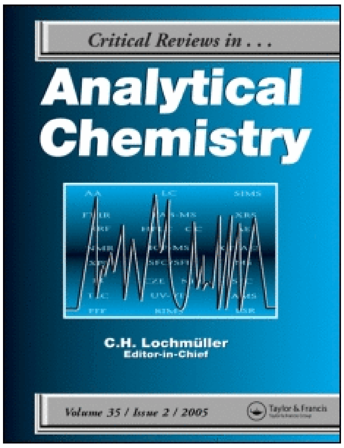
This article was downloaded by:
On: *17 January 2011*
Access details: *Access Details: Free Access*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Critical Reviews in Analytical Chemistry

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713400837

Some New Techniques for the Analysis and Interpretation of Chemical Data

Louis Meites; Bruce H. Campbell

To cite this Article Meites, Louis and Campbell, Bruce H.(1979) 'Some New Techniques for the Analysis and Interpretation of Chemical Data', Critical Reviews in Analytical Chemistry, 8: 1, 1 — 53

To link to this Article: DOI: 10.1080/10408347908542709

URL: http://dx.doi.org/10.1080/10408347908542709

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

SOME NEW TECHNIQUES FOR THE ANALYSIS AND INTERPRETATION OF CHEMICAL DATA

Author: **Louis Meites**
Department of Chemistry
Clarkson College of Technology
Potsdam, New York

Referee; Bruce H. Campbell
Research Analysis Service
J. T. Baker Chemical Company, Phillipsburg, New Jersey

TABLE OF CONTENTS

I. INTRODUCTION

In chemistry, as in any other science, there are two kinds of experiments. One involves qualitative observation, or perhaps semiquantitative measurement, and is performed to permit giving a yes-or-no answer to some question such as "does this compound contain an amino group?" or "does this substance exhibit significant bacteriostatic activity?" The other involves precise quantitative measurement and is performed so that some numerical value can be assigned to a quantity or physical parameter that characterizes the system being studied or the kind of behavior being observed.

To those engaged in experiments of the second kind, it is a sad but fundamental truism that there is scarcely any quantity of interest that can be measured directly. Distances, times, volumes, electric currents and potentials, pH values, temperatures, the radiant powers of beams of radiation, and many other kinds of quantities are amenable to direct measurement, but it is not often that any of these has any real significance. The chemist who measures the volume of reagent used in a titration has no interest in that volume and measures it merely because it can be used to compute the amount of some substance contained in the solution being titrated. The values of equilibrium and rate constants, and of constants or near-constants of proportionality (such as molar absorptivities, polarographic diffusion current constants, and equivalent or molar conductances among many others), all have to be deduced from the values of other quantities that can be measured. The same thing is true of the values of most other physical constants, including standard and half-wave potentials, wavelengths of maximum absorbance, half-lives, standard changes of enthalpy, and many others.

Fortunately, there are many cases in which it is easy to convert a value of a measured quantity into a value of the desired quantity. Instruments that combine observations of radiant powers to yield a value of the absorbance have been available for a long time; instruments that combine this result with a preset constant of proportionality to yield a value of the concentration of the absorbing species are now becoming fairly common. The distinction between measurement and calculation will become blurred as microprocessors proliferate.

This review is devoted to situations in which a number of different measurements have been made so that a much smaller number of desired quantities can be evaluated. These situations are of two kinds: those in which the evaluation is straightforward and those in which it is not.

The evaluation of a wavelength of maximum absorption is usually (however, see Eriksson et al.[1]) an example of the first type. One obtains a number of values of the absorbance, each at a different wavelength, and combines these to obtain a plot of absorbance against wavelength. On this plot, the wavelength at the maximum is easy to locate by visual inspection. The evaluation of the volume of reagent required to reach the point of maximum slope on a potentiometric titration curve is equally straightforward in principle, although it is a little more complex in execution. One obtains a number of values of the potential E of an electrochemical cell, each at a different volume V of added reagent, combines these values to obtain values of the second derivative $\Delta^2E/\Delta V^2$, and interpolates between these to find the value of V at the point where $\Delta^2E/\Delta V^2$ is equal to zero.

The characteristic feature of all cases of this first type is that one is not concerned with the nature of the relationship between the independent variable (which is of course the wavelength in the first example and the volume of reagent or the generation time in the second) and the dependent one (the absorbance or the cell potential). In locating a wavelength of maximum absorption, one rarely pays any attention whatever to whether the absorption band is Lorentzian, Gaussian or even grossly asymmetrical or malformed. The technique of locating a point of maximum slope on a potentiometric titration curve is the same regardless of whether the curve is nearly symmetrical around the equivalence point or markedly asymmetrical.

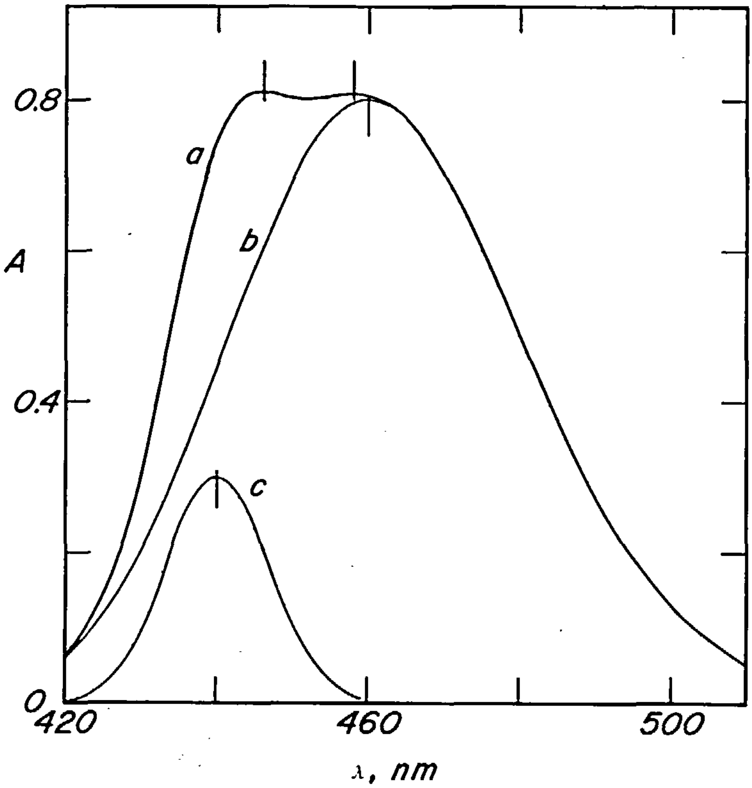


FIGURE 1. (a) Schematic absorption spectrum for a mixture of two absorbing substances; (b) and (c) spectra of these substances individually. The short vertical lines mark the wavelengths of maximum absorption, which are 440 and 460 nm for the individual components but 446 and 458 nm for the mixture.

Every silver lining has a cloud. It is convenient and simple to avoid taking the nature of the relationship between the variables into account, but doing so sometimes leads to an incorrect or meaningless result. Figure 1 shows a familiar example in which the overlapping of two absorption bands causes the wavelengths of maximum absorption of a mixture to differ from those of its separate components.

By far the most difficult and vexing problems, however, are those that arise in cases of the second type, in which the relationship between the variables must be known and used in order to evaluate the desired quantity. An especially striking example was provided by Carr and Jordan [2] and Carr.[3] These authors considered the titration curve that is obtained when a concentrated solution of reagent R is added at a constant rate to a much more dilute solution of a substance A, with which R reacts in a simple irreversible second-order reaction:

$$\mathrm{A} + \mathrm{R} \xrightarrow{k} \mathrm{P} \qquad (1)$$

If the rate constant k were extremely large, a plot of c_P, the concentration of P, against time would be /‾-shaped. Plots of this shape are common in thermometric, spectrophotometric, amperometric, and other kinds of titrations (in which the addition of reagent is usually incremental, rather than continuous, to facilitate the attainment of equilibrium after each addition). If k is finite, the shape is more complex. Carr and Jordan [2] succeeded in obtaining an equation that describes the dependence of c_P/c^0_A on time. Values of this ratio, in which c^0_A denotes the initial concentration of A, are easily obtained from experimental data (by correcting the measured increase of temperature, absorbance, limiting current, or other signal at each point for dilution and dividing the corrected value by the one that corresponds to the final horizontal segment of the curve). The equation is

$$\frac{c_P}{c_A^0} = 1 - \frac{\exp\left(\frac{\beta}{2}\right)\exp\left(-\frac{\beta}{2}\left[\frac{t}{t^*} - 1\right]^2\right)}{1 + \sqrt{\frac{\pi\beta}{2}}\exp\left(\frac{\beta}{2}\right)\left(\operatorname{erf}\sqrt{\frac{\beta}{2}}\left[\frac{t}{t^*} - 1\right] + \operatorname{erf}\sqrt{\frac{\beta}{2}}\right)} \qquad (2)$$

where β is a dimensionless parameter that combines the rate constant k and other quantities (c^0_A and the rate of addition of R) whose values are known to the experimenter; t^* is the time required to reach the equivalence point, and this can be calculated a priori for any experiment.

Consider an experiment that is performed with the aim of securing a value of k. Since the relationship between β and k will be known, this aim is equivalent to that of evaluating β. It is not difficult to obtain values of c_p/c^0_A at different times. If the reaction is known to be a simple irreversible second-order reaction, these data must conform to Equation 2, in which β is the only unknown quantity. How does one evaluate β? More fundamentally, how does one obtain assurance that the data conform to Equation 2 — and hence that the reaction does have the mechanism described by Equation 1 — within the precision that the data are believed to have? If one concludes that Equations 1 and 2 are not correct, how does one choose the next hypothesis to test?

Multiparametric curve-fitting — a term which the writer now regrets having coined[4] for what the statistician would call "nonlinear regression analysis," but which is apparently being adopted by other chemists[5] — and deviation-pattern recognition are techniques that help the chemist to cope with problems and questions like these. It is to them, and to some still newer techniques based on them, that this review is devoted. Multiparametric curve-fitting may be defined as a technique for finding the values of any number of parameters, in an equation of any form, that yield the best fit to that equation for any particular set of experimental data. Deviation-pattern recognition employs those values to generate a plot whose shape may either confirm or disprove the validity of the hypothesis that the equation is applicable to the data and, in disproving it, may point the way to another equation that will yield a better and more satisfactory fit.

As far as I know, deviation-pattern recognition was developed in my own laboratory, although the utility of the deviation plot, on which it is based, for the simpler purpose of indicating whether systematic errors are present or absent, is so widely recognized that an international organization has recommended its use for the presentation of numerical data.[6] The chemical use of nonlinear regression has a much longer history. Thanks to Sillén and his students and colleagues, much of our knowledge about the compositions and stabilities of hydroxo complexes of metal ions is based on it, and several very general computer programs for work in the area of ionic equilibria have been written and described at length.[7-10] There are a number of other computer programs for effecting nonlinear regression in a very general way, and there are also many ad hoc programs that have been written to cope with particular equations. A long list of citations might be assembled, and those given here will be illustrative rather than complete.

One of the main purposes of this review is to argue that nonlinear regression is such a powerful and versatile technique for handling chemical data that few chemists can afford the luxury of not using it. The reader is justified in wondering why, if there is anything to this argument, the technique is not already much more widely used than it is. There are several reasons why it is not. One is that textbooks and monographs on statistics and chemistry hardly mention nonlinear regression although they may discuss linear regression at considerable length. In the days when desk calculators reigned supreme, linear regression was relatively simple while nonlinear regression was forbiddingly complex, and it is quite understandable that the potentialities of the latter were ignored as long as they were so difficult to achieve. Another reason is that classical statistics was developed by mathematicians, biologists, and others who have seen the availability and utility of theoretical equations in a light quite different from that in which they would be seen by the chemist or physicist. The authors of the sentence "Faced by nonlinear regression, one often has no knowledge of a theoretical equation to use" were biological statisticians unaware of the uses that chemists could find for the technique; the author of the sentence "Chemists in this field have already published far too many 'theoretical' papers that are useless because there is no way to apply the equations to experimental data" was a chemist unaware of the techniques that statisticians could provide for accomplishing the purposes he discerned.

There are other reasons as well. For a long time after digital computers became available, the great majority of programs for effecting nonlinear regression were written for specific purposes and required large computers. Most of them would have had to be more or less drastically modified to serve other purposes, and chemists were naturally apt to be unwilling to expend the time required to make these modifications, or to write wholly new programs, unless they had already been persuaded that nonlinear regression would be useful in their own work. Moreover, though there are no longer many chemists who do not have access to a large computer, some delay and expense are often entailed in obtaining it.

To meet all these reasons head-on at once, and also in the hope of keeping the length of this review within reasonable limits, I have tried to restrict it to applications that are feasible on a minicomputer. Indeed, it is largely based on work that has been done in my own research group. This of course has the disadvantage of overemphasizing our own interests and their importance, but it has two compensating advantages. One advantage is that it simplifies my task in presenting the development and interrelationships of the different stages that have been reached. The other advantage is that it enables the reader to judge what might be done with his own programming experience and computing facilities. Few of my collaborators had any extensive knowledge of programming when their work in this area began, and our computer is a PDP8/I operated in a multiuser configuration that provides us with a user area of only 4096 words.

I shall begin with a brief review of linear regression, in which I shall try to explain the dangers and limitations that beset it. I shall argue that using it correctly is so much less simple than it appears to be that nonlinear regression usually represents an easier and more straightforward approach to the problems with which analytical and physical chemists are concerned. I shall then give a very brief description of the technique of nonlinear regression, which will be aimed at showing how simple its concepts are rather than at describing the details of any particular procedure. Then I shall describe some of its applications, beginning with its relatively straightforward use for evaluating the parameters that appear in theoretical equations and continuing through progressively more complex uses to the exciting new development of fully computerized procedures for the qualitative interpretation of quantitative data.

II. LINEAR REGRESSION

The basic tenet of least-squares calculations is that the "best" values of the parameters appearing in any equation are those that minimize the sum of the squares of the deviations of the values of the dependent variable calculated by means of that equation from the corresponding measured values. The distinction between linear and nonlinear regression is based on the form of the equation. In linear regression, the equation is linear with respect to every one of the parameters being evaluated. (Linearity with respect to the independent variable is irrelevant.) A few of the equations that are amenable to linear regression are

$$y = a + bx \tag{3a}$$

$$y = a + bx + cx^2 \tag{3b}$$

$$y = a + b/x + c \log x \tag{3c}$$

In nonlinear regression, the equation is nonlinear with respect to at least one parameter, as are the equations

$$y = a\mathrm{e}^{-bx} \tag{4a}$$

$$y = a(1 - \mathrm{e}^{-bx}) + cx \tag{4b}$$

$$y = a + b \log (x + c) \tag{4c}$$

Equation 2 is nonlinear with respect to β, and the equation that describes the pH during the titration of a strong acid with a strong base is nonlinear with respect to most of its parameters.

It is usually assumed, in both linear and nonlinear regression, that the values of the independent variable(s) are exactly known but that those of the dependent one are afflicted by random errors of measurement. For many experiments, this is a reasonable assumption. There are ways of allowing for possible random errors in the values of the independent variable(s), but these will not be described here.

There are two different kinds of random errors that can be defined exactly, and there is a third that cannot. The first kind includes errors whose absolute values are not only randomly distributed but also independent of the value of the variable they afflict. Errors in weighing are likely to be of this kind. If there is a standard error of $\pm s$ mg in weighing an object whose weight is 1g, the standard error in weighing a 50-g object is likely to be indistinguishable from s mg provided that both weighings are made under the same conditions: by the same operator exerting the same care in using the same balance, and so on. Similarly, the uncertainty in reading a deflection from a curve recorded with a strip-chart recorder does not usually depend on whether that deflection is 10 or 100 mm. The uncertainty in measuring a pH value does not usually depend on whether that value is equal to 2 or to 10.

The majority of chemical measurements are of this kind, but the requirement that the conditions must be the same is nontrivial, and whether it is satisfied is sometimes not easy to assess. The standard error of weighing begins to increase as the weight of an object becomes high enough to deform the knife edges and bend the beam of the balance being used. The standard error of a deflection on a recorded chart increases dramatically for deflections exceeding full scale. The standard error of measuring a pH value equal to 10 may become very large if the buffer capacity of the solution is very small at that point or if the electrode exhibits a large cation error in weakly alkaline solutions.

The second class comprises random errors whose values stand in some fixed proportion to the quantities they afflict. For this class, it is the relative standard error that is independent of the measured value. An example might arise in measuring the resistance of a thermistor with an auto-ranging digital voltmeter. At 25°, the resistance might be read as 5000 Ω and might have an uncertainty of ± 2 Ω. At 50°, it might be read as 500.0 Ω, and the circuitry of the meter might give rise to an uncertainty of ± 0.2 Ω; at 75°, the reading might be 50.00 Ω and the uncertainty ± 0.02 Ω. Again there is a tacit assumption that the conditions are the same, which in this example implies that the temperature coefficient of resistance of the thermistor and the precision with which its temperature is controlled do not vary with temperature.

The third class includes random errors for which neither the absolute nor the relative value is independent of the quantity being measured (or, by extension, of the independent variable). In the example just given, the resistance of the thermistor would lie between 1000 and 9999 Ω over the range of temperatures from about 17 to about 42°C. If the measurements were confined to this range, it would be reasonable to assume that the standard error was independent of resistance or temperature. However, if the measurements extended from 17 to 60°C, the absolute value of the standard error would change from 2 Ω in one portion of the range to 0.2 Ω in the other, while the relative standard error would vary from 2/9999 (or 0.2/999.9) to 2/1000. It would be necessary either to take one or the other of these variations into account in the calculations, or to disable the auto-range circuitry and accept a constant uncertainty of ± 2 Ω in each one of the measurements. Similarly, there are some spectrophotometers that yield values of transmittance for which the absolute standard error is independent of transmittance. However, if these values are converted to absorbance, they yield a set of values in which neither the absolute nor the relative standard errors conform to the fundamental assumption that a least-squares fit entails. This means that, if data obtained with such an instrument are to be used to evaluate the rate constant of a pseudo-first-order reaction in which an absorbing reactant disappears, a least-squares fit must be to the equation

$$T = \exp\left(e^{-kt} \ln T_0\right) = T_0^{\exp(-kt)} \qquad (5a)$$

rather than to the equation

$$A = A_0 e^{-kt} \qquad (5b)$$

The two forms are mathematically equivalent, but they will not yield identical values of k if there is any random error whatever in the data. Under the conditions described, nonlinear regression onto Equation (5a), assuming that the standard error of measurement is independent of T, will yield the correct value of k. Even nonlinear regression onto Equation (5b) will yield the wrong value unless great pains are taken. The nature of these pains will be mentioned in a later paragraph of this section.

Suppose that it is desired to fit a set of data x_1,y_1; x_2,y_2;...; x_n,y_n to the simple equation $y = bx$. Suppose further that a consideration of the apparatus and techniques used in making the measurements indicates that the standard error of each measurement is independent of the value of y, so that the absolute errors can reasonably be supposed to be normally distributed throughout the whole set of data. What is sought is the value of the parameter b that mininizes the sum of the squares of the deviations of the calculated values of y, represented by the symbol $y_{i,\text{calc}}$, from the corresponding measured values. This sum is $\sum_{i=1}^{n} (y_{i,\text{calc}} - y_i)^2$.

At the desired minimum

$$\frac{d\left(\sum_{i=1}^{n} [y_{i,calc} - y_i]^2\right)}{db} = 0 \qquad (6)$$

Since the values of $y_{i,calc}$ will be obtained from the equation $y_{i,calc} = bx_i$,

$$\frac{d\left(\sum_{i=1}^{n} [y_{i,\text{calc}} - y_i]^2\right)}{db} = \frac{d\left([bx_1 - y_1]^2 + [bx_2 - y_2]^2 + \ldots\right)}{db}$$

$$= 2(b\,x_1 - y_1)\,x_1 + 2(b\,x_2 - y_2)\,x_2 + \ldots$$

By equating this expression to zero, it is easy to show that the desired value of b is given by

$$b = \sum_{i=1}^{n} x_i y_i \Big/ \sum_{i=1}^{n} (x_i)^2 \qquad (7)$$

If, however, the measurements have been made in a way that makes it reasonable to suppose that the relative errors in y are normally distributed, Equation 6 must be replaced by

$$\frac{d\left(\sum_{i=1}^{n}\left[\frac{y_{i,\text{calc}} - y_i}{y}\right]^2\right)}{db} = 0 \qquad (8)$$

and this immediately raises the question whether it should be y_i (the experimental value) or $y_{i,calc}$ that appears in the denominator of the quantity whose square is summed.[11]. No matter how this question is answered, it is clear that the best value of b in this situation will not be the one described by Equation 7.

