A NEW THEORY OF DETERMINANTS EMPHASIZING GEOMETRIC INTERPRETATION AND COMPUTATION[1]

SAMUEL G. LINDLE[2] and DAVID M. ALLEN[2]

BU-534-M[3] October 1974

Abstract. A geometric definition of the determinant of a square matrix A is given. Using this definition we develop an intuitive and simplistic theory of determinants. The development suggests accurate and efficient computational procedures.

1. Introduction. The usual definition of the determinant can be stated as follows:

Definition. If $A = \begin{bmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & & \vdots \\ a_{p1} & \cdots & a_{pp} \end{bmatrix}$ is a matrix of real components, then the determinant of A, denoted $|A|$, is the number

$$|A| = \sum_{\sigma \in S_p} \text{sign}(\sigma) a_{\sigma(1),1}\, a_{\sigma(2),2} \cdots a_{\sigma(p),p}$$

where S_p is the set of all permutations of the integers $1,\cdots,p$ and sign (σ) is 1 or -1 depending on whether the permutation is even or odd respectively. (See, for example, [1] and [4]).

[1] Research was partially supported by a grant from Mead Johnson and Company, Evansville, Ind.

[2] Department of Statistics, University of Kentucky, Lexington, Ky. 40506. This paper was submitted for publication while the authors were on leave at Cornell University.

[3] The Biometrics Unit Mimeo Series, Cornell University, Ithaca, N.Y.

In practice one rarely evaluates determinants in this way because of the magnitude of the task when p gets very large [p! sums of products of p terms]. Further, the definition is void of any clear physical or geometric interpretation. We will give a more appealing definition of the determinant and develop the usual theory of determinants from it.

2. <u>Review of Elementary Reflectors</u>. Since our definition of the determinant will involve elementary reflectors (Householder transformations), we now turn our attention briefly to these. In this paper a $p \times p$ matrix R will be called an elementary reflector (ER) if R can be written as

$$R = I - 2x(x'x)^{-1}x'$$

where x is a $p \times 1$ vector of real components. (See [3], [4].) Unless otherwise indicated we will use either R or R_i throughout to symbolize elementary reflectors. It is easy to verify that R is symmetric and orthonormal. The projection matrix

$$P = x(x'x)^{-1}x$$

can be used to see the motivation in the term "elementary reflector", for

$$R = I - 2x(x'x)^{-1}x = I - 2P = (I-P) - P \quad .$$

Thus, referring to figure 1, if y is any real valued vector then it is well known that the projection of y onto x is Py and the projection of y onto the space orthogonal to x is (I-P)y. Since R is orthonormal and equals the difference (I-P) - P, it acts as an angle preserving, length preserving <u>reflection</u> of y about (I-P)y as is indicated in figure 1. The reflection is <u>elementary</u> in the sense that it occurs only in a single plane. In this case the plane is the one defined by the vectors x and y . Elementary reflectors are numerically stable operators.

