*International Journal of Latest Research in Humanities and Social Science (IJLRHSS)*
*Volume 07 - Issue 02, 2024*
*www.ijlrhss.com || PP. 11-18*

# Curve Fitting and Interpolation Polynomiale

Mohammed Abdul Rasoul Tohma
*Specialization in Biomedical Engineering*
*Graduate of the University of Technology*

**Abstract:** Matching curves is a method in which a mathematical synthesis is made through equations for a number of points through which a curve is created that goes through the specified data, forming the best equation that can go through the points. Curves are of two types: a straight line that passes through the specified points exactly. An approximation that passes through most of the points or close to each other.

The method of least squares is a standard approach to the approximate solution of over determined systems, i.e., sets of equations in which there are more equations than unknowns. "Least squares" means that the overall solution minimizes the sum of the squares of the errors made in the results of every single equation. The most important application is in data fitting. The best fit in the least-squares sense minimizes the sum of squared residuals, a residual being the difference between an observed value and the fitted value provided by a model. Least squares problems fall into two categories: linear or ordinary least squares and non-linear least squares, depending on whether or not the residuals are linear in all unknowns.

## Introduction:

Polynomials are mathematical expressions that are frequently used for problem solving and modeling in science and engineering. In many cases an equation that is written in the process of solving a problem is a polynomial, and the solution of the problem is the zero of the polynomial. Curve fitting is a process of finding a function that can be used to model data. Interpolation is the process of estimating values between data points.

The simplest kind of interpolation is done by drawing a straight line between the points. In a more sophisticated interpolation, data from additional points is used. In this chapter we focused on How MATLAB deal with these concepts.

## Curve fitting

Curve fittings the process of constructing a curve, or mathematical function, that has the best fit to a series of data points, possibly subject to constraints. Curve fitting can involve either interpolation, where an exact fit to the data is required, or smoothing, in which a "smooth" function is constructed that approximately fits the data. A related topic is regression analysis, which focuses more on questions of statistical inference such as how much uncertainty is present in a curve that is fit to data observed with random errors. Fitted curves can be used as an aid for data visualization, to infer values of a function where no data are available, and to summarize the relationships among two or more variables. Extrapolation refers to the use of a fitted curve beyond the range of the observed data, and is subject to a degree of uncertainty since it may reflect the method used to construct the curve as much as it reflects the observed data.

Two types of curve fitting:

**1-Least –Squares Regression by Linear Regression:**

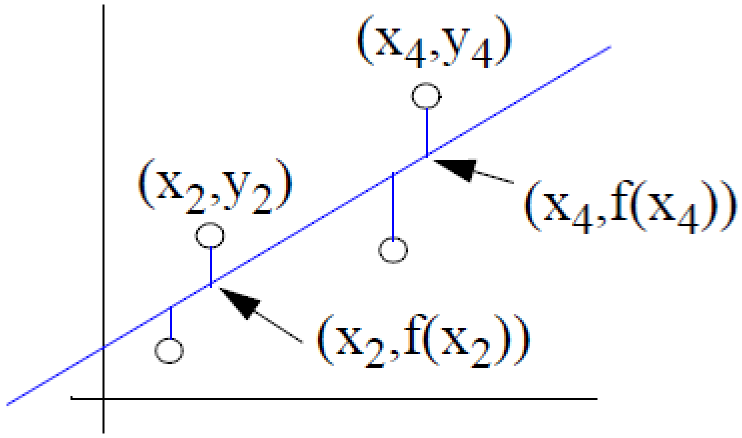
Given data for discrete values, derive a single curve that represents the general trend of the data. When the given data exhibit a significant degree of error or noise.

