LATENT ROOTS OF TRI-DIAGONAL MATRICES

by F. M. ARSCOTT

A considerable amount is known about the latent roots of matrices of the form

$$L_n \equiv \begin{bmatrix} b_1 & c_1 & & & \\ a_2 & b_2 & c_2 & & \\ & a_3 & b_3 & c_3 & \\ & & \cdots\cdots\cdots\cdots & & \\ & & & a_n & b_n \end{bmatrix},$$

in the case when each cross-product of non-diagonal elements, $a_i c_{i-1}$, is positive. One forms the sequence of polynomials $f_r(\lambda) = |L_r - \lambda I|$ for $r = 1$, $2, \ldots n$, and observes that

$$f_{r+1}(\lambda) = (b_{r+1} - \lambda) f_r(\lambda) - a_r c_{r-1} f_{r-1}(\lambda);$$

then it is easy to deduce that (i) the zeros of $f_n(\lambda)$ and $f_{n-1}(\lambda)$ interlace—that is, between two consecutive zeros of either polynomial lies precisely one zero of the other (ii) at the zeros of $f_n(\lambda)$ the values of $f_{n-1}(\lambda)$ are alternately positive and negative, (iii) all the zeros of $f_n(\lambda)$—i.e. all the latent roots of L_n—are real and different.

For the corresponding matrix in which the cross-products of non-diagonal elements are negative, nothing can be said about the real nature of the roots, but it is possible to give some bounds for the real roots. There are also parallel theorems to (ii) and (iii) in a particular case—namely, when the diagonal elements are in increasing order of magnitude.

Theorem 1. *Let M_n denote the matrix*

$$\begin{bmatrix} b_1 & -c_1 & & & \\ a_2 & b_2 & -c_2 & & \\ & a_3 & b_3 & -c_3 & \\ & & \cdots\cdots\cdots\cdots & & \\ & & & a_n & b_n \end{bmatrix}$$

(all other elements being zero) in which all the a, b, c are real, and each product $a_i c_{i-1}$ is positive. Then all the real latent roots of M_n lie between the least and the greatest of the b_i.

Proof. We first establish a lemma. Consider the matrix M_n^* obtained by replacing the diagonal elements in M_n by elements b_i^* which are all positive; we shall show that the real latent roots of this are all positive.

We observe that, because the b_i^* and all the products $a_i c_{i-1}$ are all positive,

every second-order principal minor in M_n^* is positive. Every first-order principal minor is, of course, simply a b^* and hence positive also. Now consider a principal minor of order r; expanding by its first row and column expresses it as the sum of positive multiples of two principal minors of orders $r-1$, $r-2$. Hence, inductively, all the principal minors of M_n^* are positive (including $\left| M_n^* \right|$ itself).

Now consider the equation $f_n^*(\lambda) \equiv \left| M_n^* - \lambda I \right| = 0$. Expanding the determinant according to its diagonal elements, we obtain

$$f_n^*(\lambda) = \sum_r (-)^{n+r} d_r \lambda^{n-r}$$

where all the d_r, being sums of principal minors of M_n^*, are positive. It is now obvious that no negative or zero value of λ can make $f_n^*(\lambda)$ vanish, and the lemma is proved.

Now we return to the matrix M_n and let b_l be the least of the diagonal elements. The equation $\left| M_n - \lambda I \right| = 0$ may be written

$$\left| M_n - b_l I - \lambda' I \right| = 0, \text{ where } \lambda' = \lambda - b_l, \dots\dots\dots\dots\dots\dots(1)$$

and since the matrix $(M_n - b_l I)$ has the form of the matrix M_n^* considered in the lemma, all the real λ' satisfying (1) are positive—that is, all the real λ are greater than b_l.

