OSTROWSKI, A. M.
Math. Annalen 150, 276—284 (1963)

On Positive Matrices*

To B. L. VAN DER WAERDEN for his 60-th anniversary

By

A. M. OSTROWSKI in Basel

Introduction

1. We give in this article new proofs of some Theorems of PERRON[1]) on positive matrices. Our proofs are certainly not the most elementary ones, but may be for a certain level of the mathematical background the simplest and the most instructive.

A matrix $P = (p_{\mu\nu})$ $(\mu, \nu = 1, 2, \ldots, n)$ is called *positive* if all $p_{\mu\nu}$ are positive. We have then

(1) $$M = \operatorname*{Max}_{\mu,\nu} p_{\mu\nu} \geqq p_{\mu\nu} \geqq \operatorname*{Min}_{\mu,\nu} p_{\mu\nu} = m > 0 .$$

We denote further for any $(n \times n)$-matrix A, $A = (a_{\mu\nu})$, by λ_A the maximal modulus of the characteristic roots of A.

Then the Theorems we are going to prove are:

I. *For a positive matrix P we have always*

(2) $$\lambda_P > 0 .$$

II. *If we have for an $(n \times n)$-matrix A and the positive $(n \times n)$-matrix P, $|a_{\mu\nu}| \leqq p_{\mu\nu}$ $(\mu, \nu = 1, \ldots, n)$, then*

(3) $$\lambda_A \leqq \lambda_P .$$

(This is due to G. FROBENIUS, Sitz.-Ber. Akad. Berlin 1909, pp. 515–516).

III. *For a positive matrix P every characteristic root λ with $|\lambda| = \lambda_P$ is equal to λ_P, there is only one independent characteristic vector of A corresponding to λ_P and this vector can be normed so that all its components are positive.*

IV. *For a positive matrix P, λ_P is a simple root of the fundamental equation of P, $|\lambda I - P| = 0$.*

3. The essential tool of our proof of I—III will be the following

Lemma 1. *If A is a square matrix, then necessary and sufficient for*

(4) $$\lim_{\nu \to \infty} A^\nu = 0 ,$$

is that $\lambda_A < 1$.

This Lemma is due to R. OLDENBURGER [Duke Math. J. 6, 357–361 (1940)]. We give in the sections 4—5 a new and very simple proof of it.

---

* This investigation was carried out under the contract DA-91-591-EUC-2150 with the US Army.

[1]) PERRON, O.: 1) Zur Theorie der Matrices. Math. Ann. **64**, 248—263 (1907); 2) Grundlagen für eine Theorie des Jacobischen Kettenbruch-Algorithmus. Math. Ann. **64**, 1—76 (1907).

Using this Lemma we prove very easily the Theorems I–III. As to the Theorem IV, it is contained in the following new Theorem:

V. *If* $P = (p_{\mu\nu})$ *is a positive matrix and if the notations* (1) *hold, then* P *has* $n - 1$ *characteristic roots* λ *for which*

$$|\lambda| \leqq \lambda_P \frac{M^2 - m^2}{M^2 + m^2}. \tag{5}$$ [2]

This Theorem is again a Corollary from another Theorem, VI, which gives, instead of (5), an in general closer bound for $|\lambda|$.

Then we give, in section 19, another proof of IV, deriving it from a general criterion for simplicity of characteristic roots due to I. Schur, who used it in the case of Integral Equations and proved it by means of the theory of Elementary Divisors[3]. This is the Theorem

VII. *Necessary and sufficient in order that a characteristic root* λ_0 *of the matrix* A *is a simple root of the fundamental equation of* A, *is that* A *and* A' *have each only one independent characteristic vector corresponding to* λ_0, ξ' *and* η', *and that* ξ' *and* η *are not orthogonal,* $\xi'\eta \neq 0$.

In the sections 20–21 we give a direct proof of VII independent of the theory of Elementary Divisors.

§ 1. Theorems I–III

4. *Proof of the Lemma 1.* For an $(n \times n)$-matrix A put

$$(I - \sigma A)^{-1} = B(\sigma, a_{\mu\nu});$$

$B(\sigma, a_{\mu\nu})$ is here an $(n \times n)$-matrix with the elements

$$b_{ik} = \frac{\beta_{ik}(\sigma, a_{\mu\nu})}{|I - \sigma A|},$$

where the β_{ik} are polynomials in σ and the denominator is the determinant of $I - \sigma A$.

