Pacific Journal of Mathematics

THE JACOBIAN OF A GROWTH TRANSFORMATION

DONALD STEVEN PASSMAN

Vol. 44, No. 1 May 1973

PACIFIC JOURNAL OF MATHEMATICS
Vol. 44, No. 1, 1973

THE JACOBIAN OF A GROWTH TRANSFORMATION

D. S. PASSMAN

The transformation T, described in a paper of Baum and Eagon, is frequently a growth transformation which affords an iterative technique for maximizing certain functions. In this paper, the Jacobian matrix J of T is studied. It is shown, for example, that the eigenvalues of J are real and nonnegative in a large number of cases. In addition, these eigenvalues are considered at critical points of T. One necessary assumption used throughout is that the function P to be maximized is homogeneous in the variables involved.

The author would like to thank P. Stebe for a number of stimulating conversations and useful suggestions.

1. Notation. Let P be a function of the variables x_{ij} with domain of definition D given by

$$x_{ij} > 0 \quad \text{and} \quad \sum_j x_{ij} = 1 .$$

Assume that on this domain both P and all its partial derivatives $\partial P/\partial x_{ij}$ are positive. Moreover we assume that the second partial derivatives of P exist and are continuous. Then the particular transformation T of P which we study here is given by (see [1])

$$T(x_{ij}) = \frac{x_{ij}\partial P/\partial x_{ij}}{\sum_k x_{ik}\partial P/\partial x_{ik}} . \tag{1.1}$$

Clearly T maps D into D.

We say that P is row homogeneous if for each i P is homogeneous of degree $w_i > 0$ in the variables $x_{i1}, x_{i2}, \cdots$. In this case (1.1) simplifies by means of Euler's formula and we obtain

$$T(x_{ij}) = \frac{x_{ij}\partial P/\partial x_{ij}}{w_i P} . \tag{1.2}$$

Let us assume now that P is row homogeneous. While the double subscript on the symbol x_{ij} makes the domain D easier to visualize, it turns out that a single subscript makes our later computations neater. Therefore we make the following notational change. Observe that the variables $\{x_{ij}\}$ are not all independent because of the constraints $\sum_j x_{ij} = 1$. Thus in each row of the array (x_{ij}) there is one variable which is dependent upon the others. Let us suppose now that there are a total of n' variables of which n are independent. We can then write the variables $\{x_{ij}\}$ as

$$x_1, x_2, \cdots, x_n, x_{n+1}, \cdots, x_{n'}$$

where $x_1, x_2, \cdots, x_n$ are independent and the remaining ones are dependent. Now the set $\{x_{n+1}, x_{n+2}, \cdots, x_{n'}\}$ clearly contains precisely one variable from each row so we can certainly use the subscripts $n+1, n+2, \cdots, n'$ to designate these rows. Finally we introduce the function

$$f\colon \{1, 2, \cdots, n\} \longrightarrow \{n+1,\ n+2, \cdots, n'\}$$

so that $f(i)$ indicates the row containing x_i.

In this single subscripted notation we see that for $i, j \leqq n$, x_i and x_j are in the same row if and only if $f(i) = f(j)$. Thus the fact that the sum of the variables in the row containing x_i is 1 becomes

$$(1.3) \qquad 1 - x_{f(i)} = \sum_{j=1}^{n} \delta_{f(i)f(j)} x_j \qquad \text{for } i \leqq n\,.$$

Also setting $y_i = T(x_i)$ equation (1.2) now reads

$$(1.4) \qquad y_i = \frac{x_i \partial P/\partial x_i}{w_{f(i)} p} \qquad \text{for } i \leqq n\,.$$

Now suppose that Q is a function of $x_1, \cdots, x_{n'}$. Then we use as above $\partial Q/\partial x_i$ to denote the partial derivative of Q with respect to x_i. On the other hand, Q can be viewed as a function of the independent variables $x_1, \cdots, x_n$. If we do this, then we use dQ/dx_i to denote the partial derivative of Q with respect to x_i for $i \leqq n$. It follows from (1.3) and the chain rule that

$$(1.5) \qquad \frac{dQ}{dx_i} = \frac{\partial Q}{\partial x_i} - \frac{\partial Q}{\partial x_{f(i)}} \qquad \text{for } i \leqq n\,.$$

The Jacobian of the growth transformation T is the $n \times n$ matrix

$$(1.6) \qquad J = \left[\frac{dy_i}{dx_j}\right] \qquad i, j \leqq n\,.$$

It is the matrix which we plan to study.

