REGRESSION
(Physics 1210 Notes, Partial Modified Appendix A)

HOW TO PERFORM A LINEAR REGRESSION

Consider the following data points and their graph (Table I and Figure 1):

X	Y
0	1
1	3
2	5
3	7
4	9
5	11

Table 1: Example Data

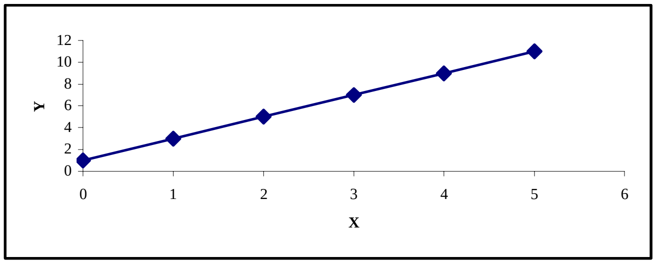

Figure 1: Graph of Example Data

In this example, the points are on a perfect straight line. The formula of a general straight line is $Y=a*X+b$ where a is the slope of the line and b is the intercept of that line with the y-axis. In this example, it is easy to verify that $a=2$ and $b=$ I. In general, with the data points you obtain in your experiments, finding a and b is not so easy. We want to use a computer to calculate a and b for us. For this, we use the regression function of Excel.

When you are in Excel, type in your data points as shown in Table I. Now, we want to do a linear regression on these data points. That will hopefully give us the value for a and b. To do this, look in the menu for [Tools], then select [Data Analysis] and finally select [Regression]. You are then faced with a dialogue-box "Regression". *For "input Y range", select the Y column of your data*
For "input X range", select the X column of your data, then click on "OK".
After a few seconds, you will see a new Excel sheet with an overkill of numbers called the "Summary Output". Of this Summary Output, the only part you need is:

	Coefficients	*Standard Error*	*t Stat*
Intercept	1	0	65535
X Variable	2	0	65535

From this, you can read the coefficient values for a and b as follows:

$$b=Intercept= 1$$

$$a=X\text{-}variable=2$$

which is what we expected. The equation for the line in this case would be:

$$Y=aX+b=2X+1$$

The standard errors in a and b are zero here because the points are on a perfect straight line. In general, this will not be the case, because experiments are not perfect, unfortunately. For example, if you were to use the following data points (they are the same points as before, except for the last one) and do a linear regression on them, you will get:

X	Y
0	1
1	3
2	5
3	7
4	9
5	15

Table 2: Not-so-perfect Example Data

	Coefficients	*Standard Error*	*t Stat*
Intercept	0.238095238	0.99886557	0.238366
X Variable	2.571428571	0.32991444	7.794229

Table 3: Regression Results of Not-so-perfect Data

And you now have standard errors, which are not zero. You would quote your results at the 95% confidence level as:

$$b = Intercept = 0.238 \pm 0.999$$

$$a = x\text{-}variable = 2.5714 \pm 0.3299$$

Of course, you must decide for yourself each time how many decimals are realistic and what the unit is. Linear regression is a very useful tool, and you will need it frequently during this course. In your report, **DO NOT** include the "regression-summary" Excel produces. Instead, when you do a linear regression on your data, all you have to give is the equation of the line (including errors) Excel calculated, and state that the calculation was a linear regression.

Velocity (m/s)	Measured Force (N)
0	4.9
1.38	4.8
1.99	4.6
2.2	4.6
2.51	4.5
3.06	4.4
3.77	4.2
4.09	4.1
4.65	3.8
5.51	3.6
6.21	3.0
7.22	2.6
7.88	2.3
8.53	2.0
9.79	1.1
10.31	0.8
10.93	0.5
11.21	0.2
11.37	0.0

