

Review of Linear Algebra Concepts and Course Introduction

Lecture Notes

April 11, 2025

Administrative Notes:

Welcome! Let's cover a few administrative points before diving into the mathematics.

Grading Scheme: Your final grade in this course will be based on the following components:

- **Homework Assignments (10%):** There will be approximately 10 assignments throughout the semester. Submission is mandatory. You are allowed to miss submitting **one** assignment without penalty. Each submitted assignment will be graded on a pass/fail/excellent scale:

- Did not submit / Unsatisfactory attempt: 0 points
- Satisfactory attempt (shows genuine effort): 80 points
- Excellent work: 100 points

A 'satisfactory' grade indicates you've seriously engaged with the problems, even if not every detail is perfect. Aim for understanding!

- **Midterm Quiz (20%):** This quiz serves as a 'shield' (or 'magen' grade), meaning it can only help your final grade (it will replace a portion of the final exam weight if it's higher than your final exam score). However, **taking the quiz is mandatory**. We anticipate the quiz will be around week 8-9, but the exact date will be confirmed later.
- **Final Exam (70%):** This comprehensive exam covers all course material. The exact weight might be adjusted slightly depending on your quiz performance (if the quiz score is higher).

Homework Logistics:

- Expect roughly one assignment per week.
- You will generally have about 10 days to complete each assignment.
- **Extensions:** If you foresee difficulty meeting a deadline due to legitimate reasons, please request an extension **in advance** through the designated private forum (see below). Don't wait until the last minute. Try your best to solve the problems yourself first, even if you need more time.
- **Collaboration and Resources:** You will likely find solutions to some homework problems online (e.g., via Google, ChatGPT, past course materials). While these resources exist, relying on them directly will hinder your learning and won't prepare you for the quiz or exam.

- **Recommended Approach:** Always attempt the problems seriously on your own first. If you get stuck, discuss concepts with peers, consult the textbook/notes, or ask for *hints* on the forum or during office hours. If you do consult an external solution, make sure you understand it deeply and then write up your own version *in your own words* without looking at the source.
- **Warning about AI Tools (e.g., ChatGPT):** While potentially helpful for generating ideas or checking simple steps, these tools can make mistakes, especially with more complex statistical or proof-based reasoning. They might provide answers that seem plausible but are incorrect. Developing your own understanding is crucial to identify such errors. Using AI to simply generate answers might get you points on homework but not the understanding needed for exams.

Getting Help and Communication:

- **Public Q&A Forum:** This is the primary place for questions about course material (definitions, concepts, lecture examples, general homework questions). Asking here benefits everyone, and you'll likely get faster responses. The forum is anonymous to other students, so please feel comfortable asking anything, no matter how basic it seems. (Note: Course staff can see your name).
 - **Private Request Forum:** Use this forum for personal matters, such as extension requests or discussing grades.
 - **Email:** Please use the forums as the primary means of communication. Email should be reserved for urgent or highly personal matters that cannot be addressed via the forums.
-

1 Motivation: Why Linear Algebra Matters

Before we dive into the technical details, let's take a moment to understand why linear algebra is such a fundamental tool, especially in statistics, data science, and machine learning. You might wonder why the first few weeks of a statistics course are dedicated to this topic. It's not just mathematical formalism; it's the language we use to efficiently represent and manipulate data.

- **Representing Data:** The most basic way to organize data is often a table, like a spreadsheet. Imagine a dataset of patients, where each row is a patient and columns represent features like height, age, and weight. In linear algebra, this table is simply a **matrix**. Each patient can be seen as a **vector** of features.
- **Beyond Simple Tables:** Linear algebra allows us to represent more complex data structures.
 - **Images:** A digital image is essentially a grid of pixels. A grayscale image can be represented as a matrix where each entry is the intensity of a pixel. A color image (RGB) can be represented as three matrices (one for each color channel) or often as a 3rd-order tensor. Operations like finding edges or features in an image often involve multiplying the image matrix by smaller matrices (filters or kernels) in a process related to convolution.

