Math 22: Linear Algebra and Applications

Farhan Sadeek

Placement Test
Last Updated: April 4, 2026

Contents

1 Linear Equations in Linear Algebra

1 Systems of Linear Equations

Definition 1.1 (Linear Equation)
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 $a_1, a_2, \dots, a_n$ and b are constants, called the **coefficients** and **constant term**, respectively.

For example, $3x_1 - 2x_2 + 5x_3 = 7$ is a linear equation in the variables x_1, x_2, x_3 with coefficients $3, -2, 5$ and constant term 7.

Definition 1.2 (System of Linear Equations)
A **system of linear equations** is a collection of one or more linear equations involving the same set of variables.

Definition 1.3 (Solution)
A **solution** to a system of linear equations is an assignment of values to the variables that satisfies all equations in the system simultaneously.

Definition 1.4 (Solution Set)
The **solution set** of a linear system is the set of all possible solutions. Two linear systems are called **equivalent** if they have the same solution set.

A system of linear equations has either:

1. **No solution** (the system is **inconsistent**), or
2. **Exactly one solution**, or
3. **Infinitely many solutions**.

A system is **consistent** if it has at least one solution (cases 2 or 3).

Definition 1.5 (Augmented Matrix)
Given a system of linear equations, the **augmented matrix** is obtained by writing the coefficients and constant terms in a rectangular array. For the system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

the augmented matrix is

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right].$$

Definition 1.6 (Elementary Row Operations)
The following three operations on the rows of a matrix are called **elementary row operations**:

1. **Replacement**: Replace one row by the sum of itself and a multiple of another row. ($R_i \leftarrow R_i + cR_j$)
2. **Interchange**: Interchange two rows. ($R_i \leftrightarrow R_j$)
3. **Scaling**: Multiply all entries in a row by a nonzero constant. ($R_i \leftarrow cR_i$, $c \neq 0$)

Two matrices are **row equivalent** if one can be obtained from the other by a sequence of elementary row operations.

Theorem 1.7
If the augmented matrices of two linear systems are row equivalent, then the two systems have the same solution set.

Example 1.8
Solve the system:

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_2 - 8x_3 &= 8 \\ 5x_1 - 5x_3 &= 10 \end{aligned}$$

Solution. The augmented matrix is

$$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{array}\right].$$

Apply $R_3 \leftarrow R_3 - 5R_1$:

$$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & 10 & -10 & 10 \end{array}\right].$$

Apply $R_3 \leftarrow R_3 - 5R_2$:

$$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & 0 & 30 & -30 \end{array}\right].$$

Scale: $R_2 \leftarrow \frac{1}{2}R_2$ and $R_3 \leftarrow \frac{1}{30}R_3$:

$$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & -1 \end{array}\right].$$

Back substitution: From R_3, $x_3 = -1$. From R_2, $x_2 - 4(-1) = 4$, so $x_2 = 0$. From R_1, $x_1 - 2(0) + (-1) = 0$, so $x_1 = 1$. The unique solution is $(x_1, x_2, x_3) = (1, 0, -1)$. □

Example 1.9

Solve the system:

$$x_1 + 3x_2 = 5$$
$$2x_1 + 5x_2 = 8$$

Solution. The augmented matrix is

$$\left[\begin{array}{cc|c} 1 & 3 & 5 \\ 2 & 5 & 8 \end{array}\right].$$

Apply $R_2 \leftarrow R_2 - 2R_1$:

$$\left[\begin{array}{cc|c} 1 & 3 & 5 \\ 0 & -1 & -2 \end{array}\right].$$

Scale $R_2 \leftarrow -R_2$:

$$\left[\begin{array}{cc|c} 1 & 3 & 5 \\ 0 & 1 & 2 \end{array}\right].$$

Apply $R_1 \leftarrow R_1 - 3R_2$:

$$\left[\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1 & 2 \end{array}\right].$$

The unique solution is $x_1 = -1$, $x_2 = 2$. □

Example 1.10 (Inconsistent System)
Show that the following system has no solution:

$$x_1 + 2x_2 = 3$$
$$2x_1 + 4x_2 = 7$$

Solution. Row reduce the augmented matrix:

$$\left[\begin{array}{cc|c} 1 & 2 & 3 \\ 2 & 4 & 7 \end{array}\right] \xrightarrow{R_2 \leftarrow R_2 - 2R_1} \left[\begin{array}{cc|c} 1 & 2 & 3 \\ 0 & 0 & 1 \end{array}\right].$$

The second row says $0 = 1$, which is impossible. The system is inconsistent. □

Example 1.11 (Infinitely Many Solutions)
Solve the system:

$$x_1 - 2x_2 + 3x_3 = 1$$
$$2x_1 - 4x_2 + 6x_3 = 2$$

Solution. Row reduce:

$$\left[\begin{array}{ccc|c} 1 & -2 & 3 & 1 \\ 2 & -4 & 6 & 2 \end{array}\right] \xrightarrow{R_2 \leftarrow R_2 - 2R_1} \left[\begin{array}{ccc|c} 1 & -2 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

There is one pivot (column 1), and x_2, x_3 are free variables. The general solution is $x_1 = 1 + 2x_2 - 3x_3$, where x_2 and x_3 can be any real numbers. This gives infinitely many solutions. □

2 Row Reduction and Echelon Forms

Definition 1.12 (Echelon Form)
A matrix is in **echelon form** (or **row echelon form**) if it satisfies:

1. All nonzero rows are above any rows of all zeros.
2. Each leading entry (leftmost nonzero 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 zero.

Definition 1.13 (Reduced Echelon Form)
A matrix is in **reduced row echelon form (RREF)** if, in addition to being in echelon form:

4. The leading entry in each nonzero row is 1 (called a **pivot**).
5. Each leading 1 is the only nonzero entry in its column.

Theorem 1.14 (Uniqueness of RREF)

Each matrix is row equivalent to one and only one reduced echelon matrix.

Definition 1.15 (Pivot Position and Pivot Column)

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 position. A column that is not a pivot column is called a **free variable column**.

Fact 1.16 (Row Reduction Algorithm)

The **row reduction algorithm** proceeds in two phases:

1. **Forward phase**: Beginning with the leftmost nonzero column, use row operations to create zeros below each pivot, moving left to right. This produces an echelon form.
2. **Backward phase**: Beginning with the rightmost pivot, use row operations to create zeros above each pivot, moving right to left. Scale each pivot to 1. This produces the reduced echelon form.

Example 1.17

Row reduce the matrix to RREF:

$$\begin{bmatrix} 0 & 3 & -6 & 6 & 4 \\ 3 & -7 & 8 & -5 & 8 \\ 3 & -9 & 12 & -9 & 6 \end{bmatrix}.$$

Solution. Swap $R_1 \leftrightarrow R_2$:

$$\begin{bmatrix} 3 & -7 & 8 & -5 & 8 \\ 0 & 3 & -6 & 6 & 4 \\ 3 & -9 & 12 & -9 & 6 \end{bmatrix}.$$

$R_3 \leftarrow R_3 - R_1$:

$$\begin{bmatrix} 3 & -7 & 8 & -5 & 8 \\ 0 & 3 & -6 & 6 & 4 \\ 0 & -2 & 4 & -4 & -2 \end{bmatrix}.$$

$R_3 \leftarrow R_3 + \frac{2}{3}R_2$:

$$\begin{bmatrix} 3 & -7 & 8 & -5 & 8 \\ 0 & 3 & -6 & 6 & 4 \\ 0 & 0 & 0 & 0 & \frac{2}{3} \end{bmatrix}.$$

Scale rows, then back-substitute to get RREF:

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

The pivot columns are $1, 2, 5$. Columns 3 and 4 correspond to free variables. □

Theorem 1.18 (Existence and Uniqueness)
A linear system is consistent if and only if the rightmost column of the augmented matrix is *not* a pivot column. In other words, the echelon form has no row of the form $[0\ 0\ \cdots\ 0\ \mid\ b]$ with $b \neq 0$. If a consistent system has free variables, it has infinitely many solutions; if it has no free variables, it has a unique solution.

3 Vector Equations

Definition 1.19 (Vectors in $\mathbb{R}^n$)
A **vector** in $\mathbb{R}^n$ is an ordered list of n real numbers, usually written as a column:

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}.$$

The set of all vectors with n entries is denoted $\mathbb{R}^n$.

Definition 1.20 (Vector Addition and Scalar Multiplication)
Given vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and a scalar $c \in \mathbb{R}$:

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 + v_1 \\ \vdots \\ u_n + v_n \end{bmatrix}, \qquad c\mathbf{u} = \begin{bmatrix} cu_1 \\ \vdots \\ cu_n \end{bmatrix}.$$

Definition 1.21 (Linear Combination)
Given vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p \in \mathbb{R}^n$ and scalars $c_1, c_2, \ldots, c_p$, the vector

$$\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p$$

is called a **linear combination** of $\mathbf{v}_1, \ldots, \mathbf{v}_p$ with **weights** $c_1, \ldots, c_p$.

Definition 1.22 (Span)
The set of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_p$ is called the **span** of $\mathbf{v}_1, \ldots, \mathbf{v}_p$, denoted $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$:

$$\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\} = \{c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p : c_1, \ldots, c_p \in \mathbb{R}\}.$$

The vector equation

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$$

has the same solution set as the linear system whose augmented matrix is

$$[\mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \cdots \;\; \mathbf{a}_n \;\mid\; \mathbf{b}].$$

In particular, $\mathbf{b}$ is a linear combination of $\mathbf{a}_1, \dots, \mathbf{a}_n$ if and only if the corresponding linear system is consistent.

Example 1.23

Let $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 5 \\ -13 \\ -3 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} -3 \\ 8 \\ 1 \end{bmatrix}$. Determine whether $\mathbf{b}$ is a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$.

Solution. We need to determine whether there exist scalars x_1, x_2 such that $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{b}$. Row reduce the augmented matrix $[\mathbf{a}_1 \;\; \mathbf{a}_2 \;\mid\; \mathbf{b}]$:

$$\left[\begin{array}{cc|c} 1 & 5 & -3 \\ -2 & -13 & 8 \\ 3 & -3 & 1 \end{array}\right] \xrightarrow[R_3 \leftarrow R_3 - 3R_1]{R_2 \leftarrow R_2 + 2R_1} \left[\begin{array}{cc|c} 1 & 5 & -3 \\ 0 & -3 & 2 \\ 0 & -18 & 10 \end{array}\right] \xrightarrow{R_3 \leftarrow R_3 - 6R_2} \left[\begin{array}{cc|c} 1 & 5 & -3 \\ 0 & -3 & 2 \\ 0 & 0 & -2 \end{array}\right].$$

The last row gives $0 = -2$, which is a contradiction. The system is inconsistent, so $\mathbf{b}$ is *not* a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$. □

Fact 1.24 (Geometric Interpretation of Span)

In $\mathbb{R}^3$:

- $\text{Span}\{\mathbf{v}\}$ (one nonzero vector) is a line through the origin.
- $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ (two linearly independent vectors) is a plane through the origin.
- $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ (three linearly independent vectors) is all of $\mathbb{R}^3$.

If the vectors are linearly dependent, the span has lower dimension than the number of vectors.

4 The Matrix Equation $A\mathbf{x} = \mathbf{b}$

Definition 1.25 (Matrix–Vector Product)

If A is an $m \times n$ matrix with columns $\mathbf{a}_1, \dots, \mathbf{a}_n$, and $\mathbf{x} \in \mathbb{R}^n$, then the **matrix–vector product** $A\mathbf{x}$ is the linear combination of the columns of A using the entries of $\mathbf{x}$ as weights:

$$A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n.$$

Theorem 1.26

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

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

Theorem 1.27 (Properties of $A\mathbf{x}$)

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

1. $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$.
2. $A(c\mathbf{u}) = c(A\mathbf{u})$.