Table 4. Free Fall Data

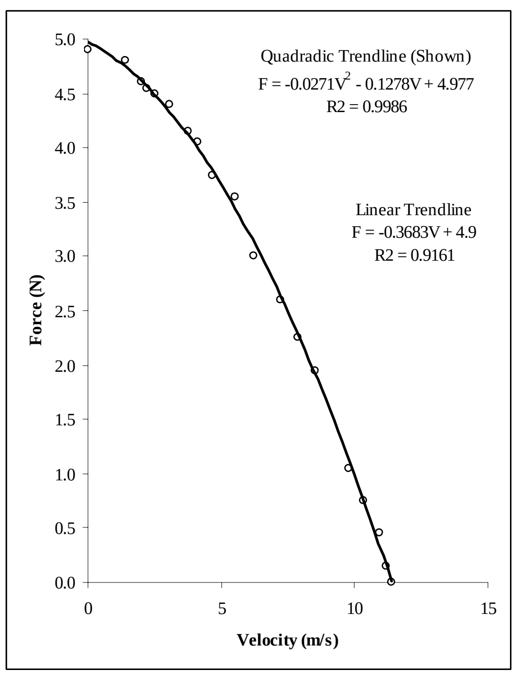

Figure 2. Velocity vs. Force

HOW TO PERFORM NONLINEAR TRENDLINE

Besides linear trendline, Excel has the capability of fitting logarithmic, polynomial of arbitrary order, power or exponential functions to data. For the data presented in Table 4, it appears that a quadratic relationship should produce an excellent fit. Figure 2 substantiates this in that this quadratic trendline has a r^2 of 0.9986 as compared to a value of 0.9161 for the linear fit when the intercept value is set to 4.90 (See Curve Fitting.xls example). Higher order polynomials may be used but any increase in r^2 that is obtained by this increased complexity is rather superficial.

HOW TO PERFORM NONLINEAR OPTIMIZING SOLVER

If we start over on this problem and apply some basic dynamics to the free fall problem, the summation of forces in this case must be equal to the gravitational body force *(m-g)* in the downward direction plus a drag force in the upward direction that is some unknown function of velocity. Therefore theory implies that the force versus velocity relationship must have the following general form:

$$F = mg - Drag(V) \tag{1}$$

but it does not supply any information about how the drag varies with velocity. Our own personal experience indicates that the drag force increases with velocity and extensive experimental testing over the years has shown that power laws can be used frequently to correlate velocity-drag data over limited

$$F = mg - aV^b \tag{2}$$

velocity ranges. If this is assumed to be the case here, then

Theory and some empirical insight has therefore been combined to obtain a possible function form between velocity and force in terms of two arbitrary constants *(a, b)* that is based upon the physics of phenomena and not just blind curve fitting as was done in the linear and quadratic (Figure 2) curve fit examples.

The values of *a* and *b* that give the best fit with the experimental data can be determine through the use of the Excel nonlinear optimizing solver. The fust requirement of using the nonlinear optimizing solver is the development of a regression function that you what to optimize in terms of minimizing or maximizing its value or obtaining a specified value. The trendlines that are presented in the previous two curve fits are based upon least square regression in which the following regression function is minimized

$$\sum_{i=1}^{n} \left(\underline{F}_i - F_i\right)^2 \tag{3}$$

where $\underline{F}_i$ is the measured force and F_i is the corresponding predicted value in the data set that contains *n* values. In this case Equation 2 would be subsituted for F_i ($F_i = mg - aV^b$). Instead of doing this, lets minimize

$$r^2 = 1 - \frac{\sum_{i=1}^{n} \left(\underline{F}_i - F_i\right)^2}{\sum_{i=1}^{n} \left(\overline{F}_i - F_i\right)^2} \tag{4}$$

r. That is

where $\overline{F}$ is the mean force of the experimental data set. Excel provides a nonlinear optimizing solver for minimizing functions such as Equation 4. However, the problem must be prepared properly to obtain an appropriate solution. Table 5 presents a copy of the spreadsheet (see file *Curve Fitting.xls* for the actual spreadsheet) that was used to determine *a* & *b*. This table contains six columns: ***column*** 1 is the independent variable (velocity); ***column*** 2 is the measured variable (acceleration a_i); ***column*** 3 is the dependent variable (force F_i) calculated from the measured variable, a_i; ***column*** 4 is the predicted dependent variable (the force calculated from Equation 2,$\underline{F}_i$); ***column*** 5 is the square of the difference between columns 3 and 4; and ***column*** 6 is the square of the difference between ***column*** 3 and the average force which is calculated at the end of ***column*** 3. The ***columns*** 5 & 6 are then summed and these values are used to calculate the r^2 value for a guess set of coefficients *(a, b)*. For instance, the guess of (1,1) produces a very poor r^2 value of -5.88.