The b_{ik} are rational functions of σ, whose poles are reciprocals of some characteristic roots of A, and they are therefore certainly developpable in powers of σ for $\sigma < \frac{1}{\lambda_A}$. We have thence

$$(I - \sigma A)^{-1} = \sum_{\nu=0}^{\infty} \sigma^\nu A_\nu$$

where the A_ν are constant $(n \times n)$-matrices. Multiplying on both sides with $I - \sigma A$ we verify that $A_\nu = A^\nu$ and obtain finally

$$(I - \sigma A)^{-1} = \sum_{\nu=0}^{\infty} \sigma^\nu A^\nu \quad \left(|\sigma| < \frac{1}{\lambda_A}\right). \tag{6}$$

5. If now $\lambda_A < 1$, (6) is convergent for $\sigma = 1$ and we see that $A^\nu \to 0$.

---

[2] E. Hopf, to whom I communicated this result, succeeded, by an entirely different method, to replace the factor $\frac{M^2 - m^2}{M^2 + m^2}$ in (5) by the factor $\frac{M - m}{M + m}$ and to prove that this factor is the best.

[3] Jentzsch, R.: Über Integralgleichungen mit positivem Kern. J. f. Math. **141**, 235–244 (1912).

If, on the other hand, we have $A^\nu \to 0$, denote by λ_0 a characteristic root of A with the modulus λ_A and by V a (non-zero) characteristic vector of A, belonging to λ_0.

Then we have

$$A\,V = \lambda_0 V\,,$$

and, applying repeatedly A on both sides of this equation,

$$A^\nu V = \lambda_0^\nu V\,.$$

But here, with $\nu \to \infty$, the left hand vector tends to 0 and it follows $\lambda_0^\nu \to 0$, $\lambda_A^\nu \to 0$, $\lambda_A < 1$. The Lemma 1 is proved.

6. *Proof of* I. Assume that for the matrix $P = (p_{\mu\nu})$ with (1) we have $\lambda_P = 0$. Then all characteristic roots of P are $= 0$. The same is then true for the matrix cP for any constant c. We have therefore

$$\lambda_{cP} = 0,\ (cP)^\nu \to 0\ \ (\nu \to \infty)\,.$$

Nut if we take $c = \frac{1}{m}$, all elements of cP are $\geqq 1$, the same is then true for all $\left(\frac{1}{m} P\right)^\nu (\nu = 1, 2, \ldots)$ and $\left(\frac{1}{m} P\right)^\nu$ cannot tend to 0.

7. *Proof of* II. Assume that (3) is wrong and that we have, under the conditions of II, $\lambda_A > \lambda_P$. Choosing c so that

$$\lambda_A > \frac{1}{c} > \lambda_P\,, \tag{7}$$

we see that for the matrix cP, $\lambda_{cP} = c\lambda_P < 1$ and therefore

$$\lim_{\nu \to \infty} (cP)^\nu = 0\,.$$

But since we have $c\,|a_{\mu\nu}| \leqq c\,p_{\mu\nu}$ we have then also

$$\lim_{\nu \to \infty} (cA)^\nu = 0\,,$$

and by the Lemma 1

$$\lambda_{cA} = c\lambda_A < 1$$

in contradiction to (7). II is proved.

8. *Proof of* III. Let λ be a characteristic root of P with $|\lambda| = \lambda_P$ and $V = (p_1, \ldots, p_n)'$ a corresponding characteristic vector, normed in such a way that one of the components, say p_k, has the value 1. Put

$$V_0 = (|p_1|, \ldots, |p_n|)'\,.$$

We have, if not all p_ν are real and $\geqq 0$,

$$\sum_{\nu=1}^{n} p_{\mu\nu}|p_\nu| > \left|\sum_{\nu=1}^{n} p_{\mu\nu}p_\nu\right| \qquad (\mu = 1, \ldots, n)\,,$$

and there exists a $\delta > 0$ such that[4])

$$\sum_{\nu=1}^{n} p_{\mu\nu}|p_\nu| > (1+\delta)\left|\sum_{\nu=1}^{n} p_{\mu\nu}p_\nu\right| \qquad (\mu = 1, \ldots, n)\,. \tag{8}$$

---

[4]) Cf. for this argument G. FROBENIUS l.c., p. 515.

