

Course Announcements

A few notes regarding the course structure and recent exercises:

- **Exercise Difficulty:** I understand that the initial exercises (Exercise 0, 1, and 2) might feel challenging, perhaps even very challenging. This is somewhat expected. The beginning of this course involves recalling concepts from linear algebra (sometimes needing to learn them more deeply) and integrating them with new ideas. The emphasis here is less on purely technical computation than you might find in a first linear algebra course.
- **Hold On!** The goal right now is to practice these foundational concepts as much as possible. Based on past experience, things tend to settle down and become more manageable after the first few weeks (around the Passover break). The core linear algebra tools, once mastered, will be used repeatedly.
- **Office Hours:** My office hours are on **Wednesdays at 9:00 AM**. This time should fit into the standard schedule. Please feel free to come – it's a great opportunity to ask questions and discuss the material. Attendance seemed low recently, so I encourage you to take advantage of this resource.
- **Exercise Solution Recording:** I plan to upload a video recording where I work through parts of one of the recent exercises. Hopefully, this will provide additional clarification.
- **Extensions:** If you find yourself needing an extension on an assignment, please reach out. We can discuss arrangements.
- **Calculator Recommendation:** For those interested, the calculator available at the Akademon store can be quite helpful for matrix operations (up to 4×4 , including inversion) and potentially for probability calculations later. While not mandatory, it can speed up computations, allowing you to focus on the concepts. (Note: Understanding the underlying algorithms is still crucial!)
- **Textbook Reference:** Some notation or approaches might differ slightly if you are referencing Prof. Pavel's book (e.g., the direction of change-of-basis matrices). We will stick to the conventions presented in these notes for consistency within this course.
- **Homework 3 Mention:** Question 3 from the homework assignment relates to some of the concepts we will discuss regarding matrix representations.

1 Review of Linear Transformations

Let's begin by recalling the fundamental concept of a linear transformation.

Definition 1.1 (Linear Transformation). Let V and W be vector spaces (or subspaces). A function $T : V \rightarrow W$ is called a **linear transformation** if for all vectors $\mathbf{u}, \mathbf{v} \in V$ and all scalars k , the following two conditions hold:

1. $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ (Additivity)
2. $T(k\mathbf{u}) = kT(\mathbf{u})$ (Homogeneity)

Equivalently, these two conditions can be combined into a single condition: for all $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ and all scalars k_1, \dots, k_n ,

$$T\left(\sum_{i=1}^n k_i \mathbf{v}_i\right) = \sum_{i=1}^n k_i T(\mathbf{v}_i).$$

This means that linear transformations "respect" the vector space operations of addition and scalar multiplication. Think about the expectation operator in probability: $E[aX + bY] = aE[X] + bE[Y]$. Expectation is a linear operator!

Proposition 1.2. *If $T : V \rightarrow W$ is a linear transformation, then $T(\mathbf{0}_V) = \mathbf{0}_W$, where $\mathbf{0}_V$ and $\mathbf{0}_W$ are the zero vectors in V and W , respectively.*

Proof. We can use the homogeneity property with the scalar $k = 0$:

$$\begin{aligned} T(\mathbf{0}_V) &= T(0 \cdot \mathbf{v}) \quad (\text{for any } \mathbf{v} \in V) \\ &= 0 \cdot T(\mathbf{v}) \quad (\text{by homogeneity}) \\ &= \mathbf{0}_W \quad (\text{scalar zero times any vector is the zero vector}) \end{aligned}$$

Alternatively, using additivity: $T(\mathbf{0}_V) = T(\mathbf{0}_V + \mathbf{0}_V) = T(\mathbf{0}_V) + T(\mathbf{0}_V)$. Subtracting $T(\mathbf{0}_V)$ from both sides yields $T(\mathbf{0}_V) = \mathbf{0}_W$. \square

Two crucial subspaces associated with any linear transformation are its kernel and image.

Definition 1.3 (Kernel and Image). Let $T : V \rightarrow W$ be a linear transformation.

- The **Kernel** (or Null Space) of T , denoted $\text{Ker}(T)$, is the set of all vectors in the domain V that map to the zero vector in the codomain W :

$$\text{Ker}(T) = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}_W\}$$

$\text{Ker}(T)$ is a subspace of V .

