Math for ML Notes

Ahmed Yasser

September 2025

Chapter 2 Linear Algebra

2.6.2 Rank

Definition (Rank)

The number of **linearly independent** columns of a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ equals the number of **linearly independent** rows of $\boldsymbol{A}$ and is called the **rank** of $\boldsymbol{A}$, and is denoted by $\mathrm{rk}(\boldsymbol{A})$

Remark (Matrix Rank Properties)

- **Rank equality**: For any matrix $\boldsymbol{A}$, the column rank equals the row rank: $\mathrm{rk}(\boldsymbol{A}) = \mathrm{rk}(\boldsymbol{A}^\top)$
- **Column space (image/range)**: The columns of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ span a subspace $U \subseteq \mathbb{R}^m$ with $\dim(U) = \mathrm{rk}(\boldsymbol{A})$. A basis for U can be found using Gaussian elimination to identify pivot columns.
- **Row space**: The rows of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ span a subspace $W \subseteq \mathbb{R}^n$ with $\dim(W) = \mathrm{rk}(\boldsymbol{A})$. A basis for W can be found by applying Gaussian elimination to $\boldsymbol{A}^\top$.
- **Invertibility condition**: For square matrices $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{A}$ is regular (invertible) if and only if $\mathrm{rk}(\boldsymbol{A}) = n$.
- **Linear system solvability**: For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$, the system $\boldsymbol{Ax} = \boldsymbol{b}$ can be solved if and only if $\mathrm{rk}(\boldsymbol{A}) = \mathrm{rk}(\boldsymbol{A}|\boldsymbol{b})$, where $\boldsymbol{A}|\boldsymbol{b}$ is the augmented matrix.
- **Null space (kernel)**: For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, the solution space of $\boldsymbol{Ax} = \boldsymbol{0}$ has dimension $n - \mathrm{rk}(\boldsymbol{A})$. This subspace is called the kernel or null space.
- **Full rank**: A matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ has full rank when $\mathrm{rk}(\boldsymbol{A}) = \min(m, n)$, meaning its rank equals the maximum possible rank for its dimensions.
- **Rank deficient**: A matrix that does not have full rank is called rank deficient.

2.7 Linear Mappings

Definition (Linear Mapping)

For vector spaces V and W, a mapping $\Phi : V \to W$ is called a **linear mapping (or linear transformation/vector space homomorphism)** if

$$\Phi(\boldsymbol{x} + \boldsymbol{y}) = \Phi(\boldsymbol{x}) + \Phi(\boldsymbol{y})$$
$$\Phi(\lambda \boldsymbol{x}) = \lambda \Phi(\boldsymbol{x})$$

$\forall \boldsymbol{x}, \boldsymbol{y} \in V$ and $\lambda \in \mathbb{R}$

Remark

Consider a mapping $\Phi : \mathcal{V} \to \mathcal{W}$, where $\mathcal{V}$ and $\mathcal{W}$ can be arbitrary sets. Then Φ is called:

- **Injective (one-to-one)** if $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V} : \Phi(\boldsymbol{x}) = \Phi(\boldsymbol{y}) \implies \boldsymbol{x} = \boldsymbol{y}$ i.e. there is no two different elements in $\mathcal{V}$ that map to the same element in $\mathcal{W}$.
- **Surjective (onto)** if $\Phi(\mathcal{V}) = \mathcal{W}$ i.e. every element in $\mathcal{W}$ can be reached from $\mathcal{V}$ using Φ.
- **Bijective** if Φ is both **injective** and **surjective**.

Remark

A bijective mapping $\Phi : \mathcal{V} \to \mathcal{W}$ is reversible: there exists a mapping $\Psi : \mathcal{W} \to \mathcal{V}$ such that $\Psi \circ \Phi(x) = x$ and $\Phi \circ \Psi(y) = y$. This mapping Ψ is the **inverse** of Φ, denoted Φ^{-1}.

