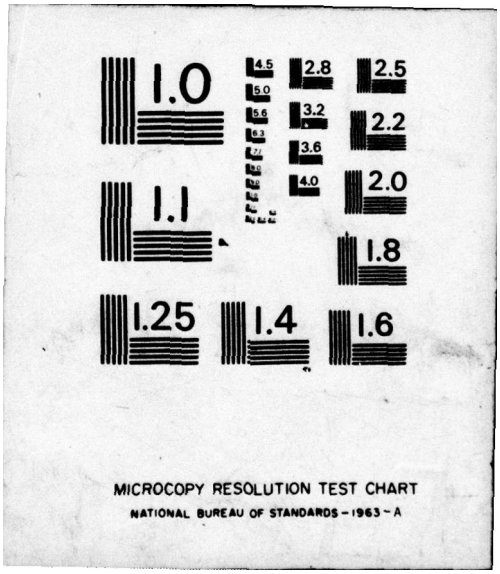
AD-A035 485 FRANK J SEILER RESEARCH LAB UNITED STATES AIR FORCE --ETC F/G 12/1
MINIMUM DISTANCE LEAST SQUARES SURFACE FITTING.(U)
DEC 76 A A FANNIN, L A KING

UNCLASSIFIED FJSRL-TR-76-0019 NL

| OF |
AD
A035485

END

DATE
FILMED
3-77


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS - 1963 - A

FRANK J. SEILER RESEARCH LABORATORY

FJSRL TECHNICAL REPORT-76-0019

DECEMBER 1976

MINIMUM DISTANCE LEAST SQUARES SURFACE FITTING

PROJECT 2303

AIR FORCE SYSTEMS COMMAND

UNITED STATES AIR FORCE

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Inquiries concerning this document should be addressed to the Frank J. Seiler Research Laboratory (FJSRL/NC), U.S. Air Force Academy, Colorado 80840. Phone - 303-472-2655.

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.

Printed in the United States of America
Available from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, Virginia 22151

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: FJSRL-TR-76-0019	2. GOVT ACCESSION NO.: ADA
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): Minimum Distance Least Squares Surface Fitting.	5. TYPE OF REPORT & PERIOD COVERED: Technical Report. April 1975-December 1976,
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Armand A. Fannin, Jr. Lowell A. King	8. CONTRACT OR GRANT NUMBER(s):
9. PERFORMING ORGANIZATION NAME AND ADDRESS: F.J. Seiler Research Laboratory (FJSRL/NC) Air Force Systems Command US Air Force Academy, CO 80840	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: New, 61102F/2303/F2/06 Old, 61102F/7903/02/06
11. CONTROLLING OFFICE NAME AND ADDRESS: F.J. Seiler Research Laboratory (FJSRL/NC) Air Force Systems Command US Air Force Academy, CO 80840	12. REPORT DATE: December 1976 13. NUMBER OF PAGES: 40
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 12 49p.	15. SECURITY CLASS. (of this report): UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

16 2303, 7903 17 F2, 02

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Data reduction
Least squares fitting

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A least squares fitting technique is described where all of the experimental variables are considered subject to error, rather than the customary single variable. The procedure is particularly valuable when the actual physical measurements contain real and estimable errors in more than one variable. The fitting procedure minimizes the closest (perpendicular) distance from the experimental point to the locus on the assumed functional relationship among the n different variables. The locus is approximated by a (n-1) dimensional surface. The technique is adaptable to any number of variables and any number of → next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

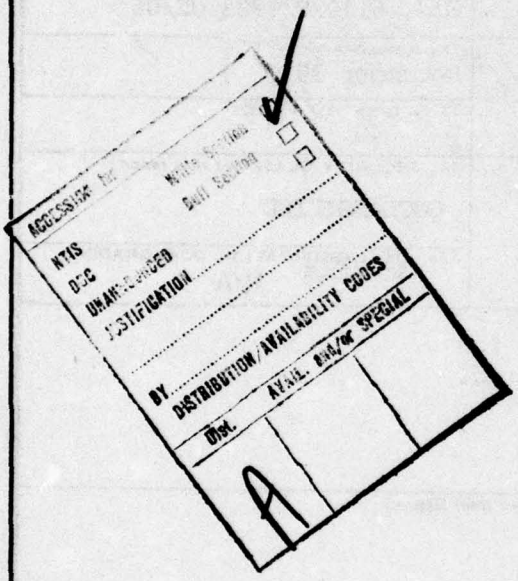
UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

319 920

#20. Abstract (continued)

adjustable constants in the assumed equation. Application of the fitting technique to vapor pressure measurements made by FJSRL and DFCBS personnel are described.



FJSRL-TR-76-0019

MINIMUM DISTANCE

LEAST SQUARES SURFACE FITTING

ARMAND A. FANNIN, JR.
Department of Chemistry and Biological Sciences

LOWELL A. KING
Frank J. Seiler Research Laboratory (AFSC)

TECHNICAL REPORT FJSRL-TR-76-0019
December 1976

Directorate of Chemical Science
Frank J. Seiler Research Laboratory
Air Force Systems Command
U.S. Air Force Academy, Colorado 80840

FOREWORD

This report was prepared by the Directorate of Chemical Sciences, Frank J. Seiler Research Laboratory, United States Air Force Academy, Colorado. The work has initiated under Project No. 2303, "Chemistry," Task No. 2303-F2, "Physical Chemistry and Electrochemistry," Work Unit No. 2303-F2-06, "Physical and Electrochemical Measurements."

The report covers work conducted from April 1975 to December 1976. The manuscript was released by the authors for publication in December 1976.

This Technical Report has been reviewed and approved.

BEN A. LOVING, LtCol, USAF
Director, Chemical Sciences

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

INTRODUCTION

The customary procedure followed in least squares fitting of experimental data to an assumed equation is to consider all of the errors to be in one of the variables, and that all other variables are exact. For example, consider a least squares fit of the equation

$$y = a + bx \tag{1}$$

Typically, x is considered to be an independent variable, not subject to error, and y is considered to be a dependent variable. If y_o are the observed values of the dependent variable, we seek to minimize with respect to a and b the quantity $\Sigma(y_o-y)^2$.

While one may indeed mathematically consider x an independent variable and not subject to error, and y to contain the error, one might equally well reverse the roles of x and y. In fact, real physical situations usually involve error in both x and y, and it may not always be stated that one variable is more suitable for use as an independent variable than the other.

It is possible to treat both x and y as subject to error in least squares fitting, but it does not often seem to be done. We describe here a technique for least squares fitting which minimizes the distance from the experimental point to the assumed equation. The technique is applicable to any number of variables subject to error and any number of adjustable constants in the assumed equation.

DISCUSSION

Consider the curve

$$y = f(x) \tag{2}$$

in Figure 1. The minimum distance from the experimental point (x,y) to the assumed equation is distance d in Figure 1(a). Providing the curvature is not too great, or expressed alternately, providing Δx and Δy are sufficiently small, the distance d may be approximated by the perpendicular distance z from point (x,y) to the chord (x,y')(x',y) in Figure 1(b). Obviously z may be smaller or larger than d, depending on the sign of the curvature, or on which side of the equation point (x,y) lies.

The least squares technique described here minimizes the quantity Σz^2.

The discussion below will treat as an example the two-dimensional case of Figure 1(b) with a set of parallel equations for the n-dimensional case. Chord (x,y')(x',y) may be considered as a one-dimensional surface. The general equation for a surface is

$$Ax + By + C = 0 \tag{3}$$

$$\sum_{i=0}^{n} a_i x_i = 0 \text{ where } x_o = 1 \tag{3'}$$

Let $(x,y)[(x_1,x_2,x_3,\ldots x_n)]$ be an experimental point, and x' and y' $[x_1',x_2',x_3',\ldots x_n']$ be the values of the variables calculated from the assumed equation and the appropriate remaining independent variables for the given experimental point. That is

$$x' = f(y \text{ and the parameters})$$
$$y' = f(x \text{ and the parameters}) \tag{4}$$

