Graphical Procedure for Fitting the Best Line to a Set of Points

J. L. DOLBY*
Lockheed Aircraft Corporation

This report provides a simple graphical procedure for obtaining the slope and intercept of the straight line of best fit to a set of points in two dimensions. The solution is obtained in the dual space (coordinate system) by use of mapping. Although this procedure is useful in itself for two dimensional problems, it may be even more useful as a teaching aid in illustrating some simple properties of mapping, dual spaces, the geometric meaning of an inverse, and the basic properties of curve fitting.

The simplest curve fitting problem in two dimensions is the fitting of a straight line, $y = a + bx$, to two points with distinct abscissas. In Figure I the two points are plotted in the xy space, and a graphical solution is obtained with a straight edge. The subscripts on a and b are to indicate that the line L connects points P_1 and P_2. While this solution is simple geometrically, it differs from the algebraic solution in that it does not use another space. This can be seen from the algebraic application as follows: After substituting the coordinates of P_1 and P_2 in the general equation of a straight line, the result is:

$$\begin{aligned} y_1 &= a + bx_1 \\ y_2 &= a + bx_2 \end{aligned} \tag{1}$$

These equations can be interpreted in a second space where the coordinates are b and a rather than x and y, as can be seen by stating (1) as:

$$\begin{aligned} a &= y_1 + (-x_1)b \\ a &= y_2 + (-x_2)b \end{aligned} \tag{2}$$

If we now plot these two lines in the ba space, the coordinates of their intersection will be the solution (b, a) of (1). *Note that we have reversed the usual choice of axes.* The reason for this will be apparent shortly.

The significance of this second representation is that mathematically it is dual to the one shown in Figure I. In particular, the two points of Figure I correspond to the two lines of Figure II and the line of Figure I corresponds to the point of intersection in Figure II. In general any point in the xy space can be represented by a line in the ba space. Because of this duality a solution to a particular problem can be obtained in whichever space is more convenient.

In the above have been given two methods for finding the straight line going

* This paper was written while the author was a member of the General Electric Laboratory, General Electric Company, Schenectady, N. Y.