Remark Special cases of linear mappings between vector spaces

- **Isomorphism**: $\Phi : V \to W$ linear and bijective (maps between different spaces, reversible)
- **Endomorphism**: $\Phi : V \to V$ linear (maps a space to itself)
- **Automorphism**: $\Phi : V \to V$ linear and bijective (maps a space to itself, reversible)
- **Identity mapping**: $\mathrm{id}_V : V \to V$, $x \mapsto x$ (leaves every vector unchanged)

Remark

Finite dimensional vector spaces V and W are **isomorphic** *if and only if* $\dim(V) = \dim(W)$

Remark

Consider vector spaces V, W, X. Then:

- If $\Phi : V \to W$ and $\Psi : W \to X$ are **linear** then $\Psi \circ \Phi : V \to X$ is **linear**.
- If $\Phi : V \to W$ is an **isomorphism** then $\Phi^{-1} : W \to V$ is an **isomorphism**.
- If $\Phi : V \to W$ and $\Psi : V \to W$ are **linear** then $\Phi + \Psi : V \to W$ and $\lambda\Phi : V \to W, \lambda \in \mathbb{R}$ are **linear**.

2.7.1 Matrix Representation of Linear Mappings

Remark (Notaion)

- $B = \{b_1, ..., b_n\}$ is an **unordered** basis
- $B = (b_1, ..., b_n)$ is an **ordered** basis
- $\boldsymbol{B} = \begin{bmatrix} b_1 & ... & b_n \end{bmatrix}$ is a **matrix** whose columns are the vectors $b_1, ..., b_n$

Definition (Coordinates)

Consider a vector space V and an **ordered** basis $B = (b_1, ..., b_n)$ of V. For any vector $x \in V$ we obtain a **unique** representaion (linear combination)

$$\boldsymbol{x} = \alpha_1 b_1 + ... + \alpha_n b_n$$

of x with respect to B. Then $\alpha_1, ..., \alpha_n$ are the *coordinates* of x with respect to B, and the vector

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^n$$

is the *coordinate vector/coordinate representaion* of x with respect to the ordered basis B.

Remark

A basis effectively defines a coordinate system and any basis of the vector space defines a valid coordinate system. The coordinates of a vector may be different between different basis.

Remark

For an n-dimensional vector space V and an ordered basis B of V, the mapping $\Phi : \mathbb{R}^n \to V, \Phi(e_i) = b_i,\ i = 1, ..., n,$ is **linear** and **bijective** (since V and $\mathbb{R}^n$ are of the same dimension), where $(\boldsymbol{e_i}, ..., \boldsymbol{e_n})$ is the **standard basis** of $\mathbb{R}^n$.

Definition (Transformation Matrix)

Consider vector spaces V and W with corresponding **ordered** bases $B = (\boldsymbol{b_1}, ..., \boldsymbol{b_n})$ and $C = (\boldsymbol{c_1}, ..., \boldsymbol{c_m})$. Also condier a **linear mapping** $\Phi : V \to W$. For $j \in \{i, ..., n\}$

$$\Phi\left(\boldsymbol{b_j}\right) = \alpha_{1j}\boldsymbol{c_1} + \alpha_{2j}\boldsymbol{c_2} + \ldots + \alpha_{mj}\boldsymbol{c_j} = \sum_{i=1}^{m} \alpha_{ij}\boldsymbol{c_i}$$

is the unique representaion (linear combination) of $\Phi\left(\boldsymbol{b_j}\right)$ with respect to the C. Then we call the $m \times n$ matrix $\boldsymbol{A}_\Phi$, whose elements are given by

$$A_{\Phi(i,j)} = \alpha_{ij},$$

The *transformation matrix* of Φ with respect to the ordered bases B of V and C of W.

Remark

From the definition of the transformation matrix we can see that the coordinates of $\Phi\left(\boldsymbol{b_j}\right)$ with respct to the ordered basis C of W are the j-th column of $\boldsymbol{A}_\Phi$

Corollary

Consider *finite dimensional* vector spaces V, W with ordered basis B, C and a linear mapping $\Phi : V \to W$ with transformation matrix $\boldsymbol{A}_\Phi$. If $\hat{\boldsymbol{x}}$ is the **coordinate vector** of $\boldsymbol{x} \in V$ with respect to B and $\hat{\boldsymbol{y}}$ is the **coordinate vector** of $\boldsymbol{y} = \Phi(\boldsymbol{x}) \in W$ with respect to C, then

$$\hat{\boldsymbol{y}} = \boldsymbol{A}_\Phi \hat{\boldsymbol{x}}.$$

This means that the transformation matrix can be used to map coordinates with respect to an ordered basis in V to coordinates with respect to an ordered basis in W.