- **Graphs and Networks:** Consider a social network. We can represent the relationships (e.g., who talks to whom) using an **adjacency matrix**, A , where $A_{ij} = 1$ if person i and person j are connected (or interact), and $A_{ij} = 0$ otherwise. We could even weight these connections. Analyzing properties of the network often involves computations on this matrix. (Think of the characters in "Friends" and mapping who talks to whom).
- **Analyzing Data:** Once data is represented using vectors and matrices, linear algebra provides the tools for analysis.
 - **Regression:** In introductory statistics, you likely encountered simple linear regression ($y = \beta_0 + \beta_1 x + \epsilon$). What if we have multiple predictors (height, age, weight predicting a disease)? We move to multiple regression, which is most naturally expressed and solved using matrix notation: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Finding the best $\boldsymbol{\beta}$ involves matrix operations.
 - **Comparing Groups (ANOVA):** You might have compared the means of two groups (e.g., men vs. women). What if you have multiple groups? Analysis of Variance (ANOVA) techniques, used to determine if there are significant differences between multiple group means, are deeply rooted in linear algebra concepts like projections and subspaces.
 - **Dimensionality Reduction:** Techniques like Principal Component Analysis (PCA) use eigenvalues and eigenvectors (which we'll review) to find the most important directions in high-dimensional data.
- **Machine Learning:** Many machine learning algorithms, including neural networks, fundamentally rely on matrix and vector operations. Training these models involves optimizing functions defined over matrices (e.g., weights in a neural network).

So, the linear algebra we cover in the first five weeks is not an arbitrary hurdle. It's the essential foundation for understanding and working with the statistical models and data analysis techniques that follow. Mastering these tools will make the rest of the course, and further studies in related fields, much more accessible.

2 Course Structure Preview

- **Weeks 1-5 (Foundation):** We start with a review of core linear algebra concepts (vector spaces, bases, orthogonality, eigenvalues, etc.). We will then introduce random vectors and random matrices, exploring their properties (mean vectors, covariance matrices). This part is mathematically intensive but crucial.
- **Subsequent Weeks (Applications):** Building on this foundation, we will delve into statistical inference in higher dimensions, multivariate analysis techniques, and more advanced regression models, all framed using the language of linear algebra.

Practice is key during the initial weeks. The concepts build on each other, and familiarity with the manipulations will make later topics much easier. You'll find that many 'tricks' and techniques reappear in different contexts.

3 Review: Vector Spaces and Subspaces

Let's refresh some fundamental concepts. We'll primarily work over the field of real numbers, \mathbb{R} .

Definition 3.1 (Vector Space). A **vector space** V over a field \mathbb{F} (for us, $\mathbb{F} = \mathbb{R}$) is a set of objects called vectors, equipped with two operations:

1. Vector Addition: $+: V \times V \rightarrow V$
2. Scalar Multiplication: $\cdot: \mathbb{F} \times V \rightarrow V$

These operations must satisfy certain axioms (associativity, commutativity, identity elements, inverses, distributivity, etc.). We won't list all the axioms here, but standard examples like \mathbb{R}^n , the space of polynomials, and the space of continuous functions should be familiar.

Definition 3.2 (Subspace). Let V be a vector space over \mathbb{F} . A subset $W \subseteq V$ is called a **subspace** of V if W is itself a vector space over \mathbb{F} with the same operations inherited from V .

It's often easier to check the subspace conditions directly:

Theorem 3.3 (Subspace Test). *A non-empty subset $W \subseteq V$ is a subspace of V if and only if it satisfies the following two closure properties:*

1. **Closure under Addition:** For all $\mathbf{w}_1, \mathbf{w}_2 \in W$, we have $\mathbf{w}_1 + \mathbf{w}_2 \in W$.
2. **Closure under Scalar Multiplication:** For all $\mathbf{w} \in W$ and all scalars $\alpha \in \mathbb{F}$, we have $\alpha \mathbf{w} \in W$.

Equivalently, a non-empty subset W is a subspace if and only if for all $\mathbf{w}_1, \mathbf{w}_2 \in W$ and all scalars $\alpha, \beta \in \mathbb{F}$, the linear combination $\alpha \mathbf{w}_1 + \beta \mathbf{w}_2$ is also in W . Note: These conditions imply that the zero vector $\mathbf{0}$ must be in W (take $\alpha = 0$).