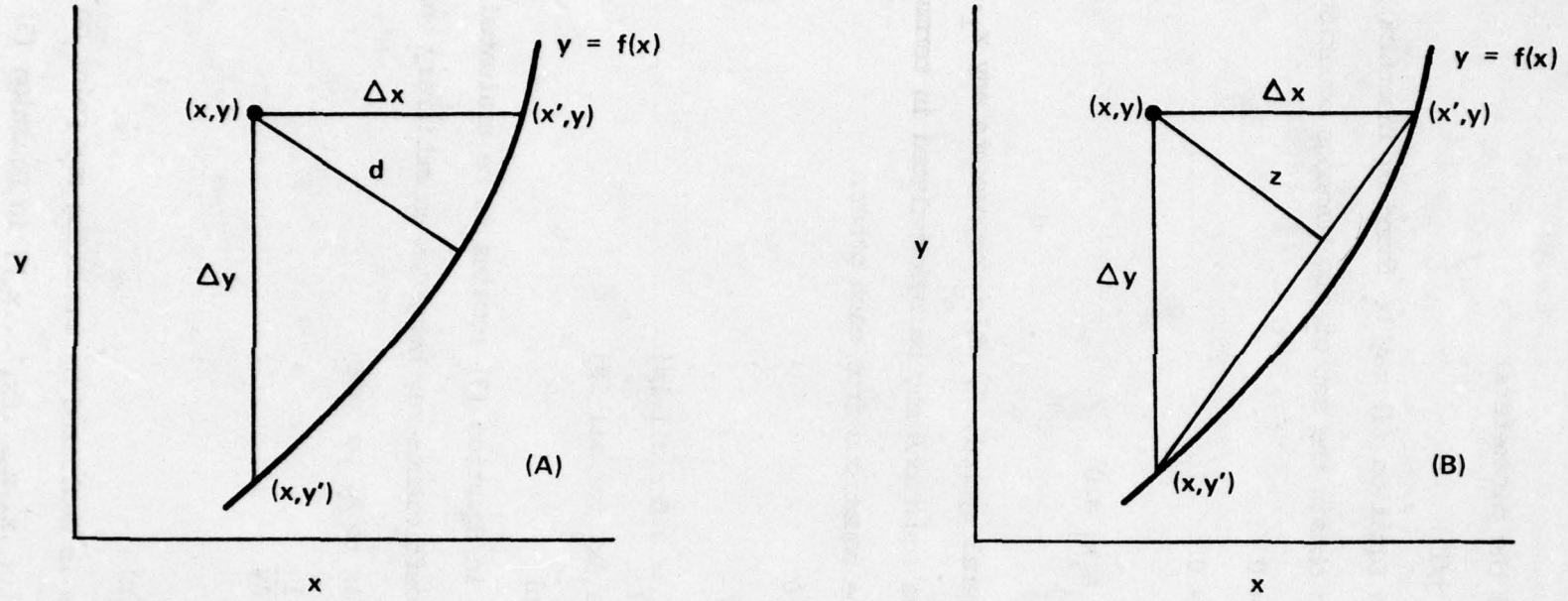


FIGURE 1. TWO DIMENSIONAL DISTANCE MINIMIZATION - I

$$x_i' = f(x_j \text{ and the parameters}) \quad (j=1,2,3,\ldots;\ j\neq i) \tag{4'}$$

The coefficients in Equation (3) may be found by inserting sufficient points on the surface to obtain the set of simultaneous equations

$$\begin{aligned} Ax' + By + C &= 0 \\ Ax + By' + C &= 0 \end{aligned} \tag{5}$$

$$a_j x_j' + \sum_{i=0,i\neq j}^{n} a_i x_i = 0 \quad (j=1,2,3,\ldots n) \tag{5'}$$

If we introduce the general symbol, ξ, which represents any x_i, and if $\Delta\xi = \xi - \xi'$ then the above equations may be reformulated in terms of the $\Delta\xi$'s by substracting the equations from each other.

$$\begin{aligned} -A\Delta x + B\Delta y &= 0 \\ \therefore\ A\Delta x &= B\Delta y \end{aligned} \tag{6}$$

$$\begin{aligned} &a_i\Delta x_i - a_j\Delta x_j = 0 \text{ for all } i\neq j \\ &\therefore\ a_i\Delta x_i = a_j\Delta x_j \text{ for all } i\neq j \end{aligned} \quad (i=1,2,3,\ldots n) \tag{6'}$$

The constant term in Equation (3) remains to be evaluated. Any one of the remaining coefficients may be given an arbitrary value. If we assign the value $1/\Delta x$ to A, we have

$$A = \frac{1}{\Delta x},\ B = \frac{1}{\Delta y} \tag{7}$$

$$a_i = \frac{1}{\Delta x_i} \quad (i=1,2,3,\ldots n) \tag{7'}$$

The constant term is evaluated by inserting any point on the surface, e.g., $(x',y)[(x_1,x_2,x_3,\ldots x_j'\ldots x_n)]$ in Equation (3)

$$\frac{x'}{\Delta x} + \frac{y}{\Delta y} + C = 0$$

$$\therefore C = -\left(\frac{x'}{\Delta x} + \frac{y}{\Delta y}\right) \tag{8}$$

$$a_o + a_j x_j' + \sum_{i=1, i\neq j}^{n} a_i x_i = 0$$

$$\therefore a_o = -a_j x_j' - \sum_{i=1, i\neq j}^{n} a_i x_i \tag{8'}$$

From analytic geometry

$$z^2 = \frac{(Ax + By + C)^2}{A^2 + B^2} \tag{9}$$

$$z^2 = \frac{\left[a_o + \sum_{i=1}^{n} a_i x_i\right]^2}{\sum_{i=1}^{n} (a_i)^2} \tag{9'}$$

Substituting Equations (7) and (8) into Equation (9), we have

$$z = \sqrt{\frac{1}{\left(\frac{1}{\Delta x}\right)^2 + \left(\frac{1}{\Delta y}\right)^2}} \tag{10}$$

$$z = \sqrt{\frac{1}{\sum_{i=1}^{n} \left(\frac{1}{\Delta x_i}\right)^2}} \tag{10'}$$

Unlike z^2, the distance z will have a sign, which must be selected when the square root is taken. Since z is a vector quantity, its magnitude, which corresponds to the positive root, is used in the fitting procedure. The procedure produces a least squares sum of these magnitudes. Signs for the corrections to the parameters arise from the partial derivatives of z with respect to the parameters and not from the sign of z itself.

The assumed functional relationship among the parameters will in general involve one or more parameters, or adjustable constants, the

values of which we must find to minimize the value Σz^2. Since we assume the true equation is very closely approximated by the surface of Equation (3), we may write for each of the parameters

$$\frac{\partial z}{\partial \theta_k} = \frac{\partial z}{\partial \Delta x}\left[-\frac{\partial x'(\theta_k)}{\partial \theta_k}\right] + \frac{\partial z}{\partial \Delta y}\left[-\frac{\partial y'(\theta_k)}{\partial \theta_k}\right] \tag{11}$$

$$\frac{\partial z}{\partial \theta_k} = \sum_{i=1}^{n}\left(\frac{\partial z}{\partial \Delta x_i}\left[-\frac{\partial x_i'(\theta_k)}{\partial \theta_k}\right]\right) \tag{11'}$$

Equation (11) simplifies to

$$\frac{\partial z}{\partial \theta_k} = -z^3\left[\frac{1}{(\Delta x)^3}\frac{\partial x'(\theta_k)}{\partial \theta_k} + \frac{1}{(\Delta y)^3}\frac{\partial y'(\theta_k)}{\partial \theta_k}\right] \tag{12}$$

$$\frac{\partial z}{\partial \theta_k} = -z^3 \sum_{i=1}^{n}\left[\frac{1}{(\Delta x_i)^3}\frac{\partial x_i'(\theta_k)}{\partial \theta_k}\right] \tag{12'}$$

The function to be minimized, given by Equation (10), and the derivatives with respect to each of the parameters, given by Equation (12), must be calculated for each experimental point, and are the quantities actually used in the least squares fitting routines. The latter are given in the next section "Mathematical Procedure."

It is better not to use the set of variables, $\{x_i\}$, in the actual units with which they were measured, for the choice of units will affect the final values of the parameters, $\{\theta_k\}$. Dimensionless variables - independent of the choice of units - may be obtained by introducing the variables in terms of the estimated uncertainty or experimental error in that variable. For example, if a particular variable, ξ, may be determined within an uncertainty of $\delta\xi$, the dimensionless variable, $\xi^* = \xi/\delta\xi$, is used in Equations (3) through (12). Equations (10) and (12) become, respectively

$$z^* = \sqrt{\frac{1}{\sum_{i=1}^{n} \left(\frac{\delta x_i}{\Delta x_i}\right)^2}} \tag{13}$$

$$\frac{\partial z^*}{\partial \theta_k} = -z^{*3} \sum_{i=1}^{n} \left[\frac{(\delta x_i)^2}{(\Delta x_i)^3} \frac{\partial x_i'(\theta_k)}{\partial \theta_k}\right] \tag{14}$$

Note that the fitted values of $\{\theta_k\}$ will still be in same units as would have been obtained from using Equations (10) and (12).

