REVIEW

Why, When, and How Biochemists Should Use Least Squares

Michael L. Johnson

Departments of Pharmacology and Internal Medicine, Box 448, University of Virginia Health Sciences Center, Charlottesville, Virginia 22908

One of the most commonly used methods for the analysis of experimental data in the biochemical literature is nonlinear least squares (regression). This group of methods are also commonly misused. The purpose of this article is to review the assumptions inherent in the use of least-squares techniques and how these assumptions govern the ways that least-squares techniques can and should be used. Since these assumptions pertain to the nature of the experimental data to be analyzed they also dictate many aspects of the data collection protocol. The examination of these assumptions includes a discussion of questions like: Why would a biochemist want to use nonlinear least-squares techniques? When is it appropriate for a biochemist to use nonlinear least-squares techniques? What confidence can be assigned to the results of a nonlinear least-squares analysis?

There are relatively few methods available for the analysis of experimental data in the biochemical laboratory. Graphical methods and least-squares (regression) methods are by far the most common. Unfortunately, both classes of analysis methods are commonly misused. The purpose of this review is to explain why, when, and how a biochemist should use least-squares techniques and what confidence can be assigned to the resulting estimated parameters.

One classic class of biochemical experiments is measuring the response of a system to an external perturbation. Temperature jump experiments perturb the chemical equilibrium of a solution by rapidly increasing the temperature of the solution and subsequently monitoring an observable, like absorbance, as a function of time. Here the absorbance is the observable (i.e., the variable) that is dependent on the experiment, and time is the variable that can be independently controlled by the experimental protocol.

Another example of this general class is the ligand-binding titration experiment. The investigator measures the amount of a ligand bound (the dependent variable) by fluorescence, absorbance, or radioactive counting. To do so, the investigator titrates the ligand concentration (the independent variable). Note that the ligand concentrations might be either the total or the free ligand concentration, depending on the experimental protocol.

In these examples, and all others of this class, the investigator has measured a response due to a perturbation of the system. The next step is to obtain the parameters of the system that characterize the chemical processes by “analyzing” the data. In the above examples these parameters, the desired answers, might be the relaxation half-lives or macroscopic binding constants. Alternatively, the desired parameters might be the microscopic forward and reverse reaction rates of the biochemical system.

Analysis of these data requires that the biochemist assume a mathematical relationship between the observed quantities, the dependent variables, and the independent variables. This relationship is the fitting function. In the past, analysis of relaxation experiments, such as temperature jump, assumed that the mathematical relationship was a single exponential decay. Based on this assumption the investigator would commonly perform a logarithmic transformation of the dependent variable and create a graph of, for example, the logarithm of absorbance as a function of time. If the original assumption of a single exponential process is correct then this graph will be a straight line with a slope related to the relaxation rate of the chemical process. A single class of binding sites is a common assumption for ligand binding experiments. This, in turn, implied that the
The desired result of the analysis of any experimental data is to obtain the set of parameters of the biochemical reaction with the maximum likelihood, highest probability, of being correct. This is the most critical lesson of this review! We do not care what the slope of a log plot is; we want the relaxation rate constants with the maximum likelihood of being correct. We do not care what the slope of a Scatchard plot is; we want the ligand binding constants with the highest probability of being correct.

Does a Scatchard plot, or a logarithmic plot, yield parameter values with the maximum likelihood of being correct? Generally they do not (1)! These methods are mathematically correct if the experimental data contain no experimental uncertainties. They fail because they do not correctly consider the experimental uncertainties present in all experimental data! Why then were these graphical methods developed and commonly reported? The evaluation of the parameters with the maximum likelihood of being correct requires a high-speed digital computer to perform the calculations. The development of these graphical methods occurred before high-speed digital computers were commonly available to the biochemical researcher. At that stage graphical methods were the only practical ones for the analysis of the experimental data. Should these methods still be used? Perhaps to aid the investigator in visualizing the data, but not for determining parameter values!

The most common alternative to graphical analysis in use in the biochemical laboratory today is nonlinear least squares (NLLS). To use a NLLS method an investigator must assume a functional form for the mathematical relationship between the dependent and independent variables of the experiments in terms of a series of desired parameters. This functional form is not restricted to a form that can be transformed into a straight line, as with the graphical procedures. NLLS is a process of "fitting" the experimental data to almost any functional form by evaluating an optimal set of parameters for the fitting function.

Does a NLLS method yield parameter values with the highest probability of being correct? Maybe, if the NLLS analysis procedure is correctly formulated and correctly used (1–3).

**WHAT IS NONLINEAR LEAST SQUARES?**

Nonlinear least squares refers to a group of different mathematical algorithms that perform a "best fit" of a fitting function to a set of experimental data. The objective of this best-fit operation is to obtain a set of "optimal" parameters for the fitting function such that the fitting function will correctly describe the original data and average out the experimental uncertainties. NLLS is a special case of a more general class of parameter estimation procedures known as maximum likelihood (ML) techniques. For linear and nonlinear least-squares procedures the definition of best fit is that the weighted sum of the squares of the differences between the dependent variables and the fitting function (WSSR) is a minimum when evaluated at the optimal parameter values and the independent variables. These differences are the deviations and/or the residuals:

\[
WSSR(\alpha) = \sum_{i=1}^{n} \left[ \frac{Y_i - F(X_i, \alpha)}{\sigma_i} \right]^2 = \sum_{i=1}^{n} \frac{r_i^2}{\sigma_i}, \tag{1}
\]

where the weighted sum of the squares of the residuals, WSSR, is a function of the parameters, represented here as the vector \( \alpha \), and the \( n \) data points, \( X_i, Y_i \). The \( \sigma_i \) refers to the statistical weight of the particular data point. This statistical weight is the standard error of the particular \( Y_i \) observation. For an unweighted analysis all the \( \sigma_i \)'s are identical and usually set to 1. The \( r_i \)'s in Eq. [1] are graphically depicted in Fig. 1. It is the weighted sum of the squares of the vertical distances.

