THE ORTHOGONAL MATRIX
AND ITS APPLICATIONS

By
Sue Harris Shugart

Submitted as an Honors Paper
in the
Department of Mathematics

THE WOMAN'S COLLEGE OF THE UNIVERSITY OF NORTH CAROLINA
GREENSBORO, NORTH CAROLINA
1953

THE ORTHOGONAL MATRIX
AND ITS APPLICATIONS

By
Sue Harris Shugart

Submitted as an Honors Paper
in the
Department of Mathematics

THE WOMAN'S COLLEGE OF THE UNIVERSITY OF NORTH CAROLINA

GREENSBORO, NORTH CAROLINA

1953

Approved by

Lila Peck Walker.

Director

Examining Committee

Helen Barton

Anna Joyce Reardon

Lila Peck Walker

Table of Contents

1. Introduction.

The subject of matrices and their applications is of great importance, for this branch of mathematics, perhaps more than any other, has been applied to numerous diversified fields, namely, to education, psychology, chemistry, physics, electrical, mechanical, and aeronautical engineering, statistics, and economics. Its importance lies in its use as a "tool" for the mathematician as well as those engaged in other fields. As one would expect, some phases of this theory, such as the theories of quadratic forms and symmetric matrices, are more highly developed than others because of their importance in the study of conic sections and quadric surfaces, in problems of maxima and minima, and in dynamics and statistics.

In this paper, the basic theorems and properties of orthogonal matrices have been set forth and discussed. However, although some theorems on general matrix theory have been used in proofs leading to and including orthogonal matrices, no proofs of these theorems have been presented but they can be found in almost any book on matrix theory.

The orthogonal matrix is approached form the standpoint of vectors, the subject of vectors and vector spaces being undertaken first. In section two some essential, basic definitions of terms in general matrix theory are given. Following this, in the third section, vectors and vector spaces are defined, the various properties of these are treated, and the orthogonal vector is introduced. In section four, the orthogonal matrix and the theorems regarding it are treated at length. Section five introduces orthogonal similarity, the main function of which is to accomplish the important orthogonal reduction of a real symmetric matrix which is treated in section six. Section seven, which deals with the applications of the orthogonal matrix to analytic geometry, begins with the definition of an orthogonal transformation followed by the correlation of this transformation with matrices. From this, we get our geometric interpretation of the orthogonal matrix in the rotation of axes in analytic geometry, and in the reduction of a real quadratic form.

2. Definitions.

The following definitions of the basic terms and concepts in matrix theory are necessary before a discussion of the orthogonal matrix can be undertaken. However, since these terms do not relate directly to orthogonal matrices, they will be listed and defined in this section.

Matrix. If $a_{11}, a_{12}, \ldots, a_{mn}$ are elements over a ring R, the mn elements arranged in a rectangular array of m rows and n columns is called an m by n algebraic matrix over the ring R and is denoted by A.

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ a_{31} & a_{32} & \cdots & a_{3n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Square Matrix. A matrix A is called a square matrix if it has as many rows as columns, that is, A is an n by n matrix.

Diagonal Matrix. An n-square matrix A which has zeros everywhere except in the principal diagonal is called a diagonal matrix.

Non-singular Matrix. A matrix A is non-singular if A is a square matrix and $|A| \neq 0$.

Singular Matrix. A matrix A is singular if A is a square matrix and $|A| = 0$.

Transpose of a Matrix A. If the rows and columns of an m by n matrix A are interchanged, the result is an n by m matrix A called the transpose of A and denoted by A'.

Conjugate of a Matrix A. If each element of an m by n matrix A is replaced by its conjugate complex number, the result is an m by n matrix called the conjugate of A and is denoted by $\bar{A}$. The conjugate of the transpose of a matrix A is called the conjugate transpose and is denoted by A^*.

Inverse of a Non-singular Matrix. If A is a non-singular square matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix};$$

the matrix

$$A^{-1} = \begin{pmatrix} \frac{A_{11}}{|A|} & \frac{A_{21}}{|A|} & \cdots & \frac{A_{n1}}{|A|} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{A_{1n}}{|A|} & \frac{A_{2n}}{|A|} & \cdots & \frac{A_{nn}}{|A|} \end{pmatrix}$$

is called the inverse of the non-singular matrix A where the element in the i-th row and j-th column is $\frac{A_{ji}}{|A|}$, where A_{ij} is the cofactor of a_{ij} in the determinant $|A|$. The inverse is denoted by A^{-1}.

Adjoint of a Square Matrix A. If A is an n-square matrix (a_{ij}) and if A_{ij} denotes the cofactor of a_{ij} in $|A|$, and if the element in the i-th row and j-th column is A_{ji}, then the n-square matrix which is composed of these elements is called the adjoint of A; that is, if $A = (a_{ij})$, then adj. $A = (A_{ji})$.

$$\text{adj. } A = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{pmatrix}$$

Characteristic Equation of a Matrix. The characteristic equation of a matrix A is $|A - \lambda I| = 0$, where λ is a scalar indeterminate and the determinant $|A - \lambda I| = f(\lambda)$ is the characteristic determinant of the characteristic function of A.

Symmetric Matrix. A matrix A which is equal to its transpose is called a symmetric matrix, that is, if $A = A'$.

Skew-symmetric Matrix. A matrix A which is equal to the negative of its transpose is called a skew-symmetric matrix, that is, $A = -A'$.

3. Vectors and Vector Spaces.

To define the orthogonal matrix, we must first develop some basic theorems and ideas regarding vectors.

The concepts of a vector as learned in physics are familiar, for example, displacement, velocity, acceleration, and forces. It is a magnitude that can be represented by a directed line segment. This representation can be thought of as points in a plane represented by ordered pairs (x,y) of real numbers; or, as points in space represented by ordered triples (x,y,z); or, generalizing still further, as points in n-dimensional space represented by ordered n-tuples. We can say then that a sequence $X=(x_1,x_2,\ldots,x_n)$ of n numbers x_i is an n-dimensional vector or, as frequently called, an n-tuple where $x_1,\ldots,x_n$ are the coordinates of X and x_i is the i-th component. If we consider ordered n-tuples $X=(x_1,\ldots,x_n)$ of elements x_i lying in a field F, each n-tuple is called a vector with n components x_i. The totality, which we shall call $V_n(F)$, of all vectors, for a fixed integer n, is called the n-dimensional vector space over F. F can be taken to be $R^{\#}$, the field of real numbers; then, the resulting vector space $V_n(R^{\#})$ is frequently called the n-dimensional Euclidean space. A vector or linear space V over a field F is a set of elements, called vectors, such that any two vectors X and Y of V determine a unique vector $X+Y$ as a sum, and that any vector X from V and any scalar c from F determine a scalar product $c.X$ in V, with the following five properties:

(1) $X+Y=Y+X$

(2) $c(X+Y)=cX+cY$

(3) $(cc')X=c(c'X)$

(4) $(c+c')X=c.X+c'.X$

(5) $1.X=X.$

A subspace of $V_n(R)$ is a subset V (containing at least one vector) of $V_n(R)$, when V is closed under addition and scalar multiplication, that is, if the sum of any two vectors of V lies in V, and if the product of any vector of V by a scalar lies in V.

A set of vectors $(X_1, X_2, \ldots, X_k)$ of a vector space V over a field R is linearly dependent over R if there are scalars c_i in R, not all 0, such that $\sum_1^k c_i X_i = 0$. Otherwise the set is linearly independent. From the statements previously made, we can say that $V_n(R)$ is a space of n-dimensions or that the dimensions of $V_n(R)$ are n. $V_n(R)$ is then the totality of 1 by n matrices over R or, writing the vector X as a column, the totality of n by 1 matrices over R. We can then say that a vector is a one column or one row matrix where a column vector is a n by 1 matrix and is denoted by X or $[x_1, x_2, \ldots, x_n]$, and a row vector is a 1 by n matrix and is denoted by X' or $(x_1, x_2, \ldots, x_n)$.

Let the field R of all real numbers be a field of scalars. If X and Y belong to $V_n(R)$, a particular scalar will be associated with them. Let $X = [x_1, x_2, \ldots, x_n]$ and $Y = [y_1, y_2, \ldots, y_n]$. Then the scalar $X'Y = x_1y_1 + x_2y_2 + \cdots + x_ny_n$ is called the inner product of X and Y, sometimes written (X,Y). It is clearly seen that the inner product of X with itself is a sum of squares:

$$X'X = (x_1, x_2, x_3, \ldots, x_n)[x_1, x_2, \ldots, x_n] = x_1^2 + x_2^2 + \cdots + x_n^2.$$

Also, if we restrict the components of X to real numbers only, this sum of squares of real numbers x_i will never be zero unless every $x_i = 0$, which leads to the theorem

<u>Theorem 3.1.</u> The inner product of a vector of $V_n(R)$ with itself is positive unless the vector is zero.

Four important algebraic properties of inner products which are immediate consequences of the definition that $(X'Y) = x_1y_1 + x_2y_2 + \ldots + x_ny_n$ are:

(1) $(X+Y,Z) = (X,Z) + (Y,Z)$

(2) $c(X,Y) = c(Y,X)$

(3) $(X,Y) = (Y,X)$

(4) $(X,X) > 0$ unless $X = 0$.

