1

Vectors, Matrices, and Systems of Linear Equations

Jane Day
San Jose State University

Throughout this chapter, F will denote a field. The references [Lay03], [Leo02], and [SIF00] are good sources for more detail about much of the material in this chapter. They discuss primarily the field of real numbers, but the proofs are usually valid for any field.

1.1 Vector Spaces

Vectors are used in many applications. They often represent quantities that have both direction and magnitude, such as velocity or position, and can appear as functions, as n-tuples of scalars, or in other disguises. Whenever objects can be added and multiplied by scalars, they may be elements of some vector space. In this section, we formulate a general definition of vector space and establish its basic properties. An element of a field, such as the real numbers or the complex numbers, is called a scalar to distinguish it from a vector.

Definitions:

A **vector space over** F is a set V together with a function $V \times V \to V$ called **addition**, denoted $(\mathbf{x},\mathbf{y}) \to \mathbf{x} + \mathbf{y}$, and a function $F \times V \to V$ called **scalar multiplication** and denoted $(c,\mathbf{x}) \to c\mathbf{x}$, which satisfy the following axioms:

1. (Commutativity) For each $\mathbf{x}, \mathbf{y} \in V$, $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
2. (Associativity) For each $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$.
3. (Additive identity) There exists a **zero vector** in V, denoted $\mathbf{0}$, such that $\mathbf{0} + \mathbf{x} = \mathbf{x}$ for each $\mathbf{x} \in V$.
4. (Additive inverse) For each $\mathbf{x} \in V$, there exists $-\mathbf{x} \in V$ such that $(-\mathbf{x}) + \mathbf{x} = \mathbf{0}$.
5. (Distributivity) For each $a \in F$ and $\mathbf{x}, \mathbf{y} \in V$, $a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}$.
6. (Distributivity) For each $a, b \in F$ and $\mathbf{x} \in V$, $(a + b)\,\mathbf{x} = a\mathbf{x} + b\mathbf{x}$.

7. (Associativity) For each $a, b \in F$ and $\mathbf{x} \in V$, $(ab)\,\mathbf{x} = a(b\mathbf{x})$.
8. For each $\mathbf{x} \in V$, $1\mathbf{x} = \mathbf{x}$.

The properties that for all $\mathbf{x}, \mathbf{y} \in V$, and $a \in F$, $\mathbf{x}+\mathbf{y} \in V$ and $a\mathbf{x} \in V$, are called **closure under addition** and **closure under scalar multiplication**, respectively. The elements of a vector space V are called **vectors**. A vector space is called **real** if $F = \mathbb{R}$, **complex** if $F = \mathbb{C}$.

If n is a positive integer, F^n denotes the set of all ordered n-tuples (written as columns). These are sometimes written instead as rows $[x_1 \quad \cdots \quad x_n]$ or $(x_1, \ldots, x_n)$. For $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in F^n$ and $c \in F$, define addition and scalar multiplication coordinate-wise: $\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}$ and $c\mathbf{x} = \begin{bmatrix} cx_1 \\ \vdots \\ cx_n \end{bmatrix}$. Let $\mathbf{0}$ denote the n-tuple of zeros. For $\mathbf{x} \in F^n$, x_j is called the j$^{\text{th}}$ **coordinate** of $\mathbf{x}$.

A **subspace** of vector space V over field F is a subset of V, which is itself a vector space over F when the addition and scalar multiplication of V are used. If S_1 and S_2 are subsets of vector space V, define $S_1 + S_2 = \{\mathbf{x}+\mathbf{y} : \mathbf{x} \in S_1 \text{ and } \mathbf{y} \in S_2\}$.

Facts:

Let V be a vector space over F.

1. F^n is a vector space over F.
2. [FIS03, pp. 11–12] (Basic properties of a vector space):
   - The vector $\mathbf{0}$ is the only additive identity in V.
   - For each $\mathbf{x} \in V$, $-\mathbf{x}$ is the only additive inverse for $\mathbf{x}$ in V.
   - For each $\mathbf{x} \in V$, $-\mathbf{x} = (-1)\mathbf{x}$.
   - If $a \in F$ and $\mathbf{x} \in V$, then $a\mathbf{x} = \mathbf{0}$ if and only if $a = 0$ or $\mathbf{x} = \mathbf{0}$.
   - (Cancellation) If $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and $\mathbf{x} + \mathbf{y} = \mathbf{x} + \mathbf{z}$, then $\mathbf{y} = \mathbf{z}$.
3. [FIS03, pp. 16–17] Let W be a subset of V. The following are equivalent:
   - W is a subspace of V.
   - W is nonempty and closed under addition and scalar multiplication.
   - $\mathbf{0} \in W$ and for any $\mathbf{x}, \mathbf{y} \in W$ and $a, b \in F$, $a\mathbf{x} + b\mathbf{y} \in W$.
4. For any vector space V, $\{\mathbf{0}\}$ and V itself are subspaces of V.
5. [FIS03, p. 19] The intersection of any nonempty collection of subspaces of V is a subspace of V.
6. [FIS03, p. 22] Let W_1 and W_2 be subspaces of V. Then $W_1 + W_2$ is a subspace of V containing W_1 and W_2. It is the smallest subspace that contains them in the sense that any subspace that contains both W_1 and W_2 must contain $W_1 + W_2$.

Examples:

1. The set $\mathbb{R}^n$ of all ordered n-tuples of real numbers is a vector space over $\mathbb{R}$, and the set $\mathbb{C}^n$ of all ordered n-tuples of complex numbers is a vector space over $\mathbb{C}$. For instance, $\mathbf{x} = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 2i \\ 4 \\ 2-3i \end{bmatrix}$ are elements of $\mathbb{C}^3$; $\mathbf{x} + \mathbf{y} = \begin{bmatrix} 3+2i \\ 4 \\ 1-3i \end{bmatrix}$, $-\mathbf{y} = \begin{bmatrix} -2i \\ -4 \\ -2+3i \end{bmatrix}$, and $i\mathbf{y} = \begin{bmatrix} -2 \\ 4i \\ 3+2i \end{bmatrix}$.
2. Notice $\mathbb{R}^n$ is a subset of $\mathbb{C}^n$ but not a subspace of $\mathbb{C}^n$, since $\mathbb{R}^n$ is not closed under multiplication by nonreal numbers.

