Chapter 9

Hermitian and Symmetric Matrices

Example 9.0.1. Let $f\colon\ D \to R$, $D \subset R^n$. The *Hessian* is defined by

$$H(x) = h_{ij}(x) \equiv \frac{\partial f}{\partial x_i \partial x_j} \in M_n.$$

Since for functions $f \in C^2$ it is known that

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

it follows that $H(x)$ is symmetric.

Definition 9.0.1. A function $f\colon\ R \to R$ is *convex* if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

for $x, y \in D$ (domain) and $0 \leq \lambda \leq 1$.

Proposition 9.0.1. *If $f \in C^2(D)$ and $f''(x) \geq 0$ on D then $f(x)$ is convex.*

Proof. Because $f'' \geq 0$, this implies that $f'(x)$ is increasing. Therefore if $x < x_m < y$ we must have

$$f(x_m) \leq f(x) + f'(x_m)(x_m - x)$$

and

$$f(x_m) \leq f(y) + f'(x_m)(x_m - y)$$

by the Mean Value Theorem. Therefore,

$$\frac{y - x_m}{y - x} f(x_m) + \frac{x_m - x}{y - x} f(x_m) \le \frac{y - x_m}{y - x} f(x) + \frac{x_m - x}{y - x} f(y)$$

or

$$f(x_m) \le \frac{y - x_m}{y - x} f(x) + \frac{(x_m - x)}{y - x} f(y).$$

It is easy to see that

$$x_m = \frac{y - x_m}{y - x} x + \frac{x_m - x}{y - x} y.$$

Defining $\lambda = \frac{y - x_m}{y - x}$, it follows that $1 - \lambda = \frac{x_m - x}{y - x}$. □

Definition 9.0.2. We say that $f\colon\ D \to R$, where $D \subset R^n$ is convex, is a convex function if f is a convex function on every line in D.

Theorem 9.0.1. *Suppose $f \in C^2(D)$ and $H(x)$ is positive definite. Then f is convex on D.*

Proof. Let $x \in D$ and η be some direction. Then $x + \lambda\eta$ is a line in D. We compute $\frac{d^2}{d\lambda^2} f(x + \lambda n)$

$$\frac{d}{d\lambda} f = \nabla f \cdot n = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \cdot [n_1 \dots n_n].$$

Now

$$\frac{d}{d\lambda}\left(\frac{\partial f}{\partial x_1}\right) = \nabla \frac{\partial f}{\partial x_1} \cdot \eta = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} \\ \\ \frac{\partial^2 f}{\partial x_n \partial x_1} \end{bmatrix} \cdot [\eta_1, \dots, \eta_n].$$

Hence we see that

$$\frac{d^2}{d\lambda^2} f(x + \lambda\eta) = \eta^T H(x)\eta \ge 0$$

by assumption. □

Example 9.0.2. Let $A = [a_{ij}] \in M_n$. Consider the quadratic form on $\mathbb{C}^n$ or R^n defined by

$$\begin{aligned} Q(x) &= x^T A x = \Sigma a_{ij} x_j x_i \\ &= \frac{1}{2}\Sigma(a_{ij} + a_{ji}) x_j x_i \\ &= x^T \frac{1}{2}(A + A^T)x. \end{aligned}$$

Since the matrix $A+A^T$ is symmetric the study of quadratic forms is reduced to the symmetric case.

Example 9.0.3. Let $Lf = \sum\limits_{i,j=1}^{n} a_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}$. L is called a partial differential operator. By the combination of devices above (assuming $f \in C^2$ for example) we can study the symmetric and equivalent partial differential operator

$$Lf = \sum_{i,j=1}^{n} \frac{1}{2}(a_{ij} + a_{ji}) \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

In particular, if $A + A^T$ is positive definite the operator is called elliptic.

Other cases are

(1) hyperbolic,

(2) degenerate/parabolic.

Characterizations of Hermitian matrices. Recall

(1) $A \in M_n$ is Hermitian if $A^* = A$.

(2) $A \in M_n$ is called skew-Hermitian if $A = -A^*$.

Here are some facts

(a) If A is Hermitian the diagonal is real.

(b) If A is skew-Hermitian the diagonal is imaginary.

(c) $A + A^*, AA^*$ and A^*A are all Hermitian if $A \in M_n$.

(d) If A is Hermitian than A^k, $k = 0, 1, \ldots$, are Hermitian. A^{-1} is Hermitian if A is invertible.

(e) $A - A^*$ is skew-Hermitian.

(f) $A \in M_n$ yields the decomposition

$$A = \underbrace{\frac{1}{2}(A + A^*)}_{\text{Hermitian}} + \underbrace{\frac{1}{2}(A - A^*)}_{\text{Skew Hermitian}}$$

(g) If A is Hermitian iA is skew-Hermitian. If A is skew-Hermitian then iA is Hermitian.

Theorem 9.0.2. *Let $A \in M_n$. Then $A = S + iT$ where S and T are Hermitian. Moreover this is unique.*

Proof.

$$\begin{aligned} A &= \frac{1}{2}(A + A^*) + \frac{1}{2}(A - A^*) \\ &= S + iT \end{aligned}$$

where $S = \frac{1}{2}(A + A^*)$ and

$$\begin{aligned} iT &= \frac{1}{2}(A - A^*) \\ \Rightarrow T &= -\frac{i}{2}(A - A^*). \quad \square \end{aligned}$$

Theorem 9.0.3. *Let $A \in M_n$ be Hermitian. Then*

(a) $x^* A x$ is real for all $x \in \mathbb{C}^n$;

(b) All the eigenvalues of A are real;

(c) $S^* A S$ is Hermitian for all $S \in M_n$.

Proof. For (a) we have

$$x^* A x = \Sigma a_{ij} x_j x_i.$$

The conjugate is

$$\begin{aligned} \overline{x^* A x} = \Sigma \bar{a}_{ij} \bar{x}_j x_i &= \Sigma a_{ji} \bar{x}_j x_i \\ &= \Sigma a_{ij} x_j \bar{x}_i + x^* A x \quad \square \end{aligned}$$

Theorem 9.0.4. *Let $A \in M_n$. Then A is Hermitian if and only if at least one of the following holds:*

(a) $\langle Ax, x\rangle = x^* Ax$ *is real* $\forall x \in \mathbb{C}^n$.

(b) A is normal with real eigenvalues.

(c) S^*AS *is Hermitian for all* $S \in M_n$.

Proof. Hermitian $\Rightarrow$ (a), (b), or (c) are obvious. (a) $\Rightarrow$ Hermitian. Prove using e_j and e_k. We have

$$\langle Ae_j + e_k, e_j + e_k\rangle = a_{jj} + a_{kk} + a_{jk} + a_{kj}$$

$\Rightarrow a_{jk} + a_{kj}$ is real $\Rightarrow$ Im $a_{jk} = -$Im a_{kj}.
Similarly

$$\langle Aie_j + e_k, ie_j + e_k\rangle - a_{jj} + a_{kk} + ia_{kj} - ia_{jk}$$

$\Rightarrow i(a_{kj} - a_{jk})$ is real $\Rightarrow$ Re $a_{kj} =$ Re a_{jk}.
All this gives $A^* = A$.
(c) $\Rightarrow$ Hermitian S^*AS is Hermitian $\Rightarrow S^*AS$ is similar to a real diagonal matrix. Therefore A is similar to a real diagonal matrix. Just let $S = I$ to get A is Hermitian. □

Theorem 9.0.5 (Spectral Theorem). *Let* $A \in M_n$ *be Hermitian. Then* A *is unitarily (similar) equivalent to a real diagonal matrix. If* A *is real Hermitian, then* A *is orthogonally similar to a real diagonal matrix.*

