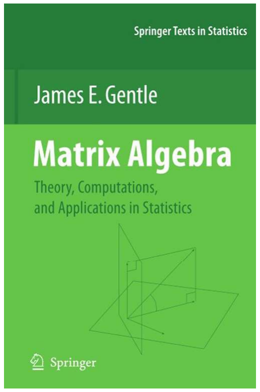
Theory of Matrices

Chapter 3. Basic Properties of Matrices
3.9. Matrix Norm—Proofs of Theorems



Theorem 3.9.2

Theorem 3.9.2. For $n \times m$ matrix $A = [a_{ij}]$, the L^1 norm satisfies $\|A\|_1 = \max_{1 \le j \le m} \{\sum_{i=1}^n |a_{ij}|\}$ and so it is also called the *column-sum norm*. The L^∞ norm satisfies $\|A\|_\infty = \max_{1 \le i \le n} \left\{\sum_{j=1}^m |a_{ij}|\right\}$ and so it is also called the *row-sum norm*.

Proof. (This proof is based on Horn and Johnson's *Matrix Analysis*, Cambridge University Press, 1985). Let the columns of A be $a_1, a_2, \ldots, a_m$ so that $a_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{bmatrix}$. Then for any $x = [x_1, x_2, \ldots, x_m]^T \in \mathbb{R}^m$ where $\|x\|_1 = 1$ we have

$$\begin{aligned} \|Ax\|_1 &= \|x_1 a_1 + x_2 a_2 + \cdots + x_m a_m\|_1 \\ &\le \sum_{i=1}^m \|x_i a_i\|_1 \text{ by the Triangle Inequality for } \|\cdot\|_1 \end{aligned}$$

Theorem 3.9.2 (continued 1)

Proof (continued).

$$\begin{aligned} \|Ax\|_1 &= \sum_{i=1}^m |x_i| \|a_i\|_1 \text{ by (2) of the definition of matrix norm} \\ &\le \sum_{i=1}^m |x_i| \max_{1 \le j \le m} \|a_j\|_1 = \max_{1 \le j \le m} \|a_j\|_1 \sum_{i=1}^m |x_i| \\ &= \|x\|_1 \max_{1 \le j \le m} \|a_j\|_1 = \max_{1 \le j \le m} \|a_j\|_1 = \max_{1 \le j \le m} \left\{ \sum_{i=1}^n |a_{ij}| \right\}. \end{aligned}$$

So $\max_{\|x\|_1 = 1} \|Ax\|_1 \le \max_{1 \le j \le m} \{\sum_{i=1}^n |a_{ij}|\}$. Let k satisfy $1 \le k \le m$ with $\max_{1 \le j \le m} \|a_j\|_1 = \|a_k\|_1$. Then with $x = e_k$ (the kth standard basis vector for $\mathbb{R}^m$) we have

$$\|Ax\|_1 = \|Ae_k\|_1 = \|a_k\|_1 = \max_{1 \le j \le m} \|a_j\|_1 = \max_{1 \le j \le m} \left\{ \sum_{i=1}^n |a_{ij}| \right\}.$$

Theorem 3.9.2 (continued 2)

Proof (continued). Therefore, $\|A\|_1 = \max_{\|x\|_1 = 1} \|Ax\|_1$ $= \max_{1 \le j \le m} \{\sum_{i=1}^n |a_{ij}|\}$, as claimed. Let the rows of A be $b_1, b_2, \ldots, b_n$, so $b_i = [a_{i1}, a_{i2}, \ldots, a_{im}]$. Then for any $x = [x_1, x_2, \ldots, x_m]^T \in \mathbb{R}^m$ where $\|x\|_\infty = 1$ we have

$$\begin{aligned} \|Ax\|_\infty &= \max_{1 \le i \le n} |b_i x| = \max_{1 \le i \le n} |\langle b_i^T, x \rangle| = \max_{1 \le i \le n} \left| \sum_{j=1}^m a_{ij} x_j \right| \\ &\le \max_{1 \le i \le n} \left\{ \sum_{j=1}^m |a_{ij}||x_j| \right\} \le \max_{1 \le i \le n} \left\{ \sum_{j=1}^m |a_{ij}| \max_{1 \le j \le m} |x_j| \right\} \\ &= \max_{1 \le i \le n} \left\{ \sum_{j=1}^m |a_{ij}| \|x\|_\infty \right\} = \|x\|_\infty \max_{1 \le i \le n} \left\{ \sum_{j=1}^m |a_{ij}| \right\} \\ &= \max_{1 \le i \le n} \left\{ \sum_{j=1}^m |a_{ij}| \right\}. \end{aligned}$$

