

OFFICIAL COURSE STUDY GUIDE

---

# MIT 18.06: Linear Algebra

---

*Gilbert Strang's world-famous MIT course on linear algebra.*

---

SCIENCE & MATH · INTERMEDIATE · 5 MODULES · 27 HR 60 MIN

Course by MIT OpenCourseWare · Curated and published by MegaCourse

**MEGACOURSE.APP**

## **MIT 18.06: Linear Algebra: The Complete Study Guide**

2026 Edition

Published by MegaCourse · [megacourse.app](https://megacourse.app)

Learn Anything. Achieve Everything.

MegaCourse curates the world's best free courses into structured learning paths with study guides, practice material, and certificates. This guide accompanies the MegaCourse course "MIT 18.06: Linear Algebra", taught by MIT OpenCourseWare. It is designed to be read alongside the video lessons and to serve as a standalone reference for review.

Study guide text copyright © 2026 MegaCourse. Course video content remains the property of its creators. You may print this guide for personal study. Redistribution or resale is not permitted.

Enroll, track progress, and earn your certificate at  
**[megacourse.app/courses/mit-18-06-linear-algebra](https://megacourse.app/courses/mit-18-06-linear-algebra)**

# Contents

---

|                                                                                |    |
|--------------------------------------------------------------------------------|----|
| <i>About This Course</i>                                                       | 4  |
| CHAPTER 1    Linear Systems, Elimination, and Vector Spaces                    | 6  |
| CHAPTER 2    The Four Fundamental Subspaces and Orthogonality                  | 12 |
| CHAPTER 3    Orthogonality, Determinants, and Eigenvalues                      | 17 |
| CHAPTER 4    Advanced Matrix Structures and Decompositions                     | 24 |
| CHAPTER 5    Basis Transformations, Generalized Inverses, and Course Synthesis | 30 |
| <i>Glossary</i>                                                                | 36 |
| <i>Answer Key</i>                                                              | 39 |
| <i>Where to Go Next</i>                                                        | 43 |

## INTRODUCTION

# About This Course

---

This guide accompanies "MIT 18.06: Linear Algebra", a world-renowned course taught by Gilbert Strang. Linear algebra is the mathematics of vectors, matrices, and linear systems. This course provides a complete understanding of these fundamental concepts.

Linear algebra underpins many modern technical fields. It is essential for machine learning, data science, computer graphics, engineering, and physics. A strong grasp of its principles will serve as a foundational skill for your academic and professional pursuits.

MegaCourse has organized this material into a structured learning path with 5 modules and 36 lessons. Professor Strang's clear explanations are complemented by MegaCourse's learning outcomes, progress tracking, and a certificate of completion. This guide reinforces the video lectures and helps you master the subject.

## Course Facts

|                      |                                                                                                                       |
|----------------------|-----------------------------------------------------------------------------------------------------------------------|
| FORMAT               | 5 modules · 36 lessons                                                                                                |
| VIDEO<br>INSTRUCTION | About 27 hr 60 min                                                                                                    |
| LEVEL                | Intermediate                                                                                                          |
| CATEGORY             | Science & Math                                                                                                        |
| INSTRUCTOR           | MIT OpenCourseWare                                                                                                    |
| CERTIFICATE          | Included with course completion on MegaCourse                                                                         |
| COURSE HOME          | <a href="https://megacourse.app/courses/mit-18-06-linear-algebra">megacourse.app/courses/mit-18-06-linear-algebra</a> |

## What You Will Learn

- Understand vectors, matrices, and linear systems.
- Master eigenvalues, eigenvectors, and diagonalization.
- Apply orthogonality, projections, and least squares.
- Build the math foundation for machine learning and engineering.

## Who This Course Is For

This course is designed for students, engineers, data scientists, and anyone seeking a rigorous mathematical foundation. It is particularly valuable for those preparing for advanced studies in STEM

fields or careers requiring quantitative analysis.

Before beginning, you should have a solid grasp of high-school algebra. This includes working with equations, variables, and basic functions. No prior exposure to linear algebra is assumed.

## Before You Start

- Possess a solid understanding of high-school algebra.

## How to Use This Guide

To get the most from this study guide, approach the course systematically. We recommend watching the video lectures for each lesson first. Then, use this guide to review key concepts, work through examples, and complete practice problems. This dual approach reinforces your learning.

Pace yourself deliberately. The course contains 5 modules, 36 lessons, and about 27 hours of instruction. Aim for consistent study sessions rather than cramming. Understanding each concept fully before moving on will build a stronger foundation.

- **Review Summaries:** Each section of this guide includes summaries of the main ideas from the video lectures. Use these for quick recall.
- **Practice Problems:** Attempt all practice problems. They are designed to test your understanding and application of the concepts.
- **Glossary:** Refer to the glossary for definitions of important terms.

## CHAPTER 1

# Linear Systems, Elimination, and Vector Spaces

---

Welcome to Chapter 1 of the official study guide for MIT 18.06 Linear Algebra. This chapter introduces the fundamental concepts of linear algebra, laying the groundwork for understanding vectors, matrices, and linear systems. You will learn to visualize linear equations, master the essential technique of Gaussian elimination, and understand the core ideas of vector spaces and their fundamental subspaces.

This material is crucial for building a solid mathematical foundation applicable across various fields, including machine learning and engineering. We begin by exploring the geometry of linear equations, then move to systematic methods for solving them, and finally introduce the abstract yet practical concepts of vector spaces.

## The Geometry of Linear Systems

---

Covers: 1. The Geometry of Linear Equations

Linear algebra begins with understanding linear equations. A linear equation in two variables, such as  $2x - y = 1$ , represents a straight line in a two-dimensional plane. When we have a system of linear equations, we are looking for values of the variables that satisfy all equations simultaneously. Geometrically, this means finding the intersection point of these lines.

Consider a system with two equations and two unknowns:

$$\begin{aligned} 2x - y &= 1 \\ x + y &= 5 \end{aligned}$$

There are two primary ways to visualize this system: the row picture and the column picture.

### The Row Picture

In the row picture, each equation is a 'row' that defines a geometric object. For two variables, each equation is a line. The solution to the system is the point where these lines intersect. If the lines are parallel, there is no solution. If they are the same line, there are infinitely many solutions.

For three variables, say  $2x - y + z = 1$ , each equation represents a plane in three-dimensional space. The solution to a system of three equations in three unknowns is the point where the three planes intersect. Planes can intersect at a single point, along a line (infinitely many solutions), or not at all (no solution).

## The Column Picture

The column picture offers a different, powerful perspective. We rewrite the system of equations as a vector equation. For the system above:

$$x \begin{bmatrix} 2 \\ 1 \end{bmatrix} + y \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$$

Here,  $x$  and  $y$  are scalars, and  $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$  and  $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$  are column vectors. The problem becomes: find the linear combination of the column vectors that equals the right-hand side vector  $\begin{bmatrix} 1 \\ 5 \end{bmatrix}$ . This means we are scaling the column vectors and adding them to reach the target vector.

Vectors are fundamental. A vector in  $\mathbb{R}^n$  is an ordered list of  $n$  numbers. We can add vectors and multiply them by scalars. These operations allow us to form linear combinations. The set of all possible linear combinations of a given set of vectors forms a vector space or subspace.

**Example:** For  $2x - y = 1$  and  $x + y = 5$ :

- Row Picture:** Draw the line  $y = 2x - 1$  and the line  $y = -x + 5$ . They intersect at  $(2, 3)$ . So  $x=2$ ,  $y=3$  is the solution.
- Column Picture:** We need  $x \begin{bmatrix} 2 \\ 1 \end{bmatrix} + y \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$ . If we take  $x=2$  and  $y=3$ :  $2 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 3 \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} + \begin{bmatrix} -3 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$ . This confirms the solution.

Understanding both pictures is vital. The row picture gives an intuitive geometric understanding of intersections, while the column picture introduces the concept of linear combinations and vector spaces, which are central to linear algebra.

**Common Pitfalls:** Students sometimes struggle to visualize systems with more than three variables. While direct visualization is impossible, the underlying algebraic principles and the column picture's concept of linear combinations still hold true in higher dimensions. Do not confuse the geometric interpretation of a single equation (a line or plane) with the geometric interpretation of the *solution* to a system (a point).

**Practice Cue:** Sketch the row and column pictures for simple  $2 \times 2$  systems. Try to predict if there will be one solution, no solutions, or infinitely many solutions before solving algebraically. For example, consider  $x + y = 2$  and  $2x + 2y = 4$  (infinitely many solutions), or  $x + y = 2$  and  $x + y = 3$  (no solution).

## Solving Linear Systems: Elimination and Matrix Operations

Covers: 2. Elimination with Matrices · 3. Multiplication and Inverse Matrices · 4. Factorization into  $A = LU$

Solving systems of linear equations algebraically is often done using Gaussian elimination. This systematic process transforms a system into an equivalent, simpler system that is easy to solve by back-substitution. The key idea is to eliminate variables one by one.

### Gaussian Elimination

Consider the system  $Ax = b$ . Gaussian elimination involves three elementary row operations:

1. Swap two rows.
2. Multiply a row by a non-zero scalar.
3. Add a multiple of one row to another row.

The goal is to transform the matrix  $A$  into an upper triangular matrix  $U$ . The leading non-zero entry in each row of  $U$  is called a pivot. If a pivot is zero, we may need to perform row swaps (using a permutation matrix  $P$ ).

**Example:** Solve  $x + 2y = 5$  and  $3x + 4y = 11$ .

We write this as an augmented matrix:

$$\begin{bmatrix} 1 & 2 & | & 5 \\ 3 & 4 & | & 11 \end{bmatrix}$$

1. Subtract 3 times Row 1 from Row 2 (to eliminate  $x$  from the second equation):  $R_2 \rightarrow R_2 - 3R_1$

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

Now we have an upper triangular form. From the second row,  $-2y = -4$ , so  $y = 2$ . Substitute  $y=2$  into the first row:  $x + 2(2) = 5$ , so  $x + 4 = 5$ , which means  $x = 1$ . The solution is  $(1, 2)$ .

### Matrix Multiplication and Inverse

Matrices are not just a convenient notation; they are mathematical objects with their own algebra. Matrix multiplication  $AB$  is defined such that the number of columns in  $A$  must equal the number of rows in  $B$ . The entry  $(AB)_{ij}$  is the dot product of the  $i$ -th row of  $A$  and the  $j$ -th column of  $B$ .

Matrix multiplication is generally not commutative (i.e.,  $AB$  is not usually equal to  $BA$ ). It is associative:  $(AB)C = A(BC)$ .

The identity matrix, denoted  $I$ , is a square matrix with ones on the main diagonal and zeros elsewhere. It acts like the number 1 in scalar multiplication:  $AI = IA = A$ .

An inverse matrix  $A^{-1}$  exists for a square matrix  $A$  if  $AA^{-1} = A^{-1}A = I$ . Not all square matrices have inverses. Matrices without inverses are called singular. We can find  $A^{-1}$  using Gauss-Jordan elimination by augmenting  $A$  with  $I$  and reducing  $A$  to  $I$ :

$$[A \mid I] \rightarrow [I \mid A^{-1}]$$

### LU Factorization

For many practical computations, especially when solving  $Ax = b$  for multiple  $b$  vectors, it is efficient to factor  $A$  into  $LU$ , where  $L$  is a lower triangular matrix and  $U$  is an upper triangular matrix. This factorization comes directly from Gaussian elimination. The entries of  $L$  are the multipliers used during elimination.

If  $A = LU$ , then  $Ax = b$  becomes  $LUx = b$ . We can solve this in two steps:

1. Let  $Ux = c$ . Solve  $Lc = b$  for  $c$  using forward substitution (easy because  $L$  is triangular).
2. Then solve  $Ux = c$  for  $x$  using back-substitution (easy because  $U$  is triangular).

**Common Pitfalls:** A common error in matrix multiplication is getting the dimensions wrong or performing element-wise multiplication instead of the row-column dot product. When finding an inverse, remember that only square matrices *can* have an inverse, and even then, many are singular. For LU factorization, ensure the multipliers are correctly placed in  $L$ .

**Practice Cue:** Perform Gaussian elimination on a  $3 \times 3$  system. Multiply two  $2 \times 2$  or  $3 \times 3$  matrices. Find the inverse of a  $2 \times 2$  matrix. Factor a  $2 \times 2$  or  $3 \times 3$  matrix into  $LU$ .

## Vector Spaces and Fundamental Subspaces

Covers: 5. Transposes, Permutations, Spaces  $\mathbb{R}^n$  · 6. Column Space and Nullspace · 7. Solving  $Ax = 0$ : Pivot Variables, Special Solutions

Linear algebra moves beyond solving specific systems to understanding the structure of solutions. This requires the concept of vector spaces and their subspaces.