By the first and second the linearity of the left-hand factor of inner products is represented. The third, together with the first and second, shows the linearity of inner products in both factors. The fourth shows the positiveness of inner products.

If X is any vector in $V_n(R)$, the square root of the inner product of X' and X, $(X'X)^{1/2}$, is defined as the length of X. Since positive real numbers have square roots in the real number system, $(X'X)^{1/2}$ may always be taken as positive. A normal vector is one whose length is unity.

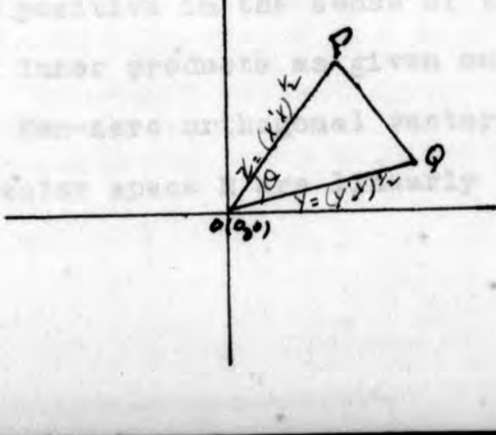
Two n-dimensional vectors, $X = [x_1, x_2, \ldots, x_n]$, $Y = [y_1, y_2, \ldots, y_n]$, of $V_n(R)$ are said to be orthogonal to each other if their inner product is zero, that is, $x_1y_1 + x_2y_2 + \ldots + x_ny_n = 0$, or, in matrix terminology, if $Y'X = X'Y = 0$. If X is orthogonal to one or more vectors, it is orthogonal to the linear system which they define. Extending this idea further, it can be said that all the vectors orthogonal to a linear system constitute a linear system called the orthogonal complement of the first linear system. The vectors, $X_1, X_2, \ldots, X_n$ are called normal orthogonal when (1) $|X_i| = 1$ for all i, (2) X_i is perpendicular to X_j where $i \neq j$.

These concepts of length and orthogonality are the same as those studied in analytic geometry; for, if we take the vector (x_1, x_2) of $V_2(R)$, it can be thought of as a line from the origin to a point whose coordinates in a rectangular coordinate system are (x_1, x_2). The length of this line is $(x_1^2 + x_2^2)^{1/2}$. If we consider the slope of this line, $\frac{x_2}{x_1}$, and the slope of another line from the origin to the point (y_1, y_2), $\frac{y_2}{y_1}$, we have for the two lines to be orthogonal, that $\frac{x_2}{x_1} \cdot \frac{y_2}{y_1} = -1$, or that $x_1 y_1 + x_2 y_2 = 0$, which is the necessary condition for two vectors to be orthogonal. The same interpretation can be extended to $V_3(R)$. Moreover, these properties of orthogonal vectors are applied in plane analytic geometry in proving that the cosine of the angle between two orthogonal vectors is zero. To prove this, let P and Q be any two points distinct from the origin and θ be the angle between the two points in rectangular Cartesian coordinates, the origin being denoted by the column vector (0,0), the point P by $X = (x_1, x_2)$, and the point Q by $Y = (y_1, y_2)$. The lengths of X and Y are :

$$|X| = (x_1^2 + x_2^2)^{1/2} = (X'X)^{1/2} = (X,X)^{1/2}$$

$$|Y| = (y_1^2 + y_2^2)^{1/2} = (Y'Y)^{1/2} = (Y,Y)^{1/2}.$$

If we join P and Q, we form the triangle OPQ with sides X,Y, $PQ = X - Y$.



From the trigonometric law of cosines, we have

(1) $|X-Y|^2 = |X|^2 + |Y|^2 - 2|X| . |Y| . \cos\theta$

which we can also write as

(2) $|X-Y|^2 = (X-Y, X-Y) = (X,X) + (Y,Y) - 2(X,Y)$.

Combining (1) and (2), we get

(3) $\cos\theta = \cos\angle(X,Y) = \dfrac{(X,Y)}{|X| . |Y|}$.

It follows immediately that if X and Y are orthogonal vectors, then $(X,Y) = 0$, and the $\cos\angle(X,Y) = 0$. Conversely, if $(X,Y) = 0$, then $\cos\angle(X,Y) = 0$ and X and Y are orthogonal.

In generalizing this case of two vectors to one of n vectors, we encounter some complications. It is relatively easy to see that in the case of three dimensional space, formula (3) becomes

$$\cos\angle(X,Y,Z) = \frac{(X,Y) + (X,Z) + (Y,Z)}{|X| . |Y| . |Z|} .$$

However, although the inner products can always be defined, the lengths $|X| = (X, X)^{\frac{1}{2}}$ are not definable unless every sum of n squares has a square root. For this reason, we shall confine ourselves to vector spaces over the real field. One such vector space, which is mentioned in the following theorems, is called Euclidean and is a vector space R with real scalars, such that, to any vectors X and Y in R, corresponds a real inner product (X,Y) which is symmetric, bilinear, and positive in the sense of the four algebraic properties of inner products as given on page seven.

Theorem 3.2. Non-zero orthogonal vectors $X_1, X_2, \ldots, X_n$ of a Euclidean vector space R are linearly independent.

If $\alpha_1 x_1 + \dots + \alpha_n x_n = 0$, then for $k = 1,\dots,m$, $0 = (0,x_k) = \alpha_1(x_1,x_k) + \dots + \alpha_m(x_m,x_k) = \alpha_k(x_k,x_k)$ where $\alpha_k(x_k,x_k)$ comes from the orthogonality assumption. But it was assumed that $x_k \neq 0$, therefore $(x_k,x_k) > 0$ and $\alpha_k = 0$. From this theorem, it follows that normal orthogonal vectors spanning R, are a basis for R. This basis is called the normal orthogonal basis.

From this discussion of normal orthogonal vectors, we are now in the position to define an orthogonal matrix. Let it be understood that the approach to the study of orthogonal matrices could be reversed, that is, the orthogonal matrix could be defined, and from that standpoint orthogonal vectors discussed.

4. Orthogonal Matrices.

A square matrix A is orthogonal as to rows if and only if each row of A has length one, and any two rows are orthogonal. $A = (a_{ij})$ is orthogonal if and only if:
(1) $\sum_{k=1}^{n} a_{ik} a_{ik} = 1$ for all i, (2) $\sum_{k=1}^{n} a_{ik} a_{jk} = 0$, if $i \neq j$.
Making use of the Kronecker delta δ_{ij}, the above equations become:

$$\sum_{k=1}^{n} a_{ik} a_{jk} = \delta_{ij} \quad (i,j = 1,2,\dots,n) \qquad \begin{array}{l}\delta_{ij} = 1, \text{ if } i = j \\ \delta_{ij} = 0, \text{ if } i \neq j.\end{array}$$

A square matrix A is orthogonal as to columns if

$$\sum_{k=1}^{n} a_{ki} a_{kj} = \delta_{ij} \quad (i,j = 1,2,\dots,n).$$

Writing A_i for the i-th row of the matrix A and A_j' for its transpose, the inner product of A_i by A_j is the matrix product $A_i \cdot A_j'$. Then the above equations can be written

$$(1) \quad A_i A_j' = \delta_{ij}$$

where δ_{ij} is the element in the i-th row and j-th column of the identity matrix $I = (\delta_{ij})$. A row-by-column interpretation of the equation (1) leads to the theorem

<u>Theorem 4.1.</u> A matrix A is orthogonal if and only if $A A' = I$.

<u>Theorem 4.2.</u> A matrix A is orthogonal if its inverse is the same as its transpose, that is, $A' = A^{-1}$.

Since $A A' = I$, A is non-singular. Multipling both sides of the equation on the left by A^{-1}, we have

$$A^{-1}AA' = A^{-1}I .$$

But by a definition on the inverse of a matrix, $A^{-1}A = I$.*
Whence, $A' = A^{-1}$.

The above definitions and theorems have defined orthogonal matrices in several ways. To summarize, a square matrix A is orthogonal if:

(1) $\sum_{k=1}^{n} a_{ik} a_{jk} = \delta_{ij} \quad (i, j = 1, 2, \ldots, n)$

(2) $\sum_{k=1}^{n} a_{ki} a_{kj} = \delta_{ij}$

(3) $AA' = I$

(4) $A' = A^{-1}$.

It may be pointed out that any one of the above implies the other three.

Two examples of orthogonal matrices are : (1) the unit matrix I which is the simplest and (2) the matrix of the rotation transformation from plane analytic geometry.

* Hereafter, all material referred to in a footnote denoted by an asterisk may be found in any of the books on matrix theory listed in the bibliography, in particular numbers 5 and 9.

(1) $I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

(2) $x' = x \cos \theta + y \sin \theta$
$y' = -x \sin \theta + y \cos \theta$ which gives the matrix

$$A = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}.$$

The latter will be discussed in more detail in section seven of this paper.

