CDF/DOC/CDF/PUBLIC/1303
11-May-1992
Version 1.0

A Straight Line Fit With Four Parameters

Hovhannes Keutelian
Fermilab

Abstract

A method of least squares fit using implicit functions is described.

1 Introduction

The standard approach of least squares fit to a line [1,2,3] (as described in textbooks and written about in statistical journals) fails when the value of the slope is much larger than one. My attempt to cure this problem has led me to a general approach of least squares fit of implicit functions.

When applying the "standard least squares" to determine the parameters of a line, almost exclusively the following form of line equations are used; $y = mx + b$ or $y = m(x - a) + b$. The advantage of the last equation is that the fit is done within the range of data. The application of this method to the case where $m = 0, b = 0$, with $\sigma \equiv \sigma_x = \sigma_y$, gives the following results: $< m >= 0$ and $< b >= 0$. The scatter plot of a sample for this problem is shown in figure 1a. But the application of this method to another sample, as shown in figure 1b, gives also $< m >= 0$ and $< b >= 0$. This is incorrect. In both samples the data is distributed uniformly along the axis of xy coordinate system; along the x axis in the first sample and along the y axis in the second sample. Each entry of both samples is smeared using Gaussian distributions with uncertainties $\sigma \equiv \sigma_x = \sigma_y$.

The reader may point out that an inappropriate form of a line equation is used for the second sample, see figure 1b. That the appropriate form to use is $x = ly + a$. Thought process such as this reveal that one has to choose a dependent or an independent variable when using the standard Least Squares method. The only way to remove the ambiguity of the state (dependent or independent) of a variable is to use implicit functions.

2 The General Method

The implicit function, $\mathrm{F} = ly + mx + b \equiv 0$, represents a line equation where both x and y variables are treated equally without referring them as independent or dependent variables. A detailed procedure that leads to solutions for l and m is given below. The detailed steps will help you to follow what is done to arrive at a general least squares expression of an implicit function.

The least squares sum for a line equation is:

$$S(x_j^0, y_j^0, \lambda_j; b, l, m) = \sum_{j=1}^{N} \left[\left(\frac{x_j - x_j^0}{\sigma_{x_j}} \right)^2 + \left(\frac{y_j - y_j^0}{\sigma_{y_j}} \right)^2 + 2\lambda_j \left[l y_j^0 + m x_j^0 + b \right] \right]. \qquad (1)$$

The above expression also contains the case where both variables (x,y) have measurement uncertainties; these uncertainties are σ_{x_j} and σ_{y_j}. The above expression is applicable only to the case where σ_{x_j} and σ_{y_j} do not depend on parameters l, m, and b. The Lagrange multipliers, λ_j, depend on the values of b, l, and m, but they do not contribute to the least squares sum because the implicit function is identically equal to zero for all values of j. The variables x_j^0 and y_j^0 are exact, these and the parameters (l, m, b) allow the implicit function to be zero for any value of j.

An informative article on the least squares fit when both variables have uncertainties can be found in reference [4]. A detailed description of the lagrange multiplier technique can be found in reference [1]. Both sources have influenced the outcome of this note.

The parameters $x_j^0, y_j^0, \lambda_j; b, l, m$ are determined by minimizing the least squares sum S. It is possible to solve for x_j^0, y_j^0, λ_j in terms of l, m, b, and hence be able to write the least squares sum without the lagrange multipliers. The least squares sum S is minimized with respect to x_j^0, y_j^0, λ_j,

$$\begin{aligned} \frac{\partial S}{\partial x_j^0} &= -2 \left(\frac{x_j - x_j^0}{\sigma_{x_j}^2} \right) + 2\lambda_j m = 0 \qquad (2) \\ \frac{\partial S}{\partial y_j^0} &= -2 \left(\frac{y_j - y_j^0}{\sigma_{y_j}^2} \right) + 2\lambda_j l = 0 \qquad (3) \\ \frac{\partial S}{\partial \lambda_j} &= 2 \left[l y_j^0 + m x_j^0 + b \right] = 0 \qquad (4) \end{aligned}$$

and simultaneous solutions of the above expressions are:

$$\begin{aligned} x_j - x_j^0 &= -\frac{m\sigma_{x_j}^2}{m^2\sigma_{x_j}^2 + l^2\sigma_{y_j}^2} \left[l y_j + m x_j + b \right] \qquad (5) \\ y_j - y_j^0 &= -\frac{l\sigma_{y_j}^2}{m^2\sigma_{x_j}^2 + l^2\sigma_{y_j}^2} \left[l y_j + m x_j + b \right] \qquad (6) \\ \lambda_j &= -\frac{1}{m^2\sigma_{x_j}^2 + l^2\sigma_{y_j}^2} \left[l y_j + m x_j + b \right]. \qquad (7) \end{aligned}$$

After substituting the above solutions in equation 1, the least squares sum is now expressed in terms of l, m, and b:

$$S(b, l, m) = \sum_{j=1}^{N} \frac{\left[l y_j + m x_j + b \right]^2}{(m^2\sigma_{x_j}^2 + l^2\sigma_{y_j}^2)}. \qquad (8)$$

Note, the implicit function is symmetric under the exchange of $l \to m$, $m \to l$, and also under the exchange of $x \to y$, $y \to x$. This symmetric property is also present in the quadrature sum $(m^2\sigma_{x_j}^2 + l^2\sigma_{y_j}^2)$. This is the first indication that no singularity exists (the case $l = m \equiv 0$ is not a physical solution to a line equation) in this least squares sum.

Two parameters completely determine the behavior of a line in a flat plane. A location parameter to carry out translation and a scale parameter to define an angular dependence about the location parameter. In the least squares sum there are two scale parameters l and m, where the location parameter is b. The procedure of finding eigenvalues and eigenvectors are used to obtain the relationship between l and m.

The problem of minimizing the least squares sum gets easier by adding an extra location parameter. With an extra location parameter it is possible to assign to each variable (x, y) a unique translation parameter. Now, the least squares sum is written as:

$$S(a,b,l,m) = \sum_{j=1}^{N} \frac{[l(y_j - b) + m(x_j - a)]^2}{(m^2\sigma_{x_j}^2 + l^2\sigma_{y_j}^2)}. \tag{9}$$

In this form the line equation is symmetric under exchange of both the location and the scale parameters. Also, the arbitrary choice of location parameter $(a = 0)$ is removed from the line equation. Minimization of $S(a,b,l,m)$ with respect to b yields:

$$\frac{\partial S}{\partial b} = -2l \sum_{j=1}^{N} \frac{[l(y_j - b) + m(x_j - a)]}{(m^2\sigma_{x_j}^2 + l^2\sigma_{y_j}^2)} = 0 \tag{10}$$

$$a = \frac{\sum_{j=1}^{N} \frac{x_j}{m^2\sigma_{x_j}^2 + l^2\sigma_{y_j}^2}}{\sum_{j=1}^{N} \frac{1}{m^2\sigma_{x_j}^2 + l^2\sigma_{y_j}^2}} \tag{11}$$

$$b = \frac{\sum_{j=1}^{N} \frac{y_j}{m^2\sigma_{x_j}^2 + l^2\sigma_{y_j}^2}}{\sum_{j=1}^{N} \frac{1}{m^2\sigma_{x_j}^2 + l^2\sigma_{y_j}^2}}. \tag{12}$$

The parameters a and b are determined by setting each term in the numerator of $\frac{\partial S}{\partial b}$ to zero. Note that a and b are functions of the scale parameters (l, m). For the case where $\sigma_{x_1} \neq \sigma_{x_2} \neq \sigma_{x_3} \neq \ldots \neq \sigma_x$ and $\sigma_{y_1} \neq \sigma_{y_2} \neq \sigma_{y_3} \neq \ldots \neq \sigma_y$, one begins an iterative procedure (initially setting $l = m$) until the values of a and b converge. Usually one iteration is sufficient to achieve convergence. When solving for l and m, the parameters a and b are treated as constants.