### Vector Spaces and Subspaces

A vector space is a collection of vectors that satisfies certain properties, allowing for vector addition and scalar multiplication while remaining within the space. The most familiar vector space is  $\mathbb{R}^n$ , the set of all  $n$ -dimensional real vectors. A subspace is a subset of a vector space that is itself a vector space. For a subset  $S$  to be a subspace, it must satisfy three conditions:

1. It contains the zero vector.
2. It is closed under vector addition (if  $v$  and  $w$  are in  $S$ , then  $v + w$  is in  $S$ ).
3. It is closed under scalar multiplication (if  $v$  is in  $S$  and  $c$  is a scalar, then  $cv$  is in  $S$ ).

Lines through the origin and planes through the origin are examples of subspaces in  $\mathbb{R}^3$ .

### Transposes and Permutations

The transpose of a matrix  $A$ , denoted  $A^T$ , is obtained by swapping its rows and columns. If  $A$  is  $m \times n$ , then  $A^T$  is  $n \times m$ . Properties include  $(A^T)^T = A$  and  $(AB)^T = B^T A^T$ .

Permutation matrices  $P$  are square matrices that have exactly one '1' in each row and column, and zeros elsewhere. Multiplying by  $P$  reorders the rows of a matrix. They are essential in Gaussian elimination when row swaps are needed to avoid zero pivots. A key property is that  $P^{-1} = P^T$ .

### Column Space and Nullspace

These are two of the four fundamental subspaces associated with a matrix  $A$ .

1. **Column Space ( $C(A)$ ):** This is the set of all possible linear combinations of the columns of  $A$ . If  $A$  is  $m \times n$ , its columns are in  $\mathbb{R}^m$ , so  $C(A)$  is a subspace of  $\mathbb{R}^m$ . The column space tells us for which vectors  $b$  the equation  $Ax = b$  has a solution. If  $b$  is in  $C(A)$ , a solution exists.

**2. Nullspace  $N(A)$ :** This is the set of all vectors  $x$  such that  $Ax = 0$ . If  $A$  is  $m \times n$ , then  $x$  is in  $\mathbb{R}^n$ , so  $N(A)$  is a subspace of  $\mathbb{R}^n$ . The nullspace describes all solutions to the homogeneous equation  $Ax = 0$ .

### Solving $Ax = 0$ : Pivot Variables, Special Solutions

To find the nullspace, we solve  $Ax = 0$  using Gaussian elimination to reduce  $A$  to its row echelon form  $U$  (or reduced row echelon form  $R$ ).

- **Pivot Variables:** These correspond to columns in  $U$  that contain pivots. These variables are determined by the free variables.
- **Free Variables:** These correspond to columns in  $U$  that do not contain pivots. We can assign any value to these variables.

By setting one free variable to 1 and others to 0, we find 'special solutions' to  $Ax = 0$ . The nullspace is the set of all linear combinations of these special solutions.

**Example:** Find the nullspace of  $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \end{bmatrix}$

1. Reduce  $A$  to  $U$  (or  $R$ ):  $R_2 \rightarrow R_2 - 2R_1$

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

2. The first column has a pivot, so  $x_1$  is a pivot variable. Columns 2 and 3 do not have pivots, so  $x_2$  and  $x_3$  are free variables.
3. Set  $x_2 = 1$ ,  $x_3 = 0$ . From  $x_1 + 2x_2 + 3x_3 = 0$ , we get  $x_1 + 2(1) + 3(0) = 0$ , so  $x_1 = -2$ . First special solution:  $s_1 = [-2, 1, 0]^T$ .
4. Set  $x_2 = 0$ ,  $x_3 = 1$ . From  $x_1 + 2x_2 + 3x_3 = 0$ , we get  $x_1 + 2(0) + 3(1) = 0$ , so  $x_1 = -3$ . Second special solution:  $s_2 = [-3, 0, 1]^T$ .

The nullspace  $N(A)$  is all linear combinations of  $s_1$  and  $s_2$ :  $c_1s_1 + c_2s_2$ .

**Common Pitfalls:** Confusing the column space (which describes when  $Ax=b$  has a solution) with the nullspace (which describes the solutions to  $Ax=0$ ). Remember the column space lives in  $\mathbb{R}^m$  and the nullspace in  $\mathbb{R}^n$  for an  $m \times n$  matrix  $A$ . Ensure you correctly identify pivot and free variables during elimination.

**Practice Cue:** For a given matrix, find its transpose. Identify pivot and free variables after performing elimination. Determine the special solutions that form a basis for the nullspace.

**KEY TAKEAWAYS**

- Linear equations can be visualized through both row pictures (intersections of lines or planes) and column pictures (linear combinations of vectors).
- Gaussian elimination systematically transforms a system of equations into an upper triangular form, allowing for easy solution by back-substitution.
- Matrix multiplication combines matrices, while the inverse matrix (if it exists) 'undoes' the original matrix's operation.
- LU factorization decomposes a matrix into lower and upper triangular forms, simplifying the solution of  $Ax = b$  into two easier steps.
- A vector space is a collection of vectors closed under addition and scalar multiplication; a subspace is a vector space contained within another.
- The column space of a matrix contains all possible output vectors  $b$  for  $Ax = b$ , while the nullspace contains all input vectors  $x$  that result in  $Ax = 0$ .

**CHECK YOUR UNDERSTANDING**

Answer these before moving on. Model answers are in the Answer Key at the back of this guide.

1. Explain the difference between the row picture and the column picture for a system of linear equations.
2. What are the three elementary row operations used in Gaussian elimination, and why are they important?
3. Under what conditions does a square matrix  $A$  have an inverse  $A^{-1}$ ?
4. What is the primary benefit of performing an LU factorization on a matrix  $A$ ?
5. Define the column space and nullspace of a matrix  $A$ .

## CHAPTER 2

# The Four Fundamental Subspaces and Orthogonality

---

Chapter 2 builds upon your understanding of solving linear systems, moving beyond simple computation to a deeper appreciation of the structure of solutions. You will explore the concepts of linear independence, basis, and dimension, which are essential for characterizing vector spaces. This chapter then introduces the four fundamental subspaces associated with any matrix, revealing their interconnectedness and importance. Finally, you will learn about orthogonality and projections, powerful tools for understanding geometric relationships between vectors and subspaces, culminating in the method of least squares.

## Solving Linear Systems and Understanding Basis

---

Covers: 8. Solving  $Ax = b$ : Row Reduced Form  $R$  · 9. Independence, Basis, and Dimension

This section deepens your understanding of solving the fundamental linear system  $Ax = b$ . You learned about Gaussian elimination in previous lessons. Now, we focus on the structure of the solutions and the underlying spaces.

### The Complete Solution to $Ax = b$

When solving  $Ax = b$ , the complete solution  $x$  is a sum of two parts: a particular solution  $x_p$  and any vector from the null space  $x_n$ . That is,  $x = x_p + x_n$ . The particular solution  $x_p$  is found by setting all free variables to zero and solving for the pivot variables. The null space  $N(A)$  consists of all vectors  $x$  such that  $Ax = 0$ . We find a basis for the null space by identifying 'special solutions'. For each free variable, set it to 1 and all other free variables to 0, then solve for the pivot variables. These special solutions form a basis for  $N(A)$ .

Consider the augmented matrix  $[A \mid b]$ . Gaussian elimination transforms this into  $[U \mid c]$ , where  $U$  is in echelon form. Further elimination produces the row reduced echelon form (RREF), denoted  $R$ . The RREF has pivots equal to 1, with zeros above and below each pivot. This form makes it straightforward to identify pivot and free variables and to write down the special solutions for the null space.

For example, if  $A$  is a  $3 \times 4$  matrix and its RREF  $R$  has pivots in columns 1 and 3, then columns 2 and 4 correspond to free variables. To find  $x_p$ , set free variables  $x_2 = 0$  and  $x_4 = 0$ , then solve for  $x_1$  and  $x_3$ . To find the special solutions for  $N(A)$ , first set  $b = 0$ . Then, set  $x_2 = 1$ ,  $x_4 = 0$  to find the first special solution. Set  $x_2 = 0$ ,  $x_4 = 1$  to find the second. Any  $x_n$  is a linear combination of these special solutions.

### Linear Independence, Basis, and Dimension

These concepts are fundamental to describing vector spaces.

**Linear Independence:** A set of vectors  $\{v_1, v_2, \dots, v_k\}$  is linearly independent if the only way to form the zero vector  $0$  as a linear combination is by setting all coefficients to zero:  $c_1 v_1 + c_2 v_2 + \dots + c_k v_k = 0$  implies  $c_1 = c_2 = \dots = c_k = 0$ . If other combinations exist, the vectors are linearly dependent.

**Span:** The span of a set of vectors is the set of all possible linear combinations of those vectors. It forms a subspace.

**Basis:** A basis for a vector space (or subspace) is a set of vectors that satisfies two conditions: they are linearly independent, and they span the entire space. All bases for a given space have the same number of vectors.

**Dimension:** The dimension of a vector space is the number of vectors in any of its bases. For example, the dimension of the null space  $N(A)$  is the number of free variables, which is  $n - r$ , where  $n$  is the number of columns in  $A$  and  $r$  is its rank (number of pivot variables).

Let's find a basis for the null space of a matrix  $A$ .

Consider the matrix:

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

This matrix is already in RREF. The pivot variables are  $x_1$  and  $x_3$  (columns 1 and 3). The free variables are  $x_2$  and  $x_4$  (columns 2 and 4).

To find the special solutions for  $Ax = 0$ :

1. Set  $x_2 = 1, x_4 = 0$ :  $x_1 + 2(1) + 0(0) = 0 \Rightarrow x_1 = -2$   $x_3 + 3(0) = 0 \Rightarrow x_3 = 0$  Special solution 1:  $s_1 = [-2, 1, 0, 0]^T$
2. Set  $x_2 = 0, x_4 = 1$ :  $x_1 + 2(0) + 1(1) = 0 \Rightarrow x_1 = -1$   $x_3 + 3(1) = 0 \Rightarrow x_3 = -3$  Special solution 2:  $s_2 = [-1, 0, -3, 1]^T$

The basis for  $N(A)$  is  $\{s_1, s_2\}$ . The dimension of  $N(A)$  is 2.

## The Four Fundamental Subspaces

Covers: 10. The Four Fundamental Subspaces · 11. Matrix Spaces; Rank 1; Small World Graphs · 12. Graphs, Networks, Incidence Matrices

Every matrix  $A$  has four fundamental subspaces, which are central to linear algebra. Understanding these subspaces provides a complete picture of how a matrix transforms vectors and the structure of its solutions.

### The Four Subspaces and Their Dimensions

Let  $A$  be an  $m \times n$  matrix with rank  $r$  (the number of pivots).

1. **Column Space,  $C(A)$**  : This is the span of the columns of  $A$ . It consists of all possible vectors  $b$  for which  $Ax = b$  has a solution. A basis for  $C(A)$  consists of the pivot columns of  $A$  (not  $R$ ). The dimension of  $C(A)$  is  $r$ .
2. **Null Space,  $N(A)$**  : This is the set of all vectors  $x$  such that  $Ax = 0$ . We discussed finding a basis for  $N(A)$  using special solutions. The dimension of  $N(A)$  is  $n - r$  (the number of free variables).
3. **Row Space,  $C(A^T)$**  : This is the span of the rows of  $A$ , or equivalently, the column space of  $A^T$ . A basis for  $C(A^T)$  consists of the non-zero rows of the RREF  $R$  (or the non-zero rows of the echelon form  $U$ ). The dimension of  $C(A^T)$  is  $r$ .
4. **Left Null Space,  $N(A^T)$**  : This is the set of all vectors  $y$  such that  $A^T y = 0$ . It is also known as the null space of  $A^T$ . The dimension of  $N(A^T)$  is  $m - r$ .

These dimensions are fundamental. Notice that  $\dim(C(A)) = \dim(C(A^T)) = r$ . Also,  $\dim(C(A)) + \dim(N(A^T)) = r + (m - r) = m$ , and  $\dim(C(A^T)) + \dim(N(A)) = r + (n - r) = n$ .

### Rank 1 Matrices

A matrix has rank 1 if its column space has dimension 1. This means all columns are multiples of a single vector. Such a matrix can always be written as the outer product of two vectors:  $A = uv^T$ . For example, if  $u = [1, 2]^T$  and  $v = [3, 4]^T$ , then  $A = \begin{bmatrix} [1] & [2] \end{bmatrix} \begin{bmatrix} [3] & [4] \end{bmatrix} = \begin{bmatrix} [3] & [4] \\ [6] & [8] \end{bmatrix}$ . Every row is a multiple of  $v^T$ , and every column is a multiple of  $u$ . Any matrix of rank  $r$  can be expressed as a sum of  $r$  rank 1 matrices.

