Johns Hopkins University
AS.110.201 Linear Algebra
Learning Notes

Jiuru Lyu

May 10, 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 $\mathbb{R}^3$)

$$\begin{cases} x + y + z = 1 \\ \quad\;\; x - z = 0 \end{cases},$$

   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^{st} row with a lower one, so a leftmost nonzero entry is in the 1st row (if necessary).
2. Scale the 1^{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^{nd} row with a lower one so that the leftmost nonzero entry is in the 2nd row.
5. Scale the 2^{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^{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. $\left[\begin{array}{cc:c} 1 & 0 & \color{red}{0} \\ 0 & 1 & \color{red}{0} \\ 0 & 0 & \color{red}{1} \end{array}\right]$ 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 $\mathbb{R}^n$.

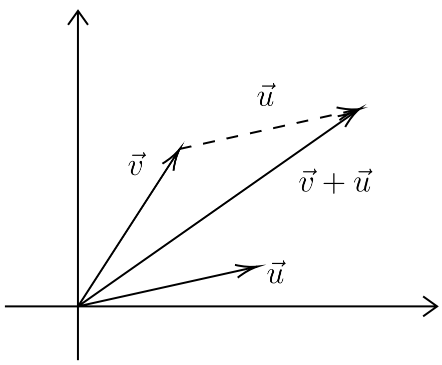
Theorem 2.1.1 (Vector Addition).

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

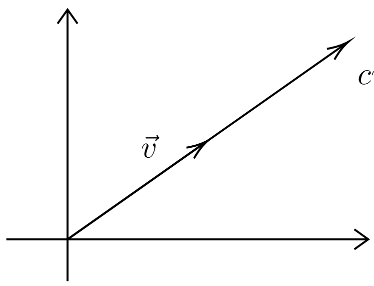
Theorem 2.1.2 (Scalar multiplication).

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

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 $\mathbb{R}^2$

$$c_1v_1 + c_2v_2 + ... + c_kv_k$$

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 $\mathbf{T} : \mathbb{R}^m \to \mathbb{R}^n$ so that:

1. $\mathbf{T}(\vec{x} + \vec{y}) = \mathbf{T}(\vec{x}) + \mathbf{T}(\vec{y})$
2. $\mathbf{T}(c \times \vec{x}) = c \times \mathbf{T}(\vec{x})$

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

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

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

are called the **standard basis vectors**.

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

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

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

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

Thus,

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

■

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

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

1. $\mathbf{T} : \mathbb{R}^m \to \mathbb{R}^n$ implies that the original parallelograms map to the transformed parallelograms
2. $\mathbf{T}(c \times \vec{x}) = c \times \mathbf{T}(\vec{x})$ 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 $\mathbf{T}\vec{x}$
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 $\mathrm{Rot}_\theta : \mathbb{R}^2 \to \mathbb{R}^2$, 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 $\vec{\mathbf{e}}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\vec{\mathbf{e}}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
The matrix of Rot_θ is

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

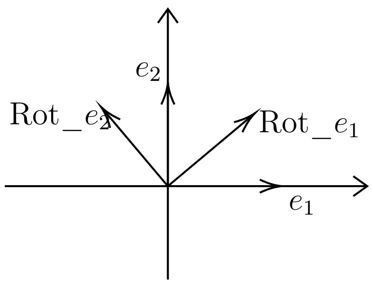
1. If $\theta = \frac{\pi}{2}$, i.e. we rotate by 90Âř counterclockwise. The matrix for rotation is

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

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

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

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



□

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 $\vec{\mathbf{e}}_i = \vec{\mathbf{e}}_i^{\parallel} + \vec{\mathbf{e}}_i^{\perp}$, and $\mathrm{Ref}(\vec{\mathbf{e}}_i) = \mathrm{Ref}(\vec{\mathbf{e}}_i^{\parallel}) + \mathrm{Ref}(\vec{\mathbf{e}}_i^{\perp})$.
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 90Âř:

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

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

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

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

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

Thus, the matrix is

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

$\square$

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 $\mathbf{A}$ and $\mathbf{B}$.
  *The number of rows of $\mathbf{A}$ equals the number of columns of $\mathbf{B}$.
- Output: The product $\mathbf{BA}$
- Procedure:
  1. View $\mathbf{A}$ as a list of its column vectors:
  $$\mathbf{A} = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}$$
  2. Multiply each column by $\mathbf{B}$:
  $$\mathbf{BA} = \begin{bmatrix} | & & | \\ \mathbf{B}v_1 & \cdots & \mathbf{B}v_n \\ | & & | \end{bmatrix}$$

Example 3.1.1. Examples of matrix multiplication.

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

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

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

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

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

Solution.

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

$\square$

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

1. A matrix encodes a linear transformation:

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

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

We can compass these maps:

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

The product **BA** encodes the composition of those transformations.

Example 3.1.2. Rotation by 90° counterclockwise:

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

Thus,

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

2. The composition **BA** is linear:

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

3. The matrix for the composition is:

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

Proof. Suppose

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

Then,

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

■

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:

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

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

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

Algebraically, we have

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

Since these are equal:

$$\boxed{\begin{aligned} \cos 2\theta &= \cos^2\theta - \sin^2\theta \\ \sin 2\theta &= 2\sin\theta\cos\theta \end{aligned}}$$

□

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

Theorem 3.1.2. Algebraic properties of matrix multiplication:

1. Matrix multiplication is associated:

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

2. Matrix multiplication is generally NOT communitive:

   (a) If $\mathbf{A}$ and $\mathbf{B}$ are matrices n rows and n columns, $\mathbf{AB} \neq \mathbf{BA}$ 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 $\vec{\mathbf{b}} = \begin{bmatrix}b_1\\b_2\end{bmatrix} \in \mathbb{R}^2$ be a fixed, arbitrary vector. Let $\mathbf{A} = \begin{bmatrix}2 & 1\\1 & 1\end{bmatrix}$. Find all solutions $\vec{\mathbf{x}} \in \mathbb{R}^2$ to the matrix equation $\mathbf{A}\vec{\mathbf{x}} = \vec{\mathbf{b}}$ (as a function of b_1 and b_2.)

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

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

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

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

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

□

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

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.

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

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

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

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

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

we call it the **identity matrix**.

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

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

1. Form augmented matrices:

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

2. Row reduce:

$$\left[\begin{array}{c:c} \operatorname{rref}(\mathbf{A}) & \mathbf{B} \end{array}\right],$$

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

Example 3.2.2.

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

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

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

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

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

Example 3.2.3. Let

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

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

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

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

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

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

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

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

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

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

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

Solution.

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

$$\xrightarrow{III/2} \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 2 & -1 & 0 \\ 0 & 1 & 2 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 & -^3/_2 & ^1/_2 \end{array}\right] \xrightarrow[I+III]{II-2III} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 3 & -^5/_2 & ^1/_2 \\ 0 & 1 & 0 & -3 & 4 & -1 \\ 0 & 0 & 1 & 1 & -^3/_2 & ^1/_2 \end{array}\right]$$

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

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

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

□

3.3 Kernel of a Matrix

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

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

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

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

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

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

Interpretation: The solution set to $\mathbf{A}\vec{\mathbf{x}} = \vec{\mathbf{b}}$ is the translation of the solution set to $\mathbf{A}\vec{\mathbf{x}} = \vec{0}$ by $\vec{\mathbf{x}}_0$.

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

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

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

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

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

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

■

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

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

Theorem 3.3.3.

$$\ker(\mathbf{A}) = \ker(\mathrm{rref}(\mathbf{A})).$$

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

1. Row reduce $\mathbf{A}$ to $\mathrm{rref}(\mathbf{A})$, compute $\ker(\mathrm{rref}(\mathbf{A}))$.

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

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

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

Solution. $\ker(\mathbf{A})$ is the solution set to $\mathbf{A}\vec{\mathbf{x}} = \vec{0}$:

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

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

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

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

Thus,

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

□

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

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

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

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

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

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

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

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

Remark.

$$\ker(\mathbf{A}) \subseteq \mathbb{R}^m \quad \text{(domain)}$$
$$\mathrm{Span}(\mathbf{A}) \subseteq \mathbb{R}^n \quad \text{(range)}$$

Proof. We know that the columns of a matrix form $\mathbf{A}\vec{\mathbf{x}}$, namely the i-th column of the matrix $\mathbf{A}$ is $\mathbf{A}\vec{e}_i$, where $\vec{\mathbf{e}}_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \rightarrow$ the i-th entry.

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

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

■

4 Spaces and Dimensions

4.1 Subspaces and Bases

Theorem 4.1.1. Spans of Sets of Vectors:

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

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

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

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

Example 4.1.1. $\vec{\mathbf{e}}_1, \cdots, \vec{\mathbf{e}}_k$ are linearly independent (L.I.) in $\mathbb{R}^n \quad \forall n \geq k$. ($\vec{\mathbf{e}}_1, \cdots, \vec{\mathbf{e}}_k$ 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 $\vec{\mathbf{v}} \in V$, then $c\vec{\mathbf{v}} \in V \quad \forall c \in \mathbb{R}$.
   Interpretation: If $\vec{\mathbf{v}} \in V$, then the line through the origin containing $\vec{\mathbf{v}}$ is in V.
3. If $\vec{\mathbf{v}}_1,\ \vec{\mathbf{v}}_2 \in V$, then $\vec{\mathbf{v}}_1 + \vec{\mathbf{v}}_2 \in V$.
   Interpretation: If $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$ are not co-linear and contained in V, then the plane through $\vec{\mathbf{v}}_1,\ \vec{\mathbf{v}}2$ 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 $\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_k \in \mathbb{R}^n$, then $\mathrm{Span}(\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_k)$ is a subspace.

   Proof.

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

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

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

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

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

   ∎

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 $\mathbf{A}$, which maps a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$ (i.e., $\mathbf{A} : \mathbb{R}^m \to \mathbb{R}^n$). Let $\ker(\mathbf{A})$ and $\mathrm{Im}(\mathbf{A})$ be the kernel and image of $\mathbf{A}$, respectively.

6. $\ker(\mathbf{A})$ is a subspace.

   Proof.

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

   $$\implies \vec{0}_n \text{ is in } \ker(\mathbf{A}).$$

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

   $$\implies \text{If } \vec{\mathbf{v}} \in \ker(\mathbf{A}), \text{ then } c\vec{\mathbf{v}} \in \ker(\mathbf{A}).$$

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

$$\implies \text{If } \vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2 \in \ker(\mathbf{A}), \text{ then } \vec{\mathbf{v}}_1 + \vec{\mathbf{v}}_2 \in \ker(\mathbf{A}).$$

■

7. $\text{Im}(\mathbf{A})$ is a subspace.

Proof.

(a) $\vec{0}_n \in \text{Im}(\mathbf{A})$

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

$$\implies c\vec{\mathbf{b}} = \mathbf{A}(c\vec{\mathbf{x}}) \in \text{Im}(\mathbf{A})$$

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

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

■

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 $\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_k \in V$, which:

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

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

Proof.

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

■

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

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

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

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

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

Remark. The coefficients -1 and 2 on the third column of $\mathbf{A}$ 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: $\mathrm{Im}(\mathbf{A}) = \mathrm{Span}(\vec{\mathbf{v}}_1, \cdots \vec{\mathbf{v}}_m)$. To produce basis, remove redundant columns. Hence, we want to show: the i-th column does not contain a pivot on row reduction *(*iff) $\vec{\mathbf{v}}_i$ is redundant:

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

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

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

Example 4.1.5.

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

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

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

■

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

1. The spanning set produced by "computing the kernel of $\mathbf{A}$" is a basis for $\ker(\mathbf{A})$.
2. Procedure:
   (a) Row reduce $\mathbf{A}$ to $\text{rref}(\mathbf{A})$, and then compute $\ker(\text{rref}(\mathbf{A}))$.
   (b) Unpack the equations encoded by matrix equation $\text{rref}(\mathbf{A}) = 0$. Solve for pivot variables in terms of free variables.
   (c) Parametrize the solution set for $\text{rref}(\mathbf{A})\vec{\mathbf{x}} = 0$ as $\{t_1\vec{\mathbf{v}}_1 + t_2\vec{\mathbf{v}}_2 + \cdots + t_d\vec{\mathbf{v}}_d;\ t_i \in \mathbb{R}\}$ and $\vec{\mathbf{v}}_i$ 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 $\vec{\mathbf{v}}_{i_j}$ is 0 if $j \neq k$; $\vec{\mathbf{v}}_{i_j}$ is 1 if $j = k$. Thus,

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

■

Example 4.1.6. Let $\vec{\mathbf{v}}_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}$ and $\vec{\mathbf{v}}_4 = \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \end{bmatrix}$. Then, $c\vec{\mathbf{v}}_3 \neq \vec{\mathbf{v}}4$ since the 4-th position of $\vec{\mathbf{v}}_3$ is 0, whereas that of $\vec{\mathbf{v}}_4$ 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 A). Let $\mathbf{A}$ be an $n \times m$ matrix (i.e., $\mathbf{A} : \mathbb{R}^m \to \mathbb{R}^n$). The **rank** of $\mathbf{A}$ is the dimension of the image of A. We denote it as $\text{rank}(\mathbf{A})$.

$$\text{rank}(\mathbf{A}) = \dim(\text{Im}(\mathbf{A}))$$

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

$$\text{nullity}(\mathbf{A}) = \dim(\ker(\mathbf{A}))$$

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

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

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

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

1. To find a basis for $\operatorname{Im}(\mathbf{A})$, we take the columns of $\mathbf{A}$ which contain a pivot upon row reduction:
$$\operatorname{Im}(\mathbf{A}) = \operatorname{Span}\left(\begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} \right).$$
$$\therefore \dim(\operatorname{Im}(\mathbf{A})) = 2.$$

2. To find a basis for $\ker(\mathbf{A})$, unpack the equation:
$$\begin{cases} x_1 - x_3 = 0 \\ x_2 + 2x_3 = 0 \end{cases}, \implies \begin{cases} x_1 = x_3 \\ x_2 = -2x_3 \end{cases}.$$
$$\therefore \ker(\mathbf{A}) = \left\{ \begin{bmatrix} x_3 \\ -2x_3 \\ x_3 \end{bmatrix} ; \ x_3 \in \mathbb{R} \right\} = \operatorname{Span}\left(\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right).$$
$$\therefore \dim(\ker(\mathbf{A})) = 1.$$

3. $\operatorname{rank}(\mathbf{A}) = \dim(\operatorname{Im}(\mathbf{A})) = 2$; $\operatorname{nullity}(\mathbf{A}) = \dim(\ker(\mathbf{A})) = 1$; $\dim(\text{domain}) = 3$
$$\therefore \operatorname{rank}(\mathbf{A}) + \operatorname{nullity}(\mathbf{A}) = 3 = \dim(\text{domain}).$$

Proof.

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

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

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

$\blacksquare$

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

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

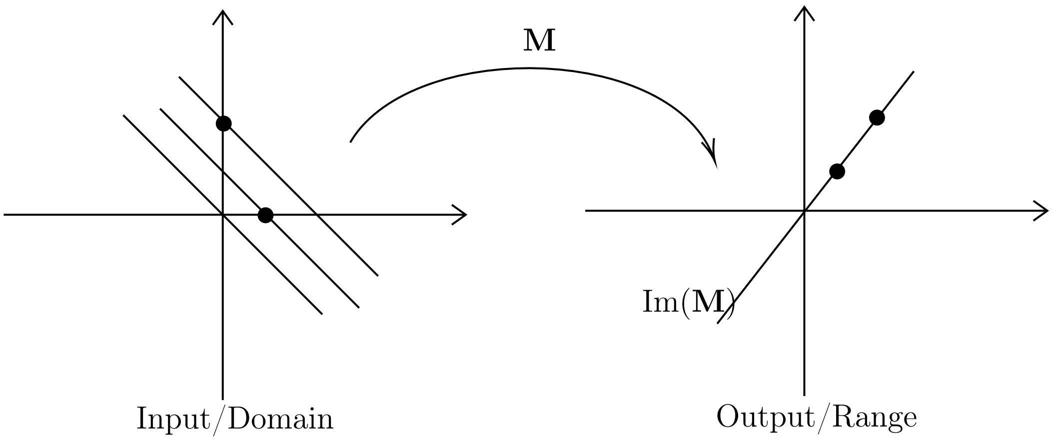
$$\therefore \operatorname{Im}(\mathbf{M}) = \operatorname{Span}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right) \text{(Line of slope 2 through the origin)} \implies \dim(\operatorname{Im}(\mathbf{M})) = 1;$$

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

If we consider the domain of $\mathbf{M}$ to be the inputs for the transformation, and range of $\mathbf{M}$ ($\operatorname{Im}(\mathbf{M})$) 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 $\dim(\ker(\mathbf{M}))$. In this specific example, $\dim(\text{inputs}) = 2$ and $\dim(\text{outputs}) = 1$, so the information loss of the linear transformation $\mathbf{M}$ is $2 - 1 = 1$.



