Johns Hopkins University

AS.110.201 Linear Algebra

Learning Notes

Jiuru Lyu

April 20, 2023

Contents

Preface

These are my personal notes for Johns Hopkins University AS.110.201 Linear Algebra course. I studied this course via Summer @ Hopkins in the summer of 2021.

As no prerequisite is required (only pre-calculus, basic algebra, and some simple knowledge from Calculus I), this course focuses on matrices. It includes systems of linear equations, basics of matrices, spaces and dimensions, determinants, eigenvalues, and singular value decomposition. The textbook used for this course is *Linear Algebra with Applications, 5th Edition* by Otto Bretscher. Another textbook by Gilbert Strang is also recommended: *Introduction to Linear Algebra, 5th Edition*.

Throughout this personal note, I use different formats to differentiate different contents, including definitions, theorems, proofs, examples, extensions, and remarks. To be more specific:

Definition 0.0.1 (Terminology). This is a **definition**.

Theorem 0.0.1 (Theorem Name). This is a **theorem**.

Example 0.0.1. This is an **example**.

Solution. This is the *answer* part of an **example**. □

Remark. This is a **remark** of a definition, theorem, example, or proof.

Proof. This is a **proof** of a theorem. ■

Extension. This is a **extension** of a theorem, proof, or example.

To better ace this course, it is recommended to do more questions than provided as examples under each section. Although each example is distinctive and representative, more questions and practice is still needed to deepen the understanding of this course. More than doing examples, using visualization tools to visualize some problems or concepts is also helpful in understanding the contents better. Videos made by **3Blue1Brown** are also recommended as a supplementary source of learning.

Even though I put efforts into making as few flaws as possible when encoding these learning notes, some errors may still exist in this note. If you find any, please contact me via email: lvjiuru@hotmail.com.

I hope you will find my notes helpful when learning Linear Algebra, a fundamental course for other Math and Computer Science courses.

Cheers,
Jiuru Lyu

1 Systems of Linear Equations

1.1 Solving Systems of Linear Equations

Definition 1.1.1 (Linear Equations). An equation in the unknowns x, y, z, ... is called **linear** if both sides of the equation are a sum of multiples of x, y, z, ..., plus an optional constant.

Example 1.1.1. Linear equations and nonlinear equations

$$\begin{cases} 3x + 4y = 2z \\ -x - z = 100 \end{cases} \text{are linear equations, but } \begin{cases} 3x + yz = 3 \\ \sin x - \cos y = 2 \end{cases} \text{are not.}$$

Definition 1.1.2 (System of Linear Equations). A **system** of linear equations is a collection of several linear equations.

Definition 1.1.3 (Solution of a System). A **solution** of a system of equations is a list of numbers x, y, z, ... that make all of the equations true simultaneously.

Definition 1.1.4 (Solution Set of a System). The **solution set** of a system of equations is the collection of all solutions.

Definition 1.1.5 (Solving a System). **Solving** the system means finding all solutions with formulas involving some number of parameters.

Definition 1.1.6 (Consistency and Inconsistency of a System). A system of equations is called **inconsistent** if it has no solutions. It is called **consistent** otherwise.

Example 1.1.2. An inconsistent system:

$$\begin{cases} x + 2y = 3 \\ x + 2y = -3 \end{cases} \text{has no solutions (the solution set is } \textit{empty}\text{).}$$

Thus, the system of equations is **inconsistent**.

Remark. A solution of equations in n variables is a list of n numbers.

Remark. We use $\mathbb{R}$ to denote the set of all real numbers.

Definition 1.1.7 ($\mathbb{R}^n$). Let n be a positive whole number. We define

$$\mathbb{R}^n = \text{all ordered } n\text{-tuples of real numbers } (x_1,\ x_2,\ x_3,\ ...,\ x_n)$$

An n-tuple of real number is called a **point** of $\mathbb{R}^n$

Example 1.1.3. Examples of $\mathbb{R}^n$

1. $\left[0, \frac{3}{2}, -\pi\right]$ and $(1, -2, 3)$ are points of $\mathbb{R}^3$
2. When $n = 1$, $\mathbb{R}^1 = \mathbb{R}$. Geometrically, this is the number line.
3. When $n = 2$, $\mathbb{R}^2$. It becomes the xy-plane.
4. When $n = 3$, $\mathbb{R}^3$. It is the *space* we live in.

Definition 1.1.8 (Line). A **line** is a ray that is *straight* and *infinite* in both directions.

Definition 1.1.9 (Plane). A **plane** is a flat sheet that is infinite in all directions.

Theorem 1.1.1. Generally, a single linear equation in n variables defines an $(n-1)$-plane in n-space.

Example 1.1.4. Examples of Lines and Planes.

1. **Lines**. For $x + y = 1$ (implicit equation), the **parametric form** is
$$(x, y) = (t, 1 - t) \text{ for any } t \in \mathbb{R}$$
We call t a **parameter** in this case.
2. For a system of two linear equations (as implicit equations in $\mathbb{R}^3$)
$$\begin{cases} x + y + z = 1 \\ x - z = 0 \end{cases},$$
the parametric form would be
$$(x, y, z) = (t, 1 - 2t, t)$$
3. **Planes**. For $x + y + z = 1$ (implicit equation), the **parametric form** is
$$(x, y, z) = (1 - t - w, t, w) \text{ for any } t, w \in \mathbb{R}$$

Theorem 1.1.2 (Elementary Operations). Since elementary operations are reversible, the solution set doesn't change:

1. Switch the order of the equation;
2. Scale the equation by a scale $c \neq 0$; (to reverse, divide equation by c)
3. Add a multiple of one equation to another. (to reverse, subtract)

1.2 Row Reduction

Theorem 1.2.1 (The Elimination Method). We can use the **elimination** method to combine the equations in various ways to eliminate as many variables as possible for each equation.

1. **Scaling**. We can multiply both sides of an equation by a nonzero number.
2. **Replacement**. We can add a multiple of one equation to another, replacing the second equation with the result.
3. **Swap**. We can swap two equations.

Definition 1.2.1 (Augmented Matrices and Row Operations). Augmented Matrix refers to the vertical line, which we draw to remind ourselves where the equals sign belongs.

Definition 1.2.2 (Matrix). A **matrix** is a grid of numbers without the vertical line.

Example 1.2.1. Augmented Matrix and Row Operations.

$$\left[\begin{array}{ccc:c} 1 & 2 & 3 & 6 \\ 2 & -3 & 2 & 14 \\ 3 & 1 & -1 & -2 \end{array}\right] \text{ is an augmented matrix.}$$

The three ways of manipulating our equations become row operations:

1. **Scaling**. multiply all entries in a row by a nonzero number.

$$\left[\begin{array}{ccc:c} 1 & 2 & 3 & 6 \\ 2 & -3 & 2 & 14 \\ 3 & 1 & -1 & -2 \end{array}\right] \xrightarrow{R_1 = R_1 \times -3} \left[\begin{array}{ccc:c} -3 & -6 & -9 & -18 \\ 2 & -3 & 2 & 14 \\ 3 & 1 & -1 & -2 \end{array}\right]$$

   Remark. Here, the notation R_1 simply means "the first row."

2. **Replacement**. add a multiple of one row to another, replacing the second row with the result.

$$\left[\begin{array}{ccc:c} 1 & 2 & 3 & 6 \\ 2 & -3 & 2 & 14 \\ 3 & 1 & -1 & -2 \end{array}\right] \xrightarrow{R_2 = R_2 - 2 \times R_1} \left[\begin{array}{ccc:c} 1 & 2 & 3 & 6 \\ 0 & -7 & -4 & 2 \\ 3 & 1 & -1 & -2 \end{array}\right]$$

3. **Swap**. Interchange two rows.

$$\left[\begin{array}{ccc:c} 1 & 2 & 3 & 6 \\ 2 & -3 & 2 & 14 \\ 3 & 1 & -1 & -2 \end{array}\right] \xrightarrow{R_1 \leftrightarrow R_3} \left[\begin{array}{ccc:c} -3 & -6 & -9 & -18 \\ 2 & -3 & 2 & 14 \\ 3 & 1 & -1 & -2 \end{array}\right]$$

Definition 1.2.3 (Row equivalent). Two matrices are called **row equivalent** if one can be obtained from the other by doing some number of row operations.

Definition 1.2.4 (Row Echelon Form (*ref*) of Matrix). A matrix is in **row echelon form** if:

1. All zero rows are at the bottom.
2. The first nonzero entry of a row is to the *right* of the first nonzero entry of the row above.
3. Below the first nonzero entry of a row, all entries are zero.

Example 1.2.2. General *ref* of matrices.

$$\left[\begin{array}{cccc:c} \color{red}{a} & b & b & b & b \\ 0 & \color{red}{a} & b & b & b \\ 0 & 0 & 0 & \color{red}{a} & b \\ 0 & 0 & 0 & 0 & 0 \end{array}\right],$$

where $b =$ is any number, and $\color{red}{a} =$ is any nonzero number.

Definition 1.2.5 (Pivot). A ***pivot*** is the first nonzero entry of a row of a matrix in row echelon form.

Definition 1.2.6 (Reduced Row Echelon Form (*rref*) of a Matrix). A matrix is in **reduced row echelon form** if it is in row echelon form, and in addition:

4. Each pivot is equal to 1.
5. Each pivot is the only nonzero entry in its column.

Example 1.2.3. Genderal *rref* of matrices

$$\left[\begin{array}{cccc:c} \color{red}{1} & 0 & b & 0 & b \\ 0 & \color{red}{1} & b & 0 & b \\ 0 & 0 & 0 & \color{red}{1} & b \\ 0 & 0 & 0 & 0 & 0 \end{array}\right],$$

where $b =$ is any number, $\color{red}{1} =$ pivot

$$\left[\begin{array}{ccc:c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 3 \end{array}\right] \xrightarrow{\textit{becomes}} \begin{cases} x = 1 \\ y = -2 \\ z = 3 \end{cases}$$

Theorem 1.2.2. Every matrix is row equivalent to one and only one matrix in reduced row echelon form.

Row reduction or ***Gaussian elimination*** demonstrates that every matrix is row equivalent to a least one matrix in reduced row echelon form.

1. Swap the 1$^{\text{st}}$ row with a lower one, so a leftmost nonzero entry is in the 1st row (if necessary).
2. Scale the 1$^{\text{st}}$ row so that its first nonzero entry equals 1.
3. Use row replacement, so all entries below this 1 are 0.
4. Swap the 2$^{\text{nd}}$ row with a lower one so that the leftmost nonzero entry is in the 2nd row.
5. Scale the 2$^{\text{nd}}$ row so that its first nonzero entry equals 1.
6. Use row replacement, so all entries below this 1 are 0.
7. Swap the 3$^{\text{rd}}$ row with a lower one so that the leftmost nonzero entry is in the 3rd row. etc.
8. Use row replacement to clear all entries above the pivots, starting with the last pivot.

Definition 1.2.7 (Pivot Position). A **pivot position** of a matrix is an entry that is a pivot of a row echelon form of that matrix.

Definition 1.2.8 (Pivot Column). A **pivot column** of a matrix is a column that contains a pivot position.

Theorem 1.2.3 (The Row Echelon Form of an Inconsistent System). An augmented matrix corresponds to an inconsistent system of equations if and only if (*iff*) the last column (i.e., the augmented column) is a pivot column.

1.3 Parametric Form

Definition 1.3.1 (Free Variable). Consider a consistent system of equations in the variables x_1, x_2,..., x_n. Let A be a row echelon form of the augmented matrix for this system. We say that x_i is a **free variable** if its corresponding column in A is not a pivot column.

Example 1.3.1. Example of free variables.

In the matrix $\left[\begin{array}{ccc:c} 1 & 0 & 5 & 1 \\ 0 & 1 & 2 & -2 \end{array}\right]$, the variable z is the free variable.

Definition 1.3.2 (Implicit Equations). The line is defined implicitly as the simultaneous solutions to those equations.

Definition 1.3.3 (Parameterized Equations). A **parameterized equation** is an expression that produces all points of the line in terms of one parameter.

Example 1.3.2. Example of implicit equations. $\begin{cases} 2x + y + 12z = 1 \\ x + 2y + 9z = -1 \end{cases}$ is an example of implicit equations in $\mathbb{R}^3$. $\begin{cases} x = 1 - 5z \\ y = 1 - 2z \end{cases}$ can be written as $(x, y, z,) = (1 - 5z, 1 - 2z, z),\ z \in \mathbb{R}$, which is a parameterized equation.

Remark. One should think of a system of equations as an implicit equation for its solution set and of the parametric form as the parameterized equation for the same set. The parametric form is much more explicit: it gives a concrete recipe for producing all solutions.

Theorem 1.3.1 (Number of Solutions). Systems of equations can have different numbers of solutions.

1. **The last column is a pivot column**. In this case, the system is inconsistent. It has zero solutions.

2. **Every column except the last column is a pivot column**. The system has a unique solution.

3. **The last column is not a pivot column, and some other column is not a pivot column either**. The system has many solutions corresponding to the infinite possible values of the free variables.

Example 1.3.3. Systems with different numbers of solutions.

1. $\left[\begin{array}{cc:c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right]$ comes form a linear system with no solutions.

2. For the matrix $\left[\begin{array}{ccc:c} 1 & 0 & 0 & a \\ 0 & 1 & 0 & b \\ 0 & 0 & 1 & c \end{array}\right]$, it has a unique solution $(x, y, z) = (a, b, c)$

2 Vector Equations and Linear Transformations

2.1 Vectors

Definition 2.1.1 (Vector). A **vector** is an array of n numbers:

$$\vec{x}(\text{or } \mathbf{x}) = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Definition 2.1.2 ($\mathbb{R}^n$). A set of all vectors of height in n is denoted in $\mathbb{R}^n$.

Theorem 2.1.1 (Vector Addition).

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} + \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} a+x \\ b+y \\ c+z \end{bmatrix}$$

Theorem 2.1.2 (Scalar multiplication).

$$c \times \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} c \times x \\ c \times y \\ c \times z \end{bmatrix}$$

Example 2.1.1. The vector $\vec{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is shown on a xy-plane as the following:

1. The vector addition of $\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$ will be:

2. The vector scalar multiplication of $3 \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ will be:

   Extension. The Parallelogram Law for Vector Addition.

   Extension. Vector Subtraction.

Definition 2.1.3 (Linear Combinations). Let $c_1, c_2, ..., c_k$ be scalars, and let $v_1, v_2, ..., v_k$ be vectors in $\mathbb{R}^2$. The vector in $\mathbb{R}^2$

$$c_1 v_1 + c_2 v_2 + ... + c_k v_k$$

is called a **linear combination** of the vectors $v_1, v_2, ..., v_k$ with **weights** or **coefficients** $c_1, c_2, ..., c_k$.

2.2 Vector Equations

Definition 2.2.1 (Vector Equation). A **vector equation** is an equation involving a linear combination of vectors with possibly unknown coefficients.

Example 2.2.1. Asking whether or not a vector equation has a solution is the same as asking if a given vector is a linear combination of some other given vector.

The equation

$$x \begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix} + y \begin{bmatrix} -1 \\ -2 \\ -6 \end{bmatrix} = \begin{bmatrix} 8 \\ 16 \\ 3 \end{bmatrix}$$

is asking if the vector $\begin{bmatrix} 8 \\ 16 \\ 3 \end{bmatrix}$ is a linear combination of the vectors $\begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ -2 \\ -6 \end{bmatrix}$.

The equation can be simplified to

$$\begin{bmatrix} x - y \\ 2x - 2y \\ 6x - y \end{bmatrix} = \begin{bmatrix} 8 \\ 16 \\ 3 \end{bmatrix} \quad \text{or} \quad \begin{cases} x - y = 8 \\ 2x - 2y = 16 \\ 6x - y = 3 \end{cases}$$

.

Then, one can use augmented matrix to solve it.

Remark. Three equivalent ways of thinking about a linear system:

1. A system of equations
2. An augmented matrix
3. A vector equation

Theorem 2.2.1. A new way to consider linear systems.

Suppose the LHS of a linear system is something we can plug a vector into to produce a list of numbers, and the RHS of a linear system shows the solution out as a vector.

