# Linear Algebra with Applications

**Lectures on YouTube:**
**https://www.youtube.com/@mathtalent**

Seongjai Kim

Department of Mathematics and Statistics
Mississippi State University
Mississippi State, MS 39762 USA
Email: skim@math.msstate.edu

Updated: December 2, 2023

*Seongjai Kim*, Professor of Mathematics, Department of Mathematics and Statistics, Mississippi State University, Mississippi State, MS 39762 USA. Email: skim@math.msstate.edu.

# Prologue

This lecture note is organized, following contents in *Linear Algebra and Its Applications, 6th Ed.*, by D. Lay, S. Lay, and J. McDonald [1].

Seongjai Kim
December 2, 2023

**Learning Objectives**

***Real-world problems*** can be approximated as and resolved by ***systems of linear equations***

$$A\mathbf{x} = \mathbf{b}, \qquad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{R}^{m\times n},$$

where one of $\{\mathbf{x}, \mathbf{b}\}$ is the input and the other is the output.

---

**What you would learn, from Linear Algebra**:

1. How to Solve Systems of Linear Equations
   - Programming with Matlab/Octave
2. Matrix Algebra (Matrix Inverse & Factorizations)
3. Determinants
4. Vector Spaces
5. Eigenvalues and Eigenvectors
   - Differential Equations (§5.7)
   - Markov Chains (§5.9)
6. Orthogonality and Least-Squares
   - Least-Squares Problems
   - Machine Learning: Regression Analysis

# Contents


**Title** ii

**Prologue** iii

**Table of Contents** vii

**1 Linear Equations** 1

1.1. Systems of Linear Equations . . . 2
1.2. Row Reduction and Echelon Forms . . . 9
1.2.1. Echelon Forms . . . 9
1.2.2. The General Solution of Linear Systems . . . 14
1.3. Vector Equations . . . 20
1.3.1. Vectors in $\mathbb{R}^n$ . . . 20
1.3.2. Linear Combinations and Span . . . 23
Programming with Matlab/Octave . . . 29
1.4. Matrix Equation $A\mathbf{x} = \mathbf{b}$ . . . 34
1.5. Solution Sets of Linear Systems . . . 38
1.5.1. Solutions of Homogeneous Linear Systems . . . 38
1.5.2. Solutions of Nonhomogeneous Linear Systems . . . 41
1.7. Linear Independence . . . 44
1.8. Linear Transformations . . . 50
1.9. The Matrix of A Linear Transformation . . . 57
1.9.1. The Standard Matrix . . . 58
1.9.2. Existence and Uniqueness Questions . . . 61

**2 Matrix Algebra** 67

2.1. Matrix Operations . . . 68
2.1.1. Sum, Scalar Multiple, and Matrix Multiplication . . . 68
2.1.2. Properties of Matrix Multiplication . . . 72
2.2. The Inverse of a Matrix . . . 77
2.3. Characterizations of Invertible Matrices . . . 82

2.5. Solving Linear Systems by Matrix Factorizations . . . . . 87
2.5.1. The $LU$ Factorization/Decomposition . . . . . 87
2.5.2. Solutions of Triangular Algebraic Systems . . . . . 92
2.8. Subspaces of $\mathbb{R}^n$ . . . . . 97
2.9. Dimension and Rank . . . . . 103
2.9.1. Coordinate Systems . . . . . 103
2.9.2. Dimension of a Subspace . . . . . 105

**3 Determinants 109**
3.1. Introduction to Determinants . . . . . 110
3.2. Properties of Determinants . . . . . 115

**4 Vector Spaces 119**
4.1. Vector Spaces and Subspaces . . . . . 120
4.2. Null Spaces, Column Spaces, and Linear Transformations . . . . . 125

**5 Eigenvalues and Eigenvectors 131**
5.1. Eigenvectors and Eigenvalues . . . . . 132
5.2. The Characteristic Equation and Similarity Transformation . . . . . 138
5.3. Diagonalization . . . . . 142
5.3.1. The Diagonalization Theorem . . . . . 142
5.3.2. Diagonalizable Matrices . . . . . 146
5.5. Complex Eigenvalues . . . . . 150
5.7. Applications to Differential Equations . . . . . 154
5.7.1. Dynamical System: The System of First-Order Differential Equations . . . . . 155
5.7.2. Trajectories for the Dynamical Systems: Attractors, Repellers, and Saddle Points 157
5.8. Iterative Estimates for Eigenvalues . . . . . 162
5.8.1. The Power Method . . . . . 162
5.8.2. Inverse Power Method . . . . . 165
5.9. Applications to Markov Chains . . . . . 168
5.9.1. Probability Vector and Stochastic Matrix . . . . . 170
5.9.2. Predicting the Distant Future: Steady-State Vectors . . . . . 173

**6 Orthogonality and Least-Squares 179**
6.1. Inner Product, Length, and Orthogonality . . . . . 180
6.1.1. Inner Product and Length . . . . . 180
6.1.2. Orthogonal Vectors . . . . . 183
6.2. Orthogonal Sets . . . . . 187

6.2.1. An Orthogonal Projection . . . . . 189
6.2.2. Orthonormal Basis and Orthogonal Matrix . . . . . 192
6.3. Orthogonal Projections . . . . . 196
6.4. The Gram-Schmidt Process and QR Factorization . . . . . 203
6.4.1. The Gram-Schmidt Process . . . . . 203
6.4.2. QR Factorization of Matrices . . . . . 206
6.5. Least-Squares Problems . . . . . 210
6.6. Machine Learning: Regression Analysis . . . . . 217
6.6.1. Regression Line . . . . . 218
6.6.2. Least-Squares Fitting of Other Curves . . . . . 221

**A Appendix** **227**
A.1. Understanding/Interpretation of Eigenvalues and Eigenvectors . . . . . 228
A.2. Eigenvalues and Eigenvectors of Stochastic Matrices . . . . . 231

**C Chapter Review** **237**
C.1. Linear Equations . . . . . 238
C.2. Matrix Algebra . . . . . 241
C.3. Determinants . . . . . 247
C.4. Vector Spaces . . . . . 250
C.5. Eigenvalues and Eigenvectors . . . . . 252
C.6. Orthogonality and Least-Squares . . . . . 257

**P Projects** **265**
P.1. Project Regression Analysis: Linear, Piecewise Linear, and Nonlinear Models . . . . . 266

**Bibliography** **275**

**Index** **277**

CHAPTER **1**

# Linear Equations

In this first chapter, we study basics of linear equations, including

- Systems of linear equations
- ***Three elementary row operations***
- Linear transformations

**Contents of Chapter 1**


1.1. Systems of Linear Equations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2
1.2. Row Reduction and Echelon Forms . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9
1.3. Vector Equations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
Programming with Matlab/Octave . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
1.4. Matrix Equation $A\mathbf{x} = \mathbf{b}$ . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34
1.5. Solution Sets of Linear Systems . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38
1.7. Linear Independence . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 44
1.8. Linear Transformations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 50
1.9. The Matrix of A Linear Transformation . . . . . . . . . . . . . . . . . . . . . . . . . . . . 57

# 1.1. Systems of Linear Equations

**Definition 1.1.** *A **linear equation** in the variables $x_1, x_2, \cdots, x_n$ is an equation that can be written in the form*

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b, \tag{1.1}$$

*where $b$ and the coefficients $a_1, a_2, \cdots, a_n$ are real or complex numbers.*

A **system of linear equations** (or a **linear system**) is a collection of one or more linear equations involving the same variables – say, $x_1, x_2, \cdots, x_n$.

**Example 1.2.**

(a) $$\begin{cases} 4x_1 - x_2 = 3 \\ 2x_1 + 3x_2 = 5 \end{cases}$$

(b) $$\begin{cases} 2x + 3y - 4z = 2 \\ x - 2y + z = 1 \\ 3x + y - 2z = -1 \end{cases}$$

- **Solution**: A **solution** of the system is a list $(s_1, s_2, \cdots, s_n)$ of numbers that makes each equation a true statement, when the values $s_1, s_2, \cdots, s_n$ are substituted for $x_1, x_2, \cdots, x_n$, respectively.
- **Solution Set**: The set of all possible solutions is called the **solution set** of the linear system.
- **Equivalent System**: Two linear systems are called **equivalent** if they have the same solution set.
- For example, above (a) is equivalent to

$$\begin{cases} 2x_1 - 4x_2 = -2 \\ 2x_1 + 3x_2 = 5 \end{cases} \qquad \textcircled{R_1} \leftarrow \textcircled{R_1} - \textcircled{R_2}$$

**Remark 1.3.** Linear systems may have

$$\left.\begin{array}{l}\text{no solution}\end{array}\right. \quad : \quad \textbf{inconsistent system}$$

$$\left.\begin{array}{l}\text{exactly one (unique) solution}\\ \text{infinitely many solutions}\end{array}\right\} \quad : \quad \textbf{consistent system}$$

**Example 1.4.** Consider the case of two equations in two unknowns.

(a) $\begin{cases} -x + y = 1 \\ -x + y = 3 \end{cases}$ (b) $\begin{cases} x + y = 1 \\ x - y = 2 \end{cases}$ (c) $\begin{cases} -2x + y = 2 \\ -4x + 2y = 4 \end{cases}$

## Existence and Uniqueness Questions

**Two Fundamental Questions about a Linear System**:

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

Most systems in real-world are consistent (existence) and
they produce the same output for the same input (uniqueness).

# Solving Linear Systems

## Matrix Form

Consider a simple system of 2 linear equations:

$$\begin{cases} -2x_1 + 3x_2 &= -1 \\ x_1 + 2x_2 &= 4 \end{cases} \tag{1.2}$$

Such a system of linear equations can be treated *much more conveniently and efficiently* with matrix form. In **matrix form**, (1.2) reads

$$\underbrace{\begin{bmatrix} -2 & 3 \\ 1 & 2 \end{bmatrix}}_{\textbf{coefficient matrix}} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 4 \end{bmatrix}. \tag{1.3}$$

The ***essential information*** of the system can be recorded compactly in a rectangular array called a **augmented matrix**:

$$\begin{bmatrix} -2 & 3 & -1 \\ 1 & 2 & 4 \end{bmatrix} \text{ or } \left[\begin{array}{cc|c} -2 & 3 & -1 \\ 1 & 2 & 4 \end{array}\right] \tag{1.4}$$

## Elementary Row Operations

**Tools 1.5. Three Elementary Row Operations (ERO)**:

- ***Replacement***: Replace one row by the sum of itself and a multiple of another row
  $R_i \leftarrow R_i + k \cdot R_j, \quad j \neq i$
- ***Interchange***: Interchange two rows
  $R_i \leftrightarrow R_j, \quad j \neq i$
- ***Scaling***: Multiply all entries in a row by a nonzero constant
  $R_i \leftarrow k \cdot R_i, \quad k \neq 0$

**Solving (1.2)**

**System of linear equations** / **Matrix form**

$$\begin{cases} -2x_1 + 3x_2 = -1 & ① \\ x_1 + 2x_2 = 4 & ② \end{cases} \qquad \begin{bmatrix} -2 & 3 & -1 \\ 1 & 2 & 4 \end{bmatrix}$$

① ↔ ②: (interchange)

$$\begin{cases} x_1 + 2x_2 = 4 & ① \\ -2x_1 + 3x_2 = -1 & ② \end{cases} \qquad \begin{bmatrix} 1 & 2 & 4 \\ -2 & 3 & -1 \end{bmatrix}$$

② ← ② + 2 · ①: (replacement)

$$\begin{cases} x_1 + 2x_2 = 4 & ① \\ 7x_2 = 7 & ② \end{cases} \qquad \begin{bmatrix} 1 & 2 & 4 \\ 0 & 7 & 7 \end{bmatrix}$$

② ← ②/7: (scaling)

$$\begin{cases} x_1 + 2x_2 = 4 & ① \\ x_2 = 1 & ② \end{cases} \qquad \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & 1 \end{bmatrix}$$

① ← ① − 2 · ②: (replacement)

$$\begin{cases} x_1 = 2 & ① \\ x_2 = 1 & ② \end{cases} \qquad \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix}$$

**At the last step**:

$$\text{LHS: solution} : \begin{cases} x_1 = 2 \\ x_2 = 1 \end{cases} \qquad \text{RHS} : \left[\begin{array}{c|c} \mathbf{I} & \begin{matrix} 2 \\ 1 \end{matrix} \end{array}\right]$$

**Definition 1.6.** Two matrices are **row equivalent** if there is a sequence of EROs that transforms one matrix to the other.

**Example 1.7.** Solve the following system of linear equations, using the 3 EROs. Then, determine if the system is consistent.

$$\begin{aligned} x_2 - 4x_3 &= 8 \\ 2x_1 - 3x_2 + 2x_3 &= 1 \\ 4x_1 - 8x_2 + 12x_3 &= 1 \end{aligned}$$

**Solution**.

*Ans*: ***Inconsistency*** means that there is no point where the three planes meet at.

**Example 1.8.** Determine the values of $h$ such that the given system is a consistent linear system

$$\begin{aligned} x + h\,y &= -5 \\ 2x - 8y &= 6 \end{aligned}$$

**Solution**.

*Ans*: $h \neq -4$

**True-or-False 1.9.**

a. Every elementary row operation is reversible.

b. Elementary row operations on an augmented matrix never change the solution of the associated linear system.

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

d. Two matrices are row equivalent if they have the same number of rows.

**Solution**.

*Ans*: T,T,T,F

You should report your homework with your work for problems. You can scan your solutions and answers, using a scanner or your phone, then try to put in a file, either in doc/docx or pdf.

## Exercises 1.1

1. Consider the augmented matrix of a linear system. State in words the next two elementary row operations that should be performed in the process of solving the system.
$$\begin{bmatrix} 1 & 6 & -4 & 0 & 1 \\ 0 & 1 & 7 & 0 & -4 \\ 0 & 0 & -1 & 2 & 3 \\ 0 & 0 & 2 & 1 & -6 \end{bmatrix}$$

2. The augmented matrix of a linear system has been reduced by row operations to the form shown. Continue the appropriate row operations and describe the solution set.
$$\begin{bmatrix} 1 & 1 & 0 & 0 & 4 \\ 0 & -1 & 3 & 0 & -7 \\ 0 & 0 & 1 & -3 & 1 \\ 0 & 0 & 0 & 2 & 4 \end{bmatrix}$$

3. Solve the systems or determine if the systems in inconsistent.

(a)
$$\begin{aligned} -x_2 - 4x_3 &= 5 \\ x_1 + 3x_2 + 5x_3 &= -2 \\ 3x_1 + 7x_2 + 7x_3 &= 6 \end{aligned}$$

(b)
$$\begin{aligned} x_1 \quad + 3x_3 \quad &= 2 \\ x_2 \quad - 3x_4 &= 3 \\ -2x_2 + 3x_3 + 2x_4 &= 1 \\ 3x_1 \quad + 7x_4 &= -5 \end{aligned}$$

4. Determine the value of $h$ such that the matrix is the augmented matrix of a consistent linear system.
$$\begin{bmatrix} 2 & -3 & h \\ -4 & 6 & -5 \end{bmatrix}$$

*Ans*: $h = 5/2$

5. An important concern in the study of heat transfer is to determine the steady-state temperature distribution of a thin plate when the temperature around the boundary is known. Assume the plate shown in the figure represents a cross section of a metal beam, with negligible heat flow in the direction perpendicular to the plate. Let $T_1, T_2, \cdots, T_4$ denote the temperatures at the four interior nodes of the mesh in the figure. The temperature at a node is approximately equal to the average of the four nearest nodes. For example, $T_1 = (10 + 20 + T_2 + T_4)/4$ or $4T_1 = 10 + 20 + T_2 + T_4$. Write a system of four equations whose solution gives estimates for the temperatures $T_1, T_2, \cdots, T_4$, and solve it.

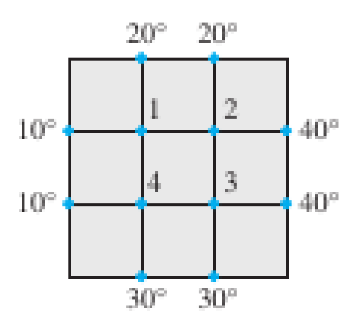


Figure 1.1

# 1.2. Row Reduction and Echelon Forms

## 1.2.1. Echelon Forms

**Terminologies**

- A **nonzero row** in a matrix is a row with at least one nonzero entry.
- A **leading entry** of a row is the left most nonzero entry in a nonzero row.
- A **leading 1** is a leading entry whose value is 1.

**Definition 1.10. Echelon form**: A rectangular matrix is in an **echelon form** if it has following properties.

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

**Row reduced echelon form**: If **a matrix in an echelon form** satisfies 4 and 5 below, then it is in the **row reduced echelon form** (**RREF**), or the **reduced echelon form** (**REF**).

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

**Example 1.11.** Check if the following matrix is in echelon form. If not, put it in echelon form.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 3 & 0 & 4 \end{bmatrix}$$

**Example 1.12.** Verify whether the following matrices are in echelon form, row reduced echelon form.

(a) $\begin{bmatrix} 1 & 0 & 2 & 0 & 1 \\ 0 & 1 & 3 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

(b) $\begin{bmatrix} 2 & 0 & 0 & 5 \\ 0 & 0 & 0 & 9 \\ 0 & 1 & 0 & 6 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 1 & 2 & 2 & 3 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$

(e) $\begin{bmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 6 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

(f) $\begin{bmatrix} 0 & 1 & 0 & 5 \\ 0 & 0 & 0 & 6 \\ 0 & 0 & 1 & 2 \end{bmatrix}$

**Uniqueness of the Reduced Echelon Form**

**Theorem 1.13.** *Each matrix is row equivalent to one and only one reduced echelon form.*

## Pivot Positions

**Terminologies**

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

**Example 1.14.** The matrix $A$ is given with its reduced echelon form. Find the pivot positions and pivot columns of $A$.

$$A = \begin{bmatrix} 1 & 1 & 0 & 2 & 0 \\ 1 & 1 & 1 & 3 & 0 \\ 1 & 1 & 0 & 2 & 4 \end{bmatrix} \xrightarrow{R.E.F} \begin{bmatrix} 1 & 1 & 0 & 2 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

**Solution**.

**Remark 1.15.** Pivot Positions. Once a matrix is in an echelon form, further row operations do not change the positions of leading entries. Thus, the ***leading entries*** become the ***leading 1's*** in the reduced echelon form.

**Terminologies**

3) **Basic variables**: In the system $A\mathbf{x} = \mathbf{b}$, the variables that correspond to pivot columns (in $[A : \mathbf{b}]$) are **basic variables**.
4) **Free variables**: In the system $A\mathbf{x} = \mathbf{b}$, the variables that correspond to non-pivotal columns are **free variables**.

**Example 1.16.** For the system of linear equations, identify its basic variables and free variables.

$$\begin{cases} -x_1 - 2x_2 & = & -3 \\ \phantom{-x_1 - 2x_2} \; 2x_3 & = & 4 \\ \phantom{-x_1 - 2x_2} \; 3x_3 & = & 6 \end{cases}$$

**Solution**. ***Hint***: You may start with its augmented matrix, and apply row operations.

*Ans*: Basic variables: $\{x_1, x_3\}$. Free variable: $\{x_2\}$.

**Why "free"?**

## The Row Reduction Algorithm

**Steps to reduce to reduced echelon form**

1. Start with ***the leftmost non-zero column***. This is a pivot column. The pivot is at the top.
2. Choose a nonzero entry in the pivot column as a pivot. If necessary, ***interchange*** rows to move a nonzero entry into the pivot position.
3. Use row ***replacement*** operations to make zeros in all positions below the pivot.
4. Ignore row and column containing the pivot and ignore all rows above it. Apply Steps 1–3 to the remaining submatrix. Repeat this until there are no more rows to modify.

---

5. Start with right most pivot and work upward and left to make zeros above each pivot. If pivot is not 1, make it 1 by a ***scaling*** operation.

**Example 1.17.** Row reduce the matrix into reduced echelon form.

$$A = \begin{bmatrix} 0 & -3 & -6 & 4 & 9 \\ -2 & -3 & 0 & 3 & -1 \\ 1 & 4 & 5 & -9 & -7 \end{bmatrix}$$

**Solution.** $\xrightarrow{R_1 \leftrightarrow R_3} \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ -2 & -3 & 0 & 3 & -1 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix} \xrightarrow{R_2 \leftarrow R_2 + 2R_1} \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 5 & 10 & -15 & -15 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix}$

$$\xrightarrow{R_2 \leftarrow R_2/5} \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 1 & 2 & -3 & -3 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix} \xrightarrow{R_3 \leftarrow R_3 + 3R_2} \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 1 & 2 & -3 & -3 \\ 0 & 0 & 0 & -5 & 0 \end{bmatrix}$$

$$\xrightarrow{R_3 \leftarrow R_3/-5;\ \text{above the pivot} \to 0} \begin{bmatrix} 1 & 4 & 5 & 0 & -7 \\ 0 & 1 & 2 & 0 & -3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \xrightarrow{R_1 \leftarrow R_1 - 4R_2} \begin{bmatrix} 1 & 0 & -3 & 0 & 5 \\ 0 & 1 & 2 & 0 & -3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Combination of Steps 1–4 is call the **forward phase** of the row reduction, while Step 5 is called the **backward phase**.

### 1.2.2. The General Solution of Linear Systems

1) For example, for an augmented matrix, its R.E.F. is given as

$$\begin{bmatrix} 1 & 0 & -5 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{1.5}$$

2) Then, the associated system of equations reads

$$\begin{array}{rrrcl} x_1 & & -5x_3 & = & 1 \\ & x_2 & +x_3 & = & 4 \\ & & 0 & = & 0 \end{array} \tag{1.6}$$

where $\{x_1, x_2\}$ are basic variables ($\because$ pivots).

3) Rewrite (1.6) as

$$\begin{cases} x_1 = 1 + 5x_3 \\ x_2 = 4 - x_3 \\ x_3 \quad \text{is free} \end{cases} \tag{1.7}$$

4) The system (1.7) can be expressed as

$$\begin{cases} x_1 = 1 + 5x_3 \\ x_2 = 4 - x_3 \\ x_3 = \phantom{4 -{}} x_3 \end{cases} \tag{1.8}$$

5) Thus, the solution of (1.6) can be written as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 5 \\ -1 \\ 1 \end{bmatrix}, \tag{1.9}$$

in which you are free to choose any value for $x_3$. (That is why it is called a "**free variable**".)

- The description in (1.9) is called a **parametric description** of solution set; the free variable $x_3$ acts as a parameter.
- The solution in (1.9) represents **all the solutions of the system (1.5)**, which is called the **general solution** of the system.

**Example 1.18.** Find the general solution of the system whose augmented matrix is

$$[A|\mathbf{b}] = \begin{bmatrix} 1 & 0 & -5 & 0 & -8 & 3 \\ 0 & 1 & 4 & -1 & 0 & 6 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

**Solution**. ***Hint***: You should first row reduce it for the reduced echelon form.

**Example 1.19.** Find the general solution of the system whose augmented matrix is

$$[A|\mathbf{b}] = \begin{bmatrix} 0 & 0 & 0 & 1 & 2 \\ 0 & 1 & 3 & 0 & 2 \\ 0 & 1 & 3 & 2 & 6 \\ 1 & 0 & -9 & 0 & -8 \end{bmatrix}$$

**Solution**.

**Properties**

1) Any nonzero matrix **may be row reduced** (i.e., transformed by elementary row operations) into **more than one matrix in echelon form**, using different sequences of row operations.
2) Once a matrix is **in an echelon form**, further row operations do not change the **pivot positions** (Remark 1.15).
3) Each matrix is row equivalent to **one and only one reduced echelon matrix** (Theorem 1.13, p. 10).

---

4) A linear system is **consistent** if and only if the rightmost column of the augmented matrix is not a pivot column
–i.e., if and only if **an echelon form of the augmented matrix has no row of the form** $[0 \ \cdots \ 0 \ b]$ with $b$ nonzero.
5) If a linear system is **consistent**, then the solution set contains either
   (a) a **unique** solution, when there are **no free variables**, or
   (b) **infinitely many** solutions, when there is **at least one free variable.**

**Example 1.20.** Choose $h$ and $k$ such that the system has

a) No solution  b) Unique solution  c) Many solutions

$$\begin{cases} x_1 - 3\,x_2 &= 1 \\ 2\,x_1 + h\,x_2 &= k \end{cases}$$

**Solution**.

*Ans*: (a) $h = -6$, $k \neq 2$

**True-or-False 1.21.**

a. The row reduction algorithm applies to only to augmented matrices for a linear system.

b. If one row in an echelon form of an augmented matrix is $[0\ 0\ 0\ 0\ 2\ 0]$, then the associated linear system is inconsistent.

c. The pivot positions in a matrix depend on whether or not row interchanges are used in the row reduction process.

d. Reducing a matrix to an echelon form is called the **forward phase** of the row reduction process.

**Solution**.

*Ans*: F,F,F,T

## Exercises 1.2

1. Row reduce the matrices to reduced echelon form. Circle the pivot positions in the final matrix and in the original matrix, and list the pivot columns.

(a) $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 5 & 6 & 7 \\ 6 & 7 & 8 & 9 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 3 & 5 & 7 \\ 3 & 5 & 7 & 9 \\ 5 & 7 & 9 & 1 \end{bmatrix}$

2. Find the general solutions of the systems (in **parametric vector form**) whose augmented matrices are given as

(a) $\begin{bmatrix} 1 & -7 & 0 & 6 & 5 \\ 0 & 0 & 1 & -2 & -3 \\ -1 & 7 & -4 & 2 & 7 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 & -5 & -6 & 0 & -5 \\ 0 & 1 & -6 & -3 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

*Ans*: (a) $\mathbf{x} = [5, 0, -3, 0]^T + x_2[7, 1, 0, 0]^T + x_4[-6, 0, 2, 1]^T$;
*Ans*: (b) $\mathbf{x} = [-9, 2, 0, 0, 0]^T + x_3[-7, 6, 1, 0, 0]^T + x_4[0, 3, 0, 1, 0]^T$[1]

3. In the following, we use the notation for matrices in echelon form: the leading entries with ■, and any values (including zero) with $*$. Suppose each matrix represents the augmented matrix for a system of linear equations. In each case, determine if the system is consistent. If the system is consistent, determine if the solution is unique.

(a) $\begin{bmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & \blacksquare & * \end{bmatrix}$ (b) $\begin{bmatrix} 0 & \blacksquare & * & * & * \\ 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 & \blacksquare \end{bmatrix}$ (c) $\begin{bmatrix} \blacksquare & * & * & * & * \\ 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & \blacksquare & * \end{bmatrix}$

4. Choose $h$ and $k$ such that the system has (a) no solution, (b) a unique solution, and (c) many solutions.

$$\begin{aligned} x_1 + hx_2 &= 2 \\ 4x_1 + 8x_2 &= k \end{aligned}$$

5. Suppose the coefficient matrix of a system of linear equations has a pivot position in every row. Explain why the system is consistent.

[1]The superscript $T$ denotes the **transpose**; for example $[a, b, c]^T = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$.

# 1.3. Vector Equations

**Definition 1.22.** A matrix with only one column is called a **column vector**, or simply a **vector**.

For example,

$$\begin{bmatrix} 3 \\ -2 \end{bmatrix} \in \mathbb{R}^2 \qquad \begin{bmatrix} 1 \\ -5 \\ 4 \end{bmatrix} \in \mathbb{R}^3$$

## 1.3.1. Vectors in $\mathbb{R}^n$

**Vectors in $\mathbb{R}^2$**

We can identify a point $(a, b)$ with a column vector $\begin{bmatrix} a \\ b \end{bmatrix}$, **position vector**.

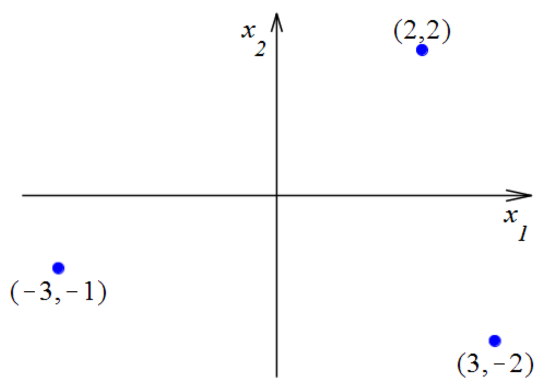



Figure 1.2: Vectors in $\mathbb{R}^2$ as points.

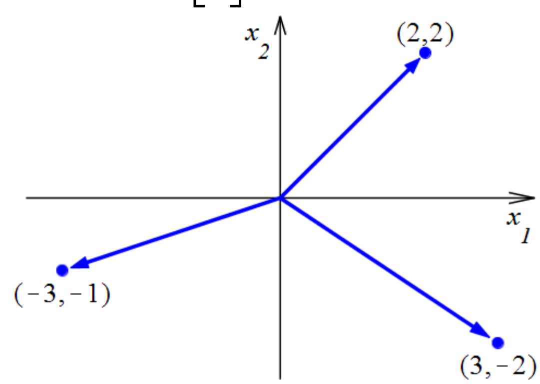



Figure 1.3: Vectors in $\mathbb{R}^2$ with arrows.

**Note**: Vectors are mathematical objects having ***direction*** and ***length***. We may try to (1) compare them, (2) add or subtract them, (3) multiply them by a scalar, (4) measure their length, and (5) apply other operations to get information related to angles.

1) **Equality of vectors**: Two vectors $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ are **equal** if and only if corresponding entries are equal, i.e., $u_i = v_i, \quad i = 1, 2.$

2) **Addition**: Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$. Then,

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \end{bmatrix}.$$

3) **Scalar multiple**: Let $c \in \mathbb{R}$, a scalar. Then

$$c\mathbf{v} = c \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} cv_1 \\ cv_2 \end{bmatrix}.$$

**Theorem 1.23. (Parallelogram Rule for Addition)** *If* u *and* v *are represented as points in the plane, then* u + v *corresponds to the fourth vertex of the parallelogram whose other vertices are* 0*,* u*, and* v*.*

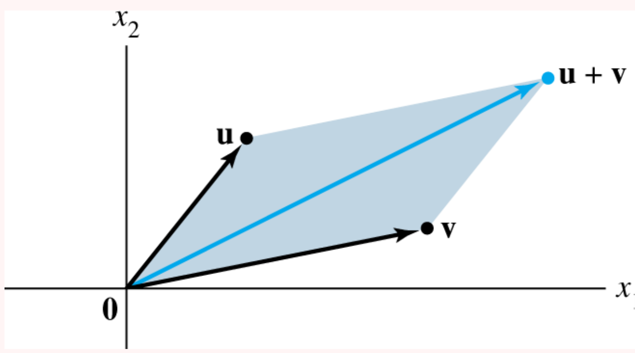



Figure 1.4: The parallelogram rule.

**Example 1.24.** Let $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$.

(a) Find $\mathbf{u} + 2\mathbf{v}$ and $3\mathbf{u} - 2\mathbf{v}$.

(b) Display them on a graph.

**Solution**.

**Remark 1.25.** Let $\mathbf{a}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\mathbf{a}_2 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$. Then

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = x_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ -3 \end{bmatrix} = \begin{bmatrix} 2x_1 + x_2 \\ x_1 - 3x_2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [\mathbf{a}_1\ \mathbf{a}_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{1.10}$$

## Vectors in $\mathbb{R}^n$

**Note**: The above vector operations, including the parallelogram rule, are also applicable for vectors in $\mathbb{R}^3$ and $\mathbb{R}^n$, in general.

**Algebraic Properties of $\mathbb{R}^n$**

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

1) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
2) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
3) $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$
4) $\mathbf{u} + (-\mathbf{u}) = (-\mathbf{u}) + \mathbf{u} = \mathbf{0}$ where $-\mathbf{u} = (-1)\mathbf{u}$
5) $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
6) $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$
7) $c(d\mathbf{u}) = (cd)\mathbf{u}$
8) $1\mathbf{u} = \mathbf{u}$

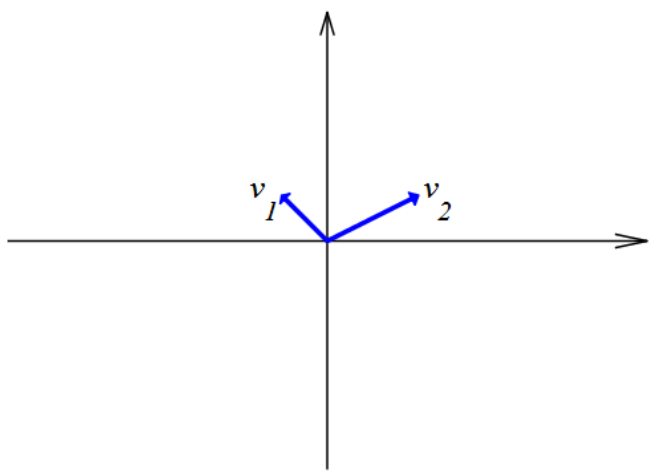
## 1.3.2. Linear Combinations and Span

**Definition 1.26.** Given vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p$ in $\mathbb{R}^n$ and scalars $c_1, c_2, \cdots, c_p$, the vector $\mathbf{y} \in \mathbb{R}^n$ defined by

$$\mathbf{y} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_p \mathbf{v}_p \tag{1.11}$$

is called the **linear combination** of $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p$ with weights $c_1, c_2, \cdots, c_p$.

**Example 1.27.** Given $\mathbf{v}_1$ and $\mathbf{v}_2$ in $\mathbb{R}^2$, as in the figure below, the collection of all linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$ must be the same as $\mathbb{R}^2$.



**Definition 1.28.** A **vector equation** is of the form

$$x_1\,\mathbf{a}_1 + x_2\,\mathbf{a}_2 + \cdots + x_p\,\mathbf{a}_p = \mathbf{b}, \tag{1.12}$$

where $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_p$, and $\mathbf{b}$ are vectors and $x_1, x_2, \cdots, x_p$ are weights.

**Example 1.29.** Let $\mathbf{a}_1 = \begin{bmatrix} 0 \\ 4 \\ -1 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 1 \\ 6 \\ 3 \end{bmatrix}$, $\mathbf{a}_3 = \begin{bmatrix} 5 \\ -1 \\ 8 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix}$.
Determine whether or not $\mathbf{b}$ can be generated as a linear combination of $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$.

**Solution**. ***Hint***: We should determine whether weights $x_1, x_2, x_3$ exist such that $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 = \mathbf{b}$, which reads $[\mathbf{a}_1\,\mathbf{a}_2\,\mathbf{a}_3] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{b}$. (See Remark 1.25 on p.22.)

**Note**:

1) The vector equation $x_1\,\mathbf{a}_1 + x_2\,\mathbf{a}_2 + \cdots + x_p\,\mathbf{a}_p = \mathbf{b}$ has the same solution set as a linear system whose augmented matrix is $[\mathbf{a}_1\,\mathbf{a}_2\,\cdots\,\mathbf{a}_p : \mathbf{b}]$.
2) $\mathbf{b}$ can be generated as a linear combination of $\mathbf{a}_1,\, \mathbf{a}_2,\, \cdots,\, \mathbf{a}_p$ if and only if the linear system $A\mathbf{x} = \mathbf{b}$ whose augmented matrix $[\mathbf{a}_1\,\mathbf{a}_2\,\cdots\,\mathbf{a}_p : \mathbf{b}]$ is consistent.

**Definition 1.30.** Let $\mathbf{v}_1,\, \mathbf{v}_2,\, \cdots,\, \mathbf{v}_p$ be $p$ vectors in $\mathbb{R}^n$. Then **Span**$\{\mathbf{v}_1,\, \mathbf{v}_2,\, \cdots,\, \mathbf{v}_p\}$ is the collection of all linear combinations of $\mathbf{v}_1,\, \mathbf{v}_2,\, \cdots,\, \mathbf{v}_p$, that can be written in the form $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p$, where $c_1,\, c_2,\, \cdots,\, c_p$ are weights. That is,

$$Span\{\mathbf{v}_1,\, \mathbf{v}_2,\, \cdots,\, \mathbf{v}_p\} = \{\mathbf{y} \mid \mathbf{y} = c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_p\,\mathbf{v}_p\} \quad (1.13)$$

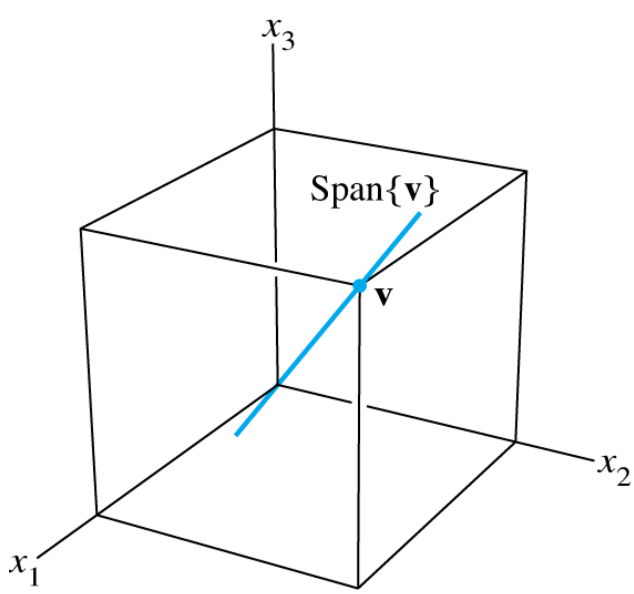



Figure 1.5: A line: $Span\{\mathbf{v}\}$ in $\mathbb{R}^3$.

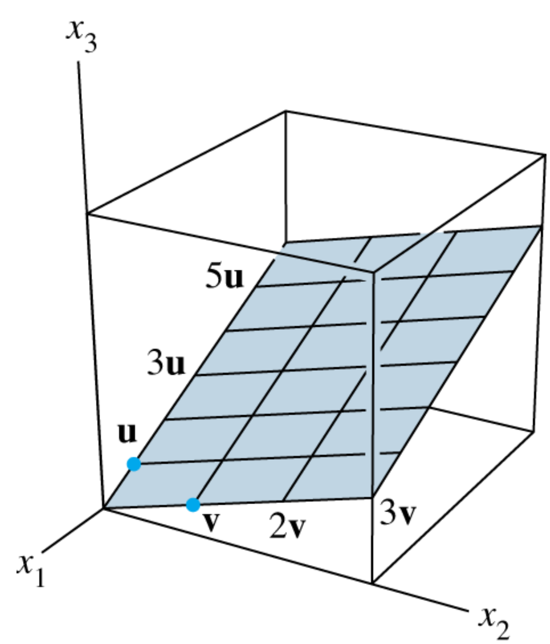



Figure 1.6: A plane: $Span\{\mathbf{u}, \mathbf{v}\}$ in $\mathbb{R}^3$.

**Example 1.31.** Determine if b $= \begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix}$ is a linear combination of the columns of the matrix $\begin{bmatrix} 1 & 0 & 5 \\ -2 & 1 & -6 \\ 0 & 2 & 8 \end{bmatrix}$. (That is, determine if b is in the span of columns of the matrix.)

**Solution**.

**Example 1.32.** Find $h$ so that $\begin{bmatrix} 2 \\ 1 \\ h \end{bmatrix}$ lies in the plane spanned by $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $\mathbf{a}_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$.

**Solution**.

$$\begin{aligned} &\mathbf{b} \in \mathit{Span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\} \\ &\Leftrightarrow x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_p\mathbf{v}_p = \mathbf{b} \text{ has a solution} \\ &\Leftrightarrow [\mathbf{v}_1\, \mathbf{v}_2\, \cdots\, \mathbf{v}_p\, : \mathbf{b}] \text{ has a solution} \end{aligned} \tag{1.14}$$

**True-or-False 1.33.**

a. Another notation for the vector $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ is $[1 \;\; -2]$.

b. The set $\mathit{Span}\{\mathbf{u}, \mathbf{v}\}$ is always visualized as a plane through the origin.

c. When $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors, $\mathit{Span}\{\mathbf{u}, \mathbf{v}\}$ contains the line through $\mathbf{u}$ and the origin.

**Solution**.

*Ans*: F,F,T

## Exercises 1.3

1. Write a system of equations that is equivalent to the given vector equation; write a vector equation that is equivalent to the given system of equations.

(a) $x_1 \begin{bmatrix} 6 \\ -1 \\ 5 \end{bmatrix} + x_2 \begin{bmatrix} -3 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -7 \\ -5 \end{bmatrix}$

(b) $\begin{aligned} x_2 + 5x_3 &= 0 \\ 4x_1 + 6x_2 - x_3 &= 0 \\ -x_1 + 3x_2 - 8x_3 &= 0 \end{aligned}$

2. Determine if $\mathbf{b}$ is a linear combination of $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$.

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 5 \\ -6 \\ 8 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix}.$$

*Ans*: Yes

3. Determine if $\mathbf{b}$ is a linear combination of the vectors formed from the columns of the matrix $A$.

(a) $A = \begin{bmatrix} 1 & -4 & 2 \\ 0 & 3 & 5 \\ -2 & 8 & -4 \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} 3 \\ -7 \\ -3 \end{bmatrix}$

(b) $A = \begin{bmatrix} 1 & -2 & -6 \\ 0 & 3 & 7 \\ 1 & -2 & 5 \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} 11 \\ -5 \\ 9 \end{bmatrix}$

4. Let $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 4 \\ -2 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} -2 \\ -3 \\ 7 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ h \end{bmatrix}$. For what value(s) of $h$ is $\mathbf{b}$ in the plane spanned by $\mathbf{a}_1$ and $\mathbf{a}_2$?

*Ans*: $h = -17$

5. Construct a $3 \times 3$ matrix $A$, with nonzero entries, and a vector $\mathbf{b}$ in $\mathbb{R}^3$ such that $\mathbf{b}$ is *not* in the set spanned by the columns of $A$. ***Hint***: Construct a $3 \times 4$ augmented matrix in echelon form that corresponds to an inconsistent system.

6. A mining company has two mines. One day's operation at mine #1 produces ore that contains 20 metric tons of copper and 550 kilograms of silver, while one day's operation at mine #2 produces ore that contains 30 metric tons of copper and 500 kilograms of silver. Let $\mathbf{v}_1 = \begin{bmatrix} 20 \\ 550 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 30 \\ 500 \end{bmatrix}$. Then $\mathbf{v}_1$ and $\mathbf{v}_2$ represent the "output per day" of mine #1 and mine #2, respectively.

   (a) What physical interpretation can be given to the vector $5\mathbf{v}_1$?

   (b) Suppose the company operates mine #1 for $x_1$ days and mine #2 for $x_2$ days. Write a vector equation whose solution gives the number of days each mine should operate in order to produce 150 tons of copper and 2825 kilograms of silver.

   (c) M[2] Solve the equation in (b).

[2]The mark M indicates that you have to solve the problem, using one of Matlab, Maple, and Mathematica. You may also try "octave" as a free alternative of Matlab. Attach a copy of your code.

# Programming with Matlab/Octave

**Note**: In computer programming, important things are

- How to deal with **objects** (variables, arrays, functions)
- How to deal with **repetition** effectively
- How to make the program **reusable**

## Vectors and matrices

The most basic thing you will need to do is to enter vectors and matrices. You would enter commands to **Matlab** or **Octave** at a prompt that looks like >>.

- Rows are separated by semicolons ( ; ) or Enter .
- Entries in a row are separated by commas ( , ) or space Space .

For example,

Vectors and Matrices

```
>> u = [1; 2; 3]       % column vector
u =
   1
   2
   3
>> v = [4; 5; 6];
>> u + 2*v
ans =
    9
   12
   15
>> w = [5, 6, 7, 8]   % row vector
w =
   5   6   7   8
>> A = [2 1; 1 2];     % matrix
>> B = [-2, 5
        1, 2]
B =
  -2   5
   1   2
>> C = A*B   % matrix multiplication
C =
   -3   12
    0    9
```

You can save the commands in a file to run and get the same results.

tutorial1_vectors.m

```
u = [1; 2; 3]
v = [4; 5; 6];
u + 2*v
w = [5, 6, 7, 8]
A = [2 1; 1 2];
B = [-2, 5
     1, 2]
C = A*B
```

## Solving equations

Let $A = \begin{bmatrix} 1 & -4 & 2 \\ 0 & 3 & 5 \\ -2 & 8 & -4 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 3 \\ -7 \\ -3 \end{bmatrix}$. Then $A\mathbf{x} = \mathbf{b}$ can be *numerically* solved by implementing a code as follows.

tutorial2_solve.m

```
A = [1 -4  2; 0  3  5; 2  8 -4];
b = [3; -7; -3];
x = A\b
```

Result

```
x =
   0.75000
  -0.97115
  -0.81731
```

## Graphics with Matlab

In Matlab, the most popular graphic command is `plot`, which creates a 2D line plot of the data in `Y` versus the corresponding values in `X`. A general syntax for the command is

```
plot(X1,Y1,LineSpec1,...,Xn,Yn,LineSpecn)
```

tutorial3_plot.m

```
1  close all
2
3  %% a curve
4  X1 = linspace(0,2*pi,10); % n=10
5  Y1 = cos(X1);
6
7  %% another curve
8  X2=linspace(0,2*pi,20); Y2=sin(X2);
9
10 %% plot together
11 plot(X1,Y1,'-or',X2,Y2,'--b','linewidth',3);
12 legend({'y=cos(x)','y=sin(x)'},'location','best',...
13        'FontSize',16,'textcolor','blue')
14 print -dpng 'fig_cos_sin.png'
```

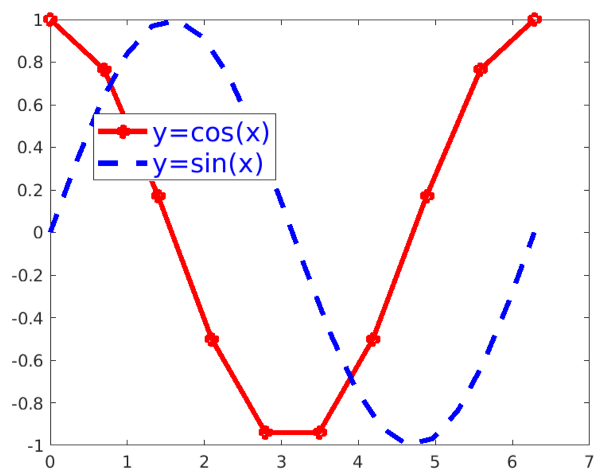



Figure 1.7: fig_cos_sin.png: `plot` of $y = \cos x$ and $y = \sin x$.

Above `tutorial3_plot.m` is a typical M-file for figuring with `plot`.

- Line 1: It closes all figures currently open.
- Lines 3, 4, 7, and 10 (comments): When the percent sign (%) appears, the rest of the line will be ignored by Matlab.
- Lines 4 and 8: The command `linspace(x1,x2,n)` returns a row vector of `n` evenly spaced points between `x1` and `x2`.
- Line 11: Its result is a figure shown in Figure 1.7.
- Line 14: it saves the figure into a png format, named `fig_cos_sin.png`.

## Repetition: iteration loops

**Note**: In scientific computation, one of most frequently occurring events is **repetition**. Each repetition of the process is also called an **iteration**. It is the act of repeating a process, to generate a (possibly unbounded) sequence of outcomes, with the aim of approaching a desired goal, target or result. Thus,

- *iteration must start with an initialization (starting point) and*
- *perform a step-by-step marching in which the results of one iteration are used as the starting point for the next iteration.*

In the context of mathematics or computer science, iteration (along with the related technique of recursion) is a very basic building block in programming. Matlab provides various types of loops: while loops, for loops, and nested loops.

### while loop

The syntax of a while loop in Matlab is as follows.

```
while <expression>
    <statements>
end
```

An expression is true when the result is nonempty and contains all nonzero elements, logical or real numeric; otherwise the expression is false. Here is an example for the while loop.

```
n1=11; n2=20;
sum=n1;
while n1<n2
    n1 = n1+1; sum = sum+n1;
end
fprintf('while loop: sum=%d\n',sum);
```

When the code above is executed, the result will be:

```
while loop: sum=155
```

## for loop

A **for loop** is a repetition control structure that allows you to efficiently write a loop that needs to execute a specific number of times. The syntax of a for loop in Matlab is as following:

```
for index = values
    <program statements>
end
```

Here is an example for the for loop.

```
n1=11; n2=20;
sum=0;
for i=n1:n2
    sum = sum+i;
end
fprintf('for loop: sum=%d\n',sum);
```

When the code above is executed, the result will be:

```
for loop: sum=155
```

## Functions: Enhancing reusability

Program scripts can be saved to **reuse later conveniently**. For example, the script for the summation of integers from `n1` to `n2` can be saved as a form of **function**.

mysum.m

```
function s = mysum(n1,n2)
% sum of integers from n1 to n2

s=0;
for i=n1:n2
    s = s+i;
end
```

Now, you can call it with e.g. `mysum(11,20)`.
Then the result reads `ans = 155`.

## 1.4. Matrix Equation $A\mathbf{x} = \mathbf{b}$

**A fundamental idea in linear algebra** is to view a **linear combination of vectors** as a **product of a matrix and a vector**.

**Definition 1.34.** Let $A = [\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_n]$ be an $m \times n$ matrix and $\mathbf{x} \in \mathbb{R}^n$, then the **product of** $A$ **and** $\mathbf{x}$ denoted by $A\,\mathbf{x}$ is the linear combination of columns of $A$ using the corresponding entries of $\mathbf{x}$ as weights, i.e.,

$$A\,\mathbf{x} = [\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\,\mathbf{a}_1 + x_2\,\mathbf{a}_2 + \cdots + x_n\,\mathbf{a}_n. \tag{1.15}$$

A **matrix equation** is of the form $A\,\mathbf{x} = \mathbf{b}$, where $\mathbf{b}$ is a column vector of size $m \times 1$.