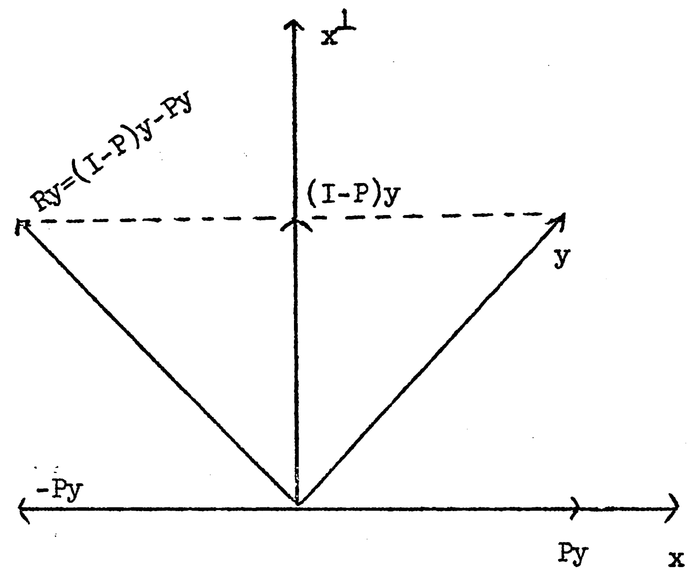


Figure 1.
Pictorial Illustration of an Elementary Reflection

Some ERs are of particular interest. First, suppose

$$x_i = 1$$

$$x_j = -1 \qquad j \neq i$$

$$x_k = 0 \qquad k \neq i,j \ ; \qquad k=1,\cdots,n$$

then Ry permutes the i^{th} and j^{th} elements of y . We denote an ER of this type by P_{ij} . (P_{ij} is often called an elementary permutation matrix.)

If $x_i = -1$ and $x_k = 0$ for all $k \neq i$ then Ry changes the sign of the i^{th} element.

If $x_i = 0 \quad i = 1,\cdots,k-1$

$$x_k = y_k \pm \sqrt{\sum_{i=k}^{p} y_\ell^2}$$

$$x_i = y_i \quad i = k+1,\cdots,p$$

then Ry retains the first k elements, the k^{th} element is $\sqrt{\sum_{\ell=k}^{p} y_\ell^2}$ and all other elements are zero. A sequence of ERs of this type can be used to triangularize any square matrix. Details are given in [3]. See also [5] for more on ERs.

3. New Definition of the Determinant. We know that there exists a sequence of ERs $\{R_i\}$ such that $R_k R_{k-1} \cdots R_1 A = T$ where T is upper triangular and $t_{ii} \geq 0$. Any product of elementary reflectors which has this property will be denoted henceforth by R. The product of an odd number of ERs cannot equal the identity matrix. We are now ready to define the determinant as follows.

Definition. Let A be a pxp matrix of real components, the determinant of A, denoted $|A|$, is the scalar quantity

$$|A| = (-1)^k v_A$$

where v_A is the volume of the parallelotope formed by the columns of A and k is the number of premultiplications by ERs that will transform A into an upper triangular matrix with nonnegative diagonal elements.

Since R is an orthonormal transformation it preserves lengths and angles, thus $v_A = v_T$. If A is singular then one of the t_{ii} is 0 so the $v_A = v_T = 0$.

Consider then the triangular matrix

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} & \cdots & t_{1p} \\ & t_{22} & t_{23} & \cdots & t_{2p} \\ & & t_{33} & \cdots & t_{3p} \\ & 0 & & \ddots & \vdots \\ & & & & t_{pp} \end{bmatrix}$$

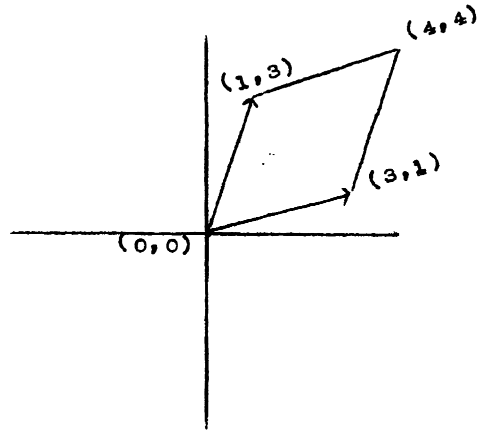
where $t_{ii} \geq 0$. Now if we take the first two columns as vectors they form a 2-dimensional parallelogram in p space. If we take the first vector to correspond to the base then the altitude of the parallelogram is t_{22} . So the area of the parallelogram is the product $t_{11}t_{22}$. Now adding the third column we form a 3-dimensional parallelotope in p space whose base can be considered to be the parallelogram just formed by the first two columns. But the altitude of the 3-dimensional parallelotope is t_{33} so its 3-dimensional volume is $t_{11}t_{22}t_{33}$. Continuing on in this way we find that the p-dimensional volume in p space of the p-dimensional parallelotope formed by all the columns of T equals