Thus, The LHS of a system is a function $\boldsymbol{T}\colon \mathbb{R}^m \to \mathbb{R}^n$, where m is the number of variables and n is the number of equations.

To solve the system, we want to find all vectors that will map to a particular group. We can record the function associated with the LHS of a system as a matrix.

Example 2.2.2. Example of converting linear systems to matrix equations.

The linear system

$$\begin{cases} 7x_1 + 3x_2 + 4x_3 = 25 \\ 2x_1 + 0x_2 + x_3 = 5 \end{cases}$$

can be recorded as

$$\begin{bmatrix} 7 & 3 & 4 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 25 \\ 5 \end{bmatrix}$$

Theorem 2.2.2. Multiplication of a vector by a matrix.

1. For each row of the matrix, multiply the entries of that row with the corresponding entries of the vector and then add.

2. The output vector is the final output.

Example 2.2.3.

$$\begin{bmatrix} 7 & 3 & 4 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \times 1 + 3 \times 1 + 4 \times 1 \\ 2 \times 1 + 0 \times 1 + 1 \times 1 \end{bmatrix} = \begin{bmatrix} 14 \\ 3 \end{bmatrix}$$

2.3 Linear Transformation

Definition 2.3.1 (Linear Transformation). A **linear transformation** is a function $\boldsymbol{T} : \mathbb{R}^m \to \mathbb{R}^n$ so that:

1. $\boldsymbol{T}(\vec{x} + \vec{y}) = \boldsymbol{T}(\vec{x}) + \boldsymbol{T}(\vec{y})$

2. $\boldsymbol{T}(c \times \vec{x}) = c \times \boldsymbol{T}(\vec{x})$

$$\forall \vec{x}, \vec{y} \in \mathbb{R}^m, \text{ and } c \in \mathbb{R}$$

Definition 2.3.2 (Standard Basis Vectors). The vectors $\vec{e}_1, \vec{e}_2, ..., \vec{e}_n \in \mathbb{R}^m$ defined by

$$\vec{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \to \text{ the } i\text{-th entry}$$

are called the **standard basis vectors**.

Theorem 2.3.1. Let $\boldsymbol{T} : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation, and

$$\boldsymbol{A} = \begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{T}\vec{e}_1 & \boldsymbol{T}\vec{e}_2 & \cdots & \boldsymbol{T}\vec{e}_n \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix}$$

Then, $\boldsymbol{T}\vec{x} = \boldsymbol{A}\vec{x}$ for all vectors $\vec{x}$

Proof. Assume $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, then $\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + ... + x_n\vec{e}_n$.

Thus,

$$\boldsymbol{T}\vec{x} = x_1\boldsymbol{T}\vec{e}_1 + x_2\boldsymbol{T}\vec{e}_2 + ... + x_n\boldsymbol{T}\vec{e}_n = \begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{T}\vec{e}_1 & \boldsymbol{T}\vec{e}_2 & \cdots & \boldsymbol{T}\vec{e}_n \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \boldsymbol{A}\vec{x}$$

■

Theorem 2.3.2. Given any sequence of elementary raw operations $s_1, s_2, ..., s_k$ involving n-rows, there exists a matrix $\boldsymbol{B}$ such that for all $\vec{v} \in \mathbb{R}^n$, $\boldsymbol{B}\vec{v}$ equals that vector obtained by applying $s_1, s_2, ..., s_k$ to $\vec{v}$.

Example 2.3.1.

$$\begin{bmatrix} x \\ y \end{bmatrix} \xrightarrow{\text{II}-\text{I}} \begin{bmatrix} x \\ y-x \end{bmatrix} \xrightarrow{\text{I}-\text{II}} \begin{bmatrix} 2x-y \\ y-x \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix},$$

where $\begin{bmatrix} 2 & -1 \\ 1 & -1 \end{bmatrix}$ is the matrix $\boldsymbol{B}$

Definition 2.3.3 (Geometric Definition of Linear Transformation). We can also think of linear transformation from a geometric perspective.

1. $\boldsymbol{T} : \mathbb{R}^m \to \mathbb{R}^n$ implies that the original parallelograms map to the transformed parallelograms
2. $\boldsymbol{T}(c\times\vec{x}) = c\times\boldsymbol{T}(\vec{x})$ means that the original lines through the origin map to the transformed lines through the origin, and the original maps the ruling defined with fundamental unit $\vec{x}$ to ruling with unit $\boldsymbol{T}\vec{x}$
3. Rotation around the origin is a linear transformation.
4. Reflection through a line through the origin is a linear transformation.
5. Translation is not a linear transformation.

Example 2.3.2. Fix $\theta \in [0, 2\pi)$. Consider the map $\text{Rot}_\theta : \mathbb{R}^2 \to \mathbb{R}^2$, which rotates a vector by angle θ around the origin counterclockwise. Rot_θ is a linear transformation. Find the matrix associated with this transformation.

Solution. Let $\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\vec{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

The matrix of Rot_θ is

$$\begin{bmatrix} | & | \\ \text{Rot}_\theta\vec{e}_1 & \text{Rot}_\theta\vec{e}_2 \\ | & | \end{bmatrix}$$

1. If $\theta = \frac{\pi}{2}$, i.e. we rotate by 90 counterclockwise. The matrix for rotation is

$$\begin{bmatrix} | & | \\ \mathrm{Rot}_\theta \vec{e}_1 & \mathrm{Rot}_\theta \vec{e}_2 \\ | & | \end{bmatrix} = \begin{bmatrix} | & | \\ \vec{e}_2 & -\vec{e}_2 \\ | & | \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

2. General case: $e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}; e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Thus,

$$\mathrm{Rot}_\theta \vec{e}_1 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}; \mathrm{Rot}_\theta \vec{e}_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}.$$

$$\Longrightarrow \begin{bmatrix} | & | \\ \mathrm{Rot}_\theta \vec{e}_1 & \mathrm{Rot}_\theta \vec{e}_2 \\ | & | \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

□

Example 2.3.3. The map $\mathrm{Ref}_L : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation that reflects a vector over the line $L : y = 2x$. Find the matrix for Ref_L.

Solution. Key idea: express $\vec{e}_i = \vec{e}_i^{\parallel} + \vec{e}_i^{\perp}$, and $\mathrm{Ref}(\vec{e}_i) = \mathrm{Ref}(\vec{e}_i^{\parallel}) + \mathrm{Ref}(\vec{e}_i^{\perp})$.

Choose $\begin{bmatrix} 1 \\ 2 \end{bmatrix} \in L$, then every parallel vector is $c \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

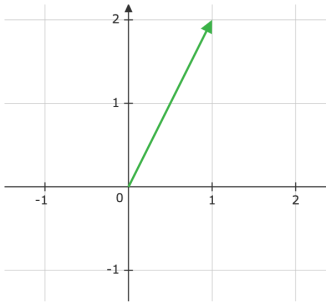
Rotate $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ by 90:

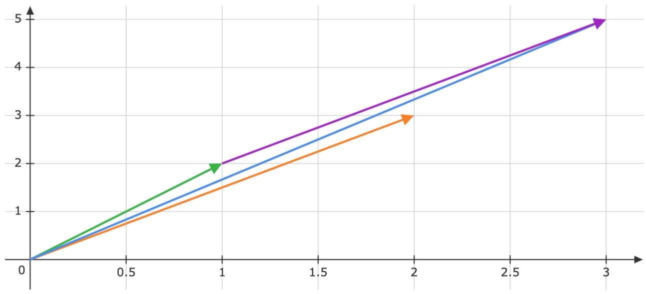
$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \end{bmatrix},$$

then are perpendicular vector is $d \begin{bmatrix} -2 \\ 1 \end{bmatrix}$.

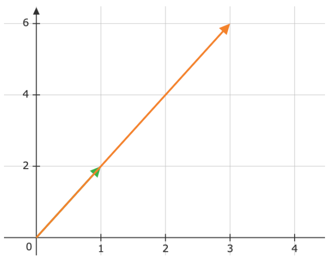
Take $\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\vec{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, then we get

$$\begin{cases} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = c \begin{bmatrix} 1 \\ 2 \end{bmatrix} + d \begin{bmatrix} -2 \\ 1 \end{bmatrix} \\ \begin{bmatrix} 0 \\ 1 \end{bmatrix} = c' \begin{bmatrix} 1 \\ 2 \end{bmatrix} + d' \begin{bmatrix} -2 \\ 1 \end{bmatrix} \end{cases} \Longrightarrow \begin{cases} c = \frac{1}{5} \\ d = -\frac{2}{5} \\ c' = \frac{2}{5} \\ d' = \frac{1}{5} \end{cases} \Longrightarrow \begin{cases} e_1 = \frac{1}{5} \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \frac{2}{5} \begin{bmatrix} -2 \\ 1 \end{bmatrix} \\ e_2 = \frac{2}{5} \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \frac{1}{5} \begin{bmatrix} -2 \\ 1 \end{bmatrix} \end{cases};$$

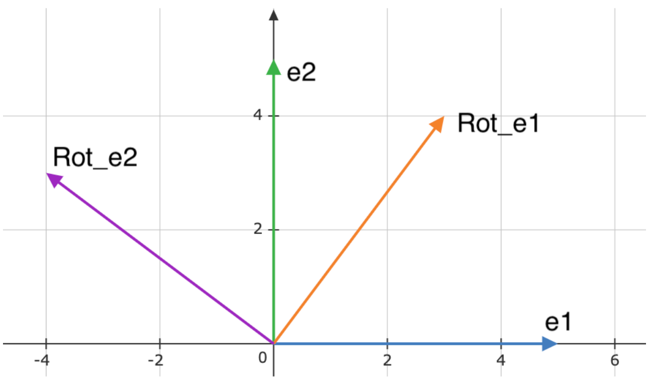




addition.png



multiplication.png



$$\xrightarrow{\mathrm{Ref}_L} \begin{cases} \mathrm{Ref}_L(\vec{e}_1) = \frac{1}{5}\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \frac{2}{5}\begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} {}^{-3}\!/_5 \\ {}^{4}\!/_5 \end{bmatrix} \\ \mathrm{Ref}_L(\vec{e}_2) = \frac{2}{5}\begin{bmatrix} 1 \\ 2 \end{bmatrix} - \frac{1}{5}\begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} {}^{4}\!/_5 \\ {}^{3}\!/_5 \end{bmatrix} \end{cases} .$$

Thus, the matrix is

$$\begin{bmatrix} {}^{-3}\!/_5 & {}^{4}\!/_5 \\ {}^{4}\!/_5 & {}^{3}\!/_5 \end{bmatrix} .$$

□

3 Matrices

3.1 Matrix Multiplication

Theorem 3.1.1 (Procedure of Matrix Multiplication). Matrix multiplication is very different from other formats of multiplication.

- Input: a pair of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.
  *The number of rows of $\boldsymbol{A}$ equals the number of columns of $\boldsymbol{B}$.

- Output: The product $\boldsymbol{BA}$

- Procedure:

  1. View $\boldsymbol{A}$ as a list of its column vectors:

  $$\boldsymbol{A} = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}$$

  2. Multiply each column by $\boldsymbol{B}$:

  $$\boldsymbol{BA} = \begin{bmatrix} | & & | \\ \boldsymbol{B}v_1 & \cdots & \boldsymbol{B}v_n \\ | & & | \end{bmatrix}$$

Example 3.1.1. Examples of matrix multiplication.

1. Let $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 3 & 5 \end{bmatrix}$. Find $\boldsymbol{BA}$.

   Solution.

   $$\boldsymbol{BA} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 \times 1 + 2 \times (-1) & 2 \times 1 + 2 \times 1 \\ 1 \times 0 + 1 \times (-1) & 0 \times 2 + 1 \times 1 \\ 3 \times 1 + 5 \times (-1) & 3 \times 1 + 5 \times 1 \end{bmatrix} = \begin{bmatrix} -1 & 4 \\ -1 & 1 \\ -2 & 11 \end{bmatrix}$$

   □

2. Let $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 3 & 5 \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}$. Find $\boldsymbol{BA}$.

   Solution. Because 2 columns is not equal to three rows, the product does not exist. □

3. Let $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} 1 & 0 & 3 \\ 2 & 1 & 5 \end{bmatrix}$. Find $\boldsymbol{AB}$.

Solution.

$$\boldsymbol{AB} = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 \\ 2 & 1 & 5 \end{bmatrix} = \begin{bmatrix} 1\times 1 + 2\times 2 & 0\times 1 + 2\times 1 & 3\times 1 + 2\times 5 \\ -1\times 1 + 2\times 1 & 0\times(-1) + 1\times 1 & 3\times(-1) + 5\times 1 \end{bmatrix} = \begin{bmatrix} 5 & 2 & 13 \\ 1 & 1 & 2 \end{bmatrix}$$

□

Remark (Conceptualizing Matrix Multiplication). There are many ways to understand matrix multiplication:

1. A matrix encodes a linear transformation:

$$\boldsymbol{A} : \mathbb{R}^m \longrightarrow \mathbb{R}^n \text{ is a } m \times n \text{ matrix.}$$

$$\boldsymbol{B} : \mathbb{R}^n \longrightarrow \mathbb{R}^k \text{ is a } n \times k \text{ matrix.}$$

We can compass these maps:

$$\begin{tikzcd} \mathbb{R}^m \arrow[r, "\boldsymbol{A}"] \arrow[rr, bend right, "\boldsymbol{BA}"] & \mathbb{R}^n \arrow[r, "\boldsymbol{B}"] & \mathbb{R}^k \end{tikzcd}$$

The product $\boldsymbol{BA}$ encodes the composition of those transformations.

Example 3.1.2. Rotation by 90° counterclockwise:

$$\boldsymbol{B} = \boldsymbol{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

Thus,

$$\boldsymbol{BA} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \text{ encodes a rotation by } 180^\circ$$

2. The composition $\boldsymbol{BA}$ is linear:

- $\boldsymbol{BA}(\vec{x} + \vec{y}) = \boldsymbol{B}(\boldsymbol{A}\vec{x} + \boldsymbol{A}\vec{y}) = \boldsymbol{BA}\vec{x} + \boldsymbol{BA}\vec{y}$
- $\boldsymbol{BA}(c\vec{x}) = \boldsymbol{B}(c\boldsymbol{A}\vec{x}) = c\boldsymbol{BA}\vec{x}$

3. The matrix for the composition is:

$$\boldsymbol{BA} = \begin{bmatrix} | & | & & | \\ \boldsymbol{B}v_1 & \boldsymbol{B}v_2 & \cdots & \boldsymbol{B}v_n \\ | & | & & | \end{bmatrix}, \text{ where } \boldsymbol{A} = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix}$$

Proof. Suppose

$$\boldsymbol{A} = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \boldsymbol{A}e_1 & \cdots & \boldsymbol{A}e_n \\ | & & | \end{bmatrix}$$

Then,

$$\boldsymbol{B}\boldsymbol{A} = \begin{bmatrix} | & & | \\ \boldsymbol{B}\boldsymbol{A}e_1 & \cdots & \boldsymbol{B}\boldsymbol{A}e_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \boldsymbol{B}v_1 & \cdots & \boldsymbol{B}v_n \\ | & & | \end{bmatrix}$$

■

Example 3.1.3 (Application: Double Angle Formulae). Find an expression for $\sin 2\theta$ and $\cos 2\theta$ in terms of $\sin\theta$ and $\cos\theta$.

Solution. For angle θ, we have rotation by θ is a linear transformation, and the matrix is:

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

Geometrically, $\boldsymbol{A}\dot{\boldsymbol{A}}$ is rotation by 2θ:

$$\boldsymbol{A}\cdot\boldsymbol{A} = \begin{bmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{bmatrix}$$

Algebraically, we have

$$\boldsymbol{A}\cdot\boldsymbol{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \cos^2\theta - \sin^2\theta & -2\sin\theta\cos\theta \\ 2\sin\theta\cos\theta & \cos^2\theta - \sin^2\theta \end{bmatrix}$$

Since these are equal:

$$\cos 2\theta = \cos^2\theta - \sin^2\theta$$

$$\sin 2\theta = 2\sin\theta\cos\theta$$

□

Remark. Generalization: $\boldsymbol{A}^3$: triple angle formulae; $\boldsymbol{A}^n$: multiple angle formulae

Theorem 3.1.2. Algebraic properties of matrix multiplication:

1. Matrix multiplication is associated:

$$(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC}), \text{ assuming the products } \boldsymbol{AB},\ \boldsymbol{BC},\ \boldsymbol{AB})\boldsymbol{C} \text{ exists.}$$

2. Matrix multiplication is generally NOT communitive:

   (a) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices n rows and n columns, $\boldsymbol{AB} \neq \boldsymbol{BA}$ in general. *View matrix multiplication as a type of function composition.

