Chapter 8
Nonnegative Matrices

In this chapter matrices have real entries in general. In a few specified cases, entries might be complex.

8.1 Nonnegative Vectors and Matrices

Definition 8.1 *A vector $x \in \mathbb{R}^n$ is* nonnegative*, and we write $x \geq 0$, if its coordinates are nonnegative. It is* positive*, and we write $x > 0$, if its coordinates are (strictly) positive. Furthermore, a matrix $A \in M_{n\times m}(\mathbb{R})$ (not necessarily square) is* nonnegative *(respectively,* positive*) if its entries are nonnegative (respectively, positive); we again write $A \geq 0$ (respectively, $A > 0$). More generally, we define an order relation $x \leq y$ whose meaning is $y - x \geq 0$.*

Definition 8.2 *Given $x \in \mathbb{C}^n$, we let $|x|$ denote the nonnegative vector whose coordinates are the numbers $|x_j|$. Likewise, if $A \in \mathbf{M}_n(\mathbb{C})$, the matrix $|A|$ has entries $|a_{ij}|$.*

Observe that given a matrix and a vector (or two matrices), the triangle inequality implies

$$|Ax| \leq |A| \cdot |x|.$$

Proposition 8.1 *A matrix is nonnegative if and only if $x \geq 0$ implies $Ax \geq 0$. It is positive if and only if $x \geq 0$ and $x \neq 0$ imply $Ax > 0$.*

Proof. Let us assume that $Ax \geq 0$ (respectively, > 0) for every $x \geq 0$ (respectively, ≥ 0 and $\neq 0$). Then the ith column $A^{(i)}$ is nonnegative (respectively, positive), since it is the image of the ith vector of the canonical basis. Hence $A \geq 0$ (respectively, > 0).

Conversely, $A \geq 0$ and $x \geq 0$ imply trivially $Ax \geq 0$. If $A > 0$, $x \geq 0$, and $x \neq 0$, there exists an index ℓ such that $x_\ell > 0$. Then

$$(Ax)_i = \sum_j a_{ij} x_j \geq a_{il} x_\ell > 0,$$

D. Serre, *Matrices,* Graduate Texts in Mathematics 216,
DOI 10.1007/978-1-4419-7683-3_8, © Springer Science+Business Media, LLC 2010

and hence $Ax > 0$. □

An important point is the following:

Proposition 8.2 *If $A \in \mathbf{M}_n(\mathbb{R})$ is nonnegative and irreducible, then $(I+A)^{n-1} > 0$.*

Proof. Let $x \neq 0$ be nonnegative, and define $x^m = (I+A)^m x$, which is nonnegative too. Let us denote by P_m the set of indices of the nonzero components of x^m: P_0 is nonempty. Because $x_i^{m+1} \geq x_i^m$, one has $P_m \subset P_{m+1}$. Let us assume that the cardinality $|P_m|$ of P_m is strictly less than n. There are thus one or more zero components, whose indices form a nonempty subset I, complement of P_m. Because A is irreducible, there exists some nonzero entry a_{ij}, with $i \in I$ and $j \in P_m$. Then $x_i^{m+1} \geq a_{ij} x_j^m > 0$, which shows that P_{m+1} is not equal to P_m, and thus $|P_{m+1}| > |P_m|$. By induction, we deduce that $|P_m| \geq \min\{m+1, n\}$. Hence $|P_{n-1}| = n$, meaning that $x^{n-1} > 0$. We conclude with Proposition 8.1. □

8.2 The Perron–Frobenius Theorem: Weak Form

The following result is not very impressive. We prove much more in the next section, with elementary calculus. It has, however, its own interest, as an elegant consequence of Brouwer's fixed point theorem.

Theorem 8.1 *Let $A \in \mathbf{M}_n(\mathbb{R})$ be a nonnegative matrix. Then $\rho(A)$ is an eigenvalue of A associated with a nonnegative eigenvector.*

Proof. Let λ be an eigenvalue of maximal modulus and v an eigenvector, normalized by $\|v\|_1 = 1$. Then

$$\rho(A)|v| = |\lambda v| = |Av| \leq A|v|.$$

Let us denote by C the subset of $\mathbb{R}^n$ (actually a subset of the unit simplex K_n) defined by the (in)equalities $\sum_i x_i = 1$, $x \geq 0$, and $Ax \geq \rho(A)x$. This is a closed convex set, nonempty, inasmuch as it contains $|v|$. Finally, it is bounded, because $x \in C$ implies $0 \leq x_j \leq 1$ for every j; thus it is compact. Let us distinguish two cases.

1. There exists $x \in C$ such that $Ax = 0$. Then $\rho(A)x \leq 0$ furnishes $\rho(A) = 0$. The theorem is thus proved in this case.
2. For every x in C, $Ax \neq 0$. Then let us define on C a continuous map f by

$$f(x) = \frac{1}{\|Ax\|_1} Ax.$$

It is clear that $f(x) \geq 0$ and that $\|f(x)\|_1 = 1$. Finally,

$$Af(x) = \frac{1}{\|Ax\|_1} AAx \geq \frac{1}{\|Ax\|_1} A\rho(A)x = \rho(A)f(x),$$

so that $f(C) \subset C$. Then Brouwer's theorem (see [3], p. 217) asserts that a continuous function from a compact convex subset of $\mathbb{R}^N$ into itself has a fixed point. Thus let y be a fixed point of f. It is a nonnegative eigenvector, associated with the eigenvalue $r = \|Ay\|_1$. Because $y \in C$, we have $ry = Ay \geq \rho(A)y$ and thus $r \geq \rho(A)$, which implies $r = \rho(A)$.
□

That proof can be adapted to the case where a real number r and a nonzero vector y are given satisfying $y \geq 0$ and $Ay \geq ry$. Just take for C the set of vectors x such that $\sum_i x_i = 1$, $x \geq 0$, and $Ax \geq rx$. We then conclude that $\rho(A) \geq r$.

