POL502: Linear Algebra

Kosuke Imai

Department of Politics, Princeton University

December 12, 2005

1 Matrix and System of Linear Equations

Definition 1 *A* $m \times n$ *matrix* A *is a rectangular array of numbers with* m *rows and* n *columns and written as*

$$
A = \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \vdots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix}
$$

a_{ij} *is called the* (i,j)*th element of* A.

Note that a special case of matrix is a *vector* where either $m = 1$ or $n = 1$. If $m = 1$ and $n > 1$, then it is called a *row vector*. If $m > 1$ and $n = 1$, then it is called a *column vector*. A vector has a nice geometric interpretation where the direction and length of the vector are determined by its elements. For example, the vector $A = \begin{bmatrix} 1 & 3 \end{bmatrix}$ has the opposite direction and twice as long as the vector $B = \begin{bmatrix} -1/2 & -3/2 \end{bmatrix}$. We will discuss vectors in more detail later in this chapter. Now, we define the basic operations of matrices.

Definition 2 *Let* A *and* B *be* $m \times n$ *matrices.*

1. *(equality)* $A = B$ *if* $a_{ij} = b_{ij}$.

2. *(addition)* $C = A + B$ *if* $c_{ij} = a_{ij} + b_{ij}$ *and* C *is an* $m \times n$ *matrix.*

3. *(scalar multiplication) Given* $k \in \mathbf{R}$, $C = kA$ *if* $c_{ij} = ka_{ij}$ *where* C *is an* $m \times n$ *matrix.*

4. *(product) Let* C *be an* $n \times l$ *matrix.* $D = AC$ *if* $d_{ij} = \sum_{k=1}^{n} a_{ik}c_{kj}$ *and* D *is an* $m \times l$ *matrix.*

5. *(transpose)* $C = A^{\top}$ *if* $c_{ij} = a_{ji}$ *and* C *is an* $n \times m$ *matrix.*

Example 1 *Calculate* $A + 2B^{\top}$, AB, *and* BA *using the following matrices.*

$$
A = \begin{bmatrix} 1 & 2 & -1 \\ 3 & 1 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} -2 & 5 \\ 4 & -3 \\ 2 & 1 \end{bmatrix},
$$

The basic algebraic operations for matrices are as follows:

Theorem 1 (Algebraic Operations of Matrices) *Let* A, B, C *be matrices of appropriate sizes.*

1. *Addition:*

(a) $A + B = B + A$ and $A + (B + C) = (A + B) + C$

(b) There exists a unique C such that $A + C = A$ and we denote $C = \mathbf{O}$.

(c) There exists a unique C such that $A + C = \mathbf{O}$ and we denote $C = -A$.

2. Multiplication:

(a) $k(lA) = (kl)A$, $k(A + B) = kA + kB$, $(k + l)A = kA + lA$, and $A(kB) = k(AB) = (kA)B$ for any $k, l \in \mathbf{R}$.

(b) $A(BC) = (AB)C$.

(c) $(A + B)C = AC + BC$ and $C(A + B) = CA + CB$.

3. Transpose:

(a) $(A^\top)^\top = A$.

(b) $(A + B)^\top = A^\top + B^\top$.

(c) $(AB)^\top = B^\top A^\top$.

(d) $(kA)^\top = kA^\top$.

Example 2 *Calculate $(ABC)^\top$ using the following matrices.*

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -2 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 2 & 2 \\ 3 & -1 \end{bmatrix}, \quad C = \begin{bmatrix} 3 & 2 \\ 3 & -1 \end{bmatrix},$$

There are some important special types of matrices.

Definition 3 *Let A be an $m \times n$ matrix.*

1. *A is called a square matrix if $n = m$.*

2. *A is called symmetric if $A^\top = A$.*

3. *A square matrix A is called a diagonal matrix if $a_{ij} = 0$ for $i \neq j$. A is called upper triangular if $a_{ij} = 0$ for $i > j$ and called lower triangular if $a_{ij} = 0$ for $i < j$.*

4. *A diagonal matrix A is called an identity matrix if $a_{ij} = 1$ for $i = j$ and is denoted by I_n.*

In particular, we have $AI_n = I_n A = A$ for any square matrix A. For a square matrix, there is another operator called trace.