4 Span, Linear Independence, Basis, and Dimension

Definition 4.1 (Linear Combination and Span). Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ be a set of vectors in a vector space V .

- A **linear combination** of vectors in S is a vector of the form $\mathbf{w} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \dots + \alpha_k \mathbf{v}_k$, where $\alpha_1, \dots, \alpha_k$ are scalars.
- The **span** of S , denoted $\text{span}(S)$, is the set of all possible linear combinations of vectors in S .

$$\text{span}(S) = \{\alpha_1 \mathbf{v}_1 + \dots + \alpha_k \mathbf{v}_k \mid \alpha_1, \dots, \alpha_k \in \mathbb{F}\}$$

If $W = \text{span}(S)$, we say that S **spans** or **generates** W .

Proposition 4.2. *For any set of vectors $S \subseteq V$, the span $\text{span}(S)$ is a subspace of V .*

Sketch based on lecture. We need to check the closure properties for $W = \text{span}(S)$.

1. Zero Vector: $\mathbf{0} = 0\mathbf{v}_1 + \dots + 0\mathbf{v}_k$, so $\mathbf{0} \in W$. (Assuming S is not empty).
2. Closure under Addition: Let $\mathbf{w}_1, \mathbf{w}_2 \in W$. Then $\mathbf{w}_1 = \sum \alpha_i \mathbf{v}_i$ and $\mathbf{w}_2 = \sum \beta_i \mathbf{v}_i$ for some scalars α_i, β_i . Their sum is $\mathbf{w}_1 + \mathbf{w}_2 = \sum (\alpha_i + \beta_i) \mathbf{v}_i$. Since $\alpha_i + \beta_i$ are scalars, this sum is also a linear combination of vectors in S , so $\mathbf{w}_1 + \mathbf{w}_2 \in W$.

3. Closure under Scalar Multiplication: Let $\mathbf{w} \in W$ and $c \in \mathbb{F}$. Then $\mathbf{w} = \sum \alpha_i \mathbf{v}_i$. So, $c\mathbf{w} = c(\sum \alpha_i \mathbf{v}_i) = \sum (c\alpha_i) \mathbf{v}_i$. Since $c\alpha_i$ are scalars, $c\mathbf{w}$ is a linear combination, hence $c\mathbf{w} \in W$.

Therefore, $\text{span}(S)$ is a subspace. \square

Definition 4.3 (Linear Dependence and Independence). Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ be a set of vectors in V .

- S is **linearly dependent** if there exist scalars $\alpha_1, \dots, \alpha_k$, *not all zero*, such that $\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \dots + \alpha_k \mathbf{v}_k = \mathbf{0}$. This means at least one vector in S can be written as a linear combination of the others.
- S is **linearly independent** if the only way to form the zero vector as a linear combination is by using all zero scalars. That is, the equation $\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \dots + \alpha_k \mathbf{v}_k = \mathbf{0}$ implies $\alpha_1 = \alpha_2 = \dots = \alpha_k = 0$.

Remark 4.4. Any set containing more than n vectors in \mathbb{R}^n must be linearly dependent.

Definition 4.5 (Basis). A set of vectors $B = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ in a vector space V is called a **basis** for V if:

1. B is linearly independent.
2. B spans V (i.e., $\text{span}(B) = V$).

Theorem 4.6 (Equivalent Conditions for Basis). *Let W be a subspace. A set $B \subseteq W$ is a basis for W if and only if it satisfies any of the following equivalent conditions:*

- B is a minimal spanning set for W . (Removing any vector makes it no longer span W).
- B is a maximal linearly independent set in W . (Adding any other vector from W makes the set dependent).

Remark 4.7 (Non-uniqueness of Basis). A given vector space (or subspace) generally has infinitely many different bases. However, all bases for a given finite-dimensional vector space have the same number of vectors.

Example 4.8 (From Lecture: Non-unique Basis). Consider the set $S = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\}$ in \mathbb{R}^2 .

