Linear or Nonlinear Least-Squares Analysis of Kinetic Data?

Charles L. Perrin*

Department of Chemistry, University of California—San Diego, La Jolla, California 92093-0358, United States

S *Supporting Information*

ABSTRACT: The disadvantages of the usual linear least-squares analysis of first- and second-order kinetic data are described, and nonlinear least-squares fitting is recommended as an alternative.

KEYWORDS: *Kinetics, Graduate Education/Research, Reactions*

A common task in chemical kinetics is to extract the rate constant from the time dependence of concentrations. We first consider a reaction where reactant A is converted to product B, as in eq 1, and where the rate of reaction (rate of product formation or rate of reactant disappearance) is proportional to the concentration of reactant A, as in eq 2:

$$A \xrightarrow{k} B \tag{1}$$

$$v = \frac{d[B]}{dt} = -\frac{d[A]}{dt} = k[A] \tag{2}$$

Such a reaction is said to follow first-order kinetics, because the rate is proportional to only one concentration.

Alternatively, the rate may also be proportional to the concentration of another species C with proportionality constant k', as in eq 3:

$$v = \frac{d[B]}{dt} = -\frac{d[A]}{dt} = k'[A][C] \tag{3}$$

However, if C is a catalyst, so that its concentration remains constant, or if C is present in large excess, so that its depletion is negligible, its concentration does not deviate from its initial concentration $[C]_0$, and eq 3 can be simplified to eq 2 by setting $k'[C]_0$ equal to k. Such a reaction is said to follow pseudo-first-order kinetics, and k is often called a rate coefficient, rather than a rate *constant*, because its value changes with $[C]_0$.

In either case, the solution to eq 2 is given by eq 4,

$$[A] = [A]_0 e^{-kt} \tag{4}$$

where $[A]_0$ is the initial concentration of A.

■ FIRST-ORDER KINETICS BY LINEAR LEAST SQUARES

The customary method for analyzing concentration–time data that follow eq 4 is to take the logarithms of both sides, to obtain eq 5:

$$\ln[A] = \ln[A]_0 - kt \tag{5}$$

This indicates that a plot of $\ln[A]$ versus t should be linear with slope equal to $-k$ and intercept equal to $\ln[A]_0$, both of which can be evaluated by linear least-squares analysis. Alternatively, even if [A] itself is not measured, but a quantity A that is

proportional to [A], such as UV−vis absorbance, conductivity, titer, or NMR intensity, is measured instead, eq 5 then becomes eq 6,

$$\ln A = \ln A_0 - kt \tag{6}$$

which is also amenable to linear least-squares analysis. The familiar equations for evaluating $[A]_0$ or A_0, k, and the error in k are given in the Supporting Information.

Linear least-squares analysis is a poor method for evaluating the rate constant, as we shall see. Nevertheless, because it is so easy to apply, it is often presented in textbooks on kinetics,[1−3] even recent ones,[4,5] and it is the classical method. The defect in applying it to eq 5 or 6 is that the values are not equally reliable. Usually [A] or A is measured with a constant error $\Delta[A]$ or ΔA, as with UV−vis absorbance, whereas the values of $\ln[A]$ or $\ln A$ do not have a constant error. Figure 1 shows the fit of eq 5 to simulated data with $[A]_0 = 100$ M and $k = 0.0016$ s^{-1} but with a random error in [A] whose average is 3.37 M. When the reaction is followed to three half-lives, as shown in this example, it becomes obvious that the error in $\ln[A]$ is not constant but increases as [A] decreases. This is the case because according to the formula for propagation of errors, $\sigma_{\ln[A]} = (d \ln[A]/$
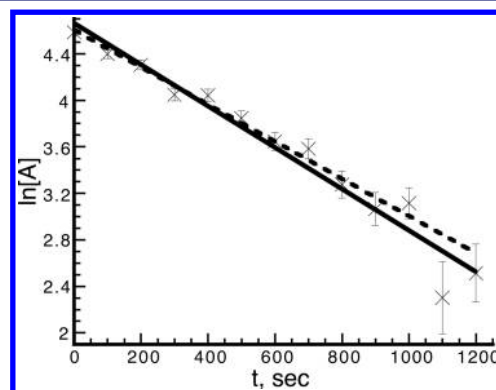
Figure 1. Plot of the best linear fit of eq 5 (solid line) to simulated data (✕), along with hypothetical values (dashed line) and error bars that increase as [A] and $\ln[A]$ decrease.