2.7.2 Basis Change

Theorem (Basis Change)

Let $\Phi : V \to W$ be a linear mapping between vector spaces with ordered bases

$$B = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n) \quad \text{and} \quad \tilde{B} = \left(\tilde{\boldsymbol{b}}_1, \ldots, \tilde{\boldsymbol{b}}_n\right)$$

of V, and

$$C = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m) \quad \text{and} \quad \tilde{C} = (\tilde{\boldsymbol{c}}_1, \ldots, \tilde{\boldsymbol{c}}_m)$$

of W.

If $\boldsymbol{A}_\Phi$ is the transformation matrix of Φ with respect to bases B and C, then the transformation matrix $\tilde{\boldsymbol{A}}_\Phi$ with respect to bases $\tilde{B}$ and $\tilde{C}$ is given by:

$$\tilde{\boldsymbol{A}}_\Phi = \boldsymbol{T}^{-1}\boldsymbol{A}_\Phi\boldsymbol{S}$$

where:

- $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ is the transformation matrix of Id_V that maps coordinates with respect to $\tilde{B}$ onto coordinates with respect to B in V
- $\boldsymbol{T} \in \mathbb{R}^{m \times m}$ is the transformation matrix of Id_W that maps coordinates with respect to $\tilde{C}$ onto coordinates with respect to C in W

Definition (Equivalence)

Two matrices $\boldsymbol{A}$ and $\tilde{\boldsymbol{A}} \in \mathbb{R}^{m \times n}$ are **equivalent** if there exists **regular** matrices $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{T} \in \mathbb{R}^{m \times m}$, such that $\tilde{\boldsymbol{A}} = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{S}$

in other words, two matrices $\boldsymbol{A}$ and $\tilde{\boldsymbol{A}}$ are **equivalent** if they represent the **same linear transformation** $\Phi : V \to W$.

Definition (Similarity)

Two matrices $\boldsymbol{A}$ and $\tilde{\boldsymbol{A}} \in \mathbb{R}^{n\times n}$ are **similar** if there exists a **regular** matrix $\boldsymbol{S} \in \mathbb{R}^{n\times n}$, such that $\tilde{\boldsymbol{A}} = \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$

in other words, two matrices $\boldsymbol{A}$ and $\tilde{\boldsymbol{A}}$ are **similar** if they represent the **same linear transformation** $\Phi : V \to V$.

Remark

Similar matrices are always **equivalent**. However, **equivalent** matrices are not necessary **similar**.

Remark

The composition of two linear transformations, $\Phi : V \to W$ and $\Psi : W \to X$, results in another linear transformation $\Psi \circ \Phi : V \to X$. The matrix representing this combined transformation, $\boldsymbol{A}_{\Psi\circ\Phi}$, is the product of the individual transformation matrices: $\boldsymbol{A}_{\Psi\circ\Phi} = \boldsymbol{A}_{\Psi}\boldsymbol{A}_{\Phi}$.

Definition (Kernel (Null Space))

For a linear transformation $\Phi : V \to W$, the **kernel** is the set of all vectors $\boldsymbol{v} \in V$ that are mapped to the zero vector of W.

$$\ker(\Phi) := \Phi^{-1}(\mathbf{0}_W) = \{\boldsymbol{v} \in V \mid \Phi(\boldsymbol{v}) = \mathbf{0}_W\}$$

Definition (Image (Range))

For a linear transformation $\Phi : V \to W$, the **image** is the set of all vectors $\boldsymbol{w} \in W$ that can be "reached" by the transformation from some vector in V.

$$\mathrm{Im}(\Phi) := \Phi(V) = \{\boldsymbol{w} \in W \mid \exists \boldsymbol{v} \in V, \Phi(\boldsymbol{v}) = \boldsymbol{w}\}$$

Intuitively, the **kernel** is the set of vector $\boldsymbol{v} \in V$ that Φ maps onto the zero vector $\mathbf{0}_W \in W$. The **image** is the set of all vectors $\boldsymbol{w} \in W$ that can be *reached* by Φ from any vector in V.