- Is S linearly independent? No, because $2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 1 \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, a non-trivial linear combination giving the zero vector.
- What is the span of S ? $\text{span}(S) = \left\{ \alpha \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \beta \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\} = \left\{ (\alpha + 2\beta) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} = \left\{ k \begin{pmatrix} 1 \\ 1 \end{pmatrix} \mid k \in \mathbb{R} \right\}$. This is the line $y = x$ in \mathbb{R}^2 . Let $W = \text{span}(S)$.
- S spans W but is not a basis because it's linearly dependent.
- What are some bases for W ?

– $B_1 = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$ is linearly independent (single non-zero vector) and spans W . So B_1 is a basis.

- $B_2 = \left\{ \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\}$ is also linearly independent and spans W . So B_2 is a basis.
- $B_3 = \left\{ \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\}$ is also a basis.

This illustrates that a subspace can have multiple bases.

Definition 4.9 (Dimension). The **dimension** of a finite-dimensional vector space V , denoted $\dim(V)$, is the number of vectors in any basis for V .

In the example above, $\dim(W) = 1$.

5 Inner Products, Orthogonality, and Orthonormality

We often want to measure lengths and angles in vector spaces. This requires the notion of an inner product.

Definition 5.1 (Inner Product). An **inner product** on a real vector space V is a function $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ that satisfies the following properties for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all scalars $\alpha \in \mathbb{R}$:

1. Symmetry: $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
2. Linearity in the first argument: $\langle \alpha \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \alpha \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
3. Positive-definiteness: $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$, and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

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

Remark 5.2 (Standard Inner Product on \mathbb{R}^n). Unless otherwise specified, the inner product we use on \mathbb{R}^n is the standard **Euclidean inner product** (or dot product): For $\mathbf{u} = (u_1, \dots, u_n)$ and $\mathbf{w} = (w_1, \dots, w_n)$,

$$\langle \mathbf{u}, \mathbf{w} \rangle = \mathbf{u}^T \mathbf{w} = \sum_{i=1}^n u_i w_i$$

We will use this inner product throughout the course.

Definition 5.3 (Norm and Orthogonality). In an inner product space V :

- The **norm** (or length) of a vector \mathbf{v} is $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.
- Two vectors $\mathbf{u}, \mathbf{v} \in V$ are **orthogonal** if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. We write $\mathbf{u} \perp \mathbf{v}$.
- A set of vectors $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is an **orthogonal set** if all pairs of distinct vectors are orthogonal: $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ for all $i \neq j$.
- An **orthonormal set** is an orthogonal set where additionally each vector has norm 1: $\|\mathbf{v}_i\| = 1$ for all i . (Equivalently, $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \delta_{ij}$, the Kronecker delta).

Theorem 5.4. Any orthogonal set of non-zero vectors is linearly independent.

Definition 5.5 (Orthogonal and Orthonormal Basis). • An **orthogonal basis** for an inner product space V is a basis that is also an orthogonal set.

- An **orthonormal basis (ONB)** for V is a basis that is also an orthonormal set.

Orthonormal bases are particularly convenient to work with (e.g., finding coordinates, projections).

Example 5.6 (From Lecture: Orthogonal Basis). Consider the set $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ in \mathbb{R}^3 , where

$$\mathbf{v}_1 = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}.$$

- **Check Orthogonality:**

$$\begin{aligned} - \langle \mathbf{v}_1, \mathbf{v}_2 \rangle &= (2)(0) + (0)(1) + (0)(0) = 0 \\ - \langle \mathbf{v}_1, \mathbf{v}_3 \rangle &= (2)(0) + (0)(0) + (0)(3) = 0 \\ - \langle \mathbf{v}_2, \mathbf{v}_3 \rangle &= (0)(0) + (1)(0) + (0)(3) = 0 \end{aligned}$$

The set is orthogonal.

- **Check Linear Independence:** Since it's an orthogonal set of non-zero vectors, it's linearly independent.
- **Check Spanning:** We have 3 linearly independent vectors in \mathbb{R}^3 . Therefore, they form a basis for \mathbb{R}^3 .
- **Conclusion:** B is an orthogonal basis for \mathbb{R}^3 .
- **Is it Orthonormal?** Let's check the norms:

$$\begin{aligned} - \|\mathbf{v}_1\| &= \sqrt{2^2 + 0^2 + 0^2} = \sqrt{4} = 2 \\ - \|\mathbf{v}_2\| &= \sqrt{0^2 + 1^2 + 0^2} = \sqrt{1} = 1 \\ - \|\mathbf{v}_3\| &= \sqrt{0^2 + 0^2 + 3^2} = \sqrt{9} = 3 \end{aligned}$$

Since not all norms are 1, the basis B is orthogonal but *not* orthonormal.