3. The vector spaces $\mathbb{R}$, $\mathbb{R}^2$, and $\mathbb{R}^3$ are the usual Euclidean spaces of analytic geometry. There are three types of subspaces of $\mathbb{R}^2$: $\{\mathbf{0}\}$, a line through the origin, and $\mathbb{R}^2$ itself. There are four types of subspaces of $\mathbb{R}^3$: $\{\mathbf{0}\}$, a line through the origin, a plane through the origin, and $\mathbb{R}^3$ itself. For instance, let $\mathbf{v} = (5, -1, -1)$ and $\mathbf{w} = (0, 3, -2)$. The lines $W_1 = \{s\mathbf{v} : s \in \mathbb{R}\}$ and $W_2 = \{s\mathbf{w} : s \in \mathbb{R}\}$ are subspaces of $\mathbb{R}^3$. The subspace $W_1 + W_2 = \{s\mathbf{v} + t\mathbf{w} : s, t \in \mathbb{R}\}$ is a plane. The set $\{s\mathbf{v} + \mathbf{w} : s \in \mathbb{R}\}$ is a line parallel to W_1, but is not a subspace. (For more information on geometry, see Chapter 65.)
4. Let $F[x]$ be the set of all polynomials in the single variable x, with coefficients from F. To add polynomials, add coefficients of like powers; to multiply a polynomial by an element of F, multiply each coefficient by that scalar. With these operations, $F[x]$ is a vector space over F. The zero polynomial z, with all coefficients 0, is the additive identity of $F[x]$. For $f \in F[x]$, the function $-f$ defined by $-f(x) = (-1)f(x)$ is the additive inverse of f.
5. In $F[x]$, the constant polynomials have degree 0. For $n > 0$, the polynomials with highest power term x^n are said to have degree n. For a nonnegative integer n, let $F[x; n]$ be the subset of $F[x]$ consisting of all polynomials of degree n or less. Then $F[x; n]$ is a subspace of $F[x]$.
6. When $n > 0$, the set of all polynomials of degree exactly n is not a subspace of $F[x]$ because it is not closed under addition or scalar multiplication. The set of all polynomials in $\mathbb{R}[x]$ with rational coefficients is not a subspace of $\mathbb{R}[x]$ because it is not closed under scalar multiplication.
7. Let V be the set of all infinite sequences $(a_1, a_2, a_3, \ldots)$, where each $a_j \in F$. Define addition and scalar multiplication coordinate-wise. Then V is a vector space over F.
8. Let X be a nonempty set and let $\mathcal{F}(X, F)$ be the set of all functions $f: X \to F$. Let $f, g \in \mathcal{F}(X, F)$ and define $f + g$ and cf pointwise, as $(f + g)(x) = f(x) + g(x)$ and $(cf)(x) = cf(x)$ for all $x \in X$. With these operations, $\mathcal{F}(X, F)$ is a vector space over F. The zero function is the additive identity and $(-1)f = -f$, the additive inverse of f.
9. Let X be a nonempty subset of $\mathbb{R}^n$. The set $C(X)$ of all continuous functions $f: X \to \mathbb{R}$ is a subspace of $\mathcal{F}(X, \mathbb{R})$. The set $\mathcal{D}(X)$ of all differentiable functions $f: X \to \mathbb{R}$ is a subspace of $C(X)$ and also of $\mathcal{F}(X, \mathbb{R})$.

1.2 Matrices

Matrices are rectangular arrays of scalars that are used in a great variety of ways, such as to solve linear systems, model linear behavior, and approximate nonlinear behavior. They are standard tools in almost every discipline, from sociology to physics and engineering.

Definitions:

An $m \times p$ **matrix** over F is an $m \times p$ rectangular array $A = \begin{bmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mp} \end{bmatrix}$, with entries from F. The notation $A = [a_{ij}]$ that displays a typical entry is also used. The element a_{ij} of the matrix A is called the (i, j) **entry** of A and can also be denoted $(A)_{ij}$. The **shape** (or **size**) of A is $m \times p$, and A is **square** if $m = p$; in this case, m is also called the size of A. Two matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ are said to be **equal** if they have the same shape and $a_{ij} = b_{ij}$ for all i, j. Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be $m \times p$ matrices, and let c be a scalar. Define **addition** and **scalar multiplication** on the set of all $m \times p$ matrices over F entrywise, as $A + B = [a_{ij} + b_{ij}]$ and $cA = [ca_{ij}]$. The set of all $m \times p$ matrices over F with these operations is denoted $F^{m \times p}$.

If A is $m \times p$, **row** i is $[a_{i1}, \ \ldots, \ a_{ip}]$ and **column** j is $\begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix}$. These are called a **row vector** and a **column vector** respectively, and they belong to $F^{n \times 1}$ and $F^{1 \times n}$, respectively. The elements of F^n are identified with the elements of $F^{n \times 1}$ (or sometimes with the elements of $F^{1 \times n}$). Let $\mathbf{0}_{mp}$ denote the $m \times p$ matrix of zeros, often shortened to $\mathbf{0}$ when the size is clear. Define $-A = (-1)A$.

Let $A = [\mathbf{a}_1 \quad \dots \quad \mathbf{a}_p] \in F^{m\times p}$, where $\mathbf{a}_j$ is the jth column of A, and let $\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_p \end{bmatrix} \in F^{p\times 1}$. The **matrix–vector product** of A and $\mathbf{b}$ is $A\mathbf{b} = b_1\mathbf{a}_1 + \cdots + b_p\mathbf{a}_p$. Notice $A\mathbf{b}$ is $m \times 1$.

If $A \in F^{m\times p}$ and $C = [\mathbf{c}_1 \quad \dots \quad \mathbf{c}_n] \in F^{p\times n}$, define the **matrix product** of A and C as $AC = [A\mathbf{c}_1 \quad \dots \quad A\mathbf{c}_n]$. Notice AC is $m \times n$.

Square matrices A and B **commute** if $AB = BA$. When $i = j$, a_{ii} is a **diagonal entry** of A and the set of all its diagonal entries is the **main diagonal** of A. When $i \neq j$, a_{ij} is an **off-diagonal entry**.

The **trace** of A is the sum of all the diagonal entries of A, $\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}$.

A matrix $A = [a_{ij}]$ is **diagonal** if $a_{ij} = 0$ whenever $i \neq j$, **lower triangular** if $a_{ij} = 0$ whenever $i < j$, and **upper triangular** if $a_{ij} = 0$ whenever $i > j$. A **unit triangular** matrix is a lower or upper triangular matrix in which each diagonal entry is 1.

The **identity matrix** I_n, often shortened to I when the size is clear, is the $n \times n$ matrix with main diagonal entries 1 and other entries 0.

A **scalar matrix** is a scalar multiple of the identity matrix.

A **permutation matrix** is one whose rows are some rearrangement of the rows of an identity matrix.

Let $A \in F^{m\times p}$. The **transpose** of A, denoted A^T, is the $p \times m$ matrix whose (i, j) entry is the (j, i) entry of A.

The square matrix A is **symmetric** if $A^T = A$ and **skew-symmetric** if $A^T = -A$.