Having arrived at the definition of the orthogonal matrix, we are now in the position to prove some interesting properties that belong to this particular type of matrix.

Theorem 4.3. The transpose of an orthogonal matrix is orthogonal.

If A' is orthogonal, then, by theorem 4.1

$$A'(A')^{-1} = A'A = I = A^{-1}A,$$

for by theorem 4.2, $A' = A^{-1}$ which is a defining property of an orthogonal matrix.

Theorem 4.4. If a matrix A is orthogonal as to its rows, it is orthogonal as to its columns; and vice versa.

This theorem follows directly from theorem 4.3, that is, the transpose of A is orthogonal.

Theorem 4.5. The inverse of an orthogonal matrix is orthogonal.

Let A be an n by n orthogonal matrix. Then $A' = A^{-1}$. By theorem 4.3 A' is an orthogonal matrix, therefore A^{-1}, the inverse of A, is also an orthogonal matrix.

Theorem 4.6. The product of the orthogonal n by n matrices A and B is orthogonal.

$$(AB)' = B'A' = B^{-1}A^{-1} = (AB)^{-1}.$$

Therefore AB is orthogonal.

At this point, it is appropriate to say a word about the group formed by orthogonal matrices. All n by n orthogonal matrices form a group, for they conform to the definition of a group of transformations which states that a group of transformations on a space V is a set of one-one transformations on V which includes the identity, with any transformation its inverse, and with any two transformations their product. For an orthogonal matrix, the n-th row identity matrix is orthogonal, the inverse A^{-1} is orthogonal, and the product of two orthogonal matrices is orthogonal. The group of all these n by n orthogonal matrices form a subgroup of the full linear group $V_n(R^{*})$ and is called the orthogonal group O_n.

<u>Theorem 4.7.</u> If A and B are orthogonal matrices, then

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \text{ is an orthogonal matrix.}$$

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}' = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}\begin{pmatrix} A' & 0 \\ 0 & B' \end{pmatrix} = \begin{pmatrix} AA' & 0 \\ 0 & BB' \end{pmatrix} = I.$$

<u>Theorem 4.8.</u> The determinant of an orthogonal matrix is equal to $\pm$ 1.

Letting P represent an orthogonal matrix, we have

$$P'P = I.$$

Then

$$|P'P| = |P'|\,|P| = |P|^2 = |I| = +1.$$

$$|P| = \pm 1$$

Before giving the theorems that concern the characteristic equation and roots of an orthogonal matrix, we must define a reciprocal equation.

An algebraic equation $f(\lambda) = 0$ of degree n is a reciprocal equation provided $f(\lambda) = \pm \lambda^n f(1/\lambda)$.

<u>Theorem</u> <u>4.9.</u> The characteristic equation of an orthogonal matrix is a reciprocal equation.

From the characteristic matrix of P, we have

$$P-\lambda I = P\lambda(1/\lambda\ I-P') = -P\lambda(P'-1/\lambda\ I).$$

On taking determinants of both sides, it becomes

$$f(\lambda) = |P-\lambda I| = |P|(-\lambda)^n|P'-1/\lambda\ I| = \pm(\lambda)^n|P-1/\lambda\ I| = \pm\lambda^n f(1/\lambda).$$

<u>Theorem</u> <u>4.10.</u> The characteristic roots of a real orthogonal matrix are of modulus unity.

For every characteristic root of a real orthogonal matrix P, there is a column vector $X \neq 0$ such that

$$(1)\quad PX = \alpha X.$$

Taking the conjugate transpose of both members of equation (1), it becomes

$$(2)\quad X^*P' = \bar{\alpha} X^*.$$

Multiplying the 1 by n matrix in (2) by the n by 1 matrix in (1),

$$X^*P'PX = \alpha\bar{\alpha} X^*X.$$

But $PP' = I$. Therefore, $X^*X = \alpha\bar{\alpha} X^*X$. Since X^*X is a non-zero 1 by 1 matrix, $\alpha\bar{\alpha} = 1$. Following directly from this theorem, we have

<u>Theorem</u> <u>4.11.</u> A real orthogonal matrix has no real characteristic roots other than ± 1.

For if $\alpha\bar{\alpha} = 1$, then α must equal ± 1, if α is real.

Two convenient terms which we shall make use of in the following discussions are proper and improper. An orthogonal matrix which does not have -1 as a characteristic root is called proper. It is equivalent to saying that $I+P$ is non-singular. An orthogonal matrix which does have -1 as a characteristic root is called improper.

Theorem 4.12. If T is any real skew-symmetric matrix and k is a real number not equal to 0, then the matrix $P=(kI+T)^{-1}(kI-T)$ is properly orthogonal.

By a theorem on skew-symmetric matrices which states that a real skew-symmetric matrix has no real characteristic root other than zero, $kI+T$ and $kI-T$ are non-singular.* Multiplying both members on the right and left of the matrix identity

$$(1)\quad (kI+T)(kI-T)=(kI-T)(kI+T)$$

by $(kI-T)^{-1}$, we have

$$(2)\quad (kI+T)(kI-T)^{-1}=(kI-T)^{-1}(kI+T).$$

Taking the inverse and transpose of

$$(3)\quad P=(kI+T)^{-1}(kI-T)$$

we have,

$$(4)\quad P^{-1}=(kI-T)^{-1}(kI+T)$$

or, on taking the inverse and using (2),

$$(5)\quad P=(kI-T)(kI+T)^{-1}.$$

Since T is real skew-symmetric, $T'=-T$, and

$$(6)\quad P'=(kI+T)(kI-T)^{-1}.$$

But the right hand members of (4) and (6) are equal. Hence $P^{-1}=P'$, so that P is orthogonal.

Theorem 4.13. If P is a real properly orthogonal matrix, there exists a real skew-symmetric matrix T such that P is given by

$$(1)\quad P=(kI+T)^{-1}(kI-T).$$

Rewriting (1), we have

$$\begin{aligned}(kI+T)P &= kI-T\\ T+TP &= kI-kIP\\ T &= kI(I-P)(I+P)^{-1}.\end{aligned}$$

Since P is properly orthogonal, $I+P$ is non-singular.

$$\begin{aligned}T' &= (I+P')^{-1}\,kI(I-P')\\ T' &= kI(I+P^{-1})(I-P^{-1})\\ T' &= kI(I+P^{-1})\quad P^{-1}P(I-P^{-1})\\ T' &= kI\,[P(I+P^{-1})]^{-1}\quad(P-I)\\ T' &= kI(P+I)^{-1}\,(P-I)\\ T' &= -kI(I+P)^{-1}\,(I-P)\end{aligned}$$

Therefore $T=-T'$ and T is a real skew-symmetric matrix.

On taking the formula $P=(kI+T)^{-1}(kI-T)$, and varying the conditions placed on the matrix T, one can arrive at several interesting results. One of these is the derivation of a formula for all 3 by 3 orthogonal matrices. Let us take $T=\begin{pmatrix}0 & c & -b\\ -c & 0 & a\\ b & -a & 0\end{pmatrix}$, find such a formula, and show that the vector (a,b,c) is an absolutely invariant vector of each matrix. Substituting the value of T and evaluating the inverse, we have

$$P=\frac{1}{k(k^2+a^2+b^2+c^2)}\begin{pmatrix}k^2+a^2 & -ck+ab & ca+kb\\ ck+ab & k^2+b^2 & -ka+ba\\ ac-bk & ak+bc & k^2+c^2\end{pmatrix}\begin{pmatrix}k & -c & b\\ c & k & -a\\ -b & a & k\end{pmatrix}.$$

From which

$$P=\frac{1}{k^2+a^2+b^2+c^2}\begin{pmatrix}k^2+a^2-b^2-c^2 & -2ck+2ab & 2bk+2ac\\ 2ck+2ab & k^2-a^2+b^2-c^2 & -2ak+2bc\\ -2bk+2ac & 2ak+2bc & k^2-a^2-b^2+c^2\end{pmatrix},$$

the desired formula.

If the vector (a,b,c) is an absolutely invariant vector of P, then $P\begin{pmatrix}a\\b\\c\end{pmatrix}=\begin{pmatrix}a\\b\\c\end{pmatrix}$.

$$P\begin{pmatrix}a\\b\\c\end{pmatrix}=\frac{1}{k^2+a^2+b^2+c^2}\begin{pmatrix}k^2+a^2-b^2-c^2 & -2ck+2ab & 2bk+2ac\\ 2ck+2ab & k^2-a^2+b^2-c^2 & -2ak+2bc\\ -2bk+2ac & 2ak+2bc & k^2-a^2-b^2+c^2\end{pmatrix}\begin{pmatrix}a\\b\\c\end{pmatrix}$$

$$=\frac{1}{k^2+a^2+b^2+c^2}\begin{pmatrix}ak^2+a^3-ab^2-ac^2-2cbk+2ab^2+2cbk+2ac^2\\ 2ack+2a^2b-bk^2-ba^2+b^3-bc^2-2ack-2bc\\ -2abk+2a^2c+2abk+2b^2c+ck^2-a^2c-cb^2+c^3\end{pmatrix}=\begin{pmatrix}a\\b\\c\end{pmatrix}.$$

Therefore, $P\begin{pmatrix}a\\b\\c\end{pmatrix}=\begin{pmatrix}a\\b\\c\end{pmatrix}$. Hence the vector (a,b,c) is an absolutely invariant vector of the matrix.

