LINEAR ALGEBRA NOTES

Philippos Dimitroglou

Table of Contents

Math 360 - Linear Algebra

Chapter 1 – Linear Equations in Linear Algebra

Section 1.1 - Systems of Linear Equations

A **linear equation** in the variables $x_1, x_2, \dots, x_n$ is an equation that can be written in the form $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$ where b and the **coefficients** $a_1, \dots, a_n$ are real or complex numbers, usually known in advance.

A **system of linear equations** (or a **linear system**) is a collection of one or more linear equations involving the same variables.

A **solution** of the system is a list $(s_1, s_2, \dots, s_n)$ of numbers that makes each equation a true statement when the values $s_1, \dots, s_n$ are substituted for $x_1, \dots, x_n$ respectively.

The set of all possible solutions is called the **solution set** of the linear system

Two linear systems are **equivalent** if they have the same solution set.

A system of linear equations has

1. No solution or
2. Exactly one solution or
3. Infinitely many solutions

A system of linear equations is said to be **consistent** if it has either one solution of infinitely many solutions; a system is **inconsistent** if it has no solution.

The essential information of a linear system can be recorded compactly in a rectangular array called a **matrix**.

With the coefficients of each variable aligned in columns, the matrix is called the **coefficient matrix** (or **matrix of coefficients**) and is called an **augmented matrix** if there is an additional column containing the constants of the right sides of the equations.

The **size** of a matrix tells how many rows and columns it has. An $m \times n$ **matrix** is a rectangular array of numbers with m rows and n columns.

The basic strategy of solving a linear system is *to replace one system with an equivalent system that is easier to solve*.

Elementary Row Operations

1. **Replacement** – replace one row by the sum of itself and a multiple of another row; in other words, add to one row a multiple of another row (ex. notated as $R_1 \leftarrow 2R_1 + 3R_2$)
2. **Interchange** – interchange two rows (ex. notated as $R_1 \leftrightarrow R_2$)
3. **Scaling** – multiply all entries in a row by a nonzero constant (ex. notated as $R_1 \leftarrow 2R_1$)

Two matrices are called **row equivalent** if there is a sequence of elementary row operations that transforms one matrix into the other.

If the augmented matrices of two linear systems are row equivalent, then the two systems have the same solution set.

Two Fundamental Questions About a Linear System

1. Is the system consistent; that is, does at least one solution *exist*?
2. If a solution exists, is it the only one; that is, is the solution *unique*?

Section 1.2 – Row Reduction and Echelon Forms

A **nonzero** row or column in a matrix means a row or column that contains at least one nonzero entry; a **leading entry** of a row refers to the leftmost nonzero entry in a nonzero row

A rectangular matrix is in **echelon form** (or **row echelon form**) if it has the following three properties:

1. All nonzero rows are above any rows of all zeros
2. Each leading entry of a row is in a column to the right of the leading entry of the row above it
3. All entries in a column below a leading entry are zeros

If a matrix in echelon form satisfies the following additional conditions, then it is in **reduced echelon form** (or **reduced row echelon form**):

4. The leading entry in each nonzero row is 1
5. Each leading 1 is the only nonzero entry in its column

An **echelon matrix** (respectively, a **reduced echelon matrix**) is one that is in echelon form (respectively, reduced echelon form)

Any nonzero matrix may be **row reduced** (that is transformed by elementary row operations) into more than one matrix in echelon form, using different sequences of row operations. However, reduced echelon form one obtains from a matrix is unique.

Theorem 1.1 (Uniqueness of the Reduced Echelon Form)
Each matrix is row equivalent to one and only one reduced echelon matrix.

If a matrix A is row equivalent to an echelon matrix U, we call U an **echelon form** (or row echelon form) **of** A; if U is in reduced echelon form, we call U the **reduced echelon form of** A.

A **pivot position** in a matrix A is a location in A that corresponds to a leading 1 in the reduced echelon form of A. A **pivot column** is a column of A that contains a pivot positon.

A **pivot** is a nonzero number in a pivot positon that is used as needed to create zeros via row operations.

The Row Reduction Algorithm

Step 1: Begin with the leftmost nonzero column. This is a pivot column. The pivot position is at the top.

Step 2: Select a nonzero entry in the pivot column as a pivot. If necessary, interchange rows to move this entry into the pivot position.

Step 3: Use row replacement operations to create zeros in all positions below the pivot.

Step 4: Cover (or ignore) the row containing the pivot position and cover all rows, if any, above it. Apply steps 1-3 to the submatrix that remains. Repeat the process until there are no more nonzero rows to modify.

Step 5: Beginning with the rightmost pivot and working upward and to the left, create zeros above each pivot. If a pivot is not 1, make it 1 by a scaling operation.

The combination of steps 1-4 is called the **forward phase** of the row reduction algorithm. Step 5, which produces the unique reduced echelon form, is called the **backward phase**.

Variables corresponding to pivot columns are called **basic variables**. Other variables are called **free variables**. Solve for basic variables in terms of free variables; free variables can take any value, and once chosen, will yield the values of basic variables.

Whenever a system is consistent and has free variables, the solution set has many parametric descriptions. Whenever a system is inconsistent, the solution set is empty, even when the system has free variables. In this case, the solution set has no parametric representation.

Theorem 1.2 (Existence and Uniqueness Theorem)
A linear system is consistent if and only if the rightmost column of the augmented matrix is not a pivot column—that is, if and only if an echelon form of the augmented matrix has *no* row of the form $[0 \quad \dots \quad 0 \quad b]$. If a linear system is consistent, then the solution set contains either (i) a unique solution, when there are no free variables, or (ii) infinitely man solutions, when there is at least one free variable.

Using Row Reduction to Solve a Linear System

1. Write the augmented matrix of the system
2. Use the row reduction algorithm to obtain an equivalent augmented matrix in echelon form. Decide whether the system is consistent. If there is no solution, stop; otherwise, go to the next step.
3. Continue row reduction to obtain the reduced echelon form
4. Write the system of equations corresponding to the matrix obtained in step 3.
5. Rewrite each nonzero equation from step 4 so that its one basic variable is expressed in terms of any free variable appearing in the equation.

Section 1.3 – Vector Equations

A matrix with only one column is called a **column vector**, or simply a **vector**.

Two vectors are **equal** if and only if their corresponding entries are equal.

The **sum** of vectors is obtained by adding their corresponding entries.

A **scalar multiple** of a vector is obtained by taking the entry of each vector and multiplying it by a scalar.

Parallelogram Rule for Addition
If $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^2$ are represented as points in the plane, then $\boldsymbol{u}+\boldsymbol{v}$ corresponds to the fourth vertex of the parallelogram whose other vertices are $\boldsymbol{u}$, $\mathbf{0}$, and $\boldsymbol{v}$.

The vector whose entries are all zero is called the **zero vector** and is denoted by $\mathbf{0}$.

Algebraic Properties of $\mathbb{R}^{\boldsymbol{n}}$
For all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ in $\mathbb{R}^n$ and all scalars c and d

i) $u+v=v+u$
ii) $(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}=\boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})$
iii) $\boldsymbol{u}+\mathbf{0}=\mathbf{0}+\boldsymbol{u}=\boldsymbol{u}$
iv) $\boldsymbol{u}+(-\boldsymbol{u})=-\boldsymbol{u}+\boldsymbol{u}-\mathbf{0}$ where $-\boldsymbol{u}$ denotes $(-1)\boldsymbol{u}$
v) $c(\boldsymbol{u}+\boldsymbol{v})=c\boldsymbol{u}+c\boldsymbol{v}$
vi) $(c+d)\boldsymbol{u}=c\boldsymbol{u}+d\boldsymbol{u}$
vii) $c(d\boldsymbol{u})=(cd)\boldsymbol{u}$
viii) $1\boldsymbol{u}=\boldsymbol{u}$

Given vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_p$ in $\mathbb{R}^n$ and given scalars $c_1, c_2, \dots, c_p$, the vector $\boldsymbol{y}$ defined by $\boldsymbol{y}=c_1\boldsymbol{v}_1+c_2\boldsymbol{v}_2+\cdots+c_p\boldsymbol{v}_p$ is called a **linear combination** of $\boldsymbol{v}_1, \dots, \boldsymbol{v}_p$ with **weights** $c_1, \dots, c_p$.

A vector equation $x_1\boldsymbol{a}_1+x_2\boldsymbol{a}_2+\cdots+x_n\boldsymbol{a}_n=\boldsymbol{b}$ has the same solution set as the linear system whose augmented matrix is $[\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \dots \quad \boldsymbol{a}_n \quad \boldsymbol{b}]$.

If $\boldsymbol{v}_1, \dots, \boldsymbol{v}_p$ are in $\mathbb{R}^n$, then the set of all linear combinations of $\boldsymbol{v}_1, \dots, \boldsymbol{v}_p$ is denoted by $Span\{v_1, \dots, v_p\}$ and is called the subset of $\mathbb{R}^n$ spanned (or generated) by $\boldsymbol{v}_1, \dots, \boldsymbol{v}_p$. That is $Span\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_p\}$ is the collection of all vectors that can be written in the form $c_1\boldsymbol{v}_1+c_2\boldsymbol{v}_2+\cdots+c_p\boldsymbol{v}_p$ with $c_1, \dots, c_p$ scalars. Asking whether a vector $\boldsymbol{b}$ is in $Span\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_p\}$ amounts to asking whether the vector equation $c_1\boldsymbol{v}_1+c_2\boldsymbol{v}_2+\cdots+c_p\boldsymbol{v}_p=\boldsymbol{b}$ has a solution.

Section 1.4 – The Matrix Equation

If A is a $m \times n$ matrix, with columns $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$, and if $\boldsymbol{x}$ is in $\mathbb{R}^n$, then **the product of A and x**, denoted by $A\boldsymbol{x}$, is **the linear combination of the columns of A using the corresponding entires in x as weights**;

$$Ax = [\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \dots \quad \boldsymbol{a}_n]\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \dots + x_n\boldsymbol{a}_n$$

An equation of the form $A\boldsymbol{x} = \boldsymbol{b}$ is called a **matrix equation**.

Theorem 1.3
If A is an $m \times n$ matrix, with columns $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$, and if $\boldsymbol{b}$ is in $\mathbb{R}^m$, the matrix equation

$$A\boldsymbol{x} = \boldsymbol{b}$$

Has the same solution set as the vector equation

$$x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \dots + x_n\boldsymbol{a}_n = \boldsymbol{b}$$

Which, in turn, has the same solution set as the system of linear equations whose augmented matrix is

$$[\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \dots \quad \boldsymbol{a}_n \quad \boldsymbol{b}]$$

The equation $A\boldsymbol{x} = \boldsymbol{b}$ has a solution if and only if $\boldsymbol{b}$ is a linear combination of the columns of A.

In the next theorem, the sentence "The columns of A span $\mathbb{R}^m$" means that every $\boldsymbol{b} \in \mathbb{R}^m$ is a linear combination of the columns of A. In general, a set of vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_p\}$ in $\mathbb{R}^m$ **spans** (or **generates**) $\mathbb{R}^m$ if every vector in $\mathbb{R}^m$ is a linear combination of $\boldsymbol{v}_1, \dots, \boldsymbol{v}_p$ – that is, if $Span\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_p\} = \mathbb{R}^m$.

Theorem 1.4
Let A be an $m \times n$ matrix. Then the following statements are logically equivalent. That is, for a particular A, either they are all true statements or they are all false

a. For each $\boldsymbol{b}$ in $\mathbb{R}^m$, the equation $A\boldsymbol{x} = \boldsymbol{b}$ has a solution
b. Each $\boldsymbol{b}$ in $\mathbb{R}^m$ is a linear combination of the columns of A
c. The columns of A span $\mathbb{R}^m$
d. A has a pivot position in every row

Row Vector Rule for Computing $A\boldsymbol{x}$
If the product $A\boldsymbol{x}$ is defined, then the ith entry in $A\boldsymbol{x}$ is the sum of the products of corresponding entries from row i of A and from the vector $\boldsymbol{x}$.