   (b) In other words, **the order matters**.

   Example 3.1.4. • Exception: $\begin{bmatrix}3\end{bmatrix}\begin{bmatrix}6\end{bmatrix} = \begin{bmatrix}18\end{bmatrix} = \begin{bmatrix}6\end{bmatrix} = \begin{bmatrix}3\end{bmatrix}$

- Consider

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} 5 & 7 \\ 6 & 8 \end{bmatrix} = \begin{bmatrix} 5+18 & 7+24 \\ 10+24 & 14+32 \end{bmatrix} = \begin{bmatrix} 23 & 31 \\ 34 & 46 \end{bmatrix}$$

$$\begin{bmatrix} 5 & 7 \\ 6 & 8 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 5+14 & 15+28 \\ 6+16 & 18+32 \end{bmatrix} = \begin{bmatrix} 19 & 43 \\ 22 & 50 \end{bmatrix}$$

Thus,

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} 5 & 7 \\ 6 & 8 \end{bmatrix} \neq \begin{bmatrix} 5 & 7 \\ 6 & 8 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$$

3.2 Invertible Matrices

Example 3.2.1 (Guiding Question). Let $\vec{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \in \mathbb{R}^2$ be a fixed, arbitrary vector. Let $\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$. Find all solutions $\vec{x} \in \mathbb{R}^2$ to the matrix equation $\boldsymbol{A}\vec{x} = \vec{b}$ (as a function of b_1 and b_2.)

Solution. Observe: $\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix}$, $\boldsymbol{A}\vec{x} = \begin{bmatrix} 2x+y \\ x+y \end{bmatrix}$. Then we want to solve

$$\begin{cases} 2x+y = b_1 \\ x+y = b_2 \end{cases}$$

$$\Rightarrow \left[\begin{array}{cc:c} 2 & 1 & b_1 \\ 1 & 1 & b_2 \end{array}\right] \xrightarrow{I \leftrightarrow II} \left[\begin{array}{cc:c} 1 & 1 & b_2 \\ 2 & 1 & b_1 \end{array}\right] \xrightarrow{II-2I} \left[\begin{array}{cc:c} 1 & 1 & b_2 \\ 0 & -1 & b_1 - 2b_2 \end{array}\right]$$

$$\xrightarrow{II/(-1)} \left[\begin{array}{cc:c} 1 & 1 & b_2 \\ 0 & 1 & 2b_2 - b_1 \end{array}\right] \xrightarrow{I-II} \left[\begin{array}{cc:c} 1 & 0 & -b_2 + b_1 \\ 0 & 1 & 2b_2 - b_1 \end{array}\right]$$

$$\therefore \vec{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -b_2 + b_1 \\ 2b_2 - b_1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

□

Definition 3.2.1 (Inverse of a Matrix). Let $\boldsymbol{A}$ be a square ($n \times n$ matrix). Assume $\boldsymbol{A}\vec{x} = \vec{b}$ has unique solution for each $\vec{b} \in \mathbb{R}^n$. Then the map $\vec{b} \longmapsto \vec{x}$, the unique solution to $\boldsymbol{A}\vec{x} = \vec{b}$, is a linear transformation and the matrix of this map is called the **inverse of $\boldsymbol{A}$**. We denote it as $\boldsymbol{A}^{-1}$.

Remark. The matrix $\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$ in the guiding question is the inverse of $\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$.

Theorem 3.2.1. Computing the inverse for a matrix.

- $\boldsymbol{A}^{-1}$ does not always exist.
- There are square matrices such that $\boldsymbol{A}\vec{x} = \vec{b}$ has infinite solutions.

- Process:

$$\left[\begin{array}{c|c} & b_1 \\ \boldsymbol{A} & \vdots \\ & b_n \end{array}\right] \xrightarrow{\text{Row reduce}} \left[\begin{array}{c|c} \text{rref}(\boldsymbol{A}) & \begin{array}{c}\text{Linear expressions} \\ \text{in terms of } b_i\end{array} \end{array}\right]$$

Check pivot over each row of rref($\boldsymbol{A}$), and the coefficient matrix is $\boldsymbol{A}^{-1}$.

Definition 3.2.2 (Identity matrix). For an $n \times n$ matrix, if it is

$$\boldsymbol{I}_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix},$$

we call it the **identity matrix**.

Remark. $\boldsymbol{I}_n$ encodes the linear transformation $\boldsymbol{I}_n: \mathbb{R}^n \to \mathbb{R}^n$ $(\vec{x} \longmapsto \vec{x})$

Theorem 3.2.2. Procedure for finding $\boldsymbol{A}^{-1}$:

1. Form augmented matrices:

$$\left[\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{I}_n \end{array}\right] = \left[\begin{array}{c|ccccc} & 1 & 0 & 0 & \cdots & 0 \\ & 0 & 1 & 0 & \cdots & 0 \\ \boldsymbol{A} & 0 & 0 & 1 & \cdots & 0 \\ & \vdots & \vdots & \vdots & \ddots & \vdots \\ & 0 & 0 & 0 & \cdots & 1 \end{array}\right]$$

2. Row reduce:

$$\left[\begin{array}{c|c} \text{rref}(\boldsymbol{A}) & \boldsymbol{B} \end{array}\right],$$

if rref($\boldsymbol{A}$)=$\boldsymbol{I}_n$, $\boldsymbol{B} = \boldsymbol{A}^{-1}$

Example 3.2.2.

$$\left[\begin{array}{cc|cc} 2 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{array}\right] \xrightarrow{I \leftrightarrow II} \left[\begin{array}{cc|cc} 1 & 1 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{array}\right] \xrightarrow{II-2I} \left[\begin{array}{cc|cc} 1 & 1 & 0 & 1 \\ 0 & -1 & 1 & -2 \end{array}\right]$$

$$\xrightarrow{II/(-1)} \left[\begin{array}{cc|cc} 1 & 1 & 0 & 1 \\ 0 & 1 & -1 & 2 \end{array}\right] \xrightarrow{I-II} \left[\begin{array}{cc|cc} 1 & 0 & 1 & -1 \\ 0 & 1 & -1 & 2 \end{array}\right]$$

$$\therefore \boldsymbol{A}^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

Theorem 3.2.3 (Function theoretic definition of A^{-1}). When $\boldsymbol{A}^{-1}$ exists, $matrix A^{-1}$ is the matrix encoding the inverse function of $\boldsymbol{A}$. Hence, $\boldsymbol{A}$ and $\boldsymbol{A}^{-1}$ always commute:

$$\boldsymbol{A}^{-1} \cdot \boldsymbol{A} = \boldsymbol{I}_n = \boldsymbol{A} \cdot \boldsymbol{A}^{-1}$$

Example 3.2.3. Let

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \ \boldsymbol{A}^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}.$$

$$\boldsymbol{A} \cdot \boldsymbol{A}^{-1} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 2-1 & -2+2 \\ 1-1 & -1+2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \boldsymbol{I}_2.$$

Theorem 3.2.4 (A new way to find A^{-1}). Solving $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}_n$

$$\boldsymbol{A}^{-1} = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}$$

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \begin{bmatrix} | & & | \\ \boldsymbol{A}v_1 & \cdots & \boldsymbol{A}v_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ e_1 & \cdots & e_n \\ | & & | \end{bmatrix} = \boldsymbol{I}_n$$

$$\boldsymbol{A}v_1 = e_1, \ \boldsymbol{A}v_2 = e_2, \cdots, \boldsymbol{A}v_n = e_n$$

$$\left[\begin{array}{c:c} \boldsymbol{A}v_1 & e_1 \end{array}\right], \left[\begin{array}{c:c} \boldsymbol{A}v_2 & e_2 \end{array}\right], \cdots, \left[\begin{array}{c:c} \boldsymbol{A}v_n & e_n \end{array}\right]$$

$$\xrightarrow{\text{Row reduce}} \left[\begin{array}{c:c} \text{rref}(\boldsymbol{A}) & v_1 \end{array}\right], \left[\begin{array}{c:c} \text{rref}(\boldsymbol{A}) & v_2 \end{array}\right], \cdots, \left[\begin{array}{c:c} \text{rref}(\boldsymbol{A}) & v_n \end{array}\right]$$

$\therefore$ To find $\boldsymbol{A}^{-1}$:

$$\left[\begin{array}{c:c} \boldsymbol{A} & \boldsymbol{I}_n \end{array}\right] \xRightarrow{\text{Row reduce}} \left[\begin{array}{c:c} \boldsymbol{I} & \boldsymbol{A}^{-1} \end{array}\right]$$

Example 3.2.4 (Problems concerning inverting matrices). Let $\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{bmatrix}$. Compute $\boldsymbol{A}^{-1}$ and use it to find all solutions to $\boldsymbol{A}\vec{x} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$.

Solution.

$$\left[\begin{array}{ccc:ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 3 & 0 & 1 & 0 \\ 1 & 4 & 9 & 0 & 0 & 1 \end{array}\right] \xrightarrow[II-I]{III-I} \left[\begin{array}{ccc:ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 2 & -1 & 1 & 0 \\ 0 & 3 & 8 & -1 & 0 & 1 \end{array}\right] \xrightarrow[I-II]{III-3II} \left[\begin{array}{ccc:ccc} 1 & 0 & 1 & 2 & -1 & 0 \\ 0 & 1 & 2 & -1 & 1 & 0 \\ 0 & 0 & 2 & 2 & -3 & 1 \end{array}\right]$$

$$\xrightarrow{III/2} \left[\begin{array}{ccc:ccc} 1 & 0 & 1 & 2 & -1 & 0 \\ 0 & 1 & 2 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 & -3/2 & 1/2 \end{array}\right] \xrightarrow[I+III]{II-2III} \left[\begin{array}{ccc:ccc} 1 & 0 & 0 & 3 & -5/2 & 1/2 \\ 0 & 1 & 0 & -3 & 4 & -1 \\ 0 & 0 & 1 & 1 & -3/2 & 1/2 \end{array}\right]$$

$$\therefore \boldsymbol{A}^{-1} = \begin{bmatrix} 3 & -^5/_2 & ^1/_2 \\ -3 & 4 & -1 \\ 1 & -^3/_2 & ^1/_2 \end{bmatrix}$$

To solve $\boldsymbol{A}\vec{x} = \vec{b}$, apply $\boldsymbol{A}^{-1}$ on both sides:

$$\begin{aligned} \boldsymbol{A}^{-1}(\boldsymbol{A}\vec{x}) &= \boldsymbol{A}^{-1}\vec{b} \\ \vec{x} &= \boldsymbol{A}^{-1}\vec{b} \\ \therefore \vec{x} &= \begin{bmatrix} 3 & -^5/_2 & ^1/_2 \\ -3 & 4 & -1 \\ 1 & -^3/_2 & ^1/_2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 +^5 /_2 +^1 /_2 \\ -3-4-1 \\ 1 +^3 /_2 +^1 /_2 \end{bmatrix} = \begin{bmatrix} 6 \\ -8 \\ 3 \end{bmatrix} \end{aligned}$$

□

3.3 Kernel of a Matrix

As $\boldsymbol{A}\vec{x} = \vec{b}$ encodes a system of linear equation, one key question of linear algebra is to find how would the solution to $\boldsymbol{A}\vec{x} = \vec{b}$ change as $\vec{b}$ varies.

Theorem 3.3.1. Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be a function, then:

1. If $\vec{b}_1,\ \vec{b}_2 \in \mathbb{R}^n$, and $\vec{b}_1 \neq \vec{b}_2$, then the sets $\left\{\vec{x} : f(\vec{x}) = \vec{b}_1\right\}$ and $\left\{\vec{x} : f(\vec{x}) = \vec{b}_2\right\}$ do not intersect.

2. Every $\vec{x}$ in the domain is an element of the solution set $\left\{\vec{x} : f(\vec{x}) = \vec{b}\right\}$ for some $\vec{b}$.

Example 3.3.1. Let $\boldsymbol{A} = \begin{bmatrix} 2 & 1 \end{bmatrix}$. Then solving $\boldsymbol{A}\vec{x} = \vec{b}$ gives $2x + y = \vec{b}$, which encodes a line of slope$= -2$ that has a $y-$intercept of $\vec{b}$.

Definition 3.3.1 (Zero Vector). The **zero vector** $\vec{0} \in \mathbb{R}^n$ (sometimes denoted as $\vec{0}_n$ if the context is unclear) is the vector all of whose entries are 0.

Example 3.3.2.

$$\vec{0}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix},\ \vec{0}_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Theorem 3.3.2. Let $\boldsymbol{A}$ be an $n \times m$ matrix (i.e., encoding a linear transformation $\boldsymbol{A} : \mathbb{R}^m \to \mathbb{R}^n$) and $\vec{b} \in \mathbb{R}^n$ such that $(s.t.)$ $\boldsymbol{A}\vec{x} = \vec{b}$ has a solution. Suppose $\vec{x}_0$ to be any fixed solution. Then, the solution set to $\boldsymbol{A}\vec{x} = \vec{b}$ is $\{\vec{x}_0 + \vec{x}' \mid \boldsymbol{A}\vec{x}' = 0\}$

Interpretation: The solution set to $\boldsymbol{A}\vec{x} = \vec{b}$ is the translation of the solution set to $\boldsymbol{A}\vec{x} = \vec{0}$ by $\vec{x}_0$. ***Proof.*** We need to prove two parts: 1. Any solution to $\boldsymbol{A}\vec{x} = \vec{b}$ is of the form $\vec{x}_0 + \vec{x}'$, where $\boldsymbol{A}\vec{x}' = \vec{0}$, and 2. $\vec{x}_0 + \vec{x}$ are solutions to $\boldsymbol{A}\vec{x} = \vec{b}$.

1. Any solution to $\boldsymbol{A}\vec{x} = \vec{b}$ is of the form $\vec{x}_0 + \vec{x}'$, where $\boldsymbol{A}\vec{x}' = \vec{0}$.
   Let $\vec{x}$ be such a solution, then $\vec{x}' := \vec{x} - \vec{x}_0$, then

$$\begin{aligned} \boldsymbol{A}\vec{x}' &= \boldsymbol{A}(\vec{x} - \vec{x}_0) \\ &= \boldsymbol{A}\vec{x} - \boldsymbol{A}\vec{x}_0 \\ &= \vec{b} - \vec{b} \\ &= \vec{0}. \end{aligned}$$

   So, $\vec{x} = \vec{x}_0 + \vec{x}'$.

2. $\vec{x}_0 + \vec{x}$ are solutions to $\boldsymbol{A}\vec{x} = \vec{b}$.

$$\begin{aligned} \boldsymbol{A}(\vec{x}_0 + \vec{x}') &= \boldsymbol{A}\vec{x}_0 + \boldsymbol{A}\vec{x}' \\ &= \vec{b} + \vec{(0)} = \vec{b}. \end{aligned}$$

■

Definition 3.3.2 (Kernel of a Matrix). The **kernel** of a linear transformation or a matrix is the solution set to $\boldsymbol{A}\vec{x} = \vec{0}$.

$$\text{i.e., } \operatorname{Ker}(\boldsymbol{A}) = \left\{ \vec{x} \in \mathbb{R}^m;\ \boldsymbol{A}\vec{x} = \vec{0} \right\}.$$

Theorem 3.3.3.

$$\operatorname{Ker}(\boldsymbol{A}) = \operatorname{Ker}(\operatorname{rref}(\boldsymbol{A})).$$

Theorem 3.3.4. Procedure of computing the kernel of a matrix:

1. Row reduce $\boldsymbol{A}$ to $\operatorname{rref}(\boldsymbol{A})$, compute $\operatorname{Ker}(\operatorname{rref}(\boldsymbol{A}))$.

