# Parameter correlations while curve fitting

**Article** *in* Methods in Enzymology · February 2000

DOI: 10.1016/S0076-6879(00)21207-X · Source: PubMed

CITATIONS
41

READS
2,095

**1 author:**

Michael L Johnson
University of Virginia
**418** PUBLICATIONS **20,640** CITATIONS

SEE PROFILE

tions, then clearly least squares will not provide the optimal parameter values. However, it cannot be demonstrated that the robust method will provide the optimal (i.e., maximum likelihood) parameter values either.

In addition, a large number of techniques have been developed that are based on the concepts of least square parameter estimations. These include goodness-of-fit[8] tests such as autocorrelation, runs test, Kolmogorov–Smirnov test, and Durbin–Watson test. It also includes methods for the evaluation of the precision of the estimated parameters[9,10] (i.e., the parameter confidence intervals). In general, these techniques apply to least squares but not to robust parameter estimation procedures.

So why use robust parameter estimation procedures? Because in many cases they *appear* to provide an analysis of the data that is less sensitive to the presence of outliers. This makes it easier to identify those points that may be outliers.

## Acknowledgments

The author acknowledges the support of the National Science Foundation Science and Technology Center for Biological Timing at the University of Virginia (NSF DIR-8920162), the General Clinical Research Center at the University of Virginia (NIH RR-00847), and the University of Maryland at Baltimore Center for Fluorescence Spectroscopy (NIH RR-08119).

[8] M. Straume and M. L. Johnson, *Methods Enzymol.* **210,** 87 (1992).
[9] D. M. Bates and D. G. Watts, "Nonlinear Regression Analysis and Its Applications." John Wiley and Sons, New York, 1988.
[10] D. G. Watts, *Methods Enzymol.* **240,** 23 (1994).

# [24] Parameter Correlations while Curve Fitting

*By* MICHAEL L. JOHNSON

## Introduction

While curve fitting experimental data, i.e., simultaneous multiple parameter estimation, an investigator will generally find that the answers are cross-correlated. This cross-correlation is evident when an increase in the variance-of-fit caused by a variation in any parameters can be partially compensated for by a variation in one, or more, of the other parameters. This is generally not a problem if the degree of this compensation, i.e., cross-correlation, is low. However, if the correlation is nearly complete, then



0076-6879/00 $30.00

the parameter estimation procedure (e.g., least squares[1-5] and maximum likelihood[1-5]) will be unable to identify unique values for the parameters being estimated and their associated standard errors. The purpose of this chapter is to review the origin of this parameter cross-correlation and discuss methods to minimize or eliminate it.

When the parameters of a simultaneous least squares analysis are correlated then the estimated uncertainties of these parameters are also correlated. Thus, if the goal is to use the parameter values and their uncertainties to test a hypothesis, it is important that both be well determined. For example, how do we test if a set of data is a straight line? One method is to fit the data to a quadratic polynomial and then ask if the quadratic term, i.e., the curvature, is zero. That test requires accurate values for the quadratic parameter and its associated standard error. However, since the intercept, the linear, and the quadratic terms (and standard errors) are probably all correlated, it is impossible to simply compare the quadratic term with zero. Uncertainties in the evaluation of the intercept and linear terms will be partially compensated by a variation of the quadratic term, but this will not be reflected in an increased uncertainty of the quadratic term. One solution would be to perform the analysis such that the parameters and their associated standard errors are not correlated.[6]

It is important to note that if a parameter estimation procedure reports that two parameters are correlated, it is generally not an indication that there is something correlated about the physiologic or biochemical mechanism underlying the data. It may be simply a consequence of performing the parameter estimation on a limited set of experimental observations. For example, consider a stopped-flow experiment (or a temperature-jump experiment) where the equilibrium position of a chemical reaction is rapidly altered and the time course of the relaxation to the new equilibrium position is monitored. These relaxations can generally be described as a sum of exponential decays. Consider the expected results if the sample consists of a mixture of two independent noninteracting chemical systems each of which is expected to exhibit a single exponential decay as it approaches its new equilibrium position. The mixture will exhibit decay characteristics that are simply the sum of the individual exponential decays. There is nothing correlated about the chemistry, but when the data are analyzed

[1] M. L. Johnson and S. G. Fraiser, *Methods Enzymol.* **117,** 301 (1985).
[2] M. L. Johnson and L. M. Faunt, *Methods Enzymol.* **210,** 1 (1992).
[3] M. L. Johnson, *Methods Enzymol.* **240,** 23 (1994).
[4] D. M. Bates and D. G. Watts, "Nonlinear Regression Analysis and Its Applications." John Wiley and Sons, New York, 1988.
[5] D. G. Watts, *Methods Enzymol.* **240,** 23 (1994).
[6] F. S. Acton, "Analysis of Straight-Line Data," p. 193. Chapman and Hall, New York, 1959.

TABLE I
DATA USED IN FIRST EXAMPLE

| Time | Amplitude |
|---|---|
| 0.01 | 5.77681 |
| 0.07 | 4.74431 |
| 0.13 | 3.91186 |
| 0.19 | 3.22843 |
| 0.25 | 2.68778 |
| 0.31 | 2.22903 |
| 0.37 | 1.86932 |
| 0.43 | 1.58560 |
| 0.49 | 1.32141 |
| 0.55 | 1.11302 |
| 0.61 | 0.960843 |
| 0.67 | 0.788441 |
| 0.73 | 0.687573 |
| 0.79 | 0.584176 |
| 0.85 | 0.495013 |
| 0.91 | 0.435635 |
| 0.97 | 0.372533 |

by any parameter estimation procedure the resulting decay rates (or half-lives) will be correlated. Examples are shown in Tables I–III and Figs. 1–3.

Examine the relaxation data shown in Table I and Fig. 1. This is an example of the classic ill-posed analysis problem of the numerical analysis literature. These data were simulated for equal amounts of two noninteracting relaxations with rate constants of two and four. These data also contain pseudo-random experimental uncertainties, i.e., noise, with a stan-

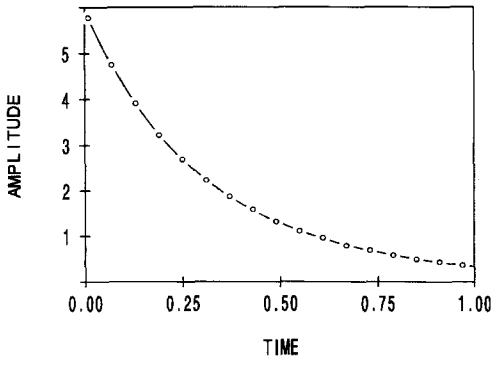


FIG. 1. Two-exponential least squares analysis of the data given in Table I.