Definition 4 *If A is an $n \times n$ matrix, then the trace of A denoted by $\mathrm{tr}(A)$ is defined as the sum of all the main diagonal elements of A. That is, $\mathrm{tr}(A) = \sum_{i=1}^n a_{ii}$.*

Some useful facts about trace operators are given below.

Theorem 2 (Trace Operator) *Let A and B be matrices of appropriate sizes.*

1. $\mathrm{tr}(kA) = k\,\mathrm{tr}(A)$ *for any $k \in \mathbf{R}$.*

2. $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$.

3. $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

4. $\mathrm{tr}(\mathrm{A}^\top) = \mathrm{tr}(\mathrm{A})$.

5. $\mathrm{tr}(\mathrm{A}^\top \mathrm{A}) \geq 0$.

If we start out with an $m \times n$ matrix and delete some, but not all, of its rows or columns, then we obtain a *submatrix*. For example, C in Example 2 is a *submatrix* of the following matrix.

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 3 & -1 & 2 \end{bmatrix}$$

A matrix can be partitioned into submatrices, and such a matrix is called *partitioned* matrices. Partitioned matrices can be manipulated in the same way (called *block manipulation*) provided that submatrices are of appropriate sizes.

Example 3

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, \quad AB = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}.$$

We now study a special type of square matrices and formulate the notion corresponding to the reciprocal of a nonzero real number.

Definition 5 *An $n \times n$ matrix A is called nonsingular or invertible if there exists an $n \times n$ matrix B such that $AB = BA = I_n$ We call such B an inverse of A. Otherwise, A is called singular or noninvertible.*

Example 4 *Show that B is an inverse of A (or A is an inverse of B).*

$$A = \begin{bmatrix} 2 & 3 \\ 2 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & \frac{3}{2} \\ 1 & -1 \end{bmatrix}.$$

We prove some important properties about the inverse of a matrix.

Theorem 3 (Uniqueness of Inverse) *The inverse of a matrix, if it exists, is unique.*

We denote the unique inverse of A by A^{-1}.

Theorem 4 (Properties of Inverse) *Let A and B be nonsingular $n \times n$ matrices.*

1. *AB is nonsingular and $(AB)^{-1} = B^{-1}A^{-1}$.*

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

3. *$(A^\top)^{-1} = (A^{-1})^\top$.*

One application of inverting a matrix is to solve a system of linear equations. In fact, matrices can be motivated in terms of linear equations. Consider a set of m linear equations of the form

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ &\vdots \qquad\quad \vdots \\ y_m &= a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{aligned}$$

3

Then, its matrix representation is $Y = AX$ where

$$
A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.
$$

We call A a *coefficient matrix*. With this notation, we can see that A^{-1} (provided that A is nonsingular) solves this system since we obtain $X = A^{-1}Y$ by premultiplying the equation by A^{-1}.

Example 5 *Confirm that*

$$
\begin{bmatrix} 2 & 3 \\ 2 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} -1 & \frac{3}{2} \\ 1 & -1 \end{bmatrix}
$$

solve the following system of linear equations by using the inverse of matrix.

$$
\begin{aligned}
2x_1 + 3x_2 &= 1 \\
2x_1 + 2x_2 &= 2
\end{aligned}
$$

Since we do not yet know how to find the inverse of a matrix in general, we rely on high-school algebra to solve a system of linear equations. To formalize what we mean by "high-school algebra", we introduce the following definitions.

Definition 6 *Elementary row (column) operations on an $m \times n$ matrix A includes the following.*

1. *Interchange rows (columns) r and s of A.*

2. *Multiply row (column) r of A by a nonzero scalar $k \neq 0$.*

3. *Add k times row (column) r of A to row (column) s of A where $r \neq s$.*

An $m \times n$ matrix A is said to be row (column) equivalent to an $m \times n$ matrix B if B can be obtained by applying a finite sequence of elementary row (column) operations to A.

Example 6 *Show that A is row equivalent to B.*

$$
A = \begin{bmatrix} 1 & 2 & 4 & 3 \\ 2 & 1 & 3 & 2 \\ 1 & -1 & 2 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 4 & 8 & 6 \\ 1 & -1 & 2 & 3 \\ 4 & -1 & 7 & 8 \end{bmatrix},
$$

Now, we can use these operations to characterize systems of linear equations.