Inspection of Equations (13) and (14) reveals that if all δx_i but one are made extremely small, the minimum distance fit reduces to a conventional least squares fit with all of the error cast into the remaining variable. A useful application of this property would be when the experimentor was not sure of the data measurement errors in one or more of the variables, and wished to selectively include or exclude individual variables from fits.

The quantities $\{\Delta x_i\} = \{x_i - x_i'\}$ were individually calculated by assuming all of the error was cast into that particular variable. Clearly the actual error in the variables $\{x_i\}$ is less than $\{\Delta x_i\}$. A new set of quantities, $\{\Delta' x_i\}$, is defined with the aid of Figure 2(a). The assumption is again made that the true locus of the equation lies negligibly far away from the surface, i.e., d=z.

From analytic geometry

$$\Delta' x_i = \frac{z^2}{\Delta x_i} \tag{15}$$

$$\Delta' x_i^* = \frac{z^{*2}}{\Delta x_i^*} \tag{16}$$

Where $\Delta' x_i^* = \Delta' x_i / \delta x_i$. If the original error estimate, δx_i, was correct, then ideally $\Delta' x_i$ and δx_i should be nearly the same size.

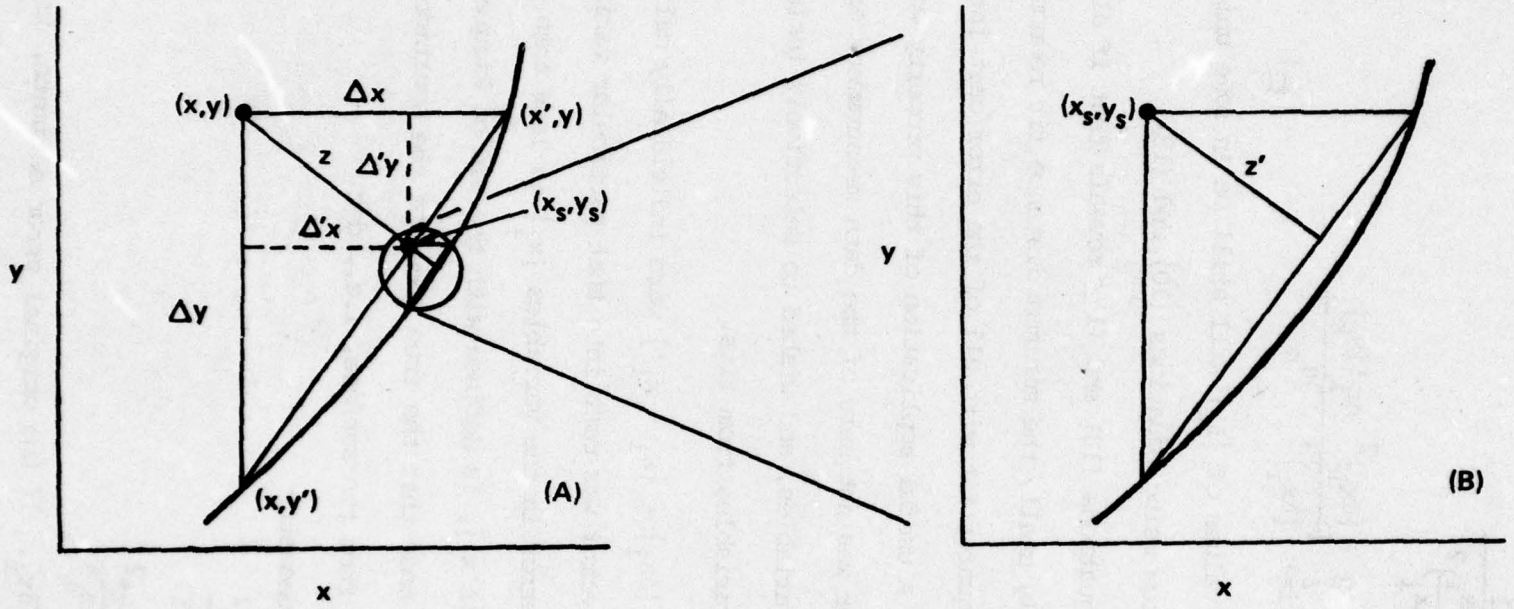


FIGURE 2. TWO DIMENSIONAL DISTANCE MINIMIZATION - II

After the point (x_s, y_s) $[x_{1s}, x_{2s}, x_{3s}, \ldots x_{ns}]$ has been identified it is possible to determine the validity of the assumptions that the surface and the true locus were very close. It may be seen in Figure 2(b) that d-z = z', and z' may be obtained by application of Equation (10) to the point (x_s, y_s) $[x_{1s}, x_{2s}, x_{3s}, \ldots x_{ns}]$.

MATHEMATICAL PROCEDURE

Physical measurements generally involve the simultaneous measurement of a number (n) of related physical properties. If the values of these properties are arranged in ordered sets of n values, vectors or state points $\{x_n\}$ are obtained. If a functional relationship exists among these n variables, containing in addition to the variables a number (p) of constants characteristic of the physical system, then a functional surface, f, may be defined such that f is identically zero.

$$0 = f(\{x_n\};\{\theta_p\}) \tag{17}$$

This ordered set of constants $\{\theta_p\}$, called parameters, which characterizes the system also forms a vector. For instance, given the typical vapor pressure equation, $0 = \log (P) - \left(\frac{A}{T} + B\right)$, the ordered set of constants (A,B) are sufficient to define the pressure, P, at any temperature, T, or the temperature at any pressure. However, each different physical system obeying the same relationship will have a different set of parameters (A,B).

In principle the evaluation of these parameters for a system should be simple. If there are p parameters, then p state points are measured producing p equations in as many unknowns which may be solved for the parameters. Two problems immediately present themselves when an attempt is made to utilize this method. First, if the functional relationship is not linear in the parameters $\{\theta_p\}$, then solution of the equations may be ambiguous, difficult, or impossible. Second, since no measurement is exact but generally yields only a few significant figures, the function f will not yield zero, but some small non-zero value, d ($\approx$z), when the values of any measured state point are inserted in the function.

The initial difficulty of non-linearity may be overcome by linearizing the equations for $\{\theta_p\}$. (To avoid confusion later, vectors and matrices will be indicated henceforth by the use of boldface instead of the subscript-in-braces notation used earlier.) This is done by expanding the equations for $\boldsymbol{\Theta}$ in a Taylor series and retaining terms in $\Delta\theta_i$ to only first order. Thus for each state vector an equation exists of the form

$$\frac{\partial z}{\partial \theta_1}\left(-\Delta\theta_1\right)+\frac{\partial z}{\partial \theta_2}\left(-\Delta\theta_2\right)+\cdots+\frac{\partial z}{\partial \theta}\left(-\Delta\theta_p\right)=z_\theta-z_{\theta+\Delta\theta} \tag{18}$$

Since $z_\theta = f = 0$, the right hand side of Equation (18) reduces to

$$z_\theta - z_{\theta+\Delta\theta} = -z_{\theta+\Delta\theta}$$

When reduced to matrix form these equations become

$$\left(\frac{\partial z}{\partial \theta_1}\ \frac{\partial z}{\partial \theta_2}\ \cdots\ \frac{\partial z}{\partial \theta_p}\right)\begin{bmatrix}-\Delta\theta_1\\-\Delta\theta_2\\\vdots\\-\Delta\theta_p\end{bmatrix}=\left(-z_{\theta+\Delta\theta}\right) \tag{19}$$

or defining $\left\{\frac{\partial z}{\partial \theta_p}\right\} \equiv \boldsymbol{P}'$ and $\{-\Delta\theta_p\} \equiv \boldsymbol{\Delta}$

$$\boldsymbol{P}'\boldsymbol{\Delta} = z_{\theta+\Delta\theta} \tag{20}$$

Using such an expansion greatly simplifies the solution of the simultaneous equations, but introduces limitations on the applicability in the following manner. Since $\boldsymbol{\Theta}$ is solved for by estimating errors in $\boldsymbol{\Theta}$ using a linear function, the solution depends on the initial estimate of $\boldsymbol{\Theta}$ and the curvature of the error surface at this point. If the function is linear in $\boldsymbol{\Theta}$ there is no curvature and any initial $\boldsymbol{\Theta}$ will converge to the correct solution. In regions where the function is linear or nearly linear, then $\boldsymbol{\Theta}$ can be solved for in one or more

iterations. Where the function deviates significantly from linearity in Θ the solutions may occur in regions where they are physically unrealistic (convergence to an incorrect solution) or the linear approximation may be sufficiently in error to move further and further from any solution with each iteration (divergence). For a well behaved and at least piece-wise continuous function then a solution generally exists. For such functions the degree of deviation from linearity in Θ in the region of the solution will determine how close to the solution the initial estimate of Θ must be for convergence to the correct solution.