TABLE II
PARAMETERS FROM WEIGHTED LEAST SQUARES ANALYSES[a]

| Parameter | Value | Confidence region[b] |
|---|---|---|
| Single-exponential analysis | | |
| $A$ | 5.87 | 5.81–5.94 |
| $K$ | 3.04 | 2.97–3.08 |
| $SSR$ | 398.42 | |
| Double-exponential analysis | | |
| $A_1$ | 3.17 | 2.11–4.41 |
| $K_1$ | 4.30 | 3.81–4.98 |
| $A_2$ | 2.80 | 1.55–3.87 |
| $K_2$ | 2.24 | 1.80–2.49 |
| $SSR$ | 9.052 | |

[a] Of the data shown in Table I and Fig. 1.
[b] A ±2 standard deviation (95.44%) confidence region as evaluated by a bootstrap method.[7]

dard deviation of 0.01 units. The best two-exponential [Eq. (1)] weighted least squares fit of these data is shown as the line in Fig. 1.

$$\text{Amplitude} = A_1 e^{-K_1 time} + A_2 e^{-K_2 time} \tag{1}$$

where the $K$ values are the relaxation decay rates and the $A$ values are the corresponding amplitudes.

The parameter values for one- and two-exponential fits of these data are shown in Table II. Table II also shows the weighted sum of squared residuals, $SSR$, for these analyses.

$$SSR = \sum_i \left[ \frac{Y_i - Curve(X_i)}{\sigma_1} \right]^2 \tag{2}$$

where the summation is over each of the $i$ data points; $Curve(X_i)$ represents the fitting function evaluated at the optimal parameter values and the particular independent variable, $X_i$; $Y_i$ is the $i$th dependent variable; and $\sigma_i$ is the standard error of the mean of $Y_i$. The residuals are simply the standard deviation weighted differences between the data points and the fitted curve. It is clear from the decrease in the $SSR$ that the two-exponential fit of these data is substantially better than the one-exponential fit.

Table II also contains estimates of the 95% confidence regions of the fitted parameters as determined by a bootstrap method.[7] Notice that these

[7] B. Efron and R. J. Tibshirani, "An Introduction to the Bootstrap." Chapman and Hall, New York, 1993.

are neither expected nor observed to be symmetrical. The bootstrap procedure is simple. First, a nonlinear regression is done to evaluate the parameters of the fitting equation with the highest probability of being correct. Second, the optimal model parameters are used to simulate a set of experimental data. This simulated set of data must contain observations at exactly the same independent variables, i.e., the $X$ axis. The simulated data set must also contain a realistic amount of random noise (e.g., experimental uncertainties). The "random noise" is generated by randomly selecting from the observed weighted residuals (the weighted differences between the data and the fitted function). For each simulated curve the residuals are randomly selected such that ~37% are used more than once and ~37% are not used. Third, the selected residuals are shuffled to produce a random order. Fourth, the nonlinear regression procedure is repeated on the simulated data to obtain apparent values for the parameters being estimated. Steps 2 to 4 are repeated many times (e.g., 500 to 1000 times) and the apparent parameter values from each cycle saved. These sets of apparent values of the parameters, from the multiple cycle of steps 2 to 4, are used to create distributions of apparent parameter values. These distributions correspond to the complete probability distributions of the estimated parameters.

The correlation matrix (defined below) for the two-exponential analysis is shown in Table III. The correlation matrix is a symmetrical matrix so only the lower diagonal is shown. This matrix presents the cross-correlation coefficients between each pair of the parameters being estimated. For example, the cross-correlation between $A_1$ and $A_2$ is $-0.999933$, an extremely correlated case. The values of the cross-correlation coefficient vary between $\pm 1$, with zero indicating no cross-correlation between the parameters. A cross-correlation above ~0.98 or below ~$-0.98$ indicated that the parameters are highly correlated. Such high cross-correlation coefficients are also an indication, as will be seen below, that the parameter estimation process is extremely difficult with resulting parameter values and associated standard

TABLE III
CROSS-CORRELATION MATRIX FOR TWO-EXPONENTIAL ANALYSIS[a]

| Parameter | Matrix | | | |
|---|---|---|---|---|
| $A_1$ | 1.00000 | | | |
| $K_1$ | −0.993626 | 1.00000 | | |
| $A_2$ | −0.999933 | 0.994455 | 1.00000 | |
| $K_2$ | −0.997424 | 0.984443 | 0.997134 | 1.00000 |

[a] Shown in Tables I and II and Fig. 1.

errors that are questionable. The diagonal elements of this matrix correspond to the correlation between each parameter and itself and are thus unity by definition.

The cross-correlation of $K_1$ and $K_2$ is shown graphically in Figs. 2 and 3. Figure 2 was generated by specifying a series of values for $K_1$ between 3.8 and 5.3 (the circles). For each specified value of $K_1$, the weighted least squares procedure was repeated to determine the apparent values of $A_1$, $A_2$, and $K_2$. The corresponding values of $K_2$ are plotted with a dashed line connecting them. The procedure was also repeated for specific values of $K_2$, while the apparent values of $A_1$, $A_2$, and $K_1$ were reevaluated by the weighted least squares procedure (shown as a solid line and diamonds). It is obvious from this example that the values of $K_1$ and $K_2$ are correlated. Altering either parameter induces a corresponding change in the apparent value of the other parameter. If $K_1$ and $K_2$ were not correlated, then the variation of one would not induce a compensating variation in the other. Also, if $K_1$ and $K_2$ were not correlated then Fig. 2 would show two perpendicular lines instead of the observed nearly superimposed lines.

Note that the curvature of the lines in Fig. 2 is a measure of the nonlinear nature of the fitting problem near the solution. If this were a linear least squares fit, then these would be straight lines.