Theorem 5 (Row Equivalence and Linear Equations) *Let $AX = B$ and $CX = D$ be two linear systems with m equations and n unknowns. If the augmented matrices $[A\ B]$ and $[C\ D]$ are row equivalent, then the linear systems have the same solutions.*

Finally, to solve systems of linear equations using high-school algebra, we need one more concept.

Definition 7 *An $m \times n$ matrix A is said to be in reduced row echelon form if it satisfies the following properties.*

1. *All rows consisting entirely of zeros, if any, are at the bottom of the matrix.*

2. *By reading from left to right, the first nonzero entry in each row that does not consist entirely of zeros is a 1, called the leading entry of its row.*

3. *If rows i and $i+1$ are two successive rows that do not consist entirely of zeros, then the leading entry of row $i+1$ is to the right of the leading entry of row i.*

4. *If a column contains a leading entry of some row, then all other entries in that column are zero.*

If A satisfies 1, 2, and 3, but not 4, then it is said to be in row echelon form. A similar definition can be applied to (reduced) column echelon form.

Example 7 *A is in row echelon form, and B is in reduced row echelon form.*

$$A = \begin{bmatrix} 1 & 5 & 0 & 2 & -2 & 4 \\ 0 & 1 & 0 & 3 & 4 & 8 \\ 0 & 0 & 0 & 1 & 7 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 0 & 0 & -2 & 4 \\ 0 & 1 & 0 & 0 & 4 & 8 \\ 0 & 0 & 0 & 1 & 7 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

Finally, we define two methods that can be used to solve systems of linear equations.

Theorem 6 (Gaussian and Gauss-Jordan Elimination) *A system of linear equations $AX = Y$ can be solved by using Gaussian (Gauss-Jordan) elimination, which consists of the following two steps:*

1. *Use elementary operations to transform the augmented matrix $[AB]$ to the matrix $[CD]$ in (reduced) row echelon form.*

2. *Solve the linear system corresponding to the augmented matrix $[CD]$ using back substitution.*

Example 8 *Solve the following system of linear equations using the Gaussian elimination.*

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 9 \\ 2x_1 - x_2 + x_3 &= 8 \\ 3x_1 - x_3 &= 3 \end{aligned}$$

2 Determinant and Inverse of Matrix

In this section, we will learn how to find the inverse of a matrix.

Definition 8 *A permutation of a finite set of integers $S = \{1, 2, \ldots, n\}$ is a bijective function $f : S \mapsto S$. A permutation is said to have an inversion if a larger integer precedes a smaller one. A permutation is called even (odd) if the total number of inversions is even (odd).*

That is, if $S = \{1, 2, 3\}$, then f defined by $f(1) = 3, f(2) = 2, f(3) = 1$ is an odd permutation. Now, we are ready to define determinant of a matrix.

Definition 9 *Let A be an $n \times n$ matrix. Then, the determinant of A denoted by $|A|$ or $\det(A)$ is $\sum(\pm)a_{1f(1)}a_{2f(2)} \ldots a_{nf(n)}$ where the summation is over all permutations $f : S \mapsto S$ with $S = \{1, 2, \ldots, n\}$. The sign is $+$ $(-)$ if the corresponding permutation is even (odd).*

Now, we compute the determinants of the following matrices. It should be noted that there is no easy method for computing determinants for $n > 3$.

Example 9 *What are the determinants of 1×1, 2×2, and 3×3 matrices?*

We examine some basic properties of determinants. In particular, there is an important relationship between the singularity and the determinant of a matrix.

Theorem 7 (Determinants) *Let A and B be $n \times n$ matrices.*

1. *$|I_n| = 1$ and $|-I_n| = (-1)^n$.*

2. *$|kA| = k^n|A|$ for $k \in \mathbf{R}$.*

3. *$|A| = |A^\top|$.*

4. *A is nonsingular if and only if $|A| \neq 0$.*

5. *$|AB| = |A||B|$.*

6. *If A is nonsingular, then $|A^{-1}| = |A|^{-1}$.*

According to Definition 9, computing the determinant of an $n \times n$ matrix can be very cumbersome if n is large. We now develop a method which reduces the problem to the computation of the determinant of an $(n-1) \times (n-1)$ matrix so that we can repeat the process until we get to a 2×2 matrix.

