Mathematics for Machine Learning

Vazgen Mikayelyan

July 28, 2020

FAST DISCOVERING THE FUTURE

Positive Definite Matrices

Positive Definite Matrices

Definition

A matrix $A \in \mathbb{R}^{n\times n}$ *is called* **positive definite**, *if*

$$\mathbf{x}^T A\mathbf{x} > 0, \quad \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{0}$$

The matrix A *is called positive semi-definite, if*

$$\mathbf{x}^T A\mathbf{x} \geq 0, \quad \mathbf{x} \in \mathbb{R}^n.$$

Positive Definite Matrices

Definition

A matrix $A \in \mathbb{R}^{n \times n}$ *is called* **positive definite**, *if*

$$\mathbf{x}^T A \mathbf{x} > 0, \quad \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{0}$$

The matrix A *is called positive semi-definite, if*

$$\mathbf{x}^T A \mathbf{x} \geq 0, \quad \mathbf{x} \in \mathbb{R}^n.$$

Example

Which of the following matrices is positive definite?

$$A = \begin{bmatrix} 4 & -4 \\ -4 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & -4 \\ -4 & 2 \end{bmatrix}$$

Positive Definite Matrices

Proposition

Let V be an n-dimensional vector space with an inner product $\langle \cdot, \cdot \rangle$ and a basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$. Prove that

$$\langle \mathbf{x}, \mathbf{y} \rangle = [\mathbf{x}]_B^T A [\mathbf{y}]_B,$$

where $A_{ij} = \langle \mathbf{b}_i, \mathbf{b}_j \rangle$. Conclude that the matrix A is positive definite.

Positive Definite Matrices

Proposition

Let V be an n-dimensional vector space with an inner product $\langle \cdot, \cdot \rangle$ and a basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$. Prove that

$$\langle \mathbf{x}, \mathbf{y} \rangle = [\mathbf{x}]_B^T A [\mathbf{y}]_B,$$

where $A_{ij} = \langle \mathbf{b}_i, \mathbf{b}_j \rangle$. Conclude that the matrix A is positive definite.

Theorem

For a real-valued, finite-dimensional vector space V and a 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

$$\langle \mathbf{x}, \mathbf{y} \rangle = [\mathbf{x}]_B^T A [\mathbf{y}]_B,$$

Positive Definite Matrices

Proposition

A symmetric matrix A is positive definite if and only if the determinants associated with all upper-left submatrices of A are positive.

Distance and metric

Distance and metric

Any inner product induces a norm

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

However, not every norm is induced by an inner product (e.g. Manhattan norm or ℓ_1 norm).

Any inner product induces a norm

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

However, not every norm is induced by an inner product (e.g. Manhattan norm or ℓ_1 norm).

Example

Compute the norm of the vector $[2\ 2]^T$ *w.r.t. the following inner products:*

a) $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$

b) $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix} \mathbf{y}$

Distance and metric

.

Definition

A function $d : V \times V \to \mathbb{R}$ *is metric, if*

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

Distance and metric

.

Definition

A function $d : V \times V \to \mathbb{R}$ *is metric, if*

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

Proposition

Every norm (inner product) induces a metric:

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

Angles and Orthogonality

By Cauchy-Schwarz inequality

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

Angles and Orthogonality

By Cauchy-Schwarz inequality

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

Therefore, there exists a unique $w \in [0, \pi]$ with

$$\cos w = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|}.$$

The number w is the **angle** between the vectors $\mathbf{x}$ and $\mathbf{y}$.

Angles and Orthogonality

By Cauchy-Schwarz inequality

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

Therefore, there exists a unique $w \in [0, \pi]$ with

$$\cos w = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|}.$$

The number w is the **angle** between the vectors $\mathbf{x}$ and $\mathbf{y}$.

Example

Find the angle between $\mathbf{x}$ *and* $\pm\mathbf{x}$.

