Curve Fitting

April 13, 2007

Suppose we have a set of data points $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$ that result from an experiment, say, and we want to find a function $f(x)$ that best suits this data. For example, if we run an experiment to verify Ohm's low (current in a resistor is propotional to the voltage across the resistor), then the set of data produced could be current readings x_k versus voltage readings y_k. In this case we are looking for the linear function $f(x) = ax + b$ that best fit this data. Theoretically, all these readings should lie on the same atraight line wilt slope equal to the resistance, but in reality it is never the case due to various sources of approximations and errors. Another example is when we run an experiment to verify Newton law of cooling (if you leave a hot body to cool in room temperature, its temperature decays exponentially with time). In this case the x_ks represent times and the y_ks represent temperatures and we are looking for an exponential function $f(x) = Ke^{\alpha x}$ that best fit this data.

Thus, the ideal situation is to find a function f such that $f(x_k) = y_k$. Since this is not possible in most cases, we look for a function f such that the errors $f(x_k) - y_k$ is as small as possible. One convenient measure of error is the mean square error:

$$E = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(f(x_k) - y_k\right)^2}.$$

Since this error is minimum if and only if NE^2 is minimum, we define the error directly as

$$E = \sum_{k=1}^{N}\left(f(x_k) - y_k\right)^2. \tag{1}$$

Our problem now is to determine the parameters of the function f (e.g., a, b in case f is liear, K, α in case f is exponential) such that E is minimum. This way of determining the parameters, and consequently the function f, is called the method of least squares.

1 The Normal Equations

Suppose now that the function f we want to find is dependent on n parameters $b_1, b_2, \ldots, b_n$. Then the least squares error equation (1) is dependent on $b_1, b_2, \ldots, b_n$ and the problem is to find $b_1, b_2, \ldots, b_n$ such that E is minimum. From the calculus of several variables, the necessary condition for minimization is that all partial derivatives with respect to the parameters $b_1, b_2, \ldots, b_n$ must vanish. That is, we must have

$$\frac{\partial E}{\partial b_j} = 0,\ j = 1, 2, \ldots, n.$$

These conditions yeild the equations

$$2\sum_{k=1}^{N}\left(f(x_k) - y_k\right)\frac{\partial f(x_k)}{\partial b_j} = 0,\ j = 1, 2, \ldots, n$$

which may be rewritten in the form

$$\sum_{k=1}^{N} f(x_k) \frac{\partial f(x_k)}{\partial b_j} = \sum_{k=1}^{N} y_k \frac{\partial f(x_k)}{\partial b_j} \quad j = 1, 2, \ldots, n.$$

These equations (called the normal equations) can be put in a matrix form as follows. Define the following vectors and matrices.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, F = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_N) \end{bmatrix}$$

and

$$F' = \begin{bmatrix} \frac{\partial f(x_1)}{\partial b_1} & \frac{\partial f(x_2)}{\partial b_1} & \cdots & \frac{\partial f(x_N)}{\partial b_1} \\ \frac{\partial f(x_1)}{\partial b_2} & \frac{\partial f(x_2)}{\partial b_2} & \cdots & \frac{\partial f(x_N)}{\partial b_2} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial f(x_1)}{\partial b_n} & \frac{\partial f(x_2)}{\partial b_n} & \cdots & \frac{\partial f(x_N)}{\partial b_n} \end{bmatrix}.$$

Then the normal equations have the matrix representation

$$F'F = F'Y. \tag{2}$$

If f depends linelealy on the paremeters $b_1, b_2, \ldots, b_n$ then (as we shall see shortly) F' is free of the paraeters $b_1, b_2, \ldots, b_n$ and the system (2) is a linear system of equations to be solved for $b_1, b_2, \ldots, b_n$. In this case we call the least squares problem *linear least squares.* Otherwise it is called *nonlinear least squares.* Examples of both types of least squares problems will be discussed in the next sections.

