

Spreadsheet Techniques in Physics Calculations

by Carlo Segre

1 Introduction

The spreadsheet is a general purpose calculation tool and has many uses in physics. In particular, it is very good for one-time data manipulation and for developing calculations which then can be made into programs if you need to do the calculations repetitively. Modern spreadsheets also will permit you to perform data fitting and plot your results. While there are better tools for each of these specific functions than a spreadsheet, the all-in-one nature of the spreadsheet makes it a good place to start. In the following exercises, you will be taking equations and experimental data and transferring them to a spreadsheet where they can be manipulated. There are four separate applications, each of which demonstrates a different aspect of spreadsheet use.

In each case, there will be a brief discussion of the problem in theoretical terms followed by the problem statement. You are expected to lay out a spreadsheet on your own and write a short report using the word processor of your choice. The report should consist of a brief description of how and why you set up the spreadsheet the way you did, followed by a discussion of the results you obtained from the calculation. The text contains questions you should answer in the report and which hopefully will help you think of other conclusions to be drawn from the exercises.

You may use any spreadsheet for this exercise, Openoffice works well on both Linux and Windows, Gnumeric also will work on the Linux cluster. The syntax for most spreadsheets is identical. Turn in the assignment as a single spreadsheet with 4 tabs, one for each exercise. Any plots should appear on the sheet where the data is located. Please turn in the spreadsheet in either .ods (Openoffice) or .xls (Excel) format.

2 Projectile Motion: a “what if” Spreadsheet

One of the most important uses of a spreadsheet is the “what if” calculation. In this application, you will be performing a simple “what if” calculation to examine the parameters in projectile motion. The problem we will be studying is that of a projectile fired from a cannon with initial velocity v_o at an angle θ above the horizontal. The question we want to answer is what initial angle θ_{max} will cause the projectile to land the furthest distance from the cannon. Some of you may already know the answer to this question, and in any case this problem can easily be solved analytically by use of Calculus, nevertheless, it is a nice straightforward example of the “what if” spreadsheet technique as applied to science. We begin by solving the analytic equations which tell how far a projectile will travel for a given v_o and θ . The x -axis is defined as horizontal and positive in the direction of projectile travel. The y -axis is positive upward. There is no component of acceleration in the x -direction, however there is an acceleration in the y -direction due to gravity:

$$a_y = -g = -9.8 \frac{\text{m}}{\text{s}}. \quad (1)$$

Remember that the acceleration is negative because gravity acts downward, in the negative y -direction. The initial velocity of the projectile is v_o , its components in the x - and y -directions are:

$$v_{ox} = v_o \cos(\theta) \quad (2)$$

$$v_{oy} = v_o \sin(\theta) \quad (3)$$

For our situation, the basic kinematic equations of motion are:

$$x(t) = v_{ox}t \quad (4)$$

$$y(t) = v_{oy}t + \frac{1}{2}a_y t^2 \quad (5)$$

The solution to the problem of how far away in the x -direction the projectile lands is as follows. Realizing that when the projectile strikes the ground, its y -coordinate must be zero, we set the equation for $y(t) = 0$ and solve for the time, t . Notice that there are two solutions possible since this is a quadratic equation in t . The first is for the time at which the projectile is fired, and the second is for the time the projectile strikes the ground.

$$t = 0 \quad (6)$$

$$t = \frac{-2v_{oy}}{a_y}. \quad (7)$$

We are, of course, interested in the second solution. The application of this to a spreadsheet is your job, let me give you a few hints now. With a spreadsheet, it is often useful to assign each variable to its own cell. Do not carry out the algebra any further than is given in the equations above. You will then have the following cells:

$$v_o, \theta, a_y, v_{ox}, v_{oy}, t, x(t).$$

Use $v_o = 100\text{m/s}$ as a good value for the initial velocity. By changing the values in the v_o cell you will see the travel distance $x(t)$ cell change its value. You can now experimentally deduce the optimal firing angle to get the largest travel distance. Does this match your previous knowledge or intuition? If you are ambitious, try adding a cell which contains the maximum height y_{max} reached by the projectile during its flight. You will be handing in the spreadsheet printout, so be sure all the cells are clearly labeled.

