Fitting Data to a Mathematical Function

Fall 2023

Introduction

In many of the laboratory exercises in this course you will be collecting data representing a phenomenon in the real world and attempting to find the mathematical function that gives the best approximation to those data. A simple version of this procedure occurs when you graph linear data and find the slope and intercept of the line that best approximates those points. In some experiments, however, it is not always easy to determine exactly which line best describes your data. Today we will look at this problem in more detail and learn how to use a graphing program to simplify this task.

The goal of this exercise is to calculate the velocity of an object from a given set of data using hand calculations, and computer analysis. In this way, you will see how the program calculates the values we need. You will also calculate the SSR (*defined below*) and uncertainty in your result using these methods.

Procedure

Consider the experiment of measuring the position x of an object as a function of time as it moves along the x-axis. If the object starts at $t = 0$ with constant x-velocity, we expect the data to be modeled by the equation $x = v_o t$, where v_o is the adjustable parameter. If you plotted your data by hand and v_o (the slope) is adjusted properly, there will be as many data points above the line you have drawn as there are below, and the sum of the differences between data points and corresponding points on the line (the *residuals*, indicated by the vertical green segments in the figures below) will be very close to zero. The **s**um of the **s**quare of the **r**esiduals (SSR – also called *Chi Square*, χ^2 – indicated by the blue shaded squares in the figures) can be thought of as a function of the parameter v_o and the goal in obtaining the best-fit line is to minimize the SSR. Compare the results of the best fit of the data points in *Fig. 1a* to a poor fit illustrated in *Fig. 1b*.

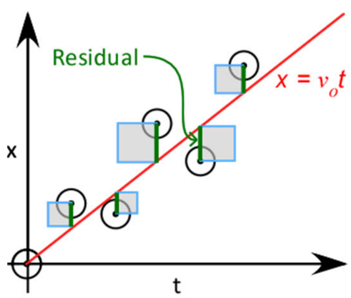

Fig. 1a: Best fit - SSR is minimized

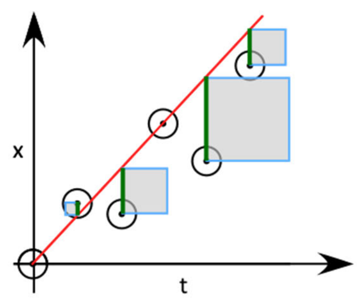

Fig. 1b: Poor fit - SSR is larger than before

The data is provided for you in today's exercise. *Carefully* copy *Data Table 1* into your journal. Remember that you will always keep a written copy of your data in your lab journal and that you should write your numbers *clearly* and *distinctly*. For example, your lab partners and your instructor should be able to tell if you have written the number '*one*' or '*seven*', and number '*five*' should not look like the letter '*s*'.

Data Table 1

t (s)	d (m)
0.0	0.0
5.0	3.0
10.0	7.5
15.0	12.3
20.0	15.8
25.0	18.8
30.0	20.2
35.0	23.3
40.0	26.6
45.0	29.7
50.0	32.1

1. *Analysis by Calculation*: Let us first derive the relationships needed for the analysis. Using calculus, the best value of v_o is calculated by minimizing the *SSR*. We define the SSR to be the sum of the square of the distance between each of the data points and the line $x = v_o t$:

$$\text{SSR} = (x_1 - v_o t_1)^2 + (x_2 - v_o t_2)^2 + \ldots \qquad \textit{Eqn. 1}$$

To get the best v_o, differentiate SSR with respect to v_o and set it equal to zero:

$$d(\text{SSR})/dv_o = 2(x_1 - v_o t_1)(-t_1) + 2(x_2 - v_o t_2)(-t_2) + \ldots = 0$$

$$= 2\{-(t_1 x_1 + t_2 x_2) + v_o(t_1^2 + t_2^2 + \ldots)\} = 0 \qquad \text{(factor out } \mathbf{-2} \text{ and clean up)}$$

The quantity inside the parentheses must equal zero if SSR is a minimum:

$$v_o(t_1^2 + t_2^2 + \ldots) - (t_1 x_1 + t_2 x_2) = 0$$

So, solving for v_o gives you the following result (note that the **Σ** symbol means "the sum of the terms"):

$$v_o = \frac{\sum t_i x_i}{\sum_i t_i^2} \qquad \textit{Eqn. 2}$$

An essential portion of our analysis of future experimental data will involve the calculation of the *uncertainty* in the slope, v_o. The uncertainty of a measurement or calculation is stated by giving a range of values that likely enclose the actual value. Calculating the uncertainty in v_o is accomplished by first calculating the *standard error* in v_o:

$$\text{Standard Error} = \sqrt{\frac{SSR}{(N-1)\sum t_i^2}} \qquad \textit{Eqn. 3}$$

Where N is the number of data points you have (e.g., $N = 11$ for your data from the table). *The uncertainty in the slope is ± **twice** the standard error in v_o.* Having the uncertainty equal to twice the standard error gives us what we call a 95% confidence interval, which means that if we look at a set of data to determine v_o, 95% of the data will fall within the calculated uncertainty range.