<u>Theorem 4.15.</u> Characteristic vectors, X,Y,... corresponding to distinct characteristic roots $\alpha_1, \alpha_2, \ldots, \alpha_n$ of a real, symmetric matrix are orthogonal.

(1) $AX=\alpha_1 X$, (2) $AY=\alpha_2 Y$ $(\alpha_1 \neq \alpha_2)$.

Multiplying each member of (1) on the left by Y' and each member of (2) on the left by X', we have

(3) $Y'AX=\alpha_1 Y'X$, (4) $X'AY=\alpha_2 X'Y$.

Taking the transpose of (4)

(5) $Y'A'X=\alpha_2 Y'X$.

Since A is symmetric, $A'=A$. Therefore $\alpha_1 Y'X=\alpha_2 Y'X$. Since $\alpha_1 \neq \alpha_2$, $Y'X=0$, or the inner product of the two vectors is zero. Therefore X and Y are orthogonal.

One cannot go far in the study of orthogonal matrices before the idea of orthogonal similarity is introduced. We are now at this point, but must refer again to the relationship of orthogonal vectors and matrices before the important orthogonal reduction of a symmetric matrix can be undertaken.

5. <u>Orthogonal Similarity.</u>

If $B = P^{-1}AP$, where P is orthogonal, B is said to be orthogonally similar to A. If we take the relationships $P^{-1}AP = D = \text{diag}(\lambda_1, \ldots, \lambda_n)$, then $AP = PD$, and $AP_i = \lambda_i P_i$, where P_i is the i-th column of P. From this, $P_1, \ldots, P_n$ are characteristic vectors of A, and from definition of an orthonormal basis, these vectors form an orthonormal basis for $V_n(R)$.

<u>Theorem 5.1.</u> If A is an n-square matrix over R and if the set of characteristic vectors of A includes an orthonormal basis for $V_n(R)$, then A is orthogonally similar to a diagonal matrix; and conversely.

By the preceding definition, the converse of the theorem was proved. The proof of the theorem follows.

Let $X_1, \ldots, X_n$ be characteristic vectors of A forming an orthonormal basis for $V_n(R)$. Then

$$AX_i = \lambda_i X_i \quad (i = 1, \ldots, n)$$

$$\text{or } AP = PD,\ D = \text{diag}(\lambda_1, \ldots, \lambda_n),$$

where P is the matrix whose i-th column is X_i, $(i = 1, \ldots, n)$. But the X_i are mutually orthonormal vectors by hypothesis; therefore, P is orthogonal, and

$$P^{-1}AP = D, \text{ and } P^{-1} = P'.$$

<u>Theorem 5.2.</u> If $X_1, \ldots, X_k$ are mutually orthogonal non-zero vectors of $V_n(R)$, they are linearly independent.

Let $a_1 X_1 + \ldots + a_k X_k = 0$, where a_i is a scalar. Since

$$X_i' X_j = 0, \quad (i \neq j)$$

$$X_i' . 0 = 0 = X_i' (a_1 X_1 + \ldots + a_k X_k)$$

$$a_1 X_i' X_1 + \ldots + a_k X_i' X_k = 0$$

$$a_i X_i' X_i = 0$$

By hypothesis, $X_i \neq 0$; therefore, $a_i = 0$, $(i = 1,\ldots,t)$. Therefore, the set of vectors $(X_1,\ldots,X_i)$ is linearly independent.

Theorem 5.3. Every set of mutually orthogonal, normal vectors of V may be extended to an orthonormal basis of V. Every non-zero vector space V over the field of real numbers has an orthonormal basis.

Let $X_1,\ldots,X_i$, where $i \geq 1$, be a set of mutually orthogonal, normal vectors of V. If $i = 1$, the X_1 must be a vector of length one. We know i cannot be greater than the dimension, t, of V. Let i be less than the dimension, t, of V. If i is less than t, the subspace spanned by $X_1,\ldots,X_i$ will not include all of V; that is, there will be some vector Y not in this subspace. Therefore, there is no vector X of the form

$$(1) \quad X = Y - a_1 X_1 - \ldots - a_i X_i,$$

where a_j is a scalar, that can equal 0. If the scalars a_j are chosen to be $a_j = c_j^{-1} X_j{}'Y$, $c_j = X_j{}'X_j$, $(j = 1,\ldots,i)$. Multiplying (1) on the left by $X_j{}'$, we have

$$X_j{}'X = X_j{}'Y - \sum_{k=1}^{i} a_k X_j{}'X_k = X_j{}'Y - a_j X_j{}'X_j = X_j{}'Y - a_j c_j = 0.$$

Therefore, all the vectors $X_1,\ldots,X_i$ are orthogonal to X and to $X_{i+1} = c^{-1}X$, where c is the length of X. By this process, $X_1,\ldots,X_i$ has been extended to $X_1,\ldots,X_i,X_{i+1}$, of mutually orthogonal, normal vectors of V. As long as $i+1 < t$, the process can be repeated and the desired basis can be constructed by a finite number of repetitions.

For proof of the second part of the theorem, it can be noted that every non-zero vector space V has a vector

$X \neq 0$, therefore, it will have a vector X_1 which will be normal. This will give a set of vectors to which the process of the first part of the theorem can be applied.

This theorem 5.3 leads to a method by which a real orthogonal matrix P, of order n, can be built up so that the elements in the first column are proportional to any set of real numbers $X_1 = (x_1, x_2, \ldots, x_n)$, not all 0. The theoretical explanation of this follows.

By hypothesis, $X_1 = (x_1, x_2, \ldots, x_n)$ is a set of real numbers. If $\sum x_i^2 = 1$, let X_1 be the first column of P. If $\sum x_i^2 \neq 1 = k$, normalize the set by dividing through by $\sqrt{k}$, and use the set obtained as the first column $[(x_1, x_2, \ldots, x_n)]$ of P. To make the second column vector orthogonal to the first column vector, find a real non-zero solution $(y_1, y_2, \ldots, y_n)$ of $x_{11} y_1 + x_{21} y_2 + \ldots + x_{n1} y_n = 0$. If necessary normalize it by dividing through by $\sqrt{\sum y_i^2}$, and use the set obtained as the second column of P. Repeat this process until $S \leqq n-1$ columns of P $[(x_{1i}, x_{2i}, \ldots, x_{ni})]$, $(i = 1, 2, \ldots, s)$ are obtained. For the $(s+1)$-th column, find a real solution $(z_1, z_2, \ldots, z_n) \neq (0, 0, \ldots, 0)$ of the set of $s < n$ homogenous linear equations.

$$x_{1i} z_1 + x_{2i} z_2 + \ldots + x_{ni} z_n = 0, \ (i = 1, 2, \ldots, s).$$

On normalizing this set and using it as the $s+1$-th column of P, the real n-square orthogonal matrix P is constructed. We know that P is orthogonal because we have built up its columns orthogonally.

For an illustration of this procedure, let us construct a real orthogonal matrix P whose first column is proportional to the set (1,2,3).

Since $\sum x_i^2 = 14 \neq 1$, the set must be normalized by dividing through by $\sqrt{\sum x_i^2} = \sqrt{14}$. The set obtained, $\left(\frac{1}{\sqrt{14}}, \frac{2}{\sqrt{14}}, \frac{3}{\sqrt{14}}\right)$, is used as the first column of P. To find the second column, obtain a solution of the equation $y_1 + 2y_2 + 3y_3 = 0$, say (1,1,-1) and normalize it. This set, $\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}\right)$, is the second column of P. To find the third column, select a solution of the two equations:

$$z_1 + 2z_2 + 3z_3 = 0$$

$$z_1 + z_2 - z_3 = 0 \quad ,$$

say (5,-4,1) and normalize it. This set, $\left(\frac{5}{\sqrt{42}}, \frac{-4}{\sqrt{42}}, \frac{1}{\sqrt{42}}\right)$, is the third column of P. The required real orthogonal matrix P is

$$P = \begin{pmatrix} \frac{1}{\sqrt{14}} & \frac{1}{\sqrt{3}} & \frac{5}{\sqrt{42}} \\ \frac{2}{\sqrt{14}} & \frac{1}{\sqrt{3}} & \frac{-4}{\sqrt{42}} \\ \frac{3}{\sqrt{14}} & \frac{-1}{\sqrt{3}} & \frac{1}{\sqrt{42}} \end{pmatrix}$$

6. <u>Orthogonal Reduction of a Symmetric Matrix.</u>

A real symmetric matrix A is orthogonally equivalent, or similar, to a real symmetric matrix B if there is an orthogonal matrix P with real elements such that

$$B = P^{-1}AP = P'AP \ .$$

The matrix $B = P^{-1}AP$ is similar to A and has the same characteristic roots as A. We now have the necessary material to state and prove the well known result about a real symmetric matrix.