Theorem 3.9.2 (continued 3)

Proof (continued). So, $\max_{\|x\|_\infty=1}\|Ax\|_\infty \le \max_{1\le i\le n}\left\{\sum_{j=1}^m |a_{ij}|\right\}$. For given $n\times m$ matrix A_*, let k be such that $1\le k\le n$ and $\max_{1\le i\le n}\left\{\sum_{j=1}^m |a_{ij}|\right\} = \sum_{j=1}^m |a_{kj}|$. Then define $x_* = [\mathrm{sgn}(a_{k1}), \mathrm{sgn}(a_{k2}), \ldots, \mathrm{sgn}(a_{km})]^T \in \mathbb{R}^m$, where $\mathrm{sgn}(a) = \begin{cases} 1 & \text{if } a>0 \\ 0 & \text{if } a=0 \\ -1 & \text{if } a<0 \end{cases}$ (so $\|x_*\|_\infty = 1$ unless $A=0$). We then have

$$\|Ax_*\|_\infty = \max_{1\le i\le n}|b_i x_*| = \max_{1\le i\le n}|\langle b_i^T, x_*\rangle| = \max_{1\le i\le n}\left\{\sum_{j=1}^m |a_{ij}|\right\}.$$

Therefore $\|A\|_\infty = \max_{\|x\|_\infty=1}\|Ax\|_\infty = \max_{1\le i\le n}\left\{\sum_{j=1}^m |a_{ij}|\right\}$, as claimed. □

Theorem 3.9.4

Theorem 3.9.4. If square matrices A and B are orthogonally similar then $\|A\|_F = \|B\|_F$.

Proof. If A and B are orthogonally similar $n\times n$ matrices then there is (by definition) orthogonal matrix Q such that $A = Q^TBQ$. Then

$$\begin{aligned}\|A\|_F^2 &= \mathrm{tr}(A^TA) \text{ as observed above} \\ &= \mathrm{tr}((Q^TB^TQ)(Q^TBQ)) \text{ since } A = Q^TBQ \\ &= \mathrm{tr}(Q^TB^TBQ) \text{ since } QQ^T = I \\ &= \mathrm{tr}(BQQ^TB^T) = \mathrm{tr}(BB^T) \text{ since } \mathrm{tr}(CD) = \mathrm{tr}(DC) \\ &\quad \text{for square } C = Q^TB^T \text{ and } D = BQ \text{ by Exercise 3.2.E} \\ &= \mathrm{tr}(B^TB) = \|B\|_F^2 \text{ by Exercise 3.2.E.}\end{aligned}$$

So $\|A\|_F = \|B\|_F$, as claimed. □

Theorem 3.9.6

Theorem 3.9.6. For any matrix norm $\|\cdot\|$ and any square matrix A, $\rho(A) \le \|A\|$.

Proof. Let (c_i, v_i) be an eigenpair for A. Consider the square matrix $V = [v_i\ 0\ 0\ \cdots\ 0]$. Then $AV = c_iV$ and so

$$\begin{aligned}|c_i|\|V\| &= \|c_iV\| \text{ by part (2) of the definition of matrix norm} \\ &= \|AV\| \text{since } AV = c_iV \\ &\le \|A\|\|V\| \text{ by the Consistency Property.}\end{aligned}$$

Since v_i is an eigenvector it is nonzero and so $\|V\| \ne 0$. Therefore $|c_i| \le \|A\|$. Since c_i is an arbitrary eigenvalue of A, $\rho(A) \le \|A\|$. □

Theorem 3.9.7

Theorem 3.9.7. Let A be a square matrix. Then $\lim_{k\to\infty} A^k = 0$ if and only if $\rho(A) < 1$.