Expressions for the best values of a and b in the equation

$$y = a + bx \qquad (3a)$$

appear in many handbooks; those for the best values of a, b, and c in the equation

$$y = a + bx + cx^2 \qquad (3b)$$

appear in a few. For any such equation, the expressions are easy to derive. With Equation 3a, one would perform two differentiations to obtain separate descriptions of

$$d\left[\sum_{i=1}^{n} (y_{i,\text{calc}} - y_i)^2\right]/da$$

and

$$d\left[\sum_{i=1}^{n} (y_{i,\text{calc}} - y_i)^2\right]/db$$

Equating these to zero yields two equations that involve both a and b and are easy to solve. The procedure becomes more complex as the number of parameters increases, but it remains simple in principle. It is this simplicity that is responsible for the widespread use of linear regression.

There are, however, some pitfalls that must be avoided in using it, and there are many scientists who have not succeeded in avoiding them. Two common pitfalls will be discussed here. I do not want to overstate the case and shall therefore say at once that both can be avoided, and those whose primary concern is with techniques for handling data are well aware of the precautions that are necessary and sufficient. My argument is that many chemists, being preoccupied with experimental and theoretical problems, are unaware of the precautions and therefore fall into error and confusion, and that the precautions are not always easy to take.

One common danger can be illustrated by supposing that a chemist wishes to evaluate the rate constant of the first-order reaction

$$A \xrightarrow{k} P \qquad (9)$$

and measures the absorbance A of a reaction mixture, under conditions such that A is proportional to the concentration of unreacted starting material, at a number of different times t. For the sake of convenience it will be assumed that the data are obtained with an instrument constructed in such a way that the absolute errors in the absorbance values are randomly distributed; as was said above in connection with Equation 5, this assumption has to do with the natures of the instrument and the experimental procedure rather than with the nature of the problem. What is desired is to find the best value of k in the equation

$$A = A_0 \ e^{-k\,t} \qquad (5b)$$

where A_0 is the absorbance at the instant of mixing.

Since Equation 5b is not linear with respect to k, linear regression is not applicable to it as it stands. Textbooks on statistics point out that it can be rewritten in the mathematically equivalent form

$$\ln A = \ln A_0 - k\,t \qquad (10)$$

which is of the form $y = a + bx$, with $y = \ln A$, $a = \ln A_0$, $b = -k$, and $x = t$. The pitfall consists of using the standard equations of linear regression to evaluate b and setting k equal to $-b$.

A chemist who was conscientious enough to analyze a set of data by both least-squares and graphical procedures might detect the problem by constructing a plot like the one in Figure 2.* This is definitely concave upward at long times, possibly because of stray light, an error in making the zero-transmittance setting, suspended matter in the solution, an unexpected complexity in the mechanism, or some other reason. The least-squares result is represented by the dashed line, and this clearly does not provide a satisfactory fit to the data. It might be observed that the experimental data do conform closely to a straight line over approximately a full decade: after some agony with a straight edge, the experimenter might decide to use only the points up to (and including) the one shown by the solid circle.

This is inefficient almost to the point of absurdity. It involves a good deal of time, trouble, and judgment, and it sacrifices the objectivity that one hopes to obtain from the least-squares procedure. If carried only a little farther, it is capable of producing a value of k that is based on the slope of a line from which the points begin to deviate after the reaction is only 50% complete — that is, the value of a pseudo-first-order rate constant for a reaction that does not obey pseudo-first-order kinetics.

To discover the source of the difficulty, one need only recall the fundamental assumption that the errors in y are randomly distributed. The widely quoted equations describing the best values of a and b in a fit to the equation $y = a + b\,x$ are obtained by assuming that the absolute error in any value of y is independent of y. This is equivalent to assuming that, if the value $y = 0$ has a standard error of ± 0.001, the value of $y = -2.3$ also has a standard error of ± 0.001. Since $y = \ln A$, these pairs of figures correspond to a standard error in A itself of ± 0.001 absorbance unit if $A = 1$ but to a standard error of only ± 0.0001 absorbance unit if $A = 0.1$. Beginning with a decision that the standard error of A is independent of A, we have been led

---

* I am indebted for these data to two former undergraduate students at the Polytechnic Institute of Brooklyn.

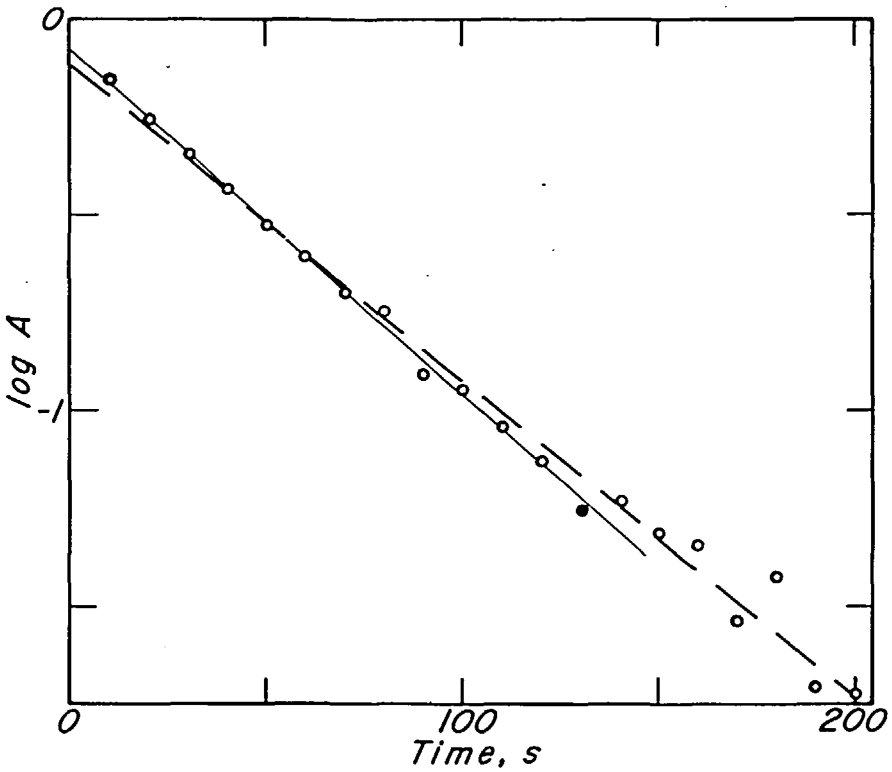


FIGURE 2. The circles show absorbance-time data conforming to Equation 5b but decaying to a finite value of the absorbance rather than to zero, and with a small superimposed random error of measurement. The dashed line shows the result of linear regression onto Equation 10 and corresponds to $k = 0.0186$. The thin solid line is obtained graphically by ignoring all the points beyond the one shown by the solid circle, and corresponds to $k = 0.0202$. Nonlinear regression onto Equation 5b, using all of the points, gives $k = 0.02029$.

into the tacit assumption that the standard error of A decreases as A itself decreases; linearizing Equation 5b has changed the precisions that are being assigned to the points, and the points for which A is smallest assume a disproportionate weight because their precisions are now grossly overestimated.

The other pitfall would arise if, in exactly the same situation, the measured absorbances were due to the product P rather than to the starting material A. They would then be described by the equation

$$A = A_\infty \left(1 - e^{-kt}\right) \qquad \text{(11a)}$$

where A_∞ is the absorbance measured at a time so long that the reaction is virtually complete. One linearized form of this equation is

$$\ln\left(A_\infty - A\right) = \ln A_\infty - k\,t \qquad \text{(11b)}$$

which is of the form $y = a + bx$ with $y = \ln(A_\infty - A)$, $a = \ln A_\infty$, $b = -k$, and $x = t$. Values of y can be obtained by measuring A after a very long time has elapsed, equating this to A_∞, subtracting each of the measured values of A from it, and taking the natural logarithm of the difference. The common equations for a and b may now be applied. The pitfall is that the value of A_∞ obtained from the fit ($= e^a$) never coincides exactly with the measured one used in computing the values of y. I have seen more than one competent chemist hiding the difficulty from himself by the simple expedient of calculating b but not a. Clearly, however, it makes no sense to assign different values to A_∞ on the two sides of Equation 11b, or to trust the value of k that results from doing this.

In these two examples, it is the second pitfall that is easier to avoid. One must guard against the possibility of assigning two different values to the same quantity. Since Equation 11b provides no protection against this, it is not a useful equivalent of Equation 11a, and instead one must write

$$\ln\left(1 - \frac{A}{A_\infty}\right) = -k\,t \qquad (11c)$$

This is of the form $y = b\,x$, and one could calculate the best value of $b\,(= -k)$ from Equation 7 if the standard error in $\ln(1 - A/A_\infty)$ were independent of A — which will not usually be true. Some linearized forms of nonlinear equations are legitimate, but others are not, and the experimenter must be alert to the difference.

The first pitfall is much less easy to avoid even in this simple case. The problem is that the points are given incorrect weights in the linear regression procedure, and it can be solved by introducing a "weighting factor" that compensates for this. The form of the weighting factor depends on the equation used and can be deduced in the following way. The value of k is sought that minimizes the sum $\sum_{i=1}^{n} (A_{i,\text{calc}} - A_i)^2$, and this is given by the simultaneous equations

$$\frac{d\left(\sum_{i=1}^{n} [A_{i,calc} - A_i]^2\right)}{d\,A_0} = 0 \qquad (12a)$$

and

$$\frac{d\left(\sum_{i=1}^{n} [A_{i,calc} - A_i]^2\right)}{d\,k} = 0 \qquad (12b)$$

The standard error of A_i, assumed to be independent of A, is given by

$$s_{A_i} = \sqrt{\frac{\sum_{i=1}^{n} (A_{i,calc} - A_i)^2}{n - p}} \qquad (13a)$$

where n is the number of values or data points employed, and p is the number of parameters being evaluated; the difference $n - p$ is called the number of degrees of freedom. If s_{A_i} is evaluated by comparing the results of a number of replicate measurements, their mean is the only parameter that can be evaluated, and consequently $p = 1$; however, in this example $p = 2$ because both A_0 and k are being evaluated.

Similarly, the standard error of y_i, which is assumed to be constant and independent of y in performing ordinary linear regression onto the equation $y = a + b\,x$, is given by

$$s_{y_i} = \sqrt{\frac{\sum_{i=1}^{n} (y_{i,calc} - y_i)^2}{n - p}} \qquad (13b)$$

and again $p = 2$ because there are two parameters (a and b) to be evaluated. Equations 13a and 13b can be combined to yield

$$\sum_{i=1}^{n} (A_{i,\mathrm{calc}} - A_i)^2 = \frac{s_{A_i}^2}{s_{y_i}^2} \sum_{i=1}^{n} (y_{i,\mathrm{calc}} - y_i)^2 \tag{14}$$

The ratio of variances on the right-hand side of Equation 14 is given by

$$\frac{s_{A_i}^2}{s_{y_i}^2} = \left(\frac{\partial A}{\partial y}\right)^2 \tag{15}$$

Since $y = \ln A$, $\partial y/\partial A = 1/A$ and $\partial A/\partial y = A$. Combining this result with Equation 14 and then with Equations 12a and b yields

$$\frac{\mathrm{d}\left(\sum_{i=1}^{n} A_i^2 \, [y_{i,\mathrm{calc}} - y_i]^2\right)}{\mathrm{d}\, A_0} = 0 \tag{16a}$$

and

$$\frac{\mathrm{d}\left(\sum_{i=1}^{n} A_i^2 \, [y_{i,\mathrm{calc}} - y_i]^2\right)}{\mathrm{d}\, k} = 0 \tag{16b}$$

Treating Equation 16b in the same fashion as Equation 6 gives

$$\frac{\mathrm{d}\left(\sum_{i=1}^{n} A_i^2 \, [y_{i,\mathrm{calc}} - y_i]^2\right)}{\mathrm{d}\, k} =$$

$$\frac{\mathrm{d}\left(A_1^2 \, [\ln A_0 - k t_1 - y_1]^2 + A_2^2 \, [\ln A_0 - k t_2 - y_2]^2 + \ldots\right)}{\mathrm{d}\ k} =$$

$$\frac{\mathrm{d}\,(A_1^2 \, [a - k t_1 - y_1]^2 + A_2^2 \, [a - k\ t_2 - y_2]^2 + \ldots)}{\mathrm{d}\ k} =$$

$$2A_1^2 \, (a - k t_1 - y_1)\,(-t_1) + 2A_2^2 \, (a - k t_2 - y_2)\,(-t_2) + \ldots = 0 \tag{17}$$

from which an equation linear in both a and k will eventually emerge. Another such equation can be obtained by treating Equation 16a in the same way, and expressions for a and k can be obtained by solving these two simultaneous equations. It is the factor A_i^2 in each term of Equation 17 that is the weighting factor. In general, as can be seen from Equation 15, the weighting factor is equal to $(\partial Y/\partial y)^2$, where Y is the dependent variable whose values are measured, and y is the independent variable resulting from the linearization.

Not all equations are as easy to linearize as Equation 5b, and the weighting factor may be very much more difficult to evaluate than it is in this rather simple example. There are some nonlinear equations that are easily amenable to linearization followed

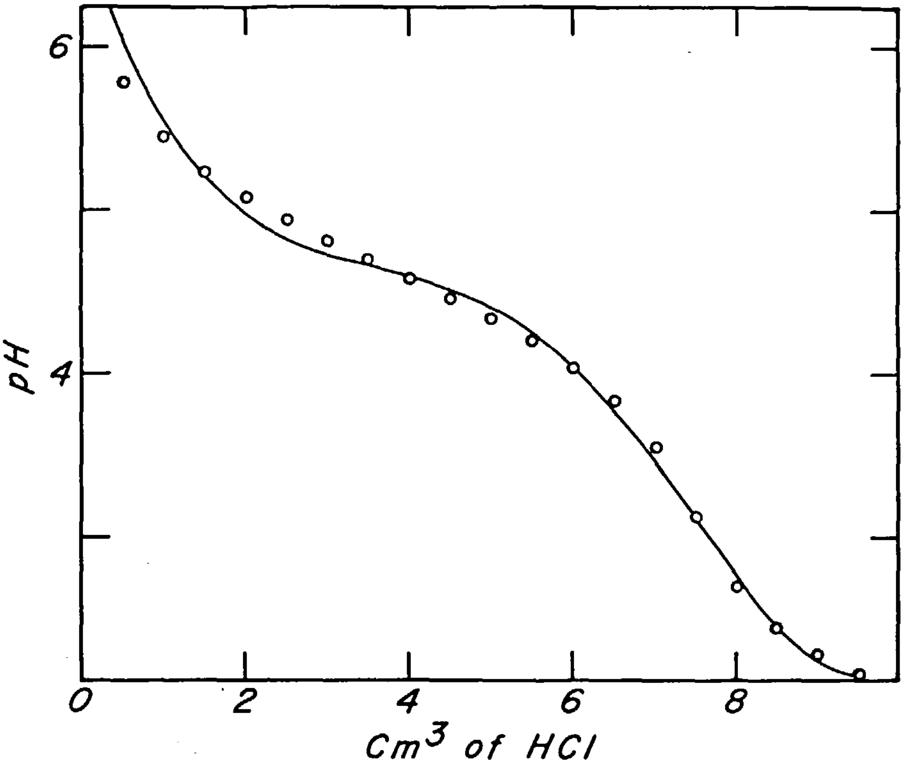


FIGURE 3. The circles show experimental data obtained by Barry in a titration of 79.98 cm^3 of 0.015849 F potassium acetate with 0.16914 F hydrochloric acid in the presence of 3 F potassium chloride. The solid line shows the best fit that can be obtained to the seven-parameter equation $\mathrm{pH} = a + b\,V + c\,V^2 + d\,V^3 + e\,V^4 + f\,V^5 + g\,V^6$. According to this best fit, the point of maximum slope is reached when $V = 7.547$ cm^3, whereas the equivalence point lies at $V = 7.494$ cm^3 (and the point of maximum slope is know to precede the equivalence point for a titration of this type.) The solid line has a positive slope for $V > 9.4$ cm^3; there are situations in which the pH of a solution can be increased by adding acid to it, but this is not one of them. (Nonlinear regression onto the titration-curve equation (cf. Reference 31 for details) gave 0.015849 F for the concentration of potassium acetate, in exact agreement with the expected value.)

by weighted linear regression, but nonlinear regression is always a simpler and more convenient approach if the skeleton of a program for effecting it is available.

With these caveats I shall leave linear regression after adding one somewhat biased paragraph. In my opinion, polynomial equations like $y = a + b\,x + c\,x^2 + \ldots$ have been used far more often than they deserve to be, for there are few theoretical relationships having this form. Figure 3 presents an especially horrid example, which was obtained by fitting data obtained in a potentiometric titration of acetate ion with hydrochloric acid to a seven-parameter equation of this form. The fit is bad, the values of the parameters have no physical significance whatever, and the equation does not yield satisfactory values for either the volume at the point of maximum slope or the pH at the start of the titration. Most chemists probably share Manfred Eigen's opinion that "with four independently adjustable parameters you can fit an elephant," but there are only two possible interpretations of Figure 3. One interpretation is that a potentiometric acid-base titration curve is not an elephant; the other interpretation is that the opinion is incorrect. Gutknecht and Perone[12] were unable to secure good fits to stationary-electrode voltammograms with an empirical equation involving eight independently adjustable parameters, and this is the common experience. Even when such fits suffice for interpolation or smoothing, they are unlikely to produce enlightenment because the values of the parameters are not amenable to physical interpretation. Of course, the same thing would be true of nonlinear regression onto empirical equations: the utility of nonlinear regression arises from the fact that it is based on theoretical equations. It makes possible the evaluation of physically meaningful parameters in equations that, like Equation 2, are so complex that it would be forbiddingly difficult and time-consuming to obtain and solve the equations that describe the best values of the parameters.

III. NONLINEAR REGRESSION

A. Basic Concepts

It is convenient to regard nonlinear regression as representing the converse of the procedure by which one calculates a theoretical curve. In doing that, one assigns certain values to the parameters appearing in the equation that is to be used and then computes the value of the dependent variable y that corresponds to each of a number of more or less arbitrarily chosen values of the independent variable x. In nonlinear regression, the values of x and y are given, and the problem is to compute the values of the parameters that gave rise to them.

Several different principles can be employed in constructing a computer program to effect nonlinear regression, and the programs that have been written vary widely in sophistication and complexity. As far as the user is concerned, they all consist of three parts. One part contains the coordinates of the data points — that is, the values of x and y that are being fitted to the equation. Another part is a subroutine to solve that equation, accepting a set of values of the parameters furnished to it by the main body of the program and combining them with the successive values of x to obtain a set of calculated values of y. These two parts of the program must be written by the user. The third part, with which they are combined, compares the calculated values of y with the measured values to find the value of the sum of the squares of the deviations or "error sum" — which may simply be the sum of the squares of the deviations, $\sum_{i=1}^{n} (y_{i,\text{calc}} - y_i)^2$, if the standard error of measurement is independent of the value of y, or the sum of the squares of the relative deviation, $\sum_{i=1}^{n} \left(\frac{y_{i,\text{calc}} - y_i}{y} \right)^2$ if the standard error of measurement is proportional to y — and adjusts the values of the parameters in ways that tend to minimize this sum.

Figure 4 shows how a very simple program might be employed to find the best value of β from a set of experimental data that conformed to Equation 2. The solid curve is a schematic representation of the "error surface," which is a $(j+1)$-dimensional plot of the error sum against the values of the parameters, j in number, that are being evaluated. This error surface is two-dimensional because β is the only parameter that governs the relationship between the independent variable t and the dependent one c_P/c^o_A. Beginning with some arbitrary initial estimate of β — such as $\beta = 1$ — the values of the error sum might be computed both for this estimated value β_1 and for a slightly smaller or larger one $\beta_1 + \Delta\beta$. Combining these, it would be found that the error surface had a positive slope at $\beta = \beta_1$. This means that the minimum lies at some value of β that is smaller than β_1. Accordingly, the original estimate β_1 might be divided by some arbitrarily chosen factor f to obtain a second estimate β_2. The slope of the error surface at $\beta = \beta_2$ would then be obtained by evaluating the error sums for $\beta = \beta_2$ and $\beta = \beta_2 + \Delta\beta$. If this were still positive, the minimum would lie at some still smaller value of β, and a third estimate β_3 might then be obtained by dividing β_2 by f. In Figure

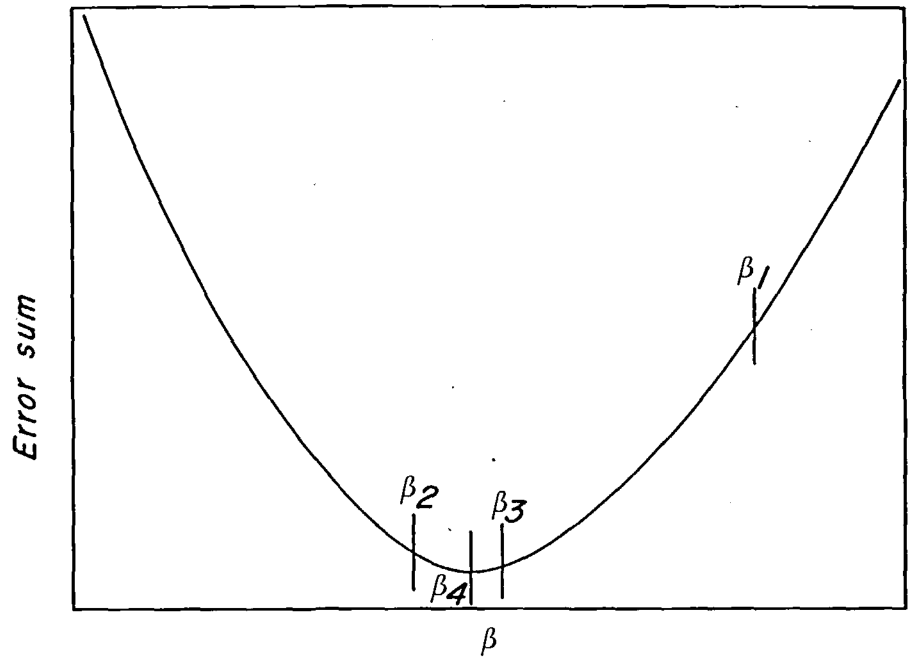


FIGURE 4. Schematic representation of a two-dimensional error surface, illustrating the operation of a simple program designed to locate the minimum.