A straight line is described generically by: $\boldsymbol{y = a_0 + a_1 x}$

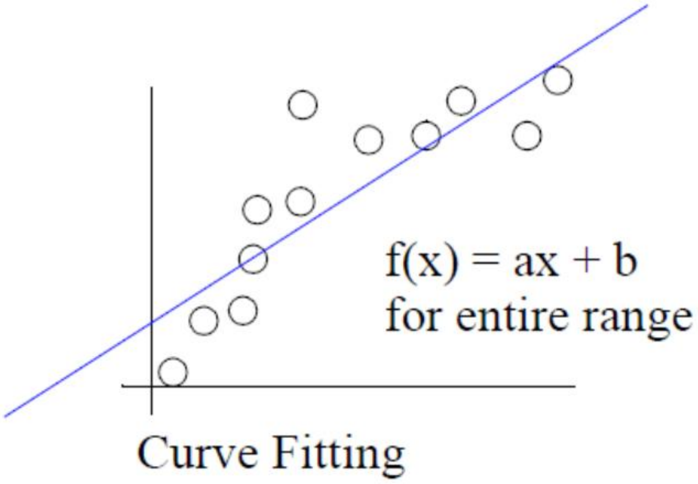
The goal is to identify the coefficients ‘$\boldsymbol{a_1}$’ and ‘$\boldsymbol{a_0}$’ such that $\boldsymbol{y}$ ‘fits’ the data well.

How can we pick the coefficients that best fits the line to the data?

The one line that provides a minimum error is then the ‘best’ straight line : Quantifying error in a curve fit assumptions:




1) positive or negative error have the same value (data point is above or  below the line.
2) Weight greater errors more heavilywe can do both of these things by squaring the distance.




To find '$a_1$' and '$a_0$ We apply the following

By table to find $\bar{y}$  and  $\bar{x}$:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \quad , \quad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \Rightarrow a_0 = \bar{y} - a_1\bar{x} \quad \dots \quad (2)$$

Then we substitute the equation of the line for values$a_1$' and '$a_0$  In order to obtain the required equation:

$$y = a_0 + a_1 x + e$$

where,
$a_0$ is intercept
$a_1$ is a slope
e is error, or residual
Rearranging it you get :

$$e = y - a_0 - a_1 x$$

Notes that:

The distance(error) between the point and the line(whatever it is curve line or it is straight line) happened only in the $\boldsymbol{y}$ axis while the $\boldsymbol{x}$ value remain the same.

The distance(error) must be normal,meaning no three sequential point or more at the end or the begnning have a very big distance(error) compering to other points.

***Exercise (1) ;***

a) Find the least square line for the data

| $x_i$ | -2 | -1 | 0 | 1 | 2 |
|---|---|---|---|---|---|
| $y_i$ | 1 | 2 | 3 | 3 | 4 |

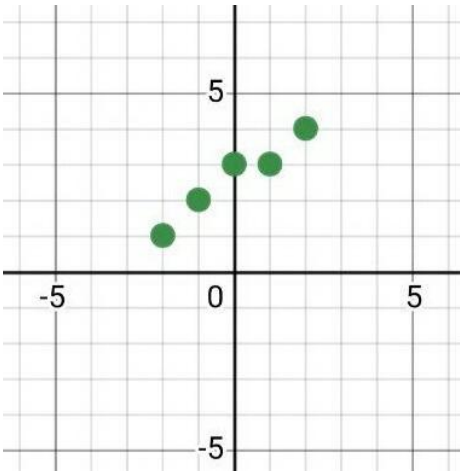
Solution:

$n = 5 \quad \Rightarrow$

| $n$ | $x_i$ | $y_i$ | ${x_i}^2$ | $y_i x_i$ |
|---|---|---|---|---|
| 1 | -2 | 1 | 4 | -2 |
| 2 | -1 | 2 | 1 | -2 |
| 3 | 0 | 3 | 0 | 0 |
| 4 | 1 | 3 | 1 | 3 |
| 5 | 2 | 4 | 4 | 8 |
| $\sum$ | 0 | 13 | 10 | 7 |

Table(1)

let $y = a_0 + a_1 x$




From the table(1) we have :

$n = 5 \;, \sum_{n=1}^{5} x_i = 0 \;, \sum_{n=1}^{5} y_i = 13 \;, \sum_{n=1}^{5} {x_i}^2 = 10 \;, \sum_{n=1}^{5} y_i x_i = 7$