2. Unpack the equations encoded by matrix equation $\operatorname{rref}(\boldsymbol{A}) = 0$, solve for pivot variables in terms of free variables.

3. Parameterize the solution set for $\operatorname{rref}(\boldsymbol{A})\vec{x} = 0$ as $\{t_1\vec{v}_1 + t_2\vec{v}_2 + \cdots + t_d\vec{v}_d :\ t_i \in \mathbb{R}\}$ and $\vec{v}_i$ tracks the coefficient of the i-th free variable.

Example 3.3.3. Let $\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix}$. Compute $\operatorname{Ker}(\boldsymbol{A})$.

Solution. $\operatorname{Ker}(\boldsymbol{A})$ is the solution set to $\boldsymbol{A}\vec{x} = \vec{0}$:

$$\left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 0 \\ 5 & 6 & 7 & 8 & 0 \\ 9 & 10 & 11 & 12 & 0 \end{array}\right] \xrightarrow[II-5I]{III-9I} \left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 0 \\ 0 & -4 & -8 & -12 & 0 \\ 0 & -8 & -16 & -24 & 0 \end{array}\right] \xrightarrow{II/-4} \left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 0 \\ 0 & 1 & 2 & 3 & 0 \\ 0 & -8 & -16 & -24 & 0 \end{array}\right]$$

$$\xrightarrow[I-2II]{III+8II} \left[\begin{array}{cccc|c} 1 & 0 & -1 & -2 & 0 \\ 0 & 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$\therefore \begin{cases} x_1 - x_3 - 2x_4 = 0 \\ x_2 + 2x_3 + 3x_4 = 0 \end{cases}$$

$$\therefore \text{ Solution set: } \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_3 + 2x_4 \\ -2x_3 - 3x_4 \\ x_3 \\ x_4 \end{bmatrix}.$$

Thus,

$$\operatorname{Ker}(\boldsymbol{A}) = \left\{ \begin{bmatrix} x_3 + 2x_4 \\ -2x_3 - 3x_4 \\ x_3 \\ x_4 \end{bmatrix} : \ x_3,\ x_4 \in \mathbb{R} \right\} = \left\{ x_3 \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \end{bmatrix} \right\}$$

□

Definition 3.3.3 (Span). Let $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_d \in \mathbb{R}$, the **span** of $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_d$ is the set:

$$\operatorname{Span}(\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_d) = \{t_1\vec{v}_1 + t_2\vec{v}_2 + \cdots + t_d\vec{v}_d;\ t_i \in \mathbb{R}\}$$

Example 3.3.4. Our procedure of finding kernels finds vectors $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_d$ which spans the kernel of the matrix.

Definition 3.3.4 (Image of a Matrix). Let $\boldsymbol{A}$ be an $n \times m$ matrix (i.e., encoding a linear transformation $\boldsymbol{A} : \mathbb{R}^m \to \mathbb{R}^n$), the **image** of $\boldsymbol{A}$ is the set:

$$\operatorname{Im}(\boldsymbol{A}) = \{\boldsymbol{A}\vec{x} \mid \vec{x} \in \mathbb{R}^m\}.$$

Interpretation: $\operatorname{Im}(\boldsymbol{A})$ is the set of $\vec{b}$ *s.t.* $\boldsymbol{A}\vec{x} = \vec{b}$ has a solution.

Theorem 3.3.5. Let $\boldsymbol{A}$ be an $n \times m$ matrix, and let $\vec{w}_1, \cdots, \vec{w}_m$ be the columns of $\boldsymbol{A}$: $\boldsymbol{A} = \begin{bmatrix} | & & | \\ \vec{w}_1 & \cdots & \vec{w}_m \\ | & & | \end{bmatrix}$. The image of $\boldsymbol{A}$ is the span of $\vec{w}_1, \cdots, \vec{w}_m$:

$$\operatorname{Im}(\boldsymbol{A}) = \operatorname{Span}(\vec{w}_1, \cdots, \vec{w}_m) = \{t_1\vec{w}_1 + \cdots + t + m\vec{w}_m;\ t_m \in \mathbb{R}\}$$

Remark.

$$\operatorname{Ker}(\boldsymbol{A}) \subseteq \mathbb{R}^m \quad \text{(domain)}$$
$$\operatorname{Span}(\boldsymbol{A}) \subseteq \mathbb{R}^n \quad \text{(range)}$$

Proof. We know that the columns of a matrix form $\boldsymbol{A}\vec{x}$, namely the i-th column of the matrix

$\boldsymbol{A}$ is $\boldsymbol{A}\vec{e}_i$, where $\vec{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \rightarrow$ the i-th entry.

Hence, $\vec{x} = \begin{bmatrix} x_1 \\ vdots \\ x_m \end{bmatrix} = x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_m\vec{e}_m$.

$$\begin{aligned} \boldsymbol{A}\vec{x} &= x_1\boldsymbol{A}\vec{e}_1 + x_2\boldsymbol{A}\vec{e}_2 + \cdots + x_m\boldsymbol{A}\vec{e}_m \\ &= x_1\vec{w}_1 + x_2\vec{w}_2 + \cdots + x_m\vec{w}_m \\ &\in \text{Span}(\vec{w}_1, \cdots, \vec{w}_m). \end{aligned}$$

■

4 Spaces and Dimensions

4.1 Subspaces and Bases

Theorem 4.1.1. Spans of Sets of Vectors:

1. In general, for $\vec{v} \in \mathbb{R}^n$, if $\vec{v} \neq 0$, then $\text{Span}(\vec{v})$ is the line through the origin containing $\vec{v}$.
2. If $\vec{v} = \vec{0}$, then $\text{Span}(\vec{v})$ is also the zero vector.
3. For vectors $\vec{v}_1,\ \vec{v}_2 \in \mathbb{R}^n$, if $\vec{v}_1 \neq \vec{v}_2$ and $\vec{v}_1,\ \vec{v}_2 \neq 0$, then $\text{Span}(\vec{v}_1, \vec{v}_2)$ is a plane through the origin containing $\vec{v}_1$ and $\vec{v}_2$.
4. If $\vec{v}_1$ and $\vec{v}_2$ are co-linear with each other, then $\text{Span}(\vec{v}_1, \vec{v}_2)$ is a line through the containing origin of $\vec{v}_1$ and $\vec{v}_2$.

Definition 4.1.1 (Redundancy). A vector $\vec{v}_k$ is called **redundant** in a list of vectors $\vec{v}_1, \cdots, \vec{v}_k \in \mathbb{R}^n$ if

$$\vec{v}_k \in \text{Span}(\vec{v}_1, \cdots, \vec{v}_{k-1})$$

Definition 4.1.2 (Span of an Empty Set). The span of the **empty set of vectors** is $\{\vec{0}\}$.

Definition 4.1.3 (Linear Independence). Let $\vec{v}_1, \cdots, \vec{v}_k \in \mathbb{R}^n$. Then vectors $\vec{v}_1, \cdots, \vec{v}_k$ are called **linearly independent** if $\vec{v}_i$ is not redundant in the list of $\vec{v}_1, \cdots, \vec{v}_i \quad \forall i \in [1, k]$.

Example 4.1.1. $\vec{e}_1, \cdots, \vec{e}_k$ are linearly independent (L.I.) in $\mathbb{R}^n \quad \forall n \geq k$. ($\vec{e}_1, \cdots, \vec{e}_k$ are the standard basis vectors.

Theorem 4.1.2. Span and Linear Independency.

1. The span of the empty set is a point $\{\vec{0}\}$.
2. The span of a single linear independent vector is a line through the origin.
3. The span of two linear independent vectors is a plane through the origin.

Definition 4.1.4 (Subspace). Let V be a subset of $\mathbb{R}^n$. V is called a **subspace** if:

1. $\vec{0} \in V$
   Interpretation: Origin is in V.
2. If $\vec{v} \in V$, then $c\vec{v} \in V \quad \forall c \in \mathbb{R}$.
   Interpretation: If $\vec{v} \in V$, then the line through the origin containing $\vec{v}$ is in V.
3. If $\vec{v}_1,\ \vec{v}_2 \in V$, then $\vec{v}_1 + \vec{v}_2 \in V$.
   Interpretation: If $\vec{v}_1$ and $\vec{v}_2$ are not co-linear and contained in V, then the plane through $\vec{v}_1,\ \vec{v}2$ and $\vec{0}$ is in V.

Example 4.1.2. Examples of subspaces.

1. $\{\vec{0}\}$ is a subspace.

2. $\mathbb{R}^n$ is a subspace.

3. If $\vec{v}_1, \cdots, \vec{v}_k \in \mathbb{R}^n$, then $\mathrm{Span}(\vec{v}_1, \cdots, \vec{v}_k)$ is a subspace.
   Proof.

   (a) $\vec{0} = 0\vec{v}_1 + 0\vec{v}_2 + \cdots + \vec{v}_k$

   (b) $\vec{v} = t_1\vec{v}_1 + t_2\vec{v}_2 + \cdots + t_k\vec{v}_k$

   $$\Longrightarrow c\vec{v} = ct_1\vec{v}_1 + ct_2\vec{v}_2 + \cdots + ct_k\vec{v}_k \in \mathrm{Span}(\vec{v}_1, \cdots, \vec{v}_k)$$

   (c) $\vec{v}' = t_1'\vec{v}_1 + t_2'\vec{v}_2 + \cdots + t_k'\vec{v}_k$

   $$\Longrightarrow \vec{v} + \vec{v}' = (t_1 + t_1')\vec{v}_1 + (t_2 + t_2')\vec{v}_2 + \cdots + (t_k + t_k')\vec{v}_k \in \mathrm{Span}(\vec{v}_1, \cdots, \vec{v}_k)$$

   ∎

4. A line through origin is a subspace.

5. A plane through the origin is a subspace. In two and three dimensions, examples 1, 2, 4, and 5 are the only examples of subspaces. For examples 6 and 7, consider an $n \times m$ matrix $\boldsymbol{A}$, which maps a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$ (i.e., $\boldsymbol{A} : \mathbb{R}^m \to \mathbb{R}^n$). Let $\mathrm{Ker}(\boldsymbol{A})$ and $\mathrm{Im}(\boldsymbol{A})$ be the kernel and image of $\boldsymbol{A}$, respectively.

6. $\mathrm{Ker}(\boldsymbol{A})$ is a subspace.
   Proof.

   (a) $\boldsymbol{A}\vec{0}_n = \vec{0}_n$

   $$\Longrightarrow \vec{0}_n \text{ is in } \mathrm{Ker}(\boldsymbol{A}).$$

   (b) $\boldsymbol{A}\vec{v} = \vec{0}$, then $\boldsymbol{A}(c\vec{v}) = c\boldsymbol{A}\vec{v} = c\vec{0} = \vec{0}$

   $$\Longrightarrow \text{ If } \vec{v} \in \mathrm{Ker}(\boldsymbol{A}), \text{ then } c\vec{v} \in \mathrm{Ker}(\boldsymbol{A}).$$

   (c) If $\boldsymbol{A}\vec{v}_1 = \vec{0}$ and $\boldsymbol{A}\vec{v}_2 = \vec{0}$, then $\boldsymbol{A}(\vec{v}_1 + \vec{v}_2) = \boldsymbol{A}\vec{v}_1 + \boldsymbol{A}\vec{v}_2 = \vec{0}$

   $$\Longrightarrow \text{ If } \vec{v}_1, \vec{v}_2 \in \mathrm{Ker}(\boldsymbol{A}), \text{ then } \vec{v}_1 + \vec{v}_2 \in \mathrm{Ker}(\boldsymbol{A}).$$

   ∎

7. $\mathrm{Im}(\boldsymbol{A})$ is a subspace.
   Proof.

(a) $\vec{0}_n \in \operatorname{Im}(\boldsymbol{A})$

(b) If $\vec{b} \in \operatorname{Im}(\boldsymbol{A})$, then $\vec{b} = \boldsymbol{A}\vec{x}$

$$\implies c\vec{b} = \boldsymbol{A}(c\vec{x}) \in \operatorname{Im}(\boldsymbol{A})$$

(c) If $\vec{b}_1, \vec{b}_2 \in \operatorname{Im}(\boldsymbol{A})$, then $\vec{b}_1 = \boldsymbol{A}\vec{x}_1$ and $\vec{b}_2 = \boldsymbol{A}\vec{x}_2$

$$\implies \vec{b}_1 + \vec{b}_2 = \boldsymbol{A}\vec{x}_1 + \boldsymbol{A}\vec{x}_2 = \boldsymbol{A}(\vec{x}_1 + \vec{x}_{@}) \in \operatorname{Im}(\boldsymbol{A}).$$

■

Remark. The same subspace can be spanned by **many** sets of vectors.

Definition 4.1.5. Let V be a subspace of $\mathbb{R}^n$. A **basis** for V is a set of vectors $\vec{v}_1, \cdots, \vec{v}_k \in V$, which:

1. Span V, and
2. Are linearly independent.

Example 4.1.3. The vectors $\vec{e}_1, \cdots, \vec{e}_n$ are a basis for $\mathbb{R}^n$

Proof.

1. $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_n\vec{e}_n$
2. $e_i \notin \operatorname{Span}(\vec{e}_1, \cdots \vec{e}_{i-1}) \to$ L.I.

■

Theorem 4.1.3 (Computing a basis for Im($\boldsymbol{A}$)). Let $\boldsymbol{A}$ be an $n \times m$ matrix with columns $\vec{v}_1, \cdots, \vec{v}_m$:

$$\boldsymbol{A} = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_m \\ | & & | \end{bmatrix}$$

The columns of $\boldsymbol{A}$ which contain a pivot upone row reduction to rref($\boldsymbol{A}$) are a basis for Im($\boldsymbol{A}$).

Example 4.1.4. Let $\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \xrightarrow{\text{Row Reduction}} \begin{bmatrix} \boxed{1} & 0 & -1 \\ 0 & \boxed{1} & 2 \\ 0 & 0 & 0 \end{bmatrix}$

$$\implies \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix} \text{ and } \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} \text{ are the basis of } \operatorname{Im}(\boldsymbol{A}).$$

Remark. The coefficients -1 and 2 on the third column of $\boldsymbol{A}$ indicates that

$$\begin{bmatrix}3\\6\\9\end{bmatrix} = -1\begin{bmatrix}1\\4\\7\end{bmatrix} + 2\begin{bmatrix}2\\5\\8\end{bmatrix}.$$

Proof. We know: $\mathrm{Im}(\boldsymbol{A}) = \mathrm{Span}(\vec{v}_1, \cdots \vec{v}_m)$. To produce basis, remove redundant columns. Hence, we want to show: the i-th column does not contain a pivot on row reduction (iff) $\vec{v}_i$ is redundant:

$$\vec{v}_i = t_1\vec{v}_1 + \cdots + t_i\vec{v}_{i-1} = \mathrm{Span}(\vec{v}_1, \cdots, \vec{v}_{i-1}) = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_{i-1} \\ | & & | \end{bmatrix}\begin{bmatrix} t_1 \\ \vdots \\ t_{i-1}\end{bmatrix} = \boldsymbol{A}_{i-1}\vec{x}$$

$\Rightarrow$ We want to show: when $\vec{v}_i = \boldsymbol{A}_{i-1}\vec{x}$ has solutions. To solve $\vec{v}_i = \boldsymbol{A}_{i-1}\vec{x}$:

$$\left[\begin{array}{c:c} \boldsymbol{A}_{i-1} & \vec{v}_i \end{array}\right] \xrightarrow{\text{Row Reduce}} \begin{cases} \text{Consistent} \Rightarrow \text{Redundant} \Rightarrow \text{Do not contain pivot in } i\text{-th column} \\ \text{Inconsistent} \Rightarrow \text{Not redundant} \Rightarrow \text{Contain a pivot} \end{cases}$$

Example 4.1.5.

$$\boldsymbol{A} = \begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix} \Rightarrow \mathrm{rref}(\boldsymbol{A}) = \begin{bmatrix}\boxed{1}&0&-1\\0&\boxed{1}&2\\0&0&0\end{bmatrix}$$

$$\left[\begin{array}{c:c}1&2\\4&5\\7&8\end{array}\right] \Rightarrow \left[\begin{array}{c:c}1&2\\0&-3\\0&-6\end{array}\right] \Rightarrow \text{Inconsistent} \Rightarrow \text{Not redundant}$$

$$\left[\begin{array}{cc:c}1&2&3\\4&5&6\\7&8&9\end{array}\right] \Rightarrow \left[\begin{array}{cc:c}1&0&-1\\0&1&2\\0&0&0\end{array}\right] \Rightarrow \text{Consistent} \Rightarrow \text{Redundant}$$

■

Theorem 4.1.4 (Computing a basis for Ker($\boldsymbol{A}$)). Recall **Theorem** 3.3.4 Procedure to find $\ker(\boldsymbol{A})$.

