

Curve fitting project

OVERVIEW

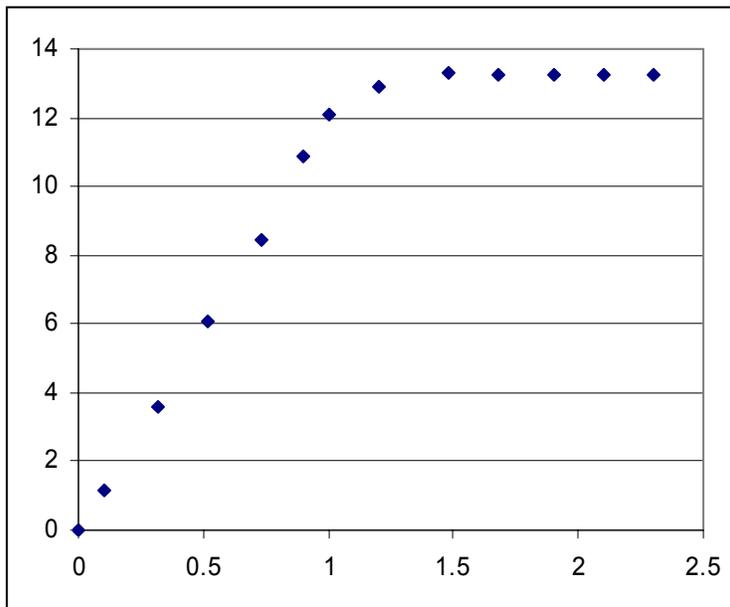
Least squares best fit of data, also called regression analysis or curve fitting, is commonly performed on all kinds of measured data. Sometimes the data is linear, but often higher-order polynomial approximations are necessary to adequately describe the trend in the data. In this project, two data sets will be analyzed using various techniques in both MATLAB and Excel. Consideration will be given to selecting which data points should be included in the regression, and what order of regression should be performed.

TOOLS NEEDED

MATLAB and Excel can both be used to perform regression analyses. For procedural details on how to do this, see Appendix A.

PART A

Several curve fits are to be performed for the following data points:



x	y
0.00	0.000
0.10	1.184
0.32	3.600
0.52	6.052
0.73	8.459
0.90	10.893
1.00	12.116
1.20	12.900
1.48	13.330
1.68	13.243
1.90	13.244
2.10	13.250
2.30	13.243

1. Using MATLAB, fit a single line through *all* of the points. Plot the result, noting the equation of the line and the R^2 value. Does this line seem to be a sensible way to describe this data?
2. Using Microsoft Excel, again fit a single line through all of the points.
3. Using hand calculations, fit a line through a subset of points (3 or 4) to confirm that the process is understood.

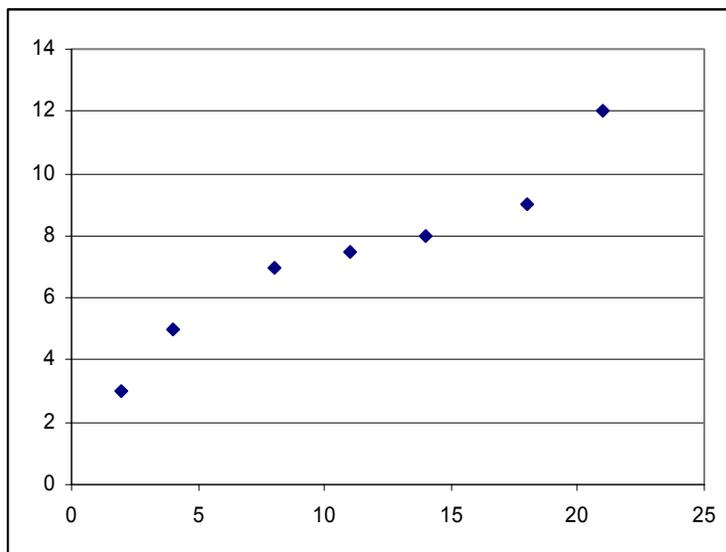
4. One should consider the physics of the problem, and how the result will be used, when deciding how to fit a line to data.

The data shown above is from the calibration of an LVDT (a transducer which produces a voltage in proportion to a displacement). The x-values of the above plot are distances in inches, and the y-values are volts of output. The slope of the line is the *sensitivity* of the LVDT—the relationship between input displacement and output voltage, in volts per inch.

An LVDT responds linearly over a certain range, and outside of that range it is non-linear. Examine the plot to see what the linear range of this particular LVDT is. Assuming that only the linear range of the LVDT is of interest, over what range of points should a line be fit? Fit a single line, using different subsets of points, until you obtain a line that makes sense and has a satisfactory R^2 value. For this line fit, what is the sensitivity of the LVDT?

PART B

Several curve fits will be performed on the following data points:



x	y
2	3
4	5
8	7
11	7.5
14	8
18	9
21	12

Either MATLAB or Excel may be used for the following exercises.

1. Perform a first-order fit on a *portion* of the data where a line seems to best describe the trend. Fit a single line through a subset of points.
2. Perform a second-order fit on a portion of the data where this seems to best describe the trend. Again, fit only a single line through a subset of points.
3. Define a single function that will best describe the measured data, over *all* data points.
4. Discuss some of the factors which should be considered when selecting the order of curve fit to be applied to a data set (in terms of simplicity, physics, ease of use in further calculations, accuracy, etc.).

APPENDIX A

Curve fitting in MATLAB and Excel

MATLAB

The command used to perform curve fitting in MATLAB is:

$$p = \text{polyfit}(x, y, n)$$

This function finds the coefficients of a polynomial, $p(x)$, which provides a least-squares best fit to the data provided. The inputs to the function are x , y and n , where x is a vector containing the x -values of the data points, y is a vector containing the y -values of the data points, and n is the order of the polynomial.

The output of `polyfit`, p in this case, is a vector containing the coefficients of the polynomial, starting with the highest order term. For example, the vector

$$p = [5 \ 12 \ 3 \ 1]$$

represents the polynomial

$$5x^3 + 12x^2 + 3x + 1$$

The command `polyval` can be used to plot the resulting polynomial. The syntax for `polyval` is

$$f = \text{polyval}(p, x)$$

where p is the array containing the polynomial's coefficients, and x is the original vector of x -values. The polynomial is therefore being evaluated at these x -values, and the result (f) is a vector of the y -values. To plot the original data points along with the regression line, simply enter

$$\text{plot}(x, y, 'o', x, f, '-')$$

This will plot the original data points with small circles, and the polynomial curve fit as a line.

To calculate the R^2 value, the mean, J value and S value must first be found. The mean is simply found using

$$\mu = \text{mean}(y)$$

The J value is

$$J = \text{sum}((f-y).^2)$$

Recall that f is the polynomial evaluated at the x -values, and y contains the original y -values of the data points. The S value is

$$S = \text{sum}((y-\mu).^2)$$

Using these, the R^2 value may then be calculated using

$$r2 = 1 - J/S$$

EXCEL

To perform curve fitting in Excel:

1. Plot the data using unconnected points.
2. Click on the chart to select it.
3. Under the **Chart** menu, select **Add Trendline**.
4. Under the **Type** tab, select the type of regression to perform, e.g. linear or polynomial.
For a polynomial fit, select the desired order.
5. Under the **Options** tab, check “Display equation on chart” and “Display R-squared value on chart.”
6. Click **Okay**.

A PDF which summarizes these menus and windows can be found here:

http://faculty.uml.edu/pavitabile/22.302/web_download/EXCEL_Plot_Trendline.PDF