8.3 The Perron–Frobenius Theorem: Strong Form

Theorem 8.2 *Let $A \in \mathbf{M}_n(\mathbb{R})$ be a nonnegative irreducible matrix. Then $\rho(A)$ is a simple eigenvalue of A, associated with a positive eigenvector. Moreover, $\rho(A) > 0$.*

8.3.1 Remarks

1. Alhough the Perron–Frobenius theorem says that $\rho(A)$ is a simple eigenvalue, it does not tell us anything about the other eigenvalues of maximal modulus. The following example shows that such other eigenvalues may exist:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

   The existence of several eigenvalues of maximal modulus is studied in Section 8.4.
2. One obtains another proof of the weak form of the Perron–Frobenius theorem by applying the strong form to $A + \alpha J$, where $J > 0$ and $\alpha > 0$, letting α tend to zero and using Theorem 5.2.
3. Without the irreducibility assumption, $\rho(A)$ may be a multiple eigenvalue, and a nonnegative eigenvector may not be positive. This holds for a matrix of size $n = 2m$ that reads blockwise

$$A = \begin{pmatrix} B & 0_m \\ I_m & B \end{pmatrix}.$$

   Here, $\rho(A) = \rho(B)$, and every eigenvalue has an even algebraic multiplicity, because $P_A = (P_B)^2$.
   Let us assume that B is nonnegative and irreducible. Then $\rho(B)$ is a simple eigenvalue of B, associated with the eigenvector $r > 0$. In addition, B^T is irreducible (Proposition 3.25) and thus has a positive eigenvector ℓ associated with $\rho(B)$.

Let
$$X=\begin{pmatrix} y \\ z \end{pmatrix}$$
belong to the kernel of $A-\rho(A)I_n$. We have
$$By=\rho(B)y, \qquad y+Bz=\rho(B)z.$$
The first equality tells us that $y=\alpha r$ for some $\alpha\in\mathbb{R}$. Multiplying the second equality by ℓ^T, we obtain $\ell^T y=0$; that is, $\alpha\ell^T r=0$. Because $\ell>0$ and $r>0$, this gives $\alpha=0$; that is, $y=0$. Then $Bz=\rho(B)z$, meaning that $z\parallel r$. Finally, the eigenspace is spanned by
$$X=\begin{pmatrix} 0_m \\ r \end{pmatrix},$$
which is nonnegative, but not positive.

4. As a matter of fact, not only the eigenvector associated with $\rho(A)$ is positive, but it is the only one to be positive. For let ℓ be the positive eigenvector of A^T associated with $\rho(A)$. Then every eigenvector x of A associated with an eigenvalue $\lambda\neq\rho(A)$ satisfies $\ell^T x=0$, which prevents x from being positive.

Proof. For $r\geq 0$, we denote by C_r the set of vectors of $\mathbb{R}^n$ defined by the conditions
$$x\geq 0, \quad \|x\|_1=1, \quad Ax\geq rx.$$
Each C_r is a convex compact set. We saw in the previous section that if λ is an eigenvalue associated with an eigenvector x of unit norm $\|x\|_1=1$, then $|x|\in C_{|\lambda|}$. In particular, $C_{\rho(A)}$ is nonempty. Conversely, if C_r is nonempty, then for $x\in C_r$,
$$r=r\|x\|_1\leq\|Ax\|_1\leq\|A\|_1\|x\|_1=\|A\|_1,$$
and therefore $r\leq\|A\|_1$. Furthermore, the map $r\mapsto C_r$ is nonincreasing with respect to inclusion, and is "left continuous" in the following sense. If $r>0$, one has
$$C_r=\bigcap_{s<r}C_s.$$
Let us then define
$$R=\sup\{r\,|\,C_r\neq\emptyset\},$$
so that $R\in[\rho(A),\|A\|_1]$. The monotonicity with respect to inclusion shows that $r<R$ implies $C_r\neq\emptyset$.

If $x>0$ and $\|x\|_1=1$, then $Ax>0$ because A is nonnegative and irreducible. Setting $r:=\min_j(Ax)_j/x_j>0$, we have $C_r\neq\emptyset$, whence $R\geq r>0$. The set C_R, being the intersection of a totally ordered family of nonempty compacts sets, is nonempty.

Let $x\in C_R$ be given. Lemma 10 below shows that x is an eigenvector of A associated with the eigenvalue R. We observe that this eigenvalue is not less than $\rho(A)$ and infer that $\rho(A) = R$. Hence $\rho(A)$ is an eigenvalue associated with the eigenvector x. Lemma 11 below ensures that $x > 0$ and $\rho(A) > 0$.

The proof of the simplicity of the eigenvalue $\rho(A)$ is given in Section 8.3.3.