Theorem 6.1. Every real symmetric matrix A is orthogonally equivalent to the diagonal matrix $D = \text{diag}(\alpha_1, \ldots, \alpha_n)$, where $\alpha_1, \ldots, \alpha_n$ are the characteristic roots of A arranged in any prescribed order. In equation form,

$$P^{-1}AP = \text{diag}(\alpha_1, \ldots, \alpha_n).$$

It is obvious that if A is a 1 by 1 matrix, it is already in diagonal form. We shall prove this theorem by the method of induction, assuming that it is true for a matrix of order n-1 and proving that it is also true for a matrix of order n.

It is known that, if A is a real symmetric matrix, its characteristic roots, $\alpha_1, \ldots, \alpha_n$ are real numbers.* Since $|A - \alpha_1 I| = 0$, the equation

$$(1) \quad AX_1 = \alpha_1 X_1$$

has a non-zero solution, which we can take, after normalizing, to be a single-column unit vector X_1 with real elements $(x_{11}, x_{21}, \ldots, x_{n1})$. We take this unit vector X_1 as the first column $(x_{11}, x_{21}, \ldots, x_{n1})$ of a real orthogonal matrix Q. The remaining columns of Q can be built up by the preceding method of constructing a real orthogonal matrix Q whose first column is proportional to a given set.

In forming the matrix product $Q'AQ$, let us first form the product AQ. But we know that the first column of Q is the vector X_1 of equation (1); therefore, the first column of the matrix AQ is the vector $\alpha_1 X_1 = [\alpha_1 x_{11}, \alpha_1 x_{21}, \ldots, \alpha_1 x_{n1}]$, and the first column of the matrix $Q'AQ$ will be the inner products of successive row vectors of Q'; that is, the successive column vectors of Q with the vector $\alpha_1 X_1$. By the definition of an orthogonal matrix that $\sum_{t=1}^{n} p_{ti}\, p_{tj} = \delta_{ij}$, $[\alpha_1, 0, \ldots, 0]$ is the first column of $Q'AQ$. Symbolically

$$Q'AQ = \left(\begin{array}{c|ccc} \alpha_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{array}\right),$$

the stars representing undetermined elements of the matrix. We now have need of a theorem on symmetric matrices which we shall state without proof. This theorem is as follows: If an n-square matrix A is symmetric and P is an n by n matrix, then $P'AP$ is symmetric.* Since A is a real symmetric matrix, and Q is a real n-square matrix, $Q'AQ$ is real symmetric. Then we can say that the starred elements are all zero, and A_1 is a real symmetric matrix of order n-1. Again, we have need of a theorem which shall be taken without proof. This theorem states that the characteristic function of a matrix A is identical with that of any of its transforms.* From this theorem, it follows immediately that the characteristic roots of $Q'AQ$ are the same as those of A, and the characteristic roots of the (n-1)-rowed real symmetric matrix A_1 are $\alpha_2, \alpha_3, \ldots, \alpha_n$. Now in the beginning of this proof, we assumed that there was a real orthogonal matrix of order n-1 which satisfied the relationship

$$R'A_1R = \operatorname{diag}(\alpha_2, \alpha_3, \ldots, \alpha_n),$$

then we can write

$$S = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & R \end{array}\right).$$

Therefore, S is an n-rowed real orthogonal matrix which satisfies the equation

$$(2)\quad S'(Q'AQ)S = \left(\begin{array}{c|c} \alpha_1 & 0 \\ \hline 0 & R'A_1R \end{array}\right).$$

Letting $P = QS$, by theorem 4.6, P will be a real orthogonal matrix. Substituting $P = QS$ in (2), we obtain

$P'AP = \text{diag}(\alpha_1, \alpha_2, \ldots, \alpha_n)$, the desired result.

The proof above is rather clumsy for use in finding an orthogonal matrix P such that $P'AP = \text{diag}(\alpha_1, \alpha_2, \ldots, \alpha_n)$. The procedure that is easier to employ is that of building up a real orthogonal matrix whose first column is proportional to any set of numbers and whose columns are pairwise orthogonal non-zero vectors. The method is explained below, followed by a numerical example.

Let A be a real symmetric matrix of order n which has the characteristic roots $\alpha_1, \alpha_2, \ldots, \alpha_t$ of multiplicities $\gamma_1, \gamma_2, \ldots, \gamma_t$, respectively (where $\sum \gamma_i = n$). By a theorem in matrix theory, we know that two invariant vectors of a real symmetric matrix arising from two distinct characteristic roots are orthogonal.* Another theorem which we shall need for proof of the next step in this method and which we shall state without proof is as follows: If A is an n-square matrix having the characteristic roots $\alpha_1, \alpha_2, \ldots, \alpha_t$, of multiplicities $\gamma_1, \gamma_2, \ldots, \gamma_t$, respectively, a necessary and sufficient condition that A be similar to a diagonal matrix is that for each root α, the matrix $A - \alpha I$ be of rank $n - \gamma$.* By this theorem, we can state that, since A is similar to a diagonal matrix, and if α is a characteristic root of multiplicity $\nu > 1$ of A, then the matrix $A-\alpha I$ must be of rank $n-\nu$. We have now to show that we can choose the ν invariant vectors of the characteristic root α so that they will be orthogonal to each other. To do this, let $X = (x_{11}, x_{21}, \ldots, x_{n1})$ be one real non-zero solution of the set of $n-\nu$ linearly independent equations

$$(1) \quad AY = \alpha Y.$$

To the equations in (1), adjoin the equation

$$(2) \quad Y'X = x_{11}y_1 + x_{21}y_2 + \ldots + x_{n1}y_n = 0.$$

Since we know that $\nu \geqq 2$, then we have at most n-1 linearly independent equations which will always have a real non-zero solution, $X_2 = (x_{12}, x_{22}, \ldots, x_{n2})$ which, by equation (2), will be orthogonal to X_1. We can proceed in this way until $S < \nu$ mutually orthogonal real vectors $X_1, X_2, \ldots, X_s$ are obtained which will satisfy (1). To the $n-\nu$ linearly independent equations in (1), adjoin the S additional equations

$$(3) \quad \begin{matrix} x_{11}y_1 & + x_{21}y_2 + \ldots + x_{n1}y_n = 0, \\ \ldots & \ldots \quad \ldots \quad \ldots \quad \ldots \\ x_{1s}y_1 & + x_{2s}y_2 + \ldots + x_{ns}y_n = 0. \end{matrix}$$

In (1) and (3), we now have at most $n-\nu + s < n$ linearly independent homogeneous linear equations that have a solution x_{s+1} which will satisfy (1) and will be orthogonal to the vectors $X_1, \ldots, X_s$. Proceeding in this way, we obtain ν

mutually orthogonal vectors of the root α. On using each of the s roots λ_i, $\sum \nu_i = n$ vectors are obtained. When normalized, these vectors will be the columns of the orthogonal matrix P such that

$$P'AP = \text{diag}(\alpha_1, \alpha_2, \ldots, \alpha_n).$$

To illustrate this method, let us take a numerical example. Our problem is to find a real orthogonal matrix P such that $P'AP = \text{diag}(\alpha_1, \alpha_2, \ldots, \lambda_n)$, where A is the real symmetric matrix

$$\begin{pmatrix} 4 & -2 & 0 \\ -2 & 3 & -2 \\ 0 & -2 & 2 \end{pmatrix}.$$

We proceed by first finding the characteristic equation of A and form it, the characteristic roots. The characteristic equation of A is

$$\lambda^3 - 9\lambda^2 + 18\lambda = 0.$$

The roots of this equation are 0,6,3.

Corresponding to the root 0, we solve the two equations

$$\begin{aligned} 4x_1 - 2x_2 &= 0 \\ -2x_2 + 2x_3 &= 0 \end{aligned}$$

and find the single invariant vector (1,2,2).

Corresponding to the root 6, we solve the two equations

$$\begin{aligned} x_1 + x_2 &= 0 \\ x_2 + 2x_3 &= 0 \end{aligned}$$

and find the single invariant vector (2,-2,1).

Corresponding to the root 3, we solve the two equations

$$\begin{aligned} x_1 - 2x_2 &= 0 \\ 2x_2 + x_3 &= 0 \end{aligned}$$

and find the single invariant vector (2,1,-2).

On normalizing the three invariant vectors obtained and using them as columns, we obtain the desired real orthogonal matrix P.

If

$$P = \begin{pmatrix} 1/3 & 2/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \\ 2/3 & 1/3 & -2/3 \end{pmatrix}, \quad P'AP = \text{diag}\,(0,6,3).$$

This can be verified by actually multiplying the matrices $P'AP$ out.

7. Applications to Analytic Geometry.[1]

The most important application of the operations performed by orthogonal matrices is that of the rotation of axes in analytic geometry. Here, too, we see the importance of the orthogonal reduction of a real symmetric

---

1. Much of this material on the applications to analytic geometry, and the order in which it is presented can be found in Solid Analytic Geometry by Adrian Albert, pp. 103-111.

matrix to diagonal form, for it is by applying this method that we perform the orthogonal reduction of a real quadratic form. We shall see that the latter is actually executed by the rotation of axes. The following section is concerned with the development of the ideas of orthogonal transformations, the rotation of axes, and the orthogonal reduction of real quadratic forms. Since the rotation of axes is dependent upon orthogonal transformations, we must first investigate this type of transformation.