2. Real eigenvalues. Let N denote the $n \times n$ symmetric matrix

$$N = \left[\frac{x_i}{w_{f(i)}}(\delta_{ij} - \delta_{f(i)f(j)}x_j)\right] . \tag{2.1}$$

The interplay of this matrix with J will prove to be of fundamental importance.

LEMMA 1. *N is a positive definite matrix.*

Proof. Since the ordering of the variables $x_1, x_2, \cdots, x_n$ does not effect the nature of N, we may assume that the variables are grouped together according to the row of the array (x_{ij}) they are contained in. Then N clearly becomes a block diagonal matrix with each block corresponding to a row of (x_{ij}). Since it clearly suffices to show that each of these blocks is a positive definite matrix, it therefore suffices to consider the case in which (x_{ij}) has only one row. Thus $n' = n + 1$ and

$$w_{n+1} N = [\delta_{ij}x_i - x_i x_j] .$$

Let z be the real row vector $z = [z_1, z_2, \cdots, z_n] \neq 0$ and set

$$u = [\sqrt{x_1}, \sqrt{x_2}, \cdots, \sqrt{x_n}]$$
$$v = [\sqrt{x_1}z_1, \sqrt{x_2}z_2, \cdots, \sqrt{x_n}z_n] .$$

Then using (,) for the usual inner product of vectors we have

$$\begin{aligned} z(w_{n+1}N)z^T &= \sum_1^n x_i z_i^2 - \left(\sum_1^n x_i z_i\right)^2 \\ &= (v, v) - (u, v)^2 > (u, u)(v, v) - (u, v)^2 \geqq 0 \end{aligned}$$

by Cauchy's inequality and the fact that

$$(u, u) = x_1 + x_2 + \cdots + x_n = 1 - x_{n+1} < 1 .$$

The lemma is proved.

THEOREM 2. *Let P be a row homogeneous function. Then*

$$JN = [x_j/w_{f(j)}\partial y_i/\partial x_j]$$

is a symmetric matrix.

Proof. From (1.4) it is clear that y_i is row homogeneous of degree zero. Thus Euler's equation yields

$$x_{f(j)}\partial y_i/\partial x_{f(j)} + \sum_{k=1}^n \delta_{f(j)f(k)}x_k\partial y_i/\partial x_k = 0 . \tag{2.2}$$

Let $JN = [h_{ij}]$. Then by (1.5) and the symmetry of N we have

$$\begin{aligned} w_{f(j)}h_{ij} &= \sum_{k=1}^{n} \frac{dy_i}{dx_k} \cdot x_k(\delta_{kj} - \delta_{f(k)f(j)}x_j) \\ &= x_j \frac{dy_i}{dx_j} - x_j \sum_{k=0}^{n} \delta_{f(k)f(j)} x_k \frac{dy_i}{dx_k} \\ &= x_j \left(\frac{\partial y_i}{\partial x_j} - \frac{\partial y_i}{\partial x_{f(j)}} \right) - x_j \sum_{k=1}^{n} \delta_{f(k)f(j)} x_k \frac{\partial y_i}{\partial x_k} \\ &\quad + x_j \frac{\partial y_i}{\partial x_{f(j)}} \sum_{k=1}^{n} \delta_{f(k)f(j)} x_k \, . \end{aligned}$$

