Let the system of N dimensional unit vectors be $e_1, e_2, \ldots, e_N$. Now consider the N vectors $y^1, e_1, e_2, \ldots, e_{N-1}$. Assume that this system of vectors is linearly independent (if this is not the case, then this will show up in the Gram-Schmidt procedure -- one of the z^k's will be 0_N - then just drop the corresponding e_{k-1} and include e_N in the set of linearly independent vectors). Use the Gram-Schmidt orthogonalization procedure to construct a system of N orthonormal vectors with the eigenvector y^1 being the first of these vectors. Denote the N orthonormal vectors as $[y^1, y^2, \ldots, y^N] \equiv U_1$, a N by N matrix with $U_1^T U_1 = I_N$. Using $Ay^1 = \lambda_1 y^1$, compute.

$$(178) \quad U_1^T A U_1 = \begin{bmatrix} y^{1T} \\ y^{2T} \\ \vdots \\ y^{NT} \end{bmatrix} A[y^1, y^2, \ldots, y^N]$$

$$= \begin{bmatrix} y^{1T} \\ y^{2T} \\ \vdots \\ y^{NT} \end{bmatrix} A[\lambda_1 y^1, Ay^2, \ldots, Ay^N] \text{ using } Ay^1 = \lambda_1 y^1$$

$$= \begin{bmatrix} \lambda_1, & y^{1T}Ay^2, & \ldots, & y^{1T}Ay^N \\ 0, & y^{2T}Ay^2, & \ldots, & y^{2T}Ay^N \\ \vdots & & & \\ 0, & y^{NT}Ay^2, & \ldots, & y^{NT}Ay^N \end{bmatrix} \text{since } y^{jT}y^1 = 0 \text{ for } j = 2,3, \ldots, N$$

$$= \begin{bmatrix} \lambda_1, & 0, & \ldots, & 0 \\ 0, & y^{2T}Ay^2, & \ldots, & y^{2T}Ay^N \\ \vdots & \vdots & & \\ 0, & y^{NT}Ay^2, & \ldots, & y^{NT}Ay^N \end{bmatrix}$$

since $(U_1^T AU_1)^T = U_1^T A^T U_1 = U_1^T AU_1$ is symmetric

$$= \begin{bmatrix} \lambda_1, & 0, & \ldots, & 0 \\ 0 & & & \\ \vdots & & A_2 & \\ 0 & & & \end{bmatrix}$$

where A_2 is an N-1 by N-1 symmetric matrix.

Now let λ_2 be an eigenvalue of A_2 (note that $A_2^T = A_2$) and let x^2 be the corresponding N-1 dimensional eigenvector, i.e., $A_2x^2 = \lambda_2x^2$ and $x^{2T}x^2 = 1$. Again use the Gram-Schdmidt orthogonalization procedure to construct a system of N-1 orthonormal N-1 dimensional vectors with x^2 being the first such vector; let the system of orthonormal vectors be denoted by $[x^2,x^3, \ldots, x^N]$. Then we may repeat the analysis given on page XXX to show that:

$$[x^2,x^3, \ldots, x^N]^T A_2 [x^2,x^3, \ldots, x^N]$$

(179) $$= \begin{bmatrix} \lambda_2 & 0 \ldots 0 \\ 0 & \\ \vdots & A_3 \\ 0 & \end{bmatrix}$$ where A_3 is an (N-2) x (N-2) matrix

Now define the N dimensional orthonormal matrix U_2 by

(180) $$U_2 = \begin{bmatrix} 1 & 0, \ldots, 0 \\ 0 & \\ \vdots & x^2, \ldots, x^N \\ 0 & \end{bmatrix}$$

Combining (178) with (179) yields the following:

$$(U_1U_2)^T A (U_1U_2) = U_2^T (U_1^T A U_1) U_2$$

$$= U_2^T \begin{bmatrix} \lambda_1 & 0 \ldots 0 \\ 0 & \\ \vdots & A_2 \\ 0 & \end{bmatrix} U_2$$

$$= \begin{bmatrix} \lambda_1 & 0 & 0 \ldots 0 \\ 0 & \lambda_2 & 0 \ldots 0 \\ \vdots & \vdots & \\ 0 & 0 & A_3 \end{bmatrix}$$

Evidently we can continue this process until we have reduced A down to a diagonal matrix by means of a product of N orthogonal matrices.

