AN APPLICATION OF ORTHOGONALIZATION PROCESS TO THE THEORY OF LEAST SQUARES

By Y. K. Wong

Introduction

The present paper is an outgrowth of the writer's attempt to fill a lacuna in the discussion of the Gauss method of substitution as given by many writers. For illustration, let us cite Brunt's *Combination of Observations*. In Chapter VI, we find:

Let the normal equations be

$$\begin{aligned} [aa]x + [ab]y + [ac]z - [al] &= 0 \\ [bb]y + [bc]z - [bl] &= 0 \\ [cc]z - [cl] &= 0\,. \end{aligned} \tag{i}$$

From this equation we find

$$x = -\frac{[ab]}{[aa]}\,y - \frac{[ac]}{[aa]}\,z + \frac{[al]}{[aa]}\,. \tag{ii}$$

Substituting, we obtain

$$\begin{aligned} [bb1]y + [bc1]z - [bl1] &= 0 \\ [cc1]z - [cl1] &= 0 \end{aligned} \tag{iii}$$

where

$$[bb1] = [bb] - [ab]\,[ab]/[aa], \text{ etc.} \tag{iv}$$

From the first equation in (iii),

$$y = -\frac{[bc1]}{[bb1]}\,z + \frac{[bl1]}{[bb1]}\,. \tag{v}$$

In connection with equations (ii) and (v), the question naturally arises as to whether or not these numbers $[aa]$, $[bb1]$, $\cdots$ are all different from zero. Since $[aa] = \Sigma a_i a_i$, one can see that $[aa] \neq 0$ if $a_i \neq 0$ for every i. However, to show the non-vanishing of $[bb.1]$, $[cc.2]$, etc. is by no means simple. Many writers do not give a demonstration on this point. We know that a system of non-homogeneous linear equations has a solution if the system of equations is linearly independent. Brunt gives a discussion of the independence of the normal equations in Chapter V, Art. 36, but he does not state clearly a condition for independence. He says: "The condition of independence is in general satisfied in the problems which arise in practice. We can then proceed to the formation and solution of the normal equations." It is one of the aims of this paper to give a necessary and sufficient condition for the independence of the normal equations and to show $[aa]$, $[bb.1]$, etc. are all different from zero when the condition is satisfied.

In the theory of least squares, there is the classical method of the derivation of normal equations by an application of the notion of minimum in differential calculus. After the normal equations are secured, the Gauss method of substitution is applied to obtain the solution. Doolittle modifies the Gauss method of substitution so as to facilitate the labor of computation. However, when the number of parameters (or unknowns) exceeds 4, Doolittle's method is quite complicated. In the present paper the writer wishes to present a mathematical discussion of a method obtained through an application of the Gram-Schmidt orthogonalization process. This method furnishes us a new procedure for determining the most probable values of the parameters (or unknowns). The formulation of the system of normal equations will be omitted in this new procedure, which is particularly effective in fitting curves to time series. The paper can be roughly divided into three parts. The first part gives an algebraic derivation of the normal equations. The second part derives a condition for a set of observation data so that the Gauss method of substitution is applicable. The third part gives a relationship between the Gauss method of substitution and the orthogonalization process. A practical application of the results of this paper will be found in a later paper.

The process of orthogonalization has been used in the 19th century, and has been applied extensively in the theory of integral equations and linear transformations in Hilbert space. In classical analysis, if $\varphi_1(x), \varphi_2(x), \cdots$, defined on $(0, 1)$, is a normally orthogonalized system, and if $f(x)$, defined on $(0, 1)$, is such that f^2 is Lebesgue integrable, then the system of Fourier coefficients

$$f_r = \int_0^1 f(x)\varphi_r(x)dx \qquad (r = 1, 2, \cdots)$$

has certain interesting properties, one of which is that

$$\frac{1}{m}\int_0^1 (f(x) - \sum_1^m f_r\varphi_r)^2\, dx = 0\,.$$

The preceding notion has a close connection with the theory of least squares as outlined in many texts on statistics. In section III, the reader will find how this notion is applied in the derivation of the normal equations. Since the number of dimensions is finite, the integration process reduces to a summation process and furthermore no limiting process is used. This new derivation of normal equations has the advantage that (1) differential calculus is not used, (2) a new form of normal equations is obtained, (3) the solution of the unknowns or parameters can be immediately obtained without further application of the

Gauss Method of Substitution or the Doolittle Method, and (4) the formula for the "quadratic residual" is obtained as a simple corollary.

From the results in section III, we see immediately what condition should be imposed upon the set of observation data so that the Gauss method of substitution may be applicable. In section VI, we find a necessary and sufficient condition for the independence of the system of normal equations (3.9), and also the fact that when this condition is fulfilled, then, due to the special nature of the coefficients of the unknowns, we see that the matrix is properly positive. It is on account of this fact that we are able to show that the numbers $[aa]$, $[bb.1]$, etc. are all different from zero. The demonstration of this point is found in section VII. In this section, we lay down a fundamental hypothesis for Gauss's method of substitution, namely, the set of observations $A_i = (a_{i1}, \cdots, a_{in})$ $i = 1, 2, \cdots, r$, is linearly independent. Lemma 7.3 may be called the fundamental lemma for Gauss's method of substitution. Some interesting properties of the numbers $[A_s A_t \cdot h]$, where $s, t = 1, \cdots, r$, and h is less than the smaller one of (s, t), are demonstrated.

From the properties of the numbers $[A_s A_t \cdot h]$, where $s, t = 1, \cdots, r$ and h is less than the smaller one of (s, t), and in comparison of the system of equations (3.7^0) with the final form of equations obtained through the application of the Gauss method of substitution, we can see the relationship between the Gauss method and the Gram-Schmidt orthogonalization process. If we should like to give some credit to Gauss, we may say that the orthogonalization process was known by him, but was stated in a different form.

The writer wishes to remark that certain theorems together with proofs in section II, IV, V and VI are obtained from E. H. Moore's lecture notes. However the writer should be responsible for any defect. Finally, I should emphasize that the use of the notion of positive matrices is only for convenience.

I. Vectors, Inner Products, and Linear Independence

In this paper, we shall consider vectors of the form[1]

$$(v_1, v_2, \cdots, v_n). \tag{1.10}$$

For convenience, we shall use capital letters to denote vectors of the type (1.10).

Let $V = (v_1, v_2, \cdots, v_n)$ and $U = (u_1, u_2, \cdots, u_n)$, then we say $V = U$ if $v_i = u_i$ for every i.

We define $V + U$ by

$$V + U = (v_1 + u_1, v_2 + u_2, \cdots, v_n + u_n), \tag{1.11}$$

and sV, where s is a number, by

$$sV = (sv_1, sv_2, \cdots, sv_n). \tag{1.12}$$

[1] If we write v_i as $v(i)$, where $i = 1, 2, \cdots, n$, then $v(i)$ may be considered as a function of one variable whose range consists of a set of positive integers, $(1, 2, \cdots, n)$. E. H. Moore defines a vector as a function of one variable.

Hence, $sV = Vs$. In particular, when $s = -1$. we shall put $-V = (-1)V$. Then $U - V$ becomes a special instance of (1.11) and (1.12).

From (1.11) and (1.12), we see that addition is commutative and associative.

INNER PRODUCTS: The inner product of two vectors $V = (v_1, \cdots, v_n)$ and $U = (u_1, \cdots, u_n)$ is defined[2] to be

$$(V, U) = \sum_1^n v_i u_i . \tag{1.2}$$

The norm of a vector V is defined by $n(V) = (V, V)$; and the modulus of a vector V is defined by mod $(V) = +\sqrt{n(V)}$.