Theorem 1.5
If A is an $m \times n$ matrix $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors in $\mathbb{R}^n$ and c is a scalar, then:

a. $A(\boldsymbol{u} + \boldsymbol{v}) = A\boldsymbol{u} + A\boldsymbol{v}$
b. $A(c\boldsymbol{u}) = c(A\boldsymbol{u})$

Section 1.5 – Solution Sets of Linear Systems

A system of linear equations is said to be **homogeneous** if it can be written in the form $A\boldsymbol{x} = \boldsymbol{0}$, where A is an $m \times n$ matrix and $\boldsymbol{0}$ is the zero vector in $\mathbb{R}^m$. Such a system $A\boldsymbol{x} = \boldsymbol{0}$ always has at least one solution, namely, $\boldsymbol{x} = \boldsymbol{0}$. This zero solution is usually called the **trivial solution**. For a given equation $A\boldsymbol{x} = \boldsymbol{0}$, the important question if whether there exists a nontrival solution, that is, a nonzero vector $\boldsymbol{x}$ that satisfies $A\boldsymbol{x} = \boldsymbol{0}$.

The homogenous equation $A\boldsymbol{x} = \boldsymbol{0}$ has a nontrivial solution if and only if the equation has at least one free variable.

A **parametric vector equation** is written as $\boldsymbol{x} = s\boldsymbol{u} + t\boldsymbol{v}$ for $s, t \in \mathbb{R}$. Such an equation gives an explicit description of the plane as the set spanned by $\boldsymbol{u}$ and $\boldsymbol{v}$. Whenever a solution set is described explicitly with vectors, we say that the solution is in **parametric vector form**.

Given $\boldsymbol{v}$ and $\boldsymbol{p}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$, the effect of adding $\boldsymbol{p}$ to $\boldsymbol{v}$ is to move $\boldsymbol{v}$ in a direction parallel to the line through $\boldsymbol{p}$ and $\boldsymbol{0}$. We say that $\boldsymbol{v}$ is **translated by $\boldsymbol{p}$** to $\boldsymbol{v} + \boldsymbol{p}$.

Consider a line of the form $\boldsymbol{x} = \boldsymbol{p} + t\boldsymbol{v}$ for $t \in \mathbb{R}$. We call this line **the equation of the line through** p **parallel to** v.

Theorem 1.6
Suppose the equation $A\boldsymbol{x} = \boldsymbol{b}$ is consistent for some given $\boldsymbol{b}$, and let $\boldsymbol{p}$ be a solution. Then the solution set of $A\boldsymbol{x} = \boldsymbol{b}$ is the set of all vectors of the form $\boldsymbol{w} = \boldsymbol{p} + \boldsymbol{v}_h$, where $\boldsymbol{v}_h$ is any solution of the homogeneous equation $A\boldsymbol{x} = \boldsymbol{0}$.

Writing a Solution Set (of a Consistent System) in Parametric Vector Form

1. Row reduce the augmented matrix to reduce echelon form.
2. Express each basic variable in terms of any free variables appearing in an equation.
3. Write a typical solution $\boldsymbol{x}$ as a vector whose entries depend on the free variables, if any.
4. Decompose $\boldsymbol{x}$ into a linear combination of vectors (with numeric entries) using free variables as parameters.

Section 1.7 – Linear Independence

An indexed set of vectors $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_p\}$ in $\mathbb{R}$ is said to be **linearly independent** if the vector equation

$$x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_p\boldsymbol{v}_p = \boldsymbol{0}$$

Has only the trivial solution. The set $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_p\}$ is said to be **linearly dependent** if there exists weights $c_1, \dots, c_p$, not all zero, such that

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_p\boldsymbol{v}_p = \boldsymbol{0}$$

Which is called a **linear dependence relation** among $\boldsymbol{v}_1, \dots, \boldsymbol{v}_p$.

The columns of a matrix A are linearly independent if and only if the equation $A\boldsymbol{x} = \boldsymbol{0}$ has only the trivial solution.

A set of two vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ is linearly dependent if at least one of the vectors is a multiple of the other. The set is linearly independent if and only if neither of the vectors is a multiple of the other.

Theorem 1.7 (Characterization of Linearly Dependent Sets)
An indexed set $S = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_p\}$ of two or more vectors is linearly dependent if and only if at least one of the vectors in S is a linear combination of the others. In fact, if S is linearly dependent and $\boldsymbol{v}_1 \neq \boldsymbol{0}$, then some $\boldsymbol{v}_j$ (with $j > 1$) is a linear combination of the preceding vectors, $\boldsymbol{v}_1, \dots, \boldsymbol{v}_{j-1}$.

Theorem 1.8
If a set contains more vectors than there are entries in each vector, then the set is linearly dependent. That is, any set $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_p\}$ in $\mathbb{R}^n$ is linearly dependent if $p > n$.

Theorem 1.9
If a set $S = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_p\}$ in $\mathbb{R}^n$ contains the zero vector, then the set is linearly dependent.

Section 1.8 – Introduction to Linear Transformations

We can think of solving the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ as finding all vectors that are transformed into the vector $\boldsymbol{b}$ under the "action" of multiplication by A.

The correspondence from $\boldsymbol{x}$ to Ax is a function from one set of vectors to another. A **transformation** (or **function** or **mapping**) T from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that assigns to each vector $\boldsymbol{x}$ in $\mathbb{R}^n$ a vector $T(\boldsymbol{x})$ in $\mathbb{R}^m$. The set $\mathbb{R}^n$ is called the **domain** of T, and $\mathbb{R}^m$ is called the **codomain** of T. The notation $T: \mathbb{R}^n \to \mathbb{R}^m$ indicates that the domain of T is $\mathbb{R}^n$ and the codomain is $\mathbb{R}^m$. For $\boldsymbol{x}$ in $\mathbb{R}^n$, the vector $T(\boldsymbol{x})$ in $\mathbb{R}^m$ is called the **image of $\boldsymbol{x}$** (under the action of T). The set of all images $T(x)$ is called the **range** of T.

For each x in $\mathbb{R}^n$, $T(\boldsymbol{x})$ is computed as $A\boldsymbol{x}$, where A is an $m \times n$ matrix. For simplicity, we sometimes denote such a matrix transformation by $\boldsymbol{x} \mapsto A\boldsymbol{x}$.

A **shear transformation** describes when T maps line segments to line segments and vertices to vertices.

A transformation (or mapping) T is a **linear transformation** if it preserves vector addition and scalar multiplication, or, IOW:

i) $T(\boldsymbol{u} + \boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$ for all $\boldsymbol{u}, \boldsymbol{v}$ in the domain of T
ii) $T(c\boldsymbol{u}) = cT(\boldsymbol{u})$ for all scalars c and all $\boldsymbol{u}$ in the domain of T

If T is a linear transformation, then $T(\mathbf{0}) = \mathbf{0}$ and $T(c\boldsymbol{u} + d\boldsymbol{v}) = cT(\boldsymbol{u}) + dT(\boldsymbol{v})$ for all vectors $\boldsymbol{u}, \boldsymbol{v}$ in the domain T and all scalars c, d. Furthermore, if a transformation satisfies the latter condition for all $\boldsymbol{u}, \boldsymbol{v}$ and c, d it must be linear. Additionally, the superposition principle tells us

$$T(c_1\boldsymbol{v}_1 + \cdots + c_p\boldsymbol{v}_p) = c_1T(\boldsymbol{v}_1) + \cdots + c_pT(\boldsymbol{v}_p)$$

Given a scalar r the function $T(\boldsymbol{x}) = r\boldsymbol{x}$ is called a **contraction** when $0 \le r \le 1$ and a **dilation** when $r > 1$.

Section 1.9 – The Matrix of a Linear Transformation

Theorem 1.10
Let $T: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there exists a unique matrix A such that

$$T(\boldsymbol{x}) = A\boldsymbol{x} \text{ for all } \boldsymbol{x} \text{ in } \mathbb{R}^n$$

In fact, A is the $m \times n$ matrix whose jth column is the vector $T(\boldsymbol{e}_j)$ where $\boldsymbol{e}_j$ is the jth column of the identity matrix in $\mathbb{R}^n$:

$$A = [T(\boldsymbol{e}_1) \quad \ldots \quad T(\boldsymbol{e}_n)]$$

Where the matrix A is called the **standard matrix for the linear transformation T**.

TABLE 1 Reflections

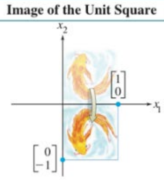
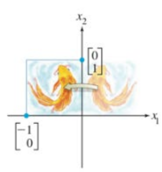
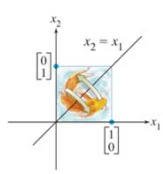
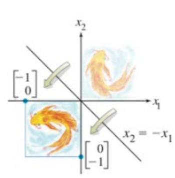
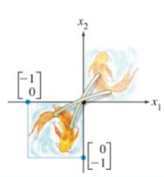
Transformation	Image of the Unit Square	Standard Matrix
Reflection through the x_1-axis		$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
Reflection through the x_2-axis		$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$
Reflection through the line $x_2 = x_1$		$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
Reflection through the line $x_2 = -x_1$		$\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$
Reflection through the origin		$\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$

TABLE 2 Contractions and Expansions

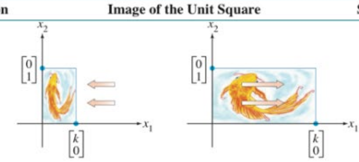
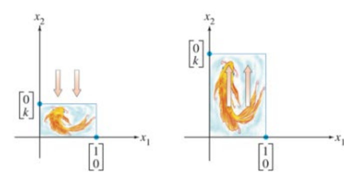
Transformation	Image of the Unit Square	Standard Matrix
Horizontal contraction and expansion		$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$
Vertical contraction and expansion		$\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$

TABLE 3 Shears

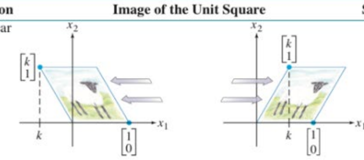
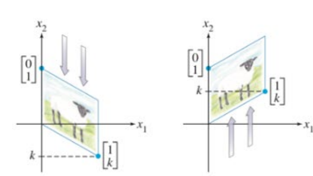
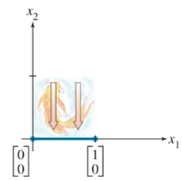
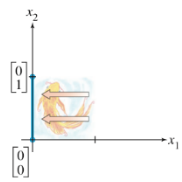
Transformation	Image of the Unit Square	Standard Matrix
Horizontal shear		$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$
Vertical shear		$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$

TABLE 4 Projections

Transformation	Image of the Unit Square	Standard Matrix
Projection onto the x_1-axis		$\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}$
Projection onto the x_2-axis		$\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}$

A mapping $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is said to be **onto** $\mathbb{R}^m$ if each $\boldsymbol{b}$ in $\mathbb{R}^m$ is the image of at least one $\boldsymbol{x}$ in $\mathbb{R}^n$

A mapping $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is said to be **one-to-one** if each $\boldsymbol{b}$ in $\mathbb{R}^m$ is the image of at most one $\boldsymbol{x}$ in $\mathbb{R}^m$

Theorem 1.11
Let $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a linear transformation. Then T is one-to-one if and only if the equation $T(\boldsymbol{x}) = \boldsymbol{0}$ has only the trivial solution.

Theorem 1.12
Let $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a linear transformation and let A be the standard matrix for T. Then

a. T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if and only if the columns of A span $\mathbb{R}^m$
b. T is one-to-one if and only if the columns of A are linearly independent

One-to-one – pivot in every column
Onto – pivot in every row

Chapter 2 – Matrix Algebra

Section 2.1 – Matrix Operations

If A is an $m \times n$ matrix then the scalar entry in the ith row and the jth column of A is denoted by a_{ij} and is called the (i, j)-entry of A.

The **diagonal entries** in an $m \times n$ matrix $A = [a_{ij}]$ are a_{11}, a_{22}, a_{33} ... and they form the **main diagonal** of A. A **diagonal matrix** is a square $n \times n$ matrix whose nondiagonal entries are zero.

Row matrices are **equal** if they have the same size and if their corresponding columns are equal.