The bootstrap process for the evaluation of the parameter confidence

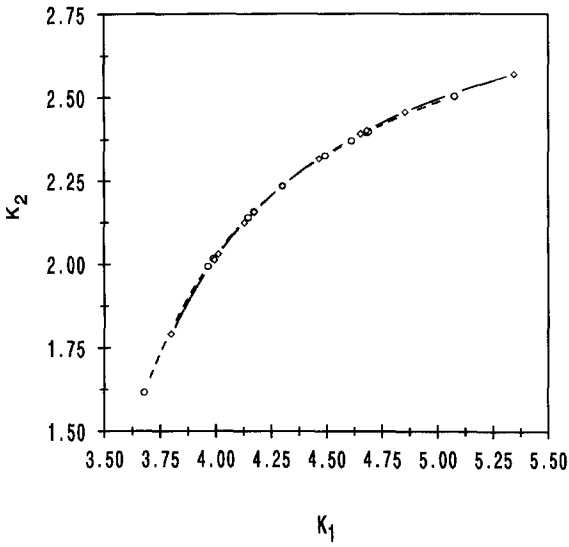


FIG. 2. A plot of the variability of $K_2$ induced by changes in $K_1$ (dashed line and circles). Also shown is a plot of the variability of $K_1$ induced by changes in $K_2$ (solid line and diamonds).

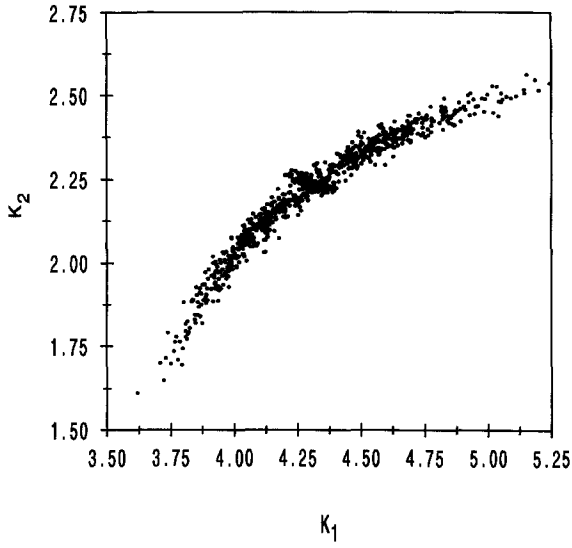


FIG. 3. An analogous plot to Fig. 2 generated from the simulations performed during the evaluation of the confidence intervals of the estimated parameters by a bootstrap procedure.[7]

intervals can be used to create a figure analogous to Fig. 2. Figure 3 presents the 1000 individual pairs of apparent values of $K_1$ and $K_2$ determined by the bootstrap process. Figure 3 also demonstrates the nonlinear highly correlated nature of this parameter estimation procedure in a manner analogous to Fig. 2. If this were an uncorrelated linear fitting problem, Fig. 3 would be circular with Gaussian distributions along all cross sections, as in Fig. 6 (shown later).

The cross-correlation coefficients observed in this example are not a result of the chemistry or physiology of the experiment. Parameter cross-correlation coefficients are a function of the following: the form of the fitting equation; the distribution of the data points, particularly the independent variables; the experimental uncertainties; and in nonlinear cases, the specific parameter values. This chapter reviews the mathematical form and origin of parameter cross-correlations and discusses methods to reduce or eliminate them.

## Definition of Parameter Cross-Correlation Coefficients

The definition of the parameter cross-correlation coefficients is based on the parameter estimation procedure. In this chapter the mathematical form of the parameter cross-correlation coefficients is derived for the

Gauss–Newton weighted nonlinear least squares algorithm.[1–5] These parameter cross-correlation coefficients are not, however, specific to this algorithm. They are not a consequence of the algorithm. They are a complication with all parameter estimation algorithms.

The Gauss–Newton procedure approximates each of the $i$ data points as a first-order series expansion of the fitting function expanded about an estimation, i.e., guess, of the parameter values, $\alpha'$.

$$\frac{1}{\sigma_i} Y_i \approx \frac{1}{\sigma_i} G(\alpha, X_i) = \frac{1}{\sigma_i} G(\alpha', X_i) + \frac{1}{\sigma_i} \sum_j \left[ \frac{\partial G(\alpha', X_i)}{\partial \alpha_j'} (\alpha_j - \alpha_j') \right] + \ldots \tag{3}$$

where $Y_i$ is the dependent variable, $X_i$ is the independent variable, $\sigma_i$ is the standard error of the mean of $Y_i$, $G$ is the fitting function, $\alpha$ is the vector of parameter values to be determined by the weighted least squares procedure, and $\alpha'$ is a guessed (i.e., the current estimates) vector of parameter values. The $j$ subscript refers to a specific fitting parameter and the $i$ subscript refers to a specific data point. Consider a two-exponential fitting function as an example:

$$G(\alpha, X_i) = A_1 e^{-K_1 X_i} + A_2 e^{-K_2 X_i} \tag{4}$$

where $\alpha$ is the vector of fitting parameters $\{A_1, K_1, A_2, K_2\}$.

The only unknown in Eqs. (3) and (4) is the vector $\alpha$. The Gauss–Newton procedure solves Eqs. (3) and (4) for the vector $\alpha$. The easiest presentation of this solution is in matrix notation. Equation (3) can be written in matrix notation as

$$Y^* = J\varepsilon \tag{5}$$

where $Y^*$ is a vector with one element for each of the $i$ data points:

$$Y^* = \begin{pmatrix} \frac{1}{\sigma_1} & [Y_1 - G(\alpha', X_1)] \\ \frac{1}{\sigma_2} & [Y_2 - G(\alpha', X_2)] \\ \frac{1}{\sigma_3} & [Y_3 - G(\alpha', X_3)] \\ & \vdots \end{pmatrix} \tag{6}$$

$\varepsilon$ is a vector with one element for each of the $j$ parameters.

$$\varepsilon = \begin{pmatrix} \alpha_1 - \alpha_1' \\ \alpha_2 - \alpha_2' \\ \alpha_3 - \alpha_3' \\ \vdots \end{pmatrix} \tag{7}$$

$J$ is a matrix of partial derivatives

$$J_{ij} = \frac{1}{\sigma_i} \frac{\partial G(\alpha', X_i)}{\partial \alpha_j'} \tag{8}$$

or

$$J = \begin{pmatrix} \frac{1}{\sigma_1}\frac{\partial G(\alpha', X_1)}{\partial \alpha_1'} & \frac{1}{\sigma_1}\frac{\partial G(\alpha', X_1)}{\partial \alpha_2'} & \frac{1}{\sigma_1}\frac{\partial G(\alpha', X_1)}{\partial \alpha_3'} & \cdots \\ \frac{1}{\sigma_2}\frac{\partial G(\alpha', X_2)}{\partial \alpha_1'} & \frac{1}{\sigma_2}\frac{\partial G(\alpha', X_2)}{\partial \alpha_2'} & \frac{1}{\sigma_2}\frac{\partial G(\alpha', X_2)}{\partial \alpha_3'} & \cdots \\ \frac{1}{\sigma_3}\frac{\partial G(\alpha', X_3)}{\partial \alpha_1'} & \frac{1}{\sigma_3}\frac{\partial G(\alpha', X_3)}{\partial \alpha_2'} & \frac{1}{\sigma_3}\frac{\partial G(\alpha', X_3)}{\partial \alpha_3'} & \cdots \\ \frac{1}{\sigma_4}\frac{\partial G(\alpha', X_4)}{\partial \alpha_1'} & \frac{1}{\sigma_4}\frac{\partial G(\alpha', X_4)}{\partial \alpha_2'} & \frac{1}{\sigma_4}\frac{\partial G(\alpha', X_4)}{\partial \alpha_3'} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{9}$$

