \frac{2}{\sqrt{\pi}}\frac{L}{a_3}$$

$$N\frac{(\sigma_{a_3}/a_3)^2}{(\sigma/a_1)^2} = \frac{2}{\sqrt{\pi}}\frac{L}{a_3}$$

For data errors proportional to the function, $\sigma_i = \beta y(x; \mathbf{a})$,

$$N\frac{\sigma_{a_1}^2}{a_1^2} = \beta^2 A_0^{(2)}(N) = \beta^2 \frac{3(3N^2-7)}{4(N^2-4)} \to \frac{9}{4}\beta^2$$

as $N \to \infty$;

$$N\frac{\sigma_{a_2}^2}{a_3^2} = \frac{\beta^2}{\sigma_t^2}$$

independent of N; and

$$N\frac{\sigma_{a_2}^2}{a_3^2} = \frac{\beta^2}{\sigma_t^4}A_4^{(2)}(N) = \frac{\beta^2}{\sigma_t^4}\frac{5(N^2-1)}{4(N^2-4)} \to \frac{5}{4}\frac{\beta^2}{\sigma_t^4}$$

as $N \to \infty$.

II. Introduction

If one measures a single quantity N times in the presence of normally distributed, random errors, then it is well known that the variance of the mean of these measurements is equal to the variance of the measurements themselves divided by N, and in the absence of systematic errors, the mean value approaches the true value of the quantity as the number of measurements increases without limit.
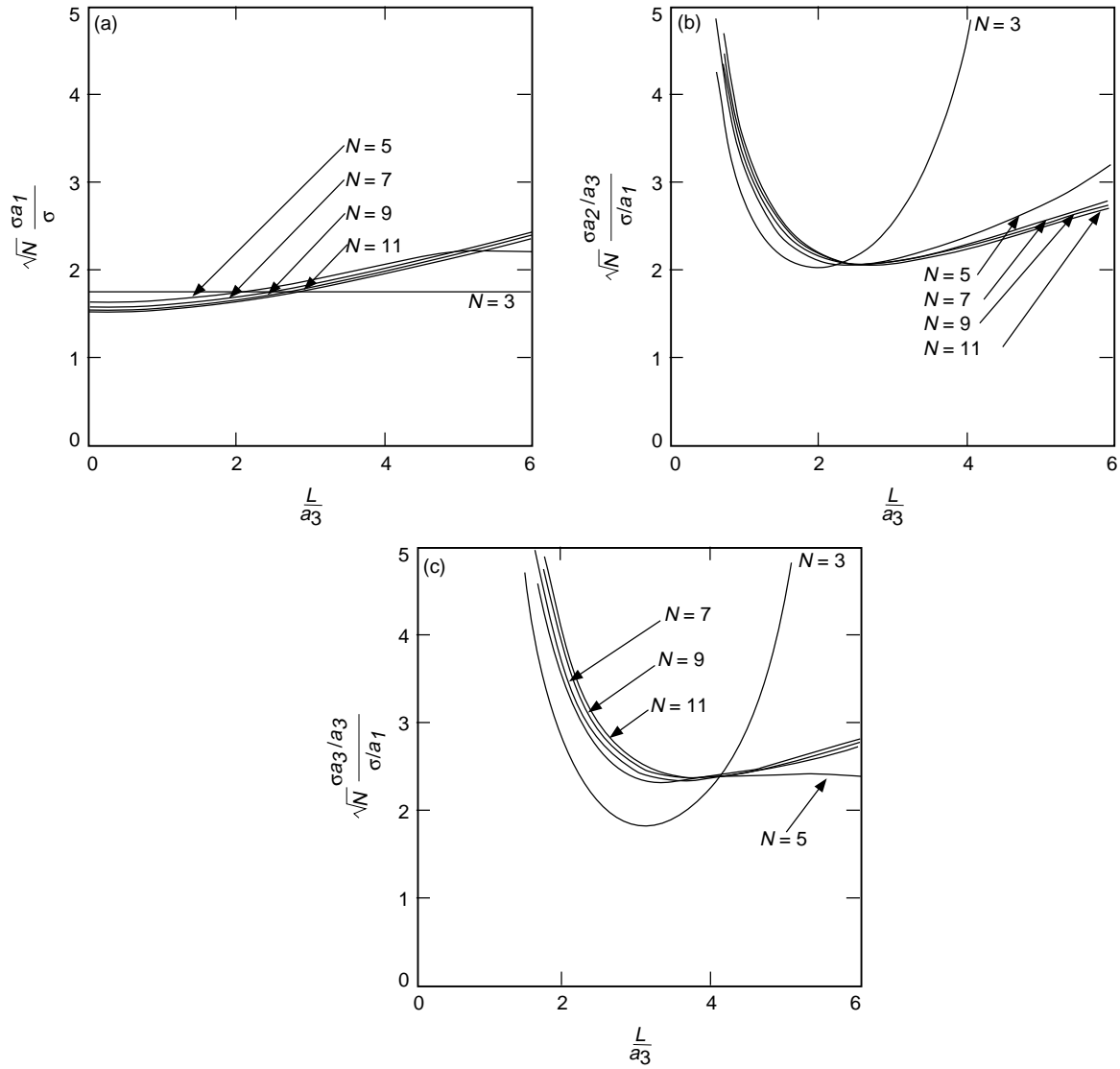
Fig. 8. Universal, normalized error curves for the three parameters of a Gaussian fit for constant data errors, $\sigma_i = \sigma$, as a function of the normalized x-axis interval L/a_3 for various values of the number of data points, N: (a) a_1, (b) a_2, and (c) a_3.

In the case of least-squares fitting of a given function to a given set of data that are likewise subject to normally distributed, random errors, the resulting fit is the mean function corresponding to the data, and the question arises as to what variance to assign to the errors of the values of this function. Here, two related concerns arise. First, the fitting function will contain a certain number of parameters, M, and one or more of these may be of interest in relation to a physical quantity whose value is sought. For example, system noise temperatures may have been recorded as an antenna is scanned through a point source of radiation, and one may be interested in the peak value or the half-width of the antenna pattern or both. Or, one may have determined a series of system noise temperatures at different elevations and wish to know what the maximum system noise temperature is and at what elevation it occurs. In each case, an appropriate fitting function could be chosen so that one or more of the parameters involved corresponds to the quantity or quantities of interest, and one would like to know, therefore, what standard error to assign to the quantities so determined. Alternatively, one may want to know what the standard error is

as a function of the independent variable, say declination or elevation in the two examples cited above. In the remainder of this article, we will refer to this standard error of the value of the fitted function as the error of the fit and designate it by the symbol $\sigma_y(x)$.

The first instance considered above, namely, determining the error of one or more fitting parameters, has a straightforward answer given in terms of the diagonal elements of the covariance matrix of the fit, and is well known. Less well known, however, particularly among nonmathematicians, is the relationship between this matrix and the error of the fit as a function of the independent variable. Some insight into this problem can be obtained by examining Fig. 9, where we show the results of sequentially fitting straight lines to a series of data sets generated by the same linear function, $y = a_1 + a_2x$, but with different random errors, corresponding, however, to the same normal distribution, i.e., having the same constant standard deviation, σ. Since each of these lines could have resulted from the same underlying function, albeit with different probability, the ensemble of all possible lines defines the error statistics of the particular fit actually obtained with the particular set of errors present during the data gathering, i.e., the data errors actually obtained correspond to but one of the infinite number of sets that could have resulted from the measurements. In the case shown in Fig. 9, one can see that the error of the fit tends to be smaller toward the centroid of the data points and larger at the extremes. In fact, it is shown below that the error curve in this case is actually a hyperbola and that the same general behavior is found for higher-degree polynomials, albeit with differing functional dependence. The standard error of the fit, $\sigma_y(x)$, derived below is shown superimposed on the ensemble of straight lines in Fig. 9.

It is the purpose of this article to discuss the above errors and, in particular, to present results that will permit one to determine the standard error of the fit as a function of the independent variable, as well as to establish confidence limits for these errors.
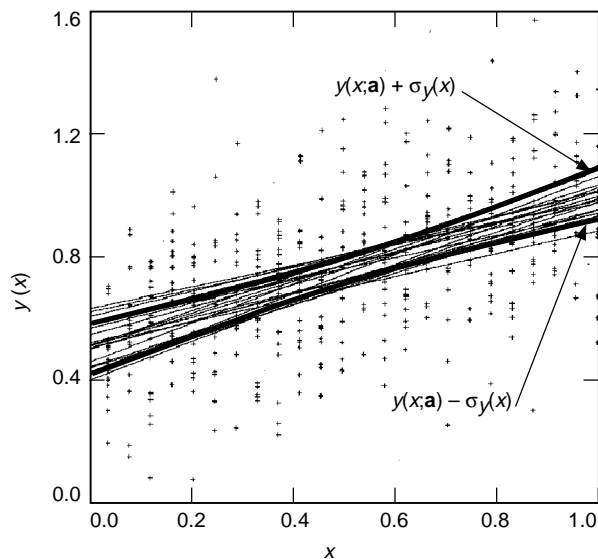
Fig. 9. Plot of 20 straight lines resulting from fits to data generated from the same parent straight line but with different random, normally distributed errors having the same statistics.