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

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

$$\iff \operatorname{Im}(\mathbf{A}) = \mathbb{R}^n \quad \text{and} \quad \ker(\mathbf{A}) = \{\vec{0}\}.$$

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

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

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

   (b) $\operatorname{rank}(\mathbf{A}) = n$

   (c) $\operatorname{nullity}(\mathbf{A}) = 0$

   (d) $\operatorname{Im}(\mathbf{A}) = \mathbb{R}^n$

(e) $\ker(\mathbf{A}) = \{\vec{0}\}$

(f) $\text{rref}(\mathbf{A}) = \mathbf{I}_n$

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

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

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

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

$$\mathbf{BA} = \mathbf{I} = \mathbf{AB} \qquad (\mathbf{B} \coloneqq \mathbf{A}^{-1})$$

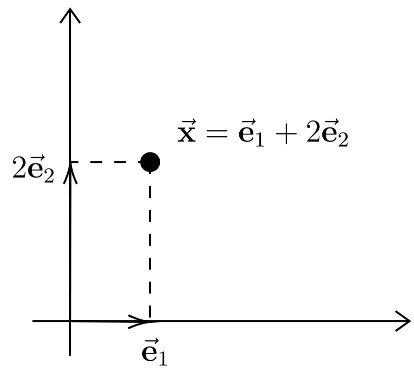
4.3 Coordinates

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 $\vec{\mathbf{x}}$ as a linear combination of the standard basis vectors.

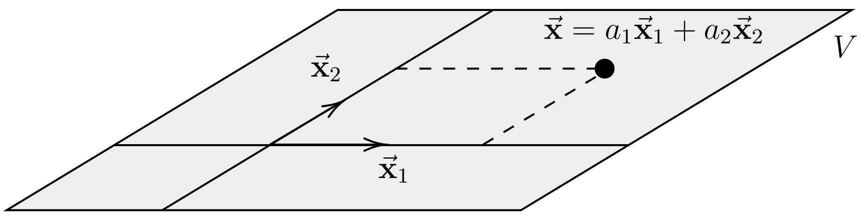
$$\text{i.e., } \vec{\mathbf{x}} = a_1 + \vec{\mathbf{e}}_1 + a_2\vec{\mathbf{e}}_2 + \cdots + a_n\vec{\mathbf{e}}_n;\ a_i \in \mathbb{R}.$$

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



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

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



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

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

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

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

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 $\vec{\mathbf{v}}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$, $\vec{\mathbf{v}}_2 = \begin{bmatrix} -3 \\ 2 \\ 3 \end{bmatrix}$. Let $\beta = (\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2)$ and $\vec{\mathbf{x}} = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}$. Find $[\vec{\mathbf{x}}]_\beta$.

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

Form augmented matrix $\left[\begin{array}{c:c} \mathbf{S} & \vec{\mathbf{x}} \end{array}\right]$:

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

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

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

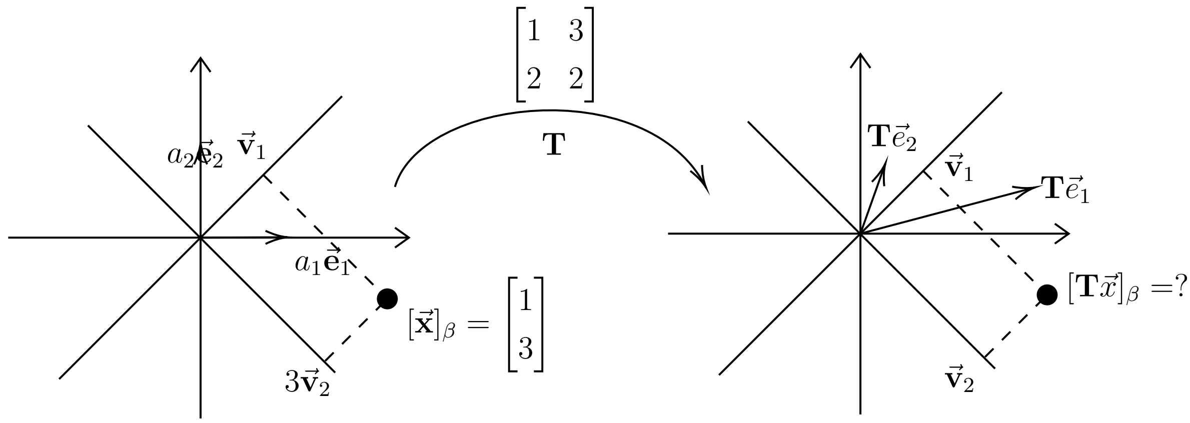
$\square$

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

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

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

Solution. First, unpack the question:



To solve this question:

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

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

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

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

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

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

□

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

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

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

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

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

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

$$[\mathbf{T}]_\beta = \begin{bmatrix} {}^1/_2 & {}^1/_2 \\ {}^1/_2 & -{}^1/_2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 4 & -1 \\ 0 & -1 \end{bmatrix}.$$

5 Approx. Solution of $\mathbf{A}\vec{\mathbf{x}} = \vec{\mathbf{b}}$

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

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

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

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

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

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

Example 5.1.2.

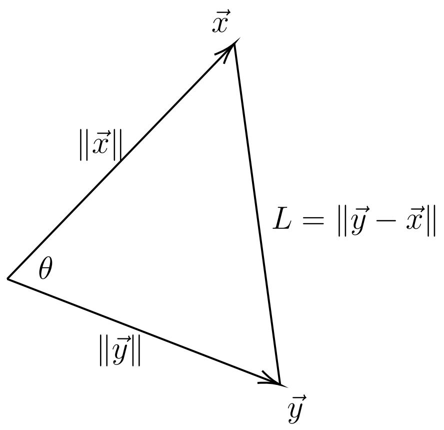
$$\left\| \begin{bmatrix} 4 \\ 3 \end{bmatrix} \right\| = \sqrt{4^2 + 3^2} = 5$$

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 $\vec{\mathbf{x}}$ and $\vec{\mathbf{y}}$. We then have

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

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



By the cosine rule, we have:

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

So,

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

■

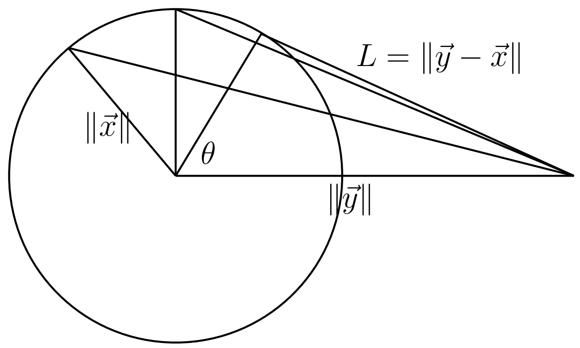
Theorem 5.1.3. Relationship of angle and dot products:

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

Proof.

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

Think of Pythagonean theorem:



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

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

■

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

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

Theorem 5.1.5.

$$\vec{\mathbf{v}} \perp \vec{\mathbf{x}} \iff \vec{\mathbf{v}} \cdot \vec{\mathbf{x}} = 0 \iff \mathbf{A}\vec{\mathbf{x}} = 0 \implies \vec{\mathbf{x}} \in \ker(\mathbf{A}).$$

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

Theorem 5.1.6.

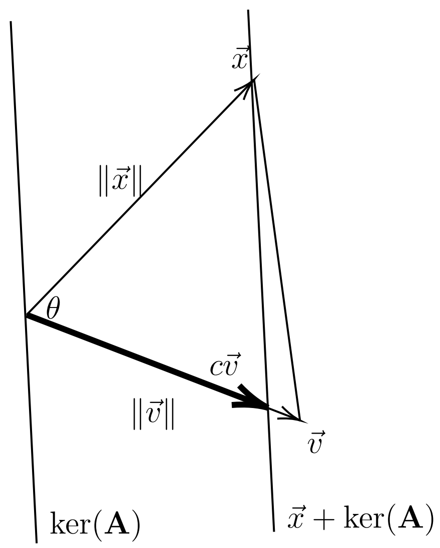
$$\vec{\mathbf{x}} \cdot \vec{\mathbf{v}} = \|\vec{\mathbf{x}}\|\|\vec{\mathbf{v}}\| \cos\theta$$

Proof.

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

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

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



3. Use trigonometry to calculate the projection:

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

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

■

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

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

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:

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

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

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

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

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

Proof.

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

■

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 $\vec{\mathbf{x}} \in \mathbb{R}^n$. Then, $\vec{\mathbf{x}}$ can be written uniquely as

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

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

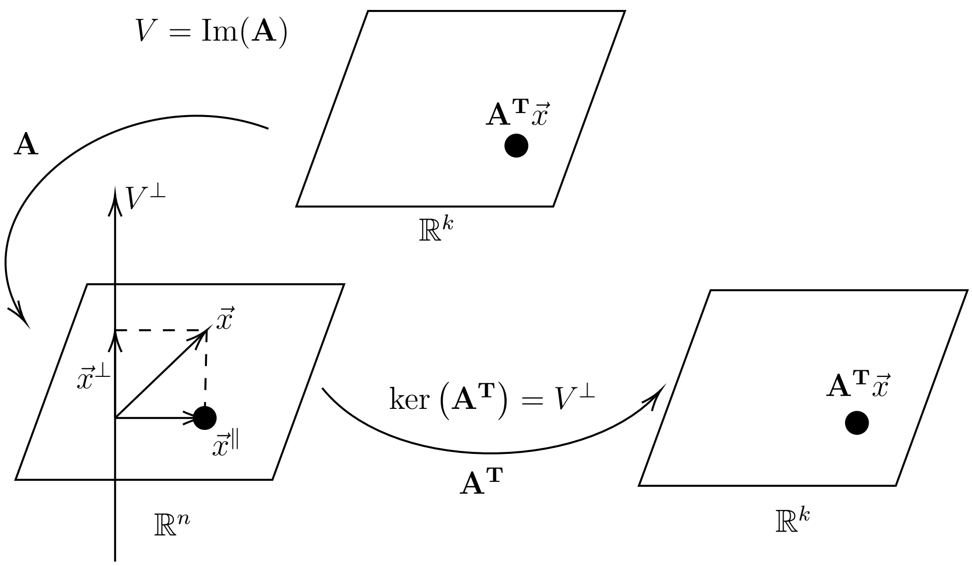
Definition 5.2.1 (Orthogonal Projection). Let $V \subseteq \mathbb{R}^n$ be a subspace. The **orthogonal projection** of $\vec{\mathbf{x}}$ into V is the vector $\vec{\mathbf{x}}^{\parallel}$. The map $\vec{\mathbf{x}} \mapsto \vec{\mathbf{x}}^{\parallel}$ is denoted as $\text{Proj}_V : \mathbb{R}^n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

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

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

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

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



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

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

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 $\vec{\mathbf{x}} = \begin{bmatrix}1\\2\\3\end{bmatrix}$. Compute the projection of $\vec{\mathbf{x}}$ onto V.

Solution.

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

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

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

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

3. Solve $\mathbf{A^T A}\vec{c} = \mathbf{A^T}\vec{\mathbf{x}}$ 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 $\mathbf{A}\vec{c} = \vec{\mathbf{x}}^{\parallel}$

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

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

$\square$

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

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

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

Theorem 5.2.3.

$$\ker(\mathbf{A^T}) = \mathrm{Im}(\mathbf{A})^{\perp}$$

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

Problem: Find $\vec{\mathbf{x}}$ *s.t.* $\|\mathbf{A}\vec{\mathbf{x}} - \vec{\mathbf{b}}\|$ is as small as possible (minimized).

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

Solution.

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

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

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

$\square$

Solution. (Advanced approach).

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

$\square$

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

Solution.

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

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

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

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:

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

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

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

□

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. □

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 $\mathbf{A}\vec{\mathbf{x}} = \vec{\mathbf{b}}$). Solve $\mathbf{A^T A}\vec{\mathbf{x}} = \mathbf{A^T}\vec{\mathbf{b}}$.

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

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

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

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

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:

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

Solve for the normal equation

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

5.4 Orthogonal Linear Transformation

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

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

Equivalently, $\mathbf{T}$ is orthogonal *iff* $\mathbf{T}$ 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

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

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

Proof.

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

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

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

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

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

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

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

■

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

Theorem 5.4.2. If $\mathbf{T} : \mathbb{R}^n \to \mathbb{R}^n$ is orthogonal, the matrix for $\mathbf{T}$ is $\begin{bmatrix} | & & | \\ \mathbf{T}(\vec{\mathbf{e}}_1) & \cdots & \mathbf{T}(\vec{\mathbf{e}}_n) \\ | & & | \end{bmatrix}$. The lengths and angles of these vectors are the same as $\vec{\mathbf{e}}_1, \cdots, \vec{\mathbf{e}}_n$ if $\mathbf{T}$ is orthogonal.

Theorem 5.4.3.

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

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

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

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

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

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

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

■

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

Solution.

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

$$\therefore \mathbf{A} \text{ is orthogonal.}$$

□

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

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

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

Theorem 5.4.6.

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

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

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

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

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

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

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

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

∎

Theorem 5.4.7. Properties of orthogonal matrices:

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

Consequences:

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

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

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

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

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

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

■

5.5 Gram-Schmidt Process, QR Factorization

Remark (Orthogonal Coordinate System). In general, a vector cannot be represented by summation of its projects, but when we have orthogonal ones, we can.

Theorem 5.5.1. Let $V = \text{Span}(\vec{\mathbf{u}}_1, \cdots, \vec{\mathbf{u}}_k)$ and $\vec{\mathbf{x}} \in V$, then there exists unique scalars $c_1, \cdots, c_k$ such that $\vec{\mathbf{x}} = c_1\vec{\mathbf{u}}_1 + c_2\vec{\mathbf{u}}_2 + \cdots + c_k\vec{\mathbf{u}}_k$. The constants $c_1, \cdots, c_k$ equal:

$$c_i = \vec{\mathbf{x}} \cdot \vec{\mathbf{u}}_i.$$

Proof. Since $\vec{\mathbf{x}} \in V$, $\vec{\mathbf{x}} = c_1\vec{\mathbf{u}}_! + \cdots + c_k\vec{\mathbf{u}}_k$ for some constants.

$$\begin{aligned} \therefore \vec{\mathbf{u}}_i \cdot \vec{\mathbf{x}} &= \vec{\mathbf{u}}_i \cdot (c_1\vec{\mathbf{u}}_1 + \cdots + c_k\vec{\mathbf{u}}_k) \\ &= c_1(\vec{\mathbf{u}}_i \cdot \vec{\mathbf{u}}_1) + \cdots + c_k(\vec{\mathbf{u}}_i \cdot \vec{\mathbf{u}}_k) \end{aligned}$$

Since $\vec{\mathbf{u}}_1, \cdots, \vec{\mathbf{u}}_k$ are orthogonal,

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

$$\therefore \vec{\mathbf{u}}_i \cdot \vec{\mathbf{x}} = c_i(\vec{\mathbf{u}}_i \cdot \vec{\mathbf{u}}_i) = c_i.$$

■

Theorem 5.5.2. Let $u_1, \cdots, \vec{\mathbf{u}}_k$ be orthogonal vectors and $V = \text{Span}(u_1, \cdots, \vec{\mathbf{u}}_k.)$ The projection of $\vec{\mathbf{x}} \in \mathbb{R}^n$ to V is given by

$$\text{Proj}_V(\vec{\mathbf{x}}) = (\vec{\mathbf{u}}_1 \cdot \vec{\mathbf{x}})\vec{\mathbf{u}}_1 + \cdots + (\vec{\mathbf{u}}_k \cdot \vec{\mathbf{x}})\vec{\mathbf{u}}_k$$

In particular, projections into a line spanned by $\vec{\mathbf{u}}$ is given by

$$\text{Proj}_L(\vec{\mathbf{x}}) = (\vec{\mathbf{u}} \cdot \vec{\mathbf{x}})\vec{\mathbf{u}}.$$

Proof. Write $\vec{\mathbf{x}} = \vec{\mathbf{x}}^{\|} + \vec{\mathbf{x}}^{\perp}$ such that $\vec{\mathbf{x}}^{\|} \in V$ and $\vec{\mathbf{x}}^{\perp} \in V^{\perp}$, and $\vec{\mathbf{x}}^{\|} = c_1\vec{\mathbf{u}}_1 + \cdots + c_k\vec{\mathbf{u}}_k$.
Note that $\vec{\mathbf{u}}_i \cdot \vec{\mathbf{x}} = \vec{\mathbf{u}}_i \cdot \vec{\mathbf{x}}^{\|} + \vec{\mathbf{u}}_i \cdot \vec{\mathbf{x}}^{\perp}$.
Since $\vec{\mathbf{u}}_i \in V$, $\vec{\mathbf{x}}^{\perp} \in V^{\perp}$, and $\vec{\mathbf{u}}_i \cdot \vec{\mathbf{x}}^{\perp} = 0$,

$$\begin{aligned}
\vec{\mathbf{u}}_i \cdot \vec{\mathbf{x}} &= \vec{\mathbf{u}}_i \cdot \vec{\mathbf{x}}^{\|} \\
&= \vec{\mathbf{u}}_i(c_1\vec{\mathbf{u}}_1 + \cdots + c_k\vec{\mathbf{u}}_k) \\
&= c_1(\vec{\mathbf{u}}_i \cdot \vec{\mathbf{u}}_1) + \cdots + c_k(\vec{\mathbf{u}}_i \cdot \vec{\mathbf{u}}_k) \\
&= c_i(\vec{\mathbf{u}}_i \cdot \vec{\mathbf{u}}_i) \\
&= c_i.
\end{aligned}$$

■

The Gram-Schmidt Process:

1. Input: $V \subseteq \mathbb{R}^n$ is a subspace with basis $\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_k$.
2. Output: $\vec{\mathbf{u}}_1, \cdots, \vec{\mathbf{u}}_k$ are orthogonal and span V.
3. Procedure:
   (a) $\vec{\mathbf{u}}_1 = \dfrac{\vec{\mathbf{v}}_1}{\|\vec{\mathbf{v}}_1\|}$
   (b) $\vec{\mathbf{v}}_k^{\perp} = \vec{\mathbf{v}}_k - \vec{\mathbf{v}}_k^{\|}$ relative to $V_{k-1} = \text{Span}(\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_{k-1})$
   (c) $\vec{\mathbf{u}}_k^{\perp} = \dfrac{\vec{\mathbf{v}}_k^{\perp}}{\|\vec{\mathbf{v}}_k^{\perp}\|}$
   (d) Compute the last $\vec{\mathbf{v}}_i^{\perp}$:
   $$\begin{aligned}
   \vec{\mathbf{v}}_i^{\perp} &= \vec{\mathbf{v}}_i - \vec{\mathbf{v}}_i^{\|} \\
   &= \vec{\mathbf{v}}_i - (\vec{\mathbf{u}}_1 \cdot \vec{\mathbf{v}}_i)\vec{\mathbf{u}}_1 - \cdots - (\vec{\mathbf{u}}_{i-1} \cdot \vec{\mathbf{v}}_i)\vec{\mathbf{u}}_{i-1} \\
   &= \vec{\mathbf{v}}_i - \sum_{j=1}^{i-1} (\vec{\mathbf{u}}_j \cdot \vec{\mathbf{v}}_i)\vec{\mathbf{u}}_j
   \end{aligned}$$

Example 5.5.1. Consider $V = \text{Span}\left(\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\0\\0\\1\end{bmatrix}, \begin{bmatrix}0\\2\\1\\-1\end{bmatrix}\right)$. Apply the Gram-Schmidt process to these vectors to find a set of vectors that are orthogonal and span V.