Proof. Suppose $\lim_{k\to\infty} A^k = 0$. Let (c_1, v_1) be an eigenpair for A where c_1 is a dominant eigenvalue of A. Let $\varepsilon > 0$. Then there exists $N \in \mathbb{N}$ such that for all $n \ge N$, $\|A^k - 0\| < \varepsilon/\|v_1\|$. Then for $n \ge N$,

$$\begin{aligned}\|A^kv_1 - 0\| &= \|A^kv_1\| \le \|A^k\|\|v_1\| \text{ by the definition of induced norm} \\ &< (\varepsilon/\|v_1\|)\|v_1\| = \varepsilon\end{aligned}$$

and so $\lim_{k\to\infty}(A^kv_1) = 0$. Since $A^kv_1 = c_1^kv_1$ then we have $\lim_{k\to\infty} c_1^kv_1 = 0$ and so $|c_1| < 1$. Therefore $\rho(A) < 1$.

Now suppose $\rho(A) < 1$. By Theorem 3.8.9, there is a Schur factorization of A such that $A = QTQ^{-1}$ where Q is orthogonal, T is upper triangular, and T has the same eigenvalues as A (see the note after Theorem 3.8.9), say $c_1, c_2, \ldots, c_n$. Fix real scalar $d > 0$ and form $n\times n$ diagonal matrix $D = \mathrm{diag}(d, d^2, d^3, \ldots, d^n)$.

Theorem 3.9.7 (continued 1)

Proof (continued). Then $D^{-1} = \text{diag}(d^{-1}, d^{-2}, d^{-3}, \ldots, d^{-n})$ and DTD^{-1} is upper triangular with the same diagonal entries as T. Now the eigenvalues of a square triangular matrix are the diagonal entries of the triangular matrix (by Example 3.1.A and the definition of eigenvalue), so the eigenvalues of DTD^{-1}, T, and A are all the same. With $T = [t_{ij}]$, $D = [d_{ij}]$, and $D^{-1} = [d'_{ij}]$, the (i,j) entry of DT is $\sum_{k=1}^n d_{ik}t_{kj} = d_{ii}t_{ij}$ and the (i,j) entry of DTD^{-1} is

$$\sum_{k=1}^{n}(d_{ii}t_{ik})(d'_{kj}) = d_{ii}t_{ij}d'_{jj} = d^i t_{ij} d^{-j} = d^{i-j}t_{ij}.$$

So the sum of the absolute values of the jth column of DTD^{-1} is

$$\sum_{i=1}^{j} d^{i-j}|t_{ij}| = |c_j| + \sum_{i=1}^{j-1} d^{i-j}|t_{ij}|.$$

Theorem 3.9.7 (continued 2)

Theorem 3.9.7. Let A be a square matrix. Then $\lim_{k\to\infty} A^k = 0$ if and only if $\rho(A) < 1$.

Proof (continued). Since $d > 0$ and $i - j < 0$ for $1 \le i \le j-1$, then $\lim_{d\to\infty} d^{i-j} = 0$ and so for any $\varepsilon > 0$ there is $M \in \mathbb{R}$ such that for $d \ge M$ we have

$$|d^{i-j} - 0| = d^{i-j} < \frac{\varepsilon}{(j-1)\max\{|t_{ij}|\}}.$$

So for given $\varepsilon > 0$ and with $d \ge M$, since $|c_j| \le \rho(A)$, we have

$$|c_j| + \sum_{i=1}^{j-1} d^{i-j}|t_{ij}| < \rho(A) + \sum_{i=1}^{j-1} \frac{\varepsilon}{(j-1)\max\{|t_{ij}|\}}|t_{ij}|$$

$$\le \rho(A) + \sum_{i=1}^{j-1}\frac{\varepsilon}{j-1} = \rho(A) + \varepsilon.$$

Theorem 3.9.7 (continued 3)

Proof (continued). So in terms of the L_1 norm for DTD^{-1} (that is, the column-sum norm)
$\|DTD^{-1}\|_1 = \max_{1\le j\le n}\left\{|c_j| + \sum_{i=1}^{j-1} d^{i-j}|t_{ij}|\right\} < \rho(A) + \varepsilon$ for $d \ge M$.

