AD ____________

APPLICATION OF THE KALMAN FILTER TO SEQUENTIAL OPTIMAL PARAMETER ESTIMATION VIA HOUSEHOLDER'S MATRIX INVERSION METHOD

(Special Report)

by

James Pappas

Approved by Guenther Hintze
GUENTHER HINTZE
Director

June 1967

U. S. ARMY TEST AND EVALUATION COMMAND

ANALYSIS AND COMPUTATION DIRECTORATE
DEPUTY FOR NATIONAL RANGE OPERATIONS
WHITE SANDS MISSILE RANGE, NEW MEXICO

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd)

ABSTRACT

A detailed derivation of the computable discrete equations for parameter estimation are developed in a geometrical vector-space setting for the state-space novice. Classical least squares curve fitting when approached with Kalman's sequential prediction-correction techniques look like state-vector feed-back control problems. It is hoped that this paper will help bridge the gap between some of the modern and classical theory of systems analysis.

PREFACE

Kalman states that recent developments in optimal control system theory are based on vector differential equations as models of physical systems. In the older literature on control theory, however the same systems are modeled by transfer functions (i.e., by the LaPlace transforms of the differential equations relating the inputs to the outputs). Two different languages have arisen, both of which purport to talk about the same problem. In the new approach, we talk about state variables, transition matrices, etc., and make constant use of abstract linear algebra. In the old approach, the key words are frequency response, pole-zero patterns, etc., and the main mathematical tool is complex function theory. (See Ref. 12)

Paul Horst states that all adequate models and methods of multivariate statistical analysis are special cases of matrix factoring techniques. (See Ref. 8)

Mahalanobis says that R. A. Fisher in a 1914 paper introduced for the first time the brilliant technique of representing a sample of size n by a point in a space of n-dimensions. (See Ref. 4, page 268.)

This report is a tutorial introductory paper on the modern approach to real time (on-line) optimal computer control of stochastic dynamical systems (processes).

By dynamical systems we mean continuously and/or discretely changing (time-varying) processes in the presence of uncertainty (stochastics).

By stochastic systems we mean real-world processes changing in the presence of process-noise and measurement-noise.

By modern we mean with respect to the mid 1960's with respect to modern mathematical methods: state-vectors in Eucleadean p-space, pxn state-transition matrices, orthogonal projections in Hilbert spaces, etc., in the mathematical spirit of the recent deluge of engineering papers in the IEEE, Siam, etc., by Kalman, Bellman, and many others; papers which are intelligible perhaps to approximately 20% of our engineering people working throughout the aero-space industry.

By "real-time" or "on-line" we have in mind the "computable decision-theoretic" aspects of process analysis and control. If man is to intelligently control any process he must make decisions based on information. The extraction of information about a process must be based on both observations and a computer model depicting the dynamical structure of the source of the information.

The multiple decisions - in the presence of process uncertainty and observational uncertainty - with respect to various criterion of optimality must be programmed onto the machines. Thus we are dealing with an on-line automated analytical system for the association of theory and observation.

The realization of "optimal test or experimental design" appears very promising today. The combination of the newly emerging mathematical theory of optimization with highly reliable computers is paving the way for this realization. Such systems are called computer-referenced optimal estimation, prediction, and control systems.

Man with his mechanisms seeks to control the states of a process (the trajectory, in state-space of the process). We have not yet developed sensors that will measure all of the states of interest. Hence we must attempt to control the states based on the observation of a subset of the states, that is observable space is always (for complex systems) a subspace of state-space. Such systems are said to be not-completely observable.

The states of a process are always being perturbed by process noises. Our instruments used to measure the states of the process are "noisy" or sensor noises are always present.

In order to control a process we must control (or design into the system) the processing of the outputs of the sensors. We must control the information extraction process.

The information contained in these observables is more easily extracted when we know the dynamics (change characteristics) of the sensors and the dynamics of the states (the source of the information).

The European editor of "Electronics" magazine in a recent article "Control Theory: Burgeoning but Baffling", states that control theorists concede that supersophisticated mathematics is outdistancing hardware-oriented engineers. He says that hardware-designing practitioners manage to keep abreast of the theorists working in their field; but that in automatic controls, a freightning gap has opened up between theory and practice. Barlow states that during the last decade, control theory had burgeoned-to the dismay of many a working engineer, the theorists have soared into the rarified atmosphere of higher mathematics - that all to often, a venture into the realm of advanced control theory leaves the hardware-oriented engineer feeling that he's in a strange country with a new language - that even the theorists recognize they've lost touch with the people who translate concepts into hardware - that at the third Congress of the International Federation of Automatic Control held in London, Cambridge University professor John Coales summed up the sentiment. Said Coales, president of the Congress, "Theory has outstripped its application and it almost always takes one or more decades to learn to apply new mathematical techniques (see Ref. 3).

Kalman estimation and predication has recently been and is continuing to be applied to range testing of missile systems at White Sands Missile Range with most promising results. (See Ref. 16 and 17.)

[illegible] information seems to reside in the [illegible] of a process. The problem [illegible] space [illegible] we have not yet extracted [illegible] measure all of the vector of interest. Hence we must [illegible] based on the [illegible] of a subset of the state [illegible] (for [illegible] systems), a subset of [illegible]. Such systems are said to be non-completely observable.

[illegible] are [illegible] in order to measure [illegible] the states [illegible] are always present.

[illegible] control [illegible] design into the system the processing of the outputs of the sensors. We must control the information extraction process.

The information contained in these derivatives is more easily extracted when we know the dynamics (change characteristics) of the sensors and the dynamics of the states (the source of the information).

[illegible] editor of "Electronics" [illegible] in a recent article [illegible] "Control theory [illegible] baffling," [illegible] control theory [illegible] sophisticated mathematics [illegible] hardware [illegible] practitioners [illegible] during the last decade, control theory has [illegible] the dismay of many a working engineer, and theorists have [illegible] the rarified atmosphere of higher mathematics - that all to [illegible] the realm of advanced control theory leaves the hard-nosed engineer feeling that he's in a strange country with a new language - that even the theorists recognize they've lost touch with [illegible] translate concepts into hardware - that at the third Congress of the International Federation of Automatic Control held in London, Cambridge University professor John Coales [illegible] up the [illegible], president of the Congress, "Theory has outstripped [illegible] one or more decades to [illegible]."

[illegible] estimation and prediction has recently been and is continuing [illegible] applied to range testing [illegible] systems at White Sands [illegible] results. (See Refs. 16 and 17.)

INTRODUCTION

This report is a detailed derivation of the modern data processing math-ware for utilization with on-line computers to sequentially compute "best estimates" of process parameters.

Detailed derivations are presented in an effort to make the matrix methods understandable to the multitude of professional people in the industry and government who are not "too" familiar with matrix methods.

Y. E. Ho at the Rand Corporation showed in his paper that the discrete data version of the Kalman filter is a least squares type of filter and leads one to a consideration of the Householder Matrix Inversion Lemma. Ho points out that the Householder-Lemma was known to numerical analysts but not to control engineers. (See Ref. 9.)

The application of the Kalman theory of optimal state-vector estimation and prediction to the optimal parameter estimation problem may be viewed as a special case with nicely definable constraints.

The continuous state-vector formulation of Kalman considers the system of equations

$$\dot{x}\rangle = \underset{pxp}{A(t)}\; x\rangle + u_d\rangle + u_n\rangle$$

and

$$z\rangle = \underset{mxp}{H(t)}\; x\rangle + v\rangle$$

where $x\rangle$ is a p-dimensional column state-vector (the variables of the process of interest), A is a time varying pxp matrix (the dynamics of the process) $u_d\rangle$ is any deterministic forcing vector, $u_n\rangle$ is the process uncertainty vector, $z\rangle$ is an m dimensional observation vector, m is generally less than p, H is the time-varying matrix describing the sensor dynamics and or geometry and $v\rangle$ is the instrumentation noise vector or measurement uncertainty.

The discrete version equations are

$$x(k+1)\rangle = \Phi(k+1, k)\; x(k)\rangle + u_d(k)\rangle + u_n(k)\rangle$$

and

$$z(k)\rangle = H(k)\; x(k)\rangle + v(k)\rangle$$

where Φ is the state transition matrix which maps the system state vector to successive states.

The constraint that there is no process noise, that is

$$u(k)\rangle = 0\rangle$$

and

$$\Phi(k+1, k) = I$$

where I is the identity matrix states that

$$x(k)\rangle = x(1)\rangle \text{ for all } k,$$

hence we can view the system as

$$x(k)\rangle = x(1)\rangle$$

$$z(k)\rangle = Hx(1)\rangle + v(k)\rangle$$

where $x(1)\rangle$ is a constant vector to be estimated.

If we change the size of the matrix H such that m is greater than p, we can interpret the problem as the vectorized version of the classical unweighted least squares problem. For example, if we are estimating the parameters of a polynomial

$$y_k = a_0 + a_1 x_k + \cdots a_{p-1} x_k^{p-1} + e_k$$

for k=1...n samples we obtain, using standard vector space methods

$$(y_1 \cdots y_n) = \langle p) a F_{pxn} + \langle n) e$$

or, as a column vector, transposing

$$y\rangle = F^T_{nxp} a(p)\rangle + e\rangle$$

If we interpret the p-dimensional column vector of parameters $a\rangle$ as $x(1)\rangle$ and F as H etc. we see the analogy.

The predict, observe, correct method allows one to sequentially up-date the estimates of the parameters as the data is fed into the computer. Naturally one desires the sensors measuring the data to automatically feed the computers. The computer can provide a running estimate of the parameters based on all of the past data.

The beauty of the theory is that it does not require the computer to store the past data thus eleminating congestive storage requirements

The method has real-time capabilities when utilized with matrix psuedo-inverses which handle non-full rank statistical data.

The report shows how one solves the classical least-squares polynomial curve fitting problem as a sequential error-corrector or discrete feed back control problem. The scalar-case equation is

$$\langle\hat{a}(k) = \langle\hat{a}(k-1) + (y_k-\hat{y}_k)\ \langle w(k),$$

where $\langle\hat{a}(k-1)$ is a p-dimensional row vector of the parameters of the curve and is the "best estimate" based on k-1 data points, y_k is the $k^{\underline{th}}$ data point and $\hat{y}_k$ is our predicted value of what the $k^{\underline{th}}$ data point should be. The difference $y_k-\hat{y}_k$ is the "feed back" error term and $\langle w(k)$ is the varying filter or "weighting matrix", in this case a row vector. $\langle\hat{a}(k)$ is the "up-dated" estimate of the parameters.

This report presents the theory for optimal parameter estimation. A follow-up report is in preparation which:

1. Extends the Householder inversion scheme to a psuedo-inversion scheme so that non-full rank data can be processed easily.

2. Develops the equations for time-varying parameter or Kalman state-vector estimation of dynamical processes.

Kalman showed the duality between the control problem and the estimation and prediction problem. It is hoped that this report will help to disseminate information leading to a deeper understanding of this new theory.

NOTATION

The notations used in the report is an effort to blend the notation of Friedman for inner-products and dyadic products with the current journal-literature on vector-spaces, psuedo-inverses, state-vectors etc.

$\underset{p\times k}{X}$ capital letters designate matrices of size p rows and k columns.

$x(k\rangle$: when p = 1, the matrix is called a column vector, and we use Friedmans symbol to distinguish this matrix.

$\langle p)x$: when k = 1, the matrix is a row vector of dimension p.

$\langle p)x\ y(p\rangle$: "inner-product" or scalar product of two vectors.

$y(p\rangle\langle p)x$: "outer-product" or dyadic product of two vectors.

$\underset{p\times k}{X} = [x(p\rangle_1, \ldots, x(p\rangle_k]$. Matrix X partitioned into a row k-tuple of column vectors from a p-space.

$\underset{p\times k}{X} = \begin{bmatrix} {}^{1}\langle k)x \\ \vdots \\ {}^{p}\langle k)x \end{bmatrix}$ Matrix X partitioned into a p-column tuple of row vectors from a k-space.

x small x is a scalar

x^i scalar from a column vector

x_j scalar from a row vector

Scalar here is a "real field" element.

Perp = perpendicular

O.N. = ortho-normal

SECTION I

INTRODUCTORY VECTOR SPACE METHODS TO DETERMINISTIC SYSTEMS OF EQUATIONS

Suppose we have a sequence of k two dimensional observation vectors about a process as shown in Figure (1) and

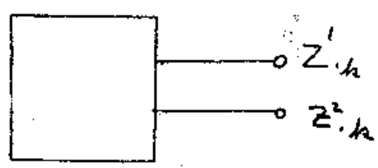

Figure 1 Process

we format the data as a two-vector

$$\begin{pmatrix} z^1 \\ z^2 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \qquad (1)$$

and plot the coordinates as shown in Figure (2). The first coordinate z^1 may be the input to a scalar device and z^2 the output.

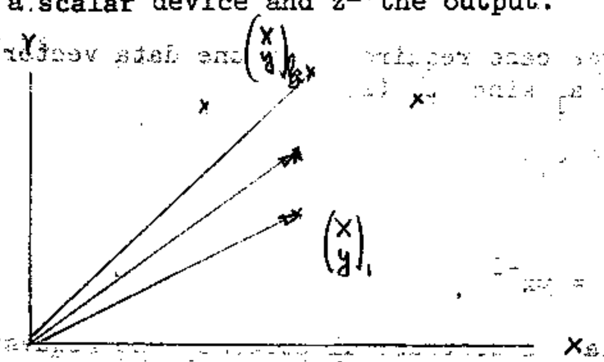

Figure (2) Two-Vector Data Graph

If the variables x and y are related by a linear (gain) constant function as shown in Figure (3) and described in Equation (2).

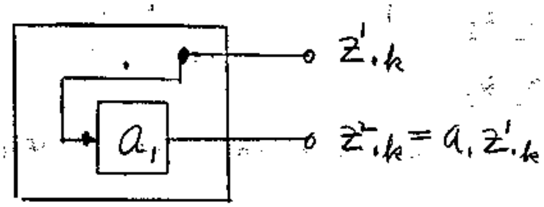

Figure (3) Constant Gain-Noise Free Block

$$\begin{pmatrix} z^1 \\ z^2 \end{pmatrix}_k = \begin{pmatrix} x \\ y \end{pmatrix}_k = \begin{pmatrix} x \\ a_1 x \end{pmatrix}_k = \begin{pmatrix} 1 \\ a_1 \end{pmatrix} x_k \tag{2}$$

then clearly the graph of the Process is a straight line through the origin as shown in Figure (4).

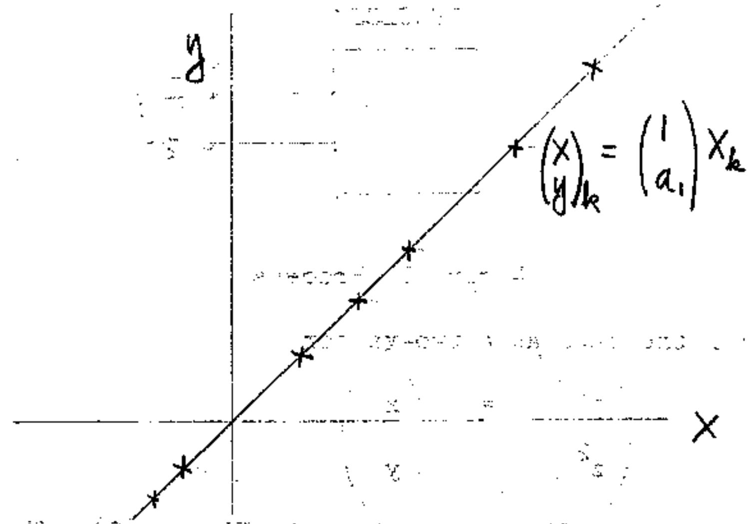

Fig. (4) Straight Line Error Free Graph

The error-free case requires only one data vector (k = 1) to solve for the parameter a_1 since by (2)

$$y = a_1 x \tag{3}$$

or

$$a_1 = yx^{-1} \; . \tag{4}$$

Suppose we have a distrustful customer who requests that we take, say, three observations of the device of Fig (3), and use all three of the data points to compute the gain a_1, that is

$$\begin{aligned} y_1 &= a_1 x_1 \\ y_2 &= a_1 x_2 \\ y_3 &= a_1 x_3, \end{aligned} \tag{5}$$

which is an over-specified system of equations - three equations and one unknown.

If we package the relations

$$(y_1 \; y_2 \; y_3) = (a_1 x_1, \; a_1 x_2, \; a_1 x_3) = a_1 (x_1, \; x_2, \; x_3) \tag{6}$$

we obtain the vector equation. Since our classical matrix theory books did not inform us how to invert the 1x3 matrix of Equation (6) we must face the advancing world and attempt to use our ingenuity. Suppose we transpose the row vector

$$(x_1, x_2, x_3)^T = \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = x\rangle \tag{7}$$

and post-multiply (6) by (7) then

$$\langle y\; x\rangle = a_1 \langle x\; x\rangle \,. \tag{8}$$

We now have one scalar equation in one unknown, hence invert the inner-product term $\langle x\; x\rangle$ and obtain

$$\frac{\langle y\; x\rangle}{\langle x\; x\rangle} = a_1 \tag{9}$$

or in open form

$$a_1 = \frac{y_1 x^1 + y_2 x^2 + y_3 x^3}{x_1 x^1 + x_2 x^2 + x_3 x^3} \,. \tag{10}$$

Since Equation (6) is a three dimensional vector equation in one unknown a_1 we could take the inner-product with the sum vector $1\rangle$ and map down to a single equation hence

$$\langle y\; 1\rangle = a_1 \langle x\; 1\rangle \tag{11}$$

or

$$\frac{\langle y\; 1\rangle}{\langle x\; 1\rangle} = a_1 \,. \tag{12}$$

Clearly the result is the same as adding all three equations of (5) to obtain a single equation, very much like averaging. We may also take the inner-product of (6) with the known data $y\rangle$ and obtain

$$\langle y\; y\rangle = a_1 \langle x\; y\rangle \tag{13}$$

or

$$a_1 = \frac{\langle y\; y\rangle}{\langle x\; y\rangle} \tag{14}$$

We may now ask the question: given the three dimensional vector equation in a scalar unknown a_1, that is

$$\langle 3)y = a_1 \langle 3)x \tag{15}$$

Can we find a right hand multiplicative element that acts like a "vector inverse" or inverse to a 1x3 matrix. Suppose this "thing" is a 3x1 matrix $v(3\rangle$, then

$$\langle y\ v\rangle = a_1 \langle x\ v\rangle \tag{16}$$

and we constrain the "inner product"

$$\langle x\ v\rangle = 1 \tag{17}$$

or matrix product to be the identity operator in one-space.

Clearly, by inspection, or guess, if

$$v\rangle = \frac{x\rangle}{\langle x\ x\rangle} \tag{18}$$

then (18) will do the job. In fact (18) is the <u>generalized inverse</u> of $\langle x$. If we designate the generized inverse by + symbol then we have

$$(\langle x)^{+} = x^{+}\rangle = (\langle x)^{T}\ [\langle x(x\rangle)^{T}]^{-1} \tag{19}$$

or in matrix form when

$$\langle x = \underset{1xk}{X} \tag{20}$$

$$\underset{kx1}{X^{+}} = \underset{kx1}{X^{T}}(\underset{(1xk)}{X}\ \underset{(kx1)}{X^{T}})^{-1} \tag{21}$$

The column vector of Equation (18) acts like a one-side inverse. Another column vector $\dfrac{z\rangle}{\langle x\ z\rangle}$ yields

$$\langle x\left(\frac{z\rangle}{\langle x\ z\rangle}\right) = \frac{\langle x\ z\rangle}{\langle x\ z\rangle} = 1 \tag{22}$$

Even though both vectors yield the identity operator 1 in one space, the vector $\dfrac{z\rangle}{\langle x\ z\rangle}$ is not a generalized inverse. Among the other requirements to qualify as the unique generalized inverse, the commuted product must be a projector (idempotent matrix of index 2), that is

$$|x\rangle\langle v| = \frac{|x\rangle\langle x|}{\langle x\, x\rangle} = P \tag{23}$$

where

$$P^2 = PP = \left(\frac{|x\rangle\langle x|}{\langle x\, x\rangle}\right)\left(\frac{|x\rangle\langle x|}{\langle x\, x\rangle}\right) = \frac{|x\rangle\langle x|}{\langle x\, x\rangle} = P. \tag{24}$$

Suppose we have two parameters, a bias a_0 plus the gain parameter a_1, or a line displaced from the origin

$$\begin{pmatrix} x \\ y \end{pmatrix}_k = \begin{pmatrix} x \\ a_0 + a_1 x \end{pmatrix}_k \tag{25}$$

The connection between the process variables becomes as shown in Figure (5)

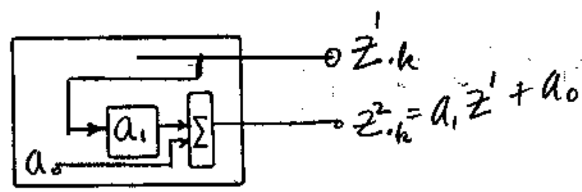

Figure (5)

or simply

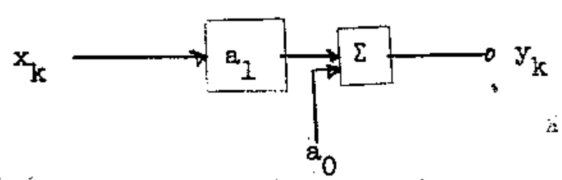

Figure (6) Two Parameter Block

and the graph is

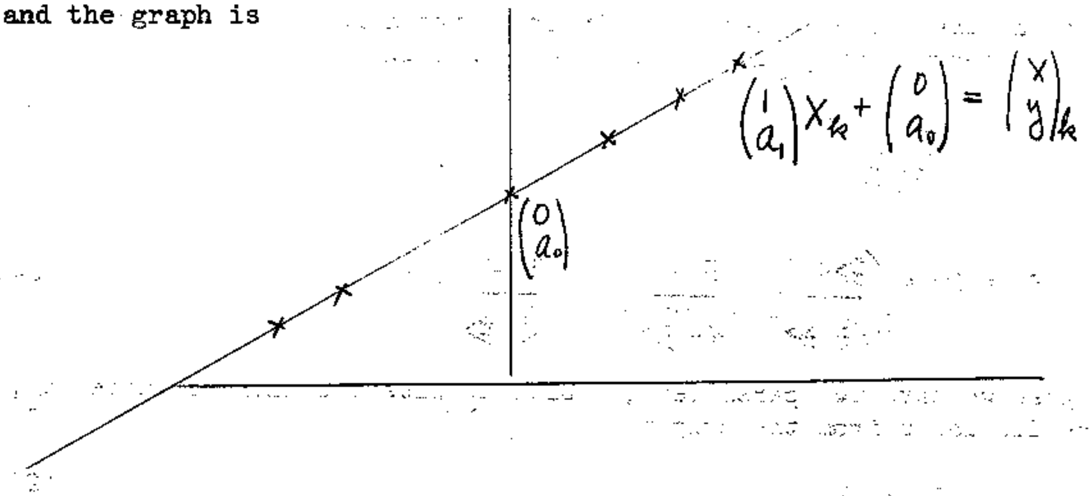

Figure (7) Translated Line Graph

By Equation (25)

$$y_k = a_0 + a_1 x_k \tag{26}$$

and separating the parameters we can write (26) as a vector inner-product as

$$y_k = (a_0, a_1) \begin{pmatrix} 1 \\ x \end{pmatrix}_k \tag{27}$$

or

$$y_k = \langle a \; f \rangle_k \tag{28}$$

where

$$f\rangle_k = \begin{pmatrix} f^1 \\ f^2 \end{pmatrix}_k \equiv \begin{pmatrix} 1 \\ x \end{pmatrix}_k . \tag{29}$$

Equation (26) is one scalar equation with two unknowns and if we have two readings (k = 2) we have

$$y_1 = a_0 + a_1 x_1 = (a_0, a_1) \begin{pmatrix} 1 \\ x \end{pmatrix}_1 = \langle a \; f \rangle_1 \tag{30}$$

$$y_2 = a_0 + a_1 x_2 = (a_0, a_1) \begin{pmatrix} 1 \\ x \end{pmatrix}_2 = \langle a \; f \rangle_2 \tag{31}$$

or two equations in two unknowns a_0 and a_1. We can, by any classical methods, without mentioning a vector, solve the two simultaneous equations in two unknowns.