1. The spanning set produced by "computing the kernel of $\boldsymbol{A}$" is a basis for $\mathrm{Ker}(\boldsymbol{A})$.

2. Procedure:

   (a) Row reduce $\boldsymbol{A}$ to $\mathrm{rref}(\boldsymbol{A})$, and then compute $\mathrm{Ker}(\mathrm{rref}(\boldsymbol{A}))$.

   (b) Unpack the equations encoded by matrix equation $\mathrm{rref}(\boldsymbol{A}) = 0$. Solve for pivot variables in terms of free variables.

(c) Parametrize the solution set for $\mathrm{rref}(\boldsymbol{A})\vec{x} = 0$ as $\{t_1\vec{v}_1 + t_2\vec{v}_2 + \cdots + t_d\vec{v}_d;\ t_i \in \mathbb{R}\}$ and $\vec{v}_i$ tracks the coefficient of the i-th free variable.

Proof. Look at the free variables $x_{i_1}, x_{i_2}, \cdots, x_{i_d}$. Then $\vec{v}_{i_j}$ is 0 if $j \neq k$; $\vec{v}_{i_j}$ is 1 if $j = k$. Thus,

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_{k-1}\vec{v}_{k-1} \neq \vec{v}_k.$$

■

Example 4.1.6. Let $\vec{v}_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}$ and $\vec{v}_4 = \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \end{bmatrix}$. Then, $c\vec{v}_3 \neq \vec{v}4$ since the 4-th position of $\vec{v}_3$ is 0, whereas that of $\vec{v}_4$ is 1.

4.2 The Rank-Nullity Theorem

Theorem 4.2.1. If V is a subspace of $\mathbb{R}^n$, then V has a basis, and all bases have the same size.

Definition 4.2.1 (The Dimension of a Subspace). Let V be a subspace, the **dimension** of V is the size of any bases. We denote it as $\dim(V)$.

Definition 4.2.2 (Rank of $\boldsymbol{A}$). Let $\boldsymbol{A}$ be an $n \times m$ matrix (i.e., $\boldsymbol{A} : \mathbb{R}^m \to \mathbb{R}^n$). The **rank** of $\boldsymbol{A}$ is the dimension of the image of A. We denote it as $\mathrm{rank}(\boldsymbol{A})$.

$$\mathrm{rank}(\boldsymbol{A}) = \dim(\mathrm{Im}(\boldsymbol{A}))$$

Definition 4.2.3 (Nullity of $\boldsymbol{A}$). Let $\boldsymbol{A}$ be an $n \times m$ matrix (i.e., $\boldsymbol{A} : \mathbb{R}^m \to \mathbb{R}^n$). The **nullity** of $\boldsymbol{A}$ is the dimension of the kernel of A. We denote it as $\mathrm{nullity}(\boldsymbol{A})$.

$$\mathrm{nullity}(\boldsymbol{A}) = \dim(\mathrm{Ker}(\boldsymbol{A}))$$

Theorem 4.2.2 (The Rank-Nullity Theorem). Suppose $\boldsymbol{A}$ to be an $n \times m$ matrix:

$$\boxed{\mathrm{rank}(\boldsymbol{A}) + \mathrm{nullity}(\boldsymbol{A}) = \dim(\text{domain of } \boldsymbol{A}) = m}.$$

Example 4.2.1. Let $\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$. To find basis for $\mathrm{Im}(\boldsymbol{A})$ and $\mathrm{Ker}(\boldsymbol{A})$:

$$\boldsymbol{A} \xrightarrow[\text{Reduce}]{\text{Row}} \begin{bmatrix} \boxed{1} & 0 & -1 \\ 0 & \boxed{1} & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

1. To find a basis for $\text{Im}(\boldsymbol{A})$, we take the columns of $\boldsymbol{A}$ which contain a pivot upon row reduction:

$$\text{Im}(\boldsymbol{A}) = \text{Span}\left(\begin{bmatrix}1\\4\\7\end{bmatrix}, \begin{bmatrix}2\\5\\8\end{bmatrix}\right).$$

$$\therefore \dim(\text{Im}(\boldsymbol{A})) = 2.$$

2. To find a basis for $\text{Ker}(\boldsymbol{A})$, unpack the equation:

$$\begin{cases} x_1 - x_3 = 0 \\ x_2 + 2x_3 = 0 \end{cases}, \implies \begin{cases} x_1 = x_3 \\ x_2 = -2x_3 \end{cases}.$$

$$\therefore \text{Ker}(\boldsymbol{A}) = \left\{\begin{bmatrix}x_3\\-2x_3\\x_3\end{bmatrix};\ x_3 \in \mathbb{R}\right\} = \text{Span}\left(\begin{bmatrix}1\\-2\\1\end{bmatrix}\right).$$

$$\therefore \dim(\text{Ker}(\boldsymbol{A})) = 1.$$

3. $\text{rank}(\boldsymbol{A}) = \dim(\text{Im}(\boldsymbol{A})) = 2;\ \text{nullity}(\boldsymbol{A}) = \dim(\text{Ker}(\boldsymbol{A})) = 1;\ \dim(\text{domain}) = 3$

$$\therefore \text{rank}(\boldsymbol{A}) + \text{nullity}(\boldsymbol{A}) = 3 = \dim(\text{domain}).$$

Proof.

1. $\text{rank}(\boldsymbol{A}) = \dim(\text{Im}(\boldsymbol{A}))$ = number of rectors in a basis of $\text{Im}(\boldsymbol{A})$ = number of pivots in $\text{rref}(\boldsymbol{A})$.

2. $\text{nullity}(\boldsymbol{A}) = \dim(\text{Ker}(\boldsymbol{A}))$ =number of rectors in a basis of $\text{Ker}(\boldsymbol{A})$ = number of free variables = number of non-pivot columns in $\text{rref}(\boldsymbol{A})$.

3. $\therefore \text{rank}(\boldsymbol{A})+\text{nullity}(\boldsymbol{A})$ = number of columns of $\text{rref}(\boldsymbol{A})$ or, simply, $\boldsymbol{A} = \dim(\text{domain of } \boldsymbol{A})$.

■

Example 4.2.2 (Geometric Perspective of Rank-Nullity Theorem). Let $\boldsymbol{M} = \begin{bmatrix}1 & 3\\2 & 6\end{bmatrix}$.

$$\therefore \text{rref}(\boldsymbol{M}) = \begin{bmatrix}1 & 3\\0 & 0\end{bmatrix}.$$

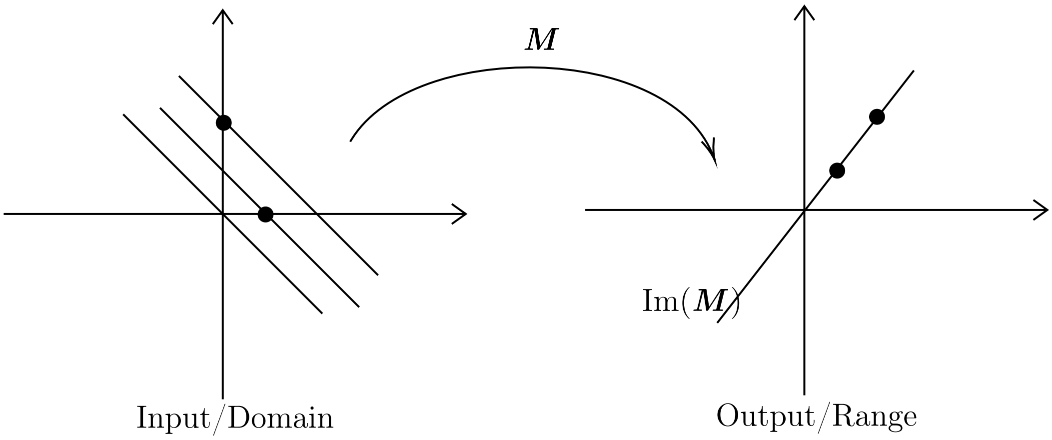
$$\therefore \text{Im}(\boldsymbol{M}) = \text{Span}\left(\begin{bmatrix}1\\2\end{bmatrix}\right) \text{(Line of slope 2 through the origin)} \implies \dim(\text{Im}(\boldsymbol{M})) = 1;$$

$$\text{Ker}(\boldsymbol{M}) = \text{Span}\left(\begin{bmatrix}-3\\1\end{bmatrix}\right) \text{(Line of slope } -\frac{1}{3} \text{ through origin)} \implies \dim(\text{Ker}(\boldsymbol{M})) = 1.$$

If we consider the domain of $\boldsymbol{M}$ to be the inputs for the transformation, and range of $\boldsymbol{M}$ ($\mathrm{Im}(\boldsymbol{M})$) to be the outputs of the linear transformation, then the rank-nullity theorem denotes that

$$\dim(\text{Inputs}) = \dim(\text{Outputs}) + \text{Information Loss}.$$

The "information loss" is given by $\dim(\mathrm{Ker}(\boldsymbol{M}))$. In this specific example, $\dim(\text{inputs}) = 2$ and $\dim(\text{outputs}) = 1$, so the information loss of the linear transformation $\boldsymbol{M}$ is $2 - 1 = 1$.



Theorem 4.2.3 (Invertibility Criteria). Let $\boldsymbol{A}$ be an $n \times m$ matrix:

1. $\boldsymbol{A}$ is invertible *iff* $\boldsymbol{A}\vec{x} = \vec{b}$ has a unique solution $\forall \vec{b} \in \mathbb{R}^n$.

$$\iff \mathrm{Im}(\boldsymbol{A}) = \mathbb{R}^n \quad \text{and} \quad \mathrm{Ker}(\boldsymbol{A}) = \{\vec{0}\}.$$

$$\iff \mathrm{rank}(\boldsymbol{A}) = n \quad \text{and} \quad \mathrm{nullity}(\boldsymbol{A}) = 0.$$

2. If $\boldsymbol{A}$ is an $n \times m$ matrix, then the following are equivalent:

   (a) $\boldsymbol{A}\vec{x} = \vec{b}$ has a unique solution for all $\vec{b}$ in $\mathbb{R}^n$.

   (b) $\mathrm{rank}(\boldsymbol{A}) = n$

   (c) $\mathrm{nullity}(\boldsymbol{A}) = 0$

   (d) $\mathrm{Im}(\boldsymbol{A}) = \mathbb{R}^n$

   (e) $\mathrm{Ker}(\boldsymbol{A}) = \{\vec{0}\}$

   (f) $\mathrm{rref}(\boldsymbol{A}) = \boldsymbol{I}_n$

   (g) The columns of $\boldsymbol{A}$ form a basis for $\mathbb{R}^n$

   (h) The columns of $\boldsymbol{A}$ span $\mathbb{R}^n$

   (i) The columns of $\boldsymbol{A}$ are L.I.

   (j) There is a matrix $\boldsymbol{B}$ *s.t.*

$$\boldsymbol{B}\boldsymbol{A} = \boldsymbol{I} = \boldsymbol{A}\boldsymbol{B} \qquad (\boldsymbol{B} \coloneqq \boldsymbol{A}^{-1})$$

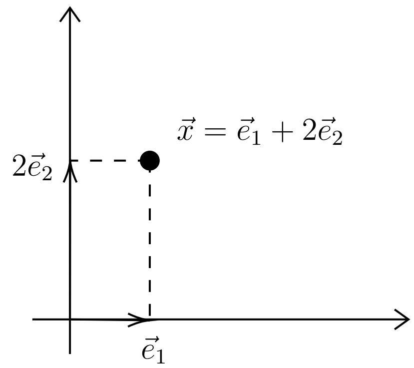
4.3 Coordinates

Remark (Goal of Coordinates). To describe the location of a vector within a subspace.

Definition 4.3.1 (Standard coordinates on $\mathbb{R}^n$). We can write $\vec{x}$ as a linear combination of the standard basis vectors.

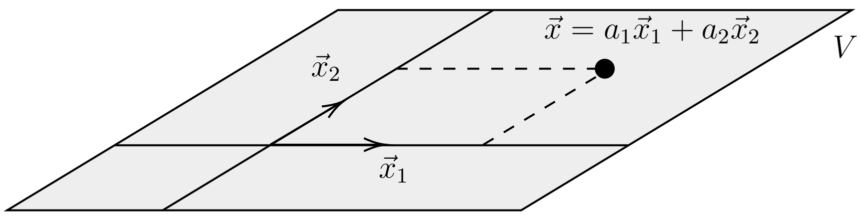
$$\text{i.e., } \vec{x} = a_1 + \vec{e}_1 + a_2\vec{e}_2 + \cdots + a_n\vec{e}_n;\ a_i \in \mathbb{R}.$$

Example 4.3.1. Suppose $\vec{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \in \mathbb{R}^2$. Then $\vec{x} = \vec{e}_1 + 2\vec{e}_2$.



Theorem 4.3.1. Let $V \subseteq \mathbb{R}^n$ be a subspace and $\beta = (\vec{x}_1, \cdots, \vec{x}_m)$ be a basis. Then every $\vec{x} \in V$ may be written as $\vec{x} = a_1\vec{x}_1 + a_2\vec{x}_2 + \cdots + a_m\vec{x}_m$ for some unique scalars $a_1, \cdots, a_m \in \mathbb{R}$.

Example 4.3.2. Suppose V is a subspace and $\beta = (\vec{x}_1, \vec{x}_2)$:



Definition 4.3.2 (β coordinates). Let $V \subseteq \mathbb{R}^n$ be a subspace and β be a basis for V. Let $\vec{x} \in V$. The **β-coordinates** for $\vec{x}$ in V is the following vector:

$$[\vec{x}]_\beta = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}$$

s.t. $\vec{x} = a_1\vec{x}_1 + \cdots + a_m\vec{x}_m$.

Example 4.3.3. Suppose $V = \text{Span}\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right)$ and $\beta = \left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right)$.

Let $\vec{x} = 2\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 1\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$. Then, $[\vec{x}]_\beta = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

Remark. V in general has many basis. The β-coordinates depend on the basis. Also, in general, coordinate axes are not perpendicular.

Example 4.3.4. Let $V \subseteq \mathbb{R}^3$ be the subspace spanned by $\vec{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} -3 \\ 2 \\ 3 \end{bmatrix}$. Let $\beta = (\vec{v}_1, \vec{v}_2)$ and $\vec{x} = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}$. Find $[\vec{x}]_\beta$.

