Analytic Geometry

Libin Jiao

Dalian University of Technology

September 27, 2022

Outline

Outline

Norms

Definition

A norm on a vector space 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 $\lambda \in \mathbf{R}$ and $\boldsymbol{x}, \boldsymbol{y} \in V$ the following hold:

- Absolutely homogeneous: $\|\lambda \boldsymbol{x}\| = |\lambda| \|\boldsymbol{x}\|$
- Triangle inequality: $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$,
- Positive definite: $\|\boldsymbol{x}\| \geqslant 0$ and $\|\boldsymbol{x}\| = 0 \iff \boldsymbol{x} = \mathbf{0}$.

Norms

Definition

A norm on a vector space 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 $\lambda \in \mathbf{R}$ and $\boldsymbol{x}, \boldsymbol{y} \in V$ the following hold:

- Absolutely homogeneous: $\|\lambda \boldsymbol{x}\| = |\lambda| \|\boldsymbol{x}\|$
- Triangle inequality: $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$,
- Positive definite: $\|\boldsymbol{x}\| \geqslant 0$ and $\|\boldsymbol{x}\| = 0 \iff \boldsymbol{x} = \mathbf{0}$.

Norms

Definition

A norm on a vector space 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 $\lambda \in \mathbf{R}$ and $\boldsymbol{x}, \boldsymbol{y} \in V$ the following hold:

- Absolutely homogeneous: $\|\lambda \boldsymbol{x}\| = |\lambda| \|\boldsymbol{x}\|$
- Triangle inequality: $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$,
- Positive definite: $\|\boldsymbol{x}\| \geqslant 0$ and $\|\boldsymbol{x}\| = 0 \iff \boldsymbol{x} = \mathbf{0}$.

Norms

Definition

A norm on a vector space 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 $\lambda \in \mathbf{R}$ and $\boldsymbol{x}, \boldsymbol{y} \in V$ the following hold:

- Absolutely homogeneous: $\|\lambda \boldsymbol{x}\| = |\lambda| \|\boldsymbol{x}\|$
- Triangle inequality: $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$,
- Positive definite: $\|\boldsymbol{x}\| \geqslant 0$ and $\|\boldsymbol{x}\| = 0 \iff \boldsymbol{x} = \mathbf{0}$.

Norms

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;

Recall that for a vector $\boldsymbol{x} \in \mathbb{R}^n$ we denote the elements of the vector using a subscript, that is, x_i is the i^{th} element of the vector x.

Norms

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;

Recall that for a vector $\boldsymbol{x} \in \mathbb{R}^n$ we denote the elements of the vector using a subscript, that is, x_i is the i^{th} element of the vector x.

Norms

Example

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

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

where $|\cdot|$ is the absolute value. The left panel of Figure 3.3 shows all vectors $\boldsymbol{x} \in \mathbf{R}^2$ with $\|\boldsymbol{x}\|_1 = 1$. The Manhattan norm is also called ℓ_1 norm.

Norms

Example

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

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

and computes the Euclidean distance of $\boldsymbol{x}$ from the origin. The Euclidean norm is also called ℓ_2 norm.

Remark: Throughout this book, we will use the Euclidean norm by default if not stated otherwise.

Norms

Example

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

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

and computes the Euclidean distance of $\boldsymbol{x}$ from the origin. The Euclidean norm is also called ℓ_2 norm.

Remark: Throughout this book, we will use the Euclidean norm by default if not stated otherwise.

Norms

Example

The Frobenius norm on $\mathbb{R}^{m\times n}$ is defined for $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ as.

$$\|\boldsymbol{A}\|_F := \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2} = \sqrt{tr(\boldsymbol{A}^\top \boldsymbol{A})}$$

Outline

Inner Products

The inner product (dot product) of two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{y}^T \boldsymbol{x} = \boldsymbol{x}^T \boldsymbol{y} = \sum_{i=1}^{n} x_i y_i.$$

- if $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$, then $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal.

The outer product (cross product) of two vectors $\boldsymbol{x} \in \mathbb{R}^m$ and $\boldsymbol{y} \in \mathbb{R}^n$ is a matrix $\boldsymbol{A} = \boldsymbol{x}\boldsymbol{y}^T$, where,

$$(a_{ij}) = (x_i y_j) = \begin{pmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_m y_1 & x_m y_2 & \cdots & x_m y_n \end{pmatrix}.$$

Inner Products

The inner product (dot product) of two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{y}^T \boldsymbol{x} = \boldsymbol{x}^T \boldsymbol{y} = \sum_{i=1}^{n} x_i y_i.$$

- if $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$, then $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal.

The outer product (cross product) of two vectors $\boldsymbol{x} \in \mathbb{R}^m$ and $\boldsymbol{y} \in \mathbb{R}^n$ is a matrix $\boldsymbol{A} = \boldsymbol{x}\boldsymbol{y}^T$, where,

$$(a_{ij}) = (x_i y_j) = \begin{pmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{pmatrix}.$$

General Inner Products

Recall the linear mapping from Section 2.7, where we can rearrange the mapping with respect to addition and multiplication with a scalar. A bilinear mapping Ω is a mapping with two arguments, and it is linear in each argument, i.e., when we look at a vector space V then it holds that for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V, \lambda, \psi \in \mathbb{R}$ that

$$\Omega(\lambda x + \psi y, z) = \lambda\Omega(x, z) + \psi\Omega(y, z) \tag{1}$$
$$\Omega(x, \lambda y + \psi z) = \lambda\Omega(x, y) + \psi\Omega(x, z) \tag{2}$$

Here, (1) asserts that Ω is linear in the first argument, and (2) asserts that Ω is linear in the second argument.

General Inner Products

Recall the linear mapping from Section 2.7, where we can rearrange the mapping with respect to addition and multiplication with a scalar. A bilinear mapping Ω is a mapping with two arguments, and it is linear in each argument, i.e., when we look at a vector space V then it holds that for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V, \lambda, \psi \in \mathbb{R}$ that

$$\Omega(\lambda x + \psi y, z) = \lambda\Omega(x, z) + \psi\Omega(y, z) \tag{1}$$

$$\Omega(x, \lambda y + \psi z) = \lambda\Omega(x, y) + \psi\Omega(x, z) \tag{2}$$

Here, (1) asserts that Ω is linear in the first argument, and (2) asserts that Ω is linear in the second argument.

General Inner Products

Definition

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

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

$$\forall x \in V \backslash \{0\} : \Omega(x, x) > 0, \quad \Omega(0, 0) = 0$$

General Inner Products

Definition

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

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

$$\forall x \in V \backslash \{0\} : \Omega(x, x) > 0, \quad \Omega(0, 0) = 0$$

General Inner Products

Definition

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

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

$$\forall x \in V \backslash \{0\} : \Omega(x, x) > 0, \quad \Omega(0, 0) = 0$$

General Inner Products

Definition

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

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

General Inner Products

Definition

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

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

General Inner Products

Definition

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

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

General Inner Products

Example

Consider $V = \mathbf{R}^2$. If we define

$$(x, y) := x_1 y_1 - (x_1 y_2 + x_2 y_1) + 2x_2 y_2$$

then $(\cdot, \cdot)$ is an inner product but different from the dot product. The proof will be an exercise.