- The **Image** (or Range) of T , denoted $\text{Im}(T)$, is the set of all vectors in the codomain W that are the image of at least one vector in V :

$$\text{Im}(T) = \{\mathbf{w} \in W \mid \text{there exists } \mathbf{v} \in V \text{ such that } T(\mathbf{v}) = \mathbf{w}\}$$

Equivalently, $\text{Im}(T) = \{T(\mathbf{v}) \mid \mathbf{v} \in V\}$. $\text{Im}(T)$ is a subspace of W .

Remark 1.4. It's important to note that $\text{Ker}(T)$ is a subset (and subspace) of the domain V , while $\text{Im}(T)$ is a subset (and subspace) of the codomain W . They live in different spaces in general!

Example 1.5 (Kernel vs. Image Location). Consider the linear transformation $T : \mathbb{R}^2 \rightarrow \mathbb{R}^1$ defined by $T(x_1, x_2) = x_1 + x_2$.

- The domain is $V = \mathbb{R}^2$.
- The codomain is $W = \mathbb{R}^1$ (which we often just write as \mathbb{R}).
- The kernel consists of all vectors (x_1, x_2) such that $T(x_1, x_2) = 0$. This means $x_1 + x_2 = 0$, or $x_2 = -x_1$. So,

$$\text{Ker}(T) = \{(x_1, -x_1) \mid x_1 \in \mathbb{R}\} = \text{span}\{(1, -1)\}$$

This is a one-dimensional subspace of \mathbb{R}^2 (a line through the origin). Note that the zero vector $(0, 0)$ is in $\text{Ker}(T)$, as required.

- The image consists of all outputs. Since for any $w \in \mathbb{R}$, we can choose $(w, 0)$ (for example) and have $T(w, 0) = w + 0 = w$, the image is all of \mathbb{R}^1 .

$$\text{Im}(T) = \mathbb{R}^1$$

This is a subspace of the codomain \mathbb{R}^1 .

Clearly, $\text{Ker}(T)$ (a subspace of \mathbb{R}^2) cannot be contained within $\text{Im}(T)$ (a subspace of \mathbb{R}^1), and vice versa. They live in different ambient spaces.

(Homework exercise suggested: Prove that $\text{Ker}(T)$ and $\text{Im}(T)$ are indeed subspaces of V and W , respectively.)

2 Coordinate Vectors

While linear transformations provide an abstract view, we often work with vectors and transformations concretely using coordinates. This requires choosing a basis.

2.1 Matrix Multiplication and Column Combinations

First, let's recall a fundamental interpretation of matrix-vector multiplication.

Proposition 2.1. *Let A be an $n \times m$ matrix with columns $\mathbf{a}_1, \dots, \mathbf{a}_m \in \mathbb{R}^n$. Let \mathbf{u} be a vector in \mathbb{R}^m with components u_1, \dots, u_m . Then the product $\mathbf{v} = A\mathbf{u}$ is the linear combination of the columns of A with coefficients given by the components of \mathbf{u} :*

$$\mathbf{v} = A\mathbf{u} = \sum_{j=1}^m u_j \mathbf{a}_j = u_1 \mathbf{a}_1 + u_2 \mathbf{a}_2 + \dots + u_m \mathbf{a}_m$$

Proof. Consider the i -th component of the product $A\mathbf{u}$, denoted $(A\mathbf{u})_i$. By the definition of matrix multiplication,

$$(A\mathbf{u})_i = \sum_{j=1}^m A_{ij} u_j$$

Now consider the i -th component of the sum $\sum_{j=1}^m u_j \mathbf{a}_j$:

$$\left(\sum_{j=1}^m u_j \mathbf{a}_j \right)_i = \sum_{j=1}^m (u_j \mathbf{a}_j)_i = \sum_{j=1}^m u_j (\mathbf{a}_j)_i$$

Since $(\mathbf{a}_j)_i$ is just the element in the i -th row and j -th column of A , which is A_{ij} , we have:

$$\left(\sum_{j=1}^m u_j \mathbf{a}_j \right)_i = \sum_{j=1}^m u_j A_{ij}$$

Comparing the two expressions, we see $(A\mathbf{u})_i = \left(\sum_{j=1}^m u_j \mathbf{a}_j \right)_i$ for all $i = 1, \dots, n$. Therefore, the vectors are equal. \square

This perspective is crucial: multiplying a matrix A by a vector \mathbf{u} combines the columns of A using the weights from \mathbf{u} .