In your report, discuss the results of your experimentation with the firing angle. Can you propose a way of graphing the trajectory? Most spreadsheets have a predefined order of calculation, that is, they either calculate row-by-row or column-by-column. Try to find out what your program does, can it be changed, in your report discuss the effect this has on how the cells of the spreadsheet must be set up.

3 Numerical Integration on a Spreadsheet

The Physics problem of motion under the influence of uniform acceleration has a straightforward solution. Suppose you were on a train which begins with a constant velocity v_o and is uniformly accelerated with an acceleration a . Since the acceleration is simply the second derivative of the position, x , the equation of motion for this one-dimensional system can be obtained simply:

$$v(t) = v_o + at \tag{8}$$

$$x(t) = x_o + v_o t + \frac{1}{2}at^2. \tag{9}$$

By knowing the initial conditions and the acceleration, we can obtain the position as a function of time, or the distance traveled in a particular amount of time. Real life cases are never this simple. Consider the case of an Elevated Train which accelerates out of the station, reaches cruising speed and then brakes to stop at the next station. The train is never under constant acceleration. If we are faced with the task of calculating how far the El Train has traveled during the trip, we certainly cannot apply the analytic equations shown above. The only way to solve this problem is numerically, using pencil and paper, a calculator or a computer. The spreadsheet is ideally suited as an electronic pencil and paper for such an application. In this exercise, you will learn how to set up a spreadsheet to perform the two numerical integrations required to solve this problem.

Suppose you have a list of ordered pairs of numbers, the first representing the time at which the measurement is taken, the second the value of the acceleration measured at that time. Each pair of numbers is identified with a subscript, the n^{th} pair would be:

$$t_n, a_n.$$

If we make the approximation that the acceleration a_n is constant over the time interval from t_n to t_{n+1} , we can now apply the simple kinematic equations of motion described above to each discrete time interval to obtain the velocity at the end of that time interval:

$$v_{n+1} = a_n(t_{n+1} - t_n) + v_n. \tag{10}$$

That is, the velocity at time t_{n+1} depends on the acceleration and velocity at time t_n and the time interval between data points. Similarly with position:

$$x_{n+1} = v_n(t_{n+1} - t_n) + x_n. \tag{11}$$

Thus, by summing the contribution from each interval, we can obtain a numerical solution for $x(t)$. In your spreadsheet, set up 4 columns for t , a , v and x . Apply the equations above to the v and x columns, noting that we are using strange units in this calculation:

acceleration	is in	$\left[\frac{\text{miles}}{\text{hr} \cdot \text{s}}\right]$
velocity	is in	$\left[\frac{\text{miles}}{\text{hr}}\right]$
position	is in	$[\text{miles}]$

Table 1: Acceleration vs. Time Data for an Elevated Train

t(s)	a($\frac{\text{mi}}{\text{hr s}}$)		t(s)	a($\frac{\text{mi}}{\text{hr s}}$)		t(s)	a($\frac{\text{mi}}{\text{hr s}}$)	
0	0.00	Station 1	33	-2.00	Station 2	66	0.00	Station 3
3	2.75		36	0.00		69	0.00	
6	2.75		39	0.00		72	-2.00	
9	1.75		42	0.00		75	-2.25	
12	1.75		45	0.00		78	-1.00	
15	1.50		48	1.75		81	-1.25	
18	0.75	51	2.75	84	-2.25			
21	-2.25	54	2.25	87	-2.25			
24	-2.50	57	1.75	90	0.00			
27	-2.25	60	1.75					
30	-2.25	63	0.75					

Check your formula entry with a simple example that you can do in your head. Remember, a computer or calculator is no good unless the number you extract from a calculation makes physical sense. Finally, insert the data in Table 1 into the first two columns of the spreadsheet and see how far the train went. Hand in a printout of the final spreadsheet and a printout of the formulae for the first row.

Plot the acceleration, velocity and position as functions of time, what is the maximum speed attained during this run? In your report, discuss the quality of the data and how it could be improved.

4 Least Squares Fitting of Thermocouple Data

The previous two exercises have demonstrated techniques of modelling and simple data analysis. Now we will use the spreadsheet to perform a more complex calculation and one that you will have many opportunities to use when analyzing experimntal data: least squares fitting of data. In this section, we will treat the most simple case, often called linear regression.