Symmetric, Positive Definite Matrices

Consider an n-dimensional vector space V with an inner product $\langle\cdot,\cdot\rangle : V \times V \to \mathrm{R}$ (see Definition 3.3) and an ordered basis $B = (b_1, \ldots, b_n)$ of V. Recall from Section 2.6.1 that any vectors $\boldsymbol{x}, y \in V$ can be written as linear combinations of the basis vectors so that $x = \sum_{i=1}^{n} \psi_i b_i \in V$ and $y = \sum_{j=1}^{n} \lambda_j b_j \in V$ for suitable $\psi_i, \lambda_j \in \mathbb{R}$.

Symmetric, Positive Definite Matrices

Consider an n-dimensional vector space V with an inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathrm{R}$ (see Definition 3.3) and an ordered basis $B = (b_1, \ldots, b_n)$ of V. Recall from Section 2.6.1 that any vectors $\boldsymbol{x}, y \in V$ can be written as linear combinations of the basis vectors so that $x = \sum_{i=1}^{n} \psi_i b_i \in V$ and $y = \sum_{j=1}^{n} \lambda_j b_j \in V$ for suitable $\psi_i, \lambda_j \in \mathbb{R}$.

Symmetric, Positive Definite Matrices

Due to the bilinearity of the inner product, it holds for all $x, y \in V$ that

$$\langle x, y\rangle = \left\langle \sum_{i=1}^{n} \psi_i b_i, \sum_{j=1}^{n} \lambda_j b_j \right\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} \psi_i \langle b_i, b_j\rangle \lambda_j = \hat{x}^\top \boldsymbol{A} \hat{y} s$$

where $A_{ij} := \langle b_i, b_j\rangle$ and $\hat{x}, \hat{y}$ are the coordinates of x and y with respect to the basis B. This implies that the inner product $\langle\cdot,\cdot\}$ is uniquely determined through A. The symmetry of the inner product also means that A is symmetric.

Symmetric, Positive Definite Matrices

Due to the bilinearity of the inner product, it holds for all $x, y \in V$ that

$$\langle x, y \rangle = \left\langle \sum_{i=1}^{n} \psi_i b_i, \sum_{j=1}^{n} \lambda_j b_j \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_i \langle b_i, b_j \rangle \lambda_j = \hat{x}^\top \boldsymbol{A} \hat{y} s$$

where $A_{ij} := \langle b_i, b_j \rangle$ and $\hat{x}, \hat{y}$ are the coordinates of x and y with respect to the basis B. This implies that the inner product $\langle \cdot, \cdot \}$ is uniquely determined through A. The symmetry of the inner product also means that A is symmetric.

Symmetric, Positive Definite Matrices

Furthermore, the positive definiteness of the inner product implies that

$$\forall \boldsymbol{x} \in V \backslash \{0\} : \boldsymbol{x}^{\top} \boldsymbol{A} x > 0 \tag{3}$$

Definition

A symmetric matrix $A \in \mathbf{R}^{n \times n}$ that satisfies (3) is called symmetric, positive definite, or just positive definite. If only $\geqslant$ holds in (3), then A is called symmetric, positive semi-definite.

Symmetric, Positive Definite Matrices

Furthermore, the positive definiteness of the inner product implies that

$$\forall \boldsymbol{x} \in V \backslash \{0\} : \boldsymbol{x}^\top \boldsymbol{A} x > 0 \tag{3}$$

Definition

A symmetric matrix $A \in \mathbf{R}^{n \times n}$ that satisfies (3) is called symmetric, positive definite, or just positive definite. If only $\geqslant$ holds in (3), then A is called symmetric, positive semi-definite.

Symmetric, Positive Definite Matrices

Example

Consider the matrices

$$\boldsymbol{A}_1 = \begin{bmatrix} 9 & 6 \\ 6 & 5 \end{bmatrix}, \quad \boldsymbol{A}_2 = \begin{bmatrix} 9 & 6 \\ 6 & 3 \end{bmatrix}$$

.

$\boldsymbol{A}_1$ is positive definite because it is symmetric and

$$\begin{aligned} \boldsymbol{x}^\top \boldsymbol{A}_1 \boldsymbol{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 9 & 6 \\ 6 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= 9x_1^2 + 12x_1x_2 + 5x_2^2 = (3x_1 + 2x_2)^2 + x_2^2 > 0 \end{aligned}$$

for all $x \in V \backslash \{0\}$.

Symmetric, Positive Definite Matrices

Example

Consider the matrices

$$\boldsymbol{A}_1 = \begin{bmatrix} 9 & 6 \\ 6 & 5 \end{bmatrix}, \quad \boldsymbol{A}_2 = \begin{bmatrix} 9 & 6 \\ 6 & 3 \end{bmatrix}$$

.

$\boldsymbol{A}_1$ is positive definite because it is symmetric and

$$\begin{aligned} \boldsymbol{x}^\top \boldsymbol{A}_1 \boldsymbol{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 9 & 6 \\ 6 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= 9x_1^2 + 12x_1x_2 + 5x_2^2 = (3x_1 + 2x_2)^2 + x_2^2 > 0 \end{aligned}$$

for all $x \in V \backslash \{0\}$.

Symmetric, Positive Definite Matrices

Example

In contrast, A_2 is symmetric but not positive definite because $x^\top \boldsymbol{A}_2\boldsymbol{x} = 9x_1^2 + 12x_1x_2 + 3x_2^2 = (3x_1 + 2x_2)^2 - x_2^2$ can be less than 0, e.g., for $x = [2, -3]^\top$.

Symmetric, Positive Definite Matrices

If $A \in \mathbb{R}^{n \times n}$ is symmetric, positive definite, then

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

defines an inner product with respect to an ordered basis B, where $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ are the coordinate representations of $\boldsymbol{x}$, $\boldsymbol{y} \in V$ with respect to B.

Symmetric, Positive Definite Matrices

If $A \in \mathbb{R}^{n\times n}$ is symmetric, positive definite, then

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

defines an inner product with respect to an ordered basis B, where $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ are the coordinate representations of $\boldsymbol{x}$, $\boldsymbol{y} \in V$ with respect to B.

Symmetric, Positive Definite Matrices

Theorem

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

Outline

Lengths and Distances

In Section 3.1, we already discussed norms that we can use to compute the length of a vector. Inner products and norms are closely related in the sense that any inner product induces a norm

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

in a natural way, such that we can compute lengths of vectors using the inner product. However, not every norm is induced by an inner product. The Manhattan norm is an example of a norm without a corresponding inner product. In the following, we will focus on norms that are induced by inner products and introduce geometric concepts, such as lengths, distances, and angles.

Lengths and Distances

In Section 3.1, we already discussed norms that we can use to compute the length of a vector. Inner products and norms are closely related in the sense that any inner product induces a norm

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

in a natural way, such that we can compute lengths of vectors using the inner product. However, not every norm is induced by an inner product. The Manhattan norm is an example of a norm without a corresponding inner product. In the following, we will focus on norms that are induced by inner products and introduce geometric concepts, such as lengths, distances, and angles.