From (1.11), (1.12), and (1.2), we can easily prove the following theorem:

THEOREM 1. *The symbol* (,) *has the following properties:*

(S) $(U, V) = (V, U)$ *for every* V, U; (symmetric property)

(L_s) $(sV, U) = s(V, U) = (V, sU)$ *for every* V, U *and every number* s;

(L_+) $(U, (V + W)) = (U, V) + (U, W)$ *for every* U, V, W; *(linear property)*

(P) $(V, V) \geqq 0$ *for every* V; *(positive property)*

(P_0) $(V, V) = 0$ *if and only if* V *is a zero vector*; *(properly positive property)*

LINEAR INDEPENDENCE. A set of vectors $V_1, \cdots, V_r$ is said to be linearly dependent in case there exist constants $c_1, \cdots, c_r$ not all equal to 0 such that

$$c_1V_1 + \cdots + c_rV_r = 0 ,$$

where 0 is a zero vector.

A set of vectors $V_1, \cdots, V_r$ is said to be linearly independent in case, if the constants $c_1, \cdots, c_r$ satisfy

$$c_1V_1 + \cdots + c_rV_r = 0 ,$$

each constant $c_i = 0$.

THEOREM 2. *If the set* $V_1, \cdots, V_r$ *is linearly independent, then none of the vectors is a zero vector, and hence the norm of every vector must be different from zero.*

For if V_s is a zero vector, then set $c_s = 1$, and $c_i = 0$ for $i \neq s$. It is obvious that

$$0\cdot V_1 + \cdots + 0\cdot V_{s-1} + 1\cdot V_s + 0\cdot V_{s+1} + \cdots + 0\cdot V_r = 0 ,$$

which show that the set of vectors $V_1, \cdots, V_r$ is linearly dependent, contradictory to the hypothesis.

A more general theorem is stated in

THEOREM 3. *If the set* $V_1, \cdots, V_r$ *is linearly independent, then every subset*[3] *is also linearly independent.*

We shall prove this theorem by a contrapositive form. The contrapositive form is as follows: *If in the set* $V_1, \cdots, V_r$, *there exists a subset which is linearly*

---

[2] The notation (,) was introduced by D. Hilbert. In treatises on least squares, the notation [ ] is used. The present writer reserves the latter notation for other purposes.

[3] Consider a set of integers $(1, 2, \cdots, n)$. Then any *combination* of this set of n distinct integers taken $r \leqq n$ at a time is called a *subset* of the set $(1, 2, \cdots, n)$. Likewise, we call any combination of the set of vectors $V_1, V_2, \cdots, V_n$ taken $r \leqq n$ at a time a *subset* of the whole set.

dependent, then the whole set is also linearly dependent. Without losing any generality, let us suppose the subset $V_1, \cdots, V_s$ $(s \leqq r)$ to be linearly dependent. Then there exist $c_1, \cdots, c_s$ such that

$$c_1V_1 + \cdots + c_sV_s = 0.$$

If $s = r$, then the whole set is linearly dependent. If $s < r$, then let $c_i = 0$ for $i = s - 1, s - 2, \cdots, r$. Then

$$\sum_1^r c_iV_i = 0,$$

which shows the whole set is linearly dependent.

THEOREM 4.[4] *A necessary and sufficient condition for the set $V_i = (v_{i1}, \cdots, v_{in})$, $i = 1, \cdots, r$ to be linearly independent is that there exists a non-vanishing determinant of order r in the array*

$$\begin{array}{l} v_{11}, v_{12}, \cdots, v_{1n} \\ v_{21}, v_{22}, \cdots, v_{2n} \\ \cdots\cdots \\ v_{r1}, v_{r2}, \cdots, v_{rn} \end{array}.$$

II. Gram-Schmidt's Orthogonalization Process

For the present section and the sequel, we shall adopt the notation $A_i = (a_{i1}, \cdots, a_{in})$, $B_i = (b_{i1}, \cdots, b_{in})$, and $C_i = (c_{i1}, \cdots, c_{in})$ for $i = 1, 2, \cdots, r$.

THEOREM 5. *For every set of vectors $A_1, \cdots, A_r$, there exists uniquely a set of vectors $B_1, \cdots, B_r$ such that*

5.1) $(B_t, B_s) = 0$ $(t \neq s)$.

5.2) *For every t satisfying the relation $1 \leqq t \leqq r$, then A_t is a linear combination of $B_1, \cdots, B_t$; and B_t is a linear combination of $A_1, \cdots, A_t$.*

5.3) $B_1 = A_1$; *and for $t > 1$, $(B_t - A_t)$ is a linear combination of $B_1, \cdots, B_{t-1}$, and is also a linear combination of $A_1, \cdots, A_{t-1}$.*

5.4) *If $t > 1$, then $(A_s, B_t) = 0$ for every $s < t$.*

5.5) $(A_t, B_t) = (B_t, B_t) = (B_t, A_t)$ *for every t.*

To prove this theorem, let us define

$$\begin{aligned} B_1 &= A_1, \\ B_2 &= A_2 \qquad \text{if} \quad n(B_1) = 0 \\ &= A_2 - \frac{(A_2, B_1)}{n(B_1)} B_1 \qquad \text{if} \quad n(B_1) \neq 0 \\ &\cdots\cdots\cdots\cdots \\ B_t &= A_t - \sum_{i=1}^{t-1} h_{ti}B_i \qquad (1 \leqq t \leqq r), \end{aligned} \tag{2.1}$$

[4] See Dickson, *Modern Algebraic Theories*, p. 55; Bocher, *Higher Algebra*, p. 36.

where

$$(2.11)\qquad \begin{aligned} h_{ti} &= (A_t, B_i)/n(B_i)\,, &\text{if}\quad n(B_i) \neq 0\,,\\ &= 0\,, &\text{if}\quad n(B_i) = 0\,. \end{aligned}$$

We proceed to show that this set has the properties stated in the theorem.

To prove 5.1), let us suppose $t < s$. This assumption is permissible since the operator (,) has the symmetric property. First, if $A_1 = 0$, then $B_1 = 0$, and

$$(B_1, B_2) = (A_1, A_2) = (0, A_2) = 0\,.$$

Secondly, if $A_1 \neq 0$, then $B_1 \neq 0$ and

$$\begin{aligned}(B_1, B_2) &= (A_1, A_2 - h_2, B_1) = (A_1, A_2) - (A_1, B_1)\frac{(A_2, B_1)}{n(B_1)}\\ &= (A_1, A_2) - (A_1, A_1)\,(A_2, A_1)/n(A_1) = 0\,.\end{aligned}$$

Assume 5.1) is true for $t = s - 1$, then

$$(B_t, B_s) = \left(B_t, A_s - \sum_1^{s-1} h_{si}B_i\right) = (B_t, A_s) - \sum_1^{s-1} h_{si}(B_t, B_i)\,.$$

The sum on the right hand side reduces to $h_{st}(B_t, B_t)$, since the other terms vanish by assumption. Now if $(B_t, B_t) \neq 0$ then by (2.11), $h_{st}(B_t, B_t) = (A_s, B_t)$, and by the symmetric property of (,), we obtain

$$(B_t, B_s) = (B_t, A_s) - (A_s, B_t) = 0\,.$$

If $(B_t, B_t) = 0$, then by the P_0-property of (,), we find that B_t is a zero vector, and hence $(B_t, B_s) = 0$.

5.2) follows from the definition of B_t.

That $(A_t - B_t)$ is a linear combination of $B_1, \cdots, B_{t-1}$ for $t > 1$ follows from the definition of B_t. Since B_s is a linear combination of $(A_1, \cdots, A_{s-1})$, we secure the second part of 5.3).

By 5.2), we can determine g_{si} such that $A_s = \sum_1^s g_{si}B_i$. Thus for every $s < t$, we have by 5.1)

$$(A_s, B_t) = \left(\sum_1^s g_{si}B_i, B_t\right) = \sum_1^s g_{si}(B_i, B_t) = .0$$

By 5.3), there exist g_{ti} such that $A_t - B_t = \sum_1^{t-1} g_{ti}B_i$ and hence $A_t = B_t + \sum_1^{t-1} g_{it}B_i$. Thus by 5.1), we have

$$\begin{aligned}(A_t, B_t) &= \left(B_t + \sum_1^{t-1} g_{ti}B_i, B_t\right) = (B_t, B_t) + \sum_1^{t-1} g_{ti}(B_i, B_t)\\ &= (B_t, B_t)\,.\end{aligned}$$

By the symmetric property of (,), we secure $(A_t, B_t) = (B_t, B_t)$.

