

A Global Strategy for Nonlinear Least Squares

JOHN A. JACQUEZ

*Department of Physiology, The Medical School, and
Department of Biostatistics, The School of Public Health
The University of Michigan
Ann Arbor, Michigan*

ABSTRACT

A general strategy for attacking problems in nonlinear least squares is developed. Parameters are classified as linear or nonlinear, depending on whether they appear linearly or nonlinearly in the functional expression being fitted to a set of data.

Basically the strategy consists of transforming the functional expression so as to maximize the number of linear parameters and then solving the problem in a two-stage process. For given values of the nonlinear parameters the linear parameters are first defined as functions of the nonlinear parameters by the solution of a linear regression. The nonlinear parameters are then found by minimizing the usual quadratic form with the use of standard search techniques.

1. INTRODUCTION

Consider the following well-known problem. We are given the function

$$y = f(x, K) \quad (1)$$

where K is an unknown vector of parameters. If $f(x, K)$ is of the form shown in Eq. (2)

$$y = k_1 g_1(x) + k_2 g_2(x) + \cdots + k_p g_p(x), \quad (2)$$

it can be rewritten as in Eq. (3)

$$y = k_1 z_1 + k_2 z_2 + \cdots + k_p z_p \quad (3)$$

and we have a problem in linear regression; the parameters k_1, \dots, k_p are said to be linear parameters. Parameters that appear in any other way are said to be nonlinear parameters. We are given a set of measurements $\{(x_i, y_i); i = 1, \dots, n\}$, $n \geq p$. The x_i are assumed to be free of

Mathematical Biosciences 7 (1970), 1-8

Copyright © 1970 by American Elsevier Publishing Company, Inc.

error but the y_i include errors of measurement. Actually we need not assume that x_i is error free but that the observed values x_i of the independent variable can be chosen and set by the experimenter [1]. Thus we assume that a measurement is given by Eq. (4),

$$y_i = f(x_i, K) + \varepsilon_i \quad (4)$$

where ε_i is a random sample from some error distribution. More generally, if replicates are taken at x_i , the j th replicate at the i th point is described by (5)

$$y_{ij} = f(x_i, K) + \varepsilon_{ij}. \quad (5)$$

Further, we assume that we take as our estimate of K the value obtained by minimizing the quadratic form (6),

$$Q = \sum_i [y_i - f(x_i, K)]^2 \quad (6)$$

or in case replicates are available, the quadratic form (7).

$$Q = \sum_i \sum_j [y_{ij} - f(x_i, K)]^2. \quad (7)$$

The minimum of Q is usually sought by setting the derivatives $\partial Q / \partial k_j$ equal to zero and solving the p simultaneous equations (8).

$$0 = \sum_i (y_i - f_i) \frac{\partial f_i}{\partial k_j} \quad (8)$$

where

$$f_i = f(x_i, K) \quad \text{and} \quad \frac{\partial f_i}{\partial k_j} = \left. \frac{\partial f(x, K)}{\partial k_j} \right|_{x=x_i}.$$

We will call the set of simultaneous equations (8) the *normal equations*.

If $f(x, K)$ is of the form of Eq. (3), we have a problem in linear regression and for such a problem a complete theory and methodology for finding the solution is available. Assume a solution exists. Let Z be the $n \times p$ matrix of elements $z_{ij} = z_j(x_i)$ and let Y be the n -component column vector $Y = [y_1, y_2, \dots, y_n]^T$. The normal equations are (9),

$$Z^T Z K = Z^T Y, \quad (9)$$

and the solution is $K = [Z^T Z]^{-1} Z^T Y$.