Put

$$\frac{1}{(1+\delta)\lambda_P} P = Q, \quad \lambda_Q = \frac{1}{1+\delta} < 1 . \tag{9}$$

Then we have from $PV = \lambda V$:

$$\left|\sum_{\nu=1}^{n} p_{\mu\nu} p_\nu\right| = \lambda_P |p_\mu| \qquad (\mu = 1, \ldots, n), \tag{10}$$

and thence by (8)

$$\sum_{\nu=1}^{n} p_{\mu\nu} |p_\nu| > (1+\delta)\lambda_P |p_\mu| \qquad (\mu = 1, \ldots, n).$$

We have therefore for V_0 the majorization

$$V_0 \ll Q V_0$$

and iterating ν times,

$$V_0 \ll Q^\nu V_0 ,$$

and this is impossible since by (9), $Q^\nu \to 0$.

We have therefore $p_\nu \geqq 0$ $(\nu = 1, \ldots, n)$ and λ is by (10) positive $= \lambda_P$. Further, if a p_μ were $= 0$, we would have from (10)

$$0 = \sum_{\nu=1}^{n} p_{\mu\nu} p_\nu \geqq p_{\mu k} > 0 .$$

We see that all p_ν are positive. If there were now two independent positive characteristic vectors corresponding to λ_P, we could find a linear combination of these two vectors in which not all components would be of the same sign, in contradiction with what has been already proved. III is now completely proved.

§ 2. Theorems IV-VI

9. In our proofs of V and VI we will use the following

Lemma 2. *If $V = (p_1, \ldots, p_n)'$ is a characteristic vector of P with positive components, belonging to the characteristic value λ_P, we have in notations* (1)

$$\frac{p_\mu}{p_\nu} \leqq \frac{M}{m} \; (\mu, \nu = 1, 2, \ldots, n)^{5)} . \tag{11}$$

This is proved, norming the p_ν by

$$p_1 + p_2 + \cdots + p_n = 1 . \tag{12}$$

Indeed, in the relation (10) the sum $\sum_{\nu=1}^{n} p_{\mu\nu} p_\nu$ lies by (1) and (12) between m and M. We have therefore

$$m \leqq \lambda_P p_\mu \leqq M \qquad (\mu = 1, \ldots, n)$$

and therefore

$$\frac{p_\mu}{p_\nu} = \frac{\lambda_P p_\mu}{\lambda_P p_\nu} \leqq \frac{M}{m} .$$

---

[5]) Cf. A. M. Ostrowski: On the Eigenvector belonging to the Maximal Root of a Non-negative Matrix. Proc. Edinb. Math. Soc. (II) Vol. 12, Part 2, pp. 107—112 (1960), Formula (10).

10. Further, we will need the following refinement of the triangle inequality:

Lemma 3. *Assume between n real or complex numbers x_ν the relation*

$$\sum_{\nu=1}^{n} p_\nu x_\nu = 0 \tag{13}$$

where the p_ν are positive and satisfy the inequalities

$$\frac{p_\mu}{p_\nu} \leqq q \qquad (\mu, \nu = 1, \ldots, n)\,; \tag{14}$$

then we have

$$\left|\sum_\nu x_\nu\right| \leqq \sigma \sum_\nu |x_\nu|, \quad \sigma = \frac{q-1}{q+1}\,. \tag{15}$$

11. *Proof.* Without loss of generality we can assume that instead of (14) we have

$$1 \leqq p_\nu \leqq q \qquad (\nu = 1, \ldots, n)\,. \tag{14⁰}$$

We consider first the case that all x_ν are real and not all vanish. Then we can assume, without loss of generality, that

$$x_1 \geqq \cdots \geqq x_m > 0 = x_{m+1} = \cdots = x_k > x_{k+1} \geqq \cdots \geqq x_n\,, \tag{16}$$

where, if no x_ν vanishes, we have $k = m$.