2 Polynomial Curve Fitting

We suppose here that we want to find the polynomial of degree m

$$\begin{aligned} P_m(x) &= a_m x^m + a_{m-1} x^{m-1} + \ldots + a_1 x + a_0 \\ &= \sum_{j=0}^{m} a_j x^j \end{aligned}$$

that best fits the data $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$. Observe that the polynomial $P_m(x)$ depends linearly on the parameters $a_m, a_{m-1}, \ldots, a_0$Then

$$F = \begin{bmatrix} \sum_{j=0}^{m} a_j x_1^j \\ \sum_{j=0}^{m} a_j x_2^j \\ \vdots \\ \sum_{j=0}^{m} a_j x_N^j \end{bmatrix} = \begin{bmatrix} x_1^m & x_1^{m-1} & \ldots & 1 \\ x_2^m & x_2^{m-1} & \ldots & 1 \\ \vdots & \vdots & \ldots & \vdots \\ x_N^m & x_N^{m-1} & \ldots & 1 \end{bmatrix} \begin{bmatrix} a_m \\ a_{m-1} \\ \vdots \\ a_0 \end{bmatrix},$$

and

$$F' = \begin{bmatrix} \frac{\partial P_m(x_1)}{\partial a_m} & \frac{\partial P_m(x_2)}{\partial a_m} & \cdots & \frac{\partial P_m(x_N)}{\partial a_m} \\ \frac{\partial P_m(x_1)}{\partial a_{m-1}} & \frac{\partial P_m(x_2)}{\partial a_{m-1}} & \cdots & \frac{\partial P_m(x_N)}{\partial a_{m-1}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial P_m(x_1)}{\partial a_0} & \frac{\partial P_m(x_2)}{\partial a_0} & \cdots & \frac{\partial P_m(x_N)}{\partial a_0} \end{bmatrix} = \begin{bmatrix} x_1^m & x_2^m & \cdots & x_N^m \\ x_1^{m-1} & x_2^{m-1} & \cdots & x_N^{m-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

Therefore,

$$
\begin{aligned}
F'F &= \begin{bmatrix} x_1^m & x_2^m & \cdots & x_N^m \\ x_1^{m-1} & x_2^{m-1} & \cdots & x_N^{m-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1^m & x_1^{m-1} & \ldots & 1 \\ x_2^m & x_2^{m-1} & \ldots & 1 \\ \vdots & \vdots & \ldots & \vdots \\ x_N^m & x_N^{m-1} & \ldots & 1 \end{bmatrix} \begin{bmatrix} a_m \\ a_{m-1} \\ \vdots \\ a_0 \end{bmatrix} \\
&= \begin{bmatrix} \sum_{k=1}^N x_k^{2m} & \sum_{k=1}^N x_k^{2m-1} & \cdots & \sum_{k=1}^N x_k^m \\ \sum_{k=1}^N x_k^{2m-1} & \sum_{k=1}^N x_k^{2m-2} & \cdots & \sum_{k=1}^N x_k^{m-1} \\ \vdots & \vdots & \cdots & \vdots \\ \sum_{k=1}^N x_k^m & \sum_{k=1}^N x_k^{m-1} & \cdots & \sum_{k=1}^N 1 \end{bmatrix} \begin{bmatrix} a_m \\ a_{m-1} \\ \vdots \\ a_0 \end{bmatrix}.
\end{aligned}
$$

and

$$
F'Y = \begin{bmatrix} x_1^m & x_2^m & \cdots & x_N^m \\ x_1^{m-1} & x_2^{m-1} & \cdots & x_N^{m-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^N x_k^m y_k \\ \sum_{k=1}^N x_k^{m-1} y_k \\ \vdots \\ \sum_{k=1}^N y_k \end{bmatrix}.
$$

Thus, to find the polynomial coefficients of $P_m(x)$ we need to solve the $(m+1) \times$l$(m+1)$ inear system

$$
\begin{bmatrix} \sum_{k=1}^N x_k^{2m} & \sum_{k=1}^N x_k^{2m-1} & \cdots & \sum_{k=1}^N x_k^m \\ \sum_{k=1}^N x_k^{2m-1} & \sum_{k=1}^N x_k^{2m-2} & \cdots & \sum_{k=1}^N x_k^{m-1} \\ \vdots & \vdots & \cdots & \vdots \\ \sum_{k=1}^N x_k^m & \sum_{k=1}^N x_k^{m-1} & \cdots & \sum_{k=1}^N 1 \end{bmatrix} \begin{bmatrix} a_m \\ a_{m-1} \\ \vdots \\ a_0 \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^N x_k^m y_k \\ \sum_{k=1}^N x_k^{m-1} y_k \\ \vdots \\ \sum_{k=1}^N y_k \end{bmatrix}. \tag{3}
$$

The following theorem tells us when this system will have a unique solution.