Remark

For $\Phi : V \to W$, we call V the **domain** and W the **codomain**

The kernel is the set of vectors $v \in V$ that Φ maps onto the *neutral element* $\mathbf{0}_W \in W$

Remark

Consider a linear mapping $\Phi : V \to W$, where V and W are vector spaces:

- It always holds that $\Phi(\mathbf{0}_V) = \mathbf{0}_W$, therefore, $\mathbf{0}_V \in \ker(\Phi)$. In particular, the null spaces is never empty.
- $\mathrm{Im}(\Phi) \subseteq W$ is a **subspace** of W, and $\ker(\Phi) \subseteq V$ is a **subspace** of V.

Theorem (Rank-Nullity Theorem or Fundamental Theorem of Linear Mappings)

For vector spaces V and W and linear mapping $\Phi : V \to W$ it holds that

$$\dim(\ker(\Phi)) + \dim(\mathrm{Im}(\Phi)) = \dim(V)$$

Direct Consequences of the Rank-Nullity Theorem:

- If $\dim(\mathrm{im}(\Phi)) < \dim(V)$, then the $\ker(\Phi)$ is **non-trivial**, i.e., the kernel contains more than $\mathbf{0}_V$ and $\dim(\ker(\Phi)) \geq 1$.
- If $\boldsymbol{A}_\Phi$ is the transformation matrix of Φ with respect to an ordered basis and $\dim(\mathrm{Im}(\Phi)) < \dim(V)$, then the SLE $\boldsymbol{A}_\Phi\boldsymbol{x} = \mathbf{0}$ has **infinitely many solutions**.

2.8.1 Affine Subspaces

Definition (Affine Subspace)

let V be a vector space, $\boldsymbol{x_0} \in V$ and $U \subseteq V$ a subspace. Then the subset

$$\begin{aligned} L = \boldsymbol{x_0} + U &:= \{\boldsymbol{x_0} + \boldsymbol{u} : u \in U\} \\ &= \{\boldsymbol{v} \in V \mid \exists \boldsymbol{u} \in U : \boldsymbol{v} = \boldsymbol{x_0} + \boldsymbol{u}\} \subseteq V \end{aligned}$$

is called **affine subspace** or **linear manifold** of V. U is called **direction** or **direction space**, and $\boldsymbol{x_0}$ is called **support point**.

Remark

The definition of an affine subspace *excludes* $\mathbf{0}$ if $\boldsymbol{x_0} \notin U$. Therefore, an affine subspace is not a **linear** subspace (vector subspace) of V for $\boldsymbol{x}_0 \notin U$.

Remark

Consider two affine subspaces $L = \boldsymbol{x_0} + U$ and $\tilde{L} = \tilde{\boldsymbol{x}}_0 + \tilde{U}$ of a vector space V. Then, $L \subseteq \tilde{L}$ if and only if $U \subseteq \tilde{U}$ and $\boldsymbol{x_0} \in \tilde{L}$

Remark

Affine subspaces are often described by parameters: Consider a k-dimensional affine space $L = \boldsymbol{x_0} + U$ of V. if $(\boldsymbol{b_1}, ..., \boldsymbol{b_k})$ is an ordered basis of U, then every element $\boldsymbol{x} \in L$ can be uniquely written as

$$\boldsymbol{x} = \boldsymbol{x_0} + \lambda_1 \boldsymbol{b_1} + ... + \lambda_k \boldsymbol{b_k}$$

where $\lambda_1, ..., \lambda_k \in \mathbb{R}$. Ths representaion is called the **parametric equation** of L with **directional vectors** $\boldsymbol{b_1}, ..., \boldsymbol{b_k}$ and **parameters** $\lambda_1, ..., \lambda_k$.

Remark

In $\mathbb{R}^n$, the $(n-1)$-dimensional affine subspaces are called **hyperplanes**, and the corresponding parametric equation is:

$$\boldsymbol{y} = \boldsymbol{x}_0 + \sum_{i=1}^{n-1} \lambda_i \boldsymbol{b}_i$$

where $\boldsymbol{b}_1, ..., \boldsymbol{b}_{n-1}$ form a basis of an $(n-1)$-dimensional subspace U of $\mathbb{R}^n$.