### Graphs, Networks, and Incidence Matrices

Linear algebra is powerful for analyzing networks, often represented as graphs. A graph consists of nodes (vertices) and edges connecting them. An **incidence matrix**  $A$  describes the connections in a graph. For a graph with  $m$  edges and  $n$  nodes,  $A$  is an  $m \times n$  matrix.

Each row of  $A$  corresponds to an edge, and each column to a node. If an edge connects node  $j$  to node  $k$ , the entry  $A_{ij}$  might be  $-1$  and  $A_{ik}$  might be  $1$  (assuming a direction for the edge). Otherwise, the entry is  $0$ .

Let  $x$  be a vector of potentials (voltages) at each node. Then  $Ax$  represents the potential differences across each edge. The null space  $N(A)$  reveals important properties:  $Ax = 0$  means all potential differences are zero, implying  $x$  is a constant vector (all nodes at the same potential). The dimension of  $N(A)$  is related to the number of separate components in the graph.

The left null space  $N(A^T)$  is crucial for understanding current flow. If  $y$  is a vector of currents flowing through the edges, then  $A^T y = 0$  represents Kirchhoff's Current Law, stating that the net current entering or leaving each node is zero. The dimension of  $N(A^T)$  is  $m - r$ , which corresponds to the number of independent loops in the graph. The column space  $C(A)$  relates to the existence of potentials that can drive given current flows.

## Orthogonality and Projections

---

Covers: 14. Orthogonal Vectors and Subspaces · 15. Projections onto Subspaces

The concepts of orthogonality and projections provide geometric insight into linear algebra. They are essential for understanding relationships between vectors and subspaces, and they form the basis for many applications, including least squares.

### Orthogonal Vectors and Subspaces

Two vectors  $\mathbf{x}$  and  $\mathbf{y}$  are **orthogonal** if their dot product is zero:  $\mathbf{x}^T \mathbf{y} = 0$ . Geometrically, this means they meet at a 90-degree angle. This generalizes the Pythagorean theorem: if  $\mathbf{x}$  and  $\mathbf{y}$  are orthogonal, then  $\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ .

Two subspaces  $V$  and  $W$  are **orthogonal subspaces** if every vector in  $V$  is orthogonal to every vector in  $W$ . The **orthogonal complement** of a subspace  $V$ , denoted  $V^\perp$ , is the set of all vectors orthogonal to every vector in  $V$ .

**Fundamental Theorem of Linear Algebra, Part 2:** This theorem establishes crucial orthogonal relationships between the four fundamental subspaces:

1. The null space  $N(A)$  is the orthogonal complement of the row space  $C(A^T)$  in  $\mathbb{R}^n$ . Every vector  $\mathbf{x}$  in  $N(A)$  is orthogonal to every row of  $A$  (and thus to every vector in  $C(A^T)$ ).
2. The left null space  $N(A^T)$  is the orthogonal complement of the column space  $C(A)$  in  $\mathbb{R}^m$ . Every vector  $\mathbf{y}$  in  $N(A^T)$  is orthogonal to every column of  $A$  (and thus to every vector in  $C(A)$ ).

These relationships mean that  $\mathbb{R}^n$  can be decomposed into  $C(A^T)$  and  $N(A)$ , which are orthogonal complements. Similarly,  $\mathbb{R}^m$  can be decomposed into  $C(A)$  and  $N(A^T)$ .

### Projections onto Subspaces

Projecting a vector  $\mathbf{b}$  onto a subspace  $S$  means finding the closest vector  $\mathbf{p}$  in  $S$  to  $\mathbf{b}$ . The error vector  $\mathbf{e} = \mathbf{b} - \mathbf{p}$  is orthogonal to the subspace  $S$ . This is the defining characteristic of a projection.

**Projection onto a Line:** To project a vector  $\mathbf{b}$  onto a line spanned by a vector  $\mathbf{a}$ , the projection  $\mathbf{p}$  is given by:

$$\mathbf{p} = (\mathbf{a}^T \mathbf{b} / \mathbf{a}^T \mathbf{a}) \mathbf{a}$$

The scalar  $(\mathbf{a}^T \mathbf{b} / \mathbf{a}^T \mathbf{a})$  is the length of the projection of  $\mathbf{b}$  onto  $\mathbf{a}$ , scaled by the length of  $\mathbf{a}$  squared. The error  $\mathbf{e} = \mathbf{b} - \mathbf{p}$  is orthogonal to  $\mathbf{a}$ .

**Projection onto a Subspace:** To project  $\mathbf{b}$  onto a subspace (e.g., the column space  $C(A)$  of a matrix  $A$  with linearly independent columns), we seek a vector  $\mathbf{p} = A \hat{\mathbf{x}}$  in  $C(A)$  such that the error  $\mathbf{e} = \mathbf{b} - \mathbf{p}$  is orthogonal to  $C(A)$ . This means  $\mathbf{e}$  must be in the left null space  $N(A^T)$ . So,  $A^T (\mathbf{b} - A \hat{\mathbf{x}}) = \mathbf{0}$ .

This leads to the **normal equations**:

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

Solving for  $\hat{\mathbf{x}}$  gives the coefficients for the linear combination of  $A$ 's columns that forms  $\mathbf{p}$ . The projection vector is  $\mathbf{p} = A \hat{\mathbf{x}}$ . If the columns of  $A$  are linearly independent, then  $A^T A$  is invertible,

and  $\hat{x} = (A^T A)^{-1} A^T b$ . The **projection matrix**  $P$  that projects any vector  $b$  onto  $C(A)$  is  $P = A(A^T A)^{-1} A^T$ .

### Least Squares

The normal equations are critical for **least squares approximation**. When  $Ax = b$  has no exact solution (i.e.,  $b$  is not in  $C(A)$ ), we want to find the 'best' approximate solution  $\hat{x}$ . The best approximation minimizes the squared error  $\|b - Ax\|^2$ . This minimum occurs when  $A\hat{x}$  is the projection of  $b$  onto  $C(A)$ . Thus, solving  $A^T A \hat{x} = A^T b$  gives the least squares solution  $\hat{x}$ .

#### KEY TAKEAWAYS

- The complete solution to  $Ax=b$  is  $x = x_p + x_n$ , where  $x_p$  is a particular solution and  $x_n$  is any vector from the null space.
- A basis for a vector space is a set of linearly independent vectors that span the space, and its size defines the dimension.
- Every  $m \times n$  matrix  $A$  has four fundamental subspaces: column space  $C(A)$ , null space  $N(A)$ , row space  $C(A^T)$ , and left null space  $N(A^T)$ .
- The dimensions of  $C(A)$  and  $C(A^T)$  are equal to the rank  $r$  of the matrix, while  $N(A)$  has dimension  $n - r$  and  $N(A^T)$  has dimension  $m - r$ .
- Orthogonal vectors have a zero dot product, and the four fundamental subspaces exhibit specific orthogonal relationships.
- Projecting a vector  $b$  onto a subspace  $S$  finds the closest vector  $p$  in  $S$  to  $b$ , with the error vector  $b - p$  being orthogonal to  $S$ .
- The normal equations  $A^T A \hat{x} = A^T b$  provide the least squares solution  $\hat{x}$  by projecting  $b$  onto the column space of  $A$ .

#### CHECK YOUR UNDERSTANDING

Answer these before moving on. Model answers are in the Answer Key at the back of this guide.

1. How do you find the complete solution to  $Ax = b$ ?
2. What are the conditions for a set of vectors to form a basis for a subspace?
3. Name the four fundamental subspaces of a matrix  $A$  and state their dimensions in terms of  $m$ ,  $n$ , and  $r$  (rank).
4. Explain the relationship between the null space of  $A$  and the row space of  $A$ .
5. What are the normal equations, and what problem do they solve?

## CHAPTER 3

# Orthogonality, Determinants, and Eigenvalues

---

Module 3 builds on the foundations of vector spaces and linear transformations. It introduces several powerful concepts critical for understanding advanced linear algebra and its applications. We begin with the geometry of projections and the process of constructing orthonormal bases, which are essential for solving inconsistent systems and simplifying calculations. Next, we explore the determinant, a scalar value that reveals fundamental properties of a matrix, including invertibility and geometric scaling. Finally, we introduce eigenvalues and eigenvectors, which describe the intrinsic behavior of linear transformations and are indispensable for analyzing dynamic systems and matrix powers.

## Projections, Least Squares, and Orthogonality

---

Covers: 16. Projection Matrices and Least Squares (video, 48 min) · 17. Orthogonal Matrices and Gram-Schmidt (video, 49 min)

This section combines Lessons 16 and 17, focusing on the geometry of vector spaces. We explore how to project vectors onto subspaces and how this concept leads to solutions for inconsistent systems. We also learn about special types of matrices with orthonormal columns and a systematic way to construct them.

### Projecting Vectors onto Subspaces

Often, a system of linear equations  $Ax = b$  has no exact solution. This happens when the vector  $b$  does not lie in the column space of  $A$ . In such cases, we seek the "best" approximate solution. This "best" solution involves projecting  $b$  onto the column space of  $A$ . The projection  $p$  is the closest vector in the column space of  $A$  to  $b$ . The error vector  $e = b - p$  is orthogonal to the column space of  $A$ .

To project a vector  $b$  onto a line spanned by a vector  $a$ , the projection  $p$  is given by:  $p = (a^T b / a^T a) a$

The projection matrix  $P$  for this line is:  $P = a a^T / a^T a$  Then  $p = P b$ .

For projecting a vector  $b$  onto a subspace spanned by the columns of a matrix  $A$  (assuming  $A$  has linearly independent columns), the projection  $p$  is  $A \hat{x}$ , where  $\hat{x}$  is the vector of coefficients that minimizes the error  $\|b - Ax\|^2$ . This minimization leads to the *normal equations*:  $A^T A \hat{x} = A^T b$

The solution  $\hat{x} = (A^T A)^{-1} A^T b$  gives the coefficients. The projection  $p$  is then  $A \hat{x}$ . The projection matrix  $P$  onto the column space of  $A$  is:  $P = A (A^T A)^{-1} A^T$  Then  $p = P b$ .

### Least Squares Approximation

The normal equations are the foundation of the *least squares* method. When  $Ax = b$  has no solution, we solve  $A^T A \hat{x} = A^T b$  instead. This finds the  $\hat{x}$  that makes  $A\hat{x}$  the projection  $p$  of  $b$  onto

the column space of  $A$ . The error  $b - Ax_{\text{hat}}$  is minimized in the sense of the Euclidean norm (length). This is widely used in data fitting, for example, finding the best-fit line or curve through a set of data points.

Consider fitting a straight line  $y = mx + c$  to data points  $(x_1, y_1), \dots, (x_n, y_n)$ . This leads to a system  $Xv = y$ , where  $v = [c, m]^T$ .

```
[1 x_1;
 1 x_2;
 ...;
 1 x_n] [c; m] = [y_1; y_2; ...; y_n]
```

This is typically an overdetermined system. We solve  $X^T X v_{\text{hat}} = X^T y$  for  $v_{\text{hat}} = [c_{\text{hat}}, m_{\text{hat}}]^T$ .

### Orthogonal Matrices and Gram-Schmidt

Vectors  $q_1, \dots, q_n$  are *orthogonal* if  $q_i^T q_j = 0$  for  $i \neq j$ . They are *orthonormal* if they are orthogonal and  $\|q_i\| = 1$  for all  $i$ . A square matrix  $Q$  whose columns are orthonormal vectors is called an *orthogonal matrix*. A key property of an orthogonal matrix is  $Q^T Q = I$ . This implies  $Q^T = Q^{-1}$ . Orthogonal matrices preserve lengths and angles, making them important for rotations and reflections. If  $Q$  has orthonormal columns, the projection onto its column space simplifies to  $P = Q Q^T$ . If  $Q$  is square,  $P = I$ .

The *Gram-Schmidt process* provides a systematic way to convert any basis of a subspace into an orthonormal basis. Given a set of linearly independent vectors  $a_1, \dots, a_n$ , we can construct an orthonormal set  $q_1, \dots, q_n$ .

1. Start with  $A_1 = a_1$ . Normalize:  $q_1 = A_1 / \|A_1\|$ .
2. For  $a_2$ , subtract its projection onto  $q_1$ :  $A_2 = a_2 - (q_1^T a_2) q_1$ . Normalize:  $q_2 = A_2 / \|A_2\|$ .
3. For  $a_k$ , subtract its projections onto  $q_1, \dots, q_{k-1}$ :  $A_k = a_k - (q_1^T a_k) q_1 - \dots - (q_{k-1}^T a_k) q_{k-1}$ . Normalize:  $q_k = A_k / \|A_k\|$ . This process yields an orthonormal basis  $q_1, \dots, q_n$ . This can be expressed as an  $A = QR$  factorization, where  $Q$  has orthonormal columns and  $R$  is an upper triangular matrix. The  $QR$  factorization is numerically stable and useful for solving least squares problems.

