Journal of Science and Technology, 4(1): 11-17, July 2014

A STUDY ON THE NON-LINEAR CURVE FITTING BY DATA LINEARIZATION AND ANALYZING THE BEST FIT

MD. BABUL HOSSAIN*[1], MD. MUSA MIAH[1], PINAKEE DEY[1] and
MD. SHARIF UDDIN[2]

[1]Department of Mathematics, Mawlana Bhasani Science and Technology University, Sontosh,Tangail, Bangladeah, [2]Department of Mathematics, Jahangirnagar University, Savar, Dhaka, Bangladesh

Abstract

This paper presents a comparative study on non-linear curve fitting between exponential function and power function. We decide on which one is the best fit by estimating "Root-Mean-Square" error using Least Square method. First we construct Least Squares line from non-linear function by data linearization, which gives the value of the parameters of the fits. Then we compare the "Root-Mean Square" error conclude that the exponential is the best fit. This analysis is performed by the implementation of the numerical scheme with the help of computer programming and numerical solution.

Keywords: Data linearization, Least Square line, Root-Mean-Square, Curve fitting, Exponential function, Power function

Introduction

Numerical analysis is the development and study of procedures to solve mathematical problems with a computer. A major advantage for numerical analysis is that a numerical answer can be obtained even when a problem has no analytical solution. The idea of curve fitting (Mathews, 2001; Burdon and Faires, 2004) is to find a mathematical model that fits our data. We assume that we have theoretical reasons for picking a function of a certain form. The curve fit finds the specific coefficients (parameters) which make that function match our data as closely as possible. A mathematical procedure for finding the best curve fitting from a given set of points to minimizing the sum of squares of the residuals. Nonlinear curve fitting also seeks to find those parametric values that minimize the deviations between the observed y values and the expected y values. In nonlinear models, however, one usually cannot solve the equation for the parameters. Various iterative procedures are used instead. This is a time- consuming computation and the iteration involved requires good starting values for the parameters. But it is possible to linearize a non linear model. We consider the nonlinear model

$y = ae^{\left(\frac{-x}{x_0}\right)}$, where a and x_0 are parameter. Taking logarithm on both sides we get, $\log(y) = \log(a) - \frac{x}{x_0}$. Hence $\log(y)$ is a linear function of x. Then we can find easily

* Author for correspondence: babulhossainh@yahoo.com

the value of parameters after solving normal equations. The least squares approximation (Burdon and Faires, 2004) is one way to compare the errors or deviations. Least square is a mathematical optimization technique, which attempts to find a function, which closely approximates the data. Then analyze the 'Root- Mean Square' error we obtain the best fit (Jupp, 1978). All of the figures are generated by computer programming code in MATLAB 7.0 (Littlefield and Hanselman, 2005).

Technique of Error analysis and Linearization

Maximum error, Average error and Root-Mean- Square:

In science and engineering it is often the case that an experiment produces a set of data points $(x_1, y_1),(x_2, y_2),... \quad ... \quad ...,(x_n, y_n)$ where abscissas $\{x_k\}$ are distinct .One goal of numerical methods is to determine a formula $y = f(x)$ that relates these variables.

Usually, a class of allows the formula is chosen and then coefficients must be determined. We emphasize the class of linear function of the form

$$y = f(x) = Ax + B \tag{1}$$

If all the numerical values $\{x_k\}$, $\{y_k\}$ are known to several significant digits of accuracy, then polynomial interpolation can be used successfully, otherwise it cannot. Some experiments are devised using specialized equipment so that the data points will have at least five digits of accuracy. However, many experiments are done with equipment that is reliable only to three or fewer digits of accuracy. Often there is an experimental error in the measurements, and although three digits are recorded for the values $\{x_k\}$ and $\{y_k\}$,

It is realized that the true value $f(xk)$ satisfies $f(x_k) = y_k + e_k$ (2)

where e_k is the measurement error (Mathews, 2001). We find the best linear approximation form (1) that goes near (not always through) the points. The error or residuals are $e_k = f(x_k) – y_k$ for $1 \le k \le n$. There are several norms that can be used with the residuals in equation (2) to measure how far the curve $y = f(x)$ lies from the data.