4, however, the slope is negative at $\beta = \beta_2$. The change of sign of the slope between $\beta = \beta_1$ and $\beta = \beta_2$ might be made to cause a decrease in the value of f — for example, by taking the cube root of the preceding value — and cause the third estimate β_3 to be produced by multiplying the preceding one β_2 by the new smaller value of f. In Figure 4, this yields a positive value of the slope of the error surface at $\beta = \beta_3$. Because the sign of the slope has changed once again, the value of f would be decreased again, and a fourth estimate β_4 would be obtained by dividing β_3 by this newly decreased value of f. The process might be stopped when f had been decreased to some value — such as 1.0001 — so close to 1 that further approximations were not worth performing, or when round-off errors caused the slope to become indistinguishable from zero.*

Some little trouble might be entailed in constructing the main body of the program so that it would do the things I have described; however, this is the province of the author, rather than the user, of the program. The same thing is true of the additional trouble that would be necessary for the procedure to cope with multidimensional error surfaces. Since my interest is in the ways in which nonlinear regression can be made to serve the chemist's purposes rather than in the ways by which it can be effected, I shall not discuss these problems or the solutions that have been found to them. Programs employing different techniques and varying widely in both sophistication and complexity (and consequently in the size and cost of the computer facility needed to run them and also in the time that their execution consumes) are already available, and new ones continue to appear. My purpose here is simply to show that, despite the rather large amount of numerical computation involved, there is no mystery about the basic concepts.

A natural reaction to the claims that have been made for nonlinear regression is to ask "How is it possible to evaluate so many parameters at once? Do computations of three — or four, or five, or even more — values have any real significance?"

The curve in Figure 5 represents a polarogram — a plot of the current flowing

---

* I am not aware of any program constructed along exactly these lines, but one resembling it in some aspects is cited in Feldberg, S., Klotz, P., and Newman, L., *Inorg. Chem.*, 11, 2860, 1972.

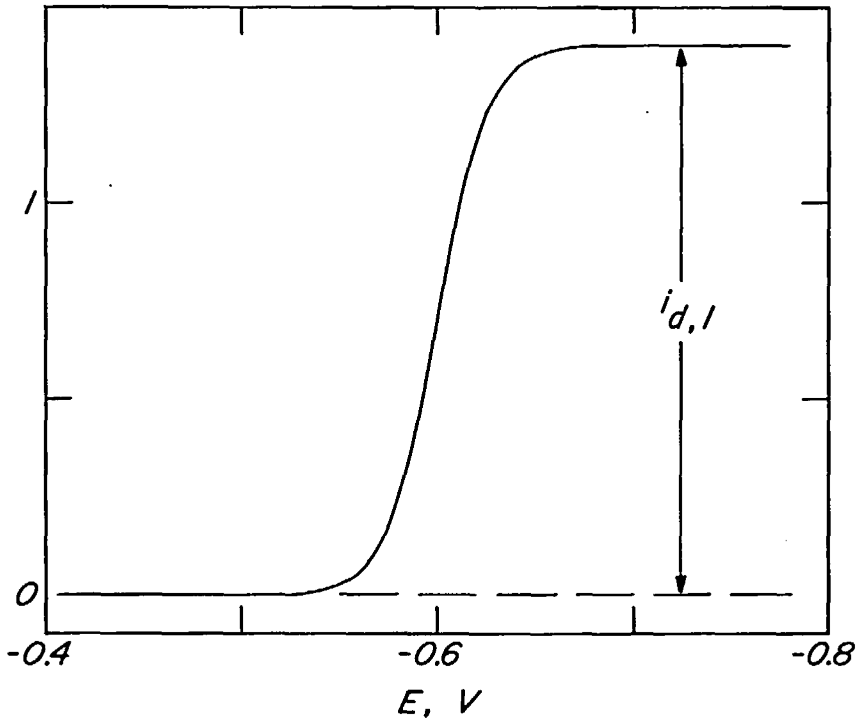


FIGURE 5. Schematic polarogram showing a single wave.

through a dropping mercury electrode against the applied potential imposed on an electrochemical cell having some reference electrode, such as a saturated calomel electrode, as its second electrode. I have simplified it in two ways: by ignoring the oscillations of current caused by the periodic growth and fall of the mercury drops and by ignoring the residual current that would flow even in the absence of the dissolved substance that is responsible for the wave.

Generally, polarographic waves obey the equation

$$E = E_{1/2} - S' \log_{10} \frac{i}{i_{d,l} - 1} \qquad (18a)$$

in which i is the current that flows at the applied potential E; $E_{1/2}$ is the half-wave potential, a parameter whose value depends on the identity of the substance responsible for the wave, the composition of the solution in which it is dissolved, and the temperature and certain other experimental conditions; $i_{d,l}$ is the limiting diffusion current, which is the limiting value that i approaches at applied potentials, so much more negative than $E_{1/2}$ that the half-reaction occurring at the dropping electrode proceeds virtually to completion; and S' is a parameter that is equal to the Nernstian factor 2.303 RT/nF if the half-reaction is thermodynamically reversible, but that differs from this if it is not.

The traditional procedure for evaluating i_d, $E_{1/2}$, and S' begins with a measurement of $i_{d,l}$ at a sufficiently negative potential. This is combined with the values of i at various points to compute the corresponding values of $\log_{10} (i/[i_{d,l} - i])$ and E is then plotted against these. The resulting "log plot" is a straight line whose slope is equal to S' and whose intercept (at $\log_{10} [i/(i_{d,l} - i)] = 0$) is equal to $E_{1/2}$. Three independent parameters have been evaluated. The equivalent procedure consists of nonlinear regression onto an alternative form of Equation 18a:

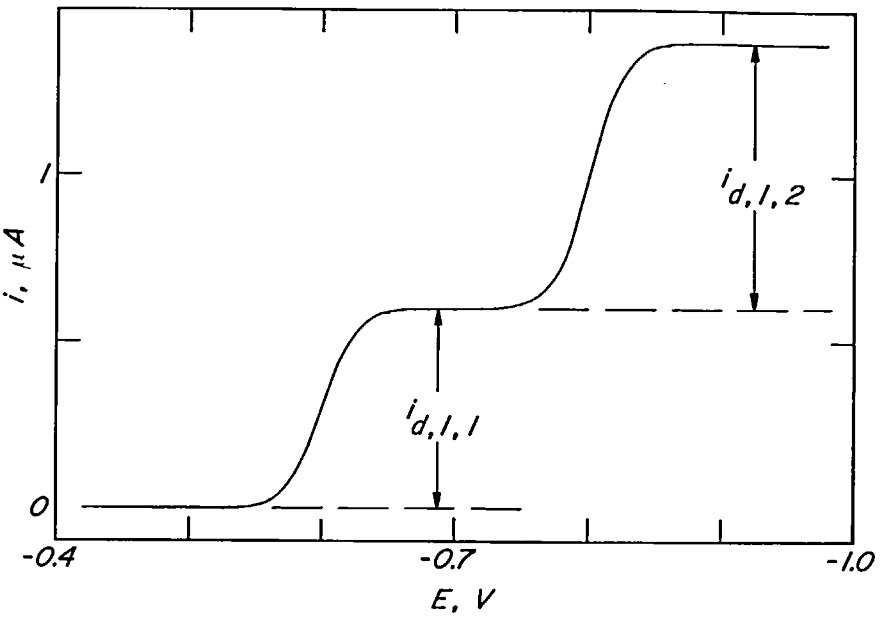


FIGURE 6. Schematic polarogram showing two well-separated waves.

$$i = \frac{i_{d,l}}{1 + \exp\left([E - E_{1/2}]/S\right)} \tag{18b}$$

where, for convenience, $S = 0.434\ S'$, E is the independent variable, i the dependent one, and $i_{d,l}$, $E_{1/2}$, and S are the three independent parameters.

Figure 6 shows a more complicated example, which consists of two waves. Here, the two limiting diffusion currents $i_{d,l,1}$ and $i_{d,l,2}$ would be evaluated separately as shown in the figure, and two separate log plots, one for each wave, would be constructed to provide values of four other parameters: $E_{1/2,1}$, S'_1, $E_{1/2,2}$, and S'_2. The nonlinear regression procedure would be based on the equation

$$i = \frac{i_{d,l,1}}{1 + \exp\left([E - E_{1/2,1}]/S_1\right)} + \frac{i_{d,l,2}}{1 + \exp\left[(E - E_{1/2,2})/S_2\right]} \tag{19}$$

which involves the same six parameters.

Finally, Figure 7 shows the result of including the residual current, which varies linearly with E over the range depicted. The traditional procedure consists of drawing a linear extrapolation of the initial segment of the recorded curve. It is true that the experimenter is unlikely to pay any attention to the slope and intercept of the line, but the assumption that it is linear and the extrapolation nevertheless have the effect of assigning values to these. Now i, $i_{d,l,1}$ and $i_{d,l,2}$ are measured in the fashion shown, and two separate log plots are constructed as before. In all, eight parameters are evaluated, just as in nonlinear regression onto the equation

$$i = \frac{i_{d,l,1}}{1 + \exp\left([E - E_{1/2,1}]/S_1\right)} + \frac{i_{d,l,2}}{1 + \exp\left([E - E_{1/2,2}]/S_2\right)} + b\,(E^* - E) \tag{20}$$

where E^* is the potential at which the residual current is equal to zero, and b is the slope of the residual-current curve. No one with any polarographic experience would

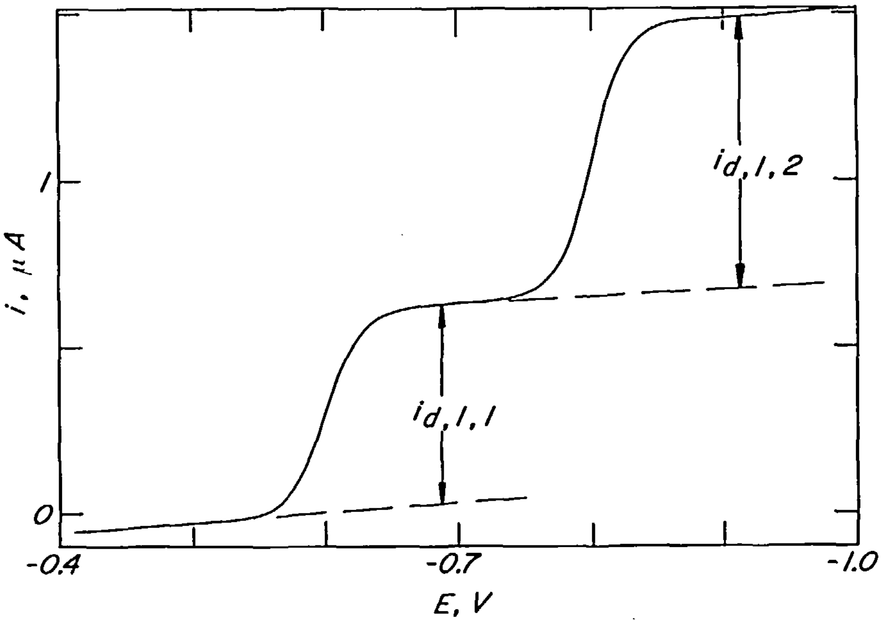


FIGURE 7. Schematic polarogram showing two well-separated waves and differing from Figure 6 by the inclusion of a residual current that increases linearly as the applied potential becomes more negative.

turn a hair at being asked to evaluate these eight parameters from the curve in Figure 7, and it should be no surprise that the same eight parameters can be evaluated in a different way.

The advantage of the nonlinear regression procedure over the graphical one would be modest in dealing with a curve like the one in Figure 7. Values of i at a number of different applied potentials E would have to be read from the curve in either case — unless a digital data-acquisition system had already stored them in the core of a computer, in which event the user of the graphical procedure would have to retrieve them from the computer, whereas the user of nonlinear regression would not. The graphical procedure would entail performing the extrapolation, which may consume a good deal of thought if the initial linear segment of the curve is short; it would also entail constructing two plots and deciding how best to represent each by a straight line. The nonlinear regression procedure would involve typing Equation 20, and possibly also the coordinates of the data points, on a key punch or teletype. This is less time-consuming, and it gives more objective results by eliminating opportunities to make wrong or dubious decisions. Of course, these advantages often have to be weighed against the cost of the computer time involved.

Thus, the two procedures are more or less comparable when they are to be applied to a curve like the one in Figure 7. For a curve resembling that in Figure 8 they are not at all comparable. It is easy to perceive that there are two waves on this curve, and it is also easy to measure the sum of their limiting diffusion currents. However, it is not at all easy to obtain reliable values of $i_{d,l,1}$ and $i_{d,l,2}$ individually; these are needed not only for analytical purposes but also for constructing the log plots needed for the evaluation of $E_{1/2}$ and S'. A log plot for the overall wave would do little more than confirm that there are two waves involved.[13] Ways of going beyond this point have been devised[14-17] but they are so complex, and most of them require such detailed information about the ways in which the currents for the two individual substances depend on potential, that they are rarely, if ever, employed in practical analysis. Worse

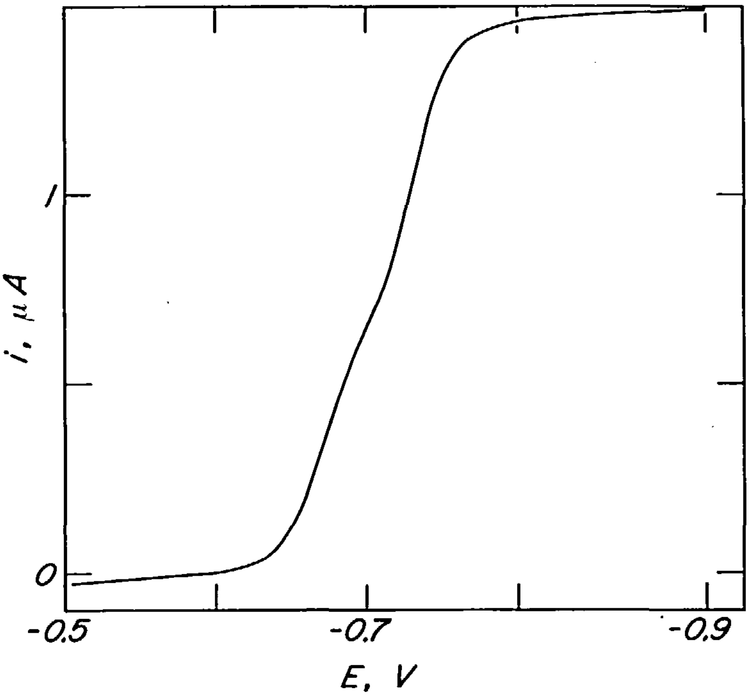


FIGURE 8. Schematic polarogram similar to Figure 7 but with a difference of only 60 mV between the two half-wave potentials, instead of 200 mV as in Figure 7.

yet, the possession of that information implies that the identities of the substances are known and, consequently, that the analyst is helpless if they are not. The difficulty has been largely responsible for the development of a host of other techniques, such as square-wave polarography, in which improved resolution has been sought at the expense of simplicity of instrumentation.

In the same situation, nonlinear regression still involves nothing more than fitting the data to Equation 20 if the values of i have not been corrected for the variation of the residual current with potential, or to Equation 19 if they have.

The straightforward applications of nonlinear regression to chemical problems can be divided, for convenience, into two groups. One group comprises situations in which the data reflect the behavior of a single substance or process and in which the problem is to evaluate a limited number of parameters in an equation so complicated that it would be difficult or impossible to manipulate in any other way. The evaluation of the rate constant contained in Equation 2 is a typical example. The other group comprises situations in which two or more substances or processes contribute to the behavior that is observed and in which the problem is to resolve their individual contributions. Applications belonging to this group usually involve the evaluation of more parameters than those belonging to the first group. The evaluation of the diffusion currents, half-wave potentials, and log-plot slopes for the two overlapping waves in Figure 8 is a typical example.

B. Some Simple Applications

To illustrate the differences among the various kinds of data treatment that I have mentioned, I shall discuss the data obtained in an experiment[18] in which a known constant current was passed through a known electrical resistance immersed in water in one of the two vessels of a differential twin calorimeter. The temperature of the

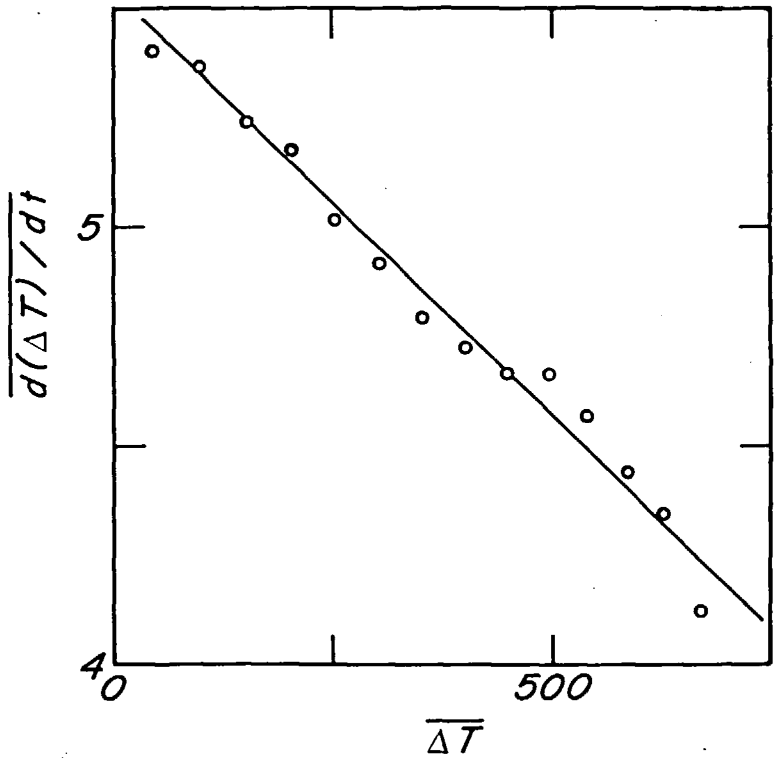


FIGURE 9. Graphical analysis of the data shown in Table 1.

water rose as time elapsed, and the difference ΔT between this temperature and that in a reference vessel was sensed by means of a pair of matched thermistors connected in a Wheatstone-bridge circuit whose output voltage was amplified and presented to a Y-T recorder. The time rate of change of ΔT is described by the differential equation

$$\frac{d(\Delta T)}{dt} = -\epsilon(\Delta T) + \frac{k}{Q} \qquad (21a)$$

in which the value of k can be computed from the known values of the current and resistance. Integration of Equation 21a gives an explicit equation for ΔT as a function of the time t:

$$\Delta T = \frac{k}{\epsilon Q}(1 - e^{-\epsilon t}) \qquad (21b)$$

These equations contain two unknown parameters: the Newtonian heat exchange constant ε, and the effective thermal capacity Q of the contents of the vessel being heated by the flow of current.

Equation 21b is nearly immune to conventional methods of data treatment, but Equation 21a can be used in either of two ways. Addicts of graphical approaches might obtain pairs of values of $d(\Delta T)/dt$ and ΔT by numerical differentiation of the data in the first two columns of Table 1. For example, the average values of $d(\Delta T)/dt$ during the interval $3 \leqslant t \leqslant 13$ is equal to $(71.24 - 17.25)/(13-3) = 5.399$ μdeg s^{-1}, while the average value of ΔT during the interval is equal to $(71.24 + 17.25)/2 = 44.25$ μdeg. On plotting $d(\Delta T)/dt$ against ΔT, one should obtain a straight line having a slope equal to $-\varepsilon$ and an intercept equal to k/Q. Such a plot is shown in Figure 9. There is some scatter of the points, because each of the values of ΔT is somewhat uncertain and

TABLE 1

Data Obtained in Calibrating a Calorimetric Cell

t,s	ΔT, µdeg	t,s	$\overline{\Delta T}$, µdeg	$\overline{d(\Delta T)/dt}$
3	17.25			
		8	44.25	5.399
13	71.24			
		18	98.08	5.367
23	124.91			
		28	151.10	5.239
33	177.30			
		38	203.18	5.175
43	229.05			
		48	254.13	5.016
53	279.21			
		58	303.80	4.919
63	328.40			
		68	352.36	4.792
73	376.32			
		78	399.96	4.728
83	423.60			
		88	446.92	4.664
93	470.24			
		98	493.56	4.664
103	516.88			
		108	539.72	4.569
113	562.57			
		118	584.77	4.440
123	606.97			
		128	628.70	4.345
133	650.42			
		138	671.02	4.121
143	691.63			

Note: Experimental conditions are outlined in the text.

because the uncertainty in each value of $d(\Delta T)/dt$ exceeds the uncertainties in the values of ΔT from which it is obtained. This scatter leads to difficulty in drawing a straight line through the points, but the one shown in Figure 9 might eventually result. Addicts of linear regression, on the other hand, would prefer computing the values of the slope and intercept after a similar numerical differentiation.