Lengths and Distances

In Section 3.1, we already discussed norms that we can use to compute the length of a vector. Inner products and norms are closely related in the sense that any inner product induces a norm

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

in a natural way, such that we can compute lengths of vectors using the inner product. However, not every norm is induced by an inner product. The Manhattan norm is an example of a norm without a corresponding inner product. In the following, we will focus on norms that are induced by inner products and introduce geometric concepts, such as lengths, distances, and angles.

Lengths and Distances

Remark: (Cauchy-Schwarz Inequality). For an inner product vector space $(V, \langle\cdot,\cdot\rangle)$ the induced norm $\|\cdot\|$ satisfies the Cauchy-Schwarz inequality

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leqslant \|\boldsymbol{x}\| \|\boldsymbol{y}\|$$

Lengths and Distances

Example

In geometry, we are often interested in lengths of vectors. We can now use an inner product to compute them. Let us take $\boldsymbol{x} = [1, 1]^\top \in \mathbb{R}^2$. If we use the dot product as the inner product, we obtain

$$\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}} = \sqrt{1^2 + 1^2} = \sqrt{2}$$

as the length of $\boldsymbol{x}$.

Lengths and Distances

Example

Let us now choose a different inner product:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \boldsymbol{x}^\top \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \boldsymbol{y} = x_1 y_1 - \frac{1}{2}(x_1 y_2 + x_2 y_1) + x_2 y_2$$

If we compute the norm of a vector, then this inner product returns smaller values than the dot product if x_1 and x_2 have the same sign (and $x_1 x_2 > 0$); otherwise, it returns greater values than the dot product. With this inner product, we obtain

$$\langle \boldsymbol{x}, \boldsymbol{x} \rangle = x_1^2 - x_1 x_2 + x_2^2 = 1 - 1 + 1 = 1 \Longrightarrow \|\boldsymbol{x}\| = \sqrt{1} = 1$$

such that x is "shorter" with this inner product than with the dot product.

Lengths and Distances

Example

Let us now choose a different inner product:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \boldsymbol{x}^\top \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \boldsymbol{y} = x_1 y_1 - \frac{1}{2}(x_1 y_2 + x_2 y_1) + x_2 y_2$$

If we compute the norm of a vector, then this inner product returns smaller values than the dot product if x_1 and x_2 have the same sign (and $x_1 x_2 > 0$); otherwise, it returns greater values than the dot product. With this inner product, we obtain

$$\langle \boldsymbol{x}, \boldsymbol{x} \rangle = x_1^2 - x_1 x_2 + x_2^2 = 1 - 1 + 1 = 1 \Longrightarrow \|\boldsymbol{x}\| = \sqrt{1} = 1$$

such that x is "shorter" with this inner product than with the dot product.

Lengths and Distances

Example

Let us now choose a different inner product:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \boldsymbol{x}^\top \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \boldsymbol{y} = x_1 y_1 - \frac{1}{2}(x_1 y_2 + x_2 y_1) + x_2 y_2$$

If we compute the norm of a vector, then this inner product returns smaller values than the dot product if x_1 and x_2 have the same sign (and $x_1 x_2 > 0$); otherwise, it returns greater values than the dot product. With this inner product, we obtain

$$\langle \boldsymbol{x}, \boldsymbol{x} \rangle = x_1^2 - x_1 x_2 + x_2^2 = 1 - 1 + 1 = 1 \Longrightarrow \|\boldsymbol{x}\| = \sqrt{1} = 1$$

such that x is "shorter" with this inner product than with the dot product.

Lengths and Distances

Definition

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 Euclidean distance.
The mapping

$$d : V \times V \to \mathbb{R}$$
$$(\boldsymbol{x}, \boldsymbol{y}) \to d(\boldsymbol{x}, \boldsymbol{y})$$

is called a metric.

Lengths and Distances

Definition

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 Euclidean distance.

The mapping

$$d : V \times V \to \mathbb{R}$$
$$(\boldsymbol{x}, \boldsymbol{y}) \to d(\boldsymbol{x}, \boldsymbol{y})$$

is called a metric.

Lengths and Distances

Definition

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 Euclidean distance.
The mapping

$$d : V \times V \to \mathbb{R}$$
$$(\boldsymbol{x},\boldsymbol{y}) \to d(\boldsymbol{x},\boldsymbol{y})$$

is called a metric.

Lengths and Distances

Remark: Similar to the length of a vector, the distance between vectors does not require an inner product: a norm is sufficient. If we have a norm induced by an inner product, the distance may vary depending on the choice of the inner product. A metric d satisfies the following:

- d is positive definite, i.e., $d(\boldsymbol{x}, \boldsymbol{y}) \geqslant 0$ for all $\boldsymbol{x}, \boldsymbol{y} \in V$ and $d(\boldsymbol{x}, \boldsymbol{y}) = 0 \Longleftrightarrow x = y$.
- 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: $d(\boldsymbol{x}, \boldsymbol{z}) \leqslant d(\boldsymbol{x}, \boldsymbol{y}) + d(\boldsymbol{y}, \boldsymbol{z})$ for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$.

Lengths and Distances

Remark: Similar to the length of a vector, the distance between vectors does not require an inner product: a norm is sufficient. If we have a norm induced by an inner product, the distance may vary depending on the choice of the inner product. A metric d satisfies the following:

- d is positive definite, i.e., $d(\boldsymbol{x}, \boldsymbol{y}) \geqslant 0$ for all $\boldsymbol{x}, \boldsymbol{y} \in V$ and $d(\boldsymbol{x}, \boldsymbol{y}) = 0 \iff x = y$.
- 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: $d(\boldsymbol{x}, \boldsymbol{z}) \leqslant d(\boldsymbol{x}, \boldsymbol{y}) + d(\boldsymbol{y}, \boldsymbol{z})$ for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$.

Lengths and Distances

Remark: Similar to the length of a vector, the distance between vectors does not require an inner product: a norm is sufficient. If we have a norm induced by an inner product, the distance may vary depending on the choice of the inner product. A metric d satisfies the following:

- d is positive definite, i.e., $d(\boldsymbol{x}, \boldsymbol{y}) \geqslant 0$ for all $\boldsymbol{x}, \boldsymbol{y} \in V$ and $d(\boldsymbol{x}, \boldsymbol{y}) = 0 \iff x = y$.
- 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: $d(\boldsymbol{x}, \boldsymbol{z}) \leqslant d(\boldsymbol{x}, \boldsymbol{y}) + d(\boldsymbol{y}, \boldsymbol{z})$ for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$.

Lengths and Distances

Remark: Similar to the length of a vector, the distance between vectors does not require an inner product: a norm is sufficient. If we have a norm induced by an inner product, the distance may vary depending on the choice of the inner product. A metric d satisfies the following:

- d is positive definite, i.e., $d(\boldsymbol{x}, \boldsymbol{y}) \geqslant 0$ for all $\boldsymbol{x}, \boldsymbol{y} \in V$ and $d(\boldsymbol{x}, \boldsymbol{y}) = 0 \iff x = y$.
- 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: $d(\boldsymbol{x}, \boldsymbol{z}) \leqslant d(\boldsymbol{x}, \boldsymbol{y}) + d(\boldsymbol{y}, \boldsymbol{z})$ for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$.