III. Least-Squares Fitting

The fitting of data of the form $(x_i, y_i), i = 1, 2, \cdots, N$ by a function $y(x; a_1, \cdots, a_M) \equiv y(x; \mathbf{a})$ depending on M coefficients, a_j, and the independent variable x is frequently used in scientific and engineering

work, either to determine the most likely values of the fitting coefficients, which may relate to some physically reasonable model of the process under study, or simply to permit the prediction of the most likely value of the dependent variable y for a given value of x, including those values where no data exist.

While many techniques for finding optimum values of the fitting coefficients exist, the one most commonly used is the least-squares method, where the coefficients are determined by minimizing the quantity χ^2, given by

$$\chi^2(\mathbf{a}) \equiv \sum_{i=1}^{N} \frac{[y_i - y(x_i; \mathbf{a})]^2}{\sigma_i^2} \tag{1}$$

The variances σ_i^2 of the data values y_i are assumed known, either from a knowledge of the experimental errors involved in the measurements or from analysis of the data itself, and it is assumed that the errors themselves are normally distributed. While this latter requirement is not essential for the derivations that follow, it is nonetheless a common assumption, valid for most measurements, and permits one to establish confidence limits, as discussed below.

This approach, also called the method of maximum likelihood, is described in numerous publications and is the basis of many curve-fitting programs available in various software packages devoted to data analysis. However, in spite of its widespread use, there are a number of aspects of the least-squares fitting problem that are not often discussed in the literature on the subject, especially those having to do with the proper evaluation and interpretation of errors. In the following section, the general linear, least-squares problem, in which the fitting coefficients a_j enter into the fitting function $y(x; \mathbf{a})$ in a linear manner, is formulated, and the solution for the coefficients is obtained. The purpose here is to establish the notation and display the main results, rather than to provide a detailed derivation.[1] This is followed by an analysis of standard errors for the various quantities encountered and a section devoted to illustration of the main ideas through consideration of some simple examples. Next, the general nonlinear, least-squares fitting problem, in which the coefficients a_j enter into the fitting function $y(x; \mathbf{a})$ in a nonlinear way, is discussed, and the standard errors once again analyzed. Two final sections provide a discussion of a related problem involving errors of the difference between a function at two values of the argument and a general discussion of the significance of the results.

IV. The Linear Least-Squares Problem

For the linear, least-squares problem, the fitting function $y(x; \mathbf{a})$ may be written

$$y(x; \mathbf{a}) \equiv \sum_{j=1}^{M} a_j X_j(x) \tag{2}$$

where $X_j(x)$ are arbitrary basis functions of the independent variable, x, and the M coefficients, a_j, are to be determined by minimizing χ^2, as given by Eq. (1). This problem is most often framed in the language of linear algebra, where the fitting function $y(x_i; \mathbf{a})$ is taken as an N-element column vector, the coefficients a_j as an M-element column vector, and the basis functions $X_j(x_i)$ are represented by an $N \times M$ matrix X_{ij}, so that letting $x = x_i$, Eq. (2) may be written

$$y(x_i; \mathbf{a}) = \sum_{j=1}^{M} X_{ij} a_j, \quad i = 1, 2, \cdots, N$$

---

[1] The latter may be found in numerous sources, such as [1], which has been used as a guide to define the notation.

or, in the matrix equation form, $\mathbf{y} = \mathbf{Xa}$.

Similarly, by defining the column vector $b_i \equiv (y_i/\sigma_i)$ and the matrix $A_{ij} \equiv (X_{ij}/\sigma_i)$, Eq. (1) may be written

$$\chi^2 = (\mathbf{b} - \mathbf{Aa})^T(\mathbf{b} - \mathbf{Aa}) \tag{3}$$

where T implies matrix transpose. Thus, the extremum condition for the coefficients,

$$\frac{\partial}{\partial a_j}\left(\chi^2\right) = 0$$

results in the matrix equation

$$\left(\mathbf{A}^T\mathbf{A}\right)\mathbf{a} = \mathbf{A}^T\mathbf{b}$$

having the solution

$$\mathbf{a} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b} = \mathbf{C}\mathbf{A}^T\mathbf{b} \tag{4a}$$

In the above, we have defined the matrix, $\mathbf{H} \equiv \mathbf{A}^T\mathbf{A}$, and its inverse, $\mathbf{C}$. The symmetric $M \times M$ matrix $\mathbf{C} = \mathbf{H}^{-1}$ is called the covariance matrix and is central to the determination of standard errors, as discussed below. In component form, Eq. (4a) may be written

$$a_j = a_j\left(y_1, \cdots, y_N\right) = \sum_{k=1}^{M} C_{jk} \sum_{i=1}^{N} y_i \frac{X_k(x_i)}{\sigma_i^2} \tag{4b}$$

where the dependence of the coefficients on the measured data has been made explicit in the notation.

Before proceeding with a discussion of errors and their evaluation, the question of the suitability of the set of basis functions chosen at the outset should be addressed. It is shown in various treatises on statistics and data analysis that the resulting fit to the function $y(x; \mathbf{a})$ is meaningful if χ^2, as given by Eq. (1), is of the order of $N - M \equiv \nu$, referred to as the number of degrees of freedom for the system [2]. Since the quantity χ^2 should be reasonably close to ν for a good fit, the related quantity $\chi_\nu \equiv \chi/\nu$, called the reduced χ^2, is often used as a measure of the suitability of the chosen basis functions for fitting to the given data, i.e., the condition $\chi_\nu^2 \approx 1$ is taken to indicate that the fit is meaningful. For $\nu \gg 1$, χ^2, and hence the reduced χ^2, is normally distributed, with the latter having a mean of 1 and a standard deviation of $\sqrt{2/\nu}$.

Figure 10 illustrates the significance of the quantity χ_ν^2 for the simple case of a polynomial fit. In Fig. 10(a), the data clearly show a quadratic dependence on the independent variable, but an attempt has been made to obtain a fit with a straight line. The formal application of Eq. (4a) results in the computation of two coefficients for the line, but since at least three coefficients are obviously required to give a reasonable fit to the given data, these coefficients are meaningless, both in terms of providing a decent fit and in terms of the model used to represent the data, and this is manifest in the large value of χ_ν^2 resulting from the fit. On the other hand, increasing the degree of the polynomial to 2 by adding a quadratic term results in a much more reasonable-looking fit and a value of χ_ν^2 close to 1.
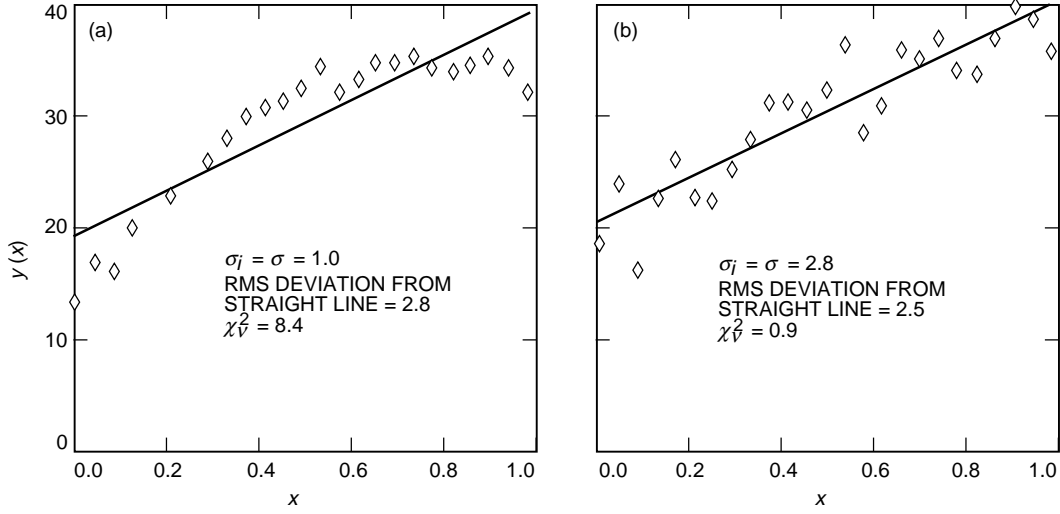
Fig. 10. The significance of the quantity χ^2_ν for a polynomial fit: (a) a straight-line fit to data having a quadratic dependence on the independent variable, illustrating the increase in χ^2_ν resulting from a poor fit to the data, and (b) a straight-line fit to straight-line data. In each case, the constant standard deviation of the data errors, σ, is relative to the parent curve, a quadratic for Fig. 10(a) and a straight line for Fig. 10(b) (see Section IV for discussion).