Angles and Orthogonality

Example

Find the angle between $\mathbf{x} = [1,\ 1]^T$ *and* $\mathbf{y} = [1, -1]^T$ *w.r.t. the inner products given by*

a) $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$

b) $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \mathbf{y}$

Angles and Orthogonality

Definition

Two vectors $\mathbf{x}$ *and* $\mathbf{y}$ *are* **orthogonal** *if and only if* $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, *and we write* $x \perp y$. *If additionally* $\|\mathbf{x}\| = 1 = \|\mathbf{y}\|$, *i.e., the vectors are unit vectors, then* $\mathbf{x}$ *and* $\mathbf{y}$ *are* **orthonormal**.

Angles and Orthogonality

Definition

Two vectors $\mathbf{x}$ *and* $\mathbf{y}$ *are* **orthogonal** *if and only if* $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, *and we write* $x \perp y$. *If additionally* $\|\mathbf{x}\| = 1 = \|\mathbf{y}\|$, *i.e., the vectors are unit vectors, then* $\mathbf{x}$ *and* $\mathbf{y}$ *are* **orthonormal**.

Remark

Geometrically, we can think of orthogonal vectors as having a right angle with respect to a specific inner product.

Angles and Orthogonality

Definition

Two vectors $\mathbf{x}$ *and* $\mathbf{y}$ *are* **orthogonal** *if and only if* $\langle \mathbf{x}, \mathbf{y} \rangle = 0$*, and we write* $x \perp y$*. If additionally* $\|\mathbf{x}\| = 1 = \|\mathbf{y}\|$*, i.e., the vectors are unit vectors, then* $\mathbf{x}$ *and* $\mathbf{y}$ *are* **orthonormal***.*

Remark

Geometrically, we can think of orthogonal vectors as having a right angle with respect to a specific inner product.

Definition

A square matrix $A \in \mathbb{R}^{n \times n}$ *is an orthogonal matrix if its columns are orthonormal, i.e.* $A^T A = I$*.*

Angles and Orthogonality

Definition

Two vectors $\mathbf{x}$ *and* $\mathbf{y}$ *are* **orthogonal** *if and only if* $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, *and we write* $x \perp y$. *If additionally* $\|\mathbf{x}\| = 1 = \|\mathbf{y}\|$, *i.e., the vectors are unit vectors, then* $\mathbf{x}$ *and* $\mathbf{y}$ *are* **orthonormal**.

Remark

Geometrically, we can think of orthogonal vectors as having a right angle with respect to a specific inner product.

Definition

A square matrix $A \in \mathbb{R}^{n \times n}$ *is an orthogonal matrix if its columns are orthonormal, i.e.* $A^T A = I$.

Note that for orthonormal matrix $AA^T = I$ and $A^{-1} = A^T$.

Theorem

Suppose S is an orthogonal set of nonzero vectors. Then S is linearly independent.

The Pythagorean Theorem

Theorem

The vectors $u, v \in \mathbb{R}^n$ *are orthogonal if and only if*

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

The Pythagorean Theorem

Theorem

The vectors $u, v \in \mathbb{R}^n$ *are orthogonal if and only if*

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

Theorem

Suppose $\{u_1, \ldots, u_n\}$ *is a set of orthogonal vectors, then*

$$\|u_1 + \ldots + u_n\|^2 = \|u_1\|^2 + \ldots + \|u_n\|^2.$$

Angles and Orthogonality

Proposition

Transformations with orthogonal matrix of transformation preserve the length and the angle w.r.t. to the dot product, i.e.

$$\|A\mathbf{x}\| = \|\mathbf{x}\| \text{ and } w_{\mathbf{x},\mathbf{y}} = w_{A\mathbf{x},A\mathbf{y}}$$

Angles and Orthogonality

Proposition

Transformations with orthogonal matrix of transformation preserve the length and the angle w.r.t. to the dot product, i.e.

