Matrix Formulation for Linear and First-Order Nonlinear Regression Analysis with Multidimensional Functions

Eduardo Fontana
Dep. de Eletrônica e Sistemas
Universidade Federal de Pernambuco
UFPE – CTG - Bloco A, Sala 421, Recife – PE 50740-550 Brasil
E-mail: fontana@ufpe.br

Published: January 20, 2014

***Abstract**:* A matrix formulation is presented to allow linear and first-order nonlinear regression analysis with multidimensional model functions. Examples of pseudocodes are presented illustrating the implementation of the formulation for numerical computation.

1. Introduction

Regression analysis is an important tool for the determination of the relationship between variables from experimental data measured in physical, biological, statistical and other phenomena. If a good model function is known to govern the experiment, parameters can be inferred by fitting the model function to the data.

The formulation broadly available in the literature employs a summation procedure that is generalized to a matrix method to derive the expressions that lead to the final set of equations for the unknown parameters [1]. In this paper, one employs a matrix formulation [2] directly to ease the development of structured algorithms. The formulation is derived for both linear and first-order nonlinear regression analysis. Examples are shown on how to determine the intermediate matrices that lead to the solution of the problem for both the linear and nonlinear cases. A pseudocode is presented for the development of simple nonlinear regression algorithms.

2. Statement of the Problem

Consider a real function $f\left(\tilde{X},\tilde{A}\right)$ modeling a physical quantity *w*, with

$$\tilde{X}=\begin{pmatrix} x_1 & x_2 & \dots & x_M \end{pmatrix}^T, \tag{1}$$

representing an independent variable, *M*-element, column matrix[1] and

---

[1] In the present formulation matrices are represented as capital letters with an appended tilde symbol. The transpose of a matrix is represented by adding the superscript *T* to the matrix. The *m*-th row is represented by appending the subscript *m* to the matrix , and the *n*-th column, by appending the superscript <*m*>.

$$\tilde{A} = \begin{pmatrix} a_1 & a_2 & \ldots & a_L \end{pmatrix}^T, \tag{2}$$

a column matrix with L parameters that define the model function.

For instance, for the model function

$$f = a + bx + cy + dxy \,,$$

M=2 and L=4 and the matrices $\tilde{X}$ and $\tilde{A}$ are

$$\tilde{X} = \begin{pmatrix} x & y \end{pmatrix}^T$$

$$\tilde{A} = \begin{pmatrix} a & b & c & d \end{pmatrix}^T.$$

In the following $\tilde{X}$ will be referred to as the **independent vector** and $\tilde{A}$, as the **parameter vector**. In practice, the physical quantity w is measured on N distinct points represented as a column matrix

$$\tilde{W} = \begin{pmatrix} w_1 & w_2 & \ldots & w_N \end{pmatrix}^T. \tag{3}$$

The problem in regression analysis is to determine the parameter vector $\tilde{A}$ yielding a function f that best describes the measured quantity w. One way to solve this problem is to determine a solution $\tilde{A}$ that minimizes the square error [1]. This parameter, in matrix notation, can be written in the form

$$\xi \equiv \left(\tilde{W} - \tilde{F}\right)^T \left(\tilde{W} - \tilde{F}\right), \tag{4}$$

with

$$\tilde{F} = \begin{pmatrix} f\left(\tilde{Y}^{\langle 1 \rangle}, \tilde{A}\right) & f\left(\tilde{Y}^{\langle 2 \rangle}, \tilde{A}\right) & \ldots & f\left(\tilde{Y}^{\langle N \rangle}, \tilde{A}\right) \end{pmatrix}^T \tag{5}$$

representing a column matrix in which the n-th element is the function f evaluated at the point

$$\tilde{X} = \tilde{Y}^{\langle n \rangle}, \tag{6}$$

with $\tilde{Y}^{\langle n \rangle}$ representing the n-th column of the $M \times N$ matrix

$$\tilde{Y} = \begin{pmatrix} \tilde{Y}^{\langle 1 \rangle} & \tilde{Y}^{\langle 2 \rangle} & \ldots & \tilde{Y}^{\langle N \rangle} \end{pmatrix} \tag{7}$$

Given that ξ is a positive definite function, if a local minimum exists, it can be obtained by imposing the condition

$$\tilde{\nabla}\xi = \tilde{0} \tag{8}$$

with

$$\tilde{\nabla} \equiv \begin{pmatrix} \dfrac{\partial}{\partial a_1} & \dfrac{\partial}{\partial a_2} & \cdots & \dfrac{\partial}{\partial a_L} \end{pmatrix}^T \tag{9}$$

representing the nabla operator, in the L-dimensional parameter space, and

$$\tilde{0} \equiv \begin{pmatrix} 0 & 0 & \dots & 0 \end{pmatrix}^T, \tag{10}$$

the L-dimensional null vector.

