Mathematics for Data Science

Lecture 5

Eva FEILLET[1]

[1]LISN
Paris-Saclay University

M1[AI], Fall 2025

Previously covered topics

- (Lecture 1) Vector spaces, subspaces, linear transormations. Rank, image, kernel
- (Lecture 2) Matrices, link with linear transformations, linear systems
- Range, rank, kernel of a matrix, inverse of a matrix
- (Lecture 3) Determinant, diagonalization, eigendecomposition (part 1)
- (Lecture 4) Diagonalization, eigendecomposition (part 2).

Previously covered topics

- (Lecture 1) Vector spaces, subspaces, linear transormations. Rank, image, kernel
- (Lecture 2) Matrices, link with linear transformations, linear systems
- Range, rank, kernel of a matrix, inverse of a matrix
- (Lecture 3) Determinant, diagonalization, eigendecomposition (part 1)
- (Lecture 4) Diagonalization, eigendecomposition (part 2).

In this lecture: Normed spaces, metric spaces, inner product spaces. Orthogonal complement, orthogonal matrix (again!). Spectral theorem ; positive (semi-)definite matrices, Gram matrix ; Singular value decomposition.

Table of Contents

Table of Contents

Inner product

Definition

An **inner product** on a real vector space V is a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ satisfying, for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and all $\alpha, \beta \in \mathbb{R}$:

1. (symmetric) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$
2. (bilinear) $\langle \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z} \rangle = \alpha\langle \mathbf{x}, \mathbf{z} \rangle + \beta\langle \mathbf{y}, \mathbf{z} \rangle$
3. (positive-definite) $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$, with equality if and only if $\mathbf{x} = \mathbf{0_V}$

Vocabulary: A vector space endowed with an inner product is called an **inner product space**, or a **pre-Hilbert space** (FR: *espaces préhilbertiens*).
NB: in the above definition we have stated linearity in the first slot; with symmetry this implies linearity in the second.

Scalar product (dot product)

The usual **scalar product** $(.|.)$ defined on $\mathbb{R}^n$ is an inner product.

$$(.|.) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, (x_1, x_2, \dots x_n), (y_1, y_2, \dots y_n) \mapsto \sum_{i=1}^{n} x_i y_i$$

The inner product on $\mathbb{R}^n$ is also often written $\mathbf{x} \cdot \mathbf{y}$ (hence the alternate name **dot product**).

NB: If $x, y \in \mathbb{R}^n$, the dot product can be expressed as:

$$\langle x, y \rangle = x^\top y.$$

i.e. this inner product is a special case of matrix multiplication where we regard the resulting 1×1 matrix as a scalar.

NB: for $\alpha_1, \dots \alpha_n \in \mathbb{R}^{*+}$ (strictly positive), the map defined on $\mathbb{R}^n$ defined as follows is also an inner product.

$$(x_1, x_2, \dots x_n), (y_1, y_2, \dots y_n) \mapsto \sum_{i=1}^{n} \alpha_i x_i y_i$$

Properties of the Dot Product ($\mathbb{R}^n$)

Using the previous definition of the scalar/dot product in $\mathbb{R}^n$, prove the following properties.

Proposition (Properties of the Dot Product)

For all $x, y, z \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$:

1. *Symmetry:* $\langle x, y\rangle = \langle y, x\rangle$.
2. *Homogeneity:* $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$.
3. *Linearity in the first argument:* $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$
4. *Linearity in the second argument:* $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$.

Proof: exercise.

Another example of inner product

Example with functions

Another example of inner product

Example with functions

Example

Take the vector space of continuous real-valued functions defined over a segment $[a, b]$ of $\mathbb{R}$, $E = \mathcal{C}^0([a, b], \mathbb{R})$. Prove that the map $(.|.)$ defined on E^2 as follows is an inner product .

$$(f|g) = \int_a^b f(x)g(x)\,dx$$

Another example of inner product

Example with functions

Example

Take the vector space of continuous real-valued functions defined over a segment $[a, b]$ of $\mathbb{R}$, $E = \mathcal{C}^0([a, b], \mathbb{R})$. Prove that the map $(.|.)$ defined on E^2 as follows is an inner product .

$$(f|g) = \int_a^b f(x)g(x)\,dx$$

NB: Symmetry, bilinearity, positive-definite.
f^2 is continuous too on $[a, b]$. If $f \in \mathcal{C}^0([a, b])$ and $\int_a^b f(x)^2\,dx = 0$, then by continuity f is the null function on $[a, b]$.

Another example of inner product

Example with matrices

Another example of inner product

Example with matrices

Example

Consider the vector space of real square matrices of size n, $\mathcal{M}_n(\mathbb{R})$. We define the following application from $\mathcal{M}_n(\mathbb{R}) \times \mathcal{M}_n(\mathbb{R})$ to $\mathbb{R}$: $(A, B) \mapsto \mathrm{trace}(A^\top B)$. Prove that this application is an inner product on $\mathcal{M}_n(\mathbb{R})$.