The Gauss–Newton least squares procedure solves Eq. (5) for $\varepsilon$. Matrix $J$ is not a square matrix so it cannot simply be inverted. However, by multiplying both sides of Eq. (5) by the transpose of $J$ the resulting $J^TJ$ matrix is a square symmetric positive definite matrix and can usually be inverted:

$$J^TY^* = (J^TJ)\varepsilon \tag{10}$$

Equation (10) can then be solved for $\varepsilon$. The major requirement for being able to invert the $J^TJ$ is that the number of independent data points is greater than or equal to the number of parameters being estimated.

$$\varepsilon = (J^TJ)^{-1}(J^TY^*) \tag{11}$$

Because $\varepsilon$ contains only the parameters being estimated and a known guess of the parameter values, it can easily be solved for the desired parameters.

The algorithm, as stated, is only guaranteed to function properly for linear least squares fitting problems. A linear fitting problem is one where all of the second and higher order partial derivatives of the fitting equation with respect to the parameters being estimated are equal to zero. A quadratic polynomial is an example of a linear fitting equation:

$$G(\alpha, X_i) = a + bX_i + cX_i^2 \tag{12}$$

In this context, linear refers to the series expansion [i.e., Eq. (3)] being linear, not the actual fitting equation being a straight line. The first derivatives of Eq. (12) are

$$\begin{aligned} \frac{\partial G(\alpha, X_i)}{\partial a} &= 1 \\ \frac{\partial G(\alpha, X_i)}{\partial b} &= X_i \\ \frac{\partial G(\alpha, X_i)}{\partial c} &= X_i^2 \end{aligned} \tag{13}$$

These are not a function of the fitting parameters (i.e., constants) so all of the second and higher partial derivatives are zero.

The two-exponential function [Eq. (4)] is an example of a nonlinear fitting function. The first derivatives of Eq. (4) are

$$\begin{aligned} \frac{\partial G(\alpha, X_i)}{\partial A_1} &= e^{-K_1 X_i} \\ \frac{\partial G(\alpha, X_i)}{\partial K_1} &= -X_i A_1 e^{-K_1 X_i} \\ \frac{\partial G(\alpha, X_i)}{\partial A_2} &= e^{-K_2 X_i} \\ \frac{\partial G(\alpha, X_i)}{\partial K_2} &= -X_i A_2 e^{-K_2 X_i} \end{aligned} \tag{14}$$

Clearly all are functions of the parameter values and as a consequence the higher order derivatives are not all equal to zero.

The distinction between linear and nonlinear fitting equations has interesting consequences for the Gauss–Newton procedure. For a linear fitting function the higher order derivatives in Eq. (3) are all zero. Thus, the first-order series expansion in Eq. (3) is exact and Eq. (11) will provide accurate answers for the parameters. The term "linear fitting equation" refers to Eq. (3) being a linear function, not the actual fitting equation. Equation (12) is clearly not a straight line but it is a linear fitting equation.

For nonlinear fitting equations, the series expansion in Eq. (3) will be only approximate. Therefore, the parameter values provided by Eq. (11) will also only be approximate. For such nonlinear problems the parameter values provided by Eq. (11) are used as guesses for Eq. (3) and the algorithm

is repeated. Thus, nonlinear least squares algorithms function by successive approximation. The results are used repeatedly as initial guesses and the process repeated until the resulting answers are the same as the current guesses.

This procedure works well if the contributions of the higher order derivatives in Eq. (3) are small. If the contributions are large, then it is possible that the iteration will diverge instead of converging. Several modifications[1–3] (e.g., Marquardt–Levenberg and damped Gauss–Newton) of the basic Gauss–Newton algorithm have been developed to ensure that the algorithm converges.

## Cross-Correlation Coefficients

The cross-correlation coefficients are defined in terms of the inverse of the $J^TJ$ matrix in Eqs. (10) and (11):

$$cross\text{-}correlation_{jk} = \frac{(J^TJ)^{-1}_{jk}}{\sqrt{(J^TJ)^{-1}_{jj}(J^TJ)^{-1}_{kk}}} \tag{15}$$

where the subscripts refer to a particular $jk$ element of the matrix. A necessary and sufficient condition to have no cross-correlation between the parameters is that all of the off-diagonal elements of the inverse of the $J^TJ$ matrix must be equal to zero. This is equivalent to having the off-diagonal elements of the $J^TJ$ matrix equal to zero. If all of the off-diagonal elements are zero, then the parameters are orthogonal to each other. If the parameters are all orthogonal, then the individual parameters can be evaluated without either a previous knowledge of the other parameters or simultaneously evaluating the other parameters.

The elements of the $J^TJ$ matrix are given by:

$$(J^TJ)_{jk} = \sum_i \left[ \frac{1}{\sigma_i^2} \frac{\partial G(\alpha', X_i)}{\partial \alpha'_j} \frac{\partial G(\alpha', X_i)}{\partial \alpha'_k} \right] \tag{16}$$

Equation (16) is the key to understanding how to reduce, or eliminate, the parameter cross-correlation. The next sections consider examples of using Eq. (16) to control cross-correlations among estimated parameters.

## Polynomials

The simplest application of reducing parameter correlation is with the use of orthogonal polynomials. First, consider the form of the elements of the $J^TJ$ matrix for the quadratic polynomial fitting function [Eq. (12)]. This can be done by inserting Eq. (13) into Eq. (16):

$$(J^T J)_{ab} = \sum_i \left[\frac{X_i}{\sigma_i^2}\right]$$

$$(J^T J)_{ac} = \sum_i \left[\frac{X_i^2}{\sigma_i^2}\right] \tag{17}$$

$$(J^T J)_{bc} = \sum_i \left[\frac{X_i^3}{\sigma_i^2}\right]$$

Clearly, in this form it is unlikely that these terms will be zero for any arbitrary data set.