Expression (8) represents a set of L-equations on the L unknown parameters. If f is a good model function for the problem in general only minima will be obtained. The best solution can then be chosen as that yielding the smallest value of ξ.

As detailed in the following sections, an exact solution can be obtained for the case of model functions that are linear in the parameters. For nonlinear model functions a numerical solution has to be obtained.

3. Linear Regression

A function that is linear in the elements of the parameter vector can be written in the form

$$f\left(\tilde{X},\tilde{A}\right) = \left[\tilde{G}^{\langle X\rangle}\right]^T \tilde{A}, \tag{11}$$

with

$$\tilde{G}^{\langle X\rangle} = \begin{pmatrix} g_1\left(\tilde{X}\right) & g_2\left(\tilde{X}\right) & \dots & g_L\left(\tilde{X}\right) \end{pmatrix}^T, \tag{12}$$

being an L-element vector that defines the distinct functions associated to the L elements of the parameter vectors. Superscript X is used to allow defining a more general rectangular G matrix from which $\tilde{G}^{\langle X\rangle}$ would represent a given column, associated with the independent vector $\tilde{X}$. This can be better understood by noticing that (5) can be cast into the form

$$\tilde{F} = \tilde{G}^T \tilde{A}, \tag{13}$$

with

$$\tilde{G} = \begin{pmatrix} \tilde{G}^{\langle 1 \rangle} & \tilde{G}^{\langle 2 \rangle} & \ldots & \tilde{G}^{\langle N \rangle} \end{pmatrix}, \tag{14}$$

representing an $L \times N$ matrix with the *n*-th column being defined according to (12), i.e.,

$$\tilde{G}^{\langle n \rangle} = \begin{pmatrix} g_1\left(\tilde{Y}^{\langle n \rangle}\right) & g_2\left(\tilde{Y}^{\langle n \rangle}\right) & \ldots & g_L\left(\tilde{Y}^{\langle n \rangle}\right) \end{pmatrix}^T \tag{15}$$

For example, for the function

$$f = a\exp(xy) + b\cos\left(x + y\right) + c\sin\left(xyz\right) + d\,,$$

one has M=3, L =4 and one could define, according to (1), (2), (11) and (12),

$$\tilde{X} = \begin{pmatrix} x & y & z \end{pmatrix}^T,$$

$$\tilde{A} = \begin{pmatrix} a & b & c & d \end{pmatrix}^T.$$

$$g_1\left(\tilde{X}\right) = \exp\left(x_1 x_2\right),$$

$$g_2\left(\tilde{X}\right) = \sin\left(x_1 + x_2\right),$$

$$g_3\left(\tilde{X}\right) = \sin\left(x_1 x_2 x_3\right),$$

$$g_4\left(\tilde{X}\right) = 1\,,$$

i.e.,

$$\tilde{G}^{\langle X \rangle} = \begin{pmatrix} \exp\left(x_1 x_2\right) & \sin\left(x_1 + x_2\right) & \sin\left(x_1 x_2 x_3\right) & 1 \end{pmatrix}^T,$$

with $x_1 = x,\ x_2 = y,\ x_3 = z$.

To determine the solution to the parameter vector, the differential operation given by (8) is first computed using (4) yielding

$$\tilde{\nabla}\xi = \tilde{\nabla}\left(\tilde{W}^T\tilde{W} - \tilde{W}^T\tilde{F} - \tilde{F}^T\tilde{W} + \tilde{F}^T\tilde{F}\right). \tag{16}$$

Noticing that $\tilde{W}$ is independent of the parameter vector, one obtains

$$\tilde{\nabla}\left(\tilde{W}^T\tilde{F} + \tilde{F}^T\tilde{W} - \tilde{F}^T\tilde{F}\right) = \tilde{0} \tag{17}$$

The *l*-th element of (17) is of the form

$$\frac{\partial}{\partial a_l}\left(\tilde{W}^T\tilde{F}+\tilde{F}^T\tilde{W}-\tilde{F}^T\tilde{F}\right)=0\,,$$

or equivalently

$$\frac{\partial \tilde{F}^T}{\partial a_l}\left(\tilde{W}-\tilde{F}\right)=0 \tag{18}$$

For the latter expression, the property

$$\tilde{A}^T\tilde{B}=\tilde{B}^T\tilde{A}\,, \tag{19}$$

was used.