For the proof of uniqueness, let us suppose there exists a second set of vectors $B'_1, \cdots, B'_r$ having the properties 5.1), 5.2), 5.3), 5.4), and 5.5). By 5.3), we see that $B_1 = A_1 = B'_1$. Assuming the uniqueness holds true for $r = t$, we proceed to show that it is also true for $r = t + 1$. By 5.3) there exist constants s_i, s'_i ($i = 1, \cdots, t$) such that

$$B_{t+1} = A_{t+1} + \sum_1^t s_i A_i$$

$$B'_{t+1} = A_{t+1} + \sum_1^t s'_i A_i .$$

Thus

$$B_{t+1} - B'_{t+1} = \sum_1^t (s_i - s'_i) A_i .$$

From this, we secure

$$(B_{t+1} - B'_{t+1}, B_{t+1} - B'_{t+1}) = \left(B_{t+1} - B'_{t+1}, \sum_1^t (s_i - s'_i)A_i\right)$$

$$= \sum_1^t (s_i - s'_i)\cdot(B_{t+1} - B'_{t+1}, A_i) = 0 ,$$

by virtue of 5.4). Hence by P_0-property of (,), we have $B_{t+1} - B'_{t+1} = 0$ and hence $B_{t+1} = B'_{t+1}$.

The set $B_1, \cdots, B_r$ with the properties stated in Theorem 5 is called the *orthogonalized set* of $A_1, \cdots, A_r$. This process is called Gram-Schmidt's orthogonalization process.

The set $B_1, \cdots, B_r$ is called the *normally orthogonalized set* of $A_1, \cdots, A_r$ in case the former set enjoys the properties 5.1), 5.2), 5.3), 5.4), and if

5.5n) $\quad (A_t, B_t) = (B_t, B_t) = (B_t, A_t) = 1$ *for every* t .

THEOREM 6. *If a subset $A_{k_1}, \cdots, A_{k_m} (1 \leqq k_1 \leqq \cdots \leqq k_m \leqq r)$ in the set $A_1, \cdots, A_r$, is linearly independent, then there is a subset $B_{k_1}, \cdots, B_{k_m}$ which has the properties stated in Theorem* 5, *and it is also linearly independent.*

Let $h = k_m - k_1 + 1$. To prove the theorem, we may assume $k_1, \cdots, k_m$ to be $1, \cdots, h \leqq r$, for otherwise, we may renumber the vectors. We construct the B vectors in the same way as given in equation (2.1) and (2.11). By Theorem 5, we have

(2.2) $$B_1 = A_1, \qquad B_s = A_s + \sum_1^{s-1} g_{si} A_i \qquad (s = 2, \cdots, h) .$$

Suppose the constants $c_1, \cdots, c_h$ be such that

$$c_1 B_1 + \cdots + c_h B_h = 0 .$$

Then by (2.2), we secure

$$0 = c_1A_1 + \sum_2^h c_sB_s = c_1A_1 + \sum_2^h c_s\left(A_s + \sum_1^{s-1} g_{si}A_i\right)$$

$$= (c_1 + c_2g_{21}, + \cdots + c_hg_{h1})\, A_1 + (c_2 + c_3g_{32} + \cdots c_hg_{h2})A_2 + \cdots + c_hA_h .$$

Since $A_1, \cdots, A_h$ are linearly independent, we have

$$(2.3) \qquad \begin{aligned} c_1 - c_2g_{21} - \cdots - c_hg_{h1} &= 0 , \\ c_2 - \cdots - c_hg_{h2} &= 0 , \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad &\cdot \\ c_h &= 0 . \end{aligned}$$

But the determinant of the coefficients of $c_i (i = 1, \cdots, h)$ is

$$\begin{vmatrix} 1 & g_{21} & g_{31} & \cdots & g_{h1} \\ 0 & 1 & g_{32} & \cdots & g_{h2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 1 \end{vmatrix} = 1 .$$

Hence by a theorem in the theory of equations,[5] the only solution that satisfies (2.3) is that $k_1 = k_2 = \cdots = k_h = 0$. Thus the subset $B_1, \cdots, B_h$ is linearly independent.

COROLLARY. *The orthogonalized set $B_1, \cdots, B_r$ is linearly independent if and only if the set $A_1, \cdots, A_r$ is linearly independent.*

THEOREM 7a. *If a set of vectors $A_1, \cdots, A_r$ is linearly independent, then the set can be normally orthogonalized.*

Let B_i be the orthogonalized set of A_i. Since A_i is a linearly independent set, then the set B_i is also linearly independent by Theorem 6. Hence by Theorem 2, the norm of every vector B_i is non-vanishing. Define $C_i = B_i/\text{mod}\,(B_i)$. Then this set C_i enjoys the properties 5.1), 5.2), 5.3), 5.4) and 5.5n).

THEOREM 7b. *If a set of vectors, $V_1, \cdots, V_r$ is normally orthogonal, i.e.* if

$$(2.4) \qquad (V_i, V_j) = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) , \end{cases}$$

then $V_1, \cdots, V_r$ is linearly independent.

For suppose

$$c_1V_1 + \cdots + c_rV_r = 0 .$$

Then

$$\sum_{i=1}^r c_i(V_i, V_j) = 0 , \qquad (j = 1,2, \cdots, r) .$$

[5] Dickson, *First Course in the Theory of Equations* (1922), p. 119.

By condition (2.4), the preceding expression reduces to

$$c_j = 0, \qquad (j = 1, 2, \cdots, r),$$

which shows the linear independence of $V_1, \cdots, V_r$.

III. Algebraic Derivation of the Normal Equations

Consider a linear function

$$l = p_1x_1 + p_2x_2 + \cdots + p_rx_r = \sum_1^r p_ix_i. \tag{3.1}$$

Let the set of observations of x_i and l be

$$A_i = (a_{i1}, \cdots, a_{in}), \qquad L = (l_1, \cdots, l_n) \qquad (i = 1, \cdots, r; n \geqq r) \tag{3.2}$$

respectively, then the residual v_i is

$$v_i = \sum_{j=1}^r p_ja_{ji} - l_i, \qquad (i, = 1, \cdots, n).$$

In vector notation,

$$V = \sum_{j=1}^r p_jA_j - L.$$

The theory of least squares requires us to find the values for $p_1, \cdots, p_r$ so as to make (V, V) a minimum, or

$$(\sum p_jA_j - L, \sum p_jA_j - L) = \text{a minimum}. \tag{3.3°}$$

Let $A_1, \cdots, A_r$ be linearly independent. By Theorem 7, the vectors $A_1, \cdots, A_r$ can be normally orthogonalized. Let $C_1, \cdots, C_r$ be the normally orthogonal set. Then every A_t $(t = 1, \cdots, r)$ is expressible as a linear combination of $C_1, \cdots, C_t$. Let us write

$$\sum_1^r p_jA_j = \sum_1^r k_jC_j. \tag{3.3}$$

Our problem now is equivalent to that of finding the values $k_i (i = 1, \cdots, r)$ so as to render the inner product

$$(\sum k_jC_j - L, \sum k_jC_j - L) \tag{3.4}$$

a minimum. Expression (3.4) can be written in the form

$$\begin{aligned} &(L, L) - 2\sum(L, C_i)k_i + \sum_{i,j} (k_iC_i, k_jC_j) \\ &= (L, L) - 2\sum(L, C_i)k_i + \sum k_i^2 \\ &= (L, L) - \sum(L, C_i)^2 + \sum(k_i - (C_i, L))^2. \end{aligned} \tag{3.5}$$

Hence (3.4) gives a minimum if and only if the last summation vanishes, i.e.,

$$k_i = (C_i, L) \qquad (i = 1, \cdots, r). \tag{3.6}$$

The Bessel's inequality

$$\sum_1^r k_i^2 \leqq (L, L)$$

is obtained from (3.6), (3.4), and (3.5).