The user of nonlinear regression would surely prefer to use the original data together with Equation 21b, and would thereby avoid the loss of precision that the numerical differentiation entails.

Table 2 shows the values of ε and Q that are obtained from these data in these three ways. The graphical procedure, based on Figure 9, yields a slope that is in error by about 5% and an intercept that is in error by about 0.7% as compared to the values obtained from nonlinear regression, whose correctness is shown by the fact that they yield the smallest standard deviation from regression. Linear regression is superior to the graphical procedure because it yields more nearly correct values of both.

At this point, the experimenter might ask how well the values obtained for the parameters served to reproduce the data — or what the internal consistency of the data was, which is really the same question in a different form when there is no uncertainty about the model employed. Using the three different sets of parameters in conjunction with Equation 21b to compute a value of ΔT at each experimental point and then

TABLE 2

Results of Different Procedures for Handling the Data in Table 1

Procedure	Values of the Parameters		Standard deviation from regression (μdeg)
	ε (s^{-1})	Q (cal deg^{-1})	
Graphical	1.80×10^{-3}	40.6_5	1.25
Linear regression	1.909×10^{-3}	40.42	0.96
Nonlinear regression	1.908×10^{-3}	40.35	0.60

obtaining the standard deviation from regression from the differences between these computed values and the measured ones, one obtains the results shown in the last column of Table 2. Although the values of ε and Q obtained from linear regression seem very close to the truth — and, indeed, would probably be indistinguishable from the truth for the majority of purposes — they lead to a serious misrepresentation of the internal consistency that has been achieved. That these values are not quite equal to the best ones is shown by the fact that they lead to a higher standard deviation from regression than the one obtained from nonlinear regression. With the values obtained graphically, the misrepresentation is even more serious. In addition, it would scarcely be possible to obtain reliable estimates of the uncertainties in the values of the parameters from the graphical procedure and something less than straightforward to obtain one from the linear regression procedure.

Reasons for these things can be discerned but will not be discussed here. My point is that nonlinear regression is advantageous, even in a very simple case where alternatives to it can be perceived, not only because it yields the best values of the parameters and the best estimate of the adequacy of the fit, but also because it does so with the least prior algebraic and numerical manipulation.

In many other cases where alternative techniques for handling the data are available, it will be found that the experiments can be greatly simplified by the use of nonlinear regression. A value of the thermodynamic dissociation constant of acetic acid was obtained by MacInnes and Shedlovsky[19] in the following way. They measured the conductances of solutions of acetic acid at various total or analytical concentrations (c_t) and calculated values of the equivalent conductance Λ_{HOAc}. (According to current IUPAC recommendations this should be called the molar conductivity.) Since there was no way in which these could be used to estimate the fraction α of the acetic acid that is dissociated at any particular value of c_t they performed similar measurements with solutions of hydrochloric acid, sodium acetate, and sodium chloride. Extrapolating the equivalent conductance of each of these electrolytes to infinite dilution made it possible to compute the equivalent conductance at infinite dilution for acetic acid

$$\Lambda^0_{HOAc} = \Lambda^0_{HCl} + \Lambda^0_{NaOAc} - \Lambda^0_{NaCl} \tag{22a}$$

A preliminary estimate of the ionic concentrations c_i in a particular solution of acetic acid could then be obtained from the Arrhenius relation

$$c_i = c_t \; (\Lambda_{HOAc} / \Lambda^0_{HOAc}) \tag{22b}$$

This is only approximate because interionic attraction and electrophoresis would cause the equivalent conductance of acetic acid at any finite concentration to be smaller than $^0_{HOAc}$ even if the acid were completely dissociated. Hence equations that represent the dependences of Λ_{HCl}, Λ_{NaOAc}, and Λ_{NaCl} on concentration were used to compute the equivalent conductance $\Lambda_{e,HOAc}$ of hypothetical completely dissociated acetic acid at the concentration c_i, and this was combined with the equation

$$c_i = c_t \quad (\Lambda_{HOAc} / \Lambda_{e,\,HOAc}) \tag{22c}$$

to yield a second and better estimate of c_i. When a value had been obtained that did not change on further iteration of this process, it was used to compute a value of the concentration constant $K_{a,C}$

$$K_{a,C} = \frac{[\mathrm{H}^+]\ [\mathrm{OAc}^-]}{[\mathrm{HOAc}]} = \frac{c_i^2}{c_t - c_i} \tag{22d}$$

Finally, the desired value of the thermodynamic constant K_a was secured by extrapolation to infinite dilution

$$K_a = \lim_{c_i \to 0} K_{a,\,C} \tag{22e}$$

Plainly, this involves a great deal of work, much of it devoted to measurements of the conductances of three electrolytes that were uninteresting in themselves. By nonlinear regression one can evaluate K_a without employing these other data, so that three fourths of the experimental work could have been avoided. The procedure is based on the equations

$$K_a = \frac{a_{H^+} a_{OAc^-}}{a_{HOAc}} = \frac{y_\pm^2\ c_i^2}{c_t - c_i} \tag{23a}$$

$$y_\pm^2 = \exp\left(-2.3432\ c_i^{1/2} / [\,1 + 0.3286\ \mathring{a}\, c_i^{1/2}\,]\right) \tag{23b}$$

in which $y_\pm$ is the mean ionic molarity activity coefficient and a the distance of closest approach, and the Shedlovsky equation[22]

$$\Lambda_{e,\,HOAc} = (\Lambda^0_{HOAc} + B c_i)\ (1 - 0.2289\ c_i^{1/2}) -$$

$$60.19\ c_i^{1/2} \tag{23c}$$

There are four parameters to be evaluated: K_a, $\mathring{a}$, Λ^0_{HOAc}, and the Shedlovsky parameter B. At the start of a calculation the preliminary estimates of K_a and a can be combined with Equations 23a and b to obtain a value of c_i by successive approximation. This can then be combined with the preliminary estimates of Λ^0_{HOAc} and B and with Equation 23c to evaluate $\Lambda_{e,HOAc}$. Finally, Λ_{HOAc} can be computed from Equation 22c for comparison with the experimental value. It is the programmer's responsibility to arrange matters so that the estimated values of the parameters are adjusted, in successive cycles of computations, in ways that will minimize the sum of the squares of the deviations of the calculated values of Λ_{HOAc} from the experimental ones.

On the molarity scale, the result obtained by MacInnes and Shedlovsky[19] was $K_a = 1.753 \times 10^{-5}$, whereas the one obtained from potentiometric measurements by Harned and Ehlers[21] at almost the same time was 1.749×10^{-5}. The difference, though not very large, cannot be attributed to experimental error — even now the literature contains few data as precise as these — and, therefore, it has generally been concealed by expressing the potentiometric results on the molality scale, which gives $K_a = 1.754 \times 10^{-5}$. The nonlinear regression procedure, which would have obviated the necessity for three quarters of the experimental effort, gives $K_a = 1.7488 \times 10^{-5}$ on the molarity scale. Extended comment hardly seems necessary.

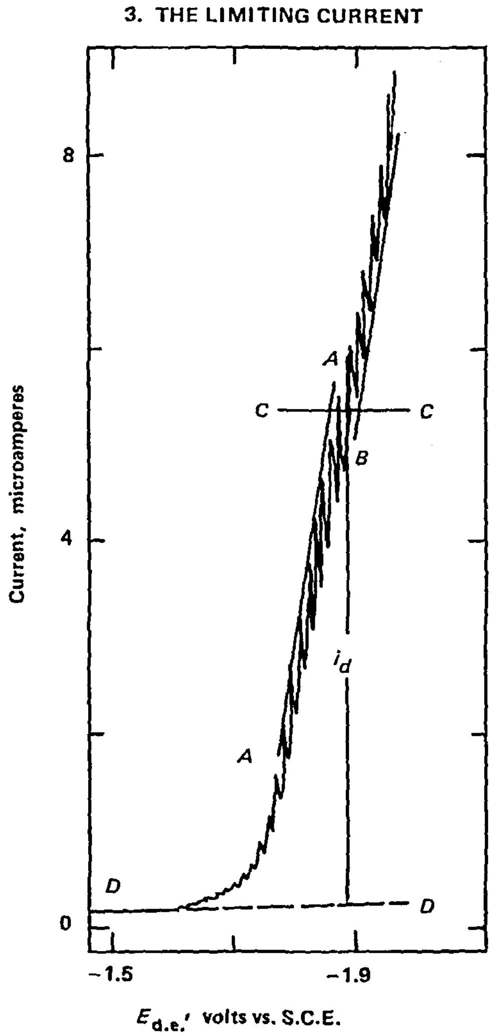


FIGURE 10. Polarogram of 0.9 m*M* vanadium (V) in 0.1 *F* sodium hydroxide. (From Meites, L., *Polarographic Techniques*, 2nd ed., *Interscience*, New York, 1965, 153. With permission.)

Situations in which unnecessary experimental work can be avoided by using nonlinear regression are so common that I shall mention two others. Figure 10[24] shows a polarogram of vanadium (V) in a supporting electrolyte containing 0.1 *M* sodium hydroxide. The plateau is poorly defined because the reduction of sodium ion begins to proceed at an appreciable rate before it is reached. One wishes to evaluate the limiting diffusion current. The lines AA,...,DD in the figure represent one graphical method of doing this. It is not very satisfactory; with a wave as badly defined as this one, it would be unlikely to yield results having a relative precision better than ± 5 to 10%. The traditional approach to this problem has been to record a separate residual-current curve, which would be a polarogram of a similar solution of sodium hydroxide containing no vanadium (V). This would be nearly identical with the curve in Figure 10 at applied potentials around −1.5 V; at slightly more negative applied potentials it would begin to rise exponentially as the rate of reduction of sodium ion increased. One could then subtract the residual current from the total current measured in the presence of vanadium (V) at each of a number of applied potentials. The difference would approach the desired value of the limiting diffusion current as the applied potential became more negative. Unfortunately, its uncertainty would increase at the same time: at very negative potentials it would be a small difference between two large numbers. A relative precision of approximately ± 3 to 5% might be reached in this way. Few polarographers would be sanguine about the reliabilities of the values of the half-wave potential and the Nernstian parameter S (Equation 18b) that might be obtained by conventional logarithmic analysis even after such corrections for the residual current had been made.

The residual current has two components: one that varies almost linearly with the applied potential and is due to charging of the electrical double layers around the successive drops of mercury, and another that varies exponentially (if the applied potential is not too negative) and is due to reduction of sodium ion. The first of these components is represented by line DD in Figure 10. Correction for it can be made easily without recording a separate residual-current curve. The resulting corrected values can then be fitted[23] to the six-parameter equation

$$i = \frac{i_{d,l}}{1 + \exp\left([E - E_{1/2}]/S\right)} + i' \exp\left(k\,[E' - E]\right) \qquad (24)$$

where the first term represents the current due to the reduction of vanadium (V) while the second represents that due to the reduction of sodium ion. In the latter, i' denotes the current due to the reduction of sodium ion at the arbitrary applied potential E'. Analyses of four replicate polarograms in this way gave values of $i_{d,l}$ having a relative standard deviation of only 1.3% and values of $E_{1/2}$ and S having absolute standard deviations of only 6 mV and 1.6 mV, respectively. These standard deviations are a little higher than those (0.6%, 4 mV, and 1.2 mV) obtained from three-parameter fits to Equation 18b after correcting for the residual current by point-by-point subtraction with the aid of separately recorded residual-current curves. However, the values themselves were so nearly identical (e.g., 0.2% in $i_{d,l}$) that the correction, and the additional experimentation that must be done to make it possible, merely save computer time at the expense of operator and instrument time.

One of our articles of faith is that a titration provides a value of the volume of reagent that is needed to reach the end point, but that the amount (or concentration) of the substance with which it has reacted cannot be found unless the concentration of reagent is determined separately. Barry et al.[24] showed that this is untrue. They titrated solutions of potassium acetate with hydrochloric acid in the presence of 3M potassium chloride to keep the activity coefficients and liquid-junction potential constant, and used a glass-electrode pH meter to obtain 12 to 20 points evenly spaced along the titration curve from its start to the point where a 25% excess of acid had been added. The titration curve is described by the equations

$$[\mathrm{H}^+]^2 + \left(K_{a,C} + \frac{V_b^0 c_b^0 - V_a c_a}{V_b^0 + V_a}\right)[\mathrm{H}^+] - \frac{V_a C_a K_{a,C}}{V_b^0 + V_a} = 0 \qquad (25a)$$

and

$$\mathrm{pH} = -\log\left(y_{\mathrm{H}^+}\,[\mathrm{H}^+]\right) \qquad (25b)$$

where V_b° is the initial volume and c_b° is the initial concentration of the acetate solution; V_a, the independent variable, is the volume of acid added; c_a is the concentration of the acid; $K_{a,C}$ is the concentration dissociation constant of acetic acid defined by the first equality in Equation 22d; and y_{H^+} is the apparent single-ion molarity activity coef-

TABLE 3

Results Obtained in Titrations of an Unknown Solution with an Unstandardized Reagent

Parameter	Found	Relative standard deviation of a single value (%)	Actual value	Relative error of the mean (%)
Concentration of OAc$^-$, M	$9.907_2 \times 10^{-3}$	0.19	$9.923_3 \times 10^{-3}$	−0.16
Concentration of HCl, M	0.16918	0.16	0.16914	+0.02

ficient of hydrogen ion. Both $K_{a,c}$ and y_{H^+} depend on the temperature and on the concentration and identity of the supporting electrolyte, and of course $K_{a,c}$ also depends on the identity of the acid being titrated.

Taking V_b^0 as the only one of these parameters whose value is known a priori, Barry et al.[26] used the date obtained in 11 titrations to evaluate the other four parameters. The results, which are shown in Table 3, should be considered in the light of the facts: that acetate ion is so weak a base, and these solutions are so dilute, that the relative standard error of locating the point of maximum slope is approximately 0.9% and that this point precedes the equivalence point by 0.4%. Both the accuracy and the precision of a determination of acetate by titration with unstandardized acid are better than can be achieved in the ordinary way. The accuracy and precision of a standardization of hydrochloric acid by titration of an unknown amount of acetate are at least as good as can be achieved by separate titrations of accurately measured amounts of pure substances whose properties make them suitable for use as primary standards. It is therefore a mere waste of time, effort, and materials to standardize the reagent separately.

These examples show that the use of nonlinear regression can make it easier to use data to obtain the values of physical quantities that those data were intended to provide and also that it can make it easier to obtain the data by decreasing the amount of experimentation that must be performed. Finally, it can enable the experimenter to design experiments of new types or to obtain kinds of information that have been impossible to extract from experiments of familiar types.

Multiple-standard-addition potentiometric analysis with ion-selective electrodes belongs to the first of these two classes.[25] The electrode is immersed in V^0 cm^3 of a solution containing an unknown concentration (c^0 M) of the ion to which it responds, and its potential is measured. Some volume ($V_{s,i}$ cm^3) of a solution containing a known concentration (c_s M) of the ion is added, the potential is measured again, and this process is repeated until n additions have been made. Assuming the potential to be described by the equation

$$E = E^{0\prime} + S \log_{10} c \tag{26a}$$

where c is the total concentration of the ion corresponding to the measured potential E, one can obtain

$$E = E^{0\prime} + S \log_{10} \frac{V^0 c^0 + \sum_{i=1}^{n} V_{s,i} c_s}{V^0 + \sum_{i=1}^{n} V_{s\,i}} \tag{26b}$$

Equation 26b can be handled by Gran's graphical method if the value of S is known[26] but both this value and that of the formal potential $E^{0'}$ are apt to vary from one unknown to another unless the compositions of their supporting electrolytes are identical. By nonlinear regression Brand and Rechnitz[25] achieved a relative error of only 0.8% in the analysis of an 0.01 M solution of chloride, using a chloride-ion selective electrode for which S was simultaneously found to be equal to −58.06 mV as against the theoretical value of −59.15 mV. Isbell et al.[27] employed an electrode selective for silver ion to determine this ion both by standard addition and by titration with standard chloride, treating the data with a nonlinear regression program especially written for the purpose. (Unlike all the other programs used in the work described in this review, this program employed a procedure originally devised by Deming[28] and Wentworth[29,30] for taking into account the effects of errors in the independent volume and other quantities — such as the concentration of a separately standardized reagent — that are commonly treated as constants free from error.) They obtained results about equally accurate by the two methods of determination: the uncertainty of the result of the standard-addition method was higher, but this could be attributed to a difference in the experimental procedures employed.

Like many other conclusions based on these techniques of data treatment, this is surprising and, indeed, even contradictory of beliefs so long and firmly held that they have become almost instinctive. It is common knowledge that the accuracy and precision of direct potentiometry are inferior to those of potentiometric titration; that potentiometric titrations are necessarily based on stoichiometric reactions; that they cannot succeed unless there is a reasonable jump of potential around the equivalence point; and that they entail knowing the concentration of the reagent in advance. The first two of these four articles of faith have been shown to be unnecessary by Brand and Rechnitz[25] and by Isbell et al;[27] Barry and Meites[31] disproved the third by achieving ing results of reasonable accuracy and precision in titrations of acetate ion with hydrochloric acid at concentrations so low that the titration curve shows no inflection point; and, as was discussed above, Barry et al.[24] demolished the fourth by showing prior standardization to be unnecessary. Clearly the whole field of potentiometry and potentiometric titrimetry needs re-examination, for its scope and limitations must be very different from those generally attributed to it.

A fundamental credo of similarly long standing in the field of polarography is that it is impossible to measure the individual heights of two waves on a single polarogram unless the two half-wave potentials are fairly well separated. The exact meaning of "fairly well separated" has been formulated differently by different authors, and special techniques have been devised[14-17] that will yield results of moderate accuracy even if the difference between the two half-wave potentials is as small as 40 mV. Most of these techniques demand that the characteristics of the substances giving rise to the waves be accurately known from prior experiments with separate solutions under exactly the same conditions as those used in obtaining the polarogram of their mixture. To perform such experiments, one must know the identities of the two substances. Much of the literature of polarography has been devoted to this problem, and a number of techniques related to polarography — including, among others, linear-sweep voltammetry and square-wave polarography — owe some of the attention they have received to the belief that they would prove advantageous in dealing with it. By nonlinear regression onto Equation 15, Meites and Lampugnani[23] were able to secure values of the limiting diffusion currents that were within 5% of the truth for a mixture of two substances having half-wave potentials that are identical within 1 mV. Without any prior calibration data whatever, they also secured values of the two half-wave potentials and the two Nernstian factors that were virtually identical with the ones reported in the prior literature, and that consequently would have served for the identification of the two substances if these had been unknown. This was an extreme and somewhat artificial case, because the two substances gave very different values of S. More often, neither the values of S nor those of $E_{1/2}$ would be much different, and then nonlinear regression would yield results too uncertain to be of any practical use. It would be foolish to suppose that nonlinear regression will provide the answer to every problem. It would be equally foolish to ignore its ability to provide answers to problems that other techniques cannot solve.

Of the same general type is the demonstration by Ingman et al.[32] that mixtures of formic and acetic acids, formic and propionic acids, and acetic and propionic acids can be analyzed for their individual constituents with satisfactory accuracy by applying nonlinear regression to data secured in potentiometric titrations with standard base. Mixtures of acetic and propionic acids are the most interesting because the values of pK_a for these acids (4.757 and 4.874, respectively, at 25°C) differ by less than 0.12 unit. Consequently, there is no inflection around the equivalence point for the titration of acetic acid, and, in classical terms, it is impossible to resolve the two overlapping stages of neutralization. With solutions containing 0.2 mmol of acetic acid and 0.8 mmol of propionic acid, three titrations gave relative mean errors of +4.5% in the amount of acetic acid, of −1.2% in the amount of propionic acid, and of −0.07% in their sum.

The following paragraphs are addressed to those who have had some experience with regression techniques and are interested in ways of applying them. Readers chiefly interested in their possible utility may turn directly to Section III.C.