By a vector space approach we may package (30) and (31) into the vector form of (28) as

$$(y_1, y_2) = [\langle a\, f\rangle_1, \langle a\, f\rangle_2] \tag{32}$$

factoring out $\langle a$

$$\langle y = \langle a\, [f\rangle_1\ f\rangle_2] \tag{33}$$

or

$$\langle y = \langle a\, F_{2x2} \tag{34}$$

which in open-form is

$$(y_1\ y_2) = (a_0,\ a_1) \begin{bmatrix} 1 & 1 \\ x_1 & x_2 \end{bmatrix} . \tag{35}$$

If the two column vectors of F are linearly independent the inverse exists and

$$\langle 2) y\, F_{2x2}^{-1} = \langle 2) a \tag{36}$$

is the solution.

Since

$$F^{-1} = \frac{\text{adj } F}{\det F} \tag{37}$$

$$\det F = x_2 - x_1 \tag{38}$$

and the adjoint matrix is

$$\text{adj } F = \begin{pmatrix} x_2 & -1 \\ -x_1 & 1 \end{pmatrix} \tag{39}$$

then

$$(a_0,\ a_1) = (y_1\ y_2) \begin{bmatrix} x_2 & -1 \\ -x_1 & 1 \end{bmatrix} \frac{1}{(x_2 - x_1)} \tag{40}$$

We are not interested in the analytical solution by conventional inversions but by computer mechanizations of iterative inversion or iterative psuedo-inversions.

We see that the simple problem has been cast into modern vector-matrix form in keeping with the current trend to state-vectorize every math model of a process.

If we factor F of Equation (35) into its row space we have

$$(y_1, y_2) = (a_0, a_1) \begin{bmatrix} (1, 1) \\ (x_1 \; x_2) \end{bmatrix} = a_0(1,1) + a_1(x_1 \; x_2) \tag{41}$$

or the vector $\langle y$ is a linear combination of the vector $\langle 1$ and $\langle x$, that is,

$$\langle y = a_0 \langle 1 + a_1 \langle x \tag{42}$$

as shown geometrically in Figure (8)

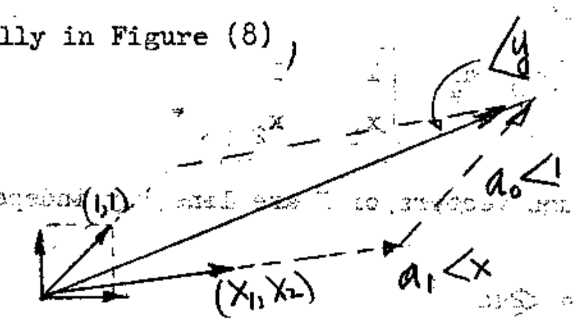

Figure (8) Two Space Version

clearly the vector $\langle y$ belongs to the two dimensional space generated by the two vectors $\langle 1$ and $\langle x$. If F of Equation (34) is invertible, then the data vector $(x_1 \; x_2)$ is not parallel to the vector (1, 1) and the row vectors of F are said to form a set of basis vectors for the two space.

Suppose we compute the two parameters a_0 and a_1 based on all k points on the line, then when k is greater than two we have the over specified system of equations

$$\begin{aligned} y_1 &= a_0 + a_1 x_1 \\ &\vdots \\ y_k &= a_0 + a_1 x_k \end{aligned} \tag{43}$$

We now have k equations and only two unknowns. Suppose we add all of the equations of (43) to obtain

$$y_1 + \ldots + y_k = ka_0 + a_1(x_1 + \ldots + x_k) \tag{44}$$

which in summation form is

$$\sum_{j=1}^{k} y_j = ka_0 + \sum_{j=1}^{k} x_j \tag{45}$$

or in vector inner-product form is

$$\langle y\ 1 \rangle = a_0 \langle 1\ 1 \rangle + a_1 \langle x\ 1 \rangle\ . \tag{46}$$

Equation (44) is one scalar equation in two unknowns. If we multiply the j^{th} equation of (43) by say the data point x_j and add all equations we obtain

$$y_1x_1 + y_2x_2 + \ldots y_kx_k = a_0(x_1 + \ldots + x_k)$$

$$+ a_1(x_1^2 + x_2^2 + \ldots + x_k^2) \tag{47}$$

which in tedious summation form is

$$\sum y_jx_j = a_0 \sum x_j + a_1 \sum x_j^2 \tag{48}$$

In vector space language (47) is written as

$$\langle y\ x \rangle = a_0 \langle 1\ x \rangle + a_1 \langle x\ x \rangle\ . \tag{49}$$

Two equations in the two unknowns are (46) and (49) which can be stated in matrix form as

$$[\langle y\ 1 \rangle, \langle y\ x \rangle] = (a_0, a_1) \begin{bmatrix} \langle 1\ 1 \rangle & \langle 1\ x \rangle \\ \langle x\ 1 \rangle & \langle x\ x \rangle \end{bmatrix} \tag{50}$$

If we use all k points of (43) as a row vector we have the analog of (34)

$$(y_1, \ldots, y_k) = (a_0, a_1) \begin{bmatrix} (1, \ldots, 1) \\ (x_1, \ldots, x_k) \end{bmatrix} \tag{51}$$

or by (33)

$$(y_1, \ldots, y_k) = \langle a\ [f(2\rangle_1, \ldots, f(2\rangle_k] \tag{52}$$

Equation (51) may be written in row space as

$$\langle k)y = (a_0,\ a_1) \begin{bmatrix} \langle k)1 \\ \langle k)x \end{bmatrix} \tag{53}$$

or

$$\langle k)y = a_0 \ \langle k)1 + a_1 \ \langle k)x \tag{54}$$

as shown in Figure (9).

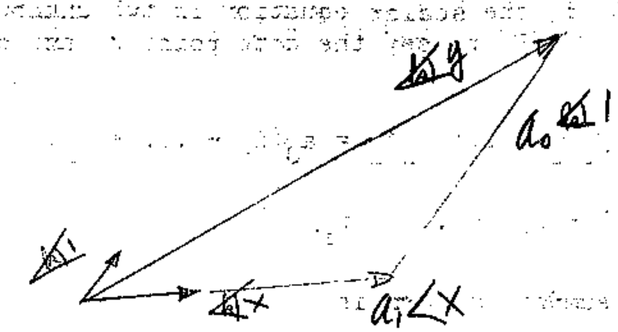

Fig. (9) Two Dimensional Parameter Subspace of k-Space

Elementary vector-space texts prove that if the vectors $\langle k)1$ and $\langle k)x$ are a basis, then the coordinates a_0 and a_1 of the vector $\langle y$ are unique. Hence (53) has a unique solution.

Equation (51) is

$$\langle k)y = \langle 2)a \underset{2xk}{F} \tag{55}$$

We cannot invert the 2xk matrix in the "conventional sense" hence we must use some ingenuity.

Consider the row space of the matrix F

$$\underset{2xk}{F} = \begin{bmatrix} \langle k)1 \\ \langle k)x \end{bmatrix} \tag{56}$$

and its transpose

$$\underset{kx2}{F^T} = [1(k\rangle,\ x(k\rangle] \tag{57}$$

Multiply Equation (55) by (57) then

$$\langle k)y \underset{kx2}{F^T} = \langle 2)a \underset{2(k)2}{F\ F^T} \tag{58}$$

If we look at the 2x2 symmetrix matrix FF^T as

$$FF^T = \begin{bmatrix} \langle 1 \\ \langle x \end{bmatrix} [1\rangle, x\rangle] \tag{59}$$

or

$$FF^T = \begin{bmatrix} \langle 1\,1\rangle & \langle 1\,x\rangle \\ \langle x\,1\rangle & \langle x\,x\rangle \end{bmatrix} \tag{60}$$

we arrive at the matrix of Equation (50) arrived at differently.

We now have a 2x2 matrix which is invertible if the two k dimensional vectors $\langle k)1$ and $\langle k)x$ are linearly independent, hence

$$\langle k)y \underset{kx2}{F^T} \underset{(2x2)}{(FF^T)^{-1}} = \langle 2)a \tag{61}$$

If we define the generalized inverse of the 2xk matrix of Equation (55) as indicated by Equation (61)

$$\underset{kx2}{F^+} = \underset{kx2}{F^T} \underset{2x2}{(FF^T)^{-1}} \tag{62}$$

then (61) becomes

$$\langle y \underset{kx2}{F^+} = \langle a \tag{63}$$

The kx2 generized inverse matrix is a one-sided inverse with respect to F since

$$\underset{(2xk)}{F} \underset{(kx2)}{F^+} = FF^T(FF^T)^{-1} = \underset{2x2}{I} \tag{64}$$

Observe that the commuted product

$$\underset{k(2)k}{F^+ F} = \underset{kxk}{P} \neq \underset{kxk}{I} \tag{65}$$

is a kxk matrix and is not the identity in k-space.

P of Equation (65) is a projection operator that is

$$P^2 = P \tag{66}$$

Observe that the vector equation in the unknown parameter vector $\langle 2)a$ of Equation (55), that is

$$\langle k)y = \langle a \underset{2xk}{F} \tag{67}$$

yields a solution for $\langle a$ if we post multiply by F^{+}, that is

$$\langle yF^{+} = \langle a\, FF^{+} = \langle a \tag{68}$$

when (64) is true.

We shall later be interested in stage-wise iterative inversion schemes on a computer, for as each new data point packs into the symmetric matrix

$$\underset{2xk}{F}\ \underset{kx2}{F^{T}}$$

we need a new inversion. Clearly FF^{T} is a function of k.

It is clear that the condition of equation (11) implies that the $\langle r$ vector of Figure (2) be perpendicular to the sum vector $\langle 1$ in k-space, hence the value of a_0 which yields a point of symmetry in one-space is the value of a_0 which makes the data vector $\langle z$ "perp" to the one-space spanned by the vector $\langle k)1$ as shown in Figure (2)

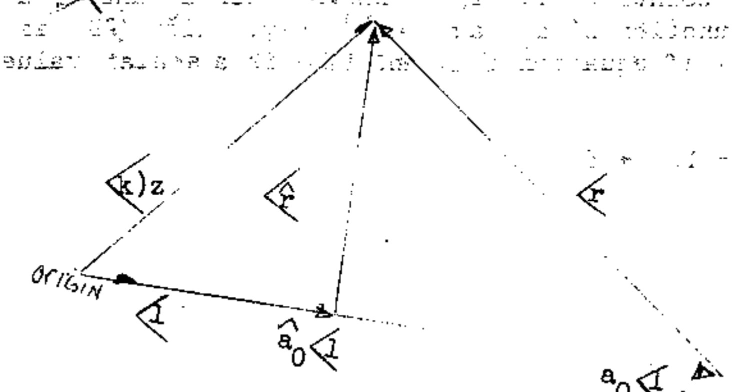

Figure (2)

The symbol $\hat{a}_0$ (hat) as before designates the value of a_0 which yields a point of symmetry in one-space or perpendicularity in k-space.

If we take the inner-product of (15) with $1\rangle$ we obtain

$$\langle z1\rangle = a_0\langle 11\rangle + \langle r1\rangle \tag{16}$$

and if

$$\langle r1\rangle = 0 \tag{17}$$

then

$$\frac{1}{k}\langle z1\rangle = a_0. \tag{18}$$

We can also look at the orthogonality condition from the view point of the minimum magnitude of the vector $\langle r$.

Solving equation (15) for $\langle r$ we have

$$\langle k)z - a_0\langle k)1 = \langle k)r \tag{19}$$

Transposing (19) into a column vector

$$z(k\rangle - 1(k\rangle\, a_0 = r(k\rangle \tag{20}$$

Taking the "inner-product" of (19) and (20)

$$\langle r\, r\rangle = [\langle z - a_0\langle 1][z\rangle - 1\rangle\, a_0] = \langle zz\rangle - 2\langle z1\rangle a_0 + a_0^2\langle 11\rangle \tag{21}$$

Define the scalar valued function $\langle r\ r\rangle$ which is the square of the magnitude of the residual vector in k-space $\langle k)r$ as

$$\phi(a_0) = \langle r\ r\rangle = r_1^{\,2} + r_2^{\,2} + \dots r_k^{\,2} \quad , \tag{22}$$

If we consider the points $z_1 \dots z_k$ as known, then a_0 and $\langle r$ are variables, hence by (19) $\langle r$ is a function of a_0 (a scalar) only. Also (22) is considered a function of a_0. Now ϕ of equation (22) and (21) is a scalar valued quadratic function in a_0, that is

$$\phi(a_0) = ka_0^{\,2} + ba_0 + c \tag{23}$$

where

$$\begin{aligned} k &= \langle 1\ 1\rangle \\ b &= -2\langle z\ 1\rangle \\ c &= \langle z\ z\rangle \end{aligned} \tag{24}$$

The quadratic function of equation (23) plots as shown in Figure (3)

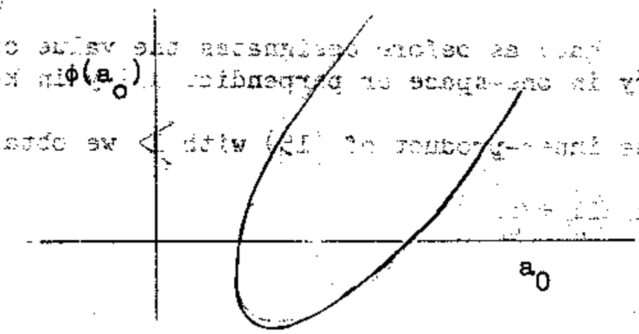

Figure (3)

We can minimize the square of the magnitude of the residual vector $\langle r$ with respect to a_0 of equation (21) by two methods:

1. partial derivatives or gradient approach
2. orthogonal projection in k-space.

We should keep in mind also that finding a point of symmetry in one-space, that is finding the geometrical (unweighted in this case) mean, is equivalent to finding an orthogonal vector in k-space.

<u>Method 1</u>: Partial Derivative Approach

Taking the partial derivative of (21) with respect to a_0 we obtain

$$\frac{\partial\phi}{\partial a_0} = -2\ \langle z 1\rangle + 2a_0 k = (2k)a_0 + b \tag{25}$$

SECTION II
SCALAR AND VECTOR MEAN

i Scalar Mean

Consider a sequence of points on a line (observations or what-ever) as shown in Figure (1) and described by

$$z_j \quad j = 1 \cdot \cdot \cdot k \tag{1}$$

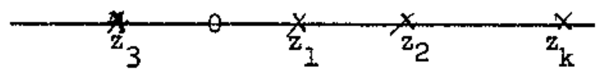

Figure (1)

Equation (1) represents k vectors in one-space and following Fisher's suggestion we can represent the k vectors in one-space as one row vector (or column vector) in k space by

$$(z_1, z_2, \cdot \cdot \cdot z_k) = \langle k) z \tag{2}$$

We are always free to "translate" to a new origin say a_0, hence

$$z_j = a_0 + r_j \qquad j = 1 \cdot \cdot \cdot k \tag{3}$$

where r_j is the position vector of the j^{th} point with respect to the point a_0 as a new origin. We may call r_j the "residual" vector to use the jargon of classical curve-fitting.

We can add the k points in one space of equation (3) as

$$z_1 + z_2 + \ldots + z_k = (a_0 + r_1) + a_0 + r_2 + \ldots + a_0 + r_k \tag{4}$$

or

$$z_1 + z_2 + \ldots + z_k = (a_0 + a_0 + \ldots + a_0) + r_1 + r_2 + \ldots r_k \tag{5}$$

or

$$z_1 + z_2 + \ldots + z_k = a_0 k + r_1 + \ldots + r_k \tag{6}$$

Expressing the sum of equation (6) as an inner-product

$$\sum_{j=1}^{k} z_j = z_1 + \ldots + z_k = (z_1, z_2, \ldots z_k) \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tag{7}$$

$$= \langle k(z \; 1(k\rangle = a_0 \langle 1 \rangle + \langle r 1 \rangle \tag{8}$$

where

$$1(k\rangle = \begin{pmatrix} 1 \\ 1 \\ \cdot \\ 1 \end{pmatrix} \tag{9}$$

and

$$\langle 1 1 \rangle = k \tag{10}$$

will be referred to as the <u>sum vector</u> in k-space.

If the sum of the residuals equals zero, i.e.,

$$r_1 + \ldots + r_k = \langle k) r \; 1(k\rangle = 0 \tag{11}$$

then

$$\frac{1}{k}(z_1 + \ldots + z_k) = \hat{a}_0 \tag{12}$$

and $\hat{a}_0$ is the conventional arithmetic mean or unweighted point of symmetry for the k points in the one-space.

In k-space equation (2) becomes

$$\langle k) z = (a_0, \ldots a_0) + (r_1, \ldots r_k) \tag{13}$$

or factoring out a_0

$$\langle k) z = a_0 (1 \ldots 1) + \langle k) r \tag{14}$$

or

$$\langle k) z = a_0 \langle k) 1 + \langle k) r \tag{15}$$

In k-space we have the geometrical vectors of Figure (2)

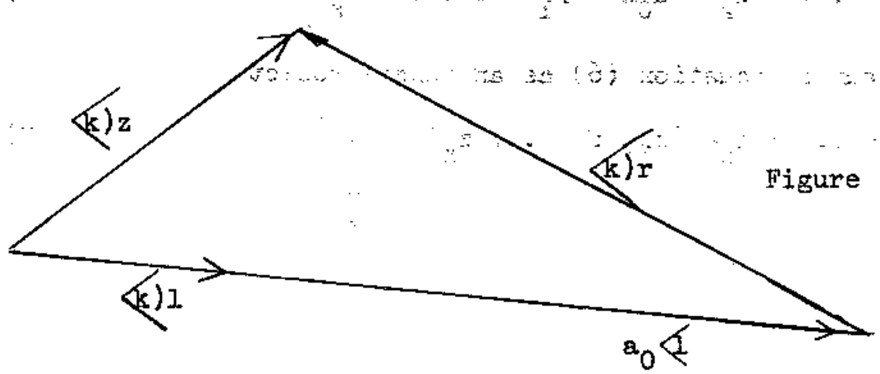

Figure (2) k-Space Vector

since

$$\langle 1 \; 1 \rangle = k \quad , \tag{26}$$

Equation (25) plots as a straight line as shown in Figure (4)

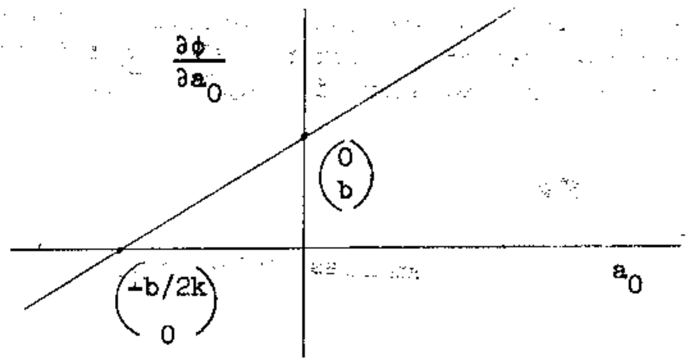

Figure (4)

Equating (25) to zero

$$\frac{\partial\phi}{\partial a_0} = 0 = - \langle z \; 1 \rangle + a_0 k \tag{27}$$

or

$$\hat{a}_0 = \frac{\langle z \; 1 \rangle}{k} \tag{28}$$

where $\hat{a}_0$ designates the solution to (27).

Observe that equation (23) is quadratic in a_0 and equation (27) is linear. The inner-product between the data vector $\langle k)z$ in k-space and the sum vector $\langle k)1$ is

$$\langle k)z \; 1(k \rangle = (z_1 \cdots z_k) \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = z_1 + \cdots + z_k \tag{29}$$

Clearly equation (28) with (29) tells us that the mean is a symmetry observation point (or origin) in one space.

Concluding by (28) and (29)

$$\hat{a}_0 = \frac{\langle z \; 1 \rangle}{k} = \frac{z_1 + \ldots + z_k}{k} \tag{30}$$

or the unweighted average, or mean.

Method 2: Orthogonal Projection

By mathematical geometrical-intuition it is clear that if the vector $\langle k)z$ is fixed and given, that as a_0 varies from $+\infty$ to $-\infty$ (that is, on the axis of $\langle k)1$) that the "closest approach" or smallest magnitude of the residual vector occurs when the residual vector is perpendicular to the one-space generated by the single vector $\langle k)1$ as shown in Figure (5).

