Theorem 6.5 Let A, B, and C be $n \times m$ matrices and λ and μ be real numbers. The following properties of addition and scalar multiplication hold:

(a) $A + B = B + A$, (b) $(A + B) + C = A + (B + C)$,

(c) $A + O = O + A = A$, (d) $A + (-A) = -A + A = 0$,

(e) $\lambda(A + B) = \lambda A + \lambda B$, (f) $(\lambda + \mu)A = \lambda A + \mu A$,

(g) $\lambda(\mu A) = (\lambda\mu)A$, (h) $1A = A$.

Theorem 6.9 Let A be an $n \times m$ matrix, B be an $m \times k$ matrix, C be a $k \times p$ matrix, D be an $m \times k$ matrix, and λ be a real number. The following properties hold:

(a) $A(BC) = (AB)C$; (Associative law for multiplication)

(b) $A(B + D) = AB + AD$; (Distributive law)

(c) $I_m B = B$ and $BI_k = B$; (Identity Element)

(d) $\lambda(AB) = (\lambda A)B = A(\lambda B)$. ■

square matrix

Definition 6.10 An $n \times n$ matrix A is **nonsingular** (or *invertible*) if an $n \times n$ matrix A^{-1} exists with $AA^{-1} = A^{-1}A = I$. The matrix A^{-1} is called the **inverse** of A. A matrix without an inverse is called **singular** (or *noninvertible*). ■

The following properties regarding matrix inverses follow from Definition 6.10. The proofs of these results are considered in Exercise 5.

Theorem 6.11 For any nonsingular $n \times n$ matrix A:

(a) A^{-1} is unique.

(b) A^{-1} is nonsingular and $(A^{-1})^{-1} = A$.

(c) If B is also a nonsingular $n \times n$ matrix, then $(AB)^{-1} = B^{-1}A^{-1}$.

Theorem 6.13 The following operations involving the transpose of a matrix hold whenever the operation is possible:

(a) $(A^t)^t = A$, (b) $(A + B)^t = A^t + B^t$,

(c) $(AB)^t = B^t A^t$, (d) if A^{-1} exists, then $(A^{-1})^t = (A^t)^{-1}$.

Definition 6.14

(a) If $A = [a]$ is a 1×1 matrix, then $\det A = a$.

(b) If A is an $n \times n$ matrix, the **minor** M_{ij} is the determinant of the $(n-1) \times (n-1)$ submatrix of A obtained by deleting the ith row and jth column of the matrix A.

(c) The **cofactor** A_{ij} associated with M_{ij} is defined by $A_{ij} = (-1)^{i+j} M_{ij}$.

(d) The **determinant** of the $n \times n$ matrix A, when $n > 1$, is given either by

$$\det A = \sum_{j=1}^{n} a_{ij} A_{ij} = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} M_{ij}, \quad \text{for any } i = 1, 2, \ldots, n,$$

or by

$$\det A = \sum_{i=1}^{n} a_{ij} A_{ij} = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} M_{ij}, \quad \text{for any } j = 1, 2, \ldots, n.$$ ■

Theorem 6.15 Suppose A is an $n \times n$ matrix:

(a) If any row or column of A has only zero entries, then $\det A = 0$.

(b) If A has two rows or two columns the same, then $\det A = 0$.

elementary row operations

(c) If $\tilde{A}$ is obtained from A by the operation $(E_i) \leftrightarrow (E_j)$, with $i \neq j$, then $\det \tilde{A} = -\det A$.

(d) If $\tilde{A}$ is obtained from A by the operation $(\lambda E_i) \rightarrow (E_i)$, then $\det \tilde{A} = \lambda \det A$.

(e) If $\tilde{A}$ is obtained from A by the operation $(E_i + \lambda E_j) \rightarrow (E_i)$ with $i \neq j$, then $\det \tilde{A} = \det A$.

(f) If B is also an $n \times n$ matrix, then $\det AB = \det A \det B$.

(g) $\det A^t = \det A$.

(h) When A^{-1} exists, $\det A^{-1} = (\det A)^{-1}$.

(i) If A is an upper triangular, or a lower triangular, or a diagonal matrix, then $\det A = \prod_{i=1}^{n} a_{ii}$. ∎

Theorem 6.16 The following statements are equivalent for any $n \times n$ matrix A:

(a) The equation $A\mathbf{x} = \mathbf{0}$ has the unique solution $\mathbf{x} = \mathbf{0}$.

(b) The system $A\mathbf{x} = \mathbf{b}$ has a unique solution for any n-dimensional column vector $\mathbf{b}$.

(c) The matrix A is nonsingular; that is, A^{-1} exists.

(d) $\det A \neq 0$.