These two properties together say that $\mathbf{x} \mapsto A\mathbf{x}$ is a linear transformation.

Example 1.28

Let $A = \begin{bmatrix} 1 & 3 & 4 \\ -4 & 2 & -6 \\ -3 & -2 & -7 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$. For which values of $\mathbf{b}$ does $A\mathbf{x} = \mathbf{b}$ have a solution?

Solution. Row reduce $[A \mid \mathbf{b}]$:

$$\left[\begin{array}{ccc|c} 1 & 3 & 4 & b_1 \\ -4 & 2 & -6 & b_2 \\ -3 & -2 & -7 & b_3 \end{array}\right] \xrightarrow[R_3 \leftarrow R_3 + 3R_1]{R_2 \leftarrow R_2 + 4R_1} \left[\begin{array}{ccc|c} 1 & 3 & 4 & b_1 \\ 0 & 14 & 10 & b_2 + 4b_1 \\ 0 & 7 & 5 & b_3 + 3b_1 \end{array}\right].$$

Apply $R_2 \leftarrow R_2 - 2R_3$:

$$\left[\begin{array}{ccc|c} 1 & 3 & 4 & b_1 \\ 0 & 0 & 0 & b_2 + 4b_1 - 2(b_3 + 3b_1) \\ 0 & 7 & 5 & b_3 + 3b_1 \end{array}\right] = \left[\begin{array}{ccc|c} 1 & 3 & 4 & b_1 \\ 0 & 0 & 0 & b_2 - 2b_1 - 2b_3 \\ 0 & 7 & 5 & b_3 + 3b_1 \end{array}\right].$$

The system is consistent if and only if $b_2 - 2b_1 - 2b_3 = 0$, i.e., $b_2 = 2b_1 + 2b_3$. Since not every $\mathbf{b}$ satisfies this condition, the columns of A do *not* span $\mathbb{R}^3$, and A does not have a pivot in every row. □

5 Solution Sets of Linear Systems

Definition 1.29 (Homogeneous System)

A system of linear equations is **homogeneous** if it can be written as $A\mathbf{x} = \mathbf{0}$, where $\mathbf{0}$ is the zero vector. A homogeneous system always has at least one solution, the **trivial solution** $\mathbf{x} = \mathbf{0}$.

Theorem 1.30

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

Fact 1.31 (Parametric Vector Form)

The solution set of a homogeneous system $A\mathbf{x} = \mathbf{0}$ can be written in **parametric vector form**:

$$\mathbf{x} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_k\mathbf{v}_k \quad (t_1, \ldots, t_k \text{ free})$$

where $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are vectors determined by the free variables. Geometrically, this is a subspace of $\mathbb{R}^n$ through the origin.

Theorem 1.32 (Structure of Solution Sets)

If the equation $A\mathbf{x} = \mathbf{b}$ is consistent and $\mathbf{p}$ is any particular solution, then the full solution set is

$$\{\mathbf{x} : A\mathbf{x} = \mathbf{b}\} = \{\mathbf{p} + \mathbf{v}_h : A\mathbf{v}_h = \mathbf{0}\}.$$

That is, every solution of $A\mathbf{x} = \mathbf{b}$ has the form $\mathbf{x} = \mathbf{p} + \mathbf{v}_h$, where $\mathbf{v}_h$ is a solution of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

Example 1.33

Describe the solution set of $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 3 & 5 & -4 \\ -3 & -2 & 4 \\ 6 & 1 & -8 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 7 \\ -1 \\ -4 \end{bmatrix}.$$

Solution. Row reduce the augmented matrix $[A \mid \mathbf{b}]$:

$$\left[\begin{array}{ccc|c} 3 & 5 & -4 & 7 \\ -3 & -2 & 4 & -1 \\ 6 & 1 & -8 & -4 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & -\frac{4}{3} & -1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

So $x_1 = -1 + \frac{4}{3}x_3$, $x_2 = 2$, x_3 is free. In parametric vector form:

$$\mathbf{x} = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 4/3 \\ 0 \\ 1 \end{bmatrix}.$$

This is a line in $\mathbb{R}^3$ through $(-1, 2, 0)$ in the direction $(4/3, 0, 1)$. □

6 Linear Independence

Definition 1.34 (Linear Independence)
An indexed set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ is **linearly independent** if the vector equation

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

has only the trivial solution $c_1 = c_2 = \cdots = c_p = 0$. The set is **linearly dependent** if there exist weights $c_1, \ldots, c_p$, not all zero, such that the equation holds.

Fact 1.35 (Practical Test for Linear Independence)
To test whether $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly independent, place the vectors as columns of a matrix and row reduce. The set is linearly independent if and only if every column is a pivot column (equivalently, there are no free variables).

Theorem 1.36 (Characterization of Linear Dependence)
An indexed set $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ with $p \geq 2$ is linearly dependent if and only if at least one of the vectors in S is a linear combination of the others.

Proof. ($\Rightarrow$) Suppose $c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$ with some $c_j \neq 0$. Then we can solve for $\mathbf{v}_j$:

$$\mathbf{v}_j = -\frac{c_1}{c_j}\mathbf{v}_1 - \cdots - \frac{c_{j-1}}{c_j}\mathbf{v}_{j-1} - \frac{c_{j+1}}{c_j}\mathbf{v}_{j+1} - \cdots - \frac{c_p}{c_j}\mathbf{v}_p,$$

so $\mathbf{v}_j$ is a linear combination of the other vectors.

($\Leftarrow$) If $\mathbf{v}_j = a_1\mathbf{v}_1 + \cdots + a_{j-1}\mathbf{v}_{j-1} + a_{j+1}\mathbf{v}_{j+1} + \cdots + a_p\mathbf{v}_p$, then

$$a_1\mathbf{v}_1 + \cdots + a_{j-1}\mathbf{v}_{j-1} + (-1)\mathbf{v}_j + a_{j+1}\mathbf{v}_{j+1} + \cdots + a_p\mathbf{v}_p = \mathbf{0}$$

is a nontrivial linear combination (the coefficient of $\mathbf{v}_j$ is $-1 \neq 0$), so the set is linearly dependent. □

Theorem 1.37
If a set contains more vectors than entries in each vector (that is, if $p > n$ in $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\} \subset \mathbb{R}^n$), then the set is linearly dependent.

Proof. Placing the vectors as columns gives an $n \times p$ matrix A with $p > n$. The matrix has at most n pivots, so at least $p - n \geq 1$ columns are non-pivot columns. Therefore $A\mathbf{x} = \mathbf{0}$ has at least one free variable, giving a nontrivial solution. □

Theorem 1.38

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

Proof. Suppose $\mathbf{v}_k = \mathbf{0}$. Then $0\mathbf{v}_1 + \cdots + 0\mathbf{v}_{k-1} + 1\mathbf{v}_k + 0\mathbf{v}_{k+1} + \cdots + 0\mathbf{v}_p = \mathbf{0}$ is a nontrivial linear combination (the coefficient of $\mathbf{v}_k$ is $1 \neq 0$). □

Fact 1.39 (Special Case: Two Vectors)

A set of two vectors $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly dependent if and only if one vector is a scalar multiple of the other. Geometrically, two vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ are linearly dependent if and only if they lie on the same line through the origin.

Fact 1.40 (Special Case: One Vector)

A set $\{\mathbf{v}\}$ containing a single vector is linearly dependent if and only if $\mathbf{v} = \mathbf{0}$. Any single nonzero vector is linearly independent by itself.

Example 1.41

Determine if the vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$ are linearly independent.

Solution. Row reduce the matrix with columns $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$:

$$\begin{bmatrix} 1 & 4 & 2 \\ 2 & 5 & 1 \\ 3 & 6 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 4 & 2 \\ 0 & -3 & -3 \\ 0 & -6 & -6 \end{bmatrix} \sim \begin{bmatrix} 1 & 4 & 2 \\ 0 & -3 & -3 \\ 0 & 0 & 0 \end{bmatrix}.$$

There are only 2 pivots, so x_3 is a free variable. The system $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = \mathbf{0}$ has nontrivial solutions, so the vectors are **linearly dependent**. Indeed, $\mathbf{v}_3 = -\mathbf{v}_2 + \mathbf{v}_1$ (i.e. $\mathbf{v}_1 - \mathbf{v}_2 + \mathbf{v}_3 = \mathbf{0}$). □

7 Linear Transformations; Matrix Representations

Definition 1.42 (Linear Transformation)
A **linear transformation** $T : \mathbb{R}^n \to \mathbb{R}^m$ is a mapping such that for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and all scalars c:

1. $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ (additivity)
2. $T(c\mathbf{u}) = cT(\mathbf{u})$ (homogeneity)

Equivalently, $T(c_1\mathbf{u}_1 + c_2\mathbf{u}_2) = c_1T(\mathbf{u}_1) + c_2T(\mathbf{u}_2)$, and more generally T preserves all linear combinations.

Note that every linear transformation satisfies $T(\mathbf{0}) = \mathbf{0}$.

Theorem 1.43 (Standard Matrix)
Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there exists a unique $m \times n$ matrix A such that

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

The matrix A is called the **standard matrix** for T, and its columns are $T(\mathbf{e}_1), T(\mathbf{e}_2), \dots, T(\mathbf{e}_n)$, where $\mathbf{e}_j$ is the jth standard basis vector.

Definition 1.44 (Onto and One-to-One)
A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is:

- **Onto** if for every $\mathbf{b} \in \mathbb{R}^m$, there exists at least one $\mathbf{x} \in \mathbb{R}^n$ such that $T(\mathbf{x}) = \mathbf{b}$.
- **One-to-one** if for each $\mathbf{b} \in \mathbb{R}^m$, the equation $T(\mathbf{x}) = \mathbf{b}$ has at most one solution.

Theorem 1.45
Let $T : \mathbb{R}^n \to \mathbb{R}^m$ with standard matrix A. Then:

1. T is onto $\iff$ the columns of A span $\mathbb{R}^m$ $\iff$ A has a pivot in every row.
2. T is one-to-one $\iff$ the columns of A are linearly independent $\iff$ A has a pivot in every column.

Example 1.46 (Geometric Transformations in $\mathbb{R}^2$)

Common linear transformations and their standard matrices:

- **Reflection through x_1-axis**: $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. This sends (x_1, x_2) to $(x_1, -x_2)$.
- **Reflection through x_2-axis**: $A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$. This sends (x_1, x_2) to $(-x_1, x_2)$.
- **Reflection through the line $x_2 = x_1$**: $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. This sends (x_1, x_2) to (x_2, x_1).
- **Rotation by angle θ counterclockwise**: $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.
- **Projection onto the x_1-axis**: $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. This sends (x_1, x_2) to $(x_1, 0)$.
- **Shear in the x_1-direction**: $A = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$. This sends (x_1, x_2) to $(x_1 + kx_2, x_2)$.
- **Contraction/dilation by factor $r > 0$**: $A = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} = rI$. This scales every vector by the factor r.

Example 1.47 (Finding the Standard Matrix)

Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be defined by $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 - x_2 \\ 2x_1 \\ x_1 + 3x_2 \end{bmatrix}$. Find the standard matrix A for T.

Solution. Compute $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$:

$$T(\mathbf{e}_1) = T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad T(\mathbf{e}_2) = T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix}.$$

Therefore $A = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 2 & 0 \\ 1 & 3 \end{bmatrix}$.