If A and B are $m \times n$ matrices, then the **sum** $A + B$ is the $m \times n$ matrix whose columns are the sums of the corresponding columns in A and B.

If r is a scalar and A is a matrix, then the **scalar multiple** rA is the matrix whose columns are r times the corresponding columns in A.

Theorem 2.1
Let A, B, and C be matrices of the same size and let r and s be scalars.

a. $A + B = B + A$
b. $(A + B) + C = A + (B + C)$
c. $A + 0 = A$
d. $r(A + B) = rA + rB$
e. $(r + s)A = rA + sA$
f. $r(sA) = (rs)A$

If A is an $m \times n$ matrix, and if B is an $n \times p$ matrix with columns $b_1, \dots, b_p$, then the product AB is the $m \times p$ matrix whose columns are $Ab_1, \dots, Ab_p$. That is,

$$AB = A[\boldsymbol{b}_1 \quad \boldsymbol{b}_2 \quad \dots \quad \boldsymbol{b}_p] = [A\boldsymbol{b}_1 \quad A\boldsymbol{b}_2 \quad \dots \quad A\boldsymbol{b}_p]$$

Each column of AB is a linear combination of the columns of A using weights from the corresponding column of B.

Row Column Rule for Computing $\boldsymbol{AB}$
If the product AB is defined, then the entry in row i and column j of AB is the sum of the products of corresponding entries from row i of A and column j of B. If $(AB)_{ij}$ denotes the (i, j)-entry in AB, and if A is an $m \times n$ matrix, then $(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$.

Theorem 2.2
Let A be an $m \times n$ matrix, and let B and C have sizes for which the indicated sums and products are defined.

a. $A(BC) = (AB)C$ (associative law of multiplication)
b. $A(B + C) = AB + AC$ (left distributive law)
c. $(B + C)A = BA + CA$ (right distributive law)
d. $r(AB) = (rA)B = A(rB)$ for any scalar r
e. $I_m A = A = AI_n$ (identity for matrix multiplication)

Warnings

1. In general, $AB \neq BA$
2. The cancellation laws do not hold for matrix multiplication. That is, if $AB = AC$, then it is not true in general that $B = C$.
3. If a product of AB is the zero matrix, you cannot conclude in general that either $A = 0$ or $B = 0$.

If A is nonzero and if $\boldsymbol{x}$ is in $\mathbb{R}^n$, then $A^k\boldsymbol{x}$ is the result of left-multiplying $\boldsymbol{x}$ by A repeatedly k times. If $k = 0$, then $A^0\boldsymbol{x}$ should be $\boldsymbol{x}$ by itself. Thus A^0 is interpreted as the identity matrix.

Given a $m \times n$ matrix A, the **transpose** of A is the $n \times m$ matrix, denoted by A^T whose columns are formed from the corresponding rows of A.

Theorem 2.3
Let A and B denote matrices whose sizes are appropriate for the following sums and products.

a. $(A^T)^T = A$
b. $(A + B)^T = A^T + B^T$
c. For any scalar r, $(rA)^T = rA^T$
d. $(AB)^T = B^T A^T$

The transpose of a product of matrices equals the product of their transposes in the *reverse* order.

Section 2.2 – The Inverse of a Matrix

An $n \times n$ matrix is said to be **invertible** if there is an $n \times n$ matrix C such that $CA = I$ and $AC = I$. In this case, C is an **inverse** of A. In fact, C is uniquely determined by A. This unique inverse is denoted by A^{-1} so that $A^{-1}A = I$ and $AA^{-1} = I$. A matrix that is not invertible is sometimes called a **singular matrix** and an invertible matrix is called a **nonsingular matrix**.

Theorem 2.4
Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. If $ad - bc \neq 0$, then A is invertible and $A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. If $ad - bc = 0$,then A is not invertible. In other words, a 2×2 matrix is invertible if and only if $\det A \neq 0$.

The quantity $ad - bc$ is called the determinant of A, and we write $\det A = ad - bc$.

Theorem 2.5
If A is an invertible $n \times n$ matrix, then for each $\boldsymbol{b}$ in $\mathbb{R}^n$, the equation $A\boldsymbol{x} = \boldsymbol{b}$ has the unique solution $\boldsymbol{x} = A^{-1}\boldsymbol{b}$.

Theorem 2.6

a. If A is an invertible matrix, then A^{-1} is invertible and $(A^{-1})^{-1} = A$
b. If A and B are $n \times n$ invertible matrices, then so is AB, and the inverse of AB is the product of the inverses of A and B in the reverse order. That is, $(AB)^{-1} = B^{-1}A^{-1}$
c. If A is an invertible matrix, then so is A^T, and the inverse of A^T is the transpose of A^{-1}. That is, $(A^T)^{-1} = (A^{-1})^T$

The product of $n \times n$ invertible matrices is invertible, and the inverse is the product of their inverses in the reverse order.

An **elementary matrix** is one that is obtained by performing a single elementary row operation on an identity matrix.

If an elementary row operation is performed on an $m \times n$ matrix A the resulting matrix can be written as EA, where $m \times m$ matrix E is created by performing the same row operation on I_m.

Each elementary matrix E is invertible. The inverse of E is the elementary matrix of the same type that transforms E back into I.

Theorem 2.7
An $n \times n$ matrix A is invertible if and only if A is row equivalent to I_n, and in this case, any sequence of elementary row operations that reduces A to I_n also transforms I_n into A^{-1}.

Algorithm for Finding A^{-1}
Row reduce the augmented matrix $[A \quad I]$. If A is row equivalent to I, then $[A \quad I]$ is row equivalent to $[I \quad A^{-1}]$. Otherwise, A does not have an inverse.

Section 2.3 – Characterizations of Invertible Matrices

Theorem 2.8 (The Invertible Matrix Theorem)
Let A be a square $n \times n$ matrix. Then the following statements are equivalent. That is, for a given A, the statements are either all true or all false.

a. A is an invertible matrix
b. A is row equivalent to the $n \times n$ identity matrix
c. A has n pivot rows
d. The equation $A\boldsymbol{x} = \boldsymbol{0}$ has only the trivial solution
e. The columns of A form a linearly independent set.
f. The linear transformation $\boldsymbol{x} \mapsto A\boldsymbol{x}$ is one-to-one
g. The equation $A\boldsymbol{x} = \boldsymbol{b}$ has at least one solution for each $\boldsymbol{b}$ in $\mathbb{R}^n$
h. The columns of A span $\mathbb{R}^n$.

i. The linear transformation $\boldsymbol{x} \mapsto A\boldsymbol{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$
j. There is an $n \times n$ matrix C such that $CA = I$
k. There is an $n \times n$ matrix D such that $AD = I$
l. A^T is an invertible matrix
m. The columns of A form a basis of $\mathbb{R}^n$
n. $Col\ A = \mathbb{R}^n$
o. $\dim Col\ A = n$
p. $rank\ A = n$
q. $Nul\ A = \{\mathbf{0}\}$
r. $\dim Nul\ A = 0$
s. The number 0 is not an eigenvalue of A
t. The determinant of A is not zero

Let A and B be square matrices. If $AB = I$, then A and B are both invertible, with $B = A^{-1}$ and $A = B^{-1}$.

A linear transformation $T: \mathbb{R}^n \to \mathbb{R}^n$ is said to be invertible if there exists a function $S: \mathbb{R}^n \to \mathbb{R}^n$ such that $S(T(\boldsymbol{x})) = \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathbb{R}^n$ and $T(S(\boldsymbol{x})) = \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathbb{R}$.

Theorem 2.9
Let $T: \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation and let A be the standard matrix for T. Then T is invertible if and only if A is an invertible matrix. In that case, the linear transformation S given by $S(\boldsymbol{x}) = A^{-1}\boldsymbol{x}$ is the unique function satisfying the equations in the above block.

From the above theorem, we call S the inverse of T and write it as T^{-1}.

Section 2.5 – Matrix Factorizations

A **factorization** of matrix A is an equation that expresses A as a product of two or more matrices.

Let A be an $m \times n$ matrix that can be row reduced to echelon form without row interchanges. Then A can be written in the form $A = LU$ where L is an $m \times m$ lower triangular matrix with 1's on the diagonal and U is an $m \times n$ echelon form of A. Such a factorization is called an LU **factorization**. The matrix L is invertible and is called a unit lower triangular matrix.

Algorithm for an LU Factorization

1. Reduce A to an echelon form by a sequence of row replacement operations, if possible. This is U.
2. Place entries in L such that the same sequence of row operations reduces L to I. Alternatively, do the row operations to get U in reverse order on I to get L.

Example of Algorithm for an LU Factorization

$$A = \begin{bmatrix} 1 & -2 & -4 & -3 \\ 2 & -7 & -7 & -6 \\ -1 & 2 & 6 & 4 \\ -4 & -1 & 9 & 8 \end{bmatrix} \xrightarrow[R_4=4R_1+R_4]{\substack{R_2=-2R_1+R_2 \\ R_3=R_1+R_3}} \begin{bmatrix} 1 & -2 & -4 & -3 \\ 0 & -3 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & -9 & -7 & -4 \end{bmatrix} \xrightarrow[R_4=-3R_2+R_4]{R_3=\frac{1}{3}R_2+R_3}$$
$$\begin{bmatrix} 1 & -2 & -4 & -3 \\ 0 & -3 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & -10 & -4 \end{bmatrix} \xrightarrow{R_4=5R_3+R_4} \begin{bmatrix} 1 & -2 & -4 & -3 \\ 0 & -3 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

From this,

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -4 & 3 & -5 & 1 \end{bmatrix} \quad U = \begin{bmatrix} 1 & -2 & -4 & -3 \\ 0 & -3 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

To get the ith column of L, take ith column of the matrix in the reduction sequence before the ith column has been reduced to the proper form for U, scale it so the entry appearing on the diagonal is one, zero out the entries above the main diagonal.

Section 2.8 – Subspaces of $\mathbb{R}^n$

Let $H \subseteq \mathbb{R}^n$. The set H is a **subspace** of $\mathbb{R}^n$ provided it has three properties:

a. The zero vector is in H
b. For each $\boldsymbol{u}$ and $\boldsymbol{v}$ in H, the sum $\boldsymbol{u} + \boldsymbol{v}$ is in H
c. For each $\boldsymbol{u}$ in H and each scalar c, the vector $c\boldsymbol{u}$ is in H

Note that $\mathbb{R}^n$ is a subspace of itself because it had the three properties required for a subspace. Another special subspace is the set consisting of only the zero vector in $\mathbb{R}^n$. This set is called the **zero subspace**.

The **column space** of a matrix A is the set $Col\ A$ of all linear combinations of the columns of A. If $A = [\boldsymbol{a}_1 \quad \cdots \quad \boldsymbol{a}_n]$ with the columns in $\mathbb{R}^m$, then $Col\ A$ is the same as $Span\{\boldsymbol{a}_1, \dots, \boldsymbol{a}_n\}$. The column space of an $m \times n$ matrix is a subspace of $\mathbb{R}^m$. When a system of linear equations is written in the form $A\boldsymbol{x} = \boldsymbol{b}$, the column space of A is the set of all $\boldsymbol{b}$ for which the system has a solution.

The **null space** of a matrix A is the set $Nul\ A$ of all solutions of the homogenous equation $A\boldsymbol{x} = \boldsymbol{0}$. The null space is the span of the coefficients of the free variables.

Theorem 2.12
The null space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^n$. Equivalently, the set of all solutions of a system $A\boldsymbol{x} = \boldsymbol{0}$ of m homogeneous linear equations in n unknowns is a subspace of $\mathbb{R}^n$.

A **basis** for a subspace H of $\mathbb{R}^n$ is a linearly independent set in H that spans H. It is the smallest set of vectors that span the subspace.

The set $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ is called the **standard basis** for $\mathbb{R}^n$ where e_i is the ith column of I_n.

Theorem 2.13
The pivot columns of a matrix A form a basis for the column space of A.

Section 2.9 – Dimension and Rank

Each vector in H can be written in only one way as a linear combination of the basis vectors.