The process of minimizing the least squares sum S(a,b,l,m) with respect to l and m leads to the following equations:

$$\frac{\partial S}{\partial l} = 2 \sum_{j=1}^{N} \frac{[l(y_j - b) + m(x_j - a)]}{(m^2\sigma_{x_j}^2 + l^2\sigma_{y_j}^2)^2} \left[m^2\sigma_{x_j}^2(y_j - b) - ml\sigma_{y_j}^2(x_j - a)\right] = 0 \tag{13}$$

$$\frac{\partial S}{\partial m} = 2 \sum_{j=1}^{N} \frac{[l(y_j - b) + m(x_j - a)]}{(m^2\sigma_{x_j}^2 + l^2\sigma_{y_j}^2)^2} \left[l^2\sigma_{y_j}^2(x_j - a) - ml\sigma_{x_j}^2(y_j - b)\right] = 0. \tag{14}$$

After some algebraic manipulations the above equations reduce down to the following sum:

$$m\left(l\sum_{j=1}^{N}\frac{\sigma_{x_j}}{\sigma_{eff}^4}(y_j-b)^2+m\sum_{j=1}^{N}\frac{\sigma_{x_j}}{\sigma_{eff}^4}(y_j-b)(x_j-a)\right)$$
$$-l\left(l\sum_{j=1}^{N}\frac{\sigma_{y_j}}{\sigma_{eff}^4}(y_j-b)(x_j-a)+m\sum_{j=1}^{N}\frac{\sigma_{y_j}}{\sigma_{eff}^4}(x_j-a)^2\right)=0 \quad (15)$$

Where the $\sigma_{eff}^2 = (m^2\sigma_{x_j}^2 + l^2\sigma_{y_j}^2)$ is the effective variance [4] of a line equation. The solutions for the following cases; 1) $\sigma_{x_j} = 0$, and 2) $\sigma_{y_j} = 0$, are trivial. They can be read off of equation 15. The solutions for cases one and two can only be written down in terms of m/l or l/m. When the measurement uncertainties of both x and y variables are greater than zero, two equations are needed to solve for l and m. A pair of equations, that are functions of parameters l and m, is obtained by setting each term in equation 15 to zero. Below, these set of equations are expressed in matrix form.

$$\begin{pmatrix} \frac{\sigma_{y_j}^2}{\sigma_{eff}^4}(x_j-a)^2 & \frac{\sigma_{y_j}^2}{\sigma_{eff}^4}(x_j-a)(y_j-b) \\ \frac{\sigma_{x_j}^2}{\sigma_{eff}^4}(x_j-a)(y_j-b) & \frac{\sigma_{x_j}^2}{\sigma_{eff}^4}(y_j-b)^2 \end{pmatrix}\begin{pmatrix} m \\ l \end{pmatrix}=0 \quad (16)$$

Where the summation signs are not shown, but sums over j are to be carried out. The procedure of finding eigenvalues and eigenvectors is used to determine l and m. As in the case of determining parameters a and b, when $\sigma_{x_1} \neq \sigma_{x_2} \neq \sigma_{x_3} \neq \ldots \neq \sigma_x$ and $\sigma_{y_1} \neq \sigma_{y_2} \neq \sigma_{y_3} \neq \ldots \neq \sigma_y$, one begins an iterative procedure (initially setting $l = m$ in σ_{eff}) until the values of parameters converge. The effective variance σ_{eff}^2 is treated as a constant to solve for l and m. The solutions are:

$$l = \frac{L}{\sqrt{(L^2+M^2)}} \quad (17)$$

$$m = \frac{M}{\sqrt{(L^2+M^2)}} \quad (18)$$

$$L = \frac{1}{2}\left[(A-D)+\sqrt{((A-D)^2+4BC)}\right]+C \quad (19)$$

$$M = \frac{1}{2}\left[(A-D)-\sqrt{((A-D)^2+4BC)}\right]-B \quad (20)$$

$$A = \sum_{j=1}^{N}\frac{\sigma_{y_j}^2}{\sigma_{eff}^4}(x_j-a)^2 \quad (21)$$

$$B = \sum_{j=1}^{N}\frac{\sigma_{y_j}^2}{\sigma_{eff}^4}(x_j-a)(y_j-b) \quad (22)$$