Definition 10 *Let A be an $n \times n$ matrix.*

1. *Let M_{ij} be the $(n-1) \times (n-1)$ submatrix of A obtained by deleting the ith row and jth column of A. Then, $|M_{ij}|$ is called the minor of a_{ij}.*

2. *The cofactor of A_{ij} of a_{ij} is defined as $A_{ij} = (-1)^{i+j}|M_{ij}|$*

Now, the following theorem gives us a new method to compute determinants.

Theorem 8 (Cofactor Expansion) *Let A be an $n \times n$ matrix. Then, for any i and j, $|A| = \sum_{j=1}^{n} a_{ij}A_{ij}$ and $|A| = \sum_{i=1}^{n} a_{ij}A_{ij}$*

Example 10 *Find the determinant of the following matrix using cofactor expansion.*

$$A = \begin{bmatrix} 1 & 2 & -3 & 4 \\ -4 & 2 & 1 & 3 \\ 3 & 0 & 0 & -3 \\ 2 & 0 & -2 & 3 \end{bmatrix}.$$

The following will show how one can use cofactors to calculate the inverse of a matrix.

Definition 11 *Let A be an $n \times n$ matrix. The adjoint of A, $adj\,A$, is the matrix whose (i,j) element is the cofactor A_{ji} of a_{ji}. That is,*

$$adj\,A = \begin{bmatrix} A_{11} & A_{21} & \ldots & A_{n1} \\ A_{12} & A_{22} & \ldots & A_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ A_{1n} & A_{2n} & \ldots & A_{nn} \end{bmatrix}.$$

Example 11 *Compute the adjoint of the following matrix.*

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

Finally, the inverse of a square matrix can be written as follows.

Theorem 9 (Inverse of a Matrix) *If A is an $n \times n$ matrix and $|A| \neq 0$, then*

$$A^{-1} = \frac{1}{|A|} adj\,A$$

The theorem illustrates why $|A| \neq 0$ is required for A^{-1} to exist.

Example 12 *Compute the inverse of A in Example 11.*

Now, you can solve a system of linear equations provided its solution exists: i.e., the inverse of the coefficient matrix exists. We introduce another method to solve a system of linear equations.

Theorem 10 (Cramer's Rule) *Consider a system of n linear equations in n unknown parameters with the coefficient matrix A so that we can write $Y = AX$*

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ &\vdots \qquad\qquad \vdots \\ y_n &= a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n \end{aligned}$$

If $|A| \neq 0$, then the system has the unique solution

$$x_1 = \frac{|A_1|}{|A|}, \quad x_2 = \frac{|A_2|}{|A|}, \quad \ldots, \quad x_n = \frac{|A_n|}{|A|},$$

where A_i is the matrix obtained from A by replacing its ith column by Y.

Example 13 *Apply the Cramer's Rule to the following system of linear equations.*

$$
\begin{aligned}
-2x_1 + 3x_2 - x_3 &= 1 \\
x_1 + 2x_2 - x_3 &= 4 \\
-2x_1 - x_2 + x_3 &= -3
\end{aligned}
$$

3   Real Vector Spaces

In the very first chapter, we learned the concept of ordered pairs. We can give a geometric representation to an ordered pair of real numbers, (x, y) with $x, y \in \mathbf{R}$. This is what we call *Euclidean 2-space*. In particular, we consider a pair of perpendicular lines called *coordinate axes* (one horizontal line called *x-axis* and the other vertical one called *y-axis*) intersecting at a point O, called the *origin*. Then, we can associate each point in the plane with each ordered pair of real numbers. The set of all such points are denoted by $\mathbf{R}^2$. We can also associate each ordered pair with the directed line segment called *vector* from the origin to a point $P = (x, y)$. Such a vector is denoted by $\overrightarrow{OP}$ where O is the *tail* of the vector and P is its *head*. We generalize this idea.

Definition 12 *A vector in the plane from the point $P = (x_1, y_1)$ to $Q = (x_2, y_2)$ is a 2×1 matrix with two components $(x_2 - x_1, y_2 - y_1)$. We denote it by $\overrightarrow{PQ}$.*

This indicates that $\overrightarrow{PQ}$ is equal to another vector $\overrightarrow{OP'}$ from the origin where $P' = (x_2 - x_1, y_2 - y_1)$. In other words, a vector can be described by its direction and its length alone. Now, it is important to develop a geometric understanding of the basic vector operations.