9.1 Variational Characterizations of Eigenvalues

Let $A \in M_n$ be Hermitian. Assume

$$\lambda_{\min} \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n-1} \leq \lambda_n = \lambda_{\max}.$$

Theorem 9.1.1 (Rayleigh–Ritz). *Let* $A \in M_n$, *and let the eigenvalues of* A *be ordered as above. Then*

$$\lambda_{\max} = \lambda_n = \max_{x\neq 0} \frac{\langle Ax, x\rangle}{\langle x, x\rangle}$$
$$\lambda_{\min} = \lambda_1 = \min_{x\neq 0} \frac{\langle Ax, x\rangle}{\langle x, x\rangle}.$$

Proof. Let $x_1 \dots x_n$ be the linearly independent and orthogonal eigenvectors of A. Any vector $x \in \mathbb{C}^n$ has the representation

$$x\Sigma\alpha_j x_j.$$

Then $\langle Ax, x\rangle = \Sigma\alpha_j^2\lambda_j$ and $\langle x, x\rangle = \Sigma\alpha_j^2$. Hence

$$\frac{\langle Ax, x\rangle}{\langle x, x\rangle} = \sum_j \frac{\alpha_j^2}{\sum_i \alpha_i^2}\lambda_j.$$

Since $\left\{\frac{\alpha_j^2}{\Sigma\alpha_i^2}\right\}$ is nonnegative and sums to 1, it follows that

$$\frac{\langle Ax, x\rangle}{\langle x, x\rangle}$$

is a convex combination of the eigenvalues, whence the theorem follows. In particular take $x = x_n(x_1)$ to achieve the maximum (minimum). □

Remark 9.1.1. This result gives just the largest and smallest eigenvalues. How can we achieve the intermediate eigenvalues? We have already considered this problem somewhat in conjunction with the power method. In that consideration we employed the bi-orthogonal eigenvectors. For a Hermitian matrix, the families are the same. So we could characterize the eigenvalues in a manner similar to that discussed previously. However, the following characterization is simpler.

Theorem 9.1.2. *Let $A \in M_n$ be Hermitian with eigenvalues as above and corresponding eigenvectors $x_1 \dots x_n$. Then*

$$(*) \qquad \lambda_{n-k} = \max_{\substack{x \perp \{x_n, \dots, x_{n-k+1}\} \\ x \neq 0}} \frac{\langle Ax, x\rangle}{\langle x, x\rangle}.$$

The following result is even more general

$$(*) \qquad \lambda_{n-k} = \min_{\{w_n, w_{n-1} \dots w_{n-k+1}\}} \max_{\substack{x \perp \{w_n, w_{n-1} \dots w_{n-k+1}\} \\ x \neq 0}} \frac{\langle Ax, x\rangle}{\langle x, x\rangle}.$$

Proof. The Rayleigh–Ritz argument above gives $(*)$ directly. □

9.2 Matrix inequalities

When the underlying matrix is symmetric or positive definite, certain powerful inequalities can be established. The first inequality is a consequence of the cofactor result Proposition 2.5.1.

Theorem 9.2.1. *Let $A \in M_n(\mathbb{C})$ be positive definite. Then* $\det A \le a_{11}a_{22}\cdots a_{nn}$

Proof. Expanding in minors, the determinant of A

$$\det A = a_{11}\begin{vmatrix} a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n2} & & a_{nn} \end{vmatrix} + \det\begin{vmatrix} 0 & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & & a_{nn} \end{vmatrix}$$

Since A is positive definite, it follows that each of it principal submatrices is also. Therefore, the first term in the right side of the equality above is postive, while the second term is negative by the cofactor result Proposition 2.5.1. To clarify , it is important to note that if A is Hermitian and invertible, so also is its inverse. In particular, it A is positive definite, we know its determinant is positive and so when we write

$$\det\begin{vmatrix} 0 & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & & a_{nn} \end{vmatrix} = -\sum a_{1j}a_{1k}a^{ik}$$

it is clear this term is negative. Therefore,

$$\det A \le a_{11}\begin{vmatrix} a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n2} & & a_{nn} \end{vmatrix}$$

whence the result follows inductively. □

Corollary 9.2.1. *Let $S_1, S_2, \ldots,\ S_r$ be any partition of the integers $\{1, \ldots, n\}$, and let $A_1, A_2, \ldots, A_r$., be the principal submatrices of A pertaining to the indices of each subdivision. Then*

$$\det A = \det A_1 \det A_2 \cdots \det A_r$$

An important consequence of Theorem 9.2.1 is one of the most famous determinantal inequalities. It is due to Hadamard.

Theorem 9.2.1. *Let B be an arbitrary nonsingular real square matrix. Then*

$$(\det B)^2 \leq \prod_{i=1}^{n} \sum_{j=1}^{n} |b_{ij}|^2$$

Proof. Define $A = B^T B$. Then

$$\operatorname{diag} A = \left(\sum_{j=1}^{n} |b_{1j}|^2, \ \ldots, \sum_{j=1}^{n} |b_{nj}|^2 \right)$$

whence the result follows from Theorem 9.2.1 since $\det A = (\det B)^2$. □

9.3 Exercises

1. Prove Corollary 9.2.1.
2. Establish Theorem 9.2.1 in the case the matrix is complex.
3. Establish a result similar to Theorem 9.2.1 for rectangular matrices.

Chapter 10

Nonnegative Matrices

10.1 Definitions

Nonnegative matrices are simply those with all nonnegative entries. We will eventually distinguish various types of such matrices substantially on how they are nonnegative or conversely where they are zero. A particular type of nonnegative matrix, the type that has no off-diagonal block to be zero, will turn out to have the greatest value to us. If nonnegative matrices distinguished themselves in only minor ways from general matrices, they would not occupy the high position of importance they enjoy in the modern literature. Indeed, many remarkable properties of nonnegative matrices have been uncovered. Owing substantially to the many applications to economics, probability, and engineering, the subject has now for more than a century been one of the hottest areas of research within the subject.

Definition 10.1.1. $A \in M_n(R)$ is called **nonnegative** if $a_{ij} \geq 0$ for $1 \leq i, j \leq n$. We use the notation $A \geq 0$ for nonnegative matrices. If $a_{ij} > 0$ for all i, j we say A is *fully* (or *strictly*) *positive*. (Notation. $A \gg 0$.)

In this section we need some special notation. Let $x \in \mathbb{C}^n$, $x = (x_1, \dots, x_n)^T$. Then

$$|x| = (|x_1|, \dots, |x_n|)^T.$$

We say that $x \in R^n$ is *positive* if $x_i \geq 0$, $1 \leq i \leq n$. We say that $x \in R^n$ is *fully* (or *strictly*) *positive* if $x_i > 0$, $1 \leq i \leq n$. We use the notation $x \geq 0$ and $x \gg 0$ for positive and strictly positive vectors, respectively.

Note that if $A \geq 0$ we have $\|A\|_\infty = \|Ae\|_\infty$ where $e = (1, 1, \dots, 1)^T$.