Suppose the set $\mathcal{B} = \{\boldsymbol{b}_1, \dots, \boldsymbol{b}_p\}$ is a basis for a subspace H. For each $\boldsymbol{x}$ in H, the **coordinates of** x **relative to the basis** $\mathcal{B}$ are the weights $c_1, \dots, c_p$ such that $x = c_1\boldsymbol{b}_1 + \cdots + c_p\boldsymbol{b}_p$ and the vector in $\mathbb{R}^p$

$$[\boldsymbol{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}$$

Is called the **coordinate vector of** x **relative to** $\mathcal{B}$ or the $\mathcal{B}$ −**coordinate vector of** x.

If $\mathcal{B} = \{\boldsymbol{b}_1, \dots, \boldsymbol{b}_p\}$ is a basis for H, then the mapping $x \longmapsto [x]_{\mathcal{B}}$ is a bijection that makes H look and act the same as $\mathbb{R}^p$. The mapping $x \longmapsto [x]_{\mathcal{B}}$ is called an **isomorphism**.
The **dimension** of a nonzero subspace H, denoted by $\dim H$ is the number of vectors in any basis for H. The dimension of the zero subspace $\{\boldsymbol{0}\}$ is defined to be zero.

The **rank** of a matrix A, denoted by $rank\ A$, is the dimension of the column space of A. We can also interpret rank as the number of pivot columns of A or the number of linearly independent columns of A.

The $\dim Nul\ A$ can be interpreted as the number of non-pivot columns of A.

Theorem 2.14 (The Rank Theorem)
If a matrix A has n columns, then $rank\ A + \dim Nul\ A = n$.

Theorem 2.15 (The Basis Theorem)
Let H be a p −dimensional subspace of $\mathbb{R}^n$. Any linear independent set of exactly p elements in H is automatically a basis for H. Also, any set of p elements of H that spans H is automatically a basis for H.

Chapter 3 – Determinants

Section 3.1 – Introduction to Determinants

For $n \geq 2$, the **determinant** of an $n \times n$ matrix $A = [a_{ij}]$ is the sum of the n terms of the form $\pm a_{1j}$ $\det A_{1j}$, with plus and minus signs alternating, where the entries $a_{11}, a_{12}, \dots, a_{1n}$ are from the first row of A. In symbols,

$$\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + \cdots + (-1)^{1+n}\ a_{1n} \det A_{1n} = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{1j}$$

Given $A = [a_{ij}]$, the $(\boldsymbol{i}, \boldsymbol{j})$**-cofactor** of A is the number C_{ij} given by $C_{ij} = (-1)^{i+j} \det\ A_{ij}$. Then

$$\det A = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}$$

Theorem 3.1
The determinant of an $n \times n$ matrix A can be computed by a **cofactor expansion** across any row or down any column. The expansion across the ith row using cofactors is

$$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$$

The cofactor expansion down the jth column is

$$\det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}$$

Theorem 3.2
If A is a triangular matrix, then $\det A$ is the product of the entries on the main diagonal of A.

Section 3.2 – Properties of Determinants

sin

Theorem 3.3 (Row Operations)
Let A be a square matrix.

a. If a multiple of one row of A is added to another row to produce a matrix B, then $\det B = \det A$.
b. If two rows of A are interchanged to produce B, then $\det B = -\det A$.
c. If one row of A is multiplied by k to produce B, then $\det B = k \cdot \det A$

Alternatively, if A is an $n \times n$ matrix and E is an $n \times n$ elementary matrix, then $\det EA = (\det E)(\det A)$ where

$$\det E = \begin{cases} 1 & \text{if } E \text{ is a row replacement} \\ -1 & \text{if } E \text{ is an interchange} \\ r & \text{if } E \text{ is a scale by } r \end{cases}$$

Suppose a square matrix A has been reduced to an echelon form U by row replacements and row interchanges. (This is always possible by the row reduction algorithm in Section 1.2). If there are r interchanges, then Theorem 3 shows that $\det A = (-1)^r \det U$. Since U is in echelon form, is triangular, and so $\det U$ is the product of the diagonal entries $u_{11}, \dots, u_{nn}$. If A is invertible, the entries u_{ii} are all pivots (because $A \sim I_n$ and the u_{ii} have not been scaled to 1's). Otherwise, at least u_{nn} is zero, and the product $u_{11} \dots u_{nn}$ is zero. Thus

$$\det A = \begin{cases} (-1)^r \cdot (\text{product of pivots in } U) & \text{when } A \text{ is invertible} \\ 0 & \text{when } A \text{ is not invertible} \end{cases}$$

Theorem 3.4
A square matrix A is invertible if and only if $\det A \neq 0$.

Theorem 3.5
If A is an $n \times n$ matrix, then $\det A^T = \det A$.

By Theorem 3.5, every statement in Theorem 3.3 is true when 'row' is replaced with 'column'.

Theorem 3.6 (Multiplicative Property)
If A and B are $n \times n$ matrices, then $\det AB = (\det A)\,(\det B)$.

Section 3.3 – Cramer's Rule, Volume, and Linear Transformations

Cramer's rule can be used to study how the solution of $A\boldsymbol{x} = \boldsymbol{b}$ is affected by changes in the entries of $\boldsymbol{b}$. For any $n \times n$ matrix A and any $\boldsymbol{b}$ in $\mathbb{R}^n$, let $A_i(\boldsymbol{b})$ be the matrix obtained from A by replacing column i by the vector $\boldsymbol{b}$.

$$A_i(b) = [\boldsymbol{a}_1 \quad \dots \quad \boldsymbol{b} \quad \dots \quad \boldsymbol{a}_n]$$

Where $\boldsymbol{b}$ is in column i.

Theorem 3.7 (Cramer's Rule)
Let A be an invertible $n \times n$ matrix. For any $\boldsymbol{b}$ in $\mathbb{R}^n$, the unique solution $\boldsymbol{x}$ of $A\boldsymbol{x} = \boldsymbol{b}$ has entries given by $x_i = \frac{\det A_i(\boldsymbol{b})}{\det A}, i = 1,2,\dots,n$

Cramer's rule leads to a general formula for the inverse of a $n \times n$ matrix A. The jth column of A^{-1} is a vector $\boldsymbol{x}$ that satisfies $A\boldsymbol{x} = \boldsymbol{e}_j$ where $\boldsymbol{e}_j$ is the jth column of the identity matrix and the ith entry of $\boldsymbol{x}$ is the (i,j) entry of A^{-1}. By Cramer's rule

$$\{(i,j) - entry\ of\ A^{-1}\} = x_i = \frac{\det A_i(\boldsymbol{e}_j)}{\det A}.$$

Recall that A_{ji} denotes the submatrix of A formed by deleting row j and column i. A cofactor expansion down column i of $A_i(\boldsymbol{e}_j)$ shows that

$$\det A_i(\boldsymbol{e})_j = (-1)^{i+j} \det A_{ji} = C_{ji}$$

where C_{ji} is a cofactor of A. By the above equation, the (i,j) –entry of A^{-1} is the cofactor C_{ji} divided by $\det A$. Thus,

$$A^{-1} = \frac{1}{\det A}\begin{bmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{bmatrix}$$

The matrix on the right side of the above expression is called the **adjugate** (or **classical adjoint**) of A denoted by $adj\ A$.

Theorem 3.8 (An Inverse Formula)
Let A be an invertible $n \times n$ matrix. Then $A^{-1} = \frac{1}{\det A} adj\ A$

Theorem 3.9
If A is a 2×2 matrix, the area of the parallelogram determined by the columns of A is $|\det A|$. If A is a 3×3 matrix, the volume of the parallelepiped determined by the columns of A is $|\det A\,|$.

Let $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ be nonzero vectors. Then for any scalar c, the area of the parallelogram determined by $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ equals the area of the parallelogram determined by $\boldsymbol{a}_1$ and $\boldsymbol{a}_2 + c\boldsymbol{a}_1$.

Theorem 3.10
Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation determined by a 2×2 matrix A. If S is the parallelogram in $\mathbb{R}^2$, then $\{area\ of\ T(S)\} = |\det A| \cdot \{area\ of\ S\}$. If T is determined by a 3×3 matrix A, and if S is a parallelepiped in $\mathbb{R}^3$, then $\{volume\ of\ T(S)\} = |\det A| \cdot \{volume\ of\ S\}$.

The conclusions of Theorem 10 hold whenever S is a region in $\mathbb{R}^2$ with finite area or a region in $\mathbb{R}^3$ with finite volume.

Chapter 4 – Vector Spaces

Section 4.1 – Vector Spaces and Subspaces

A **vector space** is a nonempty set V of objects, called vectors, on which are defined two operations called addition and multiplication by scalars subject to ten axioms listed below. These axioms must hold for all vectors $\boldsymbol{u}, \boldsymbol{v}$, and $\boldsymbol{w}$ in V and for all scalars c and d.

1. The sum of $\boldsymbol{u}$ and $\boldsymbol{v}$ denoted by $\boldsymbol{u} + \boldsymbol{v}$ is in V
2. $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$
3. $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$
4. There is a zero vector in V such that $\boldsymbol{u} + \boldsymbol{0} = \boldsymbol{u}$
5. For each $\boldsymbol{u}$ in V, there is a vector $-\boldsymbol{u}$ such that $\boldsymbol{u} + (-\boldsymbol{u}) = \boldsymbol{0}$
6. The scalar multiple of $\boldsymbol{u}$ by c denoted by $c\boldsymbol{u}$ is in V
7. $c(\boldsymbol{u} + \boldsymbol{v}) = c\boldsymbol{u} + c\boldsymbol{v}$
8. $(c + d)\boldsymbol{u} = c\boldsymbol{u} + d\boldsymbol{u}$
9. $c(d\boldsymbol{u}) = (cd)\boldsymbol{u}$
10. $1\boldsymbol{u} = \boldsymbol{u}$

The zero vector is unique, and the vector $-\boldsymbol{u}$ called the **negative of** u is unique for each u in V

For each $\boldsymbol{u}$ in V and scalr c, $0\boldsymbol{u} = \boldsymbol{0}$, $c\boldsymbol{0} = \boldsymbol{0}$, $-\boldsymbol{u} = (-1)\boldsymbol{u}$

A **subspace** of a vector space V is a subset of H of V that has three properties

a. The vector of V is in H.
b. H is closed under vector addition. That is, for each u and v in H, the sum of $u + v$ is in H.
c. H is closed under multiplication by scalars. That is, for each u in H and each scalar c, the vector $c\boldsymbol{u}$ is in H.

The set consisting of only the zero vector space V is a subspace of V, called the **zero subspace** and written as $\{\boldsymbol{0}\}$.

Theorem 4.1
If $\boldsymbol{v}_1, \ldots \boldsymbol{v}_p$ are in a vector space V, then $Span\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ is a subspace of V. We call $Span\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ **the subspace spanned** (or **generated**) by $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$. Given any subspace H of V a **spanning** (or **generating**) set for H is a set $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ in H such that $H = Span\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$.

Section 4.2 – Null Spaces, Column Spaces, and Linear Transformations

The **null space** of an $m \times n$ matrix A, written as $Nul\ A$ is the set of all solutions of the homogenous equation $A\boldsymbol{x} = \boldsymbol{0}$. In set notation, $Nul\ A = \{\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{R}^n\ and\ A\boldsymbol{x} = \boldsymbol{0}\}$.

Theorem 4.2
The null space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^n$. Equivalently, the set of all solutions to a system $A\boldsymbol{x} = \boldsymbol{0}$ of m homogeneous linear equations in n unkowns is a subspace of $\mathbb{R}^n$.

The **column space** of an $m \times n$ matrix, written as $Col\ A$, is the set of all linear combinations of the columns of A. If $A = [\boldsymbol{a}_1 \quad \cdots \quad \boldsymbol{a}_n]$, then $Col\ A = Span\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$.

Theorem 4.3
The column space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^m$.

$Col\ A = \{\boldsymbol{b} \mid \boldsymbol{b} = A\boldsymbol{x}\ for\ some\ \boldsymbol{x}\ in\ \mathbb{R}^n\}$. The notation $A\boldsymbol{x}$ for vectors in $Col\ A$ als shows that $Col\ A$ is the range of the linear transformation $\boldsymbol{x} \mapsto A\boldsymbol{x}$

The column space of an $m \times n$ matrix A is all of $\mathbb{R}^m$ if and only if the equation $A\boldsymbol{x} = \boldsymbol{b}$ has a solution for each $\boldsymbol{b}$ in $\mathbb{R}^m$.