Definition 13 *Let $u = (x_1, y_1)$ and $v = (x_2, y_2)$ be vectors in the plane. Then,*

1. *(equality) $u = v$ if $x_1 = x_2$ and $y_1 = y_2$.*

2. *(addition) $u + v = (x_1 + x_2, y_1 + y_2)$.*

3. *(scalar multiplication) $ku = (kx_1, ky_1)$ for any $k \in \mathbf{R}$.*

Example 14 *Consider $u = (2, 3)$, $v = (3, -4)$, and $w = (1, 2)$. Draw $u + v$, $u - w$, and $3w$.*

We can generalize these definitions to vectors in space, $\mathbf{R}^3$, by adding the third coordinate axis, *z-axis*. A vector in space is 3×1 matrix. It is also possible to extend beyond 3 dimensions and talk about $\mathbf{R}^n$, which is called *real vector spaces*. The following basic algebraic operations hold for any real vector spaces.

Axiom 1 *Let u, v, and w be any vectors in $\mathbf{R}^n$, and let $c, k \in \mathbf{R}$. Then,*

1. $u + v = v + u$

2. $u + (v + w) = (u + v) + w$

3. $u + \mathbf{0} = \mathbf{0} + u = u$ *where* $\mathbf{0} = (0, \ldots, 0)$.

4. $u + (-u) = \mathbf{0}$

5. $c(u + v) = cu + cv$

6. $(c + k)u = cu + ku$

7. $c(ku) = (ck)u$

8. $1u = u$

Just like we did in the first chapter for real numbers, these axioms are sufficient to give basic results such as $0u = \mathbf{0}$, $c\mathbf{0} = \mathbf{0}$, and $(-1)u = -u$. Now, we are ready to study the structure of real vector spaces. First, we introduce some key concepts. The length of a vector in plane can be defined as follows.

Theorem 11 (Pythagorean Theorem) *The length of the vector $v = (v_1, v_2)$ in $\mathbf{R}^2$ is $\sqrt{v_1^2 + v_2^2}$, and is denoted by $||v||$.*

A vector whose length is 1 is called a *unit vector*. Similarly, one can define the distance between two vectors, $u = (u_1, u_2)$ and $v = (v_1, v_2)$, as $||u - v|| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2}$. This definition can be generalized easily to real vector spaces.

Example 15 *Compute the length of the vector $v = (1, 2, 3)$, and the distance between v and $u = (-4, 3, 5)$.*

Another important concept is the inner product of two vectors.

Definition 14 *Let $u = (u_1, u_2)$ and $v = (v_1, v_2)$ be vectors in plane.*

    1. The inner product (or dot product) is defined on $\mathbf{R}^2$ as $u_1 v_1 + u_2 v_2$ and is denoted by $u \cdot v$.

    2. u and v are said to be orthogonal if $u \cdot v = 0$.

Again, it is easy to generalize this definition to real vector spaces.

Example 16 *Compute the inner product of $u = (2, 3, 2)$ and $v = (4, 2, -1)$.*

Observe that if we view the vectors u and v as $n \times 1$ matrices, then we can obtain another expression for the inner product, $u^\top v$.

Theorem 12 (Inner Product) *Let u, v, and w be vectors in $\mathbf{R}^n$, and $k \in \mathbf{R}$.*

    1. $u \cdot v = v \cdot u$.

    2. $u \cdot (v + w) = u \cdot v + u \cdot w$.

    3. $u \cdot (kv) = k(u \cdot v) = (ku) \cdot v$.

    4. $u \cdot u \geq 0$ and $u \cdot u = 0$ implies $u = 0$.

An important application of the inner product is the following theorem.

Theorem 13 (Cauchy-Schwartz Inequality) *Let u and v be vectors in $\mathbf{R}^n$. Then,*

$$(u \cdot v)^2 \leq ||u||^2 ||v||^2$$

Using this theorem, we can prove a more general version of the triangular inequalities that we proved for the real number system.

Theorem 14 (Triangular Inequalities) *Let u and v be vectors in $\mathbf{R}^n$. Then,*

    1. $||u + v|| \leq ||u|| + ||v||$.