2.2 Definition and Calculation

Definition 2.2 (Coordinate Vector). Let V be a finite-dimensional vector space, and let $B = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ be an **ordered basis** for V . This means the order in which the basis vectors are listed matters. For any vector $\mathbf{v} \in V$, there exists a unique set of scalars k_1, k_2, \dots, k_n such that

$$\mathbf{v} = k_1 \mathbf{b}_1 + k_2 \mathbf{b}_2 + \dots + k_n \mathbf{b}_n = \sum_{i=1}^n k_i \mathbf{b}_i$$

The **coordinate vector** of \mathbf{v} relative to the basis B , denoted $[\mathbf{v}]_B$, is the vector in \mathbb{R}^n formed by these unique scalars:

$$[\mathbf{v}]_B = \begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{pmatrix} \in \mathbb{R}^n$$

The scalars k_i are called the coordinates of \mathbf{v} with respect to B .

How do we find these coordinates? Let's focus on the familiar case where $V = \mathbb{R}^n$. We can form a matrix B whose columns are the ordered basis vectors from the basis B :

$$B = \begin{pmatrix} | & | & \dots & | \\ \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_n \\ | & | & \dots & | \end{pmatrix}$$

Then the equation $\mathbf{v} = \sum_{i=1}^n k_i \mathbf{b}_i$ can be rewritten using our matrix-column combination insight as:

$$\mathbf{v} = B \begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{pmatrix} = B[\mathbf{v}]_B$$

If B is a basis for \mathbb{R}^n , the matrix B is an $n \times n$ invertible matrix. Therefore, we can solve for the coordinate vector:

$$[\mathbf{v}]_B = B^{-1} \mathbf{v}$$

This gives us a direct method to compute the coordinate vector if we know the basis matrix B and the vector \mathbf{v} (in standard coordinates).

Example 2.3 (Finding Coordinates). Let $V = \mathbb{R}^2$. Consider the ordered basis $B = \{\mathbf{b}_1, \mathbf{b}_2\} = \{(1, 1), (1, 0)\}$. Let the vector be $\mathbf{v} = (2, 5)$. Find $[\mathbf{v}]_B$.

Method 1: Direct Solution We are looking for scalars k_1, k_2 such that $\mathbf{v} = k_1 \mathbf{b}_1 + k_2 \mathbf{b}_2$.

$$\begin{pmatrix} 2 \\ 5 \end{pmatrix} = k_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + k_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} k_1 + k_2 \\ k_1 \end{pmatrix}$$

This gives us a system of linear equations:

$$\begin{aligned} k_1 + k_2 &= 2 \\ k_1 &= 5 \end{aligned}$$

From the second equation, $k_1 = 5$. Substituting into the first equation gives $5 + k_2 = 2$, so $k_2 = -3$. Therefore, the coordinate vector is:

$$[\mathbf{v}]_B = \begin{pmatrix} 5 \\ -3 \end{pmatrix}$$

Let's check: $5 \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 3 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 5 \\ 5 \end{pmatrix} - \begin{pmatrix} 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$, which is \mathbf{v} . Correct.

Method 2: Matrix Inverse Form the basis matrix B :

$$B = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

We need to find B^{-1} . The determinant is $\det(B) = (1)(0) - (1)(1) = -1$. The inverse is $B^{-1} = \frac{1}{-1} \begin{pmatrix} 0 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}$. Now compute $[\mathbf{v}]_B = B^{-1}\mathbf{v}$:

$$[\mathbf{v}]_B = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 \\ 5 \end{pmatrix} = \begin{pmatrix} (0)(2) + (1)(5) \\ (1)(2) + (-1)(5) \end{pmatrix} = \begin{pmatrix} 5 \\ 2 - 5 \end{pmatrix} = \begin{pmatrix} 5 \\ -3 \end{pmatrix}$$

Both methods yield the same result, as expected.