To give an idea of the direction we are heading, let us suppose that $A \in M_n(R)$ is a nonnegative matrix. Let $\lambda \neq 0$ be an eigenvector of A with pertaining eigenvector $x \geq 0$. Thus $Ax = \lambda x$. (We will prove that this comes to pass for every nonnegative square matrix.) Suppose further that the vector Ax is zero on the index set S. That is $(Ax)_i = 0$ if $i \in S$. Since $\lambda \neq 0$ it follows that $x_i = 0$ $i \in S$. Let S^C be the complement of S in the integers $\{1, \dots, n\}$, and let P and $P^\perp$ be the projections to S and S^C, respectively. So, $Px = 0$ and $P^\perp x = x$. Also, it is easy to see that $I_n = P + P^\perp$, and hence

$$PAP^\perp x = \lambda Px = 0$$

Therefore $PAP^\perp = 0$. When we write

$$A = \left(P + P^\perp\right) A \left(P + P^\perp\right)$$

in block form

$$A = \left(P + P^\perp\right) A \left(P + P^\perp\right) = \begin{bmatrix} PAP & PAP^\perp \\ P^\perp AP & P^\perp AP^\perp \end{bmatrix}$$

we see that A has the special form

$$A = \begin{bmatrix} PAP & 0 \\ P^\perp AP & P^\perp AP^\perp \end{bmatrix}$$

This particular calculation reveals in a simple way the consequences of zero components to eigenvectors of nonnegative matrices. Zero components of eigenvectors implies zero blocks of A. It would be correct to observe that this argument requires a nonnegative eigenvector. Nonetheless, there are still some very special properties of the remaining eigenvalues of A, particularly those with the modulus equal to the spectral radius. It will be convenient to have a name for the index set where a nonnegative vector $x \in R^n$ is strictly positive.

Definition 10.1.2. Let $x \in R^n$ be nonnegative, $x \geq 0$. Let S be the index set such that $x_i \neq 0$ if $i \in S$. We call S the *support* of the vector x.

10.2 General Theory

Definition 10.2.1. For $A \in M_n(R)$ and $\lambda \notin \sigma(A)$ we define the *resolvent* of A by

$$R(\lambda) = (\lambda I - A)^{-1}.$$

This function behaves much like a rational function in complex variable theory. It has poles at each $\lambda \in \sigma(A)$. The order of the pole is the same as the order of the zero of λ in the minimal polynomial of A.

Theorem 10.2.1. *Suppose $A \in M_n(R)$ is nonnegative and $\lambda > \rho(A)$, then $R(\lambda) = (\lambda I - A)^{-1} \geq 0$.*

Proof. Apply the Neumann series

$$(\star) \qquad (\lambda I - A)^{-1} = \sum_{k=0}^{\infty} \lambda^{-(k+1)} A^k.$$

Since $\lambda > 0$ and $A \geq 0$, it is apparent that $R(\lambda) \geq 0$. □

Remark 10.2.1. It is apparent that $(\star)$ holds upon noting that

$$(\lambda I - A)^{-1} \left(1 - \left(\frac{A}{\lambda}\right)^{n+1}\right) = \sum_{0}^{n} \lambda^{-(k+1)} A^k$$

and that $|\lambda| > \rho(A)$ yields convergence.

Theorem 10.2.2. *Suppose $A \in M_n(R)$ is nonnegative. Then the spectral radius $\rho(A)$ of A is an eigenvalue of A with at least one eigenvector, $x \geq 0$.*

Proof. Suppose for each $y \geq 0$, $R(\lambda)y$ remains bounded as $\lambda \downarrow \rho$. If $x \in \mathbb{C}^n$ is arbitrary

$$\begin{aligned} |R(\lambda)x| &= \left| \sum_{n=0}^{\infty} \lambda^{-(n+1)} A^n x \right| \\ &\leq \sum_{n=0}^{\infty} |\lambda|^{-(n+1)} A^n |x| \\ &= R(|\lambda|)|x| \end{aligned}$$

for all $\lambda, |\lambda| > \rho$. It follows that $R(\lambda)x$ is uniformly bounded in the region $|\lambda| > \rho$, and this is impossible.

Now let $y_0 \geq 0$ be a vector for which $R(\lambda)y_0$ is unbounded as $\lambda \downarrow p$, and let $\| \;\; \|$ denote a vector norm. For $\lambda > r$, set

$$z(\lambda) = R(\lambda)y_0 / \|R(\lambda)y_0\|,$$

where $\|R(\lambda)y_0\| \uparrow \infty$. Now $z(\lambda) \subset \{x \mid \|x\| = 1\}$, and the latter is compact. Therefore $\{z(\lambda)\}_{\lambda > \rho}$ has a cluster point x_0 with $x_0 \geq 0$ and $\|x_0\| = 1$. Since

$$(\rho I - A)z(\lambda) = (\rho - \lambda)z(\lambda) + y_0 / \|R(\lambda)y_0\|$$

and since $\lambda \to \rho$ and $\|R(\lambda)y_0\| \uparrow \infty$ we have

$$\lim_{\lambda\to\rho}(\rho I - A)z(\lambda) = \lim_{\lambda\to\rho}(\rho I - A)x_0 = 0. \qquad \square$$

Corollary 10.2.1. *Suppose $A \in M_n(R)$ is nonnegative and $Az = \lambda z$ for some $z \gg 0$, then $\lambda = \rho(A)$.*

Proof. Since $\rho(A) \subset \sigma(A^T)$ there is a vector $y \geq 0$ for which $A^T y = \rho y$. If $\lambda \neq \rho$, we must have $\langle z, y\rangle = 0$. But $z \gg 0$ and this is impossible. $\square$

Corollary 10.2.2. *Suppose $A \in M_n(R)$ is strictly positive and $z \geq 0$ such that $Az = \rho z$, then $z \gg 0$.*

Proof. If $z_i = 0$, then $\sum\limits_{j=1}^{n} a_{ij}z_j = 0$, and thus $z_j = 0$ for all j since $a_{ij} > 0$, for all j. Hence $z = 0$, a contradiction. $\square$

This result will be strengthened in the next section by another result that has the same conclusion but a much weaker hypotheses.

Theorem 10.2.3. *Suppose $A \in M_n(R)$ is nonnegative with spectral radius $\rho(A)$ and define*

$$r_m = \min_i \sum_j a_{ij} \qquad r_M = \max_i \sum_j a_{ij}$$
$$c_m = \min_j \sum_i a_{ij} \qquad c_M = \max_j \sum_i a_{ij}$$

then both inequalities below hold.

$$r_m \leq \rho(A) \leq r_M$$
$$c_m \leq \rho(A) \leq c_M$$

Moreover, if $A, B \in M_n(R)$ then $\rho(A+B) \geq \max(\rho(A), \rho(B))$.

Proof. We prove the second set of inequalities. Let $x \in R_n$ be the eigenvector pertaining to $\rho(A)$. Since $x \geq 0$ we can normalize x so that $\sum x_i = 1$. Then

$$\sum_{j=1}^{n} a_{ij}x_j = \rho(A)x_i, \qquad i = 1, \ldots, n$$

Sum both side in i to obtain

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_j &= \sum_{i=1}^{n}\rho(A)x_i \\ \sum_{j=1}^{n}\left(\sum_{i=1}^{n} a_{ij}\right)x_j &= \rho(A)\end{aligned}$$

Replacing $\sum_{i=1}^{n} a_{ij}$ by c_M makes the sum on the left larger, that is

$$\rho(A) \le \sum_{j=1}^{n}\left(\sum_{i=1}^{n} a_{ij}\right)x_j \le c_M\sum_{i=1}^{n} x_i = c_M$$

Similarly replacing $\sum_{i=1}^{n} a_{ij}$ by c_m makes the sum on the left smaller, and therefore $\rho(A) \ge c_m$. Thus, $c_m \le \rho(A) \le c_M$. The other inequality, $r_m \le \rho(A) \le r_M$, can be easily established by applying what has just been proved to the matrix A^T noting of course that $\rho(A) = \rho(A^T)$.

The proof that $\rho(A+B) \ge \max(\rho(A), \rho(B))$ is an easy consequence of a Neumann series argument. □

We could just as well prove the first inequality and apply it to the transpose to obtain the second. This proof is given below.