$$\prod_{i=1}^{p} t_{ii} \quad .$$

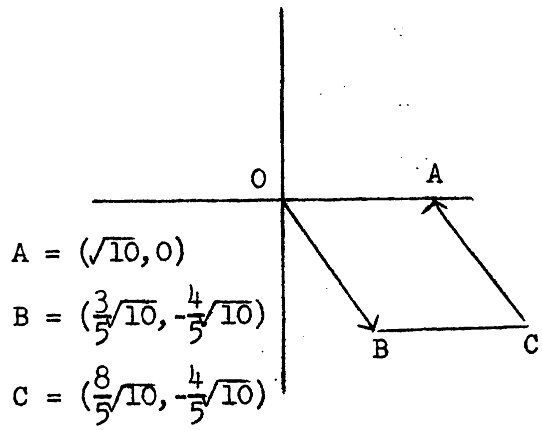
Thus in the definition above $v_A = v_T = \prod_{i=1}^{p} t_{ii}$.

The recommended method of computation is suggested by the definition. That is, premultiplying by ERs transforms A to an upper triangular matrix T with non-negative diagonal elements. Then count the number of such transformations used, say k . Finally, calculate

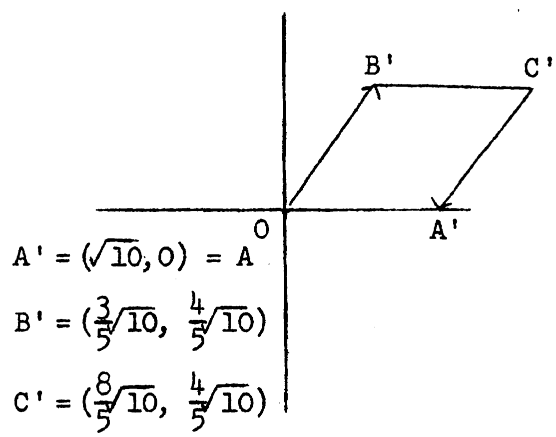
$$|A| = (-1)^k \prod_{i=1}^{p} t_{ii} \quad .$$



Original Parallelogram



After One Premultiplication by Appropriate Elementary Reflector



Final Reflection About X Axis to Put in Positive Half Plane

Figure 2.

Two-Dimensional Representation of the Effect of Premultiplication of a Matrix Successively by Elementary Reflector so as to Reduce it to an Upper Triangular Matrix.

4. Properties of Determinants. In this section, well known properties of determinants are proved using the new definition of the determinant. Let $M_i = M_i(\lambda)$ be the elementary matrix formed when the identity matrix has its i^{th} row or column multiplied by λ . That is,

$$M_i = \begin{array}{cc} & \begin{array}{ccccc} & & i & & \end{array} \\ i & \left[\begin{array}{ccccc} 1 & & \vdots & & \\ & \ddots & \vdots & \bigcirc & \\ & & 1 \vdots & & \\ \cdots & \cdots & \lambda & & \\ & & & 1 & \\ & \bigcirc & & & \ddots \\ & & & & 1 \end{array}\right] \end{array} .$$

Lemma 1. Let A be a pxp matrix of real components, if the i^{th} column of A is multiplied by a scalar λ then the determinant of the resulting matrix is $\lambda|A|$ i.e.,

$$|A\ M_i(\lambda)| = \lambda|A| .$$

Pf: First suppose $\lambda > 0$, then

$$R\ A\ M_i(\lambda) = T\ M_i(\lambda)$$

is ∇ with nonnegative diagonal elements $t_{11} \cdots \lambda t_{ii} \cdots t_{pp}$. Thus

$$|A\ M_i(\lambda)| = \lambda(-1)^k \prod_{j=1}^{p} t_{jj} = \lambda|A| .$$

For $\lambda < 0$,

$$T\ M_i(\lambda)$$

has a negative diagonal element. However,

$$M_i(-1) T\ M_i(\lambda)$$

has only the positive diagonal elements, $t_{11}, \cdots, -\lambda t_{ii}, \cdots, t_{pp}$. Thus,