$d[A])\sigma_A = \sigma_A/[A]$. Such data are said to be heteroskedastic, meaning that the error is not constant across the range. (Moreover, it may be noted that logarithms of values lower than the hypothetical ones deviate more than those higher than the hypothetical, leading to a steeper calculated slope, as can be seen in Figure 1.)

This heteroskedasticity shows that the linear least-squares analysis can be faulty. Nevertheless, this method continues to be used and published by scientists. Indeed, over the past 10 years or so I have refereed about a dozen manuscripts for prominent journals where the authors obtained faulty rate constants by analyzing the kinetics with linear least-squares. It is hoped that this article will discourage that practice.

One remedy that is often proposed for dealing with heteroskedasticity is to use a weighted linear least-squares fit.[6] Instead of pretending that all of the values are equally reliable, as is done implicitly with linear least-squares, the values are weighted inversely to the variance associated with each of them. In that way the less reliable values become less influential. The appropriate equations are given in the Supporting Information.

A further complication arises if the analytical method measures the formation of product B rather than reactant disappearance. The counterpart to eq 4 is then eq 7:

$$[B] = [A]_0[1 - e^{-kt}] \tag{7}$$

Taking the logarithms of both sides of this equation does not produce an equation linear in t. Instead, eq 7 can be solved for $[A]_0 - [B]$, whose logarithm *is* linear in t, as given in eq 8:

$$\ln([A]_0 - [B]) = \ln[A]_0 - kt \tag{8}$$

To evaluate $\ln[A]_0$ and k by linear least-squares analysis requires that $[A]_0$ be known in advance. Small errors in $[A]_0$ can cause large errors in k because of the increasing error in the logarithm of the difference $[A]_0 - [B]$ as $[A]$ decreases.

A similar situation arises if $[A]$ does not react completely, so that residual reactant remains, or if the quantity A that is proportional to $[A]$ does not decrease to zero but approaches a nonzero baseline value A_∞ at infinite time. This is well-known to produce curvature in the log plot of eq 5 or 6. Instead, eq 4 must be revised to eq 9, and eqs 5 and 6 become eqs 10 and 11, respectively:

$$[A] - [A]_\infty = ([A]_0 - [A]_\infty)e^{-kt} \tag{9}$$

$$\ln([A] - [A]_\infty) = \ln([A]_0 - [A]_\infty) - kt \tag{10}$$

$$\ln(A - A_\infty) = \ln(A_0 - A_\infty) - kt \tag{11}$$

A plot of $\ln([A] - [A]_\infty)$ or $\ln(A - A_\infty)$ versus t should be linear with slope equal to $-k$ and intercept equal to $\ln([A]_0 - [A]_\infty)$ or $\ln(A_0 - A_\infty)$, respectively. In this case, $[A]$ or A_0 need not be known in advance, but $[A]_\infty$ or A_∞ must be known. Small errors in $[A]_\infty$ or A_∞ can cause large errors in k because of the increasing error in the logarithm of the difference $[A] - [A]_\infty$ or $A - A_\infty$ as $[A]$ or A decreases and $[B]$ increases.

Finally, it should be mentioned that in contrast to the evaluation of rate constants, the problem of heteroskedasticity does not arise in evaluating $\Delta H^\ddagger$ and $\Delta S^\ddagger$ from an Eyring plot of $\ln(k/T)$ versus $1/T$. This is because the *relative* error in a rate constant is usually the same for all values, so that the error in any rate constant k is the same fraction of that rate constant. Then, although the error in k is not constant, the error in $\ln k$ *is* constant across the range of temperatures. Consequently a

linear least-squares evaluation of slope and intercept does correctly provide $\Delta H^\ddagger$ and $\Delta S^\ddagger$.

■ FIRST-ORDER KINETICS BY NONLINEAR LEAST SQUARES

A more effective method for fitting all such data to first-order kinetics is nonlinear least-squares analysis. This method has been mentioned in various books on kinetics,[7−9] but without illustrative examples. In this *Journal*, nonlinear least-squares has been recommended for curve-fitting in general,[10] for fitting the kinetics of two-step reactions,[11,12] for fitting first-order kinetics,[13−17] for fitting enzyme kinetics to the Michaelis−Menten equation,[18,19] for using Excel's Solver,[20] and for estimating the precision of the resulting parameters,[21] but none of these articles documents the advantage over linear least-squares. Two useful books present both linear and nonlinear curve fitting, along with many other examples of data analysis.[22,23] Therefore, this article presents well-known but often-ignored information, and it makes the advantages of the nonlinear least-squares method graphic.