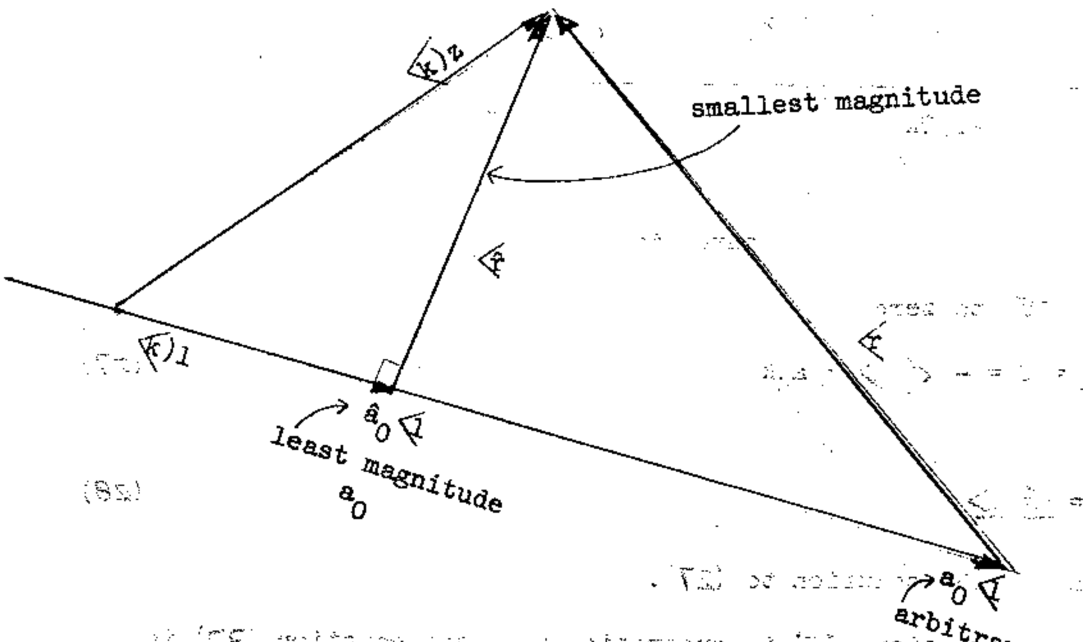

Fig. (5) Vectors In k-Space

If we select the value of the real variable a_0 to be that value such that the magnitude of the error vector is a minimum, then the residual vector $\langle r$ is perpendicular (orthogonal) to the sum vector $\langle 1$, hence

$$\langle \hat{r}\; 1\rangle = 0 \tag{31}$$

where $\hat{a}_0$ and $\langle \hat{r}$ designate the parameter and vector at this particular point.

Using (31) and taking "inner product" of (15) with $1\rangle$ we obtain

$$\langle z\; 1\rangle = \hat{a}_0 \langle 1\; 1\rangle + \langle \hat{r}\; 1\rangle \tag{32}$$

or

$$\frac{\langle z\; 1\rangle}{k} = \hat{a}_0 \tag{33}$$

which is the same result as arrived at via the partial derivative approach.

ii <u>Vector Mean</u>

Consider a given sequence of k column vectors in p-space as shown in Figure (1) and represented in equation (1)

$$z(p\rangle_j, \qquad j = 1 \cdots k \tag{1}$$

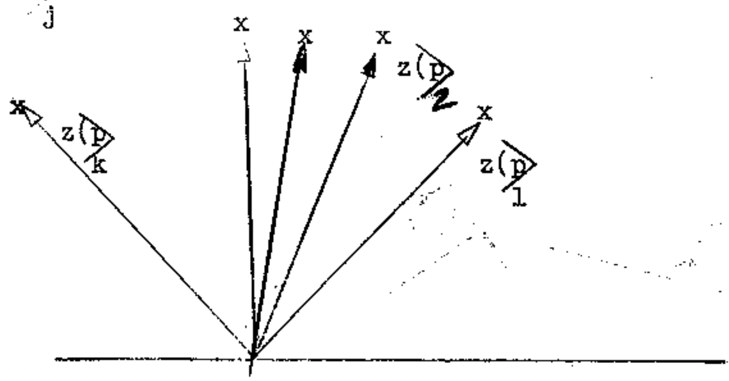

Figure (1) k Vector In p-Space

Equation (1) represents k vectors in p-space and extending Fisher's suggestion we can consider p vectors in k-space (k is the sample size).

Packaging the k vectors as a row-tuple of vectors we have

$$[z(p\rangle_1, \ldots, z(p\rangle_k] = \underset{pxk}{Z} \tag{2}$$

and upon re-partitioning the data-matrix of (2) into its row-space we have

$$\underset{pxk}{Z} = \begin{bmatrix} {}^{1}\langle k)z \\ \cdot \\ \cdot \\ \cdot \\ {}^{p}\langle k)z \end{bmatrix} \tag{3}$$

Equation (2) represents k column vectors in p-space and equation (3) represents p row vectors in k space as shown in Figure (2)

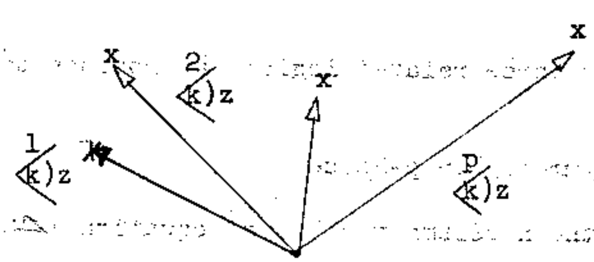

Figure (2) p Row Vectors In k-Space

We are always free to translate to any arbitrary new origin say $a(p\rangle_0$ in Fig. (3) where

$$z(p\rangle_j = a(p\rangle_0 + r(p\rangle_j \tag{4}$$

$$j = 1, \ldots, k$$

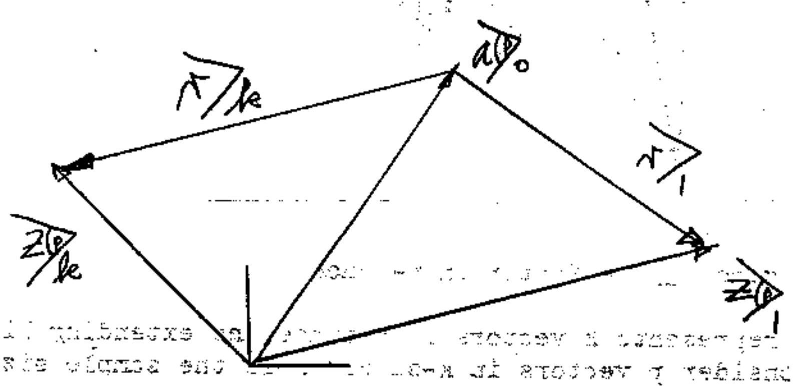

Fig (3) Translation To Arbitrary Origin $a\rangle_0$ in p-Space

Packaging (4) into a row k-tuple

$$\underset{pxk}{Z} = [a(p\rangle_0 \;\cdots\; a(p\rangle_0] + [r(p\rangle_1 \;\cdots\; r(p\rangle_k] \tag{5}$$

or factoring out $a\rangle_0$

$$\underset{pxk}{Z} = a(p\rangle_0 \langle k)1 + \underset{pxk}{R} \tag{6}$$

By equation (6) the data matrix Z is expressed as the sum of a dyad plus a pxk matrix.

We may consider three related logical approaches to the equations for the mean:

a. Point of Symmetry In p-Space

If we add the k column vectors of equation (4) Then

$$z\rangle_1 + z\rangle_2 + \ldots + z\rangle_k = a\rangle_0 + r\rangle_1 + \ldots + a\rangle_0 + r\rangle_k \tag{7}$$

or

$$z_1\rangle + \dots + z_k\rangle = a_0\rangle k + r_1\rangle + \dots + r_k\rangle \qquad (8)$$

If we translate to a new origin $a_0\rangle$ such that the sums of the "residual vectors" is zero, that is

$$r_1\rangle + \dots + r_k\rangle = 0\rangle \qquad (9)$$

then designate $a_0\rangle$ by a "hat" symbol, or

$$\frac{z_1\rangle + \dots + z_k\rangle}{k} = \hat{a}_0\rangle \qquad (10)$$

Equation (8) may be written in package form as

$$z_1\rangle + \dots + z_k\rangle = [z_1\rangle, \dots z_k\rangle] \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \underset{p\times k}{Z}\, 1(k\rangle \qquad (11)$$

Likewise equation (8) becomes

$$[z_1\rangle, \dots z_k\rangle]\, 1\rangle = a_0\rangle \langle 1\, 1\rangle + [r_1\rangle \dots r_k\rangle]\, 1\rangle \qquad (12)$$

or in matrix form as

$$Z\, 1\rangle = a_0\rangle \langle 1\, 1\rangle + R\, 1\rangle . \qquad (13)$$

Equation (13) may also be obtained by post-multiplying equation (6) on the left by $1\rangle$.

The condition of equation (9) that $a_0\rangle$ be a point of symmetry may be written as

$$r_1\rangle + \dots + r_k\rangle = [r_1\rangle, \dots r_k\rangle] \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \underset{p\times k}{R}\, 1(k\rangle = 0(p\rangle \ , \qquad (14)$$

In vector space language equation (14) says that the sum vector $1\rangle$ lies in the null-space of the matrix R. If we partition R into its row space we see that equation (14) becomes

$$R\,1\rangle = \begin{bmatrix} {}^{1}\langle k)r \\ \vdots \\ {}^{p}\langle k)r \end{bmatrix} 1(k\rangle = \begin{bmatrix} {}^{1}\langle k)r\;1\rangle \\ \vdots \\ {}^{p}\langle k)r\;1\rangle \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (15)$$

which states that all p of the k-dimensional row vectors ${}^{p}\langle k)r$ are perpendicular to the vector $1(k\rangle$. Observe that null-space notions and simultaneous orthogonality are equivalent. *Now* by (13) and (14)

$$z\rangle_{1} + \ldots + z\rangle_{k} = a\rangle_{0}\, k \qquad (16)$$

hence

$$a\rangle_{0} = \frac{z\rangle_{1} + \ldots + z\rangle_{k}}{k} = \frac{\Sigma\, 1(k\rangle}{k} \qquad (17)$$

which is the <u>unweighted mean</u> or <u>arithmetic mean</u>, a point of symmetry.

<u>b. Orthogonal Projections In k-Space</u>

Consider the row-space partitioning of equation (6)

$$\begin{bmatrix} {}^{1}\langle k)z \\ \cdot \\ \cdot \\ \cdot \\ {}^{p}\langle k)z \end{bmatrix} = \begin{bmatrix} a^{1}_{\cdot 0}\langle k)1 \\ a^{2}_{\cdot 0}\langle k)1 \\ \vdots \\ a^{p}_{\cdot 0}\langle k)1 \end{bmatrix} + \begin{bmatrix} {}^{1}\langle k)r \\ \cdot \\ \cdot \\ \cdot \\ {}^{p}\langle k)r \end{bmatrix} \qquad (18)$$

The sequence of row vectors of equation (18) is shown in Figure (4).

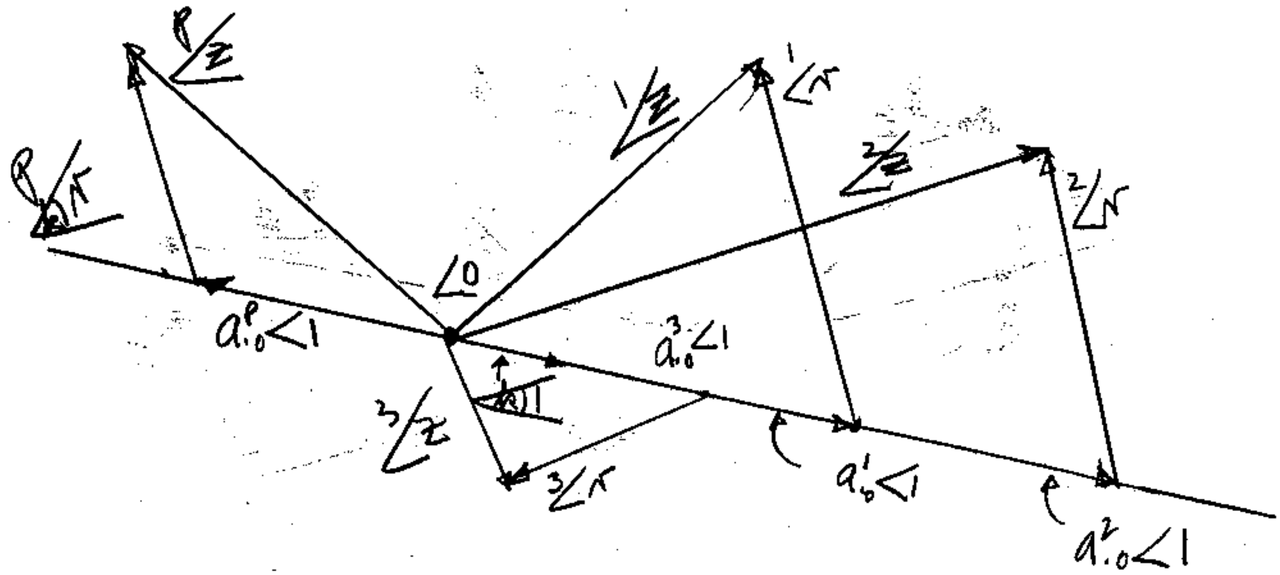

Fig. (4) Sample-Space of Dimension k.

The vector condition of a point of symmetry in p-space of equation (9) may be written in matrix form as

$$0\rangle = f(p\rangle_1 + \dots + f(p\rangle_k = [f(k\rangle_1, \dots f(p\rangle_k] \begin{pmatrix}1\\1\\ \cdot\\ \cdot\\ \cdot\\1\end{pmatrix} = \hat{R}_{p\times k}\, 1(k\rangle \qquad (19)$$

Partition the matrix of equation (19) into its row space, then

$$\hat{R}1\rangle = \begin{bmatrix} {}^{1}\langle k)f \\ \cdot \\ \cdot \\ \cdot \\ {}^{p}\langle k)f \end{bmatrix} 1(k\rangle = \begin{bmatrix} {}^{1}\langle k)f\ 1(k\rangle \\ \cdot \\ \cdot \\ \cdot \\ {}^{p}\langle k)f\ 1(k\rangle \end{bmatrix} = \begin{pmatrix}0\\ \cdot\\ \cdot\\ \cdot\\0\end{pmatrix}. \qquad (20)$$

Clearly the outer-product of inner product interpretation of (20) says that all p of the k-dimensional row vectors $\langle \hat{f}$ are orthogonal to the sum vector as shown in Figure (5)

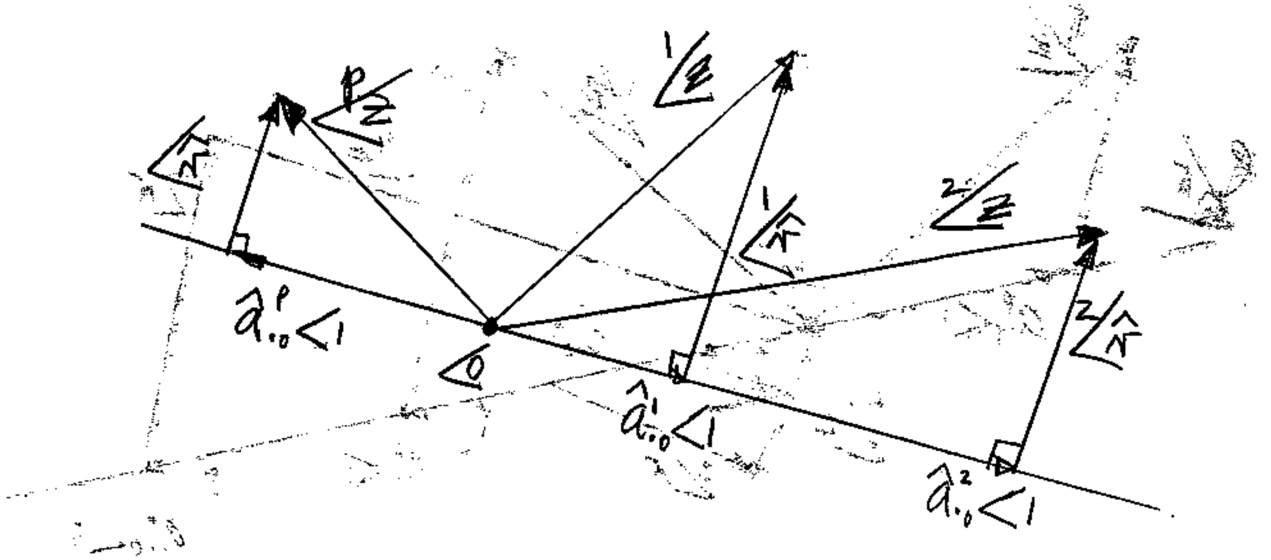

Fig (5) Orthogonality Conditions on k-Space

The row space partitioning and interpretation of equation (13) becomes

$$\underset{p\times k}{Z}\; 1(k\rangle = \hat{a}(p\rangle_{0}\, \langle k)1\; 1(k\rangle + \underset{p\times k}{\hat{R}}\; 1(k\rangle \qquad (21)$$

or

$$\begin{pmatrix} {}^{1}\langle z\,1\rangle \\ \vdots \\ {}^{p}\langle k)z\;1\rangle \end{pmatrix} = \hat{a}(p\rangle_{0}\; k, \qquad (22)$$

where

$$\underset{p\times k}{\hat{R}}\; 1(k\rangle = 0(p\rangle \qquad (23)$$

or

$$\begin{pmatrix} a^{1} \\ \cdot \\ \cdot \\ \cdot \\ a^{p} \end{pmatrix} = \frac{1}{k} \begin{pmatrix} {}^{1}\langle k)z\; 1(k\rangle \\ \cdot \\ \cdot \\ \cdot \\ {}^{p}\langle k)z\; 1(k\rangle \end{pmatrix} \qquad (24)$$

From a vector space standpoint we can say that the sum vector $\langle k)1$ lies in the null-space of the pxk matrix $\hat{R}$ when the orthogonal conditions of Figure (5) are met. Clearly the null-space notions and orthogonality of systems of vectors are equivalent concepts. In Figure (4) it is clear that all p of the k-dimensional vectors

$$a^{1}_{.0}\langle k)1,\ \ldots,\ a^{p}_{.0}\langle k)1$$

lie in a one-space. The p row-vectors

$$\overset{1}{\langle k)}z,\ \ldots,\ \overset{p}{\langle k)}z$$

can lie anywhere in the full k-space. All $\langle r$ vectors are perpendicular to the $\langle k)1$ vector.

Equation (17) and Equation (24) are equivalent statements at the matrix level since

$$\underset{pxk}{Z}\ 1(k\rangle\ \frac{1}{k} = \hat{a}(\underset{0}{p}\rangle \tag{25}$$

and (17) is the column space partitioning and equation (24) is the row-space partitioning interpretation.

C. Partial Derivative Or Minimum Sample Variance

The data-matrix sum-decomposition of equation (6), solving for R, is

$$\underset{pxk}{R} = \underset{pxk}{Z} - a(\underset{0}{p}\rangle\langle k)1 \tag{26}$$

Transposing (26)

$$\underset{kxp}{R^{T}} = \underset{kxp}{Z^{T}} - 1(k\rangle\langle p)a_{0} \tag{27}$$

Form the pxp symmetric matrix product of (26) and (27)

$$\underset{pxp}{RR^{T}} = [Z - \underset{0}{a}\rangle\langle 1]\ [Z^{T} - 1\rangle\langle \underset{0}{a}]\ . \tag{28}$$

Multiplying out the terms of (28)

$$RR^{T} = ZZ^{T} - Z\ 1(k\rangle\langle a_{0} - \underset{0}{a}\rangle\langle 1\ Z^{T} + \underset{0}{a}\rangle\langle \overset{0}{a}\ k \tag{29}$$

Equation (29) is the multivariable analog of equation (21) for the scalar case. In the scalar case we obtained the sums of the squares of the residuals $\langle k)r\ r(k\rangle$ as a scalar-valued function of a scalar argument a_0. In the vector case of equation (29) we have a matrix-valued function RR^T of a vector argument $a(\underset{0}{p}\rangle$.

We can obtain a scalar valued function of the vector variable $a(p\rangle_0$ by taking the trace of equation (29). Geometrical interpretations and properties of the trace are given in Appendix (A).

$$\mathrm{tr}(RR^T) = \mathrm{tr}(ZZ^T) - 2\langle k)1\, Z^T a(p\rangle_0 + k\langle a_0\, a\rangle_0 \qquad (30)$$

Define the scalar

$$\phi(a\rangle_0) = \mathrm{tr}(RR^T) \qquad (31)$$

Equation (31) is the multi-variable analog of the quadratic expression of equation (2-23).

Conventional minimization theory tells us that the differential of ϕ is

$$d\phi = \left\langle \frac{\partial\phi}{\partial a_0} \, da\right\rangle_0 \qquad (32)$$

The gradient of ϕ is a row vector (see Appendix (C),

$$\left\langle \frac{\partial\phi}{\partial a_0}\right. = \underline{-2\langle k)1\, Z^T + 2k\langle p)a} \qquad (33)$$

Equating the vector equation of (33) to the zero vector and designating the solution vector $a\rangle_0$ by hat

$$\langle 0 = -\langle 1\, Z^T + k\langle \hat{a}_0 \qquad (34)$$

or

$$\langle p)\hat{a}_0 = \frac{1}{k}\langle k)1\, Z^T \qquad (35)$$

Clearly if we transpose (35)

$$\hat{a}(p\rangle_0 = Z\, 1\rangle \frac{1}{k} \qquad (36)$$

we obtain equation (17).

The scalar ϕ as shown in Appendix (A) is the sums of the squares of the magnitudes of all of the residual vectors, either in p space or in k space, that is

$$\phi = \langle p)r\, r(p\rangle_1 + \ldots + \langle p)r\, r(p\rangle_k = \langle k)r\, r(k\rangle_1 + \ldots + \langle k)r\, r(k\rangle_p \qquad (37)$$

and is "related" to the sample variance.

SECTION III

NOISY LINE

Suppose now that we have the equation

$$\begin{pmatrix} z^1 \\ z^2 \end{pmatrix}_k = \begin{pmatrix} x \\ y \end{pmatrix}_k = \begin{pmatrix} x \\ a_0 + a_1 x + e \end{pmatrix}_k \qquad (1)$$

which may describe a process shown in Figure (1).