$R^{*}A = M_i(-1)RA = M_i(-1)T = T^{*}$ is upper ∇ with all diagonal elements positive and

$$|A\ M_i(\lambda)| = (-\lambda)(-1)^{k+1} \prod_{j=1}^{p} t_{jj'} = \lambda(-1)^{k} \prod_{j=1}^{p} t_{jj} = \lambda |A| \quad .$$

If the i^{th} and j^{th} columns of A are interchanged, the volume of the parallelotope formed by the columns of the resulting matrix is the same as the volume of the parallelotope formed by the columns of A since both parallelotopes are formed by the same column vectors, i.e.

$$v_{AP_{ij}} = v_A \quad .$$

This implies that

$$|A\ P_{ij}| = \pm |A| \quad .$$

Later it will be shown that indeed

$$|A\ P_{ij}| = -|A| \quad .$$

Let $E_{ij} = E_{ij}(\lambda)$ be the triangular matrix formed when the identity matrix has λ times its j^{th} row added to its i^{th} row. That is

$$E_{ij} = \begin{array}{cc} & \begin{array}{ccccccc} \ & \ & \ & \ & j & \ & \ \end{array} \\ i & \begin{bmatrix} 1 & & & & \cdot & & & \\ & 1 & & & \cdot & \bigcirc & & \\ & & \cdot & & \cdot & & & \\ \cdot & \cdot & \cdot & \cdot & \lambda & & & \\ & & & & \cdot & & & \\ & & & & \cdot & & & \\ & \bigcirc & & & & \cdot & & \\ & & & & & & 1 & \\ & & & & & & & 1 \end{bmatrix} \end{array}$$

for $i \neq j$ and $\lambda \neq 0$.

Lemma 2. *If λ times the i^{th} column of A is added to the j^{th} column, the determinant of the resultant matrix is equal to the determinant of A or*

$$|A\ E_{ij}(\lambda)| = |A| \quad .$$

Pf. First suppose, $j > i$. Then

$$R\ A\ E_{ij}(\lambda) = T\ E_{ij}(\lambda)$$

is ◹ with positive diagonal elements $t_{11} \cdots t_{pp}$. Thus

$$|A\ E_{ij}(\lambda)| = (-1)^k \prod_{\ell=1}^{p} t_{\ell\ell} = |A| \quad .$$

For $j < i$, note that

$$E_{ij}(\lambda) = P_{ij} E_{ji}(\lambda) P_{ij}$$

and

$$\begin{aligned} |A| &= \pm |A\ P_{ij}| \\ &= \pm |A\ P_{ij} E_{ji}(\lambda)| \\ &= |A\ P_{ij} E_{ji}(\lambda) P_{ij}| \\ &= |A\ E_{ij}(\lambda)| \quad . \end{aligned}$$

Corollary to Lemmas 1 and 2. $|M_i(\lambda)| = \lambda$, $|E_{ij}(\lambda)| = 1$.

Pf. Substitute $A = I$ in Lemmas 1 and 2.

<u>Theorem 1</u>. $|AB| = |A|\,|B|$.

<u>Pf</u>. If either A or B is singular AB is singular so $|AB| = |A|\,|B| = 0$.

If B is nonsingular, B can be reduced to a product of matrices of the form $M_{i_\ell}(\lambda_\ell)$ and $E_{i_\ell j_\ell}(\lambda_\ell)$ so let

$$B = \prod_{\ell=1}^{s} M_{i_\ell}(\lambda_\ell) \prod_{m=1}^{s} E_{i_m j_m}(\lambda_m) \quad .$$

But using lemma 1,

$$|A\, M_{i_1}(\lambda_1)| = \lambda_1 |A|$$

$$|A\, M_{i_1}(\lambda_1) M_{i_2}(\lambda_2)| = \lambda_1\lambda_2 |A|$$

$$\vdots$$

$$\left|A \prod_{\ell=1}^{s} M_{i_\ell}(\lambda_\ell)\right| = \left(\prod_{\ell=1}^{s} \lambda_\ell\right)|A| \quad .$$

