

Linear Algebra and Regression: Projections, Subspaces, and Diagonalization

Lecture Synthesis and Commentary

Instructor: [Your Name]

April 13, 2025

Administrative Note

Administrative Announcements:

- **Supplementary Instruction Session:** We are scheduling an additional review/recitation (minimum two hours). The plan is for Niv to conduct and record this session, likely taking place this week before Passover, preferably in-person if logistics allow. *If you have expressed interest, please make every effort to attend.*
- **Assignments:** Current assignments will be discussed in upcoming sessions. Due to the current material's density and technical nature, there is a deliberate pacing to ensure understanding. You will also have a brief period without other courses to consolidate your linear algebra knowledge.
- **General Note:** This class is uniquely positioned, as you are the first to have completed two linear algebra courses before this one. Expect an intensive, compressed review now, but be assured that the quantity of new essential results is manageable—much of what appears "overwhelming" will soon integrate into your broader understanding.
- **Other Logistics:** Administrative details regarding assignment timing, project opportunities (AMS), and TAs' responsibilities are coordinated as needed. Further announcements will be provided as scheduling solidifies.

1 Introduction and Motivation

The topics discussed in this synthesis form the mathematical backbone of the statistical linear model and, more broadly, reveal some of the most elegant structures in linear algebra. Our focus will be on:

- Projection matrices, their properties, and their geometric meaning.
- Decomposition of vectors into subspaces and their orthogonal complements.
- The uniqueness and characterization of least-squares projections.
- Diagonalization of matrices—both general and orthogonal (with an emphasis on symmetric and positive-definite matrices).

Why study projections so carefully? The answer is twofold: in statistics, projections arise naturally in regression as estimators; mathematically, projections expose the interplay between operators and geometry in vector spaces.

2 Projection Matrices and Their Properties

2.1 Notation and Setup

Let $X \in \mathbb{R}^{n \times m}$ be a matrix with *linearly independent columns*. This condition is crucial: it guarantees that $X^T X$ is invertible (full rank), which in turn ensures that the projection matrix is well-defined.

Definition 2.1 (Projection Matrix onto the Column Space). Let $X \in \mathbb{R}^{n \times m}$ with linearly independent columns. The *projection matrix* onto the column space of X (denoted by $\text{Im}(X)$) is defined as

$$P_X := X(X^T X)^{-1} X^T \in \mathbb{R}^{n \times n}.$$

Remark 2.2. It is always prudent to check the dimensions: X is $n \times m$, so X^T is $m \times n$, and $(X^T X)^{-1}$ is $m \times m$ (since $X^T X$ is square). Thus, P_X is $n \times n$, mapping \mathbb{R}^n to itself.

2.2 Key Properties of Projection Matrices

Let us proceed to understand the core properties of P_X :

Property 1: P_X is symmetric: $P_X^T = P_X$.

Property 2: P_X is idempotent: $P_X^2 = P_X$.

Property 3: P_X acts as the identity on columns of X : $P_X X = X$.

Property 4: $X^T(I - P_X) = 0$: The columns of X are orthogonal to the residuals after projection.

Property 5: Image containment: For any $v \in \mathbb{R}^n$, $P_X v \in \text{Im}(X)$.

Property 6: Full rank, square case: If $n = m$ and X is invertible, then $P_X = I_n$.

Property 7: Orthogonality of residuals: For all $v \in \mathbb{R}^n$, $(I - P_X)v \in \text{Im}(X)^\perp$.

Property 8: Projection fixes image vectors: If $w \in \text{Im}(X)$, then $P_X w = w$.

Property 9: Projection annihilates orthogonal complement: If $z \in \text{Im}(X)^\perp$, then $P_X z = 0$.

Geometric Intuition: The projection operator P_X “drops” any vector $v \in \mathbb{R}^n$ perpendicularly onto the subspace $\text{Im}(X)$. The difference $v - P_X v$ (called the *residual*) belongs to the orthogonal complement. If v already lies in $\text{Im}(X)$, projecting it changes nothing; if v is orthogonal, the projection is zero.

Example 2.3 (Action of P_X on the basis vectors). Suppose X has columns X_1, \dots, X_m . Then $P_X X_j = X_j$ for any $j = 1, \dots, m$. To see this, recall that P_X projects onto their span—any X_j is already there.