(181) $U = U_1U_2 \ldots U_N$; i.e., $$U^TAU = \begin{bmatrix} \lambda_1 & & 0 \ldots 0 \\ 0 & \lambda_2 & 0 \\ \vdots & & \vdots \\ 0 & 0 & \lambda_N \end{bmatrix}$$ where $U^TU = I_N$.

Q.E.D.

(182) **Definition:** A *projection* matrix M is a square matrix with
(i) $M \neq M^T$ (i.e., M is a symmetric); and
(ii) $M^2 = M \bullet M = M$.
(Sometimes projection matrices are called *idempotent* matrices).

(183) **Lemma:** The eigenvalues of a projection matrix are either equal to zero or unity.

Proof: Since M is a square, symmetric matrix, by (181) there exists an orthonormal matrix U such that

(184) $$U^TMU = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \\ 0 & & \lambda_N \end{bmatrix}$$

a diagonal matrix with the eigenvalues of M down the main diagonal.

Therefore,

$$\begin{bmatrix} \lambda_1 & 0 \ldots 0 \\ 0 & \lambda_2 \ldots 0 \\ \vdots & \\ 0 & 0 \ldots \lambda_N \end{bmatrix} = U^TMU = U^TM \bullet MU \qquad \text{by (182) (ii)}$$

$$= U^TMUU^TMU$$

$$= \begin{bmatrix} \lambda_1 \ldots 0 \\ 0 \ldots \lambda_N \end{bmatrix} \begin{bmatrix} \lambda_1 \ldots 0 \\ 0 \ldots \lambda_N \end{bmatrix} \text{ [using (184) twice]}$$

since U^T is an inverse for U

$$= \begin{bmatrix} \lambda_1^2 & 0 \ldots 0 \\ 0 & \lambda_2^2 \ldots 0 \\ \vdots & \\ 0 & 0 \ldots \lambda_N^2 \end{bmatrix}$$

Therefore, $\lambda_i = \lambda_i^2$ for $i = 1, 2, \ldots, N$

$\Rightarrow \lambda_i^2 - \lambda_i = 0$

$\Rightarrow \lambda_i =$ either 0 or 1.

Q.E.D.

Problem 19:

Let X be an N by K matrix where $K \leq N$ and the columns of X are linearly independent so that $(X^TX)^{-1}$ exists.

(i) Show that $M_1 \equiv X(X^TX)^{-1} X^T$ and

$$M_2 \equiv I_N - X(X^TX)^{-1} X^T$$

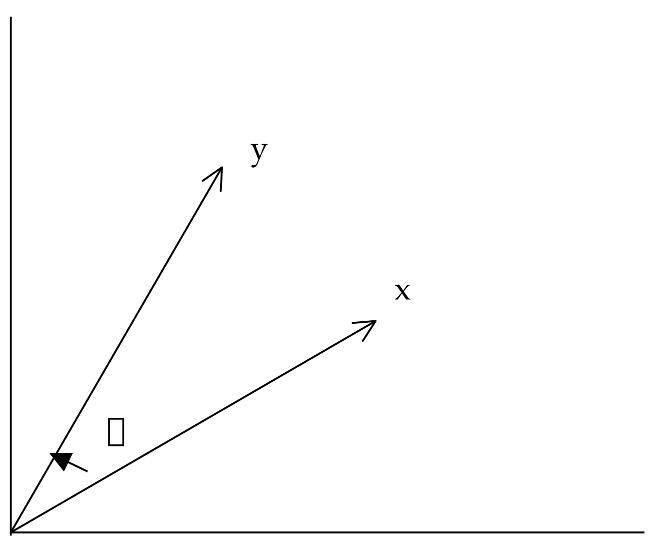
are projection matrices. (These matrices occur in the study of the linear model in econometrics).

(ii) Let U be the orthonormal matrix which diagonalizes M_1; i.e., $U^TM_1U = \Lambda_1$ a diagonal matrix. Show that the same U also diagonalizes M_2 into another diagonal matrix Λ_2 and that $\Lambda_1 + \Lambda_2 = I_N$.