Theorem 1 *If the data $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$ is such that $x_1, x_2, \ldots, x_N$ are distince and if $N \geq (m+1)$ then the system (3) has a unique solution.*

Proof. We need to show that if $x_1, x_2, \ldots, x_N$ are distince and if $N \geq (m+1)$ then the rwos of the coefficient matrix $F'F$ of the system (3) are linearly independent. If this were not the case then we could find a vector X such that

$$X^t F' F = 0.$$

Multiply on the right by X and observe that F' is exactly F^t. This gives

$$X^t F' F X = 0$$

or

$$\|FX\|^2 = 0.$$

Thus

$$FX = 0. \tag{4}$$

Therefore, the columns of the $N \times (m+1)$ matrix F are linearly dependent. Since $N \geq (m+1)$, the matrix $(m+1) \times (m+1)$

$$
F_1 = \begin{bmatrix} x_1^m & x_1^{m-1} & \ldots & 1 \\ x_2^m & x_2^{m-1} & \ldots & 1 \\ \vdots & \vdots & \ldots & \vdots \\ x_{m+1}^m & x_{m+1}^{m-1} & \ldots & 1 \end{bmatrix}
$$

is a submatrix of F. This matrix is a Holder's matrix whose determinant

$$\prod_{i>j} (x_i - x_j) \neq 0$$

since $x_1, x_2, \ldots, x_N$ are distinct. Writing F in the form $\begin{bmatrix} F_1 \\ F_2 \end{bmatrix}$ and multiplying (4) on the left by $\begin{bmatrix} F_1^{-1} & 0 \end{bmatrix}$ gives $X = 0$, which is a contradiction. ■

In the case $m = 1$, the system (3) takes the form

$$\begin{bmatrix} \sum_{k=1}^N x_k^2 & \sum_{k=1}^N x_k \\ \sum_{k=1}^N x_k & N \end{bmatrix} \begin{bmatrix} a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^N x_k y_k \\ \sum_{k=1}^N y_k \end{bmatrix} \tag{5}$$

and in the case $m = 2$, it takes the form

$$\begin{bmatrix} \sum_{k=1}^N x_k^4 & \sum_{k=1}^N x_k^3 & \sum_{k=1}^N x_k^2 \\ \sum_{k=1}^N x_k^3 & \sum_{k=1}^N x_k^2 & \sum_{k=1}^N x_k \\ \sum_{k=1}^N x_k^2 & \sum_{k=1}^N x_k & N \end{bmatrix} \begin{bmatrix} a_2 \\ a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^N x_k^2 y_k \\ \sum_{k=1}^N x_k y_k \\ \sum_{k=1}^N y_k \end{bmatrix}.$$

Example Find the least squares line that fits the data

x_k	y_k
-2	1
-1	2
0	3
1	3
2	4

For this data we have $N = 5, \sum_{k=1}^5 x_k^2 = 10, \sum_{k=1}^5 x_k = 0, \sum_{k=1}^5 x_k y_k = 7, \sum_{k=1}^5 y_k = 13$. Hence, we have the system