---

1 Abbreviations used: NLLS, nonlinear least squares; ML, maximum likelihood; LLS, linear least squares; TCSFC, time-correlated single-photon counting.
that are minimized, not the horizontal or perpendicular distances.

The temperature jump experiment mentioned above is a useful example for defining some terms used in the discussion of NLLS. In the classic Eigen temperature jump apparatus a sample solution is rapidly heated while the absorbance of the solution is recorded as a function of time. Here the absorbance \(A\) is the dependent variable \(Y_i\), time is the independent variable \(X_i\). In the example presented in Fig. 1 there are only a single dependent and a single independent variable. NLLS is equally valid for experiments with multiple dependent and/or independent variables. The analysis of these data requires a specific form for the mathematical relationship that is used to predict the absorbance as a function of time for a set of parameters that are to be determined by the analysis procedure. For the simplest temperature jump experiments this mathematical relationship, the fitting function, is of the form

\[
A = (A_0 - A_\infty)e^{-K \cdot \text{ TIME}} + A_\infty,
\]

where the parameters to be estimated by the NLLS are the chemical relaxation rate \(K\), the initial absorbance (after the temperature jump but before and relaxation has occurred) \(A_0\), and the absorbance after the relaxation process is complete (i.e., infinite time) \(A_\infty\). More complex relaxation processes can be analyzed with a fitting function that is a summation of several exponential terms.

There are many different NLLS algorithms: the Nelder-Mead (4,5), Gauss-Newton (1–3), Marquardt-Levenberg (6), and steepest descent (1,7), among others. The actual mathematical details of these algorithms are discussed elsewhere and are not repeated here (1–7). For some fitting problems a particular one of these algorithms may be preferable whereas other problems may call for a different algorithm (1–3,7). For most parameter estimation problems these algorithms have many common features. All the algorithms will find a set of parameters \(\alpha\) that minimize the weighted sum of the squares of the deviations of the fitting function and the data \(\text{WSSR}(\alpha)\) in Eq. [1]). When correctly used all the algorithms will yield the same optimal parameter values. All the algorithms require the user to provide initial estimates of the parameter values and all work by an iterative process of using the initial estimate of the parameters to provide a better estimate of the parameters. The algorithms then iteratively use the better estimate of the parameters as the initial estimate and return an even better estimate, until the parameter values do not change within some specified limit. The validity, and usefulness, of all these algorithms is based on the same set of assumptions. Does this “least-squares best fit” provide parameter values with the maximum likelihood of being correct? Only sometimes will the parameters estimated by NLLS correspond to the desired ML estimates!

Linear least-squares (LLS) is a special case of NLLS. Technically, for LLS, the second, and higher, derivatives of the fitting function with respect to the parameters are all zero, whereas for NLLS, these derivatives are not zero. An example of a linear fitting function is a simple polynomial equation like \(Y = A + BX\). The practical difference between LLS and NLLS is that if the second, and higher, derivatives are all zero then the Gauss–Newton algorithm will require only a single iteration for any initial “guesses” of the fitting parameter values. This, in turn, means that for LLS, the required initial values of the fitting parameters can all be zero. The polynomial least-squares equations found in almost every textbook on this subject (8) can be derived from the Gauss–Newton NLLS method by assuming that the initial parameter values are zero and performing only a single iteration. Consequently, the restrictions and limitations of NLLS all apply to LLS and NLLS can always be used instead of LLS. Therefore, only NLLS is discussed here.

WHY SHOULD ONE USE A NLLS ANALYSIS PROCEDURE?

There is only one valid reason: When correctly applied, NLLS will yield parameter values with the highest probability of being correct. When NLLS cannot be correctly applied it should not be used!

Some have claimed that least squares is always valid because least-squares methods will always provide a set of parameters that correspond to a minimum in the variance of fit.² Why would we want a minimum variance of fit, i.e., a minimum WSSR? We desire the parameters with the highest probability of being correct. Are the parameter values corresponding to the minimum variance of fit the parameter values with the highest probability of being correct? Not necessarily! The next section discusses the assumptions required for the parameters corresponding to a minimum variance of fit to have the highest probability of being correct. The assumptions outlined are sufficient to ensure that a least-squares procedure will yield parameter values with the highest probability of being correct. For an arbitrary fitting function, these assumptions are also necessary to demonstrate the relationship between maximum likelihood

² Variance of fit is the average of the weighted squares of the differences between the data points and the fitting function, as shown in Fig. 1. The variance of fit is calculated as the WSSR, from Eq. [1], divided by the number of data points. Thus, a minimum variance of fit corresponds to a minimum WSSR, i.e., a least-squares minimum. The variance of fit is a commonly used, and abused, measure of the quality of a fit. It is generally, and sometimes incorrectly, assumed that the lower the variance of fit, the better the fit of the data.
methods and least-squares methods. However, for a few specific fitting functions it can be demonstrated that one or more of these assumptions are not required.