a =	0.0852315	N/(m/s)^b	g = 9.8	(m/s^2)	
b =	1.6632512		m = 0.5	(kg)	

		Force (N)			
Velocity (m/s)	**Accel. (m/s^2)**	**Measured Fi**	**Predicted* <u>F</u>i**	**(<u>F</u>i - Fi)^2 N^2**	**(Fav - Fi)^2 N^2**
0	9.8	4.9	4.9	0.00E+00	3.93
1.38	9.6	4.8	4.8	2.08E-03	3.54
1.99	9.2	4.6	4.6	1.04E-03	2.83
2.2	9.1	4.6	4.6	1.13E-03	2.66
2.51	9.0	4.5	4.5	3.75E-05	2.50
3.06	8.8	4.4	4.4	2.27E-03	2.20
3.77	8.3	4.2	4.1	6.16E-04	1.52
4.09	8.1	4.1	4.0	1.39E-03	1.28
4.65	7.5	3.8	3.8	2.67E-03	0.69
5.51	7.1	3.6	3.4	1.14E-02	0.40
6.21	6.0	3.0	3.1	1.51E-02	0.01
7.22	5.2	2.6	2.6	2.79E-04	0.10
7.88	4.5	2.3	2.3	8.32E-05	0.45
8.53	3.9	2.0	1.9	3.97E-03	0.94
9.79	2.1	1.1	1.1	3.72E-03	3.49
10.31	1.5	0.8	0.8	4.18E-04	4.70
10.93	0.9	0.5	0.3	1.02E-02	6.09
11.21	0.3	0.2	0.2	1.34E-05	7.66
11.37	0.0	0.0	0.0	1.64E-03	8.52
	Fav =	2.9	Sum =	5.80E-02	53.50

R^2 = **0.9989164** = 1 - SUM(<u>F</u>i - Fi)^2/SUM(Fav - Fi)^2

* <u>**F**</u>i = **Force(m,g,V,a,b)** see Module Force(m,g,V,a,b)

Table 5. Excel Table Used to Perform Nonlinear Regression

Excel uses an iterative approach to solve the nonlinear regression problem once it has an initial guess set to start this iterative process. In this case, the program will systematically vary *a*and *b* to determine the local gradient of' r^2 and thereby determine how the *(a, b)* set should be varied to maximize r^2. In order to use the solver tool, the tool must be loaded into Excel. The solver can be loaded by:

(1) Click on Tools in the main menu bar
(2) Click on Solver in the pull down menu
    If Solver is not an option, then
    (a) Click on Add-Ins in the pull down menu
    (b) click on Solver Add-In in the Add-Ins dialog box (the check box must be checked)
    (c) Click OK
    (d) Click Solver

The Solver dialog box is now visible. The first menu item is the target cell which is r^2 in this case. The second item delineates what action is to be perform on the target cell. In this example we wish to minimize the target cell. The third item specifies which cells may have their values varied to accomplish the objective which in this case are cells containing the guess values of the regression parameters *a* and *b*. Note that named cells can be utilized in specifying the cell locations of the target cell and the adjustable cells. As an option, you can set numerical constraints on the adjustable cells. A little thought about the physics of this problem indicates that both *a* and *b* are positive and these constraints may be added. In some problems you may wish to change the default Precision and Tolerance values by first clicking the Options button. Now click OK, and Excel will attempt to find the optimum solution and replace the guess values of the regression parameters with the optimum values. Table 5 indicates that the combined theoretical/empirical correlation

$$F = 4.9 - 0.0852V^{1.663} \qquad (5)$$

produces a r^2 of 0.9989 which is slightly better than the quadratic. This correlation is also simpler than the quadratic fit and it is more physically significant. Instead of basing the curve fit on r^2, try using the least squares regression method to compute the coefficients and compare your results. This example also illustrates the use of a *function module*. To see it, click *Tools, Macro* and *Visual Basic Editor*.

One word of caution: nonlinear functions often contain more than one solution and that a given guess set may produce a local solution (in this case, a local minimum) instead of a global solution. Highly nonlinear problems may also require a fairly accurate initial guess to obtain a global solution or any solution. You may have to resort to plots to produce an accurate initial guess.

See *Nonlinear Regression.xls* **for another example.**

Reference

Physics 1210: Engineering Physics. Lab Manual, Appendix A, University of Wyoming Physics and Astronomy, Spring, 1999.