Let's apply Gram-Schmidt to  $a_1 = [1; 1; 0]$  and  $a_2 = [1; 0; 1]$ .

1.  $A_1 = [1; 1; 0]$ .  $\|A_1\| = \sqrt{1^2 + 1^2 + 0^2} = \sqrt{2}$ .  $q_1 = [1/\sqrt{2}; 1/\sqrt{2}; 0]$ .
2.  $q_1^T a_2 = (1/\sqrt{2}) * 1 + (1/\sqrt{2}) * 0 + 0 * 1 = 1/\sqrt{2}$ .  $A_2 = a_2 - (q_1^T a_2) q_1 = [1; 0; 1] - (1/\sqrt{2}) [1/\sqrt{2}; 1/\sqrt{2}; 0] = [1; 0; 1] - [1/2; 1/2; 0] = [1/2; -1/2; 1]$ .  $\|A_2\| = \sqrt{(1/2)^2 + (-1/2)^2 + 1^2} = \sqrt{1/4 + 1/4 + 1} = \sqrt{3/2}$ .  $q_2 = [1/\sqrt{6}; -1/\sqrt{6}; 2/\sqrt{6}]$ . The orthonormal basis is  $q_1$  and  $q_2$ .

A common mistake in Gram-Schmidt is forgetting to normalize each intermediate vector  $A_k$  to get  $q_k$  before using it in the next step. Another error is incorrectly calculating the projection component to subtract. Always verify orthogonality  $q_i^T q_j = 0$  and unit length  $\|q_i\| = 1$  for your final vectors.

## Determinants and Their Applications

Covers: 18. Properties of Determinants (video, 49 min) · 19. Determinant Formulas and Cofactors (video, 53 min) · 20. Cramer's Rule, Inverse Matrix, and Volume (video, 51 min)

This section covers Lessons 18, 19, and 20, focusing on the determinant as a fundamental scalar property of square matrices. We learn its definition, how to compute it, and its wide-ranging applications, from checking invertibility to calculating volumes.

### Properties of the Determinant

The determinant, denoted  $\det(A)$  or  $|A|$ , is a single number associated with every square matrix  $A$ . It reveals crucial information about the matrix. We define the determinant through three fundamental properties:

1.  $\det(I) = 1$  for the identity matrix  $I$ .
2. If two rows of  $A$  are exchanged, the determinant changes sign.
3. The determinant is linear for each row separately. This means  $\det([a+a'; b]) = \det([a; b]) + \det([a'; b])$  and  $\det([ka; b]) = k \det([a; b])$ .

From these three properties, many other important properties can be derived:

- If two rows of  $A$  are identical,  $\det(A) = 0$ .
- Subtracting a multiple of one row from another row does not change the determinant. This is a key operation in Gaussian elimination.
- If  $A$  has a row of zeros,  $\det(A) = 0$ .
- $\det(A) = 0$  if and only if  $A$  is a singular matrix (not invertible). This is one of the most important uses of the determinant.
- $\det(AB) = \det(A) \det(B)$ .
- $\det(A^T) = \det(A)$ . This implies that all row properties also apply to columns.
- $\det(A^{-1}) = 1 / \det(A)$ .
- $\det(\lambda A) = \lambda^n \det(A)$  for an  $n \times n$  matrix  $A$ .

For a  $2 \times 2$  matrix  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ ,  $\det(A) = ad - bc$ . For a  $3 \times 3$  matrix  $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ ,  $\det(A) = a(ei - fh) - b(di - fg) + c(dh - eg)$ . Notice the alternating signs.

### Determinant Formulas and Cofactors

While the  $3 \times 3$  formula can be extended, a general "big formula" for an  $n \times n$  determinant involves summing  $n!$  terms, each being a product of  $n$  entries, one from each row and column, multiplied by a sign determined by the permutation. This is computationally intensive for large  $n$ .

A more practical method for calculation is the *cofactor expansion*. The cofactor  $C_{ij}$  of an entry  $a_{ij}$  is  $(-1)^{(i+j)}$  times the determinant of the submatrix  $M_{ij}$  obtained by deleting row  $i$  and column  $j$  from  $A$ . This  $M_{ij}$  is called the *minor*. The determinant of  $A$  can be computed by expanding along any row  $i$ :  $\det(A) = a_{i1} C_{i1} + a_{i2} C_{i2} + \dots + a_{in} C_{in}$ . Or along any column  $j$ :  $\det(A) = a_{1j} C_{1j} + a_{2j} C_{2j} + \dots + a_{nj} C_{nj}$ . Choosing a row or column with many zeros simplifies the calculation considerably.

Example: Calculate the determinant of  $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$  using cofactor expansion along the first column.  $\det(A) = 1 * C_{11} + 0 * C_{21} + 0 * C_{31}$ .  $C_{11} = (-1)^{(1+1)} * \det(\begin{bmatrix} 4 & 5 \\ 0 & 6 \end{bmatrix}) = 1 * (4*6 - 5*0) = 24$ . So,  $\det(A) = 1 * 24 = 24$ . Notice that for a triangular matrix, the determinant is the product of its diagonal entries.

### Cramer's Rule, Inverse Matrix, and Volume

The determinant is also used in deriving formulas for the inverse matrix and for solving linear systems. The inverse of  $A$  can be expressed using the *adjoint matrix* (also called the adjugate matrix), which is the transpose of the matrix of cofactors  $C$ .  $A^{-1} = (1 / \det(A)) * C^T$ . This formula is theoretically important but computationally inefficient for large matrices compared to Gaussian elimination.

*Cramer's Rule* provides a formula for the solution  $x$  of  $Ax = b$  using determinants:  $x_i = \det(A_i) / \det(A)$ . Here,  $A_i$  is the matrix  $A$  with its  $i$ -th column replaced by the vector  $b$ . Cramer's Rule is also not efficient for large systems but is useful for small systems or theoretical analysis.

Geometrically, the absolute value of the determinant  $|\det(A)|$  represents the volume of the parallelepiped (or parallelogram in 2D) formed by the column vectors of  $A$ . If  $\det(A)$  is positive, the orientation of the space is preserved; if negative, it is reversed. For  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ ,  $|\det(A)| = |ad - bc|$  is the area of the parallelogram formed by  $\begin{bmatrix} a \\ c \end{bmatrix}$  and  $\begin{bmatrix} b \\ d \end{bmatrix}$ . When a linear transformation  $T(x) = Ax$  is applied to a region  $R$  in  $n$ -dimensional space, the volume of the transformed region  $T(R)$  is  $|\det(A)|$  times the volume of  $R$ . This property highlights the determinant's role as a scaling factor for volumes.

A common mistake when calculating determinants is sign errors in cofactor expansion, especially with the  $(-1)^{(i+j)}$  term. Another is forgetting that Cramer's Rule and the inverse matrix formula require  $\det(A) \neq 0$ . Practice calculating determinants of  $3 \times 3$  and  $4 \times 4$  matrices using cofactor expansion along different rows or columns, particularly those with zeros.

## Eigenvalues, Eigenvectors, and Diagonalization

Covers: 21. Eigenvalues and Eigenvectors (video, 51 min) · 22. Diagonalization and Powers of  $A$  (video, 52 min) · 23. Differential Equations and  $\exp(At)$  (video, 51 min)

This section, covering Lessons 21, 22, and 23, introduces eigenvalues and eigenvectors, which are fundamental to understanding the intrinsic behavior of linear transformations. We explore how to find them, how they enable matrix diagonalization, and their powerful applications in solving systems of differential equations.

## Understanding Eigenvalues and Eigenvectors

For a square matrix  $A$ , an *eigenvector*  $x$  is a non-zero vector that, when multiplied by  $A$ , results in a scalar multiple of itself. The scalar  $\lambda$  is called the *eigenvalue*. This relationship is expressed by the equation:  $Ax = \lambda x$

Eigenvectors represent directions that are unchanged by the linear transformation  $A$ , only scaled by the eigenvalue  $\lambda$ . To find eigenvalues, we rearrange the equation:  $Ax - \lambda x = 0$  ( $A - \lambda I$ )  $x = 0$ . For a non-zero eigenvector  $x$  to exist, the matrix  $(A - \lambda I)$  must be singular (not invertible). This means its determinant must be zero:  $\det(A - \lambda I) = 0$ . This is the *characteristic equation*. Solving this polynomial equation for  $\lambda$  yields the eigenvalues. For an  $n \times n$  matrix, there will be  $n$  eigenvalues (counting multiplicity, possibly complex).

Once an eigenvalue  $\lambda$  is found, we find its corresponding eigenvectors by solving the homogeneous system  $(A - \lambda I)x = 0$ . The eigenvectors form the null space of  $(A - \lambda I)$ . This null space is called the *eigenspace* for  $\lambda$ .

Example: Find eigenvalues and eigenvectors for  $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ .

1. Characteristic equation:  $\det(A - \lambda I) = \det(\begin{bmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{bmatrix}) = 0$ .  $(3 - \lambda)(3 - \lambda) - 1 \cdot 1 = 0$   $9 - 6\lambda + \lambda^2 - 1 = 0$   $\lambda^2 - 6\lambda + 8 = 0$   $(\lambda - 2)(\lambda - 4) = 0$  The eigenvalues are  $\lambda_1 = 2$  and  $\lambda_2 = 4$ .
2. For  $\lambda_1 = 2$ :  $(A - 2I)x = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} x = 0$ . This implies  $x_1 + x_2 = 0$ , so  $x_2 = -x_1$ . An eigenvector is  $x_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$  (or any non-zero multiple).
3. For  $\lambda_2 = 4$ :  $(A - 4I)x = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} x = 0$ . This implies  $-x_1 + x_2 = 0$ , so  $x_2 = x_1$ . An eigenvector is  $x_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$  (or any non-zero multiple).

Properties: The sum of the eigenvalues equals the trace of  $A$  (sum of diagonal entries). The product of the eigenvalues equals  $\det(A)$ . These properties can help check your calculations.

## Diagonalization and Powers of $A$

A matrix  $A$  is *diagonalizable* if it can be written as  $A = S \Lambda S^{-1}$ , where  $\Lambda$  is a diagonal matrix containing the eigenvalues of  $A$  on its main diagonal, and  $S$  is a matrix whose columns are the corresponding linearly independent eigenvectors of  $A$ . For  $A$  to be diagonalizable, it must have a full set of  $n$  linearly independent eigenvectors. If all  $n$  eigenvalues are distinct, then the corresponding eigenvectors are guaranteed to be linearly independent, and  $A$  is diagonalizable. If eigenvalues are repeated,  $A$  might or might not have  $n$  independent eigenvectors.

Diagonalization is incredibly useful for computing powers of a matrix  $A$ .  $A^2 = (S \Lambda S^{-1})(S \Lambda S^{-1}) = S \Lambda (S^{-1} S) \Lambda S^{-1} = S \Lambda I \Lambda S^{-1} = S \Lambda^2 S^{-1}$ . In general,  $A^k = S \Lambda^k S^{-1}$ . Calculating  $\Lambda^k$  is straightforward because  $\Lambda$  is diagonal: simply raise each diagonal entry (eigenvalue) to the power  $k$ .

$$\Lambda^k = \begin{bmatrix} \lambda_1^k & 0 & \dots & 0 \\ 0 & \lambda_2^k & \dots & 0 \end{bmatrix}$$

```
...;
0 0 ... lambda_n^k]
```

This makes computing  $A^k$  much faster than repeated matrix multiplication, especially for large  $k$ .

For the example matrix  $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ , we have  $\lambda_1 = 2$ ,  $x_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$  and  $\lambda_2 = 4$ ,  $x_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ .  $\Lambda = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$   $S = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$   $S^{-1} = (1/2) \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$  Then  $A = S \Lambda S^{-1}$ . To find  $A^{10}$ , we calculate  $\Lambda^{10} = \begin{bmatrix} 2^{10} & 0 \\ 0 & 4^{10} \end{bmatrix}$ .  $A^{10} = S \Lambda^{10} S^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2^{10} & 0 \\ 0 & 4^{10} \end{bmatrix} (1/2) \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ .

### Differential Equations and $\exp(At)$

Eigenvalues and eigenvectors are crucial for solving systems of linear first-order differential equations of the form  $du/dt = Au$ , where  $u(t)$  is a vector function of time. The general solution for  $du/dt = Au$  is a linear combination of exponential terms:  $u(t) = c_1 e^{\lambda_1 t} x_1 + c_2 e^{\lambda_2 t} x_2 + \dots + c_n e^{\lambda_n t} x_n$ . Here,  $\lambda_i$  are the eigenvalues and  $x_i$  are the corresponding eigenvectors of  $A$ . The constants  $c_i$  are determined by the initial conditions  $u(0)$ . If  $u(0)$  is given, we can write  $u(0) = c_1 x_1 + \dots + c_n x_n$ , which means  $u(0) = S c$ , where  $c = [c_1; \dots; c_n]$ . Thus,  $c = S^{-1} u(0)$ .

