STABILIZATION BY A DIAGONAL MATRIX

C. S. BALLANTINE

ABSTRACT. In this paper it is shown that, given a complex square matrix A all of whose leading principal minors are nonzero, there is a diagonal matrix D such that the product DA of the two matrices has all its characteristic roots positive and simple. This result is already known for real A, but two new proofs for this case are given here.

1. **The real case.** A theorem proved by Fisher and Fuller [2] is an obvious consequence of the following result (Theorem 1), which in turn is the real case of our Theorem 2 below.

THEOREM 1 (FISHER, FULLER). *Let A be an $n\times n$ real matrix all of whose leading principal minors are positive. Then there is an $n\times n$ positive diagonal matrix D such that all the roots of DA are positive and simple.*

We shall give here two proofs of Theorem 1, both of them simpler than the proof in [2]. Our first proof is the shorter of the two, but is less constructive since it makes use of the continuity of the roots (as functions of the matrix entries). Our second proof gives explicit (and relatively simple) estimates for the entries of D in terms of the entries of A.

FIRST PROOF OF THEOREM 1. Here we use induction on n. For $n=1$ the result is trivial, so suppose that $n\geqq 2$ and that the result holds for matrices of order $n-1$. Let A be an $n\times n$ real matrix all of whose lpm's (leading principal minors) are positive and let A_1 be its leading principal submatrix of order $n-1$. Then all the lpm's of A_1 are positive, so by our induction assertion there is a positive diagonal matrix D_1 of order $n-1$ such that all roots of D_1A_1 are positive and simple. Let d be a real number to be determined later (but treated as a variable for the present). Let A be partitioned as follows:

$$A=\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}$$

Received by the editors September 12, 1969.

AMS Subject Classifications. Primary 1555, 1525; Secondary 2660, 2655, 6540.

Key Words and Phrases. Diagonal matrix, stable matrix, leading principal minors, characteristic roots, positive simple roots, continuity of roots, separation of roots, parallelotope, continuous mapping.

(where A_4 is 1×1). Define an $n\times n$ diagonal matrix D (depending on d) by conformable partition:

$$D = \begin{bmatrix} D_1 & 0 \\ 0 & d \end{bmatrix}.$$

Let $DA = M(d)$, where now we emphasize the dependence on d. Then

$$M(0) = \begin{bmatrix} D_1A_1 & D_1A_2 \\ 0 & 0 \end{bmatrix},$$

so the nonzero roots of $M(0)$ are just those of D_1A_1, hence are positive and simple (and there are exactly $n-1$ of them). $M(0)$ also has a simple root at zero. Thus for all sufficiently small $d>0$ the real parts of the roots of $M(d)$ are (still) n distinct real numbers at least $n-1$ of which are positive. (This follows from the fact that the roots of $M(d)$ are continuous functions of d.) Choose some such d. Then the roots of $M(d)$ are all real and simple (since nonreal roots must occur in conjugate pairs) and at least $n-1$ of them are positive. But the determinant of $M(d)$ is positive since those of A and D are, so in fact all n roots of $M(d)$ are positive. This concludes the proof of the induction step and hence of Theorem 1.

REMARK 1. This same kind of argument can be used to prove that, when all the lpm's of A are positive and a sign pattern is given for n-tuples of real numbers ordered by their absolute values, there is a real diagonal matrix D of the prescribed sign pattern such that the roots of DA also have the prescribed sign pattern. Also, when A is an $n\times n$ complex matrix all of whose lpm's are nonzero and also an open sector containing the positive real axis is prescribed, this same kind of argument yields a complex $n\times n$ diagonal matrix D such that all the roots of DA lie in the prescribed sector (in fact, if all the lpm's of A are positive then D can be chosen positive).

For our second proof of Theorem 1 we shall use the following fact about real polynomials.

FACT 1. Let n be an integer $\geqq 2$ and let $c_0, c_1, \cdots, c_n$ be positive real numbers satisfying all the following inequalities:

$$4c_0c_2 < c_1^2, 4c_1c_3 < c_2^2, \cdots, 4c_{n-2}c_n < c_{n-1}^2.$$

Let $x_k = (c_{k+1}/c_{k-1})^{1/2}$ for $k = 1, 2, \cdots, n-1$. Then all the roots of the polynomial

$$f(x) = c_0x^n - c_1x^{n-1} + c_2x^{n-2} - \cdots + (-1)^nc_n$$

are real (hence positive) and simple, and they are separated by the $n-1$ numbers $x_1, x_2, \cdots, x_{n-1}$.