Outline

Angles and Orthogonality

In addition to enabling the definition of lengths of vectors, as well as the distance between two vectors, inner products also capture the geometry of a vector space by defining the angle ω between two vectors. We use the Cauchy-Schwarz inequality to define angles ω in inner product spaces between two vectors $\boldsymbol{x}, \boldsymbol{y}$, and this notion coincides with our intuition in R^2 and $\mathbb{R}^3$.
Assume that $\boldsymbol{x} \neq \mathbf{0}, \boldsymbol{y} \neq \mathbf{0}$. Then

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

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

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

Angles and Orthogonality

In addition to enabling the definition of lengths of vectors, as well as the distance between two vectors, inner products also capture the geometry of a vector space by defining the angle ω between two vectors. We use the Cauchy-Schwarz inequality to define angles ω in inner product spaces between two vectors $\boldsymbol{x}, \boldsymbol{y}$, and this notion coincides with our intuition in R^2 and $\mathbb{R}^3$.
Assume that $\boldsymbol{x} \neq \mathbf{0}, \boldsymbol{y} \neq \mathbf{0}$. Then

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

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

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

Angles and Orthogonality

In addition to enabling the definition of lengths of vectors, as well as the distance between two vectors, inner products also capture the geometry of a vector space by defining the angle ω between two vectors. We use the Cauchy-Schwarz inequality to define angles ω in inner product spaces between two vectors $\boldsymbol{x}, \boldsymbol{y}$, and this notion coincides with our intuition in R^2 and $\mathbb{R}^3$.
Assume that $\boldsymbol{x} \neq \mathbf{0}, \boldsymbol{y} \neq \mathbf{0}$. Then

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

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

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

Angles and Orthogonality

In addition to enabling the definition of lengths of vectors, as well as the distance between two vectors, inner products also capture the geometry of a vector space by defining the angle ω between two vectors. We use the Cauchy-Schwarz inequality to define angles ω in inner product spaces between two vectors $\boldsymbol{x}, \boldsymbol{y}$, and this notion coincides with our intuition in R^2 and $\mathbb{R}^3$.
Assume that $\boldsymbol{x} \neq \mathbf{0}, \boldsymbol{y} \neq \mathbf{0}$. Then

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

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

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

Angles and Orthogonality

The number ω is the angle between the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. Intuitively, the angle between two vectors tells us how similar their orientations are. For example, using the dot product, the angle between $\boldsymbol{x}$ and $\boldsymbol{y} = 4\boldsymbol{x}$, i.e., $\boldsymbol{y}$ is a scaled version of $\boldsymbol{x}$, is 0: Their orientation is the same.

Angles and Orthogonality

Definition

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

Angles and Orthogonality

Definition

A square matrix $A \in \mathbb{R}^{n\times n}$ orthogonal matrix if and only if its columns are orthonormal so

$$AA^\top = I = A^\top A,$$

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

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

Angles and Orthogonality

Transformations by orthogonal matrices are special because the length of a vector x is not changed when transforming it using an orthogonal matrix A. For the dot product, we obtain

$$\|Ax\|^2 = (Ax)^\top (Ax) = x^\top A^\top A x = x^\top I x = x^\top x = \|x\|^2. \quad (3.31)$$

Angles and Orthogonality

Moreover, the angle between any two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, as measured by their inner product, is also unchanged when transforming both of them using an orthogonal matrix A. Assuming the dot product as the inner product

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

which gives exactly the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$. This means that orthogonal matrices A with $A^\top = A^{-1}$ preserve both angles and distances. It turns out that orthogonal matrices define transformations that are rotations (with the possibility of flips).

Angles and Orthogonality

Moreover, the angle between any two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, as measured by their inner product, is also unchanged when transforming both of them using an orthogonal matrix A. Assuming the dot product as the inner product

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

which gives exactly the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$. This means that orthogonal matrices A with $A^\top = A^{-1}$ preserve both angles and distances. It turns out that orthogonal matrices define transformations that are rotations (with the possibility of flips).

Angles and Orthogonality

Moreover, the angle between any two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, as measured by their inner product, is also unchanged when transforming both of them using an orthogonal matrix A. Assuming the dot product as the inner product

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

which gives exactly the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$. This means that orthogonal matrices A with $A^\top = A^{-1}$ preserve both angles and distances. It turns out that orthogonal matrices define transformations that are rotations (with the possibility of flips).

Outline

Orthonormal Basis

In Section 2.6.1, we characterized properties of basis vectors and found that in an n-dimensional vector space, we need n basis vectors, i.e., n vectors that are linearly independent. In Sections 3.3 and 3.4, we used inner products to compute the length of vectors and the angle between vectors. In the following, we will discuss the special case where the basis vectors are orthogonal to each other and where the length of each basis vector is 1 . We will call this basis then an orthonormal basis.

Orthonormal Basis

In Section 2.6.1, we characterized properties of basis vectors and found that in an n-dimensional vector space, we need n basis vectors, i.e., n vectors that are linearly independent. In Sections 3.3 and 3.4, we used inner products to compute the length of vectors and the angle between vectors. In the following, we will discuss the special case where the basis vectors are orthogonal to each other and where the length of each basis vector is 1 . We will call this basis then an orthonormal basis.

Orthonormal Basis

In Section 2.6.1, we characterized properties of basis vectors and found that in an n-dimensional vector space, we need n basis vectors, i.e., n vectors that are linearly independent. In Sections 3.3 and 3.4, we used inner products to compute the length of vectors and the angle between vectors. In the following, we will discuss the special case where the basis vectors are orthogonal to each other and where the length of each basis vector is 1 . We will call this basis then an orthonormal basis.

Orthonormal Basis

Definition

Consider an n-dimensional vector space V and a basis $\{b_1, \ldots, \boldsymbol{b}_n\}$ of V. If

$$\begin{aligned} \langle b_i, \boldsymbol{b}_j \rangle &= 0 \\ \langle \boldsymbol{b}_i, \boldsymbol{b}_i \rangle &= 1 \end{aligned} \quad \text{for } i \neq j$$

for all $i, j = 1, 2, \ldots, n$, then the basis is called an orthonormal basis (ONB).

Orthonormal Basis

Definition

Consider an n-dimensional vector space V and a basis $\{b_1, \ldots, \boldsymbol{b}_n\}$ of V. If

$$\begin{aligned} \langle b_i, \boldsymbol{b}_j \rangle &= 0 \\ \langle \boldsymbol{b}_i, \boldsymbol{b}_i \rangle &= 1 \end{aligned} \quad \text{for } i \neq j$$

for all $i, j = 1, 2, \ldots, n$, then the basis is called an orthonormal basis (ONB).

Orthonormal Basis

Definition

Consider an n-dimensional vector space V and a basis $\{b_1, \ldots, \boldsymbol{b}_n\}$ of V. If

$$\begin{aligned} \langle b_i, \boldsymbol{b}_j \rangle &= 0 \\ \langle \boldsymbol{b}_i, \boldsymbol{b}_i \rangle &= 1 \end{aligned} \quad \text{for } i \neq j$$

for all $i, j = 1, 2, \ldots, n$, then the basis is called an orthonormal basis (ONB).

Inner Product of Functions

Definition

An inner product of two functions $u : \mathbb{R} \to \mathbb{R}$ and $v : \mathbb{R} \to \mathbb{R}$ can be defined as the definite integral

$$\langle u, v \rangle := \int_a^b u(x)v(x)dx$$

for lower and upper limits $a, b < \infty$, respectively.