If at least one of the k_j appears nonlinearly in Eq. (1), we have a problem in nonlinear regression. There is no general theory for finding closed solutions of problems in nonlinear regression, nor are there uniformly successful iterative numerical methods. A large number of different methods have been proposed for solving nonlinear search problems. The many methods available [2-5] can be classified under the

general headings (1) gradient methods, (2) methods that use an iterative linearization technique, (3) multivariate Newton–Raphson methods, (4) iterative methods of direct solution of the normal equations, (5) pattern searches, (6) iterative sequential paucivariate searches. The very multiplicity of methods for nonlinear least squares attests to the difficulties of the problem. Properly, these many methods represent the tactics of attacking problems in nonlinear least squares. But no matter what tactic or combination of tactics is used, two significant features stand out above all others. The difficulties in finding solutions to these problems increase enormously with increase in the dimensions of the parameter space. Furthermore, for all of these methods we need an initial estimate of the solution, and performance of these methods depends markedly on how good the initial estimate is. For these reasons it seems appropriate not to look for more tactics but to seek a more general approach, a strategy for attacking nonlinear problems in least squares.

2. A GLOBAL STRATEGY

Generally, good strategies are obtained once we can define strategic principles that focus on the major difficulties in a class of problems. The strategic principles for problems in nonlinear least squares follow almost immediately from the previous discussion:

1. Whenever possible, transform the problem to a problem in linear regression.
2. Reduce the dimensionality of any nonlinear search as much as possible.

These principles may be implemented in the following manner. Suppose that Eq. (4) describes the measurements and that the sum of squares to be minimized in K -space is Eq. (6). Of the p parameters k_1, k_2, \dots, k_p , some may appear linearly in f , others nonlinearly. The trick is to find a transformation that maximizes the number of parameters that appear linearly. No rules can be laid down on how to do this but this is the sort of facility gained from experience. Suppose the transformation is such that z is the new dependent variable, v is the new independent variable, and there are p parameters that may be functions of the parameters k_1, k_2, \dots, k_p . We label the new parameters; suppose $\alpha_1, \dots, \alpha_r$ appear linearly and $\gamma_1, \dots, \gamma_s$ appear nonlinearly, so that the transformed model is given by Eq. (10)

$$z_i = \alpha_1 g_1(v_i, \gamma) + \alpha_2 g_2(v_i, \gamma) + \dots + \alpha_r g_r(v_i, \gamma) + \eta_i \quad (10)$$

where η_i now represents the error in the transformed model. Note that

on the left-hand side of (10), z_i may also be a function of the nonlinear parameters $\gamma_1, \dots, \gamma_s$. Choose an initial estimate for γ ; then the linear parameters can be considered to be functions of γ determined as the minimum of the sum of squares.

$$Q^* = \sum_i \left[z_i - \sum_j \alpha_j g_j(v_i, \gamma) \right]^2. \quad (11)$$

Thus α is determined as an implicit function of γ in a one-step linear regression and the dimensionality of the problem is reduced by r . In effect, Q^* is used to restrict the search to a subspace of the parameter space. It is important that the minimization in γ be carried out in terms of the original sum of squares, expressed now as a function of γ .

$$Q = \sum_i [y_i - f(x_i, \gamma)]^2. \quad (12)$$

The sum of squares surface (12) in γ -space is the locus of the minima of (11). Thus the search is now reduced to a two-stage process, one stage being the solution of a simple linear regression. Starting with estimates $\gamma_{11}, \dots, \gamma_{s1}$, determine $\alpha_{11}, \dots, \alpha_{r1}$ as the minima of (11) and then calculate the original sum of squares (12). Then, use one of the standard search methods in γ -space but each time the sum of squares (12) is to be determined, the appropriate α is first found by minimizing (11). If a search method is used that requires the calculation of derivatives, it is usually easier to use finite difference approximations to the derivatives because α is determined implicitly as a function of γ .

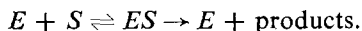
3. EXAMPLES

A. Enzyme Kinetics

The initial velocity of many one-substrate enzyme reactions is adequately described by an equation of the form of Eq. (13) [6-8].

$$v = \frac{Vx}{K + x}. \quad (13)$$

In Eq. (13) v is the initial velocity of an enzyme reaction that follows the reaction scheme