PROOF. (This result is probably known, is perhaps even classical, but it is not standard for $n \geqq 3$, so we shall give a short proof here for the sake of completeness.) Let $x_0 = c_1/c_0$ and $x_n = c_n/c_{n-1}$. Then $x_0 > x_1 > x_2 > \cdots > x_{n-1} > x_n$; in fact,

$$\left(\frac{c_1}{c_0}\right)^2 > \left(\frac{c_2}{c_0}\right) > \left(\frac{c_2}{c_1}\right)^2 > \left(\frac{c_3}{c_1}\right) > \left(\frac{c_3}{c_2}\right)^2$$
$$> \left(\frac{c_4}{c_2}\right) > \cdots > \left(\frac{c_n}{c_{n-1}}\right)^2$$

holds by hypothesis, and hence

$$x_0 > x_1 > c_2/c_1 > x_2 > c_3/c_2 > \cdots > x_{n-1} > x_n.$$

Thus it suffices to show that $(-1)^k f(x_k) > 0$ for $k = 0, 1, 2, \cdots, n$.

From the last chain of inequalities we have that $c_j x_k - c_{j+1}$ is

(1) positive if $0 \leqq k < j \leqq n-1$ (and in other cases which we shall not need here),

(2) negative if $1 \leqq j+1 < k \leqq n$, and

(3) zero if $0 = k = j$ or $j+1 = k = n$. Thus we have

$$f(x_0) = x_0^{n-1}(c_0 x_0 - c_1) + x_0^{n-3}(c_2 x_0 - c_3) + \cdots > 0,$$
$$(-1)^n f(x_n) = (c_n - c_{n-1}x_n) + x_n^2(c_{n-2} - c_{n-3}x_n) + \cdots > 0$$

since $n \geqq 2$ and the odd term at the end (when n is even) is positive in each case.

To handle the x_k for which $1 \leqq k < n$, we write

$$f(x) = x^{n-k+2}g(x) + (-1)^k x^{n-k-1}(-c_{k-1}x^2 + c_k x - c_{k+1}) + h(x),$$

where we have put

$$g(x) = c_0 x^{k-2} - c_1 x^{k-3} + c_2 x^{k-4} - \cdots + (-1)^{k-2} c_{k-2},$$
$$h(x) = (c_{k+2}x^{n-k-2} - c_{k+3}x^{n-k-3} + \cdots + (-1)^{n-k-2}c_n)(-1)^{k+2}.$$

We first show that $(-1)^k g(x_k) \geqq 0$ and $(-1)^k h(x_k) \geqq 0$. Namely,

$$(-1)^k g(x_k) = (c_{k-2} - c_{k-3}x_k) + x_k^2(c_{k-4} - c_{k-5}x_k) + \cdots \geqq 0,$$

$$(-1)^k h(x_k) = x_k^{n-k-3}(c_{k+2}x_k - c_{k+3}) + x_k^{n-k-5}(c_{k+4}x_k - c_{k+5}) + \cdots \geqq 0.$$

Thus it remains only to show that $-c_{k-1}x_k^2 + c_k x_k - c_{k+1}$ is positive. But this is a routine consequence of our hypothesis that $c_k^2 > 4c_{k-1}c_{k+1}$

and our definition of x_k. Thus, as asserted, $(-1)^k f(x_k) > 0$ for $k = 0, 1, 2, \cdots, n$, and Fact 1 is proved.

SECOND PROOF OF THEOREM 1. We first assume that all lpm's of A are 1. For $k = 1, 2, \cdots, n$ let q_k be the sum of the absolute values of the *non*leading principal minors of order k in A (thus $q_n = 0$). Choose positive real numbers $d_1, d_2, \cdots, d_n$ so that

$$2d_{k+1}q_k < d_k \quad \text{and} \quad 36d_{k+1} < d_k$$

for $k = 1, 2, \cdots, n-1$. (One can choose d_j arbitrarily > 0 for one j and choose the other d_k recursively, going outward from $k = j$). Let $D = \operatorname{diag}(d_1, d_2, \cdots, d_n)$, and define $c_0\ (=1), c_1, c_2, \cdots, c_n$ by

$$\det(xI - DA) = c_0x^n - c_1x^{n-1} + c_2x^{n-2} - \cdots + (-1)^n c_n.$$

Thus by Fact 1, in order to show that the roots of DA are positive and simple, it suffices to show that $c_1, c_2, \cdots, c_n$ are all positive and that $c_k^2 > 4c_{k-1}c_{k+1}$ for $k = 1, 2, \cdots, n-1$. This we do as follows.