Another example of inner product

Example with matrices

Example

Consider the vector space of real square matrices of size n, $\mathcal{M}_n(\mathbb{R})$. We define the following application from $\mathcal{M}_n(\mathbb{R}) \times \mathcal{M}_n(\mathbb{R})$ to $\mathbb{R}$:
$(A, B) \mapsto \mathrm{trace}(A^\top B)$. Prove that this application is an inner product on $\mathcal{M}_n(\mathbb{R})$.
NB: again prove symmetry, bilinearity, positive-definiteness.

Orthogonal vectors

Definition

Two vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be **orthogonal** if their inner product is zero

$$\langle \mathbf{x}, \mathbf{y} \rangle = 0$$

Notation : we can write $\mathbf{x} \perp \mathbf{y}$ for short.

NB: **Orthogonality** generalizes the notion of **perpendicularity** we use in 2D.

Writing the dot product using matrix notation

Let $x, y \in \mathbb{R}^n$. The dot product can be expressed as:

$$\langle x, y \rangle = x^\top y.$$

This is a special case of matrix multiplication where we regard the resulting 1×1 matrix as a scalar. Example: if $x = (x_1, x_2, x_3)^\top$ and $y = (y_1, y_2, y_3)^\top$, then

$$\langle x, y \rangle = x^\top y = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = x_1 y_1 + x_2 y_2 + x_3 y_3.$$

Example

Let $A \in \mathcal{M}_n(\mathbb{R})$ a square matrix and $x, y \in \mathbb{R}^n$. Show that $\langle Ax, y \rangle = \langle x, A^\top y \rangle$.

Writing the dot product using matrix notation

Let $x, y \in \mathbb{R}^n$. The dot product can be expressed as:

$$\langle x, y \rangle = x^\top y.$$

This is a special case of matrix multiplication where we regard the resulting 1×1 matrix as a scalar. Example: if $x = (x_1, x_2, x_3)^\top$ and $y = (y_1, y_2, y_3)^\top$, then

$$\langle x, y \rangle = x^\top y = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = x_1 y_1 + x_2 y_2 + x_3 y_3.$$

Example

Let $A \in \mathcal{M}_n(\mathbb{R})$ a square matrix and $x, y \in \mathbb{R}^n$. Show that $\langle Ax, y \rangle = \langle x, A^\top y \rangle$. $\langle Ax, y \rangle = (Ax)^\top y = x^\top A^\top y = \langle x, A^\top y \rangle$.

Table of Contents

Metric spaces

A **metric space** is a set together with a notion of **distance** between its elements, usually called **points**. The distance is measured by a function called a **metric** or **distance** function.

Definition (Metric)

A **metric** on a set S is a function $d : S \times S \to \mathbb{R}^+$ that satisfies, for all $x, y, z \in S$:

1. (Positivity) $d(x, y) \geq 0$, with equality if and only if $x = y$
2. (Symmetry) $d(x, y) = d(y, x)$
3. (Triangle inequality) $d(x, z) \leq d(x, y) + d(y, z)$

Metric spaces

A **metric space** is a set together with a notion of **distance** between its elements, usually called **points**. The distance is measured by a function called a **metric** or **distance** function.

Definition (Metric)

A **metric** on a set S is a function $d : S \times S \to \mathbb{R}^+$ that satisfies, for all $x, y, z \in S$:

1. (Positivity) $d(x, y) \geq 0$, with equality if and only if $x = y$
2. (Symmetry) $d(x, y) = d(y, x)$
3. (Triangle ineguality) $d(x, z) \leq d(x, y) + d(y, z)$

Examples of metric spaces ?

Metric spaces - Examples

Example

- The real numbers with the distance function $d(x, y) = |y - x|$ given by the absolute difference between two real numbers
- For a given $n > 0$, the **Hamming distance** is a metric on the set of the words of length n, e.g. the set of 100-character Unicode strings can be equipped with the Hamming distance, which measures the number of characters that need to be changed to get from one string to another.

NB: The Hamming distance actually comes from information theory, where it is used to count the minimum number of errors that could have transformed one string into the other.

Metric spaces - Examples

Example

The Euclidean plane $\mathbb{R}^2$ can be equipped with many different metrics:

- The usual Euclidean distance from high school :
  $d_2((x_1, y_1), (x_2, y_2)) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$
- The Manhattan (or "taxi-cab") distance :
  $d_1((x_1, y_1), (x_2, y_2)) = |x_2 - x_1| + |y_2 - y_1|$
- The maximum distance :
  $d_\infty((x_1, y_1), (x_2, y_2)) = \max\{|x_2 - x_1|, |y_2 - y_1|\}$
- A discrete metric like the following : $d(p, q) = \begin{cases} 0, & \text{if } p = q, \\ 1, & \text{otherwise.} \end{cases}$

NB: More generally, we call **Euclidean space** a vector space defined over $\mathbb{R}$, that has a finite dimension and is endowed with an inner product.