WHEN SHOULD ONE USE A NLLS ANALYSIS PROCEDURE?

Again, there is only one valid reason: Only if NLLS can be correctly applied to the data. The algebraic demonstration that NLLS will yield a set of estimated parameters that have the maximum likelihood of being correct for an arbitrary fitting function requires a series of assumptions about the characteristics of the experimental data. Specifically, it is the characteristics of the experimental uncertainties contained in the experimental data that must be assumed. Therefore, if these assumptions are valid, then NLLS should be used. Conversely, if these assumptions are invalid, then NLLS should generally not be used! The remainder of this section concentrates on these assumptions and their consequences. Several assumptions listed below are interrelated and are corollaries of other assumptions. Most of these assumptions apply to NLLS and to almost every other method of data analysis.

Assumption 1. The demonstration that NLLS is a ML method requires the assumption that the independent variables contain no experimental uncertainty. In practical terms this assumption means that the precision of the independent variables is much better than the precision of the dependent variables. It is this assumption that allows NLLS to minimize a function of the vertical deviations shown in Fig. 1. For the temperature jump experiment this assumption is that the time measurement is significantly more precise than the absorbance measurement. Here the experimental protocol can clearly be designed such that this assumption is reasonable.

Note that a Scatchard analysis generally violates this assumption. For Scatchard plots the experimental uncertainties have been transformed such that they are no longer vertical. Consequently, if an investigator represents ligand binding data as a Scatchard plot then it is usually not valid to apply LLS to calculate the best slope of the plot. ML methods other than NLLS that can be used for the analysis of experimental data with uncertainties in the independent variables are described elsewhere.

Assumption 2. The demonstration that NLLS is a ML method also requires the assumption that the experimental uncertainties of the dependent variable must follow a Gaussian (i.e., a random or bell-shaped) distribution with a mean of zero. This means that if the experiment is performed thousands of times the distribution of values of the individual data points are Gaussian distributions.

This assumption is usually reasonable for the experimental data as collected by the experimenter. In biochemistry, only two types of experimental uncertainty distributions are usually observed: Gaussian and Poisson distributions. Radioactive and photon counting and similar experiments yield Poisson uncertainty distributions. If the number of counts is high these Poisson distributions can be closely approximated by Gaussian distributions. Almost every other source of uncertainty in biochemical work will yield a Gaussian distribution. Sample handling and preparation uncertainties such as pipeting, weighing, and dilution will yield a Gaussian distribution.

The investigator should not perform any nonlinear transformations of the dependent variables, the Y-axis, that will alter the distribution of uncertainties between the collection of the data and the analysis of the data. A nonlinear transformation refers to a transformation of the variable other than a simple addition or multiplication. Logarithms, exponentials, powers, and inverses are examples of nonlinear transformations. In the previously described temperature jump experiment, the original data probably contain a Gaussian distribution of experimental uncertainties in the absorbance and comparatively little uncertainty in the time values. Due to the complexity of the NLLS fitting process an investigator might prefer to create a plot of the logarithm of the absorbance as a function of time and then subsequently evaluate the slope of the resulting straight line by LLS. There are several reasons why this is not a statistically valid procedure, but at this point consider the distribution of uncertainties in the dependent variable. The logarithmic transformation of the dependent variable changes the form of the distribution of experimental uncertainties on the absorbance. The logarithmic transformation of a Gaussian is not a Gaussian. If the experimental uncertainty distribution is not a Gaussian then LLS cannot be used to evaluate the parameters of the straight line. This problem cannot be corrected by "appropriate weighting factors." Consequently, the logarithmic transformation of the data has created a fitting equation of a significantly simpler form but, in the process, has precluded the use of LLS for the analysis.

The commonly used reciprocal plots, such as the Lineweaver-Burk plot, also violate the assumption of a Gaussian distribution of experimental uncertainties. The original enzyme velocities probably follow a Gaussian distribution, but the inverse of the velocities used in the Lineweaver-Burk plot generally does not contain a Gaussian distribution of experimental uncertainties. Consequently, the reciprocal plots generate a fitting equation of a simpler form and create a distribution of uncertainties that precludes the use of least squares as an analysis method.
An investigator should not perform nonlinear transformations of the dependent variables before proceeding with the analysis of the data (1–3,10) if the original data contain Gaussian uncertainties. However, transformations of the dependent variable are valid if the transformations are performed to convert a non-Gaussian distribution of experimental uncertainties into a Gaussian distribution of experimental uncertainties. This is the only statistically valid reason to perform nonlinear transformations of the dependent variables. The reverse hemolytic plaque assay (13) is an example of a type of experiment where the original distribution of uncertainties is a skewed distribution that is approximately an exponential of a Gaussian. For this experimental protocol it is best to perform a logarithmic transformation of the dependent variable, i.e., the plaque size. For the reverse hemolytic plaque assay this nonlinear transformation will transform the distribution of experimental uncertainties such that they are approximately Gaussian.