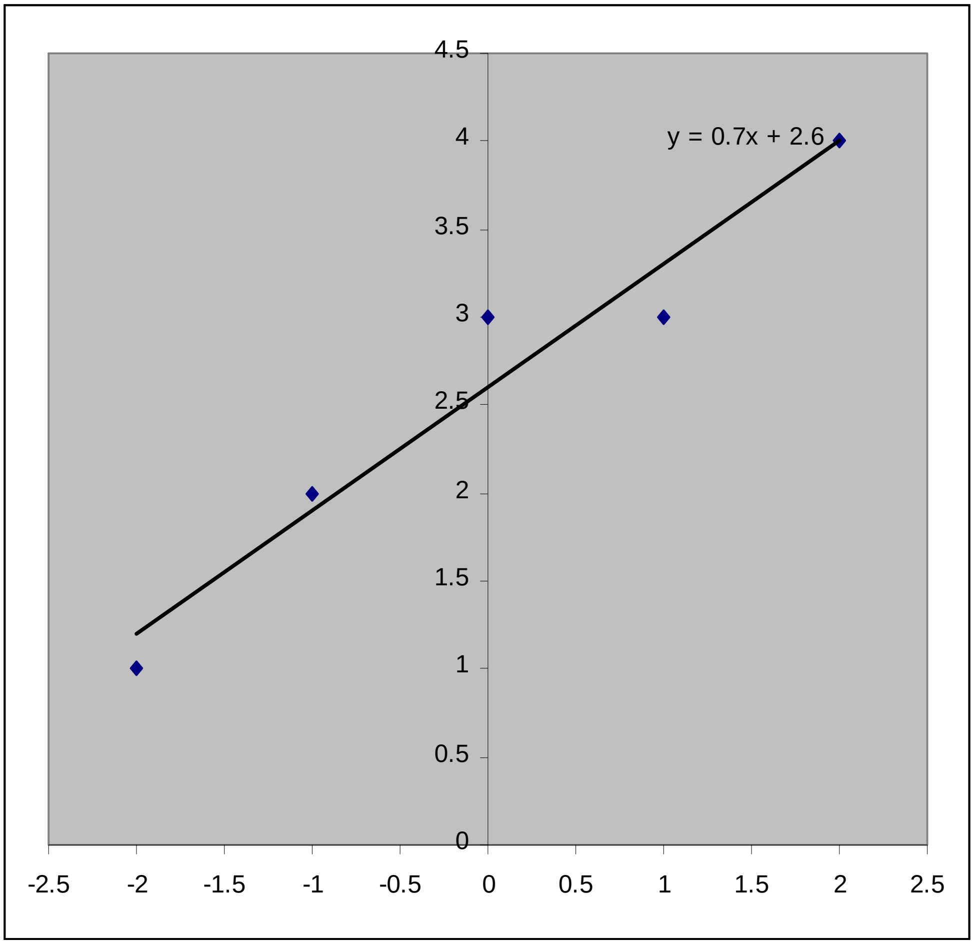
$$\begin{bmatrix} 10 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} 7 \\ 13 \end{bmatrix}$$

which gives

$$\begin{bmatrix} a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} 0.7 \\ 2.6 \end{bmatrix}$$

The linear fit is

$$f(x) = 0.7x + 13$$



Example Find the least squares paraola that fits the data

x_k	-3	-1	1	3
y_k	15	5	1	5

The system for this data is

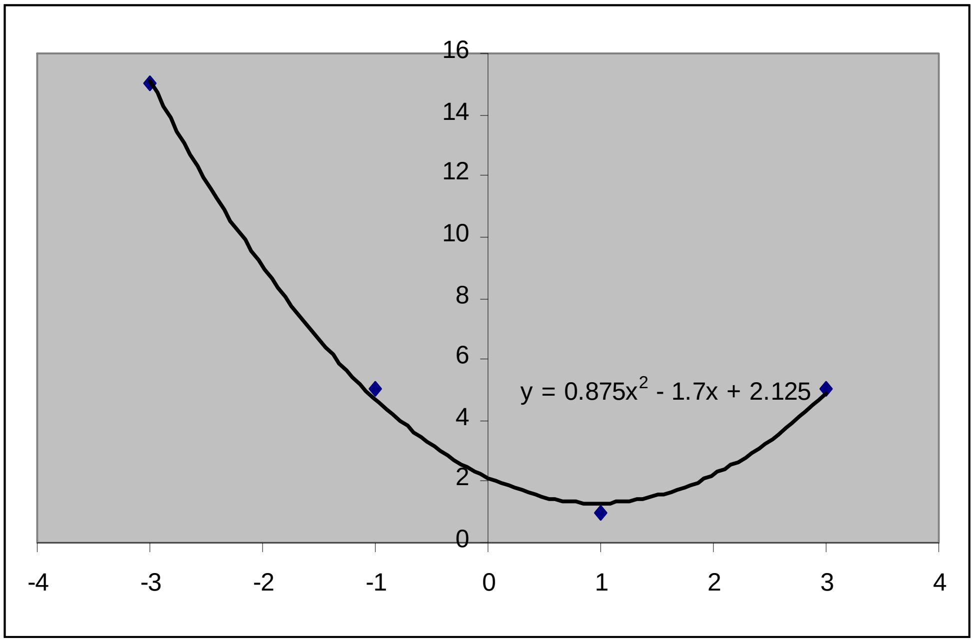
$$\begin{bmatrix} 164 & 0 & 20 \\ 0 & 20 & 0 \\ 20 & 0 & 4 \end{bmatrix} \begin{bmatrix} a_2 \\ a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} 186 \\ -34 \\ 26 \end{bmatrix}$$

and its solution is

$$\begin{bmatrix} a_2 \\ a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} 0.875 \\ -1.7 \\ 2.125 \end{bmatrix}.$$