It must be pointed out, however, that the value of χ^2_ν that one obtains with a given fitting function depends not only on the suitability of the function chosen but also on the values of the data errors assumed, as is evident from the definition given by Eq. (1). Thus, in Fig. 10(b), where a straight-line fit has been obtained for straight-line data, an accurate knowledge of the data errors, σ_i, leads to a reduced χ^2 of 0.9. However, had the data errors in Fig. 10(a) been overestimated, as could result with an overly cautious experimenter, for example, then the calculated χ^2_ν would be smaller than it should be, so that one might end up with a value close to 1 as the result of a relatively poor fit and excessively large σ_i's. This points up the importance of properly assessing the data errors, a subject we will return to in a later section.

V. Standard Errors

Since the coefficient values a_j depend on the data values y_i, as given by Eq. (4b), or its matrix equivalent, Eq. (4a), the uncertainties in the coefficients depend on the uncertainties in the data values. Thus,

$$\delta a_j = \sum_{i=1}^{N} \frac{\partial a_j}{\partial y_i} \delta y_i = \sum_{i=1}^{N} \sum_{k=1}^{M} C_{jk} \frac{X_k(x_i)}{\sigma_i^2} \delta y_i$$

or, in matrix form,

$$\delta \mathbf{a} = \mathbf{C} \mathbf{A}^T \delta \mathbf{b}$$

so that the covariance of a_j and a_k is given by

$$\sigma^2_{\mathbf{a}_{jk}} \equiv \langle \delta a_j \delta a_k \rangle = \sum_{n=1}^{M} \sum_{m=1}^{M} C_{jn} C_{km} \sum_{i=1}^{N} \frac{X_n(x_i)}{\sigma_i^2} \sum_{l=1}^{N} \frac{X_m(x_l)}{\sigma_l^2} \langle \delta y_i \delta y_l \rangle \tag{5a}$$

where $\langle\ \rangle$ indicates an ensemble average, giving the expectation value of the quantity in parenthesis. In matrix form,

$$\sigma_{\mathbf{a}}^2 \equiv \langle \delta\mathbf{a}\delta\mathbf{a}^T \rangle = \left\langle \mathbf{C}\left(\mathbf{A}^T\delta\mathbf{b}\right)\left(\mathbf{A}^T\delta\mathbf{b}\right)^T\mathbf{C}^T \right\rangle = \mathbf{C}\mathbf{A}^T\left\langle \delta\mathbf{b}\delta\mathbf{b}^T \right\rangle\mathbf{A}\mathbf{C}^T \tag{5b}$$

where use has been made of the relation $(\mathbf{A}^T\delta\mathbf{b})^T = \delta\mathbf{b}^T\mathbf{A}$. Since the data errors δy_i are assumed to be statistically uncorrelated, $\langle \delta y_i \delta y_l \rangle = \delta_{il}\sigma_i^2$, where δ_{il} is the Kronecker delta.

Thus, $\langle \delta\mathbf{b}\delta\mathbf{b}^T \rangle = \mathbf{I}$, the identity matrix, so that Eq. (5b) becomes

$$\sigma_{\mathbf{a}}^2 = \mathbf{C}(\mathbf{A}^T\mathbf{A})\mathbf{C}^T = \mathbf{C} \tag{6a}$$

where the definition of the matrix $\mathbf{C} = (\mathbf{A}^T\mathbf{A})^{-1}$ [see Eq. (4a)] and its symmetry have been used. In component form,

$$\sigma_{\mathbf{a}_{jk}}^2 = C_{jk} \tag{6b}$$

so that, unlike the data errors, the resulting errors in the coefficients are correlated, i.e., the off-diagonal terms of the matrix $\mathbf{C}$ do not, in general, vanish. The diagonal elements give the variances of the coefficients,

$$\sigma_{a_j}^2 = C_{jj} \tag{6c}$$

We are now in a position to determine the error in the value of the fitted function $y(x)$ that results from the errors in the coefficients a_j.

From Eq. (2),

$$\delta y(x; \mathbf{a}) = \sum_{j=1}^{M} \delta a_j X_j(x)$$

so that the covariance $\sigma_y^2(x, x') \equiv \langle \delta y(x)\delta y(x') \rangle$ is given by

$$\sigma_y^2(x, x') = \sum_{j=1}^{M}\sum_{k=1}^{M}\langle \delta a_j \delta a_k \rangle X_j(x)X_k(x') = \sum_{j=1}^{M}\sum_{k=1}^{M} C_{jk}X_j(x)X_k(x') \tag{7}$$

We note that this holds for any two values of the independent variable, not just the data points, and that it is independent of the parameter values, a_j.

For the special case $x' = x$, we have the variance

$$\sigma_y^2(x) = \sum_{j=1}^{M}\sum_{k=1}^{M} C_{jk}X_j(x)X_k(x) = \mathbf{x}(x)^T\mathbf{C}\,\mathbf{x}(x) \tag{8}$$

where $\mathbf{x}(x)$ is a column vector whose elements are $X_j(x)$. Since the above expression for $\sigma_y^2(x, x')$ does not, in general, vanish for $x' \neq x$, the errors in $y(x; \mathbf{a})$ at two different values of x are correlated, unlike the data errors. Some insight into the significance of the result given by Eq. (8) can be obtained by considering the weighted mean value of $\sigma_y^2(x)$ averaged over all data points $x = x_i$:

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\sigma_y^2(x_i)}{\sigma_i^2} = \frac{1}{N} \sum_{j=1}^{M} \sum_{k=1}^{M} C_{jk} \sum_{i=1}^{N} \frac{X_{ij}}{\sigma_i} \frac{X_{ik}}{\sigma_i}$$

From the definition of the matrix $\mathbf{A}$, $(X_{ij}/\sigma_i) \equiv A_{ij}$, and hence the sum over i is just $(\mathbf{A}^T \mathbf{A})_{kj} = h_{kj} = (\mathbf{C}^{-1})_{kj}$, thus giving us

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\sigma_y^2(x_i)}{\sigma_i^2} = \frac{1}{N} \sum_{j=1}^{M} \sum_{k=1}^{M} C_{jk} (\mathbf{C}^{-1})_{jk} = \frac{1}{N} \sum_{j=1}^{M} I_{jj} = \frac{M}{N} \tag{9}$$

For the special case of constant data errors, independent of the value of x, $\sigma_i = \sigma = $ constant, so that we obtain

$$\overline{\sigma_y^2} = \frac{1}{N} \sum_{i=1}^{N} \sigma_y^2(x_i) = \frac{M}{N} \sigma^2 \tag{10}$$

This result is analogous to the variance of the mean for the case of one value of x, where $M = 1$. Thus, we see that, as a result of fitting to the function $y(x)$, the variance of y, averaged over the data points, is reduced by the ratio M/N relative to the constant data variance. This implies that the higher the order of the fit required by the data to reduce χ_ν^2 to a value close to 1, the larger will be the resulting errors of the values of the fitted function.

In the following section, we examine these results for the simple case of a straight-line fit to the data, where an exact analysis enables one to bring out some of the more important consequences of the general theory.

VI. Illustrative Examples

A. Fitting to a Straight Line

In order to illustrate some of the main consequences of the theory developed above, we first consider the simplest linear, least-squares fitting problem, that of a straight line [3], for which $M = 2$. Then, the basis functions are given by $X_{ij} = x_i^{j-1}, j = 1, 2$, and the fitting function by

$$y(x_i; \mathbf{a}) = \sum_{j=1}^{2} a_j x_i^{j-1} = a_1 + a_2 x_i$$

The matrix $\mathbf{H}$ is, thus,

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} \dfrac{1}{\sigma_i^2} & \sum_{i=1}^{N} \dfrac{x_i}{\sigma_i^2} \\ \sum_{i=1}^{N} \dfrac{x_i}{\sigma_i^2} & \sum_{i=1}^{N} \dfrac{x_i^2}{\sigma_i^2} \end{bmatrix} \tag{11}$$

while its inverse, $\mathbf{C}$, is

$$\mathbf{C} = \frac{1}{\Delta} \begin{bmatrix} h_{22} & -h_{12} \\ -h_{21} & h_{11} \end{bmatrix}$$

In the above, $\Delta = h_{11}h_{22} - h_{12}^2$ and $h_{12} = h_{21}$, thus demonstrating the general result that both $\mathbf{H}$ and $\mathbf{C}$ are symmetric.

For the special case where the data variances are independent of x, the diagonal terms of the latter, which correspond to the variances of the constant and the linear coefficient, are given by

$$C_{11} = \sigma_{a_1}^2 = \frac{\sigma^2}{N - \left(\sum_{i=1}^N x_i\right)^2 / \sum_{i=1}^N x_i^2}$$

$$C_{22} = \sigma_{a_2}^2 = \frac{\sigma^2}{\sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i\right)^2 / N}$$

The variance of the data points is[2]

$$\sigma_x^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \frac{1}{N^2} \left(\sum_{i=1}^N x_i\right)^2$$

and if we choose the origin of the x-scale symmetrically, then the sum over x_i vanishes so that we may write the above in the simplified form

$$C_{11} = \sigma_{a_1}^2 = \frac{\sigma^2}{N}$$

$$C_{22} = \sigma_{a_2}^2 = \frac{1}{N} \frac{\sigma^2}{\sigma_x^2}$$

These results agree with our intuition in that the variance of the constant term is just the variance of the mean of the data values, while the variance of the slope decreases from this as the spread of the data points increases. From Eq. (8), the variance of the value of the fitted function, in this case a straight line, becomes

$$\sigma_y^2(x) = \sum_{j=1}^2 \sum_{k=1}^2 C_{jk} x^{j+k-2}$$

$$= C_{11} + 2C_{12}x + C_{22}x^2$$

$$= \frac{1}{\Delta}\left(h_{22} - 2h_{12}x + h_{11}x^2\right) \tag{12}$$

---

[2] For convenience, we are using a definition involving $1/N$ rather than $1/(N-1)$.