    2. $|||u|| - ||v||| \leq ||u - v||$.

For example, if you replace u and v with a scalar, you obtain usual triangular inequalities for the real number system.

4 Linear Independence

In this section, we further examine the structure of real vector spaces.

Definition 15 *Let $v_1, v_2, \ldots, v_k$ be vectors in a real vector space V.*

1. *A vector v in V is called a linear combination of $v_1, v_2, \ldots, v_k$ if $v = a_1 v_1 + a_2 v_2 + \ldots + a_k v_k$ for some real numbers $a_1, a_2, \ldots, a_k$.*

2. *A set of vectors $S = \{v_1, v_2, \ldots, v_k\}$ is said to span V if every vector in V is a linear combination of $v_1, v_2, \ldots, v_k$.*

3. *$v_1, v_2, \ldots, v_k$ are linearly dependent if there exist constants $a_1, a_2, \ldots, a_k$ not all zero such that $a_1 v_1 + a_2 v_2 + \ldots + a_k v_k = 0$. Otherwise, $v_1, v_2, \ldots, v_k$ are called linearly independent.*

In other words, vectors $v_1, \ldots, v_k$ in $\mathbf{R}^n$ are linearly dependent if and only if the linear system $AX = \mathbf{0}$ with $A = [v_1 v_2 \ldots v_k]$ and $X = (a_1, \ldots, a_k)$ has a nonzero solution (or equivalently $|A| \neq 0$). A set of vectors which spans a real vector space V completely describes V because every vector in V can be constructed as a linear combination of the vectors in that set. Let's apply these concepts to some examples.

Example 17 *Answer the following two questions.*

1. *Do $v_1 = (1, 2, 1)$, $v_2 = (1, 0, 2)$, and $v_3 = (1, 1, 0)$ span $\mathbf{R}^3$?*

2. *Are $v_1 = (1, 0, 1, 2)$, $v_2 = (0, 1, 1, 2)$, and $v_3 = (1, 1, 1, 3)$ in $\mathbf{R}^4$ linearly dependent?*

Notice also that any set of k vectors in $\mathbf{R}^n$ is linearly dependent if $k > n$.

Theorem 15 (Linear Independence) *Let V be a real vector space.*

1. *Let S_1 and S_2 be finite subsets of V. Suppose $S_1 \subset S_2$. If S_1 is linearly dependent, so is S_2. If S_2 is linearly independent, so is S_1.*

2. *The nonzero vectors $v_1, v_2, \ldots, v_k$ in V are linearly dependent if and only if one of the vectors v_j is a linear combination of the preceding vectors $v_1, v_2, \ldots, v_{j-1}$.*

There can be many sets of vectors that describe a given real vector space V. In particular, such a set can contain vectors which constitute a linear combination of one another. To obtain a "minimal" set of vectors which completely describes V, we develop the following concept.

Definition 16 *The vectors $v_1, v_2, \ldots, v_k$ in a real vector space V are said to form a basis if they span V and are linearly independent.*

For example, $v_1 = (1, 0, 0)$, $v_2 = (0, 1, 0)$, and $v_3 = (0, 0, 1)$ form a *natural basis* for $\mathbf{R}^3$. Note that there can be infinitely many bases for a given real vector space. For example, $c_1 v_1$, $c_2 v_2$, and $c_3 v_3$ where $c_1, c_2, c_3 \in \mathbf{R}$ also form a basis.

Example 18 *Show that the set $S = \{v_1, v_2, v_3, v_4\}$ where $v_1 = (1, 0, 1, 0)$, $v_2 = (0, 1, -1, 2)$, $v_3 = (0, 2, 2, 1)$, and $v_4 = (1, 0, 0, 1)$ forms a basis for $\mathbf{R}^4$.*

Now, we show the main result about a basis of a real vector space. Namely, although there exist infinitely many bases for a given real vector space, all the bases have the same number of vectors.