Solution.

1. $\vec{\mathbf{u}}_1 = \dfrac{\vec{\mathbf{v}}_1}{\|\vec{\mathbf{v}}_1\|}$

   Since $\|\vec{\mathbf{v}}_1\| = \sqrt{1^2+1^2+1^2+1^2} = 2,\ \vec{\mathbf{u}}_1 = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$

2. Find $\vec{\mathbf{v}}_2^{\perp}$ and $\vec{\mathbf{u}}_2$

$$\begin{aligned}
\vec{\mathbf{v}}_2^{\perp} &= \vec{\mathbf{v}}_2 - \vec{\mathbf{v}}_2^{\parallel} \\
&= \vec{\mathbf{v}}_2 - (\vec{\mathbf{u}}_1 \cdot \vec{\mathbf{v}}_2)\vec{\mathbf{u}}_1 \\
\vec{\mathbf{u}}_1 \cdot \vec{\mathbf{v}}_2 \frac{1}{2} &+ \frac{1}{2} = 1 \\
\therefore \vec{\mathbf{v}}_2^{\perp} &= \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \end{bmatrix} \\
\vec{\mathbf{u}}_2 &= \frac{\vec{\mathbf{v}}_2^{\perp}}{\|\vec{\mathbf{v}}_2^{\perp}\|} = \begin{bmatrix} 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \end{bmatrix}
\end{aligned}$$

3. Find $\vec{\mathbf{v}}_3^{\perp}$ and $\vec{\mathbf{u}}_3^{\perp}$

$$\begin{aligned}
\vec{\mathbf{v}}_3^{\perp} &= \vec{\mathbf{v}}_3 - \vec{\mathbf{v}}_3^{\parallel} \\
&= \vec{\mathbf{v}}_3 - (\vec{\mathbf{u}}_1 \cdot \vec{\mathbf{v}}_3)\vec{\mathbf{u}}_1 - (\vec{\mathbf{u}}_2 \cdot \vec{\mathbf{v}}_3)\vec{\mathbf{u}}_2 \\
\begin{bmatrix} \vec{\mathbf{u}}_1 \cdot \vec{\mathbf{v}}_3 \\ \vec{\mathbf{u}}_2 \cdot \vec{\mathbf{v}}_3 \end{bmatrix} &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \\
\therefore \vec{\mathbf{v}}_3^{\perp} &= \begin{bmatrix} 0 \\ 2 \\ 1 \\ -1 \end{bmatrix} - 1 \cdot \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} - (-2) \cdot \begin{bmatrix} 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ 2 \\ 1 \\ -1 \end{bmatrix} - \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{bmatrix} \\
\therefore \vec{\mathbf{u}}_3 &= \frac{\vec{\mathbf{v}}_3^{\perp}}{\|\vec{\mathbf{v}}_3^{\perp}\|} = \begin{bmatrix} 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{bmatrix}
\end{aligned}$$

Therefore, $\begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}, \begin{bmatrix} 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \end{bmatrix}, \begin{bmatrix} 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{bmatrix}$ are orthogonal and span V. □

Theorem 5.5.3 (QR-Decomposition). Let $\mathbf{A} = \begin{bmatrix} | & & | \\ \vec{\mathbf{v}}_1 & \cdots & \vec{\mathbf{v}}_k \\ | & & | \end{bmatrix}$ be a matrix and assume $\mathbf{A}$ has linearly independent columns. Then,

$$\mathbf{A} = \mathbf{QR},$$

where

$$\mathbf{Q} = \begin{bmatrix} | & & | \\ \vec{\mathbf{u}}_1 & \cdots & \vec{\mathbf{u}}_k \\ | & & | \end{bmatrix}$$

has orthogonal columns and

$$\mathbf{R} = \begin{bmatrix} \vec{\mathbf{u}}_1 \cdot \vec{\mathbf{v}}_1 & \vec{\mathbf{u}}_1 \cdot \vec{\mathbf{v}}_2 & \cdots & \vec{\mathbf{u}}_1 \cdot \vec{\mathbf{v}}_k \\ 0 & 0 & \vdots & \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \vec{\mathbf{u}}_k \cdot \vec{\mathbf{v}}_k \end{bmatrix}$$

is upper triangular. In particular, if $\mathbf{A}$ is a square, invertible matrix, $\mathbf{A} = \mathbf{QR}$, where $\mathbf{Q}$ is orthogonal and $\mathbf{R}$ is upper triangular.

Proof. Run G.S. process in $V = \mathrm{Im}(A)$ with basis $\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_k$, and we get $\vec{\mathbf{u}}_1, \cdots, \vec{\mathbf{u}}_k$ that are orthogonal and span V.

$$\vec{\mathbf{v}}_i = \sum_{j=1}^{i} (\vec{\mathbf{v}}_i \cdot \vec{\mathbf{u}}_j)\vec{\mathbf{u}}_j.$$

$$\therefore \begin{bmatrix} | & & | \\ \vec{\mathbf{u}}_1 & \cdots & \vec{\mathbf{u}}_k \\ | & & | \end{bmatrix} \begin{bmatrix} \vec{\mathbf{u}}_1 \cdot \vec{\mathbf{v}}_1 & \vec{\mathbf{u}}_1 \cdot \vec{\mathbf{v}}_2 & \cdots & \vec{\mathbf{u}}_1 \cdot \vec{\mathbf{v}}_k \\ 0 & \vec{\mathbf{u}}_2 \cdot \vec{\mathbf{v}}_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \vec{\mathbf{u}}_k \cdot \vec{\mathbf{v}}_k \end{bmatrix} = \begin{bmatrix} | & & | \\ \vec{\mathbf{v}}_1 & \cdots & \vec{\mathbf{v}}_k \\ | & & | \end{bmatrix}.$$

■

Find G.S. Process and QR factorization via row reduction.

1. General Idea:

   Input: A matrix $\mathbf{A}$ with linearly independent columns

Output: a factorization $\mathbf{A} = \mathbf{QR}$, where $\mathbf{Q}$ has orthogonal columns and $\mathbf{R}$ is upper triangular.

2. General Procedure:

   (a) Compute $\mathbf{A^T A}$ and form the augmented matrix $\left[\mathbf{A^T A} \ \vdots \ \mathbf{A^T}\right]$.

   (b) Row reduce the left hand side until an upper triangular only by subtracting multiples of rows from rows below them. At the conclusion of this step, left hand side is upper triangular.

   (c) Divide each row by the square root of the leading diagonal entry.

   (d) The final output is $\left[\mathbf{R} \ \vdots \ \mathbf{Q^T}\right]$.

Example 5.5.2. Find the QR factorization of $\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & 3 \end{bmatrix}$.

Solution.

1. Compute $\mathbf{A^T A}$.

$$\mathbf{A^T A} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 1 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 4 \\ 1 & 2 & 3 \\ 4 & 3 & 14 \end{bmatrix}.$$

2. Row reduce $\left[\mathbf{A^T A} \ \vdots \ \mathbf{A^T}\right]$.

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 4 & 1 & 0 & 1 \\ 1 & 2 & 3 & 1 & 1 & 0 \\ 4 & 3 & 14 & 1 & 2 & 3 \end{array}\right] \xrightarrow[\text{Reduction}]{\text{Row}} \left[\begin{array}{ccc|ccc} \sqrt{2} & 1/\sqrt{2} & 4/\sqrt{2} & 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & \sqrt{3/2} & \sqrt{2/3} & {}^1/_2\sqrt{2/3} & \sqrt{2/3} & -{}^1/_2\sqrt{2/3} \\ 0 & 0 & 2/\sqrt{3} & -{}^1/_2\sqrt{3}/3 & {}^1/_2\sqrt{3}/3 & {}^1/_2\sqrt{3}/3 \end{array}\right]$$

3. So,

$$\mathbf{R} = \begin{bmatrix} \sqrt{2} & 1/\sqrt{2} & 4/\sqrt{2} \\ 0 & \sqrt{3/2} & \sqrt{2/3} \\ 0 & 0 & 2/\sqrt{3} \end{bmatrix}$$

$$\mathbf{Q} = \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ {}^1/_2\sqrt{2/3} & \sqrt{2/3} & -{}^1/_2\sqrt{2/3} \\ -{}^1/_2\sqrt{3}/3 & {}^1/_2\sqrt{3}/3 & {}^1/_2\sqrt{3}/3 \end{bmatrix}$$

□

6 Determinant

6.1 The Definition of the Determinant

Remark. Dot products encode lengths and angles of vectors. Determinant encodes volume and orientations of subspaces.

Definition 6.1.1 (Determinant). Let $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a 2×2 matrix, the **determinant** of $\mathbf{A}$ is the quantity

$$\det(\mathbf{A}) = ad - bc.$$

Theorem 6.1.1. The image of the unit square under $\mathbf{A}$ is $|\det(\mathbf{A})|$.

Theorem 6.1.2. A matrix $\mathbf{A}$ is invertible $\iff \det(\mathbf{A}) \neq 0$.

Proof. $\mathbf{A}$ is invertible $\implies \operatorname{rank}(\mathbf{A}) = 2$, i.e., $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$ (columns of $\mathbf{A}$) are not co-linear. $\implies$ The area of the parallelogram spanned by $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$ does not have an area of 0. ■

Theorem 6.1.3. Let $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a 2×2 matrix. The sign of the determinant of $\mathbf{A}$ satisfies

$$\operatorname{sign}(\det(\mathbf{A})) = \begin{cases} 0 & \text{if } \vec{\mathbf{v}}_1 \text{ and } \vec{\mathbf{v}}_2 \text{ are colinear} \\ + & \text{if } \vec{\mathbf{v}}_2^{\perp} \text{ is a positive multiple of } \vec{\mathbf{v}}_1^{\text{rot}} = \begin{bmatrix} -c \\ a \end{bmatrix} \\ - & \text{if } \vec{\mathbf{v}}_2^{\perp} \text{ is a negative multiple of } \vec{\mathbf{v}}_1^{\text{rot}} = \begin{bmatrix} -c \\ a \end{bmatrix} \end{cases}.$$

Proof. Consider the projection of $\vec{\mathbf{v}}_2$ into the line spanned by $\vec{\mathbf{v}}_1^{\text{rot}} = \begin{bmatrix} -c \\ a \end{bmatrix}$.

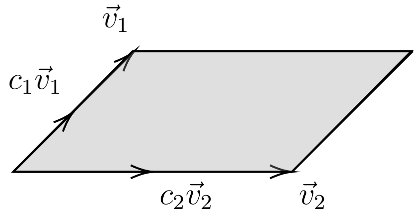
$$\begin{aligned} \vec{\mathbf{v}}_2^{\perp} = \frac{\vec{\mathbf{v}}_1^{\text{rot}} \cdot \vec{\mathbf{v}}_2}{\vec{\mathbf{v}}_1^{\text{rot}} \cdot \vec{\mathbf{v}}_1^{\text{rot}}} \cdot \vec{\mathbf{v}}_1^{\text{rot}} &= \frac{\begin{bmatrix} -c \\ a \end{bmatrix} \cdot \begin{bmatrix} b \\ d \end{bmatrix}}{\begin{bmatrix} -c \\ a \end{bmatrix} \cdot \begin{bmatrix} -c \\ a \end{bmatrix}} \cdot \begin{bmatrix} -c \\ a \end{bmatrix} \\ &= \frac{\det(\mathbf{A})}{a^2 + c^2} \cdot \vec{\mathbf{v}}_1^{\text{rot}} \end{aligned}$$

$$\begin{aligned} \therefore \det(\mathbf{A}) > 0 &\implies \vec{\mathbf{v}}_2^{\perp} > 0 \\ \det(\mathbf{A}) < 0 &\implies \vec{\mathbf{v}}_2^{\perp} < 0. \end{aligned}$$

■

Remark. The sign of the determinant describes the orientation of $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$.

Definition 6.1.2 (Parallelogram). A parallelogram is defined by the set $\{c_1\vec{\mathbf{v}}_1 + c_2\vec{\mathbf{v}}_2 \mid 0 \le c_2 \le 1\}$



Extension (K-Parallelepiped). Let $\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_k \in \mathbb{R}^n$. The **k-parallelepiped** spanned by $\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_k$ is the set

$$\{c_1\vec{\mathbf{v}}_1 + c_2\vec{\mathbf{v}}_2 + \cdots + c_k\vec{\mathbf{v}}_k \mid c_i \in [0,1]\}.$$

Extension (Unit Cube/n-parallelepiped). The **unit cube** is $\mathbb{R}^n$ is the **n-parallelepiped** spanned by $\vec{\mathbf{e}}_1, \cdots \vec{\mathbf{v}}_n$

$$\{c_1\vec{\mathbf{e}}_1 + \cdots + c_n\vec{\mathbf{e}}_n \mid c_i \in [0,1]\}.$$

Theorem 6.1.4. Let $\mathbf{A}$ be a linear transformation, then $\mathbf{A}$ maps parallelepipeds to parallelepipeds. The image of the unit cube under $\mathbf{A}$ is the parallelepipeds spanned by the columns of $\mathbf{A}$.

Theorem 6.1.5 (Volume). The **volume** of a k-parallelepiped spanned by $\vec{\mathbf{v}}_1, \cdots .\vec{\mathbf{v}}_k$ is

$$\operatorname{vol}(\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_k) = \operatorname{vol}(\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_{k-1})\|\vec{\mathbf{v}}_k^\perp\|,$$

where the $\vec{\mathbf{v}}_k^\perp$ is the perpendicular part of $\vec{\mathbf{v}}_k$ in the decomposition $\vec{\mathbf{v}}_k = \vec{\mathbf{v}}_k^{\|} + \vec{\mathbf{v}}_k^\perp$, where $\vec{\mathbf{v}}_k^{\|} \in \operatorname{Span}(\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_{k-1})$, and $\vec{\mathbf{v}}_k^\perp$ is perpendicular.

Theorem 6.1.6. The volume of the k-parallelepiped spanned by $\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_k$ equals

$$\operatorname{vol}(\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_k) = \|\vec{\mathbf{v}}_1\| \cdot \|\vec{\mathbf{v}}_2^\perp\| \cdot \|\vec{\mathbf{v}}_3^\perp\| \cdots \|\vec{\mathbf{v}}_k^\perp\|,$$

where $\vec{\mathbf{v}}_i^\perp$ is the perpendicular part of $\vec{\mathbf{v}}_i$ with respect to $V = \operatorname{Span}(\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_{i-1})$.

Example 6.1.1. Let $\vec{\mathbf{v}}_1 = \begin{bmatrix} 7 \\ 0 \\ 0 \end{bmatrix}, \vec{\mathbf{v}}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \vec{\mathbf{v}}_3 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$. Find the volume of the k-parallelepiped spanned by $\vec{\mathbf{v}}_1,\ \vec{\mathbf{v}}_2,\ \vec{\mathbf{v}}_3$.