If $P = (p_{ij})$ is an n by n matrix and $X = [x_1, \ldots, x_n]$, then $PX = Y$ is a vector, $[y_1, \ldots, y_n]$, and the equation $Y = PX$ is called a linear transformation of $V_n(R)$. If P is orthogonal, $Y = PX$ is an orthogonal transformation. This is the same as saying that a linear transformation P is orthogonal if it preserves the absolute value of every vector X, so that $|PX| = |X|$.

An orthogonal transformation P has, for every pair of vectors X,Y, the following properties.

(1) P preserves distance, or $|X-Y| = |PX-PY|$.

Since P is linear, the above definition proves (1).

(2) P preserves inner products, or $(X,Y) = (XP,YP)$.

$$(X+Y, X+Y) = (X,X) + 2(X,Y) + (Y,Y)$$

Solve for (X,Y) in terms of lengths such as $|X| = (X,X)^{\frac{1}{2}}$.

$$2(X,Y) = |X+Y|^2 - |X|^2 - |Y|^2$$

The orthogonal transformation P leaves invariant the lengths on the right, therefore also the inner product on the left of the equation. Conversely, a transformation P known to preserve all inner products must preserve length and is therefore orthogonal, for length is defined in terms of inner product,

(3) P preserves orthogonality, or X perpendicular to Y implies PX perpendicular to PY.

(4) P preserves magnitude of angles, or $\cos \angle(X,Y) = \cos \angle(PX,PY)$.

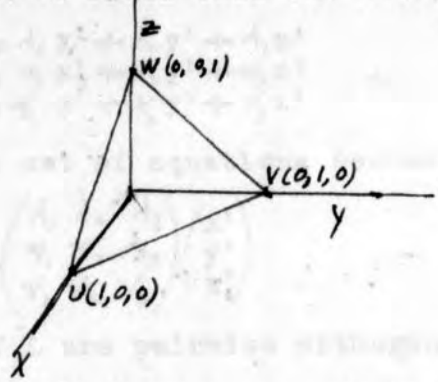
The angle between two vectors has been defined in terms of inner products (see page 8, (3)). Since X is perpendicular to Y, $(X,Y) = 0$. From (2) and the formula

$\cos \angle(X,Y) = \dfrac{(X,Y)}{|X|\cdot|Y|}$, (3) and (4) follow.

<u>Theorem</u> <u>7.1.</u> Relative to any normal orthogonal basis, a matrix A represents an orthogonal linear transformation if and only if each row of A has length one, and any two rows, regarded as vectors, are orthogonal.

Any orthogonal transformation, by theorem 3.3, carries the given basis $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ into a new normal, orthogonal basis $\alpha_1 = \epsilon_1 A, \ldots, \alpha_n = \epsilon_n A$. Conversely, if A carries a given basis $\epsilon_1, \ldots, \epsilon_n$ into a new basis $\alpha_1 = \epsilon_1 A, \ldots, \alpha_n = \epsilon_n A$, then, for any vector $X = x_1\epsilon_1 + \ldots + x_n\epsilon_n$ with a transform $XA = x_1\alpha_1 + \ldots + x_n\alpha_n$, the formula $|X| = (x_1^2 + \ldots + x_n^2)^{1/2} = |XA|$ will give the length, where A is orthogonal. The i-th row of A (regarded as a vector) represents the coordinates of $\alpha_i = \epsilon_i A$ relative to the original basis $\epsilon_1, \ldots, \epsilon_n$.

Now, to see how orthogonal transformations relate to analytic geometry, let us take points: $O = (0,0,0)$, $U = (1,0,0)$, $V = (0,1,0)$, $W = (0,0,1)$, in space. These four points will determine a tetrahedron whose vertex is the origin, and which we shall call the tetrahedron of reference.



Conversely, we can say that if we are given a tetrahedron of reference, the coordinate system is completely determined.

Let an initial coordinate system be given such that every point P has initial coordinates x,y,z, and such that P is a linear combination given by this equation

$$\text{(1)}\quad P = (x,y,z) = xU + yV + zW$$

with the coefficients the coordinates of P. With this in mind, we shall investigate all other rectangular Cartesian coordinate systems which have the same origin O(0,0,0) as the initial system. Every one of these transformed systems will be determined by three new unit vectors U', V', W' and every point will have transformed coordinates x', y', z'. In view of this we can say

$$P = x'U' + y'V' + z'W'$$

where, again, the coefficients x', y', z' are the transformed coordinates of P. If we let $U' = (\lambda_1, \mu_1, \nu_1)$, $V' = (\lambda_2, \mu_2, \nu_2)$, $W' = (\lambda_3, \mu_3, \nu_3)$ represent the initial coordinates of U', V', W', then

$$P = x'(\lambda_1, \mu_1, \nu_1) + y'(\lambda_2, \mu_2, \nu_2) + z'(\lambda_3, \mu_3, \nu_3) = (x,y,z).$$

All the relations between the initial and transformed coordinates of all points are given by the set of equations

$$(2) \qquad \begin{cases} x = \lambda_1 x' + \lambda_2 y' + \lambda_3 z' \\ y = \mu_1 x' + \mu_2 y' + \mu_3 z' \\ z = \nu_1 x' + \nu_2 y' + \nu_3 z' \end{cases} .$$

In matrix form, this set of equations becomes

$$(3) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = L \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 \\ \mu_1 & \mu_2 & \mu_3 \\ \nu_1 & \nu_2 & \nu_3 \end{pmatrix} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$$

where the columns of L are pairwise orthogonal unit vectors U', V', W', and L is an orthogonal matrix. This system of equations, where L is an orthogonal matrix is called an orthogonal transformation of coordinates. We have thus proved that an orthogonal transformation relates two rectangular coordinate systems with the same origin. Conversely, it can be said that two coordinate systems can be related by an orthogonal transformation in which the initial coordinates of the unit vectors on the transformed coordinate axes are given by the columns of the matrix L.

Since L is orthogonal, $L^{-1} = L'$ and (3) may be written

$$(4) \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = L' \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \lambda_1 & \mu_1 & \nu_1 \\ \lambda_2 & \mu_2 & \nu_2 \\ \lambda_3 & \mu_3 & \nu_3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} .$$

The relations between the transformed and the initial coordinates of all points are also given by the set of equations

$$(5) \qquad \begin{cases} x' = \lambda_1 x + \mu_1 y + \nu_1 z \\ y' = \lambda_2 x + \mu_2 y + \nu_2 z \\ z' = \lambda_3 x + \mu_3 y + \nu_3 z \end{cases} .$$

Many times the determination of vectors of integers which are scalar multiples of U', V', W' is desired. These scalar multipliers, which are square roots of rational numbers, will appear as common denominators in each equation of (5). In this discussion, it should be remembered that

$$U = (1,0,0) \rightarrow (\lambda_1, \lambda_2, \lambda_3)$$
$$V = (0,1,0) \rightarrow (\mu_1, \mu_2, \mu_3)$$
$$W = (0,0,1) \rightarrow (\nu_1, \nu_2, \nu_3)$$

where the set of x', y', z' coordinates is the set of coordinates of each of the points after the arrow. The x', y', z' coordinates are obtained by substituting the set of coordinates before the arrow, that is, $U = (1,0,0)$, $V = (0,1,0)$, $W = (0,0,1)$, for x, y, z in formula (5). The columns of L are the x, y, z coordinates of the unit vectors on the x', y', z' axes, and the rows of L are the x', y', z' coordinates of the unit vectors on the x, y, z axes.

As an example let us take the orthogonal matrix

$$A = 1/9 \begin{pmatrix} 4 & 4 & -7 \\ 8 & -1 & 4 \\ -1 & 8 & 4 \end{pmatrix}$$

and find the x', y', z' coordinates of the points whose x, y, z coordinates are $(-1,2,2)$.

From formula (4), we have

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 4/9 & 4/9 & 7/9 \\ 8/9 & -1/9 & 4/9 \\ -1/9 & 8/9 & 4/9 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} -10/9 \\ -2/9 \\ 25/9 \end{pmatrix}$$

and $x' = -10/9$, $y' = -2/9$, $z' = 25/9$.

Sometimes it is desired to apply two successive orthogonal transformations of coordinates. When this is done, the result is an orthogonal transformation of coordinates called their product.

(1) If $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = L \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$ represents one orthogonal transformation with matrix L,

(2) and $\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = M \begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix}$ represents the second orthogonal transformation with the matrix M,

(3) then $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = N \begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix}$, where the matrix N is obtained by substituting in formula (1), the equality of formula (2).

But we know that

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = L \left[M \begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix} \right] = (LM) \begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix} .$$

We have then proved that $LM = N$, or, that the matrix of a product of two orthogonal transformations is the product of the matrices of the transformations, and is also orthogonal.