Alternative Proof. Let x be the eigenvector pertaining to $\rho(A)$ and $x_k = \max_i x_i$. Then

$$\rho x_k = \sum_{j=1}^{n} a_{ij}x_j \le \left(\sum_{j=1}^{n} a_{ij}\right)\max x_j.$$

Hence

$$\rho \le \sum_{j=1}^{n} a_{ij} \le \max_k \sum_{j=1}^{n} a_{kj} = t.$$

To obtain $s \le \rho$, let $y \ge 0$ satisfy

$$A^T y = \rho y$$

and suppose $\|y\|_1 = 1$. Then we have on the one hand

$$\langle e, A^T y\rangle = \rho\langle e, y\rangle = \rho$$

and on the other hand applying convexity

$$\begin{aligned}\langle e, A^T y\rangle &= \sum_i \sum_j a_{ij}^T y_j \\ &= \sum_j y_j \left(\sum_i a_{ji}\right) \geq \min_j \sum_i a_{ji} = s\end{aligned}$$

□

Corollary 10.2.1. *(i) If $\rho(A)$ is equal to one of the quantities c_m or c_M, then all of the sums $\sum_i a_{ij}$ are equal for j in the support of the eigenvector x pertaining to ρ.*
(ii) If $\rho(A)$ is equal to one of the quantities r_m or r_M, then all of the sums $\sum_j a_{ij}$ are equal for i in the support of the eigenvector y of A^T pertaining to ρ. (iii) If all the row sums (resp. column sums) are equal, the spectral radius is equal to this value.

Proof. Assume that $\rho(A) = c_M$. Let S denote the support (recall Definition 10.1.2) of the eigenvector x. Assume also that the eigenvector x is normalized so that $\sum_{j=1}^n x_j = \sum_{j\in S} x_j = 1$. Following the proof of Theorem 10.2.3 we have

$$\sum_{j=1}^n \left(\sum_{i=1}^n a_{ij}\right) x_j = \sum_{j\in S} \left(\sum_{i=1}^n a_{ij}\right) x_j = c_M$$

Since $\dfrac{1}{c_M}\left(\sum_{i=1}^n a_{ij}\right) \leq 1$ for each $j = 1, \ldots, n$ it follows that

$$\sum_{j\in S} \frac{1}{c_M}\left(\sum_{i=1}^n a_{ij}\right) x_j \leq \sum_{j\in S} x_j$$

Clearly if for some $j \in S$ we have $\dfrac{1}{c_M}\left(\sum_{i=1}^n a_{ij}\right) < 1$ the inequality above will become a strict inequality, and the result is proved. The result is proved similarly for c_m
(ii) Apply the proof of (i) to A^T. □

10.3 Mean Ergodic Theorem

Let $A \in M_n(R)$. We have seen that understanding the nature of powers A^k, $k = 1, 2, \ldots$ leads to interesting conclusions. For example if $\lim_{k\to\infty} A^k = P$, it is easy to see that P is idempotent ($P^2 = P$) and $AP = PA = P$. It is therefore a projection to the *fixed space* of A. A less restrictive property is *mean convergence*:

$$M_k = k^{-1}(I + A + A^2 + \cdots + A^k)$$
$$\lim_{k\to\infty} M_k = P.$$

Note that if $\rho(A) < 1$ we have $A^k \to 0$. Hence $M_k \to 0$. On the other hand if $\rho(A) > 1$ the M_k become unbounded. Hence mean convergence has value only when $\rho(A) = 1$.

Theorem 10.3.1 (Mean Ergodic Theorem for matrices). *Let $A \in M_n(R)$ with $\rho(A) = 1$. If $\{A^k\}_{k=1}^{\infty}$ is bounded then $\lim_{k\to\infty} M_k = P$, where P is a projection, commuting with A, onto the fixed space of A.*

Proof. We need a norm $\|\cdot\|$, vector and matrix. We have $\|A^k\| < C$ for $k = 1, 2, \ldots$. It is easy to see that

$$(\star) \qquad AM_k - M_k = \frac{1}{k}\left(\sum_0^k A^{j+1}\right) - \frac{1}{k}\left(\sum_0^k A^j\right)$$
$$= \frac{1}{k}(I + A^{k+1})$$
$$\leq \frac{1}{k}(1 + C).$$

Since the matrices M_k are bounded they have a cluster point P. This means there is a subsequence M_{k_j} that converges to P. If there is another cluster point Q (i.e. there is a subsequence $M_{\ell j} \to Q$), then we compare P and Q. First we know

$$\|M_{k_j} - P\| < \frac{\varepsilon}{2(1+C)}$$
$$\|M_{\ell_j} - Q\| < \frac{\varepsilon}{2(1+C)}.$$

From $(\star)$ we have $AP = P$ and $AQ = Q$, whence $M_{k_j}P = P$ and $M_{l_j}Q = Q$ for all $k = 1, 2, \ldots$. Hence

$$P - Q = M_{\ell_j}(M_{k_j} - P) - M_{k_j}(M_{\ell_j} - Q)$$

or

$$\|P - Q\| \leq \varepsilon.$$

Thus $P = Q$ and M_k converges to some matrix P. We have $AP = PA = P$ hence $M_k P = P$ for all k and therefore $P^2 = P$. Clearly the range of P consists of all fixed vectors under A. □

Corollary 10.3.1. *Let $A \in M_n(R)$ and suppose that $\lim_{k\to\infty} A^k = P$, then $\lim_{k\to\infty} M_k$ exists and equals P.*

Definition 10.3.1. Let $A \in M_n(R)$ have $\rho(A) = 1$. Then A is said to be *mean ergodic* if

$$\lim k^{-1}(I + A + \cdots + A^k)$$

exists.

Clearly, if $A \in M_n(R)$ satisfies $\rho(A) = 1$ and A^k is bounded, then A is mean ergodic. (Previous theorem.) Conversely, if A is mean ergodic then the sequence A^k is bounded. This can be proved by resorting to the Jordan canonical form and showing A^k is bounded if and only if A_j^k is bounded, where A_J is the Jordan canonical form of A.

Lemma 10.3.1. *Let $A \in M_n(R)$ with $\rho(A) = 1$. Suppose $\lambda = 1$ is a simple pole of the resolvent, and the only eigenvalue of modulus 1. Then $A = P + B$ where $P = \lim_{\lambda\to 1}(\lambda - 1)R(\lambda)$ is a projection onto the fixed space of A, $PB = BP = 0$ and $\rho(B) < 1$.*

Proof. We know that P exists and using the Neumann series it is clear that $AP = PA = P$

$$\begin{aligned} A \lim_{\lambda\to 1}(\lambda - 1)(\lambda I - A)^{-1} &= A(\lambda - 1)\sum_0^\infty \lambda^{-(k+1)} A^k \\ &= \lim_{\lambda\to 1}(\lambda - 1)\sum_0^\infty \lambda^{-(k+1)} A^{k+1} \\ &= \lim_{\lambda\to 1}(\lambda - 1)\lambda\left[\sum_0 \lambda^{-(k+1)} A^k - \lambda^{-1}\right] \\ &= \lim_{\lambda\to 1}(\lambda - 1)R(\lambda) = P. \end{aligned}$$

That is, $AP = P$. The other assertions follow similarly. Also, it follows that

$$P^2 = \lim_{\lambda\to 1} PR(\lambda) = P$$

which is to say that P is a projection. We have, as well, that $Ax = x \Rightarrow Px = x$, again using the Neumann series.