Remarks

- Two different metrics will give us two different "measures" of distances.
- Not all metric spaces are vector spaces !
- A key motivation for metrics is that they allow limits to be defined for mathematical objects other than real numbers.
  e.g. we say that a sequence $\{x_n\} \subseteq S$ converges to the limit x if for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $d(x_n, x) < \epsilon$ for all $n \geq N$.

Table of Contents

Normed spaces

Some (but not all !) metric spaces are **normed spaces**. They are defined on vector spaces and endowed with a **norm** function.

Definition (Norm)

A **norm** on a real vector space V is a function $\|\cdot\| : V \to \mathbb{R}^+$ that satisfies

1. $\|\mathbf{x}\| \geq 0$, with equality if and only if $\mathbf{x} = \mathbf{0}_V$
2. $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (the **triangle inequality** again)

for all $\mathbf{x}, \mathbf{y} \in V$ and all $\alpha \in \mathbb{R}$.

A vector space endowed with a norm is called a **normed vector space**, or simply a **normed space**.

Norm induced by an inner product

Remark

Any inner product on V induces a norm on V:

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

NB: the axioms for norms are satisfied under this definition and follow from the axioms for inner products. Therefore any inner product space is also a normed space (and hence also a metric space).

Norms are useful to measure distances :

$$d(x, y) = \|\mathbf{x} - \mathbf{y}\| = \|\mathbf{y} - \mathbf{x}\|$$

Two different norms will give us two different "measures" of distances.

Norm induced by the dot product on $\mathbb{R}^n$

Example

Verify that the two-norm $\|\cdot\|_2$ (Euclidean norm) on $\mathbb{R}^2$ is induced by the dot product.

$$\langle x, x \rangle = \|x\|_2^2.$$

Examples of norms on $\mathbb{R}^n$

$$L_1 : \|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

$$L_2 : \|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \quad \text{(Euclidean norm)}$$

$$L_p : \|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \quad (p \geq 1)$$

$$L_\infty : \|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|$$

NB: The 1- and 2-norms are special cases of the p-norm, and the ∞-norm is the limit of the p-norm as p tends to infinity.

We require $p \geq 1$ because the triangle inequality fails to hold for $0 < p < 1$.

Geometric interpretation of norms in $\mathbb{R}^2$

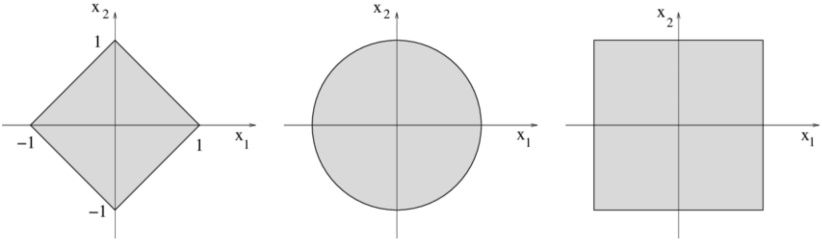

Figure 1: Unit balls for different norms in $\mathbb{R}^2$

For each subfigure, give the corresponding norm.

Geometric interpretation of norms in $\mathbb{R}^2$

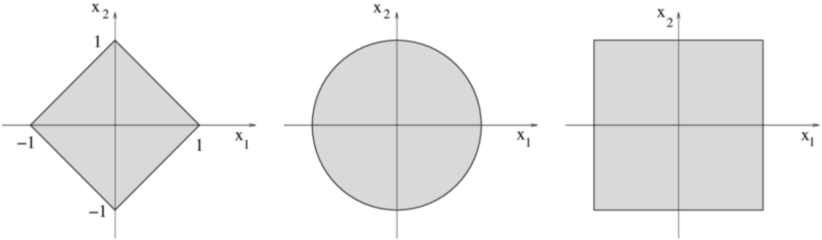

Figure 1: Unit balls for different norms in $\mathbb{R}^2$

For each subfigure, give the corresponding norm. The unit ball for the 2-norm is a circle, for the 1-norm a square rotated by 45 degrees, and for the infinity norm a square aligned with the axes.

Geometric interpretation of norms in $\mathbb{R}^2$

For the angle θ between x and y (vectors different from the null vector):

$$\langle x, y\rangle = \|x\|_2 \, \|y\|_2 \cos\theta.$$

Reminder : If $\langle x, y\rangle = 0$, the vectors x and y are said to be *orthogonal*.

Example

Let $v_1, v_2, \dots v_p \in \mathbb{R}^n$ be p orthogonal vectors. Show that they are linearly independent.