This solution can also be expressed using the matrix exponential  $e^{At}$ . The matrix exponential is defined by the infinite series  $e^{At} = I + At + (At)^2/2! + \dots$ . If  $A$  is diagonalizable,  $A = S \Lambda S^{-1}$ , then  $e^{At} = S e^{\Lambda t} S^{-1}$ .  $e^{\Lambda t}$  is a diagonal matrix:

```
e^(Lambda t) = [e^(lambda_1 t) 0 ... 0;
0 e^(lambda_2 t) ... 0;
...;
0 0 ... e^(lambda_n t)]
```

The solution to  $du/dt = Au$  with initial condition  $u(0)$  is  $u(t) = e^{At} u(0)$ . This provides a powerful and elegant way to solve such systems. The stability of the system depends on the real parts of the eigenvalues. If all real parts are negative, the system is stable and  $u(t)$  approaches zero as  $t$  goes to infinity.

A common mistake is to confuse the characteristic equation  $\det(A - \lambda I) = 0$  with solving for eigenvectors  $(A - \lambda I) x = 0$ . Remember to find all eigenvalues first, then for each eigenvalue, solve the null space problem to find its eigenvectors. When diagonalizing, ensure  $S$  contains linearly independent eigenvectors. If  $A$  does not have  $n$  independent eigenvectors, it is not diagonalizable.

**KEY TAKEAWAYS**

- Projection matrices allow us to find the closest vector in a subspace to a given vector, forming the basis for least squares solutions.
- The Gram-Schmidt process systematically transforms any basis into an orthonormal basis, simplifying many calculations involving projections and transformations.
- The determinant of a square matrix indicates its invertibility and scales volumes under linear transformations.
- Cofactor expansion and Cramer's Rule provide methods for computing determinants and solving linear systems, respectively, though they are often computationally intensive for large matrices.
- Eigenvalues and eigenvectors reveal the directions in which a linear transformation acts merely by scaling, providing insight into a matrix's fundamental behavior.
- Diagonalization simplifies matrix powers and is crucial for solving systems of linear differential equations by transforming the problem into a simpler, uncoupled form.

**CHECK YOUR UNDERSTANDING**

*Answer these before moving on. Model answers are in the Answer Key at the back of this guide.*

1. How does the concept of projection relate to finding the least squares solution to an inconsistent system  $Ax = b$ ?
2. What are the key properties of an orthogonal matrix  $Q$ , and why are they useful in linear algebra?
3. Explain how the determinant of a matrix  $A$  reveals information about its invertibility and the geometric effect of the transformation  $A$ .
4. What is the characteristic equation, and how is it used to find the eigenvalues of a matrix?
5. Under what conditions can a matrix be diagonalized, and how does diagonalization simplify the computation of  $A^k$ ?

## CHAPTER 4

# Advanced Matrix Structures and Decompositions

---

This chapter explores advanced topics in linear algebra, building upon your understanding of vectors, matrices, and eigenvalues. We examine special classes of matrices like Markov, symmetric, and positive definite matrices, uncovering their unique properties and applications. You will learn how complex numbers extend linear algebra concepts and discover the power of Fourier analysis. Finally, we work through sophisticated matrix decompositions, including the Jordan form and the Singular Value Decomposition, which provide deep insights into matrix behavior and enable powerful data analysis techniques.

This material forms a crucial bridge to higher-level mathematics, machine learning, and engineering disciplines. Mastering these concepts will solidify your foundation in linear algebra and prepare you for tackling complex real-world problems.

## Special Matrices and Their Properties

---

Covers: 24. Markov Matrices; Fourier Series · 24b. Quiz 2 Review · 25. Symmetric Matrices and Positive Definiteness · 27. Positive Definite Matrices and Minima

Linear algebra becomes particularly insightful when we study matrices with specific structures. These special matrices often arise in practical applications and possess unique properties that simplify analysis. We begin with Markov matrices, which model systems evolving over time.

### Markov Matrices and Steady States

A **Markov matrix** (also called a stochastic matrix) is a square matrix whose entries are non-negative, and each column sums to 1. These matrices are fundamental in probability and statistics, describing transitions between states in a system. For example, they can model population shifts between regions or the probability of a user moving between web pages.

A key property of Markov matrices is that they always have an eigenvalue of 1. The corresponding eigenvector, when scaled so its entries sum to 1, represents the **steady state** or equilibrium distribution of the system. This means that after many time steps, the system converges to this stable distribution, regardless of the initial state.

Consider a simple example: A population moves between two cities, City A and City B. Each year, 70% of City A residents stay in City A, and 30% move to City B. From City B, 80% stay in City B, and 20% move to City A. The Markov matrix  $P$  is:

$$P = \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix}$$

To find the steady state, we solve  $Px = x$ , which is  $(P - I)x = 0$ .

$$P - I = \begin{bmatrix} -0.3 & 0.2 \\ 0.3 & -0.2 \end{bmatrix}$$

The null space of  $(P - I)$  gives us the eigenvectors for  $\lambda = 1$ . A vector like  $x = [2; 3]$  satisfies this. Scaling it so its entries sum to 1, we get  $[2/5; 3/5]$  or  $[0.4; 0.6]$ . This means in the long run, 40% of the population will be in City A and 60% in City B.

### Symmetric Matrices and Their Spectral Properties

A matrix  $A$  is **symmetric** if it equals its transpose,  $A = A^T$ . These matrices are ubiquitous in physics, engineering, and optimization. Their eigenvalues are always real numbers, and their eigenvectors corresponding to distinct eigenvalues are orthogonal. If eigenvalues are repeated, we can always find an orthogonal basis of eigenvectors. This leads to the **Spectral Theorem** for symmetric matrices: any symmetric matrix  $A$  can be diagonalized as  $A = Q\Lambda Q^T$ , where  $Q$  is an orthogonal matrix (its columns are orthonormal eigenvectors) and  $\Lambda$  is a diagonal matrix of eigenvalues. This decomposition is incredibly powerful, allowing us to understand a symmetric matrix by its principal directions (eigenvectors) and scaling factors (eigenvalues).

### Positive Definite Matrices and Minima

A special class of symmetric matrices are **positive definite matrices**. A symmetric matrix  $A$  is positive definite if for any non-zero vector  $x$ , the quadratic form  $x^T A x > 0$ . If  $x^T A x \geq 0$ , it is positive semi-definite. Geometrically, the quadratic form  $x^T A x = \text{constant}$  describes an ellipse or ellipsoid in higher dimensions.

Positive definite matrices have several equivalent characterizations:

1. All eigenvalues of  $A$  are positive.
2. All pivots (without row exchanges) in Gaussian elimination are positive.
3. All upper-left subdeterminants (principal minors) are positive.

These tests are practical. For a  $2 \times 2$  matrix,  $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ , it is positive definite if  $a > 0$  and  $ac - b^2 > 0$ . Example: Check if  $A = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$  is positive definite.

1. Pivot test: The first pivot is 2 (positive). Eliminate 1 below it:  $R_2 \rightarrow R_2 - (1/2)R_1$ .

$$\begin{bmatrix} 2 & 1 \\ 0 & 2.5 \end{bmatrix}$$

The second pivot is 2.5 (positive). Both pivots are positive, so  $A$  is positive definite. Alternatively, check determinants:  $\det(2) = 2 > 0$ .  $\det(A) = 2 \cdot 3 - 1 \cdot 1 = 5 > 0$ .

Positive definite matrices are crucial in optimization. If a function's second derivative matrix (Hessian) is positive definite at a critical point, that point is a local minimum. This connects linear algebra directly to multivariable calculus and optimization problems.

**Practice:** For a given symmetric matrix, calculate its eigenvalues and verify they are real. Try to construct an orthogonal matrix  $Q$  from its eigenvectors. For a given matrix, apply the pivot test and the determinant test to determine if it is positive definite.

**Common mistake:** Assuming a non-symmetric matrix can be positive definite. The definition requires symmetry. Also, confusing positive definite with positive semi-definite. Positive definite requires *strict* positivity.

## Complex Numbers, Fourier Analysis, and Matrix Transformations

Covers: 26. Complex Matrices; Fast Fourier Transform · 30. Linear Transformations and Their Matrices

While much of introductory linear algebra focuses on real numbers, many important applications and theoretical constructs require the use of complex numbers. This section extends our understanding of matrices and vectors into the complex domain and connects it to powerful analytical tools like Fourier analysis and the fundamental concept of linear transformations.

### Complex Matrices and Their Analogues

When dealing with complex vectors, the standard dot product  $x^T y$  is replaced by the **Hermitian dot product**,  $x^H y$ , where  $x^H$  is the conjugate transpose of  $x$  (transpose and then conjugate each entry). This ensures that  $x^H x$  is always a non-negative real number, representing the squared length of the vector.

With complex numbers, the concept of a symmetric matrix extends to a **Hermitian matrix**, where  $A = A^H$ . Hermitian matrices share many properties with real symmetric matrices: their eigenvalues are always real, and their eigenvectors corresponding to distinct eigenvalues are orthogonal (in the Hermitian sense). This is a vital property for quantum mechanics, where Hermitian operators represent observable quantities.

Similarly, the concept of an orthogonal matrix ( $Q^T Q = I$ ) extends to a **unitary matrix**  $U$ , where  $U^H U = I$ . Unitary matrices preserve the Hermitian dot product and thus the lengths of complex vectors. Their eigenvalues always lie on the unit circle in the complex plane.

### Fourier Series and the Fast Fourier Transform

The idea of decomposing a function into a sum of simpler, orthogonal components is central to **Fourier analysis**. A **Fourier series** represents a periodic function as a sum of sines and cosines, or more compactly, as a sum of complex exponentials. These complex exponentials,  $e^{ikx}$ , form an orthogonal basis for function spaces.

For discrete data, we use the **Discrete Fourier Transform (DFT)**, which transforms a sequence of  $N$  complex numbers into another sequence of  $N$  complex numbers, representing the frequencies present in the original data. The DFT can be expressed as a matrix multiplication using the **Fourier matrix**  $F$ , where  $F_{jk} = e^{-2\pi i jk/N}$ . This matrix is symmetric and unitary (up to a scaling factor).

Computing the DFT directly takes  $O(N^2)$  operations, which can be slow for large  $N$ . The **Fast Fourier Transform (FFT)** is an incredibly efficient algorithm for computing the DFT, reducing the

computational complexity to  $O(N \log N)$ . This breakthrough, attributed to Cooley and Tukey, transformed signal processing, image compression, data analysis, and many other fields. The FFT works by recursively breaking down the DFT into smaller DFTs, exploiting the periodic nature of the complex exponentials.

### Linear Transformations and Their Matrices

A **linear transformation**  $T$  is a function that maps vectors from one vector space to another, preserving vector addition and scalar multiplication. That is,  $T(u + v) = T(u) + T(v)$  and  $T(cv) = cT(v)$ . Examples include rotations, reflections, projections, and scaling.

Every linear transformation from  $\mathbb{R}^n$  to  $\mathbb{R}^m$  can be represented by an  $m \times n$  matrix  $A$ . If  $T(x)$  is the transformation, then  $T(x) = Ax$ . The columns of  $A$  are the transformations of the standard basis vectors. For example, if  $T$  is a rotation by angle  $\theta$  in  $\mathbb{R}^2$ :

$$T\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos(\theta)x - \sin(\theta)y \\ \sin(\theta)x + \cos(\theta)y \end{bmatrix}$$

The matrix for this transformation is:

$$A = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

Understanding linear transformations provides a geometric interpretation for matrix operations. Changing the basis of the vector space changes the matrix representation of the transformation, but the underlying transformation remains the same. This concept is crucial for understanding similarity transformations, which we will explore next.

**Practice:** Given a complex vector, calculate its Hermitian dot product with another complex vector. Identify if a given complex matrix is Hermitian or unitary. For a simple  $2 \times 2$  rotation matrix, apply it to a vector and sketch the result. Consider how different bases would change the matrix representation of a simple transformation like projection onto an axis.

**Common mistake:** Forgetting to conjugate entries when calculating the Hermitian transpose or dot product. Not recognizing that a linear transformation must preserve both addition and scalar multiplication.

## Advanced Matrix Decompositions

Covers: 28. Similar Matrices and Jordan Form · 29. Singular Value Decomposition

Matrix decompositions are powerful tools for understanding and simplifying matrices. We have already seen diagonalization ( $A = SAS^{-1}$ ) for matrices with a full set of eigenvectors. This section introduces two more advanced and general decompositions: the Jordan form, which handles cases where diagonalization fails, and the Singular Value Decomposition (SVD), which applies to *any* matrix.