When $F = \mathbb{C}$, that is, when A has complex entries, the **Hermitian adjoint** of A is its conjugate transpose, $A^* = \bar{A}^T$; that is, the (i, j) entry of A^* is $\overline{a_{ji}}$. Some authors, such as [Leo02], write A^H instead of A^*.

The square matrix A is **Hermitian** if $A^* = A$ and **skew-Hermitian** if $A^* = -A$.

Let α be a nonempty set of row indices and β a nonempty set of column indices. A **submatrix** of A is a matrix $A[\alpha, \beta]$ obtained by choosing the entries of A, which lie in rows α and columns β. A **principal submatrix** of A is a submatrix of the form $A[\alpha, \alpha]$. A **leading principal submatrix** of A is one of the form $A[\{1, \dots, k\}, \{1, \dots, k\}]$.

Facts:

1. [SIF00, p. 5] $F^{m\times p}$ is a vector space over F. That is, if $\mathbf{0}$, A, B, $C \in F^{m\times p}$, and c,$d \in F$, then:
   - $A + B = B + A$
   - $(A + B) + C = A + (B + C)$
   - $A + \mathbf{0} = \mathbf{0} + A = A$
   - $A + (-A) = (-A) + A = \mathbf{0}$
   - $c(A + B) = cA + cB$
   - $(c + d)A = cA + dA$
   - $(cd)\, A = c(dA)$
   - $1A = A$
2. If $A \in F^{m\times p}$ and $C \in F^{p\times n}$, the (i, j) entry of AC is $(AC)_{ij} = \sum_{k=1}^{p} a_{ik}a_{kj}$. This is the matrix product of row i of A and column j of C.
3. [SIF00, p. 88] Let $c \in F$, let A and B be matrices over F, let I denote an identity matrix, and assume the shapes allow the following sums and products to be calculated. Then:
   - $AI = IA = A$
   - $A\mathbf{0} = \mathbf{0}$ and $\mathbf{0}A = \mathbf{0}$
   - $A(BC) = (AB)C$
   - $A(B + C) = AB + AC$
   - $(A + B)C = AC + BC$
   - $c(AB) = A(cB) = (cA)B$ for any scalar c

4. [SIF00, p. 5 and p. 20] Let $c \in F$, let A and B be matrices over F, and assume the shapes allow the following sums and products to be calculated. Then:
   - $(A^T)^T = A$
   - $(A+B)^T = A^T + B^T$
   - $(cA)^T = cA^T$
   - $(AB)^T = B^T A^T$
5. [Leo02, pp. 321–323] Let $c \in \mathbb{C}$, let A and B be matrices over $\mathbb{C}$, and assume the shapes allow the following sums and products to be calculated. Then:
   - $(A^*)^* = A$
   - $(A+B)^* = A^* + B^*$
   - $(cA)^* = \bar{c}A^*$
   - $(AB)^* = B^* A^*$
6. If A and B are $n \times n$ and upper (lower) triangular, then AB is upper (lower) triangular.

Examples:

1. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 7 \\ 8 \\ -9 \end{bmatrix}$. By definition, $A\mathbf{b} = 7\begin{bmatrix} 1 \\ 4 \end{bmatrix} + 8\begin{bmatrix} 2 \\ 5 \end{bmatrix} - 9\begin{bmatrix} 3 \\ 6 \end{bmatrix} = \begin{bmatrix} -4 \\ 14 \end{bmatrix}$. Hand calculation of $A\mathbf{b}$ can be done more quickly using Fact 2: $A\mathbf{b} = \begin{bmatrix} 1\cdot 7 + 2\cdot 8 - 3\cdot 9 \\ 4\cdot 7 + 5\cdot 8 - 6\cdot 9 \end{bmatrix} = \begin{bmatrix} -4 \\ 14 \end{bmatrix}$.
2. Let $A = \begin{bmatrix} 1 & -3 & 4 \\ 2 & 0 & 8 \end{bmatrix}$, $B = \begin{bmatrix} -3 & 8 & 0 \\ 1 & 2 & -5 \end{bmatrix}$, and $C = \begin{bmatrix} 1 & -1 & 8 \\ 1 & 3 & 0 \\ 1 & 2 & -2 \end{bmatrix}$. Then $A + B = \begin{bmatrix} -2 & 5 & 4 \\ 3 & 2 & 3 \end{bmatrix}$ and $2A = \begin{bmatrix} 2 & -6 & 8 \\ 4 & 0 & 16 \end{bmatrix}$. The matrices $A + C$, BA, and AB are not defined, but
$$AC = \left[A\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad A\begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix} \quad A\begin{bmatrix} 8 \\ 0 \\ -2 \end{bmatrix} \right] = \begin{bmatrix} 2 & -2 & 0 \\ 10 & 14 & 0 \end{bmatrix}.$$
3. Even when the shapes of A and B allow both AB and BA to be calculated, AB and BA are not usually equal. For instance, let $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$; then $AB = \begin{bmatrix} a & b \\ 2c & 2d \end{bmatrix}$ and $BA = \begin{bmatrix} a & 2b \\ c & 2d \end{bmatrix}$, which will be equal only if $b = c = 0$.
4. The product of matrices can be a zero matrix even if neither has any zero entries. For example, if $A = \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, then $AB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. Notice that BA is also defined but has no zero entries: $BA = \begin{bmatrix} 3 & -3 \\ 3 & -3 \end{bmatrix}$.
5. The matrices $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -9 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \end{bmatrix}$ are diagonal, $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 1 & 5 & -9 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 & 0 \\ 2 & -3 & 0 \end{bmatrix}$ are lower triangular, and $\begin{bmatrix} 1 & -4 & 7 \\ 0 & 1 & 2 \\ 0 & 0 & -9 \end{bmatrix}$ and $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ are upper triangular. The matrix $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 5 & 1 \end{bmatrix}$ is unit lower triangular, and its transpose is unit upper triangular.

6. Examples of permutation matrices include every identity matrix, $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, and $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.

7. Let $A = \begin{bmatrix} 1+i & -3i & 4 \\ 1+2i & 5i & 0 \end{bmatrix}$. Then $A^T = \begin{bmatrix} 1+i & 1+2i \\ -3i & 5i \\ 4 & 0 \end{bmatrix}$ and $A^* = \begin{bmatrix} 1-i & 1-2i \\ 3i & -5i \\ 4 & 0 \end{bmatrix}$.

8. The matrices $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix}$ and $\begin{bmatrix} i & 2 & 3+2i \\ 2 & 4-i & 5i \\ 3+2i & 5i & 6 \end{bmatrix}$ are symmetric.

9. The matrices $\begin{bmatrix} 0 & 2 & 3 \\ -2 & 0 & 5 \\ -3 & -5 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 2 & 3+2i \\ -2 & 0 & -5 \\ -3-2i & 5 & 0 \end{bmatrix}$ are skew-symmetric.