To solve for p_i, we make use of (3.3) and (3.6), and secure

$$\sum_1^r A_i p_i = \sum_1^r (C_i, L) C_i,$$

whence

$$\left(C_k, \sum_1^r A_j p_j\right) = \left(C_k, \sum_1^r (C_i, L)\, C_i\right).$$

On the right hand side we have

$$(C_k, \sum (C_i, L) C_i) = \sum (C_i, L)\,(C_k, C_i) = (C_k, L),$$

since $(C_k, C_i) = 0$ when $i \neq k$, and $(C_k, C_i) = 1$ when $i = k$. On the left hand side, we have

$$\left(C_k, \sum_{j=1}^r A_j p_j\right) = \sum_{j=1}^r (C_k, A_j) p_j = \sum_{j=k}^r (C_k, A_j) p_j,$$

since $(C_k, A_j) = 0$ when $j < k$. Hence the values for $p_1, \cdots, p_r$ are given by

$$\sum_{i=k}^r (C_k, A_i) p_i = (C_k, L) \qquad (k = 1, \cdots, r), \tag{3.7}$$

where $(C_i, A_i) = (C_i, C_i) = 1$.

Equations (3.7) are called the normal equations, which are derived without using any notion in differential calculus.

From (3.6) and (3.5), we secure the value for the 'quadratic residual' (V, V):

$$(V, V) = (L, L) - \sum_{i=1}^r (L, C_i)^2, \tag{3.8}$$

which is a positive quantity by virtue of the Bessel's inequality.

Let $B_1, \cdots, B_r$ be an orthogonalized set of $A_1, \cdots, A_r$. Then every vector B_i has a non-vanishing norm, and $B_i = \text{mod}\,(B_i) \cdot C_i$. Hence from (3.7) and (3.8), we have

$$\sum_{i=k}^r (B_k, A_i) p_i' = (B_k, L), \qquad (k = 1, 2, \cdots, r), \tag{3.7°}$$

$$(V, V) = (L, L) - \sum_{i=1}^r (L, B_i)^2 / n(B_i). \tag{3.8°}$$

Thus we have proved the following

THEOREM 8. *Given a linear function* (3.1). *Let the set of observations of* x_i *and* l *be*

$$A_i = (a_{i1}, \cdots, a_{in}), \qquad L = (l_1, \cdots, l_n) \qquad (i = 1, \cdots r; n \geqq r)$$

respectively. Let $A_1, \cdots, A_r$ *be linearly independent,* $B_1, \cdots, B_r$ *be the orthogonalized set, and* $C_1, \cdots, C_r$, *the normally orthogonalized set of* $A_1, \cdots, A_r$. *Then the set of values* $p_1, \cdots, p_r$ *will minimize* (3.3°) *if and only if the system of equations* (3.7°) *or* (3.7) *holds true; in other words,* $\sum_{i=1}^{r} p_i A_i - L$ *is orthogonal to* C_j *or to* B_j *for every* j. *The quadratic residual* (V, V) *is given by* (3.8°) *or* (3.8).

From (3.7), we can secure the solution for $p_1, \cdots, p_r$ immediately without further application of the Gauss method of substitution.

The proof of the following theorem does not make use of the orthogonalization process.[6]

THEOREM 8°. *Let* $F = \Sigma p_i A_i$, *where every* A_i *is not a zero vector. The set of values* $p_1, \cdots, p_r$ *will minimize* (3.3°) *if and only if* $(F - L, A_i) = 0$ *for every* i, i.e., $F - L$ *is orthogonal to every* A_i.

The condition is necessary. To prove this, we show that if $(F - L, A_i) \neq 0$ for every i, then we can find another set $q_1, \cdots, q_r$ such that $n(F - L) > n(G - L)$, where $G = \Sigma q_i A_i$. For if $(F - L, A_i) \neq 0$ for every i, then we can find a vector A_s such that $(F - L, A_s) \neq 0$. Since $A_s \neq 0$, we let $e = (F - L, A_s)/n(A_s)$ and $G = F - eA_s = \Sigma q_i A_i$. Then

$$n(G - L) = n(F - eA_s - L) = n(F - L) - (F - L, A_s)^2/n(A_s),$$

which shows that $n(G - L) < n(F - L)$.

To prove the sufficiency, we show that for every set $q_1, \cdots, q_r$ different from $p_1, \cdots, p_r$ then $n(G - L) > n(F - L)$, where $G = \Sigma q_i A_i$. Let $s_i = q_i - p_i$, and $H = \Sigma s_i A_i$. Then $G = F + H$. Now if $(F - L, A_i) = 0$ for every i, it follows that

$$(F - L, H) = \sum_{i=1}^{r} (F - L, A_i) s_i = 0.$$

Thus

$$n(G - L) = n(F - L) + n(H).$$

Since $n(H) > 0$, we have $n(G - L) > n(F - L)$.

The preceding theorem does not require the linear independence of the vectors $A_1, \cdots, A_r$. By Theorem 7a and 7b we see that it is necessary and sufficient for the set $A_1, \cdots, A_r$ to be linearly independent in order to solve the equations $(F - L, A_i) = 0$, $(i = 1, 2, \cdots, r)$, or

$$\begin{array}{c} (A_1, A_1)p_1 + (A_1, A_2)p_2 + \cdots + (A_1, A_r)p_r = (A_1, L) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ (A_r, A_1)p_1 + (A_r, A_2)p_2 + \cdots + (A_r, A_r)p_r = (A_r, L). \end{array} \tag{3.9}$$

[6] The proof is based on the same type of reasoning as used by Jackson. See Dunham Jackson's *Theory of Approximation*, pp. 151–152.

If $A_1, \cdots, A_r$ are linearly independent, the conclusion in Theorem 8° can be deduced from Theorem 8. For by Theorem 7a) $A_i = \sum_t s_{it}C_t$, and hence

$$(F - L, A_i) = \left(F - L, \sum_t s_{it}C_t\right) = \sum_t s_{it}(F - L, C_t) = 0\,.$$

Also, Theorem 8 can be deduced from Theorem 8°.

IV. Matrices and Their Reciprocals

An ordered array of numbers of the form

$$\alpha = (a_{ij}) = \begin{bmatrix} a_{11}, & a_{12}, & \cdots & a_{1m} \\ a_{21}, & a_{22}, & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}, & a_{n2}, & \cdots & a_{nm} \end{bmatrix} \tag{4.1}$$

is a matrix. If we write $a(i, j) = a_{ij}$, then the array of numbers (4.1) may be considered as a function of two variables i, j on the ranges of positive integers $(1, 2, \cdots, n)$, $(1, 2, \cdots, m)$.[7] Thus a vector is a special instance of a matrix. We shall use Greek letters to denote matrices throughout this paper unless otherwise specified. When $n = m$, i.e. the number of rows is the same as the number of columns, we have a square matrix. Associated with every n-row square matrix, κ, a determinant can be defined, and for simplicity, we shall adopt the following notation:

$$D(\kappa) = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}.$$

An identity matrix, denoted by $\delta = (d_{ij})$, is a square matrix of which the elements in the principal diagonal are 1 and elsewhere 0, i.e. $d_{ij} = 0$ $(i \neq j)$, $d_{ii} = 1$. A zero matrix, indicated by ω, is one such that every one of its elements is 0. The transposed matrix, α', of α is formed by interchanging the rows and columns. We say two matrices $\alpha = (a_{ij})$ and $\beta = (b_{ij})$ are equal in case $a_{ij} = b_{ij}$ for every i, j. A matrix α is symmetric in case $\alpha' = \alpha$. The i^{th} column of α is indicated by $\alpha(., i)$, the i^{th} row of β by $\beta(i, .)$ and the element in the i^{th} row and j^{th} column by $\alpha\ (i, j)$. Hence $\alpha(i, j) = a_{ij}$.

ADDITION: Let α be a matrix given by (1) and $\beta = (b_{ij})$ a matrix of the same number of rows and columns as α. Then

$$\alpha + \beta = (a_{ij} + b_{ij})\,.$$

[7] E. H. Moore defines a matrix as a function of two variables.

We note that $\alpha + \beta = \beta + \alpha$. If γ is a matrix of the same number of rows and columns as α, then $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$.