Before proceeding, it is instructive to introduce the generalized derivative operator, for a vector of K elements,

$$\tilde{H}'\equiv\begin{pmatrix} \tilde{\nabla}h_1 & \tilde{\nabla}h_2 & \dots & \tilde{\nabla}h_K \end{pmatrix}^T. \tag{20}$$

Expression (27) represents a $K\times L$ matrix. According to this definition, setting $\tilde{H}=\tilde{F}$, (18) can be grouped as set of L equations of the form

$$\tilde{F}'^T\left(\tilde{W}-\tilde{F}\right)=\tilde{0}\,, \tag{21}$$

with

$$\tilde{F}'^T=\begin{pmatrix} \tilde{\nabla}f\left(\tilde{Y}^{\langle 1\rangle},\tilde{A}\right) & \tilde{\nabla}f\left(\tilde{Y}^{\langle 2\rangle},\tilde{A}\right) & \dots & \tilde{\nabla}f\left(\tilde{Y}^{\langle N\rangle},\tilde{A}\right) \end{pmatrix}, \tag{22}$$

representing an $L\times N$ matrix and $\tilde{0}$ the null vector of length L.

Notice that, by use of (13),

$$\frac{\partial \tilde{F}}{\partial a_l}=\tilde{G}^T\frac{\partial \tilde{A}}{\partial a_l}\,,$$

or equivalently

$$\frac{\partial \tilde{F}^T}{\partial a_l}=\frac{\partial \tilde{A}^T}{\partial a_l}\tilde{G}\,, \tag{23}$$

which can be generalized, by use of the definition (20) to the relation

$$\tilde{F}'^T=\tilde{A}'^T\tilde{G}\,. \tag{24}$$

By applying the definition (22) to vector $\tilde{A}$, it is straightforward to show that

$$\tilde{A}'=\tilde{\Im}_L\,, \tag{25}$$

with $\tilde{\Im}_L$ representing the $L\times L$ identity matrix. From (24) and (25) one obtains

$$\tilde{F}'^T=\tilde{G}\,, \tag{26}$$

Inserting (26) into (21) yields From (13), (18) can be cast into the form

$$\tilde{G}\left(\tilde{W}-\tilde{F}\right)=\tilde{0} \tag{27}$$

By inserting (13) into (23) one obtains, after a few algebraic manipulations, the final solution for the unknown parameter vector,

$$\tilde{A}=\left(\tilde{G}\tilde{G}^{T}\right)^{-1}\left(\tilde{G}\tilde{W}\right) \tag{28}$$

Notice that $\tilde{G}\tilde{G}^{T}$ is an $L\times L$ square matrix and $\tilde{G}\tilde{W}$ is a vector of length *L*.

4. First-Order Nonlinear Regression

For a function that is nonlinear on the elements of the parameter vector, one can obtain an approximate solution by making a first-order linear approximation for the model vector. Small corrections are obtained by iteration. At the *k*-th iteration step, one obtains a set of *L* linear equations for the correction to be made on the parameter vector. In order to develop the procedure, let $\tilde{A}(k)$ the parameter vector at the *k*-th iteration step. Given this vector, one expands the model function, to first order in the form

$$f\left(\tilde{X},\tilde{A}\right)=f\left[\tilde{X},\tilde{A}(k)\right]+\tilde{\nabla}^{T}f\left[\tilde{X},\tilde{A}(k)\right]\left[\tilde{A}-\tilde{A}(k)\right]. \tag{29}$$

Using the definition (5), (29) can be generalized to the form

$$\tilde{F}\left(\tilde{A}\right)=\tilde{F}(k)+\tilde{F}'(k)\left[\tilde{A}-\tilde{A}(k)\right] \tag{30}$$

with

$$\tilde{F}'(k)\equiv\begin{pmatrix} \tilde{\nabla}f\left(\tilde{Y}^{\langle 1\rangle},\tilde{A}(k)\right) & \tilde{\nabla}f\left(\tilde{Y}^{\langle 2\rangle},\tilde{A}(k)\right) & \dots & \tilde{\nabla}f\left(\tilde{Y}^{\langle N\rangle},\tilde{A}(k)\right)\end{pmatrix}^{T} \tag{31}$$

representing the $N\times L$ generalized derivative matrix of the vector $\tilde{F}$, according to the definition (20), as per (22). In (30) one uses the notation $\tilde{F}(k)=\tilde{F}\left[\tilde{A}(k)\right]$.