10. The matrix $\begin{bmatrix} 1 & 2+i & 1-3i \\ 2-i & 0 & 1 \\ 1+3i & 1 & 6 \end{bmatrix}$ is Hermitian, and any real symmetric matrix, such as $\begin{bmatrix} 4 & 2 & 3 \\ 2 & 0 & 5 \\ 3 & 5 & -1 \end{bmatrix}$, is also Hermitian.

11. The matrix $\begin{bmatrix} i & 2 & -3+2i \\ -2 & 4i & 5 \\ 3+2i & -5 & 0 \end{bmatrix}$ is skew-Hermitian, and any real skew-symmetric matrix, such as $\begin{bmatrix} 0 & 2 & -3 \\ -2 & 0 & 5 \\ 3 & -5 & 0 \end{bmatrix}$, is also skew-Hermitian.

12. Let $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}$. Row 1 of A is $[1 \quad 2 \quad 3 \quad 4]$, column 3 is $\begin{bmatrix} 3 \\ 7 \\ 11 \\ 15 \end{bmatrix}$, and the submatrix in rows $\{1, 2, 4\}$ and columns $\{2, 3, 4\}$ is $A[\{1,2,4\},\{2,3,4\}] = \begin{bmatrix} 2 & 3 & 4 \\ 6 & 7 & 8 \\ 14 & 15 & 16 \end{bmatrix}$. A principal submatrix of A is $A[\{1, 2, 4\}, \{1, 2, 4\}] = \begin{bmatrix} 1 & 2 & 4 \\ 5 & 6 & 8 \\ 13 & 14 & 16 \end{bmatrix}$. The leading principal submatrices of A are $[1]$, $\begin{bmatrix} 1 & 2 \\ 5 & 6 \end{bmatrix}$, $\begin{bmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \\ 9 & 10 & 11 \end{bmatrix}$, and A itself.

1.3 Gaussian and Gauss–Jordan Elimination

Definitions:

Let A be a matrix with m rows.

When a row of A is not zero, its first nonzero entry is the **leading entry** of the row. The matrix A is in **row echelon form** (REF) when the following two conditions are met:

1. Any zero rows are below all nonzero rows.
2. For each nonzero row $i, i \leq m - 1$, either row $i + 1$ is zero or the leading entry of row $i + 1$ is in a column to the right of the column of the leading entry in row i.

The matrix A is in **reduced row echelon form** (RREF) if it is in row echelon form and the following third condition is also met:

3. If a_{ik} is the leading entry in row i, then $a_{ik} = 1$, and every entry of column k other than a_{ik} is zero.

Elementary row operations on a matrix are operations of the following types:

1. Add a multiple of one row to a different row.
2. Exchange two different rows.
3. Multiply one row by a nonzero scalar.

The matrix A is **row equivalent** to the matrix B if there is a sequence of elementary row operations that transforms A into B. The **reduced row echelon form** of A, RREF(A), is the matrix in reduced row echelon form that is row equivalent to A. A **row echelon form of** A is any matrix in row echelon form that is row equivalent to A. The **rank** of A, denoted rank A or rank(A), is the number of leading entries in RREF(A). If A is in row echelon form, the positions of the leading entries in its nonzero rows are called **pivot positions** and the entries in those positions are called **pivots**. A column (row) that contains a pivot position is a **pivot column** (**pivot row**).

Gaussian Elimination is a process that uses elementary row operations in a particular way to change, or reduce, a matrix to row echelon form. **Gauss–Jordan Elimination** is a process that uses elementary row operations in a particular way to reduce a matrix to RREF. See Algorithm 1 below.

Facts:

Let $A \in F^{m \times p}$.

1. [Lay03, p. 15] The reduced row echelon form of A, RREF(A), exists and is unique.
2. A matrix in REF or RREF is upper triangular.
3. Every elementary row operation is reversible by an elementary row operation of the same type.
4. If A is row equivalent to B, then B is row equivalent to A.
5. If A is row equivalent to B, then RREF(A) = RREF(B) and rank A = rank B.
6. The number of nonzero rows in any row echelon form of A equals rank A.
7. If B is any row echelon form of A, the positions of the leading entries in B are the same as the positions of the leading entries of RREF(A).
8. [Lay03, pp. 17–20] (Gaussian and Gauss–Jordan Elimination Algorithms) When one or more pivots are relatively small, using the algorithms below in floating point arithmetic can yield inaccurate results. (See Chapter 38 for more accurate variations of them, and Chapter 75 for information on professional software implementations of such variations.)

Algorithm 1. Gaussian and Gauss-Jordan Elimination

Let $A \in F^{m\times p}$. Steps 1 to 4 below do Gaussian Elimination, reducing A to a matrix that is in row echelon form. Steps 1 to 6 do Gauss–Jordan Elimination, reducing A to RREF(A).

1. Let $U = A$ and $r = 1$. If $U = \mathbf{0}$, U is in RREF.
2. If $U \neq \mathbf{0}$, search the submatrix of U in rows r to m to find its first nonzero column, k, and the first nonzero entry, a_{ik}, in this column. If $i > r$, exchange rows r and i in U, thus getting a nonzero entry in position (r, k). Let U be the matrix created by this row exchange.
3. Add multiples of row r to the rows below it, to create zeros in column k below row r. Let U denote the new matrix.
4. If either $r = m - 1$ or rows $r + 1, \ldots, m$ are all zero, U is now in REF. Otherwise, let $r = r + 1$ and repeat steps 2, 3, and 4.
5. Let $k_1, \ldots, k_s$ be the pivot columns of U, so $(1, k_1), \ldots, (s, k_s)$ are the pivot positions. For $i = s$, $s - 1, \ldots, 2$, add multiples of row i to the rows above it to create zeros in column k_i above row i.
6. For $i = 1, \ldots, s$, divide row s by its leading entry. The resulting matrix is RREF(A).

Examples:

1. The RREF of a zero matrix is itself, and its rank is zero.

2. Let $A = \begin{bmatrix} 1 & 3 & 4 & -8 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 3 & 4 & -8 \\ 0 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 \end{bmatrix}$. Both are upper triangular, but A is in REF and B is not. Use Gauss–Jordan Elimination to calculate RREF(A) and RREF(B).

   For A, add (-2)(row 2) to row 1 and multiply row 2 by $\frac{1}{2}$. This yields RREF$(A) = \begin{bmatrix} 1 & 3 & 0 & -16 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

   For B, exchange rows 2 and 3 to get $\begin{bmatrix} 1 & 3 & 4 & -8 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$, which is in REF. Then add 2(row 3) to row 1 to get a new matrix. In this new matrix, add (-4)(row 2) to row 1, and multiply row 3 by $\frac{1}{4}$. This yields RREF$(B) = \begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

   Observe that rank $(A) = 2$ and rank $(B) = 3$.