We can verify: $A\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - x_2 \\ 2x_1 \\ x_1 + 3x_2 \end{bmatrix}$. Since A has a pivot in every row (check by row reducing), T is onto $\mathbb{R}^3$? No: A is 3×2, which has at most 2 pivots, so it cannot have a pivot in all 3 rows. Therefore T is *not* onto. However, A has a pivot in every column, so T *is* one-to-one. □

Fact 1.48 (Composition of Linear Transformations)
If $T_1 : \mathbb{R}^n \to \mathbb{R}^m$ has standard matrix A and $T_2 : \mathbb{R}^m \to \mathbb{R}^p$ has standard matrix B, then the composition $T_2 \circ T_1 : \mathbb{R}^n \to \mathbb{R}^p$ has standard matrix BA (note the order).

2 Matrix Algebra

1 Matrix Operations

Definition 2.1 (Matrix)
An $m \times n$ **matrix** is a rectangular array of numbers with m rows and n columns. The entry in the ith row and jth column is denoted a_{ij}.

Definition 2.2 (Matrix Addition and Scalar Multiplication)
If A and B are $m \times n$ matrices and c is a scalar:

- $(A+B)_{ij} = a_{ij} + b_{ij}$ (add entry by entry).
- $(cA)_{ij} = ca_{ij}$ (scale each entry).

Definition 2.3 (Matrix Multiplication)
If A is $m \times n$ and B is $n \times p$, then the **product** AB is the $m \times p$ matrix whose (i,j)-entry is

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

Equivalently, the jth column of AB is $A\mathbf{b}_j$, where $\mathbf{b}_j$ is the jth column of B.

Theorem 2.4 (Properties of Matrix Multiplication)
Let A be $m \times n$, and let B and C have sizes such that the products below are defined. Then:

1. $A(BC) = (AB)C$ (associativity)
2. $A(B+C) = AB + AC$ (left distributivity)
3. $(B+C)A = BA + CA$ (right distributivity)
4. $r(AB) = (rA)B = A(rB)$ for any scalar r
5. $I_mA = A = AI_n$ (identity)

Warning: In general, $AB \neq BA$. Matrix multiplication is *not* commutative.

Example 2.5
Compute AB where $A = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{bmatrix}$.

Solution. A is 2×2 and B is 2×3, so AB is 2×3. We compute each entry using the row-column rule:

$$AB = \begin{bmatrix} 2(4)+3(1) & 2(3)+3(-2) & 2(6)+3(3) \\ 1(4)+(-5)(1) & 1(3)+(-5)(-2) & 1(6)+(-5)(3) \end{bmatrix} = \begin{bmatrix} 11 & 0 & 21 \\ -1 & 13 & -9 \end{bmatrix}.$$

Note that BA is not defined since B has 3 columns but A has only 2 rows. Even when both AB and BA are defined (e.g., both square), in general $AB \neq BA$. □

Example 2.6 (Non-commutativity)

Let $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$. Then

$$AB = \begin{bmatrix} 7 & 2 \\ 3 & 1 \end{bmatrix}, \quad BA = \begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix}.$$

Clearly $AB \neq BA$.

Definition 2.7 (Powers of a Matrix)

If A is an $n \times n$ matrix and k is a positive integer, then $A^k = \underbrace{A \cdot A \cdots A}_{k \text{ times}}$. We define $A^0 = I_n$.

Definition 2.8 (Transpose)

The **transpose** of an $m \times n$ matrix A, denoted A^T, is the $n \times m$ matrix whose (i, j)-entry is a_{ji}. That is, the rows and columns of A are interchanged.

Theorem 2.9 (Properties of Transpose)

1. $(A^T)^T = A$
2. $(A+B)^T = A^T + B^T$
3. $(rA)^T = rA^T$
4. $(AB)^T = B^T A^T$ (note the reversal of order)

2 The Inverse of a Matrix

Definition 2.10 (Invertible Matrix)

An $n \times n$ matrix A is **invertible** if there exists an $n \times n$ matrix C such that

$$AC = I_n \quad \text{and} \quad CA = I_n.$$

The matrix C is unique and is called the **inverse** of A, denoted A^{-1}. A matrix that is not invertible is called **singular**.

Theorem 2.11 (2 × 2 Inverse)

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. The quantity $ad - bc$ is the **determinant** of A.

Theorem 2.12 (Properties of Inverses)

If A and B are $n \times n$ invertible matrices:

1. $(A^{-1})^{-1} = A$
2. $(AB)^{-1} = B^{-1}A^{-1}$ (note the reversal)
3. $(A^T)^{-1} = (A^{-1})^T$

Fact 2.13 (Algorithm for Finding A^{-1})

Row reduce the augmented matrix $[A \mid I_n]$. If A is row equivalent to I_n, then $[A \mid I_n] \sim [I_n \mid A^{-1}]$. If A cannot be row reduced to I_n, then A is not invertible.

Example 2.14

Find the inverse of $A = \begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix}$.

Solution. Using the 2×2 formula: $\det(A) = 1 \cdot 7 - 2 \cdot 3 = 1 \neq 0$, so

$$A^{-1} = \frac{1}{1} \begin{bmatrix} 7 & -2 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} 7 & -2 \\ -3 & 1 \end{bmatrix}.$$

Verification: $AA^{-1} = \begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} 7 & -2 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$. ✓ □

Example 2.15 (Finding the Inverse via Row Reduction)

Find the inverse of $A = \begin{bmatrix} 1 & 0 & 2 \\ 2 & -1 & 3 \\ 4 & 1 & 8 \end{bmatrix}$.

Solution. Augment with I_3 and row reduce:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 2 & -1 & 3 & 0 & 1 & 0 \\ 4 & 1 & 8 & 0 & 0 & 1 \end{array}\right] \xrightarrow[R_3 \leftarrow R_3 - 4R_1]{R_2 \leftarrow R_2 - 2R_1} \left[\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & -1 & -1 & -2 & 1 & 0 \\ 0 & 1 & 0 & -4 & 0 & 1 \end{array}\right].$$

Apply $R_3 \leftarrow R_3 + R_2$:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & -1 & -1 & -2 & 1 & 0 \\ 0 & 0 & -1 & -6 & 1 & 1 \end{array}\right].$$

Scale $R_2 \leftarrow -R_2$ and $R_3 \leftarrow -R_3$:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & 1 & 1 & 2 & -1 & 0 \\ 0 & 0 & 1 & 6 & -1 & -1 \end{array}\right].$$

Back-substitute: $R_2 \leftarrow R_2 - R_3$ and $R_1 \leftarrow R_1 - 2R_3$:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -11 & 2 & 2 \\ 0 & 1 & 0 & -4 & 0 & 1 \\ 0 & 0 & 1 & 6 & -1 & -1 \end{array}\right].$$

Therefore $A^{-1} = \begin{bmatrix} -11 & 2 & 2 \\ -4 & 0 & 1 \\ 6 & -1 & -1 \end{bmatrix}$. □

Fact 2.16 (Using A^{-1} to Solve Systems)

If A is invertible, the unique solution of $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = A^{-1}\mathbf{b}$. This is theoretically elegant but computationally expensive for large systems. In practice, row reduction is more efficient.

Definition 2.17 (Elementary Matrix)

An **elementary matrix** is a matrix obtained by performing a single elementary row operation on the identity matrix. If E is the elementary matrix for a row operation, then EA is the result of applying that row operation to A.

Fact 2.18 (Elementary Matrices are Invertible)

Every elementary matrix is invertible, and its inverse is the elementary matrix corresponding to the reverse row operation. For example, if E corresponds to $R_2 \leftarrow R_2 + 3R_1$, then E^{-1} corresponds to $R_2 \leftarrow R_2 - 3R_1$.

3 The Invertible Matrix Theorem

Theorem 2.19 (The Invertible Matrix Theorem)

Let A be an $n \times n$ matrix. The following statements are equivalent (i.e., they are all true or all false):

1. A is an invertible matrix.
2. A is row equivalent to the $n \times n$ identity matrix I_n.
3. A has n pivot positions.
4. The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
5. The columns of A are linearly independent.
6. The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.
7. The equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b} \in \mathbb{R}^n$.
8. The columns of A span $\mathbb{R}^n$.
9. The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is onto.
10. There is an $n \times n$ matrix C such that $CA = I_n$.
11. There is an $n \times n$ matrix D such that $AD = I_n$.
12. A^T is an invertible matrix.
13. $\det(A) \neq 0$.

This is one of the most important theorems in linear algebra. It connects many different properties of a square matrix into a single equivalence. To show that a matrix A is invertible, it suffices to establish *any one* of the above conditions. Conversely, if any one condition fails, they all fail.

Fact 2.20 (How to Use the IMT)

In practice, the Invertible Matrix Theorem is used in two main ways:

1. **Checking invertibility**: To determine whether a given $n \times n$ matrix is invertible, check the easiest condition. Often this means computing $\det(A)$ or row reducing to see if there are n pivots.
2. **Deducing properties**: If you already know A is invertible, you can immediately conclude that $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$, the columns are linearly independent, the columns span $\mathbb{R}^n$, and so on.

Example 2.21

Determine whether $A = \begin{bmatrix} 1 & 0 & -2 \\ 3 & 1 & -2 \\ -5 & -1 & 9 \end{bmatrix}$ is invertible.

Solution. Row reduce:

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

There are 3 pivots in a 3×3 matrix, so by the Invertible Matrix Theorem, A is invertible. This means the columns of A are linearly independent, the columns span $\mathbb{R}^3$, $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$, $\det(A) \neq 0$, and so on. □

4 Coordinates, Dimension, and Rank

Definition 2.22 (Column Space and Row Space)

The **column space** of a matrix A, denoted $\operatorname{Col} A$, is the set of all linear combinations of the columns of A:

$$\operatorname{Col} A = \{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\}.$$

The **row space** of A, denoted $\operatorname{Row} A$, is the column space of A^T.

Definition 2.23 (Null Space)

The **null space** of a matrix A, denoted $\operatorname{Nul} A$, is the set of all solutions to $A\mathbf{x} = \mathbf{0}$:

$$\operatorname{Nul} A = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}.$$

Definition 2.24 (Rank)

The **rank** of a matrix A, denoted $\operatorname{rank} A$, is the dimension of the column space of A (equivalently, the number of pivot columns in A).

Theorem 2.25 (Rank Theorem)

If a matrix A has n columns, then

$$\operatorname{rank} A + \dim(\operatorname{Nul} A) = n.$$

That is, the number of pivot columns plus the number of free variables equals the total number of columns.

Example 2.26

If A is a 5×7 matrix with $\operatorname{rank} A = 3$, find $\dim(\operatorname{Nul} A)$ and determine whether $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b} \in \mathbb{R}^5$.

Solution. By the Rank Theorem, $\dim(\operatorname{Nul} A) = n - \operatorname{rank} A = 7 - 3 = 4$. Since $\operatorname{rank} A = 3 < 5 = m$, the columns of A do not span all of $\mathbb{R}^5$ (there are only 3 pivots, so at least 2 rows lack a pivot). Therefore $A\mathbf{x} = \mathbf{b}$ is *not* consistent for every $\mathbf{b}$. □

Example 2.27 (Finding Bases for Col, Nul, and Row)

Find bases for Col A, Nul A, and Row A where

$$A = \begin{bmatrix} 1 & 3 & 5 & 2 \\ 2 & 6 & 10 & 5 \\ 1 & 3 & 5 & 3 \end{bmatrix}.$$

Solution. Row reduce A:

$$A \sim \begin{bmatrix} 1 & 3 & 5 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 5 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The pivots are in columns 1 and 4, so rank $A = 2$.