As with our usual inner product, we can define norms and orthogonality by looking at the inner product. If $\langle u, v \rangle = 0$, we say that the functions u and v are orthogonal.

Inner Product of Functions

Definition

An inner product of two functions $u : \mathbb{R} \to \mathbb{R}$ and $v : \mathbb{R} \to \mathbb{R}$ can be defined as the definite integral

$$\langle u, v \rangle := \int_a^b u(x)v(x)dx$$

for lower and upper limits $a, b < \infty$, respectively.

As with our usual inner product, we can define norms and orthogonality by looking at the inner product. If $\langle u, v \rangle = 0$, we say that the functions u and v are orthogonal.

Inner Product of Functions

Example

If we choose $u = \sin(x)$ and $v = \cos(x)$, and define the integrand $f(x) = u(x)v(x)$, we see that this function is odd, i.e., $\sin(x)\cos(x)$. $f(-x) = -f(x)$. Therefore, the integral with limits $a = -\pi, b = \pi$ of this product evaluates to 0 . Therefore, sin and cos are orthogonal functions.

Remark: It also holds that the collection of functions

$$\{1, \cos(x), \cos(2x), \cos(3x), \dots\}$$

is orthogonal if we integrate from $-\pi$ to π, i.e., any pair of functions are orthogonal to each other. The collection of functions in $\{1, \cos(x), \cos(2x), \cos(3x), \dots\}$ spans a large subspace of the functions that are even and periodic on $[-\pi, \pi)$, and projecting functions onto this subspace is the fundamental idea behind Fourier series.

Inner Product of Functions

Example

If we choose $u = \sin(x)$ and $v = \cos(x)$, and define the integrand $f(x) = u(x)v(x)$, we see that this function is odd, i.e., $\sin(x)\cos(x)$. $f(-x) = -f(x)$. Therefore, the integral with limits $a = -\pi, b = \pi$ of this product evaluates to 0 . Therefore, sin and cos are orthogonal functions.

Remark: It also holds that the collection of functions

$$\{1, \cos(x), \cos(2x), \cos(3x), \ldots\}$$

is orthogonal if we integrate from $-\pi$ to π, i.e., any pair of functions are orthogonal to each other. The collection of functions in $\{1, \cos(x), \cos(2x), \cos(3x), \ldots\}$ spans a large subspace of the functions that are even and periodic on $[-\pi, \pi)$, and projecting functions onto this subspace is the fundamental idea behind Fourier series.

Inner Product of Functions

Example

If we choose $u = \sin(x)$ and $v = \cos(x)$, and define the integrand $f(x) = u(x)v(x)$, we see that this function is odd, i.e., $\sin(x)\cos(x)$. $f(-x) = -f(x)$. Therefore, the integral with limits $a = -\pi, b = \pi$ of this product evaluates to 0 . Therefore, sin and cos are orthogonal functions.

Remark: It also holds that the collection of functions

$$\{1, \cos(x), \cos(2x), \cos(3x), \ldots\}$$

is orthogonal if we integrate from $-\pi$ to π, i.e., any pair of functions are orthogonal to each other. The collection of functions in $\{1, \cos(x), \cos(2x), \cos(3x), \ldots\}$ spans a large subspace of the functions that are even and periodic on $[-\pi, \pi)$, and projecting functions onto this subspace is the fundamental idea behind Fourier series.

Outline

Orthogonal Projections

- Projections are an important class of linear transformations (besides rotations and reflections) and play an important role in graphics, coding theory, statistics and machine learning.
- In machine learning, we often deal with data that is high-dimensional. High-dimensional data is often hard to analyze or visualize.
- However, high-dimensional data quite often possesses the property that only a few dimensions contain most information, and most other dimensions are not essential to describe key properties of the data.

Orthogonal Projections

- Projections are an important class of linear transformations (besides rotations and reflections) and play an important role in graphics, coding theory, statistics and machine learning.
- In machine learning, we often deal with data that is high-dimensional. High-dimensional data is often hard to analyze or visualize.
- However, high-dimensional data quite often possesses the property that only a few dimensions contain most information, and most other dimensions are not essential to describe key properties of the data.

Orthogonal Projections

- Projections are an important class of linear transformations (besides rotations and reflections) and play an important role in graphics, coding theory, statistics and machine learning.
- In machine learning, we often deal with data that is high-dimensional. High-dimensional data is often hard to analyze or visualize.
- However, high-dimensional data quite often possesses the property that only a few dimensions contain most information, and most other dimensions are not essential to describe key properties of the data.

Orthogonal Projections

- When we compress or visualize high-dimensional data, we will lose information. To minimize this compression loss, we ideally find the most informative dimensions in the data.

Definition

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

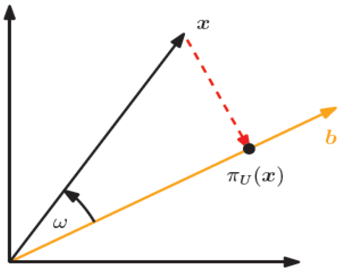
Orthogonal Projections

- When we compress or visualize high-dimensional data, we will lose information. To minimize this compression loss, we ideally find the most informative dimensions in the data.

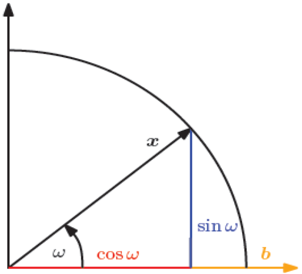
Definition

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

Projection onto One-Dimensional Subspaces (Lines)



(a) Projection of $x \in \mathbb{R}^2$ onto a subspace U with basis vector b.



(b) Projection of a two-dimensional vector x with $\|x\| = 1$ onto a one-dimensional subspace spanned by b.

Projection onto One-Dimensional Subspaces (Lines)

In the following three steps, we determine the coordinate λ, the projection $\pi_U(\boldsymbol{x}) \in U$, and the projection matrix $\boldsymbol{P}_\pi$ that maps any $\boldsymbol{x} \in \mathbb{R}^n$ onto U :

Step 1. Finding the coordinate λ. The orthogonality condition yields

$$\langle \boldsymbol{x} - \pi_U(\boldsymbol{x}), \boldsymbol{b} \rangle = 0 \overset{\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b}}{\iff} \langle \boldsymbol{x} - \lambda \boldsymbol{b}, \boldsymbol{b} \rangle = 0$$

We can now exploit the bilinearity of the inner product and arrive at With a general inner product, we get

$$\langle \boldsymbol{x}, \boldsymbol{b} \rangle - \lambda \langle \boldsymbol{b}, \boldsymbol{b} \rangle = 0 \iff \lambda = \frac{\langle \boldsymbol{x}, \boldsymbol{b} \rangle}{\langle \boldsymbol{b}, \boldsymbol{b} \rangle}$$

Projection onto One-Dimensional Subspaces (Lines)

In the following three steps, we determine the coordinate λ, the projection $\pi_U(\boldsymbol{x}) \in U$, and the projection matrix $\boldsymbol{P}_\pi$ that maps any $\boldsymbol{x} \in \mathbb{R}^n$ onto U :