S stands for the substrate, which is present at concentration x , and E stands for the enzyme; V is the maximal velocity of the reaction and K is the concentration at which $v = V/2$. Equation (13) is derived by writing the differential equations for the reaction scheme and assuming that initial conditions hold (i.e., no products present) and that the concentration of enzyme-substrate complex is constant. The latter condition is

usually referred to as the steady-state, or more often, the pseudo-steady-state, assumption [9, 10]. Data on the initial velocity of an enzyme reaction $\{(x_i, v_i); i = 1, \dots, n\}$ are usually fitted with use of a transformation to a linear regression as in Eq. (14) or (15).

$$\frac{1}{v} = \frac{1}{V} + \frac{K}{V} \left(\frac{1}{x} \right); \quad (14)$$

$$\frac{x}{v} = \frac{K}{V} + \frac{1}{V} x. \quad (15)$$

Usually K/V and $1/V$ are determined by fitting Eq. (14) or (15) to the transformed data points $(1/x_i, 1/v_i)$ or $(x_i, x_i/v_i)$ and K and V are calculated therefrom. Note that the values of K and V obtained in this way are not necessarily the same as those obtained by fitting Eq. (13) directly; in practice the difference is often small in terms of the experimental error. The difference arises from the transformation in the errors. Equation (13) can also be used to fit the concentration of binding sites per unit surface area that are combined with a chemical compound that is present at concentration x in the medium around the surface.

For multisubstrate reactions the velocity can be written as a rational function of powers and cross-products of various substrate concentrations [11–16] and hence may be written as in Eq. (16).

$$v = \frac{u_0 + \alpha_1 u_1 + \alpha_2 u_2 + \dots + \alpha_n u_n}{\beta_0 u_0 + \beta_1 u_1 + \dots + \beta_n u_n + \beta_{n+1}}. \quad (16)$$

With use of the inverse transformation $1/v_i$, there are $n + 2$ linear parameters $\beta_0, \dots, \beta_{n+1}$ and n nonlinear parameters $\alpha_1, \dots, \alpha_n$.

B. Active Transport

A simple example of a function that appears in the study of active transport [17–19] of certain compounds by cells is shown in Eq. (17).

$$y = \frac{x_1}{Ax_1 + Bx_2 + C} + k(x_1 - x_2). \quad (17)$$

The data obtained experimentally are (x_{1i}, x_{2i}, y_i) for $i = 1, \dots, n$. Let

$$v = \frac{x_1}{y - k(x_1 - x_2)}. \quad (18)$$

Then, Eq. (19)

$$v_i = Ax_{1i} + Bx_{2i} + C \quad (19)$$

is a linear regression with sum of squares Q^* .

$$Q^* = \sum_i [v_i - Ax_{1i} - Bx_{2i} - C]^2. \quad (20)$$

The original sum of squares is given by (21).

$$Q = \sum_i \left[y_i - \frac{x_{1i}}{Ax_{1i} + Bx_{2i} + C} - k(x_{1i} - x_{2i}) \right]^2. \quad (21)$$

Thus if we choose a value for k , say k_1 , this value determines v_i and A , B , and C are obtained as the minimum for Q^* , Eq. (20). Thus what appears at first to be a four-parameter nonlinear least squares problem reduces to a one-dimensional nonlinear least squares and a search over a range of values of k is easily carried out to minimize the original sum of squares (21).

C. Fitting Sums of Exponentials

Frequently it is necessary to fit a sum of exponentials of the form of Eq. (22) to a set of data.

$$y_i = \sum_{j=1}^m A_j \exp(\alpha_j x_i). \quad (22)$$

In this case the parameter set is already divided into m linear parameters A_1, \dots, A_m and m nonlinear parameters $\alpha_1, \dots, \alpha_m$ and there is no need to use any transformation. Given the data set $\{(x_i, y_i); i = 1, \dots, n\}$, choose a set of initial values for α . Then the solution of the remaining linear regression in A_1, A_2, \dots, A_m minimizes the original sum of squares for the given α .

$$Q = \sum_i \left[y_i - \sum_j A_j \exp(\alpha_j x_i) \right]^2. \quad (23)$$

Thus the problem reduces to a search in α -space.

As defined by Eq. (22) there are multiple minima due to symmetry properties but these can be avoided by redefining the parameters in the exponents. We avoid the details of this problem because it is a long digression from the main point of this article.

