

# Our Curve Fitting Computer Programs

## 1. Introduction

*KaleidaGraph* and *Graphical Analysis* are useful tools that are available to you (throughout your time at IWU) for the processing of data acquired during an experiment. You will find it quite worthwhile to learn how to use these successfully and correctly. This brief tutorial will save you a great deal of time and frustration in your lab efforts.

In a laboratory experiment, we hope to extract a general result regarding the system being investigated. This often equates to finding a mathematical model that describes the results (*i.e.* the equation for a curve that passes closest to the measured data). While it is possible to find such a “fit” by hand, this can be done quickly by software packages.

Given a number of data points that lie along an arbitrary and perhaps complicated curve, it is possible to write an equation for this curve by choosing a sufficiently large polynomial. If we have a graph of data,  $y$  vs.  $x$  such that there is a functional relationship between  $y$  and  $x$ , this may be expressed as:

$$y = c_0 + c_1x + c_2x^2 + \cdots + c_nx^n \quad (1)$$

The computer uses a method called the least squares approximation (which will be described later) to determine the values of the coefficients  $c_0 \dots c_n$ . Inserting these numerical values into the above polynomial, gives us an equation that “fits” the data, and *might* be considered to be a solution to the problem. The largest exponent of the independent variable  $x$  gives the degree of the polynomial; for example, the above polynomial is of the  $n^{\text{th}}$  degree.

There are some obvious limitations to this analysis. If  $x$  is the independent variable, the above expression is physically valid only between the smallest and largest experimental values of  $x$  on the graph. Extrapolated values beyond this range are uncertain, and no estimation of their uncertainty can be made.

In most physical applications in this course, we will have bivariate data which can be expressed either as a straight line:

$$y = c_0 + c_1x \quad (2)$$

or as a parabola:

$$y = c_0 + c_2x^2 \quad (3)$$

Rarely will higher degrees ( $x^3$ ,  $x^4$ , *etc.*) be needed. The experimenter must tell the computer which degree polynomial they want to use, a decision, which should be based upon physical knowledge of the system being investigated.

If two variables are clearly proportional to each other (*i.e.*, the graph is a straight line), then it is senseless to ask the computer to fit a quadratic, since all the coefficients after the first will be zero (or close to zero due to experimental error). Simply put, there is no substitute for common sense.

As an example, suppose  $y$  is a parabolic function of  $x$  (a 2<sup>nd</sup> degree polynomial), subject to the initial condition that at  $x = 0$ ,  $y = 0$  as well. It is clear from the general equation for a 2<sup>nd</sup> degree polynomial

$$y = c_0 + c_1x + c_2x^2 \quad (4)$$

that the first coefficient,  $c_0$ , should be zero. ***Physical reasoning should always be used to check the computer's output and to interpret what it means if  $c_0 \neq 0$  in the fit.***

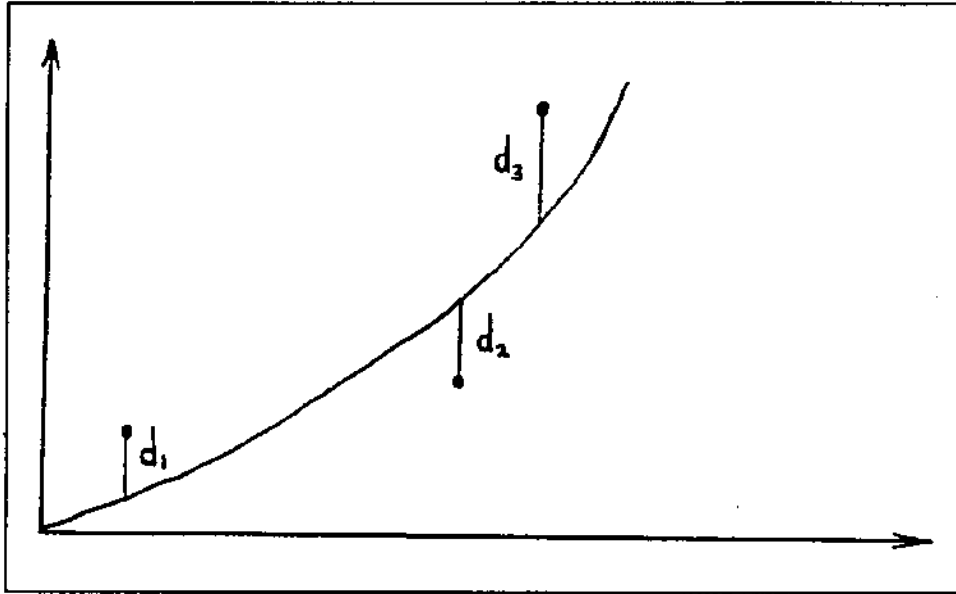
## ***2. The Least Squares Approximation used by your Software:***