- **Creating an Orthonormal Basis:** We can normalize each vector: $\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$,

$\mathbf{u}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, $\mathbf{u}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. The set $B' = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal basis (the standard basis, in this case).

Definition 5.7 (Orthogonal Matrix). A square matrix $A \in \mathbb{R}^{n \times n}$ is called an **orthogonal matrix** if its columns form an orthonormal basis for \mathbb{R}^n .

Theorem 5.8. A square matrix $A \in \mathbb{R}^{n \times n}$ is orthogonal if and only if $A^T A = I$. This also implies $A A^T = I$, and thus $A^{-1} = A^T$.

Sketch of $A^T A = I$. Let $A = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_n]$, where $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ are the columns of A . The matrix A^T has the rows $(\mathbf{a}_1)^T, \dots, (\mathbf{a}_n)^T$. Consider the (i, j) -th entry of the product $A^T A$:

$$(A^T A)_{ij} = (\text{row } i \text{ of } A^T) \cdot (\text{column } j \text{ of } A) = (\mathbf{a}_i)^T \mathbf{a}_j = \langle \mathbf{a}_i, \mathbf{a}_j \rangle$$

Since the columns $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ form an orthonormal basis by definition of an orthogonal matrix, we have:

$$\langle \mathbf{a}_i, \mathbf{a}_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} = \delta_{ij}$$

This means $(A^T A)_{ij} = \delta_{ij}$, which is precisely the definition of the identity matrix I . Therefore, $A^T A = I$. (A similar argument using the fact that the rows of an orthogonal matrix also form an ONB shows $AA^T = I$). \square

Remark 5.9 (Rank). For any matrix A , $\text{rank}(A) = \text{rank}(A^T)$. This means the row rank (dimension of the row space) equals the column rank (dimension of the column space).

6 Projections and Orthogonal Complements

Definition 6.1 (Orthogonal Complement). Let W be a subspace of an inner product space V . The **orthogonal complement** of W , denoted W^\perp (read "W perp"), is the set of all vectors in V that are orthogonal to every vector in W :

$$W^\perp = \{\mathbf{v} \in V \mid \langle \mathbf{v}, \mathbf{w} \rangle = 0 \text{ for all } \mathbf{w} \in W\}$$

Proposition 6.2. W^\perp is also a subspace of V .

Theorem 6.3 (Orthogonal Decomposition Theorem). Let W be a finite-dimensional subspace of an inner product space V . Then every vector $\mathbf{y} \in V$ can be written uniquely as

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{e}$$

where $\hat{\mathbf{y}} \in W$ and $\mathbf{e} \in W^\perp$. The vector $\hat{\mathbf{y}}$ is called the **orthogonal projection** of \mathbf{y} onto W , denoted $\text{proj}_W \mathbf{y}$. Furthermore, V decomposes as a direct sum: $V = W \oplus W^\perp$.

Remark 6.4 (Projection Formula). If $\{\mathbf{b}_1, \dots, \mathbf{b}_k\}$ is an orthonormal basis for W , then the projection of \mathbf{y} onto W is given by:

$$\hat{\mathbf{y}} = \text{proj}_W \mathbf{y} = \sum_{i=1}^k \langle \mathbf{y}, \mathbf{b}_i \rangle \mathbf{b}_i$$

The component orthogonal to W is $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$.

Remark 6.5 (Geometric Interpretation). The projection $\hat{\mathbf{y}}$ is the vector in W that is "closest" to \mathbf{y} (in terms of the norm induced by the inner product). The vector $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ represents the "error" or the part of \mathbf{y} that cannot be represented by vectors in W .