This means that a hyperplane is defined by:

- A **support point** $\boldsymbol{x}_0$
- $(n-1)$ linearly independent vectors $\boldsymbol{b}_1, ..., \boldsymbol{b}_{n-1}$ that span the **direction space**

Examples:

- In $\mathbb{R}^2$, a line is a hyperplane
- In $\mathbb{R}^3$, a plane is a hyperplane

Remark (Inhomogeneous systems of linear equations and affine subspaces)

For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$, the solution of the system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is either the **empty set** or an **affine subspace** of $\mathbb{R}^n$ of dimension $n - \mathrm{rk}(\boldsymbol{A})$.

In $\mathbb{R}^n$, every k-dimensional affine subspace is the solution of an inhomogeneous system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$ and $\mathrm{rk}(\boldsymbol{A}) = n - k$.

Note: Recall that for homogeneous equation systems $\boldsymbol{Ax} = \mathbf{0}$ the solution was a vector subspace, which we can also think of as a special affine space with support point $\boldsymbol{x}_0 = \mathbf{0}$.

2.8.2 Affine Mappings

Definition (Affine Mapping)

For two vector spaces V, W, a linear mapping $\Phi : V \to W$ and $\boldsymbol{a} \in W$, the mapping

$$\begin{aligned} \phi : V &\to W \\ \boldsymbol{x} &\mapsto \boldsymbol{a} + \Phi(\boldsymbol{x}) \end{aligned}$$

is an **affine mapping** from V to W. The vector $\boldsymbol{a}$ is called the **translation vector** of ϕ

Remark

- Every affine mapping $\phi : V \to W$ is also the composition of a linear mapping $\Phi : V \to W$ and a translation $\tau : W \to W, \boldsymbol{x} \mapsto \boldsymbol{a} + \boldsymbol{x}$ in W, such that $\phi = \tau \circ \Phi$. The mappings Φ and τ are **uniquely determined**.
- The composition $\acute{\phi} \circ \phi$ of affine mappings $\phi : V \to W$, $\acute{\phi} : W \to X$ is **affine**.
- If ϕ is bijective, affine mappings keep the geometric structure invariant. They then also preserve the dimension and parallelism.

Chapter 3 Analytic Geometry

3.1 Norms

Definition (Norm)

A **norm** on a vector spacce V is a function

$$\begin{aligned} \|\cdot\| : V &\to \mathbb{R}, \\ \boldsymbol{x} &\mapsto \|\boldsymbol{x}\|, \end{aligned}$$

which assigns each vector $\boldsymbol{x}$ its **length** $\|\boldsymbol{x}\| \in \mathbb{R}$, such that for all $\lambda \in \mathbb{R}$ and $\boldsymbol{x}, \boldsymbol{y} \in V$ the following hold:

- **Absolutely homogeneous**: $\|\lambda \boldsymbol{x}\| = |\lambda| \, \|\boldsymbol{x}\|$
  - If you scale a vector $\boldsymbol{x}$ by a number (a scalar) λ, its length is scaled by the absolute value of that number.
- **Triangle inequality**: $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$
  - The length of the sum of two vectors is less than or equal to the sum of their individual lengths.
  - The only time they would be equal is if the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ point in the **same direction**.
- **Positive definite**: $\|\boldsymbol{x}\| \geq 0$ and $\|\boldsymbol{x}\| = 0 \iff \boldsymbol{x} = \mathbf{0}$
  - The length of a vector is always non-negative.
  - A vector has zero length if and only if it is the zero vector.

Remark

In geometric terms, the triangle inequality states that for any triangle, the sum of the lengths of any two sides must be greater than or equal to the length of the remaining side.

Definition (Manhattan Norm or ℓ_1 norm)

The **Manhattan norm** on $\mathbb{R}^n$ is defined for $\boldsymbol{x} \in \mathbb{R}^n$ as

$$\|\boldsymbol{x}\|_1 := \sum_{i=1}^{n} |x_i|$$

where $|x_i|$ is the absolute value of the i-th component of $\boldsymbol{x}$.