Geometric interpretation of norms in $\mathbb{R}^2$

For the angle θ between x and y (vectors different from the null vector):

$$\langle x, y \rangle = \|x\|_2 \, \|y\|_2 \cos\theta.$$

Reminder : If $\langle x, y \rangle = 0$, the vectors x and y are said to be *orthogonal*.

Example

Let $v_1, v_2, \ldots v_p \in \mathbb{R}^n$ be p orthogonal vectors. Show that they are linearly independent. Consider a linear combination equal to zero: $\sum_{i=1}^{p} \alpha_i v_i = 0$. Taking the inner product with $v_j, 1 \leq j \leq p$ gives $\alpha_j \|v_j\|^2 = 0$, hence $\alpha_j = 0$ for all j. So the vectors are linearly independent (and $p \leq n$).

Pythagorean Theorem

The Pythagorean theorem generalizes to arbitrary inner product spaces.

Theorem (Pythagorean Theorem)

Let V a finite-dimensional inner-product vector space, and $x, y \in V$. If $\mathbf{x} \perp \mathbf{y}$, then

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$$

Proof:

Pythagorean Theorem

The Pythagorean theorem generalizes to arbitrary inner product spaces.

Theorem (Pythagorean Theorem)

Let V a finite-dimensional inner-product vector space, and $x, y \in V$. If $\mathbf{x} \perp \mathbf{y}$, *then*

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$$

Proof: Suppose $\mathbf{x} \perp \mathbf{y}$, i.e. $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. It follows:

$$\|\mathbf{x} + \mathbf{y}\|^2 = \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$$

Cauchy-Schwarz inequality

Proposition

Let V be an inner product space. For all $\mathbf{x}, \mathbf{y} \in V$,

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$$

NB: this equality holds exactly iff $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent (i.e. are scalar multiples of each other, including the case when at least one of them is null).

Proof: exercise.

Cauchy-Schwarz inequality

Proposition

Let V be an inner product space. For all $\mathbf{x}, \mathbf{y} \in V$,

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$$

NB: this equality holds exactly iff $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent (i.e. are scalar multiples of each other, including the case when at least one of them is null).

Proof: exercise. Let $\mathbf{x}, \mathbf{y}$ be two vectors in V. If $\mathbf{y} = 0$, the result is trivial. Otherwise, consider the vector $\mathbf{x} - \frac{\langle \mathbf{x},\mathbf{y} \rangle}{\|\mathbf{y}\|^2}\mathbf{y}$. Since $\langle \mathbf{x} - \frac{\langle \mathbf{x},\mathbf{y} \rangle}{\|\mathbf{y}\|^2}\mathbf{y}, \mathbf{x} - \frac{\langle \mathbf{x},\mathbf{y} \rangle}{\|\mathbf{y}\|^2}\mathbf{y} \rangle \geq 0$, we have

$$\|\mathbf{x}\|^2 - 2\frac{\langle \mathbf{x}, \mathbf{y} \rangle^2}{\|\mathbf{y}\|^2} + \frac{\langle \mathbf{x}, \mathbf{y} \rangle^2}{\|\mathbf{y}\|^4}\|\mathbf{y}\|^2 \geq 0$$

and the result follows.

Equivalence of norms in a finite-dimensional vector space (bonus)

For any given finite-dimensional vector space V, all norms on V are equivalent in the sense that for all $\mathbf{x} \in V$, for two norms $\|\cdot\|_A$ and $\|\cdot\|_B$, there exist constants $\alpha, \beta > 0$ such that

$$\alpha\|\mathbf{x}\|_A \leq \|\mathbf{x}\|_B \leq \beta\|\mathbf{x}\|_A$$

NB: the constants depend on the norms, not on the vector. **Therefore convergence in one norm implies convergence in any other norm.**

This is not a general property e.g. may not apply in infinite-dimensional vector spaces such as function spaces!

"Entry-wise" matrix norms

"Entry-wise" matrix norms treat an $m \times n$ matrix as a vector of size $m \cdot n$, and use one of the familiar vector norms we have seen before. For example, using the L_p-norm for vectors:

$$\|A\|_{p,p} = \|\mathrm{vec}(A)\|_p = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^p \right)^{1/p}$$

This is a different norm from the induced p-norm (see after)

Example

The special case $p = 2$ is the **Frobenius norm**.

$$\|A\|_{\mathrm{F}} = \sqrt{\sum_{i}^{m} \sum_{j}^{n} |a_{ij}|^2}$$

Matrix norms induced by vector norms

Definition (matrix norm induced by a vector norm)

Given a vector norm $\|\cdot\|$ on $\mathbb{R}^n$, a *matrix norm* can be defined by:

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}.$$

Consider a vector norm $\|\cdot\|_\alpha$ on $\mathbb{R}^n$ and a vector norm $\|\cdot\|_\beta$ on $\mathbb{R}^m$. Recall that any $m \times n$ matrix A induces a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ with respect to the standard basis. The corresponding **induced norm** on the space $\mathbb{R}^{m \times n}$ of all $m \times n$ matrices can be defined as follows:

$$\|A\|_{\alpha,\beta} = \sup\{\|Ax\|_\beta : x \in \mathbb{R}^n \text{ such that } \|x\|_\alpha \leq 1\}$$

where sup denotes the supremum.