When analyzing experimental data of all sorts, you are often faced with the need to draw the “best” straight line fit to the data points. While it is easy enough to graph the data on a sheet of paper and then draw such a line by eye, it is often necessary to evaluate the statistically preferred line numerically in order to extract some important quantity from the data. First we will derive the formulae used in linear regression from first principles, using Calculus, then we will apply this technique to a real, physical example where linear regression can be used. The mathematics involved in this exercise is not complex, however, the concept of partial derivatives is used and much of the algebra is left out for simplicity.

Suppose that you have a set of N experimental data points which you wish to fit with a straight line. Let us define the N data points as ordered pairs

$$x_i, y_i \quad i = 1, 2, \dots, N.$$

Suppose, further, that the ordered pairs follow an approximately linear relationship and in analyzing the data we wish to find the straight line such that the sum of the “distance” from the line to each data point is minimized. This line can be described by two parameters, the slope, m , and the y -axis intercept, b . The equation of the straight line is therefore:

$$y(x) = mx + b \tag{12}$$

The total “distance” squared of the data points to the line is simply defined as the sum of the squared differences between each data point and the straight line at that same point x_i :

$$D = \sum_{i=1}^N (y(x_i) - y_i)^2, \tag{13}$$

where

$$y(x_i) = mx_i + b \tag{14}$$

so that

$$D = \sum_{i=1}^N (mx_i + b - y_i)^2 = \sum_{i=1}^N (m^2 x_i^2 + 2mbx_i - 2mx_i y_i + b^2 - 2by_i + y_i^2) \tag{15}$$

The “best” straight line fit to the data points can be obtained by minimizing this sum with respect to both parameters (m and b). Using Calculus, in order to minimize a function with respect to a parameter, the derivative of the function, with respect to the desired parameter, is set to zero. In this case, the problem is somewhat more complex, since there are two parameters in the equation and we must minimize simultaneously with respect to both of them. Fortunately, this can be done by taking the partial derivative of the difference, D , with respect to each of the two parameters and then solving the two equations for the two unknowns, m and b . To do this, notice that the partial derivative can be taken inside the summation and that when taking the partial derivative with respect to m , b is treated as a constant, and vice versa. We then obtain the following two equations:

$$\frac{\partial D}{\partial m} = \sum_{i=1}^N (2mx_i^2 + 2bx_i - 2x_iy_i) = 0 \quad (16)$$

$$\frac{\partial D}{\partial b} = \sum_{i=1}^N (2mx_i + 2b - 2y_i) = 0 \quad (17)$$

These two equations can be solved algebraically to get the final expressions for the best fit parameters to the data.

$$m = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} \quad (18)$$

$$b = \frac{\overline{x^2y} - \bar{x}\overline{xy}}{\overline{x^2} - \bar{x}^2} \quad (19)$$

Where the bar above the quantity denotes the average of that quantity over all of the N data points. For example:

$$\overline{x^2} = \frac{\sum_{i=1}^N x_i^2}{N} \quad (20)$$

In order to perform a least squares fit to a set of experimental data points with a spreadsheet, we must have columns for each ordered pair of data points as well as the products which appear in the equations above for m and b :

$$\begin{array}{ll} x_i & \text{experimental data,} \\ y_i & \text{experimental data,} \\ x_i y_i, & \\ x_i^2, & \end{array}$$

and each column must be averaged over all N data points. The averages can then be used to calculate the slope and intercept.

Note that in order to use this method, the data points you have must be in the form of two coordinates which have a linear relationship to each other. If your coordinates do not have such a relationship, this method will not work, however, it may be possible to mathematically transform a non-linear relationship to a linear one and thus use this method. The next section, on radioactive decay, will provide a good illustration of this transformation technique. If the data points fulfill the linear condition, you must then decide which will be the x-coordinate and which will be the y-coordinate, calculate the sums and finally apply the formulae for m and b to obtain the fit.

We will now proceed to fit real experimental data in just this way. The thermocouple is a useful device for the measurement of temperatures over a wide range. The thermocouple is made by putting two dissimilar metals into mechanical contact. The different chemical potentials of the two metals results in an induced voltage which can be measured easily with a voltmeter. This thermal voltage varies approximately linearly with temperature. The thermocouple is therefore, very useful in all sorts of applications where it is necessary to measure and/or control the temperature. There are many different kinds of thermocouples, each of which is useful as a thermometer over a different temperature range. One of the most commonly used junctions is the Chromel-Alumel thermocouple. Table 2 presents data for a standard Chromel-Alumel thermocouple as a function of temperature. You will be fitting this data by the least squares fitting method described above to a straight line of the form:

$$T = T_0 + mV \quad (21)$$

Where T is the temperature and V is the voltage. As you can see by comparing Eqs. 21 and 12, the temperature, T takes the place of y and the voltage, V is equivalent to x .