Maximum error: $E_\infty(f) = \max_{1\le k\le n}\{|f(x_k)) - y_k|\}$ (3)

Average error: $E_1(f) = \frac{1}{n}\sum_{k=1}^{n}|f(x_k) - y_k|$ (4)

Root-Mean- Square Error: $E_2(f) = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\{f(x_k) - y_k\}^2}$ (5)

Least square line

The least squares method is the most convenient procedure for determining best linear approximations. The linear least squares fitting technique (Gerald and Whealthey, 1994; Burdon and Faires, 2004) is the simplest and most commonly applied form of linear regression and provides a solution to the problem of finding the best fitting straight line through a set of points. The simplest form of modeling is linear regression, which is the prediction of one variable from another when the relationship between them is assumed to be linear. The coefficients of the least square line $y = Ax + B$ are the solution of the following linear system, known as the normal equations:

$$\left.\begin{aligned}\left(\sum_{k=1}^{n} x_k^2\right)A + \left(\sum_{k=1}^{n} x_k\right)B = \left(\sum_{k=1}^{n} x_k y_k\right)\\ \left(\sum_{k=1}^{n} x_k\right)A + nB = \sum_{k=1}^{n} y_k\end{aligned}\right\} \tag{6}$$

The vertical distance d_k from the data points $(x_k y_k)$ to the point $(xk, mx_k + b)$ on the line is $dk = Ax_k + B - y_k$. Minimize the sum of the squares of the vertical distance d_k:

$$E(A,B) = \sum_{k=1}^{n}\left(Ax_k + B - y_k\right)^2 = \sum_{k=1}^{n} d_k^2 \tag{7}$$

The minimum value of $E(A, B)$ is determined by setting the partial derivatives $\frac{\partial E}{\partial A}$ and $\frac{\partial E}{\partial B}$ equal to zero (where $\{x_k\}$ and $\{y_k\}$are constant and A and B are variables). Then normal equations which minimize the errors are

$$A\sum_{k-1}^{n} x_k^2 + B\sum_{k-1}^{n} x_k - \sum_{k-1}^{n} x_k y_k = 0 \tag{8}$$

$$A\sum_{k-1}^{n} x_k + nB - \sum_{k-1}^{n} y_k = 0 \tag{9}$$

Equation (8) and (9) can be rearranged in the standard form for a system which gives the value of A and B.

Linearization of Exponential and Power Function:

Suppose an exponential function $y = ce^{Ax}$ is generated by Points. (x_1, y_1), (x_2, y_2), ..., (x_n, y_n) To linearize $y = ce^{Ax}$ (Atkinson and Kendal, 1994; Beets, 1998; Jupp, 1978; Mathews, 2001) taking logarithm and introducing new variables

$$Y = \ln(y),\ X = x \text{ and } B = \ln(c) \tag{10}$$

The linear relation between X and Y is $Y = AX + B$.

The normal equations for the Exponential function are

$$\left.\begin{aligned}\left(\sum_{k=1}^{n} x_k^2\right)A + \left(\sum_{k=1}^{n} x_k\right)B = \left(\sum_{k=1}^{n} x_k y_k\right)\\ \left(\sum_{k=1}^{n} x_k\right)A + nB = \sum_{k=1}^{n} y_k\end{aligned}\right\} \tag{11}$$

The value of A and B are determined from equation (11), now we can compute the parameter c using the value of B in equation (10) where,

$$C = e^{B} \tag{12}$$

Again we want to linearize a power function of the form

$$y = cx^{A} \tag{13}$$

Taking logarithm and introducing new variables in (13), the linear relation is

$$Y = AX + B \tag{14}$$

Where

$$Y = \ln(y),\ X = x \text{ and } B = \ln(c) \tag{15}$$

The normal equations for power function are

$$\left.\begin{aligned}\left(\sum_{k=1}^{n} x_k^2\right)A + \left(\sum_{k=1}^{n} x_k\right)B &= \left(\sum_{k=1}^{n} x_k y_k\right)\\ \left(\sum_{k=1}^{n} x_k\right)A + nB &= \sum_{k=1}^{n} y_k\end{aligned}\right\} \tag{16}$$

After have been found the value of A and B we can determine the value of parameter c from equation (15).

Results and Discussions

To demonstrate the comparison between exponential curve fit and power curve fit let us consider the data points (1.0, 2.0), (2.0, 4.0), (3.0, 8.0), (4.0, 13.0), (5.0, 21.0).