Remark 2.4 (Standard Basis). Consider the standard basis for \mathbb{R}^n , $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, where \mathbf{e}_i has a 1 in the i -th position and 0s elsewhere. The matrix formed by these basis vectors is the identity matrix I . Any vector $\mathbf{v} = (v_1, \dots, v_n)^T$ can be written as $\mathbf{v} = v_1\mathbf{e}_1 + \dots + v_n\mathbf{e}_n$. Therefore, the coordinate vector of \mathbf{v} relative to the standard basis is simply the vector \mathbf{v} itself:

$$[\mathbf{v}]_E = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \mathbf{v}$$

Using the matrix formula: $[\mathbf{v}]_E = E^{-1}\mathbf{v} = I^{-1}\mathbf{v} = I\mathbf{v} = \mathbf{v}$. The standard coordinates we usually use are just coordinates with respect to the standard basis.

2.3 Geometric Intuition

What does changing basis mean geometrically? Think of a basis as defining a coordinate system. The standard basis $E = \{(1, 0), (0, 1)\}$ in \mathbb{R}^2 defines the usual Cartesian x - y axes. A different basis defines a different set of axes (which might not be orthogonal).

Consider the previous example: $\mathbf{v} = (2, 5)$ and basis $B = \{(1, 1), (1, 0)\}$.

- **Standard Coordinates:** $[\mathbf{v}]_E = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$. This means we reach the point $(2, 5)$ by moving 2 units along the standard $\mathbf{e}_1 = (1, 0)$ axis and 5 units along the standard $\mathbf{e}_2 = (0, 1)$ axis.
- **B -Coordinates:** $[\mathbf{v}]_B = \begin{pmatrix} 5 \\ -3 \end{pmatrix}$. This means we reach the *same point* $(2, 5)$ by moving 5 units along the $\mathbf{b}_1 = (1, 1)$ axis (think of a skewed axis in the direction $(1, 1)$) and -3 units along the $\mathbf{b}_2 = (1, 0)$ axis (which happens to be the standard x -axis).

Imagine drawing the standard grid versus a grid defined by the vectors \mathbf{b}_1 and \mathbf{b}_2 . The point \mathbf{v} exists independently of the grid. The coordinate vector simply describes how to reach that point using the "steps" defined by the chosen basis vectors. Changing the basis is like changing your perspective or your measurement system, but the object being measured (\mathbf{v}) remains the same.

3 Change of Basis

Since a vector can be represented in different bases, we need a way to convert coordinates from one basis to another.

Definition 3.1 (Change of Basis Matrix). Let $B = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ and $C = \{\mathbf{c}_1, \dots, \mathbf{c}_n\}$ be two ordered bases for a vector space V . The **change of basis matrix from B to C** , denoted $P_{C \leftarrow B}$, is the unique $n \times n$ matrix such that for any vector $\mathbf{v} \in V$:

$$[\mathbf{v}]_C = P_{C \leftarrow B}[\mathbf{v}]_B$$

This matrix transforms B -coordinates into C -coordinates.

How do we find $P_{C \leftarrow B}$? Let B and C also denote the matrices whose columns are the basis vectors of the respective bases (when $V = \mathbb{R}^n$). We know:

$$\mathbf{v} = B[\mathbf{v}]_B \quad \text{and} \quad \mathbf{v} = C[\mathbf{v}]_C$$

Therefore, $C[\mathbf{v}]_C = B[\mathbf{v}]_B$. Since C is invertible (as its columns form a basis), we can multiply by C^{-1} on the left:

$$[\mathbf{v}]_C = C^{-1}(B[\mathbf{v}]_B) = (C^{-1}B)[\mathbf{v}]_B$$

Comparing this with the definition $[\mathbf{v}]_C = P_{C \leftarrow B}[\mathbf{v}]_B$, we see that:

$$P_{C \leftarrow B} = C^{-1}B$$

Proposition 3.2 (Properties of Change of Basis Matrices). *Let B, C, D be bases for V , and let E be the standard basis for \mathbb{R}^n .*

1. $P_{B \leftarrow C} = (P_{C \leftarrow B})^{-1}$. (Changing back is the inverse operation).
2. $P_{D \leftarrow B} = P_{D \leftarrow C}P_{C \leftarrow B}$. (Changes can be composed).
3. $P_{E \leftarrow B} = B$. (The matrix B transforms B -coordinates to standard coordinates).
4. $P_{B \leftarrow E} = B^{-1}$. (The matrix B^{-1} transforms standard coordinates to B -coordinates).