Remark 6.6 (Gram-Schmidt Process). Given any basis for a subspace W , the Gram-Schmidt process provides an algorithm to construct an orthogonal (and subsequently orthonormal) basis for W . This is essential for using the projection formula above.

Example 6.7 (From Lecture: Projection in Regression). Consider the linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Let $W = \text{Col}(\mathbf{X})$ be the column space of the design matrix \mathbf{X} (the subspace spanned by the predictor vectors, including a column of ones if there's an intercept).

- The vector of fitted values, $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$, is the orthogonal projection of the observed values \mathbf{y} onto the subspace W . So, $\hat{\mathbf{y}} = \text{proj}_W \mathbf{y}$.

- The vector of residuals, $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$, represents the part of \mathbf{y} that is orthogonal to the column space W . Thus, $\mathbf{e} \in W^\perp$.
- This orthogonality ($\mathbf{e} \in W^\perp$) means that \mathbf{e} is orthogonal to every column of \mathbf{X} . Mathematically, $\mathbf{X}^T \mathbf{e} = \mathbf{0}$. This is exactly what the normal equations state!
- As shown in the lecture, this implies $\langle \mathbf{e}, \hat{\mathbf{y}} \rangle = 0$. We can verify this: since $\hat{\mathbf{y}} \in W$, and $\mathbf{e} \in W^\perp$, their inner product must be zero by definition of the orthogonal complement.
- Specifically, if the model includes an intercept β_0 and a single predictor x_1 (so $W = \text{span}\{\mathbf{1}, \mathbf{x}_1\}$), then $\mathbf{X}^T \mathbf{e} = \mathbf{0}$ means:

- $\mathbf{1}^T \mathbf{e} = \sum e_i = 0$ (sum of residuals is zero).
- $\mathbf{x}_1^T \mathbf{e} = \sum x_{i1} e_i = 0$ (residuals are uncorrelated with the predictor).

The calculation $\sum \hat{y}_i e_i = \sum (\hat{\beta}_0 + \hat{\beta}_1 x_{i1}) e_i = \hat{\beta}_0 \sum e_i + \hat{\beta}_1 \sum x_{i1} e_i = \hat{\beta}_0(0) + \hat{\beta}_1(0) = 0$ directly verifies $\langle \hat{\mathbf{y}}, \mathbf{e} \rangle = 0$.

This connection between projection, orthogonal complements, and the residuals in linear regression is fundamental.

7 Sums and Direct Sums of Subspaces

Definition 7.1 (Sum of Subspaces). Let U and W be subspaces of a vector space V . Their **sum** is the set:

$$U + W = \{\mathbf{u} + \mathbf{w} \mid \mathbf{u} \in U, \mathbf{w} \in W\}$$

$U + W$ is the smallest subspace of V containing both U and W .

Theorem 7.2 (Dimension of a Sum). *If U and W are finite-dimensional subspaces of V , then*

$$\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W)$$

Definition 7.3 (Direct Sum). Let U and W be subspaces of V . The sum $U + W$ is called a **direct sum**, denoted $U \oplus W$, if $U \cap W = \{\mathbf{0}\}$ (their intersection contains only the zero vector).

Theorem 7.4 (Unique Decomposition for Direct Sums). *Let U and W be subspaces of V . Then $V = U \oplus W$ if and only if every vector $\mathbf{v} \in V$ can be written **uniquely** as $\mathbf{v} = \mathbf{u} + \mathbf{w}$, where $\mathbf{u} \in U$ and $\mathbf{w} \in W$.*

Sketch based on lecture. (\Rightarrow) Assume $V = U \oplus W$. This means $V = U + W$ and $U \cap W = \{\mathbf{0}\}$. Suppose \mathbf{v} has two representations: $\mathbf{v} = \mathbf{u}_1 + \mathbf{w}_1 = \mathbf{u}_2 + \mathbf{w}_2$, with $\mathbf{u}_1, \mathbf{u}_2 \in U$ and $\mathbf{w}_1, \mathbf{w}_2 \in W$. Rearranging gives $\mathbf{u}_1 - \mathbf{u}_2 = \mathbf{w}_2 - \mathbf{w}_1$. The left side is in U (since U is a subspace), and the right side is in W . Therefore, this vector lies in $U \cap W$. Since the intersection is only $\{\mathbf{0}\}$, we must have $\mathbf{u}_1 - \mathbf{u}_2 = \mathbf{0}$ and $\mathbf{w}_2 - \mathbf{w}_1 = \mathbf{0}$. Thus, $\mathbf{u}_1 = \mathbf{u}_2$ and $\mathbf{w}_1 = \mathbf{w}_2$, proving uniqueness.

