Fitting Data to a Straight Line

Dr. Darrel Smith
Department of Physics & Astronomy
Embry-Riddle Aeronautical University
Prescott, AZ 86301

(Dated: 5 February 2015)

This note describes how to fit a straight line to data exhibiting a linear trend in y vs. x. The x-y coordinates for the data points, along with their respective error bars are analyzed using a linear regression fit to a straight line (i.e., $y = a + bx$).

I. INTRODUCTION

The following describes a procedure for fitting a straight line to data containing error bars. An example is shown below where a straight line ($y = a + bx$) is fit a a data set x_i, y_i, σ_i where σ_i is the distance between the data point and the end of the error bar. The fitting procedure described below is used to fit data where the *error bars* can vary from one data point to the next. The fitting procedure is used to determine the following quantities:

$$\begin{aligned} a &= \text{the } y \text{ intercept} \\ \sigma_a &= \text{the uncertainty in the } y \text{ intercept} \\ b &= \text{the slope, and} \\ \sigma_b &= \text{the uncertainty in the slope} \end{aligned}$$

Assume we have 5 data points that display a linear trend as shown in Fig. 1 below. The error bars extend above and below the measured value y_i by the measured uncertainty $\pm 1\sigma_i$.

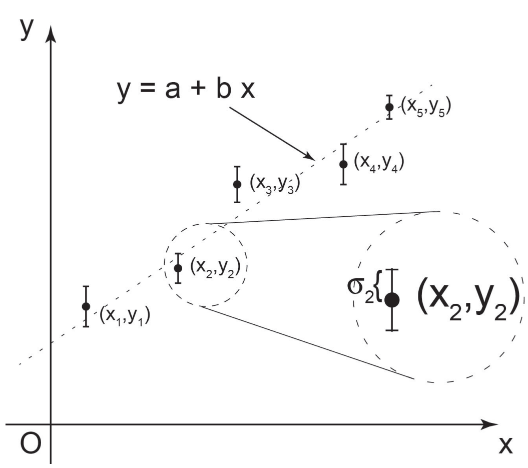

FIG. 1. Fitting five data points using a straight line fit. In general, the error bars can be different for each measurement.

II. CALCULATIONS

The four parameters a, σ_a, b, and σ_b are calculated using the following equations:

$$a = \frac{1}{\Delta} \begin{vmatrix} \sum \frac{y_i}{\sigma_i^2} & \sum \frac{x_i}{\sigma_i^2} \\ \sum \frac{x_i y_i}{\sigma_i^2} & \sum \frac{x_i^2}{\sigma_i^2} \end{vmatrix}$$

$$b = \frac{1}{\Delta} \begin{vmatrix} \sum \frac{1}{\sigma_i^2} & \sum \frac{y_i}{\sigma_i^2} \\ \sum \frac{x_i}{\sigma_i^2} & \sum \frac{x_i y_i}{\sigma_i^2} \end{vmatrix}$$

$$\sigma_a^2 = \frac{1}{\Delta} \sum \frac{x_i^2}{\sigma_i^2} \qquad \sigma_b^2 = \frac{1}{\Delta} \sum \frac{1}{\sigma_i^2}$$

$$\Delta = \sum \frac{1}{\sigma_i^2} \sum \frac{x_i^2}{\sigma_i^2} - \left(\sum \frac{x_i}{\sigma_i^2} \right)^2$$

where the sum $\sum$ is over the index i from $i = 1, lots N$ and N is the number of data points.

It's a bit cumbersome to keep writing these equations in a Mathematica program; however, one of our students, Ian Brubaker (Spring 2015) wrote a Mathematica package called `DetFit`. You can go to my website `/physicsx/` to get a copy of Ian's `DetFit` module.

1. Download the Mathematica package `DetFit5.m`
2. In Mathematica, use the `File->Install` command to install `DetFit5.m`. You only have to do this operation once.
3. Somewhere near the beginning of your Mathematica program, include the following statement: `Get["DetFit5‘"]` to load the `DetFit` package.
4. To use the `DetFit` package in your Mathematica program, use the function `DetFit[x,y,`σ`]` where x and y are separate lists of the (x, y) coordinates of your data points, and σ is a list of the error bars. All three *lists* should be the same length.
5. There is a test program, `test program.nb`, that you can download to see how this program is used to fit 8 data points, each with a different error bar.

Enjoy !!