3. Apply Gauss–Jordan Elimination to $A = \begin{bmatrix} 2 & 6 & 4 & 4 \\ -4 & -12 & -8 & -7 \\ 0 & 0 & -1 & -4 \\ 1 & 3 & 1 & -2 \end{bmatrix}$.

   Step 1. Let $U^{(1)} = A$ and $r = 1$.
   Step 2. No row exchange is needed since $a_{11} \neq 0$.

   Step 3. Add (2)(row 1) to row 2, and $(-\frac{1}{2})$(row 1) to row 4 to get $U^{(2)} = \begin{bmatrix} 2 & 6 & 4 & 4 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & -1 & -4 \end{bmatrix}$.

   Step 4. The submatrix in rows 2, 3, 4 is not zero, so let $r = 2$ and return to Step 2.

Step 2. Search the submatrix in rows 2 to 4 of $U^{(2)}$ to see that its first nonzero column is column 3 and the first nonzero entry in this column is in row 3 of $U^{(2)}$. Exchange rows 2 and 3 in $U^{(2)}$ to get

$$U^{(3)} = \begin{bmatrix} 2 & 6 & 4 & 4 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & -4 \end{bmatrix}.$$

Step 3. Add row 2 to row 4 in $U^{(3)}$ to get $U^{(4)} = \begin{bmatrix} 2 & 6 & 4 & 4 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

Step 4. Now $U^{(4)}$ is in REF, so Gaussian Elimination is finished.

Step 5. The pivot positions are $(1, 1)$, $(2, 3)$, and $(3, 4)$. Add –4(row 3) to rows 1 and 2 of $U^{(4)}$ to get $U^{(5)} = \begin{bmatrix} 2 & 6 & 4 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Add –4(row 2) of $U^{(5)}$ to row 1 of $U^{(5)}$ to get $U^{(6)} = \begin{bmatrix} 2 & 6 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

Step 6. Multiply row 1 of $U^{(6)}$ by $\frac{1}{2}$, obtaining $U^{(7)} = \begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, which is RREF$(A)$.

1.4 Systems of Linear Equations

Definitions:

A **linear equation** is an equation of the form $a_1x_1 + \cdots + a_px_p = b$ where $a_1, \ldots, a_p, b \in F$ and $x_1, \ldots, x_p$ are **variables**. The scalars a_j are **coefficients** and the scalar b is the **constant term.**

A **system of linear equations**, or **linear system**, is a set of one or more linear equations in the same variables, such as $\begin{matrix} a_{11}x_1 + \cdots + a_{1p}x_p = b_1 \\ a_{21}x_2 + \cdots + a_{2p}x_p = b_2 \\ \cdots \\ a_{m1}x_1 + \cdots + a_{mp}x_p = b_m \end{matrix}$. A **solution** of the system is a p-tuple $(c_1, \ldots, c_p)$ such that letting $x_j = c_j$ for each j satisfies every equation. The **solution set** of the system is the set of all solutions. A system is **consistent** if there exists at least one solution; otherwise it is **inconsistent**. Systems are **equivalent** if they have the same solution set. If $b_j = 0$ for all j, the system is **homogeneous**. A formula that describes a general vector in the solution set is called the **general solution**.

For the system $\begin{matrix} a_{11}x_1 + \cdots + a_{1p}x_p = b_1 \\ a_{21}x_2 + \cdots + a_{2p}x_p = b_2 \\ \cdots \\ a_{m1}x_1 + \cdots + a_{mp}x_p = b_m \end{matrix}$, the $m \times p$ matrix $A = \begin{bmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mp} \end{bmatrix}$ is the **coefficient matrix**, $\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$ is the **constant vector**, and $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$ is the **unknown vector**. The $m \times (p + 1)$ matrix $[A\ \mathbf{b}]$ is the **augmented matrix** of the system. It is customary to identify the system of linear equations with the matrix-vector equation $A\mathbf{x} = \mathbf{b}$. This is valid because a column vector $\mathbf{x} = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}$ satisfies $A\mathbf{x} = \mathbf{b}$ if and only if $(c_1, \ldots, c_p)$ is a solution of the linear system.

Observe that the coefficients of x_k are stored in column k of A. If $A\mathbf{x} = \mathbf{b}$ is equivalent to $C\mathbf{x} = \mathbf{d}$ and column k of C is a pivot column, then x_k is a **basic variable**; otherwise, x_k is a **free variable**.

Facts:

Let $A\mathbf{x} = \mathbf{b}$ be a linear system, where A is an $m \times p$ matrix.

1. [SIF00, pp. 27, 118] If elementary row operations are done to the augmented matrix $[A\ \mathbf{b}]$, obtaining a new matrix $[C\ \mathbf{d}]$, the new system $C\mathbf{x} = \mathbf{d}$ is equivalent to $A\mathbf{x} = \mathbf{b}$.
2. [SIF00, p. 24] There are three possibilities for the solution set of $A\mathbf{x} = \mathbf{b}$: either there are no solutions or there is exactly one solution or there is more than one solution. If there is more than one solution and F is infinite (such as the real numbers or complex numbers), then there are infinitely many solutions. If there is more than one solution and F is finite, then there are at least $|F|$ solutions.
3. A homogeneous system is always consistent (the zero vector $\mathbf{0}$ is always a solution).
4. The set of solutions to the homogeneous system $A\mathbf{x} = \mathbf{0}$ is a subspace of the vector space F^p.
5. [SIF00, p. 44] The system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ is not a pivot column of $[A\ \mathbf{b}]$, that is, if and only if $\text{rank}([A\ \mathbf{b}]) = \text{rank}\, A$.
6. [SIF00, pp. 29–32] Suppose $A\mathbf{x} = \mathbf{b}$ is consistent. It has a unique solution if and only there is a pivot position in each column of A, that is, if and only if there are no free variables in the equation $A\mathbf{x} = \mathbf{b}$. Suppose there are $t \geq 1$ nonpivot columns in A. Then there are t free variables in the system. If $\text{RREF}([A\ \mathbf{b}]) = [C\ \mathbf{d}]$, then the general solution of $C\mathbf{x} = \mathbf{d}$, hence of $A\mathbf{x} = \mathbf{b}$, can be written in the form $\mathbf{x} = s_1\mathbf{v}_1 + \cdots + s_t\mathbf{v}_t + \mathbf{w}$ where $\mathbf{v}_1, \ldots, \mathbf{v}_t, \mathbf{w}$ are column vectors and $s_1, \ldots, s_t$ are parameters, each representing one of the free variables. Thus $\mathbf{x} = \mathbf{w}$ is one solution of $A\mathbf{x} = \mathbf{b}$. Also, the general solution of $A\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = s_1\mathbf{v}_1 + \cdots + s_t\mathbf{v}_t$.
7. [SIF00, pp. 29–32] (General solution of a linear system algorithm)

Algorithm 2: General Solution of a Linear System $A\mathbf{x} = \mathbf{b}$

This algorithm is intended for small systems using rational arithmetic. It is not the most efficient and when some pivots are relatively small, using this algorithm in floating point arithmetic can yield inaccurate results. (For more accurate and efficient algorithms, see Chapter 38.) Let $A \in F^{m\times p}$ and $\mathbf{b} \in F^{p\times 1}$.