Geometrically speaking, an orthonormal transformation can be interpreted as a rotation of the system of co-ordinate axes. This geometric interpretation rests on the fact that an orthonormal transformation leaves the distance between two points x and y unchanged and also the angle between the two vectors is left unchanged in the new co-ordinate system.

(185) **Definition:** The distance between two N dimensional vectors x and y is defined as $D(x, y) \equiv [(x-y)^T (x-y)]^{1/2}$

(186) **Definition:** The *angle* which two vectors x and y make with each other is obtained by $\cos \theta \equiv x^Ty / (x^Tx)^{1/2}(y^Ty)^{1/2}$



(187) **Lemma:** An orthonormal transformation U leaves distances and angles between two points unchanged.

Proof: $D(Ux; Uy) \equiv [(Ux - Uy)^T(Ux - Uy)]^{1/2}$
$= [(x-y)^T U^T U(x-y)]^{1/2}$
$= [(x-y)^T I_N (x-y)]^{1/2}$
$= D(x;y)$

The angle between the points Ux and Uy is given by:

$$\cos \theta \equiv (Ux)^T(Uy) / (x^T U^T Ux)^{1/2}$$
$$= x^T U^T Uy / (x^T x)^{1/2} (y^T y)^{1/2}$$
$$= x^T y / (x^T x)^{1/2} (y^T)^{1/2}$$

which defines the angle between x and y.

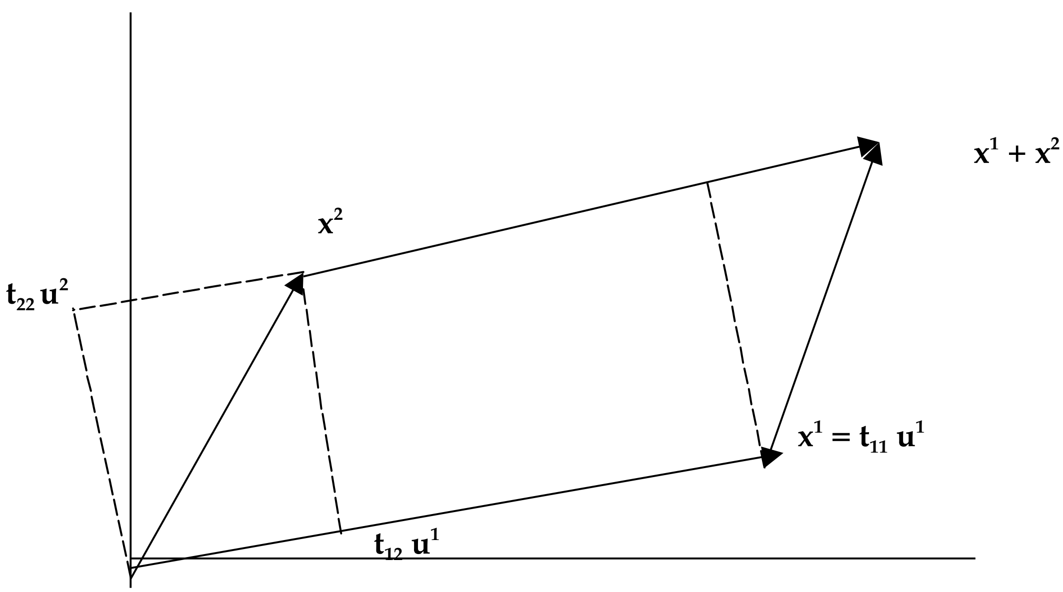
Q.E.D.

Finally, let us return to the problem of giving a geometric interpretation to the determinant of a square matrix A. Let us write A and N column vectors:

$A = [x^1, x^2, \ldots, x^N]$

We wish to show that $|A|$ = ± volume of parallelepiped generated by vectors $x^1, \ldots, x^N$.

Consider the case N = 2.



Recall the Gram-Schmidt Orthogonalization procedure (164).

$x^1 = t_{11}u^1$ where u^1 is a vector of unit length
$x^2 = t_{12}u^1 + t_{22}u^2$ u^2 is a vector of unit length perpendicular to u^1

Since the area of the parallelogram is equal to base times height, we have area = $t_{11} \bullet t_{22}$.