Proof. 1. From the definition, $[\mathbf{v}]_B = P_{B \leftarrow C}[\mathbf{v}]_C$. Substituting $[\mathbf{v}]_C = P_{C \leftarrow B}[\mathbf{v}]_B$ gives $[\mathbf{v}]_B = P_{B \leftarrow C}(P_{C \leftarrow B}[\mathbf{v}]_B)$. Since this holds for all \mathbf{v} , the matrix product must be the identity: $P_{B \leftarrow C}P_{C \leftarrow B} = I$. Thus, they are inverses. Using the formula: $P_{B \leftarrow C} = B^{-1}C = (C^{-1}B)^{-1} = (P_{C \leftarrow B})^{-1}$. 2. $[\mathbf{v}]_D = P_{D \leftarrow C}[\mathbf{v}]_C = P_{D \leftarrow C}(P_{C \leftarrow B}[\mathbf{v}]_B) = (P_{D \leftarrow C}P_{C \leftarrow B})[\mathbf{v}]_B$. By uniqueness, $P_{D \leftarrow B} = P_{D \leftarrow C}P_{C \leftarrow B}$. 3. From $\mathbf{v} = B[\mathbf{v}]_B$ and $\mathbf{v} = [\mathbf{v}]_E$, we have $[\mathbf{v}]_E = B[\mathbf{v}]_B$. By definition, $P_{E \leftarrow B} = B$. 4. From $[\mathbf{v}]_B = B^{-1}\mathbf{v}$ and $\mathbf{v} = [\mathbf{v}]_E$, we have $[\mathbf{v}]_B = B^{-1}[\mathbf{v}]_E$. By definition, $P_{B \leftarrow E} = B^{-1}$. \square

Remark 3.3 (Notation). The notation $P_{C \leftarrow B}$ emphasizes that the matrix takes coordinates relative to B (input) and produces coordinates relative to C (output). Some texts use the reverse notation. We will consistently use $P_{C \leftarrow B} = C^{-1}B$.

4 Matrix Representation of Linear Transformations

Just as we represent vectors using coordinates relative to a basis, we can represent linear transformations using matrices relative to bases for the domain and codomain.

Definition 4.1 (Matrix Representation). Let $T : V \rightarrow W$ be a linear transformation, where $\dim(V) = n$ and $\dim(W) = m$. Let $B = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be an ordered basis for V , and $C = \{\mathbf{c}_1, \dots, \mathbf{c}_m\}$ be an ordered basis for W . The **matrix representation of T with respect to bases B and C** , denoted $[T]_{C \leftarrow B}$, is the unique $m \times n$ matrix such that for any vector $\mathbf{v} \in V$:

$$[T(\mathbf{v})]_C = [T]_{C \leftarrow B}[\mathbf{v}]_B$$

This matrix takes the B -coordinates of a vector \mathbf{v} , multiplies them, and produces the C -coordinates of the transformed vector $T(\mathbf{v})$.

Remark 4.2. The j -th column of $[T]_{C \leftarrow B}$ is the C -coordinate vector of the image of the j -th basis vector of B , i.e., $[T(\mathbf{b}_j)]_C$. This provides a way to construct the matrix.

Example 4.3 (Standard Matrix Representation). Let $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be given by $F(v_1, v_2) = (v_1 + v_2, 3v_1 - v_2)$. Find the matrix representation of F with respect to the standard basis $E = \{\mathbf{e}_1, \mathbf{e}_2\}$ for both the domain and codomain, i.e., find $[F]_{E \leftarrow E}$ (often written just as $[F]_E$).

We need a matrix $A = [F]_E$ such that $[F(\mathbf{v})]_E = A[\mathbf{v}]_E$. Since coordinates relative to the standard basis are just the vectors themselves, this simplifies to $F(\mathbf{v}) = A\mathbf{v}$. We want $A \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_1 + v_2 \\ 3v_1 - v_2 \end{pmatrix}$. By inspection, or by finding the images of the basis vectors: $F(\mathbf{e}_1) = F(1, 0) = (1 + 0, 3(1) - 0) = (1, 3)$. So $[F(\mathbf{e}_1)]_E = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ (first column of A). $F(\mathbf{e}_2) = F(0, 1) = (0 + 1, 3(0) - 1) = (1, -1)$. So $[F(\mathbf{e}_2)]_E = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ (second column of A). Thus, the standard matrix representation is:

$$[F]_E = \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix}$$

This confirms that often, the "standard matrix" associated with a transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is precisely $[T]_{E \leftarrow E}$.