This is the equation of a conic section, for which the characteristic is

$$B^2 - 4AC = \frac{4h_{11}}{\Delta}$$

where $h_{11} = \sum_{i=1}^{N} \left(1/\sigma_i^2\right)$ and $\Delta = \sum_{i=1}^{N} \left(1/\sigma_i^2\right) \sum_{l=1}^{N} \left(x_l^2/\sigma_l^2\right) - \left(\sum_{i=1}^{N} [x_i/\sigma_i]\right)^2$. Letting $1/\sigma_i = a_i$ and $x_i/\sigma_i = b_i$, in the latter, and applying the Cauchy–Schwarz inequality,

$$\sum_{i=l}^{N} a_i^2 \sum_{l=1}^{N} b_l^2 \geq \left(\sum_{i=1}^{N} a_i b_i\right)^2$$

we see that $B^2 - 4AC \geq 0$, so that the equation for $\sigma_y(x)$ is that of a hyperbola. Similarly, the discriminant of the quadratic on the right-hand side of Eq. (12) is $b^2 - 4ac = -(4/\Delta) \leq 0$, so that the roots of the equation $\sigma_y^2(x) = 0$ are complex, i.e., the hyperbola is symmetric about the x-axis. The minimum value of $\sigma_y^2(x)$ occurs at $x_{min} = h_{12}/h_{11}$, for which the variance is $\sigma_y^2|_{min} = 1/h_{11}$. For constant data variances, $\sigma_i = \sigma$, $x_{min} = (1/N) \sum_{i=1}^{N} x_i = \overline{x}$, and $\sigma_y^2|_{min} = (1/h_{11}) = (\sigma^2/N)$, while at $x = \overline{x} \pm \sigma_x$,

$$\sigma_y^2 \left(\overline{x} \pm \sigma_x\right) = \frac{2}{h_{11}} = 2\frac{\sigma^2}{N} = \overline{\sigma_y^2} \tag{13}$$

showing that the mean value of $\sigma_y^2(x)$ occurs at one standard deviation from the mean value of x.

If the values of the matrix elements h_{jk} given by Eq. (11) are substituted into Eq. (12) for the special case $\sigma_i = \sigma$, it may be transformed into the simple dimensionless form

$$\eta^2 = 1 + \xi^2 \tag{14}$$

where $\eta = \sqrt{N}([\sigma_y(x)]/\sigma)$, $\xi = ([x - \overline{x}]/\sigma_x)$. The universal error curve represented by Eq. (14) is plotted, along with the higher-degree polynomial error curves discussed below, in Fig. 5.

B. More General Fits

The case of a general polynomial fit may be treated by the same methods used above, but the algebra becomes progressively more difficult for $M > 2$. However, one can see that for higher-degree polynomial fits, the general behavior illustrated by the straight-line case dealt with above is also found. For example, since the highest power present in the expansion for $\sigma_y^2(x)$ is $2(M-1)$ while the fit itself is of degree $M-1$, the equation for $\sigma_y(x)$ corresponds to a double-branched function whose asymptotic behavior for $|x - \overline{x}| \gg \sigma_x$ follows the same power law as the fit itself. Thus, a quadratic fit yields an error that increases asymptotically as the second power of $|x - \overline{x}|$, a cubic as the third power, etc. The implication of this is that higher-degree fits become less reliable than ones of lower degree as one moves away from the center of the given data.

It is shown in the Appendix that if the data variances are constant, and the data points are uniformly distributed along the x-axis, then the variance of the fitted function $\sigma_y^2(x)$ is a symmetric function of $x - \overline{x}$, i.e., $\sigma_y(\overline{x} + \Delta x) = \sigma_y(\overline{x} - \Delta x)$. Furthermore, it is possible to express $\sigma_y^2(x)$ in the same dimensionless form as for the straight-line case for the higher-degree polynomial fits. Thus, for the general polynomial of degree $M - 1$, a set of universal, normalized error curves exists, the square of which is given by

$$\eta^2 = \sum_{j=0}^{M-1} A_{2j}^{(M-1)}(N)\xi^{2j}$$

where

$$\eta = \sqrt{N}\frac{\sigma_y(x)}{\sigma}$$

$$\xi = \frac{x - \overline{x}}{\sigma_x}$$

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \frac{1}{N^2}\left(\sum_{i=1}^{N} x_i\right)^2$$

While the coefficients appearing in the above expression generally depend on the number of data points, N, they do not for the straight-line case. Furthermore, for values of N exceeding that required for a given fit by some reasonable number, say 5, they become essentially constant for polynomial fits of all degrees. These results are summarized in Table 1, which lists the coefficients $A_{2j}^{(M-1)}(N)$, including their limiting values for $N \to \infty$, for straight-line, quadratic, and cubic fits, corresponding to $M = 2, 3$, and 4. The corresponding normalized standard deviations of the fit, η, are plotted in Fig. 5.

In the general case where the basis functions are not polynomials, the standard errors are not constant, or the data points are not uniformly distributed, one must use Eq. (8) to determine the behavior of the error as a function of the independent variable, so that this equation takes on fundamental importance in determining confidence limits resulting from a given fit to a given set of data. An example of the direct use of this equation is shown in Figs. 1 through 4, where a set of antenna aperture efficiency data has been fitted to a quadratic function. The figures illustrate the behavior of the standard error of the fit over a wide range of elevation angles, including those beyond the existing data. We have assumed from the outset that the data errors are normally distributed, and under these circumstances, it may be shown that the errors in the coefficients and, consequently, also in the fitting function $y(x)$ are also normally distributed, so that Gaussian statistics may be used to determine these limits.

As a final comment, it should be observed that the above analysis assumes that the fit is meaningful, i.e., that $\chi_\nu^2 \approx 1$, and this is crucial to a proper interpretation of the fitting errors. If the fitting function has been improperly chosen, one may not assume that the errors follow Eq. (8) and, equally important, if the data errors have been incorrectly estimated, one may not assume that the fit is meaningful even if $\chi_\nu^2 \approx 1$, since these errors appear directly in the expression for χ^2 [Eq. (1)]. It is, therefore, of the utmost importance to have a reasonably accurate estimate of the actual data errors if one is to have reasonable estimates of the resulting fitting errors.

VII. The Nonlinear Least-Squares Problem

The linear, least-squares problem is characterized by Eqs. (2) and (3), which, respectively, express the fitting function $y(x; \mathbf{a})$ as a linear and χ^2 as a quadratic function of the coefficients a_j, thus allowing an explicit solution to be obtained for the latter, as shown in Section IV. In the nonlinear case, no such

simple formulation is possible, and one is led to an iterative procedure for the solution that starts with an assumed solution vector $\underline{\mathbf{a}}$ and evolves so as to produce values closer and closer to the value $\underline{\mathbf{a}_0}$, which minimizes χ^2. The iteration is based on the assumption that if $\underline{\mathbf{a}}$ is sufficiently close to $\underline{\mathbf{a}_0}$, then $y(x; \mathbf{a})$ may still be expressed as a linear and χ^2 as a quadratic function of the coefficients. This assumption rests on using Taylor expansions for these quantities and retaining only the leading terms. Thus, for $\underline{\mathbf{a}}$ near $\underline{\mathbf{a}_0}$, we may write

$$y(x; \mathbf{a}) = y(x; \mathbf{a}_0) + \sum_{j=1}^{M} \left. \frac{\partial y(x; \mathbf{a})}{\partial a_j} \right|_{\mathbf{a}_0} (a_j - a_{0j})$$

$$= y(x; \mathbf{a}_0) + (\mathbf{a} - \mathbf{a}_0)^T \, \mathbf{d}(x; \mathbf{a}_0) \tag{15}$$

where $[\mathbf{d}(x; \mathbf{a}_0)]_j \equiv d_j(x; \mathbf{a}_0) = (\partial y(x; \mathbf{a})/\partial a_j)|_{\mathbf{a}_0}$. Similarly,

$$\chi^2(\mathbf{a}) = \chi^2_{min} + \frac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{M} \left. \frac{\partial^2 \chi^2(\mathbf{a})}{\partial a_j \partial a_k} \right|_{\mathbf{a}_0} (a_j - a_{0j})(a_k - a_{0k})$$

$$= \chi^2_{min} + (\mathbf{a} - \mathbf{a}_0)^T \, \mathbf{H}(\mathbf{a}_0)(\mathbf{a} - \mathbf{a}_0) \tag{16}$$

where $\chi^2_{min} = \chi^2(\mathbf{a}_0)$, and

$$h_{jk}(\mathbf{a}_0) = \frac{1}{2} \left. \frac{\partial^2 \chi^2(\mathbf{a})}{\partial a_j \partial a_k} \right|_{\mathbf{a}_0} = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left. \frac{\partial y(x_i; \mathbf{a})}{\partial a_j} \right|_{\mathbf{a}_0} \left. \frac{\partial y(x_i; \mathbf{a})}{\partial a_k} \right|_{\mathbf{a}_0}$$

.