Contrasts Between $Nul\ A$ and $Col\ A$ for an $m \times n$ Matrix A

$Nul\ A$	$Col\ A$
1. $\boldsymbol{Nul\ A}$ is a subspace of $\mathbb{R}^{\boldsymbol{n}}$.	1. $Col\ A$ is a subspace of $\mathbb{R}^m$.
2. $\boldsymbol{Nul\ A}$ is implicitly defined, that is, you are only given a condition ($\boldsymbol{Ax} = \boldsymbol{0}$) that vectors in $\boldsymbol{Nul\ A}$ must satisfy.	2. $Col\ A$ is explicitly defined; that is, you are told how to build vectors in $Col\ A$.
3. It takes time to find vectors in $\boldsymbol{Nul\ A}$. Row operations on $[\boldsymbol{A} \quad \boldsymbol{0}]$ are required.	3. It is easy to find vectors in $Col\ A$. The columns of A are displayed; others are formed from them.
4. There is no obvious relation between $\boldsymbol{Nul\ A}$ and the entries in $\boldsymbol{A}$.	4. There is an obvious relation between $Col\ A$ and the entries in A, since each column of A is in $Col\ A$.
5. A typical vector $\boldsymbol{v}$ in $\boldsymbol{Nul\ A}$ has the property that $\boldsymbol{Av} = \boldsymbol{0}.$	5. A typical vector $\boldsymbol{v}$ in $Col\ A$ has the property that the equation $A\boldsymbol{x} = \boldsymbol{v}$ is consistent.
6. Given a specific vector $\boldsymbol{v}$ it is easy to tell if $\boldsymbol{v}$ is in $\boldsymbol{Nul\ A}$. Just compute $\boldsymbol{Av}$.	6. Given a specific vector $\boldsymbol{v}$ it may take time to tell if $\boldsymbol{v}$ is in $Col\ A$. Row operations on $[A \quad \boldsymbol{v}]$ are required.
7. $\boldsymbol{Nul\ A} = \{\boldsymbol{0}\}$ if and only if the equation $\boldsymbol{Ax} = \boldsymbol{0}$ has only the trivial solution.	7. $Col\ A = \mathbb{R}^m$ is and only if the equation $A\boldsymbol{x} = \boldsymbol{b}$ has a solution for every $\boldsymbol{b}$ in $\mathbb{R}^m$.
8. $\boldsymbol{Nul\ A} = \{\boldsymbol{0}\}$ if and only if the linear transformation $\boldsymbol{x} \longmapsto \boldsymbol{Ax}$ is one-to-one.	8. $Col\ A = \mathbb{R}^m$ if and only if the linear transformation $\boldsymbol{x} \longmapsto A\boldsymbol{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^m$.

A linear transformation T from a vector space V into a vector space W is a rule that assigns to each vector x in V a unique vector $T(x)$ in W, such that

i. $T(\boldsymbol{u}+\boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$ for all $\boldsymbol{u}, \boldsymbol{v} \in V$
ii. $T(c\boldsymbol{u}) = cT(\boldsymbol{u})$ for all $\boldsymbol{u} \in V$ and all scalars c

The **kernel** (or **null space**) of such a T is the set of all $\boldsymbol{u}$ in V such that $T(\boldsymbol{u}) = \boldsymbol{0}$. The **range** of T is the set of all vectors in W of the form $T(x)$ for some x in V. If T happens to arise as a matrix transformation—say $T(\boldsymbol{x}) = A\boldsymbol{x}$ for some matrix A—then the kernel and the range of T are just the null space and the column space of A as defined earlier.

Section 4.3 – Linearly Independent Sets; Bases

An indexed set of vectors $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_p\}$ in $V$is said to be **linearly independent** if the vector equation

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_p\boldsymbol{v}_p = \boldsymbol{0}$$

Has only the trivial solution, $c_1 = 0, \dots, c_p = 0$. The set $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_p\}$ is said to be **linearly dependent** if the above vector equation has a nontrivial solution, that is, if there exist weights $c_1, \dots, c_p$, not all zero, such that the above vector equation holds. In such a case, the vector equation with the non-trivial weights is called a **linear dependence relation** among $\boldsymbol{v}_1, \dots, \boldsymbol{v}_p$.

Theorem 4.4
An indexed set $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_p\}$ of two or more vectors, with $\boldsymbol{v}_1 \neq \boldsymbol{0}$ is linearly dependent if and only if some $\boldsymbol{v}_j$ (with $j > 1$) is a linear combination of the proceeding vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_p$.

Let H be a subspace of a vector space V. An indexed set of vectors $\mathcal{B} = \{\boldsymbol{b}_1, \dots, \boldsymbol{b}_p\}$ in V is a **basis** for H if

i. $\mathcal{B}$ is a linearly independent set, and
ii. The subspace spanned by $\mathcal{B}$ coincides with H; that is $H = Span\{\boldsymbol{b}_1, \dots, \boldsymbol{b}_p\}$

Let $S = \{1, t, t^2, \dots, t^n\}$. This basis is called **the standard basis for** $\mathbb{P}_n$.

Theorem 4.5 (The Spanning Set Theorem)
Let $S = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_p\}$ be a set in V and let $H = Span\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_p\}$**.**

a. If one of the vectors in S—say $\boldsymbol{v}_k$—is a linear combination of the remaining vectors in S, then the set formed from S by removing $\boldsymbol{v}_k$ still spans H.
b. If $H \neq \{\boldsymbol{0}\}$, some subset of S is a basis for H.

Theorem 4.6
The pivot columns of a matrix A form a basis for $Col\ A$.

The basis is a spanning set that is as small as possible and a linearly independent set that is as large as possible.

Section 4.4 – Coordinate Systems

Theorem 4.7 (The Unique Representation Theorem)
Let $\mathcal{B} = \{\boldsymbol{b}_1, \dots, \boldsymbol{b}_n\}$ be a basis for a vector space V. Then for each $\boldsymbol{x} \in V$, there exists a unique set of scalars $c_1, \dots, c_n$ such that

$$\boldsymbol{x} = c_1\boldsymbol{b}_1 + \cdots + c_n\boldsymbol{b}_n$$

Suppose $\mathcal{B} = \{\boldsymbol{b}_1, \dots, \boldsymbol{b}_n\}$ is a basis for V and $\boldsymbol{x}$ is in V. **The coordinates of** x **relative to the basis** $\mathcal{B}$ (or the $\mathcal{B}$**-coordinates of** x) are the weights $c_1, \dots, c_n$ such that $\boldsymbol{x} = c_1\boldsymbol{b}_1 + \cdots + c_n\boldsymbol{b}_n$. If $c_1, \dots, c_n$ are the $\mathcal{B}$ coordiantes of $\boldsymbol{x}$, then the vector in $\mathbb{R}^n$

$$[\boldsymbol{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

Is the **coordinate vector of** x **(relative to** $\mathcal{B}$**)** or the $\mathcal{B}$**-coordinate vector of** x. The mapping $\boldsymbol{x} \mapsto [\boldsymbol{x}]_{\mathcal{B}}$ is the **coordinate mapping (determined by** $\mathcal{B}$**)**.

Define $P_{\mathcal{B}} = [\boldsymbol{b}_1 \quad \dots \quad \boldsymbol{b}_n]$. Then we have $\boldsymbol{x} = P_{\mathcal{B}}[\boldsymbol{x}]_{\mathcal{B}}$ where $P_{\mathcal{B}}$ is called the **change-of-coordinates matrix** from $\mathcal{B}$ to the standard basis in $\mathbb{R}^n$. Furthermore, $P_{\mathcal{B}}^{-1}\boldsymbol{x} = [\boldsymbol{x}]_{\mathcal{B}}$.

Theorem 4.8
Let $\mathcal{B} = \{\boldsymbol{b}_1, \dots, \boldsymbol{b}_n\}$ be a basis for a vector space V. Then the coordinate mapping $\boldsymbol{x} \mapsto [\boldsymbol{x}]_{\mathcal{B}}$ is a one-to-one linear transformation from V onto $\mathbb{R}^n$.

A one-to-one linear transformation from a vector space V onto W is called an **isomorphism**.

Section 4.5 – The Dimension of a Vector Space

Theorem 4.9
If a vector space V has a basis $\mathcal{B} = \{\boldsymbol{b}_1, \dots, \boldsymbol{b}_n\}$, then any set in V containing more than n vectors must be linearly dependent.

Theorem 4.10
If a vector space V has a basis of n vectors, then every basis of V must consist of exactly n vectors.

If V is spanned by a finite set, then V is said to be **finite-dimensional**, and the **dimension** of V written as $\dim V$, is the number of vectors in a basis for V. The dimension of the zero vector space $\{\mathbf{0}\}$ is defined to be zero. If V is not spanned by a finite set, then V is said to be **infinite-dimensional**.

Theorem 4.11
Let H be a subspace of a finite-dimensional vector space V. Any linearly independent set in H can be expanded, if necessary, to a basis for H. Also, H is finite-dimensional and $\dim H \leq \dim V$.

Theorem 4.12 (The Basis Theorem)
Let V be a p-dimensional vector space $p \geq 1$. Any linearly independent set of exactly p elements in V is automatically a basis for V. Any set of exactly p elements that spans V is automatically a basis for V.

The dimension of $Nul\ A$ is the number of free variables in the equation $A\boldsymbol{x} = \boldsymbol{0}$ and the dimension of $Col\ A$ is the number of pivot columns in A.

Section 4.6 – Rank

If A is an $m \times n$ matrix, each row of A has n entries and thus can be identified with a vector in $\mathbb{R}^n$. The set of all linear combinations of the row vectors is called the **row space** of A and is denoted by $Row\ A$. Each row has n entries so $Row\ A$ is a subspace of $\mathbb{R}^n$. Since the rows of A are identified with the columns of A^T, we could also write $Col\ A^T$ in place of $Row\ A$.

Theorem 4.13
If two matrices A and B are row equivalent, then their row spaces are the same. If B is in echelon form, the nonzero rows of B form a basis for the row space of A as well as for that of B.

The **rank** of A is the dimension of the column space of A.

Theorem 4.14 (The Rank Theorem)
The dimensions of the column space and the row space of an $m \times n$ matrix A are equal. This common dimension, the rank of A, also equals the number of pivot positions in A and satisfies the equation $rank\ A + \dim Nul\ A = n$.

Section 4.7 – Change of Basis

Theorem 4.15
Let $\mathcal{B} = \{\boldsymbol{b}_1, \dots, \boldsymbol{b}_n\}$ and $\mathcal{C} = \{\boldsymbol{c}_1, \dots, \boldsymbol{c}_n\}$ be bases of vector space V. Then there is a unique $n \times n$ matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$ such that

$$[\boldsymbol{x}]_{\mathcal{C}} = P_{\mathcal{C}\leftarrow\mathcal{B}}[\boldsymbol{x}]_{\mathcal{B}}$$

The columns of $P_{\mathcal{C}\leftarrow\mathcal{B}}$ are the $\mathcal{C}$-coordinate vectors of the vectors in the basis $\mathcal{B}$. That is,

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = [[\boldsymbol{b}_1]_{\mathcal{C}} \quad \dots \quad [\boldsymbol{b}_n]_{\mathcal{C}}$$

The matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$ in Theorem 4.15 is called **change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$**.

$$(P_{\mathcal{C}\leftarrow\mathcal{B}})^{-1} = P_{\mathcal{B}\leftarrow\mathcal{C}}$$

$$[\boldsymbol{c}_1 \quad \boldsymbol{c}_2 \ \mid \boldsymbol{b}_1 \quad \boldsymbol{b}_2] \sim [I \quad \mid P_{\mathcal{C}\leftarrow\mathcal{B}}]$$

If $\mathcal{B} = \{\boldsymbol{b}_1, \dots, \boldsymbol{b}_n\}$ and $\mathcal{E}$ is the standard basis $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ in $\mathbb{R}^n$ then $[\boldsymbol{b}_1]_{\mathcal{E}} = \boldsymbol{b}_1$ and likewise for the other vectors in $\mathcal{B}$. In this case, $P_{\mathcal{E}\leftarrow\mathcal{B}}$ is the same as the change-of-the coordinates matrix $P_{\mathcal{B}}$ introduced in Section 4.4, namely $P_{\mathcal{B}} = [\boldsymbol{b}_1 \quad \dots \quad \boldsymbol{b}_n]$.