Put $|x_\nu| = \xi_\nu$ $(\nu = 1, \ldots, n)$,

$$u = \sum_{\nu=1}^{m} \xi_\nu \quad v = \sum_{\nu=k+1}^{n} \xi_\nu \tag{17}$$

and, using (13),

$$W = \sum_{\nu=1}^{m} p_\nu \xi_\nu = \sum_{\nu=k+1}^{n} p_\nu \xi_\nu\,. \tag{18}$$

If $u = v$ the left hand expression in (15) is 0. Assume that we have $u > v$, then from (14⁰), (17) and (18) we have

$$u \leqq W \leqq qv\,,$$

$$\frac{u}{v} \leqq q\,. \tag{19}$$

But then it follows

$$\frac{u-v}{u+v} = \frac{\dfrac{u}{v}-1}{\dfrac{u}{v}+1} \leqq \frac{q-1}{q+1} = \sigma$$

and this is the assertion (15).

If $u < v$, we have only to interchange u and v in the above argument. This proves (15) in the case of real x_ν.

12. We consider now the case where the x_ν are complex,

$$x_\nu = a_\nu + i b_\nu \qquad (\nu = 1, 2, \ldots, n)\,, \tag{20}$$

and put

$$\sum_\nu a_\nu = A\,, \quad \sum_\nu b_\nu = B\,, \tag{21}$$

$$z_\nu = |a_\nu| + i|b_\nu| \quad (\nu = 1, 2, \ldots, n), \quad \sum_{\nu=1}^{n} z_\nu = Z\,. \tag{22}$$

Then we have, since the a_ν and the b_ν satisfy separatly (13),

$$|A| \leqq \sigma \sum_{\nu=1}^{n} |a_\nu| = \sigma \Re Z$$

$$|B| \leqq \sigma \sum_{\nu=1}^{n} |b_\nu| = \sigma \mathscr{I} Z\,,$$

and it follows

$$\left|\sum_\nu x_\nu\right| = |A + iB| \leqq \sigma |Z| \leqq \sigma \sum_\nu |z_\nu| = \sigma \sum_\nu |x_\nu|$$

which proves the Lemma 3 in the general case.

13. Observe that the equality-sign in (15) is possible for any value of $q \geqq 1$. Indeed, we have the equality-sign, if we put

$$x_1 = 1,\, x_2 = -q,\, x_3 = \cdots = x_n = 0\,.$$

The following equivalent formulation, which contains $2n$ positive parameters, is useful:

Corollary to the Lemma 3. *Assume that we have $2n$ positive parameters $u_1, \ldots, u_n;\, v_1, \ldots, v_n$, satisfying the inequalities*

$$(23) \qquad \frac{u_\mu}{u_\nu} \leqq r,\, \frac{v_\mu}{v_\nu} \leqq s \qquad (\mu, \nu = 1, \ldots, n)\,.$$

Then for n real or complex numbers x_ν ($\nu = 1, 2, \ldots, n$) satisfying the relation

$$(24) \qquad \sum_{\nu=1}^{n} u_\nu x_\nu = 0$$

we have

$$(25) \qquad \left|\sum_{\nu=1}^{n} v_\nu x_\nu\right| \leqq \sigma \sum_\nu v_\nu |x_\nu|\,, \quad \sigma = \frac{rs-1}{rs+1}\,.$$

Indeed, putting

$$y_\nu = v_\nu x_\nu\,, \quad p_\nu = \frac{u_\nu}{v_\nu}\,,$$

we see that the y_ν satisfy the conditions of the Lemma 3 if q is replaced by rs.

14. We define the *relative oscillation* of the μ-th row of P, $\omega_P^{(\mu)}$, by

$$(26) \qquad \omega_P^{(\mu)} = \operatorname{Max} \frac{p_{\mu\nu_1}}{p_{\mu\nu_2}} \qquad (\nu_1, \nu_2 = 1, \ldots, n)$$

and call then

$$(27) \qquad \omega_P = \operatorname*{Max}_\mu \omega_P^{(\mu)}$$

the *relative oscillation in the rows of P*. Obviously, we have

$$(28) \qquad \omega_P \leqq \frac{M}{m}\,.$$

15. We will now prove the Theorem

VI. *If $P = (p_{\mu\nu})$ is a positive matrix and the notations* (1) *and* (27) *hold, then P has $n-1$ characteristic roots λ for which*

$$(29) \qquad |\lambda| \leqq \sigma \lambda_P\,, \quad \sigma = \frac{M\omega_P - m}{M\omega_P + m}\,.$$ [6])

[6]) A slight modification of our discussion allows to replace the value of σ in (29) by $\frac{\omega_P^2 - 1}{\omega_P^2 + 1}$. We do not insist however upon this improvement, since in the meantime E. Hopf succeeded to replace the value of σ in (29) by $\frac{\omega_P - 1}{\omega_P + 1}$.

