

UC San Diego

UC San Diego Previously Published Works

Title

Linear or Nonlinear Least-Squares Analysis of Kinetic Data?

Permalink

<https://escholarship.org/uc/item/2dz66348>

Journal

Journal of Chemical Education, 94(6)

ISSN

0021-9584

Author

Perrin, Charles L

Publication Date

2017-06-13

DOI

10.1021/acs.jchemed.6b00629

Peer reviewed

Linear or Nonlinear Least Squares Analysis of Kinetic Data?

Charles L. Perrin

Department of Chemistry, University of California—San Diego

La Jolla, CA 92093-0358

cperrin@ucsd.edu

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



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

Alternatively, the rate may also be proportional to the concentration of another species C with proportionality constant k' , as in eq 3. However, if C is a catalyst, so that its concentration remains constant, or if C is 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, where $[A]_0$ is the initial concentration of A.

$$v = \frac{d[B]}{dt} = -\frac{d[A]}{dt} = k[A][C] \quad (3)$$

$$[A] = [A]_0 e^{-kt} \quad (4)$$

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. Now a plot of $\ln[A]$ vs. t is linear, with slope $-k$ and with intercept $\ln[A]_0$, 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, eq 5 then becomes eq 6, 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 Supporting Information.

$$\ln[A] = \ln[A]_0 - kt \quad (5)$$

$$\ln A = \ln A_0 - kt \quad (6)$$

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,^{1,2,3} 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 to eq 5 of simulated data with $[A]_0 = 100$ and $k = 0.0016$ but with a random error in $[A]$ whose average is 3.37. 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 because, according to the formula for propagation of errors, $\sigma_{\ln[A]} = (d\ln[A]/d[A])\sigma_A = \sigma_A/[A]$. 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 Fig. 1.)

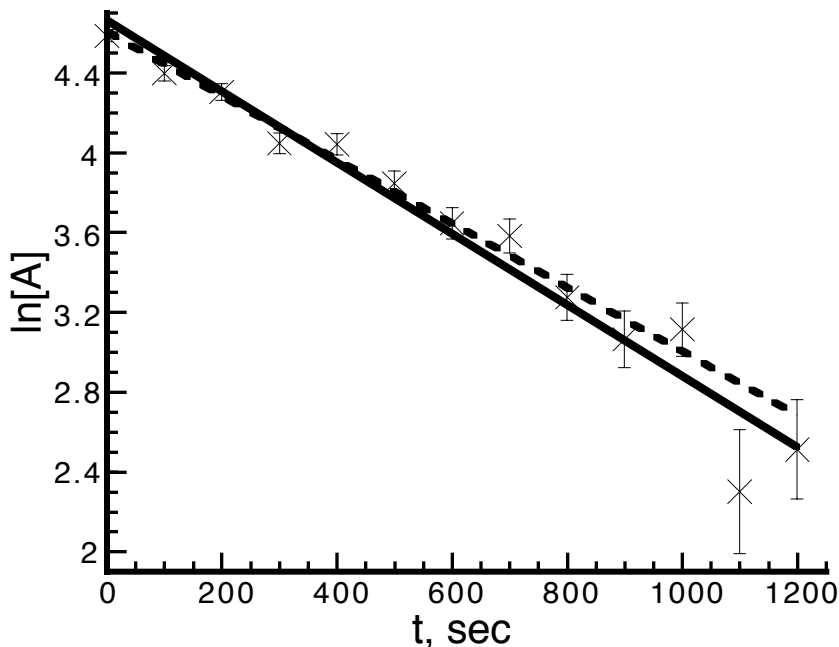


Figure 1. Plot of best linear fit (solid) to eq 5 of simulated data (x), along with hypothetical values (dashed) and error bars that increase as $[A]$ and $\ln[A]$ decrease.

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 ten or so years I have refereed about a dozen manuscripts for prominent journals where the authors obtained faulty rate constants by analyzing 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 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 formation of product B, rather than reactant disappearance. The counterpart to eq 4 is then eq 7. Taking the logarithms of both sides does not produce an equation linear in t . Instead it can be solved for $[A]_0 - [B]$, whose logarithm *is* linear in t , as given in eq 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.