MULTIPLICATION: Let $\alpha = (a_{ij})$ be defined by (1), and $\beta = (b_{jk})$ be a matrix of m row and r columns, then the product $\pi = \alpha\beta$ is defined by

$$\pi = (p_{ik}) = \left(\sum_{j=1}^{m} a_{ij} b_{jk} \right).$$

Thus π is a matrix of m rows and r columns.

The multiplication of two matrices is not necessarily commutative.

If α is a matrix of n rows and m columns, β of m rows and r columns, and γ of r rows and s columns, then $\alpha(\beta\gamma) = (\alpha\beta)\gamma$. If α is a matrix of n rows and m columns, and β, γ are matrices of m rows and r columns, then $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$.

SCALAR MULTIPLICATION: Let s be a number, and α be a matrix of n rows and m columns, then

$$s \cdot \alpha = (sa_{ij}) = \alpha \cdot s.$$

Let δ_s denote a square matrix of n rows in which the elements in the principal diagonal are s, and 0 elsewhere. Then $\delta_s = s\delta$, where δ is an n row identity matrix. We note from the associative law of multiplication that

$$s\alpha = \delta_s \cdot \alpha = \alpha \cdot \delta_s.$$

In particular, let $s = -1$, then we have -1α. For convenience, we write $-\alpha = -1\alpha$. From the definition of addition, we obtain a definition of subtraction for two matrices of the same number of rows and columns.

RECIPROCALS OF MATRICES: Let α be a matrix of n rows and m columns. Then a matrix α^{-1} of m rows and n columns is said to be a reciprocal of α in case

$$\alpha \cdot \alpha^{-1} = \delta^n, \quad \text{and} \quad \alpha^{-1} \cdot \alpha = \delta^m,$$

where δ^n, δ^m are identity matrices of order n, m respectively. If a matrix α has a reciprocal α^{-1}, we can prove α^{-1} is unique. It can be shown that *when α has a reciprocal, it must be a square matrix.*[8]

A matrix is said to be non-singular in case it has a reciprocal, otherwise it is said to be singular.[9] It is evident that every zero matrix is singular, and an identity matrix is non-singular.

Suppose α is a square matrix of order n. Let us denote the cofactor of the element a_{ij} of α by e_{ji}. Then

$$\epsilon = (e_{ij}) = \begin{pmatrix} e_{11} & \cdots & e_{1n} \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ e_{n1} & \cdots & e_{nn} \end{pmatrix}$$

is called the adjoint matrix of α.

---

[8] For the proof of this statement, see Moore, *Vector, Matrices, and Quaternions.*

[9] This definition is due to E. H. Moore.

If α is symmetric, then ϵ is also symmetric. Since $a_{i1}e_{1j} + \cdots + a_{in}e_{nj} = D(\alpha)$ or 0 according as $i = j$ or $i \neq j$, we secure the following:

THEOREM 9. *Let α be a square matrix and ϵ its adjoint, then*

$$\alpha\epsilon = \epsilon\alpha = [D(\alpha)]\delta .$$

THEOREM 10. *If the determinant of α is different from zero, then there exists a reciprocal α^{-1}, and $\alpha^{-1} = adj\ \alpha/D(\alpha)$.*

This theorem follows from theorem 5.

The converse of Theorem 6 is also true.

V. Symmetric Matrices of Positive Type[10]

Let $\alpha = (a_{ij})$ be a matrix of n rows and m columns; and let $\sigma = (k_1, \cdots, k_n)$ and $\rho = (h_1, \cdots, h_m)$ be integers among the sets $(1, \cdots, n)$ and $(1, \cdots, m)$ respectively. The subsets σ and ρ may be equal to the whole sets $(1, \cdots, n)$ and $(1, \cdots, m)$ respectively. Then

$$\alpha(\sigma, \rho) = \left\| \begin{matrix} a_{k_1h_1} & \cdots & a_{k_1h_m} \\ \cdot & \cdot\cdot\cdot & \cdot \\ a_{k_nh_1} & & a_{k_nh_m} \end{matrix} \right\| \tag{3}$$

is called a minor of α. In notation we write this minor as $\alpha(\sigma, \rho)$ indicating the ranges to be σ and ρ.

The minor $\alpha(-\sigma, -\rho)$, which is obtained by striking out all the k_ith $(i = 1, \cdots, m)$ columns and h_jth $(j = 1, \cdots, m)$ rows from α, is called the complementary minor of α (σ, ρ).

If α is a square matrix of order n, then $\alpha(\sigma, \sigma)$ is called a principal minor of α.

Let α and β be matrices of n rows and m columns; and let σ, ρ have the same meaning as above. Then $\alpha(\sigma, \rho)$, $\beta(\sigma, \rho)$ are called corresponding minors in α, β respectively.

A symmetric matrix $\alpha = (a_{ij})$ of order n is said to be of *positive type* in case the determinant of every principal minor of α is positive, and is said to be of *properly positive type* in case the determinant of every principal minor of α is greater than zero.

COROLLARY V1. *Every element in the principal diagonal of a positive, symmetric matrix is positive.*

For, let σ consist of a single integer i, then $\alpha(\sigma, \sigma) = a_{ii} \geqq 0$.

COROLLARY V2. *If a symmetric matrix is properly positive, then every element in the principal diagonal is greater than* 0.

THEOREM 11. *If a symmetric matrix α of order n is (properly) positive, then its adjoint matrix ϵ is also symmetric and (properly) positive.*

---

[10] We follow the terminology of E. H. Moore. Moore developed this notion quite extensively.

The symmetry of ϵ is evident. Let σ be a subset of $(1, \cdots, n)$ and let p be the number of integers in σ. Consider any principal minor $\epsilon(\sigma, \sigma)$ in the adjoint matrix ϵ. By a theorem in the theory of determinants, we have[11]

$$D[\epsilon(\sigma, \sigma)] = (-1)^{2k} \cdot D[\alpha(-\sigma, -\sigma)] \cdot [D(\alpha)]^{p-1},$$

where k is an integer depending on the set σ. By hypothesis α is positive (properly positive); hence $D[\alpha(-\sigma, -\sigma)]$ and $[D(\alpha)]^{p-1}$ are positive (greater than 0), and it follows that $D[\epsilon(\sigma, \sigma)]$ is positive (greater than 0).

THEOREM 12. *If a symmetric matrix is properly positive, then $D(\alpha)$ is different from zero, and α has a reciprocal α^{-1}, which is also symmetric and properly positive.*

For take σ to be the whole set $(1, \cdots, n)$ in the definition of proper positiveness, and we see that $D(\alpha) \neq 0$. The theorem now follows from Theorems 10 and 11.

VI. Gramian Matrices

In this section, we shall study the matrices of the normal equations (3.9). The main result is that if the set of observations $A_1, \cdots, A_r$ is linearly independent, then the matrix (called Gramian matrix) is properly positive and has a reciprocal which is also properly positive.

THEOREM 13. *Let $A_1, \cdots, A_r$ be a set of vectors, and let $B_1, \cdots, B_r$ be the orthogonalized set of vectors. Then the matrix*

$$(6.1) \qquad \zeta(A_1, \cdots, A_r) = \begin{pmatrix} (A_1, A_1) & \cdots & (A_1, A_r) \\ \cdots & \cdots & \cdots \\ (A_r, A_1) & \cdots & (A_r, A_r) \end{pmatrix}$$

has the following properties:

13.1) *symmetry*

13.2) $D[\zeta(A_1, \cdots, A_r)] = n(B_1)n(B_2) \cdots n(B_r)$,

13.3) *positiveness.*

A matrix of the form (6.1) is called a Gramian matrix.

In fact, the symmetric property follows from the fact that $(A_i, A_j) = (A_j, A_i)$ for every i, j.