**Example 1.35.**

| Matrix equation | Vector equation | Linear system |
|---|---|---|
| $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ | $x_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ | $\begin{aligned} x_1 + 2x_2 &= 1 \\ 3x_1 + 4x_2 &= -1 \end{aligned}$ |

**Theorem 1.36.** *Let* $A = [\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_n]$ *be an* $m \times n$ *matrix,* $\mathbf{x} \in \mathbb{R}^n$*, and* $\mathbf{b} \in \mathbb{R}^m$*. Then the matrix equation*

$$A\mathbf{x} = \mathbf{b} \tag{1.16}$$

*has the same solution set as the vector equation*

$$x_1\,\mathbf{a}_1 + x_2\,\mathbf{a}_2 + \cdots + x_n\,\mathbf{a}_n = \mathbf{b}, \tag{1.17}$$

*which, in turn, has the same solution set as the system with augmented matrix*

$$[\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_n : \mathbf{b}]. \tag{1.18}$$

**Theorem 1.37. (Existence of solutions)**: Let $A$ be an $m \times n$ matrix. Then the following statements are logically equivalent (all true, or all false).

a. For each $\mathbf{b}$ in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution.

b. Each $\mathbf{b}$ in $\mathbb{R}^m$ is a linear combination of columns of $A$.

c. The columns of $A$ span $\mathbb{R}^m$.

d. $A$ has a pivot position in every row.
(Note that $A$ is the coefficient matrix.)

**Example 1.38.** Let $\mathbf{v}_1 = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 0 \\ -3 \\ 9 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 4 \\ -2 \\ -6 \end{bmatrix}$.

Does $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ span $\mathbb{R}^3$? Why or why not?

**Example 1.39.** Do the vectors $\begin{bmatrix} 1 \\ 0 \\ 1 \\ -2 \end{bmatrix}$, $\begin{bmatrix} 3 \\ 1 \\ 2 \\ -8 \end{bmatrix}$, and $\begin{bmatrix} -2 \\ 1 \\ -3 \\ 2 \end{bmatrix}$ span $\mathbb{R}^4$?

**Solution**.

**True-or-False 1.40.**

a. The equation $A\mathrm{x} = \mathrm{b}$ is referred to as a *vector equation*.

b. Each entry in $A\mathrm{x}$ is the result of a *dot product*.

c. If $A \in \mathbb{R}^{m\times n}$ and if $A\mathrm{x} = \mathrm{b}$ is inconsistent for some $\mathrm{b} \in \mathbb{R}^m$, then $A$ cannot have a pivot position in every row.

d. If the augmented matrix $[A\ \ \mathrm{b}]$ has a pivot position in every row, then the equation $A\mathrm{x} = \mathrm{b}$ is inconsistent.

**Solution**.

*Ans*: F,T,T,F

## Exercises 1.4

1. Write the system first as a vector equation and then as a matrix equation.

$$\begin{aligned} 3x_1 + x_2 - 5x_3 &= 9 \\ x_2 + 4x_3 &= 0 \end{aligned}$$

2. Let $\mathbf{u} = \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix}$ and $A = \begin{bmatrix} 3 & -5 \\ -2 & 6 \\ 1 & 1 \end{bmatrix}$. Is $\mathbf{u}$ in the plane $\mathbb{R}^3$ spanned by the columns of $A$? (See the figure.) Why or why not?

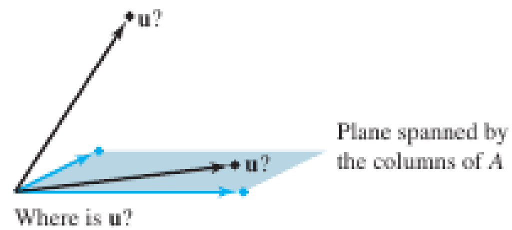



Figure 1.8

3. The problems refer to the matrices $A$ and $B$ below. Make appropriate calculations that justify your answers and mention an appropriate theorem.

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

(a) How many rows of $A$ contain a pivot position? Does the equation $A\mathbf{x} = \mathbf{b}$ have a solution for each $\mathbf{b}$ in $\mathbb{R}^4$?

(b) Can each vector in $\mathbb{R}^4$ be written as a linear combination of the columns of the matrix $A$ above? Do the columns of $A$ span $\mathbb{R}^4$?

(c) Can each vector in $\mathbb{R}^4$ be written as a linear combination of the columns of the matrix $B$ above? Do the columns of $B$ span $\mathbb{R}^4$?

*Ans*: (a) 3; (b) Theorem 1.37 (d) is not true

4. Let $\mathbf{v}_1 = \begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 0 \\ -3 \\ 8 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 4 \\ -1 \\ -5 \end{bmatrix}$. Does $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ span $\mathbb{R}^3$? Why or why not?

*Ans*: The matrix of $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ has a pivot position on each row.

5. Could a set of three vectors in $\mathbb{R}^4$ span all of $\mathbb{R}^4$? Explain. What about $n$ vectors in $\mathbb{R}^m$ when $n < m$?

6. Suppose $A$ is a $4 \times 3$ matrix and $\mathbf{b} \in \mathbb{R}^4$ with the property that $A\mathbf{x} = \mathbf{b}$ has a unique solution. What can you say about the reduced echelon form of $A$? Justify your answer. ***Hint***: How many pivot columns does $A$ have?

# 1.5. Solution Sets of Linear Systems

**Linear Systems $A\mathrm{x} = \mathrm{b}$:**

1. *Homogeneous linear systems*:
$$A\mathbf{x} = \mathbf{0}; \quad A \in \mathbb{R}^{m\times n},\ \mathbf{x} \in \mathbb{R}^n,\ \mathbf{0} \in \mathbb{R}^m. \tag{1.19}$$
   (a) It has always at least one solution: $\mathrm{x} = 0$ (**the trivial solution**)
   (b) Any nonzero solution is called a **nontrivial solution**.
2. *Nonhomogeneous linear systems*:
$$A\mathbf{x} = \mathbf{b}; \quad A \in \mathbb{R}^{m\times n},\ \mathbf{x} \in \mathbb{R}^n,\ \mathbf{b} \in \mathbb{R}^m,\ \mathbf{b} \neq \mathbf{0}. \tag{1.20}$$

**Note**: $A\mathrm{x} = \mathbf{0}$ has a nontrivial solution if and only if the system has at least one free variable.

## 1.5.1. Solutions of Homogeneous Linear Systems

**Example 1.41.** Determine if the following homogeneous system has a nontrivial solution. Then describe the solution set.

$$\begin{aligned} x_1 - 2x_2 + 3x_3 &= 0 \\ -2x_1 - 3x_2 - 4x_3 &= 0 \\ 2x_1 - 4x_2 + 9x_3 &= 0 \end{aligned}$$

**Definition 1.42.** If the solutions of $A\mathrm{x} = \mathbf{0}$ can be written in the form

$$\mathbf{x} = c_1\,\mathbf{u}_1 + c_2\,\mathbf{u}_2 + \cdots + c_r\,\mathbf{u}_r, \tag{1.21}$$

where $c_1, c_2, \cdots, c_r$ are scalars and $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r$ are vectors with size same as x, then they are said to be in **parametric vector form**.

**Note**: When solutions of $A\mathrm{x} = \mathbf{0}$ is in the form of (1.21), we may say

$$\textbf{\{The solution set of } A\mathrm{x} = \mathbf{0}\textbf{\}} = \textit{Span}\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r\}. \tag{1.22}$$

**Example 1.43.** Solve the system and write the solution in parametric vector form.

$$\begin{aligned} x_1 + 2x_2 - 3x_3 &= 0 \\ 2x_1 + x_2 - 3x_3 &= 0 \\ -x_1 + x_2 \qquad &= 0 \end{aligned}$$

**Example 1.44.** Describe all solutions of $A\mathrm{x} = \mathbf{0}$ in parametric vector form where $A$ is row equivalent to the matrix.

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

***Hint***: You should first row reduce it for the reduced echelon form.

**Solution**.

A single equation can be treated as a simple linear system.

**Example 1.45.** Solve the equation of 3 variables and write the solution in parametric vector form.

$$x_1 - 2x_2 + 3x_3 = 0$$

**Solution**. ***Hint***: $x_1$ is only the basic variable. Thus your solution would be the form of $\mathbf{x} = x_2\mathbf{v}_1 + x_3\mathbf{v}_2$, which is a parametric vector equation of the **plane**.

## 1.5.2. Solutions of Nonhomogeneous Linear Systems

**Example 1.46.** Describe all solutions of $A\mathrm{x} = \mathrm{b}$, where

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

**Solution**.

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

The solution of Example 1.46 is of the form

$$\mathbf{x} = \mathbf{p} + x_3\mathbf{v} \;\; (= \mathbf{p} + t\,\mathbf{v}), \tag{1.23}$$

where $t$ is a general parameter. Note that (1.23) is an equation of line through p parallel to v.

In the previous example, the solution of $A\mathrm{x} = \mathrm{b}$ is $\mathbf{x} = \mathbf{p} + t\,\mathbf{v}$.

**Question: What is "$t$ v"?**

**Solution**. First of all,

$$A\mathbf{x} = A(\mathbf{p} + t\,\mathbf{v}) = A\mathbf{p} + A(t\mathbf{v}) = \mathbf{b}. \tag{1.24}$$

Note that $\mathbf{x} = \mathbf{p} + t\,\mathbf{v}$ is a solution of $A\mathbf{x} = \mathbf{b}$, even when $t = 0$. Thus,

$$A(\mathbf{p} + t\,\mathbf{v})_{t=0} = A\mathbf{p} = \mathbf{b}. \tag{1.25}$$

It follows from (1.24) and (1.25) that

$$\mathbf{A(t\,v) = 0}, \tag{1.26}$$

which implies that **"$t$ v" is a solution of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.**

**Theorem 1.47.** *Suppose the equation $A\mathrm{x} = \mathrm{b}$ is consistent for some given $\mathrm{b}$, and let $\mathrm{p}$ be a solution. Then the solution set of $A\mathrm{x} = \mathrm{b}$ is the set of all vectors of the form $\{\mathrm{w} = \mathrm{p} + \mathbf{u}_h\}$, where $\mathbf{u}_h$ is the solution of the homogeneous equation $A\mathrm{x} = \mathbf{0}$.*

**Corollary 1.48.** *Let $A\mathrm{x} = \mathrm{b}$ have a solution. The solution is unique if and only if $A\mathrm{x} = \mathbf{0}$ has only the trivial solution.*

**True-or-False 1.49.**

a. The solution set of $A\mathrm{x} = \mathrm{b}$ is the set of all vectors of the form $\{\mathrm{w} = \mathrm{p} + \mathbf{u}_h\}$, where $\mathbf{u}_h$ is the solution of the homogeneous equation $A\mathrm{x} = \mathbf{0}$. (Compare with Theorem 1.47, p.42.)

b. The equation $A\mathrm{x} = \mathrm{b}$ is homogeneous if the zero vector is a solution.

c. The solution set of $A\mathrm{x} = \mathrm{b}$ is obtained by translating the solution of $A\mathrm{x} = \mathbf{0}$.

**Solution**.

*Ans*: F,T,F

## Exercises 1.5

1. Determine if the system has a nontrivial solution. Try to use as few row operations as possible.

(a)
$$\begin{aligned} 2x_1 - 5x_2 + 8x_3 &= 0 \\ 2x_1 - 7x_2 + x_3 &= 0 \\ 4x_1 - 12x_2 + 9x_3 &= 0 \end{aligned}$$

(b)
$$\begin{aligned} 3x_1 + 5x_2 - 7x_3 &= 0 \\ 6x_1 + 7x_2 + x_3 &= 0 \end{aligned}$$

***Hint***: $x_3$ is a free variable for both (a) and (b).

2. Describe all solutions of $A\mathbf{x} = \mathbf{0}$ in *parametric vector form*, where $A$ is row equivalent to the given matrix.

(a) $\begin{bmatrix} 1 & 3 & -3 & 7 \\ 0 & 1 & -4 & 5 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & -4 & -2 & 0 & 3 & -5 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

***Hint***: (b) $x_2$, $x_4$, and $x_6$ are free variables.

3. Describe and compare the solution sets of $x_1 - 3x_2 + 5x_3 = 0$ and $x_1 - 3x_2 + 5x_3 = 4$. ***Hint***: You must solve two problems each of which has a single equation, which in turn represents a plane. For both, only $x_1$ is the basic variable.

4. Suppose $A\mathbf{x} = \mathbf{b}$ has a solution. Explain why the solution is unique precisely when $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

5. (1) Does the equation $A\mathbf{x} = \mathbf{0}$ have a nontrivial solution and (2) does the equation $A\mathbf{x} = \mathbf{b}$ have at least one solution for every possible $\mathbf{b}$?

   (a) $A$ is a $3 \times 3$ matrix with three pivot positions.
   (b) $A$ is a $3 \times 3$ matrix with two pivot positions.

## 1.7. Linear Independence

**Definition 1.50.** A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ is said to be **linearly independent**, if the vector equation

$$x_1\,\mathbf{v}_1 + x_2\,\mathbf{v}_2 + \cdots + x_p\,\mathbf{v}_p = \mathbf{0} \tag{1.27}$$

has only the trivial solution (i.e., $x_1 = x_2 = \cdots = x_p = 0$). The set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$ is said to be **linearly dependent**, if there exist weights $c_1, c_2, \cdots, c_p$, not all zero, such that

$$c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_p\,\mathbf{v}_p = \mathbf{0}. \tag{1.28}$$

**Example 1.51.** Determine if the set $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent.

1) $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 5 \end{bmatrix}$

2) $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

**Remark 1.52.** Let $A = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p]$. The matrix equation $A\mathrm{x} = \mathbf{0}$ is equivalent to $x_1\,\mathbf{v}_1 + x_2\,\mathbf{v}_2 + \cdots + x_p\,\mathbf{v}_p = \mathbf{0}$.

1. Columns of $A$ are **linearly independent** if and only if $A\mathrm{x} = \mathbf{0}$ has *only* the trivial solution. ( $\Leftrightarrow$ $A\mathrm{x} = \mathbf{0}$ has no free variable $\Leftrightarrow$ Every column in $A$ is a pivot column.)
2. Columns of $A$ are **linearly dependent** if and only if $A\mathrm{x} = \mathbf{0}$ has a nontrivial solution. ( $\Leftrightarrow$ $A\mathrm{x} = \mathbf{0}$ has at least one free variable $\Leftrightarrow$ $A$ has at least one *non*-pivot column.)

**Example 1.53.** Determine if the vectors are linearly independent.

1) $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}$

2) $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}, \begin{bmatrix}2\\1\\2\end{bmatrix}$

**Solution**.

**Example 1.54.** Determine if the vectors are linearly independent.

$$\begin{bmatrix}0\\2\\3\end{bmatrix}, \begin{bmatrix}0\\0\\-8\end{bmatrix}, \begin{bmatrix}-1\\3\\1\end{bmatrix}$$

**Solution**.

**Example 1.55.** Determine if the vectors are linearly independent.

$$\begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ -6 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}$$

**Solution**.

**Note**: In the above example, vectors are in $\mathbb{R}^n$, $n = 3$; the number of vectors $p = 4$. Like this, if $p > n$ then the vectors *must* be linearly dependent.

**Theorem 1.56.** *The set of vectors* $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\} \subset \mathbb{R}^n$ *is linearly dependent, if* $p > n$.

**Proof**. Let $A = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_p] \in \mathbb{R}^{n \times p}$. Then $A\mathbf{x} = \mathbf{0}$ has $n$ equations with $p$ unknowns. When $p > n$, there are more variables than equations; this implies there is at least one free variable, which in turn means that there is a nontrivial solution. □

**Example 1.57.** Find the value of $h$ so that the vectors are linearly independent.

$$\begin{bmatrix} 3 \\ -6 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} -6 \\ 4 \\ -3 \end{bmatrix}, \quad \begin{bmatrix} 9 \\ h \\ 3 \end{bmatrix}$$

**Solution**.

**Example 1.58. (Revision of Example 1.57)**: Find the value of $h$ so that **c** is in *Span*{a, b}.

$$\mathbf{a} = \begin{bmatrix} 3 \\ -6 \\ 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -6 \\ 4 \\ -3 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} 9 \\ h \\ 3 \end{bmatrix}$$

**Solution**.

**Example 1.59.** Determine **by inspection** if the vectors are linearly dependent.

1) $\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}$ 2) $\begin{bmatrix}1\\0\\-1\end{bmatrix}, \begin{bmatrix}0\\2\\3\end{bmatrix}, \begin{bmatrix}0\\0\\0\end{bmatrix}$ 3) $\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}3\\6\\9\end{bmatrix}$

**Solution**.

**Note**: Let $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$. If $\mathcal{S}$ contains the zero vector, then it is always linearly dependent. A vector in $\mathcal{S}$ is a linear combination of other vectors in $\mathcal{S}$ if and only if $\mathcal{S}$ is linearly dependent.

**True-or-False 1.60.**

a. The columns of any $3 \times 4$ matrix are linearly dependent.

b. If u and v are linearly independent, and if $\{\mathrm{u}, \mathrm{v}, \mathrm{w}\}$ is linearly dependent, then $\mathrm{w} \in Span\{\mathrm{u}, \mathrm{v}\}$.

c. Two vectors are linearly dependent if and only if they lie on a line through the origin.

d. The columns of a matrix $A$ are linearly independent, if the equation $A\mathrm{x} = \mathbf{0}$ has the trivial solution.

**Solution**.

*Ans*: T,T,T,F

## Exercises 1.7

1. Determine if the columns of the matrix form a linearly independent set. Justify each answer.

(a) $\begin{bmatrix} -4 & -3 & 0 \\ 0 & -1 & 4 \\ 1 & 0 & 3 \\ 5 & 4 & 6 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & -3 & 3 & -2 \\ -3 & 7 & -1 & 2 \\ 0 & 1 & -4 & 3 \end{bmatrix}$

2. Find the value(s) of $h$ for which the vectors are linearly dependent. Justify each answer.

(a) $\begin{bmatrix} 1 \\ -1 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ -5 \\ 7 \end{bmatrix}, \begin{bmatrix} -1 \\ 5 \\ h \end{bmatrix}$ (b) $\begin{bmatrix} 1 \\ 5 \\ -3 \end{bmatrix}, \begin{bmatrix} -2 \\ -9 \\ 6 \end{bmatrix}, \begin{bmatrix} 3 \\ h \\ -9 \end{bmatrix}$

3. (a) For what values of $h$ is $\mathbf{v}_3$ in *Span*$\{\mathbf{v}_1, \mathbf{v}_2\}$, and (b) for what values of $h$ is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ linearly dependent? Justify each answer.

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -3 \\ 9 \\ -6 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 5 \\ -7 \\ h \end{bmatrix}.$$

*Ans*: (a) No $h$; (b) All $h$

4. Describe the possible echelon forms of the matrix. Use the notation of Exercise 3 in Section 1.2, p. 19.

(a) $A$ is a $3 \times 3$ matrix with linearly independent columns.
(b) $A$ is a $2 \times 2$ matrix with linearly dependent columns.
(c) $A$ is a $4 \times 2$ matrix, $A = [\mathbf{a}_1, \mathbf{a}_2]$ and $\mathbf{a}_2$ is not a multiple of $\mathbf{a}_1$.

# 1.8. Linear Transformations

**Example 1.61.** Let $A \in \mathbb{R}^{m\times n}$ and $\mathbf{x} \in \mathbb{R}^n$.
Then **$A\mathbf{x}$ is a new vector in $\mathbb{R}^m$**. For example,

$$A = \begin{bmatrix} 4 & -3 & 1 \\ 2 & 0 & 5 \end{bmatrix} \in \mathbb{R}^{2\times 3}, \quad \mathbf{x} = \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} \in \mathbb{R}^{3}.$$

Then

$$A\mathbf{x} \;=\; \begin{bmatrix} 4 & -3 & 1 \\ 2 & 0 & 5 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 4\cdot(-1) - 3\cdot(1) + 1\cdot(3) \\ 2\cdot(-1) + 0\cdot(1) + 5\cdot(3) \end{bmatrix} = \underbrace{\begin{bmatrix} -4 \\ 13 \end{bmatrix}}_{\text{a new vector in } \mathbb{R}^2}.$$

That is, **$A$ transforms vectors to another space.**

**Definition 1.62. Transformation (function or mapping)**

A **transformation** $T$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that **assigns to each vector $\mathbf{x} \in \mathbb{R}^n$ a vector $T(\mathbf{x}) \in \mathbb{R}^m$**. In this case, we write

$$\begin{aligned} T: \;\; \mathbb{R}^n &\to \mathbb{R}^m \\ \mathbf{x} &\mapsto T(\mathbf{x}) \end{aligned} \tag{1.29}$$

where $\mathbb{R}^n$ is the **domain** of $T$, $\mathbb{R}^m$ is the **codomain** of $T$, and $T(\mathbf{x})$ denotes **the image of** $\mathbf{x}$ under $T$. The set of all images is called the **range** of $T$.

$$\text{Range}(T) \;=\; \{T(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n\}$$

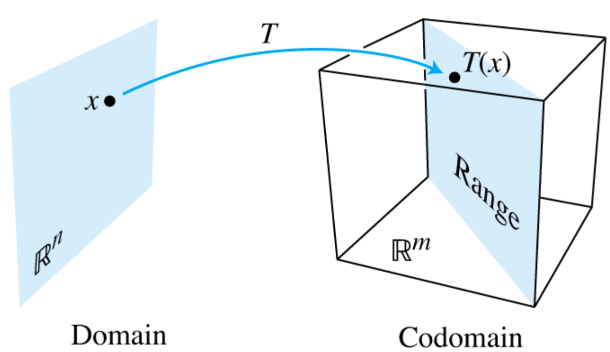



Figure 1.9: Transformation $T : \mathbb{R}^n \to \mathbb{R}^m$.

**Definition 1.63.** Transformation associated with matrix multiplication is **matrix transformation**. That is, for each $\mathbf{x} \in \mathbb{R}^n$, $T(\mathbf{x}) = A\mathbf{x}$, where $A$ is an $m \times n$ matrix. We may denote the matrix transformation as

$$\begin{aligned} T: \ \mathbb{R}^n &\to \mathbb{R}^m \\ \mathbf{x} &\mapsto A\mathbf{x} \end{aligned} \tag{1.30}$$

Here the range of $T$ is set of all linear combinations of columns of $A$.

$$\text{Range}(T) = \textit{Span}\{\text{columns of } A\}.$$

**Example 1.64.** Let $A = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$. The transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(\mathbf{x}) = A\mathbf{x}$ is called a **shear transformation**. Determine the image of a square $[0, 2] \times [0, 2]$ under $T$.

**Solution**. ***Hint***: Matrix transformations is an affine mapping, which means that they map line segments into line segments (and corners to corners).

**Example 1.65.** Let $A = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}$, $\mathbf{c} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}$, and define a transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ by $T(\mathbf{x}) = A\mathbf{x}$.

a. Find $T(\mathbf{u})$, the image of $\mathbf{u}$ under the transformation $T$.

b. Find an $\mathbf{x} \in \mathbb{R}^2$ whose image under $T$ is $\mathbf{b}$.

c. Is there more than one $\mathbf{x}$ whose image under $T$ is $\mathbf{b}$?

d. Determine if $\mathbf{c}$ is in the range of the transformation $T$.

**Solution**.

*Ans*: b. $\mathbf{x} = \begin{bmatrix} 1.5 \\ -0.5 \end{bmatrix}$; c. no; d. no

## Linear Transformations

**Definition 1.66.** A transformation $T$ is **linear** if:

(i) $\boldsymbol{T}(\mathbf{u}+\mathbf{v}) = \boldsymbol{T}(\mathbf{u}) + \boldsymbol{T}(\mathbf{v})$, for all $\mathbf{u}$, $\mathbf{v}$ in the domain of $T$

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

**Claim 1.67.** If $T$ is a linear transformation, then

$$T(\mathbf{0}) = \mathbf{0} \tag{1.31}$$

and

$$\mathbf{T}(\mathbf{c}\,\mathbf{u} + \mathbf{d}\,\mathbf{v}) = \mathbf{c}\,\mathbf{T}(\mathbf{u}) + \mathbf{d}\,\mathbf{T}(\mathbf{v}), \tag{1.32}$$

for all vectors $\mathbf{u}$, $\mathbf{v}$ in the domain of $T$ and all scalars $c, d$.
We can easily prove that if $T$ *satisfies* (1.32)*, then* $T$ *is linear*.

**Remark 1.68.** The function $\boldsymbol{f(x) = ax}$ is a linear transformation:

$$f(c\,x_1 + d\,x_2) = a(cx_1 + dx_2) = c(ax_1) + d(ax_2) = c\,f(x_1) + d\,f(x_2). \tag{1.33}$$

**Example 1.69.** Prove that a matrix transformation $T(\mathbf{x}) = A\mathbf{x}$ is linear.

**Proof**. It is easy to see that

$$T(c\,\mathbf{u} + d\,\mathbf{v}) \;=\; A(c\,\mathbf{u} + d\,\mathbf{v}) \;=\; c\,A\mathbf{u} + d\,A\mathbf{v} \;=\; c\,T(\mathbf{u}) + d\,T(\mathbf{v}),$$

which completes the proof, satisfying (1.32). □

**Remark 1.70.** Repeated application of (1.32) produces a useful generalization:

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \cdots + c_pT(\mathbf{v}_p). \tag{1.34}$$

In engineering physics, (1.34) is referred to as a **superposition principle**.

**Example 1.71.** Let $\theta$ be the angle measured from the positive $x$-axis counterclockwise. Then, the **rotation** can be defined as

$$\mathcal{R}[\theta] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{1.35}$$

1) Describe $\mathcal{R}[\pi/2]$ explicitly.
2) What are images of $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ under $\mathcal{R}[\pi/2]$.
3) Is $\mathcal{R}[\theta]$ a linear transformation?

**Solution**.

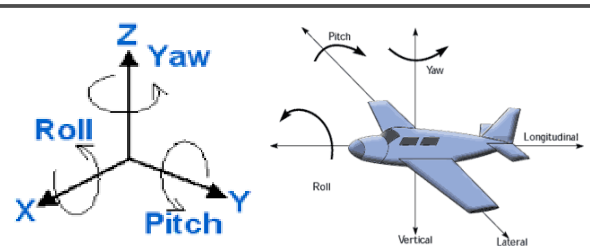



Figure 1.10: Euler angles (roll,pitch,yaw) in aerodynamics.

For example, a **yaw** is a counterclockwise rotation of $\psi$ about the $z$-axis. The rotation matrix reads

$$R_z[\psi] = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

**True-or-False 1.72.**

a. If $A \in \mathbb{R}^{3\times 5}$ and $T$ is a transformation defined by $T(\mathbf{x}) = A\mathbf{x}$, then the domain of $T$ is $\mathbb{R}^3$.

b. A linear transformation is a special type of function.

c. The *superposition principle* is a physical description of a linear transformation.

d. Every matrix transformation is a linear transformation.

e. Every linear transformation is a matrix transformation. (If it is false, can you find an example that is linear but of no matrix description?)

**Solution**.

*Ans*: F,T,T,T,F

## Exercises 1.8

1. With $T$ defined by $T\mathrm{x} = A\mathrm{x}$, find a vector x whose image under $T$ is b, and determine whether x is unique.
$$A = \begin{bmatrix} 1 & -3 & 2 \\ 0 & 1 & -4 \\ 3 & -5 & -9 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 6 \\ -7 \\ -9 \end{bmatrix}$$
*Ans*: $\mathrm{x} = \begin{bmatrix} -5 \\ -3 \\ 1 \end{bmatrix}$, unique

2. Answer the following
   (a) Let $A$ be a $6 \times 5$ matrix. What must $a$ and $b$ be in order to define $T : \mathbb{R}^a \to \mathbb{R}^b$ by by $T\mathbf{x} = A\mathbf{x}$?
   (b) How many rows and columns must a matrix $A$ have in order to define a mapping from $\mathbb{R}^4$ into $\mathbb{R}^5$ by the rule $T\mathbf{x} = A\mathbf{x}$?

*Ans*: (a) $a = 5;\ \ b = 6$

3. Let $\mathbf{b} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & -4 & 7 & -5 \\ 0 & 1 & -4 & 3 \\ 2 & -6 & 6 & -4 \end{bmatrix}$. Is b in the range of the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$? Why or why not?

*Ans*: yes

4. Use a rectangular coordinate system to plot $\mathbf{u} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}, \ \ \mathbf{u} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$, and their images under the given transformation $T$. (Make a separate and reasonably large sketch.) Describe geometrically what $T$ does to each vector x in $\mathbb{R}^2$.
$$T(\mathbf{x}) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

5. Show that the transformation $T$ defined by $T(x_1, x_2) = (2x_1 - 3x_2, x_1 + 4, 5x_2)$ is *not* linear.
***Hint***: $T(0,0) = \mathbf{0}$?

6. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the transformation that projects each vector $\mathbf{x} = (x_1, x_2, x_3)$ onto the plane $x_2 = 0$, so $T(\mathbf{x}) = (x_1, 0, x_3)$. Show that T is a linear transformation.
***Hint***: Try to verify (1.32): $T(c\mathbf{x} + d\mathbf{y}) = T(cx_1 + dy_1, cx_2 + dy_2, cx_3 + dy_3) = \cdots = cT(\mathbf{x}) + dT(\mathbf{y})$.

## 1.9. The Matrix of A Linear Transformation

**Note**: In Example 1.69 (p. 53), we proved that *every matrix transformation is linear*. The reverse is not always true. However, **a linear transformation defined in $\mathbb{R}^n$ is a matrix transformation**.

Here in this section, we will try to find matrices for linear transformations defined in $\mathbb{R}^n$. Let's begin with an example.

**Example 1.73.** Suppose $T : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear transformation such that

$$T(\mathbf{e}_1) = \begin{bmatrix} 5 \\ -7 \\ 2 \end{bmatrix}, \ T(\mathbf{e}_2) = \begin{bmatrix} -3 \\ 8 \\ 0 \end{bmatrix}, \quad \text{where } \mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \ \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

**Solution**. What we should do is to find a matrix $A \in \mathbb{R}^{3\times 2}$ such that

$$T(\mathbf{e}_1) = A\mathbf{e}_1 = \begin{bmatrix} 5 \\ -7 \\ 2 \end{bmatrix}, \ T(\mathbf{e}_2) = A\mathbf{e}_2 = \begin{bmatrix} -3 \\ 8 \\ 0 \end{bmatrix}. \tag{1.36}$$

Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$. Then

$$\mathbf{x} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2. \tag{1.37}$$

It follows from *linearity of $T$* that

$$\begin{aligned} T(\mathbf{x}) &= T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2) = x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) \\ &= [T(\mathbf{e}_1)\ T(\mathbf{e}_2)] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 5 & -3 \\ -7 & 8 \\ 2 & 0 \end{bmatrix}}_{A} \mathbf{x}. \end{aligned} \tag{1.38}$$

Now, you can easily check that $A$ satisfies (1.36). □

**Observation 1.74.** The matrix of a linear transformation is decided by **its action on the standard basis**.

### 1.9.1. The Standard Matrix

**Theorem 1.75.** *Let $T:\mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there exists a unique matrix $A \in \mathbb{R}^{m\times n}$ such that*

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

*In fact, with $\mathbf{e}_j$ denoting the $j$-th standard unit vector in $\mathbb{R}^n$,*

$$A = [T(\mathbf{e}_1)\ \ T(\mathbf{e}_2)\ \ \cdots\ \ T(\mathbf{e}_n)]\,. \tag{1.39}$$

*The matrix $A$ is called the* ***standard matrix*** *of the transformation.*

**Note**: **Standard unit vectors in $\mathbb{R}^n$ & the standard matrix**:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix},\ \ \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},\ \ \cdots,\ \ \mathbf{e}_n = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \tag{1.40}$$

Any $\mathbf{x} \in \mathbb{R}^n$ can be written as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_n \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{bmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n.$$

Thus

$$\begin{aligned} T(\mathbf{x}) &= T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n) \\ &= x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + \cdots + x_nT(\mathbf{e}_n) \\ &= [T(\mathbf{e}_1)\ \ T(\mathbf{e}_2)\ \ \cdots\ \ T(\mathbf{e}_n)]\,\mathbf{x}, \end{aligned} \tag{1.41}$$

and therefore the standard matrix reads

$$A = [T(\mathbf{e}_1)\ \ T(\mathbf{e}_2)\ \ \cdots\ \ T(\mathbf{e}_n)]\,. \tag{1.42}$$

**Example 1.76.** Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a horizontal **shear transformation** that leaves $\mathrm{e}_1$ unchanged and maps $\mathrm{e}_2$ into $\mathrm{e}_2 + 2\mathrm{e}_1$. Write the standard matrix of $T$.

**Solution**.

*Ans*: $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$.

**Example 1.77.** Write the standard matrix for the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^4$ given by

$$T(x_1, x_2) = (x_1 + 4x_2, 0, x_1 - 3x_2, x_1).$$

**Solution**.

## Geometric Linear Transformations of $\mathbb{R}^2$

**Example 1.78.** Find the standard matrices for the **reflections in** $\mathbb{R}^2$.

1) The reflection through the $x_1$-axis, defined as $R_1 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix}$.
2) The reflection through the line $x_1 = x_2$, defined as $R_2 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}$.
3) The reflection through the line $x_1 = -x_2$ **(Define** $R_3$ **first.)**

**Solution**.

*Ans*: 3) $A_3 = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$

## 1.9.2. Existence and Uniqueness Questions

**Definition 1.79.**

1) A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **surjective** (**onto** $\mathbb{R}^m$) if each b in $\mathbb{R}^m$ is the image of *at least one* x in $\mathbb{R}^n$.
2) A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **injective** (**one-to-one**) if each b in $\mathbb{R}^m$ is the image of *at most one* x in $\mathbb{R}^n$.

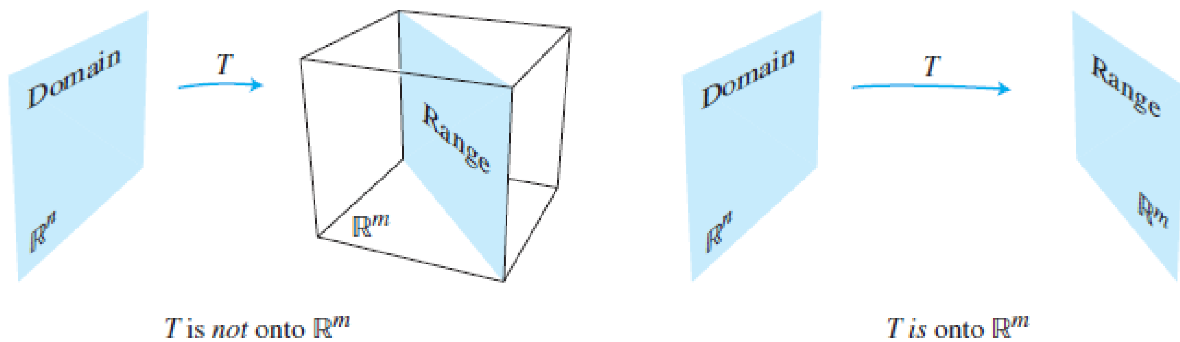



Figure 1.11: Surjective?: Is the range of $T$ all of $\mathbb{R}^m$?

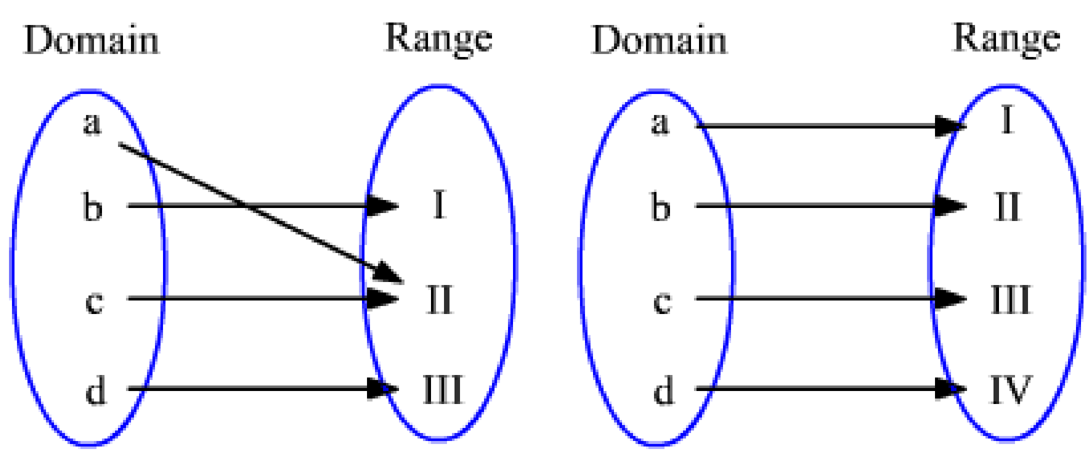



Figure 1.12: Injective?: Is each b $\in \mathbb{R}^m$ the image of one and only one x in $\mathbb{R}^n$?

**Note**: For solutions of $A\mathrm{x} = \mathrm{b}$, $A \in \mathbb{R}^{m\times n}$; existence is related to "surjective"-ness, while uniqueness is granted for "injective" mappings.

**Example 1.80.** Let $T : \mathbb{R}^4 \to \mathbb{R}^3$ be the linear transformation whose standard matrix is

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

Is $T$ onto? Is $T$ one-to-one?
**Solution**.

*Ans*: onto, but not one-to-one

**Theorem 1.81.** *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with the standard matrix $A$. Then,*

*(a) $T$ maps $\mathbb{R}^n$* **onto** *$\mathbb{R}^m$ if and only if the columns of $A$ span $\mathbb{R}^m$.*
*( $\Leftrightarrow$ every row of $A$ has a pivot position*
*$\Leftrightarrow$ $A\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b} \in \mathbb{R}^m$)*

*(b) $T$ is* **one-to-one** *if and only if the columns of $A$ are linearly independent.*
*( $\Leftrightarrow$ every column of $A$ is a pivot column*
*$\Leftrightarrow$ $A\mathbf{x} = \mathbf{0}$ has "only" the trivial solution)*

**Example 1.82.** Let $T(\mathbf{x}) = \begin{bmatrix} 1 & 3 & 0 \\ 0 & 3 & 4 \\ 0 & 0 & 4 \end{bmatrix} \mathbf{x}$. Is $T$ one-to-one? Does $T$ map $\mathbb{R}^3$ onto $\mathbb{R}^3$?
**Solution**.

**Example 1.83.** Let $T(\mathbf{x}) = \begin{bmatrix} 1 & 4 \\ 0 & 0 \\ 1 & -3 \\ 1 & 0 \end{bmatrix}\mathbf{x}$. Is $T$ one-to-one (1–1)? Is $T$ onto?

**Solution**.

**Example 1.84.** Let $T(\mathbf{x}) = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 2 & 4 & 6 \end{bmatrix}\mathbf{x}$. Is $T$ 1–1? Is $T$ onto?

**Solution**.

**Example 1.85.** Let $T : \mathbb{R}^4 \to \mathbb{R}^3$ be the linear transformation given by

$$T(x_1, x_2, x_3, x_4) = (x_1 - 4x_2 + 8x_3 + x_4, 2x_2 - 8x_3 + 3x_4, 5x_4).$$

Is $T$ 1–1? Is $T$ onto?
**Solution**.

**True-or-False 1.86.**

a. A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is *one-to-one* if each vector in $\mathbb{R}^n$ maps onto a unique vector in $\mathbb{R}^m$.
b. If $A$ is a $3 \times 2$ matrix, then the transformation $\mathbf{x} \mapsto A\mathbf{x}$ cannot map $\mathbb{R}^2$ onto $\mathbb{R}^3$.
c. If $A$ is a $3 \times 2$ matrix, then the transformation $\mathbf{x} \mapsto A\mathbf{x}$ cannot be one-to-one. (See Theorem 1.81, p.62.)
d. A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is completely determined by its action on the columns of the $n \times n$ identity matrix.

**Solution**.

*Ans*: F,T,F,T

## Exercises 1.9

1. Assume that $T$ is a linear transformation. Find the standard matrix of $T$.

   (a) $T : \mathbb{R}^2 \to \mathbb{R}^4$, $T(\mathbf{e}_1) = (3, 1, 3, 1)$ and $T(\mathbf{e}_2) = (5, 2, 0, 0)$, where $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$.

   (b) $T : \mathbb{R}^2 \to \mathbb{R}^2$ first performs a horizontal shear that transforms $\mathbf{e}_2$ into $\mathbf{e}_2 - 2\mathbf{e}_1$ (leaving $\mathbf{e}_1$ unchanged) and then reflects points through the line $x_2 = -x_1$.

   *Ans*: (b) shear: $\begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$ and reflection: $\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$; it becomes $\begin{bmatrix} 0 & -1 \\ -1 & 2 \end{bmatrix}$.

2. Show that $T$ is a linear transformation by finding a matrix that implements the mapping. Note that $x_1, x_2, \cdots$ are not vectors but are entries in vectors.

   (a) $T(x_1, x_2, x_3, x_4) = (0, x_1 + x_2, x_2 + x_3, x_2 + x_4)$

   (b) $T(x_1, x_2, x_3) = (x_1 - 5x_2 + 4x_3, x_2 - 6x_3)$

   (c) $T(x_1, x_2, x_3, x_4) = 2x_1 + 3x_3 - 4x_4$

3. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation such that $T(x_1, x_2) = (x_1 + x_2, 4x_1 + 5x_2)$. Find $\mathbf{x}$ such that $T(\mathbf{x}) = (3, 8)$.

   *Ans*: $\mathbf{x} = \begin{bmatrix} 7 \\ -4 \end{bmatrix}$

4. Determine if the specified linear transformation is (1) one-to-one and (2) onto. Justify each answer.

   (a) The transformation in Exercise 2(a).

   (b) The transformation in Exercise 2(b).

   *Ans*: (a) Not 1-1, not onto; (b) Not 1-1, but onto

5. Describe the possible echelon forms of the standard matrix for a linear transformation $T$, where $T : \mathbb{R}^4 \to \mathbb{R}^3$ is onto. Use the notation of Exercise 3 in Section 1.2.
   ***Hint***: The matrix should have a pivot position in each row. Thus there 4 different possible echelon forms.

CHAPTER **2**

# Matrix Algebra

From the elementary school, you have learned about *numbers* and *operations* such as addition, subtraction, multiplication, division, and factorization. **Matrices are also mathematical objects.** Thus you may define matrix operations, similarly done for numbers. **Matrix algebra** is a study about such matrix operations and related applications. Algorithms and techniques you will learn through this chapter are **quite fundamental and important to further develop for application tasks.**

**Contents of Chapter 2**


2.1. Matrix Operations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 68
2.2. The Inverse of a Matrix . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 77
2.3. Characterizations of Invertible Matrices . . . . . . . . . . . . . . . . . . . . . . . . 82
2.5. Solving Linear Systems by Matrix Factorizations . . . . . . . . . . . . . . . . . . . . . 87
2.8. Subspaces of $\mathbb{R}^n$ . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 97
2.9. Dimension and Rank . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 103

# 2.1. Matrix Operations

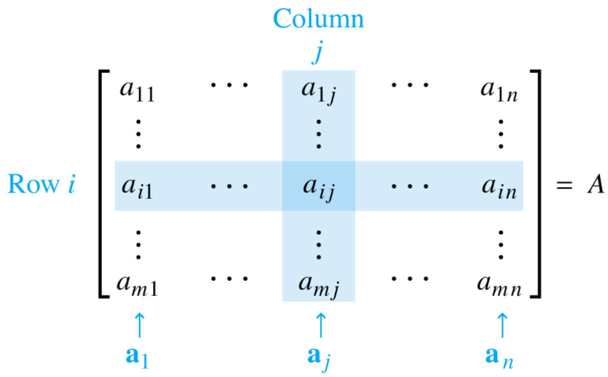



Figure 2.1: Matrix $A \in \mathbb{R}^{m\times n}$.

Let $A$ be an $m \times n$ matrix.
Let $a_{ij}$ denotes the entry in row $i$ and column $j$. Then, we write $A = [a_{ij}]$.

**Terminologies**

- If $m = n$, $A$ is called a **square matrix**.
- If $A$ is an $n \times n$ matrix, then the entries $a_{11}, a_{22}, \cdots, a_{nn}$ are called **diagonal entries**.
- A **diagonal matrix** is a square matrix (say $n \times n$) whose non-diagonal entries are zero.
  Ex: Identity matrix $I_n$.

## 2.1.1. Sum, Scalar Multiple, and Matrix Multiplication

1) **Equality**: Two matrices $A$ and $B$ of the same size (say $m \times n$) are **equal** if and only if the corresponding entries in $A$ and $B$ are equal.
2) **Sum**: The **sum of two matrices** $A = [a_{ij}]$ and $B = [b_{ij}]$ of the same size is the matrix $A + B = [a_{ij} + b_{ij}]$.
3) **Scalar multiplication**: Let $r$ be any scalar. Then, $rA = r[a_{ij}] = [ra_{ij}]$.

**Example 2.1.** Let

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

a) $A + B$, $\quad A + C$?

b) $A - 2B$

**Solution**.

## Matrix Multiplication

**Definition 2.2. Matrix Multiplication**
If $A$ is an $m \times n$ matrix and $B$ is an $n \times p$ matrix with columns $\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_p$, then the **matrix product** $AB$ is a matrix with columns $A\mathbf{b}_1, A\mathbf{b}_2, \cdots, A\mathbf{b}_p$. That is,

$$AB = A[\mathbf{b}_1\ \mathbf{b}_2\ \cdots\ \mathbf{b}_p] = [A\mathbf{b}_1\ A\mathbf{b}_2\ \cdots\ A\mathbf{b}_p] \in \mathbb{R}^{m\times p}, \tag{2.1}$$

which is a collection of matrix-vector multiplications.

**Matrix multiplication, as a Composition of Two Linear Transformations**. $AB$'s action on $\mathbf{x}$:

$$A\,B:\ \ \mathbf{x} \in \mathbb{R}^p \ \underset{B}{\mapsto}\ B\mathbf{x} \in \mathbb{R}^n \ \underset{A}{\mapsto}\ A(B\mathbf{x}) \in \mathbb{R}^m \tag{2.2}$$

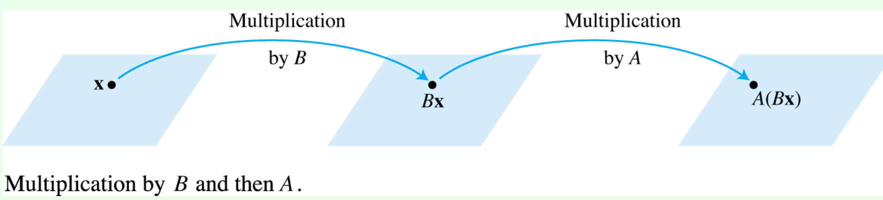



Figure 2.2

Let $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$, and $\mathbf{x} = [x_1, x_2, \cdots, x_p]^T \in \mathbb{R}^p$. Then,

$$\begin{aligned}
B\mathbf{x} &= [\mathbf{b}_1\ \mathbf{b}_2\ \cdots\ \mathbf{b}_p]\,\mathbf{x} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots + x_p\mathbf{b}_p \\
\Rightarrow\ A(B\mathbf{x}) &= A(x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots + x_p\mathbf{b}_p) \\
&= x_1A\mathbf{b}_1 + x_2A\mathbf{b}_2 + \cdots + x_pA\mathbf{b}_p \\
&= [A\mathbf{b}_1\ A\mathbf{b}_2\ \cdots\ A\mathbf{b}_p]\,\mathbf{x} \\
\Rightarrow\ AB &= [A\mathbf{b}_1\ A\mathbf{b}_2\ \cdots\ A\mathbf{b}_p]\ \in \mathbb{R}^{m\times p}
\end{aligned} \tag{2.3}$$

where $\mathbf{b}_i \in \mathbb{R}^n$.

**Example 2.3.** Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ and $S : \mathbb{R}^2 \to \mathbb{R}^2$ be such that $T(\mathbf{x}) = A\mathbf{x}$ where $A = \begin{bmatrix} 4 & -3 \\ -3 & 5 \\ 0 & 1 \end{bmatrix}$ and $S(\mathbf{x}) = B\mathbf{x}$ where $B = \begin{bmatrix} 1 & 4 \\ 3 & -2 \end{bmatrix}$. Compute the standard matrix of $T \circ S$.

**Solution.** ***Hint***: $(T \circ S)(\mathbf{x}) = T(S(\mathbf{x})) = T(B\mathbf{x}) = A(B\mathbf{x}) = (AB)\mathbf{x}$.

**Row-Column rule for Matrix Multiplication**

If the product $AB$ is defined (i.e. number of columns in $A$ = number of rows in $B$) and $A \in \mathbb{R}^{m\times n}$, then the entry in row $i$ and column $j$ of $AB$ is the **sum of products** of corresponding entries from row $i$ of $A$ and column $j$ of $B$. That is,

$$(AB)_{ij} = a_{i1}b1j + a_{i2}b2j + \cdots + a_{in}b_{nj}. \tag{2.4}$$

The *sum of products* is also called the **dot product**.