Example 2.4 (Effect in the square, full-rank case). If X is $n \times n$ and invertible, then the columns of X form a basis for \mathbb{R}^n ; so the projection matrix P_X is the identity:

$$P_X = X(X^T X)^{-1} X^T = I_n$$

because projecting onto the full space leaves any vector unchanged.

3 Projections and Decomposition of Vectors

3.1 Decomposition Theorem

The most fundamental application of orthogonal projections is that *every vector can be uniquely decomposed* as a sum of a vector in the subspace and a vector in its orthogonal complement.

Theorem 3.1 (Orthogonal Decomposition). *Let M be a subspace of \mathbb{R}^n . For every $v \in \mathbb{R}^n$, there exists a unique decomposition*

$$v = w + z$$

where $w \in M$ and $z \in M^\perp$. In particular, $w = P_M v$.

Proof. (Existence) Let $w = P_M v = X(X^T X)^{-1} X^T v$ for any matrix X with columns forming a basis for M ; define $z = v - w$. Property 5 ensures $w \in M$, and Properties 7 and 9 give $z \in M^\perp$.

(Uniqueness) Suppose also $v = w' + z'$ with $w' \in M$, $z' \in M^\perp$. Then $(w - w') = (z' - z)$ lies in $M \cap M^\perp = \{0\}$. Thus $w = w'$ and $z = z'$. \square

Example 3.2 (Orthogonal decomposition in \mathbb{R}^2). Let $M = \text{span}\{e_1\}$, the x -axis. For $v = (a, b)^T$, the projection is $w = (a, 0)^T$ and $z = (0, b)^T$; clearly $v = w + z$, $w \in M$, $z \in M^\perp$.

3.2 Least Squares Characterization

A powerful property of orthogonal projection is its solution to the *closest vector problem*:

Theorem 3.3 (Best Approximation (Least Squares)). *Let M be a subspace of \mathbb{R}^n . For any $v \in \mathbb{R}^n$,*

$$w := P_M v = \operatorname{argmin}_{y \in M} \|v - y\|^2.$$

Proof Sketch & Intuition. By the previous theorem, $v = w + z$, with z orthogonal to M . For any $y \in M$,

$$\|v - y\|^2 = \|w + z - y\|^2 = \|z\|^2 + \|w - y\|^2$$

by the Pythagorean theorem. This is minimized when $y = w$. \square

4 Advanced Properties and Uniqueness of the Projection Map

4.1 Dependence on the Subspace Only

Key Insight: The projection matrix onto a subspace M depends solely on M *itself* and not on the specific matrix X used to represent it (as long as X 's columns form a basis for M).

Theorem 4.1 (Projection Matrix Uniqueness). *Let $X, Z \in \mathbb{R}^{n \times m}$ both have linearly independent columns and span the same subspace: $\operatorname{Im}(X) = \operatorname{Im}(Z) = M$. Then $P_X = P_Z$.*

Proof. Since both X and Z span M , there exists an invertible $m \times m$ matrix H such that $Z = XH$. Then,

$$\begin{aligned} P_Z &= Z(Z^T Z)^{-1} Z^T \\ &= XH(H^T X^T X H)^{-1} H^T X^T \\ &= X(X^T X)^{-1} X^T = P_X \end{aligned}$$

where we used the invertibility of H . \square

Remark 4.2. This result justifies defining P_M as the projection operator onto M (well-defined regardless of basis).

5 Intersections, Complements, and Further Identities

Theorem 5.1 (Projection onto an Intersection). *Let L, M be subspaces of \mathbb{R}^n . Then*

$$P_L P_M = P_{L \cap M}$$

That is, projecting onto M , then onto L , is the same as projecting onto the intersection $L \cap M$.

Intuition. Each projection successively reduces the vector to reside in both subspaces; by uniqueness of the orthogonal decomposition, this is the same as a one-step orthogonal projection onto their intersection. \square

Theorem 5.2 (Projections onto Orthogonal Complements). *For any subspace $M \subseteq \mathbb{R}^n$,*

$$I - P_M = P_{M^\perp}$$

Idea. The residual left after orthogonal projection onto M is exactly the orthogonal projection onto M^\perp . \square

6 Characterization of Projection Matrices

Theorem 6.1 (Symmetric Idempotents are Projections). *Let $Q \in \mathbb{R}^{n \times n}$ be a symmetric matrix ($Q^T = Q$) satisfying $Q^2 = Q$ (idempotent). Then Q is the orthogonal projection operator onto $W = \text{Im}(Q)$, i.e.*

$$Q = P_W.$$

Remark 6.2. The two essential axioms for a projection matrix are:

- **Symmetry:** $Q^T = Q$
- **Idempotency:** $Q^2 = Q$

Any operator with these properties is an orthogonal projection onto its range.