Consider the simulated data presented in Table IV and Fig. 4. The solid line in Fig. 4 corresponds to the best weighted least squares quadratic polynomial [Eq. (12)] for these data. Figure 5 presents the correlation plot from the bootstrap confidence interval analysis of this example. Figure 5 shows an elliptically shaped region, as is expected for a linear fitting problem. The region is not aligned with the parameter axes, indicating that this was also a correlated analysis. The cross-correlation coefficient for $b$ and $c$ was $-0.95$, again indicating a correlated analysis problem.

However, if Eq. (12) is written in a slightly different form it is comparatively easy to arrange for the cross-correlations to all be zero:

$$G(\alpha, X_i) = a' + b'(X_i - \beta) + c(X_i - \gamma_1)(X_i - \gamma_2) \tag{18}$$

TABLE IV
DATA USED IN SECOND EXAMPLE

| Dependent variable[a] | Independent variable |
|---|---|
| 0.500701 | 0.0 |
| 0.569121 | 0.1 |
| 0.667800 | 0.2 |
| 0.789702 | 0.3 |
| 0.829856 | 0.4 |
| 0.893017 | 0.5 |
| 0.903147 | 0.6 |
| 0.981385 | 0.7 |
| 1.46331 | 0.8 |
| 1.47003 | 0.9 |
| 1.88120 | 1.0 |

[a] The weighting factor for the least squares fit was based on a constant 10% coefficient of variation for these observations.

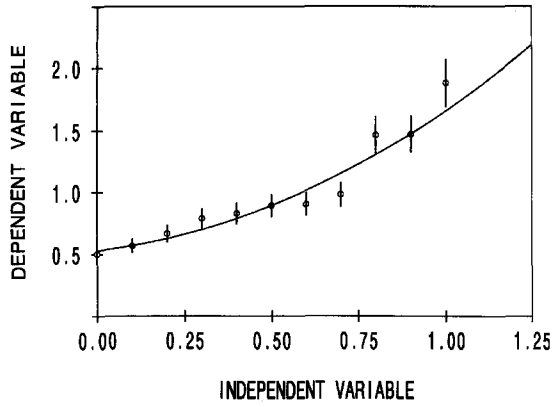


FIG. 4. Quadratic polynomial [Eq. (12)] weighted least squares analysis of the data given in Table IV. The maximum likelihood parameters for this analysis are $a = 0.529$, $b = 0.334$, and $c = 0.794$.

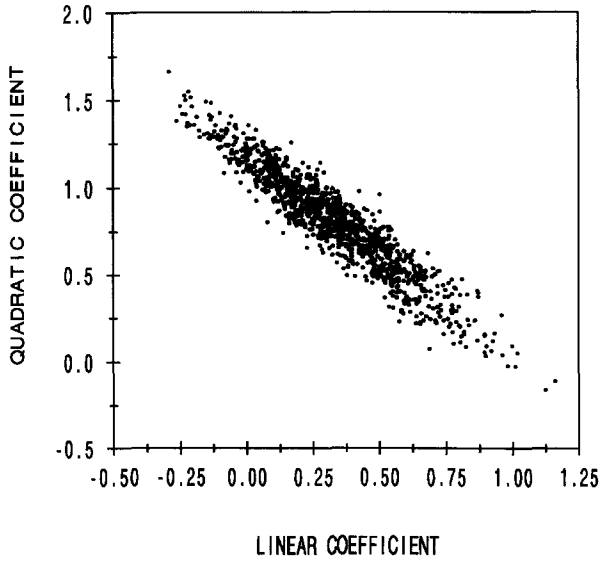


FIG. 5. A correlation plot of the simulations performed during the evaluation of the confidence intervals of the estimated parameters by a bootstrap procedure[7] for the analysis presented in Fig. 4.

where $a'$ is now a different intercept, and $b'$ is the slope at an $X \neq 0$. Terms $\beta$, $\gamma_1$, and $\gamma_2$ are evaluated such that the evaluation of $a'$, $b'$, and $c$ is an orthogonal process, i.e.:

$$(J^T J)_{a'b'} = \sum_i \left[\frac{(X_i - \beta)}{\sigma_i^2}\right] = 0$$

$$(J^T J)_{ac} = \sum_i \left[\frac{(X_i - \gamma_1)(X_i - \gamma_2)}{\sigma_i^2}\right] = 0 \tag{19}$$

$$(J^T J)_{bc} = \sum_i \left[\frac{(X_i - \beta)(X_i - \gamma_1)(X_i - \gamma_2)}{\sigma_i^2}\right] = 0$$

The following summations are over each of the $i$ data points and $\sigma_i$ is the standard error of the mean for the particular data point.

$$\beta = \frac{\sum_i \left[\frac{X_i}{\sigma_i^2}\right]}{\sum_i \left[\frac{1}{\sigma_i^2}\right]} \tag{20}$$

$$k_1 = \frac{\sum_i \left[\frac{(X_i - \beta)X_i^2}{\sigma_i^2}\right]}{\sum_i \left[\frac{(X_i - \beta)X_i}{\sigma_i^2}\right]} \tag{21}$$

$$k_2 = \frac{\sum_i \left[\frac{X_i^2}{\sigma_i^2}\right] - k_1 \sum_i \left[\frac{X_i}{\sigma_i^2}\right]}{\sum_i \left[\frac{1}{\sigma_i^2}\right]} \tag{22}$$

$$\gamma_1 = \frac{k_1 + \sqrt{k_1^2 - 4k_2}}{2} \tag{23}$$

$$\gamma_2 = \frac{k_1 - \sqrt{k_1^2 - 4k_2}}{2} \tag{24}$$

Figure 6 presents the correlation plot for a weighted least squares analysis of the data in Table IV according to an orthogonal quadratic polynomial [Eq. (18)]. From a comparison of Figs. 5 and 6, it is obvious that the simple transformation of the independent variable eliminated the parameter cross-correlations. The off-diagonal elements of $J^T J$, as evaluated by Eq. (15), for this example are all zero.