Chapter 5 – Eigenvalues and Eigenvectors

Section 5.1 – Eigenvectors and Eigen Values

An **eigenvector** of an $n \times n$ matrix A is a nonzero vector $\boldsymbol{x}$ such that $A\boldsymbol{x} = \lambda\boldsymbol{x}$ for some scalar λ. A scalar λ is called in **eigenvalue** of A if there is a nontrivial solution $\boldsymbol{x}$ of $A\boldsymbol{x} = \lambda\boldsymbol{x}$; such an $\boldsymbol{x}$ is called an eigenvector corresponding to λ.

The value λ is an eigenvalue of an $n \times n$ matrix A if and only if the equation $(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$ has a nontrivial solution. The set of all solutions to this equation is $Nul(A - \lambda I)$. This set is a subspace of $\mathbb{R}^n$ and is called the **eigenspace of A corresponding to λ**.

Theorem 5.1
The eigenvalues of a triangular matrix are the entries on its main diagonal.

The vector $\boldsymbol{0}$ is an eigenvalue of A if and only if A is not invertible.

Theorem 5.2
If $\boldsymbol{v}_1, \dots, \boldsymbol{v}_r$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1, \dots, \lambda_r$ of an $n \times n$ matrix A then the set $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_r\}$ is linearly independent.

Section 5.2 – The Characteristic Equation

Theorem 5.3 (Properties of Determinants)
Let A and B be $n \times n$ matrices.

a. A is invertible if and only if $\det A \neq 0$
b. $\det AB = (\det A)(\det B)$
c. $\det A^T = \det A$
d. If A is triangular, then $\det A$ is the product of the entries on the main diagonal of A.
e. A row replacement operation on A does not change the determinant. A row interchange changes the sign of the determinant. A row scaling also scales the determinant by the same scalar factor.

A scalar λ is an eigenvalue of an $n \times n$ matrix A if and only if λ satisfies the **characteristic equation** $\det(A - \lambda I) = 0$. The function on the LHS is called the **characteristic polynomial** of A of degree n. The multiplicity of an eigenvalue λ is its multiplicity as a root of the characteristic equation.

If A and B are $n \times n$ matrices, then A is **similar** to B if there is an invertible matrix P such that $P^{-1}AP = B$, or equivalently, $A = PBP^{-1}$. Since we can write $Q = P^{-1}$, similarity is a reflexive relation. Changing A into $P^{-1}AP$ is called a similarity transformation.

Theorem 5.4
If $n \times n$ matrices A and B are similar, then they have the same characteristic polynomial and hence the same eigenvalues (with the same multiplicities).

Section 5.3 – Diagonalization

A square matrix is said to be **diagonalizable** if A is similar to a diagonal matrix, that is, if $A = PDP^{-1}$ for some invertible matrix P and some diagonal matrix D.

Theorem 5.5 (The Diagonalization Theorem)
An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors. In fact, $A = PDP^{-1}$, with D a diagonal matrix, if and only if the columns of P are n linearly independent eigenvectors of A. In this case, the diagonal entries of D are eigenvalues of A that correspond, respectively, to the eigenvectors in P.

In other words, A is diagonalizable if and only if there are enough eigenvectors to form a basis of $\mathbb{R}^n$. We call such a basis an **eigenvector basis** of $\mathbb{R}^n$.

Steps to Diagonalize a Matrix (From Theorem 5.5)

1. Find the eigenvalues of A, an $n \times n$ matrix
2. Find n linearly independent eigenvectors of A. If this step fails, A cannot be diagonalized
3. Construct P from the vectors in step 2. The order *is not* important; simply take the linearly independent eigenvectors and group them as a matrix.
4. Construct D from the corresponding eigenvalues. The order *is* important; the order of the eigenvalues much match the order chosen for the columns of P.

Theorem 5.6
An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

Theorem 5.7
Let A be an $n \times n$ matrix whose distinct eigenvalues are $\lambda_1, \dots, \lambda_p$.

a. For $1 \leq k \leq p$ the dimension of the i for λ_k is less than or equal to the multiplicity of the eigenvalue λ_k.
b. The matrix A is diagonalizable if and only if the sum of the dimensions of the eigenspace equals n and this happens if and only if (i) the characteristic polynomial factors completely into linear factors and (ii) the dimension of the eigenspace for each λ_k equals the multiplicity of λ_k.
c. If A is diagonalizable and $\mathcal{B}_k$ is a basis for the eigenspace corresponding to λ_k for each k, then the total collection of vectors in the sets $\mathcal{B}_1, \dots, \mathcal{B}_p$ forms an eigenvector basis for $\mathbb{R}^n$.

Section 5.4 – Eigenvectors and Linear Transformations

Let V be an n-dimensional vector space, let W be an m-dimensional vector space, and let T be any linear transformation from V to W. To associate a matrix M with T, choose (ordered) bases $\mathcal{B}$ and $\mathcal{C}$ for V and W respectively. We have that

$$[T(\boldsymbol{x})]_{\mathcal{C}} = M[\boldsymbol{x}]_{\mathcal{B}}$$

where

$$M = [[T(\boldsymbol{b}_1)]_{\mathcal{C}} \quad [T(\boldsymbol{b}_2)]_{\mathcal{C}} \quad \dots \quad [T(\boldsymbol{b}_n)]_{\mathcal{C}}].$$

The matrix M is a representation of T called the matrix for T relative to bases $\mathcal{B}$ and $\mathcal{C}$.

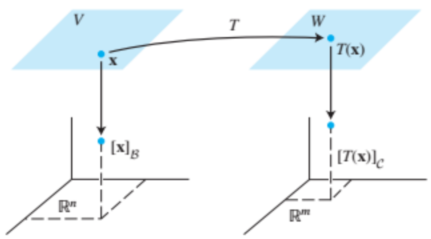


FIGURE 1 A linear transformation from V to W.

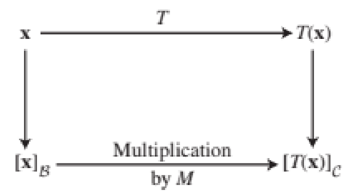


FIGURE 2

In the common case where W is the same as V and the basis $\mathcal{C}$ is the same as $\mathcal{B}$, the matrix M is called **the matrix for T relative to $\mathcal{B}$** or simply the **$\mathcal{B}$-matrix for T** and is denoted by $[T]_{\mathcal{B}}$.

Theorem 5.8 (Diagonal Matrix Representation)
Suppose $A = PDP^{-1}$, where D is a diagonal $n \times n$ matrix. If $\mathcal{B}$ is the basis for $\mathbb{R}^n$ formed from the columns of P, then D is the $\mathcal{B}$ matrix for the transformation $\boldsymbol{x} \longmapsto A\boldsymbol{x}$.

If A is similar to a matrix C with $A = PCP^{-1}$, then C is the $\mathcal{B}$-matrix for the transformation $\boldsymbol{x} \mapsto \boldsymbol{A}x$ when the basis $\mathcal{B}$ is formed from the columns of P. Conversely, if $T: \mathbb{R}^n \to \mathbb{R}^n$ is defined by $T(x) = A\boldsymbol{x}$, and if $\mathcal{B}$ is any basis for $\mathbb{R}^n$, then the $\mathcal{B}$-matrix for T is similar to A.

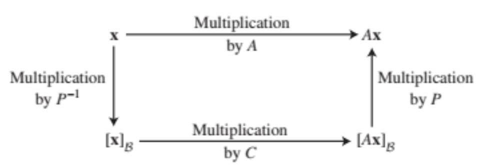


FIGURE 5 Similarity of two matrix representations: $A = PCP^{-1}$.

Chapter 6 – Orthogonality and Least Squares

Section 6.1 – Inner Product, Length, and Orthogonality

Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$. The number $\boldsymbol{u}^T\boldsymbol{v}$ is called the **inner product** of $\boldsymbol{u}$ and $\boldsymbol{v}$, often written as $\boldsymbol{u} \cdot \boldsymbol{v}$. This inner product is also referred to as a **dot product**.

Theorem 6.1
Let $\boldsymbol{u}, \boldsymbol{v}$, and $\boldsymbol{w}$ be vectors in $\mathbb{R}^n$, and let c be a scalar. Then

a. $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$
b. $(\boldsymbol{u} + \boldsymbol{v}) \cdot \boldsymbol{w} = \boldsymbol{u} \cdot \boldsymbol{w} + \boldsymbol{v} \cdot \boldsymbol{w}$
c. $(c\boldsymbol{u}) \cdot \boldsymbol{v} = c(\boldsymbol{u} \cdot \boldsymbol{v}) = \boldsymbol{u} \cdot (c\boldsymbol{v})$
d. $\boldsymbol{u} \cdot \boldsymbol{u} \geq 0$, and $\boldsymbol{u} \cdot \boldsymbol{u} = 0$ if and only if $\boldsymbol{u} = 0$

Properties (b) and (c) can be combined to produce

$$\left(c_1\boldsymbol{u}_1 + \cdots + c_p\boldsymbol{u}_p\right) \cdot \boldsymbol{w} = c_1(\boldsymbol{u}_1 \cdot \boldsymbol{w}) + \cdots + c_p\left(\boldsymbol{u}_p \cdot \boldsymbol{w}\right)$$

The **length** (or **norm**) of $\boldsymbol{v}$ is the nonnegative scalar $\|\boldsymbol{v}\|$ defined by

$$\| \boldsymbol{v} \| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}, \text{ and } \| \boldsymbol{v} \|^2 = \boldsymbol{v} \cdot \boldsymbol{v}$$

For any scalar c, the length of $c\boldsymbol{v}$ is $|c|$ times the length of $\boldsymbol{v}$. That is $\|c\boldsymbol{v}\| = |c|\|\boldsymbol{v}\|$. A vector whose length is 1 is called a unit vector. If we divide a nonzero vector v by its length—that is, multiply by $1/\|\boldsymbol{v}\|$—we obtain a **unit vector** $\boldsymbol{u}$ because the length of $\boldsymbol{u}$ is $\left(\frac{1}{\|\boldsymbol{v}\|}\right)\|\boldsymbol{v}\|$. The process of creating $\boldsymbol{u}$ from $\boldsymbol{v}$ is sometimes called **normalizing** $\boldsymbol{v}$ and we say that $\boldsymbol{u}$ is in the same direction as $\boldsymbol{v}$.

For $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, the distance between $\boldsymbol{u}$ and $\boldsymbol{v}$, written as $\operatorname{dist}(\boldsymbol{u}, \boldsymbol{v})$, is the length of the vector $\boldsymbol{u} - \boldsymbol{v}$. That is, $\operatorname{dist}(\boldsymbol{u}, \boldsymbol{v}) = \| \boldsymbol{u} - \boldsymbol{v} \|$.

Two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ are orthogonal (to each other) if $\boldsymbol{u} \cdot \boldsymbol{v} = 0$.

Theorem 6.2 (The Pythagorean Theorem)
Two vectors u and v are orthogonal if and only if $\| \boldsymbol{u} + \boldsymbol{v} \|^2 = \| \boldsymbol{u} \|^2 + \| \boldsymbol{v} \|^2$.

If a vector $\boldsymbol{z}$ is orthogonal to every vector in a subspace W of $\mathbb{R}^n$, then $\boldsymbol{z}$ is said to be orthogonal to W. The set of all vectors **orthogonal** to W is called the **orthogonal complement** of W and is denoted by $W^\perp$ (and read as "W perpendicular" or simply "W perp").