Example 6.3 (Non-example). The matrix $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is idempotent ($A^2 = A$), but not symmetric. Hence, not an orthogonal projection.

7 Diagonalization and Positive (Semi-)Definite Matrices

7.1 Motivation & Connection

Diagonalization is a powerful technique that “simplifies” matrices: in the right basis, a matrix can often be represented as diagonal, turning complicated linear operations into simple scaling along orthogonal directions (eigenvectors). This underlies much of matrix computation and theory.

Definition 7.1 (Diagonalizability). A square matrix $A \in \mathbb{R}^{n \times n}$ is *diagonalizable* over \mathbb{R} if there exists an invertible P and a diagonal matrix D such that

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

7.2 Interpretation and Mechanics

The transition matrix P represents a change of basis. Diagonalization tells us that, for a suitable choice of basis, A acts by stretching each coordinate by its respective *eigenvalue*.

Example 7.2 (Effect of Diagonal Matrices). If $D = \text{diag}(d_1, \dots, d_n)$, then D acts on $v = (v_1, \dots, v_n)^T$ via $Dv = (d_1v_1, \dots, d_nv_n)^T$.

Definition 7.3 (Orthogonal Diagonalization). $A \in \mathbb{R}^{n \times n}$ is called *orthogonally diagonalizable* if there exists an orthogonal matrix U ($U^T U = I$) and a diagonal D such that $A = UDU^T$.

Theorem 7.4 (Spectral Theorem (Finite-Dimensional Real Case)). *Every real symmetric matrix $A \in \mathbb{R}^{n \times n}$ is orthogonally diagonalizable: there exists an orthogonal U and real diagonal D such that $A = UDU^T$.*

Remark 7.5. Eigenvectors corresponding to distinct eigenvalues of a symmetric matrix are orthogonal.

7.3 Positive (Semi-)Definite Matrices

Definition 7.6 (Positive (Semi-)Definite). A real symmetric matrix A is **positive definite** if for all $v \in \mathbb{R}^n$,

$$v^T A v > 0$$

and **positive semi-definite** if

$$v^T A v \geq 0 \quad \text{for all } v \in \mathbb{R}^n.$$

Theorem 7.7 (Eigenvalues and Positive (Semi-)Definiteness). *Let A be a real symmetric $n \times n$ matrix. Then:*

- A is positive definite \iff all its eigenvalues are > 0 .
- A is positive semi-definite \iff all its eigenvalues are ≥ 0 .

Example 7.8 (Matrix Passing/Failing Positive Definiteness). $A = \text{diag}(1, 1, \dots, 1, -1)$: since -1 is an eigenvalue, A is *not* positive definite. For vectors aligned with the last coordinate, $v^T A v < 0$.

Remark 7.9. Positive definiteness is only defined for symmetric matrices.

7.4 Square Roots and Cholesky-like Decompositions

Theorem 7.10. *Let $A \in \mathbb{R}^{n \times n}$ be symmetric and positive definite. Then:*

1. *There exists $B \in \mathbb{R}^{n \times n}$ such that $B^2 = A$ (a symmetric square root).*

2. There exists $C \in \mathbb{R}^{n \times n}$ such that $C^T C = A$ (Cholesky factorization).

Sketch. Diagonalize: $A = UDU^T$ with D diagonal, U orthogonal; D has strictly positive diagonal entries.

- $B = UD^{1/2}U^T$ satisfies $B^2 = UDU^T = A$.
- $C = UD^{1/2}$ gives $C^T C = UD^{1/2}U^T UD^{1/2} = UDU^T = A$.

□

8 Concluding Remarks

These foundational results on projection matrices, orthogonal decompositions, and diagonalization provide essential tools for understanding the geometry of least-squares problems, statistical modeling, and much more. While technical at times, investing effort now will pay major dividends throughout your mathematical and applied studies.

Administrative Note

- **Upcoming:** We will soon transition to analyzing the linear statistical model in earnest, leveraging the machinery developed here. Please review the key properties above and bring questions to the next meeting.
- **Clarifications and Project Discussions:** Those interested in further projects, or encountering scheduling conflicts (seminars, TA meetings, etc.), reach out early—we will accommodate as best as possible.