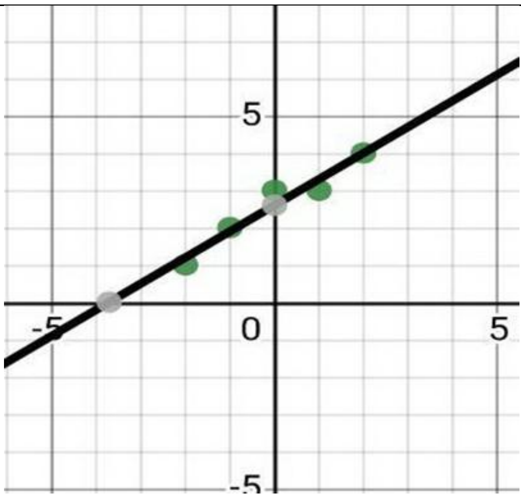
To find '$a_1$' and '$a_0$We apply the following :

$$a_1 = \frac{n\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} {x_i}^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{(5*7)-(0*13)}{(5*10)-0} = \frac{35}{50} = 0.7$$

to find $\bar{y}$ and $\bar{x}$ :$\Rightarrow \bar{y} = \frac{13}{5} = 2.6 \;, \bar{x} = \frac{0}{5} = 0$

$$a_0 = \bar{y} - a_1\bar{x} \Rightarrow a_0 = 2.6 - 0 = 2.6$$

Then : $y = a_0 + a_1 x \Rightarrow y = 2.6 + 0.7x$




b) Find the least square line for the data :

| | | | | | |
|---|---|---|---|---|---|
| $x_i$ | -4 | -2 | 0 | 2 | 4 |
| $y_i$ | 1.2 | 2.8 | 6.2 | 7.8 | 13.2 |

Solution:
$n = 5 \quad \Rightarrow$

| $n$ | $x_i$ | $y_i$ | $x_i^2$ | $y_i x_i$ |
|---|---|---|---|---|
| 1 | -4 | 1.2 | 16 | -4.8 |
| 2 | -2 | 2.8 | 4 | -5.6 |
| 3 | 0 | 6.2 | 0 | 0 |
| 4 | 2 | 7.8 | 4 | 15.6 |
| 5 | 4 | 13.2 | 16 | 52.8 |
| $\sum$ | 0 | 31.2 | 40 | 58 |

Table(2)

let $y = a_0 + a_1 x$
From the table(2) we have :

$$n = 5 \ , \sum_{n=1}^{5} x_i = 0 \ , \sum_{n=1}^{5} y_i = 31.2 \ , \sum_{n=1}^{5} x_i^2 = 40 \ , \sum_{n=1}^{5} y_i x_i = 58$$

To find '$a_1$' and '$a_0$We apply the following :

$$a_1 = \frac{n\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = 1.45$$

to find $\bar{y}$ and $\bar{x}$ :

$$\Rightarrow \bar{y} = \frac{31.2}{5} = 6.24 \ , \qquad \bar{x} = \frac{0}{5} = 0$$

$$a_0 = \bar{y} - a_1\bar{x} \Rightarrow a_0 = 6.24 - 0 = 6.24$$

Then : $y = a_0 + a_1 x \Rightarrow y = 6.24 + 1.45x$

c) Find the least squares regression line for the data on annual incomes and food expenditures of seven households, Use income as an independene variable and food expenditure as a dependent variable, All data is given in thousands of dollars.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Income | 35 | 50 | 22 | 40 | 16 | 30 | 25 |
| $Expenditure$ | 9 | 15 | 6 | 11 | 5 | 8 | 9 |

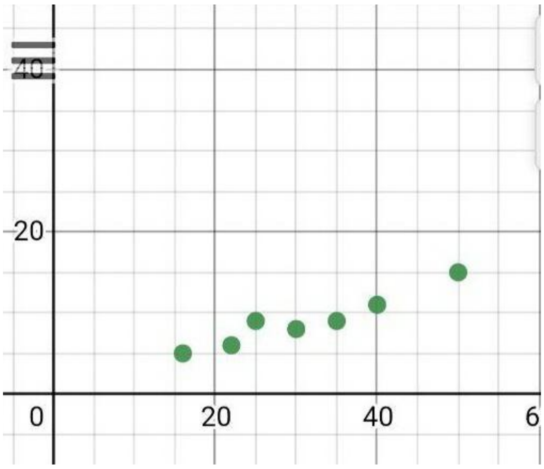
**Solution:**

$n = 7 \quad \Rightarrow$

| $n$ | $x_i$ | $y_i$ | $(x_i, y_i)$ |
|---|---|---|---|
| 1 | 16 | 5 | $(16, 5)$ |
| 2 | 22 | 6 | $(22, 6)$ |
| 3 | 25 | 9 | $(25, 9)$ |
| 4 | 30 | 8 | $(30, 8)$ |
| 5 | 35 | 9 | $(35, 9)$ |
| 6 | 40 | 11 | $(40, 11)$ |
| 7 | 50 | 15 | $(50, 15)$ |