We are now in the position to define and illustrate what is meant by the reflection and rotation of axes.

If we change the direction on a coordinate axis, we obtain an orthogonal transformation of coordinates which is defined by one of the following three matrices, depending on which coordinate axis the change in direction is made:

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} .$$

This transformation is a reflection of axes.

On the other hand, if we take a tetrahedron of reference whose vertex is O, and rotate this tetrahedron about its vertex which is held fixed in space, we perform an orthogonal transformation called a rotation of axes.

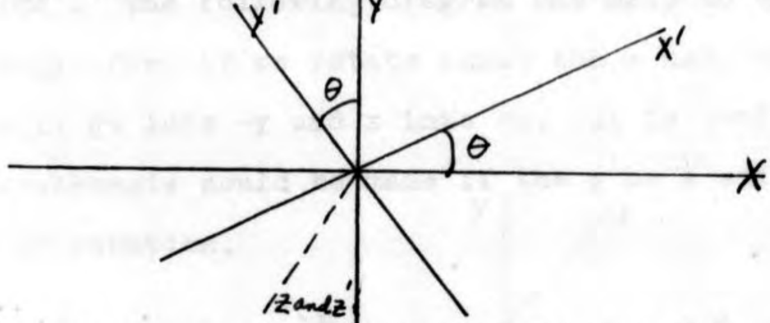
If we rotate the axes about the Z axis, we have performed a rotation of axes of plane analytic geometry, which is called a planar rotation of axes. The rotation can be represented by the equations

$$(1)\quad \begin{cases} x = x' \cos\theta - y' \sin\theta \\ y = x' \sin\theta + y' \cos\theta \\ z = z' \end{cases}$$

where the angle of rotation is measured in a counterclockwise direction form the unit point U to the unit point U'. Formula (3) gives the matrix form of this rotation where

$$L = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 \end{pmatrix}, \text{ and } |L| = 1.$$

Geometrically, the rotation is represented by the figure



If, instead of the Z axis, we had rotated the axes about the X or Y axis, the matrices of these rotations would be

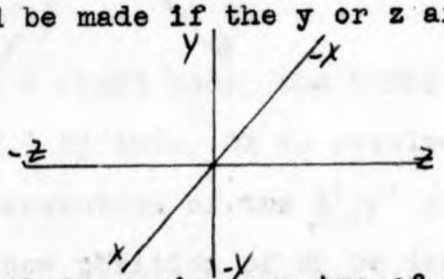
$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \text{ or } \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \text{ respectively,}$$

which are orthogonal matrices having $|L| = 1$.

It should be noted that the product of two reflections of axes is a rotation of axes, but a reflection of axes is is not a rotation of axes. For example, let us form the product of the two matrices which represent two reflections. We have

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

We see that their product is the matrix of a rotation of axes. Geometrically, it can be seen that the product of two reflections of two distinct axes is the planar rotation about the third axis through 180°. If, in the example, we replace y by -y and z by -z, the result is identical with that obtained by rotating about the x axis through 180°. The following diagram may help to visualize this analogy, for, if we rotate about the x axis through 180°, y will go into -y and z into -z. It is obvious that similar statements could be made if the y or z axis was the axis of rotation.

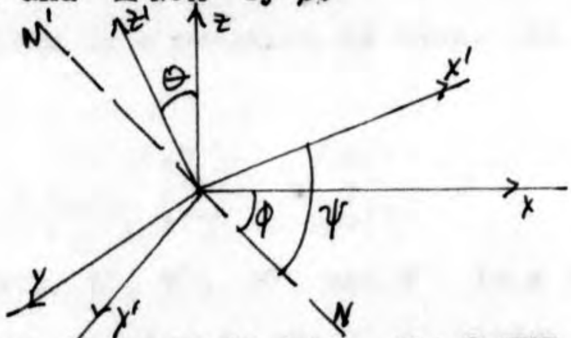


A rotation of axes that carries any two of the points U,V,W into a corresponding pair of the points U',V',W' is a rigid motion, that is, is represented by an orthogonal linear transformation of the tetrahedron, and will also carry the remaining point of U,V,W into the remaining point of U',V',W'. With this result, we shall prove the following fundamental theorem.

<u>Theorem 7.2.</u> If an orthogonal transformation is a rotation of axes, the determinant of its matrix must be 1. Every orthogonal transformation which is a rotation of axes can be expressed as a product of three planar rotations, and every orthogonal transformation which is not a

rotation of axes can be expressed as the product of a rotation and reflection of axes.

Consider an orthogonal transformation in which U,V,W and U',V',W' are the positive unit vectors on the X,Y,Z and X',Y',Z' axes respectively. Denote the line of intersection of the plane XOY and X'OY' by NN', the $\angle$ ZOZ' by θ, $\angle$ XON by ϕ, and $\angle$ NOX' by ψ.



We shall now revolve, as a rigid body, the O-XYZ axes about OZ through the $\angle \phi$. By this, OX is revolved into the position ON, the intersection of the X',Y' plane with the X,Y plane. Let the new position of OY be denoted by OY_1, so that $\angle YOY_1 = \phi$. The trihedral $\angle$ O-XYZ is brought into the position O-NY_1Z. Now, on revolving O-NY_1Z about ON through an $\angle \theta$, OZ takes a new position OZ', and OY_1, a new position OY_2. Then the $\angle ZOZ' = \angle Y_1OY_2 = \theta$. The trihedral $\angle$ O-NY_1Z is thus revolved into O-NY_2Z'. Let the trihedral angle in this last position be revolved about OZ' through an $\angle \psi$, so that ON is brought into OX'. OY is thus revolved into a direction through O perpendicular to OX' and to OZ'. OY either coincides with OY' or its direction is opposite to that of OY'. If OY coincides with OY', then the trihedral O-XYZ has been rotated into the trihedral O-X'Y'Z'. If OY is oppositely directed to OY', the rotation must be followed by a reflection, that is, the preceding planar rotation must be increased by 180°, and OY' replaced by -OY'. The angular orientation of the X'Y' plane is thus restored.

This proves that there are three planar rotations with the corresponding matrices L_1, L_2, L_3 so that the product of these three rotations is a rotation of axes. In matrix form, we have

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = L_1 L_2 L_3 \begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix} = N \begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix}$$

where $N = L_1 L_2 L_3$, $U'' = U'$, $V'' = V'$, and W'' is a point on a line through O perpendicular to the X',Y' plane. $W' = \epsilon W''$ where

$$\epsilon = \begin{vmatrix} \lambda_1 & \lambda_2 & \lambda_3 \\ \mu_1 & \mu_2 & \mu_3 \\ \nu_1 & \nu_2 & \nu_3 \end{vmatrix} = \pm 1,$$

and $\lambda_1, \lambda_2, \lambda_3;\ \mu_1, \mu_2, \mu_3;\ \nu_1, \nu_2, \nu_3$ are direction cosines of three mutually perpendicular lines.

After the rotation, if the trihedrals coincide, we have $\lambda_1 = \mu_2 = \nu_3 = 1$, the other cosines being 0, making $\epsilon = +1$. In the orthogonal transformation defined by

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = L^* \begin{pmatrix} x \\ y \\ z \end{pmatrix} = L^* N \begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix},$$

$x' = x''$, $y' = y''$, $z' = \epsilon z''$. Then we have

$$R = L^* N = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \epsilon \end{pmatrix} = N^* L = N^{-1} L$$

and

$$L = NR = L_1 L_2 L_3 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \epsilon \end{pmatrix}$$

where, as before, $N = L_1 L_2 L_3$. However, $|N| = 1$; therefore, $|L| = |R|$, and $\epsilon = |L|$. Only when $\epsilon = 1$, and therefore $R = I$, is the matrix R the matrix of a rotation of axes. Also, only when L is the matrix of rotation of axes is R one since $R = L^{*}N$, and L is one if and only if $\epsilon = |L| = +1$. If the case is $|L| = -1$, then $L = NR$ is the product of the matrix N of a rotation of axes and the matrix R of a reflection of axes. Hence, the complete theorem is proved.