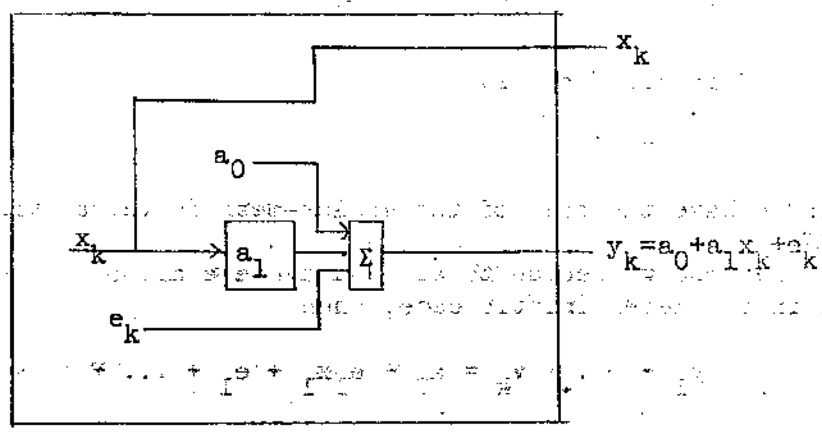

Fig. (1) Noisy Process Blocks

Equation (1) or the graph of the process variables of Fig. (1) may plot as Fig. (2)

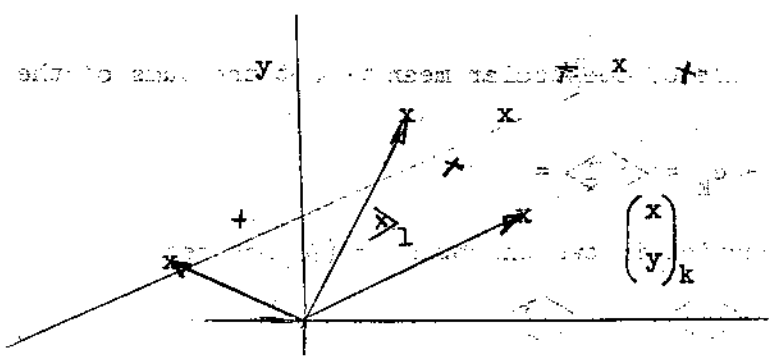

Fig. (2) Noisy Process Graph

Equating second elements of Equation (1)

$$y_k = a_0 + a_1 x_k + e_k \tag{2}$$

and writing as an inner-product

$$y_k = (a_0, a_1) \begin{pmatrix} 1 \\ x \end{pmatrix}_k + e_k \tag{3}$$

$$= \langle 2)a\ x(2 \rangle_k + e_k \quad . \tag{4}$$

If in Equation (2)

$$a_1 = 0$$

then we have the case of the scalar-mean in the previous section.

Suppose we decide to add and average all of the equations of (2) as we did in the deterministic case, then

$$y_1 + \dots + y_k = a_0 + a_1 x_1 + e_1 + \dots + a_0 + a_1 x_k + e_k \tag{5}$$

$$y_1 + \dots + y_k = k\ a_0 + a_1(x_1 + \dots + x_k) + e_1 + \dots + e_k \tag{6}$$

or

$$\langle y\ 1 \rangle = a_0 \langle 1\ 1 \rangle + a_1 \langle x\ 1 \rangle + \langle e\ 1 \rangle \tag{7}$$

Equation (7) is a scalar equation in three variables namely a_0, a_1 and $\langle e\ 1 \rangle$.

If, as in the case of the scalar mean we set the sums of the errors equal to zero or

$$e_1 + \dots + e_k = \langle e\ 1 \rangle = 0 \tag{8}$$

then we have one equation in two unknowns or (7) becomes

$$\langle y\ 1 \rangle = a_0 \langle 1\ 1 \rangle + a_1 \langle x\ 1 \rangle \tag{9}$$

$$= (a_0, a_1) \begin{bmatrix} \langle 1\ 1 \rangle \\ \langle x\ 1 \rangle \end{bmatrix}$$

$$= (a_0, a_1) \begin{bmatrix} \langle 1 \\ \langle k) x \end{bmatrix} 1\rangle \tag{10}$$

$$\langle y \; 1\rangle = \langle 2) a \underset{2xk}{F} \; 1(k\rangle \quad . \tag{11}$$

It is clear from Equation (9) that we need one additional condition besides the condition of Equation (8).

Suppose we package the k equations of (2) into a row vector

$$(y_1, \; \dots \; y_k) = a_0(1 \; \dots \; 1) + a_1(x_1, \; \dots \; x_k) \tag{12}$$

or

$$\langle y = a_0 \langle 1 + a_1 \langle x + \langle e \tag{13}$$

also

$$\langle k) y = (a_o, \; a_1) \begin{bmatrix} \langle k) 1 \\ \langle k) x \end{bmatrix} + \langle e \tag{14}$$

$$\langle k) y = \langle 2) a \underset{2xk}{F} + \langle e. \tag{15}$$

Using the same method of post-multiplying (15) by the transpose of the known data matrix $\underset{xk2}{F^T}$ as done in Section I, we have

$$\langle k) y \; F^T = \langle 2) a \underset{(2xk)}{F} \underset{(kx2)}{F^T} + \langle e F^T. \tag{16}$$

If we impose the two conditions that $\langle e$ be perpendicular to the two vectors $1(k\rangle$ and $x(k\rangle$, that is

$$\langle k) e \underset{kx2}{F^T} = \langle 2) 0 \tag{17}$$

or

$$\langle e [1\rangle, x\rangle] = [\langle e \; 1\rangle, \langle e \; x\rangle] = (0, \; 0) \tag{18}$$

then

$$\langle y \; F^T \; (F \; F^T)^{-1} = \langle \hat{a} \tag{19}$$

when $F \; F^T$ has an inverse.

We shall show that when the inverse exists and we have no constraints on the values that a_0 and a_1 can take on, that we can adjust a_0 and a_1 such that the two conditions of Equation (54)

$$\langle e\, 1\rangle = 0 \tag{20}$$

$$\langle e\, x\rangle = 0 \tag{21}$$

are satisfied.

Before doing so, let us observe that the inner-product (scalar) equation of (21)

$$\langle y\, 1\rangle = a_0 \langle 1\, 1\rangle + a_1 \langle x\, 1\rangle + \langle e\, 1\rangle \tag{22}$$

and the inner-product equation formed by operating on Equation (21) with the data vector $x\rangle$ is

$$\langle y\, x\rangle = a_0 \langle 1\, x\rangle + a_1 \langle x\, x\rangle + \langle e\, x\rangle \tag{23}$$

and the two scalar equations above can be solved to yield the two unknowns a_0 and a_1 if we set

$$\langle e\, 1\rangle = 0 = \langle e\, x\rangle \ . \tag{24}$$

The two equations of (22) and (23) namely

$$\langle y\, 1\rangle = \hat{a}_0 \langle 1\, 1\rangle + \hat{a}_1 \langle x\, 1\rangle \tag{25}$$

$$\langle y\, x\rangle = \hat{a}_0 \langle 1\, x\rangle + \hat{a}_1 \langle x\, x\rangle$$

become, in row vector form

$$\langle y [1\rangle, x\rangle] = (\hat{a}_0, \hat{a}_1) \begin{bmatrix} \langle 1 \\ \langle x \end{bmatrix} [1\rangle, x\rangle] \tag{26}$$

or

$$\langle y\, F^T = \langle \hat{a}\, F\, F^T \tag{27}$$

or

$$\langle y\, F^T (F\, F^T)^{-1} = \langle \hat{a} \tag{28}$$

which is Equation (19) showing that many arguments lead to the same results.

Consider the matrix

$$F\,F^T = \begin{pmatrix} \langle 1\;1\rangle & \langle 1\;x\rangle \\ \langle x\;1\rangle & \langle x\;x\rangle \end{pmatrix} \tag{29}$$

and its inverse

$$(F\,F^T)^{-1} = \begin{pmatrix} \langle x\;x\rangle & -\langle 1\;x\rangle \\ -\langle x\;1\rangle & \langle 1\;1\rangle \end{pmatrix} \frac{1}{\det F\,F^T} \tag{30}$$

where the determinent is

$$\Delta = \det\,(F\,F^T) = \langle 1\;1\rangle\langle x\;x\rangle - \langle 1\;x\rangle\langle x\;1\rangle \tag{31}$$

$$= \langle x\;x\rangle k - (\langle 1\;x\rangle)^2 \tag{32}$$

hence

$$\langle \hat{a} = \langle y\;[1\rangle,\;x\rangle] \begin{bmatrix} \langle x\;x\rangle & -\langle 1\;x\rangle \\ \langle x\;1\rangle & k \end{bmatrix} \frac{1}{\Delta} \tag{33}$$

or

$$\langle \hat{a} = [\langle y\;1\rangle,\;\langle y\;x\rangle] \begin{bmatrix} \langle x\;x\rangle & -\langle 1\;x\rangle \\ -\langle x\;1\rangle & k \end{bmatrix} \frac{1}{\Delta} \tag{34}$$

or

$$\langle \hat{a} = \frac{1}{\Delta}\,[\langle y\;1\rangle\langle x\;x\rangle - \langle y\;x\rangle\langle x\;1\rangle,\; -\langle y\;1\rangle\langle 1\;x\rangle + k\,\langle y\;x\rangle] \tag{35}$$

Equating elements of the two element row vectors

$$\hat{a}_0 = \frac{\langle 1\;y\rangle\langle x\;x\rangle - \langle y\;x\rangle\langle x\;1\rangle}{\langle x\;x\rangle k - (\langle 1\;x\rangle)^2} \tag{36}$$

$$\hat{a}_1 = \frac{-\langle y\;1\rangle\langle 1\;x\rangle + k\,\langle y\;x\rangle}{\langle x\;x\rangle k - (\langle 1\;x\rangle)^2} \tag{37}$$

The above equations are called the Normal Equations.

For those readers who are used to the say it with sigma summation sign of the ancient Greeks equation (36) can be written as

$$\hat{a}_o = \frac{\Sigma y_j \;\; \Sigma x_j^{\,2} - (\Sigma y_j\;x_j)\;\Sigma x_j}{k\,\Sigma x_j^{\,2} - (\Sigma\;x_j)^2} \tag{38}$$

and

$$\hat{a}_1 = \frac{k\Sigma y_j\, x_j - (\Sigma y_j)(\Sigma x_j)}{k\Sigma x_j^{\,2} - (\Sigma x_j)^2} \tag{39}$$

At this point, I draw upon the invitation of Edgar Lorch in his chapter on "The Spectral Theorem" who invites the reader to show discouragment over the complexity of his calculations and to long for a development freeing the reader of too many subscripts, too many repeated indices, and other instruments of mental torture. (R. C. Buck, page 94).

Equations (38) and (39)are called the <u>normal form</u> of the least squares equations. Clearly the two scalar equations of (36) and (37) are the inner-product version of the normal equations.

<u>GEOMETRY IN THE TWO DIMENSIONAL PARAMETER SUBSPACE OF k-SPACE.</u>

The single vector equation of (13)

$$\langle k)y = a_0 \langle k)1 + a_1 \langle k)x + \langle k)e \tag{40}$$

has the geometric model of Figure (3)

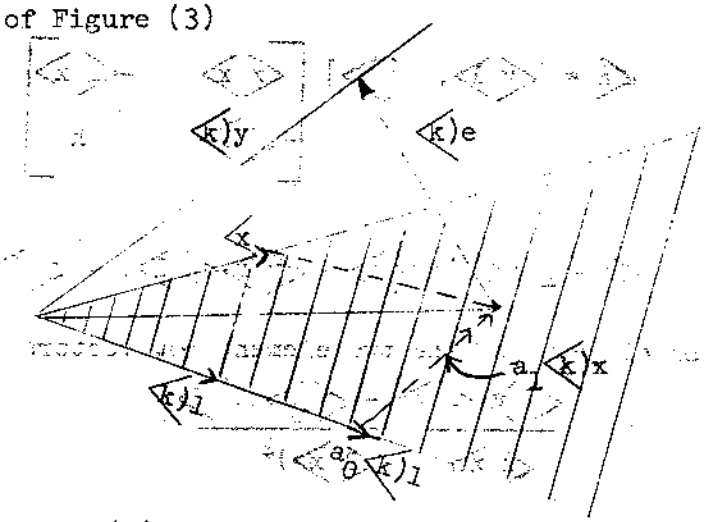

Figure (3) Vector In k-Space

The two conditions of orthogonality of Equations (20) and (21) require the $\langle e$ vector to be perpendicular to the two dimensional subspace spanned by the two base vectors $\langle 1$ and $\langle x$ (so long as $\langle x$ is not parallel to the $\langle 1$ vector). The least squares geometry is shown in Figure (4).

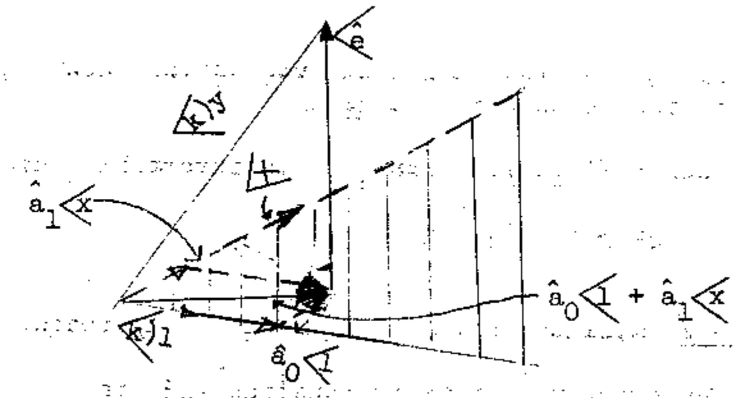

Fig. (4) Orthogonal Vectors in k-Space

The previous results are intuitive paper-talk. We shall show by two methods that the parameters (a_0, a_1) are the "best" with respect to unweighted least squares:

<u>Method 1</u>. Algebraic Approach via Orthogonal Projections in Sample-Size Space.

<u>Method 2</u>. Minimum Variance via Gradient Approach.

<u>Method 1</u>. Algebraic Approach via Orthogonal Projects in Sample-Size Space.

The error vector $(\langle k)e \neq \langle 0)$ in k-space has minimum magnitude when it is perpendicular to the sub-space generated by $\langle 1$ and $\langle x$. The truth of the statement is shown in appendix (B) on orthogonal projections. The picture is that shown in Figure (4).

By Equation (15)

$$\langle y = \langle aF + \langle e \tag{41}$$

By Equation (14)

$$F = \begin{bmatrix} \langle k)1 \\ \langle k)x \end{bmatrix} \tag{42}$$

$$F^T = [1(k\rangle, x(k\rangle] \tag{43}$$

and

$$\langle y\, F^T = \langle a\, F\, F^T + \langle eF^T \tag{44}$$

but for $\langle a$ such that

$$\langle e\, F^T = [\langle e\, 1\rangle, \langle e\, x\rangle] = [0, 0] \tag{45}$$

designate as $\langle \hat{a}$. By the orthogonal projection theory of appendix B the magnitude of the $\langle e$ vector is minimum.

When $F\, F^T$ is full rank, it is invertible, hence

$$\langle y\, F^T (F\, F^T)^{-1} = \langle \hat{a} \tag{46}$$

<u>Method 2</u>. Minimum Variance via Gradient Approach.

The error vector $\langle e$ by Equation (13) is

$$\langle e = \langle y - a_0 \langle 1 - a_1 \langle x = \langle y - \langle aF \tag{47}$$

and transposing (47)

$$e\rangle = y\rangle - 1\rangle a_0 - x\rangle\, a_1 = y\rangle - F^T a\rangle \tag{48}$$

The dyadic product of (47) and (48) yields

$$e\rangle\langle e = (y\rangle - F^T a\rangle)(\langle y - \langle aF) \tag{49}$$

or:

$$e\rangle\langle e = y\rangle\langle y - y\rangle\langle aF - F^T a\rangle\langle y + F^T a\rangle\langle aF \tag{50}$$

Equation (50) is "related" to the sample variance, the trace of (50) is

$$\text{trace}\; e\rangle\langle e = \langle e\; e\rangle \tag{51}$$

and by appendix (A-9) or (47) and (48) (inner-product)

$$\langle e\; e\rangle = \langle y\; y\rangle - \langle yF^T a\rangle - \langle aFy\rangle + \langle aF\; F^T\; a\rangle \tag{52}$$

It is clear that $\langle e\; e\rangle$ is the sums of the squares of the errors, that is

$$\phi(a\rangle) = \langle e\; e\rangle = e_1\, e^1 + e_2\, e^2 + \dots + e_k\, e^k \tag{53}$$

or the square of the magnitude of the error vector in k-space.

The differential of (52) by Appendix C is

$$d \;= \left\langle \frac{\partial\phi}{\partial a}\, da \right\rangle \tag{54}$$

$$= d\langle y\; y\rangle - 2d(\langle yF^T a\rangle) + d(\langle aFF^T a\rangle) \tag{55}$$

By Equation (C-33)

$$\langle\frac{\partial}{\partial a}\left(\langle y F^T a\rangle\right) = \langle y F^T \tag{56}$$

$$\langle\frac{\partial}{\partial a}\left(\langle a F F^T a\rangle\right) = 2\langle a F F^T \tag{57}$$

Using (56) and (57) in (54)

$$d\phi = [-2\langle y F^T + 2\langle a F F^T]\, da\rangle \tag{58}$$

and equating gradient portion of Equation (58) to zero

$$\langle\frac{\partial\phi}{\partial a} = \langle 0 = -2\langle y F^T + 2\langle \hat{a} F F^T \tag{59}$$

or

$$\langle\hat{a} = \langle y F^T (F F^T)^{-1} \tag{60}$$

Equation (60) is the same result as arrived at via orthogonal projections.

SECTION IV

POLYNOMIAL FITTING

Suppose we have a sequence of k column vectors from a two-space, that is the k^{th} vector is

$$z(2\rangle_k = \begin{pmatrix} z^1 \\ z^2 \end{pmatrix}_k \quad . \tag{1}$$

The sequence of vectors of Equation (1) may represent observation vectors of a process in which we are measuring two variables, one independent and the second dependent, assummed to be functionally related and desire to fit a polynomial to the data, then

$$\begin{pmatrix} z^1 \\ z^2 \end{pmatrix}_k = \begin{pmatrix} x \\ y \end{pmatrix}_k = \begin{pmatrix} x \\ a_0 + a_1 x + \dots + a_{p-1} x^{p-1} + e \end{pmatrix}_k \tag{2}$$

where the entries in the left hand column vector are the "data" or measurement vector.

Equating second coordinates of Equation (2)

$$y_k = (a_0 + a_1 x + \dots a_{p-1} x^{p-1})_k + e_k \quad . \tag{3}$$

Equation (3) can be written such that the unknown parameters a_0, a_1, ... a_{p-1} are separated from the data as

$$y_k = (a_0, a_1, \dots, a_{p-1}) \begin{pmatrix} 1 \\ x \\ \vdots \\ x^{p-1} \end{pmatrix}_k + e_k \tag{4}$$

or

$$y_k = \langle p) a\ f(p\rangle_k + e_k \tag{5}$$

where

$$f(p\rangle_k \equiv \begin{pmatrix} 1 \\ x \\ \vdots \\ x^{p-1} \end{pmatrix} \tag{6}$$

Using (6) in (2)

$$\begin{pmatrix} z^1 \\ z^2 \end{pmatrix}_k = \begin{pmatrix} x \\ \langle p)a\ f(p\rangle + e \end{pmatrix}_k \tag{7}$$

The vector may be expressed in terms of the mean by Equation (II-ii-4) or in terms of the approximating function of Equation (2), hence

$$z(2\rangle_k = z(2\rangle_0 + r(2\rangle_k = \begin{pmatrix} x \\ \langle p)a\ f(p\rangle + e \end{pmatrix}_k \tag{8}$$

Packaging the sequence of two dimensional column vectors into a row k-tuple we obtain

$$\underset{2xk}{Z} = [z(2\rangle_1, \ldots\ z(2\rangle_k] = z(2\rangle_0 \langle k)1 + \underset{2xk}{R} \tag{9}$$

$$= \left[\begin{pmatrix} x \\ \langle a\ f\rangle + e \end{pmatrix}_1 \cdots \begin{pmatrix} x \\ \langle a\ f\rangle + e \end{pmatrix}_k \right] \tag{10}$$

or

$$\underset{2xk}{Z} = \begin{bmatrix} x_1 & \ldots & x_k \\ \langle a\ f\rangle_1 & \ldots & \langle a\ f\rangle_k \end{bmatrix} + \begin{bmatrix} 0, & & 0 \\ e_1 & \ldots & e_k \end{bmatrix} \tag{11}$$

Partitioning (11) into its row space

$$\begin{pmatrix} \langle k)x \\ \langle k)y \end{pmatrix} = \begin{bmatrix} \langle x \\ \langle a[f(p\rangle_1\ \ldots\ f(p\rangle_k] \end{bmatrix} + \begin{bmatrix} \langle 0 \\ \langle k)e \end{bmatrix} \tag{12}$$

Equating second element vectors in Equation (12)

$$\langle k)y = \langle p)a\ [f(p\rangle_1\ \ldots\ f(p\rangle_k] + \langle k)e \tag{13}$$

or matrix wise

$$\langle k)y = \langle p)a \underset{pxk}{F} + \langle k)e. \tag{14}$$

Equation (12) by (14) becomes

$$\underset{2xk}{Z} = \begin{bmatrix} \overset{1}{\langle k)} z \\ \overset{2}{\langle k)} z \end{bmatrix} = \underset{0}{z(2\rangle}\langle k) 1 + \underset{2xk}{R} = \begin{bmatrix} \langle k) x \\ \langle p) a \underset{pxk}{F} + \langle e \end{bmatrix} \qquad (15)$$

Post-multiplying (15) by the sum vector $1(k\rangle$ in k-space

$$\underset{2xk}{Z}\, 1(k\rangle = \begin{bmatrix} \overset{1}{\langle k)} z\ 1\rangle \\ \overset{2}{\langle z}\ \ 1\rangle \end{bmatrix} = \underset{0}{z(2\rangle} k + \underset{2xk}{R}\, 1(k\rangle \qquad (16)$$

$$= \begin{bmatrix} \overset{1}{\langle k)} z\ 1(k\rangle \\ \langle p) a \underset{pxk}{F}\ 1(k\rangle + \langle k) e\ 1\rangle \end{bmatrix}$$

If $\underset{0}{z(2\rangle}$ is the mean then by Equation (II-ii-25)

$$Z 1\rangle = \underset{0}{z(2\rangle}\, k + \begin{bmatrix} \overset{1}{\langle z\ 1\rangle} \\ \langle aF1\rangle + \langle e\ 1\rangle \end{bmatrix} \qquad (17)$$

Two methods for obtaining the "least squares fit" of the parameters $(a_0, a_1, \ldots a_{p-1})$ to the sequence of vectors are presented below:

Method 1. <u>Orthogonal Projections</u>

Consider the vectors of Equation (14)

$$\langle k) y = \langle p) a \underset{pxk}{F} + \langle k) e \qquad (18)$$

and partition the matrix F into its row space

$$\underset{pxk}{F} = \begin{bmatrix} \overset{1}{\langle k)} f \\ \overset{2}{\langle k)} f \\ \overset{3}{\langle k)} f \\ \vdots \\ \overset{p}{\langle k)} f \end{bmatrix} = \begin{bmatrix} \langle k) 1 \\ \langle k) x \\ \langle k) x^2 \\ \vdots \\ \langle k) x^{p-1} \end{bmatrix} \qquad (19)$$

Using (19) in (18)

$$\langle k)y = (a_0,\ a_1,\ \ldots,\ a_{p-1}) \begin{bmatrix} \langle k)1 \\ \langle k)x \\ \langle k)x^2 \\ \vdots \\ \langle k)x^{p-1} \end{bmatrix} + \langle k)e \tag{20}$$

or

$$\langle k)y = a_0 \langle k)1 + a_1 \langle x + a_2 \langle x^2 + \ldots + a_{p-1} \langle k)x^{p-1} + \langle k)e \tag{21}$$

We know by the orthogonal projection theorem of Appendix B that when there are no constraints on the variables a_0, a_1, ... a_{p-1} that the error vector has minimum magnitude when the vector $\langle e$ is orthogonal to the p dimensional sub-space spanned by the vectors $\langle 1, \langle x, \langle x^2, \ldots \langle x^{p-1}$. By Equation (18)

$$\langle k)y = \langle p)a \underset{p\mathrm{x}k}{F} + \langle k)e \tag{22}$$

Multiplying (22) by F^T

$$\langle yF^T = \langle aFF^T + \langle eF^T \tag{23}$$

If we designate the $\langle e$ vector which is perpendicular to the linear manifold spanned by $L[\langle 1, \langle x, \langle x^2, \ldots, \langle x^{(p-1)}]$ by $\langle \hat{e}$ and the corresponding parameter vector by $\langle \hat{a}$ then (25) becomes

$$\langle yF^T = \langle \hat{a}FF^T, \tag{24}$$

since

$$\langle \hat{e}F^T = \langle p)0. \tag{25}$$

Equations (24) and (25) say that we adjust the parameter vector $\langle p)a$ such that the vector $\langle e$ lies in the null space of F^T, which is the simultaneous orthogonality condition, that is

$$\langle eF^T = (\langle e\ 1\rangle,\ \langle e\ x\rangle,\ \ldots,\ \langle ex^{p-1}\rangle) = (0,\ \ldots,\ 0). \tag{26}$$

Equation (22) is a linear algebraic equation with $\langle y$ and F known, to solve for $\langle \hat{a}$. If the pxp matrix FF^T has an inverse we can solve for $\langle \hat{a}$ as

$$\langle \hat{a} = \langle yF^T(FF^T)^{-1} \tag{27}$$

If the data-matrix F is not of full rank (in this case p) then the symmetric matrix FF^T inverse does not exist and we can resort to the methods of the "psuedo-inverse" or "generalized-inverse" to be presented in a follow-up paper.