Thus the fitting parabola has the formula

$$f(x) = 0.875x^2 - 1.7x + 2.125$$



3 Nonlinear Least Squares Fitting

Suppose we want to fit an exponential function of the form $f(x) = Ke^{\alpha x}$ to the data $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$. The normal system in this case has the form

$$F = K \begin{bmatrix} e^{\alpha x_1} \\ e^{\alpha x_2} \\ \vdots \\ e^{\alpha x_N} \end{bmatrix}, F' = \begin{bmatrix} e^{\alpha x_1} & e^{\alpha x_2} & \cdots & e^{\alpha x_N} \\ Kx_1 e^{\alpha x_1} & Kx_2 e^{\alpha x_1} & \cdots & Kx_N e^{\alpha x_N} \end{bmatrix}$$

so that

$$F'F = \begin{bmatrix} K\sum_{k=1}^{N} e^{2\alpha x_k} \\ K^2 \sum_{k=1}^{N} x_k e^{2\alpha x_k} \end{bmatrix}$$

and the right hand side is

$$\begin{bmatrix} \sum_{k=1}^{N} e^{\alpha x_k} y_k \\ K\sum_{k=1}^{N} x_k e^{\alpha x_k} y_k \end{bmatrix}.$$

So, we have to solve the system of nonlinear equations

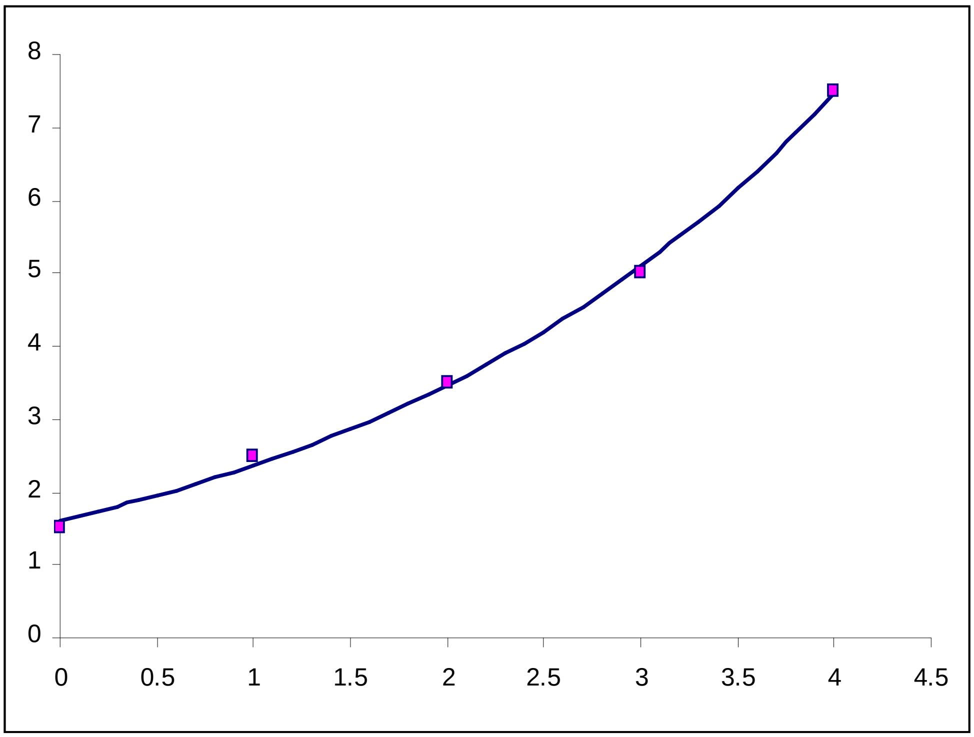
$$\begin{bmatrix} K\sum_{k=1}^{N} e^{2\alpha x_k} - \sum_{k=1}^{N} e^{\alpha x_k} y_k \\ K\sum_{k=1}^{N} x_k e^{2\alpha x_k} - \sum_{k=1}^{N} x_k e^{\alpha x_k} y_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