The second difficulty, that associated with the error of measurement is treated in the following manner.

Measurements are made of many more than p state vectors. Since, in general, no two state vectors will give the same value for z, this provides a set of k equations (k>p) which are inconsistent. The "degree" of inconsistency however is small, that is there will be a distribution of z's about zero which is assumed to be Gaussian or normal and is a function of the errors in the measurements of the state variables. The set of "overdetermined" parameters (i.e., those defined by all k equations instead of only the minimum subset p of them) are solved for by establishing a criterion for the "best" parameters. Logically the best set of parameters would be those predicted to be most probable from the Gaussian distribution of the errors, z. This would be difficult to do using functions of the distribution itself and such a procedure turns out to be unnecessary. It has been shown that the "least squares requirement" gives the most probable solution for the parameters (1). This "least squares requirement" therefore is used commonly as the criterion for the

best fit. The least squares requirement is very simple in concept. If the error is defined in terms of the distance of the state vectors from the functional surface, the error surface, T, is a function of Θ defined by the sum of the squares of all individual errors, z. This sum of the squared errors, Σz^2, is minimized by adjustment of the parameters Θ.

The adjustment is accomplished by using standard techniques from the calculus and linear algebra. The unconstrained extremum for any well-behaved function which is linear in several variables, is found by setting the partial derivatives of the function with respect to each of the variables equal to zero and solving the system of equations obtained. These are called the normal equations. The error surface is defined to be

$$T = \Sigma z^2 \tag{21}$$

The set of equations obtained by setting the partial derivatives of T with respect to each of the parameters equal to zero are the normal equations for the system. That is

$$\left(\frac{\partial T}{\partial \theta_1}\right) = 0, \left(\frac{\partial T}{\partial \theta_2}\right) = 0, \cdots \left(\frac{\partial T}{\partial \theta_p}\right) = 0 \tag{22}$$

It can be shown in the individual cases of practical interest that the extremum so obtained is a minimum. This can be seen from the fact that for any functional surface a maximum distance from the surface to the fixed set of data points does not exist, i.e., the surface can always be moved further away from all the points.

Margulis (2) also points out that the "least squares criterion" satisfies the following intuitively desirable requirements for a "best fit":

(1) the error is expressed as an unsigned value including ALL measured vector errors,

(2) the overall error is zero only if ALL errors are zero,

(3) the overall error increases if any ONE error increases.

It is necessary to build the system of normal equations and solve them. The system could be constructed by evaluation of the appropriate derivatives of T, but a simpler method exists which has been shown to be equivalent. This latter method is based on the construction of the correction Equation (18). For each measured state vector x_j, one equation like Equation (20) is obtained.

$$P_j' \Delta = z_j \tag{23}$$

The vector Δ has no subscript since a common, unique set of corrections for all the state vectors is assumed to exist. Since each vector P_j' is a row of ordered values, the set of equations for k state vectors will form a matrix with k rows (one for each state vector) and p columns (one for each parameter).

The column vector Δ is to be evaluated and added to ${}^{n}\Theta$ to obtain an improved ${}^{n+1}\Theta$. The column vector Z formed by the k values of the z_j is used in the solution. When all of the elements are collected, the set of correction equations appear as:

$$\begin{pmatrix} \frac{\partial z_1}{\partial \theta_1} & \frac{\partial z_1}{\partial \theta_2} \cdots & \frac{\partial z_1}{\partial \theta_p} \\ \frac{\partial z_2}{\partial \theta_1} & \frac{\partial z_2}{\partial \theta_2} \cdots & \frac{\partial z_2}{\partial \theta_p} \\ \vdots & \vdots & \vdots \\ \frac{\partial z_k}{\partial \theta_1} & \frac{\partial z_k}{\partial \theta_2} \cdots & \frac{\partial z_k}{\partial \theta_p} \end{pmatrix} \times \begin{pmatrix} \Delta\theta_1 \\ \Delta\theta_2 \\ \vdots \\ \Delta\theta_p \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_k \end{pmatrix} \tag{24}$$

or

$$P\Delta = Z \tag{25}$$

Premultiplying this equation on both sides by P^T the transpose of matrix P gives a matrix equation that is identical to the normal equation system.

$$P^T P\Delta = P^T Z \tag{26}$$

If the products $P^T P$ and $P^T Z$ are assigned the symbols H and B respectively then the matrix equation

$$H\Delta = B \tag{27}$$

is the one that must be solved. Inversion of the square matrix H to give H^{-1} provides this solution. Premultiplying both sides of Equation (27) by H^{-1} obtains

$$H^{-1}H\Delta = H^{-1}B \tag{28}$$

by definition $H^{-1}H\Delta = E\Delta = \Delta$

where E is the unit matrix and therefore

$$\Delta = H^{-1}B \tag{29}$$

It would appear at this point that the simplification introduced by using this method is negligible, but the construction of the matrices H and B is particularly simple.

Initially a null matrix (all elements zero) for both H and B is constructed. H is a p x p matrix (i.e., p rows and p columns) and B is a p x 1 matrix.

For each state point for 1 to k the vector P' described earlier is evaluated. The elements of this vector are the partials of z with respect to each of the parameters θ_i (i = 1 to p). z is also evaluated.

The elements of H are then augmented by adding to each element h_{ij} the product $p_i' \times p_j'$. Similarly the product $z \times p_i$ is added to b_i. This process is identical to the following matrix operation.

$$H_{(\text{augmented})} = H_{(\text{unaugmented})} + (P')^T (P') \tag{30}$$

and

$$B_{(\text{augmented})} = B_{(\text{unaugmented})} + (z)(P') \tag{31}$$

Once these matrices have been constructed, the inversion of H may be carried out using any standard procedure.

The mathematical discussion in this section has been phrased in terms of the variables $\{x_i\}$ and the minimum distance, z; however, the procedure is equally valid for the use of reduced variables $\{x_i^*\}$ and z^*.

APPLICATION AND EVALUATION

We first applied the technique described earlier to a set of temperature-vapor pressure data collected over a sample of solid aluminum chloride (3). The data are presented in Table I.

We assumed vapor pressure of aluminum chloride could be represented by an equation of the form

$$\log p = \frac{A}{T} + B \tag{34}$$

where the variables p and T represent pressure in Torr and temperature in Kelvin, respectively; and A and B are two parameters whose values are sought.

We solved for values of A and B which minimized the sum

$$\sum_{i=1}^{n} z_i^{*2} \tag{35}$$

where z^*, $\partial z^*/\partial A$ and $\partial z^*/\partial B$ may be found from Equations (13) and (14).

In the present case

$$\begin{aligned} \Delta p_i &= p_i - \exp[(A/T_i + B)\ln(10)] \\ \Delta T_i &= T_i - A/[\log(p_i) - B] \end{aligned} \tag{36}$$

and

$$\begin{aligned} \partial p'/\partial A &= p' \ln(10)/T \\ \partial T'/\partial A &= 1/[\log(p) - B] \\ \partial p'/\partial B &= p' \ln(10) \\ \partial T'/\partial B &= A/[\log(p) - B]^2 \end{aligned} \tag{37}$$

The distance function, z^*, and its derivatives were calculated for each point by the subroutine beginning at step 6000 of the program given in the next section. A careful examination and evaluation of our particular experimental technique and apparatus led to the selection of

TABLE I. VAPOR PRESSURE OF SOLID ALUMINUM CHLORIDE

Pressure, torr	Temperature, °C	
3.77	113.66	M[a]
3.80	113.66	C
3.92	114.37	M
5.52	116.70	C
5.53	116.70	M
6.30	118.50	M
6.32	118.50	C
7.14	119.37	M
7.25	119.37	C
9.36	122.87	M
10.80	125.61	M
10.91	124.23	M
10.94	124.23	C
15.09	127.97	C
16.26	128.45	M
16.36	128.45	M
16.47	128.63	C
19.03	130.92	M
33.81	137.46	C
41.13	140.42	C
58.15	144.53	C
88.42	149.66	C
103.78	152.49	C
106.05	152.44	C
168.02	158.28	C
230.13	163.40	C
230.63	163.40	C
235.66	163.40	C
376.17	169.90	C
573.20	176.80	C
574.97	176.80	C
696.79	179.66	C
716.93	179.49	C
808.52	181.84	C
934.15	184.00	C
943.10	184.08	C
952.31	184.08	C
1082.23	186.25	C
1157.09	187.27	C
1203.41	187.87	C
1344.23	189.49	C
1428.89	190.49	C
1461.56	190.82	C
1522.82	191.44	C
1567.17	191.95	C
1641.14	192.63	C
1660.37	192.79	C
1705.21	193.17	C

[a] M and C refer to the use of a McLeod Gauge or a capacitame manometer, respectively, as the pressure measuring device.