Now define $B = A - P$. Then $BP = PB = 0$. Suppose $Bx = ax$, where $|\alpha| \geq 1$. Then $Px = 0$, for

$$\alpha Px = PBx = 0.$$

Therefore $Ax = \alpha x$. Here $\alpha \neq 1$ implies $x = 0$ by hypothesis and $\alpha = 1$ implies $Px = x$, or $x = 0$, also. This contradicts the assumption that $|\alpha| \geq 1$. □

Theorem 10.3.2. *Let $A \in M_n(R)$ satisfy $A \geq 0$ and $\rho(A) = 1$. The following are equivalent:*

(a) $\lambda = 1$ is a simple pole of $R(\lambda)$.

(b) $\{A^k\}_1^\infty$ is bounded.

(c) A is mean ergodic.

Moreover

$$\lim_{\lambda \to 1} (\lambda - 1) R(\lambda) = \lim_{k \to \infty} \frac{1}{k} (I + A + \cdots + A^k)$$

if either limit exists.

10.4 Irreducible Matrices

Irreducible matrices form one of the cornerstones of positive matrix theory. Because the condition of irreducibility is both natural and often satisfied, their applications are far and wide.

Definition 10.4.1. Let $A \in M_n(\mathbb{C})$. The **zero pattern** of A is the set of ordered pairs $S = \{(i, j) \mid a_{ij} = 0\}$.

Example 10.4.1. For example with

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 0 \\ 2 & 3 & 0 \end{bmatrix}$$

$S = \{(1,1), (2,2), (2,3), (3,3)\}$.

Definition 10.4.2. A generalized permutation matrix B is any matrix having the same zero patter as a permutation matrix.

This means of course that a generalized permutation matrix has exactly one nonzero entry in each row and one nonzero entry in each column. Generalized permutation matrices are those nonnegative matrices with nonnegative inverses.

Theorem 10.4.1. *Let $A \in M_n(R)$ be invertible and $A \geq 0$. Then A is invertible with nonnegative inverse if and only if A is a generalized permutation matrix.*

Proof. First, if A is a generalized permutation matrix, define the matrix B by

$$b_{ij} = \begin{cases} 0 & \text{if} \quad a_{ij} = 0 \\ \dfrac{1}{a_{ij}} & \text{if} \quad a_{ij} \neq 0 \end{cases}$$

Then $A^{-1} = B^T \geq 0$. On the other hand suppose $A^{-1} \geq 0$ and A has on some row two non zero entries, say a_{i,j_1} and a_{i,j_2}. Since $A^{-1} \geq 0$ is invertible there is a nonzero entry in the i^{th} column, say $a^{-1}_{ki} > 0$. Now compute multiply $A^{-1}A$. It is easy to see that

$$\left(A^{-1}A\right)_{k,j_1} > 0 \quad \text{and} \quad \left(A^{-1}A\right)_{k,j_2} > 0$$

which contradicts that $A^{-1}A = I$. □

Example 10.4.1. The analysis for 2×2 matrices can be carried out directly. Suppose that

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{and} \quad A^{-1} = \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

There are two cases: (1) $\det A > 0$. In this case we conclude that $b = c = 0$. Thus A is a generalized permutation matrix. (2) $\det A < 0$. In this case we conclude that $a = d = 0$. Again A is a generalized permutation matrix. As these are the only two cases, the result is verified.

Definition 10.4.3. Let $A,\ B \in M_n(R)$. We say that A is **cogredient** to B if there is a permutation matrix P such that

$$B = P^T AP$$

Note that two cogredient matrices are similar, indeed they are unitarily similar for the very special class of unitary matrices given by permutations. It is important to note that since AP merely interchanges the columns of A and $P^T AP$ then interchanges the rows of AP. From this we observe that both A and $B = P^T AP$ have exactly the same elements, though permuted.

Definition 10.4.4. Let $A \in M_n(R)$. We say that A is **irreducible** if it is not cogredient to a matrix of the form

$$\left[\begin{array}{c|c} A_1 & 0 \\ \hline A_3 & A_4 \end{array}\right]$$

where the block matrices A_1 and A_4 are square matrices. Otherwise the matrix is called **reducible.**

Example 10.4.2. All diagonal matrices are reducible.

There is another way to describe irreducible (and reducible) matrices in terms of projections. Let $S = \{j_1, \dots, j_k\} \subset \{1, 2, \dots, n\}$ and define P_S to be the projection to the coordinate directions by

$$P_S \; e_i = \begin{cases} 1 & if \quad i \in S \\ 0 & if \quad i \notin S \end{cases}$$

Furthermore define

$$P_S^{\perp} = I - P_S$$

Then P_S and $P_S^{\perp}$ are orthogonal projections. That is, the product $P_S P_S^{\perp} = 0$ and by definition $P_S + P_S^{\perp} = I$. If S^C denotes the complement of S in $\{1, 2, \dots, n\}$, it is easy to see that $P_S^{\perp} = P_{S^C}$

Proposition 10.4.1. *Let $A \in M_n(R)$. Then A is irreducible if and only if there is no set of indices $S = \{j_1, \dots, j_k\} \subset \{1, 2, \dots, n\}$ for which $P_S A P_S^{\perp} = 0$.*

Proof. Suppose there is a set of indices $S = \{j_1, \dots, j_k\} \subset \{1, 2, \dots, n\}$ for which $P_S A P_S^{\perp} = 0$. Define the permutation matrix as from the permutation that takes the first k integers $1, \dots, k$ to the integers $j_1, \dots, j_k$ and the integers $k+1, \dots, n$ to the complement S^C. Then

$$P^T AP = \left[\begin{array}{c|c} A_1 & A_2 \\ \hline A_3 & A_4 \end{array}\right]$$

The entries in the block A_2 are those with rows given from the rows corresponding to the indices in S and columns corresponding to the indices in S^C. Thus A is reducible. Conversely, suppose that A is reducible and that P is a permutation matrix such that

$$P^T AP = \left[\begin{array}{c|c} A_1 & 0 \\ \hline A_3 & A_4 \end{array}\right]$$

For definiteness, let us assume that A_1 is $k \times k$ and A_4 is $(n-k) \times (n-k)$. Define $S = \{j_i \mid p_{j_i,i} = 1,\ i = 1, \ldots, k\}$. and $T = \{j_i \mid p_{j_i,i} = 1,\ i = k+1, \ldots, n\}$ Because P is a permutation, it follows that $T = S^C$ and $P_S A P_S^{\perp} = 0$, which proves the converse. □

As we know for a nonnegative matrix $A \in M_n(R)$ the spectral radius is an eigenvalue of A with pertaining eigenvector $x \geq 0$. When the additional assumption of irreducibility of A is added the conclusion can be strengthened to $x \gg 0$. To prove this result we establish a simple result from which the $x \gg 0$ follows almost directly.

Theorem 10.4.2. *Let $A \in M_n(R)$ be nonnegative and irreducible. Suppose that $y \in R_n$ is nonnegative with exactly $1 \leq k \leq n-1$ nonzero entries. Then $(I + A)\, y$ has strictly more nonzero entries.*

Proof. Suppose the nonzero entries of y are $S = \{j_1, \ldots, j_k\}$. Since $(I + A)\, y = y + Ay$, it follows immediately that $((I + A)\, y)_{j_i} \neq 0$ for $j_i \in S$, and therefore there are at least as many nonzero entries in $(I + A)\, y$ as there are in y. In order that the number of nonzero entries not increase, we must have for each index $i \notin S$ that $(Ay)_i = 0$. With P_S defined as the projection to the standard coordinates with indices from S we conclude therefore that $P_S^{\perp} AP = 0$, which means that A is reducible, a contradiction. □

Corollary 10.4.1. *Let $A \in M_n(R)$ be nonnegative. Then A is irreducible if and only if $(I + A)^{n-1} > 0$.*

Proof. Suppose that A is irreducible and $y \geq 0$ is any vector in R_n. Then $(I + A)\, y$ has strictly more nonzero coordinates and thus $(I + A)^2\, y$ has even more nonzero coordinates. We see that the number of nonzero coordinates of $(I + A)^k\, y$ must increase by at least one until the maximum of n nonzero coordinates is reached. This must occur by the $(n-1)^{\text{th}}$ power. Now apply this to the vectors of the standard basis $e_1, \ldots, e_n$. For example, $(I + A)^{n-1}\, e_k \gg 0$. This means that the k^{th} column of $(I + A)^{n-1}$ is strictly positive. The result follows.