8.3.2 A Few Lemmas

Lemma 10. *Let $r \geq 0$ and $x \geq 0$ such that $Ax \geq rx$ and $Ax \neq rx$. Then there exists $r' > r$ such that $C_{r'}$ is nonempty.*

Proof. Set $y := (I_n + A)^{n-1}x$. Because A is irreducible and $x \geq 0$ is nonzero, one has $y > 0$. Likewise, $Ay - ry = (I_n + A)^{n-1}(Ax - rx) > 0$. Let us define $r' := \min_j (Ay)_j / y_j$, which is strictly larger than r. We then have $Ay \geq r'y$, so that $C_{r'}$ contains the vector $y/\|y\|_1$. □

Lemma 11. *The nonnegative eigenvectors of A are positive. The corresponding eigenvalue is positive too.*

Proof. Given such a vector x with $Ax = \lambda x$, we observe that $\lambda \in \mathbb{R}^+$. Then

$$x = \frac{1}{(1+\lambda)^{n-1}}(I_n + A)^{n-1}x,$$

and the right-hand side is strictly positive, from Proposition 8.2.

Inasmuch as A is irreducible and nonnegative, we infer $Ax \neq 0$. Thus $\lambda \neq 0$; that is, $\lambda > 0$. □

Finally, we can state the following result.

Lemma 12. *Let $M, B \in \mathbf{M}_n(\mathbb{C})$ be matrices, with M irreducible and $|B| \leq M$. Then $\rho(B) \leq \rho(M)$.*

In the case of equality ($\rho(B) = \rho(M)$), the following hold.

- *$|B| = M$.*
- *For every eigenvector x of B associated with an eigenvalue of modulus $\rho(M)$, $|x|$ is an eigenvector of M associated with $\rho(M)$.*

Proof. In order to establish the inequality, we proceed as above. If λ is an eigenvalue of B, of modulus $\rho(B)$, and if x is a normalized eigenvector, then $\rho(B)|x| \leq |B| \cdot |x| \leq M|x|$, so that $C_{\rho(B)}$ is nonempty. Hence $\rho(B) \leq R = \rho(M)$.

Let us investigate the case of equality. If $\rho(B) = \rho(M)$, then $|x| \in C_{\rho(M)}$, and therefore $|x|$ is an eigenvector: $M|x| = \rho(M)|x| = \rho(B)|x| \leq |B| \cdot |x|$. Hence, $(M - |B|)|x| \leq 0$. Because $|x| > 0$ (from Lemma 11) and $M - |B| \geq 0$, this gives $|B| = M$. □

8.3.3 The Eigenvalue $\rho(\mathbf{A})$ *Is Simple*

Let $P_A(X)$ be the characteristic polynomial of A. It is given as the composition of an n-linear form (the determinant) with polynomial vector-valued functions (the columns of $XI_n - A$). If ϕ is p-linear and if $V_1(X),\dots,V_p(X)$ are polynomial vector-valued functions, then the derivative of the polynomial $P(X) := \phi(V_1(X),\dots,V_p(X))$ is given by

$$P'(X) = \phi(V_1',V_2,\dots,V_p) + \phi(V_1,V_2',\dots,V_p) + \cdots + \phi(V_1,\dots,V_{p-1},V_p').$$

One therefore has

$$P_A'(X) = \det(\mathbf{e}^1,a_2,\dots,a_n) + \det(a_1,\mathbf{e}^2,\dots,a_n) + \cdots + \det(a_1,\dots,a_{n-1},\mathbf{e}^n),$$

where a_j is the jth column of $XI_n - A$ and $\{\mathbf{e}^1,\dots,\mathbf{e}^n\}$ is the canonical basis of $\mathbb{R}^n$. Developing the jth determinant with respect to the jth column, one obtains

$$P_A'(X) = \sum_{j=1}^{n} P_{A_j}(X), \tag{8.1}$$

where $A_j \in \mathbf{M}_{n-1}(\mathbb{R})$ is obtained from A by deleting the jth row and the jth column. Let us now denote by $B_j \in \mathbf{M}_n(\mathbb{R})$ the matrix obtained from A by replacing the entries of the jth row and column by zeroes. This matrix is block-diagonal, the two diagonal blocks being $A_j \in \mathbf{M}_{n-1}(\mathbb{R})$ and $0 \in \mathbf{M}_1(\mathbb{R})$. Hence, the eigenvalues of B_j are those of A_j, together with zero, and therefore $\rho(B_j) = \rho(A_j)$. Furthermore, $|B_j| \le A$, but $|B_j| \ne A$ because A is irreducible and B_j is block-diagonal, hence reducible. It follows (Lemma 12) that $\rho(B_j) < \rho(A)$. Hence P_{A_j} does not vanish over $[\rho(A),+\infty)$. Because $P_{A_j}(t) \approx t^{n-1}$ at infinity, we deduce that $P_{A_j}(\rho(A)) > 0$. Finally, $P_A'(\rho(A))$ is positive and $\rho(A)$ is a simple root. □

8.4 Cyclic Matrices

The following statement completes Theorem 8.2.

Theorem 8.3 *Under the assumptions of Theorem* 8.2, *let p be the cardinality of the set* $\mathrm{Sp}_{max}(A)$ *of eigenvalues of A of maximal modulus* $\rho(A)$.