### Similar Matrices and the Jordan Form

Two square matrices  $A$  and  $B$  are **similar** if there exists an invertible matrix  $M$  such that  $B = M^{-1}AM$ . Similar matrices represent the same linear transformation, but with respect to different bases.

Importantly, similar matrices share the same eigenvalues, determinant, trace, and rank. Diagonalization is a special case of similarity, where  $A$  is similar to a diagonal matrix  $\Lambda$ . This is possible if and only if  $A$  has a full set of linearly independent eigenvectors.

When a matrix cannot be diagonalized, it means it does not have enough linearly independent eigenvectors to form a basis. In such cases, the matrix can still be transformed into its **Jordan canonical form** (or Jordan form). The Jordan form  $J$  is the "closest" a non-diagonalizable matrix can get to being diagonal. It is a block diagonal matrix, where each block, called a **Jordan block**, has an eigenvalue  $\lambda$  on the main diagonal and 1s just above the diagonal. For example, a  $3 \times 3$  Jordan block for eigenvalue  $\lambda$  is:

$$J_\lambda = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}$$

Every square matrix  $A$  is similar to a Jordan matrix  $J$  ( $A = M J M^{-1}$ ). The Jordan form is unique up to the ordering of the Jordan blocks. While theoretically significant for understanding the behavior of linear systems, computing the Jordan form can be numerically unstable and is less commonly used in practical applications than other decompositions like SVD.

### Singular Value Decomposition (SVD)

The **Singular Value Decomposition (SVD)** is arguably the most fundamental and widely applicable matrix factorization. It applies to *any*  $m \times n$  matrix  $A$ , regardless of whether it is square, invertible, or even real. The SVD states that  $A$  can be factored into the form  $A = U \Sigma V^T$ .

- $U$  is an  $m \times m$  orthogonal matrix whose columns are the **left singular vectors** of  $A$ .
- $\Sigma$  (Sigma) is an  $m \times n$  diagonal matrix containing the **singular values** of  $A$ , denoted  $\sigma_1, \sigma_2, \dots, \sigma_r$ , where  $r$  is the rank of  $A$ . These singular values are always non-negative and are conventionally ordered from largest to smallest. The non-diagonal entries of  $\Sigma$  are zero.
- $V$  is an  $n \times n$  orthogonal matrix whose columns are the **right singular vectors** of  $A$ .

The singular values are the square roots of the non-zero eigenvalues of  $A^T A$  (or  $A A^T$ ). The columns of  $V$  are the eigenvectors of  $A^T A$ , and the columns of  $U$  are the eigenvectors of  $A A^T$ .

Geometrically, the SVD reveals how a linear transformation represented by  $A$  maps a unit sphere in  $\mathbb{R}^n$  to an ellipsoid in  $\mathbb{R}^m$ . The right singular vectors  $v_j$  form an orthonormal basis for the domain, the left singular vectors  $u_j$  form an orthonormal basis for the codomain, and the singular values  $\sigma_j$  are the lengths of the semi-axes of the ellipsoid.

The SVD has an astonishing array of applications. For instance, in **image compression**, a grayscale image matrix  $A$  can be approximated by keeping only the largest singular values and their corresponding vectors:  $A_k = U_k \Sigma_k V_k^T$ . A small  $k$  often yields a recognizable image, demonstrating SVD's power in capturing essential information. Other uses include Principal Component Analysis (PCA) for dimensionality reduction and computing the pseudo-inverse for solving least squares problems.

**Practice:** Given a matrix, determine if it is diagonalizable. If not, consider what its Jordan form might look like. Understand the components of SVD ( $U, \Sigma, V^T$ ) and their roles. Think about how SVD could be

applied to a dataset you are familiar with to reduce its dimensionality or extract key features.

Common mistake: Confusing eigenvalues with singular values. Eigenvalues are for square matrices and can be complex. Singular values are for any matrix and are always real and non-negative. Also, confusing the columns of  $U$  and  $V$  as eigenvectors of  $A$  itself; they are eigenvectors of  $A A^T$  and  $A^T A$ , respectively.

### KEY TAKEAWAYS

- Markov matrices model state transitions and have a steady state eigenvector corresponding to eigenvalue 1.
- Symmetric matrices have real eigenvalues and orthogonal eigenvectors, allowing diagonalization into  $Q\Lambda Q^T$ .
- Positive definite matrices are symmetric matrices where  $x^T A x > 0$  for all non-zero  $x$ , characterized by positive eigenvalues, pivots, or subdeterminants.
- Complex matrices introduce Hermitian (symmetric analogue) and unitary (orthogonal analogue) matrices, which are crucial for advanced applications like quantum mechanics.
- The Fast Fourier Transform (FFT) efficiently computes the Discrete Fourier Transform, transforming data between time and frequency domains.
- Linear transformations are functions preserving vector addition and scalar multiplication, always representable by a matrix.
- The Jordan form describes matrices that cannot be diagonalized, providing a canonical form with Jordan blocks.
- The Singular Value Decomposition (SVD) factors any matrix  $A$  into  $U \Sigma V^T$ , revealing its fundamental structure and enabling applications like image compression and PCA.

### CHECK YOUR UNDERSTANDING

Answer these before moving on. Model answers are in the Answer Key at the back of this guide.

1. What are the defining characteristics of a Markov matrix, and what does its steady state eigenvector represent?
2. How can you test if a symmetric matrix is positive definite? Name at least two methods.
3. Explain the significance of the Fast Fourier Transform (FFT) in practical applications.
4. What is the primary difference between similar matrices and matrices that are simply row equivalent?
5. Describe the components of the Singular Value Decomposition (SVD) for a matrix  $A$ , and state one key application.
6. When is a square matrix guaranteed to be diagonalizable, and what alternative decomposition exists if it is not?

## CHAPTER 5

# Basis Transformations, Generalized Inverses, and Course Synthesis

This chapter, covering Module 5 of MIT 18.06, brings together advanced concepts and provides a comprehensive review of linear algebra. We begin by exploring how to represent vectors and linear transformations in different coordinate systems through the process of changing basis, a fundamental idea with applications in data compression. We then extend the concept of matrix inverses to non-square matrices, introducing left inverses, right inverses, and the powerful pseudoinverse, which is crucial for solving overdetermined or underdetermined systems and finding optimal least squares solutions. Finally, this chapter guides you through a strategic review of the entire course, emphasizing the deep interconnections between core topics like subspaces, orthogonality, eigenvalues, and the Singular Value Decomposition. Mastering these final concepts solidifies your understanding of linear algebra's breadth and utility.

## Change of Basis and Its Applications

Covers: 31. Change of Basis; Image Compression

A vector's coordinates depend entirely on the chosen basis. A basis is a set of linearly independent vectors that span a vector space. For example, in  $\mathbb{R}^2$ , the standard basis is  $e_1 = [1, 0]$  and  $e_2 = [0, 1]$ . Any vector  $v = [x, y]$  can be written as  $x \cdot e_1 + y \cdot e_2$ . If we choose a different basis, say  $b_1 = [1, 1]$  and  $b_2 = [-1, 1]$ , the same vector  $v$  will have different coordinates. The process of finding these new coordinates or transforming between coordinate systems is called a change of basis.

Consider a vector  $v$  in a vector space  $V$ . Let  $B = \{b_1, b_2, \dots, b_n\}$  be an ordered basis for  $V$ , and  $C = \{c_1, c_2, \dots, c_n\}$  be another ordered basis for  $V$ . If  $[v]_B$  represents the coordinates of  $v$  in basis  $B$ , and  $[v]_C$  represents the coordinates of  $v$  in basis  $C$ , we want to find a matrix  $P$  such that  $[v]_C = P * [v]_B$ . This matrix  $P$  is called the change of basis matrix from  $B$  to  $C$ .

To construct  $P$ , we express each vector in basis  $B$  as a linear combination of vectors in basis  $C$ . Specifically,  $b_1 = p_{11} \cdot c_1 + p_{21} \cdot c_2 + \dots + p_{n1} \cdot c_n$ ,  $b_2 = p_{12} \cdot c_1 + p_{22} \cdot c_2 + \dots + p_{n2} \cdot c_n$ , and so on. The coordinates of  $b_1$  with respect to  $C$  form the first column of  $P$ ,  $[b_1]_C$ . Similarly,  $[b_2]_C$  forms the second column, and so forth. So,  $P = [[b_1]_C \mid [b_2]_C \mid \dots \mid [b_n]_C]$ . This matrix  $P$  transforms coordinates from basis  $B$  to basis  $C$ . A more practical way to compute  $P$  is to form matrices  $B_{\text{matrix}} = [b_1 \mid b_2 \mid \dots \mid b_n]$  and  $C_{\text{matrix}} = [c_1 \mid c_2 \mid \dots \mid c_n]$ , where the basis vectors are written as columns. Then  $B_{\text{matrix}} * [v]_B = v$  (the actual vector), and  $C_{\text{matrix}} * [v]_C = v$ . Therefore,  $C_{\text{matrix}} * [v]_C = B_{\text{matrix}} * [v]_B$ . To find  $[v]_C$ , we multiply by the inverse of  $C_{\text{matrix}}$ :  $[v]_C = C_{\text{matrix}}^{-1} * B_{\text{matrix}} * [v]_B$ . So, the change of basis matrix  $P$  from  $B$  to  $C$  is  $C_{\text{matrix}}^{-1} * B_{\text{matrix}}$ . Conversely, the change of basis matrix from  $C$  to  $B$  is  $B_{\text{matrix}}^{-1} * C_{\text{matrix}}$ , which is  $P^{-1}$ .

**Example: Change of Basis in  $\mathbb{R}^2$** 

Let the standard basis be  $E = \{e_1, e_2\}$  where  $e_1 = [1, 0]$  and  $e_2 = [0, 1]$ . Let a new basis be  $B = \{b_1, b_2\}$  where  $b_1 = [1, 1]$  and  $b_2 = [-1, 1]$ . Consider a vector  $v = [3, 1]$  in standard coordinates, so  $[v]_E = [3, 1]$ . We want to find  $[v]_B$ . The matrix  $B_{\text{matrix}} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ . The matrix  $E_{\text{matrix}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ . The change of basis matrix from  $E$  to  $B$  is  $P_{BE} = B_{\text{matrix}}^{-1} * E_{\text{matrix}} = B_{\text{matrix}}^{-1}$ . First, find  $B_{\text{matrix}}^{-1}$ :  $\det(B_{\text{matrix}}) = (1)(1) - (-1)(1) = 1 + 1 = 2$ .  $B_{\text{matrix}}^{-1} = (1/2) * \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 \\ -0.5 & 0.5 \end{bmatrix}$ . Now,  $[v]_B = P_{BE} * [v]_E = \begin{bmatrix} 0.5 & 0.5 \\ -0.5 & 0.5 \end{bmatrix} * \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5*3 + 0.5*1 \\ -0.5*3 + 0.5*1 \end{bmatrix} = \begin{bmatrix} 1.5 + 0.5 \\ -1.5 + 0.5 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ . So,  $v = 2*b_1 - 1*b_2 = 2*[1, 1] - 1*[-1, 1] = [2, 2] - [-1, 1] = [3, 1]$ , which matches the original vector.

**Image Compression**

Change of basis is fundamental to image and signal compression. The idea is to transform the data (image pixels, audio samples) from a standard basis into a new basis where the information is represented more efficiently. Often, this new basis (like a Fourier basis or wavelet basis) concentrates most of the "energy" or important information into a few coefficients, while many other coefficients are close to zero. By discarding these small coefficients, we can compress the data with minimal loss of perceived quality. For example, the Discrete Cosine Transform (DCT), used in JPEG compression, transforms image blocks into a frequency domain representation. Most image information resides in low-frequency components, allowing high-frequency coefficients to be truncated or quantized, leading to significant compression. This is an application of finding a basis that provides a sparse representation of the data.

Common mistake: Mixing up the order of the bases when constructing the change of basis matrix.

Remember  $P_{CB} = C_{\text{matrix}}^{-1} * B_{\text{matrix}}$  transforms coordinates from  $B$  to  $C$ . The columns of  $P_{CB}$  are the  $B$  basis vectors expressed in  $C$  coordinates.

Practice cue: Try changing the basis for a 3D vector from the standard basis to a given orthonormal basis. Verify your result by transforming back.

**Left and Right Inverses; Pseudoinverse**

Covers: 33. Left and Right Inverses; Pseudoinverse

For a square matrix  $A$ , its inverse  $A^{-1}$  satisfies  $A * A^{-1} = A^{-1} * A = I$ , where  $I$  is the identity matrix. However, not all square matrices have an inverse (singular matrices), and rectangular matrices never have a two-sided inverse in the traditional sense. For rectangular matrices, we introduce the concepts of left and right inverses.