In fact, we can define the generalized inverse of full rank F as the kxp matrix F^+ as done in Equation (I-62).

$$\underset{kxp}{F^+} = \underset{kxp}{F^T}\underset{pxp}{(FF^T)^{-1}} \tag{28}$$

We see by Equation (28) that matrix "inversion" reduces to symmetric matrix inversion.

Equation (27) may also be written by (28) as

$$\langle y F^+ = \langle p)\hat{a} \tag{29}$$

If we partition the matrix F^+ into its row space as

$$\underset{kxp}{F^+} = \begin{bmatrix} \overset{1}{\langle f^+} \\ \vdots \\ \overset{k}{\langle f^+} \end{bmatrix} \tag{30}$$

then from a data-weighting standpoint one can consider p-dimensional parameter vector as a linear combination of the data points and the vector weights as

$$\langle p)\hat{a} = (y_1, \ldots, y_k) \begin{bmatrix} \overset{1}{\langle p)f^+} \\ \vdots \\ \overset{k}{\langle p)f^+} \end{bmatrix} \tag{31}$$

or

$$\langle p)\hat{a} = y_1 \overset{1}{\langle p)} f^+ + y_2 \overset{2}{\langle p)} f^+ + \ldots + y_k \overset{k}{\langle p)} f^+ \tag{32}$$

We have demonstrated an "n-tuple of interpretations" of the parameter estimation problem. The next question in computer-oriented society is how to iterate or **sequentially** compute the parameters as new observations roll in. For example,

$$\langle p)\hat{a}(k) = (\hat{a}_1(k) \; \ldots \; \hat{a}(k)_{p-1}) \tag{33}$$

describes the best estimate of the parameters based on k observations. We want a computable data processing scheme which up-dates the estimate when the (k+1)th observation is taken. At the matrix level we could also present the problem; given the pxk matrix F and the square fixed-size pxp symmetric matrix

$$\underset{p(k)p}{FF^T}$$

which is a function of the number of observations k how do we iterate the inversion process to solve Equation (27). These methods are developed in Section VI.

Method 2. <u>Partial Derivative Approach or Minimum Variance</u>.

This section presents the classical approach to the least squares problem, however, many modern books operate on vectors in the k-space rather than one space, hence the vector space approach via partial derivatives is presented here.

By Equation (18) we have

$$\langle k)y = \langle p)aF + \langle e \tag{34}$$

and solving for the error vector $\langle e$

$$\langle e = \langle y - \langle aF, \tag{35}$$

and transposing

$$e\rangle = y\rangle - F^T a\rangle. \tag{36}$$

The sums of the squares of the errors is the inner-product

$$\langle e\, e\rangle = [\langle y - \langle aF][y\rangle - F^T a\rangle] \tag{37}$$

We consider (37) a scalar valued function ϕ of the vector variable $a\rangle$ and upon multiplying the vectors of (37) we obtain

$$\phi(a\rangle) = \langle e\, e\rangle = \langle y\, y\rangle - 2\langle yF^T a\rangle + \langle aFF^T a\rangle \tag{38}$$

since the scalar

$$\langle yF^T a\rangle = \langle a\, F\, y\rangle. \tag{39}$$

Observe that the dyadic product of (35) and (36) yields

$$e\rangle_k\,{}_k\langle e = \begin{bmatrix} e^1e_1 & e^1e_2 & \cdots & e^1e_k \\ \vdots & & & \\ e^ke_1 & & & e^ke_k \end{bmatrix} \tag{40}$$

and that the trace of the rank one matrix of Equation (40) is the inner-product, that is by appendix (A-9)

$$\text{trace } e\rangle\langle e = \langle e\; e\rangle \;. \tag{41}$$

The trace of a matrix is a scalar invariant with respect to a change of basis. It can be seen by Equation (41) that when a matrix is a dyad, the trace is the commuted product, that is, the inner product of the outer product. Observe that cross-correlation type of information is obtained in the off-diagonal terms of Equation (40). A future report presenting the fundamentals of geometrical vector space approach to analysis of variance will be presented.

The main point to observe is that the matrix of Equation (40) is related to the sample variance matrix of a scalar process, and that to optimize or minimize the scalar of Equation (41) is a weak admission of the easy way out. Why not optimize some matrix valued measure?

If we take the differential of the inner-product by Appendix C we obtain

$$d\phi = \left\langle \frac{\partial \phi}{\partial a}\; da \right\rangle \tag{42}$$

and, as conventionally done, set the gradient of ϕ equal to zero, then

$$\left\langle \frac{\partial \phi}{\partial a} = \left(\frac{\partial \phi}{\partial a_1}, \frac{\partial \phi}{\partial a_2} \right) = \langle 0 \right. \tag{43}$$

Observe that

$$\left\langle \frac{\partial}{\partial a} \left(\langle y\; y\rangle \right) = \langle 0 \right. \tag{44}$$

$$\left\langle \frac{\partial}{\partial a} \left(\langle yF^T a\rangle \right) = \langle yF^T \right. \tag{45}$$

and

$$\left\langle \frac{\partial}{\partial a} \left(\langle aFF^T a\rangle \right) = 2\langle aFF^T \right. \tag{46}$$

Using (44) and (45) and (46) in (43)

$$\Big\langle \frac{\partial \phi}{\partial a} = \langle 0 = -\langle y F^T + \langle \hat{a} F F^T \tag{47}$$

or

$$\langle \hat{a} F F^T = \langle y F^T. \tag{48}$$

When the matrix FF^T is full rank then it is invertible and

$$\langle \hat{a} = \langle y F^T (FF^T)^{-1}. \tag{49}$$

Observe that the approach of setting the gradient to zero yielded the same result as the algebraic orthogonal projection result.

SECTION V

SOME PARAMETER OPTIMIZATION
PROBLEMS ON THE RANGE

Consider two simultaneous trackers with noise as shown in Fig. (1),

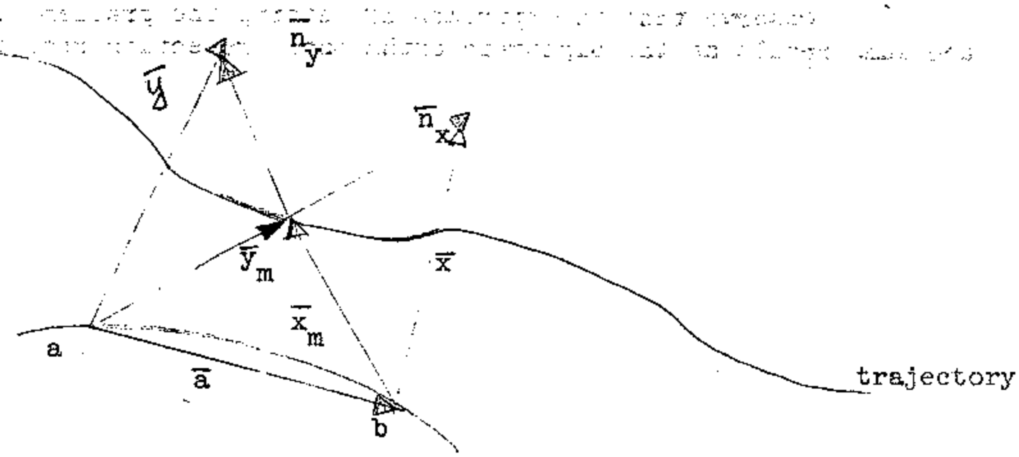

Fig. (1) Trajectory with Noise Vectors of Two
Radar Trackers

where $\bar{n}_y$ is the radar noise or error vector of tracker a and $\bar{n}_x$ is the radar noise vector of tracker b. The actual vehicle is on the trajectory. The vectors $\bar{x}_m$ and $\bar{y}_m$ are the actual vectors from the radar origins to the vehicle and satisfy

$$\bar{y}_m = \bar{a} + \bar{x}_m \quad . \tag{1}$$

The data vectors, or measurement vectors, are $\bar{x}$ and $\bar{y}$ and are corrupted by noise vectors expressed by the equations

$$\bar{x} = \bar{x}_m + \bar{n}_x \tag{2}$$

$$\bar{y} = \bar{y}_m + \bar{n}_y \tag{3}$$

Using (2) and (3) in (1)

$$\bar{y} = \bar{a} + \bar{x} + \bar{n}_y - \bar{n}_x. \tag{4}$$

Suppose that the observation vector $\bar{x}$ is given in radar rectangular coordinates

$$\bar{x} = (\bar{r}_1, \bar{r}_2, \bar{r}_3)^1 \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = {}^{1}\langle \overleftarrow{r}\; x \rangle \tag{5}$$

and that the observation vector $\bar{y}$ is given in rectangular coordinates

$$\bar{y} = {}^{2}\langle \overleftarrow{r}\; y \rangle \tag{6}$$

where the connection between the radar reference frames is

$${}^{1}\langle \overleftarrow{r} = {}^{2}\langle \overleftarrow{r}\; A \tag{7}$$

where A is an unknown matrix. Consider the position vector of site b with respect to site a, also unknown, and with coordinates to be estimated in the ${}^{2}\langle \overleftarrow{r}$ frame, that is

$$\bar{a} = {}^{2}\langle \overleftarrow{r}\; a \rangle. \tag{8}$$

The noise vectors referred to their own bases

$$\bar{n}_y = {}^{2}\langle \overleftarrow{r}\; n_y \rangle \tag{9}$$

$$\bar{n}_x = {}^{1}\langle \overleftarrow{r}\; n_x \rangle \tag{10}$$

Using (5) through (10) in Equation (4)

$${}^{2}\langle \overleftarrow{r}\; y \rangle = {}^{2}\langle \overleftarrow{r}\; a \rangle + {}^{1}\langle \overleftarrow{r}\; x \rangle + {}^{2}\langle \overleftarrow{r}\; n_y \rangle - {}^{1}\langle \overleftarrow{r}\; n_x \rangle \tag{11}$$

Operating on (11) with ${}^{2}\overrightarrow{r^*}\rangle$. the reciprical vectors of the ${}^{2}\langle \overleftarrow{r}$ bases

$$\begin{aligned} {}^{2}\overrightarrow{r^*}\rangle \cdot {}^{2}\langle \overleftarrow{r}\; y \rangle &= {}^{2}\overrightarrow{r^*}\rangle \cdot {}^{2}\langle \overleftarrow{r}\; a \rangle + \overrightarrow{r^*}\rangle \cdot \langle \overleftarrow{r}\; x \rangle \\ &+ {}^{2}\overrightarrow{r^*}\rangle \cdot {}^{2}\langle \overleftarrow{r}\; n_y \rangle - {}^{2}\overrightarrow{r^*}\rangle \cdot {}^{1}\langle \overleftarrow{r}\; n_x \rangle \end{aligned} \tag{12}$$

where the reciprical relations are

$$\overset{2}{r^*}\rangle \cdot \langle \overset{2}{r} = I \tag{13}$$

and by (7) and (13)

$$\overset{2}{r^*}\rangle \cdot \langle \overset{1}{r} = A. \tag{14}$$

Note that if the radar reference frames are ortho-normal then the reciprical vectors are the same (that is O.N. bases are self-reciprical and the inverse matrix of A is the transpose). We do not make the O.N. assumption here.

The vector-matrix equation of (12) is

$$y\rangle = a\rangle + Ax\rangle + n_y\rangle - An_x\rangle . \tag{15}$$

or

$$y\rangle = a\rangle + Ax\rangle + e\rangle \tag{16}$$

where the lumped error vector is

$$e\rangle = n_y\rangle - An_x\rangle . \tag{17}$$

If we consider k observations where k is arbitrary, then the kth observation is

$$y_k\rangle = a\rangle + Ax_k\rangle + e_k\rangle . \tag{18}$$

Equation (18) can be written as

$$y_k\rangle = [a\rangle, A] \begin{pmatrix} 1 \\ x\rangle \end{pmatrix}_k + e_k\rangle \tag{19}$$

or

$$y(3_k\rangle = \underset{3x4}{B} \, f(4_k\rangle + e(3_k\rangle \tag{20}$$

where the 3x4 matrix of parameters B is

$$\underset{3x4}{B} = [a(3\rangle, \underset{3x3}{A}] \tag{21}$$

and

$$f(4\rangle_k = \begin{pmatrix} 1 \\ x(3\rangle \end{pmatrix}_k \tag{22}$$

Packaging all k observations we have the 3xk data matrix

$$[y(3\rangle_1, \ldots\, y(3\rangle_k] = \underset{3xk}{Y} = [a\rangle_0, \ldots, a\rangle_0] + [Ay\rangle_1 \ldots\, Ay\rangle_k] + [e\rangle_1, \ldots, e\rangle_k] \tag{23}$$

The 3xk matrix $[a\rangle_0, \ldots, a\rangle_0]$ has rank one since it is a sequence of the same vector k times, hence

$$[a(3\rangle_0, \ldots\, a(3\rangle_0] = a(3\rangle_0\, (1, 1 \ldots 1)_k \tag{24}$$

Factoring out A in the matrix

$$[Ay\rangle_1, \ldots, Ay\rangle_k] = A[y\rangle_1, \ldots, y\rangle_k] = \underset{3x3}{A}\ \underset{3xk}{Y} \tag{25}$$

Using (24) and (25) in (23)

$$\underset{3xk}{Y} = a(3\rangle_o \langle k)1 + \underset{3x3}{A}\ \underset{3xk}{X} + \underset{3xk}{E} \tag{26}$$

also

$$\underset{3xk}{Y} = [a(3\rangle_0, \underset{3x3}{A}] \begin{bmatrix} \langle k)1 \\ \underset{3xk}{X} \end{bmatrix} + \underset{3xk}{E} \tag{27}$$

or

$$\underset{3xk}{Y} = \underset{3x4}{B}\ \underset{4xk}{F} + \underset{3xk}{E} \tag{28}$$

where

$$\underset{4xk}{F} = [f(4\rangle_1, \ldots, f(4\rangle_k] = \begin{bmatrix} \langle k)1 \\ \underset{3xk}{X} \end{bmatrix} \tag{29}$$

Equation (28) is a linear matrix equation in the unknown matrix of parameter B. The least squares estimate of B based on k observations is

$$\underset{3x4}{\hat{B}(k)} = \underset{(3xk)}{X} \; \underset{(kx4)}{F^T} \; \underset{(4x4)}{(FF^T)^{-1}}. \tag{30}$$

We desire to iterate B and will do so in Section VI.

Multi-Variable Polynomial Fitting

Some range problems consider three components of a missile position vector and three components of a velocity vector, or a six vector.

Suppose we are fitting time polynomials to q coordinates of a random vector, then as in section IV we have

$$y^1_{.k} = {}^1(a_0,\ a_1,\ \dots\ a_{p-1}) \begin{pmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^{p-1} \end{pmatrix}_k + e^1_k \tag{31}$$

$$\begin{matrix} . \\ . \\ . \\ . \end{matrix}$$

$$y^q_{.k} = {}^q\langle p)a \; f(p\rangle_k + e^q_{.k} \tag{32}$$

or packaging for the q coordinates

$$y(q\rangle_k = \underset{qxp}{A} \; f(p\rangle_k + e(q\rangle_k. \tag{33}$$

Packaging k-1 ovservations

$$[y(q\rangle_1,\ \dots\ y(q\rangle_{k-1}] = \underset{qxk-1}{Y} \tag{34}$$

$$Y = \underset{(qxp)}{A} \; \underset{pxk-1}{F} + \underset{qxk-1}{E} \tag{35}$$

Equation (35) is the linear equation in the unknown matrix of parameters A. The unweighted least squares estimate based on k-1 observations is

$$\underset{qxp}{\hat{A}(k-1)} = \underset{(qxk-1)}{Y} \; F^T(FF^T)^{-1}. \tag{36}$$

As before we desire to iterate A as the data rolls in.

SECTION VI

ITERATIVE COMPUTATION OF OPTIMAL PARAMETERS

This section develops the math models for the sequential computation of the least-squares estimates of parameters. The digital-computable model can estimate the parameters concurrently with the data-taking process. As new observations roll-in, the computer up-dates the estimates in a sequential manner.

When the data is full rank we can utilize a matrix inversion lemma of Householder's to up-date the estimates. When the data is not full rank we must resort to the generalized or "psuedo" inverse to do the updating.

The techniques are presented in three parts:

Part 1. Iterative Computation of the Mean.

Part 2. Iterative Computation of the Parameters of a p^{th} Degree Polynomial.

Part 3. Iterative Computation of a Matrix of Parameters.

Part 1. <u>Iterative Computation of the Mean.</u>

(i) <u>Sequential Computation of the Scalar Mean.</u>

Consider the scalar process of Equation (II-3)

$$z_j = a_0 + r_j \tag{1}$$

where we have k-1 observations or

$$\langle k-1)z = a_0 \langle k-1)1 + \langle k-1)r. \tag{2}$$

As shown in Equation (II-30)

$$\hat{a}(k-1) = \langle k-1)z\ 1(k-1\rangle \frac{1}{k-1}. \tag{3}$$

When we have k-observations we have

$$\frac{\langle k)z\ 1(k\rangle}{k} = \hat{a}_0(k) \tag{4}$$

Partition the k points such that

$$\langle k)z = (\langle k-1)z,\ z_k) \tag{5}$$

and partition

$$\langle k)1 = (\langle k-1)1,\ 1) \tag{6}$$

then

$$\langle k)z\ 1(k\rangle = [\langle k-1)z,\ z_k] \begin{bmatrix} 1(k-1\rangle \\ 1 \end{bmatrix} \tag{7}$$

or

$$\langle k)z\ 1(k\rangle = \langle k-1)z\ 1(k-1\rangle + z_k \tag{8}$$

Using Equation (3) in Equation (8)

$$\langle k)z\ 1(k\rangle = (k-1)\hat{a}(k-1) + z_k\ . \tag{9}$$

Using (4) in (9)

$$k\hat{a}(k) = (k-1)\ \hat{a}(k-1) + z_k \tag{10}$$

or

$$\hat{a}(k) = \frac{(k-1)}{k}\ \hat{a}(k-1) + \frac{z_k}{k} \tag{11}$$

which is the up-dating scheme. Note that the weighting factor of the latest data point is $\frac{1}{k}$ and that the past estimate of the mean $\hat{a}(k-1)$ based on $k-1$ points is weighted by the ratio $\frac{k-1}{k}$.

(ii) <u>Sequential Computation of the Vector Mean.</u>

A sequence of k-1 vectors from a p-dimensional column space can be represented as a row k-tuple as

$$[z(p\rangle_1,\ \ldots,\ z(p\rangle_{k-1}] = \underset{pxk-1}{Z} \equiv Z \tag{12}$$

and by Equation (II-ii-17) the mean is

$$\underset{pxk-1}{Z}\ 1(k-1\rangle \left(\frac{1}{k-1}\right) = \hat{a}(k-1)(p\rangle . \tag{13}$$

If we pack the k^{th} vector into the row-tuple, we have

$$[\,z\rangle_1, \ldots, z\rangle_{k-1}, z\rangle_k] = [\,Z_{pxk-1}, z(p\rangle_k] = Z_{pxk} \quad , \tag{14}$$

The mean-vector based on k observations is

$$Z_{pxk}\, 1(k\rangle \frac{1}{k} = \hat{a}(k)(p\rangle . \tag{15}$$

Partition the sum vector as in Equation (6), and use (14) in (15) then

$$[\,Z_{pxk-1}, z\rangle_k] \begin{pmatrix} 1(k-1\rangle \\ 1 \end{pmatrix} = Z_{pxk-1}\, 1(k-1\rangle + z\rangle_k \tag{16}$$

or

$$Z_{pxk-1}\, 1(k-1\rangle + z\rangle_k = k\, \hat{a}(k)(p\rangle \tag{17}$$

By Equation (13) in (17)

$$\hat{a}(k-1)(p\rangle \left(\frac{k-1}{k}\right) + z\rangle_k \left(\frac{1}{k}\right) = \hat{a}(k)(p\rangle \tag{18}$$

which is seen to be the vector analog of Equation (11) for the scalar case.