We shall prove 13.2) by induction. For $r = 1$, we have by Theorem 5

$$(A_1, A_1) = (B_1, B_1) = n(B_1).$$

Assume the equality is true for $r = t$, we shall show it is true for $r = t + 1$. The $(t + 1)$-row determinant is as follows:

$$(6.2) \qquad D[\zeta(A_1, \cdots, A_r)] = \begin{vmatrix} (A_1, A_1) & \cdots & (A_1, A_t) & (A_1, A_{t+1}) \\ \cdots & \cdots & \cdots & \cdots \\ (A_1, A_t) & \cdots & (A_t, A_t) & (A_t, A_{t+1}) \\ (A_1, A_{t+1}) & \cdots & (A_t, A_{t+1}) & (A_{t+1}, A_{t+1}) \end{vmatrix}$$

[11] In case $\sigma = (1, \cdots, n)$, $-\sigma$ is a null class Λ (a class which contains no element); then we define $D[\alpha(-\sigma, -\sigma)] = 1$. For the proof of this theorem, see Bocher, p. 31.

By Theorem 5, there exist constants $s_i (i = 1, \cdots, t)$ such that

$$A_{t+1} = B_{t+1} + \sum_{i=1}^{t} s_i A_i .$$

Substituting this value into the last row, we find the element in the i^{th} column is

$$(A_i, A_{t+1}) = \left(A_i, B_{t+1} + \sum_{j=1}^{t} s_j A_j\right) = (A_i, B_{t+1}) + \sum_{j=1}^{t} s_j (A_i, A_j) \qquad (i = 1, \cdots, t, t+1).$$

The second term on the right is a linear combination of the first t elements in the i^{th} column of the determinant (6.2) and hence by the theory of determinants,[12] we secure

$$D[\zeta(A_1, \cdots, A_{t+1})] = \begin{vmatrix} (A_1, A_1) & \cdots & (A_1, A_t) & (A_1, A_{t+1}) \\ \cdots & \cdots & \cdots & \cdots \\ (A_1, A_t) & \cdots & (A_t, A_t) & (A_t, A_{t+1}) \\ (A_1, B_{t+1}) & \cdots & (A_t, B_{t+1}) & (A_{t+1}, B_{t+1}) \end{vmatrix}.$$

By Theorem 5, we find that $(A_i, B_{t+1}) = 0$ for $i = 1, \cdots, t$, and $(A_{t+1}, B_{t+1}) = (B_{t+1}, B_{t+1})$, and hence the preceding determinant reduces to a form in which the first t elements in the $(t + 1)^{\text{th}}$ row are zero. Thus

$$D[\zeta(A_1, \cdots, A_{t+1})] = \begin{vmatrix} (A_1, A_1) & \cdots & (A_1, A_t) \\ \cdots & \cdots & \cdots \\ (A_1, A_t) & \cdots & (A_t, A_t) \end{vmatrix} \cdot n(B_{t+1})$$
$$= n(B_1) n(B_2) \cdots n(B_t) n(B_{t+1})$$

which proves 13.2).

Consider any subset $\sigma = (k_1, \cdots, k_m)$ of the set $(1, \cdots, r)$. By the same argument as above, we find that the determinant of any principal minor

$$(6.3) \qquad \begin{vmatrix} (A_{k1}, A_{k2}) & \cdots & (A_{k1}, A_{km}) \\ \cdots & \cdots & \cdots \\ (A_{km}, A_{k1}) & \cdots & (A_{km}, A_{km}) \end{vmatrix} = n(B_{k1}) \cdots n(B_{km}).$$

By Theorem 1, the number on the right is positive. Thus the matrix ζ is positive.

THEOREM 14. *The following three assertions are equivalent:*

14.1) *the set* $A_1, \cdots, A_r$ *is linearly independent;*

14.2) *the Gramian matrix* (6.1) *is properly positive;*

14.3) *The determinant of the Gramian matrix* (6.1) *is different from zero.*

We shall prove that 14.1) implies 14.2); 14.2) implies 14.3); and 14.3) implies 14.1). We thus prove the three statements are equivalent.

---

[12] Dickson, *First Course in the Theory of Equations* (1922), p. 113.

Let $B_1, \cdots, B_r$ be the orthogonalized set of the set $A_1, \cdots, A_r$. Since the set $A_1, \cdots, A_r$ is linearly independent, then every subset

$$A_{k_1}, \cdots, A_{k_m} (1 \leqq k_1 \leqq \cdots \leqq k_m \leqq r)$$

is also linearly independent, and hence $n(B_{k_i}) > 0$ for $i = 1, 2, \cdots, m$. By the same argument as given in the demonstration of Theorem 11, we find that the determinant of any principal minor (6.3) is greater than zero. This proves the matrix (6.1) is properly positive.

If the matrix (6.1) is properly positive, then by Theorem 10 the determinant of (6.1) is different from zero.

To prove 14.3) implies 14.1), suppose $k_i (i = 1, \cdots, r)$ are such that

$$k_1 A_1 + \cdots + k_r A_r = 0.$$

Then

$$(k_1 A_1 + \cdots + k_r A_r, A_i) = k_1(A_1, A_i) + \cdots + k_r(A_r, A_i) = 0$$

for $i = 1, \cdots, r$. Since $(A_i, A_j) = (A_j, A_i)$, and $D(\zeta) \neq 0$, the set of constants k_i must be all equal to 0.[13]

From Theorem 14, and Theorem 10, we may state the following

COROLLARY: *If the set of observations $A_1, \cdots, A_r$ is linearly independent, then the Gramian matrix ζ has a reciprocal which is properly positive.*

VII. Gauss Method of Substitution

LEMMA 7.1) *Let $\varphi = (s_{ij})$ be an r-row symmetric matrix such that $s_{11} \neq 0$. Then there exists an r-row square matrix τ whose determinant is unity such that $\psi = (r_{ij}) = \tau\varphi$ has the following properties:*

a) *$r_{i1} = 0$ for $i > 1$, and $r_{1i} = s_{1i}$ for every i;*

b) *the first minor of r_{11} is symmetric;*

c) *the determinant of every principal minor in ψ of the form*

$$(7.1) \qquad \begin{vmatrix} s_{11} & s_{1k_2} & \cdots & s_{1k_m} \\ 0 & r_{k_2k_2} & \cdots & r_{k_2k_m} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & r_{k_2k_m} & \cdots & r_{k_mk_m} \end{vmatrix}, \qquad (2 \leqq k_2 \leqq \cdots \leqq k_m \leqq r)$$

is equal to the determinant of the corresponding principal minor in φ.

To prove this lemma, let us define

$$(7.2) \qquad \tau = \delta + F_1 \cdot D_1,$$

where D_1 is the first row of an r-row identity matrix δ, and $F_1(1) = 0$,

$$F_1(n) = -s_{1n}/s_{11} \qquad (n > 1).$$

(Thus $F_1 D_1$ is an r-row square matrix in which the first column is F_1 and everywhere else 0.) It is clear that τ thus defined is a square matrix of order r, and

[13] See footnote 5.

$D(\tau) = D(\delta + F_1 D_1) = 1$. By multiplication of these two matrices, $\tau\varphi$, we obtain a new matrix such that $r_{11} = s_{11}$, $r_{i1} = 0$ for $i > 1$, and $r_{1i} = s_{1i}$ for every i, and further

$$r_{ij} = s_{ij} - s_{1i}\cdot s_{1j}/s_{11} \text{ for } i > 1, j > 1. \tag{7.3}$$

To prove property (b), we note that $s_{ij} = s_{ji}$, since φ is symmetric. Thus for $i > 1, j > 1$, we note from 7.3) that

$$r_{ij} = s_{ij} - s_{1i}s_{1j}/s_{11} = s_{ji} - s_{1j}s_{1i}/s_{11} = r_{ji}.$$

For the proof of the last property, we note that the corresponding minor of (7.1) in φ is of the form

$$\begin{bmatrix} s_{11} & s_{1k_2} & \cdots & s_{1k_m} \\ s_{1k_2} & s_{k_2k_2} & \cdots & s_{k_2k_m} \\ \cdots & \cdots & \cdots & \cdots \\ s_{1k_m} & s_{k_2k_m} & \cdots & s_{k_mk_m} \end{bmatrix} \tag{7.4}$$

Since φ is symmetric, we have by (7.3),

$$r_{k_ik_j} = s_{k_ik_j} - s_{1k_i}s_{1k_j}/s_{11} \qquad (i > 1, j > 1),$$

$$0 = s_{k_i1} - s_{1k_i}s_{11}/s_{11} \qquad (i > 1).$$

Thus by a theorem in the theory of determinants, the determinants of (7.1) and (7.4) are equal.