Basis for Col A: Take the pivot columns of the *original* A (not the reduced form):

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix} \right\}.$$

Basis for Row A: Take the nonzero rows of the echelon form:

$$\left\{ \begin{bmatrix} 1 & 3 & 5 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \right\}.$$

Basis for Nul A: Solve $A\mathbf{x} = \mathbf{0}$ from the RREF. The free variables are x_2 and x_3. From the RREF, $x_1 = -3x_2 - 5x_3$ and $x_4 = 0$. So:

$$\mathbf{x} = x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -5 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

A basis for Nul A is $\left\{ \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -5 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\}$. Note: $\operatorname{rank} A + \dim(\operatorname{Nul} A) = 2 + 2 = 4 = n$. ✓ □

3 Determinants

1 Introduction to Determinants

Definition 3.1 (Determinant)
The **determinant** of a square matrix A, denoted $\det(A)$ or $|A|$, is a scalar that encodes certain properties of the linear transformation described by A.

For a 1×1 matrix, $\det([a]) = a$. For a 2×2 matrix:

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$$

Definition 3.2 (Cofactor Expansion)
For an $n \times n$ matrix A ($n \geq 2$), let A_{ij} denote the $(n-1) \times (n-1)$ submatrix obtained by deleting the ith row and jth column of A. The **(i, j)-cofactor** of A is

$$C_{ij} = (-1)^{i+j} \det(A_{ij}).$$

The determinant can be computed by **cofactor expansion along any row or column**. Expanding along row i:

$$\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} = \sum_{j=1}^{n} a_{ij}C_{ij}.$$

Example 3.3
Compute $\det(A)$ for the matrix

$$A = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}.$$

Solution. Expand along row 3 (it has two zeros, making the computation easier):

$$\begin{aligned} \det(A) &= 0 \cdot C_{31} + (-2) \cdot C_{32} + 0 \cdot C_{33} \\ &= (-2) \cdot (-1)^{3+2} \det \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix} \\ &= (-2)(-1)(1 \cdot (-1) - 0 \cdot 2) \\ &= (-2)(-1)(-1) = -2. \end{aligned}$$

□

Theorem 3.4
If A is a triangular matrix (upper or lower triangular), then $\det(A)$ is the product of the diagonal entries:

$$\det(A) = a_{11}a_{22}\cdots a_{nn}.$$

Proof. For a triangular matrix, every cofactor expansion involves a triangular submatrix. Expanding along the column (or row) where only the diagonal entry is nonzero, the determinant reduces to the product of diagonal entries by induction on n. □

Example 3.5 (Determinant of a 3×3 Matrix via Cofactor Expansion along Column 1)
Compute $\det(B)$ where $B = \begin{bmatrix} 3 & 0 & -1 \\ 2 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}$.

Solution. Expanding along column 2 (which has two zeros):

$$\begin{aligned} \det(B) &= 0 \cdot C_{12} + 3 \cdot C_{22} + 0 \cdot C_{32} \\ &= 3 \cdot (-1)^{2+2} \det \begin{bmatrix} 3 & -1 \\ 1 & 2 \end{bmatrix} \\ &= 3 \cdot 1 \cdot (6 - (-1)) = 3 \cdot 7 = 21. \end{aligned}$$

□

Fact 3.6 (Choosing the Best Row or Column)
When computing a determinant by cofactor expansion, always expand along the row or column with the most zeros. Each zero entry eliminates a cofactor computation entirely.

Fact 3.7 (Computing Determinants via Row Reduction)
An alternative to cofactor expansion (especially for large matrices) is to row reduce to echelon form, keeping track of how row operations change the determinant:

1. Start with $\det(A)$.
2. Each row replacement ($R_i \leftarrow R_i + cR_j$) does not change the determinant.
3. Each row swap multiplies the determinant by -1.
4. Scaling R_i by k multiplies the determinant by k.

Once in echelon form, the determinant is the product of the diagonal entries (times $(-1)^s$ for s row swaps and divided by any scaling factors).

2 Properties of Determinants

Theorem 3.8 (Row Operations and Determinants)
Let A be a square matrix.

1. If a multiple of one row is added to another row, the determinant is unchanged.
2. If two rows are interchanged, the determinant changes sign.
3. If one row is multiplied by a scalar k, the determinant is multiplied by k.

Theorem 3.9 (Multiplicative Property)
If A and B are $n \times n$ matrices, then

$$\det(AB) = \det(A) \cdot \det(B).$$

Theorem 3.10
$\det(A^T) = \det(A)$.

Theorem 3.11
A square matrix A is invertible if and only if $\det(A) \neq 0$.

Theorem 3.12
If A is invertible, then $\det(A^{-1}) = \frac{1}{\det(A)}$.

Definition 3.13 (Cramer's Rule)
Let A be an invertible $n \times n$ matrix. For any $\mathbf{b} \in \mathbb{R}^n$, the unique solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ has entries

$$x_i = \frac{\det(A_i(\mathbf{b}))}{\det(A)}, \quad i = 1, 2, \ldots, n,$$

where $A_i(\mathbf{b})$ is the matrix obtained by replacing the ith column of A with $\mathbf{b}$.

Example 3.14
Use Cramer's Rule to solve: $3x_1 - 2x_2 = 6$, $-5x_1 + 4x_2 = 8$.

Solution. $A = \begin{bmatrix} 3 & -2 \\ -5 & 4 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 6 \\ 8 \end{bmatrix}$, $\det(A) = 12 - 10 = 2$.

$$x_1 = \frac{\det \begin{bmatrix} 6 & -2 \\ 8 & 4 \end{bmatrix}}{2} = \frac{24 + 16}{2} = 20, \quad x_2 = \frac{\det \begin{bmatrix} 3 & 6 \\ -5 & 8 \end{bmatrix}}{2} = \frac{24 + 30}{2} = 27.$$

□

Fact 3.15 (Geometric Interpretation)
If A is a 2×2 matrix, then $|\det(A)|$ gives the area of the parallelogram determined by the columns of A. For a 3×3 matrix, $|\det(A)|$ gives the volume of the parallelepiped determined by the columns of A. More generally, if $T : \mathbb{R}^n \to \mathbb{R}^n$ is the linear transformation with matrix A, then T multiplies all volumes by the factor $|\det(A)|$.

Example 3.16
Find the area of the parallelogram with vertices at $(0, 0)$, $(3, 1)$, $(1, 4)$, and $(4, 5)$.

Solution. The two edge vectors from the origin are $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$. The area is

$$\left| \det \begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix} \right| = |12 - 1| = 11.$$

□

Definition 3.17 (Adjugate Matrix)
The **adjugate** (or classical adjoint) of an $n \times n$ matrix A is the $n \times n$ matrix $\mathrm{adj}(A)$ whose (i, j)-entry is the (j, i)-cofactor C_{ji}. It satisfies

$$A \cdot \mathrm{adj}(A) = \det(A) \cdot I_n.$$

If A is invertible, then $A^{-1} = \frac{1}{\det(A)} \mathrm{adj}(A)$. (This formula is mainly of theoretical interest; row reduction is more efficient for computing inverses.)

4 Vector Spaces

1 Vector Spaces and Subspaces

Definition 4.1 (Vector Space)
A **vector space** is a nonempty set V of objects, called **vectors**, on which two operations are defined (addition and scalar multiplication), subject to the following ten axioms. For all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all scalars c, d:

1. $\mathbf{u} + \mathbf{v} \in V$ (closure under addition)
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (commutativity)
3. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ (associativity)
4. There exists a zero vector $\mathbf{0}$ such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$ (additive identity)
5. For each $\mathbf{u}$, there exists $-\mathbf{u}$ with $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$ (additive inverse)
6. $c\mathbf{u} \in V$ (closure under scalar multiplication)
7. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$ (distributivity)
8. $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$ (distributivity)
9. $c(d\mathbf{u}) = (cd)\mathbf{u}$ (associativity of scalar mult.)
10. $1\mathbf{u} = \mathbf{u}$ (multiplicative identity)

Key examples of vector spaces:

- $\mathbb{R}^n$: the set of all column vectors with n real entries.
- $\mathcal{P}_n$: the set of all polynomials of degree at most n, with the usual polynomial addition and scalar multiplication.
- $\mathcal{P}$: the set of all polynomials (of any degree).
- $C[a, b]$: the set of all continuous real-valued functions on the interval $[a, b]$.
- $M_{m\times n}$: the set of all $m \times n$ real matrices.
- The zero vector space $\{\mathbf{0}\}$ containing only the zero vector.

Definition 4.2 (Subspace)
A **subspace** of a vector space V is a subset H of V that satisfies:

1. The zero vector of V is in H.
2. H is closed under vector addition: if $\mathbf{u}, \mathbf{v} \in H$, then $\mathbf{u} + \mathbf{v} \in H$.
3. H is closed under scalar multiplication: if $\mathbf{u} \in H$ and c is a scalar, then $c\mathbf{u} \in H$.

Example 4.3

Show that $H = \left\{ \begin{bmatrix} s \\ 2s \end{bmatrix} : s \in \mathbb{R} \right\}$ is a subspace of $\mathbb{R}^2$.

Proof. **(1)** Setting $s = 0$ gives $\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in H$. **(2)** If $\begin{bmatrix} s \\ 2s \end{bmatrix}, \begin{bmatrix} t \\ 2t \end{bmatrix} \in H$, then $\begin{bmatrix} s+t \\ 2(s+t) \end{bmatrix} \in H$. **(3)** If $\begin{bmatrix} s \\ 2s \end{bmatrix} \in H$ and $c \in \mathbb{R}$, then $c \begin{bmatrix} s \\ 2s \end{bmatrix} = \begin{bmatrix} cs \\ 2(cs) \end{bmatrix} \in H$. Thus H is a subspace. Geometrically, H is the line $y = 2x$ through the origin. □

Example 4.4 (A Subset That Is Not a Subspace)

Let $S = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2 : x_1 x_2 \geq 0 \right\}$ (the set of points in the first and third quadrants, including the axes). Then S is *not* a subspace.

Proof. The set S contains $\mathbf{u} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ (since $1 \cdot 0 = 0 \geq 0$) and $\mathbf{v} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$ (since $0 \cdot (-1) = 0 \geq 0$). But $\mathbf{u} + \mathbf{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, and $1 \cdot (-1) = -1 < 0$, so $\mathbf{u} + \mathbf{v} \notin S$. Since S is not closed under addition, it is not a subspace. □

Fact 4.5 (Common Subspaces of $\mathbb{R}^n$)

Every subspace of $\mathbb{R}^n$ is one of the following:

- $\{\mathbf{0}\}$ (the zero subspace, dimension 0)
- A line through the origin (dimension 1)
- A plane through the origin (dimension 2)
- More generally, a k-dimensional "flat" through the origin ($0 \leq k \leq n$)
- $\mathbb{R}^n$ itself (dimension n)

Note that a line or plane that does *not* pass through the origin is not a subspace.

Theorem 4.6 (Span Is a Subspace)

If $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are vectors in a vector space V, then $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a subspace of V. In fact, it is the *smallest* subspace containing $\mathbf{v}_1, \ldots, \mathbf{v}_p$.

2 Null Space, Column Space, and Linear Transformations

Theorem 4.7
The null space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^n$. The column space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^m$.

Definition 4.8 (Kernel and Range)
For a linear transformation $T : V \to W$:

- The **kernel** (or null space) of T is $\ker(T) = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}\}$.
- The **range** (or image) of T is $\text{Range}(T) = \{T(\mathbf{v}) : \mathbf{v} \in V\}$.

The kernel is a subspace of V, and the range is a subspace of W.

Fact 4.9 (Finding Bases)

- A basis for $\text{Col}\, A$ is given by the pivot columns of A (the original columns, not the RREF columns).
- A basis for $\text{Nul}\, A$ is found by solving $A\mathbf{x} = \mathbf{0}$, writing the solution in parametric vector form, and taking the direction vectors.
- A basis for $\text{Row}\, A$ is given by the nonzero rows of the echelon form of A.