Table 2: Temperature vs. Voltage for a Chromel-Alumel Thermocouple

T (°C)	V (mV)	T (°C)	V (mV)	T (°C)	V (mV)
0	-1.000	350	13.292	700	28.128
50	1.022	400	15.395	750	30.214
100	3.095	450	17.513	800	32.277
150	5.137	500	19.640	850	34.314
200	7.137	550	21.772	900	36.325
250	9.151	600	23.902	950	38.310
300	11.207	650	26.022	1000	40.269

You should set up the spreadsheet as described above, putting the data in columns and taking averages which you will then use to compute the intercept, T_0 , and the slope, m . Once you have the “best fit” values of the linear model for this thermocouple, you can easily convert any voltage reading from the thermocouple to the temperature of the junction. This kind of fitting is often done for temperature measuring devices, even those which do not obey simple linear relationships.

The correspondence of the actual data to the linear model in this experiment is quite good but not perfect. In your report you should show not only plots of the experimental data and the linear fit, but also of the *difference* between the fit and the data.

5 Fitting Radioactive Decay Data

The method of least squares is quite general, you need not require the two coordinates to have only a linear relationship to each other. The more general technique of non-linear least squares fitting is extremely useful in analyzing scientific data and comparing it to a model theoretical prediction. In the following exercise,

Earlier this semester, you collected data from ^{137}Ba . The Geiger counter probe was placed close to a vial of liquid containing a small amount of ^{137m}Ba . This nucleus is radioactive and decays to ^{137}Ba quite rapidly. The decay process can be described by the equation:

$$R = R_0 e^{-\lambda t} \quad (22)$$

Where R is the count rate detected by the Geiger counter as a function of time, R_0 is the rate at $t = 0$ and λ is the decay constant. The important quantity used to characterize the radioactive decay of a nucleus is the half-life. That is, the time it takes for one-half of the original nuclei to decay. This condition can be expressed by the equation:

$$R = \frac{R_0}{2}, \quad (23)$$

which leads to

$$t_{\frac{1}{2}} = \frac{\ln 2}{\lambda}. \quad (24)$$

This process does not exhibit a linear relationship between the coordinates (R and t), however, with a simple mathematical manipulation, we can cast it into a linear equation. Consider the exponential relation shown above. If we were to take the natural logarithm of both sides, we would have:

$$\ln(R) = \ln(R_0 e^{-\lambda t}). \quad (25)$$

Since the log of the product is equal to the sum of the logs, we have

$$\ln(R) = \ln(R_0) + \ln(e^{-\lambda t}), \quad (26)$$

and using the fact that exponentiation and natural logarithm are inverses of one another

$$\ln(R) = -\lambda t + \ln(R_0). \quad (27)$$

Which is simply a linear equation with

$$y \equiv \ln(R), \quad x \equiv t, \quad m \equiv -\lambda, \quad b \equiv \ln(R_0).$$

Table 3: Typical Geiger Counter Readings for ^{137m}Ba Decay

t (s)	$\frac{mR}{hr}$	t (s)	$\frac{mR}{hr}$	t (s)	$\frac{mR}{hr}$
0	8.9	330	1.3	690	0.30
30	6.8	360	1.2	720	0.24
60	5.1	390	1.0	750	0.14
90	4.9	420	0.9	780	0.15
120	3.9	450	0.85	810	0.21
150	3.4	480	0.70	840	0.13
180	2.7	510	0.54	870	0.18
210	2.3	540	0.54	900	0.11
240	2.2	570	0.52	930	0.07
270	1.8	600	0.36	960	0.03
300	1.4	630	0.36		

Table 3 shows typical data for this experiment, however, you will be using the data you measured in class for this analysis.

Using the equations above, develop a spreadsheet to extract the half-life of ^{137m}Ba . Plot the raw data and the fit together. Try plotting the error as a function of time, what does it tell you about the quality of the fit and the data? Can you think of anything that is missing in the analysis above or your data?