$\delta T = 0.1°K$ and $\delta p = 0.0002\ p$ or 0.1 torr, whichever was greater for each point.

The results of the minimum distance least squares fit are shown in Table II. The fit was made on a Wang 2200 calculator in BASIC language. The program is listed in the next section. Conventional least squares fits also were made of these data for comparison, where first temperature, then pressure was made to be the independent variable. The results of all three least squares fits are summarized in Table III.

While one may not a priori state which set of A and B is "best" for these data, several interesting observations can be made. The value of dP/dT changes from 0.34 to 107 over the range of the data. This means that a small error in one variable propagates a relatively huge error in the other variable at one end of the data set. Therefore, by assuming either one of the variables to be exact, the least squares fit will reproduce the data relatively well at one extreme of the data range and relatively poorly at the other extreme. Furthermore, it may be seen that the data points are on the average only 0.18 torr and 0.24K from the line obtained from the minimum distance fit. If temperature were assumed exact (probably the conventional assumption in least squares fitting of data to Equation 34), pressures are on the average 5.5 torr in error.

In order to examine the question of which line is actually "best," we fit a set of hypothetical p,T data obtained as follows: a "true" equation was assumed for which A = -6000 and B = 16. Temperature was stepped at equal increments from 104°C to 198°C to yield 48 points, and the corresponding pressures calculated from the equation. These "true" pressures and temperatures were then randomized by $p = p + 2\delta p(0.5-R)$ and $T = T + (2)(0.1)(0.5-R)$, where δp is $0.0002p$ or 0.1, whichever is greater, and R is a random number for each datum such that $0<R<1$.

TABLE II. MINIMUM DISTANCE LEAST SQUARES FIT FOR ALUMINUM CHLORIDE

VAPOR PRESSURE OF SOLID ALUMINUM CHLORIDE

a

A	-5891.065581615	+/- 12.16813381809
B	15.8615816687	+/- 2.75345336E-02

b ITN= 5 SIGMA= 2.989E+00 DELTA SIGMA= 6.689E-11

c

NR	VARIABLE NAME	UNITS	METRIC
1	PRESSURE	TORR	COMPUTED
2	TEMP	K	1.000E-01

d

VARIABLE	DELTA	DELTA PRIME	ID/Z*
3.77	-5.126E-01	-4.521E-01	M
386.81	1.401E+00	1.654E-01	4.814E+00
3.8	-4.826E-01	-4.253E-01	C
386.81	1.314E+00	1.561E-01	4.530E+00
3.92	-6.468E-01	-5.636E-01	M
387.52	1.683E+00	2.165E-01	6.038E+00
5.52	-1.095E-01	-8.778E-02	C
389.85	2.200E-01	4.370E-02	9.806E-01
5.53	-9.954E-02	-7.974E-02	M
389.85	1.997E-01	3.973E-02	8.909E-01
6.3	-3.057E-01	-2.305E-01	M
391.65	5.351E-01	1.317E-01	2.654E+00
6.32	-2.857E-01	-2.152E-01	C
391.65	4.994E-01	1.231E-01	2.480E+00
7.14	7.165E-03	5.137E-03	M
392.52	-1.140E-02	-3.227E-03	6.067E-02
7.25	1.171E-01	8.366E-02	C
392.52	-1.851E-01	-5.294E-02	9.900E-01
9.36	-3.206E-01	-1.909E-01	M
396.02	3.890E-01	1.573E-01	2.474E+00
10.8	-1.449E+00	-7.353E-01	M
398.76	1.470E+00	7.246E-01	1.032E+01

TABLE II. Continued

10.91	2.532E-02	1.349E-02	M
397.38	-2.705E-02	-1.263E-02	1.848E-01
10.94	5.532E-02	2.945E-02	C
397.38	-5.902E-02	-2.760E-02	4.036E-01
15.09	1.262E-01	4.848E-02	C
401.12	-9.969E-02	-6.140E-02	7.824E-01
16.26	6.790E-01	2.435E-01	M
401.6	-5.078E-01	-3.256E-01	4.067E+00
16.36	7.790E-01	2.784E-01	M
401.6	-5.809E-01	-3.733E-01	4.657E+00
16.47	6.514E-01	2.297E-01	C
401.78	-4.808E-01	-3.112E-01	3.868E+00
19.03	-1.241E-01	-3.529E-02	M
404.07	7.823E-02	5.599E-02	6.618E-01
33.81	1.115E+00	1.370E-01	C
410.61	-4.174E-01	-3.661E-01	3.909E+00
41.13	-2.851E-01	-2.433E-02	C
413.57	8.711E-02	7.967E-02	8.331E-01
58.15	9.592E-01	4.549E-02	C
417.68	-2.140E-01	-2.038E-01	2.088E+00
88.42	3.608E+00	8.188E-02	C
422.81	-5.498E-01	-5.373E-01	5.435E+00
103.78	-1.196E+00	-1.926E-02	C
425.64	1.530E-01	1.506E-01	1.518E+00
106.05	1.465E+00	2.320E-02	C
425.59	-1.859E-01	-1.829E-01	1.844E+00
168.02	7.015E+00	4.861E-02	C
431.43	-5.860E-01	-5.819E-01	5.839E+00
230.13	-2.671E+00	-9.800E-03	C
436.55	1.621E-01	1.615E-01	1.618E+00
230.63	-2.171E+00	-7.950E-03	C
436.55	1.316E-01	1.311E-01	1.314E+00
235.66	2.858E+00	1.025E-02	C
436.55	-1.715E-01	-1.708E-01	1.711E+00
376.17	8.916E+00	1.351E-02	C
443.05	-3.474E-01	-3.468E-01	3.471E+00

TABLE II. Continued

573.2	-1.411E+01	-1.224E-02	C
449.95	3.627E-01	3.624E-01	3.625E+00
574.97	-1.234E+01	-1.074E-02	C
449.95	3.168E-01	3.165E-01	3.166E+00
696.79	-1.371E+01	-1.226E-02	C
452.81	2.943E-01	2.941E-01	2.942E+00
716.93	1.437E+01	1.339E-02	C
452.64	-3.061E-01	-3.058E-01	3.060E+00
808.52	-1.164E+01	-1.067E-02	C
454.99	2.180E-01	2.178E-01	2.179E+00
934.15	-1.007E+01	-9.451E-03	C
457.15	1.652E-01	1.651E-01	1.651E+00
943.1	-6.043E+00	-5.697E-03	C
457.23	9.842E-02	9.832E-02	9.837E-01
952.31	3.166E+00	3.016E-03	C
457.23	-5.134E-02	-5.129E-02	5.131E-01
1082.23	-9.690E+00	-9.285E-03	C
459.4	1.386E-01	1.385E-01	1.385E+00
1157.09	-8.644E+00	-8.370E-03	C
460.42	1.162E-01	1.161E-01	1.162E+00
1203.41	-7.890E+00	-7.687E-03	C
461.02	1.023E-01	1.022E-01	1.023E+00
1344.23	1.474E+00	1.468E-03	C
462.64	-1.731E-02	-1.729E-02	1.730E-01
1428.89	-1.522E+00	-1.526E-03	C
463.64	1.688E-02	1.686E-02	1.687E-01
1461.56	1.069E+00	1.077E-03	C
463.97	-1.161E-02	-1.160E-02	1.161E-01
1522.82	4.221E+00	4.283E-03	C
464.59	-4.417E-02	-4.412E-02	4.414E-01
1567.17	-8.346E-01	-8.477E-04	C
465.1	8.490E-03	8.482E-03	8.486E-02
1641.14	4.930E+00	5.055E-03	C
465.78	-4.812E-02	-4.807E-02	4.809E-01
1660.37	7.715E+00	7.936E-03	C
465.94	-7.455E-02	-7.447E-02	7.451E-01

TABLE II. Continued

1705.21	1.287E+01	1.333E-02 C
466.32	-1.215E-01	-1.214E-01 1.215E+00

	NR	VARIABLE NAME	RMS DELTA	RMS DELTA PRIME	UNITS
e	1	PRESSURE	5.963E+00	1.841E-01	TORR
	2	TEMP	5.008E-01	2.355E-01	K
	48	POINTS			

[a] A and B are the parameters sought in Equation (34). The "+/-" values are the estimated errors in A and B, and are calculated from the appropriate diagonal element of the inverted H matrix, h_{kk}^{-1}, using the formula σ_{θ_k} = sigma x $\sqrt{ABS(h_{kk}^{-1})}$ (4).