Then we have $\mathrm{Sp}_{max}(A) = \rho(A)\mathscr{U}_p$, *where* $\mathscr{U}_p$ *is the group of pth roots of unity. Every such eigenvalue is simple. The spectrum of A is invariant under multiplication by* $\mathscr{U}_p$. *Finally, A is conjugated via a permutation matrix to the following cyclic form. In this cyclic matrix each element is a block, and the diagonal blocks (which all vanish) are square with nonzero sizes:*

$$\begin{pmatrix} 0 & M_1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & & & \ddots & M_{p-1} \\ M_p & 0 & \cdots & \cdots & 0 \end{pmatrix}.$$

Remark

The converse is true. The characteristic polynomial of a cyclic matrix is

$$X \mapsto \det(X^p I_m - M_1 M_2 \cdots M_p),$$

up to a factor X^ν (with ν possibly negative). Its spectrum is thus stable under multiplication by $\exp(2i\pi/p)$.

Proof. Let us denote by X the unique nonnegative eigenvector of A normalized by $\|X\|_1 = 1$. If Y is a unitary eigenvector, associated with an eigenvalue μ of maximal modulus $\rho(A)$, the inequality $\rho(A)|Y| = |AY| \leq A|Y|$ implies (Lemma 12) $|Y| = X$. Hence there is a diagonal matrix $D = \mathrm{diag}(e^{i\alpha_1}, \dots, e^{i\alpha_n})$ such that $Y = DX$. Let us define a unimodular complex number $e^{i\gamma} = \mu/\rho(A)$ and set $B := e^{-i\gamma}D^{-1}AD$. One has $|B| = A$ and $BX = X$. For every j, one therefore has

$$\left|\sum_{k=1}^{n} b_{jk}x_k\right| = \sum_{k=1}^{n} |b_{jk}|x_k.$$

Because $X > 0$, one deduces that B is real-valued and nonnegative. Therefore $B = A$; that is, $D^{-1}AD = e^{i\gamma}A$. The spectrum of A is thus invariant under multiplication by $e^{i\gamma}$.

Let $\mathscr{U} = \rho(A)^{-1}\,\mathrm{Sp}_{max}(A)$, which is included in S^1, the unit circle. The previous discussion shows that $\mathscr{U}$ is stable under multiplication. Because $\mathscr{U}$ is finite, it follows that its elements are roots of unity. The inverse of a dth root of unity is its own $(d-1)$th power, therefore $\mathscr{U}$ is stable under inversion. Hence it is a finite subgroup of S^1. With p its cardinal, we have $\mathscr{U} = \mathscr{U}_p$.

Let P_A be the characteristic polynomial and let $\omega = \exp(2i\pi/p)$. One may apply the first part of the proof to $\mu = \omega\rho(A)$. One has thus $D^{-1}AD = \omega A$, and it follows that $P_A(X) = \omega^n P_A(X/\omega)$. Therefore, multiplication by ω sends eigenvalues to eigenvalues of the same multiplicities. In particular, the eigenvalues of maximal modulus are simple.

Iterating the conjugation, one obtains $D^{-p}AD^p = A$. Let us set

$$D^p = \mathrm{diag}(d_1, \dots, d_n).$$

One has thus $d_j = d_k$, provided that $a_{jk} \neq 0$. Because A is irreducible, one can link any two indices j and k by a chain $j_0 = j, \dots, j_r = k$ such that $a_{j_{s-1},j_s} \neq 0$ for every

s. It follows that $d_j = d_k$ for every j,k. But because one may choose $Y_1 = X_1$, that is $\alpha_1 = 0$, one also has $d_1 = 1$ and hence $D^p = I_n$. The α_j are thus pth roots of unity. Applying a conjugation by a permutation matrix we may limit ourselves to the case where D has the block-diagonal form $\mathrm{diag}(J_0, \omega J_1, \dots, \omega^{p-1} J_{p-1})$, where the J_ℓ are identity matrices of respective sizes $n_0, \dots, n_{p-1}$. Decomposing A into blocks A_{lm} of sizes $n_\ell \times n_m$, one obtains $\omega^k A_{jk} = \omega^{j+1} A_{jk}$ directly from the conjugation identity. Hence $A_{jk} = 0$, except for the pairs (j,k) of the form $(0,1),(1,2),\dots,(p-2,p-1),(p-1,0)$. This is the announced cyclic form. □

8.5 Stochastic Matrices

Definition 8.3 *A matrix $M \in \mathbf{M}_n(\mathbb{R})$ is said to be* stochastic *if $M \geq 0$ and if for every $i = 1,\dots,n$, one has*

$$\sum_{j=1}^{n} m_{ij} = 1.$$

One says that M is bistochastic *(or* doubly stochastic*) if both M and M^T are stochastic.*

Denoting by $\mathbf{e} \in \mathbb{R}^n$ the vector all of whose coordinates equal one, one sees that M is stochastic if and only if $M \geq 0$ and $M\mathbf{e} = \mathbf{e}$. Likewise, M is bistochastic if $M \geq 0$, $M\mathbf{e} = \mathbf{e}$, and $\mathbf{e}^T M = \mathbf{e}^T$. If M is stochastic, one has $\|Mx\|_\infty \leq \|x\|_\infty$ for every $x \in \mathbb{C}^n$, and therefore $\rho(M) \leq 1$. But because $M\mathbf{e} = \mathbf{e}$, one has in fact $\rho(M) = 1$.