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

$$\ln([A]_0 - [B]) = \ln[A]_0 - kt \quad (8)$$

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 5-6. Instead eq 4 must be revised to eq 9, and eq 5 or 6 becomes eq 10 or 11. A plot of $\ln([A] - [A]_\infty)$ or $\ln(A - A_\infty)$ vs. t is linear, with slope $-k$ and with intercept $\ln([A]_0 - [A]_\infty)$ or $\ln(A_0 - A_\infty)$. Now $[A]$ 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.

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

$$\ln([A] - [A]_\infty) = \ln([A]_0 - [A]_\infty) - kt \quad (10)$$

$$\ln(A - A_\infty) = \ln(A_0 - A_\infty) - kt \quad (11)$$

Finally, it should be mentioned that in contrast to the evaluation of rate constants the problem of heteroskedasticity does not arise in evaluating ΔH^\ddagger and ΔS^\ddagger from an Eyring plot of $\ln(k/T)$ vs. $1/T$. This is because the *relative* error in a rate constant is usually the same for all, 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 ΔH^\ddagger and ΔS^\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,^{7,8,9} 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,^{13,14,15,16,17} 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 calculated $[A]$ or A from observed. 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 , the ones that minimize S . Often it can also find the uncertainties in those values. As an example, Fig. 2 shows a plot of the same data as in Fig. 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 Fig. 1 and that therefore the rate constant k obtained here is more reliable.

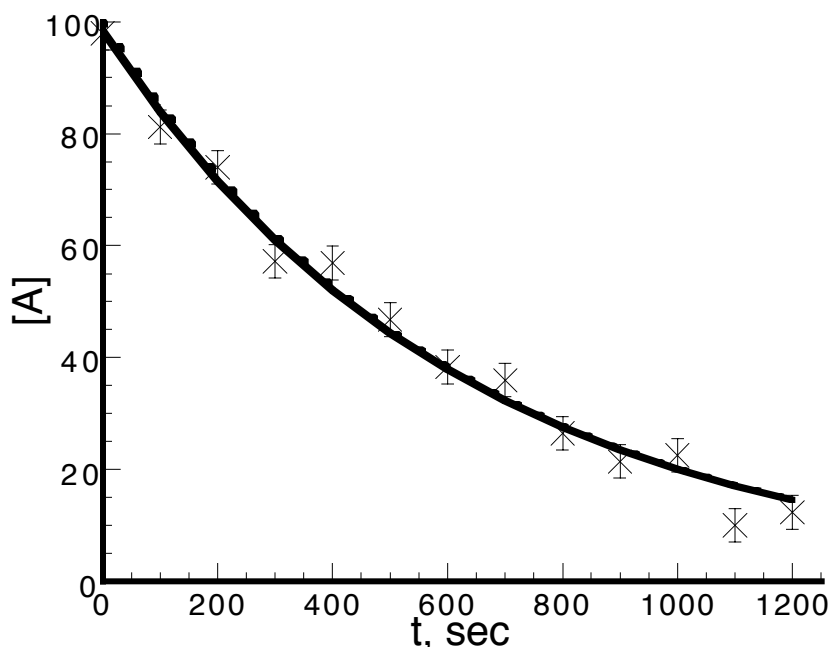


Figure 2. Plot of nonlinear least-squares fit (solid) to eq 9 of simulated data (x), along with hypothetical values (dashed) and constant error bars representing the average error in [A].