It stands to reason that decreasing the number of parameters being evaluated, by employing previously determined values of some of these, will increase the accuracy that is attained in evaluating the adjustable ones. Unfortunately, there are propositions that are no less self-evident but are nevertheless untrue, and it is therefore necessary to inquire into the truth of this one. The evidence bearing on its truth includes some observations by Barry and Meites.[31] These authors performed potentiometric titrations of acetate ion, at concentrations ranging from 0.10 to 6.5×10^{-5} M, with hydrochloric acid of known concentrations. They began by fitting their data to Equation 25, evaluating three parameters: the concentration (c_b^0) of acetate ion being titrated, the concentration dissociation constant($K_{a,c}$) of acetic acid, and the apparent single-ion molarity activity coefficient of hydrogen ion y_{H^+}. As long as c_b^0 exceeded $2.5 \times 10^{-3}M$, the values obtained for it were accurate and precise to a few hundredths of a per cent, while the values obtained for $pK_{a,c}$ (= 4.883 ± 0.003) and $\log y_{H^+}$ (= 0.239 ± 0.005) were constant within a few thousandths of a unit and independent of c_b^0. At lower concentrations, the computed values of c_b^0 exceeded the expected ones: for example, solutions known to contain 6.45×10^{-5} M acetate were found on the average to contain 1.19×10^{-4} M acetate. At the same time, the computed values of $pK_{a,c}$ and $\log y_{H^+}$ became dependent on c_b^0, changing from the ones quoted above to 4.412 and 0.283, respectively, with 6.5×10^{-5} M acetate. On the ground that the best values of these two parameters were those independent of concentration obtained at the highest concentrations, the data were again subjected to one-parameter fits in which $K_{a,c}$ and y_{H^+} were taken to be known constants, so that c_b^0 was the only unknown quantity being evaluated. These gave slightly poorer results at the highest concentrations, but yielded an average of 5.3×10^{-5} M for the concentration of the 6.45×10^{-5} M solution and thus decreased the absolute value of the relative error from 84 to 19%. On this basis they recommended replacing as many as possible of the adjustable parameters with constant values deduced from experiments under the most favorable conditions.

Since then, Murtlow[33] and Murtlow and Meites[34] also used $3M$ potassium chloride as a supporting electrolyte in potentiometric titrations of mixtures of hydrochloric and acetic acids with standard base, found it to contain a basic impurity, and identified this as hydrogen carbonate ion. It was present at a concentration of approximately 6.1

$\times 10^{-5}$ M, surprisingly close to the error in the results of the three-parameter fits performed by Barry and Meites[31] for the most dilute solutions they titrated. Both the erroneous values of c_b^0 and the drifts of $\mathrm{p}K_{a,c}$ and y_{H^+} at the lowest concentrations are probably attributable to the presence of this impurity, and it is probably accidental that better values of c_b^0 were obtained by using constant values of the other two parameters. With Murtlow's data we were unable to detect any improvement resulting from this procedure and I believe it is better not to constrain the fits under extreme conditions. Instead, I recommend very critical comparison of the values obtained in unconstrained fits with the more reliable ones corresponding to more favorable conditions. One may expect the uncertainties to increase as the conditions become less favorable, but systematic variations should be taken as a warning that the data reflect some phenomenon that is not properly described by the equations used.

I have made these remarks because Ingman et al.[32] employed a procedure about which I have some reservations. They analyzed their data on "unknown" mixtures of acetic and propionic acids by using the data obtained with similar "known" mixtures. The data for the "known" mixtures were subjected to two-parameter fits in which the two concentrations were taken as known and the values of $\mathrm{p}K_{a,c}$ for the two acids were calculated. These values were taken as fixed constants in two-parameter fits performed to evaluate the concentrations of the two acids in "unknown" mixtures. This would be unwise if there might be any difference — e.g., of temperature — between the conditions under which the "known" and "unknown" solutions were titrated of which the actual effect was to alter the values of $\mathrm{p}K_{a,c}$ to some extent. If these were constrained, the computed concentrations would have to be adjusted to compensate for that effect, and consequently they would be surely be erroneous. A striking peculiarity of the results is that the values of $\mathrm{p}K_{a,c}$ were not the same in mixtures as they were in separate solutions of the acids: for acetic acid $\mathrm{p}K_{a,c}$ was found to be equal to 4.501* when this acid was present alone but equal to 4.489 when propionic acid was present as well; for propionic acid the values were 4.649 in pure solutions and 4.681 when acetic acid was also present. The presence of either acetic acid or propionic acid appeared to decrease the acidic strength of formic acid by 0.006 pK-unit; however, the presence of formic acid appeared to increase the acidic strength of propionic acid by 0.019 pK-unit while it decreased that of acetic acid by 0.009 pK-unit. These differences were ascribed to "possible interactions between the ... acids." Different kinds of possible interactions can be imagined, although none of them would be described exactly by mere alterations of the values of $\mathrm{p}K_{a,c}$. What is surprising is that any of them should be so extensive in such dilute solutions as to have measurable effects. If the phenomenon is real, its importance would merit a much more thorough investigation. By revealing the nature of the interaction, this would provide clues to its exact description, and this would doubtless yield better results than the approximate description employed.

This discussion might be summarized by what I call Potter's Maxim:[35] "know what you are doing before you do it so that you know you are doing what you are doing." That is of course facetious, but a scientist should always know what he is doing before he does it, and nowhere is the necessity more imperative than when data are being fitted to a theoretical equation. If the wrong equation is employed, the wrong results will be obtained; this will be true no matter how much tinkering is done with the fit.

For example, the open circles in Figure 11 show data points that might be obtained from the polarogram of a substance that underwent reversible reduction to a product insoluble in mercury. To eliminate uncertainties arising from experimental errors and

---

* The numerical values given here pertain to concentration constants and are therefore not directly comparable with those given above for the thermodynamic constants.

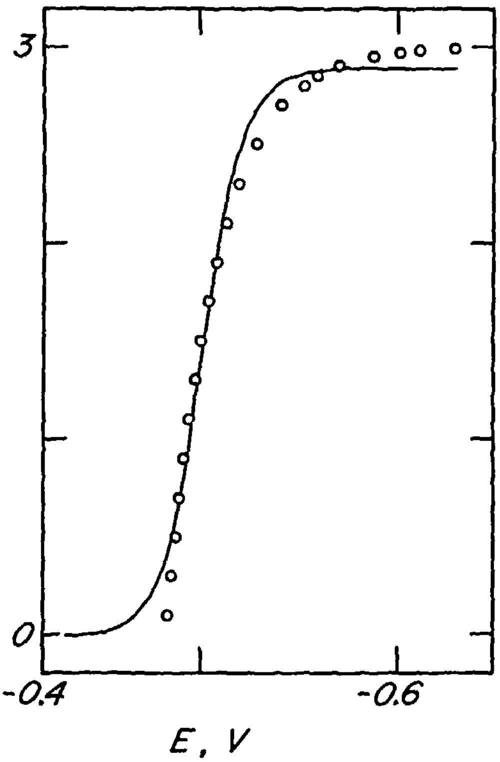


FIGURE 11. The open circles represent points on the polarogram of a substance that is reversibly reduced to give a product insoluble in mercury. Values of the parameters used to calculate these data from Equation 27 are given in the text. The solid curve represents the best three-parameter fit to Equation 18b that can be obtained.

deviations from strictly Nernstian behavior, these "data" were calculated from the familiar equation

$$E_{de} = E_{1/2} - \frac{0.05915}{n} \log \frac{i_{d,l}}{2} + \frac{0.05915}{n} \log (i_{d,l} - i) \qquad (27)$$

when n is the number of electrons involved in the half-reaction; $i_{d,l}$ and $E_{1/2}$ are the limiting diffusion current and the half-wave potential, respectively; and i is the measured current at the applied potential E_{de}. The values $i_{d,l}$ = 3.000 μA, $E_{1/2}$ = −0.5000 V, and $n = 1$ were assumed. A three-parameter fit to Equation 18b, assuming all of these to be unknown, gave $E_{1/2}$ = −0.5011 V. This is wrong, and the values obtained for $i_{d,l}$ (= 2.877 μA) and S' (= 10.50 mV) are also wrong, because Equation 18b misrepresents the behavior of the system. It assumes the half-reaction not only to be thermodynamically reversible but also to yield a product whose activity varies with the current because it is soluble in either the mercury or the liquid phase surrounding it; the second of these assumptions is untrue for data calculated from Equation 27. Section IV will describe a powerful way of recognizing that Equation 27 would provide a better fit to these data than Equation 18b. My point here is that a wrong value of $E_{1/2}$ will be obtained from a fit to Equation 18b no matter how that fit is effected. The three-parameter fit described above apportions the inapplicability of Equation 18b among all three of these parameters evaluated, with the result that although the resulting value of $E_{1/2}$ is wrong, it is not very far from the truth. A two-parameter fit to the same equation, using the fixed value $i_{d,l}$ = 3.000 μA, that might be measured independently at potentials on the plateau of the wave, gives $E_{1/2}$ = −0.5025 V and S = 11.65 mV: the error in the former is much larger than it was in the result of the three-parameter fit because the incorrectness of Equation 18b can now be divided between only two parameters. One must be sure that the equation being employed provides a description of the system that is complete and correct within the limits of error of the data.

C. Pointwise Variance Analysis

Pointwise variance analysis is an outgrowth of ordinary nonlinear regression analysis. It is designed to provide the answers to several important questions:

1. How reliable are the values of the parameters that have been obtained from any particular fit?
2. How should the data points be distributed to maximize the reliability of the information that is obtained, while minimizing the experimentation and measurement that must be performed?
3. What precision of measurement must be achieved to ensure that the calculated values of the parameters will have any desired degree of reliability?

The length of this section will reflect the importance of these questions rather than the volume of the chemical literature that has been devoted to them. The first question is of course a common one: it is easy to answer when linear regression can be employed to find the values of the parameters, and hundreds of examples might be cited. However, they are all irrelevant to nonlinear regression, for which Sillén once said that there was no general way of evaluating the uncertainties in the parameters. For the relationship

$$y = f(x, V_1, V_2, \ldots V_j) \tag{28}$$

where V_i denotes the ith one of a total of j parameters, the standard deviation of V_i is given by

$$\sigma^2_{V_i} = \sum_{m=1}^{n} \left(\frac{\partial V_i}{\partial y_m} \right)^2 \sigma^2_{y_m} \tag{29}$$

where y_m is the measured value of the dependent variable at the mth one of the total of n data points obtained, and σy_m denotes the standard error in y_m. Assuming that the data have been fitted to the correct equation, σ_{y_m} may be equated to the standard deviation from regression. One may define the finite-difference approximation $e_{i,m}$ to the partial derivative appearing in Equation 28:

$$e_{i,m} = \frac{V'_i - V_i}{y'_m - y_m} \tag{30}$$

where V_i denotes a value obtained by straightforward nonlinear regression when all the y_m have their measured values, and V'_i denotes the corresponding value obtained in the same way after one of the y_m has been changed to a slightly different value y'_m. The values of σ_{V_i} may be obtained by introducing a known error into y_1 (the first of the successive values of y_m, computing the resulting values of all the V_i, storing the corresponding values of $e_{i,1}$, returning y_1 to its original value and introducing a similar error into y_2, and so on until all the points have been treated. It is then possible to

---

* The numerical values given here pertain to concentration constants and are therefore not directly comparable with those given above for the thermodynamic constants.

sum all the values of $e_{1,m}$ to obtain the standard deviation of V_1 from Equation 29 and to identify the largest value $e_{1,m_{max}}$ that has been encountered at any point. This makes it possible to calculate the value of the normalized quantity $E_{1,m}$ defined by the equation

$$E_{1,m} = 100\, e_{1,m} / (e_{1,m})_{max} \qquad (31)$$

for each point. The standard deviation of V_2 and the values of $E_{2,m}$ can be obtained at the same time, and so on for all of the *j* parameters involved.

This is pure brute force, and it is not economical of computer time. If twenty data points have been obtained and fitted to a four-parameter equation, twenty additional similar fits have to be performed to provide answers to the questions raised at the beginning of this section. This is so time-consuming that one can hardly afford to do it routinely. After describing the information that can be obtained from such calculations, I shall therefore explain briefly how their number can be kept to a reasonable minimum.

Suppose, for the sake of a concrete illustration, that one wishes to find the half-wave potential $E_{1/2}$ and the Nernstian factor *S* for a reversible polarographic wave that is sufficiently well defined to enable the experimenter to measure its limiting diffusion current separately at a potential on the plateau. To minimize minor irregularities in the pointwise-variance-analysis plot (which will be defined and discussed below), it is advisable to begin with a set of synthetic data computed from the equation (Equation 18b) to which the data will be fitted. An interactive program designed to perform the calculations described above will request a value of the standard deviation from regression or an estimate of the standard error of measurement, and a round number may conveniently be provided. We may imagine that the data consist of thirteen points, equally spaced over the range from −0.54 to −0.66 V and computed with $i_{d,l}$ = 1.000 μA, $E_{1/2}$ = −0.6000 V, and S = 12.84 mV, which corresponds to a reversible two-electron process yielding a soluble product at 25°C, and we may take σ_{i_m} = 0.01 μA. The results will be $\sigma_{E_{1/2}} = 2.8 \times 10^{-4}$ V and σ_S = 0.26 mV.

As long as the number of data points, the way in which they are distributed, and the values of the parameters do not change, Equation 29 shows that each of these standard deviations is proportional to the standard deviation from regression. If, in a subsequent fit to a similar set of experimental data, it is found that the standard deviation from regression is, say, 0.0080 μA (as compared with the value σ_{i_m} = 0.01 μA that was used to obtain the above figures), it may be concluded that the standard deviation of the value of $E_{1/2}$ is equal to $(0.008/0.01) \times 2.8 \times 10^{-4} = 2.2 \times 10^{-4}$ V. This result can be interpreted with the aid of ordinary statistical considerations.

Let us turn from the first question to the third. Suppose that measurements like these are to be made routinely, perhaps by using a data-acquisition routine that will store values of the current and applied potential at 10-mV intervals. If thirteen such points are used in evaluating $E_{1/2}$, what standard error of measurement can be tolerated if the standard deviation of $E_{1/2}$ is not to exceed 0.5 mV? The answer follows directly from the proportionality between $\sigma_{E_{1/2}}$ and σ_{i_m}: it is equal to $(5.0 \times 10^{-4}/2.8 \times 10^{-4}) \times 0.01 = 0.018$ μA. The form of Equation 18b is such that it is really the value of the ratio $i/i_{d,l}$ that is being computed at each point, so that the same result would be obtained if both the limiting diffusion current and the standard deviation from regression were changed by identical factors — for example, to 5 and 0.09 μA, respectively.

These things are simple and straightforward: the answer to the second question is less so. Figure 12 is a "pointwise variance-analysis plot." Curve a shows how the values of $E_{i,m}$ depend on the applied potential (i.e., on the quantity x_m, where *x* is the independent variable in the experiment), if *i* denotes the parameter that corresponds to the half-wave potential. The curve is drawn on a semilogarithmic scale for conven-

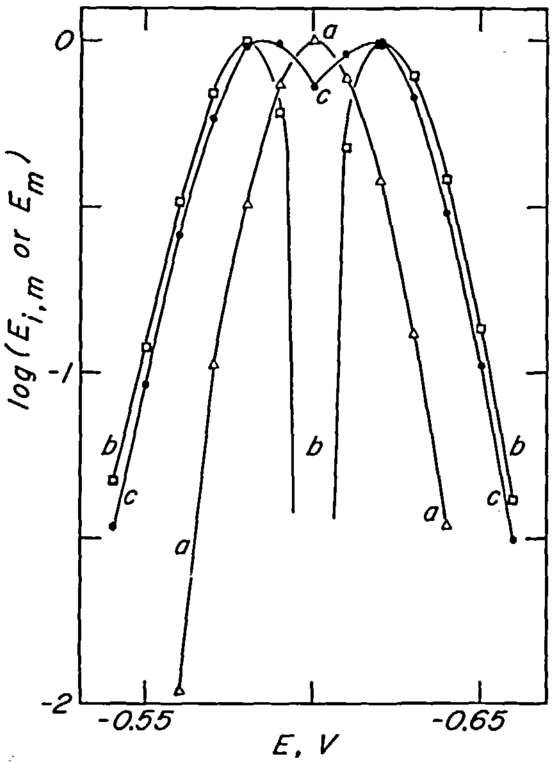


FIGURE 12. Pointwise variance-analysis plot for the evaluation of $E_{1/2}$ and S for a polarographic wave by a two-parameter fit to Equation 18b, taking the value of $i_{d,1}$ as known from separate measurements on the plateau of the wave. Curve a (triangles) shows how the relative weight $E_{i,m}$ varies with the applied potential E for the evaluation of $E_{1/2}$; Curve b (squares) is a similar plot for the evaluation of S; Curve c (solid circles) shows how the overall weight E_m varies with the applied potential on the assumption that the values of $E_{1/2}$ and S are of equal interest.

ience. At both the most positive and the most negative applied potentials, $E_{i,m}$ is very small. In these regions, the value of $E_{1/2}$ is hardly affected by errors in the measured currents, which means that data in these regions provide very little information about $E_{1/2}$. As the applied potential approaches $E_{1/2}$ from either direction, the value of $E_{i,m}$ increases, and it reaches a maximum at the half-wave potential. Curve b is a similar plot for the Nernstian parameter S. At the half-wave potential, the value of $E_{i,m}$ for this parameter is extremely small: an error in the measured current at this point affects the half-wave potential but not the Nernstian parameter. At points slightly removed from the half-wave potential, the values of $E_{i,m}$ pass through two maxima. Because the polarographic wave itself is symmetrical around its half-wave potential, both these curves are nearly symmetrical around the half-wave potential. At points still farther removed from the half-wave potential, the value of $E_{i,m}$ decreases again because the uncertainties in values of the current that are either very small or very nearly equal to the limiting diffusion currents cause these values to be nearly useless in evaluating S.

Readers who have some acquaintance with polarography can rationalize these arguments with the aid of Figure 13, which is a conventional log plot based on Equation 18a. The open circles represent "data" calculated from the same equation, and the horizontal bars through alternate open circles represent the ranges of values of log (i/

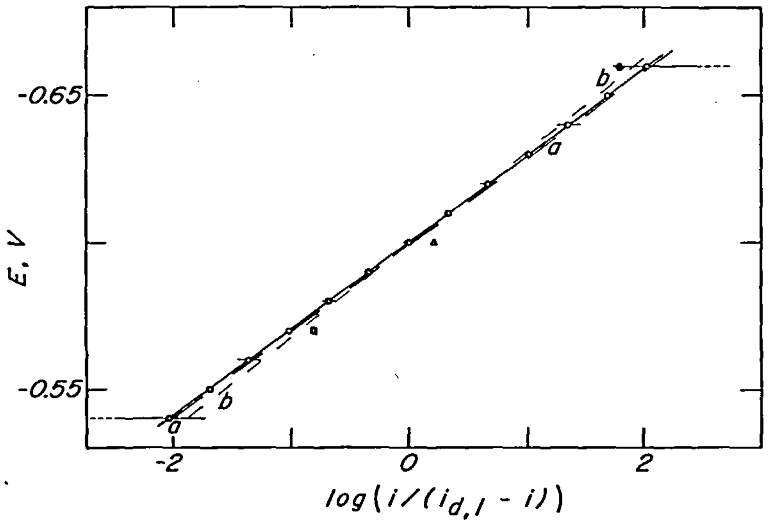


FIGURE 13. The open circles and the solid line show a conventional log plot for the evaluation of $E_{1/2}$ and S from polarographic date; the horizontal lines through alternate points represent the uncertainties in log $(i/[i_{d,l}-i])$ resulting from uncertainties of 0.01 μA in the values of i. The triange (at $E = -0.60$ V) corresponds to an erroneous value of i at the half-wave potential, and the dashed line aa shows the effect that this might have on the plot: its intercept is affected, but its slope is not. The square (at $E = -0.57$ V) corresponds to an erroneous value of i at a potential slightly different from the half-wave potential, and the line bb shows the effect of this error: both the slope and the intercept are affected. The diamond (at $E = -0.66$ V) corresponds to an erroneous value of i at a potential far removed from the half-wave potential: this has little effect on the plot because the uncertainty in log $(i/[i_{d,l}-i])$ is very large in this region.

$\{i_{d,l}-i\})$ that correspond to an uncertainty of ± 0.01 μA in the value of i. The slope of the solid straight line drawn through the open circles is equal to S' (= 2.303 S), and its intercept at $\log(i/\{i_{d,l}-i\}) = 0$ (which corresponds to $i = i_{d,l}/2$) is equal to $E_{1/2}$. An erroneous measured value of i at a potential equal to $E_{1/2}$ will replace the open circle at that point by the open triangle, and will cause the solid line to be replaced by the dashed one. This has the same slope as the solid line, but it has a different intercept. An erroneous measured value of i at a potential slightly removed from $E_{1/2}$, as represented by the open square, will affect both the slope and the intercept, although the error in the intercept will be smaller than if the same error of measurement were incurred at the half-wave potential. This is because the error of measurement can now be partitioned between the two parameters instead of affecting the half-wave potential alone. An erroneous measurement at a potential far removed from $E_{1/2}$, as represented by the diamond, will have very little influence on the line that is drawn through the points.