The stochastic matrices play an important role in the study of Markov chains. A special instance of a bistochastic matrix is a permutation matrix $P(\sigma)$ ($\sigma \in S_n$), whose entries are

$$p_{ij} = \delta^j_{\sigma(i)}.$$

The following theorem enlightens the role of permutation matrices.

Theorem 8.4 (Birkhoff) *A matrix $M \in \mathbf{M}_n(\mathbb{R})$ is bistochastic if and only if it is a center of mass (i.e., a barycenter with nonnegative weights) of permutation matrices.*

The fact that a center of mass of permutation matrices is a doubly stochastic matrix is obvious, because the set $\mathbf{DS}_n$ of doubly stochastic matrices is convex. The interest of the theorem lies in the statement that if $M \in \mathbf{DS}_n$, there exist permutation matrices $P_1, \dots, P_r$ and positive real numbers $\alpha_1, \dots, \alpha_r$ with $\alpha_1 + \cdots + \alpha_r = 1$ such that $M = \alpha_1 P_1 + \cdots + \alpha_r P_r$.

Let us recall that a point x of a convex set C is an *extremal point* if $x \in [y,z] \subset C$ implies $x = y = z$. The permutation matrices are extremal points of $\mathbf{M}_n([0,1]) \sim [0,1]^{n^2}$, thus they are extremal points of the smaller convex set $\mathbf{DS}_n$.

The Krein–Milman theorem (see [34], Theorem 3.23) says that a convex compact subset of $\mathbb{R}^n$ is the convex hull, that is, the set of centers of mass of its extremal points. Because $\mathbf{DS}_n$ is closed and bounded, hence compact, we may apply this statement. Theorem 8.4 thus amounts to saying that the extremal points of Δ_n are precisely the permutation matrices.

Proof. Let $M \in \mathbf{DS}_n$ be given. If M is not a permutation matrix, there exists an entry $m_{i_1 j_1} \in (0,1)$. Inasmuch as M is stochastic, there also exists $j_2 \neq j_1$ such that $m_{i_1 j_2} \in (0,1)$. Because M^T is stochastic, there exists $i_2 \neq i_1$ such that $m_{i_2 j_2} \in (0,1)$. By this procedure one constructs a sequence $(j_1, i_1, j_2, i_2, \ldots)$ such that $m_{i_\ell j_\ell} \in (0,1)$ and $m_{i_{\ell-1} j_\ell} \in (0,1)$. The set of indices is finite, therefore it eventually happens that one of the indices (a row index or a column index) is repeated.

Therefore, one can assume that the sequence $(j_s, i_s, \ldots, j_r, i_r, j_{r+1} = j_s)$ has the above property, and that $j_s, \ldots, j_r$ are pairwise distinct, as well as $i_s \ldots, i_r$. Let us define a matrix $B \in \mathbf{M}_n(\mathbb{R})$ by $b_{i_\ell j_\ell} = 1$, $b_{i_\ell j_{\ell+1}} = -1$, $b_{ij} = 0$ otherwise. By construction, $B\mathbf{e} = 0$ and $\mathbf{e}^T B = 0$. If $\alpha \in \mathbb{R}$, one therefore has $(M \pm \alpha B)\mathbf{e} = \mathbf{e}$ and $\mathbf{e}^T(M \pm \alpha B) = \mathbf{e}^T$. If $\alpha > 0$ is small enough, $M \pm \alpha B$ turns out to be nonnegative. Finally, $M + \alpha B$ and $M - \alpha B$ are bistochastic, and

$$M = \frac{1}{2}(M - \alpha B) + \frac{1}{2}(M + \alpha B).$$

Hence M is not an extremal point of $\mathbf{DS}_n$. □

Here is a nontrivial consequence (Stoer and Witzgall [36]):

Corollary 8.1 *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$, invariant under permutation of the coordinates. Then $\|M\| = 1$ for every bistochastic matrix (where as usual we have denoted $\|\cdot\|$ the induced norm on $\mathbf{M}_n(\mathbb{R})$).*

Proof. To begin with, $\|P\| = 1$ for every permutation matrix, by assumption. Because the induced norm is convex (true for every norm), one deduces from Birkhoff's theorem that $\|M\| \leq 1$ for every bistochastic matrix. Furthermore, $M\mathbf{e} = \mathbf{e}$ implies $\|M\| \geq \|M\mathbf{e}\|/\|\mathbf{e}\| = 1$. □

This result applies, for instance, to the norm $\|\cdot\|_p$, providing a nontrivial convex set on which the map $1/p \mapsto \log \|M\|_p$ is constant (compare with Theorem 7.2).

The bistochastic matrices are intimately related to the relation $\prec$ (see Section 6.5). In fact, we have the following theorem.

Theorem 8.5 *A matrix A is bistochastic if and only if $Ax \succ x$ for every $x \in \mathbb{R}^n$.*

Proof. If A is bistochastic, then $\|Ax\|_1 \leq \|A\|_1 \|x\|_1 = \|x\|_1$, because A^T is stochastic. Because A is stochastic, $A\mathbf{e} = \mathbf{e}$. Applying the inequality to $x - t\mathbf{e}$, one therefore has $\|Ax - t\mathbf{e}\|_1 \leq \|x - t\mathbf{e}\|_1$. Proposition 6.4 then shows that $x \prec Ax$.