Now for your calculations, use the following procedure:

a. Write a calculation table (*as shown on the next page*) **sideways and large enough to hold all your data** on a new page in your journal and complete the first five columns.
b. Use *Eqn. 2* and the calculated sums of columns 4 and 5 to calculate v_o from the data.
c. Use this value of v_o to complete columns 6 and 7.
d. Calculate the sum of column 7. This sum follows *Eqn. 1* and will give you the SSR.
e. Finally, use *Eqn. 3* (which includes the sums of columns 5 and 7) to calculate the standard error, and then the uncertainty (which is *twice the standard error*) in v_o. Write your v_o and its uncertainty as v_o = ____ ± _____ *{units}*

IT IS IMPORTANT NOT TO ROUND OFF YOUR VALUES WHILE PERFORMING THESE CALCULATIONS! DOING SO WILL INTRODUCE ROUNDING ERROR, AND YOU WILL MISS THE POINT OF THIS EXERCISE. USE THE MEMORY FUNCTION OF YOUR CALCULATOR TO SAVE RESULTS FOR SUBSEQUENT CALCULATIONS.

Calculation Table: *Note That You Should Determine the Correct Value of {units} For Each Column Title!*

i	t_i {units}	x_i {units}	$t_i x_i$ {units}	t_i^2 {units}	$v_o t_i$ {units}	(Residuals)^2 $(x_i - v_o t_i)^2$ {units}
1	0.0	0.0	0.0	0.0	0.0	0.0
2	5.0	3.0	15	25	⋮	⋮
3	10.0	7.5	75	100	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮
11	50.0	32.1	⋮	⋮	⋮	⋮
			Sum =	Sum =		Sum =

2. *Computer Analysis*: Now you will perform the same calculations using the computer. We will be using the program KaleidaGraph to analyze experimental data in this course. Follow *all* the instructions in *Graphing & Curve Analysis Using KaleidaGraph* to perform the following for your data:

   - Create a data table.
   - Create a graph and be able to explain the physical variables that each axis represents.
   - Determine the function that best fits the data (e.g., linear, polynomial, exponential, etc.)
   - Choose a "user-defined" function in Kaleidagraph that will allow you to calculate the value of the *fit parameters* from this function.
     - ➢ The *fit parameters* are the coefficients (a and b for today's data set) that are adjusted to minimize the function, giving you the lowest SSR value. You will also determine the physical quantities that each parameter represents (e.g., position, velocity, etc.)
     - ➢ *Note that you will use a fit function this week lab called* ***Linear Through Origin****, found in the General → Curve Fit menu.* *This Fit Will Only Be Used For This Week's Lab, And Not Used Again During This Course!*
   - Calculate the uncertainty in the slope.
   - Calculate the residuals and the SSR of the fit.
   - Add your name to your data as follows: Type your name (and any lab partners) in the column title of the first blank column, and then type **0** in the first cell of this column. Print the data table and your name(s) will appear on the printout. *Note that the column title will only print 15 characters, so you can enter just your first or last name or use initials when working with partners.*
   - Use the text tool to type your name(s) on the graph, and then print the graph.

   KaleidaGraph is going through essentially the same process you did to calculate the best-fit slope for the line, but obviously more quickly.

Analyzing the Graph

The first step of our analysis with graphed data will always be to look at the best-fit function and see how it compares to the expected theory for the experiment. Today you plotted data of distance (d) and time (t) for an object moving at a constant velocity. Velocity is calculated by dividing the distance traveled over the time interval. You would not want to calculate the velocity for each point in the sample data; imagine the number of calculations if the data set had 100 or 1000 points. Looking at your graph, you should note that only a few of the data points touch the best-fit function, so individual calculations would give you a variety of answers. Let us see how the best-fit function matches with the theory expected for this data.

Today you chose a linear function for this data that forced the fit through the origin. The form of a linear function is $y = (slope)\,x + b$; since we forced the fit through the origin, $b = 0$ and the linear function simplifies to $y = (slope)\,x$. The definition of velocity gives us $v = d/t$; you plotted distance on the vertical (y) axis and time on the horizontal (x) axis, so we can algebraically rearrange the velocity definition so that it follows the pattern of a linear function, as shown in the equations on the next page:

$$d = (v)\,t$$
$$y = (slope)\,x$$

Therefore, you can clearly see that the slope of your line gives you the velocity. This is how we will analyze graphed data in all future experiments, whether the data is linear or non-linear (as we will see next week.)

Discussion

- Your discussion this week will consist of a summary of the results you obtained for your data as analyzed by each method:
  - The slope, uncertainty and SSR from both your hand calculations and your KaleidaGraph plot
    - *Note #1*: write your slope and its respective uncertainty as v_o = ____ ± _____ *{units}* (e.g., v_o = 1.234 ± 0.567 *{units},* using the correct number of significant figures)
    - *Note #2*: If your results from the hand-calculations (part 2) and KaleidaGraph don't match exactly, you should find your mistake! Be sure to ask your instructor for assistance.