Thus (2.2) and (1.3) yield

$$\begin{aligned} w_{f(j)}h_{ij} &= x_j \left(\frac{\partial y_i}{\partial x_j} - \frac{\partial y_i}{\partial x_{f(j)}} \right) + x_j x_{f(j)} \frac{\partial y_i}{\partial x_{f(j)}} \\ &\quad + x_j(1 - x_{f(j)}) \frac{\partial y_i}{\partial x_{f(j)}} = x_j \frac{\partial y_i}{\partial x_j} \, . \end{aligned}$$

Therefore we have

$$h_{ij} = x_j / w_{f(j)} \partial y_i / \partial x_j \, .$$

It remains to show that $h_{ij} = h_{ji}$ and to do this we may assume that $i \neq j$. Then by (1.4)

$$h_{ij} = \frac{x_i x_j}{w_{f(i)} w_{f(j)} P^2} (P \partial^2 P / \partial x_i \partial x_j - (\partial P / \partial x_i)(\partial P / \partial x_j))$$

so the result clearly follows.

COROLLARY 3. *Let P be a row homogeneous function. Then J is diagonalizable and all eigenvalues of J are real.*

Proof. Write $JN = A$. Since N is positive definite we have $N = QQ^T$ for some real nonsingular matrix Q. Set $R = Q^{-1}$. Then we have easily

$$RJR^{-1} = RAR^T \, . \tag{2.3}$$

Since RAR^T is real and symmetric by Theorem 2, it is diagonalizable with all real eigenvalues. Thus (2.3) yields the result.

3. Critical points. In this section we study in more detail the nature of J at a critical point. It follows from (1.4) and (1.5) that at such a point we have $x_i = y_i$ and $\partial P / \partial x_i = w_{f(i)} P$. Recall that a critical point is a point at which

$$\frac{dP}{dx_i} = 0 \qquad \text{for } i \leqq n \, . \tag{3.1}$$

THEOREM 4. *At a critical point we have*

$$J = I + \frac{N}{P}\left[\frac{d^2P}{dx_i dx_j}\right]$$

where I is the $n \times n$ identity matrix.

Proof. We start with Euler's equation for the row homogeneity of P. For $i \leqq n$ we have

$$w_{f(i)}P = x_{f(i)}\partial P/\partial x_{f(i)} + \sum_{k=1}^{n} \delta_{f(i)f(k)} x_k \partial P/\partial x_k$$

and differentiating this identity with respect to x_j yields

$$w_{f(i)}\frac{dP}{dx_j} = x_{f(i)}\frac{d}{dx_j}\partial P/\partial x_{f(i)} + \sum_{k=1}^{n} \delta_{f(i)f(k)} x_k \frac{d}{dx_j}\partial P/\partial x_k + \delta_{f(i)f(j)}(\partial P/\partial x_j - \partial P/\partial x_{f(i)}) .$$

Observe that the last term is just $\delta_{f(i)f(j)}dP/dx_j$ so the above becomes

$$(w_{f(i)} - \delta_{f(i)f(j)})\frac{dP}{dx_j} = x_{f(i)}\frac{d}{dx_j}\partial P/\partial x_{f(i)} + \sum_{k=1}^{n} \delta_{f(i)f(k)} x_k \frac{d}{dx_j}\partial P/\partial x_k .$$

Now at a critical point $dP/dx_j = 0$ so

(3.2) $$0 = x_{f(i)}\frac{d}{dx_j}\partial P/\partial x_{f(i)} + \sum_{k=1}^{n} \delta_{f(i)f(k)} x_k \frac{d}{dx_j}\partial P/\partial x_k .$$

By (1.5) for $i, j \leqq n$

(3.3) $$\frac{d^2P}{dx_i dx_j} = \frac{d}{dx_j}\frac{\partial P}{\partial x_i} - \frac{d}{dx_j}\frac{\partial P}{\partial x_{f(i)}}$$

and substituting

$$\frac{d}{dx_j}\frac{\partial P}{\partial x_k} = \frac{d^2P}{dx_k dx_j} + \frac{d}{dx_j}\frac{\partial P}{\partial x_{f(k)}}$$

into (3.2) yields

(3.4) $$0 = x_{f(i)}\frac{d}{dx_j}\frac{\partial P}{\partial x_{f(i)}} + \sum_{k=1}^{n} \delta_{f(i)f(k)} x_k \frac{d}{dx_j}\frac{\partial P}{\partial x_{f(k)}} + \sum_{k=1}^{n} \delta_{f(i)f(k)} x_k \frac{d^2P}{dx_k dx_j} .$$