On this basis, one can identify the region (or regions) in which the measurements should be concentrated so as to provide the most precise value of a particular parameter. Using thirteen points equally spaced along the applied-potential axis from the point at which $i = 0.01\ i_{d,l}$ to the point at which $i = 0.99\ i_{d,l}$ leads to a standard error of 0.28 mV in the half-wave potential, if the standard error in each value of i is equal to 0.01 i_d. What must the experimenter do to achieve a standard error of 0.20 mV in the half-wave potential, if the value of S is of no interest whatever? There are three approaches. One, which was suggested by the discussion in an earlier paragraph of Section III.C., is to decrease the standard error of $E_{1/2}$ by the desired factor (0.20/

0.28) by decreasing the standard error of measurement by the same factor: that is, from 0.01 to 0.01 × (0.20/0.28) = 0.007 μA. This is not a very large decrease, but it might be very difficult to achieve. Another approach consists of saying that the precision of the result can be improved by increasing the number of data points on which it is based. A standard error of 0.28 mV in $E_{1/2}$ was attained with thirteen points or eleven degrees of freedom; a standard error of 0.20 mV should be possible to attain by increasing the number of degrees of freedom to $(0.28/0.20)^2 \times 11 = 22$, which corresponds to 24 data points. This is very nearly twice the original number.

The third approach will be adopted by those who recognize that, of the 11 points that would be added in the second approach, those near the extremes of this range have very little influence on either of the two parameters involved in the fit and that those more than about 11 mV away from the half-wave potential (corresponding to the points of intersection of curves a and b in Figure 13) influence the value of S more than they do that of $E_{1/2}$. Clearly, the most useful points are those clustered in the immediate vicinity of $E_{1/2}$, and a few points concentrated in that region should serve the experimenter's purpose just as well as, if not better than, a great many obtained from regions that contain little of the desired information. On this basis, one might go on to calculate that even nine points having the same precision as the original ones will yield a value of $E_{1/2}$ having a standard error of only 0.18 mV if they are taken at 1-mV intervals from −0.596 to −0.604 V. This is an even larger improvement than was being aimed at. It might be necessary to redesign the experimental procedure or apparatus to permit making measurements so closely spaced, but the trouble that this involved would be at least partly compensated for by the decrease in the number of measurements involved.

Similarly, Curve b in Figure 13 shows that, if S is the only parameter in which the experimenter is interested, the measurements should be concentrated in the regions very close to $E_{1/2} + 20$ mV (where $i/i_{d,l}$ is approximately equal to 0.18) and to $E_{1/2} - 20$ mV (where $i/i_{d,l}$ is approximately equal to 0.82). More generally, the regions of greatest weight in evaluating S are centered around $E_{1/2} + 1.6\,S$ mV and $E_{1/2} - 1.6\,S$ mV.

It often happens that an experimenter is more or less equally interested in the values of all the parameters involved in a fit. When this is true, it is convenient to define a relative overall weight for the mth point by the equation.

$$E_m = \frac{1}{j} \sum_{i=1}^{j} E_{i,m} \tag{32}$$

Curve c in Figure 12 shows the relative overall weight E_m depends on the applied potential for the experiment described above. As is also true for all the other curves in this section, it is actually the logarithm of the relative weight that is plotted along the ordinate axis. Points for which the weight is less than 1% of that at the maximum are omitted because they are not useful enough to deserve consideration. The variation of E_m is small over the range $E_{1/2} + 30$ mV to $E_{1/2} - 30$ mV; outside that range, the values of E_m fall off precipitously. If both parameters are of interest, the data should be evenly distributed over this range, which corresponds to $0.1 \leqslant i/i_{d,l} \leqslant 0.9$.

This example was selected because it leads to a familiar conclusion: those who are accustomed to evaluating $E_{1/2}$ and S by constructing log plots know from experience that points outside this range are much less useful than those inside it. In other situations, a little thought may be required to perceive the wisdom of the guidance provided by pointwise variance analysis. Suppose that one is attempting to evaluate the rate constant k of a pseudo-first-order reaction by following the absorbance A of the reactant. The data should conform to Equation 5b. One may assume that $k = 0.01$ sec^{-1}, that the initial absorbance $A_0 = 0.7$, that the standard error of a single measurement

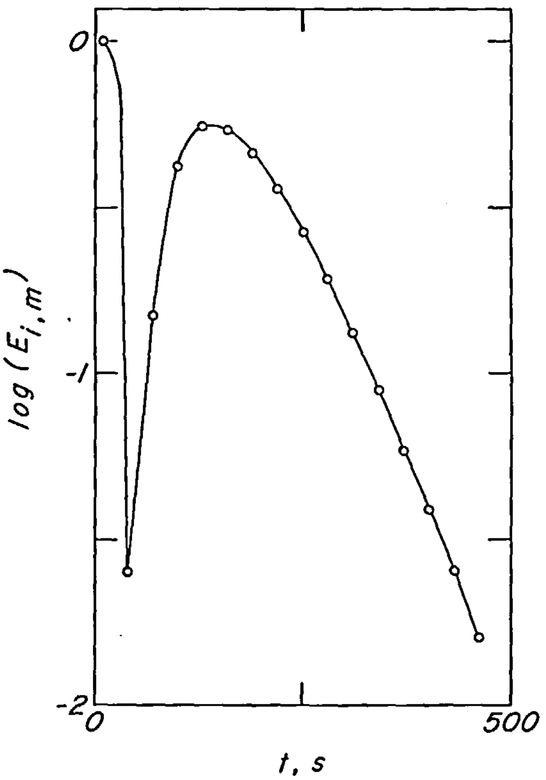


FIGURE 14. Pointwise-variance-analysis plot for the evaluation of a rate constant in the manner described in the text, showing how the relative weight $E_{i,m}$ varies with the time for the evaluation of k.

of absorbance is equal to 0.01, and that sixteen points are acquired at 30-sec intervals from t = 10 s to t = 460 s (when A/A_0 = 0.01, so that the reaction is 99% complete). According to these assumptions, the standard error in k will be equal to $2.4 \times 10^{-5}s^{-1}$, or 0.24%*. Figure 14 shows the dependence of $E_{i,m}$ on time for the rate constant; the corresponding curves for A_0 and E_m are omitted because the value of A_0 is unlikely to hold much interest. Points obtained toward the end of the experiment have little weight because the relative error of measurement becomes very large as the reaction nears completion. In interpreting the remainder of the plot, it is convenient to think of k as being numerically equal to the slope of a plot of ln A against t. If the data did not become unreliable at long times, points near the beginning and end of the experiment would have about equal weights in determining this slope whereas points near the middle of it would have much less weight. This would lead to a plot that was nearly symmetrical and ╲╱-shaped. Because the data do become unreliable, their weights on the right-hand branch decrease at very long times and this opposes, and eventually overbalances, the increasing influence that points farther away from the middle of the line would exert on its slope if they were equally reliable. Consequently the plot becomes asymmetrical. There are two regions in which the points are really useful: one is at the very beginning of the experiment, while the other extends from about 50 to 90% completion, and it is evidently in these regions that the data should be clustered.

---

* Perhaps the least certain of the tacit assumptions embodied in this value are that mixing and the attainment of thermal equilibrium are instantaneous. It would certainly be extraordinary for values of k obtained in different experiments to exhibit precision of this order.

This discussion has concentrated on examples in which the shapes of pointwise-variance-analysis plots are relatively easy to explain and understand, but of course there are many situations in which this is not at all easy, and it is in these that the technique will be most valuable. If a polarographic wave is recorded with a cell having a high resistance (R) and if the limiting diffusion current is measured separately on its plateau, R can be evaluated along with $E_{1/2}$ and S in much the same way as in the first example in this section. It turns out that points close to the half-wave applied potential are nearly useless in evaluating $E_{1/2}$, which is exactly contradictory to intuitive expectation, and that points of greatest utility for this purpose lie in three regions: one around $i/i_{d,l} = 0.3$, another around $i/i_{d,l} = 0.8$, and the third at applied potentials close to the plateau.

The relative weights calculated in this way are entirely different from the weighting factors described in connection with Equation 17. The latter are concerned with the relationships between uncertainties in the measured values of the dependent variable and uncertainties in another variable that happens to be more convenient in linear regression, whereas the relative weights are concerned with the relationships between uncertainties in the measured values and uncertainties in the values of the parameters calculated from them.

Calculations of relative weights can be made in linear as well as nonlinear regression and have been made by a number of authors including Isbell et al.[29] and Sato and Momoki,[38] but their a priori utility in deciding the questions raised at the beginning of this section was recognized only quite recently.[39] Used in the ways outlined here, pointwise variance analysis should provide important help in designing experiments and improving their efficiencies.

IV. DEVIATION PATTERNS

A. Introduction

If a solution of acetic acid is titrated with standard sodium hydroxide, the amount of acetic acid can be calculated if it is assumed that no other acid is present. If the end point is found by using phenolphthalein, or if it is taken to be the point of maximum slope on a potentiometric titration curve, the assumption is almost always made tacitly, and no information is obtained that can be used to check its accuracy. When nonlinear regression is employed, the assumption has to be made explicitly by choosing the particular titration-curve equation that embodies it. No matter whether the assumption is tacit or explicit, it may or may not be correct. One of the most important advantages of regression analysis is that it enables its user to judge the correctness of the assumptions that have been made. Another advantage is that, if these assumptions prove to have been unsatisfactory, regression analysis can be made to provide a guide to other assumptions likely to be closer to the truth. These topics will be discussed in this section.

In every experiment there are random errors in the data, and there may be systematic errors as well. Systematic errors of experiment arise from failure to control the experimental conditions adequately and may be either constant or variable. If the temperature during a kinetic experiment remains constant but is a degree higher than it is thought to be, the data will conform to the proper rate law (to an extent that reflects the precision of measurement), but they will yield the wrong value for the rate constant. If the temperature varies from a value that is half a degree higher than the desired value at the start of the experiment to one that is half a degree below it at the end, the value that is obtained for the rate constant may be approximately correct, but the conformity of the data to the proper rate law will be poorer than it should be in view of the standard error of measurement. It is often possible to correct for a systematic error of experiment, provided that its occurrence is recognized while the experiment is still in progress: one might measure a temperature that is constant or follow the dependence of a varying temperature on time and then apply corrections based on a value of the energy of activation obtained in other experiments.

Systematic errors of interpretation result from the choice of a hypothesis or equation that does not correctly describe the results of an experiment in which all of the data needed for a proper interpretation have been secured. They differ from systematic errors of experiment in that they can be rectified after the experiment is concluded. Of course chemists rarely if ever perform experiments for which the data are incomplete, but they are sometimes uncertain about whether a hypothesis or equation is correct or about which of several possible hypotheses or equations provides the best description of these data.

Assume that data free from variable systematic errors have been obtained and fitted to an equation that describes them correctly and completely. If the resulting values of the parameters are then combined with the same equation, a calculated value of the dependent variable can be obtained for each experimental point. Each calculated value will differ from the corresponding measured one because the latter contains some random error. If the difference between the measured value and the calculated one is plotted against the independent variable, the points will be randomly scattered around the horizontal axis. I call such a plot a "deviation plot."

However, if the equation is incorrect, the same sequence of operations will have a very different result. Here the difference will depend on the value of the independent variable, and the deviation plot will consist of points scattered randomly around a smooth curve rather than around the horizontal axis. The experimenter can easily recognize that the fit is unsatisfactory, and — if the experiment has been designed and performed with such care that variable systematic errors of experiment are certainly absent or negligible — he can be sure that it is unsatisfactory because the equation does not fit the data. Moreover, and at least equally importantly, the shape of the deviation plot will frequently show which of a number of equations will yield a better fit, and thus help the experimenter to find the proper interpretation of the data.

These things are so important and so common, and classical statistics provides so little help with them, that I shall illustrate them at some length with the aid of a simple example. Figure 15 shows the rising part of a polarographic wave whose limiting diffusion current could be measured separately and found to be equal to 1.00 μA. One might suppose it to be an ordinary irreversible wave: these are synthetic data and they do not conform to that supposition, but even an experienced and knowledgeable polarographer would certainly make no other one on the basis of this curve. Fitting the coordinates of these points to Equation 18b gives $E_{1/2} = -0.6500$ V and $S = 45.39$ mV; the latter figure would be taken to correspond to $2.303\ R\ T/\alpha\ n\ F = 104.5$ mV. The standard deviation from regression is 0.0164 μA.

A popular way of gauging the validity of the hypothesis that led to these numbers is to apply the F-test. This requires the experimenter to know the standard error of the measurements that have been made. Statisticians tell us that we should know this for each experiment we perform, but it is an unusual chemist who has any accurate idea about it for any experiment. Let us nevertheless, imagine that the standard error of measurement is known to be 0.01 μA, which is exactly equal to the standard deviation of the normally distributed random errors that are included in these data. We wish to assess the applicability of Equation 18b to these data by examining the ratio of the squares of the standard deviation from regression and the standard error of measurement, which is equal to 2.70. If the equation is inapplicable, the standard deviation from regression will be the larger of the two and a one-tailed F-test will be needed. Since 29 degrees of freedom are embodied in each, the value of the ratio indicates, at about the 99.5% level of confidence, that the difference between them is

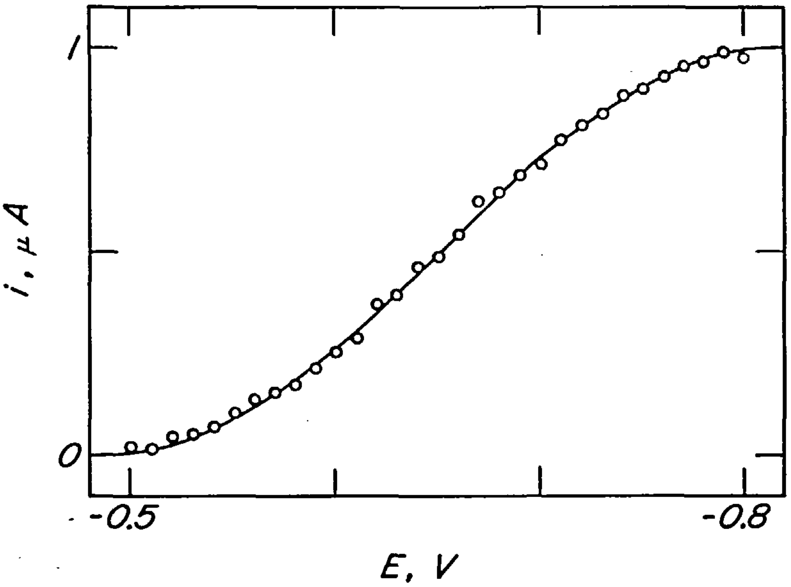


FIGURE 15. A polarographic wave for which the limiting diffusion current is equal to 1.00 μA and the standard error of a single measured value of the current is equal to 1.00×10^{-2} μA.

significant and, therefore, that Equation 18b does not correctly describe the data. This is the correct conclusion, but it is a conclusion: it provides no guidance to a better equation or a better understanding.

Figure 16 shows the deviation plot obtained on a computer terminal at the end of the fit that gave the values cited above. The horizontal axis shows the normalized deviations of the measured values of the dependent variable from the corresponding calculated ones

$$\Delta y_m = (y_{m,meas} - y_{m,calc})/s \qquad (33)$$

where s is the standard deviation from regression. The points are plotted as digits to facilitate reading their coordinates: it is the first digit to the right of the decimal point that is plotted here, so that if Δy_m has any value between 1.800 and 1.899 an 8 appears between the points labeled 1 and 2 along the ordinate axis. These deviations are certainly not randomly distributed around the abscissa axis: they are predominantly negative at the start of the plot and predominantly positive toward its end, and there are a maximum somewhere around the tenth point and a minimum somewhere around the twentieth.

Because the points are not randomly scattered around the abscissa axis, there must be a systematic error. This is the same conclusion that was obtained from the F-test, although the deviation plot does not require the user to know or assume anything about the standard error of measurement and is therefore easier for chemists to use. Whether or not a systematic error can be detected on a deviation plot depends on the relative magnitudes of the systematic and random errors. Other things being equal, the plot becomes smoother as the random errors become smaller, and detection of a systematic error becomes easier as a result. If the standard deviation from regression in this example were increased by increasing the standard error of measurement that is included in the data, it would become more difficult to perceive that the points are scattered around a curve that differs from the abscissa axis. If the standard error of measurement were sufficiently large, it would be concluded that Equation 18b pro-

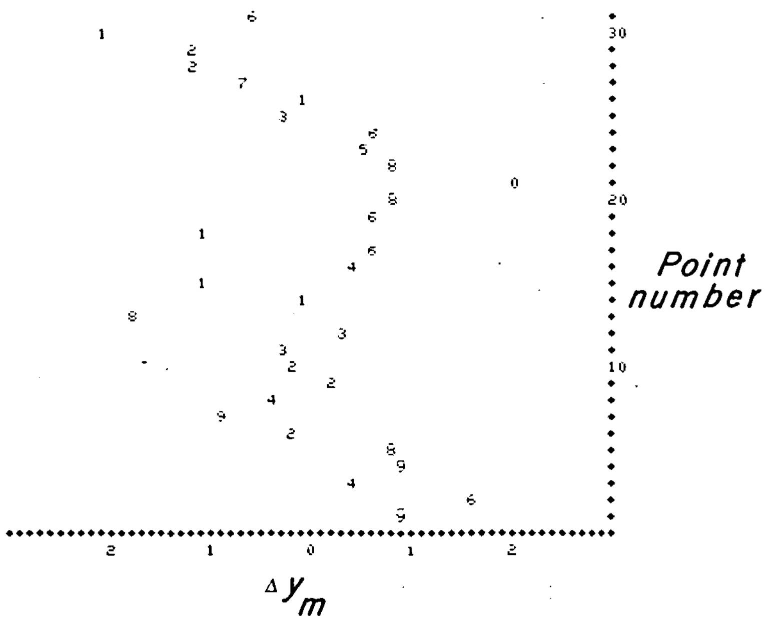


FIGURE 16. Deviation plot obtained by fitting the data of Figure 15 to Equation 18b.

vided a satisfactory representation of the data. Again, the F-test would lead to the same conclusion. Just how much random error must be present to prevent the acceptance of an incorrect equation depends on the number of data points and on the complexity of the deviation plot: As a rough approximation, a practiced user can usually detect a pattern underneath the scatter when an F-test would indicate that the equation is wrong at about the 90 to 95% level of confidence.

The deviation plot shown in Figure 16 can be interpreted with the aid of other deviation plots constructed from synthetic data free from random errors. For example, it might be imagined that the polarogram actually consists of two waves so close together that they are indistinguishable to the eye. Such a polarogram would be described by Equation 19. If data are calculated from that equation and fitted to Equation 18b, there will be a systematic error of interpretation because they do not conform to the equation to which they are fitted; this error will lead to the smooth curve shown in Figure 17. The curve is smooth because no random error has been introduced. Such a curve can be called a "deviation pattern" to distinguish it from the deviation plots that are compared with it. Deviation patterns are smooth curves obtained from synthetic data; deviation plots are obtained from experimental data and include the scatter resulting from the errors of measurement.

Another conceivable possibility is that corrections for a residual current linearly dependent on the applied potential have been ignored or wrongly applied. In this case, the polarogram should be described by the equation

$$i = \frac{i_{d,l}}{1 + \exp\left([E - E_{1/2}]/S\right)} + b\,(E^* - E) \qquad (34)$$

where E^* is the applied potential at which the necessary correction is equal to zero, and b expresses the dependence of the correction on the applied potential. Equation 34 may be used to synthesize "data" representing this hypothesis; on fitting these to Equation 18b, another deviation pattern will be obtained for comparison with Figure 16. This one is shown in Figure 18.

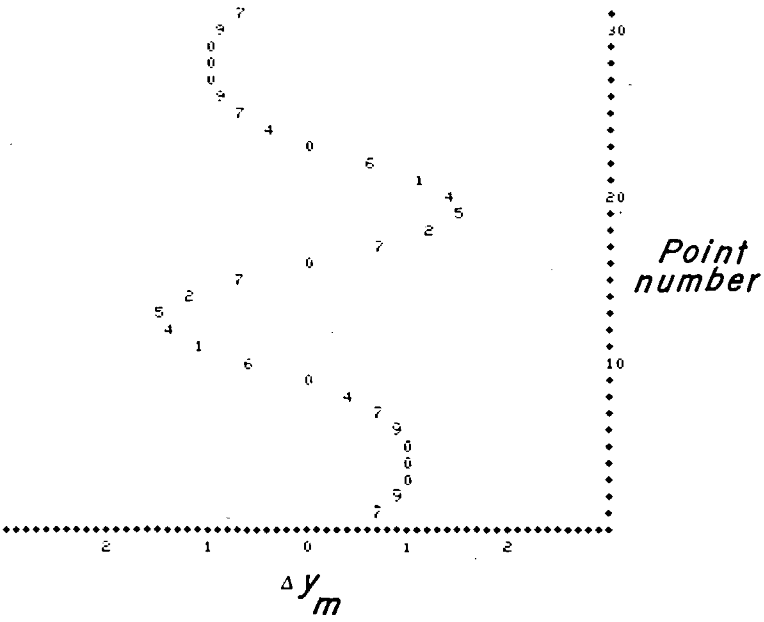


FIGURE 17. Deviation pattern obtained by fitting data that obey Equation 19, which describes two closely spaced waves, to Equation 18b.