NB: This norm can be interpreted as measuring how much the mapping induced by A “stretches” vectors.

Matrix norms induced by vector norms (bonus)

If the L_p-norm for vectors ($1 \leq p$) is used for both spaces $\mathbb{R}^n$ and $\mathbb{R}^m$ then the corresponding induced norm is:

$$\|A\|_p = \sup\{\|Ax\|_p : x \in \mathbb{R}^n \text{ such that } \|x\|_p \leq 1\}$$

Example

- Case $p = 1$: the maximum absolute *column sum* of the matrix

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |a_{ij}|$$

- Case $p = \infty$: the maximum absolute *row sum* of the matrix.

$$\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}|$$

Matrix norms induced by vector norms (bonus)

Example

- Case $p = 2$: the **spectral norm**. Not to be confounded with the **Frobenius norm** !

$$\|A\|_2 = \sqrt{\lambda_{\text{max}}\left(A^* A\right)}$$

$$\|A\|_2 = \sigma_{\max}(A) \leq \|A\|_{\mathrm{F}}$$

NB: In the above example, A^* denotes the conjugate transpose of A and $\sigma_{\text{max}}(A)$ the highest singular value of A (see section on SVD).

Table of Contents

Orthogonal basis

Definition

A set of vectors $\{e_1, \ldots, e_n\}$ from a finite-dimensional inner-product space E is an *orthogonal basis* of E if it is a basis and $\langle e_i, e_j \rangle = 0$ whenever $i \neq j$.

Theorem

Every finite-dimensional inner product space admits an orthonormal basis.

NB: *orthogonal* vs *orthonormal*.

Orthogonal complement

Definition

Let F be a subspace of a vector space E endowed with an inner product. The *orthogonal complement* of F, denoted $F^\perp$, is:

$$F^\perp = \{x \in E \mid \langle x, v \rangle = 0 \text{ for all } v \in F\}.$$

Example

In $\mathbb{R}^3$, what is the orthogonal complement of the subspace spanned by the vector $(1, 1, 0)$?

Orthogonal complement

Definition

Let F be a subspace of a vector space E endowed with an inner product. The *orthogonal complement* of F, denoted $F^\perp$, is:

$$F^\perp = \{x \in E \mid \langle x, v\rangle = 0 \text{ for all } v \in F\}.$$

Example

In $\mathbb{R}^3$, what is the orthogonal complement of the subspace spanned by the vector $(1,1,0)$? It is the plane defined by the equation $x + y = 0$, i.e. the set of vectors (x, y, z) such that $x + y = 0$. So
$span(1,1,0)^\perp = span((1,-1,0),(0,0,1))$.

Orthogonal complement

Proposition

Let F be a subspace of a finite-dimensional inner product space E. Then:

1. $F^{\perp}$ *is a subspace of* E.
2. $F \cap F^{\perp} = \{\mathbf{0}_E\}$.
3. $\dim(F) + \dim(F^{\perp}) = \dim(E)$.
4. $(F^{\perp})^{\perp} = F$.

Proof: exercise session.

Link with linear maps: Projectors

Definition (Projector)

A linear map $p \in \mathcal{L}(E)$ is called a **projector** if $p \circ p = p$.

Definition (Orthogonal projection)

Let E be a euclidean vector space, and F a subspace of E. We call **orthogonal projection** onto F the linear map $p_F : E \to F$ such that for all $\mathbf{x} \in E$, $p_F(\mathbf{x})$ is the unique vector in F satisfying:

$$\mathbf{x} - p_F(\mathbf{x}) \in F^\perp$$

NB: $p_F(\mathbf{x})$ is called the **projection** of $\mathbf{x}$ onto F.
You can check that for all $\mathbf{x} \in E$,

$$p_F(p_F(\mathbf{x})) = p_F(\mathbf{x})$$

Properties of projectors

Proposition (Bessel inequality)

For all $\mathbf{x} \in E$,

$$\|p_F(\mathbf{x})\| \leq \|\mathbf{x}\|$$

with equality if and only if $\mathbf{x} \in F$.

Proof: exercise.

Properties of projectors

Proposition (Bessel inequality)

For all $\mathbf{x} \in E$,

$$\|p_F(\mathbf{x})\| \leq \|\mathbf{x}\|$$

with equality if and only if $\mathbf{x} \in F$.