## 2.1.2. Properties of Matrix Multiplication

**Example 2.4.** Compute $AB$ if

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

**Solution**.

**Example 2.5.** Find all columns of matrix $B$ if

$$A = \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}, \quad AB = \begin{bmatrix} 8 & 7 \\ 7 & -2 \end{bmatrix}.$$

**Solution**.

*Ans*: $B = \begin{bmatrix} 2 & 5 \\ 3 & 1 \end{bmatrix}$

**Example 2.6.** Find the first column of $B$ if

$$A = \begin{bmatrix} -2 & 3 \\ 3 & 3 \\ 5 & -3 \end{bmatrix} \text{ and } AB = \begin{bmatrix} -11 & 10 \\ 9 & 0 \\ 23 & -16 \end{bmatrix}.$$

**Solution**.

*Ans*: $\mathbf{b}_1 = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$

**Remark 2.7.**

1. **(Commutativity)** Suppose both $AB$ and $BA$ are defined. Then, in general, $AB \neq BA$.
$$A = \begin{bmatrix} 3 & -6 \\ -1 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 1 \\ 3 & 4 \end{bmatrix}. \text{ Then } AB = \begin{bmatrix} -21 & -21 \\ 7 & 7 \end{bmatrix}, \quad BA = \begin{bmatrix} -4 & 8 \\ 5 & -10 \end{bmatrix}.$$
2. **(Cancellation law)** If $AB = AC$, then $B = C$ needs not be true always. (e.g., $A = 0$) $\Rightarrow$ *determinant* and *invertibility*
3. **(Powers of a matrix)** If $A$ is an $n \times n$ matrix and if $k$ is a positive integer, then $A^k$ denotes the product of $k$ copies of $A$:
$$A^k = \underbrace{A \cdots A}_{k \text{ times}}$$
If $k = 0$, then $A^k$ is identified with the identity matrix, $I_n$.

## Transpose of a Matrix

**Definition 2.8.** Given an $m \times n$ matrix $A$, the **transpose** of $A$ is the matrix, denoted by $A^T$, whose columns are formed from the corresponding rows of $A$.

**Example 2.9.** If $A = \begin{bmatrix} 1 & -4 & 8 & 1 \\ 0 & 2 & -1 & 3 \\ 0 & 0 & 0 & 5 \end{bmatrix}$, then $A^T =$

**Theorem 2.10.** *Let* $A = [a_{ij}]$.

a. $A^T = [a_{ji}]$

b. $(A^T)^T = A$

c. $(A + B)^T = A^T + B^T$

d. $(rA)^T = r\,A^T$, *for any scalar* $r$

e. $(AB)^T = B^T\,A^T$

**Note**: The transpose of a product of matrices equals the product of their transposes in the reverse order: $(ABC)^T = C^T\,B^T\,A^T$

**True-or-False 2.11.**

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

b. The second row of $AB$ is the second row of $A$ multiplied on the right by $B$.

c. The transpose of a sum of matrices equals the sum of their transposes.

**Solution.**

*Ans*: F(T if $A \leftrightarrow B$),T,T

**Challenge 2.12.**

a. Show that if the columns of $B$ are linearly dependent, then so are the columns of $AB$.

b. Suppose $CA = I_n$ (the $n \times n$ identity matrix). Show that the equation $A\mathrm{x} = \mathbf{0}$ has only the trivial solution.

***Hint***: a. The condition means that $B\mathrm{x} = \mathbf{0}$ has a nontrivial solution.

## Exercises 2.1

1. Compute the product $AB$ in two ways: (a) by the definition, where $A\mathbf{b}_1$ and $A\mathbf{b}_2$ are computed separately, and (b) by the row-column rule for computing $AB$.
$$A = \begin{bmatrix} -1 & 2 \\ 5 & 4 \\ 2 & -3 \end{bmatrix} \text{ and } B = \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}.$$

2. If a matrix $A$ is $5 \times 3$ and the product $AB$ is $5 \times 7$, what is the size of $B$?

3. Let $A = \begin{bmatrix} 2 & -3 \\ -4 & 6 \end{bmatrix}$, $B = \begin{bmatrix} 8 & 4 \\ 5 & 5 \end{bmatrix}$, and $C = \begin{bmatrix} 5 & -2 \\ 3 & 1 \end{bmatrix}$. Verify that $AB = AC$ and yet $B \neq C$.

4. If $A = \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix}$ and $AB = \begin{bmatrix} -1 & 2 & -1 \\ 6 & -9 & 3 \end{bmatrix}$, determine the first and second columns of $B$.

   *Ans*: $\mathbf{b}_1 = \begin{bmatrix} 7 \\ 4 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} -8 \\ -5 \end{bmatrix}$

5. Give a formula for $(AB\mathbf{x})^T$, where $\mathbf{x}$ is a vector and $A$ and $B$ are matrices of appropriate sizes.

6. Let $\mathbf{u} = \begin{bmatrix} -2 \\ 3 \\ -4 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$. Compute $\mathbf{u}^T\mathbf{v}$, $\mathbf{v}^T\mathbf{u}$, $\mathbf{u}\mathbf{v}^T$ and $\mathbf{v}\mathbf{u}^T$.

   *Ans*: $\mathbf{u}^T\mathbf{v} = -2a + 3b - 4c$ and $\mathbf{u}\mathbf{v}^T = \begin{bmatrix} -2a & -2b & -2c \\ 3a & 3b & 3c \\ -4a & -4b & -4c \end{bmatrix}$.

# 2.2. The Inverse of a Matrix

**Definition 2.13.** An $n \times n$ matrix $A$ is said to be **invertible (nonsingular)** if there is an $n \times n$ matrix $B$ such that $AB = I_n = BA$, where $I_n$ is the identity matrix.

**Note**: In this case, $B$ is the *unique inverse* of $A$ denoted by $A^{-1}$.
(Thus $AA^{-1} = I_n = A^{-1}A$.)

**Example 2.14.** If $A = \begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix}$ and $B = \begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix}$. Find $AB$ and $BA$.

**Solution**.

**Theorem 2.15.** **(Inverse of an $n \times n$ matrix, $n \geq 2$)** *An $n \times n$ matrix $A$ is invertible if and only if $A$ is row equivalent to $I_n$ and in this case any sequence of elementary row operations that reduces $A$ into $I_n$ will also reduce $I_n$ to $A^{-1}$.*

**Algorithm 2.16.** *Algorithm to find $A^{-1}$:*

1) *Row reduce the augmented matrix $[A : I_n]$*
2) *If $A$ is row equivalent to $I_n$, then $[A : I_n]$ is row equivalent to $[I_n : A^{-1}]$. Otherwise $A$ does not have any inverse.*

**Example 2.17.** Find the inverse of $A = \begin{bmatrix} 3 & 2 \\ 8 & 5 \end{bmatrix}$.

**Solution**. You may begin with

$$[A : I_2] = \begin{bmatrix} 3 & 2 & 1 & 0 \\ 8 & 5 & 0 & 1 \end{bmatrix}$$

*Ans*: $A^{-1} = \begin{bmatrix} -5 & 2 \\ 8 & -3 \end{bmatrix}$

**Example 2.18.** Find the inverse of $A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix}$, if it exists.

**Solution**.

**Example 2.19.** Find the inverse of $A = \begin{bmatrix} 1 & 2 & -1 \\ -4 & -7 & 3 \\ -2 & -6 & 4 \end{bmatrix}$, if it exists.

**Solution**.

**Theorem 2.20.**

a. **(Inverse of a $2 \times 2$ matrix)** *Let* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. *If* $ad - bc \neq 0$, *then* $A$ *is invertible and*

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{2.5}$$

b. *If* $A$ *is an invertible matrix, then* $A^{-1}$ *is also invertible and* $(A^{-1})^{-1} = A$.

c. *If* $A$ *and* $B$ *are* $n \times n$ *invertible matrices then* $AB$ *is also invertible and* $(AB)^{-1} = B^{-1}A^{-1}$.

d. *If* $A$ *is invertible, then* $A^T$ *is also invertible and* $(A^T)^{-1} = (A^{-1})^T$.

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

**Example 2.21.** When $A$, $B$, $C$, and $D$ are $n \times n$ invertible matrices, solve for $X$ if $C^{-1}(A+X)B^{-1} = D$.

**Solution**.

**Example 2.22.** Explain why the columns of an $n \times n$ matrix $A$ are *linearly independent* when $A$ is invertible.

**Solution**. ***Hint***: Let $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = 0$. Then, show that $\mathbf{x} = [x_1, x_2, \cdots, x_n]^T = 0$.

**Remark 2.23. (Another view of matrix inversion)** For an invertible matrix $A$, we have $AA^{-1} = I_n$. Let $A^{-1} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n]$. Then

$$AA^{-1} = A[\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n] = [\mathbf{e}_1 \ \mathbf{e}_2 \ \cdots \ \mathbf{e}_n]. \tag{2.6}$$

Thus the $j$-th column of $A^{-1}$ is the solution of

$$A\mathbf{x}_j = \mathbf{e}_j, \quad j = 1, 2, \cdots, n.$$

**True-or-False 2.24.**

a. In order for a matrix $B$ to be the inverse of $A$, both equations $AB = I_n$ and $BA = I_n$ must be true.

b. If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $ad = bc$, then $A$ is not invertible.

c. If $A$ is invertible, then elementary row operations that reduce $A$ to the identity $I_n$ also reduce $A^{-1}$ to $I_n$.

**Solution**.

*Ans*: T,T,F

## Exercises 2.2

1. Find the inverses of the matrices, if exist: $A = \begin{bmatrix} 3 & -4 \\ 7 & -8 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -2 & 1 \\ 4 & -7 & 3 \\ -2 & 6 & -4 \end{bmatrix}$

   *Ans*: $B$ is not invertible.

2. Use matrix algebra to show that if $A$ is invertible and $D$ satisfies $AD = I$, then $D = A^{-1}$. ***Hint***: You may start with $AD = I$ and then multiply $A^{-1}$.

3. Solve the equation $AB + C = BC$ for $A$, assuming that $A$, $B$, and $C$ are square and $B$ is invertible.

4. Explain why the columns of an $n \times n$ matrix $A$ span $\mathbb{R}^n$ when $A$ is invertible. ***Hint***: If $A$ is invertible, then $A\mathrm{x} = \mathrm{b}$ has a solution for all $\mathrm{b}$ in $\mathbb{R}^n$.

5. Suppose $A$ is $n \times n$ and the equation $A\mathrm{x} = \mathbf{0}$ has only the trivial solution. Explain why $A$ is row equivalent to $I_n$. ***Hint***: $A$ has $n$ pivot columns.

6. Suppose $A$ is $n \times n$ and the equation $A\mathrm{x} = \mathrm{b}$ has a solution for each $\mathrm{b}$ in $\mathbb{R}^n$. Explain why $A$ must be invertible. ***Hint***: $A$ has $n$ pivot columns.

## 2.3. Characterizations of Invertible Matrices

**Theorem 2.25. (Invertible Matrix Theorem)** *Let $A$ be an $n \times n$ matrix. Then the following are equivalent.*

a. ***$A$ is an invertible matrix.*** *(Def: There is $B$ s.t. $AB = BA = I$)*

b. *$A$ is row equivalent to the $n \times n$ identity matrix.*

c. *$A$ has $n$ pivot positions.*

d. *The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = 0$.*

e. *The columns of $A$ are linearly independent.*

f. *The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.*

g. *The equation $A\mathbf{x} = \mathbf{b}$ has unique solution for each $\mathbf{b} \in \mathbb{R}^n$.*

h. *The columns of $A$ span $\mathbb{R}^n$.*

i. *The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.*

j. *There is a matrix $C \in \mathbb{R}^{n\times n}$ such that $CA = I$*

k. *There is a matrix $D \in \mathbb{R}^{n\times n}$ such that $AD = I$*

l. *$A^T$ is invertible and $(A^T)^{-1} = (A^{-1})^T$.*

*More statements will be added in the coming sections.*

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

**Example 2.26.** Use the Invertible Matrix Theorem to decide if A is invertible:

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 3 & 1 & -2 \\ -5 & -1 & 9 \end{bmatrix}$$

**Solution**.

**Example 2.27.** Can a square matrix with two identical columns be invertible?

**Example 2.28.** An $n \times n$ **upper triangular matrix** is one whose entries below the main diagonal are zeros. When is a square upper triangular matrix invertible?

**Theorem 2.29. (Invertible linear transformations)**

1. *A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is said to be **invertible** if there exists $S : \mathbb{R}^n \to \mathbb{R}^n$ such that $S \circ T(\mathbf{x}) = T \circ S(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$. In this case, $S = T^{-1}$.*
2. *Also, if $A$ is the **standard matrix** for $T$, then $A^{-1}$ is the standard matrix for $T^{-1}$.*

**Example 2.30.** Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation such that $T\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -5x_1 + 9x_2 \\ 4x_1 - 7x_2 \end{bmatrix}$. Find a formula for $T^{-1}$.

**Solution**.

**Example 2.31.** Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be one-to-one. What can you say about $T$?

**Solution**. ***Hint***: $T$: 1-1 $\Leftrightarrow$ Columns of $A$ is linearly independent

**True-or-False 2.32.** Let $A$ be an $n \times n$ matrix.

a. If the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, then $A$ is row equivalent to the $n \times n$ identity matrix.

b. If the columns of A span $\mathbb{R}^n$, then the columns are linearly independent.

c. If A is an $n \times n$ matrix, then the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b} \in \mathbb{R}^n$.

d. If the equation $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution, then $A$ has fewer than $n$ pivot positions.

e. If $A^T$ is not invertible, then $A$ is not invertible.

f. If the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b} \in \mathbb{R}^n$, then the solution is unique for each $\mathbf{b}$.

**Solution**.

*Ans*: T,T,F,T,T,T

## Exercises 2.3

1. An $m \times n$ **lower triangular matrix** is one whose entries *above* the main diagonal are 0's. When is a square upper triangular matrix invertible? Justify your answer. ***Hint***: See Example 2.28.

2. Is it possible for a $5 \times 5$ matrix to be invertible when its columns do not span $\mathbb{R}^5$? Why or why not?

   *Ans*: No

3. If $A$ is invertible, then the columns of $A^{-1}$ are linearly independent. Explain why.

4. If $C$ is $6 \times 6$ and the equation $C\mathbf{x} = \mathbf{v}$ is consistent for every $\mathbf{v} \in \mathbb{R}^6$, is it possible that for some $\mathbf{v}$, the equation $C\mathbf{x} = \mathbf{v}$ has more than one solution? Why or why not?

   *Ans*: No

5. If the equation $G\mathbf{x} = \mathbf{y}$ has more than one solution for some $\mathbf{y} \in \mathbb{R}^n$, can the columns of $G$ span $\mathbb{R}^n$? Why or why not?

   *Ans*: No

6. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation such that $T(x_1, x_2) = (6x_1 - 8x_2,\ -5x_1 + 7x_2)$. Show that $T$ is invertible and find a formula for $T^{-1}$. ***Hint***: See Example 2.30.

# 2.5. Solving Linear Systems by Matrix Factorizations

## 2.5.1. The $LU$ Factorization/Decomposition

In industrial and business applications, the linear system is often **sparse** and to be solved for **multiple right-sides**:

$$A\mathbf{x} = \mathbf{b}_1, \quad A\mathbf{x} = \mathbf{b}_2, \quad A\mathbf{x} = \mathbf{b}_3, \; \cdots . \tag{2.7}$$

The $LU$ factorization is very useful for these common problems. (The inverse of a matrix usually becomes a full matrix.)

**Definition 2.33.** Let $A \in \mathbb{R}^{m\times n}$. The **LU factorization** of $A$ is

$$A \;=\; LU, \tag{2.8}$$

where $L \in \mathbb{R}^{m\times m}$ is a **unit lower triangular matrix** and $U \in \mathbb{R}^{m\times n}$ is **an echelon form of $A$** (**upper triangular matrix**):

$$A = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ * & 1 & 0 & 0 \\ * & * & 1 & 0 \\ * & * & * & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} \blacksquare & * & * & * & * \\ 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & \blacksquare & * \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{U}$$

**Remark 2.34.** Let $A\mathbf{x} = \mathbf{b}$ be to be solved. Then $A\mathbf{x} = LU\mathbf{x} = \mathbf{b}$, which reads

$$\begin{cases} L\mathbf{y} &= \mathbf{b}, \\ U\mathbf{x} &= \mathbf{y}. \end{cases} \tag{2.9}$$

Each algebraic equation can be solved **efficiently**, via substitutions.

**Definition 2.35.** Every **elementary row operation** can be expressed as a matrix to be left-multiplied.

- Such a matrix is called an **elementary matrix**.
- Every elementary matrix is **invertible**.

**Example 2.36.** Let $A = \begin{bmatrix} 1 & -2 & 1 \\ 2 & -2 & 3 \\ -3 & 2 & 0 \end{bmatrix}$.

a) Reduce $A$ to an echelon matrix, using **replacement operations**.

b) Express the replacement operations as **elementary matrices**.

c) Find **their inverse**.

**Solution**. **a)** **b) & c)**

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

**Algorithm 2.37.** **($LU$ Factorization Algorithm)** *The following derivation introduces an $LU$ factorization. Let $A \in \mathbb{R}^{m\times n}$. Then*

$$\begin{aligned} A &= I_m\, A \\ &= I_m E_1^{-1}\, E_1 A \\ &= I_m E_1^{-1} E_2^{-1}\, E_2 E_1 A \;=\; (E_2E_1)^{-1}\, E_2 E_1 A \\ &= \quad \vdots \\ &= I_m E_1^{-1} E_2^{-1} \cdots E_p^{-1}\, \underbrace{E_p \cdots E_2 E_1 A}_{\text{an echelon form}} \;=\; \underbrace{(E_p \cdots E_2 E_1)^{-1}}_{L}\, \underbrace{E_p \cdots E_2 E_1 A}_{U} \end{aligned} \tag{2.10}$$

*where each $E_i$ is the elementary matrix for a* **replacement** *operation.*

**Remark 2.38.** The $LU$ factorization algorithm (without pivoting) uses a sequence of **"replacement"** row operations to get

$$\begin{array}{lcl} A & \xrightarrow{E_p\cdots E_2E_1} & U = E_p \cdots E_2 E_1\, A \\ I & \xrightarrow{(E_p\cdots E_2E_1)^{-1}} & L = I\, E_1^{-1} E_2^{-1} \cdots E_p^{-1} \end{array} \tag{2.11}$$

**Example 2.39.** Find the $LU$ factorization of $A = \begin{bmatrix} 3 & -1 & 1 \\ 9 & 1 & 2 \\ -6 & 5 & -5 \end{bmatrix}$.

**Solution**. (**Forward Phase**: **Gauss Elimination**)

$$\begin{aligned} A &= \begin{bmatrix} \boxed{3} & -1 & 1 \\ 9 & 1 & 2 \\ -6 & 5 & -5 \end{bmatrix} \xrightarrow[E_2:\, R_3 \leftarrow R_3 + \mathbf{2} R_1]{E_1:\, R_2 \leftarrow R_2 - \mathbf{3} R_1} \begin{bmatrix} 3 & -1 & 1 \\ 0 & \boxed{4} & -1 \\ 0 & 3 & -3 \end{bmatrix} \\ &\xrightarrow{E_3:\, R_3 \leftarrow R_3 - \frac{\mathbf{3}}{\mathbf{4}} R_2} \begin{bmatrix} 3 & -1 & 1 \\ 0 & 4 & -1 \\ 0 & 0 & -\frac{9}{4} \end{bmatrix} = U \end{aligned} \tag{2.12}$$

Collect the "replacement" row operations and their inverse:

$$\begin{aligned} A \to U: \quad & R_2 \leftarrow R_2 - \mathbf{3} R_1 \Longrightarrow R_3 \leftarrow R_3 + \mathbf{2} R_1 \Longrightarrow R_3 \leftarrow R_3 - \tfrac{\mathbf{3}}{\mathbf{4}} R_2 \\ & E_3 E_2 E_1 A = U \Longrightarrow A = (E_3 E_2 E_1)^{-1} U \\ & L = I (E_3 E_2 E_1)^{-1} = \boldsymbol{I\, E_1^{-1} E_2^{-1} E_3^{-1}} \\ I \to L: \quad & R_2 \leftarrow R_2 + \mathbf{3} R_1 \Longleftarrow R_3 \leftarrow R_3 - \mathbf{2} R_1 \Longleftarrow R_3 \leftarrow R_3 + \tfrac{\mathbf{3}}{\mathbf{4}} R_2 \end{aligned} \tag{2.13}$$

Now we construct $\boldsymbol{L = \underline{I}\, E_1^{-1} E_2^{-1} E_3^{-1} = E_1^{-1} E_2^{-1} E_3^{-1}\, \underline{I}}$:

$$\begin{aligned} I &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{E_3^{-1}:\, R_3 \leftarrow R_3 + \frac{\mathbf{3}}{\mathbf{4}} R_2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{3}{4} & 1 \end{bmatrix} \\ &\xrightarrow[E_1^{-1}:\, R_2 \leftarrow R_2 + \mathbf{3} R_1]{E_2^{-1}:\, R_3 \leftarrow R_3 - \mathbf{2} R_1} \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -2 & \frac{3}{4} & 1 \end{bmatrix} = L. \qquad \square \end{aligned} \tag{2.14}$$

**Note**: The matrix $L$ simply collects

- all the constants of "inverse replacements",
- which are the same as **ratios to pivot values**.

**Example 2.40.** Find the $LU$ factorization of $A = \begin{bmatrix} 3 & -1 & 1 \\ 9 & 1 & 2 \\ -6 & 5 & -5 \end{bmatrix}$.

**Solution.** (**Practical Implementation**):

$$A = \begin{bmatrix} \boxed{3} & -1 & 1 \\ 9 & 1 & 2 \\ -6 & 5 & -5 \end{bmatrix} \xrightarrow[E_2:\, R_3 \leftarrow R_3 + 2R_1]{E_1:\, R_2 \leftarrow R_2 - 3R_1} \begin{bmatrix} \mathbf{3} & -1 & \mathbf{1} \\ \mathbf{3} & \boxed{4} & -1 \\ \mathbf{-2} & \mathbf{3} & -3 \end{bmatrix} \xrightarrow{E_3:\, R_3 \leftarrow R_3 - \frac{3}{4} R_2} \begin{bmatrix} \mathbf{3} & -1 & \mathbf{1} \\ \mathbf{3} & \mathbf{4} & -1 \\ \mathbf{-2} & \mathbf{\frac{3}{4}} & -\frac{9}{4} \end{bmatrix} \tag{2.15}$$

from which we can get

$$L = \begin{bmatrix} 1 & 0 & 0 \\ \mathbf{3} & 1 & 0 \\ \mathbf{-2} & \mathbf{\frac{3}{4}} & 1 \end{bmatrix}, \quad U = \begin{bmatrix} 3 & -1 & 1 \\ 0 & 4 & -1 \\ 0 & 0 & -\frac{9}{4} \end{bmatrix}. \tag{2.16}$$

**Matlab-code 2.41.** The $LU$ factorization (overwritten; without pivoting) can be implemented as

lu_nopivot_overwrite.m

```
function A = lu_nopivot_overwrite(A)

[m,n] = size(A);
for k = 1:m-1
    A(k+1:m, k) = A(k+1:n,k)/A(k,k); %ratios to pivot
    for i = k+1:m
        A(i,k+1:n) = A(i,k+1:n) - A(i,k)*A(k,k+1:n);
    end
end
```

**Self-study 2.42.** Find the $LU$ factorization of

$$A = \begin{bmatrix} 4 & 3 & -5 \\ -4 & -5 & 7 \\ 8 & 8 & -8 \end{bmatrix}$$

## 2.5.2. Solutions of Triangular Algebraic Systems

### Lower-Triangular Systems

- Consider the $n \times n$ system

$$L\,\mathbf{y} = \mathbf{b}, \tag{2.17}$$

where $L = [\ell_{ij}]$ is a nonsingular, lower-triangular matrix ($\ell_{ii} \neq 0$).
- It is easy to see how to solve this system if we write it in detail:

$$\begin{array}{lcl} \ell_{11}\,y_1 & = & b_1 \\ \ell_{21}\,y_1 + \ell_{22}\,y_2 & = & b_2 \\ \ell_{31}\,y_1 + \ell_{32}\,y_2 + \ell_{33}\,y_3 & = & b_3 \\ \quad\vdots & \vdots & \\ \ell_{n1}\,y_1 + \ell_{n2}\,y_2 + \ell_{n3}\,y_3 + \cdots + \ell_{nn}\,y_n & = & b_n \end{array} \tag{2.18}$$

- The first equation involves only the unknown $y_1$ and therefore

$$y_1 = b_1/\ell_{11}. \tag{2.19}$$

- With $y_1$ just obtained, we can determine $y_2$ from the second equation:

$$y_2 = (b_2 - \ell_{21}\,y_1)/\ell_{22}. \tag{2.20}$$

- With $y_1,\ y_2$ known, we can solve the third equation for $y_3$, and so on.

**Algorithm 2.43. Forward Substitution/Elimination**
In general, once we have $y_1,\ y_2,\ \cdots,\ y_{i-1}$, we can solve for $y_i$ using the $i$th equation of (2.18):

$$\begin{aligned} y_i &= (b_i - \ell_{i1}\,y_1 - \ell_{i2}\,y_2 - \cdots - \ell_{i,i-1}\,y_{i-1})/\ell_{ii} \\ &= \frac{1}{\ell_{ii}}\Big(b_i - \sum_{j=1}^{i-1} \ell_{ij}\,y_j\Big) \end{aligned} \tag{2.21}$$

## Upper-Triangular Systems

- Consider the system

$$U\,\mathbf{x} = \mathbf{y}, \tag{2.22}$$

  where $U = [u_{ij}] \in \mathbb{R}^{n\times n}$ is nonsingular, upper-triangular.

- Writing it out in detail, we get

$$\begin{aligned} u_{11}\,x_1 + u_{12}\,x_2 + \cdots + u_{1,n-1}\,x_{n-1} + u_{1,n}\,x_n &= y_1 \\ u_{22}\,x_2 + \cdots + u_{2,n-1}\,x_{n-1} + u_{2,n}\,x_n &= y_2 \\ \vdots &= \vdots \\ u_{n-1,n-1}\,x_{n-1} + u_{n-1,n}\,x_n &= y_{n-1} \\ u_{n,n}\,x_n &= y_n \end{aligned} \tag{2.23}$$

- It is clear that we should solve the system **from bottom to top**.

**Matlab-code 2.44. (Back Substitution)**:

```
for i=n:-1:1
   if(U(i,i)==0), error('U: singular!'); end
   x(i)=y(i)/U(i,i);
   y(1:i-1)=y(1:i-1)-U(1:i-1,i)*x(i);
end
```

(2.24)

In practice, the $LU$ factorization incorporates **partial pivoting** for an enhanced stability.

**Algorithm 2.45. Gauss Elimination with Partial Pivoting**

To get the solution of $A\mathbf{x} = \mathbf{b}$:

1. Factorize $A$ into $\boldsymbol{A = P^T LU}$ ($\Leftrightarrow \boldsymbol{PA = LU}$), where

$$\begin{aligned} P &= \text{permutation matrix} \\ L &= \text{unit lower triangular matrix (i.e., with 1's on the diagonal)} \\ U &= \text{upper-triangular matrix} \end{aligned}$$

   **Matlab**: `[L,U,P] = lu(A)`

2. Solve $A\mathbf{x} = P^T LU\mathbf{x} = \mathbf{b}$
   (a) $\mathbf{r} = P\mathbf{b}$ (permuting b)
   (b) $L\mathbf{y} = \mathbf{r}$ (forward substitution)
   (c) $U\mathbf{x} = \mathbf{y}$ (back substitution)

**Example 2.46. (Revisit of Example 2.36)** Use the $LU$ factorization to solve the linear system $A\mathbf{x} = \mathbf{b}$, where

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

**Solution**.

forward_sub.m

```
function y = forward_sub(L,b)
% function y = forward_sub(L,b)

[m,n] = size(L);
y = zeros(m,1); y(1)=b(1)/L(1,1);

for i=2:m
    y(i) = ( b(i) - L(i,1:i-1)*y(1:i-1) ) / L(i,i);
end
```

back_sub.m

```
function x = back_sub(U,y)
%function x = back_sub(U,y)

[m,n] = size(U);
x = zeros(m,1); x(m)=y(m)/U(m,m);

for i=m-1:-1:1
    x(i) = (y(i)-(U(i,i+1:end)*x(i+1:end))) / U(i,i);
end
```

lu_solve.m

```
A = [ 1 -2 1
      2 -2 3
     -3  2 0];
b = [-2 1 1]';

x = A\b

[L,U,P] = lu(A)
r = P*b;
y = forward_sub(L,r);
x = back_sub(U,y)
```

Output

```
x =
   1
   2
   1

L =
   1.00000   0.00000   0.00000
  -0.33333   1.00000   0.00000
  -0.66667   0.50000   1.00000

U =
  -3.00000   2.00000   0.00000
   0.00000  -1.33333   1.00000
   0.00000   0.00000   2.50000

P =
   0   0   1
   1   0   0
   0   1   0

x =
   1
   2
   1
```

## Exercises 2.5

1. **(Hand calculation)** Solve the equation $A\mathbf{x} = \mathbf{b}$ by using the $LU$ factorization given for $A$.
$$A = \begin{bmatrix} 4 & 3 & -5 \\ -4 & -5 & 7 \\ 8 & 6 & -8 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & 3 & -5 \\ 0 & -2 & 2 \\ 0 & 0 & 2 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 2 \\ -4 \\ 6 \end{bmatrix}.$$
Here $L\mathbf{y} = \mathbf{b}$ requires replacement operations forward, while $U\mathbf{x} = \mathbf{y}$ requires replacement and scaling operations backward.
$$Ans\text{: } \mathbf{x} = \begin{bmatrix} 1/4 \\ 2 \\ 1 \end{bmatrix}.$$

2. **(Hand calculation)** When $A$ is invertible, Matlab finds $A^{-1}$ by factoring $A = LU$, inverting $L$ and $U$, and then computing $U^{-1}L^{-1}$. You will use this method to compute the inverse of $A$ given in Exercise 1.
   (a) Find $U^{-1}$, starting from $[U \;\; I]$, reduce it to $[I \;\; U^{-1}]$.
   (b) Find $L^{-1}$, starting from $[L \;\; I]$, reduce it to $[I \;\; L^{-1}]$.
   (c) Compute $U^{-1}L^{-1}$.
$$Ans\text{: } A^{-1} = \begin{bmatrix} 1/8 & 3/8 & 1/4 \\ -3/2 & -1/2 & 1/2 \\ -1 & 0 & 1/2 \end{bmatrix}.$$

3. M[1]Use Matlab/Octave to solve the problem in Exercise 1, beginning with `[L,U,P]=lu(A)` and following steps in Algorithm 2.45.

4. Let $A = \begin{bmatrix} -1 & -5 & 8 & 4 \\ 4 & 2 & -5 & -7 \\ -2 & -4 & 7 & 5 \end{bmatrix}$.
   (a) Try to see if you can find $LU$ factorization without pivoting.
   (b) M Use Matlab/Octave to find the $LU$ factorization of $A$. Then recover $A$ from `[L,U,P]`.

5. Find $LU$ factorization (without pivoting) of $B = \begin{bmatrix} 2 & 5 & 4 & 3 \\ -4 & -9 & -6 & -2 \\ 2 & 7 & 7 & 14 \\ -6 & -14 & -10 & -2 \end{bmatrix}$.
$$Ans\text{: } U = \begin{bmatrix} 2 & 5 & 4 & 3 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

[1]All problems marked by M will have a higher credit.

## 2.8. Subspaces of $\mathbb{R}^n$

**Definition 2.47.** A **subspace** of $\mathbb{R}^n$ is any set $H$ in $\mathbb{R}^n$ that has three properties:

a) The zero vector is in $H$.

b) For each $\mathbf{u}$ and $\mathbf{v}$ in H, the sum $\mathbf{u} + \mathbf{v}$ is in $H$.

c) For each $\mathbf{u}$ in $H$ and each scalar $c$, the vector $c\mathbf{u}$ is in $H$.

That is, $H$ is **closed** under linear combinations.

**Example 2.48.**

1. A line through the origin in $\mathbb{R}^2$ is a subspace of $\mathbb{R}^2$.
2. Any plane through the origin in $\mathbb{R}^3$.

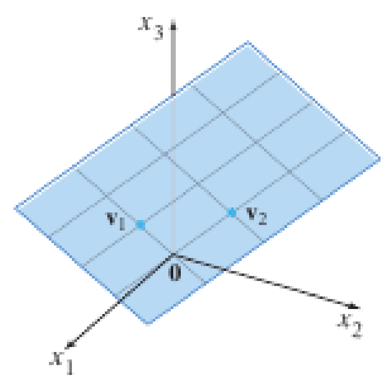



Figure 2.3: *Span*$\{\mathbf{v}_1, \mathbf{v}_2\}$ as a plane through the origin.

3. Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p \in \mathbb{R}^n$. Then *Span*$\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$ is a subspace of $\mathbb{R}^n$.

**Definition 2.49.** Let $A$ be an $m \times n$ matrix. The **column space** of $A$ is the set $Col\,A$ of all linear combinations of columns of $A$. That is, if $A = [\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_n]$, then

$$Col\,A = \{\mathbf{u} \mid \mathbf{u} = c_1\,\mathbf{a}_1 + c_2\,\mathbf{a}_2 + \cdots + c_n\,\mathbf{a}_n\}, \tag{2.25}$$

where $c_1, c_2, \cdots, c_n$ are scalars. $Col\,A$ is a subspace of $\mathbb{R}^m$.

**Example 2.50.** Let $A = \begin{bmatrix} 1 & -3 & -4 \\ -4 & 6 & -2 \\ -3 & 7 & 6 \end{bmatrix}$ and $\mathrm{b} = \begin{bmatrix} 3 \\ 3 \\ -4 \end{bmatrix}$. Determine whether b is in the column space of $A$, *Col* $A$.

**Solution**. ***Clue***: ① $\mathrm{b} \in$ *Col* $A$
⇔ ② b is a linear combination of columns of $A$
⇔ ③ $A\mathrm{x} = \mathrm{b}$ is consistent
⇔ ④ $[A \ \ \mathrm{b}]$ has a solution

**Definition 2.51.** Let $A$ be an $m \times n$ matrix. The **null space** of $A$, *Nul* $A$, is the set of all solutions of the homogeneous system $A\mathrm{x} = \mathbf{0}$.

**Theorem 2.52.** *Nul* $A$ *is a subspace of* $\mathbb{R}^n$.

**Proof**.

□

## Basis for a Subspace

**Definition 2.53.** A **basis** for a subspace $H$ in $\mathbb{R}^n$ is a set of vectors that

1. is *linearly independent*, and
2. *spans* $H$.

**Remark 2.54.**

1. $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^2$.
2. Let $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, $\cdots$, $\mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$. Then $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$ is called the **standard basis** for $\mathbb{R}^n$.

## Basis for *Nul A*

**Example 2.55.** Find a basis for the null space of the matrix

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 \\ 1 & -2 & 2 & 3 \\ 2 & -4 & 5 & 8 \end{bmatrix}.$$

**Solution**. $[A\ \ 0] \sim \begin{bmatrix} 1 & 2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

**Theorem 2.56.** *Basis for Nul* $A$ *can be obtained from the parametric vector form of solutions of* $A\mathbf{x} = \mathbf{0}$. *That is, suppose that the solutions of* $A\mathbf{x} = \mathbf{0}$ *reads*

$$\mathbf{x} = x_1\,\mathbf{u}_1 + x_2\,\mathbf{u}_2 + \cdots + x_k\,\mathbf{u}_k,$$

*where* $x_1, x_2, \cdots, x_k$ *correspond to free variables. Then, a basis for Nul* $A$ *is* $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k\}$.

### Basis for *Col A*

**Example 2.57.** Find a basis for the column space of the matrix

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

**Solution**. Observation: $\mathbf{b}_3 = -3\mathbf{b}_1 + 2\mathbf{b}_2$ and $\mathbf{b}_4 = 5\mathbf{b}_1 - \mathbf{b}_2$.

**Theorem 2.58.** *In general, non-pivot columns are linear combinations of pivot columns. Thus the pivot columns of a matrix* $A$ *form a basis for Col* $A$.

**Example 2.59.** Matrix $A$ and its echelon form is given. Find a basis for *Col* $A$ and a basis for *Nul* $A$.

$$A = \begin{bmatrix} 3 & -6 & 9 & 0 \\ 2 & -4 & 7 & 2 \\ 3 & -6 & 6 & -6 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 3 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

**Solution**.

*Ans*: $\mathcal{B}_{Col\,A} = \{\mathbf{a}_1, \mathbf{a}_3\},\ \mathcal{B}_{Nul\,A} = \{[2, 1, 0, 0]^T, [6, 0, -2, 1]^T\}$.

**True-or-False 2.60.**

a. If $\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\ \mathbf{v}_p$ are in $\mathbb{R}^n$, then $Span\{\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\ \mathbf{v}_p\} = Col\,[\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_p]$.

b. The columns of an invertible $n \times n$ matrix form a basis for $\mathbb{R}^n$.

c. Row operations do not affect linear dependence relations among the columns of a matrix.

d. The column space of a matrix $A$ is the set of solutions of $A\mathrm{x} = \mathbf{b}$.

e. If $B$ is an echelon form of a matrix $A$, then the pivot columns of $B$ form a basis for *Col* $A$.

**Solution**.

*Ans*: T,T,T,F,F

## Exercises 2.8

1. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \\ 4 \\ 3 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 4 \\ -7 \\ 9 \\ 7 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 5 \\ -8 \\ 6 \\ 5 \end{bmatrix}$, and $\mathbf{u} = \begin{bmatrix} -4 \\ 10 \\ -7 \\ -5 \end{bmatrix}$. Determine if $\mathbf{u}$ is in the subspace of $\mathbb{R}^4$ generated by $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

   *Ans*: No

2. Let $\mathbf{v}_1 = \begin{bmatrix} -3 \\ 0 \\ 6 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -2 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 0 \\ -6 \\ 3 \end{bmatrix}$, and $\mathbf{p} = \begin{bmatrix} 1 \\ 14 \\ -9 \end{bmatrix}$. Determine if $\mathbf{p}$ is in *Col* $A$, where $A = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3]$.

   *Ans*: Yes

3. Give integers $p$ and $q$ such that *Nul* $A$ is a subspace of $\mathbb{R}^p$ and *Col* $A$ is a subspace of $\mathbb{R}^q$.

$$A = \begin{bmatrix} 3 & 2 & 1 & -5 \\ -9 & -4 & 1 & 7 \\ 9 & 2 & -5 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 7 \\ -5 & -1 & 0 \\ 2 & 7 & 11 \end{bmatrix}$$

4. Determine which sets are bases for $\mathbb{R}^3$. Justify each answer.

   a) $\begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix}, \begin{bmatrix} 5 \\ -7 \\ 4 \end{bmatrix}, \begin{bmatrix} 6 \\ 3 \\ 5 \end{bmatrix}$ b) $\begin{bmatrix} 1 \\ -6 \\ -7 \end{bmatrix}, \begin{bmatrix} 3 \\ -4 \\ 7 \end{bmatrix}, \begin{bmatrix} -2 \\ 7 \\ 5 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix}$

   *Ans*: a) Yes

5. Matrix $A$ and its echelon form is given. Find a basis for *Col* $A$ and a basis for *Nul* $A$.

$$A = \begin{bmatrix} 1 & 4 & 8 & -3 & -7 \\ -1 & 2 & 7 & 3 & 4 \\ -2 & 2 & 9 & 5 & 5 \\ 3 & 6 & 9 & -5 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & 4 & 8 & 0 & 5 \\ 0 & 2 & 5 & 0 & -1 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

   ***Hint***: For a basis for *Col* $A$, you can just recognize pivot columns, while you should find the solutions of $A\mathbf{x} = \mathbf{0}$ for *Nul* $A$.

6. a) Suppose $F$ is a $5 \times 5$ matrix whose column space is not equal to $\mathbb{R}^5$. What can you say about *Nul* $F$?
   b) If $R$ is a $6 \times 6$ matrix and *Nul* $R$ is not the zero subspace, what can you say about *Col* $R$?
   c) If $Q$ is a $4 \times 4$ matrix and *Col* $Q = \mathbb{R}^4$, what can you say about solutions of equations of the form $Q\mathbf{x} = \mathbf{b}$ for $\mathbf{b}$ in $\mathbb{R}^4$?

   *Ans*: b) *Col* $R \neq \mathbb{R}^6$. Why? c) It has always a unique solution

# 2.9. Dimension and Rank

## 2.9.1. Coordinate Systems

**Remark 2.61. The main reason for selecting a basis** for a subspace $H$ (instead of merely a spanning set) is that each vector in $H$ can be written in **only one way** as a linear combination of the basis vectors.

**Example 2.62.** Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_p\}$ be a basis of $H$ and

$$\mathbf{x} = c_1\,\mathbf{b}_1 + c_2\,\mathbf{b}_2 + \cdots + c_p\,\mathbf{b}_p; \quad \mathbf{x} = d_1\,\mathbf{b}_1 + d_2\,\mathbf{b}_2 + \cdots + d_p\,\mathbf{b}_p, \quad \mathbf{x} \in H.$$

Show that $c_1 = d_1,\ c_2 = d_2,\ \cdots,\ c_p = d_p$.

**Solution**. ***Hint***: A property of a basis is that basis vectors are linearly independent

**Remark 2.63.** For example, if a vector $\mathbf{x} \in \mathbb{R}^3$ is expressed as

$$\mathbf{x} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3, \tag{2.26}$$

then $[x_1, x_2, x_3]^T$ is called the **coordinate vector** of x.

**Definition 2.64.** Suppose the set $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_p\}$ is a basis for a subspace $H$. For each $\mathbf{x} \in H$, the **coordinates of x relative to the basis** $\mathcal{B}$ are the weights $c_1, c_2, \cdots, c_p$ such that $\mathbf{x} = c_1\,\mathbf{b}_1 + c_2\,\mathbf{b}_2 + \cdots + c_p\,\mathbf{b}_p$, and the vector in $\mathbb{R}^p$

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

is called the **coordinate vector of x (relative to $\mathcal{B}$)** or the **$\mathcal{B}$-coordinate vector** of x.

**Example 2.65.** Let $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 3 \\ 12 \\ 7 \end{bmatrix}$, and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$.

Then $\mathcal{B}$ is a basis for $H = Span\{\mathbf{v}_1, \mathbf{v}_2\}$, because $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent. Determine if x is in $H$, and if it is, find the coordinate vector of x relative to $\mathcal{B}$.

**Solution**.

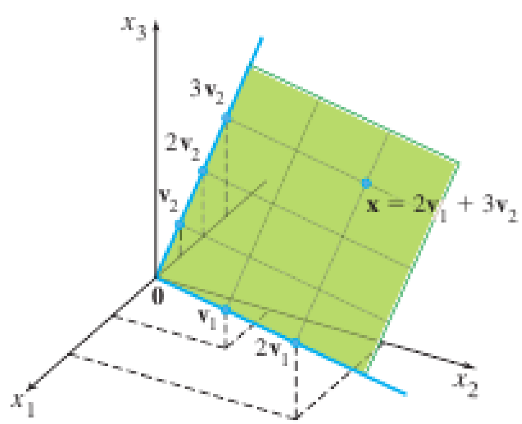



Figure 2.4: A coordinate system on a plane $H \subset \mathbb{R}^3$.

**Remark 2.66.** The grid on the plane in Figure 2.4 makes $H$ “look" like $\mathbb{R}^2$. The correspondence $\mathbf{x} \mapsto [\mathbf{x}]_{\mathcal{B}}$ is a one-to-one correspondence between $H$ and $\mathbb{R}^2$ that preserves linear combinations. We call such a correspondence an **isomorphism**, and we say that $H$ is **isomorphic** to $\mathbb{R}^2$.

## 2.9.2. Dimension of a Subspace

**Definition 2.67.** The **dimension** of a nonzero subspace $H$ ($\mathit{dim}\, H$) is the number of vectors in any basis for $H$. The dimension of zero subspace $\{0\}$ is defined to be zero.

**Example 2.68.**

a) $\mathit{dim}\,\mathbb{R}^n = n$.
b) Let $H$ be as in Example 2.65. What is $\mathit{dim}\, H$?
c) $G = \mathit{Span}\{\mathbf{u}\}$. What is $\mathit{dim}\, G$?

**Solution**.

**Remark 2.69.**

1) **Dimension of** $\mathit{Col}\, A$:

$$\mathit{dim}\,\mathit{Col}\, A = \text{The number of pivot columns in } A$$

which is called the **rank** of $A$, $\mathit{rank}\, A$.

2) **Dimension of** $\mathit{Nul}\, A$:

$$\begin{aligned} \mathit{dim}\,\mathit{Nul}\, A &= \text{The number of free variables in } A \\ &= \text{The number of non-pivot columns in } A \end{aligned}$$

**Theorem 2.70. (Rank Theorem)** *Let* $A \in \mathbb{R}^{m\times n}$. *Then*

$$\begin{aligned} \mathit{dim}\,\mathit{Col}\, A + \mathit{dim}\,\mathit{Nul}\, A &= \mathit{rank}\, A + \mathit{nullity}\, A = n \\ &= \text{(the number of columns in } A\text{)} \end{aligned}$$

*Here, "dim Nul $A$" is called the* ***nullity*** *of $A$: nullity $A$*

**Example 2.71.** A matrix and its echelon form are given. Find the bases for *Col* $A$ and *Nul* $A$ and also state the dimensions of these subspaces.

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

**Solution**.

**Example 2.72.** Find a basis for the subspace spanned by the given vectors. What is the dimension of the subspace?

$$\begin{bmatrix} 1 \\ -1 \\ -2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ -1 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 4 \\ -7 \\ 7 \end{bmatrix}, \begin{bmatrix} 3 \\ -7 \\ 6 \\ -9 \end{bmatrix}$$

**Solution**.

**Theorem 2.73. (The Basis Theorem)** *Let $H$ be a $p$-dimensional subspace of $\mathbb{R}^n$. Then*

a) *Any* linearly independent *set of exactly $p$ elements in $H$ is automatically a basis for $H$*

b) *Any set of $p$ elements of $H$ that* spans *$H$ is automatically a basis for $H$.*

**Theorem 2.74. (Invertible Matrix Theorem; continued from Theorem 2.25, p.82)**

*Let $A$ be an $n \times n$ square matrix. Then the following are equivalent.*

m. *The columns of $A$ form a basis of $\mathbb{R}^n$*

n. *Col $A = \mathbb{R}^n$*

o. *dim Col $A = n$*

p. *rank $A = n$*

q. *Nul $A = \{\mathbf{0}\}$*

r. *dim Nul $A = 0$*

**Example 2.75.**

a) If the rank of a $9 \times 8$ matrix $A$ is 7, what is the dimension of solution space of $A\mathbf{x} = \mathbf{0}$?

b) If A is a $4 \times 3$ matrix and the mapping $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one, then what is *dim Nul* $A$?

**Solution**.

**True-or-False 2.76.**

a. Each line in $\mathbb{R}^n$ is a one-dimensional subspace of $\mathbb{R}^n$.

b. The dimension of *Col* $A$ is the number of pivot columns of $A$.

c. The dimensions of *Col* $A$ and *Nul* $A$ add up to the number of columns of $A$.

d. If $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$ is a basis for a subspace $H$ of $\mathbb{R}^n$, then the correspondence $\mathbf{x} \mapsto [\mathbf{x}]_{\mathcal{B}}$ makes $H$ look and act the same as $\mathbb{R}^p$.

***Hint***: See Remark 2.66.

e. The dimension of *Nul* $A$ is the number of variables in the equation $A\mathbf{x} = \mathbf{0}$.

**Solution**.

*Ans*: F,T,T(Rank Theorem),T,F

## Exercises 2.9

1. A matrix and its echelon form is given. Find the bases for *Col* $A$ and *Nul* $A$, and then state the dimensions of these subspaces.

$$A = \begin{bmatrix} 1 & -3 & 2 & -4 \\ 3 & 9 & -1 & 5 \\ 2 & -6 & 4 & -3 \\ 4 & 12 & 2 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 2 & -4 \\ 0 & 0 & 5 & -7 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

2. Use the Rank Theorem to justify each answer, or perform construction.

   (a) If the subspace of all solutions of $A\mathbf{x} = \mathbf{0}$ has a basis consisting of three vectors and if $A$ is a $5 \times 7$ matrix, what is the rank of $A$?

   *Ans*: 4

   (b) What is the rank of a $4 \times 5$ matrix whose null space is three-dimensional?

   (c) If the rank of a $7 \times 6$ matrix $A$ is 4, what is the dimension of the solution space of $A\mathbf{x} = \mathbf{0}$?

   (d) Construct a $4 \times 3$ matrix with rank 1.

CHAPTER 3

# Determinants

In linear algebra, the **determinant** is a scalar value that can be computed for a square matrix. Geometrically, it can be viewed as a **volume scaling factor** of the linear transformation described by the matrix, $\mathbf{x} \mapsto A\mathbf{x}$.

For example, let $A \in \mathbb{R}^{2\times 2}$ and

$$\mathbf{u}_1 = A\begin{bmatrix}1\\0\end{bmatrix}, \quad \mathbf{u}_2 = A\begin{bmatrix}0\\1\end{bmatrix}$$

Then the determinant of $A$ (in modulus) is the same as the area of the parallelogram generated by $\mathbf{u}_1$ and $\mathbf{u}_2$.