[b] Five iterations were required. At each iteration the value of the standard deviation, sigma was calculated. Sigma = $(\Sigma(z^{*2})/N)^{1/2}$. Iterations were continued until the change in sigma was smaller than some arbitrary value. The convergence criterion in the present case was Δ sigma $\leq 1 \times 10^{-7}$.

[c] The "metric" in pressure, i.e., δp, was computed for each point by the subroutine at line 5000 in the program.

[d] The pairs of numbers are on the first line: p, Δp, $\Delta' p$, device code; on the second line: T, ΔT, $\Delta' T$, and z*.

[e] Root mean square values are given for Δp, $\Delta' p$, ΔT, and $\Delta' T$.

TABLE III. COMPARISON OF LEAST SQUARES FITS FOR SOLID ALUMINUM CHLORIDE

	Minimum Distance Fit	Temperature Exact	Pressure Exact
A	-5891±12	-5946±18	-5923±14
B	15.9±2.8	16.0±4.0	15.9±3.2
RMS Δp, torr	6.0	5.5	6.7
RMS $\Delta' p$, torr	0.18	5.5	-
RMS ΔT, K	0.50	0.51	0.47
RMS $\Delta' T$, K	0.24	-	0.47

The same three types of least squares fits were made on this set of hypothetical data, and the results are summarized in Table IV. As before, it is clear that the pressures and temperatures are much closer to the minimum distance fit line than to either of the other lines. It is also obvious that the exact temperature fit reproduced the "true" equation much more poorly than did the other two fits, yet this is probably just the kind of least squares fit that most investigators would choose to make if they were to select one independent variable and one dependent variable from Equation (34). The apparent agreement between the minimum distance fit and exact pressure fit values of A and B is not surprising for the random errors in pressure propagate errors in temperature smaller than the actual δT over most of the data range. Another way of expressing this is to say that pressure behaves more nearly like the independent variable and temperature the dependent one.

The fact that all three fits of hypothetical data show smaller RMS $\Delta' x_i$ and RMS $\Delta' x_i$ than do the corresponding fits of real data merits some comments. The two principal reasons are, first, Equation (34) assumes a constant enthalpy of sublimation for aluminum chloride and ideal gas behavior of the vapor. Neither assumption is entirely valid. The second reason is that the random errors introduced into the hypothetical data were constrained to not exceed δp and δT, respectively. The errors in the real data were not so constrained to not exceed these "maximum expected" error limits.

These two sets of least squares fits strongly suggest that the minimum distance criterion is valid and desirable, and that the technique described here yields results at least as good as do conventional techniques, and may do considerably better than conventional techniques

TABLE IV. COMPARISON OF LEAST SQUARES FITS FOR HYPOTHETICAL p,T DATA SET

"True" equation: $\log p = \frac{-6000}{T} + 16$

	Minimum Distance Fit	Temperature Exact	Pressure Exact
A	-5999.5±2.3	-5979.2±5.2	-6002.9±3.8
B	16.00±5.3	15.95±1.1	16.01±8.8
RMS Δp, torr	2.0	1.6	2.4
RMS $\Delta' p$, torr	0.02	1.6	-
RMS ΔT, K	0.12	0.19	0.12
RMS $\Delta' T$, K	0.04	-	0.12

if an unwise choice of which variable is the dependent one is made in the latter case. When one considers fitting more than two variables, the minimum distance technique becomes even more attractive as it becomes increasingly difficult to single out a lone dependent variable for conventional fitting.

BASIC LANGUAGE IMPLEMENTATION OF FITTING PROCEDURE

The minimum distance fitting technique described earlier has been implemented in an extended BASIC language on a Wang 2200S calculator system. The calculator had the general I/O and advanced programming options and 8K bytes of memory. Output was produced with a model 2202 plotting typewriter.

Program Concept

The program was designed to fit functions with up to eight variables and one to eight parameters. The program performs the following functions:

1. Determine the number of variables, their names, the units in which they were measured, and their uncertainty in measurement.

2. Produce on request a cassette tape containing the data in a highly packed format retaining seven decimal digit significance.

3. Ascertain the number of parameters to be determined, their names, and the fractional change in sigma necessary to terminate the fitting.

4. Generate the necessary matrices for fitting from user written routines for the calculated values of the variables $\{x'\}$ and partial derivatives $\{\frac{\partial x'}{\partial \theta}\}$. Reduction to unit independent quantities is accomplished using the uncertainties in measurement, or "metric," or, if specified, the computed uncertainty in measurement of each variable at each point as supplied by a user written routine.

5. Invert the necessary matrix and solve for improved values of the parameters, calculate the sigma and fractional change in sigma. Thirteen decimal digits significance is retained during these calculations.

6. Display the iteration number, number of points used, sigma, change in sigma, and the adjusted parameters on each iteration. Test the fractional change in sigma to determine if further iterations are necessary.

7. On termination of fitting, print the following information:

a. Title of fit.

b. The parameters, their values and estimated uncertainties.

c. The number of iterations, final value of sigma and last fractional change in sigma.

d. Each variable, its position in the data list, the units it is in, and its metric.

e. On request, the value of the measured variables, their error assuming all other variables correct (Δx), the estimate of the real error in the variable ($\Delta' x$), the reduced distance of the point from the functional surface (z^*), and a point identification of up to six alpha-numeric symbols.

f. The root-mean-square values of the Δx and $\Delta' x$ for each variable in the units of the data.

All procedures which are common to all fitting processes have been incorporated into the main program. The remaining sections, namely the calculations of $\{x'\}$, $\{\frac{\partial x'}{\partial \theta_k}\}$ and any computed metrics must be supplied for each function (or set of data for the computed metrics) by the user as described later.

Program Implementation

The program used for fitting the examples of the previous section is listed in Table V. It will be assumed that the reader is familiar

TABLE V. BASIC PROGRAM LISTING

```
10 DIM B(8),M(8),P1(8),S(8),T(8),X(8),X0(8),H(8,8),P0(8,8)
20 DIM A$6,F$1,M$9,N1$1,N2$1,R$64,T$64,Y$1
30 DIM B$(4)64,C$(12)2,M$(8)1,T$(8)8,U$(8)8,X$(8)8
40 SELECT PRINT 005: PRINT HEX(03);: INPUT "TITLE OF FIT",T$
50 INPUT "NUMBER OF VARIABLES",N
60 MAT REDIM X(N),X$(N)8,U$(N)8,X0(N): BIN(N1$)=N*5+10
70 PRINT "NAMES OF VARIABLE,UNITS,METRIC": MAT M=CON(N)
80 FOR I=1 TO N: INPUT "[]",X$(I),U$(I),M(I)
90 BIN(M$(I))=I: IF M(I)[=0 THEN 100: BIN(M$(I))=0
100 NEXT I
110 INPUT "NEW DATA TAPE",Y$:IF Y$[]"Y" THEN  120: GOSUB 1000
120 S0,N0=0:S=-99: REWIND
130 INPUT "NUMBER OF PARAMETERS",P: MAT REDIM T(P),T$(P)8
140 PRINT "ENTER PARAMETER NAMES":MAT T=ZER:T=1E-7:MAT INPUT T$
150 PRINT "INPUT DESIRED DELTA SIGMA IF OTHER THAN ";T;:INPUT T
160 FOR I=1 TO P:PRINT T$(I);" ESTIMATE";:INPUT T(I):NEXT I
170 GOTO 300
180 N0=N0+1:N9,S=0: MAT B=ZER(P): MAT H=ZER(P,P)
190 GOSUB '100(0)
200 GOSUB '100(1): IF N1]0 THEN 280
210 BIN(F$)=0: GOSUB '200: IF VAL(F$)[]0 THEN 200
220 GOSUB '201: Z1=0: N9=N9+1: PRINT HEX(01);N9: MAT P1=ZER(P)
230 FOR I=1 TO N: D=X(I)-X0(I): X=(M(I)/D)!2: Z1=Z1+X
240 FOR K=1 TO P: P1(K)=P1(K)-X*P0(I,K)/D: NEXT K: NEXT I
250 Z1=SQR(1/Z1): S=S+Z1*Z1: MAT P1=(Z1!3)*P1
260 FOR I=1 TO P: B(I)=B(I)+Z1*P1(I): FOR K=1 TO P
270 H(I,K)=H(I,K)+P1(I)*P1(K): NEXT K:NEXT I: GOTO 200
280 FOR I=1 TO P: FOR K=1 TO I: H(I,K)=H(K,I): NEXT K: NEXT I
290 MAT P0=INV(H): MAT S=P0*B: MAT T=T-S: S=SQR(S/N9)
300 S9=ABS(S-S0)/(S+S0)*2
310 PRINT : PRINT HEX(03);"ITN ";N0;" N=",N9,S,S9: MAT PRINT T
320 IF ABS(S-S0)[T*(S+S0)/2 THEN 330: S0=S: GOTO 180
330 STOP ":CONVERGENCE OCCURRED.  ENTER continue FOR PRINTOUT"
340 INPUT "PRINT ALL DATA",Y$: SELECT PRINT 213(80): REWIND
350 GOSUB 2000:STOP "ENTER continue FOR ADDN'L OUTPUT":GOTO 340
1000 N9=0: GOSUB '100(2)
1010 A$="END": N9=N9+1
1020 INPUT "DATA POINT:",A$: IF A$="END" THEN 1060
1030 MAT INPUT X: Y$="Y": INPUT "DATA OK",Y$
1040 IF Y$="Y" THEN 1050: PRINT "RE-ENTER ";A$;" ";: GOTO 1020
1050 GOSUB '100(3): GOTO 1020
1060 INPUT "ALL DATA ENTERED",Y$: IF Y$[]"Y" THEN 1010
1070 GOSUB '100(4): PRINT "DATAFILE CLOSED": RETURN
1080 DEFFN'100(Q): ON Q GOTO 1120,1190,1200,1240
1090 DATA LOAD "DATAFILE"
1100 DATA LOAD BTB$(): IF END THEN 1180
1110 N1=0: BIN(N2$)=1: IF Q=0 THEN 1170
1120 IF 252-VAL(N2$)[VAL(N1$) THEN 1100
1130 $UNPACKB$()[VAL(N2$),VAL(N1$)-4] TO R$
1140 IF STR(R$,1,3)[]"END" THEN 1150:N1=7:BACKSPACE 1F: RETURN
1150 UNPACK(+#.######!!!!)STR(R$,7) TO X()
1160 A$=STR(R$,1,6): ADD(N2$,N1$)
1170 RETURN
```