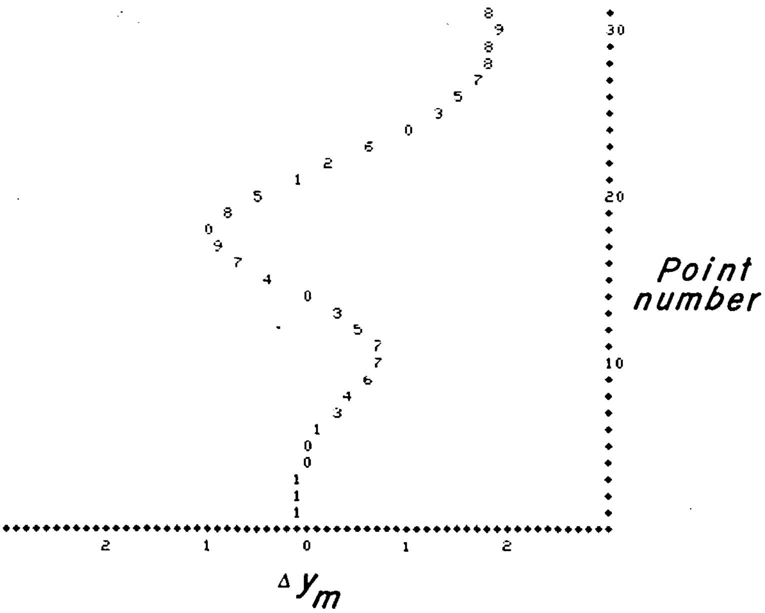


FIGURE 18. Deviation pattern obtained by fitting data that obey Equation 34, which describes the effect of an incorrectly compensated current that varies linearly with the applied potential, to Equation 18b.

Usually the deviation pattern obtained by calculating the coordinates of the "data" points from one equation and fitting them to another has a shape that is not only characteristic of the two equations but also independent of the values assigned to the parameters in the first. Changing those values will, however, affect the heights of the maxima and minima and may also affect their spacing along the axis representing the independent variable. It is as though the pattern were drawn on a sheet of gum rubber that could be stretched and compressed in various ways. This makes it possible to accumulate a library of deviation patterns that will facilitate the interpretation of many different experiments of the same general type. There are occasional exceptions in which the gross shape of the pattern is altered by changing the values of the parameters in the equation from which the data are calculated. These may involve reflection around the abscissa axis or even more far-reaching changes,[38] but fortunately occur only with very complicated equations and are associated with fairly drastic changes of the shapes of plots of the dependent variable against the independent one.

When Figure 16 is compared with the deviation patterns in Figures 17 and 18, it can be seen immediately that Figure 18, and the hypothesis that gave rise to it, cannot account for the behavior observed, but that Figure 17 and its underlying hypothesis might possibly do so. In Figure 16, the deviations are large and negative at the outset and large and positive at the end, and there are two extrema, of which the first is a maximum and the second a minimum. All of these things are also true of Figure 17, but in Figure 18 the deviations are small and positive at the start and large and negative at the end, and the first extremum is a minimum while the second is a maximum.

Although the dissimilarity between the shapes of Figures 16 and 18 proves that the hypothesis of Figure 18 cannot account for the data of Figures 15 and 16, the similarity between the shapes of Figures 16 and 17 does not prove that the hypothesis of Figure 17 is correct. Other hypotheses might be found that gave rise to deviation patterns that are difficult or impossible to distinguish from this one. In many cases, there will be small differences, as, for example, in the relative heights of the extrema and in their spacing along the abscissa axis, but these are too easily concealed by random errors and too likely to arise from differences among the values of the parameters for the experimental plot and the theoretical pattern, to be useful and reliable. Nevertheless, even this highly curtailed example shows that it is possible, and easy, to reject some of the alternative hypotheses that might be made when the initial one is found to be inadequate, and at the same time to obtain a guide to other hypotheses that might prove to be acceptable.

One must repeat the process to obtain further enlightenment. To do this it is necessary to select a second hypothesis. With the single match obtained with the very small number of deviation patterns included here, this must obviously be the hypothesis that there are two closely overlapping waves. A larger library of patterns might contain several others generally similar to Figure 17. One of the various hypotheses corresponding to those might seem more probable than the others, on chemical or experimental grounds. If these do not lead to a unique choice, one may simply select, from those that seem equally probable, the one that involves the smallest number of adjustable parameters. Fitting the data of Figure 15 to Equation 19 gives $i_{d,l,1} = 0.798$ μA (and therefore $i_{d,l,2} = 1.00 - i_{d,l,1} = 0.202$ μA), $E_{1/2,1} = -0.6298$ V, $S_1 = 39.7$ mV, $E_{1/2,2} = -0.7284$ V, and $S_2 = 22.6$ mV. The deviation plot is shown in Figure 19. The standard deviation from regression is 0.0116 μA.

To the eye, this appears to be nothing more than a random scatter around the abscissa axis. An F-test like the one described above indicates that the difference between the standard deviation from regression and the standard error of measurement is significant at only about the 75% level of confidence. Another test for the absence of systematic error, which is easier to apply if one does not have a sufficiently accurate idea about the value of the standard error of measurement and therefore cannot perform an F-test reliably, is a χ^2-test based on the number of changes of sign that occur in a list of deviations arranged in the order of increasing or decreasing values of the independent variable. If there is no systematic error, each deviation is equally likely to be positive or negative, regardless of the sign of the deviation that preceded it; but

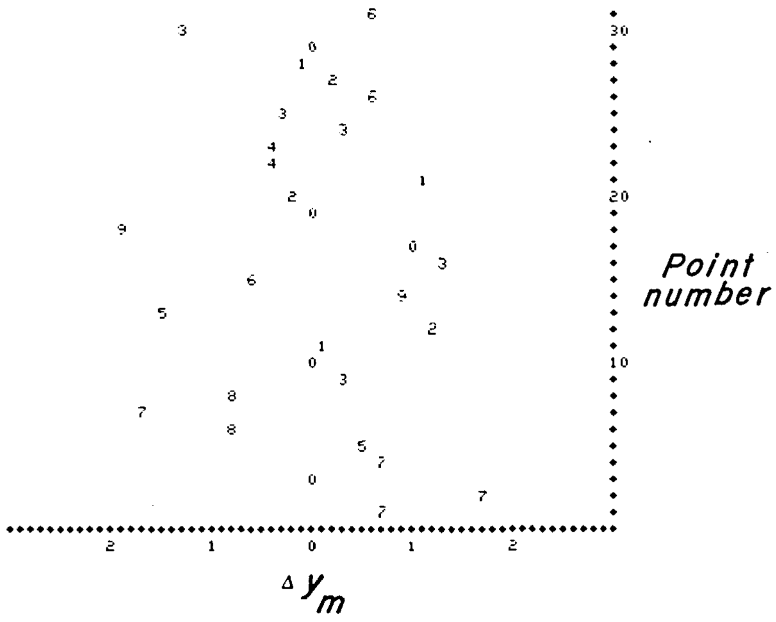


FIGURE 19. Deviation plot obtained on fitting the date of Figure 15 to Equation 19.

if there is a systematic error each deviation is much more likely to have the same sign as the preceding one. For the fit just described, such a test shows that the probability is only about 72% that there is a systematic error, which is virtually identical with the result of the F-test. On these grounds, it would be concluded that the two-wave hypothesis cannot be rejected with any certainty on the basis of this experiment.

At this point, the really conscientious experimenter should inquire whether another hypothesis might not be equally acceptable — or more acceptable — and whether there is any way in which it could be distinguished from this one. The user of the deviation-pattern approach could consider the other hypotheses that would have given patterns similar to those previously identified. Any that, when subjected to a fit to Equation 18b, would have given a pattern closely resembling the one in Figure 17 should be examined to see whether it might be applicable. A larger library of patterns would show that one similar to Figure 17 would result if there were only one wave and if the uncompensated cell resistance were high. That pattern, which is shown in Figure 20, is not quite identical with the one in Figure 17, but the two are so similar that either hypothesis would provide a satisfactory description of data having random errors as large as these. If the precision of measurement were 0.003 instead of 0.01 μA, and if this new hypothesis were correct, a pattern would barely be discernible in Figure 19.

The method is so powerful that the user must keep its limitations in mind. These are (1) that it provides no information whatever about any hypothesis that does not appear in the library of patterns available and (2) that it cannot guide the user to an interpretation guaranteed to be correct. It does provide an easy and rational guide to previously recognized hypotheses that deserve consideration, and it does serve to reject all of them that can be rejected legitimately on the basis of the data at hand. At present, it requires each prospective user to assemble his own library of patterns for the field in which he is interested. The time and effort expended doing this are soon repaid by the decrease in the number of possibilities that have to be considered in interpreting the data obtained from new experiments, and I hope that a time will come when the approach is widely enough used to enable chemists just entering new areas of work to obtain libraries from others.

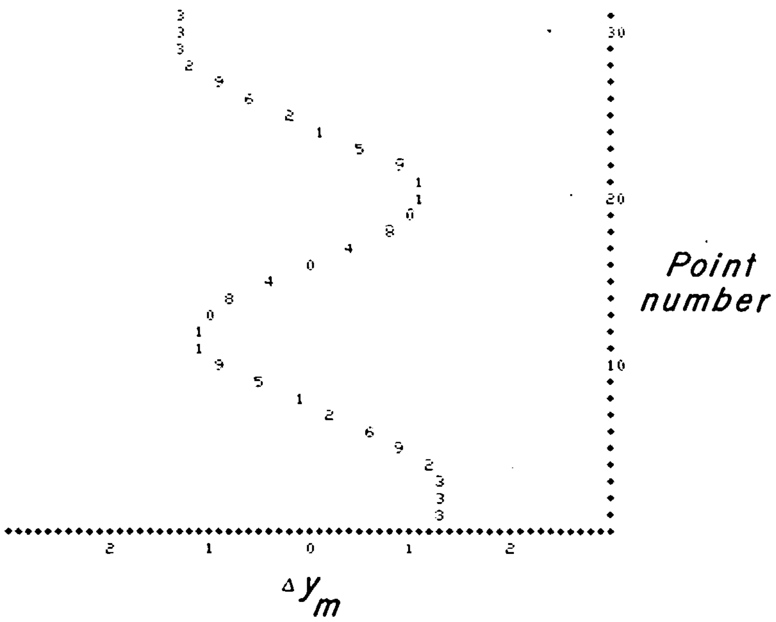


FIGURE 20. Deviation pattern obtained by fitting data that correspond to a single wave obtained in a high-resistance cell to Equation 18b.

B. Some Applications

As yet there are only two fields for which reasonably complete libraries of deviation patterns have been worked out. One is the restricted field of potentiometric acid-base titrimetry performed for the purpose of deciding the functionality of an acid that gives only a single point of maximum slope on its titration curve.[39] The other, larger field deals with current-time curves in controlled-potential electrolysis.[40,42] Since both are fairly specialized, I shall devote this section to a few more fragmentary examples.

The difficulty that arises in the polarographic analysis of a mixture of two substances giving overlapping waves was mentioned in Section III.A. One approach to this problem involved the use of matrix algebra and was tested[15] by applying it to mixtures containing two, or all three, of the isomeric nitrobenzoic acids. Polarograms of the individual acids were recorded, and measurements of the current at various potentials were made from these, with great care. The resulting data were used in calculating the heights of the closely overlapping waves obtained for mixtures. The approach was less successful than might have been hoped or expected. Some years later, wishing to see whether nonlinear regression would provide better results, we recorded polarograms of the same acids under the same conditions and fitted them to Equation 18b. A deviation plot resembling Figure 17 was obtained for each acid; this suggested that each gives two waves, and Equation 19 gave acceptable fits from which it appeared that the relative heights of these waves varied with concentration because one of them is an adsorption wave. What is impressive about this is that the deviation-pattern approach revealed the existence of two overlapping waves on polarograms where they had been invisible to a careful experimenter who was attuned to the problem while the measurements were being made.

When Carr and Jordan[2] succeeded in deriving Equation 2, they sought a system to which it could be applied and selected the titration of nitromethane with sodium hydroxide. The reaction between these substances is exothermic, and a thermometric titration curve obtained by adding sodium hydroxide at a constant rate to a solution of nitromethane has the same shape as a plot of $c_P/c_A{}^0$ against time. To evaluate β, Carr and Jordan constructed plots of $c_P/c_A{}^0$ against the dimensionless parameter t/t^* for a number of different values of β, compared each experimental curve with these, and interpolated between the values of β used in drawing the calculated curves to estimate the value that would have given rise to the experimental one. This is a clumsy and time-consuming procedure, and its precision is limited, but in view of the complexity of Equation 2, it is not at all surprising that a classical technique of data handling should have such limitations.

Shortly thereafter, the same system was reinvestigated by Campbell et al.[41] with the aid of nonlinear regression. Values of $c_P/c_A{}^0$ were obtained for comparison with the experimental ones by numerical integration of the rate equation

$$\frac{dc_P}{dt} = k\, c_A c_R = k(c_A^0 - c_P)(\rho t - c_P) \tag{35}$$

which is simpler to manipulate than its closed-form solution, Equation 2. The values obtained for k, which was the only adjustable parameter, agreed closely with those given by Carr and Jordan, but the experimental deviation plot (Figure 21, Curve a) made it impossible to accept the hypothesis that Equation 35 provided a satisfactory representation of the data. The random error in these data is much smaller in comparison to the systematic error of interpretation than it was in Figure 16, and therefore the points here lie much closer to a smooth curve. Figure 22 shows a deviation pattern that was calculated in the course of an attempt to interpret these results. It was generated by adopting the hypothesis that the mechanism is actually

$$\mathrm{A} + \mathrm{R} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{P} \tag{36}$$

(which differs from Equation 1 in that it takes a possible backward reaction into account), calculating synthetic data from the corresponding rate equation

$$\frac{dc_P}{dt} = k_1\,(c_A^0 - c_P)\,(\rho t - c_P) - k_{-1}c_P \tag{37}$$

and fitting these to Equation 35, which represents the hypothesis that had been assumed in the fit that led to Figure 21. The similarity between the shapes of Curve a in Figure 21 and the deviation pattern in Figure 22 showed the Equation 37 would provide a better fit. When that equation was employed in place of Equation 35, values of k_1 and k_{-1} were obtained that corresponded to $K_a = (4.6 \pm 1.7) \times 10^{-11}$ for the acidic dissociation of nitromethane, while the literature value is $K_a = 6.1 \times 10^{-11}$.

Carr and Jordan [2] were well aware that this acid-base reaction does not proceed to completion, as Equation 1 implies. The main reaction is

$$\mathrm{CH_3NO_2} + \mathrm{OH^-} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{CH_2{=}NO_2^-} + \mathrm{H_2O};\ K_t = k_1/k_{-1} = K_w/K_a = 6.1 \times 10^3 \tag{38}$$

Under the conditions they used, this reaction would have typically been 95% complete at the equivalence point, and more nearly complete everywhere else, if the forward reaction were instantaneous. This is to say that the deviation-pattern approach revealed the occurrence of the backward reaction even under conditions judiciously chosen so that it would be unimportant — and also that even under these unfavorable conditions nonlinear regression was able to furnish values of k_1 and k_{-1} that were in substantial accord with the results of entirely independent measurements.

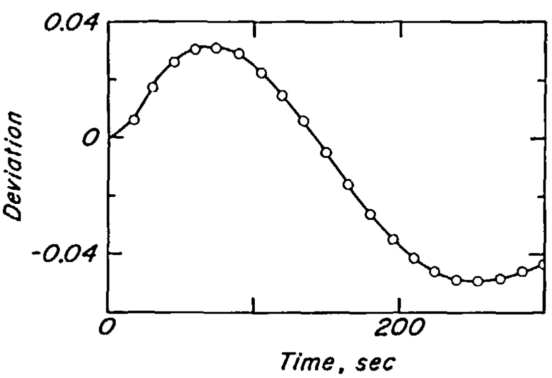


FIGURE 21. Deviation pattern obtained by fitting "data" that obey equation 37, which describes the behavior of a reversible reaction, to equation 35. (From Campbell, B. H., Meites, L., and Carr, P. W., *Anal. Chem.*, 46(3), 389, 1974. With permmision.)

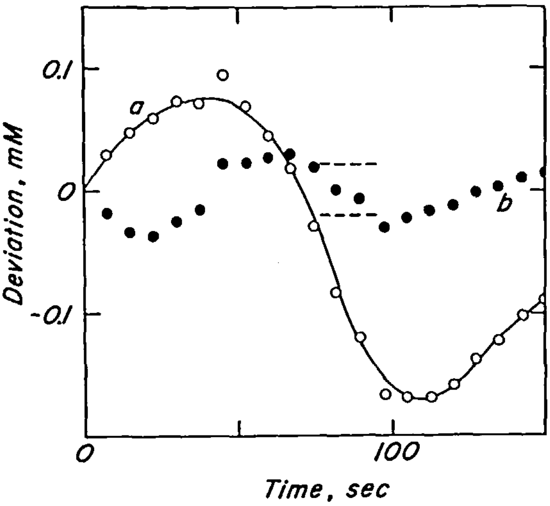


FIGURE 22. Deviation plots obtained by fitting data secured in a typical thermometric titration of 5.8×10^{-3} *M* nitromethane with 1.0 *F* sodium hydroxide to (a) (open circles) Equation 35 and (b) (solid circles) Equation 37. The distance between the horizontal dashed lines corresponds to the rated linearity (± 0.25% of full scale) of the recorder used to record the titration curve. (From Campbell, B. H., Meites, L., and Carr, P. W., *Anal. Chem.*, 46(3), 389, 1974. With permission.)

Curve b in Figure 21 shows the deviation plot resulting from fitting to Equation 37 the same data that had given Curve a when they were fitted to Equation 35. There is still some systematic error, although its magnitude is no larger than the uncertainty corresponding to the rated linearity of the recorder used in the experiment. This and other considerations peculiar to the nature of the experiment seemed to render further interpretation perilous.

Deviation patterns provide chemists and other scientists with an extremely powerful and sensitive tool for assessing the correctness of a model chosen to represent a set of experimental data.

V. AUTOMATIC CLASSIFICATION

A. Introduction

At the beginning of this review I said that some experiments are performed to obtain yes-or-no answers to chemical questions. One fairly common question, "is this model applicable to these data?", is amenable to the deviation-pattern approach. The preceding section showed how that approach can be used to identify, from a number of possible models, those for which the answer to this question would be "yes."

Automatic classification is an exciting new technique that combines the deviation-pattern approach with other techniques of nonlinear regression to make such identifications automatic. For the sake of illustration, let us suppose that a chemist wishes to know the order of a certain chemical reaction with respect to one reactant R and investigates this by following the variation of the absorbance of R with time in a reaction mixture containing a swamping or fixed concentration of every other substance that might appear in the rate law. To keep the illustration simple, let us further suppose that the reaction goes to completion and that only integral orders need to be considered. The problem is to classify the reaction as being zeroth-order, first-order, ..., with respect to R. Algebraically this corresponds to deciding which of the equations

$$
\begin{aligned}
A &= A^0 - k\,t \quad (t < A^0/k) \\
A &= 0 \qquad\quad\; (t \geq A^0/k) \qquad\qquad (39a)
\end{aligned}
$$

$$A = A^0 \exp(-k\,t) \qquad\qquad (39b)$$

$$A = A^0 / (1 - k\,t\,A^0)$$

•

•

•

etc. $\qquad\qquad$ (39c)

is obeyed by the data. In each of these equations, k denotes the product of the pseudo-nth-order rate constant by $(\varepsilon\, b)^{1-n}$, where ε is the molar absorptivity of R at the wavelength employed, and b is the thickness of the cell.

One approach would be to fit the data to as many of these equations as the experimenter's patience (and budget for computer time) permit, and examine the resulting fits, one of which will be conspicuously better than all the rest. That would be unimaginative to the point of being immoral.

Another approach would be to fit the data to Equation 39b and examine the deviation plot obtained. There are three possibilities. If the plot consists of points randomly scattered around the time axis, the corresponding hypothesis (that the reaction is first-order with respect to R) cannot be rejected on the basis of the data obtained. This is equivalent to saying that the reaction is first-order with respect to R if a plot of A against t is a straight line, but it avoids the agony that users of the graphical approach may suffer when curvature results from the increasing uncertainty in ln A as the reaction nears completion. The second possibility is that the plot will be  -shaped: a plot having this shape results from trying to fit a straight line to an equation that can only yield an  -shaped curve. In this case, the reaction must be zeroth-order with respect to R, and a fit to Equation 39a might be undertaken to evaluate k. The third possibility is that the plot will be  -shaped, in which case the order with respect to R might be 2 or any larger integer. Here, a fit to Equation 39c could be performed. If this is acceptable, the reaction will be classified as being second-order with respect to R and a value of k will be obtained, but if it is not, the data may next be fitted to the next equation in the above series to test the hypothesis that the reaction is third-order with respect to R, and so on.