Finally, let b_g be the greatest of the diagonal elements. The equation $\left| M_n - \lambda I \right| = 0$ may be written

$$\left| b_g I - M_n - \lambda'' I \right| = 0, \text{ where } \lambda'' = b_g - \lambda, \dots\dots\dots\dots\dots(2)$$

and the matrix $(b_g I - M_n)$ has again the form of the matrix M_n^* (the signs of the non-diagonal elements are reversed but the relevant products are unaltered). Thus every real λ'' satisfying (2) is positive—that is, all the λ are less than b_g. All the λ satisfying $\left| M_n - \lambda I \right| = 0$, therefore, lie between b_l and b_g, these values excluded.

Theorem 2. *Let M_n denote the matrix above, with the additional restriction imposed that $b_1 < b_2 \dots < b_n$. Then the latent roots of M_n and the latent roots of M_{n-1} cannot interlace—in fact, between any two adjacent real latent roots of M_n must lie, if any, an even number of real latent roots of M_{n-1}. If $f_n(\lambda)$ denote $\left| M_n - \lambda I \right|$, then at the zeros of $f_n(\lambda)$, the sign of $f_{n-1}(\lambda)$ is $(-)^{n-1}$ always.*

Proof. Considering again the sequence of polynomials $f_r(\lambda) \equiv \left| M_r - \lambda I \right|$ ($r = 1, 2, \dots n$), we have

$$f_1(\lambda) = b_1 - \lambda \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(3a)$$

$$f_2(\lambda) = (b_1 - \lambda)(b_2 - \lambda) + a_2 c_1 \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(3b)$$

$$f_r(\lambda) = (b_r - \lambda) f_{r-1}(\lambda) + a_r c_{r-1} f_{r-2}(\lambda), \ r = 2, 3, \dots n, \quad \dots\dots\dots(3c)$$

the last relation being established by expanding $\left| M_r - \lambda I \right|$ according to its last row and column.

For large negative λ, all the $f_r(\lambda)$ are positive; for large positive λ, $f_r(\lambda)$ has the sign $(-)^r$; for brevity, we call this the " proper " sign of $f_r(\lambda)$ and $(-)^{r+1}$ the " reversed " sign.

From Theorem 1 it follows at once that for $\lambda \geq b_r$, $f_r(\lambda)$ has its proper sign (otherwise it would have a zero in (b_r, ∞)). Let us suppose that, if possible, there is a value λ_0 of λ such that $f_n(\lambda_0) = 0$ and $f_{n-1}(\lambda_0)$ has its reversed sign. λ_0 must lie in (b_1, b_n), and since $b_1 < b_2 \ldots < b_n$, there will be some s such that λ_0 lies in the interval $b_{s-1} \leq \lambda_0 < b_s$. Then relation (3c) with $r = n$ shows that at $\lambda = \lambda_0$, $f_{n-2}(\lambda_0)$ also has its reversed sign; applying (3c) successively with $r = n-1, n-2, \ldots, s$ shows that all the $f_r(\lambda)$ from $r = n-1$ down to and including $r = s-2$ have reversed signs at $\lambda = \lambda_0$. This, however, is impossible since $f_{s-1}(\lambda)$ and $f_{s-2}(\lambda)$ certainly have their proper signs when $\lambda \geq b_{s-1}$, so the supposition is false and there can be no value of λ where $f_n(\lambda) = 0$ and $f_{n-1}(\lambda)$ has reversed sign.

It is easy to see, moreover, that if $f_n(\lambda_0) = 0$, then $f_{n-1}(\lambda_0) \neq 0$; for if this were not so, then (from (3)) $f_{n-2}(\lambda_0) = f_{n-3}(\lambda_0) = \ldots = f_1(\lambda_0) = 0$ also, hence $\lambda_0 = b_1$, which contradicts Theorem 1. At every real zero of $f_n(\lambda)$, therefore, $f_{n-1}(\lambda)$ must have its proper sign, namely $(-)^{n-1}$.

The property of the latent roots of M_n and M_{n-1} follows immediately, and the theorem is proved.

I am grateful to a referee for some comments which shortened the proof of Theorem 1.

BATTERSEA COLLEGE OF TECHNOLOGY
LONDON, S.W.11