Suppose conversely that $(I+A)^{n-1} > 0$ but that A is reducible. Then there is a permutation matrix P such that $P^T AP$ has the form

$$P^T AP = \left[\begin{array}{c|c} A_1 & 0 \\ \hline A_3 & A_4 \end{array}\right]$$

It is easy to see that all the powers of A have the same form — with respect to the same blocks. Also, $(I+A)^{n-1} = I + \sum_{i=1}^{n-1} \binom{n-1}{i} A^i$. So

$$\begin{aligned} P^T (I+A)^{n-1} P &= P^T \left(I + \sum_{i=1}^{n-1} \binom{n-1}{i} A^i \right) P \\ &= I + \sum_{i=1}^{n-1} \binom{n-1}{i} P^T A^i P^T \end{aligned}$$

has the same form

$$P^T (I+A)^{n-1} P = \left[\begin{array}{c|c} \tilde{A}_1 & 0 \\ \hline \tilde{A}_3 & \tilde{A}_4 \end{array}\right]$$

and this proves the result. □

Corollary 10.4.2. *Let $A \in M_n(R)$ be nonnegative. Then A is irreducible if and only if for each pair of indices (i,j) there a positive power $1 \leq k \leq n$ such that $\left(A^k\right)_{ij} > 0$.*

Proof. Suppose that A is irreducible. Since $(I+A)^{n-1} = I + \sum_{i=1}^{n-1} \binom{n-1}{i} A^i > 0$, it follows that $A(I+A)^{n-1} = A + \sum_{i=1}^{n-1} \binom{n-1}{i} A^{i+1} > 0$. Thus for each pair of indices (i,j), $\left(A^k\right)_{ij} > 0$ for some k.

On the other hand suppose that A is reducible. Then there is a permutation matrix P for which

$$P^T AP = \left[\begin{array}{c|c} A_1 & 0 \\ \hline A_3 & A_4 \end{array}\right]$$

Follow the same argument as above to establish that $A(I+A)^{n-1}$ must have the same form as $P^T AP$ with the same zero pattern. Thus there is no positive power $1 \leq k \leq n$ such that $\left(A^k\right)_{ij} > 0$ for any of the pairs of indices corresponding to the (upper right) zero block above. □

Another characterization of irreducibility arises when we have $Ax \leq ax$ for some nonzero vector $x \geq 0$. This type of domination-type condition appears to be quite weak. Nonetheless, it is equivalent to the others.

Corollary 10.4.3. *Let $A \in M_n(R)$ be nonnegative. Then A is irreducible if and only if $Ax \leq ax$ for some nonzero vector $x \geq 0$, then $x \gg 0$.*

Proof. If $x_i = 0$, then $a_{ij} = 0$ whenever $x_j \neq 0$. Define $S = \{1 \leq j \leq n \mid x_j \neq 0\}$. Then with P_S as the orthogonal projection to the standard vectors e_i, $i \in S$, it must follow that $P_S A P_S^{\perp} = 0$. Since S is strictly contained in $\{1, 2, \dots, n\}$, it follows that A is reducible.

Now suppose that A is reducible, which means we can assume it has the form

$$A = \left[\begin{array}{c|c} A_1 & 0 \\ \hline A_3 & A_4 \end{array}\right]$$

For definiteness, let us assume that A_1 is $k \times k$ and A_4 is $(n-k) \times (n-k)$ and of course $k \geq 1$. Let $v \in R_{n-k}$ be the nonzero nonnegative vector which satisfies $A_4 v = \rho(A_4)\, v$. Let $u = 0 \in R_k$. Define the vector $x = u \oplus v$. Then

$$(Ax)_i = \begin{cases} (A_1 u)_i = 0 = \rho(A_4)\, u_i & if \quad 1 \leq i \leq k \\ (A_1 v)_i = \rho(A_4)\, v_i & if \quad k+1 \leq i \leq n \end{cases}$$

The conditions of that the hypothesis are met without $x \gg 0$. ☐

Corollary 10.4.4 (Subinvariance). *Let $A \in M_n(R)$ be nonnegative and irreducible with spectral radius ρ. Suppose that there is a positive number s and nonnegative vector $z \geq 0$ for which*

$$(Az)_i \leq s z_i, \qquad for \quad i = 1, 2, \dots, n$$

Then (i) $z \gg 0$, and (ii) $\rho \leq s$. If in addition $\rho = s$, then $Az = \rho z$.

Proof. Clearly, the same inequality holds for powers of A. For $A(Az) \leq Az \leq sz$. Now apply induction. Next suppose that $z_i = 0$. By Corollary 10.4.2 we must have for each index pair i, j that $\left(A^k\right)_{ij} > 0$ for some positive integer k making $\left(A^k z\right)_i \leq s z_i$ impossible for some k. Thus $z \gg 0$. To prove the second assertion, let x be the eigenvector of A^T pertaining ρ. Then

$$\langle Az, x\rangle = \left\langle z, A^T x\right\rangle = \rho \langle z, x\rangle \leq s \langle z, x\rangle \tag{1}$$

Since $x \gg 0$, the innerproduct $\langle z, x\rangle > 0$. Hence $\rho \leq s$.

Finally, if $\rho = s$ and $(Az)_i < \rho z_i$ for some index i then we conclude from (1) that $\rho < \rho$, a contradiction. ☐

Theorem 10.4.3. *Let $A \in M_n(R)$ be nonnegative and irreducible. If $x \geq 0$ is the eigenvector pertaining to $\rho(A)$, i.e. $Ax = \rho(A)x$, then $x \gg 0$.*

Proof. We have that $(I + A)$ is also nonnegative and irreducible, with spectral radius $1 + \rho(A)$ and pertaining eigenvector x. Since

$$(I + A)^{n-1} x > (1 + \rho(A))^{n-1} x \gg 0$$

it follows that $x \gg 0$. □

Corollary 10.4.5. *(i) If $\rho(A)$ is equal to one of the quantities c_m or c_M from Theorem 10.2.3, then all of the sums $\sum_i a_{ij}$ are equal for $j = 1, \ldots, n$. (ii) If $\rho(A)$ is equal to one of the quantities r_m or r_M from Theorem 10.2.3, then all of the sums $\sum_j a_{ij}$ are equal for $i = 1, \ldots, n$. (iii) Conversely, if all the row sums (resp. column sums) are equal, the spectral radius is equal to this value.*

Proof. Both are direct consequences of Corollary 10.2.1 and Theorem 10.4.3 which imply that the eigenvectors (of A and A^T, resp.) pertaining to ρ are strictly positive. □

10.4.1 Sharper estimates of the maximal eigenvalue

Sharper estimates for the maximal eigenvalue (spectral radius) are possible. The following results assembles much of what we have considered above. We begin with a basic inequality that has an interesting geometric interpretation.