LEMMA 7.2) *Let $\varphi = (s_{ij})$ $(i, j = 1, \cdots, r)$ be a symmetric matrix of positive type, and $s_{11} \neq 0$. Then there exists an r-row square matrix τ whose determinant is unity such that $\psi = (r_{ij}) = \tau\varphi$ has the properties stated in Lemma* 7.1) *and furthermore the minor of r_{11} in* 7.1) *is of positive type.*

To prove the positiveness of the minor of r_{11}, let the determinant of any one of its principal minors be

$$M_1 = \begin{vmatrix} r_{k_2k_2} & \cdots & r_{k_2k_m} \\ \cdots & \cdots & \cdots \\ r_{k_2k_m} & \cdots & r_{k_mk_m} \end{vmatrix} \qquad (2 \leqq k_1 \leqq \cdots \leqq k_m \leqq r),$$

where $r_{k_ik_j} = r_{k_jk_i}$ $(i, j = 2, \cdots, m)$ due to the symmetry. Now consider the bordered determinant

$$M_2 = \begin{vmatrix} r_{11} & r_{1k_2} & \cdots & r_{1k_m} \\ 0 & r_{k_2k_2} & \cdots & r_{k_2k_m} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & r_{k_2k_m} & \cdots & r_{k_mk_m} \end{vmatrix}$$

which by property (a) in Lemma 7.1) gives $M_2 = r_{11}M_1 = s_{11}M_1$. By property (c) in Lemma 7.1), M_2 is equal to the determinant of the form (7.4), which by hypothesis is positive. Thus $s_{11}M_1 \geqq 0$. Since $s_{11} > 0$, we conclude that $M_1 = M_2/s_{11} \geqq 0$.

LEMMA 7.3). *Let* $\varphi = (s_{ij})$ $(i, j = 1, 2, \cdots, r)$ *be a symmetric matrix of properly positive type. Then there exists an r-row square matrix* τ *whose determinant is unity such that* $\psi = (r_{ij}) = \tau\varphi$ *has the properties stated in Lemma* 7.1) *and furthermore the minor of* r_{11} *in* ψ *is properly positive.*

Since φ is properly positive, we find that $s_{11} > 0$. The proof of this lemma is similar to that of Lemma 7.2).

Suppose that the set of observations $A_1, \cdots, A_r$ is linearly independent. Then by Theorem 14, the Gramian matrix (6.1) is symmetric and properly positive, and hence $(A_1, A_1) > 0$. By Lemma 7.3), the matrix (6.1) can be reduced to the form

$$\begin{bmatrix} [A_1A_1\cdot 0] & [A_1A_2\cdot 0] & \cdots\cdots\cdots & [A_1A_r\cdot 0] \\ 0 & [A_2A_2\cdot 1] & [A_2A_3\cdot 1] \;\cdots & [A_2A_r\cdot 1] \\ \cdots & \cdots & \cdots & \cdots \\ 0 & [A_2A_r\cdot 1] & [A_3A_r\cdot 1] \;\cdots & [A_rA_r\cdot 1] \end{bmatrix} \tag{7.5}$$

where

$$[A_1A_t\cdot 0] = (A_1, A_t) = [A_tA_1\cdot 0] \qquad (t = 1, \cdots, r)$$

$$[A_tA_s\cdot 1] = \frac{[A_1A_1\cdot 0]\,[A_tA_s\cdot 0] - [A_1A_t\cdot 0]\,[A_1A_s\cdot 0]}{[A_1A_1\cdot 0]}.$$

It is evident that $[A_1A_1\cdot 0] = (A_1, A_1) > 0$, since the matrix (6.1) is properly positive. By Lemma 7.3) the value of D ($\mathfrak{z}$) and the determinant of (7.5) are equal, and furthermore the minor of the element $[A_1A_1\cdot 0]$ is a symmetric matrix of properly positive type. Thus $[A_2A_2\cdot 1] > 0$, and $[A_tA_s\cdot 1] = [A_sA_t\cdot 1]$.

The minor of $[A_1A_1\cdot 0]$ surely satisfies all the conditions in Lemma 7.3). We may, therefore, apply a transformation of the form (7.2) to the minor of $[A_1A_1\cdot 0]$, and secure another matrix of the same character as (7.5). In other words, we may multiply on the left of the matrix (7.5) by

$$\tau_2 = \delta + F_2D_2 \tag{7.6}$$

where D_2 is the second row of the r row identity matrix δ, and

$$F_2(n) = 0 \quad (n \leqq 2); \qquad F_2(n) = -\frac{[A_2A_n\cdot 1]}{[A_2A_2\cdot 1]} \qquad (n > 2).$$

In general, let

$$\tau_i = \delta + F_iD_i \qquad (i = 1, \cdots, r-1), \tag{7.7}$$

where D_i is the i^{th} row of the r row identity matrix δ, and

$$F_i(n) = 0 \quad (n \leqq i); \qquad F_i(n) = -\frac{[A_iA_n\cdot i-1]}{[A_iA_i\cdot i-1]} \qquad (n > i). \tag{7.8}$$

Continuous application of this type of transformation ultimately reduces the matrix (6.1) to the form

$$(7.9)\qquad \eta = \begin{bmatrix} [A_1A_1\cdot 0] & [A_1A_2\cdot 0] & [A_1A_3\cdot 0] & \cdots & [A_1A_r\cdot 0] \\ 0 & [A_2A_2\cdot 1] & [A_2A_3\cdot 1] & \cdots & [A_2A_r\cdot 1] \\ 0 & 0 & [A_3A_3\cdot 2] & \cdots & [A_3A_r\cdot 2] \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & [A_rA_r\cdot r-1] \end{bmatrix}$$

where

$$(7.9_1)\qquad [A_tA_s\cdot h] = \frac{[A_hA_h\cdot h-1]\,[A_tA_s\cdot h\text{-}1] - [A_hA_t\cdot h-1]\,[A_hA_s\cdot h\text{-}1]}{[A_hA_h\cdot h-1]}$$

$$(t, s = 1, \cdots, r; \qquad 0 \leqq h \leqq sm(t, s)).$$[14]

In the matrix (7.9), we see by virtue of Lemma 7.3) that $[A_iA_i\cdot i-1] > 0$ for every i, and $[A_tA_s\cdot h] = [A_sA_t\cdot h]$ for every s, t and $0 \leqq h \leqq sm(t, s)$. If $h = sm(t, s)$, then $[A_tA_s\cdot h] = 0$.

Let $\tau = \tau_{r-1}\cdot\tau_{r-2}\cdots\tau_1$. Then by the associative law of multiplication of matrices, we see that

$$(7.10)\qquad \eta = (\tau_{r-1}\cdots\tau_1)\zeta = \tau\zeta.$$

Thus we prove

THEOREM 15. *If the set of vectors $A_1, \cdots, A_r$ is linearly independent, then there exists a square matrix τ of order such that $\tau\zeta$ is of the form* (7.9) *where all elements below the principal diagonal are* 0; *every element in the principal diagonal $[A_iA_i\cdot i-1]$ $(i = 1, \cdots, r)$, is greater than zero; and $[A_tA_s\cdot h] = [A_sA_t\cdot h]$ for $s, t = 1, \cdots, r$, and $h < sm(t, s)$. Furthermore the determinants of the matrices* (6.1) *and* (7.9) *are equal.*

We now prove the following lemma which will be useful in the later section.