So using lemma 2,

$$\left|A \prod_{\ell=1}^{s} M_{i_\ell}(\lambda_\ell)\cdot E_{i_1 j_1}(\lambda_1)\right| = \left(\prod_{\ell=1}^{s} \lambda_\ell\right)|A|$$

$$\vdots$$

$$|AB| = \left|A \prod_{\ell=1}^{s} M_{i_\ell}(\lambda_\ell) \prod_{m=1}^{s} E_{i_m j_m}(\lambda_m)\right| = \left(\prod_{\ell=1}^{s} \lambda_\ell\right)|A| \quad .$$

But letting A = I we get

$$|B| = \prod_{\ell=1}^{s} \lambda_\ell$$

$$\therefore |AB| = |B|\,|A| = |A|\,|B| \quad .$$

<u>Corollary to Theorem 1</u>. $|A^{-1}| = |A|^{-1}$ <u>if</u> A <u>is</u> <u>nonsingular</u>.

<u>Pf</u>. Let $B = A^{-1}$ in Theorem 1.

<u>Theorem 2</u>. $|A'| = |A|$

<u>Pf</u>. If A is singular so is A' so $|A'| = |A| = 0$.

If A is nonsingular then $A' = A\,A^{-1}A'$ and A, A', A^{-1} can be written as a product of elementary matrices as follows

$$A = \prod_{\ell=1}^{s} M_{i_\ell}(\lambda_\ell) \prod_{m=1}^{r} E_{i_m j_m}(\lambda_m)$$

$$A' = E'_{i_r j_r}(\lambda_r)\cdots E'_{i_1 j_1}(\lambda_1) M_{i_s}(\lambda_s)\cdots M_{i_1}(\lambda_1)$$

$$A^{-1} = E_{i_r j_r}(-\lambda_r)\cdots E_{i_1 j_1}(-\lambda_1) M_{i_s}(1/\lambda_s)\cdots M_{i_1}(1/\lambda_1)$$

since

$$M'_{i\ell}(\lambda_\ell) = M_{i_\ell}(\lambda_\ell), M^{-1}_{i_\ell}(\lambda_\ell) = M_{i\ell}(1/\lambda_\ell) E^{-1}_{i_\ell j_\ell}(\lambda_\ell) = E_{i_\ell j_\ell}(-\lambda_\ell) \quad .$$

So

$$|A| = |A\, E_{i_r j_r}(-\lambda_r)\cdots E_{i_1 j_1}(-\lambda_1)|$$

$$= \prod_{\ell=1}^{s} \lambda_\ell \cdot |A\, E_{i_r j_r}(-\lambda_r)\cdots E_{i_1 j_1}(-\lambda_1) M_{i_s}(1/\lambda_s)\cdots M_{i_1}(1/\lambda_1)|$$

$$= \prod_{\ell=1}^{s} \lambda_\ell |A\, A^{-1}|$$

$$= \prod_{\ell=1}^{s} \lambda_\ell |A\, A^{-1} E'_{i_r j_r}(\lambda_r)\cdots E'_{i_1 j_1}(\lambda_1)|$$

$$= |A\, A^{-1} E'_{i_r j_r}(\lambda_r)\cdots E'_{i_1 j_1}(\lambda_1) M_{i_s}(\lambda_s)\cdots M_{i_1}(\lambda_1)|$$

$$= |A\, A^{-1} A'| = |A'| \quad .$$

Theorem 2 says that row and column operations have identical effects on the value of the determinant. Thus since P_{ij} is an ER, directly from the definition we see that

$$|A\ P_{ij}| = -|A|$$

so that

$$|P_{ij}\ A| = -|A| \quad .$$

Also, because of Theorem 2, we have results analogous to lemmas 1 and 2, i.e.

$$|M_i(\lambda)A| = \lambda|A|$$

and

$$|E_{ij}(\lambda)A| = |A| \quad .$$

Further, we now know that we could have equivalently used post-multiplication to triangularize A in our definition.

With Theorems 1 and 2 we have shown that our theory of determinants is coincident with the usual definition.