Proof: exercise. Let $x \in E$. We can write

$$x = p_F(x) + (x - p_F(x))$$

with $p_F(x) \in F$ and $x - p_F(x) \in F^\perp$. By the Pythagorean theorem, we have

$$\|x\|^2 = \|p_F(x)\|^2 + \|x - p_F(x)\|^2 \geq \|p_F(x)\|^2$$

with equality if and only if $x - p_F(x) = 0$, i.e. $x \in F$.

Properties of projectors

Proposition

Let $p \in \mathcal{L}(E)$ a projector (i.e. $p \circ p = p$). The following are equivalent:

1. *p is an orthogonal projector onto $F = range(p)$ and $Ker(p) = F^{\perp}$*
2. *For all $\mathbf{x}, \mathbf{y} \in E$, $\langle p(\mathbf{x}), y - p(y) \rangle = 0$*
3. *For all $\mathbf{x}, \mathbf{y} \in E$, $\langle p(\mathbf{x}), y \rangle = \langle \mathbf{x}, p(y) \rangle$*

Proof: exercise.

Projection onto a subspace (via an orthonormal basis)

Proposition

Let E be a euclidean vector space, F a subspace of E and $\{e_1, \ldots, e_k\}$ an orthonormal basis of F. For all $\mathbf{x} \in E$, the orthogonal projection of $\mathbf{x}$ onto F is given by:

$$p_F(\mathbf{x}) = \sum_{i=1}^{k} \langle \mathbf{x}, e_i \rangle e_i$$

Distance to a subspace

Definition

Let E be a euclidean vector space, and F a subspace of E. The **distance** from a vector $\mathbf{x} \in E$ to the subspace F is defined as:

$$d(\mathbf{x}, F) = \inf_{\mathbf{y} \in F} \|\mathbf{x} - \mathbf{y}\|$$

Proposition

Let E be a euclidean vector space, F a subspace of E and p_F the orthogonal projection onto F. Then for all $\mathbf{x} \in E$, $p_F(\mathbf{x})$ is the unique vector in F satisfying:

$$d(\mathbf{x}, F) = \|\mathbf{x} - p_F(\mathbf{x})\|$$

.

Orthogonal endomorphism

Definition

An endomorphism $f \in \mathcal{L}(E)$ is said to be **orthogonal** if it preserves the inner product, i.e. for all $\mathbf{x}, \mathbf{y} \in E$,

$$\langle f(\mathbf{x}), f(\mathbf{y}) \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$$

Proposition

An endomorphism $f \in \mathcal{L}(E)$ is orthogonal if and only if it preserves the norm, i.e. for all $\mathbf{x} \in E$,

$$\|f(\mathbf{x})\| = \|\mathbf{x}\|$$

Proof: exercise.

Orthogonal endomorphism

Proof: orthogonal $\implies$ norm-preserving:

$$\|f(\mathbf{x})\|^2 = \langle f(\mathbf{x}), f(\mathbf{x}) \rangle = \langle \mathbf{x}, \mathbf{x} \rangle = \|\mathbf{x}\|^2$$

Orthogonal endomorphism

Proof: orthogonal $\implies$ norm-preserving:

$$\|f(\mathbf{x})\|^2 = \langle f(\mathbf{x}), f(\mathbf{x})\rangle = \langle \mathbf{x}, \mathbf{x}\rangle = \|\mathbf{x}\|^2$$

norm-preserving $\implies$ orthogonal: Remark

$$4\langle \mathbf{x}, \mathbf{y}\rangle = \|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2 = \|f(\mathbf{x}+\mathbf{y})\|^2 - \|f(\mathbf{x}-\mathbf{y})\|^2$$

$$= \|f(\mathbf{x}) + f(\mathbf{y})\|^2 - \|f(\mathbf{x}) - f(\mathbf{y})\|^2 = 4\langle f(\mathbf{x}), f(\mathbf{y})\rangle$$

Link with orthogonal matrices

Definition

Let $Q \in \mathcal{M}_n(\mathbb{R})$. Q is called an orthogonal matrix if the associated linear map $f_Q : \mathbb{R}^n \to \mathbb{R}^n$ defined by $f_Q(\mathbf{x}) = Q\mathbf{x}$ is an orthogonal endomorphism.

The columns of an orthogonal matrix Q are pairwise orthonormal. They form an orthonormal basis of $\mathbb{R}^n$. This definition also implies that

$$Q^\top Q = QQ^\top = I_n$$

or equivalently, $Q^\top = Q^{-1}$.
Orthogonal matrices preserve inner products:

$$(Q\mathbf{x})^\top (Q\mathbf{y}) = \mathbf{x}^\top Q^\top Q\mathbf{y} = \mathbf{x}^\top \mathbf{I}_n \mathbf{y} = \mathbf{x}^\top \mathbf{y}$$

Table of Contents

Symmetric endomorphisms and matrices

Definition

Let E be a euclidean vector space. An endomorphism $f \in \mathcal{L}(E)$ is said to be **symmetric** if for all $\mathbf{x}, \mathbf{y} \in E$,

$$\langle f(\mathbf{x}), \mathbf{y} \rangle = \langle \mathbf{x}, f(\mathbf{y}) \rangle$$

NB: If $E = \mathbb{R}^n$ with the standard inner product, then f is symmetric if and only if its matrix in the standard basis is a symmetric matrix, i.e. $A^\top = A$.