Definition (Euclidean Norm or ℓ_2 norm)

The **Euclidean norm** on $\mathbb{R}^n$ is defined for $\boldsymbol{x} \in \mathbb{R}^n$ as

$$\|\boldsymbol{x}\|_2 := \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}}$$

and computes the **Euclidean distancce** of $\boldsymbol{x}$ from the origin.

3.2.1 Dot Product

Definition (Dot Product)

The **dot product** (also called **scalar product**) is a **particular inner product** in $\mathbb{R}^n$ which is given by

$$\boldsymbol{x}^\top \boldsymbol{y} = \sum_{i=1}^{n} x_i y_i$$

3.2.2 General Inner Products

Definition (Bilinear Mapping)

A **bilinear mapping** Ω is a mapping with two arguments and it is **linear in each argument separately**,i.e., when we look at a vector spae V then it holds that for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V, \ \ \lambda, \psi \in \mathbb{R}$ the following hold:

- $\Omega(\lambda\boldsymbol{x} + \psi\boldsymbol{y}, \boldsymbol{z}) = \lambda\Omega(\boldsymbol{x}, \boldsymbol{z}) + \psi\Omega(\boldsymbol{y}, \boldsymbol{z})$ (*Linearity in first argument*)
- $\Omega(\boldsymbol{x}, \lambda\boldsymbol{y} + \psi\boldsymbol{z}) = \lambda\Omega(\boldsymbol{x}, \boldsymbol{y}) + \psi\Omega(\boldsymbol{x}, \boldsymbol{z})$ (*Linearity in second argument*)

Definition (Symmetric and Positive Definite Bilinear Mapping)

Let V be a vector space and $\Omega : V \times V \to \mathbb{R}$ be a bilinear mapping that takes two vectors in V and maps them onto a real number. Then

- Ω is called **symmetric** if $\Omega(\boldsymbol{x}, \boldsymbol{y}) = \Omega(\boldsymbol{y}, \boldsymbol{x})$ for all $\boldsymbol{x}, \boldsymbol{y} \in V$, i.e., the **order of the arguments does not matter**.
- Ω is called **positive definite** if $\forall \boldsymbol{x} \in V \setminus \{\boldsymbol{0}\} : \Omega(\boldsymbol{x}, \boldsymbol{x}) > 0, \quad \Omega(\boldsymbol{0}, \boldsymbol{0}) = 0$

Definition (Inner Product Space)

Let V be a vector space and $\Omega : V \times V \to \mathbb{R}$ be a bilinear mapping that takes two vectors in V and maps them onto a real number. Then

- A **positive definite, symmetric bilinear mapping** $\Omega : V \times V \to \mathbb{R}$ is called an **inner product** on V. We typically write $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ instead of $\Omega(\boldsymbol{x}, \boldsymbol{y})$
- The pair $(V, \langle \cdot, \cdot \rangle)$ is called an **inner product space** or **(real) vector space with inner product**. If we use the **dot product**, we call $(V, \langle \cdot, \cdot \rangle)$ a **Euclidean vector space**.

3.2.3 Symmetric, Positive Definite Matrices

Remark (Gram Matrix and Inner Product)

Consider an n-dimensional vector space V with an inner product $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ and an ordered basis $B = (\boldsymbol{b}_1, ..., \boldsymbol{b}_n)$ of V. Due to the bilinearity of the inner product, it holds for all $\boldsymbol{x}, \boldsymbol{y} \in V$ that

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \left\langle \sum_{i=1}^{n} \psi_i \boldsymbol{b}_i, \sum_{j=1}^{n} \lambda_j \boldsymbol{b}_j \right\rangle = \sum_{i=1}^{n} \psi_i \left(\sum_{j=1}^{n} \lambda_j \langle \boldsymbol{b}_i, \boldsymbol{b}_j \rangle \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_i \langle \boldsymbol{b}_i, \boldsymbol{b}_j \rangle \lambda_j = \hat{\boldsymbol{x}}^\top \boldsymbol{A} \hat{\boldsymbol{y}}$$

where $A_{ij} := \langle \boldsymbol{b}_i, \boldsymbol{b}_j \rangle$ and $\hat{\boldsymbol{x}} := (\psi_1, ..., \psi_n)^\top$, $\hat{\boldsymbol{y}} := (\lambda_1, ..., \lambda_n)^\top$ are the coordinates of $\boldsymbol{x}$ and $\boldsymbol{y}$ with respect to the basis B.