11

Theorem 16 (Basis) *Let $v_1, v_2, \ldots, v_k$ and $w_1, w_2, \ldots, w_k$ be vectors in a real vector space V.*

1. *If $S = \{v_1, v_2, \ldots, v_k\}$ is a basis for V, then every vector in V can be written in one and only one linear combination of the vectors in S.*

2. *If $S = \{v_1, v_2, \ldots, v_k\}$ is a basis for V and $T = \{w_1, w_2, \ldots, w_n\}$ is a linearly independent set of vectors in V, then $n \leq k$.*

3. *If $S = \{v_1, v_2, \ldots, v_k\}$ and $T = \{w_1, w_2, \ldots, w_n\}$ are bases for V, then $k = n$.*

The last result of the previous theorem implies that the number of vectors in two different bases for a particular real vector space is the same. In particular, every basis of $\mathbf{R}^n$ contains n vectors. This amounts to the following concept.

Definition 17 *The dimension of a real vector space V is the number of vectors in a basis for V. We often write $\dim V$.*

Example 19 *What is the dimension of the real vector space V spanned by $S = \{v_1, v_2, v_3\}$ where $v_1 = (0, 1, 1), v_2 = (1, 0, 1), v_3 = (1, 1, 2)$.*

Next, We introduce a method to find a basis for the real vector space spanned by a set of vectors.

Definition 18 *Let A be an $m \times n$ matrix. The rows of A, considered as vectors in $\mathbf{R}^n$, span a subspace of $\mathbf{R}^n$, called the row space of A. Similarly, the columns of A span a subset of $\mathbf{R}^m$ called the column space of A.*

Theorem 17 (Row and Column Spaces) *If A and B are two $m \times n$ row (column) equivalent matrices, then the row (column) spaces of A and B are identical.*

Example 20 *Find a basis for the subspace V of $\mathbf{R}^5$ that is spanned by $S = \{v_1, v_2, v_3, v_4\}$ where $v_1 = (1, -2, 0, 3, -4)$, $v_2 = (3, 2, 8, 1, 4)$, $v_3 = (2, 3, 7, 2, 3)$, and $v_4 = (-1, 2, 0, 4, -3)$.*

Definition 19 *Let A be an $m \times n$ matrix.*

1. *The dimension of the row (column) space of A is called the row (column) rank of A.*

2. *If the row (column) rank of A is equal to m (n), it is said to be of full row (column) rank.*

It follows from the definition that if A and B are row (column) equivalent, then row (column) ranks of A and B are the same.

Example 21 *Compute the row rank of a matrix $A = [v_1 \ v_2 \ v_3 \ v_4]$ where the vectors are defined in Example 20*

There is the important relationship between rank and singularity. We give the following theorem without proof.

Theorem 18 (Rank and Singularity) *Let A be a $n \times n$ square matrix. A is nonsingular if and only if A is of full rank, i.e., $\mathrm{rank} A = n$.*

5 Eigenvalues, Eigenvectors, and Definiteness

As the final topic of this chapter, we study eigenvalues and eigenvectors. Although we will not prove many of the theorems, their results are important and will be frequently applied in statistics. First, we give the definition of eigenvalues and eigenvectors.

Definition 20 *Let A be an $n \times n$ square matrix. $\lambda \in \mathbf{R}$ is called an eigenvalue of A if there exists a nonzero vector x such that $Ax = \lambda x$. Every nonzero vector $x \in \mathbf{R}^n$ satisfying this equation is called an eigenvector of A associated with the eigenvalue λ.*

Note that $x = \mathbf{0}$ always satisfies the equation, but it is not an eigenvector.

Example 22 *Confirm that $\lambda_1 = 2$ and $\lambda_2 = 3$ are the eigenvalues and $x_1 = (1, 1)$ and $x_2 = (1, 2)$ are their associated eigenvectors of $\begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}$.*

The connection between eigenvalues and singularity is critical.

Theorem 19 (Eigenvalues) *Let A be an $n \times n$ square matrix and $\lambda_1, \ldots, \lambda_n$ be its eigenvalues.*

1. *If A is diagonal, then its diagonal elements are the eigenvalues.*

2. *$\sum_{i=1}^{n} \lambda_i = \mathrm{tr}(\mathrm{A})$.*

3. *$\prod_{i=1}^{n} \lambda_i = |A|$. In particular, A is singular if and only if 0 is an eigenvalue of A.*

Before we state the key theorem, we need one more concept.