Warning: When finding a basis for $\text{Col}\, A$, you must use the pivot columns from the *original* matrix A, not from the echelon form. Row operations preserve the row space and the null space, but they change the column space.

Theorem 4.10 (Rank and the Invertible Matrix Theorem, Extended)
Let A be an $n \times n$ matrix. The following are equivalent to the statements in the Invertible Matrix Theorem:

1. $\text{Col}\, A = \mathbb{R}^n$
2. $\text{Nul}\, A = \{\mathbf{0}\}$
3. $\text{rank}\, A = n$
4. $\dim(\text{Nul}\, A) = 0$

3 Linearly Independent Sets; Bases

Definition 4.11 (Basis)
A **basis** for a vector space V is a linearly independent set that spans V.

Theorem 4.12 (Basis Theorem)

If a vector space V has a basis $\mathcal{B}$ with n vectors, then every basis of V has exactly n vectors. This number n is called the **dimension** of V, denoted $\dim V$.

Example 4.13

The standard basis for $\mathbb{R}^n$ is $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$, so $\dim \mathbb{R}^n = n$.

The standard basis for $\mathcal{P}_2$ (polynomials of degree ≤ 2) is $\{1, t, t^2\}$, so $\dim \mathcal{P}_2 = 3$.

Theorem 4.14

Let H be a subspace of a finite-dimensional vector space V. Then $\dim H \leq \dim V$, and $\dim H = \dim V$ only if $H = V$.

Definition 4.15 (Coordinate Vector)

Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis for a vector space V. For each $\mathbf{x} \in V$, the **coordinates of x relative to $\mathcal{B}$** are the unique weights $c_1, \ldots, c_n$ such that $\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_n\mathbf{b}_n$. The **coordinate vector** of $\mathbf{x}$ relative to $\mathcal{B}$ is

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

Example 4.16

Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ and $\mathbf{x} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$. Find $[\mathbf{x}]_{\mathcal{B}}$.

Solution. We need c_1, c_2 such that $c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$. This gives the system $c_1 + c_2 = 5$ and $c_2 = 3$. So $c_2 = 3$ and $c_1 = 2$.

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

In the $\mathcal{B}$-coordinate system, $\mathbf{x}$ is represented as $(2, 3)$ because $\mathbf{x} = 2\mathbf{b}_1 + 3\mathbf{b}_2$. □

Fact 4.17 (Coordinate Mapping)

The coordinate mapping $\mathbf{x} \mapsto [\mathbf{x}]_{\mathcal{B}}$ is a one-to-one and onto linear transformation from V to $\mathbb{R}^n$ (where $n = \dim V$). This is an **isomorphism**: it preserves all the algebraic structure. This is why abstract vector spaces of dimension n "behave like" $\mathbb{R}^n$.

4 The Matrix of a Linear Transformation

Definition 4.18 (Matrix Relative to Bases)
Let $T : V \to W$ be a linear transformation, $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ a basis for V, and $\mathcal{C}$ a basis for W. The **matrix for T relative to $\mathcal{B}$ and $\mathcal{C}$** is the $m \times n$ matrix M whose jth column is $[T(\mathbf{b}_j)]_{\mathcal{C}}$:

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

This matrix satisfies $[T(\mathbf{x})]_{\mathcal{C}} = M[\mathbf{x}]_{\mathcal{B}}$ for all $\mathbf{x} \in V$.

When $V = W$ and $\mathcal{B} = \mathcal{C}$, the matrix M is called the **$\mathcal{B}$-matrix for T**, often denoted $[T]_{\mathcal{B}}$.

5 Change of Basis / Composition of Linear Transformations

Definition 4.19 (Change-of-Coordinates Matrix)
Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be bases for a vector space V. The **change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$** is the $n \times n$ matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ satisfying

$$[\mathbf{x}]_{\mathcal{C}} = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}[\mathbf{x}]_{\mathcal{B}} \quad \text{for all } \mathbf{x} \in V.$$

Its columns are $[\mathbf{b}_1]_{\mathcal{C}}, [\mathbf{b}_2]_{\mathcal{C}}, \ldots, [\mathbf{b}_n]_{\mathcal{C}}$.

Theorem 4.20 (Similar Matrices)
If $A = [T]_{\mathcal{B}}$ is the $\mathcal{B}$-matrix for T and P is the change-of-coordinates matrix from $\mathcal{C}$ to $\mathcal{B}$, then the $\mathcal{C}$-matrix for T is

$$[T]_{\mathcal{C}} = P^{-1}AP.$$

The matrices A and $P^{-1}AP$ are called **similar**. Similar matrices represent the same linear transformation with respect to different bases.

Fact 4.21 (Properties of Similar Matrices)
If A and B are similar matrices (i.e., $B = P^{-1}AP$ for some invertible P), then they share many properties:

- Same determinant: $\det(B) = \det(A)$
- Same trace: $\text{tr}(B) = \text{tr}(A)$
- Same eigenvalues (with the same algebraic and geometric multiplicities)
- Same characteristic polynomial
- Same rank

However, similar matrices generally have different eigenvectors.

Example 4.22

Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$ and $\mathcal{E} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ (the standard basis). Find the change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{E}$.

Solution. The change-of-coordinates matrix $\underset{\mathcal{E} \leftarrow \mathcal{B}}{P}$ has columns $[\mathbf{b}_1]_{\mathcal{E}}$ and $[\mathbf{b}_2]_{\mathcal{E}}$. Since $\mathcal{E}$ is the standard basis, the coordinate vectors are just the vectors themselves:

$$\underset{\mathcal{E} \leftarrow \mathcal{B}}{P} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

This satisfies $\mathbf{x} = P[\mathbf{x}]_{\mathcal{B}}$. The reverse change of coordinates is $[\mathbf{x}]_{\mathcal{B}} = P^{-1}\mathbf{x}$, where $P^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$. □

6 Applications: Markov Chains

Definition 4.23 (Stochastic Matrix)

A **stochastic matrix** is a square matrix with nonneg ative entries in which each column sums to 1. It represents the transition probabilities of a **Markov chain**.

Definition 4.24 (Steady-State Vector)

A **steady-state vector** (or **equilibrium vector**) $\mathbf{q}$ for a stochastic matrix P is a probability vector (entries nonnegative and summing to 1) satisfying

$$P\mathbf{q} = \mathbf{q}.$$

Equivalently, $\mathbf{q}$ is an eigenvector of P for the eigenvalue $\lambda = 1$, normalized so its entries sum to 1.

Example 4.25

Find the steady-state vector for $P = \begin{bmatrix} 0.7 & 0.1 \\ 0.3 & 0.9 \end{bmatrix}$.

Solution. Solve $(P - I)\mathbf{q} = \mathbf{0}$:

$$\begin{bmatrix} -0.3 & 0.1 \\ 0.3 & -0.1 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This gives $-0.3q_1 + 0.1q_2 = 0$, so $q_2 = 3q_1$. With the constraint $q_1 + q_2 = 1$: $q_1 + 3q_1 = 1$, so $q_1 = \frac{1}{4}$, $q_2 = \frac{3}{4}$. The steady-state vector is $\mathbf{q} = \begin{bmatrix} 1/4 \\ 3/4 \end{bmatrix}$. □

5 Eigenvalues and Eigenvectors

1 Eigenvectors and Eigenspaces

Definition 5.1 (Eigenvector and Eigenvalue)
Let A be an $n \times n$ matrix. A nonzero vector $\mathbf{v} \in \mathbb{R}^n$ is called an **eigenvector** of A if there exists a scalar λ such that

$$A\mathbf{v} = \lambda\mathbf{v}.$$

The scalar λ is called the **eigenvalue** corresponding to the eigenvector $\mathbf{v}$.

Rearranging: $A\mathbf{v} = \lambda\mathbf{v}$ is equivalent to $(A - \lambda I)\mathbf{v} = \mathbf{0}$. So λ is an eigenvalue of A if and only if $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has a nontrivial solution.

Definition 5.2 (Eigenspace)
The **eigenspace** of A corresponding to an eigenvalue λ is the null space of $A - \lambda I$:

$$E_\lambda = \text{Nul}(A - \lambda I) = \{\mathbf{x} : (A - \lambda I)\mathbf{x} = \mathbf{0}\}.$$

It is a subspace of $\mathbb{R}^n$ consisting of the zero vector and all eigenvectors corresponding to λ.

Theorem 5.3
The eigenvalues of a triangular matrix are the entries on its main diagonal.

Proof. For simplicity, consider the 3×3 case. If A is upper triangular, then $A - \lambda I$ has the form

$$\begin{bmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ 0 & a_{22} - \lambda & a_{23} \\ 0 & 0 & a_{33} - \lambda \end{bmatrix}.$$

The scalar λ is an eigenvalue if and only if $\det(A - \lambda I) = 0$. Since $A - \lambda I$ is triangular, its determinant is the product of diagonal entries:

$$\det(A - \lambda I) = (a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda) = 0.$$

This equals zero precisely when $\lambda = a_{11}$, $\lambda = a_{22}$, or $\lambda = a_{33}$. □

Lemma 5.4
Let A be an $n \times n$ matrix. Then A is invertible if and only if 0 is not an eigenvalue of A.

Proof. 0 is an eigenvalue $\iff$ $(A - 0 \cdot I)\mathbf{x} = \mathbf{0}$ has a nontrivial solution $\iff$ $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution $\iff$ A is not invertible. Taking the contrapositive gives the result. □

Theorem 5.5
If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are eigenvectors of a matrix A corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$, then the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is linearly independent.

2 The Characteristic Equation

Definition 5.6 (Characteristic Polynomial)
The **characteristic polynomial** of an $n \times n$ matrix A is

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

The equation $\det(A - \lambda I) = 0$ is called the **characteristic equation** of A.

The eigenvalues of A are precisely the roots of the characteristic polynomial.

Example 5.7
Find the eigenvalues and eigenvectors of

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

Solution. The characteristic polynomial is

$$\det(A - \lambda I) = \det \begin{bmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{bmatrix} = (2-\lambda)(-6-\lambda) - 9 = \lambda^2 + 4\lambda - 21.$$

Factoring: $\lambda^2 + 4\lambda - 21 = (\lambda + 7)(\lambda - 3) = 0$, so $\lambda_1 = -7$ and $\lambda_2 = 3$.

Eigenspace for $\lambda_1 = -7$: Solve $(A + 7I)\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 1/3 \\ 0 & 0 \end{bmatrix}.$$

So $x_1 = -\frac{1}{3}x_2$, with x_2 free. A basis for E_{-7} is $\left\{ \begin{bmatrix} -1 \\ 3 \end{bmatrix} \right\}$.

Eigenspace for $\lambda_2 = 3$: Solve $(A - 3I)\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} -1 & 3 \\ 3 & -9 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 \\ 0 & 0 \end{bmatrix}.$$

So $x_1 = 3x_2$. A basis for E_3 is $\left\{ \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}$. □

Example 5.8 (A 3×3 Eigenvalue Problem)

Find the eigenvalues and eigenvectors of

$$A = \begin{bmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{bmatrix}.$$

Solution. The characteristic polynomial is $\det(A - \lambda I)$. By cofactor expansion (or by row reducing $A - \lambda I$ symbolically):

$$\begin{aligned} \det(A - \lambda I) &= \det \begin{bmatrix} 1-\lambda & -3 & 3 \\ 3 & -5-\lambda & 3 \\ 6 & -6 & 4-\lambda \end{bmatrix} \\ &= -\lambda^3 + 0\lambda^2 + 16\lambda + 16 \\ &= -(\lambda^3 - 16\lambda - 16). \end{aligned}$$

Testing $\lambda = 4$: $64 - 64 - 16 \neq 0$. Testing $\lambda = -2$: $-8 + 32 - 16 = 8 \neq 0$. After careful computation, the characteristic polynomial factors as $-(\lambda - 4)(\lambda + 2)^2$. So the eigenvalues are $\lambda_1 = 4$ (algebraic multiplicity 1) and $\lambda_2 = -2$ (algebraic multiplicity 2).