Now define norm $\|\cdot\|_d$ for any $n \times n$ matrix X as $\|X\|_d = \|(QD^{-1})^{-1}X(QD^{-1})\|_1$ where Q and D are the matrices above based on matrix A ($\|\cdot\|_d$ actually is a norm by Exercise 3.32). Then

$$\|A\|_d = \|(QD^{-1})^{-1}A(QD^{-1})\|_1 = \|DQ^{-1}AQD^{-1}\|_1 = \|DTD^{-1}\|_1.$$

By hypothesis, $\rho(A) < 1$, so choose $\varepsilon > 0$ (and corresponding M and d) so that $\rho(A) + \varepsilon < 1$. Then $\|A\|_d = \|DTD^{-1}\|_1 < \rho(A) + \varepsilon < 1$. So $\|A^k\|_d \le \|A\|_d^k$ by the Consistency Property and since $\|A\|_d < 1$ then $\lim_{k\to\infty}\|A^k\|_d \le \lim_{k\to\infty}\|A\|_d^k = 0$. That is, $\lim_{k\to\infty} A^k = 0$ with respect to $\|\cdot\|_d$ (and hence with respect to any matrix norm since all matrix norms are equivalent by Note 3.9.A). □

Theorem 3.9.8

Theorem 3.9.8. For square matrix A, $\lim_{k\to\infty}\|A^k\|^{1/k} = \rho(A)$.

Proof. By Theorem 3.9.6, $\rho(A) \le \|A\|$ (for any matrix norm). By Exercise 3.8.E, $\rho(A^k) = \rho(A)^k$, so $\rho(A)^k = \rho(A^k) \le \|A^k\|$ and

$$\rho(A) \le \|A^k\|^{1/k}. \qquad (*)$$

By Theorem 3.8.2(2), if c is an eigenvalue of A then bc is an eigenvalue of bA. So $\rho(bA) = b\rho(A)$. Let $\varepsilon > 0$ and consider $\dfrac{1}{\rho(A)+\varepsilon}A$. Then $\rho\left(\dfrac{1}{\rho(A)+\varepsilon}A\right) = \dfrac{\rho(A)}{\rho(A)+\varepsilon} < 1$ and so $\lim_{k\to\infty}\left(\dfrac{1}{\rho(A)+\varepsilon}A\right)^k = 0$ by Theorem 3.9.7. So

$$\lim_{k\to\infty}\left\|\frac{1}{(\rho(A)+\varepsilon)^k}A^k\right\| = \lim_{k\to\infty}\frac{\|A^k\|}{(\rho(A)+\varepsilon)^k} = 0$$

(since $A_n \to 0$ if and only if $\|A_n - 0\| = \|A_n\| \to 0$).

Theorem 3.9.8 (continued)

Theorem 3.9.8. For square matrix A, $\lim_{k\to\infty} \|A^k\|^{1/k} = \rho(A)$.

Proof (continued). So by the definition of limit, there is $M \in \mathbb{R}$ such that $k > M$ implies $\|A^k\|/(\rho(A)+\varepsilon)^k < 1$. Then $\|A^k\|^{1/k} < \rho(A)+\varepsilon$ for $k > M$. We now have from $(*)$ that

$$\rho(A) \le \|A^k\|^{1/k} < \rho(A) + \varepsilon$$

for $k > M$. Since $\varepsilon > 0$ is arbitrary, we have $\lim_{k\to\infty} \|A^k\|^{1/k} = \rho(A)$. □

Theorem 3.9.9

Theorem 3.9.9. Let A be an $n \times n$ matrix with $\|A\| < 1$. Then

$$I + \lim_{k\to\infty} \left(\sum_{n=1}^{k} A^n \right) = (I - A)^{-1}.$$

Proof. First, we introduce the sequence of partial sums $I, I + A, I + A + A^2, \ldots, S_k = I + \sum_{n=1}^{k} A^n, \ldots$. Then $(I - A)S_k = I - A^{k+1}$. Since $\|A\| < 1$ then $A^k \to 0$ (since $\|A^k - 0\| = \|A^k\| \le \|A\|^k$ by the Consistency Property) and so

$$\lim_{k\to\infty} (I - A)S_k = \lim_{k\to\infty} (I - A^{k+1})$$

or

$$(I - A) \lim_{k\to\infty} S_k = I - \lim_{k\to\infty} A^{k+1} = I$$

and so $\lim_{k\to\infty} S_k = (I - A)^{-1}$, as claimed. □