Of course this is oversimplified because the rate law may actually contain two or more terms, so that the order with respect to R may not have an integral value. It does, however, illustrate how classification problems of this general sort can be solved in a perfectly mechanical way. That being so, it follows that they can be solved automatically.

Scientific experimentation can be divided into two parts: the acquisition of the data and their subsequent interpretation. Each of these, in turn, can be further divided into two parts: one that is mechanical and one that is not. The nonmechanical part of acquiring data that will serve a useful purpose consists of imagining an experiment that is worth performing; selecting the conditions under which it may be performed; designing and improving the apparatus used in making the measurements; and identifying possible sources of error and inventing ways of eliminating or minimizing them. The mechanical part consists of the actual collection of the data. The nonmechanical part can be done only by a human brain; but great strides have been made in relieving humans of the difficulty and tedium that attend the mechanical part. Thirty years ago, polarograms were still being obtained by chemists peering through telescopes to observe the deflections of galvanometers, and the necessity of doing this consumed time and energy that would have been better devoted to the exercise of imagination and ingenuity. Since then, the advent of computerized data acquisition has freed chemists from this kind of drudgery and has helped make it possible to perform experiments that would otherwise be forbiddingly difficult or impossible.

The nonmechanical part of data interpretation consists of inventing new hypotheses, imagining their consequences, and devising ways of testing them and differentiating them from existing ones. The mechanical part consists of retracing steps that have previously been trod by others in interpreting similar data obtained in experiments of the same sort. Again the nonmechanical part cannot be done except by a human brain. To the extent that it is possible to liberate that brain from the tedium of drawing graphs, making interpretations, and reaching conclusions that have been drawn or made or reached before in similar circumstances, the uniquely human abilities of chemists should be liberated for better purposes in the same way that they have been liberated by sophisticated modern techniques of data acquisition.

B. Techniques and Applications

Automatic classification is still so new that much of the work on it has been devoted to techniques for effecting it. In this section, however, I shall attempt only to indicate that there are different possibilities without going into detail about them, and shall concentrate on the chemical situations to which they have been applied.

The simplest kind of classisication is a binary one by which a single hypothesis is either accepted or rejected. If there were only two mutually exclusive hypotheses, rejecting one would be equivalent to accepting the other: this would also be a binary classification. Binary classifications are certainly both common and important, and pattern recognition[42,43] is much concerned with them.

For data of the kinds considered in this review, simple binary classifications rarely suffice. There are often several, or many, mutually exclusive hypotheses that might be adopted to account for behavior of the type observed, and there are situations in which it is possible to frame different hypotheses that are not mutually exclusive. An example of the first sort arises in studies of stepwise complex formation between a metal ion and a ligand: the maximum coordination number j of the complexes formed at appreciable concentrations during the experiment might be 1,2,3,.... Classifying a particular system according to the value of j has been called "linear" classification[44] because the hypotheses can be arranged in such a way that each successively rejected one has a logical successor. Another example of the same sort arises in trying to identify the mechanism of a chemical reaction, which might be pseudo-first-order with respect to a reactant whose concentration is followed, which might involve a backward reaction having an appreciable rate, which might involve a slow prior equilibrium between that reactant and some other constituent of the solution, and so on. There is no a priori order in which these hypotheses could be arranged, and problems like this are said to be "branched" because, with the deviation-pattern approach, the nature of the systematic errors that cause one hypothesis to be rejected may point the way to one or another group of the ones that remain. An example of the second sort, in which the hypotheses are not mutually exclusive, arises in titrations of the alkali-metal salts of long-chain fatty acids with strong acids: micelles of the fatty acid anion A^- might be present at the start of the titration or they might not, and either the free acid HA or the "acid soap" MHA_2 might or might not precipitate during the course of the titration. In any one titration, depending on the experimental conditions and the values of the parameters that characterize these three processes, all or none of them, or any one or two of them, might occur and a decision about the occurrence of micelle formation has to be made independently of one about the formation of either of the possible precipitates. This is called "multiple binary" classification.

Four different ways have been devised for allowing the necessary decisions to be made without human intervention. They all involve the use of nonlinear regression. The programmer selects one hypothesis as the first to be tested: for a linear or branched problem this is the simplest one, while in the only multiple binary problem yet attacked it was the most complex. One technique is to allow the fit to proceed to completion and then scan the deviations of the successive calculated values from the measured ones for a group of deviations that have the same sign and are large in comparison with the standard deviation from regression or, preferably, the known standard error of measurement.[44] The existence of such a group, which can be as improbable as the programmer chooses, if systematic errors are absent, is made to result in the rejection of the hypothesis and the adoption of the next one. As the word "next" implies, this is hardly suited to any except linear problems. A more sophisticated and more useful technique is to identify the signs and locations (along the abscissa axis) of all such groups of widely deviant points.[40] This is applicable to branched problems because it permits a ∪ -shaped plot to be differentiated from a ∩-shaped one. Another technique is to reject any hypothesis that gives impossible values for any of the parameters,[47,48] such as a negative value of an equilibrium constant. A fourth is based on the behavior, while the fit is still in progress, of the values assigned to a parameter characterizing a process that does not actually occur:[47] this is applicable to multiple binary problems but probably not to others.

Two examples will suffice to show the reliability that can be attained and the extents to which the interpretation of data can be simplified and extended.

There had been much discussion in the literature of the possibility of classifying an acid as monobasic or polybasic if the titration curve obtained on titrating it potentiometrically with a strong base shows only a single point of maximum slope,* but no way of distinguishing between dibasic and tribasic acids on the basis of such data had ever been proposed. Campbell and Meites[44] automated the deviation-pattern-recognition scheme described by Meites and Barry[39] and applied it to titrations of weak bases with hydrochloric acid. This was a linear problem and a simple one: there are hardly any monomeric tetra- and higher-functional acids for which all the dissociation constants are so nearly equal that only a single break appears on the titration curve, and therefore there were only a few possibilities that had to be considered. Input to the program consisted only of the coordinates (volume of acid added and pH) of the data points, the value of the apparent single-ion activity coefficient of hydrogen ion in the supporting electrolyte employed for the titration (which can be obtained by ordinary nonlinear regression from data secured in titrating a base of known identity in the same medium), the concentration of the reagent, and an estimate of the standard error of a single pH-measurement. The hypothesis that the base is monofunctional was tested first by performing a fit to the appropriate equation and then scanning the difference between the measured and calculated pH values in the order or increasing volume of added acid. Systematic errors were considered to be present if the absolute value of this difference exceeded three times the estimated standard error of measurement for each of *n* consecutive points, where *n* was at least one tenth as large as the total number of data points being analyzed. From the previously calculated deviation patterns it was known that such a group of points would appear near the beginning of the titration if random errors were absent and the monofunctional hypothesis were incorrect. If the first such group appeared during the first half of the titration, the difunctional hypothesis was adopted next, and so on, but if it appeared during the second half (so that the monofunctional hypothesis led to a statistical improbability while a polyfunctional one could not be reasonably adopted), the logical contradiction was resolved by increasing the estimated error of measurement slightly and repeating the scan. If no such

---

* Compare the references to the prior literature cited in Reference 31.

the scan. If no such group appeared anywhere, the current hypothesis was accepted and the values obtained for its parameters were printed out.

This was applied to data secured in a number of titrations of acetate, succinate, and citrate ions. With honest estimates of the standard error of measurement, which was believed to be 0.003 pH-unit, all three of these were classified correctly in every case, and correct classifications were also obtained with any other estimate between 0.001 and 0.0075 pH-unit. Erroneously large estimates led to the acceptance of hypotheses simpler than the truth: succinate was classified as monofunctional if the estimate exceeded 0.06 pH-unit, while citrate was classified as difunctional if it lay between 0.008 and 0.1 pH-unit and as monofunctional if it equalled or exceeded 0.15 pH-unit. Results like these are to be expected. If the data are extremely imprecise, there is no sure way to reject the simplest hypothesis that can be formulated, while if they are slightly less imprecise they may serve to reject that hypothesis but may not serve to reject the next more complicated one. In such circumstances, the acceptance of a simpler hypothesis in preference to a more complex one conforms to Occam's Razor. Erroneously small estimates occasionally led to misclassifications — there were a few titrations of citrate that gave data leading to a tetrafunctional classification if the standard error of measurement was taken to be 0.0005 pH-unit (!) — but more often such estimates were adjusted to yield both a correct classification and a close approximation to the standard error actually involved. The chemical problem here is not very important, but data like these would not enable the chemist to solve it in any other known way, whereas automatic classification yields reliable solutions while the chemist is doing something else.

Titrations with strong acids of the anions of long-chain fatty acids were mentioned above in this section. Micelle formation and the precipitation of the free fatty acid HA or the "acid soap" MHA_2 may all occur independently at different stages of the titration, and curves of eight different kinds may therefore be obtained. The classification of a particular curve is a multiple binary problem. A program was developed[47] that required only the coordinates of the data points and the concentrations of the fatty-acid anion A^- in the solution being titrated, the strong acid used as reagent, and any alkali-metal (M^+) salt present in either solution. The first hypothesis was that each of the three possible processes occurs at some stage, and crude initial estimates of the parameters characterizing them were obtained automatically. For example, the critical micelle concentration C, which is the only parameter I shall discuss, was arbitrarily taken to be one third of the initial concentration of A^-. As the fit proceeds, this estimate is adjusted automatically. If micelles were not actually present initially, the value assigned to C will increase and will eventually exceed the initial concentration of A^-. The program was so constructed that, if C continued to exceed the initial concentration of A^- for some cycles of computation, the possible occurrence of micelle formation was rejected, a message to that effect was provided, and the fit was continued with the parameters that remained.

When this program was described, it had been tested with data obtained in 33 titrations of laurate ion under different conditions. The values of the parameters were known for laurate ion, so that it was possible to deduce the correct classification of each titration for comparison with the result obtained automatically. Of 33 titrations, 29 were classified correctly; one was classified incorrectly; and there were three marginal cases in which a parameter was retained but its value was used in calculating the pH values for only one or two points and was therefore identified as doubtful in the final printout. The difference among the various kinds of possible curves are so slight that the human chemists who performed this work could not correctly classify any of the three marginal cases without using additional information, and classified 13 of the

33 titrations incorrectly. The program has since been applied to about two dozen similar titrations without yielding another misclassification.

Other automatic classifications that have been devised include an automation of the well-known DeFord-Hume procedure[48] for evaluating the maximum coordination number of a reversibly reduced metal ion from polarographic data secured with various concentrations of ligand;[46] a procedure for deciding whether a one-term rate equation and an integral reaction order will account for absorbance-time data obtained in following a chemical reaction;[49] and the first part of a scheme for identifying the mechanism of a half-reaction on the basis of current-time data obtained in controlled-potential electrolysis.[40] In all of these diverse cases it has been possible to achieve classifications identical with, and usually more reliable than, those attained by human beings, and to do so without requiring human intervention at any stage once execution has begun. To my mind the greatest difficulty in constructing such programs — especially those pertaining to branched problems, in which the number of alternative hypotheses may be very large, and which are, therefore, probably of greatest importance in scientific research — arises from the necessity of providing information about all the hypotheses that cannot be distinguished from the one accepted. This should include suggestions of further experiments that will be helpful and requires the programmer to be extremely well informed about the field for which the program is being written.

There is one area in which automatic classification is proving to be conspicuously unsuccessful, and it has consumed more work than all the other areas to which automatic classification has been applied. It deals with a question that seems simple enough: "given an absorbance-wavelength envelope showing only one maximum, how many individual bands does it contain and what are the values of the parameters characterizing each?" The parameters of greatest interest are the wavelength of maximum absorption and the corresponding absorbance for each of the absorbing constituents. This question has given rise to much extremely sophisticated work,[12,50,51] to much discussion of whether real absorption bands are Gaussian, Lorentzian, hybrids, or something else, and to much disappointment resulting from the fact that one approach after another tends to underestimate the number of constituents known to be present.

Despite the undoubted practical importance of the question as formulated, I believe it reflects a fundamental misunderstanding of the philosophy of science. Its formulation is that of a linear problem. Every linear problem has alternative solutions. For example, the potentiometric titration curve for the titration of a monobasic weak acid with a strong base can be exactly duplicated by that for a dibasic, tribasic, or any other acid for which the ratios of the successive dissociation constants have certain values. It would certainly be possible to reproduce within any precision of measurement likely to be attainable in the next century any absorption envelope 100 nm wide by assuming that there are 10,000 different individual absorption bands, each having a wavelength of maximum absorption at least 0.01 nm different from that for any other. This is to say that no linear problem can be solved without invoking Occam's Razor. A more proper reformulation would be "what is the smallest number of individual bands that will yield a fit compatible with the precision of measurement?"

Thus altered, the question would be far less difficult. At the same time, unfortunately, the answer to it would be far less useful. In some restricted applications, reformulation as a multiple binary problem might make it possible to secure useful answers: "the sample may contain n substances, for which the wavelengths of maximum absorption (λ_{max}) and half-widths ($\Delta\lambda$) are equal to $\lambda_{max,1}$, $\Delta\lambda_1$, $\lambda_{max,2}$, $\Delta\lambda_2$, ..., $\lambda_{max,n}$, $\Delta\lambda_n$. Which ones are present, and what is the absorbance of each at its λ_{max}?" This will also be doomed if n is very large (or if interactions among the different components affect their values of λ_{max} and $\Delta\lambda$). However, it is a question that might be answered if the precision of measurement is high enough, whereas the one now generally being asked is not.

VI. CONCLUSION

Like pattern recognition,[44,45] though in very different ways and with very different kinds of data, the techniques described in this article provide chemists with versatile and searching methods of deciphering the significance of the data they obtain. Nonlinear regression is coming to be fairly widely used, although older techniques are still allowed to consume more time and trouble than they should. It is interesting that nonlinear regression has been more widely adopted by scientists and engineers in some other fields. Pointwise variance analysis is rarely performed, and its most important potential use — in guiding the design of experiments — still awaits a searching experimental test. Deviation-pattern recognition and automatic classification have been attempted by only a handful of cognoscenti, and our understanding of the potentialities and limitations of all these techniques is still sadly imperfect. Nevertheless, it is already clear that those potentialities are great. What is most needed is for others to join in the development and use of these techniques, and I have written this survey in the hope that it will help bring this about.

ACKNOWLEDGMENTS

Many of the ideas I have mentioned were contributed, in whole or in part, by colleagues and graduate students, and although these are too numerous to name individually I am deeply appreciative of their interest, help, and encouragement. Most of the computations described here were made with a minicomputer whose purchase was made possible by grants from the Eastman Kodak Company and the National Science Foundation. The preparation of this review was supported in part by grant number CHE 76-17247 from the National Science Foundation.

REFERENCES

1. Eriksson, K.-H., Mikiver, A., and Thorsell, W., *J. Chem. Educ.*, 54, 454 (1977)
2. Carr, P. W. and Jordan, J., *Anal. Chem.*, 45, 634 (1973).
3. Carr, P. W., *Crit. Rev. Anal. Chem.*, 2, 491 (1972).
4. Meites, T. and Meites, L., *Talanta*, 19, 1131 (1972).
5. Meloun, M. and Čermák, J., *Talanta*, 23, 15 (1976).
6. CODATA Task Group on Publication of Data in the Primary Literature, Guide for the Presentation in the Primary Literature of Numerical Data Derived from Experiments, ICSU CODATA Central Office, Frankfurt/Main, December, 1973.
7. Dyrssen, D., Jagner, D., and Wengelin, F., *Computer Calculation of Ionic Equilibria and Titration Procedures*, Almqvist and Wiksell, Stockholm, 1968.
8. Sabatini, A., Vacca, A., and Gans, P., *Talanta*, 21, 53 (1974).
9. Sayce, I. G., *Talanta*, 15, 1397 (1968).

9a. Sayce, I. G., *Talenta*, 18, 653 (1971).

9b. Sayce, I. G. and Sharma, V. S., *Talanta*, 19, 831 (1972).

10. Marquardt, D. W., *J. Soc. Ind. Appl. Math.*, 11, 431 (1963); M. H. Lietzke, ORNL-3259, Oak Ridge National Laboratory, April 4, 1962.
11. Kohmann, T. P., *J. Chem. Educ.*, 47, 657 (1970).
12. Gutknecht, W. F. and Perone, S. P., *Anal. Chem.*, 42, 906 (1970).
13. Kargin, V., Manoušek, O., and Zuman, P., *J. Electroanal. Chem.*, 12, 443 (1966).
14. Frisque, A., Meloche, V.W., and Shain, I., *Anal. Chem.*, 26, 471 (1954).
15. Israel, Y., *Talanta*, 13, 1113 (1966).
16. Ružić, I. and Branica, M., *J. Electroanal. Chem.*, 22, 243 and 422 (1969).
17. Ružić, I., *J. Electroanal. Chem.*, 25, 144 (1970).

18. Lampugnani, L. and Meites, L., *Thermochim. Acta.*, 5, 351 (1973).
19. MacInnes, D. A. and Shedlovsky, T., *J. Am. Chem. Soc.*, 54, 1429 (1932).
20. Shedlovsky, T., *J. Am. Chem. Soc.*, 54, 1405 (1932).
21. Harned, H. S. and Ehlers, R. W., *J. Am. Chem. Soc.*, 54, 1350 (1932).
22. Meites, L., *Polarographic Techniques*, 2nd ed., Interscience, New York, 1965, 153.
23. Meites, L. and Lampugnani, L., *Anal. Chem.*, 45, 1317 (1973).
24. Barry, D. M., Meites, L., and Campbell, B. H., *Anal. Chim. Acta*, 69, 143 (1974).
25. Brand, M. J. D. and Rechnitz, G. A., *Anal. Chem.*, 42, 1170 (1970).
26. Liberti, A. and Mascini, M., *Anal. Chem.*, 41, 676 (1969).
27. Isbell, A. F., Jr., Pecsok, R. L., Davies, R. H., and Purnell, J. H., *Anal. Chem.*, 45, 2363 (1973).
28. Deming, W. E., *Statistical Adjustment of Data*, John Wiley & Sons, New York, 1943.
29. Wentworth, W. E., *J. Chem. Educ.*, 42, 96 (1965).
30. Wentworth, W. E., *J. Chem. Educ.*, 42, 162 (1965).
31. Barry, D. M. and Meites, L., *Anal. Chim. Acta*, 68, 435 (1974).
32. Ingman, F., Johansson, A., Johansson, S., and Karlson, R., *Anal. Chim. Acta.*, 64, 113 (1973).
33. Murtlow, D., M.S.thesis, Clarkson College of Technology (1976).
34. Murtlow, D. and Meites, L., *Anal. Chim. Acta*, 92, 285 (1977).
35. Potter, S., *Anti-Woo*, The New American Library, New York, 1966, 50.
36. Sato, H. and Momoki, K., *Anal. Chem.*, 42, 1477 (1970).
37. Meites, L., *Anal. Chim. Acta*, 74, 177 (1975).
38. Shia, G. A., Clarkson College of Technology, unpublished research.
39. Meites, L. and Barry, D. M., *Talanta*, 20, 1173 (1973).
40. Meites, L. and Shia, G. A., in *Chemometrics: Theory and Application*, Kowalski, B. R., Ed., Symposium Series No. 52, American Chemical Society, Washington, 1977, 127.
41. Campbell, B. H., Meites, L., and Carr, P. W., *Anal. Chem.*, 46 (3), 386 (1974).
42. Isenhour, T. L., Kowalski, B. R., and Jurs, P. C., *Crit. Rev. Anal. Chem.*, 4, 1 (1974).
43. Jurs, P. C. and Isenhour, T. L., Chemical Applications of Pattern Recognition, John Wiley & Sons, New York, 1975.
44. Campbell, B. H. and Meites, L., *Talanta*, 21, 393 (1974).
45. Piljac, I., Grabarić, B., and Filipović, I., *J. Electroanal. Chem.*, 42, 433 (1973).
46. Meites, L., *Talanta*, 22, 733 (1975).
47. Meites, L. and Matijević, E., *Anal. Chim. Acta*, 76, 423 (1975).
48. DeFord, D. D. and Hume, D. N., *J. Am. Chem. Soc.*, 73, 5321 (1951).
49. Lampugnani, L., Meites, L., Papoff, P., and Rotunno, T., *Ann. Chim.* (Rome), 65, 257 (1975).
50. *Cf.*, for example, B. G. M. Vandeginste and L. DeGalan, *Anal. Chem.*, 47, 2124 (1975) and the references to the prior literature cited therein.
51. Gold, H. S., Rechsteiner, C. E., and Buck, R. P., *Anal. Chem.*, 48, 1540 (1976).