4.1 Relationship via Change of Basis (Similarity)

How does the matrix representation of a transformation $T : V \rightarrow V$ change if we change the basis from B to C ?

Theorem 4.4 (Change of Basis for Transformations). *Let $T : V \rightarrow V$ be a linear operator, and let B and C be two ordered bases for V . Let $P = P_{C \leftarrow B}$ be the change of basis matrix from B to C . Then the matrix representations of T relative to these bases are related by:*

$$[T]_C = P_{C \leftarrow B} [T]_B P_{B \leftarrow C}$$

or equivalently, using $P_{B \leftarrow C} = P_{C \leftarrow B}^{-1} = P^{-1}$:

$$[T]_C = P [T]_B P^{-1}$$

Matrices $[T]_C$ and $[T]_B$ are said to be **similar**.

Proof. Let $\mathbf{v} \in V$. We start with the definition for $[T]_C$:

$$[T(\mathbf{v})]_C = [T]_C [\mathbf{v}]_C$$

We also know $[T(\mathbf{v})]_C = P_{C \leftarrow B} [T(\mathbf{v})]_B$ and $[\mathbf{v}]_C = P_{C \leftarrow B} [\mathbf{v}]_B$. Substitute the second relation into the first equation:

$$[T(\mathbf{v})]_C = [T]_C (P_{C \leftarrow B} [\mathbf{v}]_B)$$

Now, let's work from the definition of $[T]_B$: $[T(\mathbf{v})]_B = [T]_B [\mathbf{v}]_B$. Apply $P_{C \leftarrow B}$ to both sides:

$$P_{C \leftarrow B} [T(\mathbf{v})]_B = P_{C \leftarrow B} ([T]_B [\mathbf{v}]_B)$$

The left side is $[T(\mathbf{v})]_C$. So we have two expressions for $[T(\mathbf{v})]_C$:

$$[T]_C P_{C \leftarrow B} [\mathbf{v}]_B = P_{C \leftarrow B} [T]_B [\mathbf{v}]_B$$

Since this must hold for all $[\mathbf{v}]_B$, the matrices must be equal:

$$[T]_C P_{C \leftarrow B} = P_{C \leftarrow B} [T]_B$$

Multiply by $P_{C \leftarrow B}^{-1} = P_{B \leftarrow C}$ on the right:

$$[T]_C = P_{C \leftarrow B} [T]_B P_{B \leftarrow C}$$

□

This theorem is fundamental. It tells us how the description of a linear transformation changes when we change our coordinate system. Similar matrices represent the *same* linear transformation, just viewed from different perspectives (bases).

A particularly useful case is when we want to find the representation $[T]_C$ in some basis C , and we already know the standard matrix $[T]_E$. Using $P_{C \leftarrow E} = C^{-1}$ and $P_{E \leftarrow C} = C$, the formula becomes:

$$[T]_C = P_{C \leftarrow E} [T]_E P_{E \leftarrow C} = C^{-1} [T]_E C$$

This is often easier than calculating $[T]_C$ directly from the definition.

5 Revisiting Kernel and Image (Matrix Perspective)

Now that we've connected linear transformations $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ to their standard matrix representations $A = [T]_E$, we can talk about the kernel and image of the matrix itself.

Definition 5.1 (Kernel and Image of a Matrix). Let A be an $m \times n$ matrix, representing the linear transformation $T_A(\mathbf{x}) = A\mathbf{x}$ from \mathbb{R}^n to \mathbb{R}^m .

- The **Kernel** (or Null Space) of A , denoted $\text{Ker}(A)$ or $\text{Null}(A)$, is the set of all vectors $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{0}$.

$$\text{Ker}(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\}$$

This is precisely the solution set to the homogeneous system $A\mathbf{x} = \mathbf{0}$, and it's a subspace of \mathbb{R}^n .

- The **Image** (or Column Space) of A , denoted $\text{Im}(A)$ or $\text{Col}(A)$, is the set of all vectors $\mathbf{b} \in \mathbb{R}^m$ such that $A\mathbf{x} = \mathbf{b}$ for some $\mathbf{x} \in \mathbb{R}^n$.

$$\text{Im}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$$

This is precisely the span of the columns of A , and it's a subspace of \mathbb{R}^m .

These definitions align perfectly with the definitions for the associated linear transformation T_A .