LLS and NLLS allow transformations of the independent variables, i.e., the X axis (10). This is because NLLS assumes that no experimental uncertainty exists in the independent variables.

It is also possible to convert an experimental protocol that yields experimental uncertainty distributions that are not Gaussian into a protocol that yields a Gaussian distribution by replicate measurements of the experimental data points. The central limit theorem of calculus states that the mean of a group of numbers will have a Gaussian uncertainty distribution even if the individual replicates have uncertainty distributions that are not Gaussian (15). Therefore, the mean of a group of replicate measurements will have a more Gaussian-like distribution than the individual replicates. Consider a standard radioactive labeled hormone binding experiment. Usually the amount of bound hormone is determined by radioactive counting with relative low numbers of counts. Therefore, the distribution of experimental uncertainties in the amount bound should follow a Poisson distribution. These experiments are usually performed as a series of replicate experiments at each hormone concentration and the means used for the analysis of the data. According to the central limit theorem these mean values will tend to have a Gaussian distribution, rather than a Poisson distribution. Therefore, NLLS can be used to estimate parameters from the mean values of an experiment of this type.

This does not mean that hormone binding experiments should be performed as a series of replicates. Given a choice between 10 data points measured in triplicate and 30 individual data points, it is better to measure the 30 individual data points at different hormone concentrations and count the radioactivity of each data point long enough that the Poisson distribution of uncertainties can be approximated as a Gaussian distribution. Some experimenters feel that having the triplicates will allow an obvious bad point to be eliminated. While this is true, having 30 singlet observations of hormone binding would also allow an obvious bad point to be eliminated since the observations must consistently follow a smooth binding isotherm. What is gained by more singlet observations is the ability to evaluate how well the calculated curve actually describes the data, i.e., the ability to evaluate the “goodness of fit” and test the hypothesis that the fitting equation is consistent with the experimental data.

If the experimental protocol cannot be altered, or the data manipulated, to create a Gaussian distribution of experimental uncertainties, then the NLLS method should generally not be used. The reader is referred to the more general ML methods that can be formulated without the assumption of a Gaussian distribution (1–3,10–12).

Corollary 2A. It is assumed that no systematic uncertainties exist within the data. Any type of systematic uncertainty would require either a non-Gaussian distribution of uncertainties or a nonzero mean of the uncertainties. Thus, this assumption is a corollary of Assumption 2, which states that the experimental uncertainties are Gaussian with a mean of zero. However, it is treated separately here because of its consequences. Consider the logarithmic plot of the temperature jump experiment. For this plot, it is the logarithm of the difference between the absorbance and the final absorbance \((A_i - A_\infty)\) that is plotted. Here the value of \(A_\infty\) must be estimated first, the logarithms of the differences calculated, and then the slope determined. Small errors in the determination of \(A_\infty\) will create systematic uncertainties in the values of the logarithms and will be reflected as a systematic error in the evaluation of the slope. Thus systematic errors will appear in the evaluation of the relaxation rate constants (1–3). Table 1 and Figs. 2 and 3 present an example of this problem. Fig. 2 presents a synthetic data set. Table 1 presents the results of three NLLS analyses of this data with different assumed values for \(A_\infty\). Case 4 in Table 1 is an analysis of these data with \(A_\infty\) as an additional estimated vari-

### Table 1

<table>
<thead>
<tr>
<th>Case</th>
<th>(A_\infty) \pm SD</th>
<th>(K) \pm SD</th>
<th>(A_\infty) \pm SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.0 (9.6, 10.4)*</td>
<td>1.00 (0.94, 1.06)*</td>
<td>0.50*</td>
</tr>
<tr>
<td>2</td>
<td>9.7 (9.3, 10.2)*</td>
<td>1.14 (1.07, 1.22)*</td>
<td>1.00*</td>
</tr>
<tr>
<td>3</td>
<td>10.3 (9.9, 10.7)*</td>
<td>0.89 (0.84, 0.95)*</td>
<td>0.00*</td>
</tr>
<tr>
<td>4</td>
<td>9.9 (9.4, 10.3)*</td>
<td>1.08 (0.90, 1.25)*</td>
<td>0.78 (0.10, 1.42)*</td>
</tr>
</tbody>
</table>

* These are the \(\pm 1\) SD joint confidence intervals for these parameters. See text for details.

* These values were assumed for the analysis of this case.
FIG. 2. A synthetic data set with pseudorandom experimental uncertainty added. These data were generated with $A_0 = 10.5$, $A_\infty = 0.5$, and a decay rate $K = 1.0$. Definitions are the same as those for Eq. [2]. Pseudorandom noise was added with a standard deviation of 0.26. The solid line corresponds to Case 1 in Table 1. This figure is from Johnson and Frasier (1) with permission.

Figure 3 presents the corresponding logarithmic plots with two different assumed values for $A_\infty$. There is no method by which the least-squares process can detect systematic uncertainties, or be modified to consider systematic uncertainties, like those shown in Fig. 3. These systematic errors cannot be corrected by appropriate weighting factors. As far as the least-squares parameter estimation procedure is concerned, systematic uncertainties simply do not exist. Systematic uncertainties should be eliminated by changing the data collection protocol.

This general type of problem occurs whenever a parameter is estimated and then assumed to be correct for subsequent analysis steps. The subsequent analysis does not include any possibility of considering the consequences of the uncertainties of the previous steps. The entire analysis should be performed by a single-step multiple-parameter estimation process like NLLS that considers the joint uncertainties of the parameters simultaneously.