Conversely, let us assume that $x \prec Ax$ for every $x \in \mathbb{R}^n$. Choosing x as the jth vector of the canonical basis, $\mathbf{e}^j$, the inequality $s_1(\mathbf{e}^j) \leq s_1(A\mathbf{e}^j)$ expresses that A is a nonnegative matrix, and $s_n(\mathbf{e}^j) = s_n(A\mathbf{e}^j)$ yields

$$\sum_{i=1}^{n} a_{ij} = 1. \tag{8.2}$$

One then chooses $x = \mathbf{e}$. The inequality $s_1(\mathbf{e}) \le s_1(A\mathbf{e})$ expresses[1] that $A\mathbf{e} \ge \mathbf{e}$. Finally, $s_n(\mathbf{e}) = s_n(A\mathbf{e})$ and $A\mathbf{e} \ge \mathbf{e}$ give $A\mathbf{e} = \mathbf{e}$. Hence, A is bistochastic. □

This statement is completed by the following.

Theorem 8.6 *Let $x, y \in \mathbb{R}^n$. Then $x \prec y$ if and only if there exists a bistochastic matrix A such that $y = Ax$.*

Proof. From the previous theorem, it is enough to show that if $x \prec y$, there exists A, a bistochastic matrix, such that $y = Ax$. To do so, one applies Theorem 6.8: there exists an Hermitian matrix H whose diagonal and spectrum are y and x, respectively. Let us diagonalize H by a unitary conjugation: $H = U^*DU$, with $D = \mathrm{diag}(x_1, \dots, x_n)$. Then $y = Ax$, where $a_{ij} = |u_{ij}|^2$. Because U is unitary, A is bistochastic.[2] □

Exercises

1. We consider the following three properties for a matrix $M \in \mathbf{M}_n(\mathbb{R})$.

   P1 M is nonnegative.
   P2 $M^T e = e$, where $e = (1, \dots, 1)^T$.
   P3 $\|M\|_1 \le 1$.

   a. Show that **P2** and **P3** imply **P1**.
   b. Show that **P2** and **P1** imply **P3**.
   c. Do **P1** and **P3** imply **P2** ?

2. Here is another proof of the simplicity of $\rho(A)$ in the Perron–Frobenius theorem, which does not require Lemma 12. We assume that A is irreducible and nonnegative, and we denote by x a positive eigenvector associated with the eigenvalue $\rho(A)$.

   a. Let K be the set of nonnegative eigenvectors y associated with $\rho(A)$ such that $\|y\|_1 = 1$. Show that K is compact and convex.
   b. Show that the geometric multiplicity of $\rho(A)$ equals 1. (**Hint:** Otherwise, K would contain a vector with at least one zero component.)
   c. Show that the algebraic multiplicity of $\rho(A)$ equals 1. (**Hint:** Otherwise, there would be a nonnegative vector y such that $Ay - \rho(A)y = x > 0$.)

3. Let $M \in \mathbf{M}_n(\mathbb{R})$ be either strictly diagonally dominant or irreducible and strongly diagonally dominant. Assume that $m_{jj} > 0$ for every $j = 1, \dots, n$ and $m_{ij} \le 0$ otherwise. Show that M is invertible and that the solution of $Mx = b$, when $b \ge 0$, satisfies $x \ge 0$. Deduce that $M^{-1} \ge 0$.

[1] For another vector y, $s_1(y) \le s_1(Ay)$ does not imply $Ay \ge y$.

[2] This kind of bistochastic matrix is called *orthostochastic*.

4. Here is another proof of Theorem 8.2, due to Perron himself. We proceed by induction on the size n of the matrix. The statement is obvious if $n = 1$. We therefore assume that it holds for matrices of size n. We give ourselves an irreducible nonnegative matrix $A \in \mathbf{M}_{n+1}(\mathbb{R})$, which we decompose blockwise as
$$A = \begin{pmatrix} a & \xi^T \\ \eta & B \end{pmatrix}, \quad a \in \mathbb{R}, \quad \xi, \eta \in \mathbb{R}^n, \quad B \in \mathbf{M}_n(\mathbb{R}).$$
   a. Applying the induction hypothesis to the matrix $B + \varepsilon J$, where $\varepsilon > 0$ and $J > 0$ is a matrix, then letting ε go to zero, shows that $\rho(B)$ is an eigenvalue of B, associated with a nonnegative eigenvector (this avoids the use of Theorem 8.1).
   b. Using the formula
$$(\lambda I_n - B)^{-1} = \sum_{k=1}^{\infty} \lambda^{-k} B^{k-1},$$
   valid for $\lambda \in (\rho(B), +\infty)$, deduce that the function $h(\lambda) := \lambda - a - \xi^T(\lambda I_n - B)^{-1}\eta$ is strictly increasing on this interval, and that on the same interval the vector $x(\lambda) := (\lambda I_n - B)^{-1}\eta$ is positive.
   c. Using the Schur complement formula, prove $P_A(\lambda) = P_B(\lambda)h(\lambda)$.
   d. Deduce that the matrix A has one and only one eigenvalue in $(\rho(B), +\infty)$, and that it is a simple one, associated with a positive eigenvector. One denotes this eigenvalue by λ_0.
   e. Applying the previous results to A^T, show that there exists $\ell \in \mathbb{R}^n$ such that $\ell > 0$ and $\ell^T(A - \lambda_0 I_n) = 0$.
   f. Let μ be an eigenvalue of A, associated with an eigenvector X. Show that $(\lambda_0 - |\mu|)\ell^T|X| \geq 0$. Conclusion?