(\Leftarrow) Assume every $\mathbf{v} \in V$ has a unique representation $\mathbf{v} = \mathbf{u} + \mathbf{w}$ with $\mathbf{u} \in U, \mathbf{w} \in W$.

- This immediately implies $V = U + W$.
- We need to show $U \cap W = \{\mathbf{0}\}$. Assume $\mathbf{x} \in U \cap W$. We can write \mathbf{x} in two ways as a sum from $U + W$:

1. $\mathbf{x} = \mathbf{x} + \mathbf{0}$ (where $\mathbf{x} \in U, \mathbf{0} \in W$)

2. $\mathbf{x} = \mathbf{0} + \mathbf{x}$ (where $\mathbf{0} \in U$, $\mathbf{x} \in W$)

Since the representation must be unique by assumption, we must have $\mathbf{x} = \mathbf{0}$. Therefore, $U \cap W = \{\mathbf{0}\}$, and the sum is direct.

□

8 Eigenvalues and Eigenvectors

Eigenvalues and eigenvectors are crucial for understanding linear transformations, solving differential equations, and many statistical techniques like PCA.

Definition 8.1 (Eigenvalue and Eigenvector). Let A be an $n \times n$ matrix. A scalar λ is called an **eigenvalue** of A if there exists a non-zero vector $\mathbf{u} \in \mathbb{R}^n$ such that

$$A\mathbf{u} = \lambda\mathbf{u}$$

Any such non-zero vector \mathbf{u} is called an **eigenvector** of A corresponding to the eigenvalue λ .

Intuitively, an eigenvector \mathbf{u} is a vector whose direction is unchanged (only scaled by a factor λ) when multiplied by the matrix A .

Definition 8.2 (Finding Eigenvalues: Characteristic Polynomial). Eigenvalues are found by solving the characteristic equation:

$$\det(A - \lambda I) = 0$$

The expression $p(\lambda) = \det(A - \lambda I)$ is a polynomial in λ of degree n , called the **characteristic polynomial** of A . The eigenvalues are the roots of this polynomial.

Definition 8.3 (Finding Eigenvectors: Eigenspace). For a given eigenvalue λ , the corresponding eigenvectors are the non-zero solutions \mathbf{u} to the homogeneous system $(A - \lambda I)\mathbf{u} = \mathbf{0}$. The set of all solutions (including the zero vector) forms a subspace called the **eigenspace** of λ , denoted $E_\lambda = \text{null}(A - \lambda I)$.

Definition 8.4 (Diagonalization). An $n \times n$ matrix A is **diagonalizable** if it is similar to a diagonal matrix. That is, if there exists an invertible matrix P and a diagonal matrix D such that

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

This is possible if and only if A has n linearly independent eigenvectors. In this case, the columns of P are the linearly independent eigenvectors, and the diagonal entries of D are the corresponding eigenvalues, in the same order.

8.1 Symmetric Matrices and the Spectral Theorem

Symmetric matrices have particularly nice properties regarding eigenvalues and eigenvectors.

Theorem 8.5 (Properties of Real Symmetric Matrices). *Let A be an $n \times n$ real symmetric matrix ($A = A^T$). Then:*

1. *All eigenvalues of A are real numbers.*
2. *Eigenvectors corresponding to distinct eigenvalues are orthogonal.*

3. A is always diagonalizable.

Theorem 8.6 (Spectral Theorem for Real Symmetric Matrices). *If A is an $n \times n$ real symmetric matrix, then there exists an **orthogonal** matrix U (i.e., $U^T U = I$, so $U^{-1} = U^T$) and a diagonal matrix Λ such that*

$$A = U \Lambda U^T$$

*The columns of U are an orthonormal basis of eigenvectors for A , and the diagonal entries of Λ are the corresponding real eigenvalues. This is called the **spectral decomposition** of A .*

This theorem is incredibly important in statistics and many other fields.