Corollary 2B. The validity of the use of NLLS is also dependent on the assumption that the fitting function, e.g., Eq. [2], is the correct mathematical description of the nonrandom processes contained within the data. Stated another way, NLLS assumes that the dependent variables of the data can be described as the sum of the random (Gaussian) experimental noise and the fitting function evaluated at the corresponding independent variables and the optimal parameter values. The second assumption and both of its corollaries are simply different statements of the requirement that the residuals be a measure of the random uncertainties of the data.

A self-evident, but commonly overlooked, consequence of this assumption is that an incorrect fitting equation will result in the estimated parameters having no physical meaning. For example, consider the binding of oxygen to human hemoglobin. Human hemoglobin exists in solution as an $\alpha\beta$ dimer that self-associates to form a $\alpha_2\beta_2$ tetramer (16,17). The dimer binds two oxygens and the tetramer binds four oxygens. Until recently the most common fitting equation for the analysis of oxygen binding to human hemoglobin contained only four parameters, the four Adair binding constants of the tetrameric hemoglobin (16,17). The assumption was that the hemoglobin concentration was high enough to preclude the formation of the dimeric species. Thus, the consequences of the dissociation of the tetramer into dimers and the binding of oxygen by the dimers were neglected. Recently, it was shown that a fitting equation that neglects the dimeric species will yield incorrect answers for the Adair binding constants even at the hemoglobin concentrations found within red blood cells (16,17). For other examples of this type the reader is referred to Johnson and Frasier (1). The lesson is that parameters estimated by any curve-fitting procedure are dependent on the assumed form of the fitting equation. The presentation, or publication, of parameters determined by these methods should always include a statement about the assumed molecular mechanism.

One further comment about the nature of the fitting function is in order. All least squares algorithms require that the fitting function be continuous at each of the data points. Furthermore, most least-squares algorithms also require that the first derivatives of the fitting function with respect to all of the parameters being estimated be continuous at each of the data points. The fitting functions can have discontinuities as long as they do not coincide with the experimental data.

Assumption 3. In order for a NLLS procedure to produce parameter values with the highest probability

FIG. 3. This figure presents logarithmic plots of the data shown in Fig. 2. The lower set of data points was generated by assuming that $A_\infty$ is equal to 1.0 (Case 2 in Table 1) and the upper set of points was generated by assuming that $A_\infty$ is equal to 0.0 (Case 3 in Table 1). Note that the resulting slopes are distinctly different. This figure is from Johnson and Frasier (1) with permission.
of being correct the individual data points must be independent observations. This is a standard statistical assumption for almost every type of statistical and mathematical analysis. A common source of data points with uncertainties that are not independent occurs wherever data are collected by an automated data acquisition system with a time response slow compared to the time between the data points. When data are collected in this manner the instrument response will cause an apparent serial correlation between successive data points; i.e., if one data point contains a random uncertainty then the subsequent data points will have a tendency to have an uncertainty in the same direction.

The best method to approach the analysis of data that has been perturbed by the response characteristics of an instrument is to include the instrument response function in the analysis procedure. An approach of this type is used for the analysis of time-correlated single-photon counting (TCSPC) fluorescence lifetime measurements. If the fluorescent molecules are instantaneously excited then the fluorescence intensity as a function of time is the intensity decay law, $I(t)$. The $I(t)$ can have many different forms depending on the particular mechanism for fluorescence emission; the exact form is not important for this discussion. The problem is that the fluorescent molecules cannot be instantaneously excited by the instrument. The flash lamp, or laser, pulse has a finite width. The data collected by the instrument contain information about the intensity decay, $I(t)$, and about the time dependence of the intensity of the excitation lamp, $L(t)$. The correct fitting function for TCSPC data is a combination of the lamp intensity function and the intensity decay function. The correct combination of these two functions is the convolution integral of the two functions, $L(t) \ast I(t)$. By fitting to the convolution integral, the systematic uncertainties, introduced because of the excitation lamp’s finite pulse width, are included in the fitting function. The convolution integral correctly describes the experimental data and, thus, allows the use NLLS for the analysis of the data.

**Assumption 4.** There must be sufficient data points to provide a good random sampling of the random experimental uncertainties. This assumption is not actually required to demonstrate that least squares provides a maximum likelihood estimate of the parameter values. This assumption is, however, required for the assignment of realistic measures of the accuracy/precision of the estimated parameters. The theoretical minimum number of data points is equal to the number of parameters being simultaneously estimated. Since each data point contains experimental uncertainty, significantly more data points than the minimum are required. The system is “overdetermined” when more than the minimum number of data points are used. Unfortunately, there is no method to access the actual number of data points required to provide a good random sampling of the experimental uncertainties.

**Corollary 4A.** Experimental data should never be smoothed. Data smoothing is commonly, and incorrectly, used to improve the quality of experimental data. However, once the experimental data have been smoothed it is impossible to obtain a good random sampling of the random experimental uncertainties of the original data. Furthermore, all smoothing algorithms will perturb the information within the data as well as remove noise from the data. Improving the quality of the experimental data is equivalent to increasing the information content of the data. However, the process of smoothing experimental data does not add information to the experimental data. The smoothed experimental data actually contain less information than the original data because of the perturbations caused by the smoothing process. The only method to increase the information content of an experimental data set is to collect more experimental data.