Part 2. Iterative Computation of the Parameters of a p^{th} Degree Polynomial.

Consider the polynomial equation of Equation (IV-14)

$$y_k = (a_0 + a_1 x + a_2 x^2 + \ldots + a_{p-1} x^{p-1})_k + e_k \tag{1}$$

or separating parameters

$$y_k = \langle p)a\; f(p\rangle_k + e_k \quad . \tag{2}$$

By Equation (IV-18) we have for k-1 observations

$$\langle k-1)y = \langle p)a\; F_{pxk-1} + \langle k-1)e \tag{3}$$

Conventional vector-space approach to least-squares tells us to multiply (3) by the transpose of the data matrix F^T and obtain

$$\langle k-1)y\; F^T_{(k-1)p} = \langle a\; F_{p(k-1)}\; F^T_{(k-1)xp} + \langle k-1)e\; F^T_{(k-1)xp} \tag{4}$$

It is easy to show (and is done in Section IV) via orthogonal projections in k-1 space (or via partial derivaties) that the optimal least squares fit occurs when

$$\langle k-1)\hat{e} \underset{(k-1)xp}{F^T} = \langle p)0 \tag{5}$$

hence we have when FF^T is invertible, or full rank,

$$\langle p)\hat{a}(k-1) = \langle k-1)y \underset{(k-1)xp}{F^T} \Big(\underset{p(k-1)p}{F\,F^T} \Big)^{-1}. \tag{6}$$

The row p-tuple of unweighted least-squares estimates of the p-parameters is based on k-1 observations, and it involves the computation of the inverse of the pxp matrix $\underset{pxp}{FF^T}$ which is a function of k-1 observations.

We shall show how to compute the unweighted least squares estimate of the parameters based on k observations (one more) as a function of the estimate based on k-1 when k-1 is greater than p and the data is of full rank.

Define the generalized inverse of F as the (k-1)xp matrix

$$\underset{(k-1)p}{F^+} = \underset{(k-1)xp}{F^T} \Big(\underset{p(k-1)p}{FF^T} \Big)^{-1}. \tag{7}$$

The reason for the terminology is clear for

$$\underset{p(k-1)}{F}\underset{(k-1)xp}{F^+} = \underset{p(k-1)}{F}\underset{(k-1)xp}{F^T} (FF^T)^{-1} = \underset{pxp}{I} \tag{8}$$

the one-sided product generates the pxp identity matrix (one-sided only).

Using (7) in (6)

$$\langle p)\hat{a}(k-1) = \langle k-1)y \underset{(k-1)xp}{F_{(k-1)}^{+}}. \tag{9}$$

Consider now the same expression as (3) for k observations instead of k-1

$$\langle k)y = \langle p)a \underset{(pxk)}{F} + \langle k)e . \tag{10}$$

Transpose F to the kxp matrix F^T and choose the parameters based on all k observations such that $\langle k)e$ lies in the null-space of $\underset{kxp}{F^T}$, that is

$$\langle k)\hat{e} \underset{kxp}{F^T} = \langle p)0 \tag{11}$$

and by analogy with (4)

$$\langle p)\hat{a}(k) = \langle k)y \underset{kxp}{F^T}(\underset{p(k)p}{FF^T})^{-1}. \tag{12}$$

Partition $\underset{pxk}{F}$ into

$$\underset{pxk}{F} = [\underset{pxk-1}{F} , \underset{k}{f\rangle}] \tag{13}$$

and transposing (13)

$$\underset{kxp}{F^T} = \begin{bmatrix} \underset{(k-1)xp}{F^T} \\ \overset{k}{\langle p)f} \end{bmatrix} . \tag{14}$$

Form the symmetric matrix product

$$\underset{p(k)p}{FF^T} = [\underset{p(k-1)}{F} , \underset{k}{f\rangle}] \begin{bmatrix} \underset{(k-1)xp}{F^T} \\ \overset{k}{\langle p)f} \end{bmatrix} \tag{15}$$

or

$$\underset{p(k)p}{FF^T} = \underset{p(k-1)p}{FF^T} + \underset{k}{f(p\rangle}\overset{k}{\langle p)f} . \tag{16}$$

By Equation (12) we must invert the matrix of (16), that is

$$(\underset{p(k)p}{FF^T})^{-1} = [\underset{p(k-1)p}{FF^T} + \underset{k}{f(p\rangle}\overset{k}{\langle p)f}]^{-1} \tag{17}$$

Fortunately for us, Houselholder* in his book of 1952 gives a formula for the inverse of a matrix which is the sum of an invertible matrix plus a dyad, that is when,

$$B = A + c\rangle\langle d \tag{18}$$

* Page 79.

and A^{-1} exists, then

$$B^{-1} = A^{-1} - \frac{A^{-1}c\rangle\langle dA^{-1}}{1 + \langle dA^{-1}c\rangle} \tag{19}$$

Observe that (17) is a special case (symmetric matrices) of (19).

One can verify the validity of (19) by "proof by execution". The proof by direct derivation for the symmetric case is presented in Appendix D.

Applying (19) to (17)

$$\begin{aligned}(\underset{p(k)p}{FF^T})^{-1} &= (\underset{p(k-1)p}{FF^T} + f(\underset{k}{p\rangle}\overset{k}{\langle p})f)^{-1} \\ &= (\underset{p(k-1)p}{FF^T})^{-1} - \frac{\underset{p(k-1)p}{(FF^T)^{-1}}\,\underset{k}{f\rangle}\overset{k}{\langle f}\,\underset{p(k-1)p}{(FF^T)^{-1}}}{1 + \underset{k}{\langle f}\,(\underset{p(k-1)p}{FF^T})^{-1}\,\underset{k}{f\rangle}}\end{aligned} \tag{20}$$

By Equation (IV-31)

$$\underset{kxp}{F^+} = \underset{kxp}{F^T}\,(\underset{p(k)p}{FF^T})^{-1} \tag{21}$$

or by (14) in (21)

$$\underset{kxp}{F^+} = \begin{bmatrix} \underset{(k-1)xp}{F^T} \\ \\ \overset{k}{\langle p)}f \end{bmatrix} (\underset{p(k)p}{FF^T})^{-1} \tag{22}$$

$$= \begin{bmatrix} \underset{(k-1)xp}{F^T}\;(\underset{p(k)p}{FF^T})^{-1} \\ \\ \overset{k}{\langle p)}f\;(\underset{p(k)p}{FF^T})^{-1} \end{bmatrix} \tag{23}$$

Consider the k^{th} row of (23)

$$\overset{k}{\langle f}\, (FF^T)^{-1}_{p(k)p} \tag{24}$$

$$= \overset{k}{\langle f}\, (FF^T)^{-1}_{p(k-1)p} \left[I - \frac{\underset{k}{f\rangle}\, \overset{k}{\langle f}\, (FF^T)^{-1}_{p(k-1)p}}{1 + \underset{k}{\langle f}\, (FF^T)^{-1}_{p(k-1)p}\, \underset{k}{f\rangle}} \right]$$

To simplify the notation for the following manipulations define

$$B = (FF^T)^{-1}_{p(k-1)p} \tag{25}$$

and

$$h = 1 + \underset{k}{\langle f}\, B\, \underset{k}{f\rangle} \tag{26}$$

then

$$\overset{k}{\langle f}\, (FF^T)^{-1}_{p(k)p} \tag{27}$$

$$= \overset{k}{\langle f} \left[B \left(I - \frac{\underset{k}{f\rangle}\, \overset{k}{\langle f}\, B}{h} \right) \right]$$

$$= \overset{k}{\langle f} \left[B \left(\frac{hI - \underset{k}{f\rangle}\, \overset{k}{\langle f}\, B}{h} \right) \right]$$

$$= \frac{\overset{k}{\langle f}}{h} \left[Bh - B\underset{k}{f\rangle}\, \overset{k}{\langle f} B \right]$$

$$= \frac{1}{h} \{ \langle f Bh - \langle f\, B\, f\rangle \langle f\, B \}$$

$$= \frac{1}{h} \{ \langle f\, B\, (1 + \langle f\, B\, f\rangle) - \langle f\, B\, f\rangle \langle f\, B \}$$

$$= \frac{1}{h} \langle f\, B$$

$$= \frac{1}{h} \langle f\, B$$

Using (25) and (26) back in (27)

$$\overset{k}{\langle f}\left(FF^{T} \right)^{-1}_{p(k)p} = \frac{\overset{k}{\langle f}\left(FF^{T} \right)^{-1}_{p(k-1)p}}{1 + {}_{k}\langle f \left(FF^{T} \right)^{-1}_{p(k-1)p} f\rangle_{k}} \tag{28}$$

Using Equation (20) and Equation (7) in the first (k-1) rows of Equation (23); and Equation (28) for the kth row we obtain

$$\underset{k\mathrm{x}p}{F^{+}(k)} = \begin{bmatrix} \underset{(k-1)(p)}{F^{+}(k-1)} \left[I - \dfrac{f\rangle_{k} \overset{k}{\langle f} \left(FF^{T} \right)^{-1}_{p(k-1)p}}{1 + {}_{k}\langle f \left(FF^{T} \right)^{-1}_{p(k-1)p} f\rangle_{k}} \right] \\ \\ \dfrac{\overset{k}{\langle f}\left(FF^{T} \right)^{-1}_{p(k-1)p}}{1 + {}_{k}\langle f \left(FF^{T} \right)^{-1}_{p(k-1)p} f\rangle_{k}} \end{bmatrix}$$

$$\equiv \begin{bmatrix} \underset{(k-1)\mathrm{x}p}{F^{+}(k)} \\ \\ \overset{k}{\langle f^{+}}(k) \end{bmatrix} \tag{29}$$

Returning to the expression for the optimal estimate of the parameters based on k observations of Equation (12), we have

$$\langle p)\hat{a}(k) = \langle k)y \underset{k\mathrm{x}p}{F^{T}} \left(FF^{T} \right)^{-1}_{p(k)p} \tag{30}$$

and by Equation (7) for k observations

$$\langle \hat{a}(k) = \langle k)y \underset{k\mathrm{x}p}{F^{+}(k)} \tag{31}$$

Partition the data vector $\langle k)y$

$$\langle k)y = [\langle k-1)y, \; y_{k}] \tag{32}$$

and using (32) and (29) in (31)

$$\langle p)\hat{a}(k) = [\langle k-1)y,\ y_k] \begin{bmatrix} F^{+}_{(k-1)p}(k) \\ \\ {}^{k}\langle f^{+}(k) \end{bmatrix}$$

$$= \langle k-1)y\ F^{+}_{(k-1)xp}(k) + y_k\,{}^{k}\langle f^{+}(k)$$

$$= \langle k-1)y \left[F^{+}_{(k-1)xp}(k-1)\left[I - \frac{f_k\rangle\,{}^{k}\langle f\,(FF^T)^{-1}_{p(k-1)p}}{1 + \langle f_k\,(FF^T)^{-1}_{p(k-1)p}\,f_k\rangle}\right]\right.$$

$$\left. + y_k \frac{{}^{k}\langle f\,(FF^T)^{-1}_{p(k-1)p}}{1 + \langle f_k\,(FF^T)^{-1}_{p(k-1)p}\,f_k\rangle}\right]$$

$$= \langle k-1)y\ F^{+}(k-1)\left[I - \frac{f_k\rangle\,{}^{k}\langle f\,(FF^T)^{-1}_{p(k-1)p}}{1 + \langle f_k\,(FF^T)^{-1}_{p(k-1)p}\,f_k\rangle}\right]$$

$$+ y_k \frac{{}^{k}\langle f\,(FF^T)^{-1}_{p(k-1)p}}{1 + \langle f_k\,(FF^T)^{-1}\,f_k\rangle}\ . \tag{33}$$

By equation (9)

$$\langle \hat{a}(k-1) = \langle k-1)y\ F^{+}_{(k-1)p}(k-1) \tag{34}$$

hence using (34) in (33)

$$\langle \hat{a}(k) = \langle \hat{a}(k-1)\left[I - \frac{f_k\rangle\,{}^{k}\langle f\,(FF^T)^{-1}_{p(k-1)p}}{1 + \langle f_k\,(FF^T)^{-1}_{p(k-1)p}\,f_k\rangle}\right] \tag{35}$$

Equation (35) Continued

$$+ y_k \frac{{}^{k}\langle f \,(FF^T)^{-1}_{p(k-1)p}}{1 + {}_{k}\langle f \,(FF^T)^{-1}_{p(k-1)p}\, f \rangle_k}$$

$$= \langle \hat{a}(k-1) - \langle \hat{a}(k-1)\, f \rangle_k \frac{{}^{k}\langle f \,(FF^T)^{-1}_{p(k-1)p}}{h}$$

$$+ y_k \, {}^{k}\langle f \,(FF^T)^{-1}_{p(k-1)p} \frac{1}{h}$$

Factoring ${}^{k}\langle f \, B \frac{1}{h}$ out of the last two terms on the right hand side of Eq. (35) we have

$$\langle \hat{a}(k) = \langle \hat{a}(k-1) + (y_k - \langle \hat{a}(k-1)\, f \rangle_k) \frac{{}^{k}\langle f \,(FF^T)^{-1}_{p(k-1)p}}{1 + {}_{k}\langle f \,(FF^T)^{-1}_{p(k-1)p}\, f \rangle_k} \tag{36}$$

It is clear from Equation (36) that if the scalar

$$y_k - \langle \hat{a}(k-1)\, f \rangle_k = 0 \tag{37}$$

no up-dating or correction is necessary. Suppose we are fitting a quadratic equation and the parameters based on k-1 data points is $\langle \hat{a}(k-1)$ then the equation for all points on the "fitted curve" are

$$\hat{y} = \hat{a}_0(k-1) + \hat{a}_1(k-1)x + \hat{a}_2(k-1)x^2 \tag{38}$$

where the ordered pairs $\begin{pmatrix} x \\ \hat{y} \end{pmatrix}$ lie on the curve as shown in Figure (1).

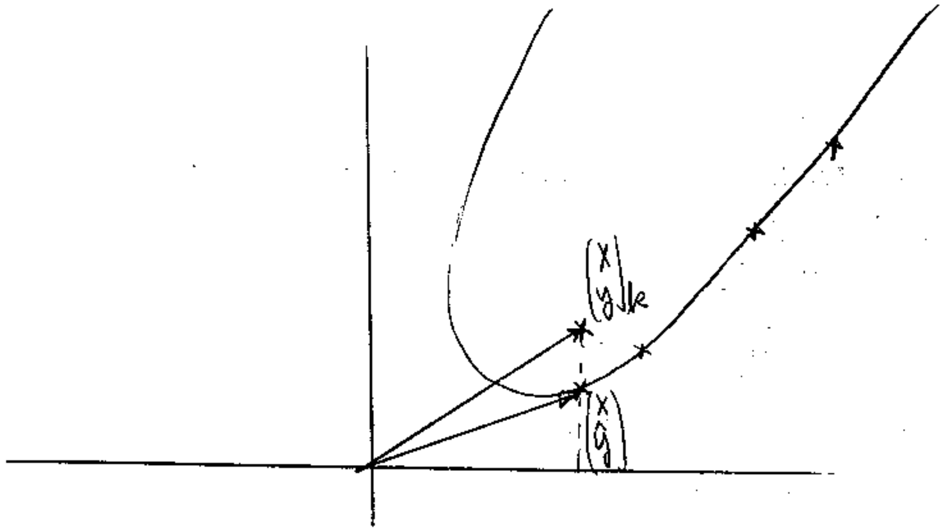

Fig. (1) Curve for Parameter Based on k-1 Observations

The set of vectors on the curve are described by

$$\begin{pmatrix} x \\ \hat{y} \end{pmatrix} = \begin{pmatrix} x \\ \hat{a}_0(k-1) + \hat{a}_1(k-1)x + \hat{a}_2(k-1)x^2 \end{pmatrix} \tag{39}$$

The k^{th} vector is shown in Fig. (1). The error in the second coordinate is

$$\tilde{y}_k = y_k - \hat{y} \; . \tag{40}$$

Using (39) in (40) we have

$$\tilde{y}_k = y_k - \langle p)\hat{a}(k-1)\; f(p\rangle_k \tag{41}$$

If the k^{th} observed value y_k falls on the $(k-1)^{th}$ curve, then

$$y_k = \hat{y}_k = \langle \hat{a}(k-1)\; f\rangle_k \tag{42}$$

and naturally we do not need to make a correction to establish a new curve. If the k^{th} value of y_k does not fall on the curve we must correct the parameters of the curve by the amount of the error

$$y_k - \langle \hat{a}(k-1)\; f\rangle_k = \tilde{y}_k \tag{43}$$

times the weighting vector

$${}^{k}\langle p)f\; (FF^T)^{-1}_{px(k-1)p} \quad \frac{1}{1 + {}_k\langle f\; (FF^T)^{-1}_{p(k-1)p}\; f\rangle_k}$$

as shown in Eq. (36).

If we define a weighting vector or a varying feed-back gain $w(k)$ as

$$\langle w(k) = \frac{\langle {}_k f \; (FF^t)^{-1}_{p(k-1)p}}{1 + \langle f_k (FF^t)^{-1}_{p(k-1)p} \; f_k \rangle} \tag{44}$$

Then Equation (36) by use of (44) and (41) becomes

$$\langle \hat{a}(k) = \langle \hat{a}(k-1) + \tilde{y}(k) \langle w(k). \tag{45}$$

A feed-back block diagram of Equation (45) is shown in Figure (2). One can construct other block diagrams or logic flow charts. The weighting vector generator is not shown in block diagram form, but can be constructed.

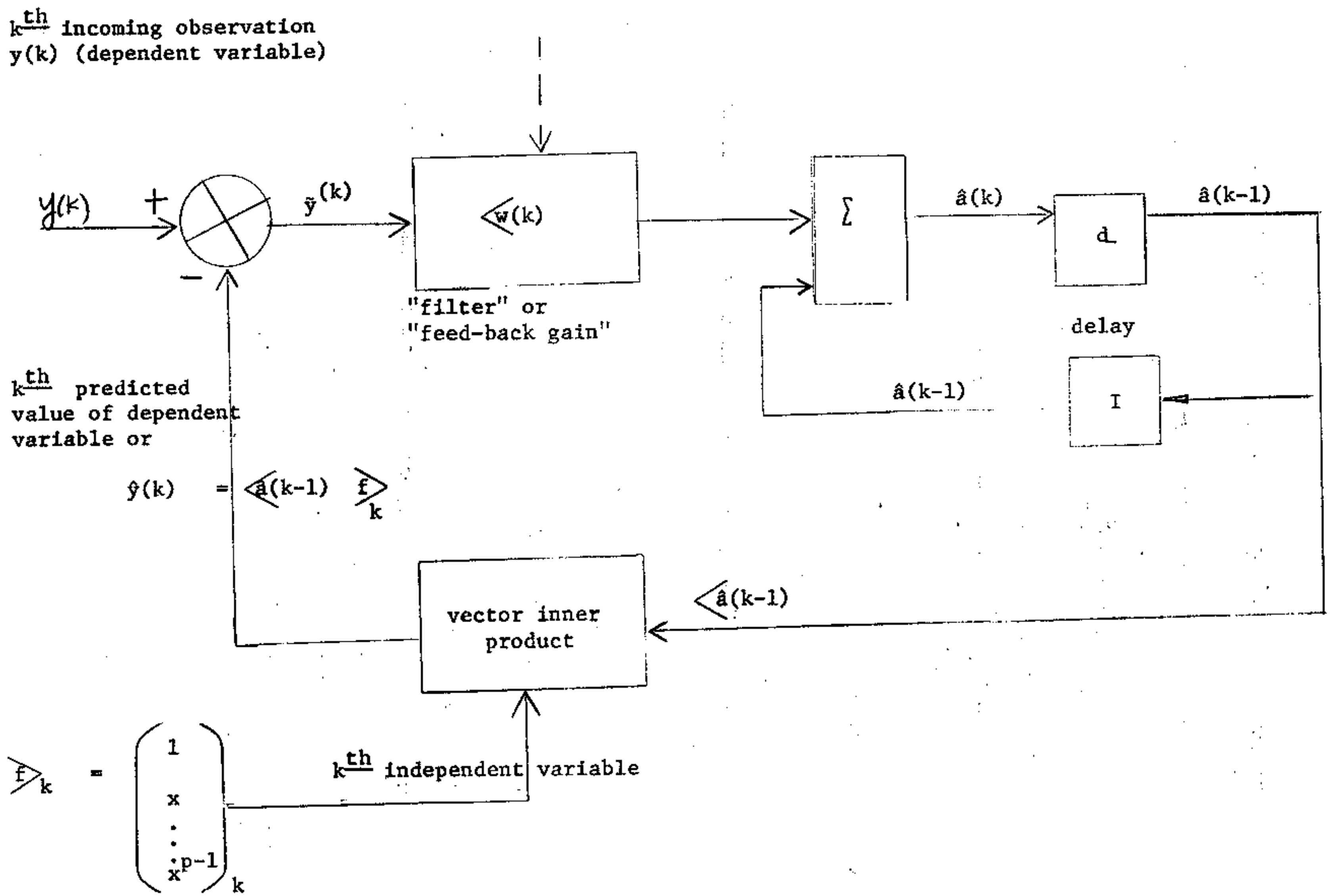

Figure (2) Feedback Diagram of Sequentially Corrective Curve-Filter.

Part 3. Iterative Computation of a Matrix of Parameters for Full Rank Data.