Solution.

$$\operatorname{vol}(\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2, \vec{\mathbf{v}}_3) = \|\vec{\mathbf{v}}_1\| \cdot \|\vec{\mathbf{v}}_2^\perp\| \cdot \|\vec{\mathbf{v}}_3^\perp\|$$

1. Since $\|\vec{\mathbf{v}}_1\| = 7$,

$$\vec{\mathbf{u}}_1 = \frac{\vec{\mathbf{v}}_1}{\|\vec{\mathbf{v}}_1\|} = \frac{1}{7}\begin{bmatrix} 7 \\ 0 \\ 0 \end{bmatrix} = \vec{\mathbf{e}}_1$$

2.

$$\begin{aligned}
\vec{\mathbf{v}}_2^{\perp} &= \vec{\mathbf{v}}_2 - \vec{\mathbf{v}}_2^{\parallel} \\
&= \vec{\mathbf{v}}_2 - (\vec{\mathbf{v}}_2 \cdot \vec{\mathbf{u}}_1)\vec{\mathbf{u}}_1 \\
&= \vec{\mathbf{v}}_2 - (\vec{\mathbf{e}}_1 \cdot \vec{\mathbf{v}}_2)\vec{\mathbf{e}}_1 \\
&= \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \\
\|\vec{\mathbf{v}}_2^{\perp}\| &= \sqrt{2} \\
\vec{\mathbf{u}}_2 &= \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}
\end{aligned}$$

3.

$$\begin{aligned}
\vec{\mathbf{v}}_3^{\perp} &= \vec{\mathbf{v}}_3 - \vec{\mathbf{v}}_3^{\parallel} \\
&= \vec{\mathbf{v}}_3 - (\vec{\mathbf{v}}_3 \cdot \vec{\mathbf{e}}_1)\vec{\mathbf{e}}_1 - (\vec{\mathbf{v}}_3 \cdot \vec{\mathbf{u}}_2)\vec{\mathbf{u}}_2 \\
&= \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\
\|\vec{\mathbf{v}}_3^{\perp}\| &= 0 \longrightarrow \vec{\mathbf{v}}_3 \in \text{Span}(\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2) \\
&\therefore \text{vol}(\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2, \vec{\mathbf{v}}_3) = 0
\end{aligned}$$

□

Theorem 6.1.7. Let $\mathbf{A}$ be an $n \times n$ matrix, then $\mathbf{A}$ is invertible $\iff$ the volume of the parallelepiped spanned by the columns of $\mathbf{A}$ is not 0.

Proof.

$$\begin{aligned}
\mathbf{A} \text{ is invertible} &\iff \text{rank}(\mathbf{A}) = n \\
&\Rightarrow \vec{\mathbf{v}}_i \notin \text{Span}(\vec{\mathbf{v}}_1, \cdots \vec{\mathbf{v}}_{i-1}) \\
&\Rightarrow \|\vec{\mathbf{v}}_i^{\perp}\| \neq 0 \\
&\Rightarrow \text{vol}(\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_k) \neq 0,
\end{aligned}$$

$\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_k$ are columns of $\mathbf{A}$. ■

Definition 6.1.3 (Formal Definition of Determinant). There is a unique function from the set of $n \times n$ matrices to real numbers called the **determinant** and denoted as

$$\det: \{n \times n \text{ matrices}\} \to \mathbb{R}$$

satisfying the following conditions:

1. $|\det(\mathbf{A})|$ =volume of the parallelepiped spanned by the columns of $\mathbf{A}$.

2. (a) $\det(\mathbf{I}) = 1$.

   Example 6.1.2. $\det\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = 1 - 0 = 1.$

   (b) The determinant is a linear function in each column of $\mathbf{A}$:

$$\det\left(\begin{bmatrix} | & & | & | \\ \vec{\mathbf{v}}_1 & \cdots & \vec{\mathbf{v}}_{n-1} & \vec{\mathbf{v}}_n + k\vec{\mathbf{v}}'_n \\ | & & | & | \end{bmatrix}\right) = \det\left(\begin{bmatrix} | & & | \\ \vec{\mathbf{v}}_1 & \cdots & \vec{\mathbf{v}}_n \\ | & & | \end{bmatrix}\right) + k\det\left(\begin{bmatrix} | & & | \\ \vec{\mathbf{v}}_1 & \cdots & \vec{\mathbf{v}}'_n \\ | & & | \end{bmatrix}\right)$$

   Example 6.1.3.

$$\det\left(\begin{bmatrix} 7 & c \\ 3 & d \end{bmatrix}\right) = 7d - 3c \text{ is a linear function.}$$

6.2 Computing the Determinant

Theorem 6.2.1 (Computing the Determinant via Row Reduction). Elementary row operations change the determinant in prescribed ways.

1. Switch rows of a matrix, the determinant changes the sign.

   Proof. Wants to show: $\det\left(\begin{bmatrix} \cdots & | & \cdots & | & \cdots \\ \cdots & \vec{\mathbf{v}}_i & \cdots & \vec{\mathbf{v}}_j & \cdots \\ \cdots & | & \cdots & | & \cdots \end{bmatrix}\right) = -\det\left(\begin{bmatrix} \cdots & | & \cdots & | & \cdots \\ \cdots & \vec{\mathbf{v}}_j & \cdots & \vec{\mathbf{v}}_i & \cdots \\ \cdots & | & \cdots & | & \cdots \end{bmatrix}\right)$.

   Consider: $\det\left(\begin{bmatrix} & | & & | & \\ \cdots & \vec{\mathbf{v}}_i + \vec{\mathbf{v}}_j & \cdots & \vec{\mathbf{v}}_i + \vec{\mathbf{v}}_j & \cdots \\ & | & & | & \end{bmatrix}\right) = 0$ because it has repeated columns:

$$\begin{aligned}
&\det\left(\begin{bmatrix} & | & & | & \\ \cdots & \vec{\mathbf{v}}_i + \vec{\mathbf{v}}_j & \cdots & \vec{\mathbf{v}}_i + \vec{\mathbf{v}}_j & \cdots \\ & | & & | & \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} & | & & | & \\ \cdots & \vec{\mathbf{v}}_i & \cdots & \vec{\mathbf{v}}_i + \vec{\mathbf{v}}_j & \cdots \\ & | & & | & \end{bmatrix}\right) + \det\left(\begin{bmatrix} & | & & | & \\ \cdots & \vec{\mathbf{v}}_j & \cdots & \vec{\mathbf{v}}_i + \vec{\mathbf{v}}_j & \cdots \\ & | & & | & \end{bmatrix}\right) \\
&= \underbrace{\det\left(\begin{bmatrix} & | & & | & \\ \cdots & \vec{\mathbf{v}}_i & \cdots & \vec{\mathbf{v}}_i & \cdots \\ & | & & | & \end{bmatrix}\right)}_{0} + \det\left(\begin{bmatrix} & | & & | & \\ \cdots & \vec{\mathbf{v}}_i & \cdots & \vec{\mathbf{v}}_j & \cdots \\ & | & & | & \end{bmatrix}\right) \\
&+ \det\left(\begin{bmatrix} & | & & | & \\ \cdots & \vec{\mathbf{v}}_j & \cdots & \vec{\mathbf{v}}_i & \cdots \\ & | & & | & \end{bmatrix}\right) + \overbrace{\det\left(\begin{bmatrix} & | & & | & \\ \cdots & \vec{\mathbf{v}}_j & \cdots & \vec{\mathbf{v}}_j & \cdots \\ & | & & | & \end{bmatrix}\right)}^{0} = 0
\end{aligned}$$

$$\therefore \det\left(\begin{bmatrix} \cdots & | & \cdots & | & \cdots \\ & \vec{\mathbf{v}}_i & \cdots & \vec{\mathbf{v}}_j & \cdots \\ & | & & | & \end{bmatrix}\right) + \det\left(\begin{bmatrix} \cdots & | & \cdots & | & \cdots \\ & \vec{\mathbf{v}}_j & \cdots & \vec{\mathbf{v}}_i & \cdots \\ & | & & | & \end{bmatrix}\right) = 0$$

$$\therefore \det\left(\begin{bmatrix} & | & & | & \\ \cdots & \vec{\mathbf{v}}_i & \cdots & \vec{\mathbf{v}}_j & \cdots \\ & | & & | & \end{bmatrix}\right) = -\det\left(\begin{bmatrix} & | & & | & \\ \cdots & \vec{\mathbf{v}}_j & \cdots & \vec{\mathbf{v}}_i & \cdots \\ & | & & | & \end{bmatrix}\right).$$

■

2. Adding a multiple of j^{th} row to i^{th} row with $i \neq j$, the determinant stays constant.

   Proof.

$$\det\left(\begin{bmatrix} & | & \\ \cdots & \vec{\mathbf{v}}_i + k\vec{\mathbf{v}}_j & \cdots \\ & | & \end{bmatrix}\right) = \det\left(\begin{bmatrix} & | & \\ \cdots & \vec{\mathbf{v}}_i & \cdots \\ & | & \end{bmatrix}\right) + k \cdot \det\left(\begin{bmatrix} & | & \\ \cdots & \vec{\mathbf{v}}_j & \cdots \\ & | & \end{bmatrix}\right)$$

   Note that $\det\left(\begin{bmatrix} & | & \\ \cdots & \vec{\mathbf{v}}_j & \cdots \\ & | & \end{bmatrix}\right) = 0$ because it has $\vec{\mathbf{v}}_j$ at both the i^{th} and j^{th} column, and thus the columns are not linearly independent.

$$\therefore \det\left(\begin{bmatrix} & | & \\ \cdots & \vec{\mathbf{v}}_i + k\vec{\mathbf{v}}_j & \cdots \\ & | & \end{bmatrix}\right) = \det\left(\begin{bmatrix} & | & \\ \cdots & \vec{\mathbf{v}}_i & \cdots \\ & | & \end{bmatrix}\right)$$

■

3. Scale a row by $k \neq 0$, the determinant scales by k.

   Proof. Note that determinant is a linear function in each column of $\mathbf{A}$. ■

Example 6.2.1. Compute the determinant of $\mathbf{A} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

Solution. Row reduce $\mathbf{A}$, keeping track of how the determinant changes.
Note that $\det(\mathbf{I}) = 1$.

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \xrightarrow[\text{Reduction}]{\text{Row}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

In this process of row reduction, we know $\dfrac{D}{4} = \det(\mathbf{I}) = 1$, so $D = 4$. □

Theorem 6.2.2. Let $\mathbf{A} = \begin{bmatrix} a_{11} & * & \cdots & * \\ 0 & a_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & a_{nn} \end{bmatrix}$ be an upper triangular matrix. The determinant of $\mathbf{A}$ is $a_{11}, a_{22}, \cdots, a_{nn}$, the product of the diagonal entries.

Proof. $\boxed{\text{Case 1}}$ All $a_{ii} \neq 0$.
Row reduce $\mathbf{A}$ to compute $\det(\mathbf{A})$ by dividing each row by a_{ii} to get the identity matrix $\mathbf{I}$.
So, $\frac{1}{a_{11}\cdots a_{nn}} D = \det(\mathbf{I}) = 1$, amd we get $D = a_{11}\cdots a_{nn}$
$\boxed{\text{Case 2}}$ Some $a_{ii} = 0$. $\Rightarrow$ Show $\det(\mathbf{A}) = a_{11}\cdots a_{nn} = 0 \Rightarrow$ Show $\mathbf{A}$ is not invertible.
Look at the first $a_{ii} = 0$, we know the i^{th} column in row reduction does not contain a pivot.
$\Rightarrow$ $\mathbf{A}$ is not invertible. ■

Computing the Determinant

1. Input: $n \times n$ matrix $\mathbf{A}$
2. Output: $\det(\mathbf{A})$
3. Procedure: Row reduce $\mathbf{A}$, keeping track of the elementary row operations until an upper triangular matrix is obtained.

Let $a_{11}, \cdots, a_{nn}$ be the diagonal entires of this matrix, $k_1, \cdots, k_m$ be the constants multiplied by in row reduction, and s be the number of switches:

$$\begin{aligned} (-1)^s(k_1\cdots k_m)\det(\mathbf{A}) &= a_{11}\cdots a_{nn} \\ \det(\mathbf{A}) &= \frac{a_{11}\cdots a_{nn}}{(k_1\cdots k_m)}(-1)^s \end{aligned}$$

Theorem 6.2.3.

$$\det(\mathbf{A}) = \det\left(\mathbf{A}^{\mathbf{T}}\right)$$

Theorem 6.2.4 (Computing Determinant via the Laplace Expansion). To find the formula for an $n \times n$ matrix determinant in terms of an $(n-1) \times (n-1)$ determinants:

$$\left[\begin{array}{c|ccc} 1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & \mathbf{A}_{n-1} & \\ 0 & & & \end{array}\right] = \det(\mathbf{A}_{n-1})$$

Proof. To row reduce the $n \times n$ matrix, we row reduce the $(n-1) \times (n-1)$ matrix.
$\Rightarrow \det(\mathbf{A}_{n-1}) = \det(\mathbf{A}_n)$. ■

Example 6.2.2.

$$\begin{aligned}\det\left(\begin{bmatrix}2&1&1\\1&2&1\\1&1&2\end{bmatrix}\right) &= 2\begin{vmatrix}1&1&1\\0&2&1\\0&1&2\end{vmatrix} + \begin{vmatrix}0&1&1\\1&2&1\\0&1&2\end{vmatrix} + \begin{vmatrix}0&1&1\\0&2&1\\1&1&2\end{vmatrix}\\ &= 2\begin{vmatrix}2&1\\1&2\end{vmatrix} - \begin{vmatrix}1&1\\1&2\end{vmatrix} + (-1)^2\begin{vmatrix}1&1\\2&1\end{vmatrix}\\ &= 6-1-1=4.\end{aligned}$$

Remark. Note that the vertical bars denote determinant.

Definition 6.2.1 (ij-Cofactor). Let $\mathbf{A}$ be an $n \times n$ matrix. The $ij-$**cofactor** of $\mathbf{A}$ is the $(n-1)\times(n-1)$ obtained by deleting the i^{th} row and j^{th} column. We denote this matrix as $\mathbf{A}_{ij}$.

Theorem 6.2.5 (Laplace Formula). Consider the i^{th} column of a matrix $\mathbf{A}$ (wasp i^{th} row), then,

$$\det(\mathbf{A}) = \sum_{j=1}^{n}(-1)^{i+j}\cdot a_{ij}\cdot\det(\mathbf{A}_{ji})$$

6.3 The Multiplicativity of the Determinant and Other Properties

Theorem 6.3.1 (Multiplicativity of the Determinant). Let $\mathbf{A}$ and $\mathbf{B}$ be $n\times n$ matrices,

$$\det(\mathbf{AB}) = \det(\mathbf{A})\times\det(\mathbf{B}).$$

Corollary 6.1.

$$\det\left(\mathbf{A}^k\right) = \det(\mathbf{A})^k$$

Corollary 6.2. If $\mathbf{A}$ is invertible,

$$\det\left(\mathbf{A}^{-1}\right) = \det(\mathbf{A})^{-1}.$$

Proof. Note that $\mathbf{AA}^{-1} = \mathbf{I}$.

$$\begin{aligned}\therefore \det\left(\mathbf{A}\cdot\mathbf{A}^{-1}\right) &= \det(\mathbf{I}) = 1\\ &= \det(\mathbf{A})\times\det\left(\mathbf{A}^{-1}\right)\\ \therefore \det\left(\mathbf{A}^{-1}\right) &= \det(\mathbf{A})^{-1}\end{aligned}$$

■

Corollary 6.3. If $\mathbf{A}$ is invertible, $\det(\mathbf{A}) =\neq 0$.

Theorem 6.3.2. If $\mathbf{Q}$ is an orthogonal transformation, then, $\det(\mathbf{Q}) = \pm1$, or $|\det(\mathbf{Q})| = 1$.

Proof. $|\det(\mathbf{Q})|$ is the volume of the unit cube under $\mathbf{Q}$.
The unit cube has a shape of volume 1, which means it reserves volumes.
Also that since $\mathbf{Q}$ is orthogonal, meaning this transformation preserves lengths and angles.
Wants to show that preserving lengths and angles means preserving volumes.

$$\mathrm{vol}(\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_n) = \|\vec{\mathbf{v}}_1\|\|\vec{\mathbf{v}}_2^{\perp}\|\|\vec{\mathbf{v}}_3^{\perp}\| \cdots \|\vec{\mathbf{v}}_n^{\perp}\|,\ \vec{\mathbf{v}}_i^{\perp} \in \mathrm{Span}(\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_{i-1})$$

Since $\mathbf{Q} = \begin{bmatrix} | & & | \\ \vec{\mathbf{u}}_1 & \cdots & \vec{\mathbf{u}}_n \\ | & & | \end{bmatrix}$,

$$|\det(\mathbf{Q})| = \mathrm{vol}(\vec{\mathbf{u}}_1, \cdots, \vec{\mathbf{u}}_n) = \|\vec{\mathbf{u}}_1\|\|\vec{\mathbf{u}}_2^{\perp}\| \cdots \|\vec{\mathbf{u}}_n^{\perp}\|$$

Since $\mathbf{Q}$ is orthogonal, $\vec{\mathbf{u}}_1, \cdots, \vec{\mathbf{u}}_n$ are perpendicular to each other and have lengths of 1,

$$|\det(\mathbf{Q})| = \|\vec{\mathbf{u}}_1\|\|\vec{\mathbf{u}}_2\| \cdots \|\vec{\mathbf{u}}_n\| = 1 \times \cdots \times 1 = 1.$$

■

Lemma 6.1. When $\mathbf{A}$ is invertible, $\det(\mathbf{A}) = \det\left(\mathbf{A}^{\mathbf{T}}\right)$

Proof. Use QR decomposition, we know that

$$\mathbf{A}^{\mathbf{T}} = (\mathbf{Q}\mathbf{R})^{\mathbf{T}} = \mathbf{R}^{\mathbf{T}}\mathbf{Q}^{\mathbf{T}}$$

$$\begin{aligned} \det(\mathbf{A}) &= \det(\mathbf{Q}) \cdot \det(\mathbf{R}) \\ \det\left(\mathbf{A}^{\mathbf{T}}\right) &= \det\left(\mathbf{R}^{\mathbf{T}}\right) \cdot \det\left(\mathbf{Q}^{\mathbf{T}}\right) \\ &= \det\left(\mathbf{Q}^{\mathbf{T}}\right) \cdot \det\left(\mathbf{R}^{\mathbf{T}}\right). \end{aligned}$$

Wants to show: $\det(\mathbf{Q}) = \det\left(\mathbf{Q}^{\mathbf{T}}\right)$ and $\det(\mathbf{R}) = \det\left(\mathbf{R}^{\mathbf{T}}\right)$.