When the experimental uncertainties contained within a data set are consistent with the above assumptions it is appropriate to use a NLLS procedure for the analysis of the data. Conversely, NLLS should probably not be used if these assumptions are not satisfied. A ML method probably can be formulated for experimental data with almost any distribution of uncertainties in the dependent and independent variables (9).

**WHAT CONFIDENCE CAN BE ASSIGNED TO THE RESULTS OF A NLLS ANALYSIS?**

There are two steps required for the analysis of experimental data by NLLS or any other method. The first is to find the set of parameters with the maximum likelihood of being correct and the second is to find realistic measures of the accuracy of those parameters. When we determined that the relative mass, $M_r$, of a protein is 90,000, what does this number mean? If the accuracy of the determination is $\pm 80,000$ then we know relatively little. However, if the accuracy is $\pm 1000$ we might be able to use the $M_r$ to increase our understanding of the protein. A functional measure of the accuracy of the determined values is actually more important than the optimal values. If we knew that the $M_r$ was probably between 89,000 and 91,000, would we care if the value with the highest probability of being correct was 90,000 or 90,001? An investigator should always provide a realistic measure of the precision of the determined values when such values are reported.

This section discusses the determination of confidence intervals for parameters determined by NLLS methods. These confidence intervals are a measure of the precision to which a group of parameters can simultaneously be determined from a limited set of data.
Confidence intervals are measures of the precision of the measurement based on a single set of data. If the assumptions required are valid, these confidence intervals will also provide a good measure of the absolute accuracy of the determined parameters.

It should be clear that there is no exact theory for the evaluation of confidence intervals for nonlinear fitting equations. All the methods are extensions of the methods developed for linear least-squares and, therefore, require a linear fitting equation. These methods all assume that the fitting equation can be approximated as a first-order series expansion in the estimated parameters. This assumption is always valid for linear fitting equations. For nonlinear fitting equations this assumption is usually reasonable for small perturbations of the parameter values from the corresponding minimum least-squares values.

There are several approximate methods for the evaluation of the confidence intervals of simultaneously estimated parameters that can be used with NLLS methods. The most commonly used one, the "asymptotic standard errors" (1,2), is both the easiest to calculate and by far the least accurate for most applications. Asymptotic standard errors nearly always provide underestimates of the actual confidence limits of the determined parameters. It is the use of the asymptotic standard errors that is responsible for the perception among many investigators that the confidence intervals reported by NLLS procedures are so inaccurate that they cannot be used for any practical purpose. This perception is correct because almost every commonly available least-squares analysis program either reports no measure of the precision of the determined parameters or reports the values of the asymptotic standard errors. There are many other published methods that provide realistic estimates of the confidence intervals of parameters determined by NLLS methods. The reason that these other methods are rarely used is that they are significantly more complex and require significantly more computer time to evaluate. They also require a considerably more complex computer program.

Most NLLS procedures require, or provide for, the evaluation of the "information matrix." The information matrix is the basis for most methods commonly used for the evaluation of the confidence intervals, or the precision, of determined parameters. This information matrix is also called the Hessian matrix, H. The individual \( j \) \( k \) elements of this matrix are defined as

\[
H_{j,k} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \left[ \frac{\partial F(X_i, \alpha)}{\partial \alpha_j} \cdot \frac{\partial F(X_i, \alpha)}{\partial \alpha_k} \right],
\]

where the summation is over the \( n \) data points. \( F(X_i, \alpha) \) is the fitting function evaluated at a particular independent variable \( X_i \) and optimal estimate of the fitting parameters \( \alpha \). The \( j \) and \( k \) subscripts refer to particular fitting parameters, i.e., particular elements of the \( \alpha \) vector and the \( H \) matrix.

The variance–covariance matrix is evaluated by multiplying the inverse of the Hessian matrix by the variance of the random uncertainties of the experimental data. Usually the variance of the residuals (variance of fit) is assumed to be a reliable estimate of the true variance of random experimental uncertainties of the data. This is Assumption 4 from the previous section. This is true only in the asymptote as the number of data points approaches infinity. In this context infinity is simply enough data points to provide a good random sampling of the experimental uncertainties of the data. The inverse of \( H \) times the variance of fit is the asymptotic variance–covariance matrix, AVC. The diagonal elements of the AVC matrix are the squares of the asymptotic standard errors of the corresponding simultaneously estimated parameters. The off-diagonal elements of AVC are the covariances of the parameters. Most NLLS procedures report the asymptotic standard errors of the parameters as the measure of the confidence, the precision, of the estimated parameters. Note that three assumptions were made to obtain these confidence estimates: We assumed that the fitting equation was linear, that the number of data points is near infinity, and that the covariance terms can be neglected. The first is probably a reasonable assumption. The second may be a reasonable assumption. The third assumption is usually unreasonable. When parameters are simultaneously determined by NLLS they will usually have a significant covariance. The consequence of neglecting the covariances is that the confidence intervals will significantly underestimate the actual range of the confidence intervals for the simultaneously determined parameters. Consequently, the resulting measures of the precision of the determined parameters that neglect the covariance are not reasonable. The reader should question the validity of computer programs that report the asymptotic standard errors of determined parameters without also reporting the corresponding covariances.