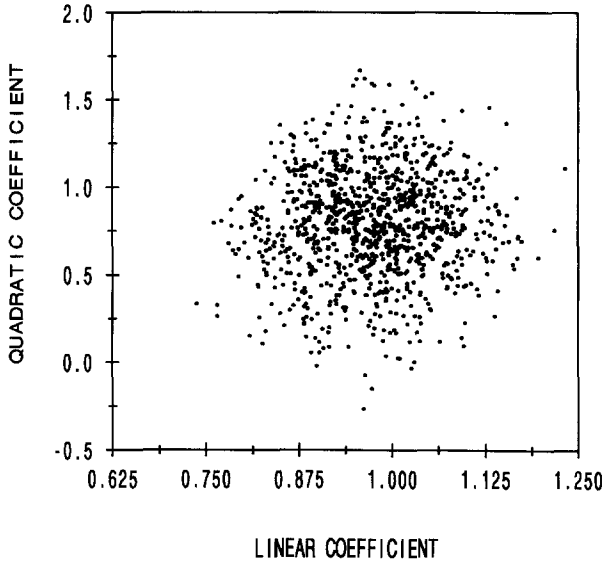


FIG. 6. A correlation plot of the simulations done during the evaluation of the confidence intervals of the estimated parameters by a bootstrap procedure[7] for the data presented in Table IV. Here the fitting function was an orthogonal quadratic polynomial [Eq. (18)]. The parameters of the maximum likelihood orthogonal polynomial are $a' = 0.752$, $b' = 0.962$, $c = 0.794$, $\beta = 0.291$, $\gamma_1 = 0.687$, and $\gamma_2 = 0.104$.

## Fourier Series Analysis

The other classic example of eliminating parameter correlation is Fourier time series analysis. The objective of Fourier series analysis is to approximate a series of data points by the sum of a series of sine and/or cosine waves with known periods. Testing for the presence of circadian rhythms is a commonly used application of Fourier time series analysis.

For a continuous series of data, a Fourier series can be written as

$$G(\alpha, X_i) = f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} c_n \sin\left(\frac{n\pi X_i}{L} - \phi_n\right) \tag{25}$$

This form of Fourier series is nonlinear and nonorthogonal in the parameters $c_n$ and $\phi_n$, and thus requires a complicated nonlinear fitting procedure. For Fourier analysis the fundamental period, $2L$, is generally a known constant.

The linear formulation of the Fourier series is given as the sum of a series of sine and cosine waves.

$$G(\alpha, X_i) = f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left( a_n \cos \frac{n\pi X_i}{L} + b_n \sin \frac{n\pi X_i}{L} \right) \tag{26}$$

The Fourier series shown in Eq. (26) is orthogonal and linear for a long continuous series of data.

The transformations between Eqs. (25) and (26) are as follows:

$$\begin{aligned} a_n &= c_n \sin \phi_n \\ b_n &= c_n \cos \phi_n \\ c_n &= \sqrt{a_n^2 + b_n^2} \\ \phi_n &= \tan^{-1} \left( \frac{a_n}{b_n} \right) \end{aligned} \tag{27}$$

The coefficients of Eq. (26) can be determined by a linear least squares procedure. However, since this is an orthogonal problem, for some conditions, an exact solution exists for $a_n$ and $b_n$. It is interesting that the exact solutions can be derived from the least squares formulation, [Eqs. (3)–(11)] for the specific conditions outlined below. This derivation is based on Eq. (26) being linear in the coefficients, which implies that only a single iteration is required, and that the initial estimated values, $\alpha'$, are zero:

$$\begin{aligned} a_n &= \frac{1}{L} \int_{-L}^{L} f(x) \cos \frac{n\pi x}{L} \, dx \\ b_n &= \frac{1}{L} \int_{-L}^{L} f(x) \sin \frac{n\pi x}{L} \, dx \end{aligned} \tag{28}$$

Does this solution apply to a discrete, noncontinuous, case? Consider the form of the off-diagonal elements of the $J^T J$ matrix:

$$\begin{aligned} (J^T J)_{a_n b_m} &= \sum_i \left[ \frac{1}{\sigma_i^2} \cos \frac{n\pi X_i}{L} \sin \frac{m\pi X_i}{L} \right], \qquad n \neq m \\ (J^T J)_{a_n a_m} &= \sum_i \left[ \frac{1}{\sigma_i^2} \cos \frac{n\pi X_i}{L} \cos \frac{m\pi X_i}{L} \right], \qquad n \neq m \\ (J^T J)_{b_n b_m} &= \sum_i \left[ \frac{1}{\sigma_i^2} \sin \frac{n\pi X_i}{L} \sin \frac{m\pi X_i}{L} \right], \qquad n \neq m \end{aligned} \tag{29}$$

Clearly, for the case of a very large number of data points spanning many times the period, $2L$, and constant experimental uncertainties, i.e., constant $\sigma_i$, all of the summations in Eq. (29) are equal to zero. Thus, for these conditions the Fourier series analysis, as in Eq. (26), is an orthogonal case.

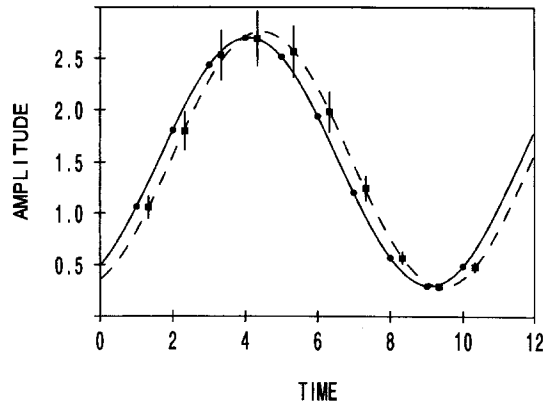


FIG. 7. Two simulated data sets for a Fourier series analysis. The squares contain experimental uncertainties that are proportional to the amplitude, whereas the circles contain uncertainties that are constant at 0.01. The lines are the least squares fits of these data according to Eq. (25).

However, if the data contain variable uncertainties, or variable spaced data points, or a small nonintegral number of periods, then these summations are not expected to be exactly equal to zero. For these conditions the exact solution [Eq. (28)] will yield results with less that the maximum likelihood of being correct, i.e., for these conditions the optimal parameters of the Fourier series cannot be evaluated by Eq. (28).

For example, consider the two least squares fits presented in Fig. 7. The solid line is the least squares fit of Eq. (25) to the solid circles. Note that the circles include constant experimental uncertainties of 0.01 units. When fit to $a_0$, $c_1$, and $\phi_1$, the analysis is orthogonal and all of the cross-correlations are zero (not shown). The dashed line in Fig. 7 is the least squares fit of Eq. (25) to the solid squares. The squares include experimental uncertainties with a magnitude that is proportional to the amplitude. Table V presents the cross-correlation matrix corresponding to the dashed line. Simply chang-

TABLE V
CROSS-CORRELATION MATRIX FOR DASHED LINE ANALYSIS[a]

| Parameter | Matrix | | |
|---|---|---|---|
| $c_1$ | 1.00000 | | |
| $\phi_1$ | 0.671532 | 1.00000 | |
| $a_0$ | −0.477350 | −0.921146 | 1.00000 |

[a] In Fig. 7.