**Left Inverse**

A matrix  $A$  has a left inverse  $A_L$  if  $A_L * A = I$ . For  $A$  to have a left inverse, it must have full column rank. This means its columns are linearly independent. If  $A$  is an  $m \times n$  matrix with  $m \geq n$  (tall or square), and its columns are linearly independent, then  $A^T * A$  is an  $n \times n$  invertible matrix. In this case, the left inverse can be found as  $A_L = (A^T * A)^{-1} * A^T$ . When does  $A_L * A = I$  apply? If  $A$  has full column rank, the system  $Ax = b$  has either a unique solution or no solution. If a solution exists, it

is unique. The left inverse is useful in finding least squares solutions for inconsistent systems. For example, if  $Ax = b$  has no exact solution, we often seek  $x$  that minimizes  $\|Ax - b\|^2$ . The normal equations are  $A^T * A * x_{\text{hat}} = A^T * b$ . If  $A$  has full column rank,  $A^T * A$  is invertible, and  $x_{\text{hat}} = (A^T * A)^{-1} * A^T * b$ . Here,  $(A^T * A)^{-1} * A^T$  acts as a left inverse.

### Right Inverse

A matrix  $A$  has a right inverse  $A_R$  if  $A * A_R = I$ . For  $A$  to have a right inverse, it must have full row rank. This means its rows are linearly independent. If  $A$  is an  $m \times n$  matrix with  $m \leq n$  (wide or square), and its rows are linearly independent, then  $A * A^T$  is an  $m \times m$  invertible matrix. In this case, the right inverse can be found as  $A_R = A^T * (A * A^T)^{-1}$ . When does  $A * A_R = I$  apply? If  $A$  has full row rank, the system  $Ax = b$  always has at least one solution. The right inverse helps find a particular solution. If  $Ax = b$  and  $A * A_R = I$ , then  $A * A_R * b = I * b = b$ . So,  $x_p = A_R * b$  is a particular solution. This solution is often the one with the smallest norm.

Common mistake: Confusing the conditions for existence. Left inverse requires full column rank. Right inverse requires full row rank. A square matrix is invertible if and only if it has both full column rank and full row rank (which means its rank equals its dimension).

### Pseudoinverse (Moore-Penrose Inverse)

The pseudoinverse, denoted  $A^+$ , is a generalization of the inverse that exists for *any* matrix, square or rectangular, full rank or rank-deficient. It is unique and satisfies four key properties (the Moore-Penrose conditions). The pseudoinverse provides a unique "best" solution to  $Ax = b$  in a least squares sense. If  $Ax = b$  has a unique solution,  $x = A^+ * b$ . If  $Ax = b$  has multiple solutions,  $A^+ * b$  gives the solution with the minimum Euclidean norm. If  $Ax = b$  has no solution,  $A^+ * b$  gives the least squares solution with the minimum Euclidean norm.

For matrices with full column rank ( $m \geq n$ , rank =  $n$ ):  $A^+ = (A^T * A)^{-1} * A^T$ . This is the same as the left inverse. For matrices with full row rank ( $m \leq n$ , rank =  $m$ ):  $A^+ = A^T * (A * A^T)^{-1}$ . This is the same as the right inverse. For a general rank-deficient matrix, the pseudoinverse is typically found using the Singular Value Decomposition (SVD). If  $A = U * \text{Sigma} * V^T$  is the SVD of  $A$ , then  $A^+ = V * \text{Sigma}^+ * U^T$ .  $\text{Sigma}$  is a diagonal matrix containing the singular values.  $\text{Sigma}^+$  is formed by taking the reciprocal of the non-zero singular values, transposing the matrix, and padding with zeros.

### Example: Pseudoinverse for a full column rank matrix

Let  $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$ . This matrix has full column rank (rank = 2).  $A^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ .  $A^T * A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$ . So,  $(A^T * A)^{-1} = I^{-1} = I$ .  $A^+ = (A^T * A)^{-1} * A^T = I * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ . Notice that  $A^+ * A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$ . This confirms  $A^+$  acts as a left inverse.

The pseudoinverse is a powerful tool in applications like least squares regression, control theory, and signal processing, where systems are often overdetermined or underdetermined.

Practice cue: Compute the pseudoinverse for a simple rank-deficient matrix, e.g.,  $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$ . (Hint: Use SVD or understand the concept of rank-deficiency to see why the full rank formulas don't

directly apply, then try to find  $\mathbf{x}$  minimizing  $\|\mathbf{Ax} - \mathbf{b}\|$  for a  $\mathbf{b}$  not in the column space.)

## Course Synthesis and Review

Covers: 32. Quiz 3 Review · 34. Final Course Review

This final module synthesizes many concepts introduced throughout the MIT 18.06 course. The "Quiz 3 Review" and "Final Course Review" lessons are opportunities to consolidate your understanding of the entire subject. Linear algebra is not a collection of isolated topics. It is a coherent framework where each concept builds upon previous ones. Effective review involves understanding these connections.

### Key Interconnections

Think about the fundamental problem  $\mathbf{Ax} = \mathbf{b}$ .

- **Existence of Solutions:** This depends on whether  $\mathbf{b}$  is in the column space of  $\mathbf{A}$ . The rank of  $\mathbf{A}$  and the augmented matrix  $[\mathbf{A}|\mathbf{b}]$  determine this.
- **Uniqueness of Solutions:** This depends on the null space of  $\mathbf{A}$ . If the null space contains only the zero vector, solutions are unique (if they exist).
- **Solving Systems:** Gaussian elimination, LU decomposition, and matrix inverses (including left, right, and pseudoinverses) are tools for solving  $\mathbf{Ax} = \mathbf{b}$ .
- **Subspaces:** The column space, null space, row space, and left null space are the four fundamental subspaces of a matrix. Understanding their dimensions and bases is crucial. The Rank-Nullity Theorem connects the dimension of the column space (rank) and the null space (nullity).
- **Orthogonality:** The row space and null space are orthogonal complements. The column space and left null space are orthogonal complements. This geometric insight is powerful, especially in projections and least squares. The Gram-Schmidt process allows us to construct orthonormal bases.
- **Eigenvalues and Eigenvectors:** These describe the "directions" in which a linear transformation acts by simple scaling. They are critical for understanding dynamics, stability, and diagonalization. Diagonalization  $\mathbf{A} = \mathbf{S} * \mathbf{\Lambda} * \mathbf{S}^{-1}$  simplifies matrix powers and differential equations.
- **Singular Value Decomposition (SVD):** This is arguably the most powerful matrix factorization.  $\mathbf{A} = \mathbf{U} * \mathbf{\Sigma} * \mathbf{V}^T$ . It reveals the underlying structure of *any* matrix, relating its four fundamental subspaces and providing a robust way to compute the pseudoinverse. SVD is essential for data compression, principal component analysis (PCA), and numerical stability.
- **Least Squares:** When  $\mathbf{Ax} = \mathbf{b}$  has no exact solution, least squares finds the "best fit" solution by projecting  $\mathbf{b}$  onto the column space of  $\mathbf{A}$ . The solution is given by  $\hat{\mathbf{x}} = (\mathbf{A}^T * \mathbf{A})^{-1} * \mathbf{A}^T * \mathbf{b}$ , which directly uses the left inverse concept.

### Review Strategies

1. **Map the Concepts:** Create a concept map or outline that links topics like subspaces, rank, nullity, orthogonality, projections, eigenvalues, and SVD. How does SVD relate to eigenvalues? How does the null space relate to the row space?

2. **Rework Examples:** Go back to key examples from each lesson and try to solve them without looking at the solution. This tests your procedural understanding.
3. **Understand "Why":** Don't just memorize formulas. Ask why a particular method works. Why does  $A^T * A$  appear in least squares? Why are orthonormal bases desirable?
4. **Practice Problem Solving:** The best way to learn linear algebra is by doing. Work through a variety of problems, including conceptual questions and computational exercises.
5. **Focus on Definitions:** A solid understanding of definitions (e.g., what is a basis, what is an eigenvector, what is rank) is foundational.

Common mistake: Treating linear algebra as a set of disconnected formulas. The beauty and power of the subject come from its interconnectedness. For example, the pseudoinverse's existence for any matrix is a direct consequence of SVD, which itself provides orthogonal bases for the four fundamental subspaces.

Practice cue: For each of the four fundamental subspaces, write down its definition, how to find a basis for it, and its relationship to the other three subspaces. Then, choose a non-square matrix and find bases for all four subspaces.

#### KEY TAKEAWAYS

- A vector's coordinates depend on the chosen basis, and a change of basis matrix transforms coordinates between different bases.
- Image compression techniques often rely on changing basis to represent data sparsely, concentrating information in fewer coefficients.
- Left inverses exist for full column rank matrices and are used in least squares solutions for overdetermined systems.
- Right inverses exist for full row rank matrices and provide particular solutions for underdetermined systems.
- The pseudoinverse (Moore-Penrose inverse) exists for any matrix and provides unique minimum-norm least squares solutions.
- The Singular Value Decomposition (SVD) is a robust method to compute the pseudoinverse for any matrix.
- A comprehensive understanding of linear algebra requires recognizing the deep interconnections between concepts like subspaces, rank, orthogonality, and eigenvalues.

#### CHECK YOUR UNDERSTANDING

Answer these before moving on. Model answers are in the Answer Key at the back of this guide.

1. Explain the purpose of a change of basis matrix.
2. Under what conditions does a matrix  $A$  have a left inverse, and how is it typically used?
3. What is the significance of the pseudoinverse compared to traditional left or right inverses?

4. How does the concept of orthogonality relate to the four fundamental subspaces of a matrix?
5. Briefly describe how SVD contributes to understanding and solving problems in linear algebra.

## REFERENCE

# Glossary

---

## Basis

A set of linearly independent vectors that span a given vector space or subspace.

## Change of basis matrix

A matrix that transforms the coordinates of a vector from one basis to another.

## Column space

The set of all possible linear combinations of the column vectors of a matrix, representing all possible  $b$  for which  $Ax = b$  has a solution.

## Determinant

A scalar value associated with a square matrix that indicates whether the matrix is invertible and represents the scaling factor of volume under the linear transformation.

## Diagonalization

The process of factoring a matrix  $A$  into  $S \Lambda S^{-1}$ , where  $S$  is a matrix of eigenvectors and  $\Lambda$  is a diagonal matrix of eigenvalues.

## Dimension

The number of vectors in any basis for a vector space or subspace.

## Eigenvalue

A scalar  $\lambda$  such that for a matrix  $A$ ,  $Ax = \lambda x$  for some non-zero vector  $x$ .

## Eigenvector

A non-zero vector  $x$  that, when multiplied by a matrix  $A$ , results in a scalar multiple of itself ( $\lambda x$ ), where  $\lambda$  is the corresponding eigenvalue.

## Fast fourier transform (fft)

An efficient algorithm for computing the Discrete Fourier Transform.

## Gaussian elimination

A systematic method for solving linear systems by applying elementary row operations to transform the augmented matrix into row echelon form.

## Hermitian matrix

A complex square matrix  $A$  such that  $A = A^H$  (conjugate transpose equals itself).

## Inverse matrix

For a square matrix  $A$ , its inverse  $A^{-1}$  satisfies  $AA^{-1} = A^{-1}A = I$ , where  $I$  is the identity matrix.

## Least squares

A method for finding the best approximate solution  $\hat{x}$  to an inconsistent system  $Ax = b$  by minimizing the squared length of the error vector  $\|b - Ax\|^2$ .

**Least squares solution**

The vector  $\hat{x}$  that minimizes  $\|Ax - b\|^2$  when  $Ax = b$  has no exact solution.

**Left inverse**

A matrix  $A_L$  such that  $A_L * A = I$ , existing for full column rank matrices.

**Linear combination**

A sum of scalar multiples of vectors, such as  $c_1*v_1 + c_2*v_2 + \dots + c_k*v_k$ .

**Linear independence**

A set of vectors is linearly independent if no vector in the set can be written as a linear combination of the others.

**Linear transformation**

A function between vector spaces that preserves vector addition and scalar multiplication.

**Lu factorization**

The decomposition of a matrix  $A$  into a lower triangular matrix  $L$  and an upper triangular matrix  $U$ , such that  $A = LU$ .

**Markov matrix**

A square matrix with non-negative entries where each column sums to 1, used to model probabilities and state transitions.

**Null space**

The set of all vectors  $x$  such that  $Ax = 0$ . It is a subspace whose dimension equals the number of free variables.

**Nullspace**

The set of all vectors  $x$  that satisfy the homogeneous equation  $Ax = 0$ .