1. Since $\mathbf{Q}$ is orthogonal, $\det(\mathbf{Q}) = \pm 1$.

   Also note that since $\mathbf{Q}$ is orthogonal, $\mathbf{Q}^{\mathbf{T}} = \mathbf{Q}^{-1}$.

   $$\det\left(\mathbf{Q}^{\mathbf{T}}\right) = \det\left(\mathbf{Q}^{-1}\right) = \det(\mathbf{Q})^{-1} = \det(\mathbf{Q}).$$

2. Note that $\mathbf{R}$ is an upper triangular matrix, and thus its determinant is the product of the entries on diagonal: $\det(\mathbf{R}) = a_{11} \cdot a_{22} \cdots a_{nn}$.

   Also note that the transpose of $\mathbf{R}$, $\mathbf{R}^{\mathbf{T}}$ is a lower triangular matrix, and thus we know that $\det\left(\mathbf{R}^{\mathbf{T}}\right) = a_{11} \cdot a_{22} \cdots a_{nn}$.

   $$\therefore \det(\mathbf{R}) = \det\left(\mathbf{R}^{\mathbf{T}}\right).$$

$$\begin{aligned}\det(\mathbf{A}) &= \det(\mathbf{Q}) \cdot \det(\mathbf{R}) \\ &= \det\left(\mathbf{Q}^{\mathbf{T}}\right) \cdot \det\left(\mathbf{R}^{\mathbf{T}}\right) = \det\left(\mathbf{A}^{\mathbf{T}}\right).\end{aligned}$$

■

Lemma 6.2. If $\mathbf{A}$ is not invertible, then $\mathbf{A}^{\mathbf{T}}$ is also not invertible.

Proof. $\mathbf{A}$ is invertible exactly when $\operatorname{rref}(\mathbf{A}) = \mathbf{I}$,
That is, $\operatorname{rank}(\mathbf{A}) = n \implies \operatorname{rank}(\mathbf{A}^{\mathbf{T}}) = n$ and thus, $\mathbf{A}^{\mathbf{T}}$ is also invertible,
If $\mathbf{A}$ is not invertible, $\operatorname{rank}(\mathbf{A}) < n$.
Thus, $\operatorname{rank}(\mathbf{A}^{\mathbf{T}}) = \operatorname{rank}(\mathbf{A}) < n$, indicating $\mathbf{A}^{\mathbf{T}}$ is also not invertible. ■

Lemma 6.3. For an $n \times n$ matrix $\mathbf{A}$, $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}^{\mathbf{T}})$.

Proof.

$$\begin{aligned}\dim(\operatorname{Im}(\mathbf{A})^{\perp}) &= n - \operatorname{rank}(\mathbf{A}) \\ \therefore \operatorname{nullity}(\mathbf{A}^{\mathbf{T}}) = \dim(\ker(\mathbf{A}^{\mathbf{T}})) &= n - \operatorname{rank}(\mathbf{A}) \\ \because \operatorname{rank}(\mathbf{A}^{\mathbf{T}}) + \operatorname{nullity}(\mathbf{A}^{\mathbf{T}}) &= n \\ \therefore \operatorname{rank}(\mathbf{A}^{\mathbf{T}}) + n - \operatorname{rank}(\mathbf{A}) &= n \\ \operatorname{rank}(\mathbf{A}^{\mathbf{T}}) &= \operatorname{rank}(\mathbf{A})\end{aligned}$$

■

Proposition 6.1. When $\mathbf{A}$ is not invertible, then $\det(\mathbf{A}) = \det\left(\mathbf{A}^{\mathbf{T}}\right)$.

Theorem 6.3.3.

$$\det(\mathbf{A}) = \det\left(\mathbf{A}^{\mathbf{T}}\right)$$

Theorem 6.3.4 (Cramer's Rule). Let $\mathbf{A}$ be an invertible $n \times n$ matrix and $\vec{\mathbf{b}} \in \mathbb{R}^n$. The unique solution to the system $\mathbf{A}\vec{\mathbf{x}} = \vec{\mathbf{b}}$ is the following vector

$$\vec{\mathbf{x}} = \frac{1}{\det(\mathbf{A})} \begin{bmatrix} \det\left(\mathbf{A}_{1,\vec{\mathbf{b}}}\right) \\ \vdots \\ \det\left(\mathbf{A}_{n,\vec{\mathbf{b}}}\right) \end{bmatrix},$$

where $\mathbf{A}_{i,\vec{\mathbf{b}}}$ is the $n \times n$ matrix obtained from $\mathbf{A}$ by replacing the i^{th} column with $\vec{\mathbf{b}}$.

Proof. Since

$$\begin{bmatrix} | & & | \\ \vec{\mathbf{v}}_1 & \cdots & \vec{\mathbf{v}}_n \\ | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\vec{\mathbf{v}}_1 + \cdots + x_n\vec{\mathbf{v}}_n = \vec{\mathbf{b}},$$

so we have

$$\det\left(\mathbf{A}_{1,\vec{\mathbf{b}}}\right) = \begin{vmatrix} | & | & & | \\ \vec{\mathbf{b}} & \vec{\mathbf{v}}_2 & \cdots & \vec{\mathbf{v}}_n \\ | & | & & | \end{vmatrix} = \begin{vmatrix} | & | & & | \\ x_1\vec{\mathbf{v}}_1 + \cdots + x_n\vec{\mathbf{v}}_n & \vec{\mathbf{v}}_2 & \cdots & \vec{\mathbf{v}}_n \\ | & | & & | \end{vmatrix}.$$

By the linearity of determinant, then

$$\begin{aligned}\det\left(\mathbf{A}_{1,\vec{\mathbf{b}}}\right) &= x_1\begin{vmatrix} | & | & & | \\ \vec{\mathbf{v}}_1 & \vec{\mathbf{v}}_2 & \cdots & \vec{\mathbf{v}}_n \\ | & | & & | \end{vmatrix} + x_2\begin{vmatrix} | & | & & | \\ \vec{\mathbf{v}}_2 & \vec{\mathbf{v}}_2 & \cdots & \vec{\mathbf{v}}_n \\ | & | & & | \end{vmatrix} + \cdots + x_n\begin{vmatrix} | & | & & | \\ \vec{\mathbf{v}}_n & \vec{\mathbf{v}}_2 & \cdots & \vec{\mathbf{v}}_n \\ | & | & & | \end{vmatrix} \\ &= x_1\begin{vmatrix} | & | & & | \\ \vec{\mathbf{v}}_1 & \vec{\mathbf{v}}_2 & \cdots & \vec{\mathbf{v}}_n \\ | & | & & | \end{vmatrix} = x_1\det(\mathbf{A}).\end{aligned}$$

$$\therefore x_1 = \frac{\det\left(\mathbf{A}_{1,\vec{\mathbf{b}}}\right)}{\det(\mathbf{A})}.$$

Similarly, we can extend this proof to an arbitrary x_i,

$$\begin{aligned}\det\left(\mathbf{A}_{i,\vec{\mathbf{b}}}\right) &= x_i\det(\mathbf{A}) \\ x_i &= \frac{\det\left(\mathbf{A}_{i,\vec{\mathbf{b}}}\right)}{\det(\mathbf{A})}\end{aligned}$$

■

Example 6.3.1. Suppose $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, then $\det(\mathbf{A}) = -2$. Let $\vec{\mathbf{b}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Therefore, $\mathbf{A}_{1,\vec{\mathbf{b}}} = \begin{bmatrix} 1 & 2 \\ 1 & 4 \end{bmatrix}$, and $\det\left(\mathbf{A}_{1,\vec{\mathbf{b}}}\right) = 2$. $\mathbf{A}_{2,\vec{\mathbf{b}}} = \begin{bmatrix} 1 & 1 \\ 3 & 1 \end{bmatrix}$, so $\det\left(\mathbf{A}_{2,\vec{\mathbf{b}}}\right) = -2$.

Then,

$$\vec{\mathbf{x}} = \frac{1}{-2}\begin{bmatrix} 2 \\ -2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Remark. For an arbitrary 2×2 matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $a, b, c, d \in \mathbb{R}$. Suppose $\vec{\mathbf{b}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, then applying Cramer's Rule, we know

$$\vec{\mathbf{x}} = \frac{1}{ad - bc}\begin{bmatrix} d \\ -c \end{bmatrix}.$$

Theorem 6.3.5 (Application of Cramer's Rule). Cramer's Rule can give formulas for $\mathbf{A}^{-1}$ in general: We saw $\mathbf{A}\vec{\mathbf{x}}_1 = \vec{\mathbf{e}}_1$ and $\mathbf{A}\vec{\mathbf{x}}_2 = \vec{\mathbf{e}}_2$, then

$$\mathbf{A}^{-1} = \begin{bmatrix} | & | \\ \vec{\mathbf{x}}_1 & \vec{\mathbf{x}}_2 \\ | & | \end{bmatrix}.$$

To be more specific, for a 2×2 matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, we have

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Theorem 6.3.6. Give a matrix $\mathbf{A}$ with integer entries and $\det(\mathbf{A}) = \pm 1$, the matrix $\mathbf{A}-1$ has integer entries.

7 Eigenvalues and Eigenvectors

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

Definition 7.1.1 (Eigenvector and Eigenvalue). Let $\mathbf{A}$ be an $n \times n$ matrix, and **eigenvector** for $\mathbf{A}$ is any non-zero vector $\vec{\mathbf{x}}$, such that $\mathbf{A}\vec{\mathbf{x}} = \lambda\vec{\mathbf{x}}$, for some $\lambda \in \mathbb{R}$. The number λ is called the **eigenvalue** for $\vec{\mathbf{x}}$. For an eigenvector $\vec{\mathbf{x}}$, $\mathbf{A}^k\vec{\mathbf{x}} = \lambda^k\vec{\mathbf{x}}$.

Definition 7.1.2 (Eigenbasis). Let $\mathbf{A}$ be an $nt \times n$ matrix, an **eigenbasis** for $\mathbf{A}$ is a basis for $\mathbb{R}^n$ consisting of eigenvectors for $\mathbf{A}$.

Theorem 7.1.1. Let $\mathbf{A}$ be an $n \times n$ matrix with an eigenbasis of $\vec{\mathbf{v}}_1, \cdots \vec{\mathbf{v}}_n$. The eigenvalues for $\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_n$ are $\lambda_1, \cdots, \lambda_n$, respectively. To compute $\mathbf{A}^k\vec{\mathbf{x}}$, we could write $\vec{\mathbf{x}} = c_1\vec{\mathbf{v}}_1 + \cdots + c_n\vec{\mathbf{v}}_n$, and then use linearity, we have

$$\begin{aligned}
\mathbf{A}^k\vec{\mathbf{x}} &= \mathbf{A}(c_1\vec{\mathbf{v}}_1 + \cdots + c_n\vec{\mathbf{v}}_n) \\
&= c_1\mathbf{A}^k\vec{\mathbf{v}}_1 + \cdots + c_n\mathbf{A}^k\vec{\mathbf{v}}_n \\
&= c_1\lambda_1^k\vec{\mathbf{v}}_1 + \cdots + c_n\lambda_n^k\vec{\mathbf{v}}_n.
\end{aligned}$$

Theorem 7.1.2. Consider $f_{\mathbf{A}} : \mathbb{R} \to \mathbb{R}$ defined by

$$f_{\mathbf{A}}(t) = \det(\mathbf{A} - t\mathbf{I}).$$

The zeros of $f_{\mathbf{A}}(t)$ are exactly the eigenvalues of $\mathbf{A}$. That is, $f_{\mathbf{A}}(\lambda) = 0$.

Definition 7.1.3 (Characteristic Polynomial). The **characteristic polynomial** of $\mathbf{A}$ is the function

$$f_{\mathbf{A}}(t) = \det(\mathbf{A} - t\mathbf{I}).$$

Definition 7.1.4 (Modified Definition of Eigenvectors). Let $\mathbf{A}$ be an $n \times n$ matrix, and λ be an eigenvalue for $\mathbf{A}$, i.e., a root of the characteristic polynomial of $\mathbf{A}$ defined by $f_{\mathbf{A}}(t) = \det(\mathbf{A} - t\mathbf{I})$. An **eigenvector** with eigenvalue λ for $\mathbf{A}$ is any non-zero solution to $\mathbf{A}\vec{\mathbf{x}} = \lambda\vec{\mathbf{x}}$. i.e., non solution to

$$(\mathbf{A} - \lambda\mathbf{I})\vec{\mathbf{x}} = 0.$$

Theorem 7.1.3. The non-zero elements for $\ker(\mathbf{A} - \lambda\mathbf{I})$ are exactly the eigenvectors with eigenvalue λ.

Definition 7.1.5 ($\boldsymbol{\lambda}$-Eigenspace). Let λ be an eigenvalue for $\mathbf{A}$, the **λ-eigenspace** is

$$E_{\mathbf{A},\lambda} = \ker(\mathbf{A} - \lambda\mathbf{I}).$$

Example 7.1.1. Compute all eigenvectors for $\mathbf{A} = \begin{bmatrix} 2 & 2 \\ 1 & 3 \end{bmatrix}$.

Solution.

$$f_{\mathbf{A}} = (t-4)(t-1) \implies \lambda = 1 \quad \text{and} \quad 4.$$

$$E_{\mathbf{A},4} = \ker(\mathbf{A}-4\mathbf{I}) = \ker\left(\begin{bmatrix}-2 & 2\\ 1 & -1\end{bmatrix}\right) = \ker\left(\begin{bmatrix}1 & -1\\ 0 & 0\end{bmatrix}\right) = \operatorname{Span}\left(\begin{bmatrix}1\\1\end{bmatrix}\right)$$

$$E_{\mathbf{A},1} = \ker(\mathbf{A}-\mathbf{I}) = \ker\left(\begin{bmatrix}1 & 2\\ 1 & 2\end{bmatrix}\right) = \ker\left(\begin{bmatrix}1 & 2\\ 0 & 0\end{bmatrix}\right) = \operatorname{Span}\left(\begin{bmatrix}-2\\1\end{bmatrix}\right)$$

□

Computing $\mathbf{A}^k\vec{\mathbf{x}}$:

1. Compute $f_{\mathbf{A}}(t) = \det(\mathbf{A}-t\mathbf{I})$
2. Find the roots of $f_{\mathbf{A}}(t)$; those are eigenvalues.
3. Compute eigenspaces
4. Ask: Is there an eigenbasis?
5. Write $\vec{\mathbf{x}}$ in form of $c_1\vec{\mathbf{v}}_1 + \cdots + c_n\vec{\mathbf{v}}_n$
6. Find the formula

Example 7.1.2. Let $\mathbf{A} = \begin{bmatrix}0 & 6\\ 1 & -1\end{bmatrix}$. Find a formula for $\mathbf{A}^k\begin{bmatrix}1\\1\end{bmatrix}$ for all k.

Solution. Find eigenvectors $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$ for $\mathbf{A}$ and express $\begin{bmatrix}1\\1\end{bmatrix} = c_1\vec{\mathbf{v}}_1 + c_2\vec{\mathbf{v}}_2$. Then,

$$\mathbf{A}^k\begin{bmatrix}1\\1\end{bmatrix} = c_1\mathbf{A}^k\vec{\mathbf{v}}_1 + c_2\mathbf{A}^k\vec{\mathbf{v}}_2 = c_1\lambda_1^k\vec{\mathbf{v}}_1 + c_2\lambda_2^k\vec{\mathbf{v}}_2.$$

1. Compute $f_{\mathbf{A}}(t)$:

$$f_{\mathbf{A}}(t) = \det(\mathbf{A}-t\mathbf{I}) = \begin{vmatrix}-t & 6\\ 1 & -1-t\end{vmatrix} = t(1+t) - 6 = t^2 = t - 6.$$

2. Find roots to the polynomial:

$$f_{\mathbf{A}}(t) = (t+3)(t-2) = 0 \implies \lambda_1 = t_1 = 2 \quad \lambda_2 = t_2 = -3.$$

3. Compute eigenspaces:

$$E_{\mathbf{A},2} = \ker\left(\mathbf{A}-2\mathbf{I}\right) = \ker\left(\begin{bmatrix}-2 & 6\\ 1 & -3\end{bmatrix}\right) = \ker\left(\begin{bmatrix}1 & -3\\ 0 & 0\end{bmatrix}\right) = \operatorname{Span}\left(\begin{bmatrix}3\\1\end{bmatrix}\right);$$

$$E_{\mathbf{A},-3} = \ker(\mathbf{A}+3\mathbf{I}) = \ker\left(\begin{bmatrix}3 & 6\\1 & 2\end{bmatrix}\right) = \ker\left(\begin{bmatrix}1 & 2\\0 & 0\end{bmatrix}\right) = \operatorname{Span}\left(\begin{bmatrix}-2\\1\end{bmatrix}\right).$$

4. Ask: IS there an eigenbasis for $\mathbf{A}$?

   YES!

$$\vec{\mathbf{v}}_1 = \begin{bmatrix}3\\1\end{bmatrix}; \qquad \vec{\mathbf{v}}_2 = \begin{bmatrix}-2\\1\end{bmatrix}; \qquad \lambda_1 = 2; \qquad \lambda_2 = -3.$$

5. Solve $\begin{bmatrix}3 & -2\\1 & 1\end{bmatrix}\begin{bmatrix}c_1\\c_2\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix}$:

$$\mathbf{S} = \begin{bmatrix}3 & -2\\1 & 1\end{bmatrix} \implies \mathbf{S}^{-1} = \frac{1}{5}\begin{bmatrix}1 & 2\\-1 & 3\end{bmatrix}$$

$$\therefore \begin{bmatrix}c_1\\c_2\end{bmatrix} = \frac{1}{5}\begin{bmatrix}1 & 2\\-1 & 3\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} = \frac{1}{5}\begin{bmatrix}3\\2\end{bmatrix} \implies \begin{bmatrix}1\\1\end{bmatrix} = \frac{3}{5}\begin{bmatrix}3\\1\end{bmatrix} + \frac{2}{5}\begin{bmatrix}-2\\1\end{bmatrix}$$

6. Find the formula:

$$\mathbf{A}^k\begin{bmatrix}1\\1\end{bmatrix} = \frac{3}{5}(2)^k\begin{bmatrix}3\\1\end{bmatrix} + \frac{2}{5}(-3)^k\begin{bmatrix}-2\\1\end{bmatrix}.$$

□

Remark. The matrix $\mathbf{A} = \begin{bmatrix}0 & -1\\1 & 0\end{bmatrix}$ has no eigenvectors.