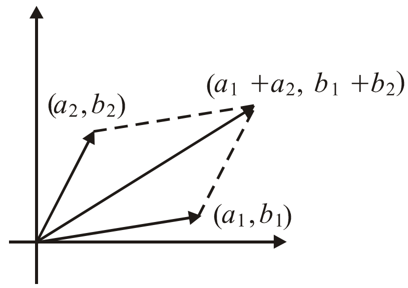
Lemma 10.4.1. *Let a_i, b_i, $i = 1, \ldots n$ be two positive sequences. Then*

$$\min_j \frac{b_j}{a_j} \leq \frac{\sum_{j=1}^n b_j}{\sum_{j=1}^n a_j} \leq \max_j \frac{b_j}{a_j}$$

Proof. We proceed by induction. The result is trivial for $n = 1$. For $n = 2$ the result is a simple consequence of vector addition as follows. Consider the ordered pairs (a_i, b_i), $i = 1, 2$. Their sum is $(a_1 + a_2,\ b_1 + b_2)$. It is a simple matter to see that the slope of this vector must lie between the slopes of the summands, which is to say

$$\min_j \frac{b_j}{a_j} \leq \frac{\sum_{j=1}^2 b_j}{\sum_{j=1}^2 a_j} \leq \max_j \frac{b_j}{a_j}$$

as is shown below.



Now assume by induction the result holds up to $n-1$. Then we must have that

$$\begin{aligned}\frac{\sum_{j=1}^{n} b_j}{\sum_{j=1}^{n} a_j} &= \frac{\sum_{j=1}^{n-1} b_j + b_n}{\sum_{j=1}^{n-1} a_j + a_n} \leq \max\left[\frac{\sum_{j=1}^{n-1} b_j}{\sum_{j=1}^{n-1} a_j}, \frac{b_n}{a_n}\right] \\ &\leq \max\left[\max_{1\leq j\leq n-1} \frac{b_j}{a_j}, \frac{b_n}{a_n}\right] = \max_{1\leq j\leq n} \frac{b_j}{a_j}\end{aligned}$$

A similar argument serves to establish the other inequality. □

Theorem 10.4.4. *Let $A \in M_n(R)$ be nonnegative with row sums r_k and spectral radius ρ. Then*

$$\min_k \left[\frac{1}{r_k}\sum_{j=1}^{n} a_{kj} r_j\right] \leq \rho \leq \max_k \left[\frac{1}{r_k}\sum_{j=1}^{n} a_{kj} r_j\right] \tag{2}$$

Proof. First assume that A is irreducible. Let $x \in R^n$ be the principal nonnegative eigenvector of A^T, so $A^T x = \rho x$ and $x \gg 0$. Assume also that x has been normalized so that $\sum x_i = 1$. By summing both sides of $A^T x = \rho x$, we see that $\sum r_k x_k = \rho$. Since ρ^2 is the spectral radius of A^2, we also have that $\left(A^T\right)^2 x = \left(A^2\right)^T x = \rho^2 x$. Summing both sides, it follows that $\sum_{k=1}^{n} \sum_{j=1}^{n} r_j a_{jk}^T x_k = \rho^2$. Now

$$\rho = \frac{\rho^2}{\rho} = \frac{\sum_{k=1}^{n} \sum_{j=1}^{n} r_j a_{jk}^T x_k}{r_k x_k} = \max_k \frac{\sum_{j=1}^{n} a_{kj} r_j}{r_k}$$

by Lemma 10.4.1. The reverse inequality is proved similarly. In the case that A is not irreducible, we take the limit $A_k \downarrow A$ where the A_k are irreducible matrices. The result holds for each A_k, and the quantities in the inequality are continuous in the limiting process. Some small care must be taken in the case that A has a zero row. □

Of course, we have not yet established that the estimates (2) are sharper that the basic inequality $\min_k r_k \leq \rho \leq \max_k r_k$. That is the content of following result, whose proof is left as an exercise.

Corollary 10.4.6. *Let* $A \in M_n(R)$ *be nonnegative with row sums* r_k. *Then*

$$\min_k r_k \leq \min_k \left[\frac{1}{r_k} \sum_{j=1}^{n} a_{kj} r_j \right] \leq \rho \leq \max_k \left[\frac{1}{r_k} \sum_{j=1}^{n} a_{kj} r_j \right] \leq \max_k r_k$$

We can continue this kind of argument even further. Let $A \in M_n(R)$ be nonnegative with row sums r_i and let x be the nonnegative eigenvector pertaining to ρ. Then $\left(A^T\right)^3 x = \rho^3 x$ or expanded

$$\sum_{m,k,j} a_{ij}^T a_{jk}^T a_{km}^T x_m = \rho^3 x_i$$

Summing both sides yields

$$\sum_{m,k,j} r_j a_{jk}^T a_{km}^T x_m = \rho^3$$

Therefore,

$$\begin{aligned} \rho &= \frac{\rho^3}{\rho^2} = \frac{\sum_m \sum_k \sum_j r_j a_{jk}^T a_{km}^T x_m}{\sum_m \sum_j r_j a_{jm}^T x_m} \\ &\leq \max_m \frac{\sum_k \sum_j r_j a_{jk}^T a_{km}^T}{\sum_j r_j a_{jm}^T} = \max_m \frac{\sum_k a_{mk} \sum_j a_{kj} r_j}{\sum_j a_{mj} r_j} \end{aligned}$$

by Lemma 10.4.1, thus yielding another estimate for the spectral radius. The estimate is two-sided as those above. This is summarized in the following result.

Theorem 10.4.5. *Let* $A \in M_n(R)$ *be nonnegative with row sums* r_i. *Then*

$$\min_m \frac{\sum_k a_{mk} \sum_j a_{kj} r_j}{\sum_j a_{mj} r_j} \leq \rho \leq \max_m \frac{\sum_k a_{mk} \sum_j a_{kj} r_j}{\sum_j a_{mj} r_j} \qquad (3)$$

Remember the complete proof as developed above does require the assumption irreducibility and the passage of the limit. Let's consider an example and see what these estimates provide.

Example 10.4.3. Let

$$A = \begin{bmatrix} 1 & 3 & 3 \\ 5 & 3 & 5 \\ 1 & 1 & 4 \end{bmatrix}$$

The eigenvalues of A are -2, 2, and 8. Thus $\rho = 8$. The row sums are $r_1 = 7$, $r_2 = 13$, $r_3 = 6$, which yields the estimates $6 \le \rho \le 13$. The estimates from (2) are $\frac{64}{7}$, 8, and $\frac{22}{3}$ giving the estimates

$$7.333333333 \le \rho \le 9.142857143$$

Finally, from (3) the estimates for ρ are

$$7.818181818 \le \rho \le 8.192307692$$

It remains to show that the estimates in (3) furnish an improvement to the estimates in (2). This is furnished by the following corollary, which is left as an exercise.

Corollary 10.4.7. *Let $A \in M_n(R)$ be nonnegative with row sums r_i. Then for each $m = 1, \ldots, n$*

$$\min_k \left[\frac{1}{r_k} \sum_{j=1}^{n} a_{kj} r_j \right] \le \frac{\sum_k a_{mk} \sum_j a_{kj} r_j}{\sum_j a_{mj} r_j} \le \max_k \left[\frac{1}{r_k} \sum_{j=1}^{n} a_{kj} r_j \right] \qquad (4)$$

These results are by no means the end of the story on the important subject of eigenvalue estimation.

10.5 Stochastic Matrices

Definition 10.5.1. A matrix $A \in M_n(R)$, $A \ge 0$, is called (row) *stochastic* if each row sum of A is 1 or equivalently $Ae = e$. (Recall, $e = (1, 1, \ldots, 1)^T$.) Similarly, a matrix $A \in M_n(R)$, $A \ge 0$, is called column stochastic if A^T is stochastic.

A direct consequence of Corollary 10.2.1(iii) is that the spectral radius of stochastic matrices must be one.