In order to determine the correction to be obtained at each iteration step, one considers the condition given by (21), re-written as

$$\left[\tilde{F}'\left(\tilde{A}\right)\right]^T\left[\tilde{W}-\tilde{F}\left(\tilde{A}\right)\right]=\tilde{0}. \quad (32)$$

Using (30) in the above equation and noticing that to first order

$$\tilde{F}'\left(\tilde{A}\right)\approx\tilde{F}'(k),$$

yields

$$\left[\tilde{F}'(k)\right]^T\left\{\tilde{W}-\tilde{F}(k)-\tilde{F}'(k)\left[\tilde{A}-\tilde{A}(k)\right]\right\}=\tilde{0}, \quad (33)$$

which after a few algebraic manipulations yields,

$$\Delta\tilde{A}(k)=\left\{\left[\tilde{F}'(k)\right]^T\tilde{F}'(k)\right\}^{-1}\left\{\left[\tilde{F}'(k)\right]^T\left[\tilde{W}-\tilde{F}(k)\right]\right\}, \quad (34)$$

with

$$\Delta\tilde{A}\left(k\right)\equiv\tilde{A}-\tilde{A}\left(k\right), \quad (35)$$

representing the differential correction in the parameter vector.

Expression (34) is calculated iteratively until differential correction becomes smaller than a certain preset error parameter.

In the following section, examples on how to define the matrices for both the linear and nonlinear cases are shown. A pseudocode is shown for the development of algorithms for the case of nonlinear regression analysis of data.

5. Examples

5.1 Linear regression example

Consider once again the example of Section 3 with

$$f=a\exp(xy)+b\cos\left(x+y\right)+c\sin\left(xyz\right)+d,$$

one has M=3, L =4. Assume that N values are obtained for the quantity w, i.e.

$$\tilde{W} = \begin{pmatrix} w_1 & w_2 & \dots & w_N \end{pmatrix}^T. \tag{36}$$

From (7)

$$\tilde{Y} = \begin{pmatrix} x_1 & x_2 & \dots & x_N \\ y_1 & y_2 & \dots & y_N \\ z_1 & z_2 & \dots & z_N \end{pmatrix}. \tag{37}$$

From (12)

$$\tilde{G}^{\langle X \rangle} = \begin{pmatrix} \exp(xy) & \sin(x+y) & \sin(xyz) & 1 \end{pmatrix}^T \tag{38}$$

and from (15),

$$\tilde{G} = \begin{pmatrix} \exp(x_1 y_1) & \exp(x_2 y_2) & \dots & \exp(x_N y_N) \\ \sin(x_1 + y_1) & \sin(x_2 + y_2) & \dots & \sin(x_N + y_N) \\ \sin(x_1 y_1 z_1) & \sin(x_2 y_2 z_2) & \dots & \sin(x_N y_N z_N) \\ 1 & 1 & 1 & 1 \end{pmatrix} \tag{39}$$

One obtains a $4 \times N$ G matrix and expressions (36) and (39) are sufficient to determine the solution given by (28).

5.2 Nonlinear regression example

For the sake of simplicity, consider the 3-parameter gaussian model for a single variable function

$$f(x, \tilde{A}) = a \exp\left\{ -\left[(x - \overline{x}) / w \right]^2 \right\}, \tag{40}$$

where a is the amplitude, $\overline{x}$ is the centroid and w is the halfwidth of the gaussian. The parameter vector is

$$\tilde{A} = \begin{pmatrix} a & \overline{x} & w \end{pmatrix}^T \tag{41}$$

Assume once again that a set of N data points is obtained, represented by (36). According to (34) all that is necessary is to determine the matrices $\tilde{F}$ and $\tilde{F}'$. Assuming that at the k-th iteration the parameter vector is given by

$$\tilde{A}(k) = \begin{pmatrix} a(k) & \overline{x}(k) & w(k) \end{pmatrix}^T. \tag{42}$$