$$C = \sum_{j=1}^{N}\frac{\sigma_{x_j}^2}{\sigma_{eff}^4}(x_j-a)(y_j-b) \quad (23)$$

$$D = \sum_{j=1}^{N} \frac{\sigma_{x_j}^2}{\sigma_{eff}^4}(y_j - b)^2 \tag{24}$$

The above solutions are obtained using the eigenvalue with the magnitude closest to zero. This choice allows for correct solutions when σ_x or σ_y tend to zero. The quadrature sum of parameters l and m is equal to one, $m^2 + l^2 = 1$. This relationship conveys a process that is fundamental when solving implicit functions. Details of writing down the correct form of least squares sum for implicit functions is given in the next section.

A few comments on the solution of a line equation. 1) A line equation can also be written in terms of sine and cosine functions, $ycos\phi + xsin\phi = 0$. A comparison of this equation with $ly + mx = 0$, and $y = m'x$, shows that the problem of solving $m' = tan\phi$ is turned into solving $m = sin\phi$ and $l = cos\phi$. Hence l and m have the same properties that of $sin\phi$ and $cos\phi$ functions. 2) In the standard procedure the values of either l or m is set to one. Where in the general method, as described in this note, the quadrature sum $(m^2 + l^2)$ is set to one. 3) To obtain an optimum minimization it is better to solve equation 9 with location parameters defined as in equations 11 and 12. The iterative solutions give an excellent approximate results for most practical problems, but for a puritan the solution of equation 9 is the correct result. 4) Depending on whetter A is larger than D or vice versa, it is possible to rewrite equations 19 and 20 to avoid roundoff errors. 5) Before starting an iterative procedure, set $\sigma_{x_j} \equiv 1$, $\sigma_{y_j} \equiv 1$, and $\sigma_{eff} \equiv 1$, and solve for a, b, l, and m. Next start iterating until the values of parameters converge. 6) Uncertainties on the parameters using parabolic [3] approximation has a complicated form, and are not given. 7) Also the least squares sum can be written in terms of already calculated quantities A, B, C, and D, thus saving computing time.

3 Implicit Functions and Least Squares

The solution described in the previous section can be applied to any implicit function that is differentiable. The required differentiation is of first order and it is with respect to the measured variables. If z is considered a measured quantity, it is always possible (but not necessarily desirable) to write an implicit function using the following form, $q(z - c)$. In this form 'c' is the location parameter and 'q' is the scale parameter. When fitting an implicit function always set the quadrature sum of scale parameters to one, $l^2 + m^2 + p^2 + \ldots = 1$.

Given an implicit function $F(x, y, z, \ldots; l, m, q, \ldots, a) \equiv 0$, the least squares sum is:

$$S(a, l, m, q, \ldots) = \sum_{j=1}^{N} \frac{[F(x_j, y_j, z_j, \ldots; l, m, q, \ldots, a)]^2}{((\frac{\partial F}{\partial x})^2\sigma_{x_j}^2 + (\frac{\partial F}{\partial y})^2\sigma_{y_j}^2 + (\frac{\partial F}{\partial z})^2\sigma_{z_j}^2 + \ldots)}. \tag{25}$$

The derivatives are evaluated by using measured values, $x_j, y_j, z_j, \ldots$. This is a very good approximation compared to evaluating with ideal values; $x_j^0, y_j^0, z_j^0, \ldots$. In the above sum the measurement uncertainties are treated as having no dependence on parameters whatsoever.

The above sum can be minimized with respect to parameters $a, l, m, q, \ldots$ using a well established minimization software package. Without the constraint $l^2 + m^2 + p^2 + \ldots = 1$, the least squares sum will not converge.

The advantage of using the general form of effective variance, $\sigma^2_{eff} = ((\frac{\partial F}{\partial x})^2\sigma^2_{x_j} + (\frac{\partial F}{\partial y})^2\sigma^2_{y_j} + (\frac{\partial F}{\partial z})^2\sigma^2_{z_j} + \ldots)$. can be appreciated by considering a case where the first partial derivative with respect to a variable is quite large compared to others. Since the measurement uncertainty of that particular variable dominates, the scale parameter associated with that variable will have the largest magnitude. Thus, the above form of effective variance allows an easy transition among variables.