Proof. $\boxed{\text{Algebraic}}$

$$f_{\mathbf{A}}(t) = \begin{vmatrix}-t & -1\\1 & -t\end{vmatrix} = t^2 + 1 > 0 \forall t \in \mathbb{R}.$$

So, $f_{\mathbf{A}}(t)$ has no zeros $\implies$ no eigenvalues $\implies$ no eigenvectors.
$\boxed{\text{Geometric}}$ $\mathbf{A}$ encodes rotation counterclockwise by 90°.
The condition $\mathbf{A}\vec{\mathbf{x}} = \lambda\vec{\mathbf{x}}$ implies that $\mathbf{A}\vec{\mathbf{x}}$ and $\vec{\mathbf{x}}$ have to be on the same line.
Yet, rotation by 90° preserves no lines. ■

7.2 Diagonalization

Definition 7.2.1 (Diagonal Matrix). Let $\mathbf{D} = \begin{bmatrix}a_{11} & \cdots & a_{1n}\\ \vdots & \ddots & \vdots\\ a_{n1} & \cdots & a_{nn}\end{bmatrix}$, we say $\mathbf{D}$ is a **diagonal matrix** if $a_{ij} = 0$ for all $i \neq j$.

Theorem 7.2.1. $\mathbf{D}$ is diagonal if and only if $\mathbf{D}\vec{\mathbf{e}}_i = \lambda_i\vec{\mathbf{e}}_i$, $\lambda_i \in \mathbb{R}$. i.e., $e_1, \cdots, e_n$ are eigenvectors for $\mathbf{D}$. That is, eigenvalues are the diagonal entries: $\begin{bmatrix}\lambda_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \lambda_n\end{bmatrix}$.

Theorem 7.2.2. Properties of Diagonal Matrices

- Computing $\mathbf{D}^k$

$$\mathbf{D}^k = \begin{bmatrix} | & | & & | \\ \mathbf{D}^k\vec{\mathbf{e}}_1 & \mathbf{D}^k\vec{\mathbf{e}}_2 & \cdots & \mathbf{D}^k\vec{\mathbf{e}}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1^k\vec{\mathbf{e}}^1 & \lambda_2^k\vec{\mathbf{e}}^2 & \cdots & \lambda_n^k\vec{\mathbf{e}}^n \\ | & | & & | \end{bmatrix}$$
$$= \begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n^k \end{bmatrix}$$

- Computing $\mathbf{D}^{-1}$

$$\mathbf{D}^{-1} = \begin{bmatrix} \lambda_1^{-1} & 0 & \cdots & 0 \\ 0 & \lambda_2^{-1} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n^{-1} \end{bmatrix}$$

- Rank of $\mathbf{D}$

$$\text{rank}(\mathbf{D}) = \text{number of non-zero diagonal entries.}$$

- Nullity of $\mathbf{D}$

$$\text{nullity}(\mathbf{D}) = \text{number of zeros along the diagonal.}$$

- Determinant of $\mathbf{D}$

$$\det(\mathbf{D}) = \lambda_1\lambda_2\cdots\lambda_n$$

Definition 7.2.2 (Diagonalizable). Let $\mathbf{A}$ be an $n \times n$ matrix. $\mathbf{A}$ is said to be **diagonalizable** if there is an eigenbasis for $\mathbf{A}$. i.e., there is a basis $\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_n$ of $\mathbb{R}^n$ such that $\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_n$ are eigenvectors of $\mathbf{A}$.

Theorem 7.2.3. $\mathbf{A}$ is diagonalizable if and only if

$$\mathbf{A} = \mathbf{SDS}^{-1},$$

where $\mathbf{D}$ is diagonal with diagonal entries the eigenvalues of $\mathbf{A}$ ($\lambda_1 \cdots \lambda_n$), and $\mathbf{S}$ is invertible with column vectors the eigenvectors of $\mathbf{A}$ ($\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_n$). **Diagonalizing** a matrix means to find an invertible matrix $\mathbf{S}$ and a diagonal matrix $\mathbf{D}$ such that $\mathbf{A} = \mathbf{SDS}^{-1}$.

Example 7.2.1. Let $\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$. Diagonalize $\mathbf{A}$.

Solution. By definition, we know

$$f_{\mathbf{A}}(t) = \det\left(\begin{bmatrix} 1-t & 1 & 1 \\ 1 & 1-t & 1 \\ 1 & 1 & 1-t \end{bmatrix}\right) = 3t^2 - t^3 \implies t_1 = t_2 = 0, \quad t_3 = 3.$$

Therefore, we know

$$E_{\mathbf{A},0} = \ker(\mathbf{A}) = \ker\left(\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\right) = \operatorname{Span}\left(\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}\right);$$

$$E_{\mathbf{A},3} = \ker(\mathbf{A} - 3\mathbf{I}) = \ker\left(\begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix}\right) = \operatorname{Span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right)$$

Note that since $\vec{\mathbf{v}}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \vec{\mathbf{v}}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \vec{\mathbf{v}}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ span $\mathbb{R}^3$, they are eigenbasis of $\mathbf{A}$.

$$\begin{aligned} \therefore \mathbf{A} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1} &= \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1/3 & -2/3 & 1/3 \\ 1/3 & 1/3 & -2/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}^{-1}. \end{aligned}$$

□

Corollary 7.1. Linear Algebra Becomes Easy for Diagonalized Matrices

1. $\mathbf{A}^k = \mathbf{S}\mathbf{D}^k\mathbf{S}^{-1}$

   Proof.

   $$\mathbf{A}^k = \underbrace{\mathbf{S}\mathbf{D}\mathbf{S}^{-1}\mathbf{S}\mathbf{D}\mathbf{S}^{-1}\cdots\mathbf{S}\mathbf{D}\mathbf{S}^{-1}}_{k \text{ times}} = \mathbf{S}\mathbf{D}^k\mathbf{S}^{-1}$$

   ■

2. $\mathbf{A}^{-1} = \mathbf{S}\mathbf{D}^{-1}\mathbf{S}^{-1}$

   Proof. Since

   $$\mathbf{A}(\mathbf{S}\mathbf{D}^{-1}\mathbf{S}^{-1}) = \mathbf{S}\mathbf{D}\mathbf{S}^{-1}\mathbf{S}\mathbf{D}^{-1}\mathbf{S}^{-1} = \mathbf{S}\mathbf{D}\mathbf{D}^{-1}\mathbf{S}^{-1} = \mathbf{S}\mathbf{S}^{-1} = \mathbf{I},$$

   so we know that $\mathbf{A}^{-1} = \mathbf{S}\mathbf{D}^{-1}\mathbf{S}^{-1}$. ■

3. $\det(\mathbf{A}) = \det(\mathbf{D}) = \lambda_1 \cdots \lambda_n$

   Proof.

   $$\det(\mathbf{A}) = \det\big(\mathbf{SDS}^{-1}\big) = \det(\mathbf{S})\det(\mathbf{D})\det\big(\mathbf{S}^{-1}\big) = \det(\mathbf{S})\det(\mathbf{D})\det(\mathbf{S})^{-1} = \det(\mathbf{D}).$$

   ■

4. $f_{\mathbf{A}}(t) = \det(\mathbf{A} - t\mathbf{I}) = \det(\mathbf{D} - t\mathbf{I}) = f_{\mathbf{D}}(t) = \prod_{i=1}^{n} (\lambda_i - t)$

   Proof. Let's fix $t \in \mathbb{R}$. then

   $$\mathbf{S}(\mathbf{D} - t\mathbf{I})\mathbf{S}^{-1} = \mathbf{SDS}^{-1} - t\mathbf{SIS}^{-1} = \mathbf{A} - t\mathbf{I}.$$

   ■

5. $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{D}) =$ number of non-zero λ_i.

6. $\text{nullity}(\mathbf{A}) = \text{nullity}(\mathbf{D}) =$ number of zero λ_i.

7. If $f_{\mathbf{A}}(t)$ is not a polynomial with all real roots, then $\mathbf{A}$ is not diagonalizable.

7.3 Procedure of Finding an Eigenbasis

Definition 7.3.1 (Revisit Definition of Characteristic Polynomials). For an $n \times n$ matrix $\mathbf{A}$, its characteristic polynomial is a function $f_{\mathbf{A}} : \mathbb{R} \to \mathbb{R}$ defined by $f_{\mathbf{A}}(t) = \det(\mathbf{A} - t\mathbf{I})$.

Theorem 7.3.1. $f_{\mathbf{A}}(\lambda) = 0$ if and only if λ is an eigenvalue of $\mathbf{A}$.

Theorem 7.3.2. $f_{\mathbf{A}}(t)$ is a polynomial. i.e., $f_{\mathbf{A}}(t) = a_d t^d + ad - 1t^{d-1} + \cdots + a_0$, where $a_d \neq 0$ and $a_i \in \mathbb{R}$. We say d is the degree of $f_{\mathbf{A}}(t)$.

Remark. Determinant can be calculated without division: Laplace Expansion

Proof. Note that $\mathbf{A} - t\mathbf{I}$ is an $n \times n$ matrix with polynomial entries.

By Laplace expansion, $\det(\mathbf{A} - t\mathbf{I})$ is the sums and products of those polynomial entries, and thus it is a polynomial. ■

Proposition 7.1. If $k \in \mathbb{R}$ and $\mathbf{B}$ is an $n \times n$ matrix, then

$$\det(k\mathbf{B}) = k^n \det(\mathbf{B}).$$

Proof.

$$\det\left(\begin{bmatrix} - & k\vec{\mathbf{r}}_1 & - \\ & \vdots & \\ - & k\vec{\mathbf{r}}_n & - \end{bmatrix}\right) = k^n \cdot \det\left(\begin{bmatrix} - & \vec{\mathbf{r}}_1 & - \\ & \vdots & \\ - & \vec{\mathbf{r}}_n & - \end{bmatrix}\right)$$

■

Theorem 7.3.3. If $\mathbf{A}$ is an $n \times n$ matrix, then the degree of $f_{\mathbf{A}}$ is n.

Proof. Since $f_{\mathbf{A}}(t)$ is a polynomial, $f_{\mathbf{A}}(t) = a_d t^d + ad - 1t^{d-1} + \cdots + a_0$, where $a_d \neq 0$ and $a_i \in \mathbb{R}$.

If we can prove $f_{\mathbf{A}}(t)$ *and* $\det(-t\mathbf{I})$ *has the same growth rate, and since* $\det(-t\mathbf{I}) = (-t)^n \det(\mathbf{I}) = (-t)^n$, *we can say* $f_{\mathbf{A}}(t)$ *has a degree of* n. *Therefore, we want to show* $\lim\limits_{t\to\infty} \dfrac{f_{\mathbf{A}}(t)}{t^n}$ *is finite and non-zero.*

$$\lim_{t\to\infty} \frac{f_{\mathbf{A}}(t)}{t^n} = \lim_{t\to\infty} \frac{\det(\mathbf{A} - t\mathbf{I})}{t^n} = \lim_{t\to\infty} \det\left(\frac{\mathbf{A}}{t} - \mathbf{I}\right) = \lim_{t\to 0} \det(t\mathbf{A} - \mathbf{I}) = \det(-\mathbf{I}) = (-1)^n$$

$\therefore a_d = (-1)^n$, and $f_{\mathbf{A}}(t)$ has a degree of n. ■

Remark. If $\lambda_1, \cdots \lambda_k$ are roots of $f_{\mathbf{A}}(t)$, them

$$f_{\mathbf{A}}(t) = (t - \lambda_1)^{M_1}(t - \lambda_2)^{M_2} \cdots (t - \lambda_k)^{M_k} g(x),$$

where $g(x)$ has no real roots. Then,

$$n = M_1 + M_2 + \cdots + M_k + \text{degree}(g(x)).$$

Counted with multiplicity (this power M_k), $\mathbf{A}$ has <u>at most</u> n eigenvalues.

Theorem 7.3.4. $\mathbf{A}$ has exactly n roots (counted with multiplicity) when $\mathbf{A}$ is diagonalizable.

Definition 7.3.2 (Algebraic Multiplicity). The **algebraic multiplicity** of a matrix $\mathbf{A}$ is the multiplicity of an eigenvalue λ in the characteristic polynomial of $\mathbf{A}$.

Theorem 7.3.5. If we write $f_{\mathbf{A}}(t)$ as $f_{\mathbf{A}}(t) = a_d t^d + ad - 1t^{d-1} + \cdots + a_0$, where $a_d \neq 0$ and $a_i \in \mathbb{R}$, then $a_0 = \det(\mathbf{A})$

Proof.

$$f_{\mathbf{A}}(0) = a_0 = \det(\mathbf{A} - 0 \cdot \mathbf{I}) = \det(\mathbf{A}).$$

■

Example 7.3.1. Prove that matrix $\mathbf{B} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ is not diagonalizable.

Proof. Note that $f_{\mathbf{B}}(t) = \begin{vmatrix} -t & 1 \\ 0 & -t \end{vmatrix} = t^2 \implies \lambda_1 = 0, \text{ multiplicity} = 2.$

$\boxed{\text{Method 1}}$ Assume for the sake of contradiction that $\mathbf{B}$ is diagonalizable.

Therefore, $\mathbf{B} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1}$, where $\mathbf{S} = \begin{bmatrix} | & | \\ \vec{\mathbf{v}}_1 & \vec{\mathbf{v}}_2 \\ | & | \end{bmatrix}$ and $\mathbf{D} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$ for eigenvectors $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$ with eigenvalues λ_1 and λ_2, respectively.

Then,

$$\mathbf{D} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = 0 \cdot \mathbf{I}$$

$$\mathbf{B} = \mathbf{SDS}^{-1} = \mathbf{S}(0 \cdot \mathbf{I})\mathbf{S}^{-1} = 0 \cdot \mathbf{S} \cdot \mathbf{I} \cdot \mathbf{S}^{-1} = 0 \cdot \mathbf{S} \cdot \mathbf{S}^{-1} = 0$$

⨳ This is a contradiction that $\mathbf{B} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} =\neq 0$.

Therefore, **B** cannot be diagonalizable.

$\boxed{\text{Method 2}}$ From Method 1, we know that

$$E_{\mathbf{B},0} = \ker(\mathbf{B} - 0 \cdot \mathbf{I}) = \ker\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) = \text{Span}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)$$

Since $\ker(\mathbf{B})$ is 1-dimensional, it doesn't contain a basis for $\mathbb{R}^2$.

So, **B** doesn't have an eigenbasis. ■

Theorem 7.3.6. A matrix $\mathbf{C} = \begin{bmatrix} \lambda & * & \cdots & \cdots \\ 0 & \lambda & & \vdots \\ \vdots & & \ddots & * \\ \cdots & \cdots & 0 & \lambda \end{bmatrix}$ is diagonalizable if and only if **C** is diagonal.

Proof. Since $f_{\mathbf{C}}(t) = \begin{vmatrix} \lambda - t & * & \cdots & \cdots \\ 0 & \lambda - t & & \vdots \\ \vdots & & \ddots & * \\ \cdots & \cdots & 0 & \lambda - t \end{vmatrix} = (\lambda - t)^n$, we know **C** has λ as the only eigenvalue with algebraic multiplicity of n.

$$\therefore \mathbf{D} = \begin{bmatrix} \lambda & 0 & \cdots & \cdots \\ 0 & \lambda & & \vdots \\ \vdots & & \ddots & 0 \\ \cdots & \cdots & 0 & \lambda \end{bmatrix} = \lambda \cdot \mathbf{I}.$$

$$\therefore \mathbf{C} = \mathbf{SDS}^{-1} = \mathbf{S}(\lambda\mathbf{I})\mathbf{S}^{-1} = \lambda\mathbf{SIS}^{-1} = \lambda\mathbf{I}.$$

■

Theorem 7.3.7. For matrix $\mathbf{C} = \begin{bmatrix} \lambda & * & \cdots & \cdots \\ 0 & \lambda & & \vdots \\ \vdots & & \ddots & * \\ \cdots & \cdots & 0 & \lambda \end{bmatrix}$, λ is the only eigenvalue. Also, $E_{\mathbf{C},\lambda} = \ker(\mathbf{C} - \lambda\mathbf{I})$ contains a basis if and only if $\ker(\mathbf{C} - \lambda\mathbf{I})$ is the entire space. i.e., $\mathbf{C} - \lambda\mathbf{I} = 0$, or $\mathbf{C} = \lambda\mathbf{I}$.