LEMMA 7.4). *If $[A_iA_i\cdot i-1]$ is different from zero for every $i \geqq 0$, then for every pair of integers (s, t), where $s, t = 1, \cdots, r$, and $n \leqq sm(t, s)$, we have*

$$\text{a)}\qquad [A_tA_s\cdot n] = (A_t, A_s) - \sum_{i=1}^{n-1} \frac{[A_iA_t\cdot i-1]}{[A_iA_i\cdot i-1]}[A_iA_s\cdot i-1].$$

$$\text{b)}\qquad [A_t(A_s + A_u)\cdot n] = [A_tA_s\cdot n] + [A_tA_u\cdot n], \quad (u = 1, \cdots, r).$$

$$\text{c)}\qquad [(cA_t)A_s\cdot n] = c\,[A_tA_s\cdot n], \qquad (c = a \text{ constant}).$$

To prove a), take every pair (s, t). We find the lemma is true for $n = 0$. Assuming it is true for every (s, t) and for $n = h < sm(s, t)$, we find that $h + 1 \leqq sm(s, t)$, and

$$(A_t, A_s) - \sum_{i=1}^{h+1} \frac{[A_iA_t\cdot i-1]}{[A_iA_i\cdot i-1]}[A_iA_s\cdot i-1]$$

[14] sm (s, t) read "the smaller one of (t, s)."

$$= (A_t, A_s) - \sum_{i=1}^{h} \frac{[A_iA_t \cdot i - 1]}{[A_iA_i \cdot i - 1]} [A_iA_s \cdot i - 1] - \frac{[A_{h+1}A_t \cdot h]}{[A_hA_h \cdot h]} [A_{h+1}A_s \cdot h]$$

$$= [A_tA_s \cdot h] - \frac{[A_{h+1}A_t \cdot h]}{[A_hA_h \cdot h]} [A_{h+1}A_s \cdot h] = [A_tA_s \cdot h + 1],$$

for every s, t.

Parts b) and c) are true for $n = 0$. Now make use of the equality in a) and prove by induction.

VIII. Gauss's Method of Substitution and its Relation to Gramian Schmidt's Orthogonalization Process

Let us write the set of observations in the form:

$$\alpha \equiv \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{r1} & a_{r2} & \cdots & a_{rn} \end{pmatrix}.$$

Let the orthogonalized set also be written in the form

$$\beta = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \cdots & \cdots & \cdots \\ b_{r1} & \cdots & b_{rn} \end{pmatrix}.$$

From Theorems 5 and 6, we find that there exists a transformation κ given by an r-row square matrix such that $\beta = \kappa\alpha$. Thus by the associative law of multiplication of matrices, we have

$$\beta\alpha' = (\kappa\alpha)\alpha' = \kappa(\alpha\alpha').$$

Now the matrix $\alpha\alpha'$ is the Gramian matrix (6.1). Thus

$$\beta\alpha' = \kappa\zeta. \tag{8.1}$$

The composite matrix $\beta\alpha'$ is of the form

$$\begin{bmatrix} (B_1, A_1)(B_1, A_2) & \cdots & (B_1, A_r) \\ (B_2, A_1)(B_2, A_2) & \cdots & (B_2, A_r) \\ \cdots & \cdots & \cdots \\ (B_r, A_1)(B_r, A_2) & \cdots & (B_r, A_r) \end{bmatrix}. \tag{8.2}$$

By Theorems 5 and 6, we note that $(B_s, A_t) = 0$ for $s > t$, and $(B_s, A_s) = (B_s, B_s)$ for every s. Thus the preceding matrix can be written in the form

$$\begin{bmatrix} (B_1, B_1)(B_1, A_2)(B_1, A_3) & \cdots & (B_1, A_r) \\ 0 \quad (B_2, B_2)(B_2, A_3) & \cdots & (B_2, A_r) \\ \cdots & \cdots & \cdots \\ 0 \quad\quad 0 \quad\quad 0 & \cdots & (B_r, B_r) \end{bmatrix}. \tag{8.3}$$

We have proved the following theorem:

THEOREM 16. *Let $A_1, \cdots, A_r$ be a set of vectors, and $B_1, \cdots, B_r$ be the orthogonalized set; and let $\alpha = (a_{ij})$, $\beta = (b_{ij})$. Then there exists a square r-row matrix κ such that $\beta = \kappa\alpha$, and $\kappa\alpha\alpha'$ is a matrix of the form* (8.3) *where all the elements below the principal diagonal are zeros and every element in the principal diagonal is positive. If the set $A_1, \cdots, A_r$ is linearly independent, then every element in the principal diagonal is greater than zero.*

THEOREM 17. *Let $A_1, \cdots, A_r$ be a set of vectors and $B_1, \cdots, B_r$ be the orthogonalized set; and let $\alpha = (a_{ij})$, $\beta = (b_{ij})$. Then $D(\beta\alpha') = D(\alpha\alpha')$.*

For by equations (2.1), we note that $D(\kappa) = 1$. Thus

$$D(\beta\alpha') = D(\kappa\alpha\alpha') = D(\kappa)D(\alpha\alpha') = D(\alpha\alpha') .$$

THEOREM 18. *If the set of vectors, $A_1, \cdots, A_r$ is linearly independent, the matrix κ arising from Gram-Schmidt's orthogonalization process is identical with the matrix τ defined by* (7.10).

To prove this theorem, we first establish the following

LEMMA 8.5): *If the set $A_1, \cdots, A_r$ be linearly independent, and $B_1, \cdots, B_r$ be the orthogonalized set, then for every t, h, we have*

$$(B_h, A_t) = [A_h A_t \cdot h - 1] .$$

By Theorem 10, the set B_i is linearly independent, and hence $n(B_i) > 0$ for every i. The lemma is evidently true for every t and $h = 1$. Assuming it is true for every t and $h = s$, we shall prove it is also true for every t and $h = s - 1$. Now

$$B_{s+1} = A_{s+1} - \sum_{i=1}^{s} \frac{(A_s, B_i)}{(B_i, B_i)} B_i = A_{s+1} - \sum_{i=1}^{s} \frac{[A_i A_s \cdot i - 1]}{[A_i A_i \cdot i - 1]} B_i .$$

Thus by the linear property of (,) we secure, for every t

$$\begin{aligned}(B_{s+1}, A_t) &= \left(A_{s+1} - \sum_{i=1}^{s} \frac{[A_i A_s \cdot i - 1]}{[A_i A_i \cdot i - 1]} B_i, A_t\right) \\ &= (A_{s+1}, A_t) - \sum_{i=1}^{s} \frac{[A_i A_s \cdot i - 1]}{[A_i A_i \cdot i - 1]} (B_i, A_t) \\ &= (A_{s+1}, A_t) - \sum_{i=1}^{s} \frac{[A_i A_s \cdot i - 1]}{[A_i A_i \cdot i - 1]} [A_i A_t \cdot i - 1] \\ &= [A_{s+1} A_t \cdot s]\end{aligned}$$

by virtue of lemma 4.4).

From this lemma, we conclude at once that the matrices (7.9) and (8.3) are equal. Thus by (8.1), we have

$$\kappa\zeta = \beta\alpha' = \tau\zeta, \quad \text{or} \quad (\kappa - \tau)\zeta = \omega .$$

Since ζ is non-singular (by Theorem 12), we have

$$\omega = (\kappa - \tau)\zeta\zeta^{-1} = (\kappa - \tau)\delta = \kappa - \tau,$$

which proves the theorem.

From Lemma 8.5), we have

LEMMA 8.6). Let $L = (l_1, \cdots, l_n)$. Suppose the set $A_1, \cdots, A_r$ to be linearly independent, and $B_1, \cdots, B_r$ to be the orthogonalized set. Then for every h,

$$[A_h L \cdot h - 1] = (B_h, L).$$

Theorems 16, 17, and 18 furnish us a new method for finding the most probable values of the unknowns in the theory of least squares. The formulation of the system of normal equations may be omitted in this new procedure, which may be described briefly as follows: After we obtain a set of observations $A_1, \cdots, A_r$, we orthogonalize this set by means of Gram-Schmidt's process. Let L be a non-zero vector. The product

$$\begin{pmatrix} b_{11} \cdots b_{1n} \\ \cdots\cdots \\ b_{r1} \cdots b_{rn} \end{pmatrix} \begin{pmatrix} a_{11} \cdots a_{r1}, & -l_1 \\ \cdots\cdots & \\ a_{1n} \cdots a_{rn}, & -l_n \end{pmatrix}$$

will give us the result as desired by Gauss's method of substitution.

ACADEMIA SINICA,
PEIPING, CHINA.