The above statements can be easily verified using implicit polynomial functions of any order. By using orthonormal polynomials [2,3] and the procedures in section 2, one can derive the constraint $l^2 + m^2 + p^2 + \ldots = 1$ in a straight forward fashion. Other implicit functions (with linear parameters) can also be used to verify the above conjecture.

In cases where one has to deal with more than one implicit function, the procedure of writing down the least squares sum is not as straight forward as given above. The least squares sum will contain cross products of implicit functions. In addition, each term in the sum is multiplied by weights which depend on the functional forms of implicit functions. Also, each term in the effective variance contains products of partial derivatives. Least squares problems that make use of more than one implicit functions occur naturally in three or higher dimensions. An example is fitting a curve in three or higher dimensions. When there are more than one implicit functions, then there are more than one constraint. The quadrature sum of scale parameters for each implicit sum is set to one. The steps given in section 2 can be used (highly tedious) to solve problems with more than one implicit functions. There is no choice but to expand each implicit functions about the known values of parameters to obtain effective variance of a problem.

Uncertainties on the parameters using parabolic approximation has a complicated form, and is misleading for a general implicit function. An efficient thing to do is to use uncertainties given by a minimization package, such as MINUIT [5]. In case of MINUIT, the results do reliably agree with theoretical definition.

The constraint $l^2 + m^2 + p^2 + \ldots = 1$ is also valid when carrying out minimization using other than least squares method. For example, the quadrature sum of sine and cosine functions is independent of any fitting procedure. Above, it was presupposed that the measurement uncertainties belong to Gaussian distributions. The constraint does not depend on the distribution of measurement uncertainties. The distribution of measurement uncertainties do AFFECT the functional form of effective variance.

4 Examples and Applications

Example 1) A plane in xyz coordinates. The implicit function is $F(x, y, z; b, m, l, p) = mx + ly + pz + b \equiv 0$. Hence, the constraint is $l^2 + m^2 + p^2 = 1$, and the effective variance $\sigma^2_{eff} = \sigma^2_{y_j} l^2 + \sigma^2_{x_j} m^2 + \sigma^2_{z_j} p^2$.

Example 2) A second degree polynomial in xy plane. The general form of implicit function is $G(x, y; A, B, C, D, E, F) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F \equiv 0$. The constraint is $A^2 + B^2 + C^2 + D^2 + E^2 = 1$, and the effective variance $\sigma^2_{eff} = \sigma^2_{x_j}(2Ax_j + By_j + D)^2 + \sigma^2_{y_j}(2Cy_j + Bx_j + E)^2$.

Example 3) The function of interest is $y = ae^{mx}$. The implicit function for this problem is $G(x, y; l, m, b, a) = l(y - b) + e^{m(x-a)} \equiv 0$. The constraint is $l^2 + m^2 = 1$, and the effective variance $\sigma^2_{eff} = \sigma^2_{x_j}(me^{m(x-a)})^2 + \sigma^2_{y_j}l^2$. Note that the transition to two extreme situations, $(x - a) \rightarrow -\infty$ and $(x - a) \rightarrow \infty$ is achieved smoothly via the effective variance.

Application 1) Relative offset of two electronic channels. This can be a time offset or a charge offset among channels. When each channel behavior is linear then, the solution in section 1 is applicable. But if the response of each channel is none linear, then a consistent way to obtain relative offsets is to carry out implicit function fit.

Suppose the response of each channel is of the following form: $y = b + mx + kx^2$. The implicit function of interest than is: $kx^2 + mx + lz + qz^2 + d \equiv 0$, where the relative offset is d.

Application 2) Fitting of two curves that intersect one another in xy plane. A straight forward case is when two line segments intersect. The implicit functions are: $F = l(y - b) + m(x - a) \equiv 0$ and $G = p(y - b) + q(x - a) \equiv 0$, with the following the constraints: $l^2 + m^2 = 1$ and $p^2 + q^2 = 1$.