In this chapter, you will study the determinant and its properties.

**Contents of Chapter 3**

3.1. Introduction to Determinants . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 110
3.2. Properties of Determinants . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 115

# 3.1. Introduction to Determinants

**Definition 3.1.** Let $A$ be an $n \times n$ square matrix. Then **determinant** is a scalar value denoted by $\mathit{det}\, A$ or $|A|$.

1) Let $A = [a] \in \mathbb{R}^{1 \times 1}$. Then $\mathit{det}\, A = a$.

2) Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbb{R}^{2 \times 2}$. Then $\mathit{det}\, A = ad - bc$.

**Example 3.2.** Let $A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$. Consider a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(\mathbf{x}) = A\mathbf{x}$.

1) Find the determinant of $A$.
2) Determine the image of a rectangle $R = [0, 2] \times [0, 1]$ under $T$.
3) Find the area of the image.
4) Figure out how $\mathit{det}\, A$, the area of the rectangle ($= 2$), and the area of the image are related.

**Solution**.

*Ans*: 3) 12

**Note**: The determinant can be viewed as a **volume scaling factor**.

**Definition 3.3.** Let $A_{ij}$ be the *submatrix* of $A$ obtained by deleting row $i$ and column $j$ of $A$. Then the $(i,j)$**-cofactor** of $A = [a_{ij}]$ is the scalar $C_{ij}$, given by

$$C_{ij} = (-1)^{i+j} \textbf{det}\, A_{ij}. \tag{3.1}$$

**Definition 3.4.** For $n \geq 2$, the **determinant** of an $n \times n$ matrix $A = [a_{ij}]$ is given by the following formulas:

1. The *cofactor expansion* across the first row:

$$\textbf{det}\, A \;=\; a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n} \tag{3.2}$$

2. The *cofactor expansion* across the row $i$:

$$\textbf{det}\, A \;=\; a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} \tag{3.3}$$

3. The *cofactor expansion* down the column $j$:

$$\textbf{det}\, A \;=\; a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj} \tag{3.4}$$

**Example 3.5.** Find the determinant of $A = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}$, by expanding across the first row and down column 3.

**Solution**.

*Ans*: $-2$

**Example 3.6.** Compute the determinant of $A = \begin{bmatrix} 1 & -2 & 5 & 2 \\ 2 & -6 & -7 & 5 \\ 0 & 0 & 3 & 0 \\ 5 & 0 & 4 & 4 \end{bmatrix}$ by a cofactor expansion.

**Solution**.

**Note**: If $A$ is a triangular (upper or lower) matrix, then *det* $A$ is the product of entries on the main diagonal of $A$.

**Example 3.7.** Compute the determinant of $A = \begin{bmatrix} 1 & -2 & 5 & 2 \\ 0 & -6 & -7 & 5 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$.

**Solution**.

**True-or-False 3.8.**

a. An $n \times n$ determinant is defined by determinants of $(n-1) \times (n-1)$ submatrices.

b. The $(i, j)$-cofactor of a matrix $A$ is the matrix $A_{ij}$ obtained by deleting from $A$ its $i$-th row and $j$-th column.

c. The cofactor expansion of *det* $A$ down a column is equal to the cofactor expansion along a row.

d. The determinant of a triangular matrix is the sum of the entries on the main diagonal.

**Solution**.

*Ans*: T,F,T,F

## Exercises 3.1

1. Compute the determinants in using a cofactor expansion across the *first row* and down the *second column*.

a) $\begin{bmatrix} 3 & 0 & 4 \\ 2 & 3 & 2 \\ 0 & 5 & -1 \end{bmatrix}$ b) $\begin{bmatrix} 2 & 3 & -3 \\ 4 & 0 & 3 \\ 6 & 1 & 5 \end{bmatrix}$

*Ans*: a) $1$, b) $-24$

2. Compute the determinants by cofactor expansions. At each step, choose a row or column that involves the least amount of computation.

a) $\begin{bmatrix} 3 & 5 & -6 & 4 \\ 0 & -2 & 3 & -3 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 3 \end{bmatrix}$ b) $\begin{bmatrix} 4 & 0 & -7 & 3 & -5 \\ 0 & 0 & 2 & 0 & 0 \\ 7 & 3 & -6 & 4 & -8 \\ 5 & 0 & 5 & 2 & -3 \\ 0 & 0 & 9 & -1 & 2 \end{bmatrix}$

*Ans*: a) $-18$, b) $6$

3. The expansion of a $3 \times 3$ determinant can be remembered by the following device. Write a second copy of the first two columns to the right of the matrix, and compute the determinant by multiplying entries on six diagonals:

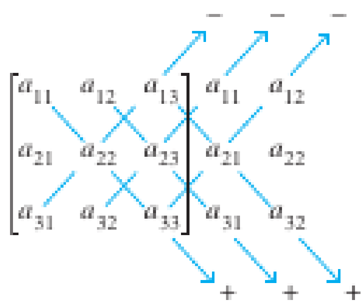



Figure 3.1

Then, add the downward diagonal products and subtract the upward products. Use this method to compute the determinants for the matrices in Exercise 1. **Warning**: *This trick does not generalize in any reasonable way to* $4 \times 4$ *or larger matrices.*

4. Explore the effect of an elementary row operation on the determinant of a matrix. In each case, state the row operation and describe how it affects the determinant.

a) $\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \begin{bmatrix} c & d \\ a & b \end{bmatrix}$

b) $\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \begin{bmatrix} a+kc & b+kd \\ c & d \end{bmatrix}$

*Ans*: b) Replacement does not change the determinant

5. Let $A = \begin{bmatrix} 3 & 1 \\ 4 & 2 \end{bmatrix}$. Write $5A$. Is $\det(5A) = 5\det A$?

# 3.2. Properties of Determinants

## Determinants under Elementary Row Operations

**Theorem 3.9.** *Let $A$ be an $n \times n$ square matrix.*

a) **(Replacement)**: *If $B$ is obtained from $A$ by a row replacement, then $\det B = \det A$.*

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

b) **(Interchange)**: *If two rows of $A$ are interchanged to form $B$, then $\det B = -\det A$.*

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

c) **(Scaling)**: *If one row of $A$ is multiplied by k ($\neq 0$), then $\det B = k \cdot \det A$.*

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

**Example 3.10.** Compute $\det A$, where $A = \begin{bmatrix} 1 & -4 & 2 \\ -1 & 7 & 0 \\ -2 & 8 & -9 \end{bmatrix}$, after applying a couple of steps of replacement operations.

**Solution**.

*Ans*: 15

**Theorem 3.11. Invertible Matrix Theorem (p.82)**

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

**Claim 3.12.** Let $A$ and $B$ be $n \times n$ matrices.

a) $\det A^T = \det A$.

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}, \quad A^T = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}$$

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

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 4 & 2 \end{bmatrix}; \text{then } AB = \begin{bmatrix} 13 & 7 \\ 6 & 4 \end{bmatrix}.$$

c) If $A$ is invertible, then $\det A^{-1} = \dfrac{1}{\det A}$. ($\because \det I_n = 1$.)

**Example 3.13.** Suppose the sequence $5 \times 5$ matrices $A$, $A_1$, $A_2$, and $A_3$ are related by following elementary row operations:

$$A \xrightarrow{R_2 \leftarrow R_2 - 3R_1} A_1 \xrightarrow{R_3 \leftarrow (1/5)\, R_3} A_2 \xrightarrow{R_4 \leftrightarrow R_5} A_3$$

Find $\det A$, if $A_3 = \begin{bmatrix} 1 & 2 & 3 & 4 & 1 \\ 0 & -2 & 1 & -1 & 1 \\ 0 & 0 & 3 & 0 & 1 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

**Solution**.

*Ans*: $-30$

## A Linearity Property of the Determinant Function

**Note**: Let $A \in \mathbb{R}^{n\times n}$, $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_j \ \cdots \ \mathbf{a}_n]$. Suppose that the $j$-th column of $A$ is allowed to vary, and write

$$A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_{j-1} \ \mathbf{x} \ \mathbf{a}_{j+1} \ \cdots \ \mathbf{a}_n].$$

Define a transformation $T$ from $\mathbb{R}^n$ to $\mathbb{R}$ by

$$T(\mathbf{x}) = \textbf{\textit{det}}\,[\mathbf{a}_1 \ \cdots \ \mathbf{a}_{j-1} \ \mathbf{x} \ \mathbf{a}_{j+1} \ \cdots \ \mathbf{a}_n]. \tag{3.5}$$

Then,

$$\begin{aligned} T(c\mathbf{x}) &= cT(\mathbf{x}) \\ T(\mathbf{u}+\mathbf{v}) &= T(\mathbf{u}) + T(\mathbf{v}) \end{aligned} \tag{3.6}$$

This (multi-) linearity property of the determinant turns out to have many useful consequences that are studied in more advanced courses.

**True-or-False 3.14.**

a. If the columns of $A$ are linearly dependent, then ***det*** $A = 0$.

b. ***det*** $(A + B) =$ ***det*** $A +$ ***det*** $B$.

c. If three row interchanges are made in succession, then the new determinant equals the old determinant.

d. The determinant of $A$ is the product of the diagonal entries in $A$.

e. If ***det*** $A$ is zero, then two rows or two columns are the same, or a row or a column is zero.

**Solution**.

*Ans*: T,F,F,F,F

## Exercises 3.2

1. Find the determinant by row reduction to echelon form.
$$\begin{bmatrix} 1 & 3 & 0 & 2 \\ -2 & -5 & 7 & 4 \\ 3 & 5 & 2 & 1 \\ 1 & -1 & 2 & -3 \end{bmatrix}$$
*Ans*: 0

2. Use determinants to find out if the matrix is invertible.
$$\begin{bmatrix} 2 & 0 & 0 & 6 \\ 1 & -7 & -5 & 0 \\ 3 & 8 & 6 & 0 \\ 0 & 7 & 5 & 4 \end{bmatrix}$$
*Ans*: Invertible

3. Use determinants to decide if the set of vectors is linearly independent.
$$\begin{bmatrix} 7 \\ -4 \\ -6 \end{bmatrix}, \begin{bmatrix} -8 \\ 5 \\ 7 \end{bmatrix}, \begin{bmatrix} 7 \\ 0 \\ -5 \end{bmatrix}$$
*Ans*: linearly independent

4. Compute $\det B^6$, where $B = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix}$.
*Ans*: 64

5. Show or answer with justification.
   a) Let $A$ and $P$ be square matrices, with $P$ invertible. Show that $\det(PAP^{-1}) = \det A$.
   b) Suppose that $A$ is a square matrix such that $\det A^3 = 0$. Can $A$ can be invertible?
   *Ans*: No
   c) Let $U$ be a square matrix such that $U^T U = I$. Show that $\det U = \pm 1$.

6. Compute $AB$ and verify that $\det AB = \det A \cdot \det B$.
$$A = \begin{bmatrix} 3 & 0 \\ 6 & 1 \end{bmatrix}, B = \begin{bmatrix} 2 & 0 \\ 5 & 4 \end{bmatrix}$$

CHAPTER 4

# Vector Spaces

A **vector space** (also called a *linear space*) is a nonempty set of objects, called **vectors**, which is closed under **two operations**:

- **addition** and
- **scalar multiplication**.

In this chapter, we will study basic concepts of such *general* vector spaces and their subspaces.

**Contents of Chapter 4**

4.1. Vector Spaces and Subspaces . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 120
4.2. Null Spaces, Column Spaces, and Linear Transformations . . . . . . . . . . . . . . . . . 125

# 4.1. Vector Spaces and Subspaces

**Definition 4.1.** A **vector space** is a nonempty set $V$ of objects, called **vectors**, on which are defined **two operations**, called **addition** and **multiplication by scalars** (real numbers), subject to the ten axioms (or rules) listed below. The axioms must hold for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and for all scalars $c$ and $d$.

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

**Example 4.2.** Examples of Vector Spaces:

a) $\mathbb{R}^n$, $n \geq 1$, are the premier examples of vector spaces.

b) Let $\mathbb{P}_n = \{\mathbf{p}(t) = a_0 + a_1t^1 + \cdots + a_nt^n\}$, $n \geq 1$. For $\mathbf{p}(t) = a_0 + a_1t^1 + \cdots + a_nt^n$ and $\mathbf{q}(t) = b_0 + b_1t^1 + \cdots + b_nt^n$, define

$$\begin{aligned} (\mathbf{p} + \mathbf{q})(t) &= \mathbf{p}(t) + \mathbf{q}(t) = (a_0 + b_0) + (a_1 + b_1)t^1 + \cdots + (a_n + b_n)t^n \\ (c\mathbf{p})(t) &= c\mathbf{p}(t) = ca_0 + ca_1t^1 + \cdots + ca_nt^n \end{aligned}$$

Then $\mathbb{P}_n$ is a vector space, with the usual polynomial addition and scalar multiplication.

c) Let $V = \{$all real-valued functions defined on a set $D\}$. Then, $V$ is a vector space, with the usual function addition and scalar multiplication.

**Definition 4.3.** A **subspace** of a vector space $V$ is a subset $H$ of $V$ that has three properties:

a) $\mathbf{0} \in H$, where $\mathbf{0}$ is the zero vector of $V$

b) $H$ is closed under vector addition: for each $\mathbf{u}, \mathbf{v} \in H$, $\mathbf{u} + \mathbf{v} \in H$

c) $H$ is closed under scalar multiplication: for each $\mathbf{u} \in H$ and each scalar $c$, $c\mathbf{u} \in H$

**Example 4.4.** Examples of Subspaces:

a) $H = \{\mathbf{0}\}$: the **zero subspace**

b) Let $\mathbb{P} = \{$all polynomials with real coefficients defined on $\mathbb{R}\}$. Then, $\mathbb{P}$ is a subspace of the space $\{$all real-valued functions defined on $\mathbb{R}\}$.

c) The vector space $\mathbb{R}^2$ is not a subspace of $\mathbb{R}^3$, because $\mathbb{R}^2 \not\subset \mathbb{R}^3$.

d) Let $H = \left\{ \begin{bmatrix} s \\ t \\ 0 \end{bmatrix} : s, t \in \mathbb{R} \right\}$. Then $H$ is a subspace of $\mathbb{R}^3$.

**Example 4.5.** Determine if the given set is a subspace of $\mathbb{P}_n$ for an appropriate value of $n$.

a) $\{at^2 \mid a \in \mathbb{R}\}$

b) $\{\mathbf{p} \in \mathbb{P}_3$ with integer coefficients$\}$

c) $\{a + t^2 \mid a \in \mathbb{R}\}$

d) $\{\mathbf{p} \in \mathbb{P}_n \mid \mathbf{p}(0) = 0\}$

**Solution**.

*Ans*: a) Yes, b) No, c) No, d) Yes

## A Subspace Spanned by a Set

**Example 4.6.** Let $\mathbf{v}_1, \mathbf{v}_2 \in V$, a vector space. Prove that $H = \mathit{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ is a subspace of $V$.

**Solution**.

a) $\mathbf{0} \in H$, where $\mathbf{0}$ is the zero vector of $V$

b) For each $\mathbf{u}, \mathbf{w} \in H$, $\mathbf{u} + \mathbf{w} \in H$

c) For each $\mathbf{u} \in H$ and each scalar $c$, $c\mathbf{u} \in H$

**Theorem 4.7.** *If $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p$ are in a vector space $V$, then $\mathit{Span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$ is a subspace of $V$.*

**Example 4.8.** Let $H = \{(a - 3b, b - a, a, b) \mid a, b \in \mathbb{R}\}$. Show that $H$ is a subspace of $\mathbb{R}^4$.

**Solution**.

**Self-study 4.9.** Let $H$ and $K$ be subspaces of $V$. Define the **sum** of $H$ and $K$ as

$$H + K = \{\mathbf{u} + \mathbf{v} \mid \mathbf{u} \in H, \, \mathbf{v} \in K\}.$$

Prove that $H + K$ is a subspace of $V$.
**Solution**.

**True-or-False 4.10.**

a. A vector is an arrow in three-dimensional space.

b. A subspace is also a vector space.

c. $\mathbb{R}^2$ is a subspace of $\mathbb{R}^3$.

d. A subset $H$ of a vector space $V$ is a subspace of $V$ if the following conditions are satisfied: (i) the zero vector of $V$ is in $H$, (ii) $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u} + \mathbf{v}$ are in $H$, and (iii) $c$ is a scalar and $c\mathbf{u}$ is in $H$.

**Solution**.

*Ans*: F,T,F,F (In (ii), there is no statement that $\mathbf{u}$ and $\mathbf{v}$ represent all possible elements of $H$)

## Exercises 4.1

*You may use Definition 4.3, p. 121, or Theorem 4.7, p. 122.*

1. Let $V$ be the first quadrant in the $xy$-plane; that is, let $V = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \geq 0, y \geq 0 \right\}$.

   a) If $\mathbf{u}$ and $\mathbf{v}$ are in $V$, is $\mathbf{u} + \mathbf{v}$ in $V$? Why?

   b) Find a specific vector $\mathbf{u}$ in $V$ and a specific scalar $c$ such that $c\mathbf{u}$ is *not* in $V$. (This is enough to show that $V$ is *not* a vector space.)

2. Let $H$ be the set of all vectors of the form $\begin{bmatrix} s \\ 3s \\ 2s \end{bmatrix}$.

   a) Find a vector $\mathbf{v}$ in $\mathbb{R}^3$ such that $H = \mathit{Span}\{\mathbf{v}\}$.

   b) Why does this show that $H$ is a subspace of $\mathbb{R}^3$?

   *Ans*: a) $\mathbf{v} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$

3. Let $W$ be the set of all vectors of the form $\begin{bmatrix} 5b + 2c \\ b \\ c \end{bmatrix}$.

   a) Find vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^3$ such that $W = \mathit{Span}\{\mathbf{u}, \mathbf{v}\}$.

   b) Why does this show that $W$ is a subspace of $\mathbb{R}^3$?

4. Let $W$ be the set of all vectors of the form $\begin{bmatrix} s + 3t \\ s - t \\ 2s - t \\ 4t \end{bmatrix}$. Show that $W$ is a subspace of $\mathbb{R}^4$.

For fixed positive integers $m$ and $n$, the set $M_{m \times n}$ of all $m \times n$ matrices is a vector space, under the usual operations of addition of matrices and multiplication by real scalars.

5. Determine if the set $H$ of all matrices of the form $\begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$ is a subspace of $M_{2 \times 2}$.

# 4.2. Null Spaces, Column Spaces, and Linear Transformations

**Note**: In applications of linear algebra, **subspaces of $\mathbb{R}^n$** usually arise in one of two ways:

a) as **the set of all solutions** to a homogeneous linear system, or

b) as **the set of all linear combinations** of certain vectors.

In this section, we study these two descriptions of subspaces.

- The section looks like a duplication of *Section 2.8. Subspaces of* $\mathbb{R}^n$.
- It aims **to practice to use the concept of a subspace.**

## The Null Space of a Matrix

**Definition 4.11.** The **null space** of an $m \times n$ matrix $A$, written as *Nul* $A$, is the set of all solutions of the homogeneous equation $A\mathbf{x} = \mathbf{0}$. In set notation,

$$\text{Nul}\, A = \{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n \text{ and } A\mathbf{x} = \mathbf{0}\}$$

**Example 4.12.** Let $A = \begin{bmatrix} 1 & -3 & -2 \\ -5 & 9 & 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix}$. Determine if $\mathbf{v}$ belongs to the null space of $A$.

**Solution**.

**Theorem 4.13.** *The null space of an $m \times n$ matrix $A$ is a subspace of $\mathbb{R}^n$. Equivalently, the set of all solutions to a system $A\mathbf{x} = \mathbf{0}$ of $m$ homogeneous linear equations in $n$ unknowns is a subspace of $\mathbb{R}^n$.*

**Proof**.

**Example 4.14.** Let $H$ be the set of all vectors in $\mathbb{R}^4$, whose coordinates $a, b, c, d$ satisfy the equations $\begin{cases} a - 2b + 5c &= d \\ c - a &= b \end{cases}$. Show that H is a subspace of $\mathbb{R}^4$.

**Solution**. Rewrite the above equations as

$$\begin{cases} a - 2b + 5c - d &= 0 \\ -a - b + c &= 0 \end{cases}$$

Then $\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$ is the solution of $\begin{bmatrix} 1 & -2 & 5 & -1 \\ -1 & -1 & 1 & 0 \end{bmatrix} \mathbf{x} = \mathbf{0}$. Thus **the collection of these solutions** is a subspace.

## An Explicit Description of *Nul A*

**Example 4.15.** Find a spanning set for the null space of the matrix
$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$

**Solution**. $[A \ \ 0] \sim \begin{bmatrix} 1 & -2 & 0 & -1 & 3 & 0 \\ 0 & 0 & 1 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ **(R.E.F)**

## The Column Space of a Matrix

**Definition 4.16.** The **column space** of an $m \times n$ matrix $A$, written as *Col* $A$, is the set of all linear combinations of the columns of $A$. If $A = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_n]$, then

$$\mathit{Col}\, A = \mathit{Span}\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n\} \tag{4.1}$$

**Theorem 4.17.** *The column space of an $m \times n$ matrix $A$ is a subspace of $\mathbb{R}^m$.*

**Example 4.18.** Find a matrix $A$ such that $W = \mathit{Col}\, A$.

$$W = \left\{ \begin{bmatrix} 6a - b \\ a + b \\ -7a \end{bmatrix} : a, b \in \mathbb{R} \right\}$$

**Solution**.

**Remark 4.19.** The column space of an $m \times n$ matrix $A$ is all of $\mathbb{R}^m$ if and only if the equation $A\mathrm{x} = \mathrm{b}$ has a solution for each $\mathrm{b} \in \mathbb{R}^m$.

## The Contrast Between *Nul A* and *Col A*

**Remark 4.20.** Let $A \in \mathbb{R}^{m\times n}$.

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

## Kernel (Null Space) and Range of a Linear Transformation

**Definition 4.21.** A **linear transformation** $T$ from a vector space $V$ into a vector space $W$ is a rule that assigns to each vector $\mathbf{v} \in V$ a unique vector $T(\mathbf{x}) \in W$, such that

$$\begin{array}{lll} \text{(i)} & T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) & \text{for all } \mathbf{u}, \mathbf{v} \in V\text{, and} \\ \text{(ii)} & T(c\mathbf{u}) = cT(\mathbf{u}) & \text{for all } \mathbf{u} \in V \text{ and scalar } c \end{array} \tag{4.2}$$

**Example 4.22.** Let $T : V \to W$ be a linear transformation from a vector space $V$ into a vector space $W$. Prove that the range of $T$ is a subspace of $W$.

***Hint***: Typical elements of the range have the form $T(\mathbf{u})$ and $T(\mathbf{v})$ for $\mathbf{u}, \mathbf{v} \in V$. See Definition 4.3, p. 121; you should check if the three conditions are satisfied.

**Solution**.

**True-or-False 4.23.**

a. The column space of A is the range of the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

b. The kernel of a linear transformation is a vector space.

c. *Col* $A$ is the set of all vectors that can be written as $A\mathbf{x}$ for some $\mathbf{x}$. That is, *Col* $A = \{\mathbf{b} \mid \mathbf{b} = A\mathbf{x},\ \text{for } \mathbf{x} \in \mathbb{R}^n\}$.

d. *Nul* $A$ is the kernel of the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

e. *Col* $A$ is the set of all solutions of $A\mathbf{x} = \mathbf{b}$.

**Solution**.

*Ans*: T,T,T,T,F

## Exercises 4.2

1. Either show that the given set is a vector space, or find a specific example to the contrary.

   (a) $\left\{ \begin{bmatrix} a \\ b \\ c \end{bmatrix} : a + b + c = 2 \right\}$  (b) $\left\{ \begin{bmatrix} b - 2d \\ b + 3d \\ d \end{bmatrix} : b, d \in \mathbb{R} \right\}$

   ***Hint***: See Definition 4.3, p. 121, and Example 4.18.

2. Let $A = \begin{bmatrix} -8 & -2 & -9 \\ 6 & 4 & 8 \\ 4 & 0 & 4 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$.

   (a) Is w in *Col* $A$? If yes, express w as a linear combination of columns of $A$.

   (b) Is w in *Nul* $A$? Why?

   *Ans*: yes, yes

3. Let $V$ and $W$ be vector spaces, and let $T : V \to W$ be a linear transformation. Given a subspace $U$ of $V$, let $T(U)$ denote the set of all images of the form $T(\mathbf{x})$, where $\mathbf{x} \in U$. Show that $T(U)$ is a subspace of $W$.

   ***Hint***: You should check if the three conditions in Definition 4.3 are satisfied for all elements in $T(U)$. For example, for the second condition, let's first select two arbitrary elements in $T(U)$: $T(\mathbf{u}_1)$ and $T(\mathbf{u}_2)$, where $\mathbf{u}_1, \mathbf{u}_2 \in U$. Then what you have to do is to show $\underline{T(\mathbf{u}_1) + T(\mathbf{u}_2)} \in T(U)$. To show the underlined, you may use the assumption that $T$ is linear. That is, $T(\mathbf{u}_1) + T(\mathbf{u}_2) = T(\mathbf{u}_1 + \mathbf{u}_2)$. Is the term in blue in $T(U)$? Why?

   *Advice from an old man: I know some of you may feel that the last problem is crazy. It is related to **mathematical logic** and understandability. Just try to beat your brain out for it.*

CHAPTER 5

# Eigenvalues and Eigenvectors

In this chapter, for square matrices, you will study

- How to find eigenvalues and eigenvectors
- Similarity transformation & diagonalization
- How to estimate eigenvalues by computation
- Applications to differential equations & Markov Chains

**Contents of Chapter 5**


5.1. Eigenvectors and Eigenvalues . . . . . . . . . . 132
5.2. The Characteristic Equation and Similarity Transformation . . . . . . . . . . 138
5.3. Diagonalization . . . . . . . . . . 142
5.5. Complex Eigenvalues . . . . . . . . . . 150
5.7. Applications to Differential Equations . . . . . . . . . . 154
5.8. Iterative Estimates for Eigenvalues . . . . . . . . . . 162
5.9. Applications to Markov Chains . . . . . . . . . . 168

## 5.1. Eigenvectors and Eigenvalues

**Definition 5.1.** Let $A$ be an $n \times n$ matrix. An **eigenvector** of $A$ is a **nonzero** vector x such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar $\lambda$. In this case, the scalar $\lambda$ is an **eigenvalue** of $A$ and x is the *corresponding* **eigenvector**.

**Example 5.2.** Is $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ an eigenvector of $\begin{bmatrix} 5 & 2 \\ 3 & 6 \end{bmatrix}$? What is the eigenvalue?

**Solution**.

**Example 5.3.** Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$.

a) Are u and v eigenvectors of $A$?

b) Show that 7 is an eigenvalue of matrix $A$, and find the corresponding eigenvectors.

**Solution**. ***Hint***: b) Start with $A\mathbf{x} = 7\mathbf{x}$.

**Definition 5.4.** The set of all solutions of $(A - \lambda I)\,\mathrm{x} = \mathbf{0}$ is called the **eigenspace** of $A$ corresponding to eigenvalue $\lambda$.

**Remark 5.5.** Let $\lambda$ be an eigenvalue of $A$. Then

a) Eigenspace is a subspace of $\mathbb{R}^n$ and the eigenspace of A corresponding to $\lambda$ is ***Nul***$(A - \lambda I)$.

b) The homogeneous equation $(A - \lambda I)\,\mathrm{x} = \mathbf{0}$ has at least one free variable.

**Example 5.6.** Find a basis for the eigenspace and hence the dimension of the eigenspace of $A = \begin{bmatrix} 4 & 0 & -1 \\ 3 & 0 & 3 \\ 2 & -2 & 5 \end{bmatrix}$, corresponding to the eigenvalue $\lambda = 3$.

**Solution**.

**Example 5.7.** Find a basis for the eigenspace and hence the dimension of the eigenspace of $A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$, corresponding to the eigenvalue $\lambda = 2$.

**Solution**.

**Theorem 5.8.** *The eigenvalues of a triangular matrix are the entries on its main diagonal.*

**Example 5.9.** Let $A = \begin{bmatrix} 3 & 6 & -8 \\ 0 & 0 & 6 \\ 0 & 0 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 0 & 0 \\ -2 & 1 & 0 \\ 5 & 3 & 4 \end{bmatrix}$. What are eigenvalues of $A$ and $B$?

**Solution.**

**Remark 5.10.** Zero (0) is an eigenvalue of $A \Leftrightarrow A$ is not invertible.

$$A\mathbf{x} = 0\mathbf{x} = \mathbf{0}. \tag{5.1}$$

Thus the eigenvector $\mathbf{x} \neq \mathbf{0}$ is a nontrivial solution of $A\mathbf{x} = \mathbf{0}$.

**Theorem 5.11.** *If $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_r$ of $n \times n$ matrix $A$, then the set $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\}$ is linearly independent.*

**Proof**.

- Assume that $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\}$ is *linearly dependent*.
- One of the vectors in the set is a linear combination of the preceding vectors.
- $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$ is linearly independent; $\mathbf{v}_{p+1}$ is a linear combination of the preceding vectors.
- Then, there exist scalars $c_1, c_2, \cdots, c_p$ such that

$$c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_p\,\mathbf{v}_p = \mathbf{v}_{p+1} \tag{5.2}$$

- Multiplying both sides of (5.2) by $A$, we obtain

$$c_1\,A\mathbf{v}_1 + c_2\,A\mathbf{v}_2 + \cdots + c_p\,A\mathbf{v}_p = A\mathbf{v}_{p+1}$$

  and therefore, using the fact $A\mathbf{v_k} = \lambda_k\mathbf{v_k}$:

$$c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \cdots + c_p\lambda_p\mathbf{v}_p = \lambda_{p+1}\mathbf{v}_{p+1} \tag{5.3}$$

- Multiplying both sides of (5.2) by $\lambda_{p+1}$ and subtracting the result from (5.3), we have

$$c_1(\lambda_1 - \lambda_{p+1})\mathbf{v}_1 + c_2(\lambda_2 - \lambda_{p+1})\mathbf{v}_2 + \cdots + c_p(\lambda_p - \lambda_{p+1})\mathbf{v}_p = \mathbf{0}. \tag{5.4}$$

- Since $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$ is linearly independent,

$$c_1(\lambda_1 - \lambda_{p+1}) = 0,\;\; c_2(\lambda_2 - \lambda_{p+1}) = 0,\;\; \cdots,\;\; c_p(\lambda_p - \lambda_{p+1}) = 0.$$

- Since $\lambda_1, \lambda_2, \cdots, \lambda_r$ are distinct,

$$c_1 = c_2 = \cdots = c_p = 0 \;\;\Rightarrow\;\; \mathbf{v}_{p+1} = \mathbf{0},$$

  which is a contradiction. □

**Example 5.12.** Show that if $A^2$ is the zero matrix, then the only eigenvalue of $A$ is $0$.

**Solution.** ***Hint***: You may start with $A\mathbf{x} = \lambda\mathbf{x}$, $\mathbf{x} \neq \mathbf{0}$.

**True-or-False 5.13.**

a. If $A\mathbf{x} = \lambda\mathbf{x}$ for some vector $\mathbf{x}$, then $\lambda$ is an eigenvalue of $A$.

b. A matrix $A$ is not invertible if and only if 0 is an eigenvalue of $A$.

c. A number $c$ is an eigenvalue of $A$ if and only if the equation $(A-cI)\mathbf{x} = \mathbf{0}$ has a nontrivial solution.

d. If $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent eigenvectors, then they correspond to distinct eigenvalues.

e. An eigenspace of $A$ is a null space of a certain matrix.

**Solution.**

*Ans*: F,T,T,F,T

## Exercises 5.1

1. Is $\lambda = -2$ an eigenvalue of $\begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix}$? Why or why not?

   *Ans*: Yes

2. Is $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$ an eigenvector of $\begin{bmatrix} -3 & 1 \\ -3 & 8 \end{bmatrix}$? If so, find the eigenvalue.

3. Find a basis for the eigenspace corresponding to each listed eigenvalue.

   (a) $A = \begin{bmatrix} 1 & 0 & -1 \\ 1 & -3 & 0 \\ 4 & -13 & 1 \end{bmatrix}, \lambda = -2$

   (b) $B = \begin{bmatrix} 3 & 0 & 2 & 0 \\ 1 & 3 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}, \lambda = 4$

   *Ans*: (a) $\begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$ (b) $\begin{bmatrix} 2 \\ 3 \\ 1 \\ 0 \end{bmatrix}$ and another vector

4. Find the eigenvalues of the matrix $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 5 \\ 0 & 0 & -1 \end{bmatrix}$.

5. For $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}$, find one eigenvalue, with no calculation. Justify your answer.

   *Ans*: 0. Why?

6. Prove that $\lambda$ is an eigenvalue of $A$ if and only if $\lambda$ is an eigenvalue of $A^T$. ($A$ and $A^T$ have exactly the same eigenvalues, which is frequently used in engineering applications of linear algebra.)

   ***Hint***: ① $\lambda$ is an eigenvalue of $A$
   $\Leftrightarrow$ ② $(A - \lambda I)\mathbf{x} = 0$, for some $\mathbf{x} \neq \mathbf{0}$
   $\Leftrightarrow$ ③ $(A - \lambda I)$ is not invertible.
   Now, try to use the Invertible Matrix Theorem (Theorem 2.25) to finish your proof. Note that $(A - \lambda I)^T = (A^T - \lambda I)$.

## 5.2. The Characteristic Equation and Similarity Transformation

**Recall**: Let $A$ be an $n \times n$ matrix. An **eigenvalue** $\lambda$ of $A$ and its corresponding **eigenvector** x are defined to satisfy

$$A\mathbf{x} = \lambda\mathbf{x}, \quad \mathbf{x} \neq 0.$$

Thus $(A - \lambda I)$ is not invertible and therefore $\boldsymbol{det\,(A - \lambda I) = 0}$.

**Definition 5.14.** The scalar equation $det\,(A - \lambda I) = 0$ is called the **characteristic equation** of $A$; the polynomial $p(\lambda) = det\,(A - \lambda I)$ is called the **characteristic polynomial** of $A$.

The solutions of $\boldsymbol{det\,(A - \lambda I) = 0}$ are the **eigenvalues of $A$**.

**Example 5.15.** Find the characteristic polynomial and all eigenvalues of $A = \begin{bmatrix} 8 & 2 \\ 3 & 3 \end{bmatrix}$.

**Solution**.

**Example 5.16.** Find the characteristic polynomial and all eigenvalues of $A = \begin{bmatrix} 1 & 1 & 0 \\ 6 & 0 & 5 \\ 0 & 0 & 2 \end{bmatrix}$

**Solution**.

**Theorem 5.17.** **(Invertible Matrix Theorem; continued from Theorem 2.25, p.82, and Theorem 2.74, p.107)**

*Let $A$ be $n \times n$ square matrix. Then the following are equivalent.*

*s. The number $0$ is not an eigenvalue of $A$.*

*t. det $A \neq 0$*

**Example 5.18.** Find the characteristic equation and all eigenvalues of

$$A = \begin{bmatrix} 5 & -2 & 6 & -1 \\ 0 & 3 & -8 & 0 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

**Solution**.

**Theorem 5.19.** *The eigenvalues of a triangular matrix are the entries on its main diagonal.*

**Definition 5.20.** The **algebraic multiplicity** (or, **multiplicity**) of an eigenvalue is its multiplicity as a root of the characteristic equation.

**Remark 5.21.** Let $A$ be an $n \times n$ matrix. Then the characteristic equation of $A$ is of the form

$$\begin{aligned} p(\lambda) = \textbf{det}\,(A - \lambda I) &= (-1)^n\,(\lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0) \\ &= (-1)^n\,\Pi_{i=1}^n(\lambda - \lambda_i), \end{aligned} \tag{5.5}$$

where some of eigenvalues $\lambda_i$ can be complex-valued numbers. Thus

$$\textbf{det}\,A = p(0) \;=\; (-1)^n\,\Pi_{i=1}^n(0 - \lambda_i) \;=\; \Pi_{i=1}^n\lambda_i. \tag{5.6}$$

That is, ***det A* is the product of all eigenvalues of *A*.**

## Similarity

**Definition 5.22.** Let $A$ and $B$ be $n \times n$ matrices. Then, $A$ is **similar** to $B$, if there is an invertible matrix $P$ such that

$$A = PBP^{-1}, \quad \text{or equivalently,} \quad P^{-1}AP = B.$$

Writing $Q = P^{-1}$, we have $B = QAQ^{-1}$. So $B$ is also similar to $A$, and we say simply that *$A$ and $B$ are similar*. The map $A \mapsto P^{-1}AP$ is called a **similarity transformation**.

The next theorem illustrates one use of the characteristic polynomial, and it provides **the foundation for the computation of eigenvalues**.

**Theorem 5.23.** *If $n \times n$ matrices $A$ and $B$ are similar, then they have the same characteristic polynomial and hence the same eigenvalues (with the same multiplicities).*

**Proof**. $B = P^{-1}AP$. Then,

$$\begin{aligned} B - \lambda I &= P^{-1}AP - \lambda I \\ &= P^{-1}AP - \lambda P^{-1}P \\ &= P^{-1}(A - \lambda I)P, \end{aligned}$$

from which we conclude

$$\textbf{det}\,(B - \lambda I) = \textbf{det}\,(P^{-1})\,\textbf{det}\,(A - \lambda I)\,\textbf{det}\,(P) = \textbf{det}\,(A - \lambda I).$$

**True-or-False 5.24.**

a. The determinant of $A$ is the product of the diagonal entries in $A$.

b. An elementary row operation on $A$ does not change the determinant.

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

d. If $\lambda + 5$ is a factor of the characteristic polynomial of $A$, then 5 is an eigenvalue of $A$.

e. The multiplicity of a root $r$ of the characteristic equation of $A$ is called the algebraic multiplicity of $r$ as an eigenvalue of $A$.

**Solution**.

*Ans*: F,F,T,F,T

## Exercises 5.2

1. Find the characteristic polynomial and the eigenvalues of $\begin{bmatrix} 5 & -3 \\ -4 & 3 \end{bmatrix}$.

   *Ans*: $\lambda = 4 \pm \sqrt{13}$

2. Find the characteristic polynomial of matrices. [*Note*. Finding the characteristic polynomial of a $3 \times 3$ matrix is not easy to do with just row operations, because the variable is involved.]

   (a) $\begin{bmatrix} 0 & 3 & 1 \\ 3 & 0 & 2 \\ 1 & 2 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} -1 & 0 & 1 \\ -3 & 4 & 1 \\ 0 & 0 & 2 \end{bmatrix}$

   *Ans*: (b) $-\lambda^3 + 5\lambda^2 - 2\lambda - 8 = -(\lambda + 1)(\lambda - 2)(\lambda - 4)$

3. M Report the matrices and your conclusions.

   (a) Construct a random integer-valued $4 \times 4$ matrix $A$, and verify that $A$ and $A^T$ have the same characteristic polynomial (the same eigenvalues with the same multiplicities). Do $A$ and $A^T$ have the same eigenvectors?

   (b) Make the same analysis of a $5 \times 5$ matrix.

   *Note*. Figure out by yourself how to generate random integer-valued matrices, how to make its transpose, and how to get eigenvalues and eigenvectors.

# 5.3. Diagonalization

## 5.3.1. The Diagonalization Theorem

**Definition 5.25.** An $n \times n$ matrix $A$ is said to be **diagonalizable** if there exists an invertible matrix $P$ and a diagonal matrix $D$ such that

$$A = PDP^{-1} \quad (\text{or } P^{-1}AP = D) \tag{5.7}$$

**Remark 5.26.** Let $A$ be diagonalizable, i.e., $A = PDP^{-1}$. Then

$$\begin{array}{lcl} A^2 & = & (PDP^{-1})(PDP^{-1}) = PD^2P^{-1} \\ A^k & = & PD^kP^{-1} \\ A^{-1} & = & PD^{-1}P^{-1} \quad (\text{when } A \text{ is invertible}) \\ \det A & = & \det D \end{array} \tag{5.8}$$

Diagonalization enables us to compute $A^k$ and $\det A$ quickly.

**Example 5.27.** Let $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. Find a formula for $A^k$, given that $A = PDP^{-1}$, where $P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}$ and $D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$.

**Solution**.

*Ans*: $A^k = \begin{bmatrix} 2\cdot 5^k - 3^k & 5^k - 3^k \\ 2\cdot 3^k - 2\cdot 5^k & 2\cdot 3^k - 5^k \end{bmatrix}$

**Theorem 5.28. (The Diagonalization Theorem)**

1. *An $n \times n$ matrix $A$ is diagonalizable if and only if $A$ has $n$ linearly independent eigenvectors $\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\ \mathbf{v}_n$.*
2. *In fact, $A = PDP^{-1}$ if and only if columns of P are $n$ linearly independent eigenvectors of $A$. In this case, the diagonal entries of $D$ are the corresponding eigenvalues of $A$. That is,*

$$\begin{aligned} P &= [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n], \\ D &= \textit{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n) = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}, \end{aligned} \tag{5.9}$$

*where $A\mathbf{v}_k = \lambda_k \mathbf{v}_k$, $k = 1, 2, \cdots, n$.*

The Diagonalization Theorem can be proved using the following remark.

**Remark 5.29. $AP = PD$ with $D$ Diagonal**

Let $P = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n]$ and $D = \text{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n)$ be **arbitrary $n \times n$ matrices**. Then,

$$AP = A[\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n] = [A\mathbf{v}_1\ A\mathbf{v}_2\ \cdots\ A\mathbf{v}_n], \tag{5.10}$$

while

$$PD = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = [\lambda_1\mathbf{v}_1\ \lambda_2\mathbf{v}_2\ \cdots\ \lambda_n\mathbf{v}_n]. \tag{5.11}$$

**If $AP = PD$ with $D$ diagonal, then the nonzero columns of $P$ are eigenvectors of $A$.**

**Example 5.30.** Diagonalize the following matrix, if possible.

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

**Solution**. For the computation of *det* $(A - \lambda I)$, apply $R_3 \leftarrow R_3 + R_2$ to $A - \lambda I$.

1. Find the eigenvalues of $A$.
2. Find three linearly independent eigenvectors of $A$.
3. Construct $P$ from the vectors in step 2.
4. Construct $D$ from the corresponding eigenvalues.
   *Check*: $AP = PD$?

*Ans*: $P = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$ and $D = \text{diag}(1, -2, -2)$.

**Note**: A matrix is not always diagonalizable.

**Example 5.31.** Diagonalize the following matrix, if possible.

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

for which $\textit{det}\,(B - \lambda I) = -(\lambda - 1)(\lambda + 2)^2$.
**Solution**.

## 5.3.2. Diagonalizable Matrices

**Example 5.32.** Diagonalize the following matrix, if possible.

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

**Solution**.

***Ans***: $p(\lambda) = (\lambda - 2)^3(\lambda - 4)$. $P = \begin{bmatrix} 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ **and** $D =$ **diag**$(2, 2, 2, 4)$.

**Theorem 5.33.** *An $n \times n$ matrix with $n$ distinct eigenvalues is diagonalizable.*

**Example 5.34.** Determine if the following matrix is diagonalizable.

$$A = \begin{bmatrix} 5 & -8 & 1 \\ 0 & 0 & 7 \\ 0 & 0 & -2 \end{bmatrix}$$

**Solution**.

## Matrices Whose Eigenvalues Are Not Distinct

**Theorem 5.35.** *Let $A$ be an $n \times n$ matrix whose distinct eigenvalues are $\lambda_1,\, \lambda_2,\, \cdots,\, \lambda_p$. Let $E_A(\lambda_k)$ be the eigenspace for $\lambda_k$.*

1. *dim $E_A(\lambda_k) \le$ (the multiplicity of the eigenvalue $\lambda_k$), for $1 \le k \le p$*
2. *The matrix A is diagonalizable*
   $\Leftrightarrow$ *the sum of the dimensions of the eigenspaces equals $n$*
   $\Leftrightarrow$ *dim $E_A(\lambda_k)$ = the multiplicity of $\lambda_k$, for each $1 \le k \le p$*
   *(and the characteristic polynomial factors completely into linear factors)*
3. *If $A$ is diagonalizable and $\mathcal{B}_k$ is a basis for $E_A(\lambda_k)$, then the total collection of vectors in the sets $\{\mathcal{B}_1,\, \mathcal{B}_2,\, \cdots,\, \mathcal{B}_p\}$ forms an eigenvector basis for $\mathbb{R}^n$.*

**Example 5.36.** Diagonalize the following matrix, if possible.

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

**Solution**.

**True-or-False** **5.37.**

a. A is diagonalizable if $A = PDP^{-1}$, for some matrix $D$ and some invertible matrix $P$.

b. If $\mathbb{R}^n$ has a basis of eigenvectors of $A$, then $A$ is diagonalizable.

c. If $A$ is diagonalizable, then $A$ is invertible.

d. If $A$ is invertible, then $A$ is diagonalizable.

e. If $AP = PD$, with $D$ diagonal, then the nonzero columns of $P$ must be eigenvectors of $A$.

**Solution**.

*Ans*: F,T,F,F,T

## Exercises 5.3

1. The matrix $A$ is factored in the form $PDP^{-1}$. Find the eigenvalues of $A$ and a basis for each eigenspace.

$$A = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/4 & 1/2 & -3/4 \\ 1/4 & -1/2 & 1/4 \end{bmatrix}$$

2. Diagonalize the matrices, if possible.

(a) $\begin{bmatrix} 2 & 2 & -1 \\ 1 & 3 & -1 \\ -1 & -2 & 2 \end{bmatrix}$ (b) $\begin{bmatrix} 7 & 4 & 16 \\ 2 & 5 & 8 \\ -2 & -2 & -5 \end{bmatrix}$ (c) $\begin{bmatrix} 4 & 0 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix}$

***Hint***: Use (a) $\lambda = 5, 1$. (b) $\lambda = 3, 1$. (c) Not diagonalizable. Why?

3. $A$ is a $3 \times 3$ matrix with two eigenvalues. Each eigenspace is one-dimensional. Is $A$ diagonalizable? Why?

4. Construct and verify.
   (a) A nonzero $2 \times 2$ matrix that is invertible but not diagonalizable.
   (b) A nondiagonal $2 \times 2$ matrix that is diagonalizable but not invertible.

## 5.5. Complex Eigenvalues

**Definition 5.38.** Let $A$ be an $n \times n$ matrix with real entries. A **complex eigenvalue** $\lambda$ is a complex-valued scalar such that

$$A\mathbf{x} = \lambda\mathbf{x}, \quad \text{where } \mathbf{x} \neq \mathbf{0}.$$

As usual, we determine $\lambda$ by solving the **characteristic equation**

$$\boldsymbol{det}\,(A - \lambda I) = 0.$$

**Example 5.39.** Let $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ and consider the linear transformation

$$\mathbf{x} \mapsto A\mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^2.$$

Then

a) It rotates the plane counterclockwise through a quarter-turn.
b) The action of $A$ is periodic, since after four quarter-turns, a vector is back where it started.
c) Obviously, no nonzero vector is mapped into a multiple of itself, so $A$ has no eigenvectors in $\mathbb{R}^2$ and hence no real eigenvalues.

Find the eigenvalues of $A$, and find a basis for each eigenspace.
**Solution**.

## Eigenvalues and Eigenvectors of a Real Matrix That Acts on $\mathbb{C}^n$

**Remark 5.40.** Let $A$ be an $n \times n$ matrix whose entries are real. Then

$$\overline{A\mathbf{x}} = \overline{A}\overline{\mathbf{x}} = A\overline{\mathbf{x}}.$$

Thus, if $\lambda$ is an eigenvalue of $A$, i.e., $A\mathbf{x} = \lambda\mathbf{x}$, then

$$A\overline{\mathbf{x}} = \overline{A\mathbf{x}} = \overline{\lambda\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}.$$

That is,

$$A\mathbf{x} = \lambda\mathbf{x} \quad \Leftrightarrow \quad A\overline{\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}} \tag{5.12}$$

If $\lambda$ is an eigenvalue of $A$ with corresponding eigenvector $\mathbf{x}$, then the **complex conjugate** of $\lambda$, $\overline{\lambda}$, is also an eigenvalue with eigenvector $\overline{\mathbf{x}}$.

**Example 5.41.** Let $A = \begin{bmatrix} 1 & 5 \\ -2 & 3 \end{bmatrix}$.

(a) Find all eigenvalues and the corresponding eigenvectors for $A$.

(b) Let $(\lambda, \mathbf{v})$, with $\lambda = a - bi$, be an eigen-pair. Let $P = [\mathrm{Re}\,\mathbf{v} \;\; \mathrm{Im}\,\mathbf{v}]$ and $C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. Show that $AP = PC$. (This implies that $A = PCP^{-1}$.)

**Solution**.

*Ans*: (a) $\lambda = 2 \pm 3i$.

**Theorem 5.42. (Factorization)**: *Let $A$ be $2 \times 2$ real matrix with complex eigenvalue $\lambda = a - bi$ ($b \neq 0$) and an associated eigenvector* $\mathbf{v}$*. If* $P = [Re\,\mathbf{v}\ \ Im\,\mathbf{v}]$ *and* $C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$*, then*

$$A = PCP^{-1}. \tag{5.13}$$

**Example 5.43.** Find eigenvalues of $C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$.

**Solution**.

**Example 5.44.** Find all eigenvalues and the corresponding eigenvectors for

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

Also find an invertible matrix $P$ and a matrix $C$ such that $A = PCP^{-1}$.
**Solution**.