Now clearly

$$\delta_{f(i)f(k)}x_k \frac{d}{dx_j}\frac{\partial P}{\partial x_{f(k)}} = \delta_{f(i)f(k)}x_k \frac{d}{dx_j}\frac{\partial P}{\partial x_{f(i)}}$$

so (3.4) becomes

$$0 = \frac{d}{dx_j}\frac{\partial P}{\partial x_{f(i)}}\left(x_{f(i)} + \sum_{k=1}^{n} \delta_{f(i)f(k)}x_k\right) + \sum_{k=1}^{n} \delta_{f(i)f(k)}x_k \frac{d^2P}{dx_k dx_j}\,.$$

Hence by (1.3) we have

$$\frac{d}{dx_j}\frac{\partial P}{\partial x_{f(i)}} = -\sum_{k=1}^{n} \delta_{f(i)f(k)}x_k \frac{d^2P}{dx_k dx_j}\,. \tag{3.5}$$

We now compute J at the critical point. By (1.4) and (3.1)

$$\begin{aligned}\frac{dy_i}{dx_j} &= \delta_{ij}\frac{\partial P/\partial x_i}{w_{f(i)}P} + \frac{x_i}{w_{f(i)}P}\frac{d}{dx_j}\frac{\partial P}{\partial x_i} \\ &= \delta_{ij} + \frac{x_i}{w_{f(i)}P}\frac{d}{dx_j}\frac{\partial P}{\partial x_i}\end{aligned}$$

since at a critical point $\partial P/\partial x_i = w_{f(i)}P$. Thus

$$J = I + \frac{1}{P}\left[\frac{x_i}{w_{f(i)}}\frac{d}{dx_j}\frac{\partial P}{\partial x_i}\right]. \tag{3.6}$$

Let $E = [e_{ij}]$ denote the latter matrix. Then using (3.3) and (3.5) we have

$$\begin{aligned}e_{ij} &= \frac{x_i}{w_{f(i)}}\left(\frac{d^2P}{dx_i dx_j} + \frac{d}{dx_j}\frac{\partial P}{\partial x_{f(i)}}\right) \\ &= \frac{x_i}{w_{f(i)}}\sum_{k=1}^{n}(\delta_{ik} - \delta_{f(i)f(k)}x_k)\frac{d^2P}{dx_k dx_j}\end{aligned}$$

and this is the (i, j)th entry in the matrix product

$$\left[\frac{x_i}{w_{f(i)}}(\delta_{ij} - \delta_{f(i)f(j)}x_j)\right]\left[\frac{d^2P}{dx_i dx_j}\right].$$

In view of (3.6), the result follows.

Let B denote the matrix

$$B = \left[\frac{d^2P}{dx_i dx_j}\right]. \tag{3.7}$$

COROLLARY 5. *Suppose that at a critical point we have* $\det B \neq 0$ *and let* λ *be an eigenvalue of* J. *Then*

(i) *at a minimum*, $\lambda > 1$
(ii) *at a maximum*, $\lambda < 1$.

Proof. By Theorem 4, $\lambda = 1 + \mu/P$ where μ is an eigenvalue of NB and by Corollary 3, λ is real. Thus it suffices to show that μ is positive at a minimum and negative at a maximum.

Let v be a real column eigenvector for μ. Then $NBv = \mu v$ yields easily

$$v^T Bv = \mu(u^T Nu) \tag{3.8}$$

where $v = Nu$. Since N is positive definite by Lemma 1 we have $u^T Nu > 0$.

Now at a minimum, since $\det B \neq 0$, we see that B is a positive definite matrix. Thus $v^T Bv > 0$ and $\mu > 0$ by (3.8). Similarly at a maximum, B is negative definite so $\mu < 0$. This completes the proof.