Consider the sequence of vectors $y\rangle_j$ of dimension q of Eq. (V-33) where,

$$y(q)\rangle_j = \underset{q\times p}{A}\; f(p)\rangle_j + e(q)\rangle_j \tag{1}$$

$$j = 1, k-1, k, k+1 \;.\;.\;.\;.\;.$$

Suppose we have k-1 observations as in Eq. (V-35)

$$\underset{q\times k-1}{Y} = \underset{(q\times p)}{A}\;\underset{(p\times k-1)}{F} + \underset{q\times k-1}{E} \tag{2}$$

Multiplying Eq. (2) by F^T

$$\underset{q(k-1)p}{YF^T} = \underset{q\times p}{A}\;\underset{p(k-1)p}{F\,F^T} + \underset{q(k-1)p}{E\,F^T} \tag{3}$$

When the data matrix F is full rank, the conventional inverse $(FF^T)^{-1}$ exists, hence

$$\underset{q(k-1)p}{Y\,F^T}\;\underset{p(k-1)p}{(\;F\,F^T)^{-1}} = \hat{A}(k-1) \tag{4}$$

where, as before, orthogonality holds

$$\underset{q(k-1)p}{\hat{E}\,F^T} = \underset{q\times p}{[\;0\;]}. \tag{5}$$

Suppose now that the k^{th} observation vector "comes in", then

$$\underset{p\times k}{F} = [\;\underset{p\times(k-1)}{F}\;,\; f(p)\rangle_k\;] \tag{6}$$

and the transpose of (6) is

$$\underset{k\times p}{F^T} = \begin{bmatrix} \underset{(k-1)\times p}{F^T} \\ {}_k\langle(p)f \end{bmatrix} \tag{7}$$

Equation (2) for k points is

$$\underset{q\times k}{Y} = \underset{(q\times p)}{A}\,\underset{(p\times k)}{F} + \underset{q\times k}{e} \tag{8}$$

and

$$\underset{q(k)p}{Y F^T} = \underset{(q\times p)}{\hat{A}(k)}\;\underset{p(k)p}{F F^T} \tag{9}$$

or

$$\underset{q(k)p}{Y F^T}\;\underset{p\times p}{(FF^T)^{-1}} = \underset{q\times p}{\hat{A}(k)} \tag{10}$$

The data matrix Y can be partitioned thusly,

$$\underset{q\times k}{Y} = [\underset{q\times(k-1)}{Y}\;,\;\underset{k}{y\langle q\rangle}]. \tag{11}$$

Using (7) and (11) in (10)

$$[\underset{q\times k-1}{Y}\;,\;\underset{k}{y\rangle}] \begin{bmatrix} \underset{(k-1)\times p}{F^T} \\ \overset{k}{\langle p)f} \end{bmatrix} \underset{p\times p}{(FF^T)^{-1}} = \hat{A}(k) \tag{12}$$

or

$$\hat{A}(k) = [\underset{q\times k-1}{Y}\;,\;\underset{k}{y\langle q\rangle}] \begin{bmatrix} \underset{(k-1)\times p}{F^T}\;\underset{p(k)p}{(FF^T)^{-1}} \\ \overset{k}{\langle p)f}\;\underset{p(k)p}{(FF^T)^{-1}} \end{bmatrix}. \tag{13}$$

By the Householder matrix inversion of Equation (D-29)

$$\underset{p(k)p}{(FF^T)^{-1}} = \underset{p(k-1)p}{(FF^T)^{-1}} \left[\underset{p\times p}{I} - \frac{\underset{k}{f\rangle}\,\overset{k}{\langle f}\,\underset{p(k-1)p}{(FF^T)^{-1}}}{1 + \underset{k}{\langle f}\,\underset{p(k-1)p}{(FF^T)^{-1}}\,\underset{k}{f\rangle}} \right] \tag{14}$$

where

$$H = \underset{p(k)p}{FF^T} = \left[\underset{p(k-1)p}{FF^T} + f\rangle_k \, {}^k\langle f \right]$$

and

$$a = 1.$$

The column of matrices terms of Equation (13) can be written as

$$\begin{bmatrix} \underset{(k-1)p}{F^T} \left(\underset{p(k)p}{FF^T} \right)^{-1} \\ \\ {}^k\langle p)f \left(\underset{p(k)p}{FF^T} \right)^{-1} \end{bmatrix} \tag{15}$$

$$= \begin{bmatrix} \underset{(k-1)p}{F^T} \left(\underset{p(k-1)p}{FF^T} \right)^{-1} \left[I - \dfrac{f\rangle_k \, {}^k\langle f \left(\underset{p(k-1)p}{FF^T} \right)^{-1}}{1 + {}_k\langle f \left(\underset{p(k-1)p}{FF^T} \right)^{-1} f\rangle_k} \right] \\ \\ {}^k\langle p)f \left(\underset{p(k)p}{FF^T} \right)^{-1} \end{bmatrix}$$

Now the last row of Equation (15) can be expressed as

$${}^k\langle p)f \left(\underset{p(k)p}{FF^T} \right)^{-1} \tag{16}$$

$$= {}^k\langle f \left(\underset{p(k-1)p}{FF^T} \right)^{-1} \left[\underset{pxp}{I} - \dfrac{f\rangle_k \, {}^k\langle f \left(\underset{p(k-1)p}{FF^T} \right)^{-1}}{1 + {}_k\langle f \left(\underset{p(k-1)p}{FF^T} \right)^{-1} f\rangle_k} \right]$$

In order to simplify the notation let

$$\left(\underset{p(k-1)p}{FF^T} \right)^{-1} = B \tag{17}$$

then

$$\overset{k}{\langle f}\,(FF^T)^{-1}_{p(k)p} = \overset{k}{\langle f} B \left[I - \frac{f\rangle_k \overset{k}{\langle f} B}{1 + {}_k\langle f B f\rangle_k} \right] \tag{18}$$

$$= \langle f B - \frac{\langle f B f\rangle \langle f B}{1 + \langle f B f\rangle}$$

Obtaining a common scalar denominator in Equation (18)

$$\overset{k}{\langle p)} f\,(FF^T)^{-1}_{p(k)p} = \frac{(1 + \langle f B f\rangle)\ \langle f B - \langle f B f\rangle \langle f B}{1 + \langle f B f\rangle} \tag{19}$$

$$= \frac{\langle f B + \langle f B f\rangle \langle f B - \langle f B f\rangle \langle f B}{1 + \langle f B f\rangle} \tag{20}$$

$$= \frac{\langle f B}{1 + \langle f B f\rangle} \tag{21}$$

Replacing (17) in (21)

$$\overset{k}{\langle f}\,(FF^T)^{-1}_{p(k)p} = \frac{\overset{k}{\langle f}\,(FF^T)^{-1}_{p(k-1)p}}{1 + {}_k\langle f\,(FF^T)^{-1}_{p(k-1)p}\, f\rangle_k}\ . \tag{22}$$

Using Equation (22) for the last row term of Equation (15) we obtain

$$\begin{bmatrix} F^T_{(k-1)p}\,(FF^T)^{-1}_{p(k-1)p} \\ \\ \overset{k}{\langle p)} f\,(FF^T)^{-1}_{p(k)p} \end{bmatrix} =$$

Equation (23) Cont'd

$$= \left[\begin{array}{c} F^T_{(k-1)xp} \, (FF^T)^{-1}_{p(k-1)p} \left[I - \dfrac{f_k\rangle \langle{}^k f \, (FF^T)^{-1}_{p(k-1)p}}{1 + \langle f_k \, (FF^T)^{-1}_{p(k-1)p} \, f_k\rangle} \right] \\ \\ \dfrac{\langle{}^k f \, (FF^T)^{-1}_{p(k-1)p}}{1 + \langle f_k \, (FF^T)^{-1}_{p(k-1)p} \, f_k\rangle} \end{array} \right] , \qquad (23)$$

Now by (13)

$$\hat{A}(k)_{qxp} = Y_{qx(k-1)} \, F^T_{(k-1)xp} \, (FF^T)^{-1}_{p(k)p} + y(q\rangle_k \langle{}^k p) f \, (FF^T)^{-1}_{p(k)p} \qquad (24)$$

Using the expressions of Eq. (23) for $F^T(FF^T)^{-1}_{p(k)p}$ and for $\langle{}^k f \, (FF^T)^{-1}_{p(k)p}$ in Equation (24)

$$\hat{A}(k) = Y_{qx(k-1)} \, F^T_{(k-1)xp} \, (FF^T)^{-1}_{p(k-1)p} \left[I - \frac{f_k\rangle \langle{}^k f \, (FF^T)^{-1}_{p(k-1)p}}{1 + \langle f_k \, (FF^T)^{-1}_{p(k-1)p} \, f_k\rangle} \right] \qquad (25)$$

$$+ \frac{y(q\rangle_k \langle{}^k p) f \, (FF^T)^{-1}_{p(k-1)p}}{1 + \langle f_k \, (FF^T)^{-1}_{p(k-1)p} \, f_k\rangle} .$$

By Equation (4) for $\hat{A}(k-1)$

$$\hat{A}(k) = \hat{A}(k-1)\left[I - \frac{f_k\rangle \langle f_k \left(FF^T\right)^{-1}_{p(k-1)p}}{1 + \langle f_k \left(FF^T\right)^{-1}_{p(k-1)p}\, f_k\rangle} \right] + \frac{y_k\rangle \langle f_k \left(FF^T\right)^{-1}_{p(k-1)p}}{1 + \langle f_k \left(FF^T\right)^{-1} f_k\rangle} \tag{26}$$

or

$$\hat{A}(k) = \hat{A}(k-1) + \left[\, y_k\rangle - \hat{A}(k-1)\, f_k\rangle \right] \frac{\langle f_k \left(FF^T\right)^{-1}_{p(k-1)p}}{1 + \langle f_k \left(FF^T\right)^{-1}_{p(k-1)p}\, f_k\rangle} \tag{27}$$

If there were no error vector on the k^{th} observation then Equation (1) would be

$$\hat{y}_k\rangle = \hat{A}(k-1)\, f_k\rangle \tag{28}$$

where $\hat{y}_k\rangle$ is the expected k^{th} vector. Since the k^{th} observation has noise, define,

$$\tilde{y}_k\rangle = y_k\rangle - \hat{y}_k\rangle \tag{29}$$

or

$$\tilde{y}_k\rangle = y_k\rangle - \hat{A}(k-1)\, f_k\rangle \tag{30}$$

where the tilde (˜) designates error. Using (30) in (27)

$$\hat{A}(k) = \hat{A}(k-1) + \frac{\tilde{y}_k\rangle \langle f_k \left(FF^T\right)^{-1}_{p(k-1)p}}{1 + \langle f_k \left(FF^T\right)^{-1}_{p(k-1)p}\, f_k\rangle} \,. \tag{31}$$

Equation (31) is the recursive equations for up-dating the estimates of the parameters. It is clear that when the error vector in the dependent vectory $\tilde{y}_k$ is zero, that the new estimate of the parametric matrix $\hat{A}$ is the same as the past.

Equation (27) states the new estimate of the parameters $\hat{A}(k)$ based on k observations as a matrix function of the estimate based on k-1 observations and the k^{th} observation vector.

APPENDIX A

TRACE OF A MATRIX

The trace of a square matrix A is defined in elementary vector-space texts as the sum of the elements on the main diagonal, for example if A is 2x2

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{1}$$

then

$$\operatorname{tr} A = a_{11} + a_{22} \tag{2}$$

is the scalar of Equation (2).

Consider the sum of two matrices A and B

$$\begin{aligned} A + B &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \\ &= \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{pmatrix} \end{aligned} \tag{3}$$

and the corresponding trace of the sum is the sum of the traces

$$\begin{aligned} \operatorname{tr}(A+B) &= a_{11} + b_{11} + a_{22} + b_{22} = (a_{11} + a_{22}) + (b_{11} + b_{22}) \\ &= \operatorname{tr} A + \operatorname{tr} B. \end{aligned} \tag{4}$$

When A is pxp, analogous results occur.

Consider a square rank-one matrix, a single dyad, say

$$D = b(p \times p)c = \begin{pmatrix} b^1 \\ b^2 \\ \vdots \\ b^p \end{pmatrix} (c_1, c_2, \cdots c_p) \tag{5}$$

or

$$D = \begin{pmatrix} b^1c_1 & \cdots & b^1c_p \\ \vdots & \ddots & \vdots \\ b^pc_1 & \cdots & b^pc_p \end{pmatrix}. \tag{6}$$

The sums of the main diagonal terms of Equation (6) is

$$\text{tr } D = b^1c_1 + \ldots + b^pc_p = c_1b^1 + \ldots + c_pb^b \tag{7}$$

Separating the b's and the c's and writing (7) as an inner-product we have

$$\text{tr } D = (c_1, \ldots, c_p) \begin{pmatrix} b^1 \\ \vdots \\ b^p \end{pmatrix} . \tag{8}$$

or

$$\text{tr}[b\rangle\langle c] = \langle c\ b\rangle . \tag{9}$$

By Equation (9) we see that the trace of an outer-product (dyadic) is an inner-product, or the "commuted product".

Equation (9) holds only when the vectors $b\rangle$ and $\langle c$ are of the same dimension or size, since the inner-product is not defined other-wise.

Consider a square matrix A of size pxp and of rank equal to or less than p. Partition A into its column space

$$\underset{pxp}{A} = [a(p\rangle_1, \ldots, a(p\rangle_p] \tag{10}$$

or into its row space

$$A = \begin{pmatrix} \langle a^1 \\ \vdots \\ \langle a^p \end{pmatrix} . \tag{11}$$

If the p column vectors of A are linearly independent then the rank of A is p. If they are linearly dependent, and only $m < p$ are linearly independent then all p column vectors lie in an m-dimensional subspace. Suppose we take any basis for the sub-space, such a basis could be an orthonormal Gram-Schmidt basis. Suppose we have a basis consisting of m vectors in the p-space expressed as a row of column vectors as

$$\underset{pxm}{B} = [b(p\rangle_1, \ldots, b(p\rangle_m] \tag{12}$$

then each vector can be expressed in the new basis as

$$\begin{aligned} a(p\rangle_1 &= b(p\rangle_1 \, c^{1}_{.1} + \ldots + b(p\rangle_m \, c^{m}_{.1} = \underset{pxm}{B} c(m\rangle_1 \\ &\vdots \\ a(p\rangle_p &= b(p\rangle_1 \, c^{1}_{.p} + \ldots + b(p\rangle_p \, c^{m}_{.p} = Bc(m\rangle_p \end{aligned} \tag{13}$$

or

$$A = [Bc\rangle_1, \ldots Bc(m\rangle_p] \tag{14}$$

$$= \underset{pxm}{B} [c(m\rangle_1, \ldots c(m\rangle_p] \tag{15}$$

$$\underset{pxp}{A} = \underset{(pxm)}{B} \underset{(mxp)}{C} \tag{16}$$

Partition B and C into column and row spaces, we obtain

$$\underset{pxm}{B} = [b(p\rangle_1, \ldots, b(p\rangle_m] = \begin{pmatrix} {}^{1}\langle m)b \\ \vdots \\ {}^{p}\langle m)b \end{pmatrix} \tag{17}$$

and

$$C = [c(m\rangle_1, \ldots, c(m\rangle_p] = \begin{pmatrix} {}^{1}\langle p)c \\ \vdots \\ {}^{m}\langle p)c \end{pmatrix} . \tag{18}$$

The product of (16) can be written by (17) and (18) as an "outer-product" of inner-products

$$A = BC = \begin{pmatrix} {}^{1}\langle m)b \\ \vdots \\ {}^{p}\langle m)b \end{pmatrix} [c(m\rangle_1, \ldots, c(m\rangle_p] \tag{19}$$

$$= \begin{pmatrix} {}^{1}\langle m)b \; c(m\rangle_{1}, & \ldots, & {}^{1}\langle m)b \; c(m\rangle_{p} \\ \vdots & & \vdots \\ {}^{p}\langle m)b \; c(m\rangle_{1}, & \ldots, & {}^{p}\langle m)b \; c(m\rangle_{p} \end{pmatrix} \tag{20}$$

Equation (20) is the conventional row-by-column product rule for matrices.

The product can also be written as an "inner product" of "outer products" as

$$A = BC = [b(p\rangle_{1}, \; \ldots, \; b(p\rangle_{m}] \begin{pmatrix} {}^{1}\langle p)c \\ \vdots \\ {}^{m}\langle p)c \end{pmatrix} \tag{21}$$

$$A = BC = b(p\rangle_{1}\,{}^{1}\langle p)c \; + \ldots + \; b(p\rangle_{m}\,{}^{m}\langle p)c. \tag{22}$$

Clearly Equation (22) relates A to its minimum rank factors B and C and to its minimum number of rank-one sums.

The sums of the diagonal elements of Equation (20) is

$$\mathrm{tr}\; A = \mathrm{tr}(BC) = {}^{1}\langle m)b \; c(m\rangle_{1} + \ldots + {}^{p}\langle m)b \; c(m\rangle_{p}. \tag{23}$$

If we take the trace of the dyadic decomposition of A in Equation (22) we obtain

$$\mathrm{tr}\; A = \mathrm{tr}\; \{b(p\rangle_{1}\,{}^{1}\langle p)c + \ldots + b(p\rangle_{m}\,{}^{m}\langle p)c\} \tag{24}$$

By Equation (4), the trace of sums is the sums of traces

$$\begin{aligned} \mathrm{tr}(BC) &= \mathrm{tr}\; \{b(p\rangle_{1}\,{}^{1}\langle p)c + \ldots + b(p\rangle_{m}\,{}^{m}\langle p)c\} \\ &= \mathrm{tr}\;\; b(p\rangle_{1}\,{}^{1}\langle p)c \; + \ldots + \mathrm{tr}\;\; b(p\rangle_{m}\,{}^{m}\langle p)c \end{aligned} \tag{25}$$

and by Equation (9) and Equation (25)

$$\mathrm{tr}\;(BC) = {}^{1}\langle p)c \; b(p\rangle_{1} + \; \ldots + {}^{m}\langle p)c \; b(p\rangle_{m}\;. \tag{26}$$

Equation (23) expresses the trace as the summation of p inner-product terms of m dimensional vectors. Equation (26) expresses the trace as the summation of m inner-product terms of p-dimensional vectors.

If we look at the commuted product expressed as an "outer-product of inner-products"

$$\underset{(mxp)(pxm)}{C \quad B} = \begin{pmatrix} \overset{1}{\langle p)}c \\ \vdots \\ \overset{m}{\langle p)}c \end{pmatrix} \quad [b\underset{1}{(p\rangle}, \ \ldots, \ b\underset{m}{(p\rangle}]$$

$$= \begin{pmatrix} \overset{1}{\langle p)}c \ b\underset{1}{(p\rangle}, & \cdots & , \overset{1}{\langle p)}c \ b\underset{m}{(p\rangle} \\ \vdots & \ddots & \vdots \\ \overset{m}{\langle p)}c \ b\underset{1}{(p\rangle} & & \overset{m}{\langle p)}c \ b\underset{m}{(p\rangle} \end{pmatrix} \tag{27}$$

or as "inner-product of outer product"

$$\underset{mxm}{CB} = [c\underset{1}{(m\rangle}, \ \ldots \ , \ c\underset{p}{(m\rangle}] \begin{pmatrix} \overset{1}{\langle m)}b \\ \vdots \\ \overset{p}{\langle m)}b \end{pmatrix} \tag{28}$$

or

$$CB = c\underset{1}{(m\rangle}\overset{1}{\langle m)}b + \ . \ . \ . \ + c\underset{p}{(m\rangle}\overset{p}{\langle m)}b \tag{29}$$

The trace of CB represented by Equation (27) and Equation (29) is

$$\mathrm{tr}(CB) = \overset{1}{\langle p)}c \ b\underset{1}{(p\rangle} + \ldots + \overset{m}{\langle p)}c \ b\underset{m}{(p\rangle} \tag{30}$$

$$= \overset{1}{\langle m)}b \ c\underset{1}{(m\rangle} + \ldots + \overset{p}{\langle m)}b \ c\underset{p}{(m\rangle} \ . \tag{31}$$

By Equation (26) and Equation (30) we see that

$$\underset{pxp}{\mathrm{tr}(BC)} = \underset{mxm}{\mathrm{tr}(CB)}. \tag{32}$$

APPENDIX B

ORTHOGONAL PROJECTIONS

Let $\bar{x}$ be a vector in a p-dimensional vector space $L^{(p)}$ and let $L^{(m)}$ $(m<p)$ be a linear subspace of $L^{(p)}$ that is $L^{(m)} \subset L^{(p)}$. Any arbitrary vector $\bar{x}$ in $L^{(p)}$ can be decomposed into a sum of two vectors

$$\bar{x} = \hat{x} + \tilde{x} \tag{1}$$

such that $\hat{x}$ lies in $L^{(m)}$ and $\tilde{x}$ is orthogonal to $L^{(m)}$. The vector $\hat{x}$ is called the projection of $\bar{x}$ on $L^{(m)}$ and the vector $\tilde{x}$ is called the projection of $\bar{x}$ on $\tilde{L}^{(m)}$ where $\tilde{L}^{(m)}$ is the orthogonal complement subspace to $L^{(m)}$ (sometimes designated $L^{(m)\perp}$.) The geometry is shown in Fig. (1)

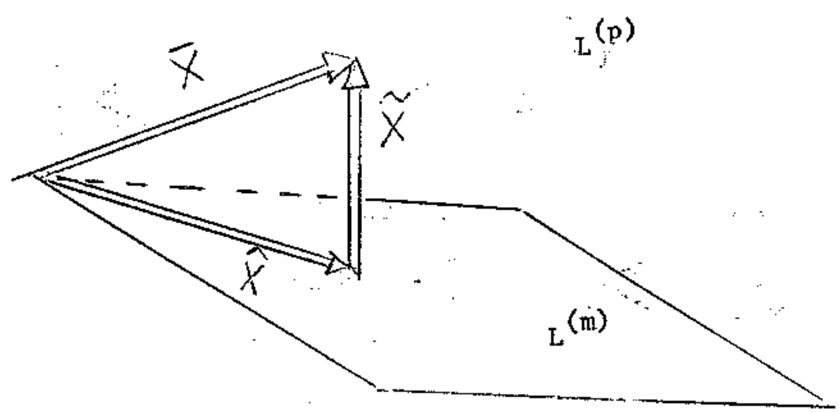

Fig. (1) P-Space and M-Dimensional Subspace

The vector $\hat{x}$ is sometimes said to be the vector in $L^{(m)}$ which is "nearest" to $\bar{x}$.

Theorem:

If $\bar{x} = \hat{x} + \tilde{x}$ (as used in previous discussion) then for any other vector

$$\bar{x}_m \epsilon L^{(m)}$$

$$\bar{x} = \bar{x}_m + \bar{r} = \hat{x} + \tilde{x} \tag{2}$$

APPENDIX C

DIFFERENTIALS AND GRADIENTS OF BI-LINEAR AND QUADRATIC FORMS

A bi-linear form is here taken to be a scalar valued (form) function of a row vector and a column vector, that is

$$\psi = \langle u \; B \; w \rangle \tag{1}$$

where perhaps $\langle u$ is an ~~n~~ m-dimensional row vector, $w\rangle$ is an n-dimensional column vector and B is a mxn matrix.