Solution. Find $[\vec{x}]_\beta = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ *s.t.* $\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2$. (Find an expression for $\vec{x}$ in the span of $\vec{v}_1$ and $\vec{v}_2$, which is the image of $\boldsymbol{S} = \begin{bmatrix} | & | \\ \vec{v}_1 & \vec{v}_2 \\ | & | \end{bmatrix}$. Hence, we need to find $\vec{x} = \boldsymbol{S}\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ (i.e., solve $\boldsymbol{S}\vec{c} = \vec{x}$).
Form augmented matrix $\left[\begin{array}{c:c} \boldsymbol{S} & \vec{x} \end{array}\right]$:

$$\left[\begin{array}{c:c} \boldsymbol{S} & \vec{x} \end{array}\right] = \left[\begin{array}{cc:c} | & | & \\ \vec{v}_1 & \vec{v}_2 & \vec{x} \\ | & | & \end{array}\right] = \left[\begin{array}{cc:c} 1 & -3 & -1 \\ 2 & 2 & 2 \\ 1 & 3 & 2 \end{array}\right]$$

$$\xrightarrow[\text{reduce}]{\text{Row}} \left[\begin{array}{cc:c} 1 & 0 & 1/2 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{array}\right]$$

$$\therefore [\vec{x}]_\beta = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}.$$

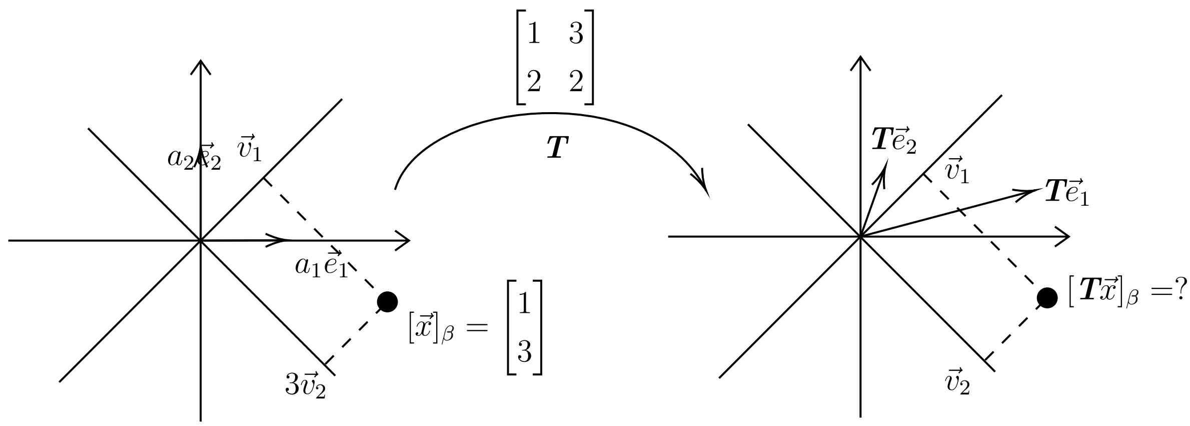
□

Remark. If $(\vec{v}_1, \cdots, \vec{v}_m) = \beta$ is a basis for a subspace V, and $\boldsymbol{S} := \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_m \\ | & & | \end{bmatrix}$, then $\boldsymbol{S}$ converts β-coordinates to standard coordinates.

$$\text{i.e., } \boldsymbol{S}[\vec{x}]_\beta = \vec{x}.$$

Example 4.3.5 ($\boldsymbol{\beta}$-coordinates Under Linear Transformation). Consider $\boldsymbol{T} : \mathbb{R}^2 \to \mathbb{R}^2$ defined by matrix $\begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix}$. Let $\vec{x} \in \mathbb{R}^2$ be the vector whose β-coordinates are $[\vec{x}]_\beta = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$, where $\beta = \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)$. Find $[\boldsymbol{T}\vec{x}]_\beta$.

Solution. First, unpack the question:



To solve this question:

1. Find standard coordinates for $\vec{x}$:

$$\boldsymbol{S} = \begin{bmatrix}1 & 1\\1 & -1\end{bmatrix} \Longrightarrow [\vec{x}]_{st} = \boldsymbol{S}[\vec{x}]_\beta = \begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}\begin{bmatrix}1\\3\end{bmatrix} = \begin{bmatrix}4\\-2\end{bmatrix}$$

2. Multiply $[\vec{x}]_{st}$ by $\boldsymbol{T}$:

$$\boldsymbol{T}[\vec{x}]_{st} = \begin{bmatrix}1 & 3\\2 & 2\end{bmatrix}\begin{bmatrix}4\\-2\end{bmatrix} = \begin{bmatrix}-2\\-4\end{bmatrix}$$

3. Compute $[\boldsymbol{T}\vec{x}]_\beta$.

$$\boldsymbol{T}[\vec{x}]_{st} = \boldsymbol{S}[\boldsymbol{T}\vec{x}]_\beta \Longrightarrow \begin{bmatrix}-2\\-4\end{bmatrix} = \begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}[\boldsymbol{T}\vec{x}]_\beta \Longrightarrow [\boldsymbol{T}\vec{x}]_\beta = \begin{bmatrix}1\\3\end{bmatrix}.$$

□

Theorem 4.3.2. Let $\boldsymbol{T} : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and $\beta = (\vec{v}_1, \cdots, \vec{v}_n)$ be a basis for $\mathbb{R}^n$. Let $\vec{x} \in \mathbb{R}^n$:

$$[\boldsymbol{T}\vec{x}]_\beta = \boldsymbol{S}^{-1}\boldsymbol{T}\boldsymbol{S}[\vec{x}]_\beta, \qquad \text{where} \quad \boldsymbol{S} = \begin{bmatrix}| & & |\\ \vec{v}_1 & \cdots & \vec{v}_n\\ | & & |\end{bmatrix}.$$

$$[\boldsymbol{T}\vec{x}]_\beta = [\boldsymbol{T}]_\beta[\vec{x}]_\beta.$$

Theorem 4.3.3. The matrix for $\boldsymbol{T}$ with respect to the basis β is

$$[\boldsymbol{T}]_\beta = \boldsymbol{S}^{-1}\boldsymbol{T}\boldsymbol{S}.$$

Example 4.3.6. Let $\boldsymbol{T} = \begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix}$ and $\beta = \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right)$. Then

$$[\boldsymbol{T}]_\beta = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 4 & -1 \\ 0 & -1 \end{bmatrix}.$$

5 Approximate Solution of $\boldsymbol{A}\vec{x} = \vec{b}$

5.1 Lengths and Angles in $\mathbb{R}^n$

Definition 5.1.1 (Dot Product). Let $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n$. The **dot product** of $\vec{x}$ and $\vec{y}$ is the following number:

$$\vec{x} \cdot \vec{y} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Example 5.1.1.

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 7 \\ 5 \\ 2 \end{bmatrix} = 1 \times 7 + 2 \times 5 + 3 \times 2 = 23.$$

Theorem 5.1.1. Algebraic property of dot products:

1. $\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}$
2. $\vec{x} \cdot (\vec{y}_1 + \vec{y}_2) = \vec{x} \cdot \vec{y}_1 + \vec{x} \cdot \vec{y}_2$
   $(\vec{x}_1 + \vec{x}_2) \cdot \vec{y} = \vec{x}_1 \cdot \vec{y} + \vec{x}_2 \cdot \vec{y}$
3. $\vec{x} \cdot (c\vec{y}) = c(\vec{x} \cdot \vec{y}) = (c\vec{x}) \cdot \vec{y}$

Definition 5.1.2 (Length). Let $\vec{x} \in \mathbb{R}^n$. The **length** of $\vec{x}$ is the following number:

$$\|\vec{x}\| := \sqrt{\vec{x} \cdot \vec{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}, \text{ where } \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Example 5.1.2.

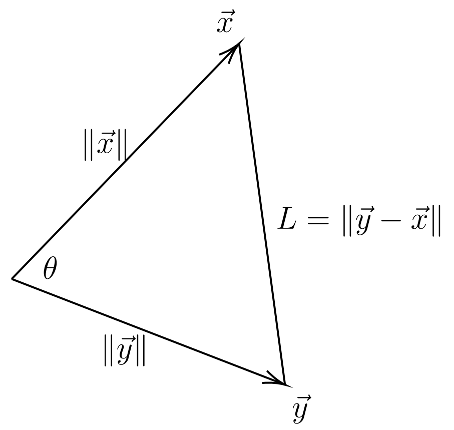
$$\left\| \begin{bmatrix} 4 \\ 3 \end{bmatrix} \right\| = \sqrt{4^2 + 3^2} = 5$$

Remark. In $\mathbb{R}^2$, the definition of length is the Pythagorean theorem.

Theorem 5.1.2 (Angle Between Vectors). Let θ be the angle between $\vec{x}$ and $\vec{y}$. We then have

$$\cos\theta = \frac{\|\vec{x}\|^2 + \|\vec{y}\|^2 - \|\vec{y} - \vec{x}\|^2}{2\|\vec{x}\|\|\vec{y}\|}$$

Proof. Assume vectors $\vec{x}$ and $\vec{y}$ are drawn as below.



By the cosine rule, we have:

$$L^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2 - 2\|\vec{x}\|\|\vec{y}\|\cos\theta$$

So,

$$\cos\theta = \frac{\|\vec{x}\|^2 + \|\vec{y}\|^2 - L^2}{2\|\vec{x}\|\|\vec{y}\|} = \frac{\|\vec{x}\|^2 + \|\vec{y}\|^2 - \|\vec{y} - \vec{x}\|^2}{2\|\vec{x}\|\|\vec{y}\|}.$$

■

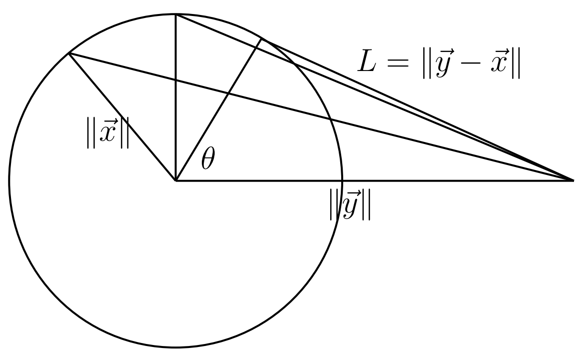
Theorem 5.1.3. Relationship of angle and dot products:

1. $\vec{x} \cdot \vec{y} > 0$ if $\theta < 90°$
2. $\vec{x} \cdot \vec{y} = 0$ if $\theta = 90°$
3. $\vec{x} \cdot \vec{y} < 0$ if $\theta > 90°$

Proof.

$$\begin{aligned}
\|\vec{y} - \vec{x}\|^2 &= (\vec{y} - \vec{x}) \cdot (\vec{y} - \vec{x}) \\
&= (\vec{y} - \vec{x}) \cdot \vec{y} - (\vec{y} - \vec{x}) \cdot \vec{x} \\
&= \vec{y} \cdot \vec{y} - \vec{x} \cdot \vec{y} - \vec{y} \cdot \vec{x} + \vec{x} \cdot \vec{x} \\
&= \vec{y} \cdot \vec{y} - 2\vec{x} \cdot \vec{y} + \vec{x} \cdot \vec{x} \\
&= \|\vec{y}\|^2 - 2\vec{x} \cdot \vec{y} + \|\vec{x}\|^2
\end{aligned}$$

Think of Pythagonean theorem:



- If $\theta < 90°$, $\|\vec{y} - \vec{x}\|^2 < \|\vec{y}\| + \|\vec{}\|^2 \implies \vec{x} \cdot \vec{y} > 0$.

- If $\theta = 90°$, $\|\vec{y} - \vec{x}\|^2 = \|\vec{y}\| + \|\vec{}\|^2 \implies \vec{x} \cdot \vec{y} = 0$.
- If $\theta > 90°$, $\|\vec{y} - \vec{x}\|^2 > \|\vec{y}\| + \|\vec{}\|^2 \implies \vec{x} \cdot \vec{y} < 0$.

■

Definition 5.1.3 (Perpendicular). Let $\vec{x}, \vec{y} \in \mathbb{R}^n$. Then, $\vec{x}$ and $\vec{y}$ are **perpendicular** *iff* $\vec{x} \cdot \vec{y} = 0$. (Equivalently: orthogonal)

Theorem 5.1.4. Suppose $\boldsymbol{A}$ is an $1 \times n$ matrix *s.t.* $\boldsymbol{A} = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}$. Then, $\boldsymbol{A}^{\boldsymbol{T}} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \vec{v}$. Thus, $\boldsymbol{A}\vec{x} = \vec{v} \cdot \vec{x}$.

Theorem 5.1.5.

$$\vec{v} \perp \vec{x} \iff \vec{v} \cdot \vec{x} = 0 \iff \boldsymbol{A}\vec{x} = 0 \implies \vec{x} \in \operatorname{Ker}(\boldsymbol{A}).$$

- Let $\vec{v} \neq \vec{0}$. The set $\{\vec{x} \mid \vec{x} \perp \vec{v}\}$ is a subspace of dimension $m - 1$.
- Let $\boldsymbol{A} : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. Then, the kernel of $\boldsymbol{A}$ is the set of all vectors $\vec{x} \in \mathbb{R}^m$, which are perpendicular to the row of the matrix for $\boldsymbol{A}$.

Theorem 5.1.6.

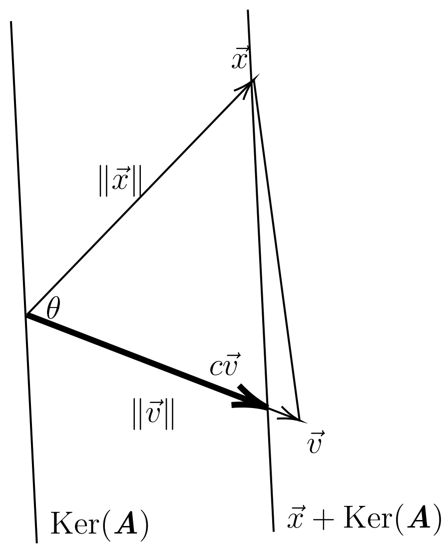
$$\vec{x} \cdot \vec{v} = \|\vec{x}\|\|\vec{v}\| \cos\theta$$

Proof.

1. $\vec{v} \cdot \vec{x}$ is constant along translates of the subspace perpendicular to the line spanned by $\vec{v}$:

$$\vec{v} \cdot \vec{x} = \boldsymbol{A}\vec{x} = \vec{b}$$

2. Project $\vec{x}$ into the line spanned by $\vec{v}$:



3. Use trigonometry to calculate the projection:

$$c\vec{v} = (\|\vec{x}\| \cos\theta) \left(\frac{\vec{v}}{\|\vec{v}\|} \right)$$

$$\begin{aligned}
\vec{v} \cdot \vec{x} &= \vec{v} \cdot c\vec{v} \\
&= \frac{\|\vec{x}\| \cos\theta}{\|\vec{v}\|} \|\vec{v}\|^2 = \|\vec{x}\| \cdot \|\vec{v}\| \cos\theta \\
&\Rightarrow \; \theta = \arccos\left(\frac{\vec{x}\vec{v}}{\|\vec{x}\| \|\vec{v}\|} \right)
\end{aligned}$$

■

Theorem 5.1.7. Projection of $\vec{x}$ into line spanned by $\vec{v}$ is given by the following formula:

$$\begin{aligned}
\text{Projection} = c\vec{v} &= \frac{\|\vec{x}\| \cos\theta}{\|\vec{v}\|} \vec{v} \\
&= \frac{\vec{x} \cdot \vec{v}}{\|\vec{v}\|^2} \vec{v} \\
&= \frac{\vec{x} \cdot \vec{v}}{\vec{v} \cdot \vec{v}} \vec{v}.
\end{aligned}$$

Definition 5.1.4 (Orthogonal Complement). Let $V \subseteq \mathbb{R}^n$ be a subspace. The **orthogonal complement** of V is the set of vectors perpendicular to all vectors in V:

$$V^\perp = \left\{ \vec{x} \in \mathbb{R}^n ; \; \vec{v} \cdot \vec{x} = 0 \quad \forall \vec{v} \in \mathbb{R}^n \right\}.$$

Example 5.1.3. The orthogonal complement of a line with a slope m through the origin is a line through the origin with a slop of $-\dfrac{1}{m}$.

Theorem 5.1.8. Let V be a subspace. If $\vec{v}_1, \cdots, \vec{v}_k \in V$ is a spanning set, (i.e., $V = \text{Span}(\vec{v}_1, \cdots, \vec{v}_k)$), then $\vec{x} \in V^\perp$ *iff* $\vec{v}_1 \cdot \vec{x} = 0, \; \vec{v}_2 \cdot \vec{x} = 0, \cdots, \vec{v}_k \cdot \vec{x} = 0$.