4. Polynomials.

We assume here that P is a row homogeneous function and use the notation of the preceding sections. In addition, we assume that P is a polynomial with positive coefficients so that (in single subscripted variables)

$$P = \sum_a m_a \tag{4.1}$$

where

$$m_a = e_a x_1^{a_1} x_2^{a_2} \cdots x_{n'}^{a_{n'}}, \quad e_a > 0 . \tag{4.2}$$

Here, of course, a designates the n'-tuple. $a = (a_1, a_2, \cdots, a_{n'})$. Let $\mathscr{A}$ denote the set of all such a's which occur in P.

Fix some ordering of the a's and let α denote a subset $\{a, b\}$ of $\mathscr{A}$ with $b > a$. For each such α set

$$m_\alpha = m_a m_b , \quad \alpha_i = b_i - a_i . \tag{4.3}$$

Since P is row homogeneous we have for $i \leqq n$

$$a_{f(i)} + \sum_{j=1}^{n} \delta_{f(i)f(j)} a_j = w_{f(i)}$$

and hence (4.3) yields

$$\alpha_{f(i)} + \sum_{j=1}^{n} \delta_{f(i)f(j)} \alpha_j = 0 . \tag{4.4}$$

For the row homogeneous polynomial P we define the vector subspace $V(P) \subseteqq R^{n'}$ to be the subspace of $R^{n'}$ spanned by all n'-tuples $(\alpha_1, \alpha_2, \cdots, \alpha_{n'})$ with $\alpha = \{a, b\}$, $a, b \in \mathscr{A}$. In view of (4.4) we have certainly

$$\dim V(P) \leqq n \, .$$

THEOREM 6. *Let P be a row homogeneous polynomial. Then JN is a positive semi-definite symmetric matrix with*

$$\operatorname{rank} JN = \dim V(P) \, .$$

Proof. Let P be given by (4.1) and (4.2). Then by (1.4)

$$w_{f(i)} y_i = \frac{\Sigma a_i m_a}{\Sigma m_a}$$

and we have

$$\begin{aligned} w_{f(i)} P^2 x_j \partial y_i / \partial x_j &= (\sum_a m_a)(\sum_b b_i b_j m_b) - (\sum_a a_j m_a)(\sum_b b_i m_b) \\ &= \sum_{a,b} m_a m_b (b_i b_j - b_i a_j) \, . \end{aligned}$$

Observe that the inner summand vanishes at $a = b$. Thus if we sum over $a < b$ then we obtain

$$\begin{aligned} w_{f(i)} P^2 x_j \frac{\partial y_i}{\partial x_j} &= \sum_{a<b} m_a m_b (b_i b_j - b_i a_j + a_i a_j - a_i b_j) \\ &= \sum_{\alpha} m_\alpha \alpha_i \alpha_j \end{aligned}$$

in the notation of (4.3). Thus by Theorem 2

$$P^2 JN = [\sum_\alpha m_\alpha \alpha_i \alpha_j / w_{f(i)} w_{f(j)}] \, . \tag{4.5}$$

Let $z = [z_1, z_2, \cdots, z_n]$ be a row vector of real entries. Then

$$\begin{aligned} z(P^2 JN) z^T &= \sum_{i,j} \sum_\alpha m_\alpha \alpha_i \alpha_j z_i z_j / w_{f(i)} w_{f(j)} \\ &= \sum_\alpha m_\alpha (\sum_i \alpha_i z_i / w_{f(i)})^2 \, . \end{aligned} \tag{4.6}$$

Thus clearly $P^2 JN$ and hence JN is positive semi-definite.