For $k = 0, 1, 2, \cdots, n$ we can write $c_k = p_k + R_k$, where $p_k = d_1d_2 \cdots d_k$ (hence $p_0 = 1$) and R_k is the sum of the nonleading principal $k \times k$ minors in DA (hence $R_0 = 0 = R_n$).

Now, each term in R_k is of the form

$$(*) \qquad d_{j_1}d_{j_2} \cdots d_{j_k}m,$$

where m is a nonleading principal $k \times k$ minor of A and $j_1 < j_2 < \cdots < j_k$ and $k < j_k$. Since $d_1 > d_2 > \cdots > d_n > 0$, the absolute value of the term (*) is therefore less than or equal to

$$d_1d_2 \cdots d_{k-1}d_{k+1} \mid m \mid = (d_{k+1}/d_k)p_k \mid m \mid.$$

By the Triangle Inequality and the definition of q_k we thus have (for $1 \leqq k \leqq n-1$)

$$\mid R_k \mid \leqq (d_{k+1}/d_k)p_kq_k < \tfrac{1}{2}p_k,$$

the last inequality coming from the way we chose d_k and d_{k+1}. Thus $c_k = p_k + R_k$ satisfies $p_k/2 < c_k < 3p_k/2$ (for $1 \leqq k \leqq n-1$, but also for $k = 0$ and for $k = n$), so

$$c_k^2 > \tfrac{1}{4}p_k^2 = \tfrac{1}{4}(d_k/d_{k+1})p_{k+1}p_{k-1} > 9p_{k+1}p_{k-1} > 4c_{k+1}c_{k-1}$$

for $k = 1, 2, \cdots, n-1$. This completes the proof for the case where all the lpm's of A are 1.

Returning now to the general case, where the lpm's of A are not necessarily all 1 (but are all positive), we let m_k be the $k \times k$ lpm of A (hence $m_0 = 1$) and let

$$E = \text{diag } (m_0/m_1, m_1/m_2, \cdots, m_{n-1}/m_n).$$

Then all the lpm's of EA are 1, so we can now apply the proof for that case and get the required matrix $D = \text{diag}(d_1, \cdots, d_n)$ for this general case by choosing $d_1, \cdots, d_n$ so that

$$2d_{k+1}m_{k+1}m_{k-1}q_k < d_k m_k^2 \quad \text{and} \quad 0 < 36d_{k+1}m_{k+1}m_{k-1} < d_k m_k^2$$

for $k = 1, 2, \cdots, n-1$, where now q_k is the sum of the absolute values of the nonleading principal $k \times k$ minors of EA. This completes our second proof of Theorem 1.

2. **The complex case.** Here we adapt our second proof of Theorem 1 to yield a proof of the complex case (Theorem 2 below). However, the rest of this latter proof is not constructive, depending as it does on the following result from algebraic topology. (This result is a special case of [1, Lemma 2, p. 232].)

FACT 2. Let P and Q be n-parallelotopes in R^n which are parallel to each other, and let f be a continuous mapping of P into R^n such that f takes each hyperface of P into the closed supporting halfspace of Q at the corresponding hyperface of Q. Then $f(P)$ includes Q.

Using Fact 2, we can now prove the next theorem, which is the main result of this section.

THEOREM 2. *Let A be an $n \times n$ complex matrix all of whose leading principal minors are nonzero. Then there is an $n \times n$ complex diagonal matrix D such that all the roots of DA are positive and simple.*

PROOF. We follow the lines of our second proof of Theorem 1 where we can. Without loss of generality we may assume all lpm's of A are 1. For $k = 1, 2, \cdots, n$, again let q_k be the sum of the absolute values of the $k \times k$ nonleading principal minors of A. Choose positive real numbers $r_1, r_2, \cdots, r_n$ so that

$$2r_{k+1}q_k < r_k \quad \text{and} \quad 36r_{k+1} < r_k$$

for $k = 1, 2, \cdots, n-1$. Now let $d_k = r_k \exp i\theta_k$ for $k = 1, 2, \cdots, n$, where $\theta_1, \cdots, \theta_n$ are real numbers yet to be determined and are treated for now as variables. Let

$$D = \text{diag } (d_1, d_2, \cdots, d_n),$$

$$\det (xI - DA) = c_0x^n - c_1x^{n-1} + \cdots + (-1)^n c_n$$

($c_0 = 1$), as before. Here we define new "variables" $\phi_1, \cdots, \phi_n$, related to $\theta_1, \cdots, \theta_n$ by means of the linear transformation

$$\phi_k = \theta_1 + \theta_2 + \cdots + \theta_k \qquad \text{for} \quad k = 1, 2, \cdots, n,$$
$$\theta_k = \phi_k - \phi_{k-1} \qquad \text{for} \quad k = 2, \cdots, n, \quad \theta_1 = \phi_1,$$

which is one-one of R^n onto R^n.