4. SOME UNFINISHED PROBLEMS

An interesting problem arises when a transformation can be found to convert the problem to one in linear regression, as for the initial velocity of one-substrate enzyme reactions (Eq. (13)). The minimum in the transformed problem is usually different from but near to the minimum in the original nonlinear problem. This arises because the transformation

involves a transformation in the distribution of errors and the least squares method gives equal weight to each experimental point. This can be minimized by taking replicates at each point and using the inverse of the estimated error variance to weight each point in the sum of squares. In practice the easiest way around this difficulty is to use the parameter estimates obtained from the transformed problem as initial estimates for the original nonlinear least squares problem. Surprisingly, the two minima often do not differ by much.

It seems intuitively obvious that if the sum of squares surface is unimodal in the parameter space, the two-stage process must converge on the minimum. However, some more theoretical work is needed on defining as generally as possible the conditions under which the two-stage process must converge to the minimum of the sum of squares surface of the parameter space in the original nonlinear least squares problem.

ACKNOWLEDGMENTS

This work was supported in part by grants GM 00892 and CA 06734 from the National Institutes of Health.

REFERENCES

- 1 J. Berkson, Are there two regressions? *J. Amer. Statist. Assoc.* **45**(1950), 164–180.
- 2 N. Draper and H. Smith, *Applied regression analysis*. Wiley, New York, 1966.
- 3 D. J. Wilde, *Optimum seeking methods*. Prentice-Hall, New Jersey, 1964.
- 4 D. J. Wilde, *Foundations of optimization*. Prentice-Hall, New Jersey, 1967
- 5 J. A. Jacquez, *A first course in computing and numerical methods*. Addison-Wesley, Reading, Massachusetts, 1970 (in press).
- 6 L. Michaelis and M. L. Menten, Die Kinetic der Invertinwirkung, *Biochem. Z.* **49**(1913), 333–369.
- 7 G. E. Briggs and J. B. S. Haldane. A note on the kinetics of enzyme action, *Biochem. J.* **19**(1925), 338–339.
- 8 M. Dixon and E. C. Webb, *Enzymes*. Academic Press, New York, 1964.
- 9 J. O. Hirschfelder, Pseudostationary state approximation in chemical kinetics, *J. Chem. Phys.* **26**(1957), 271–273.
- 10 F. G. Heineken, H. M. Tsuchiya, and R. Aris, On the mathematical status of the pseudo-steady state hypothesis of biochemical kinetics, *Math. Biosci.* **1**(1967), 95–113.
- 11 E. L. King and C. Altman, A schematic method of deriving the rate laws for enzyme-catalyzed reactions, *J. Phys. Chem.* **60**(1956), 1375–1378.
- 12 J. T. F. Wong and C. S. Hanes, Kinetic formulations for enzymic reactions involving two substrates, *Canad. J. Biochem. Physiol.* **40**(1962), 763–804.
- 13 W. W. Cleland, The kinetics of enzyme-catalyzed reactions with two or more substrates or products, I, *Biochim. Biophys. Acta* **67**(1963), 104–137.

- 14 M. V. Volkenstein and B. N. Goldstein, A new method for solving the problems of the stationary kinetics of enzymological reactions, *Biochim. Biophys. Acta* **115**(1966), 471–477.
- 15 S. Cha, A simple method for derivation of rate equations for enzyme-catalyzed reactions under the rapid equilibrium assumption or combined assumptions of equilibrium and steady state, *J. Biol. Chem.* **243**(1968), 820–825.
- 16 D. D. Fisher and A. R. Schulz, Connection matrix representation of enzyme reaction sequence, *Math. Biosci.* **4**(1969), 189–200.
- 17 J. A. Jacquez, The kinetics of carrier-mediated active transport of amino acids, *Proc. Natl. Acad. USA* **47**(1961), 153–163.
- 18 J. A. Jacquez, The kinetics of carrier-mediated transport: Stationary state approximations, *Biochim. Biophys. Acta* **79**(1964), 318–328.
- 19 J. T. F. Wong, The possible role of polyvalent carriers in cellular transports, *Biochim. Biophys. Acta* **94**(1965), 102–113.