In the preceding theorem let the coordinates of a point P referred to O-XYZ be (x,y,z), referred to $O\text{-}NY_1Z$ be (x_1, y_1, z_1), referred to $O\text{-}NY_2Z$ be (x_2, y_2, z_2) and referred to $O\text{-}X'Y'Z'$ be (x', y', z'). Then, in the rotation through $\angle \phi$, z remains fixed, and from plane analytic geometry,

$$\begin{aligned} x &= x_1 \cos \phi - y_1 \sin \phi \\ y &= x_1 \sin \phi + y_1 \cos \phi \\ z &= z_1 \end{aligned} \quad .$$

In the rotation through the $\angle \theta$, x_1 remains fixed and we have

$$\begin{aligned} x_1 &= x_2 \\ y_1 &= y_2 \cos \theta - z_2 \sin \theta \\ z_1 &= y_2 \sin \theta + z_2 \cos \theta \ . \end{aligned}$$

If we can obtain O-X'Y'Z' from O-XYZ by rotation, z_2 will remain fixed, and we have

$$\begin{aligned} x_2 &= x' \cos \psi - y' \sin \psi \\ y_2 &= x' \sin \psi + y' \cos \psi \\ z_2 &= z' \end{aligned} \quad .$$

Eliminating x_2, y_2, z_2; x_1, y_1, z_1, we have

$$x = x'(\sin \phi \cos \psi - \sin \phi \sin \psi \cos \theta) - y'(\cos \phi \sin \psi - \sin \phi \cos \psi \cos \theta) - z' \sin \phi \sin \theta.$$

$$y = x'(\sin \phi \cos \psi + \cos \phi \sin \psi \cos \theta) - y'(\sin \phi \sin \psi - \cos \phi \cos \psi \cos \theta) - z' \cos \phi \sin \theta.$$

$z = x' \sin \phi \sin \theta + y' \cos \phi \sin \theta + z' \cos \theta.$

If we cannot obtain O-X'Y'Z' from O-XYZ by rotation, the sign of y' must be changed. These last three formulas are what are known as Euler's formulas.

The rotation of axes by an orthogonal transformation is a fundamental idea in the orthogonal reduction of a real quadratic form. For, if we are given a real quadratic form in x,y,z as the polynomial

$$(1) \quad f(x,y,z) = ax^2 + by^2 + cz^2 + 2dxy + 2exz + 2gyz,$$

with a,b,c,d,e,g real numbers, we know from analytic geometry that the axes of the conic represented by equation (1) is at an angle with the rectangular axes. It is only when we eliminate the cross product terms that we have a conic which is not at an angle with its axes of reference. This elimination of the cross product terms leads to a rotation of axes.

Considering the quadric (1), we have

$$f(x,y,z) = PAP^*$$

where $P = (x,y,z)$ and A is the real symmetric matrix

$$\begin{pmatrix} a & d & e \\ d & b & g \\ e & g & c \end{pmatrix}.$$

By the previous theorem, we have shown that the rotation of axes is a linear transformation from which we have

$$P^* = LQ^*$$

where L is a real orthogonal matrix whose determinant is 1 and $Q = (x',y',z')$. Taking the conjugate transpose of both sides, we have

$$P = QL^*.$$

By a rotation of axes, we have replaced $f(x,y,z)$ by

$\phi(x',y',z') = a'x'^2 + b'y'^2 + c'z'^2 + 2d'x'y' + 2e'x'z' + 2f'y'z'$

where $g(x',y',z') = f(x,y,z) = QL^*ALQ^* = QBQ^*$.

Therefore,

$$B = \begin{pmatrix} a' & d' & e' \\ d' & b' & g' \\ e' & g' & c' \end{pmatrix} = L^*AL .$$

Now if we let α, β, γ be the characteristic roots of the real symmetric matrix A, by our theorem on the orthogonal reduction of a real symmetric matrix to diagonal form, there will be an orthogonal matrix L such that

$$L_1^*AL_1 = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix} .$$

If $|L_1| = 1$, let $L = L_1$ and the orthogonal transformation with matrix L will be a rotation of axes. If $|L_1| = -1$, change the sign of one column of L_1 and replace L_1 by an orthogonal matrix L so that $|L| = -|L_1| = 1$. Then $L^*AL = L_1^*AL_1$. From this we see that the rotation of axes with matrix L replaces $f(x,y,z)$ by

$$\phi(x',y',z') = Q(L^*AL)Q^* = (x',y',z')\begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix}\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \alpha x'^2 + \beta y'^2 + \gamma z'^2 .$$

We have then, by the rotation of axes, reduced the given quadratic form to one involving only square terms.

Let us close this section and this paper by two illustrative problems: one, the orthogonal reduction of a given real quadratic form; two, a problem relating this orthogonal reduction of a real quadratic form to the various conics in plane analytic geometry.

For problem one, we are to reduce the quadratic form $4x + 3y - z - 12xy + 4xz - 8yz$ to diagonal form by a rotation of axes, and to give the equation of this rotation. The matrix of the given form is

$$A = \begin{pmatrix} 4 & -6 & 2 \\ -6 & 3 & -4 \\ 2 & -4 & -1 \end{pmatrix}.$$

From the characteristic determinant of the characteristic function of A, we have, x being a characteristic root of A

$$|A-xI| = \begin{vmatrix} 4-x & -6 & 2 \\ -6 & 3-x & -4 \\ 2 & -4 & -1-x \end{vmatrix} = f(x),$$

$f(x) = (4-x)(3-x)(-1-x) + 96-12 + 4x-64 + 16x + 36 + 36x,$

$f(x) = (x + 1)(x + 4)(x-11).$

Then, the diagonal is $11x'^2 - y'^2 - 4z'^2$.

We now find the characteristic vectors of A.

Corresponding to the root 11, we solve the three equations

$$\begin{cases} -7x-6y + 2z = 0 \\ -6x-8y - 4z = 0 \\ 2x-4y - 12z = 0 \end{cases}$$

and find the invariant vector (2,-2,1).

Corresponding to the root -1, we solve the three equations

$$\begin{cases} 5x - 6y + 2z = 0 \\ -6x + 4y - 4z = 0 \\ 2x - 4y = 0 \end{cases}$$

and find the invariant vector (2,1,-2).

Corresponding to the root -4, we solve the three equations

$$\begin{cases} 8x - 6y + 2z = 0 \\ -6x + 7y - 4z = 0 \\ 2x - 4y + 3z = 0 \end{cases}$$

and find the invariant vector (1,2,2).

On normalizing the three invariant vectors obtained and using them as columns, we obtain the real orthogonal matrix L, which is the matrix of a rotation of axes.

$$L = 1/3\begin{pmatrix} 2 & 2 & 1 \\ -2 & 1 & 2 \\ 1 & -2 & 2 \end{pmatrix}$$

The equations of rotation given by the form

$$(x',y',z') = (x,y,z)L,$$

are

$$\begin{cases} 3x' = 2x - 2y + z \\ 3y' = 2x + y - 2z \\ 3z' = 2x + 2y + 2z \end{cases} .$$

In relating the orthogonal reduction of a real quadratic form to plane analytic geometry, consider the locus in real two-dimensional space of the equation $X'AX = 1$, where $X'AX$ is a quadratic form over the field R and $X' = (x,y)$. Let α and β denote the characteristic roots of the real, symmetric matrix A. Our problem is to show that the locus is an ellipse if α and β are positive; a hyperbola if α and β have opposite signs; nonexistent if α and β are negative; a circle if α and β are positive and equal; that there is no locus if α and β are zero or if one root is zero and the other is negative; and, that the locus is two straight lines if one root is zero and the other is positive.

We have

$$X'AX = (x,y)\begin{pmatrix} a & b \\ b & c \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 1 = ax^2 + 2bxy + cy^2 .$$

We can reduce the real symmetric matrix A to the diagonal (α, β). Then

$$P'AP = P'\begin{vmatrix} a & b \\ b & c \end{vmatrix} P = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$$

where P is a real orthogonal matrix.

Then $(x',y')\begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}\begin{pmatrix} x' \\ y' \end{pmatrix} = 1.$

The rotation of axes with the matrix P replaces $f(x,y)$ by $\phi(x',y')$ so that

$$(1) \quad \alpha x'^2 + \beta y'^2 = 1.$$

From this equation, we can see the following conditions verified:

1. If α equal β and both have positive signs, (1) is the locus of a circle.
2. If α does not equal β and both have positive signs, (1) is the locus of an ellipse.
3. If α and β have opposite signs, (1) is the locus of a hyperbola.
4. If α and β have negative signs, (1) is nonexistent.
5. If α and β are zero, (1) does not represent a locus.
6. If α is negative and β is zero, (1) does not represent a locus.
7. If α is positive and β is zero, (1) is the locus of two straight lines.

BIBLIOGRAPHY

1. Aitken, A.C., Determinants and Matrices. New York: Interscience Publishers, 1948.

2. Albert, A.A., Solid Analytic Geometry. New York: McGraw-Hill Book Company, 1949.

3. Birkhoff, G., and MacLane, S., A Survey of Modern Algebra. New York:Macmillan, 1941.

4. Bocher, M., Introduction to Higher Algebra. New York:Macmillan, 1931.

5. Browne, E.T., "Lecture Notes on the Theory of Matrices." North Carolina:University of North Carolina.

6. Dresden, A., Solid Analytic Geometry and Determinants. New York:Wiley, 1930.

7. Ferrar, W.L., Finite Matrices. Oxford:Oxford University Press, 1951.

8. MacDuffee, C.C., An Introduction to Abstract Algebra. New York:Wiley, 1940.

9. Perlis, S., Theory of Matrices. Cambridge:Wesley Press, 1952.

10. Sisam,C., and Snyder, V., Analytic Geometry of Space. New York:Henry Holt,1938.

11. Stoll,R., Linear Algebra and Matrix Theory. New York:McGraw-Hill Book Company, 1952.

12. Weiss, M., Higher Algebra for the Undergraduate. New York:Wiley, 1949.