Note that:

- This implies that the inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ is **uniquely determined** through $\boldsymbol{A}$.
- The **symmetry** of the inner product also means that $\boldsymbol{A}$ is **symmetric**.
- The **positive definiteness** of the inner product imples that $\forall \boldsymbol{x} \in V \setminus \{\boldsymbol{0}\} : \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} > 0$.

Definition (Symmetric Positive Definite Matrix)

A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is classified as:

- **Positive definite**: if $\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\} : \boldsymbol{x}^\top A \boldsymbol{x} > 0$
- **Positive semidefinite**: if $\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\} : \boldsymbol{x}^\top A \boldsymbol{x} \geq 0$

Theorem (Matrix Representation of Inner Product)

For a real-valued, finite-dimensional vector space V and an ordered basis B of V, it holds that $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ is an **inner product** *if and only if* there exists a **symmetric, positive definite matrix** $A \in \mathbb{R}^{n \times n}$ with

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \hat{\boldsymbol{x}}^\top A \hat{\boldsymbol{y}}$$

Remark

The following properties hold if $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is **symmetric positive definite**:

- The null space (kernel) of $\boldsymbol{A}$ consists only of $\boldsymbol{0}$ because $\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}$. This implies that $\boldsymbol{A}\boldsymbol{x} \neq \boldsymbol{0}$ if $\boldsymbol{x} \neq \boldsymbol{0}$.
- The **diagonal elements** a_{ii} of $\boldsymbol{A}$ are positive because $a_{ii} = \boldsymbol{e_i}^\top \boldsymbol{A} \boldsymbol{e_i} > 0$, where $\boldsymbol{e_i}$ is the i-th vector of the standard basis of $\mathbb{R}^n$.

3.3 Lengths and Distances

Remark

Any inner product induces a norm

$$\|\boldsymbol{x}\| := \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$$

However, not every norm is induced by an inner product. The **Manhattan norm** is an example of a norm without a corresponding inner product.

Remark (Cauchy-Schwarz Inequality)

For an inner product vector space $(V, \langle\cdot,\cdot\rangle)$, the induced norm $\|\cdot\|$ satisfies the *Cauchy-Schwarz inequality*: $|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|$

Definition (Distance and Metric)

Consider an inner product space $(V, \langle \cdot, \cdot \rangle)$. Then

$$d(\boldsymbol{x}, \boldsymbol{y}) := \|\boldsymbol{x} - \boldsymbol{y}\| = \sqrt{\langle \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{y} \rangle}$$

is called the **distance** between $\boldsymbol{x}$ and $\boldsymbol{y}$ for $\boldsymbol{x}, \boldsymbol{y} \in V$. If we use the **dot product** as the inner product, then the distance is called the **Euclidean distance**.

The mapping

$$\begin{aligned} d &: V \times V \to \mathbb{R}, \\ &(\boldsymbol{x}, \boldsymbol{y}) \mapsto d(\boldsymbol{x} - \boldsymbol{y}) \end{aligned}$$

is called a **metric**

Remark

Similar to the length of a vector, the distance between vectors does not require an inner product: a norm is sufficient

Remark

A metric $d : V \times V \to \mathbb{R}$ satisfies the following properties:

- d is **positive definite**, i.e., $d(\boldsymbol{x}, \boldsymbol{y}) \geq 0$ for all $\boldsymbol{x}, \boldsymbol{y} \in V$ and $d(\boldsymbol{x}, \boldsymbol{y}) = 0 \iff \boldsymbol{x} = \boldsymbol{y}$.
- d is **symmetric**, i.e., $d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})$ for all $\boldsymbol{x}, \boldsymbol{y} \in V$.
- **Triangle inequality**: $d(\boldsymbol{x}, \boldsymbol{z}) \leq d(\boldsymbol{x}, \boldsymbol{y}) + d(\boldsymbol{y}, \boldsymbol{z})$ for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$.