Note that the gradient term vanishes in the above expression for $\chi^2(\mathbf{a})$ since it is evaluated at the minimum. Also, the final expression for $h_{jk}(\mathbf{a}_0)$ involves only first derivatives because of the assumed form for $y(x; \mathbf{a})$ given by Eq. (15). If we now define

$$A_{ij}(\mathbf{a}_0) = \frac{1}{\sigma_i} \left. \frac{\partial y(x_i; \mathbf{a})}{\partial a_j} \right|_{\mathbf{a}_0} = \frac{d_{ij}(\mathbf{a}_0)}{\sigma_i}$$

then $\mathbf{H}(\mathbf{a}_0) = \mathbf{A}^T \mathbf{A}$ just as in the linear case and, hence, the derivation leading up to Eq. (6) carries through just as before, giving us

$$\sigma_a^2(\mathbf{a}_0) = \mathbf{C}(\mathbf{a}_0) = \mathbf{H}^{-1}(\mathbf{a}_0)$$

Also, the derivation leading up to Eq. (8) follows through as before when $d_j(x; \mathbf{a}_0)$ is substituted for $X_j(x)$, giving the result

$$\sigma_y^2(x; \mathbf{a}_0) \;=\; \sum_{j=1}^{M} \sum_{k=1}^{M} C_{jk}(\mathbf{a}_0) d_j(x; \mathbf{a}_0) d_k(x; \mathbf{a}_0)$$

$$= \mathbf{d}(x; \mathbf{a}_0)^T \mathbf{C}(\mathbf{a}_0) \mathbf{d}(x; \mathbf{a}_0)$$

as does that leading up to Eqs. (9) and (10).

Thus, the results

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\sigma_y^2(x_i)}{\sigma_i^2} = \frac{M}{N}$$

and

$$\overline{\sigma_y^2} = \frac{1}{N} \sum_{i=1}^{M} \sigma_y^2(x_i) = \frac{M}{N} \sigma^2$$

continue to hold in the general, nonlinear case so that with the appropriate definitions for the quantities involved, the nonlinear and linear, least-squares problems lead to the same formal results. Note, however, that all of the quantities now explicitly depend on the solution vector, $\underline{\mathbf{a}}_0$.

Furthermore, within the assumptions made in writing down Eqs. (15) and (16), i.e., assuming that the deviations δa_j are not so large as to invalidate the linear and quadratic approximations involved, one may assume that all of the distributions are normal and, hence, that the usual confidence limits derived from such distributions hold. Thus, the limits for a confidence level of 68.3 percent are $\pm\sigma_X$, those for 95.4 percent $\pm2\sigma_X$, for 99.73 percent $\pm3\sigma_X$, etc., where X corresponds to either $\underline{\mathbf{a}}$ or y. As an example of the application of the above theory to a specific case, the nonlinear fitting problem for the Gaussian function

$$y(x; \mathbf{a}) = a_1 \exp\left[\frac{-(x - a_2)^2}{2a_3^2}\right]$$

is treated in the Appendix, where it is shown that the normalized variance of the fit for the case of constant data errors is given by

$$\eta^2 = \sqrt{\frac{3}{\pi}} \sigma_t e^{-(\xi\sigma_t)^2} \left(3 + 4\sigma_t^4 \xi^4\right)$$

under the assumptions that the number of data points is reasonably large, say 20 or more, that they are uniformly and symmetrically distributed about the peak $x = a_2$, and that they extend out to a distance of $x - a_2 = \pm2a_3$ or more. In the above expression,

$$\eta = \sqrt{N} \frac{\sigma_y(x)}{\sigma}$$

and

124

$$\sigma_t = \frac{\sigma_x}{a_3}$$

where $\sigma_x^2 = (1/N)\sum_{i=1}^{N} x_i^2 - (1/N^2)(\sum_{i=1}^{N} x_i)^2$, as before. This equation is plotted in Fig. 6 for $\sigma_t = 2/\sqrt{3}$ and $4/\sqrt{3}$, corresponding to $L/a_3 = 4$ and 8, where L is the full range of the data points.

If the data errors, instead of being constant, are proportional to the value of the function, $\sigma_i = \sigma(x_i) \propto y(x_i; \mathbf{a})$, then the normalized error of the fit is given by

$$\eta^2 = A_0^{(2)}(N) + A_2^{(2)}(N)\xi^2 + A_4^{(2)}(N)\xi^4$$

where now $\eta = \sqrt{N}\sigma_y(x)/\sigma(x)$ and the only restriction on the data points is that they be uniformly and symmetrically distributed relative to the peak. The coefficients $A_{2j}^{(2)}(N)$ appearing in the above expression are the same as those found for the quadratic polynomial fit, and their values, including those for $N \to \infty$, are give in Table 1. The equation is plotted in Fig. 7.

The functional dependence of the parameter errors, σ_{a_j}, on the data interval, L, and the number of points, N, is frequently of importance in determining an optimum sampling strategy, and this is derived in Section V of the Appendix, with the results for constant data errors plotted in Fig. 8.

VIII. A Related Problem

It is sometimes of interest to determine the variance of the difference of the fit at two points x_1 and x_2, $\Delta y(x_1, x_2) = y(x_1; \mathbf{a}) - y(x_2; \mathbf{a})$. Thus, from Eq. (13), we have

$$\delta\left[\Delta y(x_1, x_2)\right] = \delta y(x_1; \mathbf{a}) - \delta y(x_2; \mathbf{a})$$

$$= \sum_{j=1}^{M} [d_j(x_1) - d_j(x_2)]\,\delta a_j$$

so that

$$\sigma_{\Delta_{1,2}}^2 \equiv \left\langle \{\delta\left[\Delta y(x_1, x_2)\right]\}^2 \right\rangle = \sum_{j=1}^{M}\sum_{k=1}^{M} \langle \delta a_j \delta a_k \rangle (d_{1j} - d_{2j})(d_{1k} - d_{2k})$$

$$= \sum_{j=1}^{M}\sum_{k=1}^{M} C_{jk}(d_{1j} - d_{2j})(d_{1k} - d_{2k})$$

For the case of a polynomial expansion, $d_{ij} = X_{ij} = x_i^{j-1}$, so

$$\sigma_{\Delta_{1,2}}^2 = \sum_{j=1}^{M}\sum_{k=1}^{M} C_{jk}\left(x_1^{j-1} - x_2^{j-1}\right)\left(x_1^{k-1} - x_2^{k-1}\right)$$

and, hence, using the results of Section VI, we obtain

$$\sigma^2_{\Delta_{1,2}} = C_{22}(x_1 - x_2)^2 = \sigma^2_{a_2}(x_1 - x_2)^2$$

for a straight-line fit. For constant data errors,

$$\sigma^2_{\Delta_{1,2}} = \frac{1}{N} \frac{\sigma^2}{\sigma^2_x}(x_1 - x_2)^2$$

IX. Discussion

In applying the above results, it is important to be aware that the variances of the fit obtained in Eq. (8) and related equations are valid only if one is reasonably confident of the fit itself. If the fitting function being considered does not accurately represent the data, then one must find one that does before attempting to assign appropriate errors, and the best test for this is the value of the reduced χ^2. In this connection, it is worth mentioning that while fits using ordinary polynomials are often used for convenience and simplicity, they seldom correspond to physical reality. They may approximate a given physical process reasonably well over a limited range of the independent variable x, but one should exercise care when attempting to go beyond this, because the hallmark of these functions is that they diverge for large absolute values of x, and the variances of the fit derived for the polynomials in the Appendix and discussed in Section VI offer a warning of this by doing likewise in the region beyond the existing data.

Clearly, one is better off with a fitting function that reflects the underlying physics reasonably well, but of course it is not always possible to find such a function, especially if the system being studied has a large number of basic processes going on at the same time, some of which may be unknown to the observer. A good example is trying to determine the microwave spectrum of a complex astrophysical source of radiation, such as a galaxy, or even a planet, where not only is the physics not known with certainty, but also the data have typically been gathered by many different workers using different measuring systems with different accuracies at different frequencies. Under such circumstances, one cannot be sure that some hidden absorption or emission feature has not gone undetected because of a gap in the measurements and knowledge, over some small range of frequencies, possibly at one of the extremes of the existing measurements. A case in point is the HII region DR21, where early measurements by workers identified the basic radiation process as thermal bremsstrahlung and, on that basis, predicted its microwave spectrum out to frequencies of 100 GHz, unaware of the fact that the region was surrounded by a dust cloud that converted intense visible and ultraviolet radiation to near-infrared radiation, which resulted in significant departure from the predicted spectrum at frequencies as low as 85 GHz. The existing measurements of the time did not extend beyond 31.4 GHz, so considerable extrapolation was involved, even on a logarithmic scale.