The assumption that the covariance terms can be neglected is equivalent to assuming that the fitting parameters are all orthogonal. Parameters are mathematically orthogonal if the corresponding off-diagonal elements of the inverse of the Hessian matrix are zero, i.e., if the cross-correlation and covariance of the parameters are zero. Operationally, if the parameters are orthogonal then the evaluation of the parameters does not depend on the values of the other parameters. This means that the values of the parameters can be evaluated separately and a simultaneous NLLS procedure is not required. Note that the orthogonality of fitting parameters is dependent on both the actual form of the fitting equation and the individual data points being fit. For
example, a Fourier series is an orthogonal equation, but the Fourier coefficients are orthogonal only if there are $2m + 1$ equally spaced data points per primary period of the sine and cosine functions.

A Fourier analysis is one of the few cases that a biochemist is likely to encounter in which these assumptions are valid. A Fourier analysis is equivalent to a least-squares fit of the experimental data to the function

$$Y_i = \sum_{l=0}^{m} \left[ a_l \cos \left( \frac{2\pi X_l}{\text{period}} \right) + b_l \sin \left( \frac{2\pi X_l}{\text{period}} \right) \right], \quad [4]$$

where the parameters to be estimated are the coefficients of the sine and cosine terms, $a_l$ and $b_l$, and $b_0$ is fixed at zero. There are $2m + 1$ parameters estimated in a Fourier analysis. Since all the second derivatives of $Y$ with respect to the parameters to be estimated are zero, this is a linear fitting problem. If the data points are equally spaced in the independent variable $X$, if the number of data points is equal to the number of estimated parameters $n = 2m + 1$, and if the period is equal to $(n + 1)/n$ times the difference between the largest and smallest independent variable, then the off-diagonal elements of the inverse Hessian matrix $H$ are zero and the parameters are orthogonal. If these assumptions about the spacing of the data points are not met, then the coefficients from a Fourier analysis will not be orthogonal even though the basis functions are orthogonal! If a classic Fourier series analysis is performed without these assumptions being met then it will yield incorrect estimates of the Fourier coefficients. For almost every other fitting equation (including a simple straight line like $Y = A + BX$) the parameters will not be orthogonal. If the fitting equation is not orthogonal in the parameters the covariance terms will not be neglected for the estimation of the uncertainties of the estimated parameters. If the covariances cannot be neglected then the asymptotic standard errors do not provide the investigator with reasonable estimates of the uncertainties of the fitted parameters. The consequence of neglecting the covariances is that the confidence intervals for the determined parameters will be significantly underestimated. This underestimate can commonly be a factor of two or three. Thus the investigator might significantly underestimate the standard errors of the determined parameters and reach incorrect conclusions about the significance of the results. Asymptotic standard errors should not be used as an estimate of the confidence of parameters determined by either LLS or NLLS.

What, then, can be used to evaluate confidence intervals of simultaneously determined parameters? Monte-Carlo methods are the very best, but they require a tremendous amount of computer time (18) and, therefore, are usually impractical and are not discussed here. One could create a large grid of all combinations of the fitting parameters and then search for where the increase of the variance is statistically significant. These grid search methods will usually provide good estimates of the regions of the parameter grid where the parameters are not significantly different. These regions are the joint confidence regions for the parameters and can usually be approximated as a multidimensional ellipse. Figure 4 presents a typical elliptical-shaped confidence region obtained by a grid search. Because grid search methods also require a large amount of computer time, they are generally not used.

The question of what increase in the variance (or WSSR) is statistically significant provides the groundwork for the following discussion. The standard definition of statistical significance in this context is

$$\frac{\text{WSSR}(a')}{\text{WSSR}(a)} = 1 + \frac{p}{n - p} F(p, n - p, 1 - \text{PROB}), \quad [5]$$

where $p$ is the number of parameters being simultaneously estimated, $n$ is the number of data points, and $F$
is the upper 1-PROB quantile for Fisher's $F$ distribution with $p$ and $n - p$ degrees of freedom (12). This equation can be used to compare the probability, PROB, that any set of parameters $a'$ is statistically different from the optimal parameters $a$. The validity of this equation is based on two assumptions. It assumes that the observations are independent and, therefore, that the number of degrees of freedom for the problem is $n - p$. It also assumes a linear fitting equation. The derivation of the right-hand side of Eq. [5] requires that the WSSR at any point $a'$ be the sum of the WSSR at the point $a$ and a WSSR due to the change of the parameters from $a$ to $a'$. This separation of the WSSR into component parts is valid only for linear equations. However, the assumption that the fitting equation is approximately linear for small changes in $a$ is usually reasonable.

The functional form of the elliptical-shaped joint confidence region is available for linear equations (1-3,11,12,19). The joint confidence region for a particular PROB is the ellipsoid $a'$, 

$$(a' - a)^T H^T H (a' - a) \leq ps^2 F(p, n - p, 1 - PROB), \ [6]$$

where

$$s^2 = \frac{WSSR(a)}{n - p} \ [7]$$

and the other variables are as previously defined. The derivation of this equation makes the assumption that the parameters are not orthogonal. This equation models the variance as a quadratic-shaped space near the point of minimum variance where the parameters have the maximum likelihood of being correct. Therefore, the joint confidence intervals derived from this equation only make the assumption that the fitting equation is linear. This assumption is usually reasonable for small perturbations of the estimated parameters. The use of Eqs. [6] and [7] for the evaluation of the joint confidence intervals provides a significantly better estimate of the precision of the determined parameters than the asymptotic standard errors.