## Exercises 5.5

1. Let each matrix act on $\mathbb{C}^2$. Find the eigenvalues and a basis for each eigenspace in $\mathbb{C}^2$.

   (a) $\begin{bmatrix} 1 & -2 \\ 1 & 3 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 5 \\ -2 & 3 \end{bmatrix}$

   *Ans*: (a) An eigen-pair: $\lambda = 2+i$, $\begin{bmatrix} -1+i \\ 1 \end{bmatrix}$

2. Let $A = \begin{bmatrix} 5 & -2 \\ 1 & 3 \end{bmatrix}$. Find an invertible matrix $P$ and a matrix $C$ of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ such that $A = PCP^{-1}$.

   *Ans*: $P = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix}$, $C = \begin{bmatrix} 4 & -1 \\ 1 & 4 \end{bmatrix}$.

3. Let $A$ be an $n \times n$ real matrix with the property that $A^T = A$. Let x be any vector in $\mathbb{C}^n$, and let $q = \overline{\mathbf{x}}^T A\mathbf{x}$. The equalities below show that $q$ is a real number by verifying that $\overline{q} = q$. Give a reason for each step.

$$\overline{q} = \overline{\overline{\mathbf{x}}^T A\mathbf{x}} \underset{(a)}{=} \mathbf{x}^T \overline{A\mathbf{x}} \underset{(b)}{=} \mathbf{x}^T A\overline{\mathbf{x}} \underset{(c)}{=} (\mathbf{x}^T A\overline{\mathbf{x}})^T \underset{(d)}{=} \overline{\mathbf{x}}^T A^T \mathbf{x} \underset{(e)}{=} q$$

## 5.7. Applications to Differential Equations

**Recall**: For solving the first-order **differential equation**:

$$\frac{dx}{dt} = a\,x, \tag{5.14}$$

we rewrite it as

$$\frac{dx}{x} = a\,dt. \tag{5.15}$$

By integrating both sides, we have

$$\ln|x| = a\,t + K.$$

Thus, for $x = x(t)$,

$$x = \pm e^{a\,t+K} = \pm e^{K} e^{at} = C \cdot e^{at}. \tag{5.16}$$

**Example 5.45.** Consider the first-order **initial-value problem**

$$\frac{dx}{dt} = 5x, \quad x(0) = 2. \tag{5.17}$$

(a) Find the solution $x(t)$.

(b) Check if our solution satisfies both the differential equation and the **initial condition**.

**Solution**.

### 5.7.1. Dynamical System: The System of First-Order Differential Equations

Consider a system of first-order differential equations:

$$\begin{cases} x_1' &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ x_2' &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ & \vdots \\ x_n' &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{cases} \tag{5.18}$$

where $x_1, x_2, \cdots, x_n$ are functions of $t$ and $a_{ij}$'s are constants.
Then the system can be written as

$$\mathbf{x}'(t) = A\mathbf{x}(t), \tag{5.19}$$

where

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \quad \mathbf{x}'(t) = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix}, \quad \text{and} \quad A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}.$$

**How to solve (5.19): $\mathbf{x}'(t) = A\mathbf{x}(t)$?**

**Observation 5.46.** Let $(\lambda, \mathbf{v})$ be an eigen-pair of $A$, i.e., $A\mathbf{v} = \lambda\mathbf{v}$. Then, for an arbitrary constant $c$,

$$\mathbf{x}(t) = c\mathbf{v}e^{\lambda t} \tag{5.20}$$

is a solution of (5.19).

**Proof**. Let's check it:

$$\begin{aligned} \mathbf{x}'(t) &= c\mathbf{v}(e^{\lambda t})' &= c\lambda\mathbf{v}e^{\lambda t} \\ A\mathbf{x}(t) &= Ac\mathbf{v}e^{\lambda t} &= c\lambda\mathbf{v}e^{\lambda t}, \end{aligned} \tag{5.21}$$

which completes the proof. □

**Example 5.47.** Solve the initial-value problem:

$$\mathbf{x}'(t) = \begin{bmatrix} -4 & -2 \\ 3 & 1 \end{bmatrix} \mathbf{x}(t), \quad \mathbf{x}(0) = \begin{bmatrix} -2 \\ 4 \end{bmatrix}. \tag{5.22}$$

**Solution**. The two eigen-pairs are $(\lambda_i, \mathbf{v}_i)$, $i = 1, 2$. Then the general solution is $\mathbf{x} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 = c_1\mathbf{v}_1 e^{\lambda_1 t} + c_2\mathbf{v}_2 e^{\lambda_2 t}$.

*Ans*: $\mathbf{x}(t) = -2\begin{bmatrix} 2 \\ -3 \end{bmatrix} e^{-t} + 2\begin{bmatrix} 1 \\ -1 \end{bmatrix} e^{-2t}$

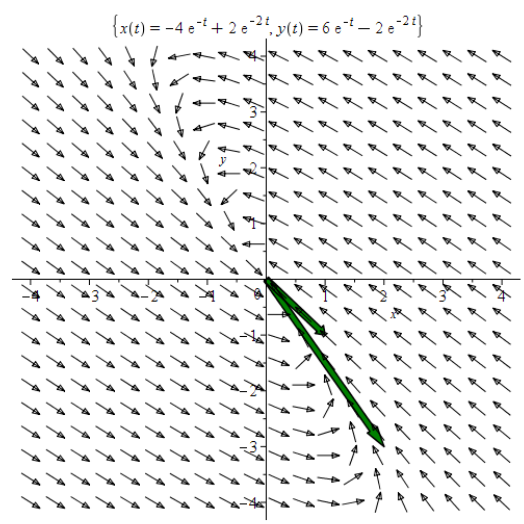



Figure 5.1: Trajectories for the dynamical system (5.22).

### 5.7.2. Trajectories for the Dynamical Systems: Attractors, Repellers, and Saddle Points

**Summary 5.48.** For a dynamical system

$$\mathbf{x}'(t) = A\,\mathbf{x}(t), \quad A \in \mathbb{R}^{2\times 2}, \tag{5.23}$$

let $(\lambda_i, \mathbf{v}_i)$, $i = 1, 2$, be eigen-pairs of $A$. Then the general solution of (5.23) reads

$$\mathbf{x}(t) = c_1\mathbf{v}_1 e^{\lambda_1 t} + c_2\mathbf{v}_2 e^{\lambda_2 t}, \tag{5.24}$$

for arbitrary scalars $c_1, c_2 \in \mathbb{R}$.

**Definition 5.49. Two Distinct Real Eigenvalues**

1. If the eigenvalues of $A$ are **both negative**, the origin is called an **attractor** or **sink**, since all trajectories (solutions x(t)) are drawn to the origin.
   - For the solution in the last example, the **direction of greatest attraction** is along the trajectory of the eigenfunction $\mathrm{x}_2$ (along the line through 0 and $\mathrm{v}_2$), corresponding to the more negative eigenvalue $\lambda = -2$.
2. If the eigenvalues of $A$ are **both positive**, the origin is called a **repeller** or **source**, since all trajectories (solutions x(t)) are traversed away from the origin.
3. If $A$ has both positive and negative eigenvalues, the origin is called **saddle point** of the dynamical system.

The larger the eigenvalue is in modulus, the greater attraction/repulsion.

**Example 5.50.** Solve the initial-value problem:

$$\mathbf{x}'(t) = \begin{bmatrix} 2 & 3 \\ -1 & -2 \end{bmatrix} \mathbf{x}(t), \quad \mathbf{x}(0) = \begin{bmatrix} 3 \\ 2 \end{bmatrix}. \tag{5.25}$$

What are the direction of greatest attraction and the direction of greatest repulsion?
**Solution**.

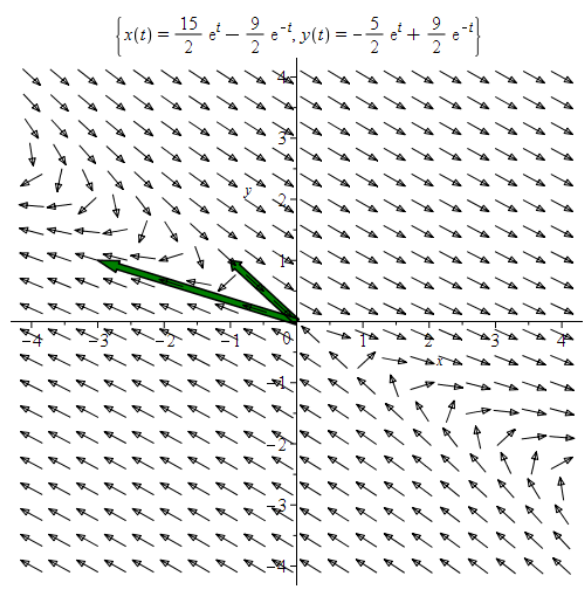



Figure 5.2: Trajectories for the dynamical system (5.25).

*Ans*: $\mathbf{x}(t) = -\frac{5}{2}\begin{bmatrix} -3 \\ 1 \end{bmatrix} e^{t} + \frac{9}{2}\begin{bmatrix} -1 \\ 1 \end{bmatrix} e^{-t}$. The direction of greatest attraction = $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

**Summary 5.51.** *For the dynamical system $\mathbf{x}'(t) = A\mathbf{x}(t)$, $A \in \mathbb{R}^{2\times 2}$, let $(\lambda_1, \mathbf{v}_1)$ and $(\lambda_2, \mathbf{v}_2)$ be eigen-pairs of $A$.*

- **Case (i)**: *If $\lambda_1$ and $\lambda_2$ are real and distinct, then*

$$\mathbf{x} = c_1\mathbf{v}_1 e^{\lambda_1 t} + c_2\mathbf{v}_2 e^{\lambda_2 t}. \quad (5.26)$$

- **Case (ii)**: *If $A$ has a **double eigenvalue** $\lambda$ (with $\mathbf{v}$), then you should find a second **generalized eigenvector** by solving, e.g.,*

$$(A - \lambda I)\,\mathbf{w} = \mathbf{v}. \quad (5.27)$$

*(The above can be derived from a guess: $\mathbf{x} = t\mathbf{v}e^{\lambda t} + \mathbf{w}e^{\lambda t}$.) Then the general solution becomes [2]*

$$\mathbf{x}(t) = c_1\,\mathbf{v}e^{\lambda t} + c_2\,(t\mathbf{v} + \mathbf{w})e^{\lambda t} \quad (5.28)$$

***($\mathbf{w}$ is a simple shift vector, which may not be unique.)***

- **Case (iii)**: *If $\lambda = a + bi$ ($b \neq 0$) and $\overline{\lambda}$ are complex eigenvalues with eigenvectors $\mathbf{v}$ and $\overline{\mathbf{v}}$, then*

$$\mathbf{x} = c_1\mathbf{v}e^{\lambda t} + c_2\overline{\mathbf{v}}e^{\overline{\lambda} t}, \quad (5.29)$$

*from which two linearly independent real-valued solutions must be extracted.*

**Note**: Let $\lambda = a + bi$ and $\mathbf{v} = \mathbf{Re}\mathbf{v} + i\,\mathbf{Im}\mathbf{v}$. Then

$$\begin{aligned}\mathbf{v}e^{\lambda t} &= (\mathbf{Re}\mathbf{v} + i\,\mathbf{Im}\mathbf{v}) \cdot e^{at}(\cos bt + i\sin bt)\\ &= [(\mathbf{Re}\mathbf{v})\cos bt - (\mathbf{Im}\mathbf{v})\sin bt]e^{at}\\ &\quad + i\,[(\mathbf{Re}\mathbf{v})\sin bt + (\mathbf{Im}\mathbf{v})\cos bt]e^{at}.\end{aligned}$$

Let

$$\begin{aligned}\mathbf{y}_1(t) &= [(\mathbf{Re}\mathbf{v})\cos bt - (\mathbf{Im}\mathbf{v})\sin bt]e^{at}\\ \mathbf{y}_2(t) &= [(\mathbf{Re}\mathbf{v})\sin bt + (\mathbf{Im}\mathbf{v})\cos bt]e^{at}\end{aligned}$$

Then they are linearly independent and satisfy the dynamical system. Thus, the real-valued general solution of the dynamical system reads

$$\mathbf{x}(t) = C_1\mathbf{y}_1(t) + C_2\mathbf{y}_2(t). \quad (5.30)$$

**Example 5.52.** Construct the general solution of $\mathbf{x}'(t) = A\mathbf{x}(t)$ when
$A = \begin{bmatrix} -3 & 2 \\ -1 & -1 \end{bmatrix}$

**Solution**.

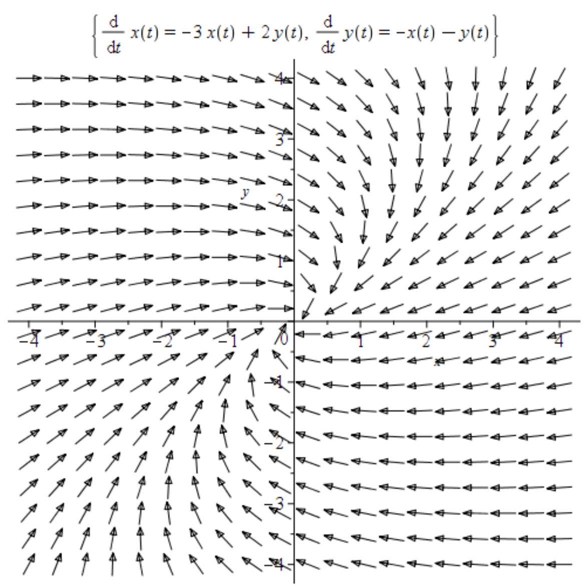



Figure 5.3: Trajectories for the dynamical system.

*Ans*: (complex solution) $c_1 \begin{bmatrix} 1-i \\ 1 \end{bmatrix} e^{(-2+i)t} + c_2 \begin{bmatrix} 1+i \\ 1 \end{bmatrix} e^{(-2-i)t}$

*Ans*: (real solution) $c_1 \begin{bmatrix} \cos t + \sin t \\ \cos t \end{bmatrix} e^{-2t} + c_2 \begin{bmatrix} \sin t - \cos t \\ \sin t \end{bmatrix} e^{-2t}$

## Exercises 5.7

1. (i) Solve the initial-value problem $\mathbf{x}' = A\mathbf{x}$ with $\mathbf{x}(0) = (3, 2)$. (ii) Classify the nature of the origin as an attractor, repeller, or saddle point of the dynamical system. (iii) Find the directions of greatest attraction and/or repulsion. (iv) When the origin is a saddle point, sketch typical trajectories.

   (a) $A = \begin{bmatrix} -2 & -5 \\ 1 & 4 \end{bmatrix}$ (b) $A = \begin{bmatrix} 7 & -1 \\ 3 & 3 \end{bmatrix}$.

   *Ans*: (a) The origin is a saddle point.
   The direction of G.A. = $\begin{bmatrix} -5 \\ 1 \end{bmatrix}$. The direction of G.R. = $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.
   *Ans*: (b) The origin is a repeller. The direction of G.R. = $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

2. Use the strategies in (5.27) and (5.28) to solve

$$\mathbf{x}' = \begin{bmatrix} 7 & 1 \\ -4 & 3 \end{bmatrix}\mathbf{x}, \quad \mathbf{x}(0) = \begin{bmatrix} 2 \\ -5 \end{bmatrix}.$$

   *Ans*: $\mathbf{x}(t) = 2\begin{bmatrix} 1 \\ -2 \end{bmatrix}e^{5t} - \left(t\begin{bmatrix} 1 \\ -2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)e^{5t}$

# 5.8. Iterative Estimates for Eigenvalues

## 5.8.1. The Power Method

The **power method** is an **iterative algorithm**:

Given a square matrix $A \in \mathbb{R}^{n\times n}$, the algorithm finds a number $\lambda$, which is **the largest eigenvalue of $A$** (in modulus), and **its corresponding eigenvector** v.

**Assumption**. To apply the power method, we assume that $A \in \mathbb{R}^{n\times n}$ has

- **$n$ linearly independent eigenvectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$**, and
- **exactly one eigenvalue** that is largest in magnitude, $\lambda_1$:

$$|\lambda_1| > |\lambda_2| \ge |\lambda_3| \ge \cdots \ge |\lambda_n|. \tag{5.31}$$

The power method approximates the largest eigenvalue $\lambda_1$ and its associated eigenvector $\mathbf{v}_1$.

**Derivation of Power Iteration**

- Since eigenvectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ are linearly independent, any vector $\mathbf{x} \in \mathbb{R}^n$ can be expressed as

$$\mathbf{x} = \sum_{j=1}^{n} \beta_j \mathbf{v}_j, \tag{5.32}$$

  for some constants $\{\beta_1, \beta_2, \cdots, \beta_n\}$.

- Multiplying both sides of (5.32) by $A$ and $A^2$ gives

$$\begin{aligned} A\mathbf{x} &= A\Big(\sum_{j=1}^{n} \beta_j \mathbf{v}_j\Big) = \sum_{j=1}^{n} \beta_j A\mathbf{v}_j = \sum_{j=1}^{n} \beta_j \lambda_j \mathbf{v}_j, \\ A^2\mathbf{x} &= A\Big(\sum_{j=1}^{n} \beta_j \lambda_j \mathbf{v}_j\Big) = \sum_{j=1}^{n} \beta_j \lambda_j^2 \mathbf{v}_j. \end{aligned} \tag{5.33}$$

- In general,

$$A^k\mathbf{x} = \sum_{j=1}^{n} \beta_j \lambda_j^k \mathbf{v}_j, \quad k = 1, 2, \cdots, \tag{5.34}$$

which gives

$$A^k\mathbf{x} = \lambda_1^k \cdot \sum_{j=1}^{n} \beta_j \Big(\frac{\lambda_j}{\lambda_1}\Big)^k \mathbf{v}_j = \lambda_1^k \cdot \Big[\beta_1 \Big(\frac{\lambda_1}{\lambda_1}\Big)^k \mathbf{v}_1 + \beta_2 \Big(\frac{\lambda_2}{\lambda_1}\Big)^k \mathbf{v}_2 + \cdots + \beta_n \Big(\frac{\lambda_n}{\lambda_1}\Big)^k \mathbf{v}_n\Big]. \tag{5.35}$$

- For $j = 2, 3, \cdots, n$, since $|\lambda_j/\lambda_1| < 1$, we have $\lim_{k\to\infty} |\lambda_j/\lambda_1|^k = 0$, and

$$\lim_{k\to\infty} A^k\mathbf{x} = \lim_{k\to\infty} \lambda_1^k \beta_1 \mathbf{v}_1. \tag{5.36}$$

**Remark 5.53.** The sequence in (5.36) converges to 0 if $|\lambda_1| < 1$ and diverges if $|\lambda_1| > 1$, provided that $\beta_1 \neq 0$.

- The entries of $A^k\mathbf{x}$ will grow with $k$ if $|\lambda_1| > 1$ and will go to 0 if $|\lambda_1| < 1$.
- In either case, it is hard to decide the largest eigenvalue $\lambda_1$ and its associated eigenvector $\mathbf{v}_1$.
- **To take care of that possibility**, we scale $A^k\mathbf{x}$ in an appropriate manner to ensure that the limit in (5.36) is finite and nonzero.

**Algorithm 5.54. (The Power Iteration)** Given $\mathbf{x} \neq 0$:

$$\begin{aligned} &\textbf{initialization}: \quad \mathbf{x}^0 = \mathbf{x}/||\mathbf{x}||_\infty \\ &\textbf{for} \;\; k = 1, 2, \cdots \\ &\qquad \mathbf{y}^k = A\mathbf{x}^{k-1}; \;\; \mu_k = ||\mathbf{y}^k||_\infty \\ &\qquad \mathbf{x}^k = \mathbf{y}^k/\mu_k \\ &\textbf{end for} \end{aligned} \tag{5.37}$$

**Claim 5.55.** Let $\{\mathbf{x}^k, \mu_k\}$ be a sequence produced by the power method. Then,

$$\mathbf{x}^k \to \mathbf{v}_1, \;\; \mu_k \to |\lambda_1|, \quad \text{as} \;\; k \to \infty. \tag{5.38}$$

More precisely, the power method converges as

$$\mu_k = |\lambda_1| + \mathcal{O}(|\lambda_2/\lambda_1|^k). \tag{5.39}$$

**Example 5.56.** The matrix $A = \begin{bmatrix} -4 & 1 & -1 \\ 1 & -3 & 2 \\ -1 & 2 & -3 \end{bmatrix}$ has eigenvalues and eigenvectors as follows

$$\texttt{eig(A)} = \begin{bmatrix} -6 \\ -3 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 & -2 & 0 \\ -1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Verify that the sequence produced by the power method converges to the largest eigenvalue and its associated eigenvector.

**Solution**.

power_iteration.m

```
A = [-4 1 -1; 1 -3 2; -1 2 -3];
%[V,D] = eig(A)

fmt = ['k=%2d: x = [',repmat('%.5f, ',1,numel(x)-1),'%.5f], ',...
       'mu=%.5f (error = %.5f)\n'];
x = [1 0 0]';
for k=1:10
    y = A*x;
    [~,ind] = max(abs(y)); mu = y(ind);
    x =y/mu;
    fprintf(fmt,k,x,mu,abs(-6-mu))
end
```

Output

```
k= 1: x = [1.00000, -0.25000, 0.25000], mu=-4.00000 (error = 2.00000)
k= 2: x = [1.00000, -0.50000, 0.50000], mu=-4.50000 (error = 1.50000)
k= 3: x = [1.00000, -0.70000, 0.70000], mu=-5.00000 (error = 1.00000)
k= 4: x = [1.00000, -0.83333, 0.83333], mu=-5.40000 (error = 0.60000)
k= 5: x = [1.00000, -0.91176, 0.91176], mu=-5.66667 (error = 0.33333)
k= 6: x = [1.00000, -0.95455, 0.95455], mu=-5.82353 (error = 0.17647)
k= 7: x = [1.00000, -0.97692, 0.97692], mu=-5.90909 (error = 0.09091)
k= 8: x = [1.00000, -0.98837, 0.98837], mu=-5.95385 (error = 0.04615)
k= 9: x = [1.00000, -0.99416, 0.99416], mu=-5.97674 (error = 0.02326)
k=10: x = [1.00000, -0.99708, 0.99708], mu=-5.98833 (error = 0.01167)
```

Notice that $|-6-\mu_k| \approx \frac{1}{2}|-6-\mu_{k-1}|$, for which $|\lambda_2/\lambda_1| = \frac{1}{2}$.

## 5.8.2. Inverse Power Method

Some applications require to find an eigenvalue of the matrix $A$, near a prescribed value $q$. The **inverse power method** is a variant of the Power method to solve such a problem.

- We begin with the eigenvalues and eigenvectors of $(A-qI)^{-1}$. Let

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad i = 1, 2, \cdots, n. \tag{5.40}$$

- Then it is easy to see that

$$(A-qI)\mathbf{v}_i = (\lambda_i - q)\mathbf{v}_i. \tag{5.41}$$

  Thus, we obtain

$$(A-qI)^{-1}\mathbf{v}_i = \frac{1}{\lambda_i - q}\mathbf{v}_i. \tag{5.42}$$

- That is, when $q \notin \{\lambda_1, \lambda_2, \cdots, \lambda_n\}$, the eigenvalues of $(A-qI)^{-1}$ are

$$\frac{1}{\lambda_1 - q}, \frac{1}{\lambda_2 - q}, \cdots, \frac{1}{\lambda_n - q}, \tag{5.43}$$

  with the same eigenvectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ of $A$.

**Algorithm 5.57. (Inverse Power Method)** Applying the power method to $(A-qI)^{-1}$ gives the **inverse power method**. Given $\mathbf{x} \neq 0$:

$$\begin{aligned}
&\textbf{set}: \quad \mathbf{x}^0 = \mathbf{x}/||\mathbf{x}||_\infty \\
&\textbf{for } k = 1, 2, \cdots \\
&\qquad \mathbf{y}^k = (A-qI)^{-1}\mathbf{x}^{k-1}; \quad \mu_k = ||\mathbf{y}^k||_\infty \\
&\qquad \mathbf{x}^k = \mathbf{y}^k/\mu_k \\
&\qquad \lambda_k = 1/\mu_k + q \\
&\textbf{end for}
\end{aligned} \tag{5.44}$$

**Example 5.58.** The matrix $A$ is as in Example 5.56: $A = \begin{bmatrix} -4 & 1 & -1 \\ 1 & -3 & 2 \\ -1 & 2 & -3 \end{bmatrix}$.

Find the the eigenvalue of $A$ nearest to $q = -5/2$, using the inverse power method.

**Solution**.

inverse_power.m

```
A = [-4 1 -1; 1 -3 2; -1 2 -3];
%[V,D] = eig(A)

fmt = ['k=%2d: x = [',repmat('%.5f, ',1,numel(x)-1),'%.5f], ',...
        'lambda=%.7f (error = %.7f)\n'];
q = -5/2; x = [1 0 0]';
B = inv(A-q*eye(3));
for k=1:10
    y = B*x;
    [~,ind] = max(abs(y)); mu = y(ind);
    x =y/mu;
    lambda = 1/mu + q;
    fprintf(fmt,k,x,lambda,abs(-3-lambda))
end
```

Output

```
k= 1: x = [1.00000, 0.40000, -0.40000], lambda=-3.2000000 (error = 0.2000000)
k= 2: x = [1.00000, 0.48485, -0.48485], lambda=-3.0303030 (error = 0.0303030)
k= 3: x = [1.00000, 0.49782, -0.49782], lambda=-3.0043668 (error = 0.0043668)
k= 4: x = [1.00000, 0.49969, -0.49969], lambda=-3.0006246 (error = 0.0006246)
k= 5: x = [1.00000, 0.49996, -0.49996], lambda=-3.0000892 (error = 0.0000892)
k= 6: x = [1.00000, 0.49999, -0.49999], lambda=-3.0000127 (error = 0.0000127)
k= 7: x = [1.00000, 0.50000, -0.50000], lambda=-3.0000018 (error = 0.0000018)
k= 8: x = [1.00000, 0.50000, -0.50000], lambda=-3.0000003 (error = 0.0000003)
k= 9: x = [1.00000, 0.50000, -0.50000], lambda=-3.0000000 (error = 0.0000000)
k=10: x = [1.00000, 0.50000, -0.50000], lambda=-3.0000000 (error = 0.0000000)
```

**Note**: Eigenvalues of $(A - qI)^{-1}$ are $\{-2/7,\ -2,\ 2/3\}$.

- The initial vector: $\mathbf{x}_0 = [1, 0, 0]^T = \dfrac{1}{3}(\mathbf{v}_1 - \mathbf{v}_2)$; see Example 5.56.
- Thus, each iteration must reduce the error **by a factor of 7**.

## Exercises 5.8

1. M The matrix in Example 5.58 has eigenvalues $\{-6, -3, -1\}$. We may try to find the eigenvalue of $A$ nearest to $q = -3.1$.

   (a) Estimate (mathematically) the convergence speed of the inverse power method.

   (b) Verify it by implementing the inverse power method, with $\mathbf{x}_0 = [0, 1, 0]^T$.

2. M Let $A = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & 0 & -1 \\ 0 & 0 & 4 & -2 \\ 0 & -1 & -2 & 4 \end{bmatrix}$. Use indicated methods to approximate eigenvalues and their associated eigenvectors of $A$ within to $10^{-12}$ accuracy.

   (a) The power method, the largest eigenvalue.

   (b) The inverse power method, an eigenvalue near $q = 3$.

   (c) The inverse power method, the smallest eigenvalue.

# 5.9. Applications to Markov Chains

**Theory**

- Markov chains are useful tools in some probabilistic models.
- The basic idea is the following:
  **Suppose that you are watching some collection of objects that are changing through time.**
- **Assumptions** (on states & changes):
  - The **total number of objects is not changing**, but their **states** (position, colour, disposition, etc) are changing.
  - The **proportion of changing states is constant** and these changes occur at discrete times, one after the next.
- Then we are in a good position to model changes by a **Markov chain**.

**Example 5.59.** Consider a **three storey aviary at a local zoo** which houses 300 small birds.

- The aviary has three levels.
- The birds spend their day, flying from one level to another.

Our problem is to determine what the probability is of a given bird being at a given level of the aviary at a given time.

*Continued on the next page* ⇒

**Data**

- Observe a vector

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}, \tag{5.45}$$

  where $p_i$ is the proportion (probability) of birds on the $i$-th level.
  **Note**: $p_1 + p_2 + p_3 = 1$.

- After 10 minutes, we have a new distribution of the birds

$$\mathbf{p}' = \begin{bmatrix} p'_1 \\ p'_2 \\ p'_3 \end{bmatrix}. \tag{5.46}$$

**Model**

- We assume that the change from p to p′ is given by a linear operator on $\mathbb{R}^3$. In other words, there is a matrix $T \in \mathbb{R}^{3\times 3}$ such that

$$\mathbf{p}' = T\mathbf{p}. \tag{5.47}$$

  The matrix $T$ is called the **transition matrix** for the Markov chain.

- Another 10 minutes later, we observe another distribution

$$\mathbf{p}'' = T\mathbf{p}'. \tag{5.48}$$

**Note**: The same matrix $T$ is used in (5.47) and (5.48), because we assume that ***the probability of a bird moving to another level is independent of time***.

- In other words, the probability of a bird moving to a particular level depends only on the present state of the bird, and not on any past states.
- This type of model is known as a **finite Markov chain**.

### 5.9.1. Probability Vector and Stochastic Matrix

**Definition 5.60. Probability Vector and Stochastic Matrix**

- A vector $\mathbf{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix}$ with nonnegative entries that add up to 1 is called a **probability vector**.
- A **(left) stochastic matrix** is a square matrix whose columns are probability vectors.

A **stochastic matrix** is also called a **probability matrix**, **transition matrix**, **substitution matrix**, or **Markov matrix**.

**Lemma 5.61.** Let $T$ be a stochastic matrix. If $\mathbf{p}$ is a probability vector, then so is $\mathbf{q} = T\mathbf{p}$.

**Proof**. Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ be the columns of $T$. Then

$$\mathbf{q} = T\mathbf{p} = p_1\mathbf{v}_1 + p_2\mathbf{v}_2 + \cdots p_n\mathbf{v}_n.$$

Clearly $\mathbf{q}$ has nonnegative entries; their sum reads

$$\text{sum}(\mathbf{q}) = \text{sum}(p_1\mathbf{v}_1 + p_2\mathbf{v}_2 + \cdots p_n\mathbf{v}_n) = p_1 + p_2 + \cdots + p_n = 1.$$

**Definition 5.62. Markov Chain**

In general, a finite **Markov chain** is a sequence of **probability vectors** $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \cdots$, together with a **stochastic matrix** $T$, such that

$$\mathbf{x}_1 = T\mathbf{x}_0, \quad \mathbf{x}_2 = T\mathbf{x}_1, \quad \mathbf{x}_3 = T\mathbf{x}_2, \quad \cdots \tag{5.49}$$

We can rewrite the above conditions as a recurrence relation

$$\mathbf{x}_{k+1} = T\mathbf{x}_k, \quad k = 0, 1, 2, \cdots \tag{5.50}$$

The vector $\mathbf{x}_k$ is often called a **state vector**.

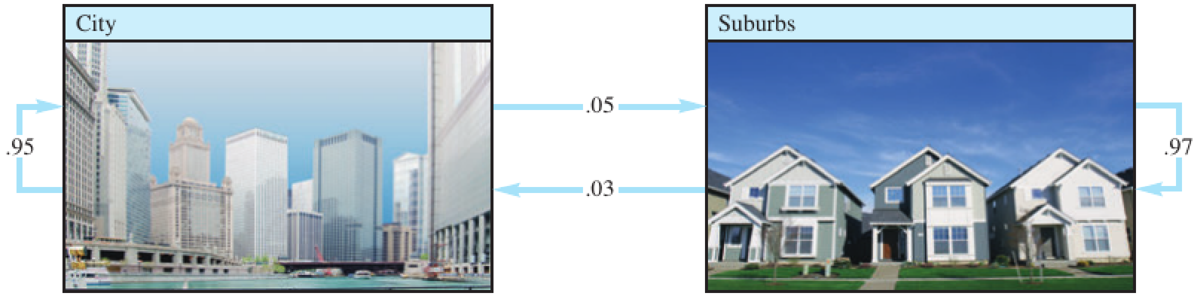



Figure 5.4: Annual percentage migration between a city and its suburbs.

**Example 5.63.** Figure 5.4 shows population movement between a city and its suburbs. Then, the **annual migration** between these two parts of the metropolitan region can be expressed by the migration matrix M:

$$M = \begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix}. \tag{5.51}$$

Suppose the 2023 population of the region is 60,000 in the city and 40,000 in the suburbs. What is the distribution of the population in 2024? In 2025?

**Solution**.

annual_migration.m

```
M = [0.95 0.03
     0.05 0.97];

x0 = [60000
      40000];

x1 = M*x0
x2 = M*x1
```

Output

```
x1 =
       58200
       41800

x2 =
       56544
       43456
```

**Example 5.64. (Revisit to the aviary example: Example 5.59)**
Assume that

- Whenever a bird is on any level of the aviary, the probability of that bird **staying on the same level 10 min later** is 1/2.
- If the bird is on **the first level**, the probability of moving to the second level in 10 min is 1/3 and of moving to the third level in 10 min is 1/6.
- For a bird on **the second level**, the probability of moving to either the first or third level is 1/4.
- For a bird on **the third level**, the probability of moving to the second level is 1/3 and of moving to the first is 1/6.

(a) Find the transition matrix for this example.

(b) Suppose that after breakfast, all the birds are in the dining area on the first level. Where are they in 10 min? In 20 min? In 30 min?

**Solution**. (a) From the information given, we derive the **transition matrix**:

$$T = \begin{bmatrix} 1/2 & 1/4 & 1/6 \\ 1/3 & 1/2 & 1/3 \\ 1/6 & 1/4 & 1/2 \end{bmatrix} \tag{5.52}$$

(b) The probability matrix at time 0 is $\mathbf{p} = [1, 0, 0]^T$.

birds_on_aviary.m

```
T = [1/2 1/4 1/6
     1/3 1/2 1/3
     1/6 1/4 1/2];

p0 = [1 0 0]';

p1 = T*p0
p2 = T*p1
p3 = T*p2
```

Output

```
p1 =
    0.5000
    0.3333
    0.1667
p2 =
    0.3611
    0.3889
    0.2500
p3 =
    0.3194
    0.3981
    0.2824
```

## 5.9.2. Predicting the Distant Future: Steady-State Vectors

**The most interesting aspect of Markov chains** is the study of the chain's long term behavior.

**Example 5.65. (Revisit to Example 5.64)**
What can be said in Example 5.64 about the bird population "in the long run"? What happens if the chain starts with other initial vectors?

**Solution**.

birds_on_aviary2.m
```
T = [1/2 1/4 1/6
     1/3 1/2 1/3
     1/6 1/4 1/2];

p = [1 0 0]'; q = [0 0 1]';
fprintf('p_%-2d = [%.5f %.5f %.5f];  q_%-2d = [%.5f %.5f %.5f]\n',0,p,0,q)
fprintf('%s\n',repelem('-',1,68))

n=12;
for k=1:n
    p = T*p; q = T*q;
    fprintf('p_%-2d = [%.5f %.5f %.5f];  q_%-2d = [%.5f %.5f %.5f]\n',k,p,k,q)
end
```

Output
```
p_0  = [1.00000 0.00000 0.00000];  q_0  = [0.00000 0.00000 1.00000]
--------------------------------------------------------------------
p_1  = [0.50000 0.33333 0.16667];  q_1  = [0.16667 0.33333 0.50000]
p_2  = [0.36111 0.38889 0.25000];  q_2  = [0.25000 0.38889 0.36111]
p_3  = [0.31944 0.39815 0.28241];  q_3  = [0.28241 0.39815 0.31944]
p_4  = [0.30633 0.39969 0.29398];  q_4  = [0.29398 0.39969 0.30633]
p_5  = [0.30208 0.39995 0.29797];  q_5  = [0.29797 0.39995 0.30208]
p_6  = [0.30069 0.39999 0.29932];  q_6  = [0.29932 0.39999 0.30069]
p_7  = [0.30023 0.40000 0.29977];  q_7  = [0.29977 0.40000 0.30023]
p_8  = [0.30008 0.40000 0.29992];  q_8  = [0.29992 0.40000 0.30008]
p_9  = [0.30003 0.40000 0.29997];  q_9  = [0.29997 0.40000 0.30003]
p_10 = [0.30001 0.40000 0.29999];  q_10 = [0.29999 0.40000 0.30001]
p_11 = [0.30000 0.40000 0.30000];  q_11 = [0.30000 0.40000 0.30000]
p_12 = [0.30000 0.40000 0.30000];  q_12 = [0.30000 0.40000 0.30000]
```

## Steady-State Vectors

**Definition 5.66.** If $T$ is a stochastic matrix, then a **steady-state vector** for $T$ is a probability vector $\mathbf{q}$ such that

$$T\mathbf{q} = \mathbf{q}. \tag{5.53}$$

**Note**: The steady-state vector $\mathbf{q}$ can be seen as an eigenvector of $T$, of which the corresponding eigenvalue $\lambda = 1$.

**Strategy 5.67. How to Find a Steady-State Vector**

(a) First, solve for $\mathbf{x} = [x_1, x_2, \cdots, x_n]^T$:

$$T\mathbf{x} = \mathbf{x} \;\Leftrightarrow\; T\mathbf{x} - \mathbf{x} = \mathbf{0} \;\Leftrightarrow\; (T - I)\mathbf{x} = \mathbf{0}. \tag{5.54}$$

(b) Then, set

$$\mathbf{q} = \frac{1}{x_1 + x_2 + \cdots + x_n}\mathbf{x}. \tag{5.55}$$

**Example 5.68.** Let $T = \begin{bmatrix} 0.6 & 0.3 \\ 0.4 & 0.7 \end{bmatrix}$. Find a steady-state vector for $T$.

**Solution**.

*Ans*: $\mathbf{q} = \begin{bmatrix} 3/7 \\ 4/7 \end{bmatrix}$

**Definition 5.69.** A stochastic matrix $T$ is **regular** if some matrix power $T^k$ contains only strictly positive entries.

**Interpretation 5.70.** If the transition matrix of a Markov chain is *regular*, then for some $k$ it is possible to go from any state to any other states (including remaining in the current state) in exactly $k$ steps.

**Example 5.71.** Let $T = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.2 & 0 & 0.4 \end{bmatrix}$.

(a) Is $T$ regular?

(b) Find a steady-state vector for $T$, using the **power method**.

**Solution**.

regular_stochastic.m

```
T = [0.5 0.2 0.3
     0.3 0.8 0.3
     0.2  0  0.4];

T2 = T*T;
disp('T^2 ='); disp(T2)

% The Power Method
x = [1 0 0]';
for k = 1:20
 x = T*x;
 fprintf('x_%-2d=[%.5f %.5f %.5f]\n',k,x)
end
```

Output

```
T^2 =
    0.3700    0.2600    0.3300
    0.4500    0.7000    0.4500
    0.1800    0.0400    0.2200

x_1 =[0.50000 0.30000 0.20000]
x_2 =[0.37000 0.45000 0.18000]
x_3 =[0.32900 0.52500 0.14600]
x_4 =[0.31330 0.56250 0.12420]
x_5 =[0.30641 0.58125 0.11234]
x_6 =[0.30316 0.59062 0.10622]
x_7 =[0.30157 0.59531 0.10312]
x_8 =[0.30078 0.59766 0.10156]
x_9 =[0.30039 0.59883 0.10078]
x_10=[0.30020 0.59941 0.10039]
x_11=[0.30010 0.59971 0.10020]
x_12=[0.30005 0.59985 0.10010]
x_13=[0.30002 0.59993 0.10005]
x_14=[0.30001 0.59996 0.10002]
x_15=[0.30001 0.59998 0.10001]
x_16=[0.30000 0.59999 0.10001]
x_17=[0.30000 0.60000 0.10000]
x_18=[0.30000 0.60000 0.10000]
x_19=[0.30000 0.60000 0.10000]
x_20=[0.30000 0.60000 0.10000]
```

**Theorem 5.72.** *If $T$ is an $n \times n$ regular stochastic matrix, then $T$ has a unique steady-state vector* $\mathbf{q}$.

(a) *The entries of* $\mathbf{q}$ *are strictly positive.*

(b) *The steady-state vector*

$$\mathbf{q} = \lim_{k\to\infty} T^k \mathbf{x}_0, \tag{5.56}$$

*for* **any initial probability vector** $\mathbf{x}_0$.

**Remark 5.73.** Let $T \in \mathbb{R}^{n\times n}$ be a regular stochastic matrix. Then

- If $T\mathbf{v} = \lambda\mathbf{v}$, then $|\lambda| \le 1$.
  (The above is true for every stochastic matrix; see § A.2.)
- Every column of $T^k$ converges to $\mathbf{q}$ as $k \to \infty$, i.e.,

$$T^k \to [\mathbf{q}\ \ \mathbf{q}\ \ \cdots\ \ \mathbf{q}] \in \mathbb{R}^{n\times n}, \quad \text{as } k \to \infty. \tag{5.57}$$

See Exercise 3 on p. 178.

**Example 5.74.** Let a regular stochastic matrix be given as in Example 5.71: $T = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.2 & 0 & 0.4 \end{bmatrix}$.

(a) Find the steady-state vector $\mathbf{q}$, by deriving the RREF.

(b) Find $T^{10}$ and $T^{20}$.

**Solution**. (a) Use Strategy 5.67.

regular_stochastic_Tk.m

```
T = [0.5 0.2 0.3
     0.3 0.8 0.3
     0.2  0  0.4];
Tk = eye(3);
rref(T-Tk)

for k = 1:20
    Tk = Tk*T;
    if(mod(k,10)==0), fprintf('T^%d =\n',k); disp(Tk) end
end
```

Output

```
ans =
    1.0000         0   -3.0000
         0    1.0000   -6.0000
         0         0         0

T^10 =
    0.3002    0.2999    0.3002
    0.5994    0.6004    0.5994
    0.1004    0.0997    0.1004

T^20 =
    0.3000    0.3000    0.3000
    0.6000    0.6000    0.6000
    0.1000    0.1000    0.1000
```

**True-or-False 5.75.** Let $T$ be a stochastic matrix.

a. The steady-state vector is an eigenvector of $T$.
b. Every eigenvector of $T$ is a steady-state vector.
c. The all-ones vector is an eigenvector of $T^T$.
d. The number 2 can be an eigenvalue of $T$ or $T^T$.
e. All stochastic matrices are regular.

*Ans*: T, F, T, F, F

## Exercises 5.9

1. Find the steady-state vector, deriving the RREF.

(a) $\begin{bmatrix} .1 & .6 \\ .9 & .4 \end{bmatrix}$ (b) $\begin{bmatrix} .7 & .1 & .1 \\ .2 & .8 & .2 \\ .1 & .1 & .7 \end{bmatrix}$

*Ans*: (b) $[1/4, 1/2, 1/4]^T$

2. The weather in Starkville, MS, is either good, indifferent, or bad on any given day.

- If the weather is good today, there is a 60% chance the weather will be good tomorrow, a 30% chance the weather will be indifferent, and a 10% chance the weather will be bad.
- If the weather is indifferent today, it will be good tomorrow with probability .40 and indifferent with probability .30.
- Finally, if the weather is bad today, it will be good tomorrow with probability .40 and indifferent with probability .50.

(a) What is the stochastic matrix for this situation?
(b) Suppose there is a 50% chance of good weather today and a 50% chance of indifferent weather. What are the chances of bad weather tomorrow?
(c) Suppose the predicted weather for Monday is 40% indifferent weather and 60% bad weather. What are the chances for good weather on Wednesday?

*Ans*: (b) 20%

3. Let $T \in \mathbb{R}^{n\times n}$ be a regular stochastic matrix. Prove (5.57).

***Hint***: Lrt $T = [\mathbf{t}_1, \mathbf{t}_2, \cdot, \mathbf{t}_n]$, where $\mathbf{t}_j$ is the $j$-th column of $T$. Take $\mathbf{x}_0 = \mathbf{e}_i$, for some $i$. Then

$$\mathbf{x}_1 = T\mathbf{x}_0 = T\mathbf{e}_i = \mathbf{t}_i,$$

which implies that $\mathbf{x}_1$ is the $i$-th column of $T$. From the above we have

$$\mathbf{x}_k = T^k\mathbf{x}_0 = T^k\mathbf{e}_i. \tag{5.58}$$

Thus $\mathbf{x}_k$ is the $i$-th column of $T^k$. Now, use Theorem 5.72.

4. M Generate a regular stochastic matrix of dimension 5.

(a) Find all eigenvalues and corresponding eigenvectors, using e.g. `eig` in Matlab.
(b) Express the eigenvector corresponding to $\lambda = 1$ as a probability vector $\mathbf{p}$.
(c) Use the power method to find a steady-state vector $\mathbf{q}$, beginning with $\mathbf{x}_0 = \mathbf{e}_1$.
(d) Compare $\mathbf{p}$ with $\mathbf{q}$.

CHAPTER **6**

# Orthogonality and Least-Squares

In this chapter, we will learn

- Inner product, length, and orthogonality,
- Orthogonal projections,
- The Gram-Schmidt process, which is an algorithm to produce an orthogonal basis for any nonzero subspace of $\mathbb{R}^n$, and
- Least-Squares problems, with applications to linear models.

**Contents of Chapter 6**


6.1. Inner Product, Length, and Orthogonality . . . . . 180
6.2. Orthogonal Sets . . . . . 187
6.3. Orthogonal Projections . . . . . 196
6.4. The Gram-Schmidt Process and QR Factorization . . . . . 203
6.5. Least-Squares Problems . . . . . 210
6.6. Machine Learning: Regression Analysis . . . . . 217

# 6.1. Inner Product, Length, and Orthogonality

## 6.1.1. Inner Product and Length

**Definition 6.1.** Let $\mathbf{u} = [u_1, u_2, \cdots, u_n]^T$ and $\mathbf{v} = [v_1, v_2, \cdots, v_n]^T$ are vectors in $\mathbb{R}^n$. Then, the **inner product** (or **dot product**) of u and v is given by

$$\begin{aligned} \mathbf{u} \bullet \mathbf{v} &= \mathbf{u}^T \mathbf{v} = [u_1 \ u_2 \ \cdots \ u_n] \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \\ &= u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{k=1}^{n} u_k v_k. \end{aligned} \tag{6.1}$$

**Example 6.2.** Let $\mathbf{u} = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ -4 \end{bmatrix}$. Find $\mathbf{u} \bullet \mathbf{v}$.

**Solution**.

**Theorem 6.3.** *Let* $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *be vectors in* $\mathbb{R}^n$, *and* $c$ *be a scalar. Then*

*a.* $\mathbf{u} \bullet \mathbf{v} = \mathbf{v} \bullet \mathbf{u}$

*b.* $(\mathbf{u} + \mathbf{v}) \bullet \mathbf{w} = \mathbf{u} \bullet \mathbf{w} + \mathbf{v} \bullet \mathbf{w}$

*c.* $(c\mathbf{u}) \bullet \mathbf{v} = c(\mathbf{u} \bullet \mathbf{v}) = \mathbf{u} \bullet (c\mathbf{v})$

*d.* $\mathbf{u} \bullet \mathbf{u} \ge 0$, *and* $\mathbf{u} \bullet \mathbf{u} = 0 \Leftrightarrow \mathbf{u} = 0$

**Definition 6.4.** The **length** (**norm**) of $\mathbf{v}$ is nonnegative scalar $\|\mathbf{v}\|$ defined by

$$\|\mathbf{v}\| = \sqrt{\mathbf{v}\bullet\mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} \quad \text{and} \quad \|\mathbf{v}\|^2 = \mathbf{v}\bullet\mathbf{v}. \tag{6.2}$$

**Note**: For any scalar $c$, $\|c\mathbf{v}\| = |c|\,\|\mathbf{v}\|$.

**Example 6.5.** Let $W$ be a subspace of $\mathbb{R}^2$ spanned by $\mathbf{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$. Find a **unit vector** $\mathbf{u}$ that is a basis for $W$.
**Solution**.

## Distance in $\mathbb{R}^n$

**Definition 6.6.** For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, the **distance** between $\mathbf{u}$ and $\mathbf{v}$ is

$$\textit{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|, \tag{6.3}$$

the length of the vector $\mathbf{u} - \mathbf{v}$.

**Example 6.7.** Compute the distance between the vectors $\mathbf{u} = (7, 1)$ and $\mathbf{v} = (3, 2)$.
**Solution**.

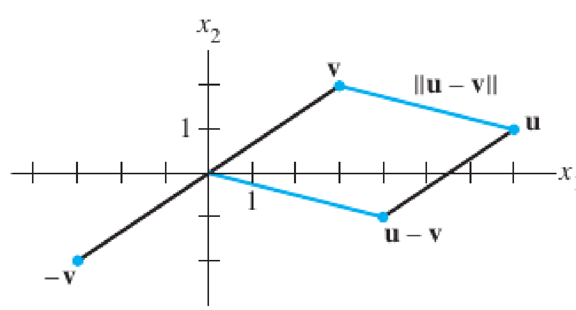



Figure 6.1: The distance between $\mathbf{u}$ and $\mathbf{v}$ is the length of $\mathbf{u} - \mathbf{v}$.

**Example 6.8.** Let $\mathbf{u} = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ -4 \end{bmatrix}$. Find the distance between $\mathbf{u}$ and $\mathbf{v}$.