$$\|A\mathbf{x}\| = \|\mathbf{x}\| \text{ and } w_{\mathbf{x},\mathbf{y}} = w_{A\mathbf{x},A\mathbf{y}}$$

Proof.

$$\|A\mathbf{x}\|^2 = A\mathbf{x} \cdot A\mathbf{x} =$$

Angles and Orthogonality

Proposition

Transformations with orthogonal matrix of transformation preserve the length and the angle w.r.t. to the dot product, i.e.

$$\|A\mathbf{x}\| = \|\mathbf{x}\| \text{ and } w_{\mathbf{x},\mathbf{y}} = w_{A\mathbf{x},A\mathbf{y}}$$

Proof.

$$\|A\mathbf{x}\|^2 = A\mathbf{x} \cdot A\mathbf{x} = (A\mathbf{x})^T A\mathbf{x} = \mathbf{x}^T A^T A\mathbf{x} =$$

Proposition

Transformations with orthogonal matrix of transformation preserve the length and the angle w.r.t. to the dot product, i.e.

$$\|A\mathbf{x}\| = \|\mathbf{x}\| \text{ and } w_{\mathbf{x},\mathbf{y}} = w_{A\mathbf{x},A\mathbf{y}}$$

Proof.

$$\|A\mathbf{x}\|^2 = A\mathbf{x} \cdot A\mathbf{x} = (A\mathbf{x})^T A\mathbf{x} = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T I\mathbf{x} =$$

Angles and Orthogonality

Proposition

Transformations with orthogonal matrix of transformation preserve the length and the angle w.r.t. to the dot product, i.e.

$$\|A\mathbf{x}\| = \|\mathbf{x}\| \text{ and } w_{\mathbf{x},\mathbf{y}} = w_{A\mathbf{x},A\mathbf{y}}$$

Proof.

$$\|A\mathbf{x}\|^2 = A\mathbf{x} \cdot A\mathbf{x} = (A\mathbf{x})^T A\mathbf{x} = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T I\mathbf{x} = \|\mathbf{x}\|^2$$

Proposition

Transformations with orthogonal matrix of transformation preserve the length and the angle w.r.t. to the dot product, i.e.

$$\|A\mathbf{x}\| = \|\mathbf{x}\| \text{ and } w_{\mathbf{x},\mathbf{y}} = w_{A\mathbf{x},A\mathbf{y}}$$

Proof.

$$\|A\mathbf{x}\|^2 = A\mathbf{x} \cdot A\mathbf{x} = (A\mathbf{x})^T A\mathbf{x} = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T I\mathbf{x} = \|\mathbf{x}\|^2$$

and

$$\cos w_{A\mathbf{x},A\mathbf{y}} = \frac{(A\mathbf{x})^T A\mathbf{y}}{\|A\mathbf{x}\|\|A\mathbf{y}\|} =$$

Angles and Orthogonality

Proposition

Transformations with orthogonal matrix of transformation preserve the length and the angle w.r.t. to the dot product, i.e.

$$\|A\mathbf{x}\| = \|\mathbf{x}\| \text{ and } w_{\mathbf{x},\mathbf{y}} = w_{A\mathbf{x},A\mathbf{y}}$$

Proof.

$$\|A\mathbf{x}\|^2 = A\mathbf{x} \cdot A\mathbf{x} = (A\mathbf{x})^T A\mathbf{x} = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T I\mathbf{x} = \|\mathbf{x}\|^2$$

and

$$\cos w_{A\mathbf{x},A\mathbf{y}} = \frac{(A\mathbf{x})^T A\mathbf{y}}{\|A\mathbf{x}\|\|A\mathbf{y}\|} = \frac{\mathbf{x}^T A^T A\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} =$$

Proposition

Transformations with orthogonal matrix of transformation preserve the length and the angle w.r.t. to the dot product, i.e.

$$\|A\mathbf{x}\| = \|\mathbf{x}\| \text{ and } w_{\mathbf{x},\mathbf{y}} = w_{A\mathbf{x},A\mathbf{y}}$$