Eigenspace for $\lambda_1 = 4$: Solve $(A - 4I)\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} -3 & -3 & 3 \\ 3 & -9 & 3 \\ 6 & -6 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}.$$

Free variable $x_3 = t$: $\mathbf{x} = t \begin{bmatrix} 1/2 \\ 1/2 \\ 1 \end{bmatrix}$, or equivalently $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$.

Eigenspace for $\lambda_2 = -2$: Solve $(A + 2I)\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} 3 & -3 & 3 \\ 3 & -3 & 3 \\ 6 & -6 & 6 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Two free variables $x_2 = s$, $x_3 = t$: $\mathbf{x} = s \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$. The eigenspace E_{-2} has dimension 2 (geometric multiplicity equals algebraic multiplicity), and a basis is $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$. □

Definition 5.9 (Algebraic and Geometric Multiplicity)
The **algebraic multiplicity** of an eigenvalue λ is its multiplicity as a root of the characteristic polynomial. The **geometric multiplicity** of λ is $\dim E_\lambda = \dim \mathrm{Nul}(A - \lambda I)$.

Theorem 5.10
For each eigenvalue λ, the geometric multiplicity is at most the algebraic multiplicity:

$$1 \leq \dim E_\lambda \leq \text{algebraic multiplicity of } \lambda.$$

Fact 5.11 (Trace and Determinant)
For an $n \times n$ matrix A with eigenvalues $\lambda_1, \ldots, \lambda_n$ (counted with algebraic multiplicity):

- $\mathrm{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn} = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.
- $\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n$.

3 Diagonalization

Definition 5.12 (Diagonalizable)
An $n \times n$ matrix A is **diagonalizable** if there exists an invertible matrix P and a diagonal matrix D such that

$$A = PDP^{-1}.$$

Equivalently, $D = P^{-1}AP$.

Theorem 5.13 (The Diagonalization Theorem)
An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors. In that case, $P = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n]$ and $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, where $A\mathbf{v}_j = \lambda_j \mathbf{v}_j$.

Theorem 5.14
An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

Theorem 5.15
An $n \times n$ matrix A is diagonalizable if and only if the geometric multiplicity of each eigenvalue equals its algebraic multiplicity.

Fact 5.16 (Powers of a Diagonalizable Matrix)
If $A = PDP^{-1}$, then $A^k = PD^kP^{-1}$, where $D^k = \text{diag}(\lambda_1^k, \dots, \lambda_n^k)$.

Example 5.17
Diagonalize $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$.

Solution. Characteristic polynomial: $\det(A-\lambda I) = (7-\lambda)(1-\lambda)+8 = \lambda^2 - 8\lambda + 15 = (\lambda-3)(\lambda-5) = 0$. Eigenvalues: $\lambda_1 = 3$, $\lambda_2 = 5$.

For $\lambda_1 = 3$: $(A-3I)\mathbf{x} = \mathbf{0}$ gives $\begin{bmatrix} 4 & 2 \\ -4 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & 1/2 \\ 0 & 0 \end{bmatrix}$, so $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$.

For $\lambda_2 = 5$: $(A-5I)\mathbf{x} = \mathbf{0}$ gives $\begin{bmatrix} 2 & 2 \\ -4 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, so $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

Therefore $A = PDP^{-1}$ with

$$P = \begin{bmatrix} -1 & -1 \\ 2 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}.$$

To verify: $P^{-1} = \frac{1}{-1+2}\begin{bmatrix} 1 & 1 \\ -2 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -2 & -1 \end{bmatrix}$. Then $PDP^{-1} = \begin{bmatrix} -1 & -1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -2 & -1 \end{bmatrix} = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix} = A$. ✓ □

Example 5.18 (A Matrix That Is Not Diagonalizable)
Show that $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ is not diagonalizable.

Solution. The characteristic polynomial is $(2-\lambda)^2 = 0$, so $\lambda = 2$ is the only eigenvalue with algebraic multiplicity 2.

Find the eigenspace E_2: Solve $(A-2I)\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\mathbf{x} = \mathbf{0} \implies x_2 = 0,\ x_1 \text{ free}.$$

The eigenspace is $E_2 = \text{Span}\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\}$, which has dimension 1. Since the geometric multiplicity (1) is less than the algebraic multiplicity (2), there are not enough linearly independent eigenvectors to form a basis for $\mathbb{R}^2$, so A is not diagonalizable. □

Fact 5.19 (Diagonalization Procedure)

To diagonalize an $n \times n$ matrix A:

1. Find the eigenvalues by solving $\det(A - \lambda I) = 0$.
2. For each eigenvalue λ_k, find a basis for the eigenspace $E_{\lambda_k} = \text{Nul}(A - \lambda_k I)$.
3. Check: if the total number of basis vectors across all eigenspaces equals n, then A is diagonalizable. If not, A is not diagonalizable.
4. Form P by placing the eigenvectors as columns, and D by placing the corresponding eigenvalues on the diagonal (in the same order).

Example 5.20 (Application: Computing A^k)

Using the diagonalization $A = PDP^{-1}$ from the previous example (with $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$), compute A^{10}.

Solution. $A^{10} = PD^{10}P^{-1}$. Since $D = \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}$, we have $D^{10} = \begin{bmatrix} 3^{10} & 0 \\ 0 & 5^{10} \end{bmatrix} = \begin{bmatrix} 59049 & 0 \\ 0 & 9765625 \end{bmatrix}$.

Then $A^{10} = \begin{bmatrix} -1 & -1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 59049 & 0 \\ 0 & 9765625 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -2 & -1 \end{bmatrix}$. Without diagonalization, computing A^{10} directly would require 9 matrix multiplications. Diagonalization reduces this to computing powers of scalars. □

Fact 5.21 (Complex Eigenvalues)

A real $n \times n$ matrix may have complex eigenvalues. They always occur in conjugate pairs: if $\lambda = a + bi$ is an eigenvalue (with $b \neq 0$), then so is $\bar{\lambda} = a - bi$. For example, the rotation matrix $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ (rotation by 90) has eigenvalues $\lambda = \pm i$. A real matrix with complex eigenvalues is not diagonalizable over $\mathbb{R}$, but it can be put into a real block diagonal form.

4 Eigenvalues and Linear Transformations

Theorem 5.22

Let $T : V \to V$ be a linear transformation on a finite-dimensional vector space V, and let $\mathcal{B}$ be any basis for V. Then λ is an eigenvalue of T if and only if λ is an eigenvalue of the matrix $[T]_{\mathcal{B}}$. The eigenvalues of T are independent of the choice of basis.

Proof. Let $A = [T]_{\mathcal{B}}$. Then $T(\mathbf{x}) = \lambda\mathbf{x}$ if and only if $[T(\mathbf{x})]_{\mathcal{B}} = \lambda[\mathbf{x}]_{\mathcal{B}}$, which is $A[\mathbf{x}]_{\mathcal{B}} = \lambda[\mathbf{x}]_{\mathcal{B}}$. Since $\mathbf{x} \mapsto [\mathbf{x}]_{\mathcal{B}}$ is an isomorphism, nontrivial solutions correspond bijectively. If $\mathcal{C}$ is another basis, then $[T]_{\mathcal{C}} = P^{-1}AP$ for some invertible P, and similar matrices have the same characteristic polynomial. □

5 Iteration Method for Approximating Eigenvalues

Fact 5.23 (The Power Method)
The **power method** is an iterative algorithm for finding the dominant eigenvalue (the eigenvalue with the largest absolute value) of a matrix A.

Starting with an initial guess $\mathbf{x}_0$, compute:

$$\mathbf{x}_{k+1} = \frac{A\mathbf{x}_k}{\|A\mathbf{x}_k\|}$$

Under mild conditions (the dominant eigenvalue is unique in absolute value, and $\mathbf{x}_0$ has a component in the direction of the corresponding eigenvector), the sequence $\mathbf{x}_k$ converges to an eigenvector for the dominant eigenvalue, and $\mathbf{x}_k^T A\mathbf{x}_k$ converges to the dominant eigenvalue (the **Rayleigh quotient**).

Fact 5.24 (Inverse Power Method)
To find the eigenvalue of A closest to a target μ, apply the power method to $(A-\mu I)^{-1}$. This converges to the eigenvector for the eigenvalue closest to μ.

6 Orthogonality and Least Squares

1 Inner Product, Length, and Orthogonality

Definition 6.1 (Inner Product (Dot Product))

The **inner product** (or **dot product**) of $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Theorem 6.2 (Properties of the Inner Product)

For $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and scalar c:

1. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
2. $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
3. $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v})$
4. $\mathbf{u} \cdot \mathbf{u} \geq 0$, and $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$

Definition 6.3 (Length (Norm))

The **length** (or **norm**) of $\mathbf{v} \in \mathbb{R}^n$ is

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

A vector of length 1 is called a **unit vector**. To **normalize** $\mathbf{v} \neq \mathbf{0}$, form $\hat{\mathbf{v}} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$.

Definition 6.4 (Distance)

The **distance** between $\mathbf{u}$ and $\mathbf{v}$ is $\text{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$.

Definition 6.5 (Orthogonal Vectors)

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if $\mathbf{u} \cdot \mathbf{v} = 0$.

Theorem 6.6 (Pythagorean Theorem)

If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, then $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

Proof. $\|\mathbf{u} + \mathbf{v}\|^2 = (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) = \|\mathbf{u}\|^2 + 2(\mathbf{u} \cdot \mathbf{v}) + \|\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$ since $\mathbf{u} \cdot \mathbf{v} = 0$. □

Theorem 6.7 (Cauchy–Schwarz Inequality)
For any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

$$|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\| \cdot \|\mathbf{v}\|.$$

Equality holds if and only if one vector is a scalar multiple of the other (i.e., $\mathbf{u}$ and $\mathbf{v}$ are parallel).

Definition 6.8 (Angle Between Vectors)
The **angle** θ between two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ is defined by

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \cdot \|\mathbf{v}\|}, \quad 0 \leq \theta \leq \pi.$$

This is well-defined by the Cauchy–Schwarz inequality (the fraction lies in $[-1, 1]$). Two vectors are orthogonal when $\theta = \pi/2$.

Theorem 6.9 (Triangle Inequality)
For any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

Definition 6.10 (Orthogonal Complement)
The **orthogonal complement** of a subspace W of $\mathbb{R}^n$ is

$$W^\perp = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} \cdot \mathbf{w} = 0 \text{ for all } \mathbf{w} \in W\}.$$

Theorem 6.11
Let A be an $m \times n$ matrix. Then:

1. $(\text{Row } A)^\perp = \text{Nul } A$
2. $(\text{Col } A)^\perp = \text{Nul } A^T$

2 Orthogonal Sets

Definition 6.12 (Orthogonal Set)
A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is an **orthogonal set** if each pair of distinct vectors is orthogonal: $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ for $i \neq j$. If, additionally, each vector is a unit vector ($\|\mathbf{u}_i\| = 1$), the set is **orthonormal**.

Theorem 6.13
An orthogonal set of nonzero vectors is linearly independent.

Definition 6.14 (Orthogonal Basis)
An **orthogonal basis** for a subspace W is a basis that is also an orthogonal set.

Theorem 6.15
If $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an orthogonal basis for a subspace W, then for any $\mathbf{y} \in W$,

$$\mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 + \cdots + \frac{\mathbf{y}\cdot\mathbf{u}_p}{\mathbf{u}_p\cdot\mathbf{u}_p}\mathbf{u}_p.$$

Definition 6.16 (Orthogonal Matrix)
An $n \times n$ matrix U is an **orthogonal matrix** if $U^T U = I$, i.e., $U^{-1} = U^T$. Equivalently, the columns of U form an orthonormal set.