Obviously, by (28) the inequality (5) of V follows from (29).

16. *Proof of* VI. Assume that λ is a characteristic root of P, different from λ_P, and let $X = (x_1, \ldots, x_n)'$ be a characteristic vector of P corresponding to λ. Take further a characteristic vector with positive components corresponding to λ_P for the *transpose matrix* P', $(u_1, \ldots, u_n)'$; then we have the relation

$$\sum_{\nu=1}^{n} u_\nu x_\nu = 0 ,$$

where the u_ν satisfy the condition (23) with $r = \frac{M}{m}$.

We have therefore, by the inequality (25) and the definition (26) of $\omega_P^{(\mu)}$, for $\mu = 1, 2, \ldots, n$:

$$|\lambda| \, |x_\mu| = \left| \sum_{\nu=1}^{n} p_{\mu\nu} x_\nu \right| \leqq \sigma_\mu \sum_{\nu=1}^{n} p_{\mu\nu} |x_\nu| , \quad \sigma_\mu = \frac{M \omega_P^{(\mu)} - m}{M \omega_P^{(\mu)} + m} .$$

Indeed, the second inequality (23) is here satisfied with $s = \omega_P^{(\mu)}$. As, by (27), we have, for σ from (29), $\sigma_\mu \leqq \sigma$ $(\mu = 1, \ldots, n)$, we obtain finally

$$|\lambda| \, |x_\mu| \leqq \sigma \sum_{\nu=1}^{n} p_{\mu\nu} |x_\nu| \qquad (\mu = 1, \ldots, n) . \tag{30}$$

17. Introducing the vector

$$T = (|x_1|, \ldots, |x_n|)'$$

and the matrix

$$Q = \frac{\sigma}{|\lambda|} P , \quad \lambda_Q = \sigma \frac{\lambda_P}{|\lambda|} , \tag{31}$$

we obtain from (30) the majorization

$$T \ll QT ,$$

and iterating,

$$T \ll Q^\nu T \qquad (\nu = 1, 2, \ldots) .$$

Therefore, we cannot have $Q^\nu \to 0$ and, by the Lemma 1, $\lambda_Q \geqq 1$ and by (31)

$$|\lambda| \leqq \sigma \lambda_P , \tag{32}$$

which is the relation (29), now proved for any characteristic root λ of P, *different* from λ_P. And the same holds, of course, for the relation (5).

18. In order to prove completely the Theorem VII, it is now sufficient to prove that λ_P is a *simple root* of the characteristic equation of P.

Assume that λ_P is a multiple root. If we replace then the p by $p_{\mu\nu} + \delta_{\mu\nu}$, the discriminant of the characteristic equation of the new matrix is a non identically vanishing polynomial of the n^2 variables $\delta_{\mu\nu}$.

Indeed, by an affine transformation of the $\delta_{\mu\nu}$ this discriminant becomes the discriminant of the characteristic equation of the matrix $\Delta = (\delta_{\mu\nu})$, and this characteristic equation can be made an arbitrarily prescribed equation of degree n, choosing appropriately the $\delta_{\mu\nu}$.

If we now choose in the matrix $(p_{\mu\nu} + \delta_{\mu\nu})$ the increments $\delta_{\mu\nu}$ appropriately as arbitrarily small positive quantities we obtain a sequence of matrices P_ν $(\nu = 1, 2, \ldots)$ which tend to P and have no multiple roots. For the matrices P_ν

the quantities corresponding to λ_P, M, m have then respectively λ_P, M, m as their limits. Applying to these matrices the inequality (5), and denoting by D a positive number with

$$\lambda_P \frac{M^2 - m^2}{M^2 + m^2} < D < \lambda_P ,$$

we see that, from a certain ν on, each P_ν has $n - 1$ of its characteristic roots in the circle $|\lambda| < D$.