Proof.

$$\|A\mathbf{x}\|^2 = A\mathbf{x} \cdot A\mathbf{x} = (A\mathbf{x})^T A\mathbf{x} = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T I\mathbf{x} = \|\mathbf{x}\|^2$$

and

$$\cos w_{A\mathbf{x},A\mathbf{y}} = \frac{(A\mathbf{x})^T A\mathbf{y}}{\|A\mathbf{x}\|\|A\mathbf{y}\|} = \frac{\mathbf{x}^T A^T A\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \cos w_{\mathbf{x},\mathbf{y}}.$$

□

Orthonormal Basis

Definition

Consider an n-dimensional vector space V and a basis $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ of V. The basis B is called an **orthogonal basis**, *if*

$$\langle \mathbf{b}_i, \mathbf{b}_j \rangle = 0, \;\; i \neq j.$$

If additionally the vectors $\mathbf{b}_i$ are unit, then the basis B is called an **orthonormal basis (ONB).**

Orthonormal Basis

Definition

Consider an n-dimensional vector space V and a basis $B = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ of V. The basis B is called an **orthogonal basis**, *if*

$$\langle \mathbf{b}_i, \mathbf{b}_j \rangle = 0, \ i \neq j.$$

If additionally the vectors $\mathbf{b}_i$ are unit, then the basis B is called an **orthonormal basis (ONB).**

Example

The canonical/standard basis for a Euclidean vector space R^n is an orthonormal basis, where the inner product is the dot product of vectors.

Orthonormal Basis

Example

Check that the vectors

$$\mathbf{b}_1 = \frac{1}{2}\begin{bmatrix} 1 \\ \sqrt{3} \end{bmatrix}, \quad \mathbf{b}_2 = \frac{1}{2}\begin{bmatrix} -\sqrt{3} \\ 1 \end{bmatrix}$$

form an orthonormal basis for $\mathbb{R}^2$.

Projection operators

Projection operators

Definition

Let V be a vector space and $W \subset V$ a subspace of V. A linear mapping $\pi : V \to W$ is called a **projection** *if $\pi^2 = \pi \circ \pi = \pi$.*

Projection operators

Definition

Let V be a vector space and $W \subset V$ a subspace of V . A linear mapping $\pi : V \to W$ is called a **projection** *if $\pi^2 = \pi \circ \pi = \pi$.*

Remark

The transformation matrix P_π of a projection π is called projection matrix and satisfies $P_\pi^2 = P_\pi$.

Orthogonal Projection onto 1-Dimensional Subspaces

Let $U \subset \mathbb{R}^n$ be a line passing through the origin parallel to a vector $\mathbf{b} \in \mathbb{R}^n$.

Orthogonal Projection onto 1-Dimensional Subspaces

Let $U \subset \mathbb{R}^n$ be a line passing through the origin parallel to a vector $\mathbf{b} \in \mathbb{R}^n$. Then

- $U = \operatorname{span}\{\mathbf{b}\}$,

Orthogonal Projection onto 1-Dimensional Subspaces

Let $U \subset \mathbb{R}^n$ be a line passing through the origin parallel to a vector $\mathbf{b} \in \mathbb{R}^n$. Then

- $U = \operatorname{span}\{\mathbf{b}\}$,
- $\pi_U(\mathbf{x}) \in U$, hence $\pi_U(\mathbf{x}) = \lambda\mathbf{b}$ for some $\lambda \in \mathbb{R}$,

Orthogonal Projection onto 1-Dimensional Subspaces

Let $U \subset \mathbb{R}^n$ be a line passing through the origin parallel to a vector $\mathbf{b} \in \mathbb{R}^n$. Then