Table(3)

We define the points in the table(3) at the coordinate level until we find an equation close to drawing the points with the least error:




We suggest the function$y = a_0 + a_1 x$ as the nearest function of the points above:

$n = 7 \quad , \sum_{n=1}^{7} x_i = 218 \quad , \sum_{n=1}^{5} y_i = 63$

$, \sum_{n=1}^{5} x_i^2 = 7590 \quad , \sum_{n=1}^{5} y_i x_i = 2182$

To find '$a_1$' and '$a_0$We apply the following:

$$a_1 = \frac{n \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = 0.2747$$

to find $\bar{y}$ and $\bar{x}$ :

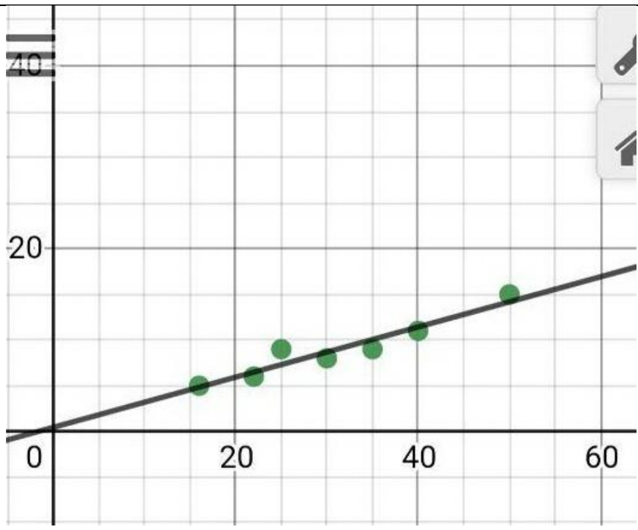
$$\Rightarrow \bar{y} = \frac{31.2}{5} = 8 \quad , \qquad \bar{x} = 31.14$$

$$a_0 = \bar{y} - a_1 \bar{x} \Rightarrow a_0 = 0.4458$$

Then: $y = a_0 + a_1 x \Rightarrow$

$$y = 0.4458 + 0.2747x$$

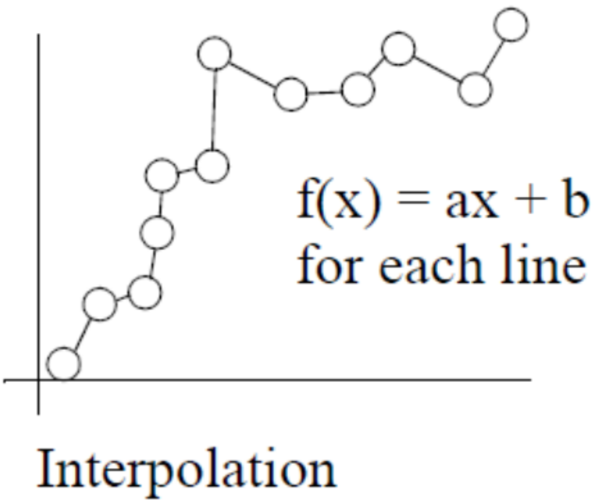
We draw the line on the coordinate plane to see the points that are somewhat close to the line:




**2- Polynomial Interpolation:**

The word interpolation refers to approximating, predicting, or estimating some unknown information from a given set of known information. The technique of interpolation is widely used as a valuable tool in science and engineering. The problem is a classical one and dates back to the time of Newton and Kepler, who needed to solve such a problem in analyzing data on the position of stars and planets.

Mathematical applications of interpolation include derivation of computational techniques for Numerical differentiation, Numerical integration, and Numerical solutions of differential equations.