5. Let $A \in \mathbf{M}_n(\mathbb{R})$ be a matrix satisfying $a_{ij} \geq 0$ for every pair (i, j) of distinct indices.

   a. Let us define
$$\sigma := \sup\{\Re\lambda; \lambda \in \operatorname{Sp} A\}.$$
   Among the eigenvalues of A whose real parts equal σ, let us denote by μ the one with the largest imaginary part. Show that for every positive large enough real number τ, $\rho(A + \tau I_n) = |\mu + \tau|$.
   b. Deduce that $\mu = \sigma = \rho(A)$ (apply Theorem 8.1).

6. Let $B \in \mathbf{M}_n(\mathbb{R})$ be a matrix whose off-diagonal entries are positive and such that the eigenvalues have strictly negative real parts. Show that there exists a nonnegative diagonal matrix D such that $B' := D^{-1}BD$ is strictly diagonally dominant, namely,
$$b'_{ii} < -\sum_{j \neq i} b'_{ij}.$$

7. a. Let $B \in \mathbf{M}_n(\mathbb{R})$ be given, with $\rho(B) = 1$. Assume that the eigenvalues of B of modulus one are (algebraically) simple. Show that the sequence $(B^m)_{m \geq 1}$ is bounded.
   b. Let $M \in \mathbf{M}_n(\mathbb{R})$ be a nonnegative irreducible matrix, with $\rho(M) = 1$. We denote by x and y^T the left and right eigenvectors for the eigenvalue 1 ($Mx = x$ and $y^T M = y^T$), normalized by $y^T x = 1$. We define $L := xy^T$ and $B = M - L$.
      i. Verify that $B - I_n$ is invertible. Determine the spectrum and the invariant subspaces of B by means of those of M.
      ii. Show that the sequence $(B^m)_{m \geq 1}$ is bounded. Express M^m in terms of B^m.
      iii. Deduce that
$$\lim_{N \to +\infty} \frac{1}{N} \sum_{m=0}^{N-1} M^m = L.$$
      iv. Under what additional assumption do we have the stronger convergence
$$\lim_{N \to +\infty} M^N = L?$$
8. Let $B \in \mathbf{M}_n(\mathbb{R})$ be a nonnegative irreducible matrix and let $C \in \mathbf{M}_n(\mathbb{R})$ be a nonzero nonnegative matrix. For $t > 0$, we define $r_t := \rho(B + tC)$ and we let X_t denote the nonnegative unitary eigenvector associated with the eigenvalue r_t.

   a. Show that $t \mapsto r_t$ is strictly increasing.
   Define $r := \lim_{t \to +\infty} r_t$. We wish to show that $r = +\infty$. Let X be a cluster point of the sequence X_t. We may assume, up to a permutation of the indices, that
$$X = \begin{pmatrix} Y \\ 0 \end{pmatrix}, \quad Y > 0.$$
   b. Suppose that in fact, $r < +\infty$. Show that $BX \leq rX$. Deduce that $B'Y = 0$, where B' is a matrix extracted from B.
   c. Deduce that $X = Y$; that is, $X > 0$.
   d. Show, finally, that $CX = 0$. Conclude that $r = +\infty$.
   e. Assume, moreover, that $\rho(B) < 1$. Show that there exists one and only one $t \in \mathbb{R}$ such that $\rho(B + tC) = 1$.
9. Verify that $\mathbf{DS}_n$ is stable under multiplication. In particular, if M is bistochastic, the sequence $(M^m)_{m \geq 1}$ is bounded.
10. Let $M \in \mathbf{M}_n(\mathbb{R})$ be a bistochastic irreducible matrix. Show that
$$\lim_{N \to +\infty} \frac{1}{N} \sum_{m=0}^{N-1} M^m = \frac{1}{n} \begin{pmatrix} 1 & \dots & 1 \\ \vdots & & \vdots \\ 1 & \dots & 1 \end{pmatrix} =: J_n$$
   (use Exercise 7). Show by an example that the sequence $(M^m)_{m \geq 1}$ may or may not converge.

11. Show directly that for every $p \in [1,\infty]$, $\|J_n\|_p = 1$, where J_n was defined in the previous exercise.
12. Let $P \in \mathbf{GL}_n(\mathbb{R})$ be given such that $P, P^{-1} \in \mathbf{DS}_n$. Show that P is a permutation matrix.
13. Let $A, B \in \mathbf{H}_n$ be given and $C := A + B$.