5. Application of New Definition to Jacobians. We will use our new definition of the determinant to help motivate the Jacobian. Although the language is given in the context of probability and statistics, it is clear that it carries through for Jacobians of transformations used in change of variable for integration.

Suppose that $y_1 = u_1(x_1,x_2)$ and $y_2 = u_2(x_1,x_2)$ is a one-to-one transformation which maps a (two-dimensional) set $\mathcal{A}$ in the x_1x_2-plane onto a (two-dimensional) set $\mathcal{B}$ in the y_1y_2-plane. Let $x_1 = w_1(y_1,y_2)$ and $x_2 = w_2(y_1,y_2)$ represent the inverse transformation. The determinant of order two

$$J = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} \\ \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} \end{vmatrix}$$

is commonly called the Jacobian of the transformation. If X_1 and X_2 are continuous random variables having joint density function $f(x_1,x_2)$ and $Y_1 = u_1(X_1,X_2)$ and $Y_2 = u_2(X_1,X_2)$ then the joint density function of Y_1 and Y_2 is

$$g(y_1,y_2) = f(w_1(y_1,y_2),w_2(y_1,y_2))|J| \quad .$$

(See [2]).

We wish to provide motivation for this via our definition of the determinant. Consider a small rectangular region $\mathcal{S}$ in the y_1y_2-plane with vertices as shown in Figure 3.

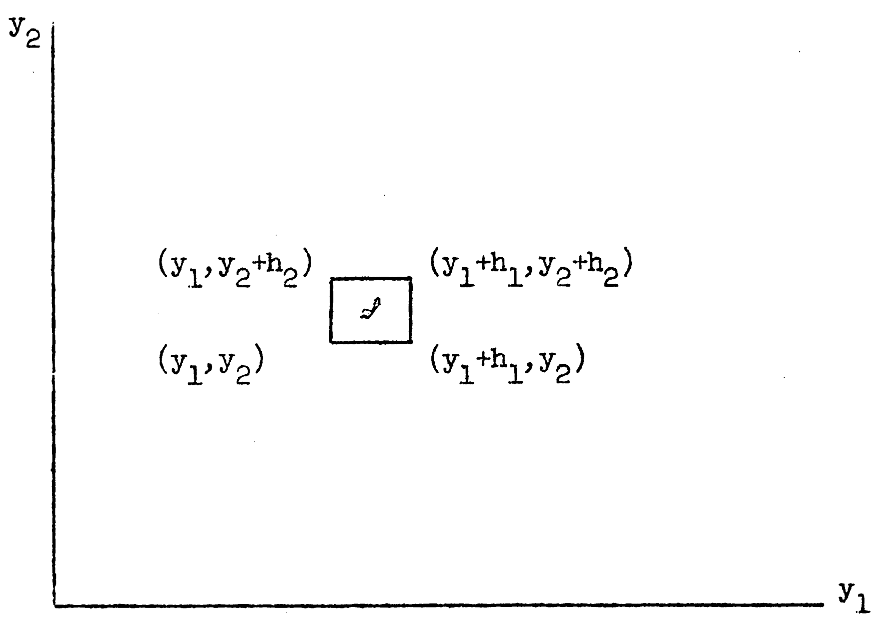
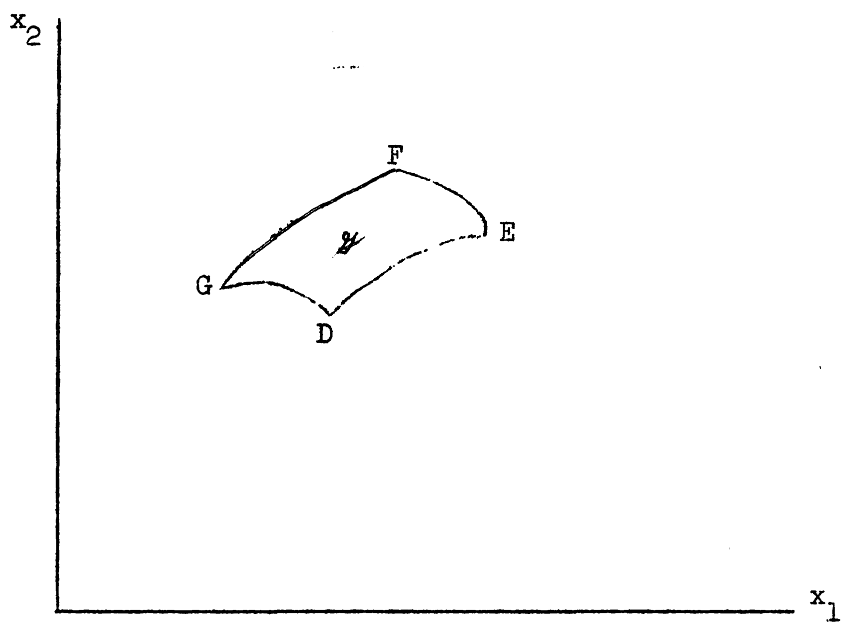