Theorem 10.5.1. *Let $A \in M_n(R)$ row or column stochastic. Then the spectral radius of A is $\rho(A) = 1$.*

Lemma 10.5.1. *Let $A,\ B \in M_n(R)$ be nonnegative row stochastic matrices. Then for any number $0 \le c \le 1$, the matrices $cA + (1-c)B$ and AB are also row stochastic. The same result holds for column stochastic matrices.*

Theorem 10.5.2. *Every stochastic matrix is mean ergodic, and the peripheral spectrum is fully cyclic.*

Proof. If $A \ge 0$ is stochastic, so also is A^k, $k = 1, 2, \ldots$. Therefore A^k is bounded. Also $\rho(A) \le \max_i \sum_k a_{ij} = 1$. Hence the conditions of the previous theorems are met. This proves A is mean ergodic. □

Stochastic matrices—as a set—have some interesting properties. Let $S_n \subset M_n(R)$ of stochastic matrices. Then S_n is a convex set, it is also closed. Whenever a closed convex set is determined, the characterization of its extreme points is desired.

Recall, the (matrix) point $A \in S_n$ is called an *extreme point* if whenever

$$A = \Sigma \lambda_j A_j$$

$\Sigma\lambda_j = 1$, $\lambda_j \ge 0$ and $A_j \in S_n$ it follows that one of the A_j's equals A and all λ_j but one are 0.

Examples. We can also view $S_n \subset R^{n^2}$. It is a convex subset of R^{n^2} and is moreover the intersection of the positive orthant of R^{n^2} with the n hyperplanes defined by $\sum_j a_{ij} = 1$, $i = 1, \ldots, n$. [Note that we are viewing a vector $x \in R^{n^2}$ as

$$x = (\alpha_{11}, \ldots, \alpha_{1n}, \alpha_{21}, \ldots, \alpha_{2n}, \ldots, \alpha_{n1}, \ldots, \alpha_{nn})^T]$$

S_n is therefore a *convex polyhedron* in R^{n^2}. The dimension of S_n is $n^2 - n$.

Theorem 10.5.3. *$A \in S_n$ is an extreme point of S_n if and only if A has exactly one 1 in each row, and all other entries are zero.*

Proof. Let $C = c_{ij}$ be a (0,1)-matrix. (That is a matrix $c_{ij} = \begin{cases} 1 \\ 0 \end{cases}$ or for all $1 \le i, j \le n$.) Suppose that $C \in S_n$ then there is a unique j for which $c_{ij} = 1$. If $C = \lambda A + (1-\lambda)B$ it is easy to see that $a_{1j} = b_{1j} = 1$ and moreover that $a_{1k} = b_{1k} = 0$ for all other $k \ne j$. It follows that C is an extreme point. Conversely suppose $C \in S_n$ is *not* a (0,1) matrix. Fix one row. Let

$$A_j = \begin{bmatrix} e_j & \\ c_2 & \rightarrow \\ \vdots & \\ c_n & \rightarrow \end{bmatrix}$$

where $e_j = (0, 0, \underset{j^{\text{th}}\text{position}}{\ldots, 1}, 0 \ldots 0)$ and $c_2 \ldots c_n$ are the respective rows of C. Then

$$C = \sum_1^n C_{1j} A_j.$$

If C_1 is not a (0,1) row we are finished, since C cannot be an extreme point. Otherwise select any non (0,1) row and repeat this argument with respect to that row. □

Corollary 10.5.1. *Permutation matrices are extreme points.*

Definition 10.5.2. A *lattice homomorphism* of $\mathbb{C}^n$ is a linear map $A\colon\ \mathbb{C}^n \rightarrow \mathbb{C}^n$ such that $|Ax| = A|x|$, for all $x \in \mathbb{C}^n$. A is a lattice isomorphism if both A and A^{-1} are lattice homomorphisms.

Theorem 10.5.4. *$A \in S_n$ is a lattice homomorphism if and only if A is an extreme point of S_n. $A \in S_n$ is a lattice isomorphism if and only if A is a permutation matrix.*

Theorem 10.5.5. *$A \subset S_n$ is a permutation matrix if and only if $\sigma(A) \subset \Gamma$.*

$\Gamma = \{\lambda \in \mathbb{C} \mid |\lambda| = 1\}$.**Exercise.**

1. Show that any row stochastic matrix (i.e. nonnegative entries with row sums are equal 1) with at least one column having equal entries is singular.

2. Show that the only nonnegative matrices with nonnegative inverses must be diagonal matrices.

3. Adapt the argument from Section 10.1 to prove that the nonnegative eigenvector x pertaining to the spectral radius of an irreducible matrix must be strictly positive.

4. Suppose that $A, B \in M_n(R)$ are nonnegative matrices. Prove or disprove

   (a) If A is irreducible then A^T is irreducible.

   (b) If A is irreducible then A^p is irreducible, for every positive power $p \geq 1$.

   (c) If A^2 is irreducible then A is irreducible.

   (d) If A, B are irreducible, then AB is irreducible.

   (e) If A, B are irreducible, then $aA + bB$ is irreducible for all nonnegative constants $a, b \geq 0$.

5. Suppose that $A, B \in M_n(R)$ are nonnegative matrices. Prove that if A is irreducible, then $aA + bB$ is irreducible for all nonnegative constants $a, b > 0$.

6. Suppose that $A \in M_n(R)$ is nonnegative and irreducible and that the trace of A is positive, $\mathrm{tr}A > 0$. Prove that $A^k > 0$ for some sufficiently large integer k.

7. Prove Corollary 10.2.1(ii) for $\rho(A) = r_m$.

8. Let $A \in M_n(R)$ be nonnegative with row sums $r_i(A)$ and column sums $c_i(A)$. Show that $\sum_{i=1}^n r_i(A^2) = \sum_{i=1}^n c_i(A)\, r_i(A)$.

9. Prove Corollary 10.4.6.

10. Prove Corollary 10.4.7. (Hint. Use Lemma 10.4.1.)

11. Let p be any positive integer and suppose that $A \in M_n(R)$ is nonnegative with row sums r_i. Define $r = (r_{1,\ldots,}r_n)^T$. Recall that $(A^p r)_m$ refers to the m^{th} coordinate of the vector $A^p r$. Show that

$$\min_m \frac{(A^p r)_m}{(A^{p-1} r)_m} \leq \rho \leq \max_m \frac{(A^p r)_m}{(A^{p-1} r)_m}$$

   (Hint. Assume first that A is irreducible.)

12. Use the estimates (2) and (3) to estimate the maximal eigenvalue of

$$A = \begin{bmatrix} 5 & 1 & 4 \\ 3 & 1 & 5 \\ 1 & 2 & 2 \end{bmatrix}$$

   The maximal eigenvalue is approximately 7.640246936.

13. Use the estimates (2) and (3) to estimate the maximal eigenvalue of

$$A = \begin{bmatrix} 2 & 1 & 4 & 4 \\ 5 & 7 & 3 & 2 \\ 5 & 0 & 6 & 0 \\ 5 & 5 & 5 & 7 \end{bmatrix}$$

The maximal eigenvalue is approximately 14.34259731.

14. Suppose that $A \in M_n(R)$ is nonnegative and irreducible with spectral radius ρ, and suppose $x \in R^n$ is any strictly positive vector, $x \gg 0$. Define $\tau = \max_i \frac{(Ax)_i}{x_i}$. Show that $\rho \le \tau$.

15.

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \sigma(A) = \{1\} \qquad C = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \sigma(B) = \{-1, 1\} \qquad \rho_C(\lambda) = \lambda^3 + 1 = 0$$

$$\lambda = -1, e^{i\pi/3}, e^{-i\pi/3}$$