Spectral determinations are particularly treacherous, especially at the extremes of the data, because, in principle, the range of the independent variable, frequency, is open at both ends, thus precluding a knowledge of the asymptotic behavior. At the opposite extreme, however, one may be dealing with a relatively uncomplicated situation involving a finite and known range for the independent variable and a system that can reliably be modeled in a fairly simple manner. In such a case, it pays to determine the errors as accurately as one can, i.e., to make maximum use of the given data by applying the variance of the fit theory discussed above.

The question of how best to determine the actual data errors is not an easy one. If the measuring system and the system under study are simple, it may be possible to determine these from first principles, combined with the performance data for the instrumentation used, and it might further be possible to check this by a series of measurements on a known system. Often, however, one does not have such a simple circumstance to deal with, nor does one have the time to make the requisite test measurements.

In this case, one might obtain a reasonable estimate of the data errors by breaking up the range of the independent variable into a small number of segments over each of which the functional dependence may be reasonably well approximated by a quadratic, and fitting each of these restricted data sets by a quadratic assuming $\sigma = 1$. Then, $\chi_\nu^2(\sigma = 1) = (1/\nu) \sum_{i=1}^{N}[y_i - y(x_i)]^2$ and $\chi_\nu^2(\sigma) = (1/\nu\sigma^2) \sum_{i=1}^{N}[y_i - y(x_i)]^2 \approx 1$ so that $\sigma^2 \approx \chi_\nu^2(\sigma = 1)$. A plot of the resulting values of σ as a function of the mean value of the independent variable for each segment would then provide an estimate for $\sigma(x)$, which would in turn permit an assignment of the appropriate values for the σ_i's for the final overall fit. In the final analysis, as always, common sense and experience must be the ultimate guide in determining how best to proceed.

Acknowledgments

References

[1] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*, Chapter 14, New York: Cambridge University Press, 1992.

[2] P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, New York: McGraw-Hill, 1969.

[3] F. S. Acton, *Analysis of Straight-Line Data*, New York: John Wiley & Sons, 1959.

[4] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Corrected and Enlarged Edition, New York: Academic Press, 1980.

Appendix

Mathematical Details

I. Symmetry Considerations in Polynomial Fits

If the basis functions chosen are the ordinary polynomials and constant data errors are assumed, the curvature matrix elements are given by

$$h_{jk} = \frac{1}{\sigma^2} \sum_{i=1}^{N} x_i^{j+k-2}$$

Furthermore, if the data points are uniformly distributed, we may choose our y-axis so that $\sum_{i=1}^{N} x_i^{j+k} = 0$ for $j + k =$ odd integer.[3] Then, the curvature matrix has the general appearance

$$\mathbf{H} = \begin{bmatrix} h_{11} & 0 & h_{13} & 0 & \cdots \\ 0 & h_{22} & 0 & h_{24} & \cdots \\ h_{13} & 0 & h_{33} & 0 & \cdots \\ 0 & h_{24} & 0 & h_{44} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{A-1}$$

where the symmetry has been explicitly indicated, i.e., $h_{jk} = h_{kj}$. The elements of the covariance matrix are given by

$$C_{jk} = \frac{1}{\Delta} \times \text{cofactor}(h_{jk})$$

where $\Delta = \det \mathbf{H}$ and use has been made of the symmetry of $\mathbf{H}$. By an inductive process, one can show that if $j + k$ is an odd integer, then the cofactor matrix will, for a given rank of the curvature matrix, have $n + 1$ rows of identical sequences of zero and nonzero elements, where the number of the latter in a given such row is n. Thus, if the rank of $\mathbf{H}$ is 4 or 5, $n = 1$, if it is 6 or 7, $n = 2$, etc. In view of this structure of the cofactor matrix, one may perform a Gauss elimination on these rows and thereby end up with one row consisting of nothing but zeroes, in consequence of which the corresponding cofactor and, hence, also the covariance matrix element, C_{jk}, vanish for $j + k$ equal to an odd integer. The covariance matrix, therefore, has the same structure as the curvature matrix, so that Eq. (8) for the variance of the fit becomes $\sigma_y^2(x) = \sum_{j=1}^{M} \sum_{k=1}^{M} C_{jk} x^{j+k-2}$ for $j + k =$ even integer. Hence, $\sigma_y^2(x)$ is an even function of x, i.e., the variance of the fit for polynomials is symmetric about the mean of the data points if these are uniformly distributed and the data errors are constant.

II. Polynomial Fits for *M* = 2, 3, and 4 With Constant Data Errors and Uniformly Distributed Data Points

A. Straight-Line Fits

While straight-line fits have already been dealt with in Section VI, the simplicity of this case makes it a useful starting point for discussion of the above ideas. Thus, assuming constant data errors and

---

[3] The y-axis location affects the covariance matrix, but not the variance of the fit.

uniformly distributed data points, we may introduce the variable $x' = x - \overline{x}$, where $\overline{x} = (1/N) \sum_{i=1}^{n} x_i$, in terms of which the curvature and covariance matrices have the sparse form shown in Eq. (A-1) above. Then, for $M = 2$, we have

$$\mathbf{H} = \begin{bmatrix} h_{11} & 0 \\ 0 & h_{22} \end{bmatrix}$$

so that $\Delta = \det \mathbf{H} = h_{11} h_{22}$, and

$$\mathbf{C} = \begin{bmatrix} \dfrac{1}{h_{11}} & 0 \\ 0 & \dfrac{1}{h_{22}} \end{bmatrix} = \dfrac{\sigma^2}{N} \begin{bmatrix} 1 & 0 \\ 0 & \dfrac{1}{\sigma_x^2} \end{bmatrix}$$

where $\sigma_x^2 = (1/N) \sum_{i=1}^{N} (x_i - \overline{x})^2 = (1/N) \sum_{i=1}^{N} x_i'^2$. Substituting into Eq. (8), we thus have

$$\sigma_y^2(x) = \frac{\sigma^2}{N} [1 \ \ x'] \begin{bmatrix} 1 & 0 \\ 0 & \dfrac{1}{\sigma_x^2} \end{bmatrix} \begin{bmatrix} 1 \\ x' \end{bmatrix} = \frac{\sigma^2}{N} \left[1 + \left(\frac{x'}{\sigma_x} \right)^2 \right]$$

This may be written in the normalized form

$$\eta^2 = 1 + \left(\frac{x'}{\sigma_x} \right)^2 = 1 + \xi^2 \tag{A-2}$$

where $\eta = \sqrt{N}([\sigma_y(x)]/\sigma)$ and $\xi = ([x - \overline{x}]/\sigma_x)$, in agreement with the results of Section VI.A [Eq. (14)].

Note that while the above symmetric result has been obtained under the restrictive condition that the data points are uniformly distributed, no such assumption was made in the original derivation given in Section VI, i.e., as long as the data variances are constant, the data points can be distributed in any manner and the result of Eq. (A-2) still holds. As we shall see in what follows, this is a special feature of the straight-line fit that does not carry over to polynomial fits of higher degree.

B. Quadratic Fits

The general form for the variance of the fit for the quadratic case is

$$\sigma_y^2(x) = [1 \ \ x' \ \ x'^2] \begin{bmatrix} C_{11} & 0 & C_{13} \\ 0 & C_{22} & 0 \\ C_{13} & 0 & C_{33} \end{bmatrix} \begin{bmatrix} 1 \\ x' \\ x'^2 \end{bmatrix} \tag{A-3}$$

where $x' = x - \overline{x}$ as before, and

$$\mathbf{H} = \begin{bmatrix} h_{11} & 0 & h_{13} \\ 0 & h_{22} & 0 \\ h_{13} & 0 & h_{33} \end{bmatrix}$$