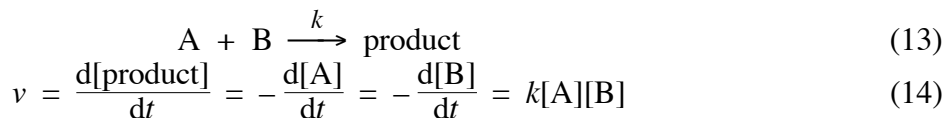
This procedure is also effective if the analytical method measures B , proportional to the extent of formation of product [B]. Instead of eq 7, data can be fit to three-parameter eq 12, where $\Delta = B_{\infty} - B_0$.

$$B = B_{\infty} - \Delta e^{-kt} \quad (12)$$

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 product of the concentrations of the two reactants, as in eq 14. When $[A]_0 \neq [B]_0$ the solution to the differential equation can be written as eq 15. The customary method for analyzing second-order kinetics is linear least-squares fitting of the left-hand side of this equation, whose slope vs. t is conveniently $([A]_0 - [B]_0)k$. 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.



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

Alternatively, eq 15 can be solved for $[A]$, as in eq 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 calculated $[A]$ from observed.

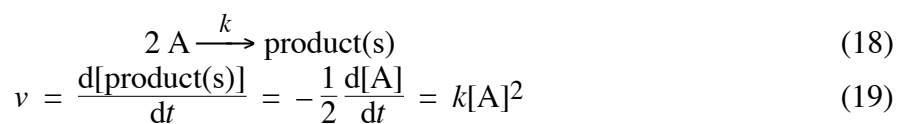
$$[A] = \frac{[A]_0([A]_0 - [B]_0)}{[A]_0 - [B]_0 e^{-([A]_0 - [B]_0)kt}} \quad (16)$$

Simpler equations result if $[A]_0 = [B]_0$. Then eq 16 becomes eq 17, whose plot vs. t is linear, with a 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.

$$\frac{1}{[A]} = \frac{1}{[A]_0} + kt \quad (17)$$

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. Eq. 16 then becomes eq. 20, and eq 17 becomes eq 21, which can be fit by linear least squares to obtain slope k and intercept $1/[A]_0$. Figure 3 shows the fit to eq 17 of simulated data with $[A]_0 = 100$ and $k = 0.000016$ but with a random error in $[A]$ whose average is the same 3.37

as in Figs. 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 calculated $[A]$ from observed. The result of this fit, for the same simulated data, is depicted in Fig. 4.



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

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

$$\frac{1}{[A]} = \frac{1}{[A]_0} + 2kt \quad (21)$$

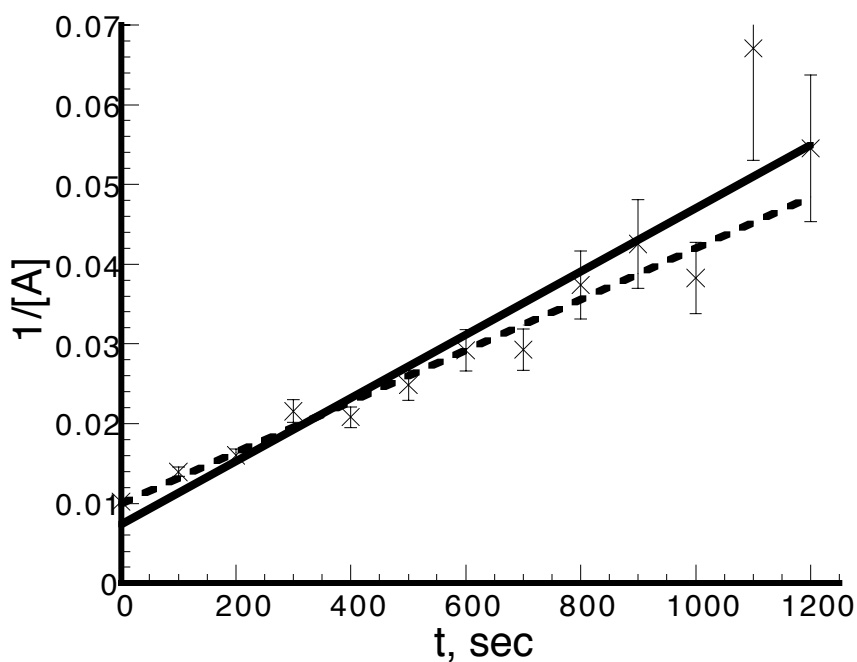


Figure 3. Plot of best linear least-squares fit (solid) to eq 21 of simulated data (x), along with hypothetical values (dashed) and error bars representing the average error in $[A]$.