It remains to compute the rank of JN. Let $W(P) \subseteqq R^n$ be the subspace of R^n spanned by all n-tuples $\tilde{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_n)$. In view of (4.4) we have clearly

$$\dim W(P) = \dim V(P) \, . \tag{4.7}$$

Let $[\ ,\ ]$ be the inner product on R^n defined by

$$[(u_1, u_2, \cdots, u_n), (v_1, v_2, \cdots, v_n)] = \sum_{i=1}^n u_i v_i / w_{f(i)} \, .$$

Then (4.6) becomes

$$z(P^2JN)z^T = \sum_\alpha m_\alpha [\tilde{\alpha}, z]^2 .$$

Thus $z(P^2JN)z^T = 0$ if and only if $z \in W(P)^\perp$, the orthogonal compliment of $W(P)$. Finally since P^2JN is positive semi-definite we have

$$\operatorname{rank} P^2JN = n - \dim W(P)^\perp = \dim W(P)$$

and the result follows from (4.7).

COROLLARY 7. *Let P be a row homogeneous polynomial. Then all eigenvalues of J are non-negative real numbers and hence* $\det J \geqq 0$.

Proof. This follows immediately from (2.3) and Theorem 6.

Observe that Theorem 6 implies that $\det J > 0$ if $\dim V(P) = n$ and $\det J = 0$ otherwise. This fact is an unpublished result of L. Baum.

5. **Examples.** In this section, we consider a number of examples with P not homogeneous. Suppose y_1 is given by

$$y_1 = \frac{x_1 \partial P/\partial x_1}{x_1 \partial P/\partial x_1 + x_2 \partial P/\partial x_2} .$$

Then

$$\frac{1}{1 - y_1} = 1 + \frac{x_1 \partial P/\partial x_1}{x_2 \partial P/\partial x_2} .$$

Differentiating with respect to some variable x then yields

$$\frac{dy_1}{dx} = (1 - y_1)^2 \frac{d}{dx}\left(\frac{x_1 \partial P/\partial x_1}{x_2 \partial P/\partial x_2}\right) .$$

This formula enables the following computations to be done easily.

Let $P(x_1, x_2) = x_1 + x_1^2 x_2$. Then

$$\det J = \frac{dy_1}{dx_1} = \frac{(1 - y_1)^2}{(x_1 x_2)^2}(2x_1 - 1)$$

and this changes sign at $x_1 = 1/2$. Thus Corollary 7 requires that P be homogeneous.

Now let

$$P(x_1, x_2, x_3, x_4) = x_1 x_2^2 + x_1^2 x_3 + x_2^2 x_4$$

subject to the constraints $x_1 + x_3 = 1$, $x_2 + x_4 = 1$. Then

$$J = \begin{bmatrix} \dfrac{(1-y_1)^2}{x_1^2 x_3^2} x_2^2(2x_1 - 1) & \dfrac{(1-y_1)^2 2x_2}{x_1 x_3} \\ \dfrac{(1-y_2)^2\, 2}{x_4} & \dfrac{(1-y_2)^2 2x_1}{x_4^2} \end{bmatrix}.$$

Thus

$$\frac{x_4^2}{(1-y_1)^2(1-y_2)^2} \det J = \begin{vmatrix} \dfrac{x_2^2}{x_1^2 x_3^2}(2x_1 - 1) & \dfrac{2x_2}{x_1 x_3} \\ 2x_4 & 2x_1 \end{vmatrix}.$$

Finally let $x_2 \sim 1$ so $x_4 \sim 0$ and the right hand determinant is approximately equal to

$$\frac{2(2x_1 - 1)}{x_1 x_3^2}$$

which changes sign at $x_1 = 1/2$. Thus we see that even though P is homogeneous, Corollary 7 can still fail unless P is row homogeneous.

References

1. L. Baum and J. Eagon, *An inequality with application to statistical estimation for probabilistic functions of Markov processes and to a model for ecology*, Bull. Amer. Math. Soc., **73** (1967), 360-363.

Received August 31, 1971.

Institute for Defense Analyses
PRINCETON, NEW JERSEY
AND
UNIVERSITY OF WISCONSIN, MADISON

Printed in Japan by International Academic Printing Co., Ltd., Tokyo, Japan

Pacific Journal of Mathematics

Vol. 44, No. 1 May, 1973