**Solution**.

**Note**: The **inner product** can be defined as

$$\mathbf{u} \bullet \mathbf{v} = \|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta, \tag{6.4}$$

where $\theta$ is the angle between $\mathbf{u}$ and $\mathbf{v}$.

**Example 6.9.** Let $\mathbf{u} = \begin{bmatrix} 1 \\ \sqrt{3} \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -1/2 \\ \sqrt{3}/2 \end{bmatrix}$. Use (6.4) to find the angle between $\mathbf{u}$ and $\mathbf{v}$.

**Solution**.

## 6.1.2. Orthogonal Vectors

**Definition 6.10.** Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are **orthogonal** if $\mathbf{u} \bullet \mathbf{v} = 0$.

**Theorem 6.11. The Pythagorean Theorem**: *Two vectors* $\mathbf{u}$ *and* $\mathbf{v}$ *are orthogonal if and only if*

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2. \tag{6.5}$$

**Proof**. For all $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$,

$$\|\mathbf{u}+\mathbf{v}\|^2 = (\mathbf{u}+\mathbf{v})\bullet(\mathbf{u}+\mathbf{v}) = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\mathbf{u}\bullet\mathbf{v}. \tag{6.6}$$

Thus, $\mathbf{u}$ and $\mathbf{v}$ are orthogonal $\Leftrightarrow$ (6.5) holds □

### Orthogonal Complements

**Definition 6.12.** Let $W \subset \mathbb{R}^n$ be a subspace. A vector $\mathbf{z} \in \mathbb{R}^n$ is said to be **orthogonal** to $W$ if $\mathbf{z}\bullet\mathbf{w} = 0$ for all $\mathbf{w} \in W$. The set of all vectors $\mathbf{z}$ that are orthogonal to $W$ is called the **orthogonal complement** of $W$ and is denoted by $W^\perp$ (and read as "$W$ perpendicular" or simply "$W$ perp"). That is,

$$W^\perp = \{\mathbf{z} \mid \mathbf{z}\bullet\mathbf{w} = 0, \ \forall\, \mathbf{w} \in W\}. \tag{6.7}$$

**Example 6.13.** Let $W$ be a plane through the origin in $\mathbb{R}^3$, and let $L$ be the line through the origin and perpendicular to $W$. If $\mathbf{z} \in L$ and and $\mathbf{w} \in W$, then

$$\mathbf{z}\bullet\mathbf{w} = 0.$$

In fact, $L$ consists of all vectors that are orthogonal to the $\mathbf{w}$'s in $W$, and $W$ consists of all vectors orthogonal to the $\mathbf{z}$'s in $L$. That is,

$$L = W^\perp \quad \text{and} \quad W = L^\perp. \tag{6.8}$$

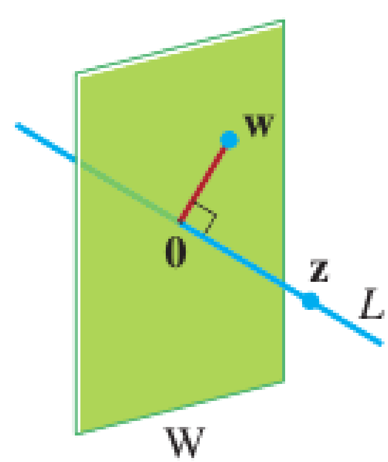



Figure 6.2: A plane and line through the origin as orthogonal complements.

**Example 6.14.** Let $W$ is a subspace of $\mathbb{R}^n$. Prove that if $\mathrm{x} \in W$ and $\mathrm{x} \in W^\perp$, then $\mathrm{x} = 0$.

**Solution**. ***Hint***: Let $\mathrm{x} \in W$. The condition $\mathrm{x} \in W^\perp$ implies that x is perpendicular to every element in $W$, particularly to itself.

**Remark 6.15.** Let $W$ be a subspace of $\mathbb{R}^n$, with ***dim*** $W = m \le n$.

(a) Consider a basis for $W$. Let $A$ be the collection of the basis vectors. Then $A \in \mathbb{R}^{n\times m}$, $W =$ ***Col*** $A$, and

$$W^\perp = \{\mathbf{x} \in \mathbb{R}^n \mid A^T\mathbf{x} = 0\} = \textbf{\textit{Nul}}\, A^T. \tag{6.9}$$

Note that $A^T \in \mathbb{R}^{m\times n}$, $m \le n$, and

$$m = \textbf{\textit{dim}}\, W = \textbf{\textit{dim}}\, \textbf{\textit{Col}}\, A = \textbf{\textit{dim}}\, \textbf{\textit{Col}}\, A^T. \tag{6.10}$$

(b) A vector x is in $W^\perp \;\Leftrightarrow\;$ x is orthogonal to **every vector in a spanning set of *W***.

(c) $W^\perp$ is a subspace of $\mathbb{R}^n$.

**Example 6.16.** Let $W$ be a subspace of $\mathbb{R}^n$. Prove that

$$\textbf{\textit{dim}}\, W + \textbf{\textit{dim}}\, W^\perp = n. \tag{6.11}$$

**Solution**. ***Hint***: Use Remark 6.15 (a).

**Definition 6.17.** Let $A$ be an $m \times n$ matrix. Then the **row space** is the set of all linear combinations of of the rows of $A$, denoted by *Row* $A$. That is, $Row\, A = Col\, A^T$.

**Theorem 6.18.** *Let $A$ be an $m \times n$ matrix. The orthogonal complement of the row space of $A$ is the null space of $A$, and the orthogonal complement of the column space of $A$ is the null space of $A^T$:*

$$(Row\, A)^\perp = Nul\, A \quad \text{and} \quad (Col\, A)^\perp = Nul\, A^T. \tag{6.12}$$

**True-or-False 6.19.**

a. For any scalar $c$, $\mathbf{u}\bullet(c\mathbf{v}) = c(\mathbf{u}\bullet\mathbf{v})$.

b. If the distance from $\mathbf{u}$ to $\mathbf{v}$ equals the distance from $\mathbf{u}$ to $-\mathbf{v}$, then $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

c. For a square matrix $A$, vectors in *Col* $A$ are orthogonal to vectors in *Nul* $A$.

d. For an $m \times n$ matrix $A$, vectors in the null space of $A$ are orthogonal to vectors in the row space of $A$.

**Solution**.

*Ans*: T,T,F,T

## Exercises 6.1

1. Let $\mathbf{x} = \begin{bmatrix} 3 \\ -1 \\ -5 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 6 \\ -2 \\ 3 \end{bmatrix}$. Find $\mathbf{x}\bullet\mathbf{x}$, $\mathbf{x}\bullet\mathbf{w}$, and $\dfrac{\mathbf{x}\bullet\mathbf{w}}{\mathbf{x}\bullet\mathbf{x}}$.

   *Ans*: 35, 5, 1/7

2. Find the distance between $\mathbf{u} = \begin{bmatrix} 0 \\ -5 \\ 2 \end{bmatrix}$ and $\mathbf{z} = \begin{bmatrix} -4 \\ -1 \\ 8 \end{bmatrix}$.

   *Ans*: $2\sqrt{17}$

3. Verify the **parallelogram law** for vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$:

$$\|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2 = 2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2. \tag{6.13}$$

   ***Hint***: Use (6.6).

4. Let $\mathbf{u} = \begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -7 \\ -4 \\ 6 \end{bmatrix}$.

   (a) Compute $\mathbf{u}\bullet\mathbf{v}$, $\|\mathbf{u}\|$, $\|\mathbf{v}\|$, $\|\mathbf{u}+\mathbf{v}\|$, and $\|\mathbf{u}-\mathbf{v}\|$.

   (b) Verify (6.6) and (6.13).

5. Suppose $\mathbf{y}$ is orthogonal to $\mathbf{u}$ and $\mathbf{v}$. Show that $\mathbf{y}$ is orthogonal to every $\mathbf{w}$ in *Span*$\{\mathbf{u}, \mathbf{v}\}$.

# 6.2. Orthogonal Sets

**Definition 6.20.** A set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is said to be an **orthogonal set** if each pair of distinct vectors from the set is orthogonal. That is $\mathbf{u}_i \bullet \mathbf{u}_j = 0$, for $i \neq j$.

**Example 6.21.** Let $\mathbf{u}_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$, and $\mathbf{u}_3 = \begin{bmatrix} -5 \\ -2 \\ 1 \end{bmatrix}$. Is the set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ orthogonal?
**Solution**.

**Theorem 6.22.** *If $S = \{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then $S$ is linearly independent and therefore forms a basis for the subspace spanned by $S$.*

**Proof**. It suffices to prove that $S$ is linearly independent. Suppose

$$c_1\,\mathbf{u}_1 + c_2\,\mathbf{u}_2 + \cdots + c_p\,\mathbf{u}_p = \mathbf{0}.$$

Take the dot product with $\mathbf{u}_1$. Then the above equation becomes

$$c_1\mathbf{u}_1 \bullet \mathbf{u}_1 + c_2\mathbf{u}_1 \bullet \mathbf{u}_2 + \cdots + c_p\mathbf{u}_1 \bullet \mathbf{u}_p = 0,$$

from which we conclude $c_1 = 0$. Similarly, by taking the dot product with $\mathbf{u}_i$, we can get $c_i = 0$. That is,

$$c_1 = c_2 = \cdots = c_p = 0,$$

which completes the proof. □

**Definition 6.23.** An **orthogonal basis** for a subspace $W$ of $\mathbb{R}^n$ is a basis for $W$ that is also an orthogonal set.

The following theorem shows one of reasons why orthogonality is a useful property in vector spaces and matrix algebra.

**Theorem 6.24.** *Let $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace $W$ of $\mathbb{R}^n$. For each $\mathbf{y}$ in $W$, the weights in the linear combination*

$$\mathbf{y} = c_1\,\mathbf{u}_1 + c_2\,\mathbf{u}_2 + \cdots + c_p\,\mathbf{u}_p \tag{6.14}$$

*are given by*

$$c_j = \frac{\mathbf{y}\bullet\mathbf{u}_j}{\mathbf{u}_j\bullet\mathbf{u}_j} \quad (j = 1, 2, \cdots, p). \tag{6.15}$$

**Proof**. $\mathbf{y}\bullet\mathbf{u}_j = (c_1\,\mathbf{u}_1 + c_2\,\mathbf{u}_2 + \cdots + c_p\,\mathbf{u}_p)\bullet\mathbf{u}_j = c_j\mathbf{u}_j\bullet\mathbf{u}_j = c_j\|\mathbf{u}_j\|^2$. □

**Example 6.25.** Let $\mathbf{u}_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$, and $\mathbf{u}_3 = \begin{bmatrix} -5 \\ -2 \\ 1 \end{bmatrix}$. In Example 6.21, we have seen that $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is orthogonal. Express the vector $\mathbf{y} = \begin{bmatrix} 11 \\ 0 \\ -5 \end{bmatrix}$ as a linear combination of the vectors in $S$.

**Solution**.

*Ans*: $\mathbf{y} = \mathbf{u}_1 - 2\mathbf{u}_2 - 2\mathbf{u}_3$.

### 6.2.1. An Orthogonal Projection

Given a nonzero vector u in $\mathbb{R}^n$, consider the problem of **decomposing a vector** $\mathbf{y} \in \mathbb{R}^n$ **into sum of two vectors**, one a multiple of u and the other orthogonal to u:

$$\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{z}, \quad \widehat{\mathbf{y}} \,//\, \mathbf{u} \text{ and } \mathbf{z} \perp \mathbf{u}. \tag{6.16}$$

Let $\widehat{\mathbf{y}} = \alpha\mathbf{u}$. Then $\mathbf{z} = \mathbf{y} - \alpha\mathbf{u}$ and

$$0 = \mathbf{z}\bullet\mathbf{u} = (\mathbf{y}-\alpha\mathbf{u})\bullet\mathbf{u} = \mathbf{y}\bullet\mathbf{u}-\alpha\mathbf{u}\bullet\mathbf{u}.$$

Thus $\alpha = \mathbf{y}\bullet\mathbf{u}/\mathbf{u}\bullet\mathbf{u}$.

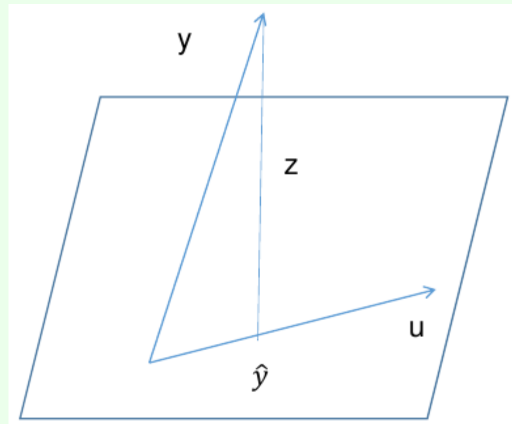


Figure 6.3: Orthogonal projection: $\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{z}$.

**Definition 6.26.** Given a nonzero vector u in $\mathbb{R}^n$, for $\mathbf{y} \in \mathbb{R}^n$, let

$$\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{z}, \quad \widehat{\mathbf{y}} \,//\, \mathbf{u} \text{ and } \mathbf{z} \perp \mathbf{u}. \tag{6.17}$$

Then

$$\widehat{\mathbf{y}} = \alpha\mathbf{u} = \frac{\mathbf{y}\bullet\mathbf{u}}{\mathbf{u}\bullet\mathbf{u}}\mathbf{u}, \quad \mathbf{z} = \mathbf{y} - \widehat{\mathbf{y}}. \tag{6.18}$$

The vector $\widehat{\mathbf{y}}$ is called the **orthogonal projection of y onto u**, and z is called the **component of y orthogonal to u**.

- Let $L = Span\{\mathbf{u}\}$. Then we denote

$$\widehat{\mathbf{y}} = \frac{\mathbf{y}\bullet\mathbf{u}}{\mathbf{u}\bullet\mathbf{u}}\mathbf{u} = \text{proj}_L\mathbf{y}, \tag{6.19}$$

  which is called the **orthogonal projection of y onto** $L$.
- It is meaningful whether the angle between y and u is **acute or obtuse**.

**Example 6.27.** Let $\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$.

(a) Find the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$.

(b) Write $\mathbf{y}$ as the sum of two orthogonal vectors, one in $L = Span\{\mathbf{u}\}$ and one orthogonal to $\mathbf{u}$.

(c) Find the distance from $\mathbf{y}$ to $L$.

**Solution.**

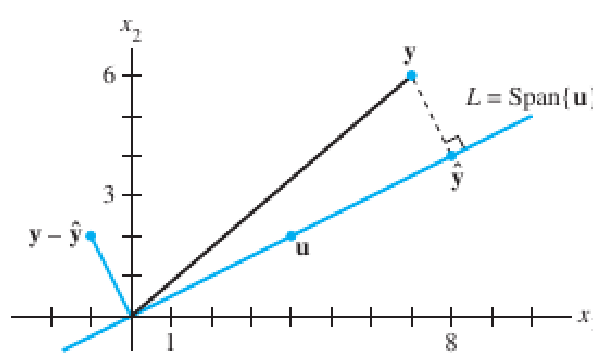



Figure 6.4: The orthogonal projection of $\mathbf{y}$ onto $L = Span\{\mathbf{u}\}$.

**Example 6.28.** Let $\mathbf{y} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$.

(a) Find the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$.

(b) Write $\mathbf{y}$ as the sum of a vector in *Span*$\{\mathbf{u}\}$ and one orthogonal to $\mathbf{u}$.

**Solution**.

**Example 6.29.** Let $\mathbf{v} = \begin{bmatrix} 4 \\ -12 \\ 8 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}$. Find the distance from $\mathbf{v}$ to *Span*$\{\mathbf{w}\}$.

**Solution**.

## 6.2.2. Orthonormal Basis and Orthogonal Matrix

**Definition 6.30.** A set $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p\}$ is an **orthonormal set**, if it is an orthogonal set of *unit* vectors. If $W$ is the subspace spanned by such a set, then $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p\}$ is an **orthonormal basis** for $W$, since the set is automatically linearly independent.

**Example 6.31.** In Example 6.21, p. 187, we know $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} -5 \\ -2 \\ 1 \end{bmatrix}$ form an *orthogonal* basis for $\mathbb{R}^3$. Find the corresponding *orthonormal* basis.

**Solution**.

**Theorem 6.32.** *An $m \times n$ matrix $U$ has orthonormal columns if and only if $U^TU = I$.*

**Proof**. To simplify notation, we suppose that $U$ has only three columns: $U = [\mathbf{u}_1\ \ \mathbf{u}_2\ \ \mathbf{u}_3]$, $\mathbf{u}_i \in \mathbb{R}^m$. Then

$$U^TU = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \mathbf{u}_3^T \end{bmatrix} [\mathbf{u}_1\ \ \mathbf{u}_2\ \ \mathbf{u}_3] = \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \mathbf{u}_1^T\mathbf{u}_3 \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \mathbf{u}_2^T\mathbf{u}_3 \\ \mathbf{u}_3^T\mathbf{u}_1 & \mathbf{u}_3^T\mathbf{u}_2 & \mathbf{u}_3^T\mathbf{u}_3 \end{bmatrix}.$$

Thus, $U$ has orthonormal columns $\Leftrightarrow U^TU = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. The proof of the general case is essentially the same. □

**Theorem 6.33.** *Let $U$ be an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then*

*(a)* $\|U\mathbf{x}\| = \|\mathbf{x}\|$ (length preservation)

*(b)* $(U\mathbf{x})\bullet(U\mathbf{y}) = \mathbf{x}\bullet\mathbf{y}$ (dot product preservation)

*(c)* $(U\mathbf{x})\bullet(U\mathbf{y}) = 0 \;\Leftrightarrow\; \mathbf{x}\bullet\mathbf{y} = 0$ (orthogonality preservation)

**Proof**.

Theorems 6.32 & 6.33 are particularly useful when applied to square matrices.

**Definition 6.34.** An **orthogonal matrix** is a square matrix $U$ such that $U^T = U^{-1}$, i.e.,

$$U \in \mathbb{R}^{n\times n} \quad \text{and} \quad U^T U = I. \tag{6.20}$$

Let's generate a **random orthogonal matrix** and test it.

orthogonal_matrix.m

```
n = 4;

[Q,~] = qr(rand(n));
U = Q;

disp("U =");   disp(U)
disp("U'*U ="); disp(U'*U)

x = rand([n,1]);
fprintf("\nx' ="); disp(x')
fprintf("||x||_2  =");disp(norm(x,2))
fprintf("||U*x||_2=");disp(norm(U*x,2))
```

Output

```
U =
   -0.3770    0.6893    0.2283   -0.5750
   -0.3786   -0.2573   -0.8040   -0.3795
   -0.6061    0.3149   -0.1524    0.7143
   -0.5892   -0.5996    0.5274   -0.1231
U'*U =
    1.0000    0.0000    0.0000   -0.0000
    0.0000    1.0000   -0.0000    0.0000
    0.0000   -0.0000    1.0000   -0.0000
   -0.0000    0.0000   -0.0000    1.0000

x' =   0.4709    0.2305    0.8443    0.1948
||x||_2  =    1.0128
||U*x||_2=    1.0128
```

**True-or-False 6.35.**

a. If $\mathbf{y}$ is a linear combination of nonzero vectors from an orthogonal set, then the weights in the linear combination can be computed without row operations on a matrix.

b. If the vectors in an orthogonal set of nonzero vectors are normalized, then some of the new vectors may not be orthogonal.

c. A matrix with orthonormal columns is an orthogonal matrix.

d. If $L$ is a line through $\mathbf{0}$ and if $\widehat{\mathbf{y}}$ is the orthogonal projection of $\mathbf{y}$ onto $L$, then $\|\widehat{\mathbf{y}}\|$ gives the distance from $\mathbf{y}$ to $L$.

e. Every orthogonal set in $\mathbb{R}^n$ is linearly independent.

f. If the columns of an $m \times n$ matrix $A$ are orthonormal, then the linear mapping $\mathbf{x} \mapsto A\mathbf{x}$ preserves lengths.

**Solution**.

*Ans*: T,F,F,F,F,T

## Exercises 6.2

1. Determine which sets of vectors are orthogonal.

   (a) $\begin{bmatrix} 2 \\ -7 \\ 1 \end{bmatrix}, \begin{bmatrix} -6 \\ -3 \\ 9 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}$ (b) $\begin{bmatrix} 2 \\ -5 \\ -3 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ -2 \\ 6 \end{bmatrix}$

2. Let $\mathbf{u}_1 = \begin{bmatrix} 3 \\ -3 \\ 0 \end{bmatrix}, \ \mathbf{u}_2 = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}, \ \mathbf{u}_3 = \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} 5 \\ -3 \\ 1 \end{bmatrix}$.

   (a) Check if $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal basis for $\mathbb{R}^3$.

   (b) Express x as a linear combination of $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.

   *Ans*: $\mathbf{x} = \frac{4}{3}\mathbf{u}_1 + \frac{1}{3}\mathbf{u}_2 + \frac{1}{3}\mathbf{u}_3$

3. Compute the orthogonal projection of $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ onto the line through $\begin{bmatrix} -1 \\ 3 \end{bmatrix}$ and the origin.

4. Determine which sets of vectors are orthonormal. If a set is only orthogonal, normalize the vectors to produce an orthonormal set.

   (a) $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$ (b) $\begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \begin{bmatrix} -1/2 \\ 0 \\ 1/2 \end{bmatrix}$ (c) $\begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -2/3 \\ 1/3 \\ -2/3 \end{bmatrix}$

5. Let $U$ and $V$ be $n \times n$ orthogonal matrices. Prove that $UV$ is an orthogonal matrix. ***Hint***: See Definition 6.34, where $U^{-1} = U^T \ \Leftrightarrow \ U^T U = I$.

# 6.3. Orthogonal Projections

**Recall**: **(Definition 6.26)** Given a nonzero vector $\mathbf{u}$ in $\mathbb{R}^n$, for $\mathbf{y} \in \mathbb{R}^n$, let

$$\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{z}, \quad \widehat{\mathbf{y}} \,/\!/\, \mathbf{u} \text{ and } \mathbf{z} \perp \mathbf{u}. \tag{6.21}$$

Then

$$\widehat{\mathbf{y}} = \alpha \mathbf{u} = \frac{\mathbf{y} \bullet \mathbf{u}}{\mathbf{u} \bullet \mathbf{u}} \mathbf{u}, \quad \mathbf{z} = \mathbf{y} - \widehat{\mathbf{y}}. \tag{6.22}$$

The vector $\widehat{\mathbf{y}}$ is called the ***orthogonal projection* of y onto u**, and $\mathbf{z}$ is called the component of $\mathbf{y}$ orthogonal to $\mathbf{u}$. Let $L = \mathit{Span}\{\mathbf{u}\}$. Then we denote

$$\widehat{\mathbf{y}} = \frac{\mathbf{y} \bullet \mathbf{u}}{\mathbf{u} \bullet \mathbf{u}} \mathbf{u} = \text{proj}_L \mathbf{y}, \tag{6.23}$$

which is called the orthogonal projection of $\mathbf{y}$ onto $L$.

We generalize this orthogonal projection to subspaces.

**Theorem 6.36. (The Orthogonal Decomposition Theorem)** *Let $W$ be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y} \in \mathbb{R}^n$ can be written uniquely in the form*

$$\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{z}, \tag{6.24}$$

*where $\widehat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^\perp$. In fact, if $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p\}$ is an* **orthogonal basis** *for $W$, then*

$$\begin{aligned} \widehat{\mathbf{y}} &= \text{proj}_W \mathbf{y} = \frac{\mathbf{y} \bullet \mathbf{u}_1}{\mathbf{u}_1 \bullet \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y} \bullet \mathbf{u}_2}{\mathbf{u}_2 \bullet \mathbf{u}_2} \mathbf{u}_2 + \cdots + \frac{\mathbf{y} \bullet \mathbf{u}_p}{\mathbf{u}_p \bullet \mathbf{u}_p} \mathbf{u}_p, \\ \mathbf{z} &= \mathbf{y} - \widehat{\mathbf{y}}. \end{aligned} \tag{6.25}$$

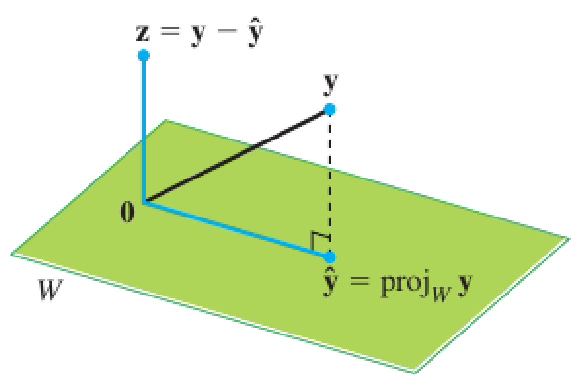



Figure 6.5: Orthogonal projection of $\mathbf{y}$ onto $W$.

**Example 6.37.** Let $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Observe that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for $W = \mathit{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$.

(a) Write $\mathbf{y}$ as the sum of a vector in $W$ and a vector orthogonal to $W$.

(b) Find the distance from $\mathbf{y}$ to $W$.

**Solution**. $\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{z} \;\Rightarrow\; \widehat{\mathbf{y}} = \dfrac{\mathbf{y} \bullet \mathbf{u_1}}{\mathbf{u_1} \bullet \mathbf{u_1}} \mathbf{u}_1 + \dfrac{\mathbf{y} \bullet \mathbf{u_2}}{\mathbf{u_2} \bullet \mathbf{u_2}} \mathbf{u}_2$ and $\mathbf{z} = \mathbf{y} - \widehat{\mathbf{y}}$.

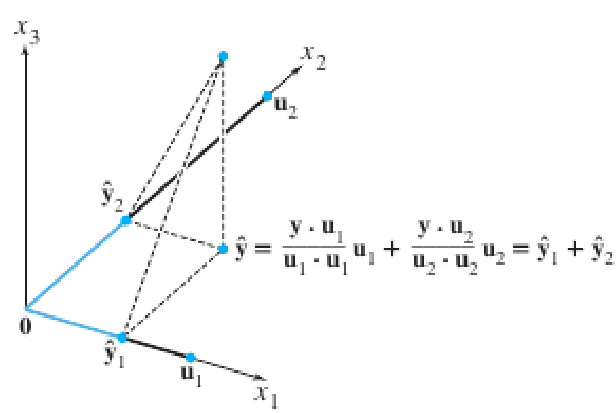



Figure 6.6: A geometric interpretation of the orthogonal projection.

**Remark 6.38. (Properties of Orthogonal Decomposition)**
Let $\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{z}$, where $\widehat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^\perp$. Then

1. $\widehat{\mathbf{y}}$ is called the **orthogonal projection** of $\mathbf{y}$ onto $W$ ($= \text{proj}_W \mathbf{y}$)
2. $\widehat{\mathbf{y}}$ is the **closest point** to $\mathbf{y}$ in $W$.
   (in the sense $\|\mathbf{y} - \widehat{\mathbf{y}}\| \le \|\mathbf{y} - \mathbf{v}\|$, for all $\mathbf{v} \in W$)
3. $\widehat{\mathbf{y}}$ is called the **best approximation** to $\mathbf{y}$ by elements of $W$.
4. If $\mathbf{y} \in W$, then $\text{proj}_W \mathbf{y} = \mathbf{y}$.

**Proof**. 2. For an arbitrary $\mathbf{v} \in W$, $\mathbf{y} - \mathbf{v} = (\mathbf{y} - \widehat{\mathbf{y}}) + (\widehat{\mathbf{y}} - \mathbf{v})$, where $(\widehat{\mathbf{y}} - \mathbf{v}) \in W$. Thus, by the *Pythagorean theorem*,

$$\|\mathbf{y} - \mathbf{v}\|^2 = \|\mathbf{y} - \widehat{\mathbf{y}}\|^2 + \|\widehat{\mathbf{y}} - \mathbf{v}\|^2,$$

which implies that $\|\mathbf{y} - \mathbf{v}\| \ge \|\mathbf{y} - \widehat{\mathbf{y}}\|$. □

**Example 6.39.** Find the closest point to $\mathbf{y}$ in the subspace $Span\{\mathbf{u}_1, \mathbf{u}_2\}$ and hence find the distance from $\mathbf{y}$ to $W$. (Notice that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal.)

$$\mathbf{y} = \begin{bmatrix} 3 \\ -1 \\ 1 \\ 13 \end{bmatrix}, \quad \mathbf{u}_1 = \begin{bmatrix} 1 \\ -2 \\ -1 \\ 2 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -4 \\ 1 \\ 0 \\ 3 \end{bmatrix}$$

**Solution**.

**Example 6.40.** Find the distance from y to the plane in $\mathbb{R}^3$ spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$.

$$\mathbf{y} = \begin{bmatrix} 5 \\ -9 \\ 5 \end{bmatrix}, \quad \mathbf{u}_1 = \begin{bmatrix} -3 \\ -5 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}$$

**Solution**.

**Example 6.41.** Let $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ be an orthogonal basis for a subspace $W$ of $\mathbb{R}^4$ and $\mathbf{v} \in W$:

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 1 \\ 1 \\ -2 \\ -1 \end{bmatrix}, \quad \text{and } \mathbf{v} = \begin{bmatrix} -3 \\ 7 \\ -6 \\ 2 \end{bmatrix}.$$

Write v as the sum of two vectors: one in $Span\{\mathbf{u}_1, \mathbf{u}_2\}$ and the other in $Span\{\mathbf{u}_3\}$.

**Solution**.

*Ans*: $\mathbf{u}_1 + 3\mathbf{u}_2$ and $2\mathbf{u}_3$.

**Theorem 6.42.** *If* $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p\}$ *is an orthonormal basis for a subspace* $W$ *of* $\mathbb{R}^n$*, then*

$$\text{proj}_W \mathbf{y} = (\mathbf{y}\bullet\mathbf{u_1})\,\mathbf{u}_1 + (\mathbf{y}\bullet\mathbf{u_2})\,\mathbf{u}_2 + \cdots + (\mathbf{y}\bullet\mathbf{u}_p)\,\mathbf{u}_p. \tag{6.26}$$

*If* $U = [\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_p]$*, then*

$$\text{proj}_W \mathbf{y} = UU^T\mathbf{y}, \quad \textit{for all } \mathbf{y} \in \mathbb{R}^n. \tag{6.27}$$

*The orthogonal projection can be viewed as a* ***matrix transformation****.*

**Proof**. Notice that

$$\begin{aligned}&(\mathbf{y}\bullet\mathbf{u_1})\,\mathbf{u}_1 + (\mathbf{y}\bullet\mathbf{u_2})\,\mathbf{u}_2 + \cdots + (\mathbf{y}\bullet\mathbf{u}_p)\,\mathbf{u}_p\\ &\quad= (\mathbf{u}_1^T\mathbf{y})\,\mathbf{u}_1 + (\mathbf{u}_2^T\mathbf{y})\,\mathbf{u}_2 + \cdots + (\mathbf{u}_p^T\mathbf{y})\,\mathbf{u}_p\\ &\quad= U(U^T\mathbf{y}).\end{aligned}$$

**Example 6.43.** Let $\mathbf{y} = \begin{bmatrix}7\\9\end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix}1/\sqrt{10}\\-3/\sqrt{10}\end{bmatrix}$, and $W = \mathit{Span}\{\mathbf{u}_1\}$.

(a) Let $U$ be the $2\times 1$ matrix whose only column is $\mathbf{u}_1$. Compute $U^TU$ and $UU^T$.

(b) Compute $\text{proj}_W \mathbf{y} = (\mathbf{y}\bullet\mathbf{u_1})\,\mathbf{u}_1$ and $UU^T\mathbf{y}$.

**Solution**.

*Ans*: (a) $UU^T = \frac{1}{10}\begin{bmatrix}1 & -3\\-3 & 9\end{bmatrix}$ (b) $\begin{bmatrix}-2\\6\end{bmatrix}$

**True-or-False 6.44.**

a. If $\mathbf{z}$ is orthogonal to $\mathbf{u}_1$ and $\mathbf{u}_2$ and if $W = \mathit{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, then $\mathbf{z}$ must be in $W^\perp$.

b. The orthogonal projection $\widehat{\mathbf{y}}$ of $\mathbf{y}$ onto a subspace $W$ can sometimes depend on the orthogonal basis for $W$ used to compute $\widehat{\mathbf{y}}$.

c. If the columns of an $n \times p$ matrix $U$ are orthonormal, then $UU^T\mathbf{y}$ is the orthogonal projection of $\mathbf{y}$ onto the column space of $U$.

d. If an $n \times p$ matrix $U$ has orthonormal columns, then $UU^T\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

**Solution**.

*Ans*: T,F,T,F

## Exercises 6.3

1. (i) Verify that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal set, (ii) find the orthogonal projection of $\mathbf{y}$ onto $W = Span\{\mathbf{u}_1, \mathbf{u}_2\}$, and (iii) write $\mathbf{y}$ as a sum of a vector in $W$ and a vector orthogonal to $W$.

   (a) $\mathbf{y} = \begin{bmatrix} -1 \\ 2 \\ 6 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}$ (b) $\mathbf{y} = \begin{bmatrix} -1 \\ 4 \\ 3 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -1 \\ 3 \\ -2 \end{bmatrix}$

   *Ans*: (b) $\mathbf{y} = \frac{1}{2}\begin{bmatrix} 3 \\ 7 \\ 2 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} -5 \\ 1 \\ 4 \end{bmatrix}$

2. Find the best approximation to $\mathbf{z}$ by vectors of the form $c_1\mathbf{v}_1 + c_2\mathbf{v}_2$.

   (a) $\mathbf{z} = \begin{bmatrix} 3 \\ -1 \\ 1 \\ 13 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \\ -1 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -4 \\ 1 \\ 0 \\ 3 \end{bmatrix}$ (b) $\mathbf{z} = \begin{bmatrix} 3 \\ -7 \\ 2 \\ 3 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 2 \\ -1 \\ -3 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}$

   *Ans*: (a) $\widehat{\mathbf{z}} = 3\mathbf{v}_1 + \mathbf{v}_2$

3. Let $\mathbf{z}$, $\mathbf{v}_1$, and $\mathbf{v}_2$ be given as in Exercise 2. Find the distance from $\mathbf{z}$ to the subspace of $\mathbb{R}^4$ spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$.

   *Ans*: (a) 8

4. Let $W$ be a subspace of $\mathbb{R}^n$. A transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is defined as
   $\mathbf{x} \mapsto T(\mathbf{x}) = \text{proj}_W \mathbf{x}$.

   (a) Prove that $T$ is a linear transformation.
   (b) Prove that $T(T(\mathbf{x})) = T(\mathbf{x})$.

   ***Hint***: Use Theorem 6.42.

# 6.4. The Gram-Schmidt Process and QR Factorization

## 6.4.1. The Gram-Schmidt Process

The **Gram-Schmidt process** is an algorithm to produce an **orthogonal or orthonormal basis** for any nonzero subspace of $\mathbb{R}^n$.

**Example 6.45.** Let $W = Span\{\mathbf{x}_1, \mathbf{x}_2\}$, where $\mathbf{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$. Find an orthogonal basis for $W$.

**Main idea: Orthogonal projection**

$$\left\{ \begin{matrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{matrix} \right\} \Rightarrow \left\{ \begin{matrix} \mathbf{x}_1 \\ \mathbf{x}_2 = \alpha \mathbf{x}_1 + \mathbf{v}_2 \end{matrix} \right. \Rightarrow \left\{ \begin{matrix} \mathbf{v}_1 = \mathbf{x}_1 \\ \mathbf{v}_2 = \mathbf{x}_2 - \alpha \mathbf{x}_1 \end{matrix} \right.$$

where $\mathbf{x}_1 \bullet \mathbf{v}_2 = 0$. Then $W = Span\{\mathbf{x}_1, \mathbf{x}_2\} = Span\{\mathbf{v}_1, \mathbf{v}_2\}$.

**Solution**.

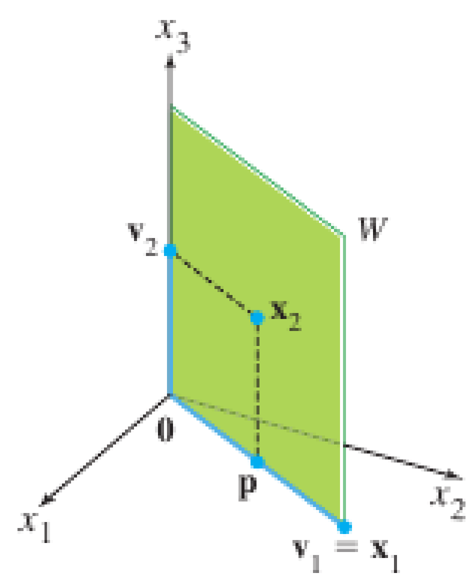



Figure 6.7: Construction of an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$.

**Example 6.46.** Find an *orthonormal basis* for a subspace whose basis is $\left\{ \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 8 \\ 5 \\ -6 \end{bmatrix} \right\}$.

**Solution.**

**Theorem 6.47. (The Gram-Schmidt Process)** *Given a basis* $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p\}$ *for a nonzero subspace* $W$ *of* $\mathbb{R}^n$*, define*

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{x}_1 \\ \mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x_2} \bullet \mathbf{v_1}}{\mathbf{v_1} \bullet \mathbf{v_1}} \mathbf{v}_1 \\ \mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x_3} \bullet \mathbf{v_1}}{\mathbf{v_1} \bullet \mathbf{v_1}} \mathbf{v}_1 - \frac{\mathbf{x_3} \bullet \mathbf{v_2}}{\mathbf{v_2} \bullet \mathbf{v_2}} \mathbf{v}_2 \\ &\vdots \\ \mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x_p} \bullet \mathbf{v_1}}{\mathbf{v_1} \bullet \mathbf{v_1}} \mathbf{v}_1 - \frac{\mathbf{x_p} \bullet \mathbf{v_2}}{\mathbf{v_2} \bullet \mathbf{v_2}} \mathbf{v}_2 - \cdots - \frac{\mathbf{x_p} \bullet \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \bullet \mathbf{v}_{p-1}} \mathbf{v}_{p-1} \end{aligned} \tag{6.28}$$

*Then* $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$ *is an* ***orthogonal basis*** *for* $W$. *In addition,*

$$\text{Span}\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k\} = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}, \quad \text{for } 1 \le k \le p. \tag{6.29}$$

**Remark 6.48.** For the result of the Gram-Schmidt process, define

$$\mathbf{u}_k = \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}, \quad \text{for } 1 \le k \le p. \tag{6.30}$$

Then $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p\}$ is an **orthonormal basis** for $W$. In practice, it is often implemented with the **normalized Gram-Schmidt process**.

**Example 6.49.** Find an *orthonormal basis* for $W = Span\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$, where

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} -2 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \text{ and } \mathbf{x}_3 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}.$$

**Solution**.

## 6.4.2. QR Factorization of Matrices

**Theorem 6.50. (The QR Factorization)** *If $A$ is an $m \times n$ matrix with linearly independent columns, then $A$ can be factored as*

$$A = QR, \tag{6.31}$$

*where*

- *$Q$ is an $m \times n$ matrix whose columns are orthonormal and*
- *$R$ is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.*

**Proof**. The columns of A form a basis $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$ for $W = Col\, A$.

1. Construct an **orthonormal basis** $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$ for $W$ (the **Gram-Schmidt process**). Set

$$Q \stackrel{\text{def}}{=\!=} [\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_n]. \tag{6.32}$$

2. **(Expression)** Since $Span\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k\} = Span\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k\}$, $1 \le k \le n$, there are constants $r_{1k}, r_{2k}, \cdots, r_{kk}$ such that

$$\mathbf{x}_k = r_{1k}\mathbf{u}_1 + r_{2k}\mathbf{u}_2 + \cdots + r_{kk}\mathbf{u}_k + 0\cdot\mathbf{u}_{k+1} + \cdots + 0\cdot\mathbf{u}_n. \tag{6.33}$$

   We may assume hat $r_{kk} > 0$. (If $r_{kk} < 0$, multiply both $r_{kk}$ and $\mathbf{u}_k$ by $-1$.)

3. Let $\mathbf{r}_k = [r_{1k}, r_{2k}, \cdots, r_{kk}, 0, \cdots, 0]^T$. Then

$$\mathbf{x}_k = Q\mathbf{r}_k \tag{6.34}$$

4. Define

$$R \stackrel{\text{def}}{=\!=} [\mathbf{r}_1\ \mathbf{r}_2\ \cdots\ \mathbf{r}_n]. \tag{6.35}$$

Then we see $A = [\mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_n] = [Q\mathbf{r}_1\ Q\mathbf{r}_2\ \cdots\ Q\mathbf{r}_n] = QR$. □

We can summarize the QR Factorization as follows.

**Algorithm 6.51. (QR Factorization)** Let $A = [\mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_n]$.

- Apply the Gram-Schmidt process to obtain an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$.
- Then

$$\begin{aligned} \mathbf{x}_1 &= (\mathbf{u}_1 \bullet \mathbf{x}_1)\mathbf{u}_1 \\ \mathbf{x}_2 &= (\mathbf{u}_1 \bullet \mathbf{x}_2)\mathbf{u}_1 + (\mathbf{u}_2 \bullet \mathbf{x}_2)\mathbf{u}_2 \\ \mathbf{x}_3 &= (\mathbf{u}_1 \bullet \mathbf{x}_3)\mathbf{u}_1 + (\mathbf{u}_2 \bullet \mathbf{x}_3)\mathbf{u}_2 + (\mathbf{u}_3 \bullet \mathbf{x}_3)\mathbf{u}_3 \\ &\ \vdots \\ \mathbf{x}_n &= \textstyle\sum_{j=1}^n (\mathbf{u}_j \bullet \mathbf{x}_n)\mathbf{u}_j. \end{aligned} \tag{6.36}$$

- Thus

$$A = [\mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_n] = QR \tag{6.37}$$

  implies that

$$\begin{aligned} Q &= [\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_n], \\ R &= \begin{bmatrix} \mathbf{u}_1 \bullet \mathbf{x}_1 & \mathbf{u}_1 \bullet \mathbf{x}_2 & \mathbf{u}_1 \bullet \mathbf{x}_3 & \cdots & \mathbf{u}_1 \bullet \mathbf{x}_n \\ 0 & \mathbf{u}_2 \bullet \mathbf{x}_2 & \mathbf{u}_2 \bullet \mathbf{x}_3 & \cdots & \mathbf{u}_2 \bullet \mathbf{x}_n \\ 0 & 0 & \mathbf{u}_3 \bullet \mathbf{x}_3 & \cdots & \mathbf{u}_3 \bullet \mathbf{x}_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathbf{u}_n \bullet \mathbf{x}_n \end{bmatrix} = Q^T A. \end{aligned} \tag{6.38}$$

- In practice, the coefficients $r_{ij} = \mathbf{u}_i \bullet \mathbf{x}_j$, $i < j$, can be saved during the (normalized) Gram-Schmidt process.

**Example 6.52.** Find the QR factorization for $A = \begin{bmatrix} 4 & -1 \\ 3 & 2 \end{bmatrix}$.

**Solution**.

*Ans*: $Q = \begin{bmatrix} 0.8 & -0.6 \\ 0.6 & 0.8 \end{bmatrix}$ $R = \begin{bmatrix} 5 & 0.4 \\ 0 & 2.2 \end{bmatrix}$

**True-or-False 6.53.**

a. If $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal basis for $W$, then multiplying $\mathbf{v}_3$ by a scalar $c$ gives a new orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, c\mathbf{v}_3\}$. ***Clue***: $c =$?

b. The Gram-Schmidt process produces from a linearly independent set $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p\}$ an orthogonal set $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$ with the property that for each $k$, the vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ span the same subspace as that spanned by $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k$ .

c. If $A = QR$, where $Q$ has orthonormal columns, then $R = Q^T A$.

d. If $\mathbf{x}$ is not in a subspace $W$, then $\widehat{\mathbf{x}} = \text{proj}_W \mathbf{x}$ is not zero.

e. In a QR factorization, say $A = QR$ (when $A$ has linearly independent columns), the columns of $Q$ form an orthonormal basis for the column space of $A$.

**Solution**.

*Ans*: F,T,T,F,T

## Exercises 6.4

1. The given set is a basis for a subspace $W$. Use the Gram-Schmidt process to produce an orthogonal basis for $W$.

(a) $\begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 8 \\ 5 \\ -6 \end{bmatrix}$

(b) $\begin{bmatrix} 1 \\ -4 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 7 \\ -7 \\ -4 \\ 1 \end{bmatrix}$

*Ans*: (b) $\mathbf{v}_2 = (5, 1, -4, -1)$

2. Find an orthogonal basis for the column space of the matrix

$$\begin{bmatrix} -1 & 6 & 6 \\ 3 & -8 & 3 \\ 1 & -2 & 6 \\ 1 & -4 & -3 \end{bmatrix}$$

*Ans*: $\mathbf{v}_3 = (1, 1, -3, 1)$

3. M Let $A = \begin{bmatrix} -10 & 13 & 7 & -11 \\ 2 & 1 & -5 & 3 \\ -6 & 3 & 13 & -3 \\ 16 & -16 & -2 & 5 \\ 2 & 1 & -5 & -7 \end{bmatrix}$

(a) Use the Gram-Schmidt process to produce an orthogonal basis for the column space of $A$.

(b) Use the method in this section to produce a QR factorization of $A$.

*Ans*: (a) $\mathbf{v}_4 = (0, 5, 0, 0, -5)$

# 6.5. Least-Squares Problems

**Note**: Let $A$ is an $m \times n$ matrix. Then $A\mathbf{x} = \mathbf{b}$ may have no solution, particularly when $m > n$. **In real-world,**

- $m \gg n$, where $m$ represents the number of data points and $n$ denotes the dimension of the points
- Need to find a best solution for $A\mathbf{x} \approx \mathbf{b}$

**Definition 6.54.** Let $A$ is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$.
A **least-squares (LS) solution** of $A\mathbf{x} = \mathbf{b}$ is an $\widehat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|\mathbf{b} - A\widehat{\mathbf{x}}\| \le \|\mathbf{b} - A\mathbf{x}\|, \quad \text{for all } \mathbf{x} \in \mathbb{R}^n. \tag{6.39}$$

**Note**: The **information matrix** $A$ and the **observation vector** $\mathbf{b}$ are often formulated from a certain dataset.

- Finding a **best approximation/representation** is a major subject in research level.
- Here we assume that the dataset is acquired appropriately.

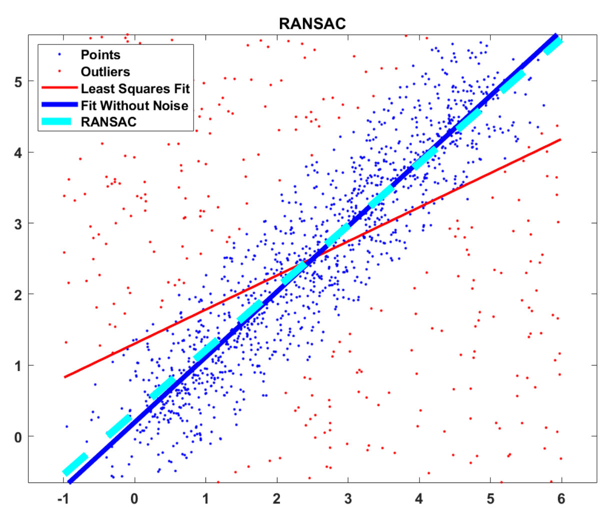



Figure 6.8: Least-Squares approximation for noisy data. The dashed line in cyan is the linear model from **random sample consensus** (**RANSAC**). The data has 1,200 and 300 points respectively for inliers and outliers.

## Solution of the General Least-Squares Problem

**Recall**: **(Definition 6.54)** Let $A$ is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$.
A **least-squares (LS) solution** of $A\mathbf{x} = \mathbf{b}$ is an $\widehat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|\mathbf{b} - A\widehat{\mathbf{x}}\| \le \|\mathbf{b} - A\mathbf{x}\|, \quad \text{for all } \mathbf{x} \in \mathbb{R}^n. \tag{6.40}$$

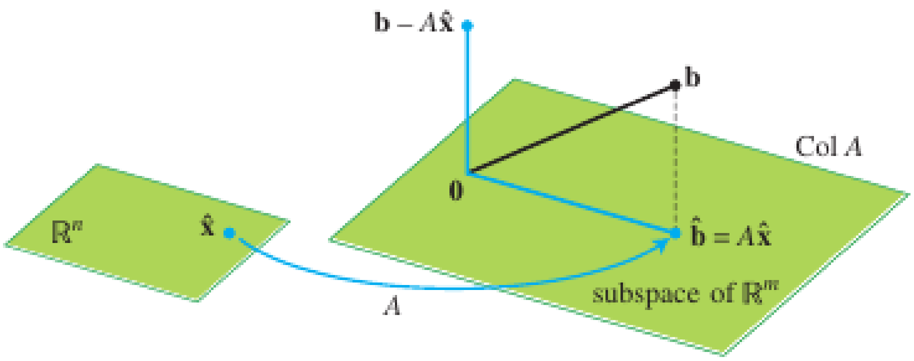



Figure 6.9: The LS solution $\widehat{\mathbf{x}}$ is in $\mathbb{R}^n$.