Step 1. Finding the coordinate λ. The orthogonality condition yields

$$\langle \boldsymbol{x} - \pi_U(\boldsymbol{x}), \boldsymbol{b} \rangle = 0 \overset{\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b}}{\iff} \langle \boldsymbol{x} - \lambda \boldsymbol{b}, \boldsymbol{b} \rangle = 0$$

We can now exploit the bilinearity of the inner product and arrive at With a general inner product, we get

$$\langle \boldsymbol{x}, \boldsymbol{b} \rangle - \lambda \langle \boldsymbol{b}, \boldsymbol{b} \rangle = 0 \iff \lambda = \frac{\langle \boldsymbol{x}, \boldsymbol{b} \rangle}{\langle \boldsymbol{b}, \boldsymbol{b} \rangle}$$

Projection onto One-Dimensional Subspaces (Lines)

In the following three steps, we determine the coordinate λ, the projection $\pi_U(\boldsymbol{x}) \in U$, and the projection matrix $\boldsymbol{P}_\pi$ that maps any $\boldsymbol{x} \in \mathbb{R}^n$ onto U :

Step 1. Finding the coordinate λ. The orthogonality condition yields

$$\langle \boldsymbol{x} - \pi_U(\boldsymbol{x}), \boldsymbol{b} \rangle = 0 \overset{\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b}}{\iff} \langle \boldsymbol{x} - \lambda \boldsymbol{b}, \boldsymbol{b} \rangle = 0$$

We can now exploit the bilinearity of the inner product and arrive at With a general inner product, we get

$$\langle \boldsymbol{x}, \boldsymbol{b} \rangle - \lambda \langle \boldsymbol{b}, \boldsymbol{b} \rangle = 0 \iff \lambda = \frac{\langle \boldsymbol{x}, \boldsymbol{b} \rangle}{\langle \boldsymbol{b}, \boldsymbol{b} \rangle}$$

Projection onto One-Dimensional Subspaces (Lines)

In the following three steps, we determine the coordinate λ, the projection $\pi_U(\boldsymbol{x}) \in U$, and the projection matrix $\boldsymbol{P}_\pi$ that maps any $\boldsymbol{x} \in \mathbb{R}^n$ onto U :

Step 1. Finding the coordinate λ. The orthogonality condition yields

$$\langle \boldsymbol{x} - \pi_U(\boldsymbol{x}), \boldsymbol{b} \rangle = 0 \overset{\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b}}{\Longleftrightarrow} \langle \boldsymbol{x} - \lambda \boldsymbol{b}, \boldsymbol{b} \rangle = 0$$

We can now exploit the bilinearity of the inner product and arrive at With a general inner product, we get

$$\langle \boldsymbol{x}, \boldsymbol{b} \rangle - \lambda \langle \boldsymbol{b}, \boldsymbol{b} \rangle = 0 \Longleftrightarrow \lambda = \frac{\langle \boldsymbol{x}, \boldsymbol{b} \rangle}{\langle \boldsymbol{b}, \boldsymbol{b} \rangle}$$

Projection onto One-Dimensional Subspaces (Lines)

In the last step, we exploited the fact that inner products are symmetric. If we choose $\langle \cdot, \cdot \rangle$ to be the dot product, we obtain

$$\lambda = \frac{\boldsymbol{b}^\top \boldsymbol{x}}{\boldsymbol{b}^\top \boldsymbol{b}} = \frac{\boldsymbol{b}^\top \boldsymbol{x}}{\|\boldsymbol{b}\|^2}$$

If $\|\boldsymbol{b}\| = 1$, then the coordinate λ of the projection is given by $\boldsymbol{b}^\top \boldsymbol{x}$.

Projection onto One-Dimensional Subspaces (Lines)

In the last step, we exploited the fact that inner products are symmetric. If we choose $\langle \cdot, \cdot \rangle$ to be the dot product, we obtain

$$\lambda = \frac{\boldsymbol{b}^\top \boldsymbol{x}}{\boldsymbol{b}^\top \boldsymbol{b}} = \frac{\boldsymbol{b}^\top \boldsymbol{x}}{\|\boldsymbol{b}\|^2}$$

If $\|\boldsymbol{b}\| = 1$, then the coordinate λ of the projection is given by $\boldsymbol{b}^\top \boldsymbol{x}$.

Projection onto One-Dimensional Subspaces (Lines)

Step 2. Finding the projection point $\pi_U(\boldsymbol{x}) \in U$. Since $\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b}$, we immediately obtain with (3.40) that

$$\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b} = \frac{\langle \boldsymbol{x}, \boldsymbol{b} \rangle}{\|\boldsymbol{b}\|^2} \boldsymbol{b} = \frac{\boldsymbol{b}^\top \boldsymbol{x}}{\|\boldsymbol{b}\|^2} \boldsymbol{b}$$

where the last equality holds for the dot product only. We can also compute the length of $\pi_U(\boldsymbol{x})$ by means of Definition 3.1 as

$$\|\pi_U(\boldsymbol{x})\| = \|\lambda \boldsymbol{b}\| = |\lambda| \|\boldsymbol{b}\|$$

Hence, our projection is of length $|\lambda|$ times the length of b. This also adds the intuition that λ is the coordinate of $\pi_U(\boldsymbol{x})$ with respect to the basis vector b that spans our one-dimensional subspace U.

Projection onto One-Dimensional Subspaces (Lines)

If we use the dot product as an inner product, we get

$$\|\pi_U(\boldsymbol{x})\| \overset{(3.42)}{=} \frac{|\boldsymbol{b}^\top \boldsymbol{x}|}{\|\boldsymbol{b}\|^2} \|\boldsymbol{b}\| \overset{(3.25)}{=} |\cos\omega| \|\boldsymbol{x}\| \|\boldsymbol{b}\| \frac{\|\boldsymbol{b}\|}{\|\boldsymbol{b}\|^2} = |\cos\omega| \|\boldsymbol{x}\|$$

Here, ω is the angle between $\boldsymbol{x}$ and $\boldsymbol{b}$. This equation should be familiar from trigonometry: If $\|\boldsymbol{x}\| = 1$, then $\boldsymbol{x}$ lies on the unit circle. It follows that the projection onto the horizontal axis spanned by b is exactly is a one-dimensional $\cos\omega$, and the length of the corresponding vector $\pi_U(\boldsymbol{x}) = |\cos\omega|$.

Projection onto One-Dimensional Subspaces (Lines)

Step 3. Finding the projection matrix $\boldsymbol{P}_\pi$. We know that a projection is a linear mapping (see Definition 3.10). Therefore, there exists a projection matrix $\boldsymbol{P}_\pi$, such that $\pi_U(\boldsymbol{x}) = \boldsymbol{P}_\pi \boldsymbol{x}$. With the dot product as inner product and

$$\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b} = \boldsymbol{b}\lambda = \boldsymbol{b}\frac{\boldsymbol{b}^\top \boldsymbol{x}}{\|\boldsymbol{b}\|^2} = \frac{\boldsymbol{b}\boldsymbol{b}^\top}{\|\boldsymbol{b}\|^2}\boldsymbol{x}$$

we immediately see that $\boldsymbol{P}_\pi = \frac{\boldsymbol{b}\boldsymbol{b}^\top}{\|\boldsymbol{b}\|^2}$ Note that $\boldsymbol{b}\boldsymbol{b}^\top$ (and, consequently, $\boldsymbol{P}_\pi$) is a symmetric matrix (of rank 1), and $\|\boldsymbol{b}\|^2 = \langle \boldsymbol{b}, \boldsymbol{b} \rangle$ is a scalar.