**Orthogonal matrix**

A square matrix  $Q$  whose columns are orthonormal vectors, satisfying  $Q^T Q = I$ , which means  $Q^T = Q^{-1}$ .

**Orthogonal subspaces**

Two subspaces are orthogonal if every vector in one subspace is orthogonal to every vector in the other.

**Orthogonal vectors**

Two vectors are orthogonal if their dot product is zero.

**Pivot**

The first non-zero entry in a row of a matrix in row echelon form, used to eliminate entries below it.

**Positive definite matrix**

A symmetric matrix  $A$  where  $x^T A x > 0$  for all non-zero vectors  $x$ .

**Projection matrix**

A matrix  $P$  that transforms any vector  $b$  into its projection  $p$  onto a specific subspace, such that  $p = Pb$ .

**Pseudoinverse (moore-penrose inverse)**

A generalized inverse  $A^+$  that exists for any matrix and provides optimal least squares solutions.

**Rank (of a matrix)**

The number of pivot variables in a matrix, which is equal to the dimension of its column space and row space.

**Right inverse**

A matrix  $A_R$  such that  $A * A_R = I$ , existing for full row rank matrices.

**Row reduced echelon form (rref)**

A unique form of a matrix resulting from Gaussian elimination, where pivot entries are 1, all other entries in pivot columns are 0, and pivot entries are in a staircase pattern.

**Singular value decomposition (svd)**

A factorization of any matrix  $A$  into  $U \Sigma V^T$ , where  $U$  and  $V$  are orthogonal matrices and  $\Sigma$  is a diagonal matrix of singular values.

**Symmetric matrix**

A square matrix  $A$  such that  $A = A^T$ .

**Unitary matrix**

A complex square matrix  $U$  such that  $U^H U = I$ , preserving vector lengths and angles in complex space.

**Vector space**

A set of vectors that is closed under vector addition and scalar multiplication, containing the zero vector.

## REFERENCE

# Answer Key

---

## Chapter 1: Linear Systems, Elimination, and Vector Spaces

1. **Explain the difference between the row picture and the column picture for a system of linear equations.**

The row picture visualizes each equation as a geometric object, like a line or plane. The solution is the intersection of these objects. The column picture represents the system as a linear combination of column vectors, where the goal is to find scalar multiples of the column vectors that sum to the right-hand side vector.

2. **What are the three elementary row operations used in Gaussian elimination, and why are they important?**

The three elementary row operations are swapping two rows, multiplying a row by a non-zero scalar, and adding a multiple of one row to another. They are important because they transform a system into an equivalent system with the same solutions, allowing us to systematically simplify the matrix into an upper triangular form for easy solving.

3. **Under what conditions does a square matrix  $A$  have an inverse  $A^{-1}$ ?**

A square matrix  $A$  has an inverse  $A^{-1}$  if and only if its determinant is non-zero. Such a matrix is called non-singular. Geometrically, this means its columns are linearly independent, and  $Ax=b$  has a unique solution for every  $b$ .

4. **What is the primary benefit of performing an LU factorization on a matrix  $A$ ?**

The primary benefit of LU factorization is to efficiently solve  $Ax = b$  for multiple different  $b$  vectors. Once  $A$  is factored into  $L$  and  $U$ , solving  $Lc = b$  and then  $Ux = c$  is much faster than performing full Gaussian elimination repeatedly for each new  $b$ .

5. **Define the column space and nullspace of a matrix  $A$ .**

The column space of  $A$  is the set of all possible linear combinations of the columns of  $A$ . It represents all vectors  $b$  for which the equation  $Ax = b$  has a solution. The nullspace of  $A$  is the set of all vectors  $x$  that satisfy the homogeneous equation  $Ax = 0$ .

## Chapter 2: The Four Fundamental Subspaces and Orthogonality

1. **How do you find the complete solution to  $Ax = b$ ?**

The complete solution is  $x = x_p + x_n$ , where  $x_p$  is a particular solution and  $x_n$  is any vector from the null space. To find  $x_p$ , set all free variables to zero and solve for the pivot variables. To find  $x_n$ , find the special solutions to  $Ax = 0$  by setting one free variable to 1 and others to 0 at a time, then forming linear combinations of these special solutions.

2. **What are the conditions for a set of vectors to form a basis for a subspace?**

For a set of vectors to form a basis for a subspace, two conditions must be met: the vectors must be linearly independent, and they must span the entire subspace. This means no vector in the set can be

expressed as a linear combination of the others, and every vector in the subspace can be expressed as a linear combination of the basis vectors.

**3. Name the four fundamental subspaces of a matrix  $A$  and state their dimensions in terms of  $m$ ,  $n$ , and  $r$  (rank).**

The four fundamental subspaces are the column space  $C(A)$  with dimension  $r$ , the null space  $N(A)$  with dimension  $n - r$ , the row space  $C(A^T)$  with dimension  $r$ , and the left null space  $N(A^T)$  with dimension  $m - r$ .

**4. Explain the relationship between the null space of  $A$  and the row space of  $A$ .**

The null space  $N(A)$  and the row space  $C(A^T)$  are orthogonal complements in  $\mathbb{R}^n$ . This means every vector in  $N(A)$  is orthogonal to every vector in  $C(A^T)$ . Their dimensions sum to  $n$ , specifically  $(n - r) + r = n$ .

**5. What are the normal equations, and what problem do they solve?**

The normal equations are  $A^T A \hat{x} = A^T b$ . They arise when finding the least squares solution  $\hat{x}$  to an inconsistent system  $Ax = b$ . They solve the problem of finding the  $\hat{x}$  that minimizes the squared error  $\|b - Ax\|^2$  by projecting  $b$  onto the column space of  $A$ .

### Chapter 3: Orthogonality, Determinants, and Eigenvalues

**1. How does the concept of projection relate to finding the least squares solution to an inconsistent system  $Ax = b$ ?**

When  $Ax = b$  has no exact solution, it means  $b$  is not in the column space of  $A$ . The least squares solution  $\hat{x}$  finds the vector  $A\hat{x}$  that is the projection of  $b$  onto the column space of  $A$ . This projection  $p = A\hat{x}$  is the closest vector to  $b$  within the column space, minimizing the error  $\|b - A\hat{x}\|^2$ .

**2. What are the key properties of an orthogonal matrix  $Q$ , and why are they useful in linear algebra?**

An orthogonal matrix  $Q$  has orthonormal columns, meaning  $Q^T Q = I$ , which implies  $Q^T = Q^{-1}$ . These matrices preserve lengths and angles, making them ideal for rotations and reflections. Their inverse being simply their transpose simplifies many calculations, especially in projections and transformations.

**3. Explain how the determinant of a matrix  $A$  reveals information about its invertibility and the geometric effect of the transformation  $A$ .**

The determinant  $\det(A)$  is zero if and only if the matrix  $A$  is singular (not invertible). Geometrically,  $|\det(A)|$  represents the scaling factor of volume when a linear transformation  $A$  is applied to a region. A positive determinant preserves orientation, while a negative one reverses it.

**4. What is the characteristic equation, and how is it used to find the eigenvalues of a matrix?**

The characteristic equation is  $\det(A - \lambda I) = 0$ . It is derived from the eigenvalue equation  $Ax = \lambda x$ , rewritten as  $(A - \lambda I)x = 0$ . For non-trivial solutions (eigenvectors) to exist, the matrix  $(A - \lambda I)$  must be singular, hence its determinant must be zero. Solving this polynomial equation for  $\lambda$  yields the eigenvalues.

**5. Under what conditions can a matrix be diagonalized, and how does diagonalization simplify the computation of  $A^k$ ?**

A matrix  $A$  can be diagonalized if it has a full set of linearly independent eigenvectors. If  $A$  is diagonalizable, it can be written as  $A = S \Lambda S^{-1}$ , where  $\Lambda$  is a diagonal matrix of eigenvalues and  $S$  is a matrix of eigenvectors. This factorization simplifies  $A^k$  to  $S \Lambda^k S^{-1}$ , which is much faster to compute because  $\Lambda^k$  simply involves raising each diagonal eigenvalue to the power  $k$ .

## Chapter 4: Advanced Matrix Structures and Decompositions

**1. What are the defining characteristics of a Markov matrix, and what does its steady state eigenvector represent?**

A Markov matrix has non-negative entries, and each of its columns sums to 1. Its steady state eigenvector, corresponding to the eigenvalue 1, represents the long-term equilibrium distribution of the system the matrix models. This state is reached after many transitions, regardless of the initial conditions.

**2. How can you test if a symmetric matrix is positive definite? Name at least two methods.**

A symmetric matrix is positive definite if all its eigenvalues are positive. Alternatively, it is positive definite if all its pivots (without row exchanges during Gaussian elimination) are positive. Another method involves checking if all its upper-left subdeterminants (principal minors) are positive.

**3. Explain the significance of the Fast Fourier Transform (FFT) in practical applications.**

The FFT is significant because it provides an extremely efficient algorithm for computing the Discrete Fourier Transform. It reduces the computational complexity from  $O(N^2)$  to  $O(N \log N)$ , making Fourier analysis feasible for large datasets in signal processing, image compression, and scientific computing.

**4. What is the primary difference between similar matrices and matrices that are simply row equivalent?**

Similar matrices ( $A$  and  $B$  where  $B = M^{-1} A M$ ) represent the same linear transformation under different bases and share properties like eigenvalues, determinant, and trace. Row equivalent matrices are obtained through elementary row operations and share the same row space and null space, but generally do not share eigenvalues or determinant.

**5. Describe the components of the Singular Value Decomposition (SVD) for a matrix  $A$ , and state one key application.**

For  $A = U \Sigma V^T$ ,  $U$  is an orthogonal matrix of left singular vectors,  $\Sigma$  is a diagonal matrix of singular values, and  $V$  is an orthogonal matrix of right singular vectors. A key application is image compression, where approximating the image matrix with fewer singular values and vectors significantly reduces data size while maintaining visual quality.

**6. When is a square matrix guaranteed to be diagonalizable, and what alternative decomposition exists if it is not?**

A square matrix is guaranteed to be diagonalizable if it has a full set of linearly independent eigenvectors. This occurs if all its eigenvalues are distinct, or if for any repeated eigenvalue, its

algebraic multiplicity equals its geometric multiplicity. If a matrix is not diagonalizable, it can still be transformed into its Jordan canonical form.

## Chapter 5: Basis Transformations, Generalized Inverses, and Course Synthesis

### 1. Explain the purpose of a change of basis matrix.

A change of basis matrix allows us to convert the coordinate representation of a vector from one basis to another. This is useful for simplifying problems or for representing data more efficiently in a different coordinate system, such as in image compression.

### 2. Under what conditions does a matrix $A$ have a left inverse, and how is it typically used?

A matrix  $A$  has a left inverse if it has full column rank. It is typically used in solving overdetermined linear systems  $Ax = b$  that have no exact solution, providing the least squares solution  $\hat{x} = (A^T A)^{-1} A^T b$ .

### 3. What is the significance of the pseudoinverse compared to traditional left or right inverses?

The pseudoinverse is significant because it exists for *any* matrix, regardless of its shape or rank. It provides a unique, minimum-norm least squares solution to  $Ax = b$ , making it a universally applicable tool for solving linear systems, even inconsistent or underdetermined ones.

### 4. How does the concept of orthogonality relate to the four fundamental subspaces of a matrix?

The row space and the null space of a matrix are orthogonal complements in  $\mathbb{R}^n$ . Similarly, the column space and the left null space are orthogonal complements in  $\mathbb{R}^m$ . This means every vector in one subspace is orthogonal to every vector in its complement, and their dimensions sum to the dimension of the ambient space.

### 5. Briefly describe how SVD contributes to understanding and solving problems in linear algebra.

SVD decomposes any matrix  $A$  into  $U * \Sigma * V^T$ , revealing its singular values and orthonormal bases for its four fundamental subspaces. It is crucial for understanding matrix rank, computing the pseudoinverse, performing data compression, and is numerically stable for many computational tasks.

## CLOSING

## Where to Go Next

---

After completing "MIT 18.06: Linear Algebra", consolidate your skills through continued practice. Apply the concepts learned to new problems. Consider working through additional exercises from textbooks or online resources.

You can apply linear algebra to various practical areas. Try implementing simple linear regression models in Python or analyzing small datasets. Explore how linear transformations affect geometric shapes in programming environments.

This course prepares you for many advanced subjects. These include multivariable calculus, differential equations, statistics, and machine learning algorithms. You are now equipped with a powerful mathematical toolkit for further study.



### MEGACOURSE

*Learn Anything. Achieve Everything.*

Track your progress, take the quizzes, and earn your certificate for this course at  
**[megacourse.app/courses/mit-18-06-linear-algebra](https://megacourse.app/courses/mit-18-06-linear-algebra)**