Proof. "Perpendicular to everything" implies $\vec{v}_i \cdot \vec{x} = 0 \quad \forall \vec{v} \in V$, then $\vec{v} = c_1\vec{v}_1 + \cdots + c_k\vec{v}_k \implies \vec{x} \cdot \vec{v} = c_1(\vec{x} \cdot \vec{v}_1) + \cdots + c_k(\vec{x} \cdot \vec{v}_k) = 0 \implies \therefore \vec{x} \perp \vec{v}$. ■

Theorem 5.1.9. Let $V \subseteq \mathbb{R}^n$ be a subspace, $V^\perp$ is a subspace. Specifically, if $V = \text{Span}(\vec{v}_1, \cdots, \vec{v}_k)$, then

$$V^\perp = \text{Ker}\left(\begin{bmatrix} - & \vec{v}_1 & - \\ - & \vec{v}_2 & - \\ & \vdots & \\ - & \vec{v}_k & - \end{bmatrix} \right)$$

Proof.

$$\begin{bmatrix} - & \vec{v}_1 & - \\ - & \vec{v}_2 & - \\ & \vdots & \\ - & \vec{v}_k & - \end{bmatrix} \begin{bmatrix} | \\ \vec{x} \\ | \end{bmatrix} = \begin{bmatrix} \vec{x}\cdot\vec{v}_1 \\ \vdots \\ \vec{x}\cdot\vec{v}_k \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

■

Example 5.1.4. Let $V = \text{Span}\left(\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \end{bmatrix}\right)$. Compute $V^{\perp}$. *Solution.*

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & -1 & 0 & 1 \end{bmatrix} \xrightarrow[\text{reduce}]{\text{Row}} \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & -1 & -1 & 0 \end{bmatrix} \Longrightarrow \text{rref} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

Unpack, we have

$$\begin{cases} x_1 = -x_3 - x_4 \\ x_2 = -x_3 \end{cases}$$

$$\therefore V^{\perp} = \text{Kernel} = \left\{ \begin{bmatrix} -x_3 - x_4 \\ -x_3 \\ x_3 \\ x_4 \end{bmatrix}; \ x_{3,4} \in \mathbb{R} \right\} = \text{Span}\left(\begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}\right)$$

□

5.2 Orthogonal Projection

Theorem 5.2.1. Let $V \subseteq \mathbb{R}^n$ be a subspace and $\vec{x} \in \mathbb{R}^n$. Then, $\vec{x}$ can be written uniquely as

$$\vec{x} = \vec{x}^{\parallel} + \vec{x}^{\perp},$$

when $\vec{x}^{\parallel} \in V$ and $\vec{x}^{\perp} \in V$.

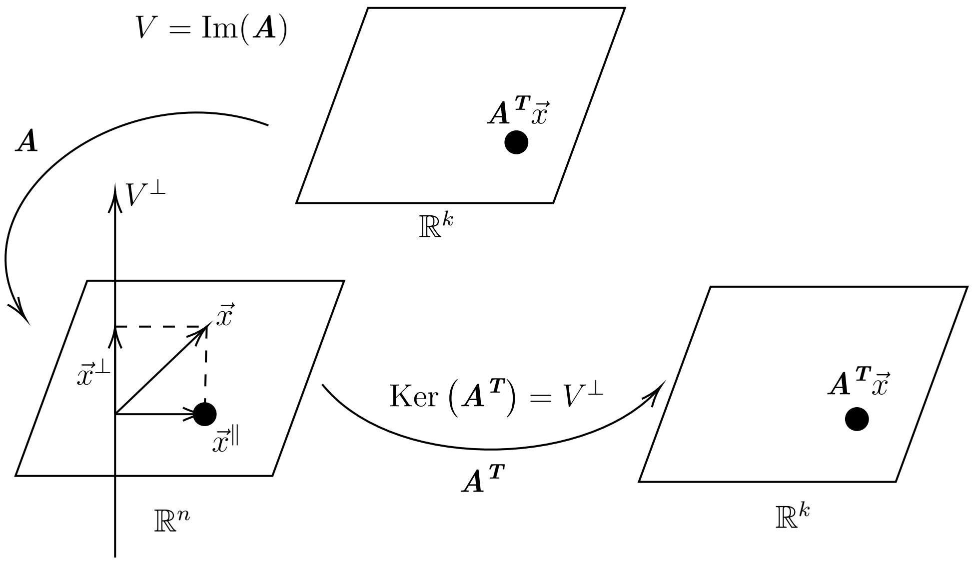
Definition 5.2.1 (Orthogonal Projection). Let $V \subseteq \mathbb{R}^n$ be a subspace. The **orthogonal projection** of $\vec{x}$ into V is the vector $\vec{x}^{\parallel}$. The map $\vec{x} \mapsto \vec{x}^{\parallel}$ is denoted as $\text{Proj}_V : \mathbb{R}^n \to \mathbb{R}^n$.

Theorem 5.2.2. Computing $\text{Proj}_V(\vec{x}) \coloneqq \vec{x}^{\parallel}$:

1. Let $\vec{v}_1, \cdots, \vec{v}_k$ be a basis for V:

$$\boldsymbol{A^T} = \begin{bmatrix} - & \vec{v}_1 & - \\ & \vdots & \\ - & \vec{v}_k & - \end{bmatrix} \quad \text{and} \quad \boldsymbol{A} = \begin{bmatrix} | & & | \\ \vec{v}_1 & cdots & \vec{v}_k \\ | & & | \end{bmatrix}$$

$$\boxed{\boldsymbol{A^T}\vec{x}^{\parallel} = \boldsymbol{A^T}\vec{x}}$$



2. Slove $\boldsymbol{A^T A}\vec{c} = \boldsymbol{A^T}\vec{x}$ for $\vec{c}$.

3. $\vec{x}^{\parallel} = \boldsymbol{A}\vec{c}$

Example 5.2.1. Let $V = \mathrm{Span}\left(\begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix}\right)$ and $\vec{x} = \begin{bmatrix}1\\2\\3\end{bmatrix}$. Compute the projection of $\vec{x}$ onto V.

Solution.

1.
$$\boldsymbol{A} = \begin{bmatrix}1 & 0\\1 & 1\\0 & 1\end{bmatrix} \quad \text{and} \quad \boldsymbol{A^T} = \begin{bmatrix}1 & 1 & 0\\0 & 1 & 1\end{bmatrix}$$

2. Compute $\boldsymbol{A^T A}$ and $\boldsymbol{A^T}\vec{x}$:

$$\boldsymbol{A^T A} = \begin{bmatrix}1 & 1 & 0\\0 & 1 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\1 & 1\\0 & 1\end{bmatrix} = \begin{bmatrix}2 & 1\\1 & 2\end{bmatrix}$$

$$\boldsymbol{A^T}\vec{x} = \begin{bmatrix}1 & 1 & 0\\0 & 1 & 1\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix} = \begin{bmatrix}3\\5\end{bmatrix}$$

3. Solve $\boldsymbol{A^T A}\vec{c} = \boldsymbol{A^T}\vec{x}$ for $\vec{c}$:

$$\begin{bmatrix}2 & 1\\1 & 2\end{bmatrix}\vec{c} = \begin{bmatrix}3\\5\end{bmatrix} \Longrightarrow \left[\begin{array}{cc:c}2 & 1 & 3\\1 & 2 & 5\end{array}\right] \xrightarrow[\text{reduce}]{\text{Row}} \left[\begin{array}{cc:c}1 & 0 & {}^1/_3\\0 & 1 & {}^7/_3\end{array}\right]$$

$$\therefore \vec{c} = \begin{bmatrix} 1/3 \\ 7/3 \end{bmatrix}$$

4. Compute $\boldsymbol{A}\vec{c} = \vec{x}^{\parallel}$

$$\vec{x}^{\parallel} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/3 \\ 7/3 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 8/3 \\ 7/3 \end{bmatrix}$$

$$\therefore \vec{x}^{\perp} = \vec{x} - \vec{x}^{\parallel} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 1/3 \\ 8/3 \\ 7/3 \end{bmatrix} = \begin{bmatrix} 2/3 \\ -2/3 \\ 2/3 \end{bmatrix}$$

□

Definition 5.2.2 (Transpose of a Matrix). Let $\boldsymbol{A}$ be an $n \times m$ matrix. The **transpose** of $\boldsymbol{A}$ is the $m \times n$ matrix $\boldsymbol{A^T}$ whose rows are the columns of $\boldsymbol{A}$:

$$\boldsymbol{A} = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_k \\ | & & | \end{bmatrix}; \ \boldsymbol{A^T} = \begin{bmatrix} - & \vec{v}_1 & - \\ & \vdots & \\ = & \vec{v}_k & - \end{bmatrix}$$

Equivalently, the ij-entry of $\boldsymbol{A}$ is the ji-entry of $\boldsymbol{A^T}$.
Equivalently, whose columns are rows of $\boldsymbol{A}$.

Theorem 5.2.3.

$$\mathrm{Ker}(\boldsymbol{A^T}) = \mathrm{Im}(\boldsymbol{A})^{\perp}$$

Remark. In general, if rank$(\boldsymbol{A})$ is less than the dimension of range, small perturbations of any $\vec{b} \in \mathrm{Im}(\boldsymbol{A})$ lie outside the image of $\boldsymbol{A}$. In such cases, rather than try to find $\vec{x}$ *s.t.* $\boldsymbol{A}\vec{x} = \vec{b}$, try to find $\vec{x}$ *x.t.* $\boldsymbol{A}\vec{x}$ is as close as to $\vec{b}$ as possible. Problem: Find $\vec{x}$ *s.t.* $\|\boldsymbol{A}\vec{x} - \vec{b}\|$ is as small as possible (minimized).

- The solution agrees with solving $\boldsymbol{A}\vec{x} = \vec{b}$ when there are solutions.
- This question always has solutions.

Solution.

1. Find $\vec{b}* \in \mathrm{Im}(\boldsymbol{A})$ which are as close as to $\vec{b}$ as possible.

   Theorem 5.2.4. Let $\boldsymbol{A}$ be an $n \times m$ matrix and $\vec{b} \in \mathbb{R}^m$. The closest vector to $\vec{b}$ in $\mathrm{Im}(\boldsymbol{A})$ is $\vec{b}* = \mathrm{Proj}_{\mathrm{Im}(\boldsymbol{A})}(\vec{b}) = \vec{b}^{\parallel}$

2. Solve $\boldsymbol{A}\vec{x} = \vec{b}*$

$\square$

Solution. (Advanced approach).

1. Approximate solutions to $\boldsymbol{A}\vec{x} = \vec{b}$
   $\Longleftrightarrow$ Solutions $\boldsymbol{A}\vec{x} = \vec{b}^{\parallel}$ where $\vec{b}^{\parallel} \in \mathrm{Im}(\boldsymbol{A})$
   $\longrightarrow \boldsymbol{A}\vec{x} - \vec{b} = \vec{b}^{\perp}$ equivalently $\boldsymbol{A}\vec{x} - \vec{b}$ is perpendicular to $\mathrm{Im}(\boldsymbol{A})$
   $\longrightarrow \boldsymbol{A^T}(\boldsymbol{A}\vec{x} - \vec{b}) = 0$
   $$\boxed{\text{i.e., } \boldsymbol{A^T}\boldsymbol{A}\vec{x} = \boldsymbol{A}^T\vec{b}}$$

2. The approximate solutions to $\boldsymbol{A}\vec{x} = \vec{b}$ are exactly the solutions to $\boldsymbol{A^T}\boldsymbol{A}\vec{x} = \boldsymbol{A^T}\vec{b}$.

$\square$

Example 5.2.2. Let $\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 1 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. Find all approximate solutions to $\boldsymbol{A}\vec{x} = \vec{b}$.

Solution.

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 1 \end{bmatrix} \vec{x} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 4 \\ 4 & 6 \end{bmatrix} \vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Solve the equation, we have $\vec{x} = \begin{bmatrix} 1 \\ -\frac{1}{2} \end{bmatrix}$ as the unique approximate solution to $\boldsymbol{A}\vec{x} = \vec{b}$. $\square$

5.3 Graph Fitting

Example 5.3.1. Consider the following data set:

x	y
0	0
1	0
2	1

Find a quadratic polynomial $f(x) = Ax^2 + Bx + C$ (i.e., find $A, B, C \in \mathbb{R}$) $s.t.$ $f(x) = y \quad \forall x$ in the data set.

Solution. Plug-in data points to $f(x) = Ax^2 + Bx + C$ to obtain algebraic relations between A, B, and C.

$$\begin{cases} 0A + 0B + C = f(0) = 0 \\ 1A + 1B + C = f(1) = 0 \\ 4A + 2B + C = f(2) = 1 \end{cases}$$

We can form a system of linear equations:

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \Longrightarrow \left[\begin{array}{ccc:c} 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 4 & 2 & 1 & 1 \end{array}\right] \xrightarrow[\text{reduce}]{\text{Row}} \text{rref} = \left[\begin{array}{ccc:c} 1 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & -\frac{1}{2} \\ 0 & 0 & 1 & 0 \end{array}\right]$$

$$\therefore A = \frac{1}{2}, B = -\frac{1}{2}, C = 0$$

$$\therefore f(x) = \frac{1}{2}x^2 - \frac{1}{2}x = \frac{1}{2}x(x-1)$$

□

Theorem 5.3.1 (Fundamental Problem of Graph fitting). Given some data set $(x_1, y_1), \cdots, (x_m, y_m) \in \mathbb{R}^2$ and functions $f_1, \cdots, f_n : \mathbb{R} \to \mathbb{R}$. Find a function $f : \mathbb{R} \to \mathbb{R}$ *s.t.*: 1. $f(x_i) = y_i$, and 2. $f = A_1 f_1 + \cdots + A_n f_n$.

To solve this, plug-in data points and get a matrix equation as following:

$$\begin{bmatrix} f_1(x_1) & \cdots & f_n(x_1) \\ \vdots & \ddots & \vdots \\ f_1(x_m) & \cdots & f_n(x_m) \end{bmatrix} \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

Example 5.3.2. Consider the following data set:

x	y
0	0
1	0
2	0
3	1

Find a quadratic polynomial $f(x) = Ax^2 + Bx + C$ (i.e., find $A, B, C \in \mathbb{R}$) *s.t.* $f(x) = y \quad \forall x, y$ in the data set.

Solution. Plug-in data points:

$$\begin{cases} 0A + 0B + C = f(0) = 0 \\ 1A + 1B + C = f(1) = 0 \\ 4A + 2B + C = f(2) = 0 \\ 9A + 3B + C = f(3) = 1 \end{cases}$$

Form a matrix equation:

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \Rightarrow \left[\begin{array}{ccc:c} 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 4 & 2 & 1 & 0 \\ 9 & 3 & 1 & 1 \end{array}\right] \xrightarrow[\text{reduce}]{\text{Row}} \left[\begin{array}{ccc:c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

$\therefore$ There's no solution. $\square$

Example 5.3.3. Using the same data set from Example 5.3.2, find a quadratic polynomial *s.t.* the distance between $\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} f(0) \\ f(1) \\ f(2) \\ f(3) \end{bmatrix}$ is minimized.

Solution. This problem is equivalent to the least squares problems (finding the best approximate solution to $\boldsymbol{A}\vec{x} = \vec{b}$). Solve $\boldsymbol{A^T A}\vec{x} = \boldsymbol{A^T}\vec{b}$.

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{bmatrix}; \ \boldsymbol{A^T} = \begin{bmatrix} 0 & 1 & 4 & 9 \\ 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\boldsymbol{A^T A} = \begin{bmatrix} 0 & 1 & 4 & 9 \\ 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 98 & 36 & 14 \\ 36 & 24 & 6 \\ 14 & 6 & 4 \end{bmatrix}$$

Remark. $\boldsymbol{A^T A}$ is **symmetric across diagonal**, meaning a_ij entry is equal to a_ji entry.

$$\vec{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \Longrightarrow \boldsymbol{A^T}\vec{b} = \begin{bmatrix} 0 & 1 & 4 & 9 \\ 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 9 \\ 3 \\ 1 \end{bmatrix}$$

Form a matrix equation:

$$\left[\begin{array}{ccc:c} 98 & 36 & 14 & 9 \\ 36 & 24 & 6 & 3 \\ 14 & 6 & 4 & 1 \end{array}\right] \xrightarrow[\text{reduce}]{\text{Row}} \left[\begin{array}{ccc:c} 1 & 0 & 0 & \frac{1}{4} \\ 0 & 1 & 0 & -\frac{9}{20} \\ 0 & 0 & 1 & \frac{1}{20} \end{array}\right]$$

$$\therefore f(x) = \frac{1}{4}x^2 - \frac{9}{20}x + \frac{1}{20}$$

$$\therefore \begin{bmatrix} f(0) \\ f(1) \\ f(2) \\ f(3) \end{bmatrix} = \begin{bmatrix} 0.05 \\ -0.15 \\ 0.15 \\ 0.95 \end{bmatrix}$$

The distance between these vectors is minimized:

$$d = \sqrt{(0 - 0.05)^2 + (0 + 0.15)^2 + (0 - 0.15)^2 + (1 - 0.95)^2} \approx 0.2236$$

That is, **error**≈ 0.2236. $\square$

Theorem 5.3.2 (General Problem of Graph Fitting). Given a data set $(x_1, y_1), \cdots, (x_m, y_m) \in \mathbb{R}^2$ and functions $f_1, \cdots, f_n : \mathbb{R} \to \mathbb{R}$. Find a function $f : \mathbb{R} \to \mathbb{R}$ *s.t.*: 1. $f = A_1 f_1 + \cdots + A_n f_n$, and 2. $\begin{bmatrix} f(x_1) \\ \vdots \\ f(x_m) \end{bmatrix}$ and $\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$ are as close as possible.