If we define a vector

$$z\rangle = Bw\rangle \tag{2}$$

then

$$\psi = \langle u \; z\rangle \tag{3}$$

looks like the inner-product or dot product of two vectors.

The differential of (1) can be written as

$$d\psi = \left\langle du \; \frac{\partial(\psi)}{\partial u} \right\rangle + \left\langle \frac{\partial(\psi)}{\partial w} \; dw \right\rangle \tag{4}$$

where the gradient of ψ with respect to the row vector $\langle u$ is a column vector, that is

$$\frac{\partial(\psi)}{\partial u}\Big\rangle = \begin{pmatrix} \dfrac{\partial\psi}{\partial u^1} \\ \vdots \\ \dfrac{\partial\psi}{\partial u^n} \end{pmatrix} = \frac{\partial}{\partial u}\Big\rangle \psi \; . \tag{5}$$

Using (1) in (5)

$$\frac{\partial(\psi)}{\partial u}\Big\rangle = \frac{\partial}{\partial u}\Big\rangle \langle u \; B \; w\rangle \; . \tag{6}$$

If the n coordinates $(u_1, \ldots, u_n)$ are independent then

$$\frac{\partial}{\partial u}\rangle\langle u = \begin{pmatrix} \frac{\partial}{\partial u^1} \\ \vdots \\ \frac{\partial}{\partial u^n} \end{pmatrix} (u_1, \ldots, u_n) \tag{7}$$

or

$$\frac{\partial}{\partial u}\rangle\langle u = \begin{pmatrix} \frac{\partial u^1}{\partial u} & \frac{\partial u_2}{\partial u_1} & \cdots & \frac{\partial u^n}{\partial u_1} \\ & & & \\ \frac{\partial u^n}{\partial u^n} & \cdots & \cdots & \frac{\partial u^n}{\partial u^n} \end{pmatrix} \tag{8}$$

$$= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \\ \cdot & & \ddots & 0 \\ \cdot & & & \\ \cdot & & & 1 \\ & 0 & & 1 \end{pmatrix} = I_{nxm} \tag{9}$$

Using (9) in (6)

$$\frac{\partial(\psi)}{\partial u}\rangle = IBw\rangle = Bw\rangle \; . \tag{10}$$

The scalar ψ is a function also of the column vector $w\rangle$ hence the gradient of ψ with respect to $w\rangle$ is a row vector, that is $w\rangle$ is a column vector and $dw\rangle$ is a column vector, hence to have a contraction process we need an inner product as indicated in Equation (4). The gradient of ψ with respect to $w\rangle$ is

$$\langle\frac{\partial(\psi)}{\partial w} = \psi\langle\frac{\partial}{\partial w} = \psi\left(\frac{\partial}{\partial w_1} \quad \cdots \quad \frac{\partial}{\partial w_n}\right) \tag{11}$$

and

$$||\bar{r}|| = ||\bar{x}-\bar{x}_m|| > ||\bar{x} - \hat{x}|| = ||\tilde{x}|| \tag{3}$$

Proof

Express $\bar{x}$ as

$$\bar{x} = \hat{x} + \tilde{x} = \bar{x}_m + \bar{r} \tag{4}$$

as shown in Figure (2), where

$$\bar{x}_m \neq \hat{x}$$

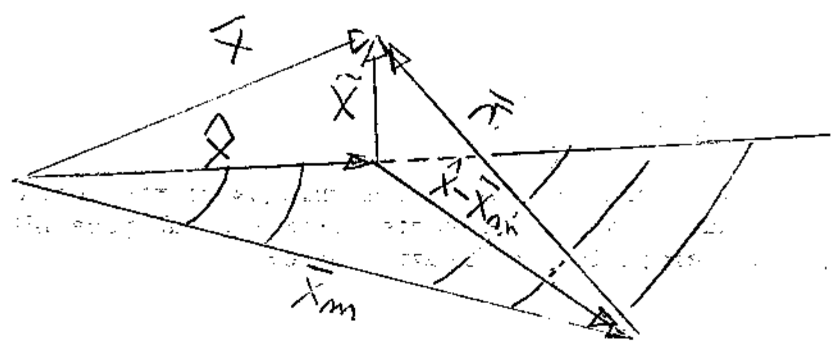

Fig. (2) Minimum Magnitude Vector

By Equation (2)

$$\bar{r} = \bar{x} - \bar{x}_m \tag{5}$$

$$\bar{r} \cdot \bar{r} = (\bar{x} - \bar{x}_m) \cdot (\bar{x} - \bar{x}_m) = ||\bar{r}||^2 \tag{6}$$

or

$$||\bar{r}||^2 = ||\bar{x} - \bar{x}_m||^2 \tag{7}$$

but

$$\bar{x} = \hat{x} + \tilde{x} \tag{8}$$

therefore

$$||\bar{r}||^2 = [\hat{x} + \tilde{x} - \bar{x}_m] \cdot [\hat{x} + \tilde{x} - \bar{x}_m] \tag{9}$$

$$= [(\hat{x} - \bar{x}_m) + \tilde{x}] \cdot [(\hat{x} - \bar{x}_m) + \tilde{x}]$$

$$||\bar{r}||^2 = (\hat{x} - \bar{x}_m) \cdot (\hat{x} - \bar{x}_m) + \tilde{x} \cdot \tilde{x}$$

since

$$(\hat{x} - \bar{x}_m) \cdot \tilde{x} = 0 \tag{10}$$

therefore

$$||\bar{r}||^2 = ||\hat{x} - \bar{x}_m||^2 + ||\tilde{x}||^2. \tag{11}$$

By Equation (7) and (11)

$$||\bar{x} - \bar{x}_m||^2 = ||\hat{x} - \bar{x}_m||^2 + ||\tilde{x}||^2 \tag{12}$$

since

$$||\hat{x} - \bar{x}_m|| > 0 \tag{13}$$

$$||\bar{x} - \bar{x}_m||^2 > ||\tilde{x}||^2$$

and

$$||\bar{x} - \bar{x}_m|| > ||\bar{x} - \hat{x}|| \tag{14}$$

Equation (14) tells us that the magnitude of the vector $\tilde{x}$ is smaller than the magnitude of any other vector from the subspace $L^{(m)}$ to the point at $\underline{x}$, hence the terminology of least squares.

or by (1) in (11)

$$\psi \left\langle \frac{\partial}{\partial w} \right. = \left\langle u \; B \; w \right\rangle \left\langle \frac{\partial}{\partial w} \right. \tag{12}$$

As before when the w's are independent variables, then

$$\left. w \right\rangle \left\langle \frac{\partial}{\partial w} \right. = \begin{pmatrix} w^1 \\ \cdot \\ \cdot \\ \cdot \\ w^n \end{pmatrix} \left(\frac{\partial}{\partial w_1} \; . \; . \; . \; \frac{\partial}{\partial w_n} \right)$$

$$= \begin{bmatrix} \frac{\partial w^1}{\partial w_1} & \frac{\partial w^1}{\partial w_2} & \cdots & \frac{\partial w^1}{\partial w_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial w^n}{\partial w_1} & \cdots & \cdots & \frac{\partial w^n}{\partial w_n} \end{bmatrix} \tag{13}$$

or

$$\left. w \right\rangle \left\langle \frac{\partial}{\partial w} \right. = \underset{nxm}{I} \; . \tag{14}$$

Using (14) in (12)

$$\psi \left\langle \frac{\partial}{\partial w} \right. = \left\langle uBI \right. = \left\langle uB \right. \tag{15}$$

hence the row vector gradient is given by (15).

Substituting (10) and (15) into Equation (4).

$$d\psi = \left\langle du \; Bw \right\rangle + \left\langle uB \; dw \right\rangle \tag{16}$$

It is clear that the differential product rule applied to Equation (1) yields

$$d\psi = d\left(\left\langle uBw \right\rangle \right) = \left\langle duBw \right\rangle + \left\langle uBdw \right\rangle \tag{17}$$

when

$$dB = [0]. \tag{18}$$

Transposing the row vector $\langle u$ to a column vector

$$(\langle u)^T = u\rangle \tag{19}$$

and if the condition

$$u\rangle = w\rangle \tag{20}$$

is imposed, we have the <u>quadratic form</u>

$$\psi = \langle u \; B \; u\rangle \, . \tag{21}$$

By (20)

$$du\rangle = dw\rangle \tag{22}$$

and using (20) and (22) in (17)

$$d\psi = \langle du B u\rangle + \langle u B du\rangle \tag{23}$$

The transpose of a 1x1 matrix of reals, or a scalar, equals itself, that is

$$(\langle du B u\rangle)^T = \langle u B^T du\rangle \tag{24}$$

hence (23) can be written as

$$d\psi = \langle u B^T du\rangle + \langle u B du\rangle = \langle u(B^T + B)du\rangle \tag{25}$$

If the matrix B is symmetric, that is

$$B = B^T$$

then

$$B^T + B = B + B = 2B \tag{26}$$

hence (25) becomes

$$d\psi = 2\langle u B \; du\rangle \, . \tag{27}$$

We can "loosely" say that if $u\rangle$ is a column vector (contravariant form) so is the differential $du\rangle$ and the gradient is a (covariant vector) that is a row vector, or

$$\left\langle \frac{\partial(\psi)}{\partial u} = \left(\frac{\partial \psi}{\partial u_1}, \; \ldots, \; \frac{\partial \psi}{\partial u_n}\right) = 2\langle u B \right. \tag{28}$$

If $w\rangle$ of Equation (1) is a constant vector

$$w\rangle = c\rangle \tag{29}$$

then

$$\psi = \langle uBc\rangle \tag{30}$$

and

$$d\psi = \left\langle du \, \frac{\partial\psi}{\partial u}\right\rangle \tag{31}$$

where

$$\left.\frac{\partial\psi}{\partial u}\right\rangle = Bc\rangle \, . \tag{32}$$

Or transposing (32)

$$\langle cB = \left\langle \frac{\partial(\psi)}{\partial u}\right. \, . \tag{33}$$

APPENDIX D

HOUSEHOLDER MATRIX INVERSION LEMMA FOR SYMMETRIC NON-SINGULAR MATRIX PLUS A SYMMETRIC DYAD

The sequential computation on a digital computer of optimal (least **squares**) parameter estimation can be derived for full rank data matrices via Householder's Matrix Inversion Lemma.

When the data is not of full rank, then we must utilize a generalization of the lemma in terms of a generalized or psuedo inverse. **Clearly** when the data is first started to be collected it is not of full rank and we can not achieve "real time" without the "generalized lemma".

This section derives in a manner different from Householder's derivation that portion necessary for the sequential computation of the generalized inverse for full rank kxp data matrices.

Consider a non-singular symmetric pxp matrix H which is equal to a non-singular symmetric matrix G plus a symmetric dyad $b\rangle\langle b$ times a real scalar a, that is

$$H = G + a\, b\rangle\langle b\,. \tag{1}$$

We seek H^{-1} as a function of G^{-1} plus other simple terms.

Multiply (1) by H^{-1} on left then

$$H^{-1}H = I = H^{-1}G + a\,H^{-1}b\rangle\langle b \tag{2}$$

and multiply on right by H^{-1} (commutativity property)

$$HH^{-1} = I = GH^{-1} + a\,b\rangle\langle b\,H^{-1}. \tag{3}$$

Multiply (2) on right (post) by G^{-1} then

$$G^{-1} = H^{-1} + aH^{-1}b\rangle\langle b\,G^{-1}\,. \tag{4}$$

Pre-multiply (3) by G^{-1} then

$$G^{-1} = H^{-1} + aG^{-1}b\rangle\langle b\,H^{-1}\,. \tag{5}$$

Subtract (5) from (4) then

$$[0] = aH^{-1}b\rangle\langle b\,G^{-1} - aG^{-1}b\rangle\langle b\,H^{-1} \tag{6}$$

or

$$[0] = H^{-1}b\rangle\langle b\,G^{-1} - G^{-1}b\rangle\langle b\,H^{-1} \tag{7}$$

for $a \neq 0$.

Define the vectors

$$c\rangle = H^{-1} b\rangle \tag{8}$$

and

$$\langle f = \langle b G^{-1} . \tag{9}$$

Transposing (8) and (9) and using symmetry property of H^{-1} and G^{-1} we obtain

$$\langle c = \langle b (H^{-1})^T = \langle b H^{-1} \tag{10}$$

and

$$f\rangle = (G^{-1})^T b\rangle = G^{-1} b\rangle \ . \tag{11}$$

Using (8), (9), (10) and (11) in (6)

$$[0] = c\rangle\langle f - f\rangle\langle c \tag{12}$$

or

$$c\rangle\langle f = f\rangle\langle c \ , \tag{13}$$

Since the right hand side of (13) is the transpose of the left hand side, that is

$$[c\rangle\langle f]^T = f\rangle\langle c \tag{14}$$

we have the skew-symmetry property of

$$c\rangle\langle f - [c\rangle\langle f]^T = [0] \tag{15}$$

If the dyad is to be symmetric as required by (13), then clearly $c\rangle$ and $f\rangle$ are parallel vectors or

$$c\rangle = f\rangle \beta = G^{-1} b\rangle \beta . \tag{16}$$

By (5) and (10)

$$H^{-1} = G^{-1} - aG^{-1} b\rangle\langle b H^{-1} \tag{17}$$

$$H^{-1} = G^{-1} - aG^{-1} b\rangle\langle c \tag{18}$$

By (16) and (18)

$$H^{-1} = G^{-1} - aG^{-1} b\rangle\langle b G^{-1} \beta \tag{19}$$

By (1)

$$H = G + a b\rangle\langle b \tag{20}$$

By (1), (2), and (19)

$$H^{-1}H = I = [G^{-1} - aG^{-1}b\rangle\langle bG^{-1}\beta][G + a b\rangle\langle b] \tag{21}$$

or

$$I = GG^{-1} + aG^{-1}b\rangle\langle b - aG^{-1}b\rangle\langle b\beta - a^2G^{-1}b\rangle\langle bG^{-1}b\rangle\langle b\, \beta \tag{22}$$

and using property

$$GG^{-1} = I \tag{23}$$

and associating terms in last term of (22)

$$a^2G^{-1}b\rangle\langle bG^{-1}b\rangle\langle b \tag{24}$$

$$= a^2G^{-1}b\rangle\left(\langle bG^{-1}b\rangle\right)\langle b \tag{25}$$

$$= a^2G^{-1}b\rangle\langle b\left(\langle bG^{-1}b\rangle\right) \tag{26}$$

we obtain

$$I = I + G^{-1}b\rangle\langle b[a - a\beta - a^2\beta\langle bG^{-1}b\rangle]. \tag{27}$$

The scalar coefficient of the second matrix term of (27) must be zero, or β must satisfy

$$0 = a - a\beta - a^2 \langle bG^{-1}b\rangle\beta \tag{28}$$

or

$$a = a\beta + a^2\beta \langle bG^{-1}b\rangle \tag{29}$$

$$1 = \beta + a\beta\langle bG^{-1}b\rangle \tag{30}$$

$$1 = \beta(1 + a \langle bG^{-1}b\rangle) \tag{31}$$

or

$$\beta = \frac{1}{1 + a\langle bG^{-1}b\rangle} \tag{32}$$

Using (32) in (19)

$$H^{-1} = G^{-1} - \frac{aG^{-1}|b\rangle\langle b|G^{-1}}{1 + a\langle b|G^{-1}|b\rangle} = [G + a|b\rangle\langle b|]^{-1} \tag{33}$$

We may factor out G^{-1} in (33)

$$H^{-1} = G^{-1} [I - \frac{a|b\rangle\langle b|G^{-1}}{1 + a\langle b|G^{-1}|b\rangle}]. \tag{34}$$

One may perform a proof by execution by multiplying Equation (34) by Equation (1) to obtain the identity matrix.

REFERENCES

1. Albert, Arthur and Sittler, Robert; "Real Time Computation of Constrained Least Squares Estimators", Arcon, Lexington, Mass., June 1965.

2. Bodewig, E., "Matrix Calculus", North Holland Publishing Co., Amsterdam, 1956.

3. European Editor, Electronics Magazine, July 25, 1966, page 149.

4. Fisher, R. A., "Contributions To Mathematical Statistics", John Wiley & Sons, Inc., New York, 1950, page 273 of "Biography" by P. C. Mahalanobis.

5. Friedman, Bernard, "Principles and Techniques of Applied Mathematics", John Wiley, New York, 1956.

6. Graybill, F. A., An Introduction to Linear Statistical Models", McGraw-Hill Book Co. Inc., 1961.

7. Greville, T.N.E., "Some Applications of the Psuedo Inverse of a Matrix", Siam Review, Vol. 2, No. 1, January 1960.

8. Horst, Paul, "Factor Analysis of Data Matrices", Part I, Univ. of Washington, Seattle, Washington, Nov 1963.

9. Ho, Y. C., "A Method of Least Squares and Optimal Filtering Theory", Memo RM-3329-PR, Rand Corp., Oct 1962.

10. Householder, A. S., "Principles of Numerical Analysis", McGraw-Hill Book Co. Inc., New York, 1953.

11. Kalman, R.E., "New Methods In Wiener Filtering Theory" in Proceedings of the First Symposium on Engineering Applications of Random Function Theory and Probability, John Wiley and Sons, New York, 1963.

12. Kalman, R. E., "Mathematical Description of Linear Dynamical Systems", J. Siam Control Ser. A. Vol. I, No. 2, 1963.

13. M.A.A. "Studies in Modern Mathematics", Vol. I. "Studies in Modern Analysis", R. C. Buck Editor - Prentice Hall Inc. 1962.

14. Scheffe, Henry, "The Analysis of Variance", John Wiley, 1959.

15. Tou, Julius T., "Optimum Design of Digital Control Systems", Academic Press, New York, 1963.

16. Willis, A.P., "Vector Analysis with an Introduction to Tensor Analysis", Dover Publications, New York, 1931.

REFERENCES

17. Scott, Maceo T. and Rede, Edmundo, "Recursive Estimation: An Approximate Solution to the Geodesy Problem Associated with WSMR", Instrumentation Internal Memo Nr. 26, Analysis & Computation Directorate, White Sands Missile Range, New Mexico, July 1966.

18. Scott, Maceo T., Rede, Edmundo, Wygant. M., "Descrete Recursive Estimation: An Optimum Automated Data Processing Technique for Multi-Radar Tracking Systems." Technical Report, 5 April 1967. WSMR, New Mexico.

DISTRIBUTION LIST

Copy Nos.	Distribution
215	Mr. D. R. Hamilton Pan American Airways Cape Kennedy, Florida
216	Mr. Richard B. Holst Code 341 Pacific Missile Range Point Mugu, California
217	Mr. Jim Jantz Vandenberg Air Force Base, California 93437
218	Mr. Art Raine NASA Range Safety Wallops Island, Virginia
219	Mr. Arnold H. RObinson Federal Electric Corp Western Test Range Vandenberg Air Force Base, California 93437
220	Mr. Olaf E. W. Heindahl U.S. Nâval Ordnance Test Station China Lake, California
221	Mr. Clarence M. Vick Sandia Corp. Albuquerque, New Mexico 87115
222	Mr. William G. Gyertson Kentron Hawaii Ltd. 2003 Byrd Spring Rd. Huntsville, Alabama 35802

DISTRIBUTION LIST

Copy Nos.	Distribution
1	Commanding General, WSMR, N. M.
2	Deputy for Air Force, WSMR, N.M.
3	Deputy for Navy, WSMR, N.M.
4-9	National Range Engineering (1 - Chief, RE) (2 - RE-I) (1 - RE-S) (1 - RE-Q) (1 - RE-T)
10-116	National Range Operations (1 - Chief, NR) (100 - NR-A - author) (4 - NR-D) (1 - NR-E) (1 - NR-L)
117-128	Army Missile Test and Evaluation (1 - Chief) (1 - TE-M) (4 - TE-T) (4 - TE-X)
129	Missile Flight Surveillance
130-139	Technical Library
140-159	Defense Documentation Center Cameron Station, Alexandria, Virginia 22314
160-173	Department of Defense, The Pentagon Washington 25, D.C. (ATTN: Technical Library)
174-179	Department of the Army, The Pentagon Washington 25, D.C. (ATTN: Technical Library)
180-209	HQ, U.S. Army Test & Evaluation Command, Aberdeen Proving Ground, Maryland 21005
210	Range Commanders Council Secretariat White Sands Missile Range, N. M.
211	U.S. Army Research Office, Durham Box C.M. Duke Station Durham, North Carolina 27706 (ATTN: Tech Lib)
212	U.S. Army Research Office Arlington, Virginia
213	U.S. Army Mathematics Research Center, University of Wisconsin (ATTN: Tech. Library) Madison, Wisconsin
214	Cpt. Ray H. Lee, USAF AFETR Patrick Air Force Base, Florida

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
1. Sequential						
2. Optimum						
3. Least squares						
4. Matrix						
5. Parameters						
6. State-vector						
7. Predict-correct						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* ANALYSIS AND COMPUTATION DIRECTORATE DEPUTY FOR NATIONAL RANGE OPERATIONS WHITE SANDS MISSILE RANGE, NEW MEXICO 88002	**2a. REPORT SECURITY CLASSIFICATION** UNCLASSIFIED **2b. GROUP**

3. REPORT TITLE
APPLICATION OF THE KALMAN FILTER TO SEQUENTIAL OPTIMAL PARAMETER ESTIMATION VIA HOUSEHOLDER'S MATRIX INVERSION METHOD

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
SPECIAL REPORT

5. AUTHOR(S) *(Last name, first name, initial)*
PAPPAS, JAMES

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
JUNE 1967	91	18

8a. CONTRACT OR GRANT NO.	**9a. ORIGINATOR'S REPORT NUMBER(S)**
b. PROJECT NO.	
c.	**9b. OTHER REPORT NO(S)** *(Any other numbers that may be assigned this report)*
d.	

10. AVAILABILITY/LIMITATION NOTICES
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT
A detailed derivation of the computable discrete equations for parameter estimation are developed in a geometrical vector-space setting for the state-space novice. Classical least squares curve fitting when approached with Kalman's sequential prediction-correction techniques look like state-vector feed back control problems. It is hoped that this paper will help bridge the gap between some of the modern and classical theory of systems analysis.

DD FORM 1473 1 JAN 64