**Remark 6.55. Geometric Interpretation of the LS Problem**

- For all $\mathbf{x} \in \mathbb{R}^n$, **$A\mathbf{x}$ will necessarily be in *Col A*,** a subspace of $\mathbb{R}^m$.
  - So we seek an x that makes $A\mathbf{x}$ the closest point in *Col* $A$ to b.
- Let $\widehat{\mathbf{b}} = \text{proj}_{Col\,A}\mathbf{b}$. Then $A\mathbf{x} = \widehat{\mathbf{b}}$ has a solution and there is an $\widehat{\mathbf{x}} \in \mathbb{R}^n$ such that
$$A\widehat{\mathbf{x}} = \widehat{\mathbf{b}}. \tag{6.41}$$
- $\widehat{\mathbf{x}}$ is an LS solution of $A\mathbf{x} = \mathbf{b}$.
- The quantity $\|\mathbf{b}-\widehat{\mathbf{b}}\|^2 = \|\mathbf{b}-A\widehat{\mathbf{x}}\|^2$ is called the **least-squares error**.

**Note**: If $A \in \mathbb{R}^{n\times n}$ is invertible, then $A\mathbf{x} = \mathbf{b}$ has a unique solution $\widehat{\mathbf{x}}$ and therefore

$$\|\mathbf{b} - A\widehat{\mathbf{x}}\| = 0. \tag{6.42}$$

## The Method of Normal Equations

**Theorem 6.56.** *The set of LS solutions of* $A\mathbf{x} = \mathbf{b}$ *coincides with the nonempty set of solutions of the* ***normal equations***

$$A^T A\mathbf{x} = A^T\mathbf{b}. \tag{6.43}$$

**Proof**. Suppose $\widehat{\mathbf{x}}$ satisfies $A\widehat{\mathbf{x}} = \widehat{\mathbf{b}}$

$\Leftrightarrow \mathbf{b} - \widehat{\mathbf{b}} = \mathbf{b} - A\widehat{\mathbf{x}} \;\perp\; \mathit{Col}\, A$

$\Leftrightarrow \mathbf{a}_j \bullet (\mathbf{b} - A\widehat{\mathbf{x}}) = 0$ for all columns $\mathbf{a}_j$

$\Leftrightarrow {\mathbf{a}_j}^T(\mathbf{b} - A\widehat{\mathbf{x}}) = 0$ for all columns $\mathbf{a}_j$ (Note that ${\mathbf{a}_j}^T$ is a row of $A^T$)

$\Leftrightarrow A^T(\mathbf{b} - A\widehat{\mathbf{x}}) = 0$

$\Leftrightarrow A^T A\widehat{\mathbf{x}} = A^T\mathbf{b}$ □

**Example 6.57.** Let $A = \begin{bmatrix} 1 & 1 \\ 2 & 0 \\ -2 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} -4 \\ 8 \\ 1 \end{bmatrix}$.

(a) Find an LS solution of $A\mathbf{x} = \mathbf{b}$.

(b) Find the **least-squares error**, $\|\mathbf{b} - A\widehat{\mathbf{x}}\|^2$.

**Solution**.

*Ans*: (a) $\widehat{\mathbf{x}} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

**Remark 6.58.** Theorem 6.56 implies that LS solutions of $A\mathbf{x} = \mathbf{b}$ are solutions of the normal equations $A^TA\widehat{\mathbf{x}} = A^T\mathbf{b}$.

- When $A^TA$ is not invertible, the normal equations have either no solution or infinitely many solutions.
- So, data acquisition is important, to make it invertible.

**Theorem 6.59.** *Let A be an $m \times n$ matrix. The following statements are logically equivalent:*

a. *The equation $A\mathrm{x} = \mathrm{b}$ has a* **unique LS solution** *for each* $\mathrm{b} \in \mathbb{R}^m$.
b. *The columns of $A$ are linearly independent.*
c. *The matrix $A^TA$ is invertible.*

*When these statements are true, the unique LS solution $\widehat{\mathrm{x}}$ is given by*

$$\widehat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}. \tag{6.44}$$

**Definition 6.60.** The matrix

$$A^+ := (A^TA)^{-1}A^T \tag{6.45}$$

is called the **pseudoinverse** of $A$.

**Example 6.61.** Describe all least squares solutions of the equation $A\mathrm{x} = \mathrm{b}$, given

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 1 \\ 3 \\ 8 \\ 2 \end{bmatrix}.$$

**Solution**.

## Alternative Calculations of Least-Squares Solutions

**Theorem 6.62.** *Given an $m \times n$ matrix $A$ with linearly independent columns, let $A = QR$ be a QR factorization of $A$ as in Algorithm 6.51. Then, for each $\mathbf{b} \in \mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique LS solution, given by*

$$\widehat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}. \tag{6.46}$$

**Proof**. Let $A = QR$. Then the **pseudoinverse** of $A$ reads

$$\begin{aligned} (A^TA)^{-1}A^T &= ((QR)^TQR)^{-1}(QR)^T = (R^TQ^TQR)^{-1}R^TQ^T \\ &= R^{-1}(R^T)^{-1}R^TQ^T = R^{-1}Q^T, \end{aligned} \tag{6.47}$$

which completes the proof. □

**Self-study 6.63.** Find the LS solution of $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix}, \text{ where } A = QR = \begin{bmatrix} 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix}$$

**Solution**.

*Ans*: $Q^T\mathbf{b} = (6, -6, 4)$ and $\widehat{\mathbf{x}} = (10, -6, 2)$

**True-or-False 6.64.**

a. The general least-squares problem is to find an $\mathbf{x}$ that makes $A\mathbf{x}$ as close as possible to $\mathbf{b}$.

b. Any solution of $A^T A\mathbf{x} = A^T\mathbf{b}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

c. If $\widehat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$, then $\widehat{\mathbf{x}} = (A^T A)^{-1}A^T\mathbf{b}$.

d. The normal equations always provide a reliable method for computing least-squares solutions.

**Solution**.

*Ans*: T,T,F,F

## Exercises 6.5

1. Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ by (i) constructing the normal equations and (ii) solving for $\widehat{\mathbf{x}}$. Also (iii) compute the least-squares error ($\|\mathbf{b} - A\widehat{\mathbf{x}}\|$) associated with the least-squares solution.

   (a) $A = \begin{bmatrix} -1 & 2 \\ 2 & -3 \\ -1 & 3 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}$ (b) $A = \begin{bmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 3 \\ 2 & 5 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ 1 \\ -4 \\ 2 \end{bmatrix}$

   *Ans*: (b) $\widehat{\mathbf{x}} = \begin{bmatrix} 4/3 \\ -1/3 \end{bmatrix}$

2. Find (i) the orthogonal projection of $\mathbf{b}$ onto $Col\,A$ and (ii) a least-squares solution of $A\mathbf{x} = \mathbf{b}$. Also (iii) compute the least-squares error associated with the least-squares solution.

   (a) $A = \begin{bmatrix} 4 & 0 & 1 \\ 1 & -5 & 1 \\ 6 & 1 & 0 \\ 1 & -1 & -5 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 9 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ (b) $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \\ -1 & 1 & -1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ 5 \\ 6 \\ 6 \end{bmatrix}$

   *Ans*: (b) $\widehat{\mathbf{b}} = (5, 2, 3, 6)$ and $\widehat{\mathbf{x}} = (1/3, 14/3, -5/3)$

3. Describe all least-squares solutions of the system and the associated least-squares error.

   $$\begin{cases} x + y &= 1 \\ x + 2y &= 3 \\ x + 3y &= 3 \end{cases}$$

   *Ans*: $\widehat{\mathbf{x}} = (1/3, 1)$

For the above problems, you may use either pencil-and-paper or computer programs. For example, for the last problem, a code can be written as

exercise-6.5.3.m

```
A = [1 1; 1 2; 1 3];
b = [1;3;3];

ATA = A'*A; ATb = A'*b;

xhat = ATA\ATb
error = norm(b-A*xhat)^2
```

# 6.6. Machine Learning: Regression Analysis

**Recall**: **(Section 6.5)**

- **(Definition 6.54)** Let $A$ is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$.
  A **least-squares (LS) solution** of $A\mathbf{x} = \mathbf{b}$ is an $\widehat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|\mathbf{b} - A\widehat{\mathbf{x}}\| \le \|\mathbf{b} - A\mathbf{x}\|, \quad \text{for all } \mathbf{x} \in \mathbb{R}^n. \tag{6.48}$$

- **(Theorem 6.56)** The set of LS solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the **nonempty** set of solutions of the **normal equations**

$$A^T A\mathbf{x} = A^T \mathbf{b}. \tag{6.49}$$

- **(Theorem 6.59)** The normal equations have a **unique** solution, if and only if **the columns of $A$ are linearly independent.**
- **(Definition 6.60)** The matrix

$$A^+ := (A^T A)^{-1} A^T$$

  is called the **pseudoinverse** of $A$.
- **(Theorem 6.62)** Given an $m \times n$ matrix $A$ with **linearly independent columns**, let $A = QR$ be a QR factorization of $A$ as in Algorithm 6.51. Then, for each $\mathbf{b} \in \mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique LS solution, given by

$$\widehat{\mathbf{x}} = R^{-1} Q^T \mathbf{b}. \tag{6.50}$$

### 6.6.1. Regression Line

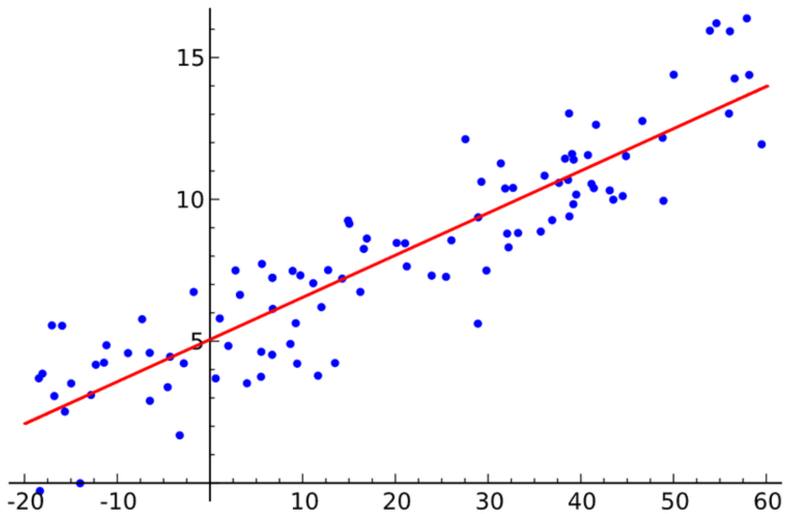



Figure 6.10: A regression line.

**Definition 6.65.** Suppose a set of experimental data points are given as

$$(x_1, y_1),\ (x_2, y_2),\ \cdots,\ (x_m, y_m)$$

such that the graph is close to a line. We determine a line

$$y = \beta_0 + \beta_1 x \tag{6.51}$$

that is as close as possible to the given points. This line is called the **least-squares line**; it is also called **regression line** of $y$ on $x$ and $\beta_0$, $\beta_1$ are called **regression coefficients**.

## Calculation of Least-Squares Lines

**Remark 6.66.** Consider a least-squares (LS) model of the form $y = \beta_0 + \beta_1 x$, for a given data set $\{(x_i, y_i) \mid i = 1, 2, \cdots, m\}$.

- Then

| Predicted $y$-value | | Observed $y$-value |
|---|---|---|
| $\beta_0 + \beta_1 x_1$ | $=$ | $y_1$ |
| $\beta_0 + \beta_1 x_2$ | $=$ | $y_2$ |
| $\vdots$ | | $\vdots$ |
| $\beta_0 + \beta_1 x_m$ | $=$ | $y_m$ |

(6.52)

- It can be equivalently written as

$$X\boldsymbol{\beta} = \mathbf{y}, \tag{6.53}$$

  where

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

  Here we call $X$ the **design matrix**, $\boldsymbol{\beta}$ the **parameter vector**, and y the **observation vector**.

- **(Method of Normal Equations)** Thus the LS solution can be determined as

$$X^T X\boldsymbol{\beta} = X^T\mathbf{y} \;\Rightarrow\; \boldsymbol{\beta} = (X^T X)^{-1} X^T \mathbf{y}, \tag{6.54}$$

  provided that $X^T X$ is invertible.

**Example 6.67.** Find the equation $y = \beta_0 + \beta_1 x$ of least-squares line that best fits the given points:

$(-1, 0)$, $(0, 1)$, $(1, 2)$, $(2, 4)$

**Solution.**

**Remark 6.68.** It follows from (6.53) that

$$\begin{aligned}
X^T X &= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_m \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} = \begin{bmatrix} m & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix}, \\
X^T \mathbf{y} &= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_m \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{bmatrix}.
\end{aligned} \tag{6.55}$$

Thus the normal equations for the regression line read

$$\begin{bmatrix} \Sigma 1 & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{bmatrix}. \tag{6.56}$$

**Example 6.69.** Find the equation $y = \beta_0 + \beta_1 x$ of least-squares line that best fits the given points:

$(0,1),\ (1,1),\ (2,2),\ (3,2)$

**Solution**.

## 6.6.2. Least-Squares Fitting of Other Curves

**Remark 6.70.** Consider a regression model of the form

$$y = \beta_0 + \beta_1 x + \beta_2 x^2,$$

for a given data set $\{(x_i, y_i) \mid i = 1, 2, \cdots, m\}$.

- As for the regression line, we will get a **linear system** and try to find **LS solutions of the system**.
- **Linear System**:

$$\begin{array}{ccc} \boxed{\text{Predicted } y\text{-value}} & & \boxed{\text{Observed } y\text{-value}} \\ \hline \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 & = & y_1 \\ \beta_0 + \beta_1 x_2 + \beta_2 x_2^2 & = & y_2 \\ \vdots & & \vdots \\ \beta_0 + \beta_1 x_m + \beta_2 x_m^2 & = & y_m \\ \hline \end{array} \tag{6.57}$$

- It is equivalently written as

$$X\boldsymbol{\beta} = \mathbf{y}, \tag{6.58}$$

where

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

- The system can be solved by the **method of normal equations**:

$$X^T X\boldsymbol{\beta} = \begin{bmatrix} \Sigma 1 & \Sigma x_i & \Sigma x_i^2 \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 \\ \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i^4 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \\ \Sigma x_i^2 y_i \end{bmatrix} = X^T \mathbf{y} \tag{6.59}$$

**Example 6.71.** Find an LS curve of the form $y = \beta_0 + \beta_1 x + \beta_2 x^2$ that best fits the given points:

$(0, 1)$, $(1, 1)$, $(1, 2)$, $(2, 3)$.

**Solution**. The normal equations are $\begin{bmatrix} \Sigma 1 & \Sigma x_i & \Sigma x_i^2 \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 \\ \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i^4 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \\ \Sigma x_i^2 y_i \end{bmatrix}$

*Ans*: $y = 1 + 0.5x^2$

**Self-study 6.72.** Find an LS curve of the form $y = \beta_0 + \beta_1 x + \beta_2 x^2$ that best fits the given points:

$(-2, 1),\ (-1, 0),\ (0, 1),\ (1, 4),\ (2, 9)$

**Solution**.

***Ans***: $y = 1 + 2x + x^2$

## Further Applications

**Example 6.73.** Find an LS curve of the form $y = a\cos x + b\sin x$ that best fits the given points:

$(0, 1),\ (\pi/4, 2),\ (\pi, 0)$.

**Solution**.

***Ans***: $(a, b) = (1/2, -1/2 + 2\sqrt{2}) = (0.5, 2.32843)$

## Nonlinear Models: Linearization

**Strategy 6.74.** For nonlinear models, **change of variables** can be applied for a linear model.

$$\begin{array}{lll} \text{Model} & \text{Change of Variables} & \text{Linearization} \\ \hline y = A + \dfrac{B}{x} & \widetilde{x} = \dfrac{1}{x},\ \ \widetilde{y} = y & \Rightarrow\ \ \widetilde{y} = A + B\widetilde{x} \\ y = \dfrac{1}{A + Bx} & \widetilde{x} = x,\ \ \widetilde{y} = \dfrac{1}{y} & \Rightarrow\ \ \widetilde{y} = A + B\widetilde{x} \\ y = Ce^{Dx} & \widetilde{x} = x,\ \ \widetilde{y} = \ln y & \Rightarrow\ \ \widetilde{y} = \ln C + D\widetilde{x} \\ \hline \end{array} \tag{6.60}$$

**The Idea**: Transform the nonlinear model to produce a **linear system**.

**Example 6.75.** Find an LS curve of the form $y = Ce^{Dx}$ that best fits the given points:

$(0, e),\ (1, e^3),\ (2, e^5)$.

**Solution**.

$$\begin{array}{c|c} x & y \\ \hline 0 & e \\ 1 & e^3 \\ 2 & e^5 \\ \hline \end{array} \quad \Rightarrow \quad \begin{array}{c|c} \widetilde{x} & \widetilde{y} = \ln y \\ \hline 0 & 1 \\ 1 & 3 \\ 2 & 5 \\ \hline \end{array} \quad \Rightarrow \quad X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix},\ \ \mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}$$

*Ans*: $y = ee^{2x} = e^{2x+1}$

## Exercises 6.6

1. Find an LS curve of the form $y = \beta_0 + \beta_1 x$ that best fits the given points.

   (a) $(1,0),\ (2,1),\ (4,2),\ (5,3)$ (b) $(2,3),\ (3,2),\ (5,1),\ (6,0)$

   *Ans*: (a) $y = -0.6 + 0.7x$

2. M A certain experiment produces the data

$$(1, 1.8),\ (2, 2.7),\ (3, 3.4),\ (4, 3.8),\ (5, 3.9).$$

   For these points, we will try to find the best-fitting model of the form $y = \beta_1 x + \beta_2 x^2$.

   (a) Find and display the design matrix and the observation vector.
   (b) Find the unknown parameter vector.
   (c) Find the LS error.
   (d) Plot the associated LS curve along with the data.

APPENDIX **A**

# Appendix

**Contents of Chapter A**


A.1. Understanding/Interpretation of Eigenvalues and Eigenvectors . . . . . . . . . . . . . 228
A.2. Eigenvalues and Eigenvectors of Stochastic Matrices . . . . . . . . . . . . . . . . . . . . 231

# A.1. Understanding / Interpretation of Eigenvalues and Eigenvectors

**Recall**: Let $A$ be an $n \times n$ matrix. An **eigenvalue** $\lambda$ of $A$ and its corresponding **eigenvector** v are defined as

$$A\mathbf{v} = \lambda\mathbf{v}, \quad \mathbf{v} \neq 0. \tag{A.1.1}$$

**Observation A.1. (Matrix Transformation)**
Let $A$ be an $n \times n$ matrix. Consider the **matrix multiplication**

$$A\mathbf{x} = \mathbf{y}. \tag{A.1.2}$$

- It **scales** the vector.
- It **rotates** the vector.

**Remark A.2.** Historically, eigenvalues and eigenvectors appeared in the study of quadratic forms and differential equations:

> *In the 18th century,* **Leonhard Euler** *studied the rotational motion of a rigid body, and discovered the importance of the principal axes.* **Joseph-Louis Lagrange** *realized that the principal axes are the eigenvectors of the inertia matrix [3].*

- **(Having Principal Axes)** There are ***favored* directions/vectors**, for square matrices.

  When the matrix acts on these favored (principal) vectors, the action results in **scaling the vectors, without rotation**.

  - These **favored vectors** are the eigenvectors;
  - the **scaling factor** is the eigenvalue.

- **(Forming a Basis)** In various real-world interesting applications, **the eigenvectors form a basis**, which makes **matrix methods and algorithms much more effective and useful**.

**Example A.3.** Let $A \in \mathbb{R}^{n\times n}$ and its eigenvectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ form a **basis** for $\mathbb{R}^n$, $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$, $i = 1, 2, \cdots, n$. Then for an arbitrary $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{x} = \xi_1\mathbf{v}_1 + \xi_2\mathbf{v}_2 + \cdots + \xi_n\mathbf{v}_n = \sum_{i=1}^{n} \xi_i \mathbf{v}_i. \tag{A.1.3}$$

and therefore

$$A\mathbf{x} = \sum_{i=1}^{n} \xi_i A\mathbf{v}_i = \sum_{i=1}^{n} \xi_i \lambda_i \mathbf{v}_i. \quad \Rightarrow \quad \boldsymbol{A^k\mathbf{x} = \sum_{i=1}^{n} \xi_i \lambda_i^k \mathbf{v}_i.} \tag{A.1.4}$$

The formulation is applicable for many tasks in scientific computing, e.g. convergence analysis for **various iterative procedures**. □

**Example A.4.** **(Geometric Interpretation)** Consider a $2 \times 2$ matrix

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}. \tag{A.1.5}$$

Then its eigenvalues and eigenvectors are

$$\lambda_1,\ \lambda_2 = 3,\ 1 \qquad \mathbf{v}_1,\ \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{A.1.6}$$

The action of $A$ on the **unit circle** $(S^1)$ results in the following figure.

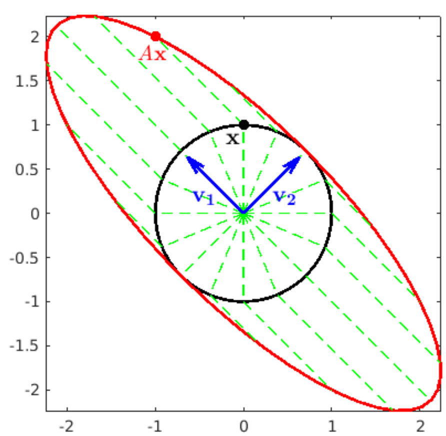



Figure A.1: Action of $A$ at $\mathbf{x} \in S^1$.

Find the area of the **solid ellipse**, **the image of the unit disk** by $A$.

*Ans*: $\pi \cdot |\boldsymbol{\lambda_1 \cdot \lambda_2}| = \pi \cdot 3 \cdot 1 = 3\pi$

## Singular Value Decomposition

**Theorem A.5. (SVD Theorem).** *Let $A \in \mathbb{R}^{m\times n}$ with $m \ge n$. Then*

$$A = U\Sigma V^T, \tag{A.1.7}$$

*where $U \in \mathbb{R}^{m\times n}$ and satisfies $U^TU = I$, $V \in \mathbb{R}^{n\times n}$ and satisfies $V^TV = I$, and $\Sigma = diag(\sigma_1, \sigma_2, \cdots, \sigma_n)$, where*

$$\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_n \ge 0. \tag{A.1.8}$$

**Remark A.6.** **The SVD**

- The **singular values** are the square root of **eigenvalues of $A^TA$**:

$$\sigma_i = \sqrt{\lambda_i}, \quad A^TA\mathbf{v}_i = \lambda_i\mathbf{v}_i, \tag{A.1.9}$$

  the **right singular vectors** $V$ is the collection of **eigenvectors of $A^TA$**:

$$V = [\mathbf{v}_1\ \ \mathbf{v}_2 \cdots\ \ \mathbf{v}_n], \tag{A.1.10}$$

  the **left singular vectors** $U$ are the collection of $\mathbf{u}_j$'s:

$$\mathbf{u}_j = A\mathbf{v}_j/\sigma_j, \quad \sigma_j \ne 0, \tag{A.1.11}$$

  and the **principal components** are

$$AV = U\Sigma. \tag{A.1.12}$$

- (**Dyadic Decomposition**) Given $A = U\Sigma V^T$, the matrix $A \in \mathbb{R}^{m\times n}$ can be expressed as

$$A = \sum_{j=1}^{n} \sigma_j\mathbf{u}_j\mathbf{v}_j^T. \tag{A.1.13}$$

- **(Data Compression)** The matrix $A$ can be approximated by $A_k$:

$$A \approx A_k \overset{\text{def}}{=\!=} \sum_{j=1}^{k} \sigma_j\mathbf{u}_j\mathbf{v}_j^T, \quad k < n, \tag{A.1.14}$$

  with error $||A - A_k||_2 = \sigma_{k+1}$.

- The SVD plays crucial roles in various applications, including **regression analysis** and **principal component analysis**.

## A.2. Eigenvalues and Eigenvectors of Stochastic Matrices

**Definition A.7. Probability Vector and Stochastic Matrix**

- A vector $\mathbf{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix}$ with nonnegative entries that add up to 1 is called a **probability vector**.
- A **(left) stochastic matrix** is a square matrix whose columns are probability vectors.

A **stochastic matrix** is also called a **probability matrix**, **transition matrix**, **substitution matrix**, or **Markov matrix**.

**Lemma A.8.** If $\mathbf{p}$ is a probability vector and $T$ is a stochastic matrix, then $T\mathbf{p}$ is a probability vector.

**Proof**. Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ be the columns of $T$. Then

$$\mathbf{q} := T\mathbf{p} = p_1\mathbf{v}_1 + p_2\mathbf{v}_2 + \cdots + p_n\mathbf{v}_n \in \mathbb{R}^n.$$

Clearly $\mathbf{q}$ has nonnegative entries; their sum reads

$$\text{sum}(\mathbf{q}) = \text{sum}(p_1\mathbf{v}_1 + p_2\mathbf{v}_2 + \cdots + p_n\mathbf{v}_n) = p_1 + p_2 + \cdots + p_n = 1.$$

**Definition A.9. Markov Chain**

In general, a finite **Markov chain** is a sequence of **probability vectors** $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \cdots$, together with a **stochastic matrix** $T$, such that

$$\mathbf{x}_1 = T\mathbf{x}_0, \;\; \mathbf{x}_2 = T\mathbf{x}_1, \;\; \mathbf{x}_3 = T\mathbf{x}_2, \;\; \cdots \qquad \text{(A.2.1)}$$

We can rewrite the above conditions as a recurrence relation

$$\mathbf{x}_{k+1} = T\mathbf{x}_k, \quad k = 0, 1, 2, \cdots \qquad \text{(A.2.2)}$$

The vector $\mathbf{x}_k$ is often called a **state vector**.

### The Maximum of Eigenvalues

**Definition A.10.** For a vector $\mathbf{x} = [x_1, x_2, \cdots, x_n]^T \in \mathbb{R}^n$, the **p-norm** is defined as

$$||\mathbf{x}||_p = \big(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\big)^{1/p}, \quad p > 0. \tag{A.2.3}$$

For example,

$$\begin{aligned} ||\mathbf{x}||_1 &= |x_1| + |x_2| + \cdots + |x_n| && \text{(1-norm)} \\ ||\mathbf{x}||_2 &= \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2} && \text{(2-norm)} \end{aligned} \tag{A.2.4}$$

**Theorem A.11.** *Let $T \in \mathbb{R}^{n\times n}$ be a stochastic matrix. Then*

$$||T\mathbf{x}||_1 \le ||\mathbf{x}||_1, \quad \forall\, \mathbf{x} \in \mathbb{R}^n. \tag{A.2.5}$$

**Proof**. Consider the case, $n = 2$: $T = \begin{bmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{bmatrix}$. Then, for $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$,

$$\begin{aligned} ||T\mathbf{x}||_1 &= |t_{11}x_1 + t_{12}x_2| + |t_{21}x_1 + t_{22}x_2| \\ &\le t_{11}|x_1| + t_{12}|x_2| \;+\; t_{21}|x_1| + t_{22}|x_2| \\ &= (t_{11} + t_{21})|x_1| \;+\; (t_{12} + t_{22})|x_2| \\ &= |x_1| + |x_2| = ||\mathbf{x}||_1. \end{aligned} \tag{A.2.6}$$

For general $n \ge 2$, use the same argument to complete the proof. □

**Corollary A.12.** *Let $T \in \mathbb{R}^{n\times n}$ be a stochastic matrix. Then every eigenvalue of $T$ is bounded by 1 in modulus. That is,*

$$T\mathbf{v} = \lambda\mathbf{v} \quad \Rightarrow \quad |\lambda| \le 1. \tag{A.2.7}$$

**Proof**. Let $T\mathrm{v} = \lambda\mathrm{v}$. Then it follows from Theorem A.11 that

$$||T\mathbf{v}||_1 = ||\lambda\mathbf{v}||_1 = |\lambda|\; ||\mathbf{v}||_1 \le ||\mathbf{v}||_1, \tag{A.2.8}$$

which completes the proof. □

## The Eigenvalue 1 and Its Corresponding Eigenvector

**Theorem A.13.** *Let $T \in \mathbb{R}^{n\times n}$ be a stochastic matrix. Then the number 1 is an eigenvalue of $T$.*

**Proof**. We prove the theorem in two different ways.

(a) Note that $det\,(T^T - \lambda I) = det\,(T - \lambda I)$, which implies that ***$T$* and *$T^T$* have exactly the same eigenvalues.** Consider the **all-ones vector** $\mathbf{1} = [1, 1, \cdots, 1]^T \in \mathbb{R}^n$. Then

$$T^T\mathbf{1} = \mathbf{1},$$

which implies that the number 1 is an eigenvalue of $T^T$ and therefore it is an eigenvalue of $T$.

(b) Construct $T - \lambda I$ for $\lambda = 1$:

$$T - I = \begin{bmatrix} t_{11}-1 & t_{12} & \cdots & T_{1n} \\ t_{21} & t_{22}-1 & \cdots & T_{2n} \\ \vdots & & \ddots & \vdots \\ t_{n1} & t_{12} & \cdots & T_{1n}-1 \end{bmatrix}. \tag{A.2.9}$$

We apply **replacement** operations so that all rows are added to the bottom row. Then **the resulting bottom row must become a zero row**, which implies

$$det\,(T - I) = 0, \tag{A.2.10}$$

and therefore the number 1 is an eigenvalue of $T$. □

**Definition A.14.** The **eigenvector** v corresponding to the eigenvalue 1 is called a **steady-state vector** of $T$. It is also called a **Perron-Frobenius eigenvector** or a **stable equilibrium distribution**.

The steady-state vector represents a **long term behavior** of a Markov chain.

## The Steady-State Vector

**Theorem A.15.** *If $T$ is an $n \times n$* ***regular stochastic matrix****, then $T$ has a unique steady-state vector* $\mathbf{q}$.

*(a) The entries of* $\mathbf{q}$ *are strictly positive.*

*(b) The steady-state vector can be computed by the* ***power method***

$$\mathbf{q} = \lim_{k\to\infty} T^k \mathbf{x}_0, \tag{A.2.11}$$

*where* $\mathbf{x}_0$ *is* ***a probability vector***.

**Example A.16.** Find eigenvalues and corresponding eigenvectors of the transition matrix.

$$T = \begin{bmatrix} 1/2 & 1/4 & 1/6 \\ 1/3 & 1/2 & 1/3 \\ 1/6 & 1/4 & 1/2 \end{bmatrix}. \tag{A.2.12}$$

**Solution.**

stochastic_eigen.py

```python
import numpy as np
np.set_printoptions(precision=4,suppress=True)

T = [[1/2,1/4,1/6],
     [1/3,1/2,1/3],
     [1/6,1/4,1/2]]
T = np.array(T)

D,V = np.linalg.eig(T)

print('Eigenvalues:');  print(D)
print('Eigenvectors:'); print(V)

print('----- steady-state vector ----')
v1 = V[:,0]; v1 /= sum(v1)
print('v1 = ',v1)

print('----- power method -----------')
x = np.array([1,0,0]);
print('k = %2d; '%(0),x)
for k in range(10):
    x = T.dot(x);
    print('k = %2d; '%(k+1),x)
```

Output

```
Eigenvalues:
[1.     0.3333 0.1667]
Eigenvectors:
[[-0.5145 -0.7071  0.4082]
 [-0.686  -0.     -0.8165]
 [-0.5145  0.7071  0.4082]]
----- steady-state vector ----
v1 =  [0.3 0.4 0.3]
----- power method -----------
k =  0;  [1 0 0]
k =  1;  [0.5    0.3333 0.1667]
k =  2;  [0.3611 0.3889 0.25  ]
k =  3;  [0.3194 0.3981 0.2824]
k =  4;  [0.3063 0.3997 0.294 ]
k =  5;  [0.3021 0.3999 0.298 ]
k =  6;  [0.3007 0.4    0.2993]
k =  7;  [0.3002 0.4    0.2998]
k =  8;  [0.3001 0.4    0.2999]
k =  9;  [0.3 0.4 0.3]
k = 10;  [0.3 0.4 0.3]
```

**Remark A.17.** *Some eigenvalues of a stochastic matrix can be negative.* For example, apply a **row interchange operation** to $T$ in (A.2.12):

```
T[[1,2]] = T[[2,1]]
```

Then

- The resulting matrix is still a stochastic matrix.
- Its eigenvalues become

```
[ 1.  0.281 -0.1977].
```

APPENDIX **C**

# Chapter Review

Sections selected for the review:

§1.7. Linear Independence, p. 44

§1.9. The Matrix of A Linear Transformation, p. 57

§2.3. Characterizations of Invertible Matrices, p. 82

§2.8. Subspaces of $\mathbb{R}^n$, p. 97

§2.9. Dimension and Rank, p. 103

§3.2. Properties of Determinants, p. 115

§4.1. Vector Spaces and Subspaces, p. 120

§5.3. Diagonalization, p. 142

§5.9. Applications to Markov Chains, p. 168

§6.3. Orthogonal Projections, p. 196

§6.4. The Gram-Schmidt Process and QR Factorization, p. 203

§6.6. Machine Learning: Regression Analysis, p. 217

**Contents of Chapter Reviews**

C.1. Linear Equations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 238

C.2. Matrix Algebra . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 241

C.3. Determinants . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 247

C.4. Vector Spaces . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 250

C.5. Eigenvalues and Eigenvectors . . . . . . . . . . . . . . . . . . . . . . . . 252

C.6. Orthogonality and Least-Squares . . . . . . . . . . . . . . . . . . . . . . 257

# C.1. Linear Equations

## §1.7. Linear Independence

**Definition 1.50**. A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ is said to be **linearly independent**, if the vector equation

$$x_1\,\mathbf{v}_1 + x_2\,\mathbf{v}_2 + \cdots + x_p\,\mathbf{v}_p = \mathbf{0} \tag{C.1.1}$$

has only the trivial solution (i.e., $x_1 = x_2 = \cdots = x_p = 0$). The set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$ is said to be **linearly dependent**, if there exist weights $c_1, c_2, \cdots, c_p$, not all zero, such that

$$c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_p\,\mathbf{v}_p = \mathbf{0}. \tag{C.1.2}$$

**Remark 1.52**. Let $A = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p]$. The matrix equation $A\mathbf{x} = \mathbf{0}$ is equivalent to $x_1\,\mathbf{v}_1 + x_2\,\mathbf{v}_2 + \cdots + x_p\,\mathbf{v}_p = \mathbf{0}$.

1. Columns of $A$ are **linearly independent** if and only if $A\mathbf{x} = \mathbf{0}$ has *only* the trivial solution. ( $\Leftrightarrow$ $A\mathbf{x} = \mathbf{0}$ has no free variable $\Leftrightarrow$ Every column in $A$ is a pivot column.)

2. Columns of $A$ are **linearly dependent** if and only if $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution. ( $\Leftrightarrow$ $A\mathbf{x} = \mathbf{0}$ has at least one free variable $\Leftrightarrow$ $A$ has at least one *non*-pivot column.)

**Example C.1.** Determine if the vectors are linearly independent.

$$\begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ -8 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix}$$

**Solution**.

## §1.9. The Matrix of A Linear Transformation

**Theorem 1.75**. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there exists a unique matrix $A \in \mathbb{R}^{m\times n}$ such that

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

In fact, with $\mathbf{e}_j$ denoting the $j$-th standard unit vector in $\mathbb{R}^n$,

$$A = [T(\mathbf{e}_1)\ \ T(\mathbf{e}_2)\ \ \cdots\ \ T(\mathbf{e}_n)]\,. \tag{C.1.3}$$

The matrix $A$ is called the **standard matrix** of the transformation.

**Note**: **Standard unit vectors in $\mathbb{R}^n$ & the standard matrix**:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix},\ \ \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},\ \ \cdots,\ \ \mathbf{e}_n = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \tag{C.1.4}$$

Any $\mathbf{x} \in \mathbb{R}^n$ can be written as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_n \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{bmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n.$$

Thus

$$\begin{aligned} T(\mathbf{x}) &= T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n) \\ &= x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + \cdots + x_nT(\mathbf{e}_n) \\ &= [T(\mathbf{e}_1)\ \ T(\mathbf{e}_2)\ \ \cdots\ \ T(\mathbf{e}_n)]\,\mathbf{x}, \end{aligned} \tag{C.1.5}$$

and therefore the standard matrix reads

$$A = [T(\mathbf{e}_1)\ \ T(\mathbf{e}_2)\ \ \cdots\ \ T(\mathbf{e}_n)]\,. \tag{C.1.6}$$

**Example C.2.** Write the standard matrix for the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^4$ given by

$$T(x_1, x_2) = (x_1 + 4x_2, 5x_1, -3x_2, x_1 - x_2).$$

**Solution**.

**Theorem 1.81**. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with the standard matrix $A$. Then,

(a) $T$ maps $\mathbb{R}^n$ **onto** $\mathbb{R}^m$ if and only if the columns of $A$ span $\mathbb{R}^m$.
( $\Leftrightarrow$ every row of $A$ has a pivot position
$\Leftrightarrow A\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b} \in \mathbb{R}^m$)

(b) $T$ is **one-to-one** if and only if the columns of $A$ are linearly independent.
( $\Leftrightarrow$ every column of $A$ is a pivot column
$\Leftrightarrow A\mathbf{x} = \mathbf{0}$ has "*only*" the trivial solution)

**Example C.3.** Let $T : \mathbb{R}^4 \to \mathbb{R}^3$ be the linear transformation whose standard matrix is

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

Is $T$ onto? Is $T$ one-to-one?
**Solution**.

*Ans*: onto, but not one-to-one

# C.2. Matrix Algebra

## §2.3. Characterizations of Invertible Matrices

**Theorem 2.25. (Invertible Matrix Theorem)**
Let $A$ be an $n \times n$ matrix. Then the following are equivalent.

a. ***A* is an invertible matrix.** (Def: There is $B$ s.t. $AB = BA = I$)
b. $A$ is row equivalent to the $n \times n$ identity matrix.
c. $A$ has $n$ pivot positions.
d. The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = 0$.
e. The columns of $A$ are linearly independent.
f. The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.
g. The equation $A\mathbf{x} = \mathbf{b}$ has unique solution for each $\mathbf{b} \in \mathbb{R}^n$.
h. The columns of $A$ span $\mathbb{R}^n$.
i. The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.
j. There is a matrix $C \in \mathbb{R}^{n\times n}$ such that $CA = I$
k. There is a matrix $D \in \mathbb{R}^{n\times n}$ such that $AD = I$
l. $A^T$ is invertible and $(A^T)^{-1} = (A^{-1})^T$.

**Theorem 2.74 (Invertible Matrix Theorem); §2.9**

m. The columns of $A$ form a basis of $\mathbb{R}^n$
n. $\mathit{Col}\, A = \mathbb{R}^n$
o. $\mathit{dim}\, \mathit{Col}\, A = n$
p. $\mathit{rank}\, A = n$
q. $\mathit{Nul}\, A = \{\mathbf{0}\}$
r. $\mathit{dim}\, \mathit{Nul}\, A = 0$

**Theorem 5.17 (Invertible Matrix Theorem); §5.2**

s. The number $0$ is not an eigenvalue of $A$.
t. $\mathit{det}\, A \neq 0$

**Example C.4.** An $n \times n$ **upper triangular matrix** is one whose entries below the main diagonal are zeros. When is a square upper triangular matrix invertible?

**Theorem 2.29 (Invertible linear transformations)**

1. A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is said to be **invertible** if there exists $S : \mathbb{R}^n \to \mathbb{R}^n$ such that $S \circ T(\mathbf{x}) = T \circ S(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$. In this case, $S = T^{-1}$.
2. Also, if $A$ is the **standard matrix** for $T$, then $A^{-1}$ is the standard matrix for $T^{-1}$.

**Example C.5.** Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation such that $T\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -5x_1 + 9x_2 \\ 4x_1 - 7x_2 \end{bmatrix}$. Find a formula for $T^{-1}$.

**Solution**.

## §2.8. Subspaces of $\mathbb{R}^n$

**Definition 2.47**. A **subspace** of $\mathbb{R}^n$ is any set $H$ in $\mathbb{R}^n$ that has three properties:

a) The zero vector is in $H$.

b) For each $\mathbf{u}$ and $\mathbf{v}$ in H, the sum $\mathbf{u}+\mathbf{v}$ is in $H$.

c) For each $\mathbf{u}$ in $H$ and each scalar $c$, the vector $c\mathbf{u}$ is in $H$.

That is, $H$ is **closed** under linear combinations.

**Definition 2.49**. Let $A$ be an $m\times n$ matrix. The **column space** of $A$ is the set *Col* $A$ of all linear combinations of columns of $A$. That is, if $A=[\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_n]$, then

$$\mathit{Col}\,A=\{\mathbf{u}\mid \mathbf{u}=c_1\,\mathbf{a}_1+c_2\,\mathbf{a}_2+\cdots+c_n\,\mathbf{a}_n\}, \tag{C.2.1}$$

where $c_1,\ c_2,\ \cdots,\ c_n$ are scalars. ***Col A* is a subspace of $\mathbb{R}^m$.**

**Definition 2.51**. Let $A$ be an $m\times n$ matrix. The **null space** of $A$, *Nul* $A$, is the set of all solutions of the homogeneous system $A\mathbf{x}=\mathbf{0}$.

**Theorem 2.52**. *Nul* $A$ is a subspace of $\mathbb{R}^n$.

**Example C.6.** Let $A=\begin{bmatrix}1&-2&-3\\2&4&2\\-3&5&6\end{bmatrix}$ and $\mathbf{b}=\begin{bmatrix}2\\-4\\-7\end{bmatrix}$. Determine whether b is in the column space of $A$, *Col* $A$.

**Solution**. ***Clue***: ① $\mathbf{b}\in \mathit{Col}\,A$
⇔ ② b is a linear combination of columns of $A$
⇔ ③ $A\mathbf{x}=\mathbf{b}$ is consistent
⇔ ④ $[A\ \ \mathbf{b}]$ has a solution

**Definition 2.53.** A **basis** for a subspace $H$ in $\mathbb{R}^n$ is a set of vectors that

1. is *linearly independent*, and
2. *spans* $H$.

**Theorem 2.56.** Basis for *Nul* $A$ can be obtained from the parametric vector form of solutions of $A\mathrm{x} = \mathbf{0}$. That is, suppose that the solutions of $A\mathrm{x} = \mathbf{0}$ reads

$$\mathbf{x} = x_1\,\mathbf{u}_1 + x_2\,\mathbf{u}_2 + \cdots + x_k\,\mathbf{u}_k,$$

where $x_1,\ x_2,\ \cdots,\ x_k$ correspond to free variables. Then, a basis for *Nul* $A$ is $\{\mathbf{u}_1,\ \mathbf{u}_2,\ \cdots,\ \mathbf{u}_k\}$.

**Theorem 2.58.** In general, non-pivot columns are linear combinations of pivot columns. Thus the **pivot columns of a matrix *A*** form a basis for *Col* $A$.

**Example C.7.** Matrix $A$ and its echelon form is given. Find a basis for *Col* $A$ and a basis for *Nul* $A$.

$$A = \begin{bmatrix} 3 & -6 & 9 & 0 \\ 2 & -4 & 7 & 2 \\ 3 & -6 & 6 & -6 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 3 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

**Solution.**

*Ans*: $\mathcal{B}_{\mathit{Col}\,A} = \{\mathbf{a}_1, \mathbf{a}_3\},\ \mathcal{B}_{\mathit{Nul}\,A} = \{[2,1,0,0]^T, [6,0,-2,1]^T\}$.

## §2.9. Dimension and Rank

**Definition 2.64.** Suppose the set $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_p\}$ is a basis for a subspace $H$. For each $\mathbf{x} \in H$, the **coordinates of** $\mathbf{x}$ **relative to the basis** $\mathcal{B}$ are the weights $c_1, c_2, \cdots, c_p$ such that $\mathbf{x} = c_1\,\mathbf{b}_1 + c_2\,\mathbf{b}_2 + \cdots + c_p\,\mathbf{b}_p$, and the vector in $\mathbb{R}^p$

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

is called the **coordinate vector of** $\mathbf{x}$ **(relative to** $\mathcal{B}$**)** or the $\mathcal{B}$-**coordinate vector** of $\mathbf{x}$.

**Self-study C.8.** Let $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \\ -2 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 2 \\ 6 \\ -2 \end{bmatrix}$, and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$. Then $\mathcal{B}$ is a basis for $H = \mathit{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$, because $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent. Determine if $\mathbf{x}$ is in $H$, and if it is, find the coordinate vector of $\mathbf{x}$ relative to $\mathcal{B}$.

**Solution**.

**Theorem 2.70.** **(Rank Theorem)** Let $A \in \mathbb{R}^{m\times n}$. Then

$$\begin{aligned} \textit{dim}\,\textit{Col}\,A + \textit{dim}\,\textit{Nul}\,A \;=\; \textit{rank}\,A + \textit{nullity}\,A \;=\; n \\ = \text{(the number of columns in } A) \end{aligned}$$

Here, "*dim Nul* $A$" is called the **nullity** of $A$: *nullity* $A$

**Theorem 2.73.** **(The Basis Theorem)**
Let $H$ be a $p$-dimensional subspace of $\mathbb{R}^n$. Then

a) Any *linearly independent* set of exactly $p$ elements in $H$ is automatically a basis for $H$

b) Any set of $p$ elements of $H$ that *spans* $H$ is automatically a basis for $H$.

**Example C.9.** Find a basis for the subspace spanned by the given vectors. What is the dimension of the subspace?

$$\begin{bmatrix} 1 \\ -1 \\ -2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ -1 \\ 4 \end{bmatrix}, \begin{bmatrix} -3 \\ 5 \\ 0 \\ -5 \end{bmatrix}, \begin{bmatrix} -4 \\ 6 \\ 2 \\ -8 \end{bmatrix}$$

**Solution**.

# C.3. Determinants

## §3.2. Properties of Determinants

**Definition 3.1**. Let $A$ be an $n \times n$ square matrix. Then **determinant** is a scalar value denoted by ***det*** $A$ or $|A|$.

1) Let $A = [a] \in \mathbb{R}^{1 \times 1}$. Then ***det*** $A = a$.

2) Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbb{R}^{2 \times 2}$. Then ***det*** $A = ad - bc$.

**Definition 3.3**. Let $A_{ij}$ be the *submatrix* of $A$ obtained by deleting row $i$ and column $j$ of $A$. Then the $(i, j)$**-cofactor** of $A = [a_{ij}]$ is the scalar $C_{ij}$, given by

$$C_{ij} = (-1)^{i+j} \, \textbf{\textit{det}} \, A_{ij}. \tag{C.3.1}$$

**Definition 3.4**. For $n \geq 2$, the **determinant** of an $n \times n$ matrix $A = [a_{ij}]$ is given by the following formulas:

1. The *cofactor expansion* across the first row:

$$\textbf{\textit{det}} \, A \; = \; a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n} \tag{C.3.2}$$

2. The *cofactor expansion* across the row $i$:

$$\textbf{\textit{det}} \, A \; = \; a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} \tag{C.3.3}$$

3. The *cofactor expansion* down the column $j$:

$$\textbf{\textit{det}} \, A \; = \; a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj} \tag{C.3.4}$$

**Note**: The determinant can be viewed as a **volume scaling factor**.

**Theorem 3.9**. Let $A$ be an $n \times n$ square matrix.

a) **(Replacement)**: If $B$ is obtained from $A$ by a row replacement, then
$\boldsymbol{det}\, B = \boldsymbol{det}\, A$.

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

b) **(Interchange)**: If two rows of $A$ are interchanged to form $B$, then
$\boldsymbol{det}\, B = -\boldsymbol{det}\, A$.

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

c) **(Scaling)**: If one row of $A$ is multiplied by k ($\neq 0$), then
$\boldsymbol{det}\, B = k \cdot \boldsymbol{det}\, A$.

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

**Example C.10.** Compute $\boldsymbol{det}\, A$, where $A = \begin{bmatrix} 1 & -4 & 2 \\ -1 & 7 & 0 \\ -2 & 8 & -9 \end{bmatrix}$, after applying a couple of steps of replacement operations.

**Claim 3.12**. Let $A$ and $B$ be $n \times n$ matrices.

a) $\boldsymbol{det}\, A^T = \boldsymbol{det}\, A$.

b) $\boldsymbol{det}\,(AB) = \boldsymbol{det}\, A \cdot \boldsymbol{det}\, B$.

c) If $A$ is invertible, then $\boldsymbol{det}\, A^{-1} = \dfrac{1}{\boldsymbol{det}\, A}$. ($\because$ $\boldsymbol{det}\, I_n = 1$.)

**Example C.11.** Find the **determinant** of $A^2$, when

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

**Solution**.

***Ans***: $\boldsymbol{det}\, A = 12 \;\Rightarrow \boldsymbol{det}\,(A^2) = (\boldsymbol{det}\, A)^2 = (12)^2 = 144$.

# C.4. Vector Spaces

## §4.1. Vector Spaces and Subspaces

**Definition 4.1**. A **vector space** is a nonempty set $V$ of objects, called **vectors**, on which are defined **two operations**, called **addition** and **multiplication by scalars** (real numbers), subject to the ten axioms (or rules) listed below. The axioms must hold for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and for all scalars $c$ and $d$.

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

**Definition 4.3**. A **subspace** of a vector space $V$ is a subset $H$ of $V$ that has three properties:

a) $\mathbf{0} \in H$, where $\mathbf{0}$ is the zero vector of $V$

b) $H$ is closed under vector addition: for each $\mathbf{u}, \mathbf{v} \in H$, $\mathbf{u} + \mathbf{v} \in H$

c) $H$ is closed under scalar multiplication: for each $\mathbf{u} \in H$ and each scalar $c$, $c\mathbf{u} \in H$

**Theorem 4.7**. If $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p$ are in a vector space $V$, then *Span*$\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$ is a subspace of $V$.