Inverting this, the matrix elements of $\mathbf{C}$ are

$$C_{11} = \frac{h_{33}}{h_{11}h_{33} - h_{13}^2}$$

$$C_{13} = -\frac{h_{13}}{h_{11}h_{33} - h_{13}^2}$$

$$C_{22} = \frac{1}{h_{22}}$$

$$C_{33} = \frac{h_{11}}{h_{11}h_{33} - h_{13}^2}$$

(A-4)

where $h_{11} = N/\sigma^2$, $h_{13} = h_{22} = (1/\sigma^2)\sum_{i=1}^{N} x_i'^2 = N(\sigma_x^2/\sigma^2)$, and $h_{33} = (1/\sigma^2)\sum_{i=1}^{N} x_i'^4$. The last term may be evaluated by making use of the uniformity of the distribution of the points. Thus, designating the spacing between adjacent points by δ, we have

$$\sum_{i=1}^{N} x_i'^4 = 2\delta^4 \sum_{n=1}^{(N-1)/2} n^4 = \frac{\delta^4 N}{240}(N^2 - 1)(3N^2 - 7)$$

where the last step has used the relation

$$\sum_{k=1}^{n} k^4 = \frac{1}{30}n(n+1)(2n+1)(3n^2 + 3n - 1)$$

from [4, Eq. (0.121.4)], and for definiteness we have assumed an odd number of points.[4] Similarly, applying the relation $\sum_{k=1}^{n} k^2 = (1/6)n(n+1)(2n+1)$ to the expression for h_{22} above, we obtain the following relationship between the point spacing, δ, and the standard deviation of the data points, σ_x:

$$\delta^2 = \frac{12}{N^2 - 1}\sigma_x^2$$

so that we may write h_{33} in the form

$$h_{33} = \frac{3(3N^2 - 7)}{5(N^2 - 1)}\left(N\frac{\sigma_x^4}{\sigma^2}\right)$$

Substituting these results into Eq. (A-3), we finally obtain

$$\eta^2 = \frac{1}{4(N^2 - 4)}\left[3(3N^2 - 7) - 6(N^2 + 1)\xi^2 + 5(N^2 - 1)\xi^4\right]$$

$$= A_0^{(2)}(N) + A_2^{(2)}(N)\xi^2 + A_4^{(2)}(N)\xi^4$$

---

[4] The results obtained do not depend on whether N is even or odd.

where $\xi = ([x - \overline{x}]/\sigma_x)$, as before. In the limit as $N \to \infty$, and this limit is approached quite rapidly as N increases, this reduces to the simple fourth-degree polynomial

$$\eta^2 = \frac{9}{4} - \frac{3}{2}\xi^2 + \frac{5}{4}\xi^4$$

C. Cubic Fits

For the cubic case, the variance of the fit has the form

$$\sigma_y^2(x) = \begin{bmatrix} 1 & x'^2 & x'^4 & x'^6 \end{bmatrix} \begin{bmatrix} C_{11} & 0 & C_{13} & 0 \\ 0 & C_{22} & 0 & C_{24} \\ C_{13} & 0 & C_{33} & 0 \\ 0 & C_{24} & 0 & C_{44} \end{bmatrix} \begin{bmatrix} 1 \\ x'^2 \\ x'^4 \\ x'^6 \end{bmatrix}$$

$$= C_{11} + (C_{22} + 2C_{13})x'^2 + (2C_{24} + C_{33})x'^4 + C_{44}x'^6 \tag{A-5}$$

and the curvature matrix is of the form

$$\mathbf{H} = \begin{bmatrix} h_{11} & 0 & h_{13} & 0 \\ 0 & h_{22} & 0 & h_{24} \\ h_{13} & 0 & h_{33} & 0 \\ 0 & h_{24} & 0 & h_{44} \end{bmatrix}$$

Inverting this, we have

$$C_{11} = \frac{h_{33}}{h_{11}h_{33} - h_{13}^2}$$

$$C_{13} = -\frac{h_{13}}{h_{11}h_{33} - h_{13}^2}$$

$$C_{22} = \frac{h_{44}}{h_{22}h_{44} - h_{24}^2}$$

$$C_{24} = -\frac{h_{24}}{h_{22}h_{44} - h_{24}^2}$$

$$C_{33} = \frac{h_{11}}{h_{11}h_{33} - h_{13}^2}$$

$$C_{44} = \frac{h_{22}}{h_{22}h_{44} - h_{24}^2}$$

where

$$h_{11} = \frac{N}{\sigma^2}$$

$$h_{13} = h_{22} = \frac{1}{\sigma^2} \sum_{i=1}^{N} x_i'^2 = N\frac{\sigma_x^2}{\sigma^2}$$

$$h_{24} = h_{33} = \frac{1}{\sigma^2} \sum_{i=1}^{N} x_i'^4 = \frac{3(3N^2 - 7)}{5(N^2 - 1)} N\frac{\sigma_x^4}{\sigma^2}$$

$$h_{44} = \frac{1}{\sigma^2} \sum_{i=1}^{N} x_i'^6 = \frac{9(3N^2 - 18N^2 + 31)}{7(N^2 - 1)^2} N \frac{\sigma_x^6}{\sigma^2}$$

and the last term has been evaluated with the use of the formula $\sum_{k=1}^{n} k^6 = (1/42)n(n+1)(2n+1)(3n^4 + 6n^3 - 3n + 1)$, taken from [4, Eq. (0.121.6)].

Substituting these results into Eq. (A-5), we obtain

$$\eta^2 = A_0^{(3)} + A_2^{(3)}\xi^2 + A_4^{(3)}\xi^4 + A_6^{(3)}\xi^6$$

where the coefficients are given by

$$A_0^{(3)} = \frac{3(3N^2 - 7)}{4(N^2 - 4)}$$

$$A_2^{(3)} = \frac{5}{12}\frac{9N^4 - 12N^2 - 61}{(N^2 - 4)(N^2 - 9)}$$

$$A_4^{(3)} = -\frac{5}{36}\frac{33N^4 - 23N^2 - 226}{(N^2 - 4)(N^2 - 9)}$$

$$A_6^{(3)} = \frac{175}{108}\frac{(N^2 - 1)^2}{(N^2 - 4)(N^2 - 9)}$$

thus giving us the limiting result $\eta^2 = (9/4) + (15/4)\xi^2 - (55/12)\xi^4 + (175/108)\xi^6$ as $N \to \infty$.

III. Gaussian Fits

A. Constant Data Errors

The general Gaussian fitting function is given by

$$y(x; \mathbf{a}) = a_1 \exp\left[\frac{-(x - a_2)^2}{2a_3^2}\right] = a_1 \exp\left(\frac{-t^2}{2}\right)$$

where $t = (x - a_2)/a_3$. Thus, the required derivatives are

$$d_1(x) = \frac{\partial y}{\partial a_1} = \exp\frac{-t^2}{2}$$

$$d_2(x) = \frac{\partial y}{\partial a_2} = \frac{a_1 t}{a_3}\exp\frac{-t^2}{2} \qquad\qquad \text{(A-6)}$$

$$d_3(x) = \frac{\partial y}{\partial a_3} = \frac{a_1 t^2}{a_3}\exp\frac{-t^2}{2}$$

For $\sigma = $ constant, $h_{jk} = \sum_{i=1}^{N} A_{ij}A_{ik} = (1/\sigma^2)\sum_{i=1}^{N} d_j(x_i)d_k(x_i)$, so the elements of the curvature matrix are given by

$$h_{11} = \frac{1}{\sigma^2}\sum_{i=1}^{N}\exp\left(-t_i^2\right)$$

$$h_{12} = \frac{r}{\sigma^2}\sum_{i=1}^{N}t_i\exp\left(-t_i^2\right)$$

$$h_{13} = \frac{r}{\sigma^2}\sum_{i=1}^{N}t_i^2\exp\left(-t_i^2\right)$$

$$h_{22} = \frac{r^2}{\sigma^2}\sum_{i=1}^{N}t_i^2\exp\left(-t_i^2\right)$$

$$h_{23} = \frac{r^2}{\sigma^2}\sum_{i=1}^{N}t_i^3\exp\left(-t_i^2\right)$$

$$h_{33} = \frac{r^2}{\sigma^2}\sum_{i=1}^{N}t_i^4\exp\left(-t_i^2\right)$$

where $r = a_1/a_3$.

We see from this that unless the data points are located symmetrically relative to the peak of the Gaussian, all six matrix elements will be nonzero, and the resulting variance of the fit will be a function of the amount of offset. Therefore, we consider only the simplest, and in fact most common, case where such a symmetry exists, at least approximately, and furthermore assume that the data points are uniformly distributed, as in the polynomial case. Then $h_{12} = h_{23} = 0$, and the curvature and covariance matrices have the same form dealt with above [see Eq. (A-1)]. Thus, proceeding as in the polynomial case, we have

$$h_{11} = \frac{1}{\sigma^2}(2S_0 + 1)$$

$$h_{13} = \frac{2}{\sigma^2}\delta_t^2 r S_2$$

$$h_{22} = \frac{2}{\sigma^2}\delta_t^2 r^2 S_2 \qquad\qquad\text{(A-7)}$$

$$h_{33} = \frac{2}{\sigma^2}\delta_t^4 r^2 S_4$$

where $S_0 = \sum_{i=1}^{(N-1)/2} \exp(-i^2\delta_t^2)$, $S_2 = \sum_{i=1}^{(N-1)/2} i^2 \exp(-i^2\delta_t^2)$, $S_4 = \sum_{i=1}^{(N-1)/2} i^4 \exp(-i^2\delta_t^2)$, and $\delta_t^2 = (12/[N^2-1])(\sigma_x^2/a_3^2) = (12/[N^2-1])\sigma_t^2$ is the spacing between points, normalized with respect to the standard deviation of the Gaussian function, a_3, and $\sigma_t = \sigma_x/a_3$.