6 Diagonalization and Symmetric Matrices

A key goal in linear algebra is to find a basis in which the matrix representation of a linear transformation is particularly simple, ideally diagonal.

Definition 6.1 (Eigenvalues and Eigenvectors). Let A be an $n \times n$ matrix. A scalar λ is an **eigenvalue** of A if there exists a non-zero vector $\mathbf{v} \in \mathbb{R}^n$ such that

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

Any such non-zero vector \mathbf{v} is called an **eigenvector** of A corresponding to the eigenvalue λ . Eigenvectors are vectors whose direction is unchanged (only scaled) by the transformation A .

Definition 6.2 (Diagonalizable Matrix). An $n \times n$ matrix A is **diagonalizable** if it is similar to a diagonal matrix D . That is, if there exists an invertible matrix P and a diagonal matrix D such that

$$A = PDP^{-1} \quad (\text{or equivalently, } D = P^{-1}AP)$$

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

A particularly important class of matrices is guaranteed to be diagonalizable.

Definition 6.3 (Symmetric and Orthogonal Matrices). • An $n \times n$ matrix A is **symmetric** if $A^T = A$.

- An $n \times n$ matrix U is **orthogonal** if its columns form an orthonormal basis for \mathbb{R}^n . This is equivalent to the condition $U^T U = I$, which also implies $U U^T = I$, so $U^{-1} = U^T$.

Theorem 6.4 (Spectral Theorem for Symmetric Matrices). *Let A be an $n \times n$ symmetric matrix with real entries. Then:*

1. *All eigenvalues of A are real.*
2. *A is orthogonally diagonalizable. That is, there exists an orthogonal matrix U and a diagonal matrix Λ such that*

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

The columns of U are orthonormal eigenvectors of A , and the diagonal entries of Λ are the corresponding real eigenvalues.

This theorem is incredibly powerful. Symmetric matrices (which arise frequently in applications) always have this nice structure.

7 Application: Matrix Representation in Eigenbasis

Let's connect diagonalization with matrix representations. Consider a symmetric matrix $A = U \Lambda U^T$. What is the matrix representation of the transformation $T_A(\mathbf{v}) = A\mathbf{v}$ with respect to the eigenbasis $U = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$?

Let $B = U$ be the eigenbasis. We want to find $[A]_U = [T_A]_U$. We know the standard representation is $[A]_E = A$. Using the change of basis formula:

$$[A]_U = P_{U \leftarrow E} [A]_E P_{E \leftarrow U}$$

We know $P_{E \leftarrow U} = U$ and $P_{U \leftarrow E} = U^{-1}$. Since A is symmetric, U is orthogonal, so $U^{-1} = U^T$.

$$[A]_U = U^{-1} A U = U^T A U$$

Now substitute $A = U \Lambda U^T$:

$$[A]_U = U^T (U \Lambda U^T) U = (U^T U) \Lambda (U^T U) = I \Lambda I = \Lambda$$

This confirms a beautiful fact: The matrix representation of a transformation T_A with respect to a basis of its eigenvectors is the diagonal matrix Λ containing the corresponding eigenvalues. In the eigenbasis, the transformation simply scales the coordinate axes.

Example 7.1 (Representation of $A = X^T X$). Let $A = X^T X$ for some $n \times p$ matrix X . First, note that A is symmetric: $A^T = (X^T X)^T = X^T (X^T)^T = X^T X = A$. (Also, A is $p \times p$). By the Spectral Theorem, A is orthogonally diagonalizable as $A = U \Lambda U^T$. The previous derivation shows that the diagonal matrix Λ (containing the eigenvalues of A) is precisely the matrix representation of the transformation $T_A(\mathbf{v}) = A\mathbf{v}$ with respect to the orthonormal basis U (whose columns are the eigenvectors of A).

$$[A]_U = \Lambda$$

8 Geometric Interpretation of Multiplication by a Symmetric Matrix

What does the operation $\mathbf{v} \mapsto A\mathbf{v}$ look like geometrically when A is symmetric, using its diagonalization $A = U\Lambda U^T$? Let's break down the product $A\mathbf{v} = U\Lambda U^T\mathbf{v}$:

1. **Step 1:** $\mathbf{v}' = U^T\mathbf{v}$ Since U is orthogonal, $U^T = U^{-1}$. This matrix $U^T = P_{U \leftarrow E}$ changes coordinates from the standard basis E to the orthonormal eigenbasis U . Geometrically, multiplying by an orthogonal matrix corresponds to a **rotation** or a **reflection** (or a combination). It preserves lengths and angles:

- **Length Preservation:** $\|U^T\mathbf{v}\|^2 = (U^T\mathbf{v})^T(U^T\mathbf{v}) = \mathbf{v}^T U U^T \mathbf{v} = \mathbf{v}^T I \mathbf{v} = \mathbf{v}^T \mathbf{v} = \|\mathbf{v}\|^2$.