Consider the following problem which the scientist often encounters: How to mathematically describe my experimental data, that is, what is the equation for the functional form of my data? Given some data points we could try to draw a curve, by hand, whose shape has the same general pattern as the data. This method is quick and easy and you will be employing it in the analysis of some of your experiments; however, there are two disadvantages to this method. First, every individual has his or her own notions of what constitutes a "best fit" and consequently the curve lacks the objectivity, which any scientific endeavor attempts to embrace. Secondly, except for a straight line, it is difficult if not impossible to write down an equation for a given curve that the human hand has drawn. This is a major drawback when we wish to know actual numerical parameters of our curve and not simply its shape.

The method of least squares addresses the above difficulties: it precisely and objectively defines what is meant by the "best fit" to a group of data points, so that the parameter of the curve can be adjusted until an equation for the fitted curve is found.

If we were fitting a polynomial to some points, our intuitive notion of a good fit would probably be a curve in which the distance from any data point to curve was small. Hence, the best fit to our data might be the curve that minimizes the sum of the distances between the data points and curve.

The method of least squares adopts a similar, yet slightly *different* criterion for a good fit. Instead of concerning the distance from the data point to the curve, we consider only the distances in the  $y$ -direction, as seen in Figure 2.



**Figure 2**

The least squares method minimizes the sum of the squares of these distances. This tends to place as much importance on minor adjustments of small discrepancies (within the uncertainty of the measurement) as large discrepancies for which the deviation sense is well known.

$$\sum_i d_i^2 = \text{minimum} \quad (5)$$

When a function,  $y(x)$ , is fitted to a data set, the software begins (with trial values of the coefficients) to calculate the deviation of the curve from the data. For each data point  $(x_i, y_i)$ , the software calculates the difference in the corresponding  $y$ -values of the polynomial:

$$d_i = y_i - (c_0 + c_1x_i + c_2x_i^2) \quad (6)$$

The quantity  $\sum_i d_i^2$  is then computed. This process is repeated with different values for the coefficients ( $c_0$ ,  $c_1$ , and  $c_2$ ), until it finds a set of coefficients for which the quantity  $\sum_i d_i^2$  is less than it is for any set. This set of coefficients is the least squares “fit” to the data.

### 3. Graphing Program Exercises:

- a. Familiarize yourself with *KaleidaGraph* or *Graphical Analysis* by reading the information available in the lab.
- b. Use the following data to prepare a graph\* and determine the relationship between the two variables. The data comes from a special set of aluminum rectangles. Length and width for each triangle has been measured in centimeters and tabulated below. Create a graph of the data and show the graph to your instructor. Be certain to include the title, label the axes, put units on the axes, and give the mathematical relationship between the variables. Give the equation that states the relationship between the variables. State the units and physical significance of the constant.

Length (cm)	Width (cm)
1.0	10
2.0	5.0
4.0	2.5
5.0	2.0
6.0	1.67
8.0	1.25

- c. Use the following data to prepare a graph\* and determine the relationship between the two variables. The data comes from a set of aluminum disks. Diameters and circumferences for each disk has been measured in centimeters and tabulated below. Create a graph of the data, plotting diameter on the X-axis and circumference on the Y-axis. Show the completed graph to your instructor. Be certain to include the title, label the axes, put units on the axes, and give the mathematical relationship between the variables. Give the equation that states the relationship between the variables. State the units and physical significance of the constant. How certain are you of the value of the constant? How many significant figures should it have? How can you tell how reliable the curve fit is?

Diameter (cm)	Circumference (cm)
4.0	12.6
6.0	18.8
8.0	25.1
9.0	28.3

- \* Each graph should contain data points and the fitted line or curve only. Do not connect data points with lines. Do not plot the calculated “independent variable” and “dependent variable” points.