- $U = \operatorname{span}\{\mathbf{b}\}$,
- $\pi_U(\mathbf{x}) \in U$, hence $\pi_U(\mathbf{x}) = \lambda \mathbf{b}$ for some $\lambda \in \mathbb{R}$,
- $\pi_U(\mathbf{x})$ is closest to $\mathbf{x}$ in U, and hence

$$\pi_U(\mathbf{x}) - \mathbf{x} \perp \mathbf{b} \Leftrightarrow \langle \pi_U(\mathbf{x}) - \mathbf{x}, \mathbf{b} \rangle = 0,$$

Orthogonal Projection onto 1-Dimensional Subspaces

Let $U \subset \mathbb{R}^n$ be a line passing through the origin parallel to a vector $\mathbf{b} \in \mathbb{R}^n$. Then

- $U = \text{span}\{\mathbf{b}\}$,
- $\pi_U(\mathbf{x}) \in U$, hence $\pi_U(\mathbf{x}) = \lambda \mathbf{b}$ for some $\lambda \in \mathbb{R}$,
- $\pi_U(\mathbf{x})$ is closest to $\mathbf{x}$ in U, and hence

$$\pi_U(\mathbf{x}) - \mathbf{x} \perp \mathbf{b} \Leftrightarrow \langle \pi_U(\mathbf{x}) - \mathbf{x}, \mathbf{b} \rangle = 0,$$

- $\pi_U(\mathbf{x}) = \lambda \mathbf{b} = \dfrac{\langle \mathbf{x}, \mathbf{b} \rangle}{\|\mathbf{b}\|^2} \mathbf{b}$,

Orthogonal Projection onto 1-Dimensional Subspaces

Let $U \subset \mathbb{R}^n$ be a line passing through the origin parallel to a vector $\mathbf{b} \in \mathbb{R}^n$. Then

- $U = \operatorname{span}\{\mathbf{b}\}$,
- $\pi_U(\mathbf{x}) \in U$, hence $\pi_U(\mathbf{x}) = \lambda \mathbf{b}$ for some $\lambda \in \mathbb{R}$,
- $\pi_U(\mathbf{x})$ is closest to $\mathbf{x}$ in U, and hence

$$\pi_U(\mathbf{x}) - \mathbf{x} \perp \mathbf{b} \Leftrightarrow \langle \pi_U(\mathbf{x}) - \mathbf{x}, \mathbf{b} \rangle = 0,$$

- $\pi_U(\mathbf{x}) = \lambda \mathbf{b} = \dfrac{\langle \mathbf{x}, \mathbf{b} \rangle}{\|\mathbf{b}\|^2} \mathbf{b}$,
- $\|\pi_U(\mathbf{x})\| = |\cos w| \|\mathbf{x}\|$, where w is the angle between $\mathbf{x}$ and $\mathbf{b}$.

Orthogonal Projection onto 1-Dimensional Subspaces

Let $U \subset \mathbb{R}^n$ be a line passing through the origin parallel to a vector $\mathbf{b} \in \mathbb{R}^n$. Then

- $U = \operatorname{span}\{\mathbf{b}\}$,
- $\pi_U(\mathbf{x}) \in U$, hence $\pi_U(\mathbf{x}) = \lambda\mathbf{b}$ for some $\lambda \in \mathbb{R}$,
- $\pi_U(\mathbf{x})$ is closest to $\mathbf{x}$ in U, and hence

$$\pi_U(\mathbf{x}) - \mathbf{x} \perp \mathbf{b} \Leftrightarrow \langle \pi_U(\mathbf{x}) - \mathbf{x}, \mathbf{b} \rangle = 0,$$

- $\pi_U(\mathbf{x}) = \lambda\mathbf{b} = \dfrac{\langle \mathbf{x}, \mathbf{b} \rangle}{\|\mathbf{b}\|^2}\mathbf{b}$,
- $\|\pi_U(\mathbf{x})\| = |\cos w|\|\mathbf{x}\|$, where w is the angle between $\mathbf{x}$ and $\mathbf{b}$.
- With the dot product as inner product the projection matrix P_π satisfying the condition $\pi_U(\mathbf{x}) = P_\pi\mathbf{x}$ is given by

$$P_\pi = \frac{\mathbf{b}\mathbf{b}^T}{\|\mathbf{b}\|^2} = \frac{\mathbf{b}\mathbf{b}^T}{\mathbf{b}^T\mathbf{b}}.$$

Clearly, $\mathbf{b}\mathbf{b}^T$ is a symmetric matrix.