5 Conclusion

A method to fit the parameters in implicit functions is shown to have no singularities. A straight line equation is used to demonstrate the advantage of this method. In this method each variable has a corresponding scale parameter. The standard least squares method has one less scale parameter compared to the number of variables. In the standard least squares method one of the scale parameters is set to one, where in the general method the quadrature sum of scale parameters is set to one. Since this note covers only the overall properties of an implicit function fit, the interested reader may come across features that are not mentioned in this note. It is intellectually worth the time of an individual to to use the general least squares method to explore problems and compare the results with that of the standard least squares method.

6 Inquiry

If you recall reading about such a method elsewhere, please let me know. Please send your comments to

FNALD::HK, FNAL::HK

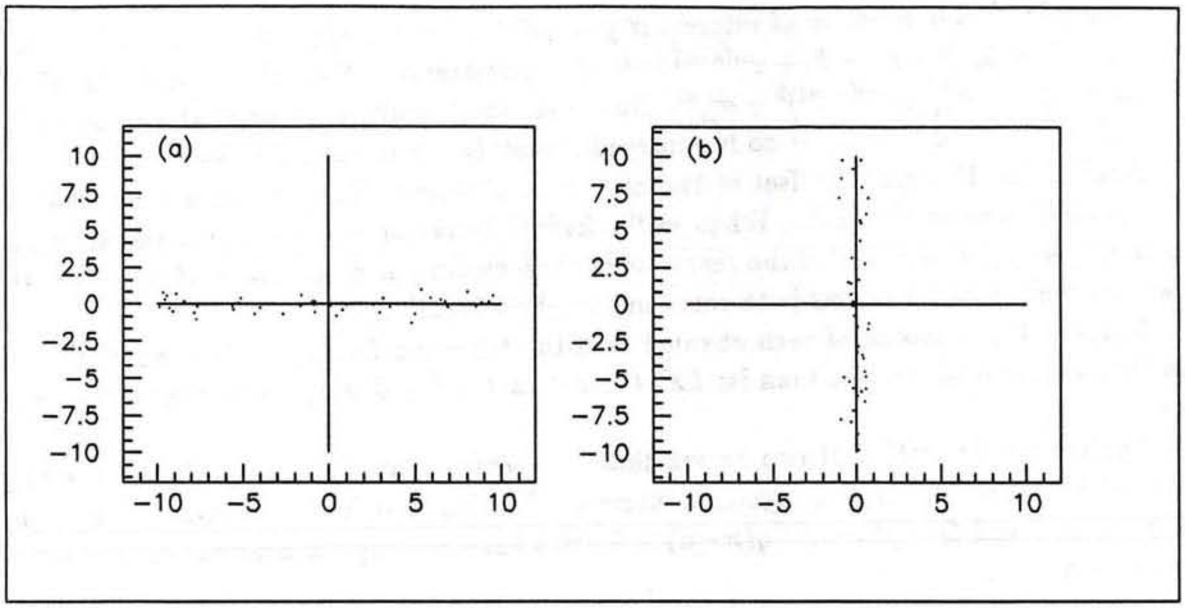

Figure 1: These scatter plots ('a' and 'b') are prepared using fifty events in each sample. The Least squares fits (using a line equation $y = b + mx$) for many such samples lead to average values of slope m=0 and intercept b=0. These values are correct for samples such as shown in plot 'a', and incorrect for distributions such as shown in plot 'b'. The General Least Squares Method does give correct fit results for the above distributions. Details of this method are given in section 2 of this note.

References

[1] A. G. Frodesen, O. Skjeggestad, H. Tøfte. **Probability and Statistics in Particle Physics.** UNIVERSITETSFORLAGET 1979.

[2] W. T. Eadie, D. Drijard, F. E. James, M. Roos, B. Sadoulet. **STATISTICAL METHODS in Experimental Physics.** North-Holland Publishing Company, 1971.

[3] P. R. Bevington, **Data Reduction and Error Analysis for the Physical Sciences.** McGraw-Hill, New York, 1969.

[4] Jay Orear. **Least Squares When Both Variables Have Uncertainties.** *American Journal of Physics.* **50**, 912(1982).

[5] F. James, M. Roos. **MINUIT, Function Minimization and Error Analysis.** CERN Computer Center, 1983.