For the case of general N we have a similar result:

(188) $$A \equiv [x^1, x^2, \ldots, x^N] = [u^1, u^2, \ldots, u^N] \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1N} \\ 0 & t_{22} & \cdots & t_{2N} \\ \vdots & \vdots & & \\ 0 & 0 & & t_{NN} \end{bmatrix}$$

↑ orthogonal vectors   ↑ all zeros below main diagonal.

- the length of the vector x^1 is t_{11},
- the area of the parallelogram generated by x^1 and x^2 in the x_1, x_2 plane is $t_{11} \bullet t_{22}$,
- the area of the parallelepiped generated by x^1, x^2, and x^3 in x_1, x_2, x_3 space is $t_{11} \bullet t_{22} \bullet t_{33}$ (or the absolute value of this number if it turns out to be negative). Thus

$$\pm \text{ volume of parallelipied} = \pm\ t_{11}t_{22}\ldots t_{NN} = \begin{vmatrix} t_{11} & t_{12} & \cdots & t_{1N} \\ 0 & t_{22} & \cdots & t_{2N} \\ & & & \vdots \\ 0 & 0 & \cdots & t_{NN} \end{vmatrix}$$

$= \pm\ |T|\ |U|$ since $|U| = \pm$ by lemma (159)

$= \pm\ |UT|$ since $|T| \bullet |U| = |U| \bullet |T| = |UT|$

$= \pm\ |A|$ by (188).

Q.E.D.

12. Additional Useful Properties of Square Invertible Matrices

The following three results are very useful in applications.

(189) **Lemma:** If $A = A^T$ and A^{-1} exists, then $A^{-1} = (A^{-1})^T$; i.e., if A is symmetric and A^{-1} exists, then A^{-1} is also symmetric.

Proof: We have, using the associative law for matrix multiplication:

(190) $A^{-1} A(A^{-1})^T = [A^{-1}A](A^{-1})^T = I_N (A^{-1})^T = (A^{-1})^T$ and

$A^{-1}A(A^{-1})^T = A^{-1}[A(A^{-1})^T]$

$= A^{-1}[A^T (A^{-1})^T]$ using $A = A^T$

$= A^{-1}[A^{-1}A]^T$ using $C^TB^T = (BC)^T$

(191) $= A^{-1}I_N^T = A^{-1} I_N = A^{-1}$.

Equating (190) and (191) yields the desired result.

Q.E.D.

(192) **Lemma:** If A is positive definite, then so is A^{-1}.

Proof: Since A is positive definite, we have $|A| > 0$ and hence A^{-1} exists. Let

(193) $y \neq 0_N$ and define x by
(194) $x \equiv A^{-1} y$.

Suppose the x defined by (194) were 0_N. Then

$$A^{-1}y = 0_N \quad \text{or}$$
$$y = A\, 0_N = 0_N$$

which contradicts (193). Thus our *supposition* is false and we must have $x \neq 0_N$. Hence, since A is positive definite, we have:

$$\begin{aligned} 0 &< x^TAx \\ &= (A^{-1}y)^TA(A^{-1}y) && \text{using (194), } x = A^{-1}y \\ &= y^T(A^{-1})^TA\,A^{-1}y \\ &= y^T(A^{-1})^Ty \\ &= y^T A^{-1} y && \text{using Lemma (189).} \quad (195) \end{aligned}$$

(193) and (195) show that A^{-1} is positive definite.

Q.E.D.

(196) **Corollary:** If A is negative definite, then so is A^{-1}.

Proof: Adapt the above proof.

(197) **Lemma:** Suppose A is N by N and A^{-1} exists. Then $(A^{-1})^T = (A^T)^{-1}$; i.e., we can interchange the order of transposition and inversion and obtain the same result.

Proof: Again use the associative law for matrix multiplication:

$$(198) \quad (A^{-1})^T A^T (A^T)^{-1} = (A^{-1})^T [A^T (A^T)^{-1}] = (A^{-1})^T I_N = (A^{-1})^T \quad \text{and}$$
$$\begin{aligned} (A^{-1})^T A^T (A^T)^{-1} &= [(A^{-1})^T A^T] (A^T)^{-1} \\ &= [A\,A^{-1}]^T (AT)^{-1} && \text{using} \quad C^TB^T = (BC)^T \\ &= I_N^T (A^T)^{-1} \\ &= (A^T)^{-1}. \quad (199) \end{aligned}$$

Equating (198) and (199) yields the desired result. Q.E.D.