(e) Gaussian elimination with row interchanges can be performed on the system $A\mathbf{x} = \mathbf{b}$ for any n-dimensional column vector $\mathbf{b}$. ∎

Definition 7.1 A **vector norm** on $\mathbb{R}^n$ is a function, $\|\cdot\|$, from $\mathbb{R}^n$ into $\mathbb{R}$ with the following properties:

(i) $\|\mathbf{x}\| \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$,

(ii) $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$,

(iii) $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for all $\alpha \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$,

(iv) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. ∎

$\|x\|$ defines a measure for the magnitude of vector x

Definition 7.2 The l_2 and l_∞ norms for the vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)^t$ are defined by

$$\|\mathbf{x}\|_2 = \left\{\sum_{i=1}^{n} x_i^2\right\}^{1/2} \quad \text{and} \quad \|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

Theorem 7.3 **(Cauchy–Bunyakovsky–Schwarz Inequality for Sums)**

For each $\mathbf{x} = (x_1, x_2, \ldots, x_n)^t$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)^t$ in $\mathbb{R}^n$,

$$\mathbf{x}^t\mathbf{y} = \sum_{i=1}^{n} x_i y_i \leq \left\{\sum_{i=1}^{n} x_i^2\right\}^{1/2} \left\{\sum_{i=1}^{n} y_i^2\right\}^{1/2} = \|\mathbf{x}\|_2 \cdot \|\mathbf{y}\|_2.$$

Definition 7.4 If $\mathbf{x} = (x_1, x_2, \ldots, x_n)^t$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)^t$ are vectors in $\mathbb{R}^n$, the l_2 and l_∞ distances between $\mathbf{x}$ and $\mathbf{y}$ are defined by

$$\|\mathbf{x} - \mathbf{y}\|_2 = \left\{\sum_{i=1}^{n} (x_i - y_i)^2\right\}^{1/2} \quad \text{and} \quad \|\mathbf{x} - \mathbf{y}\|_\infty = \max_{1 \leq i \leq n} |x_i - y_i|. \quad ∎$$

Definition 7.5 A sequence $\{\mathbf{x}^{(k)}\}_{k=1}^{\infty}$ of vectors in $\mathbb{R}^n$ is said to **converge** to $\mathbf{x}$ with respect to the norm $\|\cdot\|$ if, given any $\varepsilon > 0$, there exists an integer $N(\varepsilon)$ such that

$$\|\mathbf{x}^{(k)} - \mathbf{x}\| < \varepsilon, \quad \text{for all } k \geq N(\varepsilon).$$

■

Theorem 7.6 The sequence of vectors $\{\mathbf{x}^{(k)}\}$ converges to $\mathbf{x}$ in $\mathbb{R}^n$ with respect to $\|\cdot\|_\infty$ if and only if $\lim_{k\to\infty} x_i^{(k)} = x_i$, for each $i = 1, 2, \ldots, n$. ■

Theorem 7.7 For each $\mathbf{x} \in \mathbb{R}^n$,

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \sqrt{n}\|\mathbf{x}\|_\infty.$$

Definition 7.8 A **matrix norm** on the set of all $n \times n$ matrices is a real-valued function, $\|\cdot\|$, defined on this set, satisfying for all $n \times n$ matrices A and B and all real numbers α:

(i) $\|A\| \geq 0$;

(ii) $\|A\| = 0$, if and only if A is O, the matrix with all 0 entries;

(iii) $\|\alpha A\| = |\alpha|\|A\|$;

(iv) $\|A + B\| \leq \|A\| + \|B\|$;

(v) $\|AB\| \leq \|A\|\|B\|$. ■

The **distance between $n \times n$ matrices** A and B with respect to this matrix norm is $\|A - B\|$.

Theorem 7.9 If $\|\cdot\|$ is a vector norm on $\mathbb{R}^n$, then

$$\|A\| = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|$$

is a matrix norm.

Corollary 7.10 For any vector $\mathbf{z} \neq \mathbf{0}$, matrix A, and any natural norm $\|\cdot\|$, we have

$$\|A\mathbf{z}\| \leq \|A\| \cdot \|\mathbf{z}\|.$$

Theorem 7.11 If $A = (a_{ij})$ is an $n \times n$ matrix, then

$$\|A\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|.$$

Definition 7.12 If A is a square matrix, the **characteristic polynomial** of A is defined by

$$p(\lambda) = \det(A - \lambda I).$$

Definition 7.13 If p is the characteristic polynomial of the matrix A, the zeros of p are **eigenvalues**, or characteristic values, of the matrix A. If λ is an eigenvalue of A and $\mathbf{x} \neq \mathbf{0}$ satisfies $(A - \lambda I)\mathbf{x} = \mathbf{0}$, then $\mathbf{x}$ is an **eigenvector**, or characteristic vector, of A corresponding to the eigenvalue λ. ■

Definition 7.14 The **spectral radius** $\rho(A)$ of a matrix A is defined by

$$\rho(A) = \max |\lambda|, \quad \text{where } \lambda \text{ is an eigenvalue of } A.$$

(Recall that for complex $\lambda = \alpha + \beta i$, we have $|\lambda| = (\alpha^2 + \beta^2)^{1/2}$.)