Definition 21 *Let A be an $n \times n$ square matrix. Then, $|\lambda I_n - A|$ is called the characteristic polynomial of A. The equation $|\lambda I_n - A| = 0$ is called the characteristic equation of A.*

The next theorem shows how one can find eigenvalues.

Theorem 20 (Characteristic Polynomial) *Let A be an $n \times n$ matrix. The eigenvalues of A are the real roots of the characteristic polynomial of A. A is said to be diagonalizable if all the roots of its characteristic polynomial are real and distinct.*

The word "diagonalizable" comes from the fact that the diagonal matrix whose nonzero elements are the eigenvalues of A represent a linear transformation, a function mapping from one real vector space to another. $L : V \mapsto W$ is a linear transformation if it satisfies $L(v + w) = L(v) + L(w)$ and $L(cv) = cL(v)$ for any vector $v \in V$ and $w \in W$. One important linear transformation is what is called *projection* defined by $L : \mathbf{R}^3 \mapsto \mathbf{R}^2$ and represented by the matrix $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ so that for any vector $v = (v_1, v_2, v_3) \in \mathbf{R}^3$ we have $L(v) = (v_1, v_2)$.

Example 23 *Find the eigenvalues and eigenvectors, if they exist, of $\begin{bmatrix} 1 & 2 & -1 \\ 1 & 0 & 1 \\ 4 & -4 & 5 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.*

Definition 22 *An $n \times n$ square matrix A is called orthogonal if $A^\top A = I_n$.*

In particular, the orthogonal matrix is invertible, and hence $A^{-1} = A^\top$. We end this chapter with the concept of definiteness of a matrix.

Definition 23 *Let A be an $n \times n$ symmetric matrix A and x be a column vector in $\mathbf{R}^n$.*

1. *A is positive definite (positive semidefinite) if $x^\top A x > 0$ ($x^\top A x \geq 0$) for any $x \neq \mathbf{0}$.*

2. *A is negative definite (negative semidefinite) if $x^\top A x < 0$ ($x^\top A x \leq 0$) for any $x \neq \mathbf{0}$.*

If A fits none of the above definitions, then it is called indefinite. From the definition, one immediately sees that A is positive (negative) definite if and only if $-A$ is negative (positive) definite. Positive (negative) definite matrices have some important properties.

Theorem 21 (Positive and Negative Definiteness) *Let A be an $n \times n$ symmetric matrix A*

1. *A is positive (negative) definite if and only if all of its eigenvalues are positive (negative).*

2. *If A is positive (negative) definite, then A is invertible and A^{-1} is postive (negative) definite.*

3. *If A is positive definite, then $|A| > 0$.*

4. *If A is positive definite, then there exists a upper triangular matrix U such that $A = U^\top U$ (Cholesky decomposition).*

The first part of the theorem implies that if A is a diagonal matrix and positive definite, then all of its diagonal elements must be positive. The Cholesky decomposition can be seen as a "square root" of a symmetric matrix. One way to check the definiteness of matrices is to compute eigen values and use Theorem 21 (1). However, there is an easier way to check the definiteness.

Definition 24 *Let A be an $n \times n$ square matrix and k be an integer with $1 \leq k \leq n$.*

1. *The kth order principal submatrix of A is an $k \times k$ submatrix of A which can be constructed by deleting the $n - k$ columns and the same $n - k$ rows. If the last $n - k$ columns and rows are deleted, then the resulting submatrix is called the kth order leading principal submatrix.*

2. *The determinant of a kth order (leading) principal submatrix of A is called a kth order (leading) principal minor of A.*

There is a special relationship between principal minors and definiteness of a matrix.

Theorem 22 (Principal Minor and Definiteness) *Let A be an $n \times n$ symmetric matrix and k be an integer with $1 \leq k \leq n$.*

1. *A is positive definite if and only if the kth order leading principal minor is positive for all k.*

2. *A is positive semi-definite if and only if every principal minor is greater than or equal to zero for all k.*

3. *A is negative definite if and only if the kth order leading principal minor has the same sign as $(-1)^k$ for all k.*

4. *A is negative semi-definite if and only if every kth order principal minor is zero or has the same sign as $(-1)^k$ for all k.*

An example is given below.

Example 24 *Check the definiteness of the following matrices.*

$$A = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 5 \\ 0 & 5 & 6 \end{bmatrix}$$

We are done with linear algebra!