1. Calculate $\text{RREF}([A\ \mathbf{b}])$, obtaining $[C\ \mathbf{d}]$.
2. If there is a pivot in the last column of $[C\ \mathbf{d}]$, stop. There is no solution.
3. Assume the last column of $[C\ \mathbf{d}]$ is not a pivot column, and let $\mathbf{d} = [d_1, \ldots, d_m]^T$.
   a. If $\text{rank}(C) = p$, so there exists a pivot in each column of C, then $\mathbf{x} = \mathbf{d}$ is the unique solution of the system.
   b. Suppose $\text{rank}\, C = r < p$.
      i. Write the system of linear equations represented by the nonzero rows of $[C\ \mathbf{d}]$. In each equation, the first nonzero term will be a basic variable, and each basic variable appears in only one of these equations.
      ii. Solve each equation for its basic variable and substitute parameter names for the $p - r$ free variables, say $s_1, \ldots, s_{p-r}$. This is the general solution of $C\mathbf{x} = \mathbf{d}$ and, thus, the general solution of $A\mathbf{x} = \mathbf{b}$.
      iii. To write the general solution in vector form, as $\mathbf{x} = s_1\mathbf{v}^{(1)} + \cdots + s_{p-r}\mathbf{v}^{(p-r)} + \mathbf{w}$, let (i, k_i) be the i^{th} pivot position of C. Define $\mathbf{w} \in F^p$ by $w_{k_i} = d_i$ for $i = 1, \ldots, r$, and all other entries of $\mathbf{w}$ are 0. Let x_{u_j} be the j^{th} free variable, and define the vectors $\mathbf{v}^{(j)} \in F^p$ as follows:

         For $j = 1, \ldots, p - r$,
         the u_j-entry of $\mathbf{v}^{(j)}$ is 1,
         for $i = 1, \ldots, r$, the k_i-entry of $\mathbf{v}^{(j)}$ is $-c_{iu_j}$,
         and all other entries of $\mathbf{v}^{(j)}$ are 0.

Examples:

1. The linear system $\begin{array}{r} x_1 + x_2 = 0 \\ -x_1 + x_2 = 0 \end{array}$ has augmented matrix $\begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \end{bmatrix}$. The RREF of this is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, which is the augmented matrix for the equivalent system $\begin{array}{l} x_1 = 0 \\ x_2 = 0 \end{array}$. Thus, the original system has a unique solution in $\mathbb{R}^2$, (0,0). In vector form the solution is $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.
2. The system $\begin{array}{l} x_1 + x_2 = 2 \\ x_1 - x_2 = 0 \end{array}$ has a unique solution in $\mathbb{R}^2$, $(1, 1)$, or $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.
3. The system $\begin{array}{r} x_1 + x_2 + x_3 = 2 \\ x_2 + x_3 = 2 \\ x_3 = 0 \end{array}$ has a unique solution in $\mathbb{R}^3$, $(0, 2, 0)$, or $\mathbf{x} = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$.
4. The system $\begin{array}{r} x_1 + x_2 = 2 \\ 2x_1 + 2x_2 = 4 \end{array}$ has infinitely many solutions in $\mathbb{R}^2$. The augmented matrix reduces to $\begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$, so the only equation left is $x_1 + x_2 = 2$. Thus x_1 is basic and x_2 is free. Solving for x_1 and letting $x_2 = s$ gives $x_1 = -s + 2$. Then the general solution is $\begin{array}{l} x_1 = -s + 2 \\ x_2 = s \end{array}$, or all vectors of the form $(-s + 2, s)$. Letting $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, the vector form of the general solution is $\mathbf{x} = \begin{bmatrix} -s + 2 \\ s \end{bmatrix} = s \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \end{bmatrix}$.
5. The system $\begin{array}{r} x_1 + x_2 + x_3 + x_4 = 1 \\ x_2 + x_3 - x_4 = 3 \end{array}$ has infinitely many solutions in $\mathbb{R}^4$. Its augmented matrix $\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & -1 & 3 \end{bmatrix}$ reduces to $\begin{bmatrix} 1 & 0 & 0 & 2 & -2 \\ 0 & 1 & 1 & -1 & 3 \end{bmatrix}$. Thus, x_1 and x_2 are the basic variables, and x_3 and x_4 are free. Write each of the new equations and solve it for its basic variable to see $\begin{array}{l} x_1 = -2x_4 - 2 \\ x_2 = -x_3 + x_4 + 3 \end{array}$. Let $x_3 = s_1$ and $x_4 = s_2$ to get the general solution $\begin{array}{l} x_1 = -2s_2 - 2 \\ x_2 = -s_1 + s_2 + 3 \\ x_3 = s_1 \\ x_4 = s_2 \end{array}$, or $\mathbf{x} = s_1\mathbf{v}^{(1)} + s_2\mathbf{v}^{(2)} + \mathbf{w} = s_1 \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + s_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} -2 \\ 3 \\ 0 \\ 0 \end{bmatrix}$.
6. These systems have no solutions: $\begin{array}{l} x_1 + x_2 = 0 \\ x_1 + x_2 = 1 \end{array}$ and $\begin{array}{r} x_1 + x_2 + x_3 = 0 \\ x_1 - x_2 - x_3 = 0 \\ x_2 + x_3 = 1 \end{array}$. This can be verified by inspection, or by calculating the RREF of the augmented matrix of each and observing that each has a pivot in its last column.

1.5 Matrix Inverses and Elementary Matrices

Invertibility is a strong and useful property. For example, when a linear system $A\mathbf{x} = \mathbf{b}$ has an invertible coefficient matrix A, it has a unique solution. The various characterizations of invertibility in Fact 10 below are also quite useful. Throughout this section, F will denote a field.

Definitions:

An $n \times n$ matrix A is **invertible**, or **nonsingular**, if there exists another $n \times n$ matrix B, called the **inverse** of A, such that $AB = BA = I_n$. The inverse of A is denoted A^{-1} (cf. Fact 1). If no such B exists, A is **not invertible**, or **singular**.

For an $n \times n$ matrix and a positive integer m, the ***m*th power** of A is $A^m = \underbrace{AA\dots A}_{m \text{ copies of } A}$. It is also convenient to define $A^0 = I_n$. If A is invertible, then $A^{-m} = (A^{-1})^m$.