Two facts about $W^\perp$

1. A vector x is in $W^\perp$ if and only if $\boldsymbol{x}$ is orthogonal to every vector in a set that spans W
2. $W^\perp$ is a subspace of $\mathbb{R}^n$

Theorem 6.3
Let A be an $m \times n$ matrix. The orthogonal complement of the row space of A is the null space of A, and the orthogonal complement of the column space of A is the null space of A^T:

$$(Row\ A)^{\perp} = Nul\ A \text{ and } (Col\ A)^{\perp} = Nul\ A^T$$

Section 6.2 – Orthogonal Sets

A set of vectors $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_p\}$ in $\mathbb{R}^n$ is said to be an **orthogonal set** if each pair of distinct vectors from the set is orthogonal, that is, if $\boldsymbol{u}_i \cdot \boldsymbol{u}_j = 0$ whenever $i \neq j$.

Theorem 6.4
If $S = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent and hence is a basis for the subspace spanned by S.

An **orthogonal basis** for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

Theorem 6.5
Let $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\boldsymbol{y}$ in W, the weights in the linear combination $\boldsymbol{y} = c_1\boldsymbol{u}_1 + \cdots + c_p\boldsymbol{u}_p$ are given by $c_j = \frac{\boldsymbol{y}\cdot\boldsymbol{u}_j}{\boldsymbol{u}_j\cdot\boldsymbol{u}_j} (j = 1, \dots, p)$.

Given a nonzero vector $\boldsymbol{u}$ in $\mathbb{R}^n$, consider the problem of decomposing a vector $\boldsymbol{y}$ in $\mathbb{R}^n$ into the sum of two vectors, one a multiple of $\boldsymbol{u}$ and the other orthogonal to $\boldsymbol{u}$. We wish to write $\boldsymbol{y} = \widehat{\boldsymbol{y}} + \boldsymbol{z}$ where $\hat{y} = \alpha\boldsymbol{u}$ for some scalar α and $\boldsymbol{z}$ orthogonal to $\boldsymbol{u}$. Given any scalar α, let $\boldsymbol{z} = \boldsymbol{y} - \alpha\boldsymbol{u}$. Then $y - \hat{y}$ is orthogonal to u if and only if $\alpha = \frac{\boldsymbol{y}\cdot\boldsymbol{u}}{\boldsymbol{u}\cdot\boldsymbol{u}}$ and $\widehat{\boldsymbol{y}} = \frac{\boldsymbol{y}\cdot\boldsymbol{u}}{\boldsymbol{u}\cdot\boldsymbol{u}}\boldsymbol{u}$. The vector $\widehat{\boldsymbol{y}}$ is called the orthogonal projection of $\boldsymbol{y}$ onto $\boldsymbol{u}$ and the vector $\boldsymbol{z}$ is called the component of $\boldsymbol{y}$ orthogonal to $\boldsymbol{u}$.

The orthogonal projection of $\boldsymbol{y}$ onto a subspace L spanned by $\boldsymbol{u}$ is notated as $\widehat{\boldsymbol{y}}$ and denoted by $proj_L\boldsymbol{y}$ and is called **the orthogonal projection of** y **onto** L.

A set $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_p\}$ is an **orthonormal set** if it is an orthogonal set of unit vectors. If W is the subspace spanned by such a set, then $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_p\}$ is an **orthonormal basis** for W, since the set is automatically linearly independent by Theorem 6.4.

Theorem 6.6
An $m \times n$ matrix U has orthonormal columns if and only if $U^TU = I$.

Theorem 6.7
Let U be an $m \times n$ matrix with orthonormal columns, and let $\boldsymbol{x}$ and $\boldsymbol{y}$ be in $\mathbb{R}^n$. Then

a. $\|U\boldsymbol{x}\| = \|\boldsymbol{x}\|$
b. $(U\boldsymbol{x}) \cdot (U\boldsymbol{y}) = \boldsymbol{x} \cdot \boldsymbol{y}$
c. $(U\boldsymbol{x}) \cdot (U\boldsymbol{y}) = 0$ if and only if $\boldsymbol{x} \cdot \boldsymbol{y} = 0$

An **orthogonal matrix** is a *square* invertible matrix U such that $U^{-1} = U^T$. By Theorem 6, such a matrix has orthonormal columns. Any *square* matrix with orthonormal columns is an orthogonal matrix.

An orthogonal matrix must have orthonormal rows too. To show this, we must show that for an orthogonal matrix U, the matrix U^T is orthogonal. Note that $(U^{-1})^{-1} = U$. But since $U^{-1} = U^T$ (by U as an orthogonal matrix) we substitute to get $(U^T)^{-1} = U$. But since $U = (U^T)^T$, we substitute further to get $(U^T)^{-1} = (U^T)^T$. Hence, by definition, U^T is an orthogonal matrix, meaning if U is an orthogonal matrix, then the rows of U are orthonormal.

Section 6.3 – Orthogonal Projections

Theorem 6.8 (The Orthogonal Decomposition Theorem)
Let W be a subspace of $\mathbb{R}^n$. Then each $\boldsymbol{y}$ in $\mathbb{R}^n$ can be written uniquely in the form $\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{z}$ where $\hat{y}$ is in W and z is in $W^{\perp}$. In fact, if $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_p\}$ is any orthogonal basis of W, then

$$\hat{\boldsymbol{y}} = \frac{\boldsymbol{y} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1}\boldsymbol{u}_1 + \cdots + \frac{\boldsymbol{y} \cdot \boldsymbol{u}_p}{\boldsymbol{u}_p \cdot \boldsymbol{u}_p}\boldsymbol{u}_p$$

And $\boldsymbol{z} = \boldsymbol{y} - \hat{\boldsymbol{y}}$.

The vector $\hat{\boldsymbol{y}}$ is called the **orthogonal projection of** y **onto** W and often is written as $proj_W\boldsymbol{y}$. When W is a one-dimensional subspace, the formula for $\hat{\boldsymbol{y}}$ matches the formula given in Section 6.2.

If $\boldsymbol{y}$ is in $W = Span\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_p\}$, then $proj_W\boldsymbol{y} = \boldsymbol{y}$

Theorem 6.9 (The Best Approximation Theorem)
Let W be a subspace of $\mathbb{R}^n$, let $\boldsymbol{y}$ be any vector in $\mathbb{R}^n$, and let $\hat{\boldsymbol{y}}$ be the orthogonal projection of $\boldsymbol{y}$ onto W. Then $\hat{\boldsymbol{y}}$ is the closest point in W to $\boldsymbol{y}$, in the sense that $\| \boldsymbol{y} - \hat{\boldsymbol{y}} \| < \| \boldsymbol{y} - \boldsymbol{v} \|$ for all $\boldsymbol{v}$ in W distinct from $\hat{\boldsymbol{y}}$.

The vector $\hat{\boldsymbol{y}}$ in Theorem 9 is called **the best approximation to** y **by elements of** W.

Theorem 6.10
If $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

$$proj_W\ \boldsymbol{y} = (\boldsymbol{y} \cdot \boldsymbol{u}_1)\boldsymbol{u}_1 + (\boldsymbol{y} \cdot \boldsymbol{u}_2)\boldsymbol{u}_2 + \cdots + (\boldsymbol{y} \cdot \boldsymbol{u}_p)\boldsymbol{u}_p$$

In other words, if $U = [\boldsymbol{u}_1 \quad \cdots \quad \boldsymbol{u}_p]$, then $proj_W\ \boldsymbol{y} = UU^T\boldsymbol{y}$ for all $\boldsymbol{y}$ in $\mathbb{R}^n$.

Section 6.4 – The Gram-Schmidt Process

The Gram-Schmidt is a simple algorithm for producing an orthogonal or orthonormal basis for any nonzero subspace of $\mathbb{R}^n$.

Theorem 6.11 (The Gram-Schmidt Process)
Given a basis $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$\begin{aligned}
\boldsymbol{v}_1 &= \boldsymbol{x}_1 \\
\boldsymbol{v}_2 &= \boldsymbol{x}_2 - \frac{\boldsymbol{x}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 \\
\boldsymbol{v}_3 &= \boldsymbol{x}_3 - \frac{\boldsymbol{x}_3 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 - \frac{\boldsymbol{x}_3 \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2}\boldsymbol{v}_2 \\
&\vdots \\
\boldsymbol{v}_p &= \boldsymbol{x}_p - \frac{\boldsymbol{x}_p \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 - \frac{\boldsymbol{x}_p \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2}\boldsymbol{v}_2 - \cdots - \frac{\boldsymbol{x}_p \cdot \boldsymbol{v}_{p-1}}{\boldsymbol{v}_{p-1} \cdot \boldsymbol{v}_{p-1}}\boldsymbol{v}_{p-1}
\end{aligned}$$

Then $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ is an orthogonal basis for W. In addition $Span\,\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\} = Span\,\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$ for $1 \leq k \leq p$.

An orthonormal basis is easily constructed from an orthogonal basis $\{\boldsymbol{v}_{1,\ldots,}\ \boldsymbol{v}_p\}$; simply normalize all the $\boldsymbol{v}_k$.

If an $m \times n$ matrix A has linearly independent columns $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, then applying the Gram-Schmit process (with normalization) to $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ amounts to factoring A as described in the next theorem.

Theorem 6.12 (The QR Factorization)
If A is an $m \times n$ matrix with linearly independent columns, then A can be factored as $A = QR$, where Q is an $m \times n$ matrix whose columns form an orthonormal basis for $Col\ A$ and R is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

By construction, the first k columns of Q are an orthonormal basis of $Span\,\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$. From Theorem 6.12, $A = QR$ for some R. To find R, observe that $Q^TQ = I$, because the columns of Q are orthonormal. Hence $Q^TA = Q^T(QR) = IR = R$. Alternatively, note R has entries $r_{ii} = ||\vec{v}_i||$ and $r_{ij} = \vec{u}_i \cdot \vec{b}_j$.

Section 6.5 – Least-Squares Problems

If a situation arises where an application problem has its corresponding system $A\boldsymbol{x} = \boldsymbol{b}$ as inconsistent, the best one can do is find an $\boldsymbol{x}$ that makes $A\boldsymbol{x}$ as close as possible to $\boldsymbol{b}$. Hence, $A\boldsymbol{x}$ is an approximation to $\boldsymbol{b}$. The smallest the distance between $\boldsymbol{b}$ and $A\boldsymbol{x}$, given by $\|\boldsymbol{b} - A\boldsymbol{x}\|$, the better the approximation. The **general least-squares problem** is to find $\boldsymbol{x}$ that makes $\|\boldsymbol{b} - A\boldsymbol{x}\|$ as small as possible.

If A is $m \times n$ and $\boldsymbol{b}$ is in $\mathbb{R}^n$, a **least-squares solution** of $A\boldsymbol{x} = \boldsymbol{b}$ is an $\hat{\boldsymbol{x}}$ in $\mathbb{R}^n$ such that $\|\boldsymbol{b} - A\hat{\boldsymbol{x}}\| \leq \|\boldsymbol{b} - A\boldsymbol{x}\|$ for all $\boldsymbol{x}$ in $\mathbb{R}^n$.

Theorem 6.13
The set of least-squares solutions of $A\boldsymbol{x} = \boldsymbol{b}$ coincides with the nonempty set of solutions of the normal equations $A^T A\boldsymbol{x} = A^T\boldsymbol{b}$**.**

Theorem 6.14
Let A be an $m \times n$ matrix. The following statements are logically equivalent

a. The equation $A\boldsymbol{x} = \boldsymbol{b}$ has a unique least-squares solution for each $\boldsymbol{b}$ in $\mathbb{R}^m$
b. The columns of A are linearly independent
c. The matrix $A^T A$ is invertible

When the statements are true, the least-squares solution $\widehat{\boldsymbol{x}}$ is given by $\widehat{\boldsymbol{x}} = (A^T A)^{-1} A^T \boldsymbol{b}$.

When a least-squares solution $\widehat{\boldsymbol{x}}$ is used to produce $A\widehat{\boldsymbol{x}}$ as an approximation to $\boldsymbol{b}$, the distance from $\boldsymbol{b}$ to $A\widehat{\boldsymbol{x}}$ is called the **least-squares error** of this approximation.