Remark

Inner products and metrics have **opposing behaviors** despite similar property lists:

- **Similar** x and y $\to$ **large** inner product value
- **Similar** x and y $\to$ **small** metric value

3.4 Angles and Orthogonality

Remark (Angle between Vectors)

We use the **Cauchy-Schwarz inequality** to define the angle ω in inner product spaces between two vectors $\boldsymbol{x}, \boldsymbol{y}$. Assume that $\boldsymbol{x} \neq \boldsymbol{0}$ and $\boldsymbol{y} \neq \boldsymbol{0}$. Then

$$-1 \leq \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|} \leq 1$$

Therefore, there exits a unique angle $\omega \in [0, \pi]$ between the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, with

$$\cos(\omega) = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|}$$

Note that:

- We restict ω to $[0, \pi]$ so that $\cos(\omega)$ returns a unique number in the interval $[-1, 1]$.
- The angle between two vectors tells us **how similar** their **orientations** are.

Definition (Orthogonal Vectors)

Two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are **orthogonal** *if and only if* $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$, and we write $\boldsymbol{x} \perp \boldsymbol{y}$.

Definition (Orthonormal Vectors)

Two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are **orthonormal** *if and only if* $\boldsymbol{x} \perp \boldsymbol{y}$ and $\|\boldsymbol{x}\| = \|\boldsymbol{y}\| = 1$ (i.e., the vectors are **unit** vectors).

Remark

The **0**-vector is **orthogonal** to every vector in the vector space, because $\langle \mathbf{0}, \boldsymbol{x} \rangle = 0$ for all $\boldsymbol{x} \in V$.

Remark

Orthogonality is the generalization of the concept of **perpendicularity** to bilinear forms that do not have to be the dot product.

Remark

Vectors that are orthogonal with respect to one inner product **do not have to be orthogonal** with respect to a different inner product.

Definition (Orthogonal Matrix)

A square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is an **orthogonal matrix** *if and only if*

$$\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I} = \boldsymbol{A}^\top \boldsymbol{A},$$

which implies that

$$\boldsymbol{A}^{-1} = \boldsymbol{A}^\top$$

i.e., the inverse is obtained by simply transposing the matrix.

Note that, it is convention to call these matrices **orthogonal** but a more precise description would be **orthonormal**.

Remark (Orthogonal Matrices)

Transformations by orthogonal matrices are special because the length of a vector $\boldsymbol{x}$ is not changed when transforming it using an orthogonal matrix $\boldsymbol{A}$. Moreover, the angle between any two vectors $\boldsymbol{x}, \boldsymbol{y}$, as measured by their inner product, is also unchanged when transforming them using $\boldsymbol{A}$.

Assume the dot product is the inner product:

- **length preservation**:

$$\|\boldsymbol{A}\boldsymbol{x}\|_2 = (\boldsymbol{A}\boldsymbol{x})^\top (\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^\top \boldsymbol{I}\boldsymbol{x} = \boldsymbol{x}^\top \boldsymbol{x} = \|\boldsymbol{x}\|_2$$

- **angle preservation**:

$$\cos(\omega) = \frac{(\boldsymbol{A}\boldsymbol{x})^\top (\boldsymbol{A}\boldsymbol{y})}{\|\boldsymbol{A}\boldsymbol{x}\|_2 \|\boldsymbol{A}\boldsymbol{y}\|_2} = \frac{\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{y}}{\sqrt{(\boldsymbol{A}\boldsymbol{x})^\top \boldsymbol{A}\boldsymbol{x}}\sqrt{(\boldsymbol{A}\boldsymbol{y})^\top \boldsymbol{A}\boldsymbol{y}}} = \frac{\boldsymbol{x}^\top \boldsymbol{y}}{\sqrt{\boldsymbol{x}^\top \boldsymbol{x}}\sqrt{\boldsymbol{y}^\top \boldsymbol{y}}} = \frac{\boldsymbol{x}^\top \boldsymbol{y}}{\|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2}$$

Which means that the orthogonal matrices $\boldsymbol{A}$ with $\boldsymbol{A}^\top = \boldsymbol{A}^{-1}$ preserve both **angles** and **distances**.

Remark

Orthogonal matrices define transformations that are **rotations** (with the possibility of flips).