Projection onto One-Dimensional Subspaces (Lines)

Step 3. Finding the projection matrix $\boldsymbol{P}_\pi$. We know that a projection is a linear mapping (see Definition 3.10). Therefore, there exists a projection matrix $\boldsymbol{P}_\pi$, such that $\pi_U(\boldsymbol{x}) = \boldsymbol{P}_\pi \boldsymbol{x}$. With the dot product as inner product and

$$\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b} = \boldsymbol{b}\lambda = \boldsymbol{b}\frac{\boldsymbol{b}^\top \boldsymbol{x}}{\|\boldsymbol{b}\|^2} = \frac{\boldsymbol{b}\boldsymbol{b}^\top}{\|\boldsymbol{b}\|^2}\boldsymbol{x}$$

we immediately see that $\boldsymbol{P}_\pi = \frac{\boldsymbol{b}\boldsymbol{b}^\top}{\|\boldsymbol{b}\|^2}$ Note that $\boldsymbol{b}\boldsymbol{b}^\top$ (and, consequently, $\boldsymbol{P}_\pi$) is a symmetric matrix (of rank 1), and $\|\boldsymbol{b}\|^2 = \langle \boldsymbol{b}, \boldsymbol{b} \rangle$ is a scalar.

Projection onto One-Dimensional Subspaces (Lines)

Step 3. Finding the projection matrix $\boldsymbol{P}_\pi$. We know that a projection is a linear mapping (see Definition 3.10). Therefore, there exists a projection matrix $\boldsymbol{P}_\pi$, such that $\pi_U(\boldsymbol{x}) = \boldsymbol{P}_\pi \boldsymbol{x}$. With the dot product as inner product and

$$\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b} = \boldsymbol{b}\lambda = \boldsymbol{b}\frac{\boldsymbol{b}^\top \boldsymbol{x}}{\|\boldsymbol{b}\|^2} = \frac{\boldsymbol{b}\boldsymbol{b}^\top}{\|\boldsymbol{b}\|^2}\boldsymbol{x}$$

we immediately see that $\boldsymbol{P}_\pi = \frac{\boldsymbol{b}\boldsymbol{b}^\top}{\|\boldsymbol{b}\|^2}$ Note that $\boldsymbol{b}\boldsymbol{b}^\top$ (and, consequently, $\boldsymbol{P}_\pi$) is a symmetric matrix (of rank 1), and $\|\boldsymbol{b}\|^2 = \langle \boldsymbol{b}, \boldsymbol{b} \rangle$ is a scalar.

Projection onto One-Dimensional Subspaces (Lines)

The projection matrix $\boldsymbol{P}_\pi$ projects any vector $\boldsymbol{x} \in \mathbb{R}^n$ onto the line through the origin with direction $\boldsymbol{b}$ (equivalently, the subspace U spanned by $\boldsymbol{b}$).

Remark: The projection $\pi_U(\boldsymbol{x}) \in \mathbb{R}^n$ is still an n-dimensional vector and not a scalar. However, we no longer require n coordinates to represent the projection, but only a single one if we want to express it with respect to the basis vector b that spans the subspace U : λ.

Projection onto One-Dimensional Subspaces (Lines)

The projection matrix $\boldsymbol{P}_\pi$ projects any vector $\boldsymbol{x} \in \mathbb{R}^n$ onto the line through the origin with direction $\boldsymbol{b}$ (equivalently, the subspace U spanned by $\boldsymbol{b}$).

Remark: The projection $\pi_U(\boldsymbol{x}) \in \mathbb{R}^n$ is still an n-dimensional vector and not a scalar. However, we no longer require n coordinates to represent the projection, but only a single one if we want to express it with respect to the basis vector b that spans the subspace $U : \lambda$.

Projection onto General Subspaces

As in the 1D case, we follow a three-step procedure to find the projection $\pi_U(\boldsymbol{x})$ and the projection matrix $\boldsymbol{P}_\pi$:

Step 1. Find the coordinates $\lambda_1, \dots, \lambda_m$ of the projection (with respect to the basis of U), such that the linear combination

$$\pi_U(\boldsymbol{x}) = \sum_{i=1}^{m} \lambda_i \boldsymbol{b}_i = B\boldsymbol{\lambda},$$

$$B = [\boldsymbol{b}_1, \dots, \boldsymbol{b}_m] \in \mathbb{R}^{n \times m}, \quad \boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_m]^\top \in \mathbb{R}^m$$

is closest to $x \in \mathbb{R}^n$.

Projection onto General Subspaces

As in the 1D case, we follow a three-step procedure to find the projection $\pi_U(\boldsymbol{x})$ and the projection matrix $\boldsymbol{P}_\pi$:
Step 1. Find the coordinates $\lambda_1, \ldots, \lambda_m$ of the projection (with respect to the basis of U), such that the linear combination

$$\pi_U(\boldsymbol{x}) = \sum_{i=1}^{m} \lambda_i \boldsymbol{b}_i = B\boldsymbol{\lambda},$$

$$B = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m] \in \mathbb{R}^{n\times m}, \quad \boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_m]^\top \in \mathbb{R}^m$$

is closest to $x \in \mathbb{R}^n$.

Projection onto General Subspaces

As in the 1D case, "closest" means "minimum distance", which implies that the vector connecting $\pi_U(\boldsymbol{x}) \in U$ and $x \in \mathbb{R}^n$ must be orthogonal to all basis vectors of U. Therefore, we obtain m simultaneous conditions (assuming the dot product as the inner product)

$$\langle \boldsymbol{b}_1, \boldsymbol{x} - \pi_U(\boldsymbol{x}) \rangle = \boldsymbol{b}_1^\top (\boldsymbol{x} - \pi_U(\boldsymbol{x})) = 0$$
$$\vdots$$
$$\langle \boldsymbol{b}_m, \boldsymbol{x} - \pi_U(\boldsymbol{x}) \rangle = \boldsymbol{b}_m^\top (\boldsymbol{x} - \pi_U(\boldsymbol{x})) = 0$$

which, with $\pi_U(\boldsymbol{x}) = B\lambda$, can be written as

$$\boldsymbol{b}_1^\top (\boldsymbol{x} - \boldsymbol{B}\boldsymbol{\lambda}) = 0$$
$$\vdots$$
$$\boldsymbol{b}_m^\top (\boldsymbol{x} - \boldsymbol{B}\boldsymbol{\lambda}) = 0$$

such that we obtain a homogeneous linear equation system.