TABLE VI
CROSS-CORRELATION MATRIX FOR SOLID LINE ANALYSIS[a]

| Parameter | Matrix | | | |
|---|---|---|---|---|
| $c_1$ | 1.00000 | | | |
| $\phi_1$ | 0.062116 | 1.00000 | | |
| $L$ | −0.069469 | −0.891111 | 1.00000 | |
| $a_0$ | −0.035931 | 0.464825 | −0.521835 | 1.00000 |

[a] In Fig. 7 when the period is also determined.

ing the noise characteristics contained within the data changed one cross-correlation coefficient from 0.0 to 0.92. Clearly, the dashed line analysis is no longer an orthogonal case and thus the exact solution for the Fourier series does not apply to the analysis of these data.

These Fourier series analyses are also not orthogonal for the determination of the period. Table VI presents the cross-correlation for the Fourier series analysis of the closed circles in Fig. 7 when the period is also determined by the least squares procedure. Here the cross-correlation between the phase shift and the period is −0.89. Clearly, the inclusion of the determination of the period into the Fourier series analysis introduced nonorthogonality into the analysis.

Note that the inclusion of the evaluation of the period in the analysis of the variable uncertainty data (squares in Fig. 7) further increased the cross-correlation of all of the parameters. These are shown in Table VII. The effects are cumulative.

Also, although not shown here, the magnitude of the cross-correlation increases if the data points are irregularly spaced or some data points are missing.

TABLE VII
CROSS-CORRELATION MATRIX FOR DASHED LINE ANALYSIS[a]

| Parameter | Matrix | | | |
|---|---|---|---|---|
| $c_1$ | 1.00000 | | | |
| $\phi_1$ | 0.309562 | 1.00000 | | |
| $L$ | −0.336108 | −0.916876 | 1.00000 | |
| $a_0$ | −0.920338 | 0.315724 | −0.346446 | 1.00000 |

[a] In Fig. 7 when the period is also determined.

## Exponential Fitting Functions

First, note that exponential fitting functions are neither linear nor orthogonal. They are the quintessential example of an ill-posed numerical analysis problem. There is no trick to eliminating the difficulties caused by the parameter cross-correlation. However, some things can be done to improve the ill-conditioned nature of the analysis of exponential decay data.

Consider a single exponential fitting function with an additive constant. This function can be written in at least four ways:

$$G(\alpha, X_i) = Ae^{-KX_i} + C \tag{30}$$

$$G(\alpha, X_i) = Ae^{-X_i/\tau} + C \tag{31}$$

$$G(\alpha, X_i) = Area\, Ke^{-KX_i} + C \tag{32}$$

$$G(\alpha, X_i) = \frac{Area}{\tau} e^{-X_i/\tau} + C \tag{33}$$

where

$$Area = \frac{A}{K} = A\tau$$

$$K = \frac{1}{\tau} \tag{34}$$

Equations (30) and (32) are formulated with elimination rate constants, whereas Eqs. (31) and (33) are formulated with elimination lifetimes. The preexponential factors in Eqs. (30) and (31) are amplitudes, whereas the preexponential factors in Eqs. (32) and (33) are in terms of the area under the exponential decay curve. Each of these equations has three adjustable parameters. If the data in Table I and Fig. 1 are least squares fit by Eqs. (30) through (33) while $C = 0$, the cross-correlation coefficients are 0.6892, −0.6802, −0.7649, and 0.7672, respectively. For this example it appears that the choice of rate constants versus lifetimes does not affect the cross-correlation coefficients. However, the magnitude of the cross-correlation is increased substantially when the fitting function is formulated with a preexponential factor in terms of the area under the decay curve as compared with an amplitude.

Does the number of data points change the apparent cross-correlation coefficients? The first matrix in Table VIII presents the cross-correlation matrix for the analysis of the data in Table I according to Eq. (30). Here all three parameters are varied. Also shown in Table VIII is an analogous analysis using only the odd numbered data points. Thus, the second matrix in Table VIII corresponds to the same range of data points, just fewer data points. For this example, decreasing the number of data points makes only a small difference in the cross-correlation coefficients.

TABLE VIII
CROSS-CORRELATION MATRIX FOR SUBSETS OF DATA[a]

| Parameter | Matrix | | |
|---|---|---|---|
| All 17 data points | | | |
| $A$ | 1.00000 | | |
| $K$ | −0.119816 | 1.00000 | |
| $C$ | −0.454369 | 0.882387 | 1.00000 |
| Data points 1–3–7– ··· –17 | | | |
| $A$ | 1.00000 | | |
| $K$ | −0.257120 | 1.00000 | |
| $C$ | −0.569803 | 0.87336 | 1.00000 |
| First 9 data points | 1.00000 | | |
| $A$ | −0.851445 | | |
| $K$ | −0.940692 | 1.00000 | |
| $C$ | | 0.968304 | 1.00000 |

[a] Shown in Table I when fit to Eq. (30).

Does the range of the data points change the apparent cross-correlation coefficients? The third matrix in Table VIII presents the analogous analysis using only the first nine data points. Here the magnitude of the cross-correlation increases significantly, and dramatically, as compared with the complete data set or the data set truncated to every odd-numbered data point. The conclusion drawn is that the wider the range of the independent variable, the lower the cross-correlation coefficients yielded. And as a consequence the fitting procedure will perform better.