TABLE V. Continued

```
1180 BACKSPACE 1F: RETURN CLEAR : RETURN
1190 REWIND : DATA LOAD "DATAFILE": BIN(N2$)=1: RETURN
1200 STR(R$,1,6)=A$: PACK(+#.######!!!!)STR(R$,7) FROM X()
1210 $PACKB$()[VAL(N2$),VAL(N1$)]FROM STR(R$,1,VAL(N1$)-4)
1220 ADD(N2$,N1$): IF 252-VAL(N2$)[VAL(N1$) THEN 1230: RETURN
1230 DATA SAVE BTB$(): BIN(N2$)=1: RETURN
1240 A$,STR(R$,1,6)="END": PACK(+#.######!!!!)STR(R$,7)FROM X()
1250 $PACKB$()[VAL(N2$),VAL(N1$)]FROM STR(R$,1,VAL(N1$)-4)
1260 DATA SAVE BTB$(): DATA SAVE END :BACKSPACE 1F: RETURN
2000 GOSUB '100(0):N9=0:MAT B=ZER(N):MAT S=ZER(N):MAT P1=ZER(N)
2010 PRINT HEX(0D0A0A0A);T$;HEX(0A): FOR I=1 TO P
2020 PRINT T$(I),T(I)," +/-";S*SQR(ABS(P0(I,I)))
2030 NEXT I: PRINT : PRINTUSING 2040,N0,S,S9: PRINT : PRINT
2040% ITN= ###    SIGMA= #.###!!!! DELTA SIGMA= -#.###!!!!
2050 PRINTUSING 2060: PRINT
2060% NR   VARIABLE NAME    UNITS     METRIC
2070%  #   ########         ######## #########
2080 FOR I=1 TO N
2090 IF VAL(M$(I))=0 THEN 2100: M$="COMPUTED": GOTO 2110
2100 CONVERT M(I) TO M$,(#.###!!!!)
2110 PRINTUSING 2070,I,X$(I),U$(I),M$: NEXT I: PRINT : PRINT
2120 IF Y$[]"Y" THEN 2130: PRINTUSING 2240: PRINT
2130 GOSUB '100(1): IF N1[]0 THEN 2270: BIN(F$)=0: GOSUB '200
2140 IF VAL(F$)[]0 THEN 2130
2150 Z1=0: N9=N9+1:MAT X0=X-X0: FOR I=1 TO N
2160 Z1=Z1+(M(I)/X0(I))!2: NEXT I: Z1=SQR(1/Z1)
2170 FOR I=1 TO N: B(I)=(Z1*M(I))!2/X0(I)
2180 S(I)=S(I)+X0(I)!2: P1(I)=P1(I)+B(I)!2
2190 IF Y$[]"Y" THEN 2260: PRINT X(I),
2200 IF I[]1 THEN 2210: M$=A$
2210 IF I[]N THEN 2220: CONVERT Z1 TO M$,(#.###!!!!)
2220 IF I[2 THEN 2230: IF I=N THEN 2230: M$=" "
2230 PRINTUSING 2250,X0(I),B(I),M$
2240% VARIABLE              DELTA    DELTA PRIME    ID/Z*
2250%  -#.###!!!!  -#.###!!!!    ##########
2260 NEXT I: IF Y$[]"Y" THEN 2130: PRINT : GOTO 2130
2270 PRINT HEX(0A0A0A0A): PRINTUSING 2280: PRINT
2280% NR   VARIABLE NAME   RMS DELTA   RMS DELTA PRIME  UNITS
2290%  #   ########        #.###!!!!      #.###!!!!     ########
2300 FOR I=1 TO N
2310 PRINTUSING 2290,I,X$(I),SQR(S(I)/N9),SQR(P1(I)/N9),U$(I)
2320 NEXT I: PRINT : PRINT N9;" POINTS": RETURN
5100 M(J)=.0002*X(1): IF M(J)].1 THEN 5110: M(J)=.1
5110 RETURN
6000 DEFFN'200
6010 X(2)=X(2)+273.15: X0(1)=10!(T(2)+T(1)/X(2))
6020 X0(2)=T(1)/(LOG(X(1))/LOG(10)-T(2))
6030 FOR J=1 TO N:ON VAL(M$(J)) GOSUB 5100: NEXT J: RETURN
6040 BIN(F$)=7: RETURN
6050 DEFFN'201: P0(1,2)=X0(1)*LOG(10)
6060 P0(1,1)=P0(1,2)/X(2)
6070 P0(2,1)=1/(LOG(X(1))/LOG(10)-T(2)): P0(2,2)=T(1)*P0(2,1)!2
6080 RETURN
```

with standard Dartmouth BASIC and the extensions provided for the Wang system used. The character set of the typewriter used required three nonstandard symbols in the listing. They are

Listing	Standard Symbol
!	↑
[	<
]	>

The extensions to BASIC used may be found in the System 2200S Wang BASIC Language Programming Manual published by Wang Laboratories, Inc., Tewksbury, MA.

Statements 10-30 allocate space for matrices and alphanumeric variables. The symbols have been kept consistent with earlier parts of this report. Primed quantities have had a zero appended and starred quantities have had a one appended since primes and asterisks are not legal parts of a BASIC variable (e.g., x_i' becomes XØ(I) and z* becomes Z1).

Statements 40-170 perform the functions indicated in 1, 2 and 3 above. The sequence of statements to generate a tape appears at 1000-1070.

Statements 180-330 perform the functions described in 4, 5, and 6 above. These operations are repeated using the data from cassette until convergence of the values is achieved.

Statements 340-350 with the associated subroutine sequence in statements 2000-2320 perform the output as specified in 7 above.

Statements 1080-1260 are a sequence of statements for cassette handling--to read and write the data on tape. The argument Q passed to the subroutine determines the operation to be performed.

Value of Q (integer)	Operation
Q<1 or Q>4	Rewind tape and open "DATAFILE" for reading
Q = 1	Read a data point from tape
Q = 2	Rewind tape and open "DATAFILE" for writing
Q = 3	Write a data point to tape
Q = 4	Write any data points stored in buffer to tape and close "DATAFILE" on tape

The following three sets of statements are user-supplied and tailored to each fit:

Statements 5100-5110 - compute a metric for the first variable

Statements 6000-6040 - compute the values of the primed variables $\{x'\}$

Statements 6050-6080 - compute the partial derivatives $\{\frac{\partial x_i'}{\partial \theta_k}\}$

Program Modification for Fitting Other Functions

To fit another function the statements from 6000 on must be replaced. A new subroutine to calculate $\{x_i'\}$ must be written (DEFFN '200), and a new subroutine to calculate the necessary partial derivatives, $\{\frac{\partial x_i'}{\partial \theta_k}\}$, must replace the old DEFFN '201.