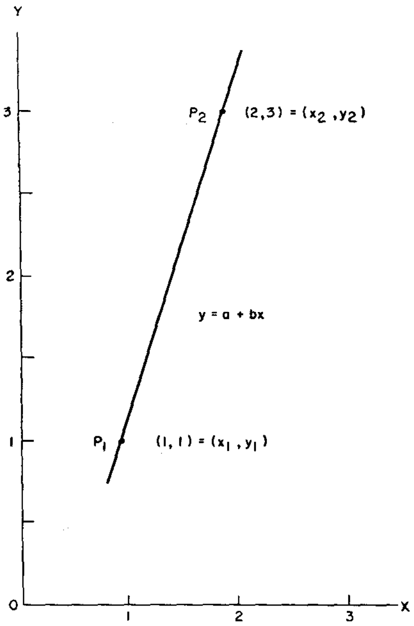

FIGURE I

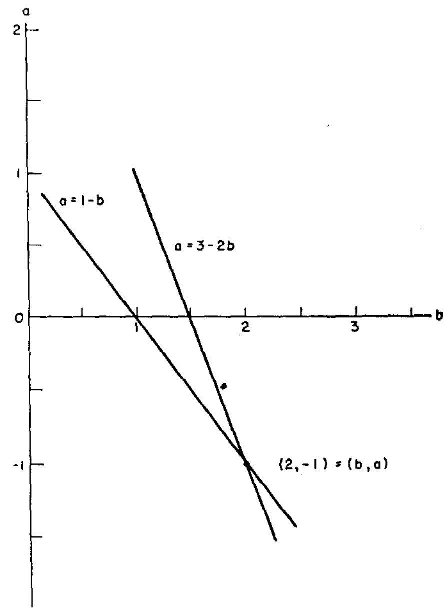

FIGURE II

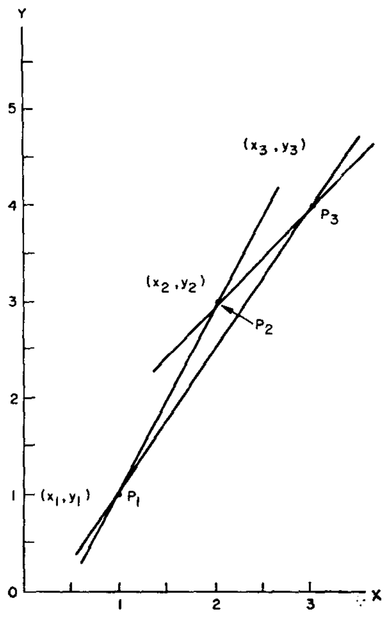

FIGURE IIIa

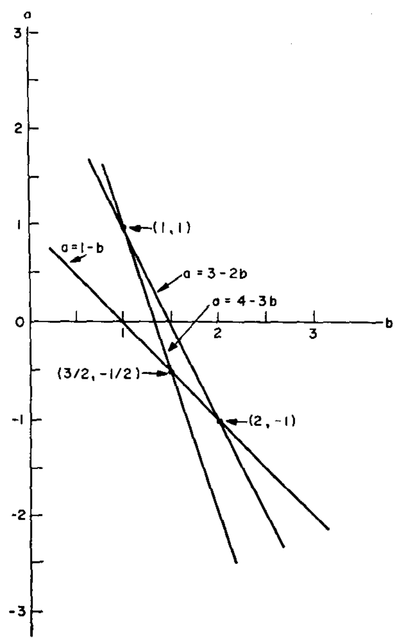

FIGURE III b

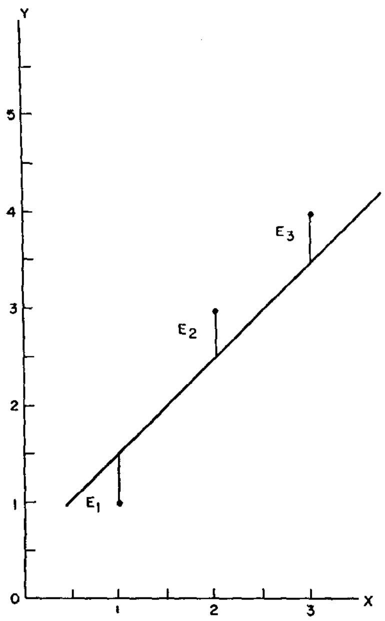

FIGURE IV a

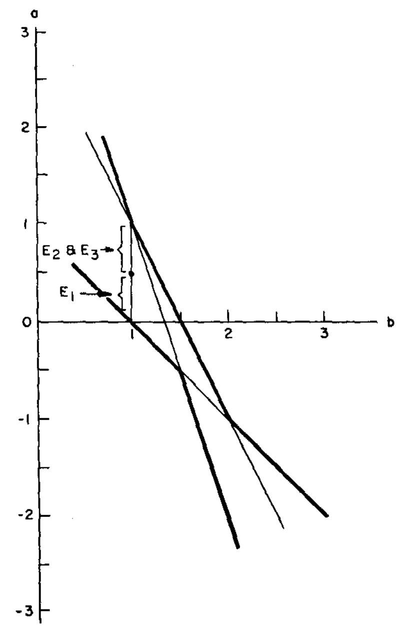

FIGURE IV b

through two points; namely, the obvious one of using a straight edge and connecting the two points in the xy space and the more difficult one of replotting the two points in the ba space as lines and reading the coordinates of their intersection. Assuming that we wish to work geometrically, it is clear that in this case we would use the xy space to obtain our solution. However, there are cases where it is more advantageous to use the ba space.

To illustrate this let us add a third point in the xy space and the corresponding third line in the ba space as in Figure III. Suppose however, that the problem is to find the particular straight line that minimizes the error (in some sense) at the three points. Before considering the various ways of measuring the error, an arbitrary line in the xy space is considered (and the corresponding point in the ba space) so that the relationship between the error measurements in the two diagrams can be studied. In the xy space the distance from a particular point to the given line (indicated by the heavy line segments in Figure IV) is given very simply by the quantity $E_i = Y_i - a - bx_i$. That this distance is identical to the corresponding vertical distance in the ba space follows immediately. To be more specific: the distance from a point to a given line measured in a vertical direction in the xy space is identical to the distance from the corresponding line to the corresponding point also measured in the vertical direction in the ba space.

This fact is now used to find the line that minimizes the maximum error at the given point. Such a line is known as the "best" line or sometimes as the Tchebycheff line. An attempt to find this line in the xy space causes certain difficulties. The error distances can be easily determined for any line. However, since they are located at different points in the diagram it is not immediately