Theorem 6.17 (Properties of Orthogonal Matrices)
If U is orthogonal, then:

1. $\|U\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$ (length-preserving)
2. $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}, \mathbf{y}$ (inner-product-preserving)
3. $\det(U) = \pm 1$

3 Orthogonal Projections

Fact 6.18 (Projection onto a Line)
As a special case, the orthogonal projection of $\mathbf{y}$ onto the line $L = \text{Span}\{\mathbf{u}\}$ through a single nonzero vector $\mathbf{u}$ is:

$$\text{proj}_L\,\mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}.$$

The component of $\mathbf{y}$ orthogonal to $\mathbf{u}$ is $\mathbf{y} - \text{proj}_L\,\mathbf{y}$.

Example 6.19

Find the projection of $\mathbf{y} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ onto the line through $\mathbf{u} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$.

Solution.

$$\text{proj}_L\,\mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u} = \frac{12+2}{9+1}\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \frac{14}{10}\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 42/10 \\ 14/10 \end{bmatrix} = \begin{bmatrix} 21/5 \\ 7/5 \end{bmatrix}.$$

The orthogonal component is $\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} - \begin{bmatrix} 21/5 \\ 7/5 \end{bmatrix} = \begin{bmatrix} -1/5 \\ 3/5 \end{bmatrix}$.

Verification: $\hat{\mathbf{y}} \cdot (\mathbf{y} - \hat{\mathbf{y}}) = \frac{21}{5} \cdot \frac{-1}{5} + \frac{7}{5} \cdot \frac{3}{5} = \frac{-21+21}{25} = 0.$ ✓ □

Theorem 6.20 (Orthogonal Decomposition Theorem)
Let W be a subspace of $\mathbb{R}^n$. Each $\mathbf{y} \in \mathbb{R}^n$ can be written uniquely as

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z},$$

where $\hat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^\perp$.

Definition 6.21 (Orthogonal Projection)
The **orthogonal projection of $\mathbf{y}$ onto W**, denoted $\operatorname{proj}_W \mathbf{y}$, is the unique vector $\hat{\mathbf{y}} \in W$ such that $\mathbf{y} - \hat{\mathbf{y}} \in W^\perp$. If $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthogonal basis for W:

$$\operatorname{proj}_W \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\mathbf{u}_p.$$

Theorem 6.22 (Best Approximation Theorem)
Let W be a subspace of $\mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^n$, and $\hat{\mathbf{y}} = \operatorname{proj}_W \mathbf{y}$. Then $\hat{\mathbf{y}}$ is the closest point in W to $\mathbf{y}$:

$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\| \quad \text{for all } \mathbf{v} \in W,\ \mathbf{v} \neq \hat{\mathbf{y}}.$$

Example 6.23
Let $\mathbf{y} = \begin{bmatrix} 6 \\ 3 \\ 2 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and $W = \operatorname{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Find $\operatorname{proj}_W \mathbf{y}$.

Solution. Note $\mathbf{u}_1 \cdot \mathbf{u}_2 = 0$, so $\{\mathbf{u}_1, \mathbf{u}_2\}$ is orthogonal.

$$\operatorname{proj}_W \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 = \frac{8}{2}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \frac{3}{1}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 4 \end{bmatrix}.$$

□

4 The Gram–Schmidt Process / Least Squares Problems

Theorem 6.24 (The Gram–Schmidt Process)
Given a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ for a subspace W, define:

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_k &= \mathbf{x}_k - \sum_{j=1}^{k-1} \frac{\mathbf{x}_k \cdot \mathbf{v}_j}{\mathbf{v}_j \cdot \mathbf{v}_j}\mathbf{v}_j.
\end{aligned}$$

Then $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is an orthogonal basis for W. Normalizing each vector gives an orthonormal basis.

Example 6.25

Apply Gram–Schmidt to $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$.

Solution. $\mathbf{v}_1 = \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$. Then:

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}.$$

Clearing fractions, we can write $\mathbf{v}_2 = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ (scalar multiples preserve orthogonality). Check: $\mathbf{v}_1 \cdot \mathbf{v}_2 = -3 + 1 + 1 + 1 = 0$. ✓ □

Fact 6.26 (Intuition for Gram–Schmidt)
The idea behind Gram–Schmidt is simple: at each step, take the next vector $\mathbf{x}_k$ and subtract off its projections onto all the previously constructed orthogonal vectors $\mathbf{v}_1, \dots, \mathbf{v}_{k-1}$. What remains is the component of $\mathbf{x}_k$ that is perpendicular to all of them.

Theorem 6.27 (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 $\operatorname{Col} A$, and R is an $n \times n$ upper triangular invertible matrix with positive diagonal entries.

Proof. Apply Gram–Schmidt to the columns of A to obtain orthogonal vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$. Normalize them to get $\mathbf{q}_1, \dots, \mathbf{q}_n$ where $\mathbf{q}_k = \mathbf{v}_k / \|\mathbf{v}_k\|$. Set $Q = [\mathbf{q}_1 \ \cdots \ \mathbf{q}_n]$. Since Q has orthonormal columns, $Q^T Q = I_n$, and $R = Q^T A$ is upper triangular with positive diagonal entries. □

Fact 6.28 (Using QR for Least Squares)
If $A = QR$, then the normal equation $A^T A\hat{\mathbf{x}} = A^T \mathbf{b}$ simplifies to $R\hat{\mathbf{x}} = Q^T \mathbf{b}$. This is easier to solve because R is upper triangular, so we just use back-substitution. The QR method is also numerically more stable than directly computing $A^T A$.

Definition 6.29 (Least Squares Solution)
When $A\mathbf{x} = \mathbf{b}$ is inconsistent (has no exact solution), a **least-squares solution** $\hat{\mathbf{x}}$ minimizes $\|A\mathbf{x} - \mathbf{b}\|$. Geometrically, $\hat{\mathbf{x}}$ is the value of $\mathbf{x}$ that makes $A\mathbf{x}$ as close as possible to $\mathbf{b}$.

Theorem 6.30 (Normal Equations)
The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ is the solution set of the **normal equations**:

$$A^T A\hat{\mathbf{x}} = A^T \mathbf{b}.$$

If the columns of A are linearly independent, then $A^T A$ is invertible and the unique least-squares solution is

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}.$$

Proof. The projection of $\mathbf{b}$ onto $\operatorname{Col} A$ is $\hat{\mathbf{b}} = A\hat{\mathbf{x}}$, and $\mathbf{b} - A\hat{\mathbf{x}} \perp \operatorname{Col} A$. This means $\mathbf{b} - A\hat{\mathbf{x}} \in (\operatorname{Col} A)^\perp = \operatorname{Nul}(A^T)$. So $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$, which gives $A^T A\hat{\mathbf{x}} = A^T \mathbf{b}$. □

Example 6.31 (Linear Regression)
Find the least-squares line $y = \beta_0 + \beta_1 x$ for the data points $(1, 0)$, $(2, 1)$, $(3, 3)$.

Solution. We want $\beta_0 + \beta_1 x_i \approx y_i$, i.e. $A\boldsymbol{\beta} \approx \mathbf{y}$ where

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}.$$

Compute:

$$A^TA = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}, \quad A^T\mathbf{y} = \begin{bmatrix} 4 \\ 11 \end{bmatrix}.$$

Solve $A^TA\hat{\boldsymbol{\beta}} = A^T\mathbf{y}$:

$$\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 4 \\ 11 \end{bmatrix}.$$

From the first equation: $\beta_0 = \frac{4-6\beta_1}{3}$. Substituting into the second: $6 \cdot \frac{4-6\beta_1}{3} + 14\beta_1 = 11$, which gives $8 - 12\beta_1 + 14\beta_1 = 11$, so $\beta_1 = \frac{3}{2}$ and $\beta_0 = \frac{4-9}{3} = -\frac{5}{3}$.

The least-squares line is $y = -\frac{5}{3} + \frac{3}{2}x$. □

Fact 6.32 (Least Squares for Polynomial Fitting)

The same method works for fitting higher-degree polynomials. To fit $y = \beta_0 + \beta_1 x + \beta_2 x^2$ to data points $(x_1, y_1), \ldots, (x_N, y_N)$, use the design matrix

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_N & x_N^2 \end{bmatrix}$$

and solve the normal equations $A^TA\hat{\boldsymbol{\beta}} = A^T\mathbf{y}$.

Fact 6.33 (Geometric Interpretation of Least Squares)

The least-squares solution finds the point $A\hat{\mathbf{x}}$ in Col A that is closest to $\mathbf{b}$. Equivalently, $A\hat{\mathbf{x}} = \text{proj}_{\text{Col } A}\,\mathbf{b}$. The residual vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to Col A, and $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ is the distance from $\mathbf{b}$ to Col A.

Fact 6.34 (Least-Squares Error)

The **least-squares error** is $\|\mathbf{b} - A\hat{\mathbf{x}}\|$. This is the minimum possible value of $\|A\mathbf{x} - \mathbf{b}\|$ over all $\mathbf{x}$. It measures how well the model fits the data. If the error is zero, then $\mathbf{b} \in \text{Col } A$ and the system $A\mathbf{x} = \mathbf{b}$ has an exact solution.

7 Symmetric Matrices and Quadratic Forms

1 Diagonalization of Symmetric Matrices

Definition 7.1 (Symmetric Matrix)
A matrix A is **symmetric** if $A^T = A$.

Theorem 7.2
If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.

Proof. Let $\lambda_1 \neq \lambda_2$ be eigenvalues with eigenvectors $\mathbf{v}_1, \mathbf{v}_2$. Then:

$$\lambda_1(\mathbf{v}_1 \cdot \mathbf{v}_2) = (\lambda_1 \mathbf{v}_1) \cdot \mathbf{v}_2 = (A\mathbf{v}_1) \cdot \mathbf{v}_2 = (A\mathbf{v}_1)^T \mathbf{v}_2 = \mathbf{v}_1^T A^T \mathbf{v}_2 = \mathbf{v}_1^T A \mathbf{v}_2 = \mathbf{v}_1 \cdot (A\mathbf{v}_2) = \mathbf{v}_1 \cdot (\lambda_2 \mathbf{v}_2) = \lambda_2(\mathbf{v}_1 \cdot \mathbf{v}_2).$$

So $(\lambda_1 - \lambda_2)(\mathbf{v}_1 \cdot \mathbf{v}_2) = 0$. Since $\lambda_1 \neq \lambda_2$, we conclude $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. □

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

1. All eigenvalues of A are real.
2. Eigenvectors from different eigenspaces are orthogonal.
3. A is orthogonally diagonalizable: there exists an orthogonal matrix Q and a diagonal matrix D such that $A = QDQ^T$.
4. The geometric multiplicity of each eigenvalue equals its algebraic multiplicity.

Definition 7.4 (Spectral Decomposition)
If $A = QDQ^T$ where $Q = [\mathbf{q}_1 \ \cdots \ \mathbf{q}_n]$ has orthonormal columns and $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, then the **spectral decomposition** of A is

$$A = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^T + \cdots + \lambda_n \mathbf{q}_n \mathbf{q}_n^T.$$

Each $\mathbf{q}_j \mathbf{q}_j^T$ is the projection matrix onto $\operatorname{Span}\{\mathbf{q}_j\}$.

Fact 7.5 (Procedure for Orthogonal Diagonalization)

To orthogonally diagonalize a symmetric matrix A:

1. Find all eigenvalues of A.
2. For each eigenvalue, find a basis for the corresponding eigenspace.
3. If an eigenspace has dimension ≥ 2, apply Gram–Schmidt within that eigenspace to get an orthogonal basis. (Eigenvectors from *different* eigenspaces are automatically orthogonal.)
4. Normalize all eigenvectors to unit length.
5. Assemble the orthonormal eigenvectors as columns of Q.