Theorem 6.15
Given an $m \times n$ matrix A with linearly independent columns, let $A = QR$ be a QR factorization of A as in Theorem 6.12. Then, for each $\boldsymbol{b}$ in $\mathbb{R}^m$, the equation $A\boldsymbol{x} = \boldsymbol{b}$ has a unique least-squares solution, given by $\widehat{\boldsymbol{x}} = R^{-1} Q^T \boldsymbol{b}$.

Section 6.7 – Inner Product Spaces

An inner product on a vector space V is a function that, to each pair of vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in V, associates a real number $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ and satisfies the following axioms, for all $\boldsymbol{u}, \boldsymbol{v}$, and $\boldsymbol{w}$ in V and all scalars c.

1. $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle$
2. $\langle \boldsymbol{u} + \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{u}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}, \boldsymbol{w} \rangle$
3. $\langle c\boldsymbol{u}, \boldsymbol{v} \rangle = c\langle \boldsymbol{u}, \boldsymbol{v} \rangle$
4. $\langle \boldsymbol{u}, \boldsymbol{u} \rangle \geq 0$ and $\langle \boldsymbol{u}, \boldsymbol{u} \rangle = 0$ if and only if $\boldsymbol{u} = \boldsymbol{0}$

A vector space with an inner product is called an **inner product space**.

Let V be an inner product space, with the inner product denoted by $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$. Just as in $\mathbb{R}^n$, we define **length**, or **norm**, of a vector $\boldsymbol{v}$ to be the scalar $\|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$. Equivalently, $\|\boldsymbol{v}\|^2 = \langle \boldsymbol{v}, \boldsymbol{v} \rangle$. A **unit vector** is one whose length is 1. The **distance between** u **and** v is $\|\boldsymbol{u} - \boldsymbol{v}\|$. Vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are **orthogonal** if $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$.

Theorem 6.16 (The Cauchy-Schwarz Inequality)
For all $\boldsymbol{u}, \boldsymbol{v}$ in V, $|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \|\boldsymbol{u}\| \|\boldsymbol{v}\|$.

Theorem 6.17 (The Triangle Inequality)
For all $\boldsymbol{u}, \boldsymbol{v}$ in V, $\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$.

Chapter 7 – Symmetric Matrices and Quadratic Forms

Section 7.1 – Diagonalization of Symmetric Matrices

A **symmetric matrix** is a matrix A such that $A^T = A$. Such a matrix is necessarily square. Its main diagonal entries are arbitrary, but its other entries occur in pairs—on opposite sides of the main diagonal.

Theorem 7.1
If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.

An $n \times n$ matrix A is said to be **orthogonally diagonalizable** if there are an orthogonal matrix P with $P^{-1} = P^T$ and a diagonal matrix D such that $A = PDP^T = PDP^{-1}$. Such a diagonalization requires n linearly independent and orthonormal eigenvectors. This happens when A is orthogonally diagonalizable meaning $A^T = (PDP^T)^T = P^{TT}D^TP^T = PDP^T = A$ meaning A is symmetric.

Theorem 7.2
An $n \times n$ matrix A is orthogonally diagnolizable if and only if A is a symmetric matrix.

The set of eigenvalues of a matrix A is sometimes called the **spectrum of** A.

Theorem 7.3 (The Spectral Theorem for Symmetric Matrices)
An $n \times n$ symmetric matrix A has the following properties.

a. A has n real eigenvalues, counting multiplicities
b. The dimension of the eigenspace for each eigenvalue λ equals the multiplicity of λ as a root of the characteristic equation
c. The eigenspaces are mutually orthogonal, in the sense that eigenvectors corresponding to different eigenvalues are orthogonal
d. A is orthogonally diagonalizable

Suppose $A = PDP^{-1}$ where the columns of P are orthonormal eigenvectors $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ of A and the corresponding eigenvalues $\lambda_1, \dots, \lambda_n$ are in the diagonal matrix D. Then, since $P^{-1} = P^T$,

$$\begin{aligned} A = PDP^T &= [\boldsymbol{u}_1 \quad \cdots \quad \boldsymbol{u}_n] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1^T \\ \vdots \\ \boldsymbol{u}_n^T \end{bmatrix} \\ &= [\lambda_1\boldsymbol{u}_1 \quad \cdots \quad \lambda_n\boldsymbol{u}_n] \begin{bmatrix} \boldsymbol{u}_1^T \\ \vdots \\ \boldsymbol{u}_n^T \end{bmatrix} \end{aligned}$$

Using a column-row expansion of a product, we can write

$$A = \lambda_1\boldsymbol{u}_1\boldsymbol{u}_1^T + \lambda_2\boldsymbol{u}_2\boldsymbol{u}_2^T + \cdots + \lambda_n\boldsymbol{u}_n\boldsymbol{u}_n^T$$

This representation of A is called the **spectral decomposition** of A because it breaks up A into pieces determined by the spectrum of A. Each term in the equation above is an $n \times n$ or rank 1. Furthermore, each matrix $\boldsymbol{u}_j\boldsymbol{u}_j^T$ is a **projection matrix** in the sense that for each $\boldsymbol{x}$ in $\mathbb{R}^n$, the vector $(\boldsymbol{u}_j\boldsymbol{u}_j^T)\boldsymbol{x}$ is the orthogonal projection of $\boldsymbol{x}$ onto the subspace spanned by $\boldsymbol{u}_j$.

Section 7.2 – Quadratic Forms

A quadratic form on $\mathbb{R}^n$ is a function Q defined on $\mathbb{R}^n$ whose value at a vector $\boldsymbol{x}$ in $\mathbb{R}^n$ can be computed by an expression of the form $Q(\boldsymbol{x}) = \boldsymbol{x}^T A\boldsymbol{x}$, where A is an $n \times n$ symmetric matrix. The matrix A is called the **matrix of the quadratic** form.

If x represents a variable vector in $\mathbb{R}^n$, the a change of variable is an equation of the form $\boldsymbol{x} = P\boldsymbol{y}$ or equivalently $\boldsymbol{y} = P^{-1}\boldsymbol{x}$ where P is an invertible matrix and $\boldsymbol{y}$ is a new variable vector in $\mathbb{R}^n$. Here, $\boldsymbol{y}$ is the coordinate vector of $\boldsymbol{x}$ relative to the basis of $\mathbb{R}^n$ determined by the columns of P. If the change of variable is made in a quadratic form $\boldsymbol{x}^T A\boldsymbol{x}$, then $\boldsymbol{x}^T A\boldsymbol{x} = (P\boldsymbol{y})^T A(P\boldsymbol{y}) = \boldsymbol{y}^T P^T AP\boldsymbol{y} = \boldsymbol{y}^T(P^T AP)\boldsymbol{y}$ and the new matrix of the quadratic form is $P^T AP$. Since A is symmetric, Theorem 7.2 guarantees that there is an orthogonal matrix P such that $P^T AP$ is a diagonal matrix D, and the quadratic form becomes $\mathbf{y}^T D\mathbf{y}$.

Theorem 7.4 (The Principal Axis Theorem)
Let A be an $n \times n$ symmetric matrix. Then there is an orthogonal change of variable, $\boldsymbol{x} = P\boldsymbol{y}$, that transforms the quadratic form $\boldsymbol{x}^T A\boldsymbol{x}$ into a quadratic form $\boldsymbol{y}^T D\boldsymbol{y}$ with no cross-product term.

The columns of P in Theorem 7.4 are called the principal axes of the quadratic form $\boldsymbol{x}^T A\boldsymbol{x}$. The vector $\boldsymbol{y}$ is the coordinate vector of $\boldsymbol{x}$ relative to the orthonormal basis of $\mathbb{R}^n$ given by these principal axes.

Suppose $Q(\boldsymbol{x}) = \boldsymbol{x}^T A\boldsymbol{x}$ where A is an invertible 2×2 symmetric matrix, and let c be a constant. It can be shown that the set of all $\boldsymbol{x}$ in $\mathbb{R}^2$ that satisfy $\boldsymbol{x}^T A\boldsymbol{x} = c$ either corresponds to an ellipse, circle, hyperbola, two intersecting lines, or a single point, or no points. If A is a diagonal matrix, then the graph is in standard position, meaning it is symmetric with respect to both axes. If A is not a diagonal matrix, then the graph is rotated out of standard position where the principal axis, determined by the eigenvectors of A, form a new coordinated system with respect to which the graph is in standard positon.

A quadratic form Q is

a. **Positive definite** if $Q(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \neq 0$
b. **Negative definite** if $Q(\boldsymbol{x}) < 0$ for all $\boldsymbol{x} \neq 0$
c. **Indefinite** if $Q(\boldsymbol{x})$ assumes both positive and negative values

Also, Q is sid to be **positive semidefinite** if $Q(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x}$ and **negative semidefinite** if $Q(\boldsymbol{x}) \leq 0$ for all $\boldsymbol{x}$.

Theorem 7.5 (Quadratic Forms and Eigenvalues)
Let A be an $n \times n$ symmetric matrix. Then a quadratic form $\boldsymbol{x}^T A\boldsymbol{x}$ is

a. Positive definite if and only if the eigenvalues of A are all positive
b. Negative definite if and only if the eigenvalues of A are all negative
c. Indefinite if and only if A has both positive and negative eigenvalues

Section 7.3 – Constrained Optimization

Often, it is needed to find the maximum or minimum value of a quadratic form $Q(\boldsymbol{x})$ for $\boldsymbol{x}$ in some specified set. Typically, the problem is arranged so that $\boldsymbol{x}$ varies over the set of unit vectors. This is known as the **constrained optimization problem**. The requirement that a vector $\boldsymbol{x}$ in $\mathbb{R}^n$ be a unit vector can be stated in several equivalent ways:

$$\|\boldsymbol{x}\| = 1 \text{ or } \|\boldsymbol{x}\|^2 = 1 \text{ or } \boldsymbol{x}^T\boldsymbol{x} = 1$$

which expands to

$$x_1^2 + x_2^2 + \cdots + x_n^2 = 1.$$

It can be shown that for any symmetric matrix A, the set of all possible values of $\boldsymbol{x}^T A\boldsymbol{x}$ for $\|\boldsymbol{x}\| = 1$ is a closed interval on the real axis. Denote the left and the right endpoints of this interval by m and M respectively. That is, let $m = \min\{\boldsymbol{x}^T A\boldsymbol{x} \mid \|\boldsymbol{x}\| = 1\}$ and $M = \max\{\boldsymbol{x}^T A\boldsymbol{x} \mid \|\boldsymbol{x}\| = 1\}$.

Theorem 7.6
Let A be a symmetric matrix and define m and M as above. Then M is the greatest eigenvalue λ_1 of A and m is the least eigenvalue of A. The value $\boldsymbol{x}^T A\boldsymbol{x}$ is M when $\boldsymbol{x}$ is a unit eigenvector $\boldsymbol{u}_1$ corresponding to M. The value of $\boldsymbol{x}^T A\boldsymbol{x}$ is m when $\boldsymbol{x}$ I a unit eigenvector corresponding to m.

Theorem 7.7
Let A, λ_1, and $\boldsymbol{u}_1$ be as in Theorem 6. Then the maximum value of $\boldsymbol{x}^T A\boldsymbol{x}$ subject to the constraints $\boldsymbol{x}^T\boldsymbol{x} = 1, \boldsymbol{x}^T\boldsymbol{u}_1 = 0$ is the second greatest eigenvalue λ_2 and this maximum is attained when $\boldsymbol{x}$ is an eigenvector $\boldsymbol{u}_2$ corresponding to λ_2.

Theorem 7.8
Let A be a symmetric $n \times n$ matrix with an orthogonal diagonalization $A = PDP^{-1}$ where the entries on the diagonal of D are arranged so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and where the columns of P are corresponding unit eigenvectors $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$. Then, for $k = 2, \dots, n$, the maximum value of $\boldsymbol{x}^T A\boldsymbol{x}$ subject to the constraints $\boldsymbol{x}^T\boldsymbol{x} = 1, \boldsymbol{x}^T\boldsymbol{u}_1 = 0, \dots, \boldsymbol{x}^T\boldsymbol{u}_{k-1} = 0$ is the eigenvalue λ_k and this maximum is attained at $\boldsymbol{x} = \boldsymbol{u}_k$.