**Example C.12.** Let $H = \{(a-b, 3b-a, a+b, b) \mid a, b \in \mathbb{R}\}$. Show that $H$ is a subspace of $\mathbb{R}^4$.

**Solution**.

**Example C.13.** Determine if the given set is a subspace of $\mathbb{P}_n$ for an appropriate value of $n$.

a) $\left\{at^2 \mid a \in \mathbb{R}\right\}$

b) $\{\mathbf{p} \in \mathbb{P}_3$ with integer coefficients$\}$

c) $\left\{a + t^2 \mid a \in \mathbb{R}\right\}$

d) $\{\mathbf{p} \in \mathbb{P}_n \mid \mathbf{p}(0) = 0\}$

**Solution**.

*Ans*: a) Yes, b) No, c) No, d) Yes

**Self-study C.14.** Let $H$ and $K$ be subspaces of $V$. Define the **sum** of $H$ and $K$ as

$$H + K = \{\mathbf{u} + \mathbf{v} \mid \mathbf{u} \in H, \mathbf{v} \in K\}.$$

Prove that $H + K$ is a subspace of $V$.

**Solution**.

# C.5. Eigenvalues and Eigenvectors

## §5.3. Diagonalization

**Definition 5.25**. An $n \times n$ matrix $A$ is said to be **diagonalizable** if there exists an invertible matrix $P$ and a diagonal matrix $D$ such that

$$A = PDP^{-1} \quad (\text{or } \; P^{-1}AP = D) \tag{C.5.1}$$

**Theorem 5.28**. **(The Diagonalization Theorem)**

1. An $n \times n$ matrix $A$ is diagonalizable if and only if $A$ has $n$ linearly independent eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$.
2. In fact, $A = PDP^{-1}$ if and only if columns of P are $n$ linearly independent eigenvectors of $A$. In this case, the diagonal entries of $D$ are the corresponding eigenvalues of $A$. That is,

$$\begin{aligned} P &= [\mathbf{v}_1 \; \mathbf{v}_2 \; \cdots \; \mathbf{v}_n], \\ D &= \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}, \end{aligned} \tag{C.5.2}$$

where $A\mathbf{v}_k = \lambda_k \mathbf{v}_k$, $k = 1, 2, \cdots, n$.

The Diagonalization Theorem can be proved using the following remark.

**Remark 5.29.** ***$AP = PD$* with *$D$* Diagonal**

Let $P = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n]$ and $D = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ be **arbitrary $n \times n$ matrices**. Then,

$$AP = A[\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n] = [A\mathbf{v}_1\ A\mathbf{v}_2\ \cdots\ A\mathbf{v}_n], \tag{C.5.3}$$

while

$$PD = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = [\lambda_1\mathbf{v}_1\ \lambda_2\mathbf{v}_2\ \cdots\ \lambda_n\mathbf{v}_n]. \tag{C.5.4}$$

**If $AP = PD$ with $D$ diagonal, then the nonzero columns of $P$ are eigenvectors of $A$.**

**Self-study C.15.** Diagonalize the following matrix, if possible.

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

for which *det* $(B - \lambda I) = -(\lambda - 1)(\lambda + 2)^2$.

**Solution**.

## §5.9. Applications to Markov Chains

**Definition 5.60**. **Probability Vector and Stochastic Matrix**

- A vector $\mathbf{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix}$ with nonnegative entries that add up to 1 is called a **probability vector**.
- A **(left) stochastic matrix** is a square matrix whose columns are probability vectors.

A **stochastic matrix** is also called a **probability matrix**, **transition matrix**, **substitution matrix**, or **Markov matrix**.

**Lemma 5.61**. Let $T$ be a stochastic matrix. If $\mathbf{p}$ is a probability vector, then so is $\mathbf{q} = T\mathbf{p}$.

**Proof**. Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ be the columns of $T$. Then

$$\mathbf{q} = T\mathbf{p} = p_1\mathbf{v}_1 + p_2\mathbf{v}_2 + \cdots p_n\mathbf{v}_n.$$

Clearly $\mathbf{q}$ has nonnegative entries; their sum reads

$$\text{sum}(\mathbf{q}) = \text{sum}(p_1\mathbf{v}_1 + p_2\mathbf{v}_2 + \cdots p_n\mathbf{v}_n) = p_1 + p_2 + \cdots + p_n = 1.$$

**Definition 5.62**. **Markov Chain**

In general, a finite **Markov chain** is a sequence of **probability vectors** $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \cdots$, together with a **stochastic matrix** $T$, such that

$$\mathbf{x}_1 = T\mathbf{x}_0, \;\; \mathbf{x}_2 = T\mathbf{x}_1, \;\; \mathbf{x}_3 = T\mathbf{x}_2, \;\; \cdots \tag{C.5.5}$$

We can rewrite the above conditions as a recurrence relation

$$\mathbf{x}_{k+1} = T\mathbf{x}_k, \quad k = 0, 1, 2, \cdots \tag{C.5.6}$$

The vector $\mathbf{x}_k$ is often called a **state vector**.

## Steady-State Vectors

**Definition 5.66**. If $T$ is a stochastic matrix, then a **steady-state vector** for $T$ is a probability vector q such that

$$T\mathbf{q} = \mathbf{q}. \tag{C.5.7}$$

**Note**: The steady-state vector **q** can be seen as an eigenvector of $T$, of which the corresponding eigenvalue $\lambda = 1$.

**Strategy 5.67**. **How to Find a Steady-State Vector**

(a) First, solve for $\mathbf{x} = [x_1, x_2, \cdots, x_n]^T$:

$$T\mathbf{x} = \mathbf{x} \;\Leftrightarrow\; T\mathbf{x} - \mathbf{x} = \mathbf{0} \;\Leftrightarrow\; (T - I)\mathbf{x} = \mathbf{0}. \tag{C.5.8}$$

(b) Then, set

$$\mathbf{q} = \frac{1}{x_1 + x_2 + \cdots + x_n}\mathbf{x}. \tag{C.5.9}$$

**Example C.16.** Let $T = \begin{bmatrix} 0.4 & 0.3 \\ 0.6 & 0.7 \end{bmatrix}$. Find a steady-state vector for $T$.

**Definition 5.69**. A stochastic matrix $T$ is **regular** if some matrix power $T^k$ contains only strictly positive entries.

**Theorem 5.72**. If $T$ is an $n \times n$ regular stochastic matrix, then $T$ has a unique steady-state vector $\mathbf{q}$.

(a) The entries of $\mathbf{q}$ are strictly positive.

(b) The steady-state vector

$$\mathbf{q} = \lim_{k\to\infty} T^k \mathbf{x}_0, \tag{C.5.10}$$

for **any initial probability vector** $\mathbf{x}_0$.

**Remark 5.73**. Let $T \in \mathbb{R}^{n\times n}$ be a regular stochastic matrix. Then

- If $T\mathbf{v} = \lambda\mathbf{v}$, then $|\lambda| \le 1$.
  (The above is true for every stochastic matrix; see § A.2.)
- Every column of $T^k$ converges to $\mathbf{q}$ as $k \to \infty$, i.e.,

$$T^k \to [\mathbf{q}\ \ \mathbf{q}\ \ \cdots\ \ \mathbf{q}] \in \mathbb{R}^{n\times n}, \quad \text{as } k \to \infty. \tag{C.5.11}$$

**Example C.17.** Let $T = \begin{bmatrix} 0 & 0.5 \\ 1 & 0.5 \end{bmatrix}$.

(a) Is $T$ regular?

(b) What is the first column of $\lim_{k\to\infty} T^k$?

# C.6. Orthogonality and Least-Squares

## §6.3. Orthogonal Projections

**Theorem 6.36. (The Orthogonal Decomposition Theorem)**
Let $W$ be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y} \in \mathbb{R}^n$ can be written uniquely in the form

$$\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{z}, \tag{C.6.1}$$

where $\widehat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^\perp$. In fact, if $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p\}$ is an **orthogonal basis** for $W$, then

$$\begin{aligned} \widehat{\mathbf{y}} &= \text{proj}_W \mathbf{y} = \frac{\mathbf{y} \bullet \mathbf{u}_1}{\mathbf{u}_1 \bullet \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y} \bullet \mathbf{u}_2}{\mathbf{u}_2 \bullet \mathbf{u}_2}\mathbf{u}_2 + \cdots + \frac{\mathbf{y} \bullet \mathbf{u}_p}{\mathbf{u}_p \bullet \mathbf{u}_p}\mathbf{u}_p, \\ \mathbf{z} &= \mathbf{y} - \widehat{\mathbf{y}}. \end{aligned} \tag{C.6.2}$$

**Remark 6.38. (Properties of Orthogonal Decomposition)**
Let $\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{z}$, where $\widehat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^\perp$. Then

1. $\widehat{\mathbf{y}}$ is called the **orthogonal projection** of $\mathbf{y}$ onto $W$ ($= \text{proj}_W \mathbf{y}$)
2. $\widehat{\mathbf{y}}$ is the **closest point** to $\mathbf{y}$ in $W$.
   (in the sense $\|\mathbf{y} - \widehat{\mathbf{y}}\| \le \|\mathbf{y} - \mathbf{v}\|$, for all $\mathbf{v} \in W$)
3. $\widehat{\mathbf{y}}$ is called the **best approximation** to $\mathbf{y}$ by elements of $W$.
4. If $\mathbf{y} \in W$, then $\text{proj}_W \mathbf{y} = \mathbf{y}$.

**Example C.18.** Find the distance from $\mathbf{y}$ to the plane in $\mathbb{R}^3$ spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$.

$$\mathbf{y} = \begin{bmatrix} 5 \\ -9 \\ 5 \end{bmatrix}, \quad \mathbf{u}_1 = \begin{bmatrix} -3 \\ -5 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}$$

**Theorem 6.42.** If $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace $W$ of $\mathbb{R}^n$, then

$$\text{proj}_W \mathbf{y} = (\mathbf{y} \bullet \mathbf{u_1})\, \mathbf{u}_1 + (\mathbf{y} \bullet \mathbf{u_2})\, \mathbf{u}_2 + \cdots + (\mathbf{y} \bullet \mathbf{u}_p)\, \mathbf{u}_p. \tag{C.6.3}$$

If $U = [\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_p]$, then

$$\text{proj}_W \mathbf{y} = UU^T\mathbf{y}, \quad \text{for all } \mathbf{y} \in \mathbb{R}^n. \tag{C.6.4}$$

The orthogonal projection can be viewed as a **matrix transformation**.

**Example C.19.** Let $\mathbf{y} = \begin{bmatrix} 7 \\ 9 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, and $W = \mathit{Span}\{\mathbf{v}\}$.

(a) Find the projection matrix $UU^T$.

(b) Compute $\text{proj}_W \mathbf{y} = \dfrac{\mathbf{y} \bullet \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}}\mathbf{v}$ and $UU^T\mathbf{y}$.

(c) Find the distance from $\mathbf{y}$ to the subspace $W$.

**Solution**.

## §6.4. The Gram-Schmidt Process and QR Factorization

The **Gram-Schmidt process** is an algorithm to produce an **orthogonal or orthonormal basis** for any nonzero subspace of $\mathbb{R}^n$.

**Theorem 6.47. (The Gram-Schmidt Process)** Given a basis $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p\}$ for a nonzero subspace $W$ of $\mathbb{R}^n$, define

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x_2} \bullet \mathbf{v_1}}{\mathbf{v_1} \bullet \mathbf{v_1}} \mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x_3} \bullet \mathbf{v_1}}{\mathbf{v_1} \bullet \mathbf{v_1}} \mathbf{v}_1 - \frac{\mathbf{x_3} \bullet \mathbf{v_2}}{\mathbf{v_2} \bullet \mathbf{v_2}} \mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x_p} \bullet \mathbf{v_1}}{\mathbf{v_1} \bullet \mathbf{v_1}} \mathbf{v}_1 - \frac{\mathbf{x_p} \bullet \mathbf{v_2}}{\mathbf{v_2} \bullet \mathbf{v_2}} \mathbf{v}_2 - \cdots - \frac{\mathbf{x_p} \bullet \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \bullet \mathbf{v}_{p-1}} \mathbf{v}_{p-1}
\end{aligned} \tag{C.6.5}$$

Then $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$ is an **orthogonal basis** for $W$. In addition,

$$\mathit{Span}\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k\} = \mathit{Span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}, \quad \text{for } 1 \le k \le p. \tag{C.6.6}$$

**Remark 6.48.** For the result of the Gram-Schmidt process, define

$$\mathbf{u}_k = \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}, \quad \text{for } 1 \le k \le p. \tag{C.6.7}$$

Then $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p\}$ is an **orthonormal basis** for $W$. In practice, it is often implemented with the **normalized Gram-Schmidt process**.

**Example C.20.** Find an *orthogonal basis* for $W = \mathit{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$ and $\mathrm{proj}_W \mathbf{y}$, when

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}, \text{ and } \mathbf{y} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}.$$

**Algorithm 6.51. (QR Factorization)** Let $A = [\mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_n]$.

- Apply the Gram-Schmidt process to obtain an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$.
- Then

$$\begin{aligned} \mathbf{x}_1 &= (\mathbf{u}_1 \bullet \mathbf{x}_1)\mathbf{u}_1 \\ \mathbf{x}_2 &= (\mathbf{u}_1 \bullet \mathbf{x}_2)\mathbf{u}_1 + (\mathbf{u}_2 \bullet \mathbf{x}_2)\mathbf{u}_2 \\ \mathbf{x}_3 &= (\mathbf{u}_1 \bullet \mathbf{x}_3)\mathbf{u}_1 + (\mathbf{u}_2 \bullet \mathbf{x}_3)\mathbf{u}_2 + (\mathbf{u}_3 \bullet \mathbf{x}_3)\mathbf{u}_3 \\ &\vdots \\ \mathbf{x}_n &= \textstyle\sum_{j=1}^{n} (\mathbf{u}_j \bullet \mathbf{x}_n)\mathbf{u}_j. \end{aligned} \tag{C.6.8}$$

- Thus

$$A = [\mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_n] = QR \tag{C.6.9}$$

  implies that

$$\begin{aligned} Q &= [\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_n], \\ R &= \begin{bmatrix} \mathbf{u}_1 \bullet \mathbf{x}_1 & \mathbf{u}_1 \bullet \mathbf{x}_2 & \mathbf{u}_1 \bullet \mathbf{x}_3 & \cdots & \mathbf{u}_1 \bullet \mathbf{x}_n \\ 0 & \mathbf{u}_2 \bullet \mathbf{x}_2 & \mathbf{u}_2 \bullet \mathbf{x}_3 & \cdots & \mathbf{u}_2 \bullet \mathbf{x}_n \\ 0 & 0 & \mathbf{u}_3 \bullet \mathbf{x}_3 & \cdots & \mathbf{u}_3 \bullet \mathbf{x}_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathbf{u}_n \bullet \mathbf{x}_n \end{bmatrix} = Q^T A. \end{aligned} \tag{C.6.10}$$

- In practice, the coefficients $r_{ij} = \mathbf{u}_i \bullet \mathbf{x}_j$, $i < j$, can be saved during the (normalized) Gram-Schmidt process.

**Self-study C.21.** Find the QR factorization for $A = \begin{bmatrix} 4 & -1 \\ 3 & 2 \end{bmatrix}$.

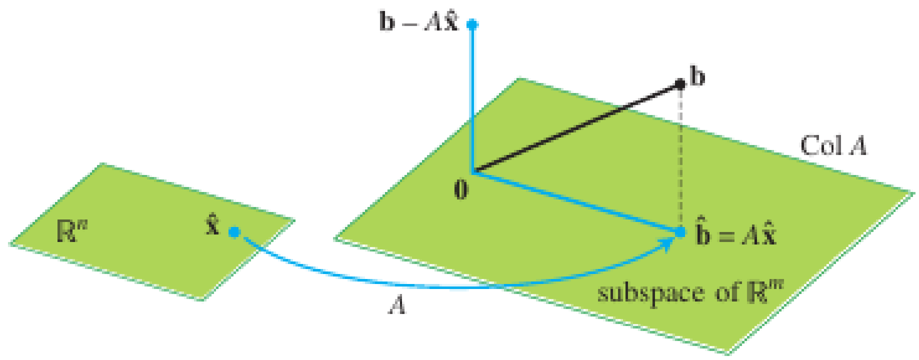
**Solution**.

*Ans*: $Q = \begin{bmatrix} 0.8 & -0.6 \\ 0.6 & 0.8 \end{bmatrix}$ $R = \begin{bmatrix} 5 & 0.4 \\ 0 & 2.2 \end{bmatrix}$

## §6.6. Machine Learning: Regression Analysis

**Definition 6.54.** Let $A$ is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$.
A **least-squares (LS) solution** of $A\mathbf{x} = \mathbf{b}$ is an $\widehat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$\|\mathbf{b} - A\widehat{\mathbf{x}}\| \le \|\mathbf{b} - A\mathbf{x}\|, \quad \text{for all } \mathbf{x} \in \mathbb{R}^n. \tag{C.6.11}$$




**Remark 6.55. Geometric Interpretation of the LS Problem**

- For all $\mathbf{x} \in \mathbb{R}^n$, **$A\mathbf{x}$ will necessarily be in *Col A*,** a subspace of $\mathbb{R}^m$.
  - So we seek an $\mathbf{x}$ that makes $A\mathbf{x}$ the closest point in $Col\,A$ to $\mathbf{b}$.
- Let $\widehat{\mathbf{b}} = \text{proj}_{Col\,A}\mathbf{b}$. Then $A\mathbf{x} = \widehat{\mathbf{b}}$ has a solution and there is an $\widehat{\mathbf{x}} \in \mathbb{R}^n$ such that
$$A\widehat{\mathbf{x}} = \widehat{\mathbf{b}}. \tag{C.6.12}$$
- $\widehat{\mathbf{x}}$ is an LS solution of $A\mathbf{x} = \mathbf{b}$.
- The quantity $\|\mathbf{b} - \widehat{\mathbf{b}}\|^2 = \|\mathbf{b} - A\widehat{\mathbf{x}}\|^2$ is called the **least-squares error**.

## The Method of Normal Equations

**Theorem 6.56**. The set of LS solutions of $A\mathrm{x} = \mathrm{b}$ coincides with the nonempty set of solutions of the **normal equations**

$$A^T A\mathbf{x} = A^T\mathbf{b}. \tag{C.6.13}$$

**Theorem 6.59**. Let A be an $m \times n$ matrix. The following statements are logically equivalent:

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

When these statements are true, the unique LS solution $\widehat{\mathrm{x}}$ is given by

$$\widehat{\mathbf{x}} = (A^T A)^{-1} A^T\mathbf{b}. \tag{C.6.14}$$

## Regression Line

**Definition 6.65**. Suppose a set of experimental data points are given as

$$(x_1, y_1),\ (x_2, y_2),\ \cdots\ ,\ (x_m, y_m)$$

such that the graph is close to a line. We determine a line

$$y = \beta_0 + \beta_1 x \tag{C.6.15}$$

that is as close as possible to the given points. This line is called the **least-squares line**; it is also called **regression line** of $y$ on $x$ and $\beta_0$, $\beta_1$ are called **regression coefficients**.

## Calculation of Least-Squares Lines

**Remark 6.66**. Consider a least-squares (LS) model of the form $y = \beta_0 + \beta_1 x$, for a given data set $\{(x_i, y_i) \mid i = 1, 2, \cdots, m\}$.

- Then

| Predicted $y$-value | | Observed $y$-value |
|---|---|---|
| $\beta_0 + \beta_1 x_1$ | $=$ | $y_1$ |
| $\beta_0 + \beta_1 x_2$ | $=$ | $y_2$ |
| $\vdots$ | | $\vdots$ |
| $\beta_0 + \beta_1 x_m$ | $=$ | $y_m$ |

(C.6.16)

- It can be equivalently written as

$$X\boldsymbol{\beta} = \mathbf{y}, \tag{C.6.17}$$

where

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

Here we call $X$ the **design matrix**, $\boldsymbol{\beta}$ the **parameter vector**, and y the **observation vector**.

- **(Method of Normal Equations)** Thus the LS solution can be determined as

$$X^T X\boldsymbol{\beta} = X^T\mathbf{y} \;\Rightarrow\; \boldsymbol{\beta} = (X^T X)^{-1} X^T\mathbf{y}, \tag{C.6.18}$$

provided that $X^T X$ is invertible.

**Self-study C.22.** Find the equation $y = \beta_0 + \beta_1 x$ of least-squares line that best fits the given points:

$(-1, 1)$, $(0, 1)$, $(1, 2)$, $(2, 3)$

**Solution**.

## Further Applications

**Example C.23.** Find an LS curve of the form $y = a\cos x + b\sin x$ that best fits the given points:

$(0, 1),\ (\pi/2, 1),\ (\pi, -1).$

**Solution.**

## Nonlinear Models: Linearization

**Strategy 6.74**. For nonlinear models, **change of variables** can be applied for a linear model.

| Model | Change of Variables | | Linearization |
|---|---|---|---|
| $y = A + \dfrac{B}{x}$ | $\widetilde{x} = \dfrac{1}{x},\ \widetilde{y} = y$ | $\Rightarrow$ | $\widetilde{y} = A + B\widetilde{x}$ |
| $y = \dfrac{1}{A + Bx}$ | $\widetilde{x} = x,\ \widetilde{y} = \dfrac{1}{y}$ | $\Rightarrow$ | $\widetilde{y} = A + B\widetilde{x}$ |
| $y = Ce^{Dx}$ | $\widetilde{x} = x,\ \widetilde{y} = \ln y$ | $\Rightarrow$ | $\widetilde{y} = \ln C + D\widetilde{x}$ |

(C.6.19)

**The Idea**: Transform the nonlinear model to produce a **linear system**.

**Self-study C.24.** Find an LS curve of the form $y = Ce^{Dx}$ that best fits the given points:

$(0, e),\ (1, e^3),\ (2, e^5).$

**Solution.**

| $x$ | $y$ |
|---|---|
| 0 | $e$ |
| 1 | $e^3$ |
| 2 | $e^5$ |

$\Rightarrow$

| $\widetilde{x}$ | $\widetilde{y} = \ln y$ |
|---|---|
| 0 | 1 |
| 1 | 3 |
| 2 | 5 |

$$\Rightarrow \quad X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix},\quad \mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}$$

*Ans*: $y = ee^{2x} = e^{2x+1}$

APPENDIX **P**

# Projects

Finally we add projects.

**Contents of Projects**

P.1. Project Regression Analysis: Linear, Piecewise Linear, and Nonlinear Models . . . . . 266

# P.1. Project Regression Analysis: Linear, Piecewise Linear, and Nonlinear Models

**Regression analysis** is a set of statistical processes for estimating the relationships between independent variables and dependent variables.

- Regression analysis is **a way to find trends in data**.
- There are variations: **linear, multiple linear, and nonlinear.**
  - The most common models are simple linear and multiple linear.
  - Nonlinear regression analysis is commonly used when the dataset shows a nonlinear relationship.
- Choosing an appropriate regression model is often a difficult task. **In this project**: we'll try to **find best regression models**, for given datasets.

**Strategy P.1. Determination of the Best Model**
Suppose we are given a dataset as in the figure.

1. We may plot the dataset for a **visual inspection**.
2. One or more good models can be selected.
3. Then, the best model is determined through analysis.

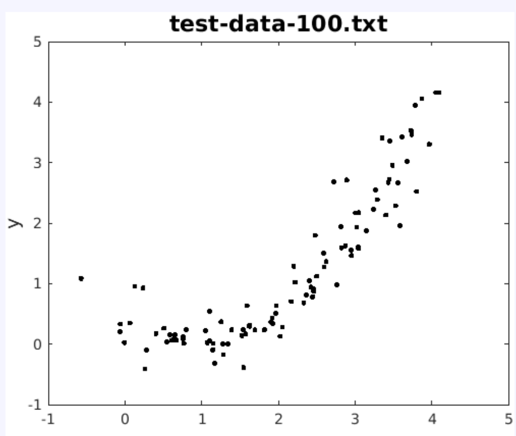



Figure P.1: A test dataset of 100 points $\{(x_i, y_i) \mid i = 1, 2, \cdots, 100\}$.

**It looks like a quadratic polynomial!**

## Polynomial Fitting: A Review

For the dataset $\{(x_i, y_i) \mid i = 1, 2, \cdots, m\}$ in Figure P.1, consider a regression model of the form

$$y = a_0 + a_1 x + a_2 x^2.$$

- Then

$$\begin{array}{ccc} \boxed{\text{Predicted } y\text{-value}} & & \boxed{\text{Observed } y\text{-value}} \\ \hline a_0 + a_1 x_1 + a_2 x_1^2 & = & y_1 \\ a_0 + a_1 x_2 + a_2 x_2^2 & = & y_2 \\ \vdots & & \vdots \\ a_0 + a_1 x_m + a_2 x_m^2 & = & y_m \\ \hline \end{array} \tag{P.1.1}$$

- It is equivalently written as

$$X\mathbf{p} = \mathbf{y}, \tag{P.1.2}$$

where

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 \end{bmatrix}, \quad \mathbf{p} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

- The system can be solved using the **method of normal equations**:

$$X^T X\mathbf{p} = \begin{bmatrix} \Sigma 1 & \Sigma x_i & \Sigma x_i^2 \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 \\ \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i^4 \end{bmatrix} \mathbf{p} = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \\ \Sigma x_i^2 y_i \end{bmatrix} = X^T\mathbf{y} \tag{P.1.3}$$

**Note**: The above **polynomial fitting** is well implemented in most programming languages.

**Matlab**: `p = polyfit(x,y,deg);`

**Python**: `p = np.polyfit(x,y,deg)`

where `x` and `y` are arrays of $x$- and $y$-coordinates, respectively, and `deg` denotes the degree of the regression polynomial. **We will use it!**

test_data_100.m

```
close all; clear all

DATA = readmatrix('test-data-100.txt');
x = DATA(:,1); y = DATA(:,2);
p = polyfit(x,y,2);
yhat = polyval(p,x);                    % predicted y-values
LS_error = norm(y-yhat)^2/length(y);   % variance

%-----------------------------------------------------
fprintf('LS_error= %.3f; p=',LS_error); disp(p)
% Output: LS_error= 0.130; p= 0.3944 -0.6824 0.3577

%-----------------------------------------------------
x1 = linspace(min(x),max(x),100);
y1 = polyval(p,x1);                     % regression curve

figure, plot(x,y,'k.','MarkerSize',8); hold on
xlim([-1,5]); ylim([-1,5]);
xlabel('x','fontsize',15); ylabel('y','fontsize',14);
title('test-data-100: Regression','fontsize',14);
plot(x1,y1,'r-','linewidth',2)
exportgraphics(gcf,'test-data-100-regression.png','Resolution',100);
```


Figure P.2: test-data-100-regression.png: $y = 0.3944 * x^2 - 0.6824 * x + 0.3577.$

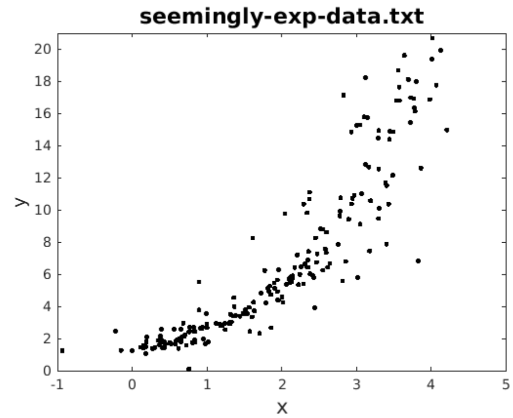
## Nonlinear Regression

**Example P.2.** See the dataset $\{(x_i, y_i)\}$ shown in the figure. When we try to find the best fitting model of the form

$$y = c\,e^{dx}, \tag{P.1.4}$$

the corresponding **nonlinear least-squares problem** reads

$$\min_{c,d} \sum_{i=1}^{m} \left(y_i - c\,e^{dx_i}\right)^2. \tag{P.1.5}$$




The problem can be solved by applying a nonlinear iterative solver such as the **Newton's method** with a good initialization $(c_0, d_0)$.

We can solve it **much more easily** through **linearization**.

**Linearization by Change of Variables**

$$y = c\,e^{dx}. \tag{P.1.6}$$

**The Goal: find the best-fitting $(c, d)$ for the dataset.**

- **(Transform)** Apply the logarithmic function to have

$$\ln y = \ln(c\,e^{dx}) = \ln c + \ln e^{dx} = dx + \ln c. \tag{P.1.7}$$

- **(Change of Variables)** Define

$$X = x; \quad Y = \ln y; \quad a = \ln c. \tag{P.1.8}$$

- **(Linear Model)** Then the model in (P.1.7) reads

$$Y = d\,X + a, \tag{P.1.9}$$

  which is a linear model; we can get best $(d, a)$, by `polyfit(X,Y,1)`.
- Finally, we recover $c = e^a$.

nonlinear_regression.m

```
lose all; clear all

FILE = 'seemingly-exp-data.txt';
DATA = readmatrix(FILE);
% fitting: y =c*exp(d*x) ---> ln(y) = d*x + ln(c)
%--------------------------------------------------

x = DATA(:,1); y = DATA(:,2);
lny = log(y);                    % data transform
p = polyfit(x,lny,1);            % [p1,p2] = [d, ln(c)]
d = p(1); c = exp(p(2));

LS_error = norm(y-c*exp(d*x))^2/length(y);  % variance

%--------------------------------------------------
fprintf('c=%.3f; d=%.3f; LS_error=%.3f\n',c,d,LS_error)
% Outout: c=1.346; d=0.669; LS_error=4.212

% figure
```

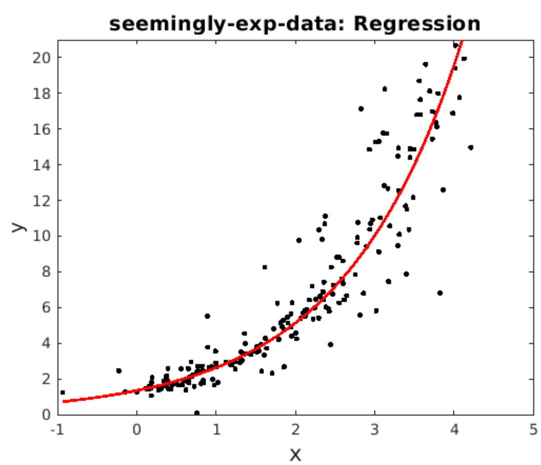



Figure P.3: The data and a nonlinear regression via linearization: `y=1.346*exp(0.669*x)`.

**Note**: If you want to test `nonlinear_regression.m`, you may download

https://skim.math.msstate.edu/LectureNotes/data/nonlinear_regression.m

https://skim.math.msstate.edu/LectureNotes/data/seemingly-exp-data.txt

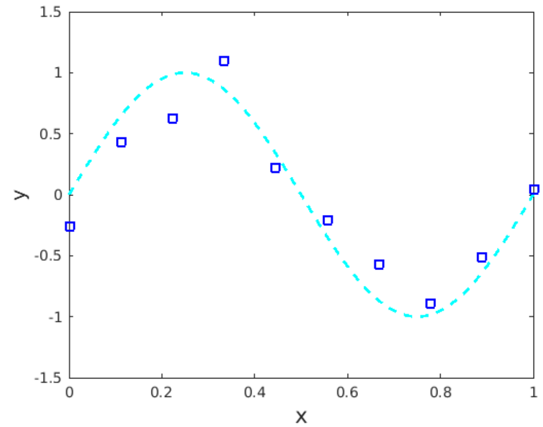
The dataset consists of 200 points, generated by `y=1.3*exp(0.7*x)` with a random positioning and random noise.

## Finding the Best Regression Model

**Example P.3.** Consider a simple dataset: 10 points generated from a sine function, with noise.

**Wanted**: Find the best regression model for the dataset:

- Let's select a model from $\mathbb{P}_n$, polynomials of degree $\leq n$




Sine_Noisy_Data_Regression.m

```
close all; clear all

a=0; b=1; m=10;
f = @(t) sin(2*pi*t);
DATAFILE = 'sine-noisy-data.txt';
renew_data = 0;

%%-----------------------------------------------
if isfile(DATAFILE) && renew_data == 0
   DATA = readmatrix(DATAFILE);        % np.loadtxt()
else
   X = linspace(a,b,m); Y0 = f(X);
   noise = rand([1,m]); noise = noise-mean(noise(:));
   Y = Y0 + noise; DATA = [X',Y'];
   writematrix(DATA,DATAFILE);         % np.savetxt()
end

%%-----------------------------------------------
x = linspace(a,b,101); y = f(x);
x1 = DATA(:,1); y1 = DATA(:,2);
E = zeros(1,m);
for n = 0:m-1
    p = polyfit(x1,y1,n);              % np.polyfit()
    yhat = polyval(p,x1);              % np.polyval()
    E(n+1) = norm(y1-yhat,2)^2;
    %savefigure(x,y,x1,y1,polyval(p,x),n)
end

% figure
```

## Which One is the Best?

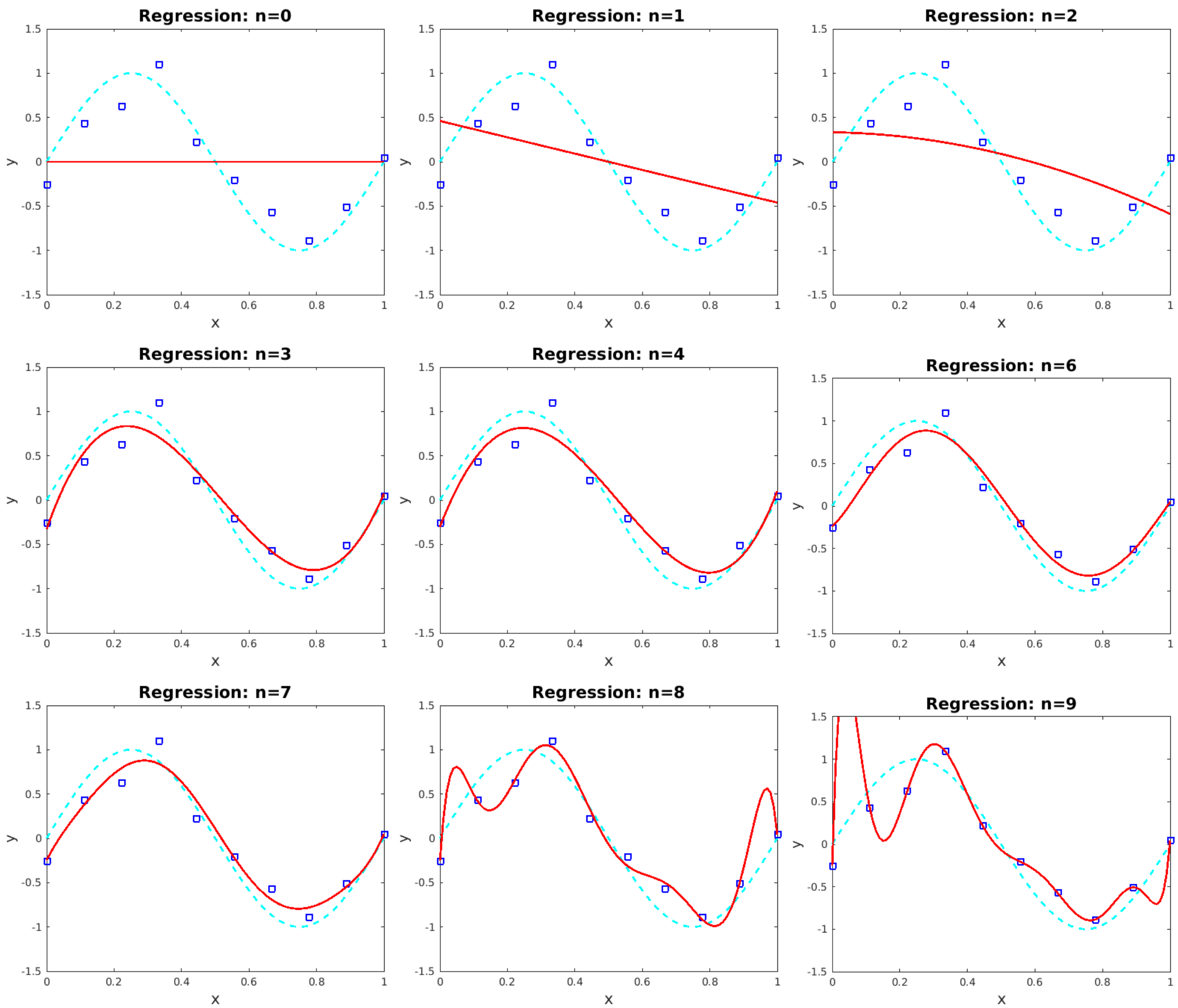



Figure P.4: Regression models $P_n$, $n = 0, 1, \cdots, 9$.

**Strategy P.4.** Given several models with similar explanatory ability, **the simplest is most likely to be the best choice.**

- Start simple, and only make the model more complex as needed.

## The LS Error

Given the dataset $\{(x_i, y_i) \mid i = 1, 2, \cdots, m\}$ and the model $P_n$, define the LS-error

$$E_n = \sum_{i=1}^{m} \big(y_i - P_n(x_i)\big)^2, \qquad (m = 10), \tag{P.1.10}$$

which is also called the **mean square error**.

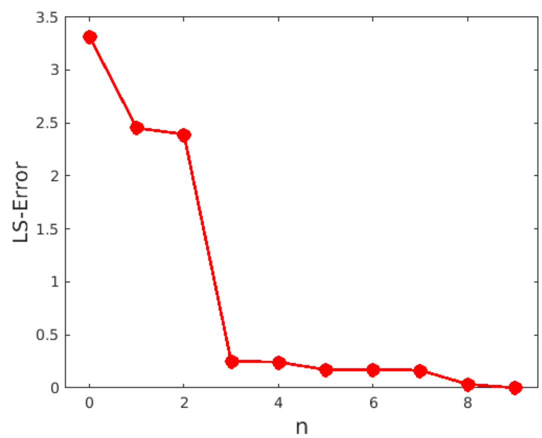
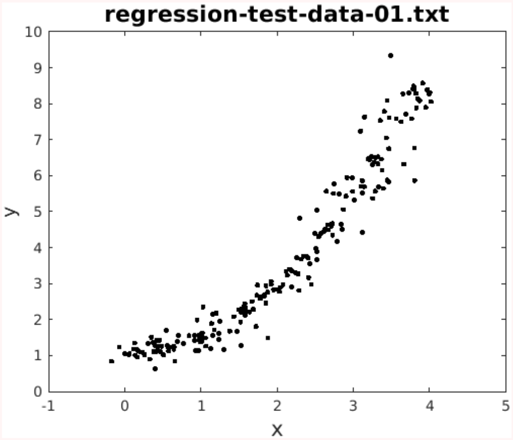


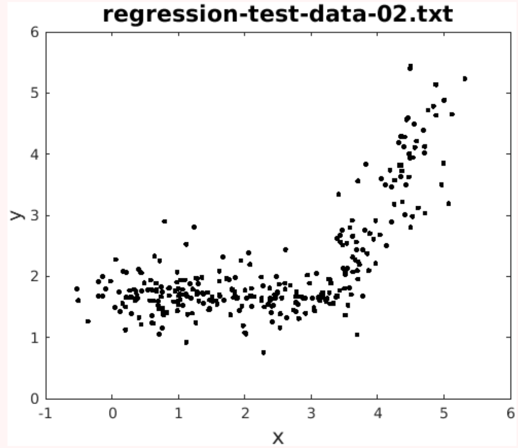

Figure P.5: **The best choice is $P_3$, the third-order polynomial.**

**Summary P.5.** Let's summarize what we have done.

- Review: method of normal equations
- Example: polynomial fitting
- Example: nonlinear regression & its linearization
- Strategy: determination of the best model

**Project Objective**: To find the best model for each of the datasets:




**What to Do**

First download two datasets:

https://skim.math.msstate.edu/LectureNotes/data/regression-test-data-01.txt
https://skim.math.msstate.edu/LectureNotes/data/regression-test-data-02.txt

1. **Finding Best Models**
   (a) For **regression-test-data-01.txt**, which model is better,
   `y=a0+a1*x+a2*x.^2` or `y = c*exp(d*x)`?
   (b) For **regression-test-data-02.txt** (Data-02), which order of polynomial is fitting the best? Your claim must be supported pictorially as in Figure P.4.
2. **Verification**: For all models, measure the LS-errors. ***Show them in tabular form.***
3. **Figuring**: For Data-02, display the LS-error as in Figure P.5.
4. **Extra Credit**: Find a **piecewise regression** model for Data-02. Is it better than polynomial models? (You must verify your answer.)

**Note**: You may use parts of the codes shown in this project. Report your code, numerical outputs, and figures, with a summary.

- A code itself will not give you any credit; include outputs or figures.
- The summary will be worth 20% the full credit.
- Include all into a single file in pdf or doc/docx format.

# Bibliography

[1] D. Lay and S. L. abd Judi McDonald, *Linear Algebra and Its Applications, 6th Ed.*, Pearson, 2021.

[2] Paul, *Paul's Online Notes*. https://tutorial.math.lamar.edu/Classes/DE/RepeatedEigenvalues.aspx.

[3] Wikipedia, *Eigenvalues and eigenvectors*. https://en.wikipedia.org/wiki/Eigenvalues_and_eigenvectors.

# Index

M, 28, 96, 141, 167, 178, 209, 225
1-norm, 232
2-norm, 232

acute angle, 189
addition, 120, 250
algebraic multiplicity, 139
all-ones vector, 233
annual_migration.m, 171
attractor, 157
augmented matrix, 4

back substitution, 93
back_sub.m, 95
backward phase, 13
basic variables, 12
basis, 99, 244
basis theorem, 107, 246
best approximation, 198, 257
birds_on_aviary.m, 172
birds_on_aviary2.m, 173

change of variables, 224, 264, 269
Chapter Review, 237
characteristic equation, 138, 150
characteristic polynomial, 138
closed, 97, 243
closest point, 198, 257
codomain, 50
coefficient matrix, 4
cofactor, 111, 247
cofactor expansion, 111, 247
column space, 97, 127, 243
column vector, 20
complex conjugate, 151
complex eigenvalue, 150
consistent system, 3
coordinate vector, 103, 245
coordinates, 103, 245

data compression, 230
design matrix, 219, 263
determinant, 73, 109–111, 247
diagonal entries, 68
diagonal matrix, 68
diagonalizable, 142, 252
diagonalization theorem, 143, 252
differential equation, 154
dimension, 105
direction of greatest attraction, 157
distance, 181
domain, 50
dot product, 36, 71, 180
dot product preservation, 193
double eigenvalue, 159
dyadic decomposition, 230

echelon form, 9
eigenspace, 133, 147
eigenvalue, 132, 138, 228
eigenvalues of stochastic matrices, 231
eigenvector, 132, 138, 228
elementary matrix, 87
elementary row operation, 87
elementary row operations, 4, 115
ellipse, 229
equal, 68
equality of vectors, 21
equivalent system, 2
Euler angles, 54
Euler, Leonhard, 228
exercise-6.5.3.m, 216
existence and uniqueness, 3

finite Markov chain, 169
for loop, 33

forward elimination, 92
forward phase, 13, 18, 90
forward substitution, 92
forward_sub.m, 94
free variable, 14
free variables, 12
function, 33
fundamental questions, two, 3

Gauss Elimination, 90
Gauss elimination with partial pivoting, 94
general solution, 14
generalized eigenvector, 159
geometric linear transformations of $\mathbb{R}^2$, 60
Gram-Schmidt process, 203, 206, 259
Gram-Schmidt process, normalized, 204, 259

homogeneous linear system, 38

inconsistent system, 3
information matrix, 210
initial condition, 154
initial-value problem, 154
injective, 61
inner product, 180, 182
interchange operation, 235
inverse power method, 165
inverse_power.m, 166
invertibility, 73
invertible, 84, 242
invertible linear transformation, 84, 242
invertible matrix, 77
invertible matrix theorem, 82, 107, 116, 139, 241
isomorphic, 104
isomorphism, 104
iteration, 32
iterative algorithm, 162

kernel, 129

Lagrange, Joseph-Louis, 228
leading 1, 9
leading entry, 9
least-squares (LS) problem, 210
least-squares (LS) solution, 210, 211, 217, 261
least-squares error, 211, 212, 261
least-squares line, 218, 262
left singular vectors, 230
length, 181
length preservation, 193
linear combination, 23
linear equation, 2
linear space, 119
linear system, 2
linear transformation, 53, 129
linearization, 269
linearly dependent, 44, 238
linearly independent, 44, 80, 238
linearly independent eigenvectors, 162
linspace, in Matlab, 31
lower triangular matrix, 86
lower-triangular system, 92
LU decomposition, 87
LU factorization, 87
LU factorization algorithm, 89
lu_nopivot_overwrite.m, 91
lu_solve.m, 95

machine learning, 217
Markov chain, 168, 170, 231, 254
Markov matrix, 170, 231, 254
Matlab, 29
matlab: fileparts, 270
matlab: polyfit, 270
matlab: readmatrix, 271
matlab: writematrix, 271
matrix algebra, 67
matrix equation, 34
matrix form, 4
matrix multiplication, 70, 228
matrix product, 70
matrix transformation, 51, 200, 228, 258
mean square error, 273
method of normal equations, 212, 221, 262, 267
multiplication by scalars, 120, 250
multiplicity, 139

mysum.m, 33

Newton's method, 269
nonhomogeneous linear system, 38
nonlinear least-squares problem, 269
nonlinear_regression.m, 270
nonsingular matrix, 77
nontrivial solution, 38
nonzero row, 9
norm, 181
normal equations, 212, 217, 262
np.linalg.eig, 234
np.loadtxt, 271
np.polyfit, 271
np.polyval, 271
np.savetxt, 271
np.set_printoptions, 234
null space, 98, 125, 129, 243
nullity, 105, 246

objects, 29
observation vector, 210, 219, 263
obtuse angle, 189
Octave, 29
one-to-one, 61, 64
onto, 61
orthogonal, 183
orthogonal basis, 188, 203, 204, 259
orthogonal complement, 183
orthogonal decomposition theorem, 196, 257
orthogonal matrix, 193
orthogonal projection, 189, 196, 198, 257
orthogonal set, 187
orthogonal_matrix.m, 193
orthogonality preservation, 193
orthonormal basis, 192, 203, 204, 259
orthonormal set, 192

p-norm, 232
parallelogram law, 186
parallelogram rule for addition, 21
parameter vector, 219, 263
parametric description, 14
parametric vector form, 19, 39, 43
partial pivoting, 94

Perron-Frobenius eigenvector, 233
piecewise regression, 274
pivot column, 11
pivot position, 11
plot, in Matlab, 30
polynomial fitting, 267
position vector, 20
power iteration, 162
power method, 162, 175, 234
power_iteration.m, 164
principal component analysis, 230
principal components, 230
probability matrix, 170, 231, 254
probability vector, 170, 231, 254
programming, 29
pseudoinverse, 213, 214, 217
Pythagorean theorem, 183, 198

QR factorization, 206
QR factorization algorithm, 207, 260

random orthogonal matrix, 193
random sample consensus, 210
range, 50, 129
rank, 105
rank theorem, 105, 246
RANSAC, 210
reduced echelon form, 9
REF, 9
reflections in $\mathbb{R}^2$, 60
regression analysis, 217, 230, 266
regression coefficients, 218, 262
regression line, 218, 262
regular, 175, 256
regular stochastic matrix, 234
regular_stochastic.m, 175
regular_stochastic_Tk.m, 176
repeller, 157
repetition, 29, 32
replacement, 233
reusability, 33
reusable, 29
right singular vectors, 230
roll-pitch-yaw, 54
rotation, 54

row equivalent, 5
row reduced echelon form, 9
row space, 185
row-column rule, 71
RREF, 9

saddle point, 157
scalar multiple, 21
scalar multiplication, 68, 119
shear transformation, 51, 59
similar, 140
similarity, 140
similarity transformation, 140
Sine_Noisy_Data_Regression.m, 271
singular value decomposition, 230
singular values, 230
sink, 157
solution, 2
solution set, 2
source, 157
span, 25
sparse system, 87
square matrix, 68
stable equilibrium distribution, 233
standard basis, 99
standard matrix, 58, 84, 239, 242
standard unit vectors in $\mathbb{R}^n$, 58, 239
state vector, 170, 231, 254
states, 168
steady-state vector, 174, 233, 255
stochastic matrix, 170, 231, 254
stochastic_eigen.py, 234
submatrix, 111, 247
subspace, 97, 121, 243, 250
substitution matrix, 170, 231, 254
sum, 123, 251
sum of products, 71
sum of two matrices, 68
superposition principle, 53, 55
surjective, 61
SVD theorem, 230
system of linear equations, iii, 2

test_data_100.m, 268
transformation, 50
transition matrix, 169, 170, 172, 231, 254
transpose, 19, 74
trivial solution, 38

unique inverse, 77
unit circle, 229
unit lower triangular matrix, 87
unit vector, 181, 192
upper triangular matrix, 84, 87, 242
upper-triangular system, 93

vector, 20
vector addition, 21
vector equation, 24, 36
vector space, 119, 120, 250
vectors, 119, 120, 250
visual inspection, 266
volume scaling factor, 109, 110, 247

yaw, 54

zero subspace, 121