Projection onto General Subspaces

As in the 1D case, "closest" means "minimum distance", which implies that the vector connecting $\pi_U(\boldsymbol{x}) \in U$ and $x \in \mathbb{R}^n$ must be orthogonal to all basis vectors of U. Therefore, we obtain m simultaneous conditions (assuming the dot product as the inner product)

$$\begin{aligned} \langle \boldsymbol{b}_1, \boldsymbol{x} - \pi_U(\boldsymbol{x}) \rangle &= \boldsymbol{b}_1^\top (\boldsymbol{x} - \pi_U(\boldsymbol{x})) = 0 \\ &\vdots \\ \langle \boldsymbol{b}_m, \boldsymbol{x} - \pi_U(\boldsymbol{x}) \rangle &= \boldsymbol{b}_m^\top (\boldsymbol{x} - \pi_U(\boldsymbol{x})) = 0 \end{aligned}$$

which, with $\pi_U(\boldsymbol{x}) = B\lambda$, can be written as

$$\begin{aligned} \boldsymbol{b}_1^\top (\boldsymbol{x} - \boldsymbol{B\lambda}) &= 0 \\ &\vdots \\ \boldsymbol{b}_m^\top (\boldsymbol{x} - \boldsymbol{B\lambda}) &= 0 \end{aligned}$$

such that we obtain a homogeneous linear equation system.

Projection onto General Subspaces

As in the 1D case, "closest" means "minimum distance", which implies that the vector connecting $\pi_U(\boldsymbol{x}) \in U$ and $x \in \mathbb{R}^n$ must be orthogonal to all basis vectors of U. Therefore, we obtain m simultaneous conditions (assuming the dot product as the inner product)

$$\begin{aligned} \langle \boldsymbol{b}_1, \boldsymbol{x} - \pi_U(\boldsymbol{x}) \rangle &= \boldsymbol{b}_1^\top (\boldsymbol{x} - \pi_U(\boldsymbol{x})) = 0 \\ &\vdots \\ \langle \boldsymbol{b}_m, \boldsymbol{x} - \pi_U(\boldsymbol{x}) \rangle &= \boldsymbol{b}_m^\top (\boldsymbol{x} - \pi_U(\boldsymbol{x})) = 0 \end{aligned}$$

which, with $\pi_U(\boldsymbol{x}) = B\lambda$, can be written as

$$\begin{aligned} \boldsymbol{b}_1^\top(\boldsymbol{x} - \boldsymbol{B}\boldsymbol{\lambda}) &= 0 \\ &\vdots \\ \boldsymbol{b}_m^\top(\boldsymbol{x} - \boldsymbol{B}\boldsymbol{\lambda}) &= 0 \end{aligned}$$

such that we obtain a homogeneous linear equation system.

Projection onto General Subspaces

$$\begin{bmatrix} b_1^\top \\ \vdots \\ b_m^\top \end{bmatrix} [x - B\lambda] = 0 \iff B^\top (x - B\lambda) = 0$$

$$\iff B^\top B\lambda = B^\top x$$

The last expression is called normal equation. Since $b_1, \ldots, b_m$ are a basis of U and, therefore, linearly independent, $B^\top B \in \mathbb{R}^{m \times m}$ is regular and can be inverted. This allows us to solve for the coefficients/coordinates

$$\boldsymbol{\lambda} = \left(\boldsymbol{B}^\top \boldsymbol{B}\right)^{-1} \boldsymbol{B}^\top \boldsymbol{x}$$

Projection onto General Subspaces

The matrix $(\boldsymbol{B}^\top \boldsymbol{B})^{-1} \boldsymbol{B}^\top$ is also called the pseudo-inverse of $\boldsymbol{B}$, which can be computed for non-square matrices B. It only requires that $B^\top B$ is positive definite, which is the case if B is full rank. In practical applications (e.g., linear regression), we often add a "jitter term" ϵI to $B^\top B$ to guarantee increased numerical stability and positive definiteness. This "ridge" can be rigorously derived using Bayesian inference. See Chapter 9 for details.

Projection onto General Subspaces

The matrix $(\boldsymbol{B}^{\top}\boldsymbol{B})^{-1}\boldsymbol{B}^{\top}$ is also called the pseudo-inverse of $\boldsymbol{B}$, which can be computed for non-square matrices B. It only requires that $B^{\top}B$ is positive definite, which is the case if B is full rank. In practical applications (e.g., linear regression), we often add a "jitter term" ϵI to $B^{\top}B$ to guarantee increased numerical stability and positive definiteness. This "ridge" can be rigorously derived using Bayesian inference. See Chapter 9 for details.

Projection onto General Subspaces

The matrix $(\boldsymbol{B}^{\top}\boldsymbol{B})^{-1}\boldsymbol{B}^{\top}$ is also called the pseudo-inverse of $\boldsymbol{B}$, which can be computed for non-square matrices B. It only requires that $B^{\top}B$ is positive definite, which is the case if B is full rank. In practical applications (e.g., linear regression), we often add a "jitter term" ϵI to $B^{\top}B$ to guarantee increased numerical stability and positive definiteness. This "ridge" can be rigorously derived using Bayesian inference. See Chapter 9 for details.

Projection onto General Subspaces

Step 2. Find the projection $\pi_U(\boldsymbol{x}) \in U$. We already established that $\pi_U(\boldsymbol{x}) = B\lambda$. Therefore, with (3.57)

$$\pi_U(\boldsymbol{x}) = \boldsymbol{B}\left(\boldsymbol{B}^\top \boldsymbol{B}\right)^{-1} \boldsymbol{B}^\top \boldsymbol{x}$$

Projection onto General Subspaces

Step 3. Find the projection matrix $\boldsymbol{P}_\pi$. From (3.58), we can immediately see that the projection matrix that solves $\boldsymbol{P}_\pi \boldsymbol{x} = \pi_U(\boldsymbol{x})$ must be

$$\boldsymbol{P}_\pi = \boldsymbol{B}\left(\boldsymbol{B}^\top \boldsymbol{B}\right)^{-1} \boldsymbol{B}^\top$$

Remark: The solution for projecting onto general subspaces includes the 1D case as a special case: If $\dim(U) = 1$, then $B^\top \boldsymbol{B} \in \mathbb{R}$ is a scalar and we can rewrite the projection matrix in (3.59) $\boldsymbol{P}_\pi = \boldsymbol{B}\left(\boldsymbol{B}^\top \boldsymbol{B}\right)^{-1} \boldsymbol{B}^\top$ as $\boldsymbol{P}_\pi = \frac{B^B}{B^T B}$, which is exactly the projection matrix.

Projection onto General Subspaces

Step 3. Find the projection matrix $\boldsymbol{P}_\pi$. From (3.58), we can immediately see that the projection matrix that solves $\boldsymbol{P}_\pi \boldsymbol{x} = \pi_U(\boldsymbol{x})$ must be

$$\boldsymbol{P}_\pi = \boldsymbol{B}\left(\boldsymbol{B}^\top \boldsymbol{B}\right)^{-1} \boldsymbol{B}^\top$$

Remark: The solution for projecting onto general subspaces includes the 1D case as a special case: If $\dim(U) = 1$, then $B^\top \boldsymbol{B} \in \mathbb{R}$ is a scalar and we can rewrite the projection matrix in (3.59) $\boldsymbol{P}_\pi = \boldsymbol{B}\left(\boldsymbol{B}^\top \boldsymbol{B}\right)^{-1} \boldsymbol{B}^\top$ as $\boldsymbol{P}_\pi = \frac{B^B}{B^T B}$, which is exactly the projection matrix.

Thanks for your attention!