Exponential fit ($y = Ce^{Ax}$) : Our transformed equation is $Y = AX + B$, where $Y = \ln(y)$, $X = x$ and $B = \ln(c)$

The normal equations for finding A and B are

$$\left.\begin{aligned}\left(\sum_{k=1}^{n} x_k^2\right)A + \left(\sum_{k=1}^{n} x_k\right)B &= \left(\sum_{k=1}^{n} x_k y_k\right)\\ \left(\sum_{k=1}^{n} x_k\right)A + nB &= \sum_{k=1}^{n} y_k\end{aligned}\right\} \tag{17}$$

Table 1. Obtaining coefficients of the normal equations for the transformed data points {(X_k, Y_k)}

x_k	y_k	x_k	$y_k = \ln(y_k)$	X_k^2	X_kY_k
1	2	1	0.69315	1	0.69315
2	4	2	1.38630	4	2.77260
3	8	3	2.07945	9	6.23835
4	13	4	2.56495	16	10.25980
5	21	5	3.04452	25	15.22260
		$\sum X_k = 15$	$\sum X_k = 9.76834$	$\sum X_k^2 = 55$	$\sum X_k Y_k = 35.1868$

Now the normal equations are reduced to linear system which can be written as

$$\left.\begin{aligned} 55A + 15B &= 35.1868 \\ 15A + 5B &= 9.76834 \end{aligned}\right\} \tag{18}$$

Solving the above linear system we found the value of A and B are 0.588178 and 0.189134 respectively. Then C is obtained by exp (0.189134) = 1.2082 and the exponential fit $y = 1.2082e^{0.58818x}$.

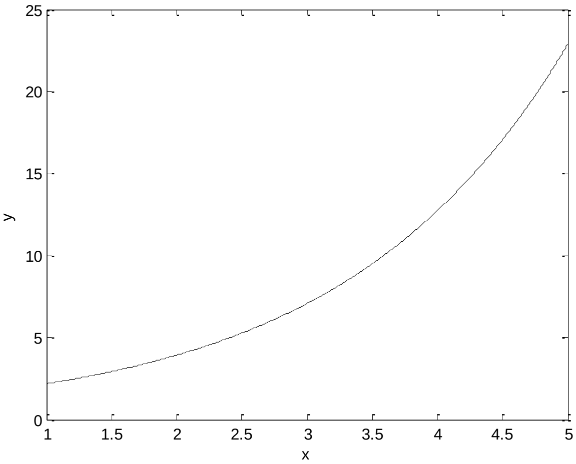

Fig. 1. Exponential fit.

Power fit ($y = Cx^A$) : Our transformed equation is $Y = AX + B$; where $Y = \ln(y)$, $X = x$ and $B = \ln(c)$.

The normal equations for finding A and B are

$$\left.\begin{aligned} \left(\sum_{k=1}^{n} x_k^2\right)A + \left(\sum_{k=1}^{n} x_k\right)B &= \left(\sum_{k=1}^{n} x_k y_k\right) \\ \left(\sum_{k=1}^{n} x_k\right)A + nB &= \sum_{k=1}^{n} y_k \end{aligned}\right\} \tag{19}$$

Table 2. Obtaining coefficients of the normal equations for the transformed data points $\{(X_k, Y_k)\}$

x_k	y_k	x_k	$y_k = \ln(y_k)$	X_k^2	X_kY_k
1	2	0	0.69315	0	0
2	4	0.69315	1.38630	0.48045	0.96091
3	8	1.09861	2.07945	1.20695	2.28451
4	13	1.38629	2.56495	1.92180	3.55577
5	21	1.60943	3.04452	2.59027	4.89995
		4.78748 = $\sum X_k$	9.76834 = $\sum X_k$	6.19947 = $\sum X_k^2$	11.70114 = $\sum X_k Y_k$

Consequently the resulting linear system is

$$\left.\begin{aligned} 6.1994\,A + 4.78748\,B &= 11.70114 \\ 7.78748\,A + 5B &= 9.76834 \end{aligned}\right\} \qquad (20)$$

The solution of the linear system is A = 1.45343 and B = 0.56202. Then C is obtained by C = exp (0.56202) = 1.75421, and we obtain the power fit $y = 1.75421x^{1.45343}$.