8.2 Examples and Properties

Example 8.7 (From Lecture: Eigenvalues/vectors of $A + cI$). Let A be a symmetric $n \times n$ matrix with spectral decomposition $A = U \Lambda U^T$. Let $c \in \mathbb{R}$ be a scalar, and define $S = A + cI$.

1. **Is S symmetric?** $S^T = (A + cI)^T = A^T + (cI)^T = A + cI^T = A + cI = S$. Yes, S is symmetric.
2. **What are the eigenvalues and eigenvectors of S ?** Let \mathbf{u} be an eigenvector of A corresponding to eigenvalue λ , so $A\mathbf{u} = \lambda\mathbf{u}$. Consider $S\mathbf{u}$:

$$S\mathbf{u} = (A + cI)\mathbf{u} = A\mathbf{u} + cI\mathbf{u} = \lambda\mathbf{u} + c\mathbf{u} = (\lambda + c)\mathbf{u}$$

This shows that \mathbf{u} is also an eigenvector of S , but with eigenvalue $(\lambda + c)$.

3. **Spectral Decomposition of S :** Since S is symmetric and has the same eigenvectors as A (which form the columns of U) and eigenvalues $\lambda_i + c$, its spectral decomposition is:

$$S = U(\Lambda + cI)U^T$$

where $\Lambda + cI$ is the diagonal matrix with entries $\lambda_i + c$.

Example 8.8 (From Lecture: Powers and Inverses via Spectral Decomposition). Let A be a symmetric matrix with $A = U \Lambda U^T$.

1. **Powers of A :** Consider A^2 :

$$A^2 = AA = (U \Lambda U^T)(U \Lambda U^T) = U \Lambda (U^T U) \Lambda U^T$$

Since $U^T U = I$ (because U is orthogonal), this simplifies to:

$$A^2 = U \Lambda I \Lambda U^T = U \Lambda^2 U^T$$

By induction, for any positive integer k , $A^k = U \Lambda^k U^T$. (Λ^k is the diagonal matrix with entries λ_i^k).

2. **Inverse of A :** Assume A is invertible. This means all eigenvalues λ_i must be non-zero. We look for A^{-1} . Consider the matrix $X = U \Lambda^{-1} U^T$, where Λ^{-1} is the diagonal matrix with entries $1/\lambda_i$. Let's check if $AX = I$:

$$AX = (U \Lambda U^T)(U \Lambda^{-1} U^T) = U \Lambda (U^T U) \Lambda^{-1} U^T = U \Lambda I \Lambda^{-1} U^T = U (\Lambda \Lambda^{-1}) U^T$$

Since $\Lambda \Lambda^{-1} = I$ (the identity matrix), we have:

$$AX = U I U^T = U U^T = I$$

Therefore, $A^{-1} = U \Lambda^{-1} U^T$.

3. **Applying to $S = A + cI$:** Assuming S is invertible (i.e., all $\lambda_i + c \neq 0$), its inverse is:

$$S^{-1} = (A + cI)^{-1} = U(\Lambda + cI)^{-1}U^T$$

where $(\Lambda + cI)^{-1}$ is the diagonal matrix with entries $1/(\lambda_i + c)$.

These properties are extremely useful for computing functions of symmetric matrices.

Administrative Notes:

Homework Assignments:

- **Homework 0 (Optional):** This assignment covers background material and serves as a refresher. It is highly recommended but not for submission/grading. You can find it on the course website.
- **Homework 1 (Required):** This assignment is due [Insert Due Date if known, otherwise state 'Next Week' or similar]. Please submit it via the course website by the deadline.
 - It covers material from this week's review.
 - It includes topics like positive definite matrices (the definition is provided).
 - **Note on Question 3:** This question might be more challenging, partly because we haven't covered all related concepts in detail yet and the problem itself is a bit more involved. Please attempt it, try your best, and don't hesitate to ask for clarification or hints on the forum or during office hours if you get stuck.

My office hours are [Insert Day/Time/Location]. Please come prepared with specific questions.