As before, we write $p_k = d_1 d_2 \cdots d_k$ and $c_k = p_k + R_k$; hence, putting $\phi_0 = 0$, we have

$$p_k = r_1 r_2 \cdots r_k \exp i(\theta_1 + \theta_2 + \cdots + \theta_k) = r_1 r_2 \cdots r_k \exp i\phi_k$$

for $k = 0, 1, 2, \cdots, n$. Again we find that

$$|R_k| < |p_k|/2, \ |p_k|/2 < |c_k| < 3|p_k|/2 \qquad \text{for} \quad 0 \leqq k \leqq n,$$

and

$$|c_k|^2 > 4|c_{k-1}| \cdot |c_{k+1}| \qquad \text{for} \quad 1 \leqq k \leqq n - 1.$$

Thus by Fact 1 our proof will be complete when we have shown that we can choose $\phi_1, \phi_2, \cdots, \phi_n$ as real numbers such that $c_1, c_2, \cdots, c_n$ are all positive.

To show that we can do this, let $1 \leqq k \leqq n$. For $\phi_k = \frac{1}{2}\pi$, $p_k = ir_1 r_2 \cdots r_k$ is positive imaginary, so

$$\text{Im } c_k = \text{Im } p_k + \text{Im } R_k = |p_k| + \text{Im } R_k \geqq |p_k| - |R_k| > \tfrac{1}{2}|p_k|,$$

and likewise, for $\phi_k = -\frac{1}{2}\pi$, $\text{Im } c_k < -\frac{1}{2}|p_k|$. Thus we have a continuous mapping

$$(\phi_1, \phi_2, \cdots, \phi_n) \to (\text{Im } c_1, \text{Im } c_2, \cdots, \text{Im } c_n)$$

from the rectangular n-parallelotope (which here is actually an n-cube)

$$-\tfrac{1}{2}\pi \leqq \phi_k \leqq \tfrac{1}{2}\pi, \qquad k = 1, 2, \cdots, n,$$

into real n-space, and this mapping satisfies the hypotheses of Fact 2 relative to the rectangular n-parallelotope

$$|\text{Im } c_k| \leqq \tfrac{1}{2}|p_k| = \tfrac{1}{2} r_1 r_2 \cdots r_k, \qquad k = 1, 2, \cdots, n.$$

Thus the range of this mapping includes the latter parallelotope and in particular contains the origin. Therefore we can choose $\phi_1, \cdots, \phi_n$ all in the interval $[-\frac{1}{2}\pi, \frac{1}{2}\pi]$ so as to yield

$$(\text{Im } c_1, \cdots, \text{Im } c_n) = (0, \cdots, 0)$$

and hence for this choice of $\phi_1, \cdots, \phi_n$ we have $c_1, \cdots, c_n$ all real. It is evident geometrically (and routine to show analytically) that c_k cannot be $\leqq 0$ for $-\frac{1}{2}\pi \leqq \phi_k \leqq \frac{1}{2}\pi$, so in fact $c_1, \cdots, c_n$ are all posi-

tive for the above choice of $\phi_1, \cdots, \phi_n$. This completes the proof of Theorem 2.

ACKNOWLEDGMENT. The author wishes to thank his colleague Dr. D. H. Carlson and the referee for helpful suggestions concerning this paper, and particularly wishes to thank the referee for the reference [1] below.

REFERENCES

1. K. Fan, *Topological proofs for certain theorems on matrices with non-negative elements*, Monatsh. Math. **62** (1958), 219–237. MR **20** #2354.

2. M. E. Fisher and A. T. Fuller, *On the stabilization of matrices and the convergence of linear iterative processes*, Proc. Cambridge Philos. Soc. **54** (1958), 417–425. MR **20** #2086.

OREGON STATE UNIVERSITY, CORVALLIS, OREGON 97331