for K, α. One way of solving this nonlinear system is by Newton method. However, usually the iterations converge slowly and require a good guess for the initial values of K, α. In cases like these, one tries to minimize the least squares error directly without using the normal equations. The methods belong to the topic of optimization which is not part of this course. However, Matlab provides a function that finds the parameters corresponding to the least square error. This function is FMINSEARCH(FUN,X0,OPTIONS,P1,P2,...). For more on this function type "Help FMINSEARCH" in the Matlab command window.

Example Find the least square ecponential function $f(x) = Ke^{\alpha x}$ that best fit the data

x_k	0	1	2	3	4
y_k	1.5	2.5	3.5	5	7.5

The Matlab function requires in initial guess of the parameters (K, α). A good guess of K is 1.5, which would correspond to where the exponential function would meet the y axis and a good guess for α is obtained by approxiamting the slope of the tangent at 0. Using the two points $(0, 1.5)$ and $(1, 2.5)$ the slope of the tangent at zero is approximately 1. This should equal to $f'(0) = \alpha K$. Therefore, a good guess for α is $1/1.5 = 2/3$. Calling the function FMINSEARCH with this initial guess gives the results $K = 1.6108$ and $\alpha = 0.3836$. The following figure shows the data and the fitted function $f(x) = 1.6108e^{0.3836x}$.



4 Linearizing Transformations

Since nonliear least squares problems are numerically hard to solve and require a good initial guess of the fitting function parameters, we should look for alternative ways, if possible to change the dependence on the fitting function parameters into linear dependence. For example, for the exponential fitting function

$$f(x) = Ke^{\alpha x}$$

we could define the new function F by

$$\begin{aligned} F(x) &= \ln f(x) = \ln K + \alpha x \\ &= C + \alpha x. \end{aligned}$$

The function F depends linearly on the "new" parameters C and α. It is actually a polynomial in x. Now since f is required to fit the values y_k, F will by required to fit the values $Y_k = \ln y_k$.

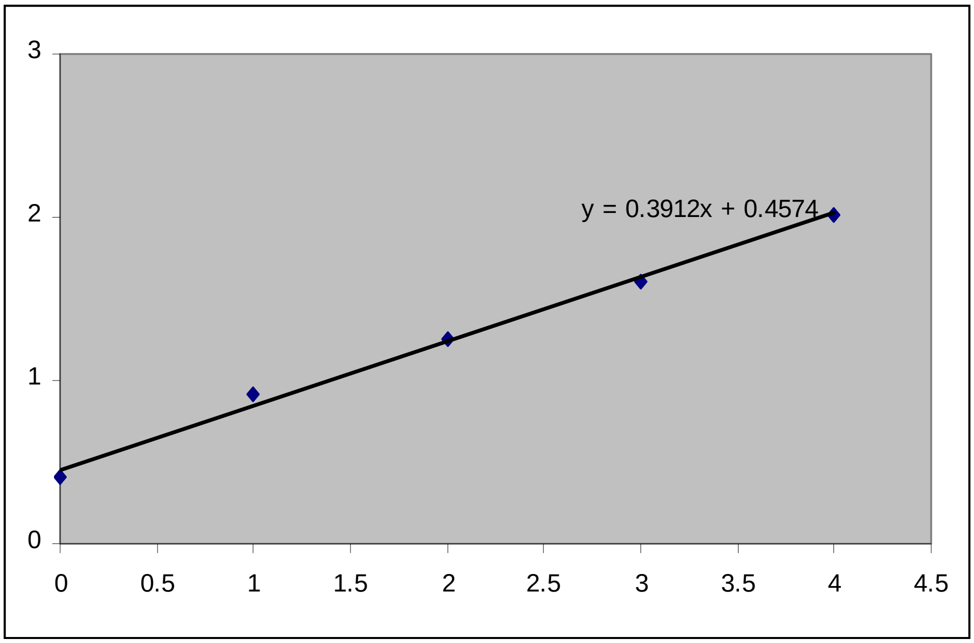
Example For the data of the previous example

x_k	0	1	2	3	4
y_k	1.5	2.5	3.5	5	7.5

we compute

x_k	0	1	2	3	4
$Y_k = \ln y_k$	0.405 47	0.916 29	1. 252 8	1. 609 4	2. 014 9

Applying the system (5) to this problem we get $F(x) = 0.3912x + 0.4574$. Thus $\alpha = 0.3912$ and $C = 0.4574$. From $C = \ln K$, we get $K = e^{0.4574} = 1.580\,0$.