Theorem 7.15 If A is an $n \times n$ matrix, then

(i) $\|A\|_2 = [\rho(A^t A)]^{1/2}$,

(ii) $\rho(A) \leq \|A\|$, for any natural norm $\|\cdot\|$.

Definition 7.16 We call an $n \times n$ matrix A **convergent** if

$$\lim_{k\to\infty} (A^k)_{ij} = 0, \quad \text{for each } i = 1, 2, \ldots, n \quad \text{and} \quad j = 1, 2, \ldots, n.$$

Theorem 7.17 The following statements are equivalent.

(i) A is a convergent matrix.

(ii) $\lim_{n\to\infty} \|A^n\| = 0$, for some natural norm.

(iii) $\lim_{n\to\infty} \|A^n\| = 0$, for all natural norms.

(iv) $\rho(A) < 1$.

(v) $\lim_{n\to\infty} A^n \mathbf{x} = \mathbf{0}$, for every $\mathbf{x}$.

The method of Example 1 is called the **Jacobi iterative method**. It consists of solving the ith equation in $A\mathbf{x} = \mathbf{b}$ for x_i to obtain (provided $a_{ii} \neq 0$)

$$x_i = \sum_{\substack{j=1 \\ j\neq i}}^{n} \left(-\frac{a_{ij}x_j}{a_{ii}} \right) + \frac{b_i}{a_{ii}}, \qquad \text{for } i = 1, 2, \ldots, n$$

and generating each $x_i^{(k)}$ from components of $\mathbf{x}^{(k-1)}$ for $k \geq 1$ by

$$x_i^{(k)} = \frac{\sum_{\substack{j=1 \\ j\neq i}}^{n} \left(-a_{ij}x_j^{(k-1)} \right) + b_i}{a_{ii}}, \qquad \text{for } i = 1, 2, \ldots, n. \tag{7.4}$$

A possible improvement in Algorithm 7.1 can be seen by reconsidering Eq. (7.4). The components of $\mathbf{x}^{(k-1)}$ are used to compute $x_i^{(k)}$. Since, for $i > 1$, $x_1^{(k)}, \ldots, x_{i-1}^{(k)}$ have already been computed and are probably better approximations to the actual solutions $x_1, \ldots, x_{i-1}$ than $x_1^{(k-1)}, \ldots, x_{i-1}^{(k-1)}$, it seems more reasonable to compute $x_i^{(k)}$ using these most recently calculated values. That is, we can use

$$x_i^{(k)} = \frac{-\sum_{j=1}^{i-1}(a_{ij}x_j^{(k)}) - \sum_{j=i+1}^{n}(a_{ij}x_j^{(k-1)}) + b_i}{a_{ii}}, \tag{7.7}$$

for each $i = 1, 2, \ldots, n$, instead of Eq. (7.4). This modification is called the **Gauss–Seidel iterative technique** and is illustrated in the following example.

Lemma 7.18 If the spectral radius $\rho(T)$ satisfies $\boxed{\rho(T) < 1}$ then $(I - T)^{-1}$ exists, and

$$(I - T)^{-1} = I + T + T^2 + \cdots = \sum_{j=0}^{\infty} T^j.$$

Theorem 7.19 For any $\mathbf{x}^{(0)} \in \mathbb{R}^n$, the sequence $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ defined by

(Proof skipped)

$$\mathbf{x}^{(k)} = T\mathbf{x}^{(k-1)} + \mathbf{c}, \quad \text{for each } k \geq 1, \tag{7.10}$$

converges to the unique solution of $\mathbf{x} = T\mathbf{x} + \mathbf{c}$ if and only if $\rho(T) < 1$. ■

Corollary 7.20 If $\|T\| < 1$ for any natural matrix norm and $\mathbf{c}$ is a given vector, then the sequence $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ defined by $\mathbf{x}^{(k)} = T\mathbf{x}^{(k-1)} + \mathbf{c}$ converges, for any $\mathbf{x}^{(0)} \in \mathbb{R}^n$, to a vector $\mathbf{x} \in \mathbb{R}^n$, and the following error bounds hold:

(i) $\|\mathbf{x} - \mathbf{x}^{(k)}\| \leq \|T\|^k \|\mathbf{x}^{(0)} - \mathbf{x}\|$;

(ii) $\|\mathbf{x} - \mathbf{x}^{(k)}\| \leq \dfrac{\|T\|^k}{1 - \|T\|}\|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|$. ■