Orthogonal Projection onto General Subspaces

Let $U \subset \mathbb{R}^n$ be a subspace with $\dim(U) = m$. Assume $(\mathbf{b}_1, \ldots, \mathbf{b}_m)$ is a basis for U.

Orthogonal Projection onto General Subspaces

Let $U \subset \mathbb{R}^n$ be a subspace with $\dim(U) = m$. Assume $(\mathbf{b}_1, \dots, \mathbf{b}_m)$ is a basis for U. Then

- $\pi_U(\mathbf{x}) = \sum_{i=1}^m \lambda_i \mathbf{b}_i = B\lambda$, where
  $B = [\mathbf{b}_1, \dots, \mathbf{b}_m] \in \mathbb{R}^{n \times m},\ \lambda = [\lambda_1, \dots, \lambda_m]^T \in \mathbb{R}^m$.

Orthogonal Projection onto General Subspaces

Let $U \subset \mathbb{R}^n$ be a subspace with $\dim(U) = m$. Assume $(\mathbf{b}_1, \dots, \mathbf{b}_m)$ is a basis for U. Then

- $\pi_U(\mathbf{x}) = \sum_{i=1}^m \lambda_i \mathbf{b}_i = B\lambda$, where $B = [\mathbf{b}_1, \dots, \mathbf{b}_m] \in \mathbb{R}^{n \times m}, \; \lambda = [\lambda_1, \dots, \lambda_m]^T \in \mathbb{R}^m$.
- The condition $\mathbf{x} - \pi_U(\mathbf{x}) \perp U$ is equivalent to $\langle \mathbf{b}_i, \mathbf{x} - \pi_U(\mathbf{x}) \rangle = \mathbf{b}_i^T(\mathbf{x} - \pi_U(\mathbf{x})) = 0$ for $i = 1, \dots, m$. Hence

$$\begin{bmatrix} \mathbf{b}_1^T \\ \vdots \\ \mathbf{b}_m^T \end{bmatrix} [\mathbf{x} - B\lambda] = \mathbf{0} \Leftrightarrow B^T(\mathbf{x} - B\lambda) = \mathbf{0} \Leftrightarrow B^T B\lambda = B^T\mathbf{x}.$$

Therefore

$$\lambda = (B^T B)^{-1} B^T \mathbf{x}.$$

The matrix $(B^T B)^{-1} B^T$ is also called **pseudo-inverse** of B, which can be computed for non-square matrices B.

Orthogonal Projection onto General Subspaces

- By $\pi_U(\mathbf{x}) = B\lambda$ we have the projection $\pi_U(\mathbf{x})$ is computed by

$$\pi_U(\mathbf{x}) = B(B^T B)^{-1} B^T \mathbf{x}.$$

Orthogonal Projection onto General Subspaces

- By $\pi_U(\mathbf{x}) = B\lambda$ we have the projection $\pi_U(\mathbf{x})$ is computed by

$$\pi_U(\mathbf{x}) = B(B^T B)^{-1} B^T \mathbf{x}.$$

- The projection matrix P_π satisfying $P_\pi \mathbf{x} = \pi_U(\mathbf{x})$ is given by

$$P_\pi = B(B^T B)^{-1} B^T.$$

Orthogonal Projection onto General Subspaces

- By $\pi_U(\mathbf{x}) = B\lambda$ we have the projection $\pi_U(\mathbf{x})$ is computed by

$$\pi_U(\mathbf{x}) = B(B^T B)^{-1} B^T \mathbf{x}.$$

- The projection matrix P_π satisfying $P_\pi \mathbf{x} = \pi_U(\mathbf{x})$ is given by