For generality, lest $[A]$ or A does not approach zero at infinite time, we seek to fit data to eq 9. We seek values of the three parameters k, $[A]_\infty$ or A_∞, and $[A]_0$ or A_0 (or $\Delta = [A]_0 - [A]_\infty$ or $A_0 - A_\infty$) that minimize S, the sum of the squares of the deviations of the calculated $[A]$ or A values from the observed ones. In past years, finding that minimum was a formidable problem.

With modern computers, even small ones, it is possible to find that minimum numerically. Numerous programs are available. A set of initial values is needed, and guesses or values from a linear least-squares analysis are adequate. The program will then find the "best" values of k, $[A]_\infty$ or A_∞, and Δ, i.e., the ones that minimize S. Often it can also find the uncertainties in those values. As an example, Figure 2 shows a
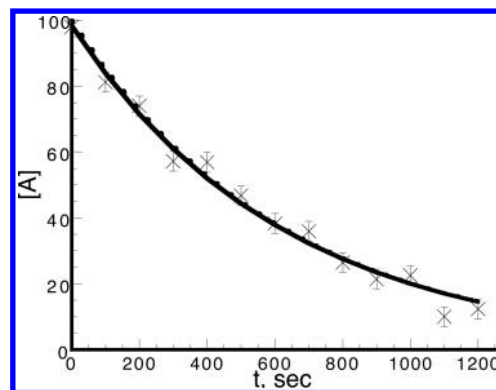
Figure 2. Plot of the nonlinear least-squares fit of eq 9 (solid curve) to simulated data (×), along with hypothetical values (dashed curve) and constant error bars representing the average error in $[A]$.

plot of the same data as in Figure 1, along with the best nonlinear least-squares fit to eq 9. It is clear that the fitted line is closer to the hypothetical one than in Figure 1 and therefore that the rate constant k obtained here is more reliable.

This procedure is also effective if the analytical method measures some property B that is proportional to the extent of formation of product B. Instead of eq 7, data can be fit to the three-parameter equation

$$B = B_\infty - \Delta e^{-kt} \tag{12}$$

where $\Delta = B_\infty - B_0$.

■ SECOND-ORDER KINETICS BY LINEAR AND NONLINEAR LEAST SQUARES

We next consider a reaction where reactants A and B are converted to product, as in eq 13, and where the rate of reaction is proportional to the concentrations of the two reactants, multiplied together, as in eq 14:

$$A + B \xrightarrow{k} \text{product} \tag{13}$$

$$v = \frac{d[\text{product}]}{dt} = -\frac{d[A]}{dt} = -\frac{d[B]}{dt} = k[A][B] \tag{14}$$

When $[A]_0 \neq [B]_0$, the solution to the differential equation can be written as eq 15:

$$\ln \frac{[A][B_0]}{[A]_0([A] - [A]_0 + [B]_0)} = ([A]_0 - [B]_0)kt \tag{15}$$

The customary method for analyzing second-order kinetics is linear least-squares fitting of the left-hand side of this equation as a function of t, whose slope is conveniently equal to $([A]_0 - [B]_0)k$. Unfortunately, the error in the logarithm is again not constant but increases as $[A]$ and $[B]$ decrease. As with eq 6, the data are heteroskedastic. Moreover, the data are subject to the uncertainty in $[A]_0 - [B]_0$, which is the small difference between two large concentrations.

Alternatively, eq 15 can be solved for $[A]$, as in eq 16:

$$[A] = \frac{[A]_0([A]_0 - [B]_0)}{[A]_0 - [B]_0 e^{-([A]_0 - [B]_0)kt}} \tag{16}$$

Experimental data can then be fit by nonlinear least-squares to find the values of the three parameters k, $[A]_0$, and $[B]_0$ that minimize the sum of the squares of the deviations of the calculated $[A]$ values from the observed ones.