Theorem 7.3.8. Let $\mathbf{A}$ be an $n \times n$ matrix, and $\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_k$ be eigenvectors of $\mathbf{A}$. The vectors $\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_k$ are linearly independent if for every eigenvalue λ of $\mathbf{A}$, the set of these vectors with eigenvalue λ. i.e., $\{\vec{\mathbf{v}}_i \mid \mathbf{A}\vec{\mathbf{v}}_i = \lambda\vec{\mathbf{v}}_i\}$ is linearly independent.

Finding an eigenbasis/diagonalizing A as an $n \times n$ matrix.

1. Find eigenvalues of $\mathbf{A}$
   (a) Compute $f_{\mathbf{A}}(t)$
   (b) Find the roots of $f_{\mathbf{A}}(t)$ and the multiplicity $M_1, \cdots, M_k$

   Remark. If $M_1 + M_2 + \cdots + M_k \neq n$, then STOP. $\mathbf{A}$ is not diagonalizable.

   (c) Form matrix $\mathbf{D}$.
2. Find basis for eigenspaces:
   (a) Form $\mathbf{S} = \begin{bmatrix} | & & | \\ \vec{\mathbf{v}}_1 & \cdots & \vec{\mathbf{v}}_n \\ | & & | \end{bmatrix}$, where $\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_n$ are linearly independent.
   (b) For each λ_i, compute a basis for $\ker(\mathbf{A} - \lambda_i\mathbf{I})$.

   Remark. If $\dim(\ker(\mathbf{A} - \lambda_i\mathbf{I})) < M_i$, then STOP. There is no enough eigenvectors and $\mathbf{A}$ is not diagonalizable.
3. By theorem, the concatenation of the lists of bases is an eigenbasis.

Definition 7.3.3 (Geometric Multiplicity). Let λ be an eigenvalue of $\mathbf{A}$.. The **geometric multiplicity** of λ is $\dim(\ker(\mathbf{A} - \lambda\mathbf{I}))$, the number of linearly independent vector in an eigenspace.

Theorem 7.3.9. For a matrix to be diagonalizable,

$$\text{geometric multiplicity} = \text{algebraic multiplicity}.$$

7.4 Multiplicity

Definition 7.4.1 (Multiplicity). Let $\mathbf{A}$ be an $n \times n$ matrix and λ be an eigenvalue of $\mathbf{A}$:

1. The **algebraic multiplicity** of λ is the largest k such that $f_{\mathbf{A}}(t) = (t - \lambda)^k g(t)$, where $g(t)$ is a polynomial. We denote the algebraic multiplicity of λ as $\text{almu}(\lambda) = k$. $\text{almu}(\lambda)$ is the multiplicity of λ as a root of $f_{\mathbf{A}}(t)$.
2. The **geometric multiplicity** of λ is $\text{gemu}(\lambda) = \dim(\ker(\mathbf{A} - \lambda\mathbf{I}))$. $\text{gemu}(\lambda)$ is the maximum number of linearly independent eigenvectors with eigenvalue λ.

Theorem 7.4.1. $\text{gemu}(\lambda) \leq \text{almu}(\lambda)$.

Remark. Note that in $\mathbf{A}-\lambda\mathbf{I}$, every non-zero diagonal entry contributes a pivot to $\text{rref}(\mathbf{A}-\lambda\mathbf{I})$. Then, $\text{rank}(\mathbf{A}-\lambda\mathbf{I}) \geq$ the number of diagonal entries that is not λ.

Therefore, $\text{nullity}(\mathbf{A}-\lambda\mathbf{I}) \leq$ the number of diagonal entries that equals λ.

Hence, $\text{gemu}(\lambda) \leq \text{almu}(\lambda)$.

Proof. Assume $\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_g$ is a basis of $E_{\mathbf{A},\lambda}$. Then, $\text{gemu}(\lambda) = g$.

Choose $\vec{\mathbf{v}}_{g+1}, \cdots, \vec{\mathbf{v}}_n$ such that $\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_g, \vec{\mathbf{v}}_{g+1}, \cdots, \vec{\mathbf{v}}_n$ is a basis for $\mathbb{R}^n$. Then,

$$\mathbf{S} = \begin{bmatrix} | & & | & & | \\ \vec{\mathbf{v}}_1 & \cdots & \vec{\mathbf{v}}_g & \cdots & \vec{\mathbf{v}}_n \\ | & & | & & | \end{bmatrix}$$

$\mathbf{S}$ *is invertible since* $\vec{\mathbf{v}}_1, \cdots \vec{\mathbf{v}}_n$ *is a basis.*

Claim 7.1. $\mathbf{B} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$, where $\mathbf{B} = \left[\begin{array}{ccc|c} \lambda & & & \\ & \ddots & & * \\ & & \lambda & \\ \hline & \underbrace{0}_{g-\text{columns}} & & \mathbf{C} \end{array}\right]$

$$f_{\mathbf{B}}(t) = \det\big(\mathbf{S}\mathbf{A}\mathbf{S}^{-1} - t\mathbf{I}\big) = \det\big(\mathbf{S}\mathbf{A}\mathbf{S}^{-1} - t\mathbf{S}^{-1}\mathbf{I}\mathbf{S}\big) = \det\big(\mathbf{S}^{-1}(\mathbf{A}-t\mathbf{I})\mathbf{S}\big)$$
$$= \det(\mathbf{A}-t\mathbf{I}) = f_{\mathbf{A}}(t).$$

Since $f_{\mathbf{B}}(t) = (\lambda - t)^g f_{\mathbf{C}}(t) = f_{\mathbf{A}}(t)$, we know that gemu $\leq$ almu. ■

Theorem 7.4.2. For a matrix to be diagonalizable, it is necessary that $\text{almu}(\lambda) = \text{gemu}(\lambda)$ for all λ.

Theorem 7.4.3. Let $\mathbf{A}$ be an $n \times n$ matrix. If $f_{\mathbf{A}}(t)$ has n distinct real roots, then $\mathbf{A}$ is diagonalizable.

Proof. Every eigenvalue has an eigenvector: $\det(\mathbf{A}-\lambda\mathbf{I}) = 0 \implies (\mathbf{A}-\lambda\mathbf{I})$ is not invertible.

$$\therefore \ker(\mathbf{A}-\lambda\mathbf{I}) =\neq 0.$$

Therefore, there are eigenvectors $\vec{\mathbf{v}}_1, \cdots \vec{\mathbf{v}}_n$ for eigenvalues $\lambda_1, \cdots, \lambda_n$, respectively.

Since eigenvectors with distinct eigenvalues are linearly independent, $\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_n$ is an eigenbasis. ■

8 Singular Value Decomposition

8.1 The Spectral Theorem

Definition 8.1.1 (Symmetry). A matrix $\mathbf{A}$ is called **symmetric** if $\mathbf{A} = \mathbf{A}^{\mathbf{T}}$. A symmetric matrix is symmetric across the diagonal. That is, $a_{ij} = a_{ji}$.

Theorem 8.1.1 (Spectral Theorem). $\mathbf{A}$ is symmetric if and only if $\mathbf{A}$ has an orthogonal eigenbasis. Equivalently, $\mathbf{A} = \mathbf{SDS}^{-1}$, where $\mathbf{D}$ is diagonal and $\mathbf{S}$ is orthogonal (having orthogonal columns). That is, $\mathbf{A} = \mathbf{SDS}^{\mathbf{T}}$ because $\mathbf{S}^{-1} = \mathbf{S}^{\mathbf{T}}$ if $\mathbf{S}$ is orthogonal.

Proof. *Given* $\mathbf{A} = \mathbf{SDS}^{\mathbf{T}}$, *we want to show that* $\mathbf{A} = \mathbf{A}^{\mathbf{T}}$.
Note that since $\mathbf{D}$ is diagonal, we have $\mathbf{D} = \mathbf{D}^{\mathbf{T}}$. Then

$$\mathbf{A}^{\mathbf{T}} = (\mathbf{SDS}^{\mathbf{T}})^{\mathbf{T}} = (\mathbf{S}^{\mathbf{T}})^{\mathbf{T}}(\mathbf{D})^{\mathbf{T}}\mathbf{S}^{\mathbf{T}} = \mathbf{SD}^{\mathbb{R}}\mathbf{S}^{\mathbf{T}} = \mathbf{SDS}^{\mathbf{T}} = \mathbf{A}.$$

■

Theorem 8.1.2. Orthogonal projection is symmetric.

Proof. Let V to be a subspace of $\mathbb{R}^n$. Define $\mathrm{Proj}_{\vec{\mathbf{v}}} : \mathbb{R}^n \to \mathbb{R}^n$ as $\vec{\mathbf{x}} \longmapsto \vec{\mathbf{x}}^{\parallel} \in V$, where $\vec{\mathbf{x}} = \vec{\mathbf{x}}^{\parallel} + \vec{\mathbf{x}}^{\perp}$ and $\vec{\mathbf{x}}^{\perp} \in V^{\perp}$. Finding eigenspaces of $\mathrm{Proj}_{\vec{\mathbf{v}}}$, we get

$$E_{\mathrm{Proj}_{\vec{\mathbf{v}}},1} = V, \qquad \text{and} \qquad E_{\mathrm{Proj}_{\vec{\mathbf{v}}},0} = V^{\perp}.$$

Since eigenspaces are perpendicular, $\mathrm{Proj}_{\vec{\mathbf{v}}}$ is symmetric. ■

Corollary 8.1. Let

$$V = \mathrm{Span}\left(\begin{bmatrix} \vec{\mathbf{v}}_1 \\ \vdots \\ \vec{\mathbf{v}}_n \end{bmatrix}\right) = \mathrm{Span}(\vec{\mathbf{v}}),$$

then $\vec{\mathbf{x}} \longmapsto V$ is $\dfrac{\vec{\mathbf{x}} \cdot \vec{\mathbf{v}}}{\vec{\mathbf{v}} \cdot \vec{\mathbf{v}}} \cdot \vec{\mathbf{v}}$. That is, $\dfrac{1}{\vec{\mathbf{v}} \cdot \vec{\mathbf{v}}} \begin{bmatrix} \vec{\mathbf{v}} \\ \vdots \\ vecv_n \end{bmatrix} \left(\begin{bmatrix} \vec{\mathbf{v}}_1 & \cdots & \vec{\mathbf{v}}_n \end{bmatrix} \vec{\mathbf{x}}\right)$.

$$\therefore \mathrm{Proj}_{\vec{\mathbf{v}}}(\vec{\mathbf{x}}) = \frac{1}{\vec{\mathbf{v}} \cdot \vec{\mathbf{v}}} \begin{bmatrix} v_1v_1 & v_1v_2 & \cdots & \cdots \\ v_1v_2 & v_2v_2 & & \vdots \\ \cdots & & \ddots & \vdots \\ \cdots & \cdots & \cdots & v_nv_n \end{bmatrix}$$

Theorem 8.1.3 (Adjoint Property of Transpose). If $\mathbf{A}$ is an $n \times n$ matrix, $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2 \in \mathbb{R}^n$. Then, $(\mathbf{A}\vec{\mathbf{v}}_1) \cdot \vec{\mathbf{v}}_2 = \vec{\mathbf{v}}_1 \cdot \left(\mathbf{A}^{\mathbf{T}}\vec{\mathbf{v}}_2\right)$. That is, bringing a matrix through a dot product, transpose it. Specially, if $\mathbf{A}$ is symmetric, $\mathbf{A} = \mathbf{A}^{\mathbf{T}}$, and $(\mathbf{A}\vec{\mathbf{v}}_1) \cdot \vec{\mathbf{v}}_2 = \vec{\mathbf{v}}_1 \cdot (\mathbf{A}\vec{\mathbf{v}}_2)$.

Proof.

$$(\mathbf{A}\vec{\mathbf{v}}_1)\cdot\vec{\mathbf{v}}_2 = (\mathbf{A}\vec{\mathbf{v}}_1)\vec{\mathbf{v}}_2 = (\vec{\mathbf{v}}_1)^{\mathbf{T}}\big(\mathbf{A}^{\mathbf{T}}\vec{\mathbf{v}}_2\big) = \vec{\mathbf{v}}_1\cdot\big(\mathbf{A}^{\mathbf{T}}\vec{\mathbf{v}}_2\big).$$

■

Theorem 8.1.4 (Spectral Theorem – Continued 1). If $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2$ are eigenvectors for $\mathbf{A}$ with eigenvalues $\lambda_1 \neq \lambda_2$, then $\vec{\mathbf{v}}_1 \perp \vec{\mathbf{v}}_2$.

Proof. Note that

$$\lambda_1\vec{\mathbf{v}}_1\cdot\vec{\mathbf{v}}_2 = \mathbf{A}\vec{\mathbf{v}}_1\cdot\vec{\mathbf{v}}_2 = \vec{\mathbf{v}}_1\cdot(\mathbf{A}\vec{\mathbf{v}}_2) = \lambda_2\vec{\mathbf{v}}_1\cdot\vec{\mathbf{v}}_2,$$

However, by our assumption we have $\lambda_1 \neq \lambda_2$. So it must be $\vec{\mathbf{v}}_1\cdot\vec{\mathbf{v}}_2 = 0$.
That is, exactly, $\vec{\mathbf{v}}_1 \perp \vec{\mathbf{v}}_2$. ■

Corollary 8.2. Distinct eigenspaces are perpendicular.

Theorem 8.1.5 (Spectral Theorem – Continued 2). If $\mathbf{A}$ is a symmetric matrix, then $\mathbf{A}$ has an orthogonal eigenbasis.

Proof.

Claim 8.1. $f_{\mathbf{A}}(t)$ has all real roots.

If $\lambda = x + \mathrm{i}y$ is a root of $f_{\mathbf{A}}(t)$, show $y = 0$.

Let $\vec{\mathbf{v}} + \mathrm{i}\vec{\mathbf{w}} \in \mathbb{C}^n$ be an eigenvector for $\mathbf{A}$ with eigenvalue λ. Then, $\vec{\mathbf{v}} - \mathrm{i}\vec{\mathbf{w}} \in \mathbb{C}^n$ is also an eigenvector for $\mathbf{A}$ with eigenvalue $\lambda^* = x - \mathrm{i}y$.

$$\therefore \begin{cases} \mathbf{A}(\vec{\mathbf{v}} - \mathrm{i}\vec{\mathbf{w}}) = \lambda^*(\vec{\mathbf{v}} - \mathrm{i}\vec{\mathbf{w}}) \\ \mathbf{A}(\vec{\mathbf{v}} + \mathrm{i}\vec{\mathbf{w}}) = \lambda(\vec{\mathbf{v}} + \mathrm{i}\vec{\mathbf{w}}) \end{cases}$$

$$\begin{aligned}
\therefore (\vec{\mathbf{v}} - \mathrm{i}\vec{\mathbf{w}})\cdot\mathbf{A}(\vec{\mathbf{v}} + \mathrm{i}\vec{\mathbf{w}}) &= \lambda(\vec{\mathbf{v}} - \mathrm{i}\vec{\mathbf{w}})\cdot(\vec{\mathbf{v}} + \mathrm{i}\vec{\mathbf{w}}) \\
&= \lambda(\vec{\mathbf{v}}\cdot\vec{\mathbf{v}} + \vec{\mathbf{w}}\cdot\vec{\mathbf{w}}) \\
&= \lambda(\underbrace{\|\vec{\mathbf{v}}\|^2 + \|\vec{\mathbf{w}}\|^2}_{\text{greater than 0}}) \\
\mathbf{A}(\vec{\mathbf{v}} - \mathrm{i}\vec{\mathbf{w}})\cdot(\vec{\mathbf{v}} + \mathrm{i}\vec{\mathbf{w}}) &= \lambda^*(\vec{\mathbf{v}} - \mathrm{i}\vec{\mathbf{w}})\cdot(\vec{\mathbf{v}} + \mathrm{i}\vec{\mathbf{w}}) \\
&= \lambda^*(\underbrace{\|\vec{\mathbf{v}}\|^2 + \|\vec{\mathbf{w}}^2\|}_{\text{greater than 0}}) \\
\therefore \lambda(\|\vec{\mathbf{v}}\|^2 + \|\vec{\mathbf{w}}\|^2) &= \lambda^*(\|\vec{\mathbf{v}}\|^2 + \|\vec{\mathbf{w}}\|^2)
\end{aligned}$$

Since $\|\vec{\mathbf{v}}\|^2 + \|\vec{\mathbf{w}}\|^2 > 0$, it must be $\lambda = \lambda^*$. That is, $x + \mathrm{i}y = x - \mathrm{i}y$. So, $y = 0$.
Thus, all the roots are real.

Claim 8.2. $\mathbf{A}$ has an eigenbasis. That is, $\mathrm{gemu}(\lambda) = \mathrm{almu}(\lambda)\forall\lambda$.

We can write $\mathbf{A} = \mathbf{S}_1 \underbrace{\left[\begin{array}{c|ccc} * & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \mathbf{A}_{n-1} & \\ 0 & & & \end{array}\right]}_{\text{symmetric matrix}} \mathbf{S}_1^{-1}$

We can do the same thing over and over again, and eventually, we will get a diagonal matrix. So we know $\text{gemu}(\lambda) = \text{almu}(\lambda)$. ■

Corollary 8.3. We can always diagonalize $\mathbf{A}$.