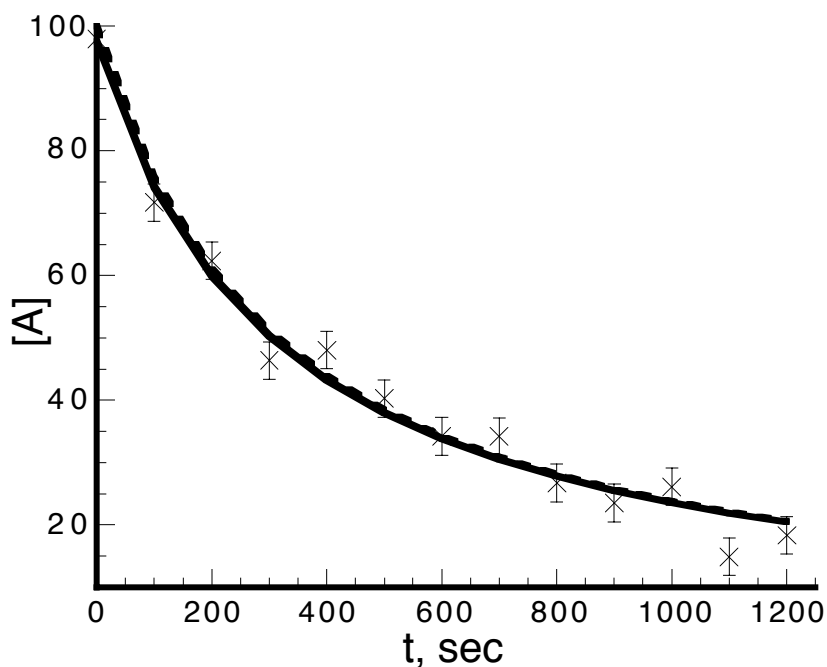


Figure 4. Plot of best nonlinear least-squares fit (solid) to eq 20 of simulated data (x), along with hypothetical values (dashed) and constant error bars representing the average error in [A].

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.

Supporting Information Equations S1-S3 for evaluating slope, intercept, and error in intercept by linear least squares analysis, and equations S4-S6 for evaluating slope, intercept, and error in intercept by weighted linear least squares analysis. Simulated data used for construction of Figs. 1-4, with results of linear, weighted, 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.

Acknowledgment. Preparation of this manuscript and its presentation at the Gordon

Research Conference on Physical Organic Chemistry were supported by NSF Grant CHE11-48992.

References

1. Frost, A. A.; Pearson, R. G. *Kinetics and Mechanism*, 2ed John Wiley: New York, 1953; p. 48.
2. Laidler, K. J. *Chemical Kinetics* 3ed Harper & Row: New York 1987; pp. 11-12.
3. Pilling, M. J.; Seakins, P. W. *Reaction Kinetics* Oxford University Press: Oxford, 1995; p. 11.
4. Wright, M. R. *Introduction to Chemical Kinetics* Wiley: Chichester W. Sussex, 2004; p. 63.
5. House, J. E. *Principles of Chemical Kinetics*, 2ed 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*, 2ed Marcel Dekker: New York NY, 1990; p. 5.
8. Connors, K. A. *Chemical Kinetics: The Study of Reaction Rates in Solution*, Wiley-VCH: New York NY, 1990; pp.49-51.
9. Maskill, H. In *The Investigation of Organic Reactions and Their Mechanisms*, H. Maskill, Ed., Blackwell: Oxford, 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*, 3ed John Wiley: Hoboken NJ, 2011; Chapters 14-15.
23. Bevington, P. R.; Robinson, D. K. *Data Reduction and Error Analysis for the Physical Sciences*, 3ed, McGraw-Hill: Dubuque IA, 2003; Chapters 6, 8.