In addition, if the method of computing a metric for any variable is changed, suitable changes in statement 6030 and the subroutine at 5100 would have to be made also.

1. Calculation of the $\{x_i'\}$ is done in subroutine DEFFN '200. When the subroutine DEFFN '200 is entered the measured values for a point are available in X(I), where I = 1 to N (N is the number of variables), and the current parameter estimates are available in T(J) where J = 1 to P (P is the number of parameters). The values of the $\{x_i'\}$ are to be

calculated and stored in XØ(I), I = 1 to N. In the listing the pressure p' ≡ XØ(1) is calculated using the parameters T(1), T(2) and the temperature t ≡ X(2). The temperature t' ≡ XØ(2) is calculated from the parameters T(1), T(2) and the measured pressure p ≡ X(2).

2. Calculation of partial derivatives $\{\frac{\partial x_i'}{\partial \theta_k}\}$ is done in subroutine DEFFN '201. The partial derivative of every variable with respect to each parameter must be calculated and stored in an N x P array. The partial derivative $(\frac{\partial x_1}{\partial \theta_2})$ is calculated and stored in PØ(1,2), and $(\frac{\partial x_2}{\partial \theta_1})$ in PØ(2,1)

3. Calculation of computed metrics is done by a normal subroutine internally in the marked subroutine DEFFN '200. The matrix element M$(I) contains a zero for fixed metric variables, or the index of the variable for variables whose metric is to be computed. This index may be used to call the correct subroutine as in the example of statement 6030 and the subroutine that statement calls at statements 5100-5110. The computer metric for variable I is stored in M(I).

Operation

Write the necessary subroutines and metric calculating procedures. Insure that the statement numbers for these routines begin with 3000 or a larger number (9999 is the largest statement number allowed on the Wang 2200). Save the program on tape, if desired.

(a) Mount the tape containing GENFIT and rewind it.

(b) Enter LOAD "GENFIT" ↓ where ↓ stands for RETURN (EXEC).

(c) Enter RUN ↓.

(D) The program will respond with: TITLE OF FIT?

(d) Enter any character string up to 64 characters long which does not contain a comma, then ↓.

(F) The program will respond with: NAMES OF VARIABLE, UNITS, METRIC, <>?

(f) For each variable, when the <>? is displayed, enter the requested information, separated by commas, then ↓. After the last entry program response is:

(G) NEW DATA TAPE?

(g) Enter any word starting with a Y and ending with a ↓ for an affirmative answer. Any other response will be interpreted as NO ↓. If a data tape is already in existence, mount and rewind it, proceed to (M) below. If a new tape is to be created, insure a tape with a data file header is mounted and rewound. (Note: A data file header may be created on a new tape by mounting and rewinding the tape with the write tabs on, then entering the command DATASAVE OPEN "DATAFILE" ↓.)

(H) Program response after affirmative response to new tape is DATA POINT:?

(h) Enter any point identifier (up to 6 letters) then ↓ or if last data point has been entered, enter END ↓.

(I) For program response to END ↓ see (K) below. For any other identifier: ?

(i) Enter variable values in order separated by commas or carriage returns.

(J) Program response after last value: DATA OK?

(j) Enter ↓ or Y↓ if data is correct. If not all right, enter any string of characters not starting with a Y. Program response if data is OK is to return to output (H). If data is not OK, data will not be put on tape and return to (H) for re-entry of data point.

(K) Program response: ALL DATA ENTERED?

(k) If all entered, enter Y↓ or ↓ ; more data to be entered any response not beginning with Y will return program to (H) above.

(L) Program response to an affirmative to all data entered is to close the data file and print, DATAFILE CLOSED; no response is required.

(M) Program will ask: NUMBER OF PARAMETERS?

(m) Enter the number of parameters the program is designed to fit, (this number is called P), then ↓.

(N) Program will respond: ENTER PARAMETER NAMES ?

(n) Enter the parameter names in order separated by commas, then ↓.

(O) Program will respond: INPUT DESIRED DELTA SIGMA IF OTHER THAN 1.00000000E-07?

(o) If a different value of the fractional change in sigma to terminate the fit is desired enter it, then ↓, otherwise just ↓.

(P) Program will respond with requests for each of the parameters _ _ _ ESTIMATE? (Where _ _ _ is the parameter name.)

(p) Enter the estimate for each parameter as requested, followed by ↓.

(Q) After last estimate is entered, the program will proceed to iteratively fit the data on the casette until the fractional change in sigma is smaller than the entered delta sigma. Initially and after each iteration, the iteration number, number of points used, sigma, the fractional change in sigma, and all of the current parameters are displayed. On convergence (fractional change in sigma < delta sigma) the program will respond: STOP: CONVERGENCE OCCURRED. ENTER CONTINUE FOR PRINTOUT.

(q) Enter CONTINUE ↓

(R) Program will respond PRINT ALL DATA?

(r) If all data is to be printed with error indicators respond Y↓; if not and only a summary is desired give any response not starting with Y and only the summary will be printed.

(S) Program terminate with the message STOP ENTER CONTINUE FOR ADDN'L OUTPUT

(s) If further output is desired, enter CONTINUE ↓ and program will resume at (R) above.

Index of Variables Used in Program

Single variables (numeric)

D (temporary) Used to store $\Delta x_i = (x_i - x_i')$

I (temporary) Used as an index in FOR loops

J (temporary) Used as an index in FOR loops

K (temporary) Used as an index in FOR loops

N - Number of variables

NØ - Iteration number

N1 - End of file indicator for datafile = 0:no EOF; ≠ 0:EOF

N9 - Number of points entered into fit (incremental as each point is processed)

P - Number of parameters

S - Sigma; used to collect $\Sigma(z^*)^2$ during each iteration

S0 - Sigma of previous iteration

S9 - Fractional change in sigma on last iteration $\frac{ABS(S-S0)*2}{(S+S0)}$

T - Delta sigma; termination of fit indicated by S9<T

X (temporary) Used to calculate z*

Z1 - z*; the distance from the functional surface in the metric space

Single variables (alphanumeric) [final digits indicate length in characters]

A$6 (temporary) Point identification for current point

F$1 (temporary) Flag used to indicate valid point F$=0, valid point; F$≠0, invalid point program will not use for fitting

M$9 (temporary) Used to display

N1$1 - String length indicator for packing and unpacking data points

N2$1 - Pointer for packing and unpacking data points

R$64 (temporary) String variable used to pack and unpack data points

T$64 - String variable used to hold title of fit

Y$1 - Response indicator for various portions of program (YES/NO)

Array variables (numeric)

B( ) - Used for error vector during fitting; used to store $(\Delta'x)$ during print out

M( ) - Used to store metrics for variables

P1( ) - Used to store $\{\frac{\partial z^*}{\partial \theta_i}\}$ during fitting; $\Sigma(\Delta'x)^2$ during printout

S( ) - Used to store $\Sigma(\Delta x)^2$ during printout

T( ) - Used to store the parameter values

X( ) - Used to store the measured values for each point

X0( ) - Used to store calculated values for each point during fit; used to store $\{\Delta x\}$ during printout

H(,) - Curvature matrix for least squares fit

P0(,) - Used to store partial derivatives $\{\frac{\partial x_i}{\partial \theta_k}\}$ during initial part of fit; used to store the error matrix (inverse of H) during later part of fit

Array variables (alphanumeric) [final digits are lengths of each element in characters]

B$( )64 - Buffer used to store data for tape operations

M$( )1 - Used to store computed metric indices

T$( )8 - Used to store names of parameters

U$( )8 - Used to store name of units for variables

X$( )8 - Used to store name of variables

REFERENCES

1. B.V. Gnedenko and A. Ya. Khinchin, An Elementary Introduction to the Theory of Probability, W.H. Freeman and Co., San Francisco, 1961.

2. B.E. Margules, Systems of Linear Equations, MacMillan, New York, 1964.

3. J.T. Viola, D.W. Seegmiller, A.A. Fannin, and L.A. King, unpublished data, Frank J. Seiler Research Laboratory (AFSC), US Air Force Academy, Colorado.

4. P.R. Bevington, Data Reduction and Error Analyses for the Physical Sciences, McGraw-Hill, New York, 1969, p. 242ff.