The orthogonal polynomials were generated by simply shifting the independent variable axis slightly. Can a similar trick be used to lower the cross-correlation coefficients for the exponential decay case? Consider the modified version of Eq. (1) as shown in Eq. (35):

$$G(\alpha, X_i) = A_1 e^{-K_1(X_i - Q_1)} + A_2 e^{-K_2(X_i - Q_2)} \tag{35}$$

The first derivatives of Eq. (35) with respect to the fitting parameters are

$$\begin{aligned}
\frac{\partial G(\alpha, X_i)}{\partial A_1} &= e^{-K_1(X_i - Q_1)} \\
\frac{\partial G(\alpha, X_i)}{\partial K_1} &= -A_1(X_i - Q_1)e^{-K_1(X_i - Q_1)} \\
\frac{\partial G(\alpha, X_i)}{\partial A_2} &= e^{-K_2(X_i - Q_2)} \\
\frac{\partial G(\alpha, X_i)}{\partial K_2} &= A_2(X_i - Q_2)e^{-K_2(X_i - Q_2)}
\end{aligned} \tag{36}$$

These can be substituted into Eq. (16) to obtain the mathematical form for the individual elements of the $J^TJ$ matrix. It is this matrix that is inverted during the least squares parameter estimation procedures to obtain the values of the cross-correlation coefficients. To obtain an orthogonal system of equations, each of the off-diagonal elements of this matrix must be equal to zero.

$$\begin{aligned}(J^TJ)_{A_1K_1} &= \sum_i \left[\frac{1}{\sigma_i^2}(e^{-K_1(X_i-Q_1)})(-A_1(X_i-Q_1)e^{-K_1(X_i-Q_1)})\right] = 0 \\ (J^TJ)_{A_2K_2} &= \sum_i \left[\frac{1}{\sigma_i^2}(e^{-K_2(X_i-Q_2)})(-A_2(X_i-Q_2)e^{-K_2(X_i-Q_2)})\right] = 0\end{aligned} \tag{37}$$

Clearly, a pair of $Q$ values can be found for each particular data set such that these two elements of the $J^TJ$ matrix are equal to zero. However, the evaluation of these $Q$ values requires a prior knowledge of the $K$ values. They do not depend on the $A$ values since they can be canceled.

$$\begin{aligned}Q_1 &= \frac{\sum_i \frac{1}{\sigma_i^2} X_i e^{-2K_1X_i}}{\sum_i \frac{1}{\sigma_i^2} e^{-2K_1X_i}} \\ Q_2 &= \frac{\sum_i \frac{1}{\sigma_i^2} X_i e^{-2K_2X_i}}{\sum_i \frac{1}{\sigma_i^2} e^{-2K_2X_i}}\end{aligned} \tag{38}$$

The other four off-diagonal elements of the $J^TJ$ matrix cannot be made to be equal to zero if the $Q$ values have been obtained from Eq. (37).

$$\begin{aligned}(J^TJ)_{A_1K_2} &= \sum_i \left[\frac{1}{\sigma_i^2}(e^{-K_1(X_i-Q_1)})(-A_2(X_i-Q_2)e^{-K_2(X_i-Q_2)})\right] \neq 0 \\ (J^TJ)_{A_2K_1} &= \sum_i \left[\frac{1}{\sigma_i^2}(e^{-K_2(X_i-Q_2)})(-A_1(X_i-Q_1)e^{-K_1(X_i-Q_1)})\right] \neq 0 \\ (J^TJ)_{K_1K_2} &= \sum_i \left[\frac{1}{\sigma_i^2}(-A_1(X_i-Q_1)e^{-K_1(X_i-Q_1)})\right. \\ &\qquad \left.(-A_2(X_i-Q_2)e^{-K_2(X_i-Q_2)})\right] \neq 0 \\ (J^TJ)_{A_1A_2} &= \sum_i \left[\frac{1}{\sigma_i^2}(e^{-K_1(X_i-Q_2)})(e^{-K_2(X_i-Q_2)})\right] \neq 0\end{aligned} \tag{39}$$

Table IX presents the cross-correlation matrix for an analysis of the data in Table I with Eq. (35) as a fitting function. For this analysis the values of $Q_1$ and $Q_2$ were specified as 0.098625 and 0.194224. These were evaluated according to Eq. (38). While this procedure improved the cross-correlation, these values are still large and problematic.

It is possible that a different pair of $Q$ values could be derived based on any two of Eq. (39) that would provide a lower parameter cross-correlation coefficient. To test this, a grid search was done to test all possible pairs of $Q$ within the range of $\pm 5$ to a resolution of 0.01. This grid search required a least squares parameter estimation for each of the $10^6$ pairs of $Q$. It was found that for $Q$ values of $-0.97$ and $+1.63$, the largest cross-correlation coefficient was 0.9845. Consequently, it appears that simply scaling the independent variable, i.e., the $X$ axis, by an additive constant has some small effects, but not enough to warrant the effort.

## Conclusion

The objective of this chapter is to acquaint the reader with the origins and consequences of parameter cross-correlations. These commonly are simply a consequence of the fitting process and a limited amount of data. They need not be a consequence of the underlying physical process being studied. It was shown that these parameter cross-correlations can depend on the form of the fitting equation, the distribution of the data points along the independent variables, the distribution of the experimental uncertainties, and for nonlinear cases, the specific parameter values.

When the magnitude of these parameter cross-correlations approaches unity, the parameter estimation process becomes impossible. Consequently, it is important to consider methods to decrease the apparent parameter cross-correlations. Parameter cross-correlations can be modulated by altering the actual form of the fitting equation. They can also be altered by changing the distribution and range of the data points along the independent variable, the $X$ axis.

TABLE IX
CROSS-CORRELATION MATRIX FOR TWO-EXPONENTIAL ANALYSIS OF DATA[a]

| Parameter | Matrix | | | |
|---|---|---|---|---|
| $A_1$ | 1.00000 | | | |
| $K_1$ | −0.995215 | 1.00000 | | |
| $A_2$ | −0.999967 | 0.995605 | 1.00000 | |
| $K_2$ | −0.996475 | 0.984554 | 0.996124 | 1.00000 |

[a] In Table I, according to Eq. (35), $Q_1 = 0.098625$ and $Q_2 = 0.194224$.

The parameter cross-correlation coefficients are usually not a function of the dependent variable, i.e., the $Y$ axis. Therefore, it is not expected that transformation of the $Y$ axis, such as logarithmic plots, will alter the parameter cross-correlation coefficients. Such transformation may allow alterations of the fitting equation that will improve the parameter cross-correlation, but these are not without a price. One basic assumption of the least squares method is that all of the experimental uncertainties are contained in the dependent variable. Thus, nonlinear transformations of the $Y$ axis will also alter the apparent distribution of the experimental uncertainties contained within the data. Occasionally this can be useful,[8] but usually it is not a good idea.[9]

## Acknowledgments

The author acknowledges the support of the National Science Foundation Science and Technology Center for Biological Timing at the University of Virginia (NSF DIR-8920162), the General Clinical Research Center at the University of Virginia (NIH RR-00847), and the University of Maryland at Baltimore Center for Fluorescence Spectroscopy (NIH RR-08119).

[8] R. Abbot and H. P. Gutgesell, *Methods Enzymol.* **240,** 37 (1994).
[9] F. S. Acton, "Analysis of Straight-Line Data," p. 219. Chapman and Hall, New York, 1959.