An **elementary matrix** is a square matrix obtained by doing one elementary row operation to an identity matrix. Thus, there are three types:

1. A multiple of one row of I_n has been added to a different row.
2. Two different rows of I_n have been exchanged.
3. One row of I_n has been multiplied by a nonzero scalar.

Facts:

1. [SIF00, pp. 114–116] If $A \in F^{n\times n}$ is invertible, then its inverse is unique.
2. [SIF00, p. 128] (Method to compute A^{-1}) Suppose $A \in F^{n\times n}$. Create the matrix $[A\ I_n]$ and calculate its RREF, which will be of the form $[\text{RREF}(A)X]$. If $\text{RREF}(A) = I_n$, then A is invertible and $X = A^{-1}$. If $\text{RREF}(A) \neq I_n$, then A is not invertible. As with the Gaussian algorithm, this method is theoretically correct, but more accurate and efficient methods for calculating inverses are used in professional computer software. (See Chapter 75.)
3. [SIF00, pp. 114–116] If $A \in F^{n\times n}$ is invertible, then A^{-1} is invertible and $(A^{-1})^{-1} = A$.
4. [SIF00, pp. 114–116] If $A, B \in F^{n\times n}$ are invertible, then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.
5. [SIF00, pp. 114–116] If $A \in F^{n\times n}$ is invertible, then A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.
6. If $A \in F^{n\times n}$ is invertible, then for each $\mathbf{b} \in F^{n\times 1}$, $A\mathbf{x} = \mathbf{b}$ has a unique solution, and it is $\mathbf{x} = A^{-1}\mathbf{b}$.
7. [SIF00, p. 124] If $A \in F^{n\times n}$ and there exists $C \in F^{n\times n}$ such that either $AC = I_n$ or $CA = I_n$, then A is invertible and $A^{-1} = C$. That is, a left or right inverse for a square matrix is actually its unique two-sided inverse.
8. [SIF00, p. 117] Let E be an elementary matrix obtained by doing one elementary row operation to I_n. If that same row operation is done to an $n \times p$ matrix A, the result equals EA.
9. [SIF00, p. 117] An elementary matrix is invertible and its inverse is another elementary matrix of the same type.
10. [SIF00, pp. 126] (Invertible Matrix Theorem) (See Section 2.5.) When $A \in F^{n\times n}$, the following are equivalent:
    - A is invertible.
    - $\text{RREF}(A) = I_n$.
    - $\text{Rank}(A) = n$.
    - The only solution of $A\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$.
    - For every $\mathbf{b} \in F^{n\times 1}$, $A\mathbf{x} = \mathbf{b}$ has a unique solution.
    - For every $\mathbf{b} \in F^{n\times 1}$, $A\mathbf{x} = \mathbf{b}$ has a solution.
    - There exists $B \in F^{n\times n}$ such that $AB = I_n$.
    - There exists $C \in F^{n\times n}$ such that $CA = I_n$.
    - A^T is invertible.
    - There exist elementary matrices whose product equals A.
11. [SIF00, p. 148] and [Lay03, p.132] Let $A \in F^{n\times n}$ be upper (lower) triangular. Then A is invertible if and only if each diagonal entry is nonzero. If A is invertible, then A^{-1} is also upper (lower) triangular, and the diagonal entries of A^{-1} are the reciprocals of those of A. In particular, if L is a unit upper (lower) triangular matrix, then L^{-1} is also a unit upper (lower) triangular matrix.

12. Matrix powers obey the usual rules of exponents, i.e., when A^s and A^t are defined for integers s and t, then $A^sA^t = A^{s+t}, (A^s)^t = A^{st}$.

Examples:

1. For any n, the identity matrix I_n is invertible and is its own inverse. If P is a permutation matrix, it is invertible and $P^{-1} = P^T$.
2. If $A = \begin{bmatrix} 7 & 3 \\ 2 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -3 \\ -2 & 7 \end{bmatrix}$, then calculation shows $AB = BA = I_2$, so A is invertible and $A^{-1} = B$.
3. If $A = \begin{bmatrix} 0.2 & 4 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & -1 \end{bmatrix}$, then $A^{-1} = \begin{bmatrix} 5 & -10 & -5 \\ 0 & 0.5 & 0.5 \\ 0 & 0 & -1 \end{bmatrix}$, as can be verified by multiplication.
4. The matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ is not invertible since $\text{RREF}(A) \neq I_2$. Alternatively, if B is any 2×2 matrix, AB is of the form $\begin{bmatrix} r & s \\ 2r & 2s \end{bmatrix}$, which cannot equal I_2.
5. Let A be an $n \times n$ matrix A with a zero row (zero column). Then A is not invertible since $\text{RREF}(A) \neq I_n$. Alternatively, if B is any $n \times n$ matrix, AB has a zero row (BA has a zero column), so B is not an inverse for A.
6. If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is any 2×2 matrix, then A is invertible if and only if $ad - bc \neq 0$; further, when $ad - bc \neq 0$, $A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. The scalar $ad - bc$ is called the determinant of A. (The determinant is defined for any $n \times n$ matrix in Section 4.1.) Using this formula, the matrix $A = \begin{bmatrix} 7 & 3 \\ 2 & 1 \end{bmatrix}$ from Example 2 (above) has determinant 1, so A is invertible and $A^{-1} = \begin{bmatrix} 1 & -3 \\ -2 & 7 \end{bmatrix}$, as noted above. The matrix $\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ from Example 3 (above) is not invertible since its determinant is 0.
7. Let $A = \begin{bmatrix} 1 & 3 & 0 \\ 2 & 7 & 0 \\ 1 & 1 & 1 \end{bmatrix}$. Then $\text{RREF}([A \quad I_n]) = \begin{bmatrix} 1 & 0 & 0 & 7 & -3 & 0 \\ 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & 0 & 1 & -5 & 2 & 1 \end{bmatrix}$, so A^{-1} exists and equals $\begin{bmatrix} 7 & -3 & 0 \\ -2 & 1 & 0 \\ -5 & 2 & 1 \end{bmatrix}$.

1.6 LU Factorization

This section discusses the *LU* and *PLU* factorizations of a matrix that arise naturally when Gaussian Elimination is done. Several other factorizations are widely used for real and complex matrices, such as the QR, Singular Value, and Cholesky Factorizations. (See Chapter 5 and Chapter 38.) Throughout this section, F will denote a field and A will denote a matrix over F. The material in this section and additional background can be found in [GV96, Sec. 3.2].

Definitions:

Let A be a matrix of any shape.