Example

We have seen that the orthogonal projection onto a given subspace is a symmetric endomorphism.

Spectral theorem

Theorem (Spectral Theorem for endomorphisms)

Let E be a finite-dimensional euclidean vector space, and $f \in \mathcal{L}(E)$ a symmetric endomorphism. Then there exists an orthonormal basis of E consisting of eigenvectors of f.

Theorem (Spectral Theorem for matrices)

If A is a symmetric $n \times n$ matrix, then A is **orthogonally diagonalizable**, *i.e. there exists an orthogonal matrix V and a diagonal matrix D such that:*

$$A = VDV^T,$$

where the diagonal entries of D are the n real eigenvalues of A.

Practical application: any symmetric matrix is diagonalizable in an orthonormal basis, a decomposition called **spectral decomposition**.

Spectral theorem

Proof: exercise.

Spectral theorem

Proof: exercise. Let E be a finite-dimensional euclidean vector space, and $f \in \mathcal{L}(E)$ a symmetric endomorphism. We want to prove that there exists an orthonormal basis of E consisting of eigenvectors of f.

The proof proceeds in three steps:

- First, we prove that the eigenspectrum of a symmetric endomorphism contains at least one eigenvalue that is real.

$$sp(f)_{\mathbb{R}} \neq \emptyset$$

- Then, we prove that if F is a stable subspace of E under f, then its orthogonal complement $F^{\perp}$ is also stable under f.
- Finally, we prove the theorem by induction on n.

Rayleigh quotients

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a symmetric matrix.

Definition

The expression $\mathbf{x}^\top \mathbf{A}\mathbf{x}$ is called a **quadratic form**.
The following quantity is called a **Rayleigh quotient**:

$$R_{\mathbf{A}}(\mathbf{x}) = \frac{\mathbf{x}^\top \mathbf{A}\mathbf{x}}{\mathbf{x}^\top \mathbf{x}}$$

NB: the Rayleigh quotient is defined for all $\mathbf{x} \neq \mathbf{0}$.

Proposition

(i) **Scale invariance**: *for any vector* $\mathbf{x} \neq \mathbf{0}$ *and any scalar* $\alpha \neq 0$, $R_{\mathbf{A}}(\mathbf{x}) = R_{\mathbf{A}}(\alpha\mathbf{x})$.

(ii) *If* $\mathbf{x}$ *is an eigenvector of* $\mathbf{A}$ *with eigenvalue* λ, *then* $R_{\mathbf{A}}(\mathbf{x}) = \lambda$.

Rayleigh quotients

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a symmetric matrix.

Proposition

For any $\mathbf{x}$ *such that* $\|\mathbf{x}\|_2 = 1$,

$$\lambda_{\min}(\mathbf{A}) \leq \mathbf{x}^\top \mathbf{A}\mathbf{x} \leq \lambda_{\max}(\mathbf{A})$$

with equality if and only if $\mathbf{x}$ *is a corresponding eigenvector.*

Corollary: since $\mathbf{x}^\top \mathbf{A}\mathbf{x} = R_{\mathbf{A}}(\mathbf{x})$ for unit $\mathbf{x}$, the Rayleigh quotient is bounded by the largest and smallest eigenvalues of $\mathbf{A}$.

Theorem (Min-max theorem)

For all $\mathbf{x} \neq \mathbf{0}$,

$$\lambda_{\min}(\mathbf{A}) \leq R_{\mathbf{A}}(\mathbf{x}) \leq \lambda_{\max}(\mathbf{A})$$

with equality if and only if $\mathbf{x}$ *is a corresponding eigenvector.*

Positive (semi-)definite matrices

Definition

An $n \times n$ matrix A is *positive definite* if:

1. A is symmetric ($A^T = A$),
2. For all $x \in \mathbb{R}^n \setminus \{0\}$, $x^T A x > 0$.

It is *positive semi-definite* if A is symmetric and $x^T A x \geq 0$ for all x.

NB: Similarly, A is *negative definite* if $x^T A x < 0$ for all nonzero x.