To solve this problem, form a matrix equation and solve for its best approximate solutions:

$$\underbrace{\begin{bmatrix} f_1(x_1) & \cdots & f_n(x_1) \\ \vdots & \ddots & \vdots \\ f_1(x_m) & \cdots & f_n(x_m) \end{bmatrix}}_{\boldsymbol{A}} \underbrace{\begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix}}_{\vec{x}} = \underbrace{\begin{bmatrix} y_1 \\ vdots \\ y_m \end{bmatrix}}_{\vec{b}}$$

Solve for the normal equation

$$\boldsymbol{A^T A}\vec{x} = \boldsymbol{A^T}\vec{b}$$

5.4 Orthogonal Linear Transformation

Definition 5.4.1 (Orthogonal Transformation). Let $\boldsymbol{T} : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. $\boldsymbol{T}$ is called an **orthogonal transformation** if

$$\boldsymbol{T}(\vec{x}) \cdot \boldsymbol{T}(\vec{y}) = \vec{x} \cdot \vec{y} \quad \forall \vec{x}, \vec{y} \in \mathbb{R}^n.$$

Equivalently, $\boldsymbol{T}$ is orthogonal *iff* $\boldsymbol{T}$ preserves lengths and angles.

Example 5.4.1. Rotations and reflections in $\mathbb{R}^2$ are orthogonal. Reflections through a subspace $V \subseteq \mathbb{R}^n$ is also orthogonal.

Definition 5.4.2. Let $V \subseteq \mathbb{R}^n$ be a subspace and $\text{Proj}_V : \mathbb{R}^n \to \mathbb{R}^n$ and $\text{Proj}_{V^\perp} : \mathbb{R}^{\to}\mathbb{R}^n$ be the orthogonal projections into V and $V^\perp$, respectively. We define $\text{Ref}_V : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\text{Ref}_V(\vec{x}) = \text{Proj}_V(\vec{x}) - \text{Proj}_{V^\perp}(\vec{x})$$

Theorem 5.4.1 (Property of $\mathbf{Ref}_{\boldsymbol{V}}$). Ref_V is an orthogonal linear transformation.

Proof.

1. It's linear because the projections are linear:

$$\begin{aligned} \text{Ref}_V(\vec{x} + \vec{y}) &= \text{Proj}_V(\vec{x} + \vec{y}) - \text{Proj}_{V^\perp}(\vec{x} + \vec{y}) \\ &= \text{Proj}_V(\vec{x}) + \text{Proj}_V(\vec{y}) - \text{Proj}_{V^\perp}(\vec{x}) - \text{Proj}_{V^\perp}(\vec{y}) = \text{Ref}_V(\vec{x}) + \text{Ref}_V(\vec{y}) \end{aligned}$$

$$\text{Ref}_V(c\vec{x}) = \text{Proj}_V(c\vec{x}) - \text{Proj}_{V^\perp}(c\vec{x}) = c\text{Proj}_V(\vec{x}) - c\text{Proj}_{V^\perp}(\vec{x}) = c\text{Ref}_V(\vec{x})$$

2. It's orthogonal $\longleftrightarrow$ preserve lengths and angles

$$\begin{aligned}\vec{x}\cdot\vec{y} &= (\vec{x}^{\parallel}+\vec{x}^{\perp})\cdot(\vec{y}^{\parallel})+\vec{y}^{\perp})\\ &= \vec{x}^{\parallel}\cdot\vec{y}^{\parallel}+\underbrace{\vec{x}^{\perp}\cdot\vec{y}^{\parallel}}_{0}+\underbrace{\vec{x}^{\parallel}\cdot\vec{y}^{\perp}}_{0}+\vec{x}^{\perp}\cdot\vec{y}^{\perp}=\vec{x}^{\parallel}\cdot\vec{y}^{\parallel}+\vec{x}^{\perp}\cdot\vec{y}^{\perp}\end{aligned}$$

$$\begin{aligned}\mathrm{Ref}_V(\vec{x})\cdot\mathrm{Ref}_V(\vec{y}) &= (\vec{x}^{\parallel})-\vec{x}^{\perp})\cdot(\vec{y}^{\parallel}-\vec{y}^{\perp})\\ &= \vec{x}^{\parallel}\cdot\vec{y}^{\parallel}-\underbrace{\vec{x}^{\parallel}\cdot\vec{y}^{\perp}}_{0}-\underbrace{\vec{x}^{\perp}\cdot\vec{y}^{\parallel}}_{0}+\vec{x}^{\perp}\cdot\vec{y}^{\perp}=\vec{x}^{\parallel}\cdot\vec{y}^{\parallel}+\vec{x}^{\perp}\cdot\vec{y}^{\perp}\end{aligned}$$

$$\therefore \mathrm{Ref}_V(\vec{x})\cdot\mathrm{Ref}_V(\vec{y}) = \vec{x}\cdot\vec{y}.$$

■

Definition 5.4.3 (Orthogonal Matrices). Orthogonal Matrices are matrices encoding orthogonal linear transformations.

Theorem 5.4.2. If $\boldsymbol{T}:\mathbb{R}^n\to\mathbb{R}^n$ is orthogonal, the matrix for $\boldsymbol{T}$ is $\begin{bmatrix} | & & | \\ \boldsymbol{T}(\vec{e}_1) & \cdots & \boldsymbol{T}(\vec{e}_n) \\ | & & | \end{bmatrix}$.

The lengths and angles of these vectors are the same as $\vec{e}_1,\cdots,\vec{e}_n$ if $\boldsymbol{T}$ is orthogonal.

Theorem 5.4.3.

$$\vec{e}_i\cdot\vec{e}_j = \begin{cases} 1, & i=j \\ 0, & i\neq j \end{cases}.$$

Equivalently, $\vec{e}_i\perp\vec{e}_j$ if $i\neq j$ and $\|\vec{e}_i\|=\sqrt{\vec{e}_i\cdot\vec{e}_i}=1$.

Extension. Let $\vec{v}_1,\cdots,\vec{v}_k$ be vectors in $\mathbb{R}^n$, we say $\vec{v}_1,\cdots,\vec{v}_k$ are orthogonal if $\vec{v}_i\cdot\vec{v}_j = \begin{cases} 1, & i=j \\ 0, & i\neq j \end{cases}$.

Theorem 5.4.4. A matrix $\boldsymbol{A}$ is orthogonal *iff* its columns are an orthogonal set of vectors.

Proof. Suppose $\boldsymbol{A} = \begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_n \\ | & & | \end{bmatrix}$, in which $\vec{u}_1,\cdots,\vec{u}_n$ are orthogonal.

Let $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$.

$$\begin{aligned}
\therefore \boldsymbol{A}(\vec{x}) \cdot \boldsymbol{A}(\vec{y}) &= \boldsymbol{A}(x_1\vec{e}_1 + \cdots + x_n\vec{e}_n) \cdot \boldsymbol{A}(y_1\vec{e}_1 + \cdots + y_n\vec{e}_n) \\
&= (x_1\vec{u}_1 + \cdots + x_n\vec{u}_n) \cdot (y_1\vec{u}_1 + \cdots + y_n\vec{u}_n) \\
&= \sum_{1 \le i,j \le n} (x_i\vec{u}_i) \cdot (y_j\vec{u}_j) \\
&= \sum_{1 \le i,j \le n} x_i y_j (\vec{u}_i \cdot \vec{u}_j) \\
&= \sum_{1 \le i,j \le n} x_i y_j \qquad \left[\vec{u}_i \cdot \vec{u}_j = \begin{cases} 1, & i = j \\ 0, & i \ne j \end{cases} \right] \\
&= \vec{x} \cdot \vec{y}.
\end{aligned}$$

■

Example 5.4.2. Consider $\boldsymbol{A} = \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \end{bmatrix}$. Is $\boldsymbol{A}$ orthogonal?

Solution.

$$\begin{aligned}
\vec{v}_1 \cdot \vec{v}_1 &= \left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2 + \left(\frac{2}{3}\right)^2 = 1 \\
\vec{v}_1 \cdot_v ecv_2 &= \left(\frac{2}{3}\right)\left(-\frac{2}{3}\right) + \left(\frac{1}{3}\right)\left(\frac{2}{3}\right) + \left(\frac{1}{3}\right)\left(\frac{2}{3}\right) = 0 \\
\vec{v}_1 \cdot_v ecv_3 &= \left(\frac{2}{3}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{3}\right)\left(\frac{2}{3}\right) + \left(\frac{2}{3}\right)\left(-\frac{2}{3}\right) = 0 \\
\vec{v}_2 \cdot \vec{v}_2 &= \left(-\frac{2}{3}\right)^2 + \left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2 = 1 \\
\vec{v}_2 \cdot_v ecv_3 &= \left(-\frac{2}{3}\right)\left(\frac{1}{3}\right) + \left(\frac{2}{3}\right)\left(\frac{2}{3}\right) + \left(\frac{1}{3}\right)\left(-\frac{2}{3}\right) = 0 \\
\vec{v}_3 \cdot \vec{v}_3 &= \left(\frac{1}{3}\right)^2 + \left(\frac{2}{3}\right)^2 + \left(-\frac{2}{3}\right)^2 = 1
\end{aligned}$$

$\therefore \boldsymbol{A}$ is orthogonal.

□

Theorem 5.4.5. To compute lots of dot products, we can encode them as a matrix product:

$$\begin{bmatrix} - & \vec{u}_1 & - \\ & \vdots & \\ - & \vec{u}_n & - \end{bmatrix} \begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} \vec{u}_1 \cdot \vec{u}_1 & \cdots & \vec{u}_n \cdot \vec{u}_1 \\ \vdots & \ddots & \vdots \\ \vec{u}_1 \cdot \vec{u}_n & \cdots & \vec{u}_n \cdot \vec{u}_n \end{bmatrix}$$

Extension. An $n \times n$ matrix $\boldsymbol{A}$ is orthogonal *iff* $\boldsymbol{A^T A} = \boldsymbol{I}$. Consequently, all orthogonal matrices are invertible, and $\boldsymbol{A}^{-1} = \boldsymbol{A^T}$.

Theorem 5.4.6.

$$(\boldsymbol{AB})^{\boldsymbol{T}} = \boldsymbol{B^T} \cdot \boldsymbol{A^T}.$$

Example 5.4.3. Consider $\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

$$(\boldsymbol{AB})^{\boldsymbol{T}} = \left(\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right)^{\boldsymbol{T}} = \begin{bmatrix} 6 \end{bmatrix}^{\boldsymbol{T}} = \begin{bmatrix} 6 \end{bmatrix}, \quad \boldsymbol{B^T A^T} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \end{bmatrix}.$$

$$\therefore (\boldsymbol{AB})^{\boldsymbol{T}} = \boldsymbol{B^T A^T}.$$

Proof. Suppose $\boldsymbol{A} = \begin{bmatrix} - & \vec{a}_1 & - \\ & \vdots & \\ - & \vec{a}_n & - \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} | & & | \\ \vec{b}_1 & \cdots & \vec{b}_m \\ | & & | \end{bmatrix}$

$$\boldsymbol{AB} = \begin{bmatrix} - & \vec{a}_1 & - \\ & \vdots & \\ - & \vec{a}_n & - \end{bmatrix} \begin{bmatrix} | & & | \\ \vec{b}_1 & \cdots & \vec{b}_m \\ | & & | \end{bmatrix} = \begin{bmatrix} \vec{b}_1 \cdot \vec{a}_1 & \cdots & \vec{b}_m \cdot \vec{a}_1 \\ \vdots & \ddots & \vdots \\ \vec{b}_1 \cdot \vec{a}_n & \cdots & \vec{b}_m \cdot \vec{a}_n \end{bmatrix}$$

$$\boldsymbol{B^T A^T} = \begin{bmatrix} - & \vec{b}_1 & - \\ & \vdots & \\ - & \vec{b}_m & - \end{bmatrix} \begin{bmatrix} | & & | \\ \vec{a}_1 & \cdots & \vec{a}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} \vec{a}_1 \cdot \vec{b}_1 & \cdots & \vec{a}_n \cdot \vec{b}_1 \\ \vdots & \ddots & \vdots \\ \vec{a}_1 \cdot \vec{b}_m & \cdots & \vec{a}_n \cdot \vec{b}_m \end{bmatrix}$$

$$\therefore (\boldsymbol{AB})^{\boldsymbol{T}} = \boldsymbol{B^T} \cdot \boldsymbol{A^T}.$$

■

Theorem 5.4.7. Properties of orthogonal matrices:

1. The inverse $\boldsymbol{A}^{-1} = \boldsymbol{A^T}$ of an orthogonal matrix $\boldsymbol{A}$ is orthogonal.
2. The product $\boldsymbol{AB}$ of orthogonal matrices is orthogonal.

Consequences:

- $\boldsymbol{A}$ is orthogonal $\Longleftrightarrow$ columns of $\boldsymbol{A}$ are an orthogonal basis.
- $\boldsymbol{A^T}$ is orthogonal $\Longleftrightarrow$ rows of $\boldsymbol{A}$ are an orthogonal basis.

Proof. We know if $\boldsymbol{A}$ is orthogonal, then $\boldsymbol{A^T A} = \boldsymbol{I}$.

1. To show $\boldsymbol{A^T}$ is orthogonal, we need to show $(\boldsymbol{A^T})^{\boldsymbol{T}}\boldsymbol{A^T} = \boldsymbol{I}$.

$$(\boldsymbol{A^T})^{\boldsymbol{T}} = \boldsymbol{A} \Rightarrow (\boldsymbol{A^T})^{\boldsymbol{T}}\boldsymbol{A^T} = \boldsymbol{A}\boldsymbol{A^T} = \boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}.$$

2. To show $\boldsymbol{AB}$ is orthogonal, we need to show $(\boldsymbol{AB})^{\boldsymbol{T}}(\boldsymbol{AB}) = \boldsymbol{I}$

$$(\boldsymbol{AB})^{\boldsymbol{T}} = \boldsymbol{B^T}\boldsymbol{A^T} \Rightarrow (\boldsymbol{AB})^{\boldsymbol{T}}(\boldsymbol{AB}) = \boldsymbol{B^T}\boldsymbol{A^T}(\boldsymbol{AB}) = \boldsymbol{B^T}(\boldsymbol{A^T}\boldsymbol{A})\boldsymbol{B} = \boldsymbol{B^T}\boldsymbol{I}\boldsymbol{B} = \boldsymbol{B^T}\boldsymbol{B} = \boldsymbol{I}.$$

■

5.5 Gram-Schmidt Process and QR Factorization

6 Determinant

6.1 The Definition of the Determinant

6.2 Computing the Determinant

6.3 The Multiplicativity of the Determinant and Other Properties

7 Eigenvalues and Eigenvectors

7.1 Computing $\boldsymbol{A}^k\vec{x}$

7.2 Diagonalization

7.3 Procedure of Finding an Eigenbasis

7.4 Multiplicity

8 Singular Value Decomposition

8.1 The Spectral Theorem

8.2 Quadratic Forms and the Principal Axis Theorem