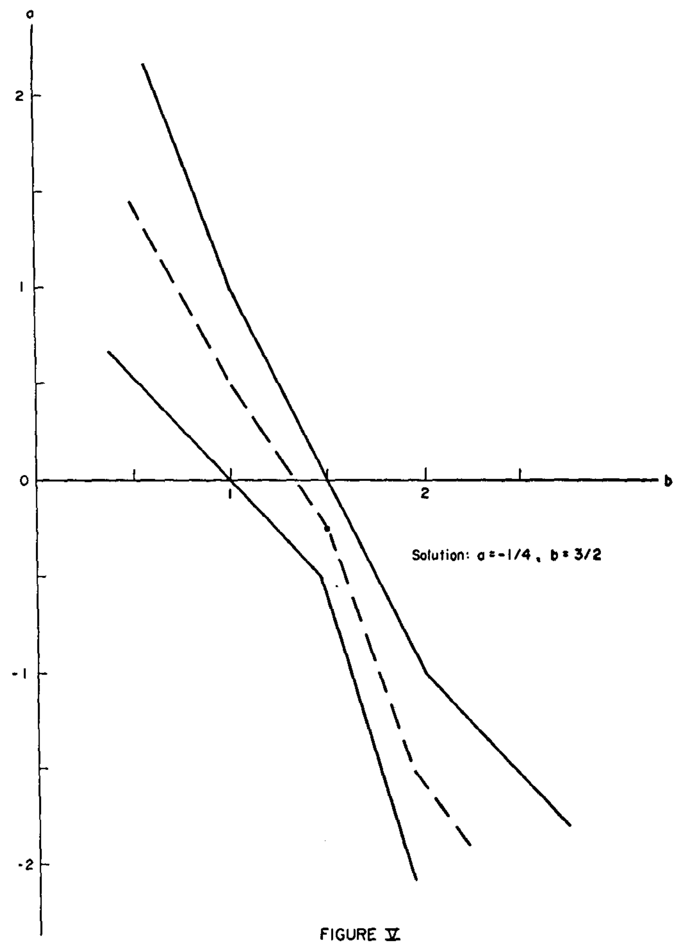

FIGURE VI

obvious how a particular trial solution should be rotated or translated to reduce the maximum error. Although trial and error solution can be obtained geometrically it is desirable to have another means of accomplishing this. The *ba* space provides this means. A demonstration is made by examining the consequences of choosing a particular slope b and then finding the value of a that minimizes the maximum error for this choice of b. In Figure IV the choice of $b = 1$ is given by the heavy vertical line. For any choice of a on this line the error is given by the distance from the point determined by this choice of a to the lines on the diagram (measured vertically). Consequently the optimum choice of a is such as to place the point midway between the extremal lines ($a = 1/2$). If a is greater than 1/2, the distance to the lower line is increased whereas, if a is less than 1/2 the distance to the higher line is increased. The internal line (that is: the line or lines that for a particular value of b do not represent the maximum or minimum value of a) is not used in such a procedure and may be eliminated from consideration (as is done in Figure VI). The solution will then lie somewhere on the bisector of the two extremal lines indicated by

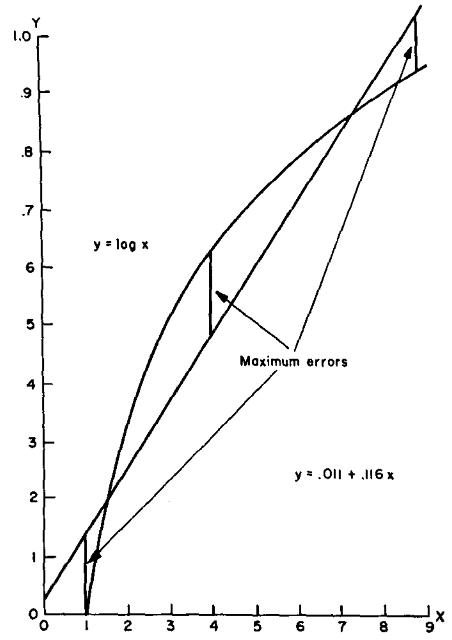

FIGURE VIa

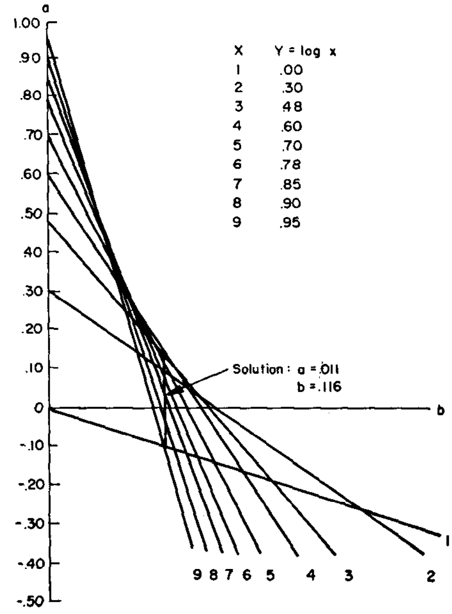

FIGURE VIb

the dashed line in Figure V. From this it follows that the solution can be found immediately by choosing the value of b for which the extremal lines are closest together ($b = 3/2$) and the value of $a (a = -1/4)$ which is midway between the extremal lines at that value of b.

As an example, let us apply this procedure to find the line of best fit (in the Tchebycheff sense) to the logrithmic curve. The original function is $y = \log x$ which provides the table of data given in Figure VIb. To these data we now fit the straight line $y = a + bx$. Our solution is found by plotting the nine straight lines $a = y + (-x)b$ shown in Figure VIb, determining the extremal boundary of these lines and we observe the solution to be $b = .116$ and $a = .011$. This line has in turn been plotted in the xy space of Figure VIa with the maximum errors again indicated heavily.

This procedure is relatively quick in that it only requires the plotting of as many lines in the ba space as there are points in the xy space. The solution is obvious on inspection. A similar solution can be derived for the straight line that minimizes the sum of the ab values of the errors though in this case the internal lines do play an important part in the proceedings. However, this example may be even more useful from the pedagalogical point of view for the light it may shed on the use of mapping in curve fitting work.