From (5)

$$\tilde{F}(k) = \begin{pmatrix} a(k)\exp\left\{-\left\{\left[x_1 - \bar{x}(k)\right]/w(k)\right\}^2\right\} \\ a(k)\exp\left\{-\left\{\left[x_2 - \bar{x}(k)\right]/w(k)\right\}^2\right\} \\ \dots \\ a(k)\exp\left\{-\left\{\left[x_N - \bar{x}(k)\right]/w(k)\right\}^2\right\} \end{pmatrix}, \tag{43}$$

and from (31), one obtains, after a few algebraic manipulations

$$\tilde{F}'(k) = \begin{pmatrix} e^{-\left\{\left[x_1 - \bar{x}(k)\right]/w(k)\right\}^2} & 2a(k)\left\{\left[x_1 - \bar{x}(k)\right]/\left[w(k)\right]^2\right\} e^{-\left\{\left[x_1 - \bar{x}(k)\right]/w(k)\right\}^2} & 2a(k)\left\{\left[x_1 - \bar{x}(k)\right]^2/\left[w(k)\right]^3\right\} e^{-\left\{\left[x_1 - \bar{x}(k)\right]/w(k)\right\}^2} \\ e^{-\left\{\left[x_2 - \bar{x}(k)\right]/w(k)\right\}^2} & 2a(k)\left\{\left[x_2 - \bar{x}(k)\right]/\left[w(k)\right]^2\right\} e^{-\left\{\left[x_2 - \bar{x}(k)\right]/w(k)\right\}^2} & 2a(k)\left\{\left[x_3 - \bar{x}(k)\right]^2/\left[w(k)\right]^3\right\} e^{-\left\{\left[x_3 - \bar{x}(k)\right]/w(k)\right\}^2} \\ \dots & \dots & \dots \\ e^{-\left\{\left[x_N - \bar{x}(k)\right]/w(k)\right\}^2} & 2a(k)\left\{\left[x_N - \bar{x}(k)\right]/\left[w(k)\right]^2\right\} e^{-\left\{\left[x_N - \bar{x}(k)\right]/w(k)\right\}^2} & 2a(k)\left\{\left[x_N - \bar{x}(k)\right]^2/\left[w(k)\right]^3\right\} e^{-\left\{\left[x_N - \bar{x}(k)\right]/w(k)\right\}^2} \end{pmatrix}. \tag{44}$$

Together with (36), (43) and (44) are sufficient to determine the differential correction for the parameter vector.

5.3 Pseudocode for nonlinear regression

Table I shows a pseudocode for the computational implementation of the procedure, for the case of nonlinear regression analysis. As can be noticed from the program structure, by use of the matrix formulation presented in this paper one can organize the algorithm using a rather simple and modular scheme.

References

[1] Douglas M. Bates and Donald G. Watts, "Nonlinear regression analysis and its applications," John Wiley & Sons, Inc. (1988).

[2] G. B. Arfken and H. J. Weber, "Mathematical Methods for Physicists," Chapter 3, 5th. ed., San Diego, CA, Academic Press (2001).

Table I – Pseudocode for nonlinear regression analysis.

1. Initialization:
- Define model function $f(X,A)$
- Read data and store into matrix W
- Define parameter vector A
- Set initial guess to the parameter vector: $A \leftarrow A0$
- Set a high value for an error parameter: $err \leftarrow 100$
- Define **function** $F(A)$
- Define system matrix function $S(A) \leftarrow [F'(A)]^T F'(A)$
- Define input vector function $V(A) \leftarrow [F'(A)]^T [W - F(A)]$
- Set a maximum value for the change in magnitude of the parameter vector ε

2. Calculation:

$$Aout \leftarrow \left| \begin{array}{l} \text{while err} > \varepsilon \\ \quad \left| \begin{array}{l} Aold \leftarrow A \\ A \leftarrow A + S(A)V(A) \\ err \leftarrow \max\left(\overrightarrow{|A - \text{Aold}|}\right) \end{array} \right. \\ A \end{array} \right.$$

Remarks:
- Mathcad programming style is used as a model.
- The vectorize operation $|A - A\text{old}|$ produces a matrix in which each element is the absolute value of the difference between vectors
- The function max( ) calculates the maximum element of the vector.
- The last value of A stores the approximate solution to the parameter vector.