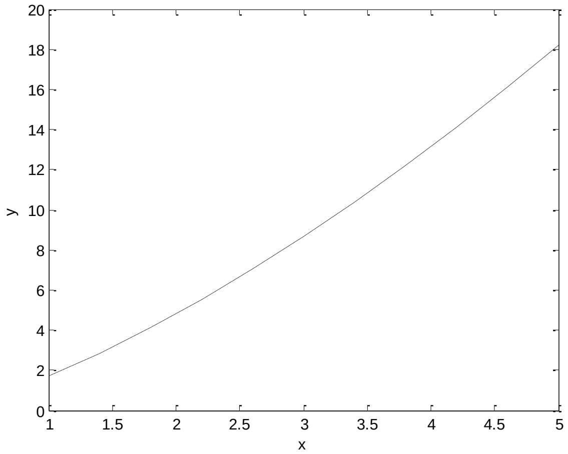

Fig. 2. Power Fit.

Table 3. Calculation of Root-Mean-Square errors

x_k	y_k	$y_k = 1.2082e^{0.58818x_k}$	e^k	$\|e_k\|^2$	$y_k = 1.75421x^{1.45343}$	e^k	$\|e_k\|^2$
1	2	2.17562	0.17562	0.03084	1.75421	-0.24579	0.06041
2	4	3.91765	-0.08235	0.00678	4.80405	0.80405	0.6465
3	8	7.05454	-0.94546	0.89389	8.66052	0.66052	0.43629
4	13	12.70317	-0.29683	0.08811	13.1563	0.1563	0.02443
5	21	22.87470	1.8747	3.5145	18.1964	-2.8036	7.86017
				$\Sigma\|e_k\|^2$ = 4.5341			$\Sigma\|e_k\|^2$ = 9.0278

The Root-Mean-Square errors of the exponential fit and power fit are respectively ≈ 0.952273and ≈ 1.3412.

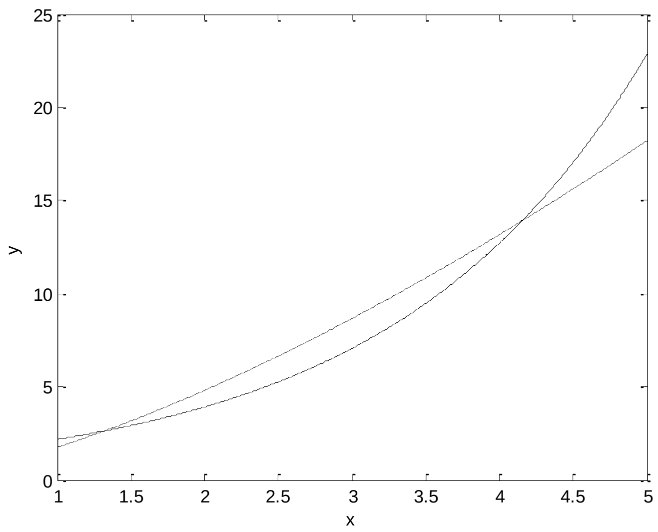

Fig. 3. Graphical Comparison between Exponential and Power fit.

Conclusion

The nonlinear equations can be solved using various software packages such as Newton's method, Nelder-Mead simplex algorithm. These are time-consuming computation and iterations involved require good starting value for the parameters. By data linearization method we can easily find the value of parameters and drafting the curves. In this paper we have analyzed the nonlinear curve fitting by data linearization for exponential function and power function as well as analyzing the error by "Root-Mean-Square" and we have seen that the "Root-Mean-Square" for exponential fit smaller than the power fit. It is found that the exponential fit is the best fitting for the data linearization.

References

Atkinson and E. Kendal, (1994). Elementary Numerical Analysis (2nd ed.). Singapore: John Wiley and Sons.

Beets, J., (1998). Survey of numerical method for trajectory optimization. *Journal of Guidance,Control and Dynamics,* **21**(2): 193-207.

Burdon, R. L., and J. D. Faires, (2004). Numerical Analysis. Boston: Cole.

Gerald, C., and P. Whealthey (1994). *Applied Numerical Analysis*, **5,** Wesley.

Jupp, D., (1978). Approximations to data by splines with free knots. STAM Journal of Numerical Analysis, **15:** 328-343.

Littlefield, B., and D. C. Hanselman, (2005). Bruce Mastering MATLAB 7.0. India: Pearson Education.

Mathews, J. H., (2001). Numerical Methods for Mathematics Science and Engineering (2nd ed.). New Delhi: Prentice Hall.