Notice that these values are different from the ones we obtained by the optimization technique (see the previous example). The following figure shows a comparison of both curves and the data they fit.

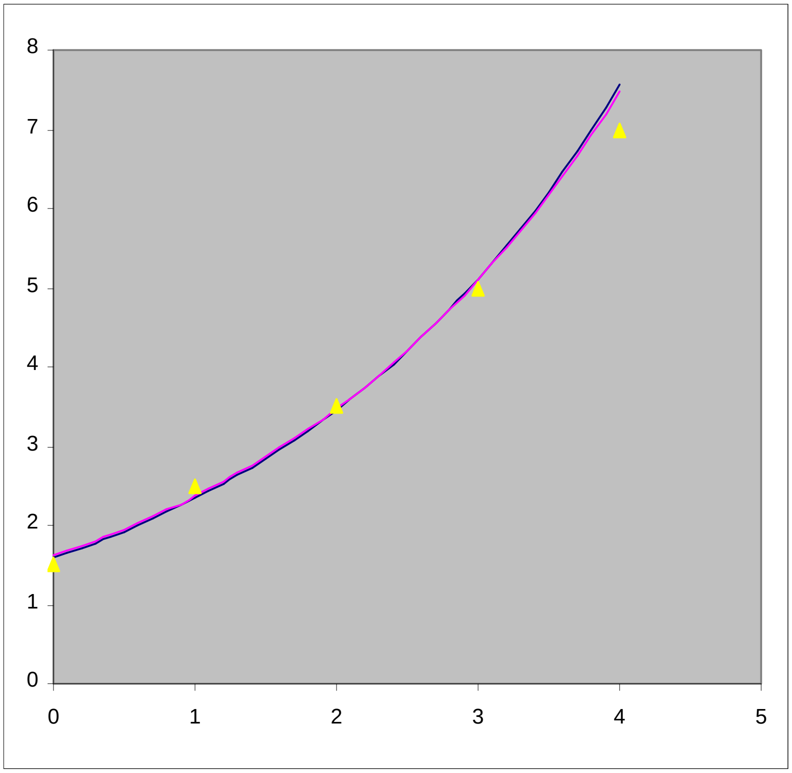
As another example, if we want to fit a function of the form $f(x) = \frac{A}{x+B}$, one can linearize as follows.

$$\begin{aligned} y &= \frac{A}{x+B} \\ y(x+B) &= A \\ xy &= A - By \\ Y &= A - BX, \end{aligned}$$

where the new independent variable $X = y$ and the new dependent variable $Y = xy$.

The following table gives the some linearizing transformations for commonly used fitting functions.

fuction $y = f(x)$	Linearized form $Y = AX + B$	Transformation
$y = \frac{1}{Ax+B}$	$\frac{1}{y} = Ax + B$	$X = x, Y = \frac{1}{y}$
$y = Cx^A$	$\ln y = \ln C + A \ln x$	$X = \ln x, Y = \ln y$
$y = (Ax+B)^{-2}$	$y^{-1/2} = Ax + B$	$X = x,\ Y = y^{-1/2}$
$y = Cxe^{-Dx}$	$\ln\left(\frac{y}{x}\right) = \ln C - Dx$	$X = x,\ Y = \ln\left(\frac{y}{x}\right)$
$y = \frac{L}{1+Ce^{Ax}}$	$\ln\left(\frac{L}{y} - 1\right) = Ax + \ln C$	$X = x,\ Y = \ln\left(\frac{L}{y} - 1\right)$ (L must be given)
$y = \frac{A}{x} + B$	$y = A\frac{1}{x} + B$	$X = \frac{1}{x},\ Y = y$
$y = A \ln x + B$	$y = A \ln x + B$	$X = \ln x,\ Y = y$



Notice that the last two functions are already linear in the parameters A, B, However, the transformation given changes them into polynomial functions which can be hadled by the system (5).