Equations [6] and [7] predict the quadratic shape of the variance space from the Hessian matrix evaluated at $a$. This is possible because of the assumption of a linear fitting equation. My preference is to use Eq. [5] for an actual search for all parameters $a'$ corresponding to any desired probability (1-3). This search can be limited to specific directions from the optimal values $a$ to save computer time. If $p$ parameters are being estimated the search can be limited to $4p$ direction. First, each $a_i$ is searched, in both directions, while holding the remaining $a_j$, $i \neq j$, at their optimal values. Second, Eqs. [6] and [7] are used to evaluate the directions of the axes of the multidimensional ellipse of the joint confidence intervals. These ellipse axes are also searched in both directions for values of $a'$ that are different at the same probability levels. The evaluation of these directions simply involves the rotation of the coordinate system such that the off-diagonal elements of the inverse of the Hessian matrix in the new coordinate system are all zero (1,2). In this new coordinate system the new parameters are orthogonal. The joint confidence regions are the extreme values of statistically acceptable parameters found by the search. The search for statistically significant sets of parameters $a'$ eliminates some, but not all, of the consequences of the assumption that the fitting equation is linear. Therefore, this search procedure will provide joint confidence intervals that are more precise than the joint confidence intervals predicted by Eqs. [6] and [7].

It is interesting that the joint confidence intervals for nonlinear problems are not symmetrical. Suppose that we have determined a free energy change for some biochemical process. Further, suppose that we have evaluated the joint confidence region for this free energy change and that it is symmetrical. We can then express the value of the free energy change as some value $\pm$ a value of the uncertainty. If we also want to express the value of the corresponding equilibrium constant we can perform the appropriate nonlinear transformation. However, when we attempt to transform the joint confidence interval of the free energy change into a joint confidence interval for the equilibrium constant we find that the interval is no longer symmetrical and cannot be expressed as $\pm$ a single value. A careful examination of Fig. 4 and Table 1 shows that the elliptically shaped confidence region is not quite symmetrical and/or not centered at the optimal values $a$. Therefore, the reader should question the validity of any NLLS computer program that provides a symmetrical estimate of the confidence intervals of the determined parameters.

CONCLUSIONS

Our choice of methods for the analysis of experimental data is extremely limited. The methods that are available always make assumptions about the nature of the experimental data being analyzed. An investigator needs to be aware of the requirements placed upon the data by these assumptions before collecting the data. It is while the experiment is being designed that the data collection protocol can most readily be altered to be compatible with the available data analysis methods.

When publishing results, a realistic measure of the precision of the determined values should accompany the published values. These are essential for the reader to evaluate the significance of the values reported. Asymptotic standard errors should not be used as an estimate of the confidence of parameters simultaneously determined by either LLS or NLLS. Joint con-
fidence intervals are preferred since they are more accurate than asymptotic standard errors.

Some investigators consider the results of a computer analysis as gospel. But computers are not oracles and computer programmers sometimes make inappropriate assumptions. Programmers commonly use approximations to speed either the programming or the time of execution of programs, and they do make mistakes. Some computer programs are correct for one application but when used for different applications, the methods no longer apply. It is necessary to be aware of the assumptions made by the programmer about the nature of the experimental data being analyzed and one must be aware of the basic assumptions of the method of analysis. The investigator must always question the applicability of any method of analysis for each particular problem. After results are obtained from a computer the next question should be “does this result have any physical meaning?” Do not assume that the values are correct because they come from a computer analysis of the data!

Whenever possible investigators should devise methods to “test” their analysis programs. These tests might be with real sets of data that have known answers, for example, measuring the fluorescence lifetime of a compound with a known lifetime. These tests also might involve simulated experiments with realistic amounts of pseudorandom experimental uncertainties added (1–3,18). The need to include realistic experimental uncertainties in simulated data cannot be overemphasized. Many analysis methods work well for test cases without experimental noise and fail with even small amounts of experimental noise present.

This review has attempted to present the basic ideas and assumptions of linear and nonlinear least-squares analysis methods. It does not include the rigorous mathematical descriptions of the methods nor does it include a discussion of topics like the propagation of errors based on joint confidence intervals (1–3), the analysis of the randomness of residuals (3,20), goodness-of-fit criteria (3,20), global analysis (1–3), weighting functions (21), and the advantages of alternate sets of fitting parameters (1–3). For the actual methods the reader is referred to other articles (1–6,9,14,18,20). More complete general discussions of these topics are available for the beginner (1–3,5,7,9,10,14,18,20,21) and for the mathematician (11,12).

The author’s NLLS software is available upon written request.

ACKNOWLEDGMENTS

This work was supported, in part, by the University of Virginia Diabetes Endocrinology Research Center Grant USPHS DK-38942, the University of Virginia National Science Foundation Science and Technology Center for Biological Timing, and National Institutes of Health Grants GM-28928 and GM-35154. I acknowledge and thank Dr. Ludwig Brand, Dr. D. Wayne Bolen, and Dr. Dima Toptygin for comments on this review.

REFERENCES