Find orthogonal eigenbasis of a symmetric matrix A

1. Find an eigenbasis for $\mathbf{A}$.
2. Run Gram-Schudt on eigenbasis. The result is orthogonal eigenbasis.

8.2 Quadratic Form, Principal Axis Theorem

Definition 8.2.1 (Quadratic Form). A **quadratic form** is a function $f : \mathbb{R}^n \to \mathbb{R}$ of the form

$$f(x_1, \cdots, x_n) = \sum_{1 \le i,j \le n} a_{ij} x_i x_j$$

for some constants $a_{ij} \in \mathbb{R}$.

Example 8.2.1. $7x^2 + 3xy + 4y^2$ and $7x^2 + 3xy + 4xz + 2y^2 + 3yz + 7z^2$ are quadratic forms.

Definition 8.2.2 (Diagonal form). A quadratic form is called a **diagonal form** if

$$f(x_1, \cdots, x_n) = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2,$$

where $\lambda_i \in \mathbb{R}$.

Example 8.2.2. x^2+y^2, x^2-y^2, and $-x^2-y^2$ are examples of quadratic forms in the diagonal form. But xy or $x^2 + 7xy + 3y^2$ are not examples of diagonal forms.

Definition 8.2.3 (Degenerate). A diagonal form is called **degenerate** if $\lambda_i = 0$ for some i. If $\lambda_i \neq 0 \forall i$, then the diagonal form is called **non-degenerate**.

Theorem 8.2.1. Let f be a non-degenerate diagonal form:

1. If all $\lambda_i > 0$, then $f(\vec{\mathbf{x}}) \ge 0$ and $f(\vec{\mathbf{x}}) = 0$ if and only if $\vec{\mathbf{x}} = 0$. That is, $\vec{\mathbf{0}}$ is a global minimum, and f is positive definite.
2. If all $\lambda_i < 0$, then $f(\vec{\mathbf{x}}) \le 0$ and $f(\vec{\mathbf{x}}) = 0$ if and only if $\vec{\mathbf{x}} = 0$. That is, $\vec{\mathbf{0}}$ is a global maximum, and f is negative definite.

3. If some λ_i are positive and some are negative, there is no local maxima or minima, and we say f is indefinite.

Remark.

$$\vec{\mathbf{x}}^{\mathbf{T}}\vec{\mathbf{x}} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1^2 + \cdots + x_n^2$$

Remark. Let $\mathbf{D} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \lambda_n \end{bmatrix}$, then we have

$$\vec{\mathbf{x}}^{\mathbf{T}}\mathbf{D}\vec{\mathbf{x}} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \lambda_1 x_1 \\ \vdots \\ \lambda_n x_n \end{bmatrix} = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2.$$

Theorem 8.2.2. Let $\mathbf{A}$ be an $n \times n$ matrix, then $f(\vec{\mathbf{x}}) = \vec{\mathbf{x}}^{\mathbf{T}}\mathbf{A}\vec{\mathbf{x}}$ is a quadratic form. In general,

$$\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{1 \le i,j \le n} a_{ij} x_i x_j.$$

Theorem 8.2.3. Let $f(x_1, \cdots, x_n) = \sum_{i \le j} c_{ij} x_i x_j$, then there is a unique symmetric matrix such that $f = \vec{\mathbf{x}}^{\mathbf{T}}\mathbf{A}\vec{\mathbf{x}}$, where the $ij-$th entry of $\mathbf{A} = \begin{cases} c_{ii}, & i = j \\ c_{ij} & i \ne j \end{cases}$

Example 8.2.3. $7x^2 + 11xy + y^2$ can be written as $\begin{bmatrix} 7 & 11/2 \\ 11/2 & 1 \end{bmatrix}$.

Theorem 8.2.4 (Principal Axes Theorem). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a quadratic form, then there exists an orthogonal matrix $\mathbf{S}$ and a diagonal quadratic form $d : R^n \to \mathbb{R}$ such that

$$f(\vec{\mathbf{x}}) = d \circ \mathbf{S}^{\mathbf{T}}\vec{\mathbf{x}} = d\big(\mathbf{S}^{\mathbf{T}}\vec{\mathbf{x}}\big).$$

Any quadratic form looks diagonal in some coordinate $\lambda_1 x_1^2$ system.

Proof. $f(\vec{\mathbf{x}}) = \vec{\mathbf{x}}^{\mathbf{T}}\mathbf{A}\vec{\mathbf{x}}$, where $\mathbf{A}$ is symmetric.
Note that $\mathbf{A} = \mathbf{SDS}^{-1}$, where $\mathbf{S}$ is orthogonal and $\mathbf{D}$ is diagonal.
So, $f(\vec{\mathbf{x}}) = \vec{\mathbf{x}}^{\mathbf{T}}(\mathbf{SDS}^{-1})\vec{\mathbf{x}} = \big(\mathbf{S}^{\mathbf{T}}\vec{\mathbf{x}}\big)^{\mathbf{T}}\mathbf{D}\big(\mathbf{S}^{\mathbf{T}}\vec{\mathbf{x}}\big)$.
Since $d(\vec{\mathbf{x}}) = \vec{\mathbf{x}}^{\mathbf{T}}\mathbf{D}\vec{\mathbf{x}}$, we know $f(\vec{\mathbf{x}}) = d\big(\mathbf{S}^{\mathbf{T}}\big)\vec{\mathbf{x}}$. ∎

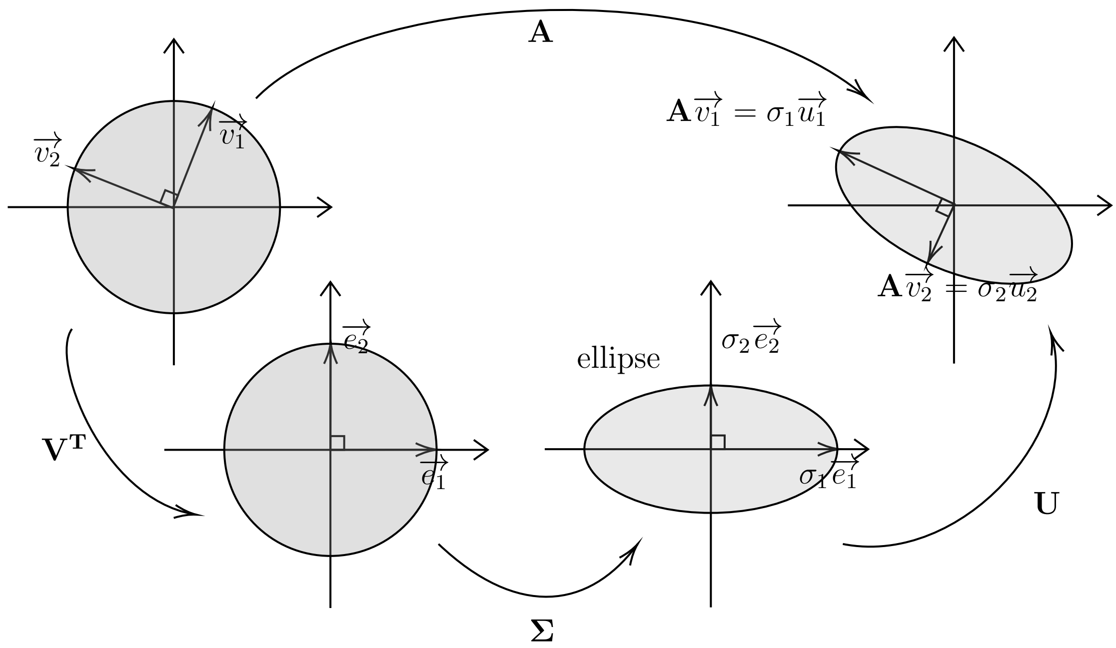
Corollary 8.4. f is positive definite if $\lambda_i > 0$ and f is negative definite if $\lambda_i < 0$.

8.3 Singular Value Decomposition

Definition 8.3.1 (Sigular Value Decomposition). The **singular value decomposition (SVD)** is a recipe to write a general matrix $\mathbf{A}$ as a product of matrices which is easy to understand geometrically.

Theorem 8.3.1. Let $\mathbf{A}$ to be an $n \times m$ matrix, then $\mathbf{A} = \mathbf{U\Sigma V^T}$, where $\mathbf{V}$ is an orthogonal matrix (so $\mathbf{V^T} = \mathbf{V}^{-1}$), $\mathbf{\Sigma}$ is an $n \times m$ matrix, whose $ij-$elements are all zero, and whose ii-entries satisfy $a_{11} \geq a_{22} \geq a_{33} \geq \cdots \geq 0$, and $\mathbf{U}$ is an orthogonal matrix.

Example 8.3.1. Suppose $\mathbf{A} = \begin{bmatrix} | & | \\ \vec{\mathbf{u}}_1 & \vec{\mathbf{u}}_2 \\ | & | \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} - & \vec{\mathbf{v}}_1 & - \\ - & \vec{\mathbf{v}}_2 & - \end{bmatrix}$:



1. $\sigma\vec{\mathbf{u}}_1$ is a longest vector on the ellipse (image of the unit circle under $\mathbf{A}$).
   (a) σ_1 is its length
   (b) $\vec{\mathbf{u}}_1$ is a unit vector pointing in direction.
   (c) $\vec{\mathbf{v}}_1$ is a vector on unit circle such $\|\mathbf{A}\vec{\mathbf{v}}_1\|$ is maximized.
2. $\mathbf{A}\vec{\mathbf{v}}_2 = \sigma_2\vec{\mathbf{u}}_2$ is a shortest vector on the ellipse (image of unit circle under $\mathbf{A}$).
   (a) σ_2 is its length
   (b) $\vec{\mathbf{u}}_2$ is a vector on unit circle such that $\|\mathbf{A}\vec{\mathbf{v}}_2\|$ is minimized.

Remark. SVD of $\mathbf{A}$ encodes information about lengths change under $\mathbf{A}$. Let $\vec{\mathbf{x}} \in \mathbb{R}^m$ and consider $\|\mathbf{A}\vec{\mathbf{x}}\| = \sqrt{\mathbf{A}\vec{\mathbf{x}} \cdot \mathbf{A}\vec{\mathbf{x}}}$, then

$$\begin{aligned}\mathbf{A}\vec{\mathbf{x}} \cdot \mathbf{A}\vec{\mathbf{x}} = \vec{\mathbf{x}} \cdot \mathbf{A}^{\mathbf{T}}\mathbf{A}\vec{\mathbf{x}} &= \left(\sum c_i \vec{\mathbf{v}}_i\right) \cdot \left(\mathbf{A}^{\mathbf{T}}\mathbf{A}\right)\left(\sum c_i \vec{\mathbf{v}}_i\right) = \left(\sum_{i=1}^{m} c_i k \vec{\mathbf{v}}_i\right) \cdot \left(\sum_{j=1}^{m} c_j \lambda_j \vec{\mathbf{v}}_j\right) \\ &= \sum_{i,j} c_i c_j (\vec{\mathbf{v}}_i \cdot \vec{\mathbf{v}}_j)\lambda_j = \sum c_i^2 \lambda_i.\end{aligned}$$

So, $\|\mathbf{A}\vec{\mathbf{x}}\| = \sqrt{\sum c_i^2 \lambda_i}$ and $\|\vec{\mathbf{x}}\| = \sqrt{\sum c_i^2}$. Therefore, $\boxed{\|\mathbf{A}\vec{\mathbf{v}}_i\| = \sqrt{\lambda_i} = \sigma_i}$

Since $\mathbf{V}$ is orthogonal, $\mathbf{V}^{\mathbf{T}}\vec{\mathbf{x}} = \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}$. Since $\mathbf{\Sigma} = \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_m} \end{bmatrix}$, then

$$\mathbf{\Sigma}\mathbf{V}^{\mathbf{T}}\vec{\mathbf{x}} = \begin{bmatrix} \sqrt{\lambda_1}c_1 \\ \vdots \\ \sqrt{\lambda_m}c_m \end{bmatrix}.$$

Definition 8.3.2 (Singular Value). Let $\mathbf{A}$ be a matrix, the **singular value** of $\mathbf{A}$ are the square roots of the positive eigenvalues of $\mathbf{A}^{\mathbf{T}}\mathbf{A}$. i.e., $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ are eigenvalues of $\mathbf{A}^{\mathbf{T}}\mathbf{A}$, and $\sigma_i = \sqrt{\lambda_i}$ whenever $\lambda_i > 0$.

Remark. To find $\mathbf{U}$, we can consider the following:

$$\vec{\mathbf{u}}_i = \begin{bmatrix} | & | & & | & | & & | \\ \dfrac{\mathbf{A}\vec{\mathbf{v}}_1}{\sigma_1} & \dfrac{\mathbf{A}\vec{\mathbf{v}}_2}{\sigma_2} & \cdots & \dfrac{\mathbf{A}\vec{\mathbf{v}}_r}{\sigma_r} & \vec{\mathbf{x}}_1 & \cdots & \vec{\mathbf{x}}_k \end{bmatrix},$$

where $\vec{\mathbf{x}}_1, \cdots, \vec{\mathbf{x}}_k$ are orthogonal and span orthogonal component to the space spanned by first r columns.

Also, note that $\dfrac{\mathbf{A}\vec{\mathbf{v}}_1}{\sigma_1}, \dfrac{\mathbf{A}\vec{\mathbf{v}}_2}{\sigma_2}, \cdots, \dfrac{\mathbf{A}\vec{\mathbf{v}}_r}{\sigma_r}$ are image of $\mathbf{A}$ with

$$\frac{\mathbf{A}\vec{\mathbf{v}}_i}{\sigma_i} \cdot \frac{\mathbf{A}\vec{\mathbf{v}}_j}{\sigma_j} = \frac{\vec{\mathbf{v}}_i \cdot \mathbf{A}^{\mathbf{T}}\mathbf{A}\vec{\mathbf{v}}_j}{\sigma_i\sigma_j} = \frac{\vec{\mathbf{v}}_i \cdot \lambda_j \vec{\mathbf{v}}_i}{\sigma_i\sigma_j} = 0.$$

Procedure to find the SVD for an $n \times m$ matrix A.

1. Compute $\mathbf{A}^{\mathbf{T}}\mathbf{A}$ and find orthogonal eigenbasis $\vec{\mathbf{v}}_1, \cdots, \vec{\mathbf{v}}_m$ such that the eigenvalues satisfy $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$.

$$\mathbf{V} = \begin{bmatrix} | & & | \\ \vec{\mathbf{v}}_1 & \cdots & \vec{\mathbf{v}}_m \\ | & & | \end{bmatrix}; \qquad \mathbf{\Sigma} = \begin{bmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_m} \end{bmatrix}.$$

2. Define
$$\mathbf{U} = \begin{bmatrix} \dfrac{\mathbf{A}\vec{\mathbf{v}}_1}{\sigma_1} & \dfrac{\mathbf{A}\vec{\mathbf{v}}_2}{\sigma_2} & \cdots & \dfrac{\mathbf{A}\vec{\mathbf{v}}_r}{\sigma_r} & \vec{\mathbf{x}}_1 & \cdots & \vec{\mathbf{x}}_k \end{bmatrix},$$
where $\vec{\mathbf{x}}_1, \cdots, \vec{\mathbf{x}}_k$ are choices of orthogonal basis of $\mathrm{Im}(\mathbf{A})^\perp$.

3. $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathbf{T}$.

Example 8.3.2. Compute the SVD of $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}$.

Solution.

1. Compute $\mathbf{A}^\mathbf{T}\mathbf{A}$, and find orthogonal eigenbasis:

$$\mathbf{A}^\mathbf{T}\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \implies f_\mathbf{A}(t) = (2-t)^2 - 1 = (t-1)(t-3).$$

Therefore, $\lambda_1 = 3, \quad \sigma_1 = \sqrt{3}, \quad \lambda_2 = 1, \quad \sigma_2 = \sqrt{1} = 1.$

$$E_{\mathbf{A}^\mathbf{T}\mathbf{A},3} = \mathrm{Span}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) \implies \vec{\mathbf{v}}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$$E_{\mathbf{A}^\mathbf{T}\mathbf{A},1} = \mathrm{Span}\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) \implies \vec{\mathbf{v}}_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

$$\therefore \mathbf{V}^\mathbf{T} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}; \qquad \mathbf{\Sigma} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

2. Find $\mathbf{U} = \begin{bmatrix} \dfrac{\mathbf{A}\vec{\mathbf{v}}_1}{\sigma_1} & \dfrac{\mathbf{A}\vec{\mathbf{v}}_2}{\sigma_2} & \vec{\mathbf{x}}_1 \end{bmatrix}$

$$\frac{\mathbf{A}\vec{\mathbf{v}}_1}{\sigma_1} = \frac{1}{\sqrt{3}} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$$\frac{\mathbf{A}\vec{\mathbf{v}}_2}{\sigma_2} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

$$\ker(\mathbf{A}^\mathbf{T}) = \ker\left(\begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}\right) = \mathrm{Span}\left(\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}\right) \implies \vec{\mathbf{x}}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

$$\therefore \mathbf{U} = \begin{bmatrix} 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ 2/\sqrt{6} & 0 & -1\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix}$$

3. $\mathbf{A} = \mathbf{U\Sigma V^T}$

$$\mathbf{A} = \begin{bmatrix} 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ 2/\sqrt{6} & 0 & -1\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \frac{1}{\sqrt{2}}$$

$\square$