Is a common method to determine intermediate values between data points where the General equation for $n^{th}$ order polynomial is:

$$f(x) = a_0 + a_1 x + a_2 x^2 \dots + a n x^n$$

So Polynomial interpolation consists of determining the unique nth-order polynomial that fits n+1 data point and For $n+1$ data points, there is only one polynomial of order n that passes through all the points.
There are two ways:

**a-Lagrange Interpolation**

The Lagrange interpolating polynomial is simply a reformulation of the Newton's polynomial that avoids the computation of divided difference.

**b-Newton DD:**

The most popular and useful in polynomial forms. Which have many order but we going to cover the first- and second-order versions:

**1-Linear Interpolation**

This linear interpolation technique can be depicted in which, the similar triangles can berearranged to yield a linear-interpolation formulaIn general, the smaller the interval between the data points, the better the approximation.

**2-Quadratic Interpolation**

In quadratic interpolation of sinusoidal spectrum-analysis peaks, we replace the main lobe of our window transform by a quadratic polynomial, or ``parabola''. This is valid for any practical window transform in a sufficiently small neighborhood about the peak, because the higher order terms in a Taylor series expansion about the peak converge to zero as the peak is approached.

Note that, the Gaussian window transform magnitude is precisely a parabola on a dB scale. As a result, quadratic spectral peak interpolation is exact under the Gaussian window. Of course, we must somehow remove the infinitely long tails of the Gaussian window in practice, but this does not cause much deviation from a parabola, as shown:

***<u>Exercise (2 ;</u>***

a) Estimate the common logarithm of 10 using linear interpolation.
1. Interpolate between $log_8 = 0.9030900$ and $log_{12} = 1.0791812$
2. Interpolate between $log_9 = 0.9542425$ and $log_{11} = 1.0413927$. For each of the interpolations, compute the percent error.

By using equation a linear interpolation:

| $x$ | $f(x)$ |
|---|---|
| $x_0 = 8$ | $f(x_0)$=0.9030900 |
| $x = 10$ | $f_1(x)$=? |
| $x_1 = 12$ | $f(x_1)$=1.0791812 |

$$f_1(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0)$$

$$f_1(x) = 0.9030900 + \frac{1.0791812 - 0.9030900}{12 - 4}(10 - 1)$$

$$f_1(x) = 1.1011926$$

$$\log_{10} 10 = 1 \Rightarrow erorr = \left|\frac{1 - 1.1011926}{1}\right| * 100 = 10.11 \%$$

***<u>Exercise (3) ;</u>***

Fit a second order Newtons interpolating polynomial to estimate $log_{10}$ using the data from exercise (2) above at x=8 , 9 , and 11.

Use the polynomial to evaluate log 10:

$x_0 = 8$ , $f(x_0)$=0.9030900 ,

$x_1 = 9$ , $f(x_1) = 0.9542425$

$x_2 = 11$ , $f(x_2) = 1.0413927$

Abblying equation yielde $f(x_0) = b_0 = 0.9030900$

$$b_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{0.9542425 - 0.9030900}{9 - 8} = 0.0152425$$

$$b_2 = \frac{\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0}$$

$$b_2 = \frac{\frac{1.0413927 - 0.9542425}{11 - 9} - 0.0152425}{11 - 8} = 0.0094442$$

Substituting these values into equation :

$$f_2(x) = b_0 + b_1(x - x_0) + b_2(x - x_1)$$

$$f_2(x) = 0.9030900 + 0.0152425 (x - 8) + 0.0094442 (x - 12)$$

which can be evaluated at $x = 10$ for :

$$f_2(x) = 0.9146866$$

$$\log_{10} 10 = 1 \Rightarrow$$

which represents a relative $error = \left| \frac{1 - 0.9146866}{1} \right| * 100 = 8.53 \%$

Thus, the curvature introduced by the quadratic Interpolation is more accurate (have less error) than the Linear Interpolation

## Reference

[1]. Lancaster P, Salkauskas K. Curve and Surface Fitting, an Introduction. San Diego (CA), Academic Press Ltd, 1987.
[2]. F. Gerald & P. O. Wheatley, Applied numerical analysis, Pearson, 2004.
[3]. Roberto Cavoretto, A numerical algorithm for multidimensional modeling of scattering data points, Computational and Applied Mathematics. 2013
[4]. Salomon, David 2005, Curves and Surfaces for Computer Graphics. Springer Verlag, August 2005.