    a. If $t \in [0,1]$ consider the matrix $S(t) := A + tB$, so that $S(0) = A$ and $S(1) = C$. Arrange the eigenvalues of $S(t)$ in increasing order $\lambda_1(t) \leq \cdots \leq \lambda_n(t)$. For each value of t there exists an orthonormal eigenbasis $\{X_1(t), \ldots, X_n(t)\}$. We admit the fact that it can be chosen continuously with respect to t, so that $t \mapsto X_j(t)$ is continuous with a piecewise continuous derivative (see [24], Chapter 2, Section 6.) Show that $\lambda_j'(t) = (BX_j(t), X_j(t))$.
    b. Let $\alpha_j, \beta_j, \gamma_j$ ($j = 1, \ldots, n$) be the eigenvalues of A, B, C, respectively. Deduce from part (a) that
    $$\gamma_j - \alpha_j = \int_0^1 (BX_j(t), X_j(t))\, dt.$$
    c. Let $\{Y_1, \ldots, Y_n\}$ be an orthonormal eigenbasis for B. Define
    $$\sigma_{jk} := \int_0^1 |(X_j(t), Y_k)|^2 dt.$$
    Show that the matrix $\Sigma := (\sigma_{jk})_{1 \leq j,k \leq n}$ is bistochastic.
    d. Show that $\gamma_j - \alpha_j = \sum_k \sigma_{jk} \beta_k$. Deduce (Lidskiĭ's theorem) that the vector $(\gamma_1 - \alpha_1, \ldots, \gamma_n - \alpha_n)$ belongs to the convex hull of the vectors obtained from the vector $(\beta_1, \ldots, \beta_n)$ by all possible permutations of the coordinates.

14. Let $a \in \mathbb{R}^n$ be given, $a = (a_1, \ldots, a_n)$.

    a. Show that $C(a) := \{b \in \mathbb{R}^n \,|\, b \succ a\}$ is a convex compact set. Characterize its extremal points.
    b. Show that $Y(a) := \{M \in \mathbf{Sym}_n(\mathbb{R}) \,|\, \operatorname{Sp} M \succ a\}$ is a convex compact set. Characterize its extremal points.
    c. Deduce that $Y(a)$ is the closed convex hull (actually the convex hull) of the set $X(a) := \{M \in \mathbf{Sym}_n(\mathbb{R}) \,|\, \operatorname{Sp} M = a\}$.
    d. Set $\alpha = s_n(a)/n$ and $a' := (\alpha, \ldots, \alpha)$. Show that $a' \in C(a)$, and that $b \in C(a) \Longrightarrow b \prec a'$.
    e. Characterize the set $\{M \in \mathbf{Sym}_n(\mathbb{R}) \,|\, \operatorname{Sp} M \prec a'\}$.

15. Use Exercise 14 to prove the theorem of Horn and Schur. The set of diagonals $(h_{11}, \ldots, h_{nn})$ of Hermitian matrices with given spectrum $(\lambda_1, \ldots, \lambda_n)$ is the convex hull of the points $(\lambda_{\sigma(1)}, \ldots, \lambda_{\sigma(n)})$ as σ runs over the permutations of $\{1, \ldots, n\}$.
16. (Boyd, Diaconis, Sun and Xiao.) Let P be a symmetric stochastic $n \times n$ matrix:

$$p_{ij} = p_{ji} \geq 0, \qquad \sum_j p_{ij} = 1 \quad (i = 1, \dots, n).$$

We recall that $\lambda_1 = 1$ is an eigenvalue of P, which is the largest in modulus (Perron–Frobenius). We are interested in the second largest modulus $\mu(P) = \max\{\lambda_2, -\lambda_n\}$ where $\lambda_1 \geq \cdots \geq \lambda_n$ is the spectrum of P; $\mu(P)$ is the second singular value of P.

a. Let $y \in \mathbb{R}^n$ be such that $\|y\|_2 = 1$ and $\sum_j y_j = 1$. Let $z \in \mathbb{R}^n$ be such that

$$(p_{ij} \neq 0) \Longrightarrow \left(\frac{1}{2}(z_i + z_j) \leq y_i y_j\right).$$

Show that $\lambda_2 \geq \sum_j z_j$. **Hint:** Use Rayleigh ratio.

b. Likewise, if y is as above and w such that

$$(p_{ij} \neq 0) \Longrightarrow \left(\frac{1}{2}(w_i + w_j) \geq y_i y_j\right),$$

show that $\lambda_n \leq \sum_j w_j$.

c. Taking

$$y_j = \sqrt{\frac{2}{n}} \cos\frac{(2j-1)\pi}{2n}, \quad z_j = \frac{1}{n}\left(\cos\frac{\pi}{n} + \frac{\cos\frac{(2j-1)\pi}{n}}{\cos\frac{\pi}{n}}\right),$$

deduce that $\mu(P) \geq \cos\dfrac{\pi}{n}$ for every symmetric stochastic $n \times n$ matrix.

d. Find a symmetric stochastic $n \times n$ matrix Q such that $\mu(Q) = \cos\dfrac{\pi}{n}$. **Hint:** Exploit the equality case in the analysis, with the y and z given above.

e. Prove that $P \mapsto \mu(P)$ is a convex function over symmetric stochastic $n \times n$ matrices.

17. Prove the equivalence of the following properties for real $n \times n$ matrices A:

Strong Perron–Frobenius. The spectral radius is a simple eigenvalue of A, the only one of this modulus; it is associated with positive left and right eigenvectors.

Eventually positive matrix. There exists an integer $k \geq 1$ such that $A^k > 0_n$.

18. Let A be a cyclic matrix, as in Theorem 8.3. If $1 \leq j \leq p$, prove that

$$P_A(X) = X^{n - pm_j} \det(X^p I_{m_j} - M_j M_{j+1} \cdots M_p M_1 \cdots M_{j-1}),$$

where m_j is the number of rows of M_j (or columns in M_{j-1}) (**Hint:** argue by induction over p, using the Schur complement formula). Deduce a lower bound of the multiplicity of the eigenvalue $\lambda = 0$.