An LU **factorization**, or **triangular factorization**, of A is a factorization $A = LU$ where L is a square unit lower triangular matrix and U is upper triangular. A ***PLU*** **factorization** of A is a factorization of the form $PA = LU$ where P is a permutation matrix, L is square unit lower triangular, and U is upper triangular. An ***LDU* factorization** of A is a factorization $A = LDU$ where L is a square unit lower triangular matrix, D is a square diagonal matrix, and U is a unit upper triangular matrix.

A ***PLDU* factorization** of A is a factorization $PA = LDU$ where P is a permutation matrix, L is a square unit lower triangular matrix, D is a square diagonal matrix, and U is a unit upper triangular matrix.

Facts: [GV96, Sec. 3.2]

1. Let A be square. If each leading principal submatrix of A, except possibly A itself, is invertible, then A has an LU factorization. When A is invertible, A has an LU factorization if and only if each leading principal submatrix of A is invertible; in this case, the LU factorization is unique and there is also a unique LDU factorization of A.
2. Any matrix A has a PLU factorization. Algorithm 1 (Section 1.3) performs the addition of multiples of pivot rows to lower rows and perhaps row exchanges to obtain an REF matrix U. If instead, the same series of row exchanges are done to A before any pivoting, this creates PA where P is a permutation matrix, and then PA can be reduced to U without row exchanges. That is, there exist unit lower triangular matrices E_j such that $E_k \dots E_1(PA) = U$. It follows that $PA = LU$, where $L = (E_k \dots E_1)^{-1}$ is unit lower triangular and U is upper triangular.
3. In most professional software packages, the standard method for solving a square linear system $A\mathbf{x} = \mathbf{b}$, for which A is invertible, is to reduce A to an REF matrix U as in Fact 2 above, choosing row exchanges by a strategy to reduce pivot size. By keeping track of the exchanges and pivot operations done, this produces a PLU factorization of A. Then $A = P^T LU$ and $P^T LU\mathbf{x} = \mathbf{b}$ is the equation to be solved. Using forward substitution, $P^T L\mathbf{y} = \mathbf{b}$ can be solved quickly for $\mathbf{y}$, and then $U\mathbf{x} = \mathbf{y}$ can either be solved quickly for $\mathbf{x}$ by back substitutution, or be seen to be inconsistent. This method gives accurate results for most problems. There are other types of solution methods that can work more accurately or efficiently for special types of matrices. (See Chapter 7.)

Examples:

1. Calculate a PLU factorization for $A = \begin{bmatrix} 1 & 1 & 2 & 3 \\ -1 & -1 & -3 & 1 \\ 0 & 1 & 1 & 1 \\ -1 & 0 & -1 & 1 \end{bmatrix}$. If Gaussian Elimination is performed on A, after adding row 1 to rows 2 and 4, rows 2 and 3 must be exchanged and the final result is $U = E_3PE_2E_1A = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & -1 & 4 \\ 0 & 0 & 0 & 3 \end{bmatrix}$ where E_1, E_2, and E_3 are lower triangular unit matrices and P is a permutation matrix. This will not yield an LU factorization of A. But if the row exchange is done to A first, by multiplying A by $P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, one gets $PA = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 0 & 1 & 1 & 1 \\ -1 & -1 & -3 & 1 \\ -1 & 0 & -1 & 1 \end{bmatrix}$; then Gaussian Elimination can proceed without any row exchanges. Add row 1 to rows 3 and 4 to get $F_2F_1PA = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & -1 & 4 \\ 0 & 1 & 1 & 4 \end{bmatrix}$ where $F_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ and $F_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$. Then add (-1)(row 2) to row 4 to get $U = F_3\, F_2\, F_1\, PA = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & -1 & 4 \\ 0 & 0 & 0 & 3 \end{bmatrix}$, where $F_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}$.

Note that U is the same upper triangular matrix as before. Finally, $L = (F_3 F_2 F_1)^{-1}$ is unit lower triangular and $PA = LU$ is true, so this is a PLU factorization of A. To get a $PLDU$ factorization, use the same P and L, and define $D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$ and $U = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

2. Let $A = LU = \begin{bmatrix} 1 & 3 & 4 \\ -1 & -1 & -5 \\ 2 & 12 & 3 \end{bmatrix}$. Each leading principal submatrix of A is invertible so A has both LU and LDU factorizations:
$A = LU = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 4 \\ 0 & 2 & -1 \\ 0 & 0 & -2 \end{bmatrix}$. This yields an LDU factorization of A, $\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{bmatrix}$ $\begin{bmatrix} 1 & 3 & 4 \\ 0 & 1 & -0.5 \\ 0 & 0 & 1 \end{bmatrix}$. With the LU factorization, an equation such as $A\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ can be solved efficiently as follows. Use forward substitution to solve $L\mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, getting $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ -8 \end{bmatrix}$, and then backward substitution to solve $U\mathbf{x} = \mathbf{y}$, getting $\mathbf{x} = \begin{bmatrix} -24 \\ 3 \\ 4 \end{bmatrix}$.

3. Any invertible matrix whose $(1, 1)$ entry is zero, such as $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ or $\begin{bmatrix} 0 & -1 & 5 \\ 1 & 1 & 1 \\ 1 & 0 & 3 \end{bmatrix}$, does not have an LU factorization.

4. The matrix $A = \begin{bmatrix} 1 & 3 & 4 \\ -1 & -3 & -5 \\ 2 & 6 & 6 \end{bmatrix}$ is not invertible, nor is its leading principal 2×2 submatrix, but it does have an LU factorization: $A = LU = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 4 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix}$. To find out if an equation such as $A\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ is consistent, notice $L\mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ yields $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ -8 \end{bmatrix}$, but $U\mathbf{x} = \mathbf{y}$ is inconsistent, hence $A\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ has no solution.

5. The matrix $A = \begin{bmatrix} 0 & -1 & 5 \\ 1 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix}$ has no LU factorization, but does have a PLU factorization with
$P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$, and $U = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & 5 \\ 0 & 0 & -4 \end{bmatrix}$.

References

[FIS03] S.H. Friedberg, A.J. Insel, and L.E. Spence. *Linear Algebra*, 3rd ed. Pearson Education, Upper Saddle River, NJ, 2003.

[GV96] G.H. Golub and C.F. Van Loan. *Matrix Computations*, 3rd ed. Johns Hopkins Press, Baltimore, MD, 1996.

[Lay03] David C. Lay. *Linear Algebra and Its Applications*, 3rd ed. Addison Wesley, Boston, 2003.

[Leo02] Steven J. Leon. *Linear Algebra with Applications*, 6th ed. Prentice Hall, Upper Saddle River, NJ, 2003.

[SIF00] L.E. Spence, A.J. Insel, and S.H. Friedberg. *Elementary Linear Algebra*. Prentice Hall, Upper Saddle River, NJ, 2000.