It follows then that P has $n - 1$ of its characteristic roots in the circle $|\lambda| \leq D$, so that λ_P is certainly a simple characteristic root of P. The Theorems IV, V and VI are now proved.

§ 3. Theorem VII

19. If follows from the Theorem III that both P and P' have for λ_P each only one independent characteristic vector and that these vectors can be assumed with positive components so that their inner product certainly does not vanish. But then the conditions of Schur's criterion VII are satisfied and we see directly that λ_P is a simple root of the characteristic equation of P. This is another proof of the Theorem IV.

20. We have now to prove the Theorem VII. Assume that we have under the conditions of VII,

$$A\,\xi' = \lambda_0 \xi', \ \eta A = \lambda_0 \eta, \ \eta\,\xi' = 1 \tag{33}$$

where ξ', η' are the characteristic vectors of A and A' and the product can obviously be assumed as $= 1$. Denote by ε the vector $(1, 0, \ldots, 0)$ and form an $(n \times n)$-matrix S, such that

$$|S| \neq 0, \quad S\,\xi' = \varepsilon' .$$

If we replace A by $SAS^{-1} = B$ and ξ by ε, η by $\eta S^{-1} = \eta_1$, the fundamental equation of A is the same as that of B, and the hypotheses of VII are verified for B, since we have

$$B\varepsilon' = SAS^{-1}S\xi' = SA\xi' = \lambda_0\varepsilon' , \qquad \eta_1 B = \eta S^{-1}SAS^{-1} = \eta_1\lambda_0 , \qquad \eta_1\varepsilon' = 1 .$$

We will therefore assume from the beginning that $\xi = \varepsilon$. But then it follows from $A\varepsilon' = \lambda_0\varepsilon'$, putting $A = (a_{\mu\nu})$ that $a_{11} = \lambda_0$, $a_{\mu 1} = 0$ $(\mu = 2, \ldots, n)$, so that A can be decomposed in the form

$$A = \begin{pmatrix} \lambda_0 & L \\ 0 & A_1 \end{pmatrix} , \tag{34}$$

where A_1 is a square matrix of the order $n - 1$ and L a row vector with $n - 1$ components. And we have now to prove that λ_0 cannot be a characteristic root of A_1.

21. In this proof we will sometimes write an n-dimensional vector $\alpha = (a_1, a_2, \ldots, a_n)$ in the form

$$\alpha = a_1 \dot{+} \alpha_1 , \quad \alpha_1 = (a_2, \ldots, a_n) ,$$

where α_1 is the $(n - 1)$-dimensional vector formed by the $n - 1$ last components of α.

Write now $\eta = (y_1, y_2, \ldots, y_n)$ in the form

$$\eta = y_1 \dot{+} \eta_1, \quad \eta_1 = (y_2, \ldots, y_n) .$$

Then we have from $\eta \varepsilon' = 1 : y_1 = 1$, and from (33) and (34)

$$(35) \qquad \lambda_0 \eta = \eta A = (1 \dot{+} \eta_1) A = \lambda_0 \dot{+} (L + \eta_1 A_1), \lambda_0 \eta_1 = L + \eta_1 A_1 .$$

Suppose now that there exists a characteristic vector of dimension $n - 1$, $\zeta_1 = (z_2, \ldots, z_n)$ of A_1 corresponding to λ_0, so that we have $A_1 \zeta_1' = \lambda_0 \zeta_1'$. Then it follows from (35) multiplying it from the right by ζ_1':

$$(36) \qquad \lambda_0 \eta_1 \zeta_1' = L \zeta_1' + \eta_1 A_1 \zeta_1' = L \zeta_1' + \eta_1 \lambda_0 \zeta_1', L \zeta_1' = 0 .$$

If we form now the n-dimensional vector

$$\zeta = 0 \dot{+} \zeta_1 = (0, z_2, \ldots, z_n)$$

we have from (34) by (36)

$$A \zeta' = L \zeta_1' \dot{+} A_1 \zeta_1' = L \zeta_1' \dot{+} \lambda_0 \zeta_1' = \lambda_0 \zeta'$$

so that ζ' is a characteristic vector of A corresponding to λ_0. But ζ is obviously independent of ε contrary to the hypotheses of VII. VII is proved.

(Received January 21, 1963)