$$P_\pi = B(B^T B)^{-1} B^T.$$

Remark

If the basis $(\mathbf{b}_1, \ldots, \mathbf{b}_m)$ is an ONB, i.e. $\mathbf{b}_1, \ldots, \mathbf{b}_m$ are unit and orthogonal vectors, then the projection equation simplifies greatly to

$$\pi_U(\mathbf{x}) = BB^T \mathbf{x}.$$

Orthogonal Projection onto General Subspaces

Projections allow us to look at situations where we have a linear system $A\mathbf{x} = \mathbf{b}$ without a solution. Recall that this means that $\mathbf{b} \notin \mathrm{col}(A)$. Given that the linear equation cannot be solved exactly, we can find an **approximate solution.**

Orthogonal Projection onto General Subspaces

Projections allow us to look at situations where we have a linear system $A\mathbf{x} = \mathbf{b}$ without a solution. Recall that this means that $\mathbf{b} \notin \mathrm{col}(A)$. Given that the linear equation cannot be solved exactly, we can find an **approximate solution.**
The idea is to find the vector in $\mathrm{col}(A)$ that is closest to $\mathbf{b}$, i.e., we compute the orthogonal projection of $\mathbf{b}$ onto the subspace spanned by the columns of A. This problem arises often in practice, and the solution is called the **least squares solution**.

Trace

Definition

The trace of a square matrix $A \in \mathbb{R}^{n \times n}$ *is given by*

$$tr(A) := \sum_{i=1}^{n} a_{ii}$$

in other words, the trace is the sum of the diagonal elements of A.

Trace

Definition

The trace of a square matrix $A \in \mathbb{R}^{n\times n}$ *is given by*

$$tr(A) := \sum_{i=1}^{n} a_{ii}$$

in other words, the trace is the sum of the diagonal elements of A.

Remark

For $A; B \in \mathbb{R}^{n\times n}$ *the trace satisfies the following properties:*

1. $tr(A + B) = tr(A) + tr(B)$
2. $tr(\alpha A) = \alpha tr(A)$
3. $tr(I_n) = n$
4. $tr(AB) = tr(BA)$

Trace

Corollary

- *In particular, for two vectors* $x, y \in \mathbb{R}^n$

$$tr(\mathbf{x}\mathbf{y}^T) = tr(\mathbf{y}^T\mathbf{x}) = \mathbf{y}^T\mathbf{x} \in \mathbb{R}.$$

- *If* $A \sim B$ *(i.e.* $\exists\, Q$ *s.t.* $A = QBQ^{-1}$*), then* $trA = trB$
- *All transformation matrices* A_T *of a linear mapping* $T : V \to V$ *have the same trace.*

Trace

Corollary

- *In particular, for two vectors* $x, y \in \mathbb{R}^n$

$$tr(\mathbf{x}\mathbf{y}^T) = tr(\mathbf{y}^T\mathbf{x}) = \mathbf{y}^T\mathbf{x} \in \mathbb{R}.$$

- *If* $A \sim B$ *(i.e.* $\exists\, Q$ *s.t.* $A = QBQ^{-1}$*), then* $tr A = tr B$
- *All transformation matrices* A_T *of a linear mapping* $T : V \to V$ *have the same trace.*

Definition

For $\lambda \in \mathbb{R}$ *and a square matrix* $A \in \mathbb{R}^{n\times n}$

$$p_A(\lambda) = \det(A - \lambda I) = c_0 + c_1\lambda + c_2\lambda^2 + \ldots + c_{n-1}\lambda^{n-1} + (-1)^n\lambda^n$$

$c_0, \ldots, c_{n-1} \in \mathbb{R}$, *is the* **characteristic polynomial of** A. *In particular,*

$$c_0 = \det(A), \quad c_{n-1} = (-1)^{n-1} tr(A).$$