Figure 3

By definition $g(y_1,y_2) = \lim\limits_{\substack{h_1\to 0\\ h_2\to 0}} \dfrac{P((Y_1,Y_2) \in \mathcal{S})}{h_1h_2}$.

We need an expression for $P((Y_1,Y_2) \in \mathcal{S})$. $\mathcal{A}$ is the region in the x_1x_2-plane that is transformed onto $\mathcal{S}$. It is represented symbolically with its vertices in Figure 4.



$D = (w_1(y_1,y_2), w_2(y_1,y_2))$

$E = (w_1(y_1,y_2+h_2), w_2(y_1,y_2+h_2))$

$F = (w_1(y_1+h_1,y_2+h_2), w_2(y_1+h_1,y_2+h_2))$

$G = (w_1(y_1+h_1,y_2), w_2(y_1+h_1,y_2))$

Figure 4

We know that $P((Y_1,Y_2) \in \mathcal{S}) \doteq f(w_1(y_1,y_2), w_2(y_1,y_2))v_A$ if h_1 and h_2 are sufficiently small. Here v_A is the area of $\mathcal{A}$. v_A is approximated by the area of a parallelogram defined by the vectors from D to G and from D to E. These vectors adjoined to form a matrix are

$$A = \begin{bmatrix} w_1(y_1+h_1,y_2) - w_1(y_1,y_2) & w_1(y_1,y_2+h_2) - w_1(y_1,y_2) \\ w_2(y_1+h_1,y_2) - w_2(y_1,y_2) & w_2(y_1,y_2+h_2) - w_2(y_1,y_2+h_2) \end{bmatrix} .$$

Thus

$$\frac{P((Y_1,Y_2) \in \mathscr{S})}{h_1h_2} \doteq f(w_1(y_1,y_2),\ w_2(y_1,y_2)) \frac{|A|}{h_1h_2} \quad .$$

When we take the limits as h_1 and h_2 approach zero the expressions are equal and we have

$$g(y_1y_2) = f(w_1(y_1y_2),\ w_2(y_1y_2))|J| \quad .$$

In summary, the Jacobian is simply a scaling factor that gives us the area of a region in the x_1x_2-plane that corresponds to an infinitesimal rectangular region in the y_1y_2-plane.

In conclusion, we note that our definition can be used to motivate any area of mathematics where determinants are used.

REFERENCES

[1] G. BIRKHOFF AND S. MACLANE. A Survey of Modern Algebra, Third Edition, The Macmillan Company, London, 1965, pages 280-291.

[2] R. V. HOGG AND A. T. CRAIG. Introduction to Mathematical Statistics, Third Edition, The Macmillan Company, London, 1970, page 128.

[3] H. R. SCHWARZ, H. RUTSHAUSER AND E. STIEFEL. Numerical Analysis of Symmetric Matrices, Translated by P. Hertelendy, Prentice-Hall, Englewood Cliffs, N. J., 1973, pages 139-144.

[4] G. W. STEWART. Introduction to Matrix Computations, Academic Press, N. Y. and London, 1973, pages 230-244 and 396-404.