Simpler equations result if $[A]_0 = [B]_0$. Then eq 16 becomes eq 17,

$$\frac{1}{[A]} = \frac{1}{[A]_0} + kt \tag{17}$$

which indicates that a plot of $1/[A]$ versus t should be linear with slope equal to k. Unfortunately, the data are again heteroskedastic because the error in the reciprocal is not constant but increases as $[A]$ and $[B]$ decrease. Moreover, the data remain subject to any error in the difference $[A]_0 - [B]_0$, which is assumed to be zero but whose magnitude increases relative to the decreasing values of $[A]$ or $[B]$. Consequently, it is not generally advisable to assume that samples can be prepared with identical initial concentrations of A and B.

The only situation where equal initial concentrations can be guaranteed is when A and B are identical, as in a dimerization or disproportionation (eq 18), so that the rate of reaction is given by eq 19:
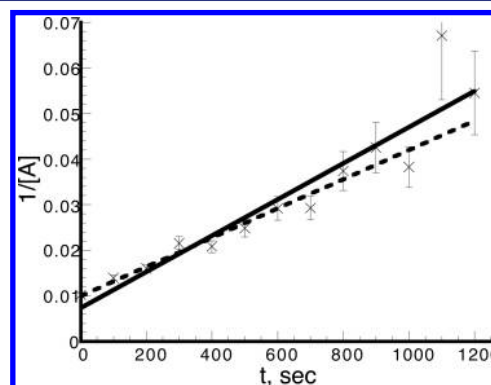
$$2A \xrightarrow{k} \text{product(s)} \tag{18}$$

$$v = \frac{d[\text{product(s)}]}{dt} = -\frac{1}{2}\frac{d[A]}{dt} = k[A]^2 \tag{19}$$

Equation 16 then becomes eq 20 and eq 17 becomes eq 21,

$$[A] = \frac{[A]_0}{1 + 2k[A]_0 t} \tag{20}$$

$$\frac{1}{[A]} = \frac{1}{[A]_0} + 2kt \tag{21}$$

which can be fit by linear least-squares to obtain the slope k and intercept $1/[A]_0$. Figure 3 shows the fit of eq 21 to simulated

Figure 3. Plot of the best linear least-squares fit of eq 21 (solid line) to simulated data (×), along with hypothetical values (dashed line) and error bars representing the average error in $1/[A]$.

data with $[A]_0 = 100$ M and $k = 0.000016$ M^{-1} s^{-1} but with a random error in $[A]$ whose average is the same 3.37 M as in Figures 1 and 2. Again the data are heteroskedastic, with errors that increase as $[A]$ decreases. A better method is to use nonlinear least-squares to fit eq 20 with the values of the two parameters k and $[A]_0$ that minimize the sum of the squares of the deviations of the calculated $[A]$ values from the observed ones. The result of this fit, for the same simulated data, is depicted in Figure 4.
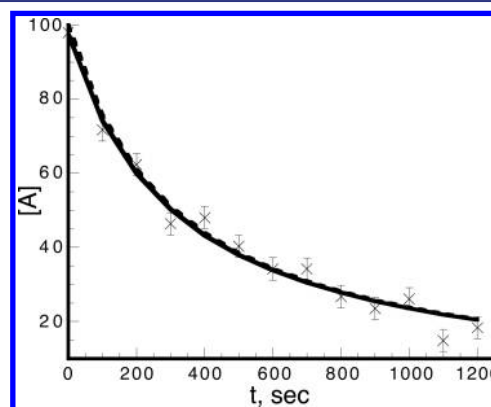
Figure 4. Plot of best nonlinear least-squares fit of eq 20 (solid curve) to simulated data (×), along with hypothetical values (dashed curve) and constant error bars representing the average error in $[A]$.

In summary, the familiar linear least-squares fit to kinetic data for first- and second-order reactions is defective because the data are heteroskedastic, i.e., of unequal reliability. This is especially the case if the analytical method measures product formation rather than disappearance of reactant or if the analytical method reports residual reactant. The remedy is to use nonlinear least-squares fitting, which is now readily accomplished with modern computer programs.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available on the ACS Publications website at DOI: 10.1021/acs.jchemed.6b00629.

> Equations S1−S3 for evaluating the slope, intercept, and error in the intercept by linear least-squares analysis and eqs S4−S6 for evaluating the slope, intercept, and error in the intercept by weighted linear least-squares analysis; simulated data used for the construction of Figures 1−4, with results of linear, weighted linear, and nonlinear least-squares fits of those simulated data showing the influence of experimental error on the rate constant as evaluated by the different methods (PDF, DOC)

AUTHOR INFORMATION

Corresponding Author

*E-mail: cperrin@ucsd.edu.