- **Angle Preservation:** The angle θ between $\mathbf{v}_1, \mathbf{v}_2$ satisfies $\cos \theta = \frac{\mathbf{v}_1^T \mathbf{v}_2}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|}$. The angle $\tilde{\theta}$ between $U^T \mathbf{v}_1, U^T \mathbf{v}_2$ satisfies $\cos \tilde{\theta} = \frac{(U^T \mathbf{v}_1)^T (U^T \mathbf{v}_2)}{\|U^T \mathbf{v}_1\| \|U^T \mathbf{v}_2\|} = \frac{\mathbf{v}_1^T U U^T \mathbf{v}_2}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|} = \frac{\mathbf{v}_1^T \mathbf{v}_2}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|} = \cos \theta$. (Angles or their negatives are preserved).

So, U^T rotates/reflects \mathbf{v} to express it in the coordinate system defined by the eigenvectors.

2. **Step 2:** $\mathbf{v}'' = \Lambda \mathbf{v}'$ Since Λ is diagonal with eigenvalues $\lambda_1, \dots, \lambda_n$ on the diagonal, this multiplication scales the coordinates of \mathbf{v}' (which are coordinates in the eigenbasis U) along the eigenbasis axes:

$$\mathbf{v}'' = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \begin{pmatrix} v'_1 \\ \vdots \\ v'_n \end{pmatrix} = \begin{pmatrix} \lambda_1 v'_1 \\ \vdots \\ \lambda_n v'_n \end{pmatrix}$$

This is an axis-aligned **stretching or compression** along the eigenvector directions.

3. **Step 3:** $A\mathbf{v} = U\mathbf{v}''$ The matrix $U = P_{E \leftarrow U}$ changes coordinates from the eigenbasis U back to the standard basis E . Since $U = (U^T)^{-1}$ is also orthogonal, this is another **rotation or reflection**, effectively reversing the coordinate change from Step 1.

In summary: Multiplication by a symmetric matrix $A = U\Lambda U^T$ can be interpreted geometrically as: 1. A rotation/reflection into the coordinate system defined by the orthonormal eigenvectors (U^T). 2. A scaling along these new coordinate axes, dictated by the eigenvalues (Λ). 3. A rotation/reflection back to the original coordinate system (U).

9 Concluding Remarks: Trace and Determinant

For a diagonalizable matrix $A = PDP^{-1}$, two important invariants are related to the eigenvalues on the diagonal of D :

- **Determinant:** $\det(A) = \det(PDP^{-1}) = \det(P) \det(D) \det(P^{-1}) = \det(P) \det(D) \frac{1}{\det(P)} = \det(D)$. Since D is diagonal with eigenvalues λ_i on the diagonal, $\det(D) = \prod_{i=1}^n \lambda_i$. Thus,

$$\det(A) = \prod_{i=1}^n \lambda_i$$

The determinant is the product of the eigenvalues.

- **Trace:** The trace of a square matrix is the sum of its diagonal elements, $\text{Tr}(A) = \sum_{i=1}^n A_{ii}$. Trace has the property $\text{Tr}(XY) = \text{Tr}(YX)$. Therefore, $\text{Tr}(A) = \text{Tr}(PDP^{-1}) = \text{Tr}((PD)P^{-1}) = \text{Tr}(P^{-1}(PD)) = \text{Tr}((P^{-1}P)D) = \text{Tr}(ID) = \text{Tr}(D)$. Since D is diagonal, $\text{Tr}(D) = \sum_{i=1}^n \lambda_i$. Thus,

$$\text{Tr}(A) = \sum_{i=1}^n \lambda_i$$

The trace is the sum of the eigenvalues.

These hold for any diagonalizable matrix, including symmetric matrices.