The above sums, which depend on δ_t, cannot be expressed in closed form in the general case, but in the limit as $N \to \infty$ and $\delta_t \to 0$, they may be expressed in terms of integrals. Thus, if we further restrict ourselves to the most useful case where the data points extend out to at least two standard deviations of the Gaussian, i.e., $L \geq 4a_3$, where L is the full width of the interval of the data points along the x-axis, then we find

$$\lim_{N,\sigma_x \to \infty} \delta_t S_0 = \frac{\sqrt{\pi}}{2}$$

$$\lim_{N,\sigma_x \to \infty} \delta_t^3 S_2 = \frac{\sqrt{\pi}}{4} \qquad\qquad\text{(A-8)}$$

$$\lim_{N,\sigma_x \to \infty} \delta_t^5 S_4 = 3\frac{\sqrt{\pi}}{8}$$

From the general form of the covariance matrix for symmetric data points with $M = 3$ [see Eq. (A-3), for example] and the derivatives given by Eq. (A-6), the variance of the fit becomes

$$\sigma_y^2(x) = \mathbf{d}^T(x)\mathbf{C}\mathbf{d}(x) = e^{-t^2} \begin{bmatrix} 1 & rt & rt^2 \end{bmatrix} \begin{bmatrix} C_{11} & 0 & C_{13} \\ 0 & C_{22} & 0 \\ C_{13} & 0 & C_{33} \end{bmatrix} \begin{bmatrix} 1 \\ rt \\ rt^2 \end{bmatrix}$$

$$= e^{-t^2}\left[C_{11} + (C_{22}r + 2C_{13})rt^2 + C_{33}r^2t^4 \right]$$

Substituting Eqs. (A-7) and (A-8) into Eq. (A-4), we finally obtain

$$\eta^2 = \sqrt{\frac{3}{\pi}}\sigma_t e^{-(\xi\sigma_t)^2}\left(3 + 4\sigma_t^4\xi^4\right)$$

where $(L_\infty/a_3) \equiv \lim_{N \to \infty}(L/a_3) = \lim_{N \to \infty}(L/\sigma_x)\sigma_t = 2\sqrt{3}\sigma_t$ has been used.

B. Data Errors Proportional to *y* (*x*;a)

When the dependent variable extends over a very large range, as in the case of the Gaussian function, it often happens that the measurement error scales with the measurement itself, at least in a piecewise sense, due to the changing of instrument ranges as the variable values change. Thus, letting $\sigma_i = \beta y(x_i; \mathbf{a})$, the curvature matrix elements become

$$h_{jk} = \frac{1}{\beta^2} \sum_{i=1}^{N} \frac{d_j(x_i) d_k(x_i)}{y^2(x_i; \mathbf{a})}$$

and using Eq. (A-6) for the derivatives, we have

$$
\left.
\begin{aligned}
h_{11} &= \frac{N}{(a_1\beta)^2} \\[2ex]
h_{12} &= \frac{r}{(a_1\beta)^2} \sum_{i=1}^{N} t_i \\[2ex]
h_{13} &= \frac{r}{(a_1\beta)^2} \sum_{i=1}^{N} t_i^2 \\[2ex]
h_{22} &= \frac{r^2}{(a_1\beta)^2} \sum_{i=1}^{N} t_i^2 \\[2ex]
h_{23} &= \frac{r^2}{(a_1\beta)^2} \sum_{i=1}^{N} t_i^3 \\[2ex]
h_{33} &= \frac{r^2}{(a_1\beta)^2} \sum_{i=1}^{N} t_i^4
\end{aligned}
\right\}
\tag{A-9}
$$

These are of the same form as was found for the case of the quadratic fit, so the derivation follows through in much the same way, giving the result

$$\sigma_y^2(x) = \frac{(a_1\beta)^2}{N} e^{-t^2} \left[A_0^{(2)}(N) + A_2^{(2)}(N) \left(\frac{t}{\sigma_t}\right)^2 + A_4^{(2)}(N) \left(\frac{t}{\sigma_t}\right)^4 \right]$$

where we have again assumed symmetric data, and the coefficients that appear are the same ones appearing in the quadratic fit. This may be simplified by noting that $a_1\beta = a_1\sigma_i/y(x_i; \mathbf{a}) = a_1\sigma(x)/y(x; \mathbf{a}) = \sigma(x)e^{t^2/2}$ so that we finally obtain $\eta^2 = A_0^{(2)}(N) + A_2^{(2)}(N)\xi^2 + A_4^{(2)}(N)\xi^4$, where $\eta = \sqrt{N}([\sigma_y(x)]/[\sigma(x)])$ and $t/\sigma_t = (x - \overline{x})/\sigma_x = \xi$, in exact agreement with the expression for the quadratic fit.

IV. Monte Carlo Simulations

As a check on the above derivations, a series of Monte Carlo simulations has been carried out for the various cases treated by creating ensembles of data of the form $y_{ni} = y(x_i; \mathbf{a}) + r_n(x_i)$, where $r_n(x_i)$

is a zero-mean, normally distributed random variable having variance σ_i, fitting the function $y(x; \mathbf{a})$ to each of these to obtain an ensemble of fitted functions $y(x; \mathbf{a}_n)$, and computing the variance $\sigma_y^2(x_m) = (1/N_e) \sum_{n=1}^{N_e} y^2(x_m; \mathbf{a}_n) - [(1/N_e) \sum_{n=1}^{N_e} y(x_m; \mathbf{a}_n)]^2$ for a series of values of x_m, where N_e is the number of ensemble members, so that comparisons can be made with the theoretical results. These are shown in Figs. 5 through 7, and it can be seen that the agreement is excellent.

V. Parameter Errors for Gaussian Fitting Functions

The results obtained above permit the calculation of standard errors for the parameters a_j appearing in the various cases treated, i.e., $\sigma_{a_j}^2 = C_{jj}$. However, since these errors will depend on the location of the y-axis, the results are not of general interest unless the data are located symmetrically relative to the fitting function, as, for example, in the case of the Gaussian function. This function is of sufficient interest to warrant a separate discussion, which we present below for the two cases considered earlier, namely, constant data errors and data errors proportional to the function.

A. Constant Data Errors

When Eq. (A-7) is substituted into Eq. (A-4), the following expressions are obtained for the normalized standard errors:

$$N\frac{\sigma_{a_1}^2}{\sigma^2} = \frac{NS_4}{(2S_0+1)S_4 - 2S_2^2}$$

$$N\frac{(\sigma_{a_2}/a_3)^2}{(\sigma/a_1)^2} = \frac{1}{2}\frac{N}{\delta_t^2 S_2}$$

$$N\frac{(\sigma_{a_3}/a_3)^2}{(\sigma/a_1)^2} = \frac{1}{2}\frac{N(2S_0+1)}{\delta_t^4[(2S_0+1) - 2S_2^2]}$$

where the sums S_0, S_2, and S_4 are given by Eq. (A-7), and $\delta_t^2 = (12/[N^2-1])(\sigma_x^2/a_3^2) = (12/[N^2-1])\sigma_t^2$, as before. These results are valid for any odd value of $N \geq 3$, and in Fig. 8 we show the general behavior as a function of the variable $L/a_3 = (N-1)\delta_t$ for $N =$ 3, 5, 7, 9, and 11. In the limit $N \to \infty$ and $L/a_3 \geq 4$, the above normalized standard errors become

$$N\frac{\sigma_{a_1}^2}{\sigma^2} = \frac{3}{2\sqrt{\pi}}\frac{L}{a_3}$$

$$N\frac{(\sigma_{a_2}/a_3)^2}{(\sigma/a_1)^2} = \frac{2}{\sqrt{\pi}}\frac{L}{a_3}$$

$$N\frac{(\sigma_{a_3}/a_3)^2}{(\sigma/a_1)^2} = \frac{2}{\sqrt{\pi}}\frac{L}{a_3}$$

B. Proportional Data Errors

Substituting Eq. (A-9) into Eq. (A-4), we obtain the following normalized standard errors:

$$N\frac{\sigma_{a_1}^2}{a_1^2} = \beta^2 A_0^{(2)}(N) = \beta^2 \frac{3(3N^2 - 7)}{4(N^2 - 4)} \rightarrow \frac{9}{4}\beta^2$$

as $N \rightarrow \infty$;

$$N\frac{\sigma_{a_2}^2}{a_3^2} = \frac{\beta^2}{\sigma_t^2}$$

independent of N; and

$$N\frac{\sigma_{a_2}^2}{a_3^2} = \frac{\beta^2}{\sigma_t^4} A_4^{(2)}(N) = \frac{\beta^2}{\sigma_t^4} \frac{5(N^2 - 1)}{4(N^2 - 4)} \rightarrow \frac{5}{4}\frac{\beta^2}{\sigma_t^4}$$

as $N \rightarrow \infty$, where β is the error proportionality constant, i.e., $\sigma_i = \beta y(x_i; \mathbf{a})$.