ORCID Ⓘ

Charles L. Perrin: 0000-0001-5732-5330

Notes

The author declares no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Frost, A. A.; Pearson, R. G. *Kinetics and Mechanism*, 2nd ed.; Wiley: New York, 1953; p 48.

(2) Laidler, K. J. *Chemical Kinetics*, 3rd ed.; Harper & Row: New York, 1987; pp 11−12.

(3) Pilling, M. J.; Seakins, P. W. *Reaction Kinetics*; Oxford University Press: Oxford, U.K., 1995; p 11.

(4) Wright, M. R. *Introduction to Chemical Kinetics*; Wiley: Chichester, U.K., 2004; p 63.

(5) House, J. E. *Principles of Chemical Kinetics*, 2nd ed.; Elsevier/Academic Press: Amsterdam, 2007; pp 5−6.

(6) Sands, D. E. Weighting Factors in Least Squares. *J. Chem. Educ.* **1974**, *51*, 473−474.

(7) Ritchie, C. D. *Physical Organic Chemistry: The Fundamental Concepts*, 2nd ed.; Marcel Dekker: New York, 1990; p 5.

(8) Connors, K. A. *Chemical Kinetics: The Study of Reaction Rates in Solution*; Wiley-VCH: New York, 1990; pp 49−51.

(9) Maskill, H. In *The Investigation of Organic Reactions and Their Mechanisms*; Maskill, H., Ed.; Blackwell: Oxford, U.K., 2007; p 54.

(10) Dye, J. L.; Nicely, V. A. A General Purpose Curvefitting Program for Class and Research Use. *J. Chem. Educ.* **1971**, *48*, 443−448.

(11) Bisby, R. H.; Thomas, E. W. Kinetic Analysis by the Method of Nonlinear Least Squares. *J. Chem. Educ.* **1986**, *63*, 990−992.

(12) Bluestone, S.; Yan, K. Y. A Method to Find the Rate Constants in Chemical Kinetics of a Complex Reaction. *J. Chem. Educ.* **1995**, *72*, 884−886.

(13) Copeland, T. G. The Use of Non-Linear Least Squares Analysis. *J. Chem. Educ.* **1984**, *61*, 778−779.

(14) McNaught, I. J. Nonlinear Fitting to First-Order Kinetic Equations. *J. Chem. Educ.* **1999**, *76*, 1457.

(15) Tellinghuisen, J. Nonlinear Least-Squares Using Microcomputer Data Analysis Programs: KaleidaGraph in the Physical Chemistry Teaching Laboratory. *J. Chem. Educ.* **2000**, *77*, 1233−1239.

(16) Silverstein, T. P. Using a Graphing Calculator To Determine a First-Order Rate Constant. *J. Chem. Educ.* **2004**, *81*, 485.

(17) Silverstein, T. P. Nonlinear and Linear Regression Applied to Concentration versus Time Kinetic Data from Pinhas's Sanitizer Evaporation Project. *J. Chem. Educ.* **2011**, *88*, 1589−1590.

(18) Martin, R. B. Disadvantages of Double Reciprocal Plots. *J. Chem. Educ.* **1997**, *74*, 1238−1240.

(19) Barton, J. S. A Comprehensive Enzyme Kinetic Exercise for Biochemistry. *J. Chem. Educ.* **2011**, *88*, 1336−1339.

(20) Harris, D. C. Nonlinear Least-Squares Curve Fitting with Microsoft Excel Solver. *J. Chem. Educ.* **1998**, *75*, 119−121.

(21) de Levie, R. Estimating Parameter Precision in Nonlinear Least Squares with Excel's Solver. *J. Chem. Educ.* **1999**, *76*, 1594−1598.

(22) Billo, E. J. *Excel for Chemists: A Comprehensive Guide, with CD-ROM*, 3rd ed.; Wiley: Hoboken, NJ, 2011; Chapters 14 and 15.

(23) Bevington, P. R.; Robinson, D. K. *Data Reduction and Error Analysis for the Physical Sciences*, 3rd ed.; McGraw-Hill: Dubuque, IA, 2003; Chapters 6 and 8.