Definition 7.6 (Quadratic Form)

A **quadratic form** on $\mathbb{R}^n$ is a function $Q : \mathbb{R}^n \to \mathbb{R}$ defined by $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ for some symmetric matrix A.

Example 7.7

The quadratic form associated with $A = \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix}$ is

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3x_1^2 - 4x_1x_2 + 7x_2^2.$$

Note the coefficient of x_1x_2 is $-2 + (-2) = -4$ (twice the off-diagonal entry). The eigenvalues of A are $\lambda_1 = \frac{10-\sqrt{32}}{2}$ and $\lambda_2 = \frac{10+\sqrt{32}}{2}$, both positive. Therefore A is positive definite, meaning $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$.

Fact 7.8 (Principal Axis Theorem)

If A is symmetric with orthogonal diagonalization $A = QDQ^T$, then the change of variable $\mathbf{x} = Q\mathbf{y}$ transforms the quadratic form $\mathbf{x}^T A\mathbf{x}$ into $\mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$, which has no cross terms. The columns of Q are called the **principal axes** of the quadratic form.

Definition 7.9 (Positive/Negative Definite)

A symmetric matrix A (and its associated quadratic form) is:

- **Positive definite** if $\mathbf{x}^T A\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$ (equivalently, all eigenvalues are positive).
- **Positive semidefinite** if $\mathbf{x}^T A\mathbf{x} \geq 0$ for all $\mathbf{x}$ (eigenvalues ≥ 0).
- **Negative definite** if $\mathbf{x}^T A\mathbf{x} < 0$ for all $\mathbf{x} \neq \mathbf{0}$ (all eigenvalues negative).
- **Indefinite** if $\mathbf{x}^T A\mathbf{x}$ takes both positive and negative values.

Example 7.10

Orthogonally diagonalize $A = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$.

Solution. Characteristic polynomial: $\det(A - \lambda I) = (4-\lambda)(1-\lambda) - 4 = \lambda^2 - 5\lambda = \lambda(\lambda - 5)$. Eigenvalues: $\lambda_1 = 0$, $\lambda_2 = 5$.

For $\lambda_1 = 0$: $A\mathbf{x} = \mathbf{0}$ gives $\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 1/2 \\ 0 & 0 \end{bmatrix}$, so $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$.

For $\lambda_2 = 5$: $(A - 5I)\mathbf{x} = \mathbf{0}$ gives $\begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 \\ 0 & 0 \end{bmatrix}$, so $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

Check orthogonality: $\mathbf{v}_1 \cdot \mathbf{v}_2 = -2 + 2 = 0$. ✓

Normalize: $\mathbf{q}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} -1 \\ 2 \end{bmatrix}$, $\mathbf{q}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

$$A = QDQ^T = \frac{1}{\sqrt{5}} \begin{bmatrix} -1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 5 \end{bmatrix} \frac{1}{\sqrt{5}} \begin{bmatrix} -1 & 2 \\ 2 & 1 \end{bmatrix}^T.$$

Since $\lambda_1 = 0$ and $\lambda_2 = 5 > 0$, the matrix is positive semidefinite. □

2 Singular Value Decomposition (SVD)

Definition 7.11 (Singular Values)

Let A be an $m \times n$ matrix. The **singular values** of A are the square roots of the eigenvalues of $A^T A$:

$$\sigma_i = \sqrt{\lambda_i(A^T A)}, \quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0.$$

Note that $A^T A$ is symmetric and positive semidefinite, so all eigenvalues are nonneg ative.

Theorem 7.12 (Singular Value Decomposition)

Let A be an $m \times n$ matrix with rank r. Then there exist:

- An $m \times m$ orthogonal matrix U,
- An $n \times n$ orthogonal matrix V,
- An $m \times n$ "diagonal" matrix Σ with entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ on the diagonal and zeros elsewhere,

such that $A = U\Sigma V^T$.

Fact 7.13 (Computing the SVD)

1. Compute A^TA and find its eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ and corresponding orthonormal eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Set $V = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_n]$.
2. The singular values are $\sigma_i = \sqrt{\lambda_i}$.
3. For each $\sigma_i > 0$, compute $\mathbf{u}_i = \frac{1}{\sigma_i} A\mathbf{v}_i$. Extend $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ to an orthonormal basis for $\mathbb{R}^m$. Set $U = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_m]$.
4. Form Σ with $\sigma_1, \ldots, \sigma_r$ on the diagonal.

Fact 7.14 (Interpretations of the SVD)

- $\operatorname{rank}(A)$ = number of nonzero singular values.
- The columns of V corresponding to zero singular values form an orthonormal basis for $\operatorname{Nul} A$.
- The columns $\mathbf{u}_1, \ldots, \mathbf{u}_r$ of U form an orthonormal basis for $\operatorname{Col} A$.
- $\|A\| = \sigma_1$ (the operator norm / largest singular value).
- The **condition number** of A is $\kappa(A) = \sigma_1/\sigma_r$ (measures sensitivity to perturbation).

Fact 7.15 (Geometric Interpretation of the SVD)

The SVD says that every linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ can be decomposed into three steps:

1. **Rotate/reflect in** $\mathbb{R}^n$: Apply V^T (orthogonal transformation in the domain).
2. **Scale and embed**: Apply Σ (stretch along coordinate axes by the singular values, and embed into $\mathbb{R}^m$ if $m \neq n$).
3. **Rotate/reflect in** $\mathbb{R}^m$: Apply U (orthogonal transformation in the codomain).

In other words, every matrix, no matter how complicated, is "just a rotation, a scaling, and another rotation."

Fact 7.16 (Pseudoinverse)

The SVD gives a natural way to compute the **pseudoinverse** (or Moore–Penrose inverse) A^+ of any $m \times n$ matrix A: if $A = U\Sigma V^T$, then $A^+ = V\Sigma^+U^T$, where Σ^+ is obtained by taking the reciprocal of each nonzero diagonal entry and then transposing. The pseudoinverse gives the least-squares solution: $\hat{\mathbf{x}} = A^+\mathbf{b}$.

Theorem 7.17 (Eckart–Young Theorem / Low-Rank Approximation)

The best rank-k approximation to A (in the Frobenius or operator norm) is

$$A_k = \sigma_1\mathbf{u}_1\mathbf{v}_1^T + \sigma_2\mathbf{u}_2\mathbf{v}_2^T + \cdots + \sigma_k\mathbf{u}_k\mathbf{v}_k^T = U_k\Sigma_kV_k^T.$$

Example 7.18

Find the SVD of $A = \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}$.

Solution. First, $A^T A = \begin{bmatrix} 4 & 3 \\ 0 & -5 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix} = \begin{bmatrix} 25 & -15 \\ -15 & 25 \end{bmatrix}$.

Eigenvalues of $A^T A$: $\det(A^T A - \lambda I) = (25-\lambda)^2 - 225 = \lambda^2 - 50\lambda + 400 = (\lambda - 10)(\lambda - 40) = 0$. So $\lambda_1 = 40$, $\lambda_2 = 10$, giving $\sigma_1 = \sqrt{40} = 2\sqrt{10}$, $\sigma_2 = \sqrt{10}$.

Eigenvectors of $A^T A$: For $\lambda_1 = 40$: $(A^T A - 40I)\mathbf{v} = \mathbf{0}$ gives $\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. For $\lambda_2 = 10$: $\mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Compute $\mathbf{u}_i = \frac{1}{\sigma_i} A\mathbf{v}_i$: $\mathbf{u}_1 = \frac{1}{2\sqrt{10}} \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{2\sqrt{20}} \begin{bmatrix} 4 \\ 8 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$.

$\mathbf{u}_2 = \frac{1}{\sqrt{10}} \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{20}} \begin{bmatrix} 4 \\ -2 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{bmatrix}$.

The SVD is $A = U\Sigma V^T$ with

$$U = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2\sqrt{10} & 0 \\ 0 & \sqrt{10} \end{bmatrix}, \quad V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

□

3 Principal Component Analysis (PCA) and Eigenfaces

Definition 7.19 (Covariance Matrix)

Given a data matrix X (each column is an observation, each row a variable), the **sample covariance matrix** is

$$S = \frac{1}{N-1} BB^T,$$

where $B = X - \bar{X}$ is the mean-centered data matrix (subtract the mean of each row from every entry in that row) and N is the number of observations.

Definition 7.20 (Principal Components)

The **principal components** of a data set are the eigenvectors of the covariance matrix S, ordered by decreasing eigenvalue. The first principal component is the direction of greatest variance in the data, the second is the direction of greatest variance orthogonal to the first, and so on.

Fact 7.21 (PCA Procedure)

1. Compute the mean of each variable and form the mean-centered data matrix B.
2. Compute the covariance matrix $S = \frac{1}{N-1} BB^T$.
3. Find the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ and corresponding unit eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_p$ of S.
4. The fraction of variance explained by the first k components is $\frac{\lambda_1 + \cdots + \lambda_k}{\lambda_1 + \cdots + \lambda_p}$.
5. To reduce to k dimensions, project: $\mathbf{y}_i = U_k^T(\mathbf{x}_i - \bar{\mathbf{x}})$, where $U_k = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_k]$.

Fact 7.22 (PCA via SVD)

PCA can also be computed directly from the SVD of the mean-centered data matrix B. If $B = U\Sigma V^T$, then the principal components are the columns of U, and the singular values satisfy $\sigma_i = \sqrt{(N-1)\lambda_i}$.

Example 7.23 (PCA with 2D Data)

Suppose we have 4 data points in $\mathbb{R}^2$: $(2, 1)$, $(4, 3)$, $(6, 5)$, $(8, 7)$. Perform PCA.

Solution. **Step 1**: The means are $\bar{x}_1 = 5$, $\bar{x}_2 = 4$. The mean-centered data matrix (observations as columns) is:

$$B = \begin{bmatrix} -3 & -1 & 1 & 3 \\ -3 & -1 & 1 & 3 \end{bmatrix}.$$

Step 2: The covariance matrix is

$$S = \frac{1}{3} BB^T = \frac{1}{3} \begin{bmatrix} 20 & 20 \\ 20 & 20 \end{bmatrix} = \begin{bmatrix} 20/3 & 20/3 \\ 20/3 & 20/3 \end{bmatrix}.$$

Step 3: The eigenvalues are $\lambda_1 = 40/3$ and $\lambda_2 = 0$. The first principal component is $\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ (the direction along the line $y = x$).

Step 4: The fraction of variance explained by the first component is $\frac{40/3}{40/3+0} = 100\%$. This makes sense because all data points lie on the line $x_2 = x_1 - 1$ (after centering, on $x_2 = x_1$), so one dimension captures all the variation. □

Fact 7.24 (Choosing the Number of Components)

In practice, one chooses k so that the first k principal components capture a sufficiently large fraction of the total variance (often 90% or 95%). This is sometimes visualized with a **scree plot**: a plot of λ_i versus i, looking for an "elbow" where the eigenvalues drop off sharply.

Fact 7.25 (Eigenfaces)

In the **eigenfaces** approach to face recognition:

1. Represent each face image as a high-dimensional vector (one entry per pixel).
2. Perform PCA on the set of training face images.
3. The resulting principal components are called **eigenfaces**. They resemble ghostly faces and capture the main modes of variation in the training set.
4. Project each face onto the first k eigenfaces to get a low-dimensional representation.
5. Recognize new faces by comparing their projections to those of known faces (e.g., using nearest neighbor in the reduced space).

This dramatically reduces dimensionality while preserving the most important facial features.