Link with eigenvalues

Proposition

A symmetric matrix is

- *positive semi-definite if and only if all of its eigenvalues are nonnegative,*
- *positive definite if and only if all of its eigenvalues are positive.*

Positive definite matrices are invertible (since their eigenvalues are nonzero), whereas positive semi-definite matrices might not be.
NB: if you already have a positive semi-definite matrix, it is possible to perturb its diagonal slightly to get a positive definite matrix. e.g. with $\mathbf{A}$ a positive semi-definite and $\epsilon > 0$, $\mathbf{A} + \epsilon \mathbf{I}_n$ is positive definite, as for any $\mathbf{x} \neq \mathbf{0}$ we have:

$$\mathbf{x}^\top(\mathbf{A} + \epsilon \mathbf{I}_n)\mathbf{x} = \mathbf{x}^\top \mathbf{A}\mathbf{x} + \epsilon \mathbf{x}^\top \mathbf{I}_n \mathbf{x} = \underbrace{\mathbf{x}^\top \mathbf{A}\mathbf{x}}_{\geq 0} + \underbrace{\epsilon \|\mathbf{x}\|_2^2}_{>0} > 0$$

Example: the Gram Matrix

Definition (Gram Matrix)

If A is an $m \times n$ matrix, its *Gram matrix* is:

$$G = A^T A.$$

Proposition

The Gram matrix is symmetric and positive semi-definite.

Proof: exercise.

Example: the Gram Matrix

Definition (Gram Matrix)

If A is an $m \times n$ matrix, its *Gram matrix* is:

$$G = A^T A.$$

Proposition

The Gram matrix is symmetric and positive semi-definite.

Proof: exercise. We have, $G^T = (A^T A)^T = A^T (A^T)^T = A^T A = G$. Moreover, for any $\mathbf{x} \neq \mathbf{0}$,

$$\mathbf{x}^\top G\mathbf{x} = \mathbf{x}^\top A^\top A\mathbf{x} = (A\mathbf{x})^\top (A\mathbf{x}) = \|A\mathbf{x}\|_2^2 \geq 0$$

with equality if and only if $A\mathbf{x} = \mathbf{0}$, i.e. $\mathbf{x} \in \mathrm{Ker}(A)$.

Following the previous remark, the matrix $\mathbf{A}^\top\mathbf{A} + \epsilon\mathbf{I}_n$ is positive definite (and in particular, invertible) for *any* matrix $\mathbf{A}$ and any $\epsilon > 0$. It is often used in practice in data science.

Polar decomposition (bonus)

Theorem (Polar Decomposition)

For any $n \times n$ matrix A, there exists an orthogonal matrix R and a symmetric positive semi-definite matrix F such that:

$$A = RF.$$

One can take $F = (A^T A)^{1/2}$.

NB: If A is singular (non invertible), R is not unique; if A is invertible, R is unique.

Table of Contents

Singular Value Decomposition

Theorem (Singular Value Decomposition)

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$*. Then there exist orthogonal matrices* $\mathbf{U} \in \mathbb{R}^{m\times m}$ *and* $\mathbf{V} \in \mathbb{R}^{n\times n}$*, and a diagonal matrix* $\mathbf{\Sigma} \in \mathbb{R}^{m\times n}$ *with nonnegative diagonal entries* $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(m,n)} \geq 0$ *such that*

$$\mathbf{A} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^\top.$$

The diagonal entries (σ_i) *are the* **singular values** *of* $\mathbf{A}$*. The columns of* $\mathbf{U}$ *(resp.* $\mathbf{V}$*) are the* **left** *(resp.* **right***) singular vectors of* $\mathbf{A}$*.*

NB: *every matrix* $\mathbf{A} \in \mathbb{R}^{m\times n}$ has an SVD (even non-square matrices)!
Shapes: $\mathbf{U} = [u_1 \ \ldots \ u_m]$, $\mathbf{V} = [v_1 \ \ldots \ v_n]$ orthogonal,

$\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots) \in \mathbb{R}^{m\times n}$ with $r = \mathrm{rank}(\mathbf{A})$.

Remarks on SVD

By convention, the singular values are given in non-increasing order, i.e.

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(m,n)} \geq 0$$

Only the first r singular values are nonzero, where r is the rank of $\mathbf{A}$. We can write the rank-r part as $\mathbf{A} = \sum_{i=1}^{r} \sigma_i \, u_i v_i^\top$. Remark that the SVD factors provide eigendecompositions for $\mathbf{A}^\top\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\top$:

$$\mathbf{A}^\top\mathbf{A} = (\mathbf{U\Sigma V}^\top)^\top\mathbf{U\Sigma V}^\top = \mathbf{V\Sigma}^\top\mathbf{U}^\top\mathbf{U\Sigma V}^\top = \mathbf{V\Sigma}^\top\mathbf{\Sigma V}^\top$$
$$\mathbf{A}\mathbf{A}^\top = \mathbf{U\Sigma V}^\top(\mathbf{U\Sigma V}^\top)^\top = \mathbf{U\Sigma V}^\top\mathbf{V\Sigma}^\top\mathbf{U}^\top = \mathbf{U\Sigma\Sigma}^\top\mathbf{U}^\top$$

Hence the columns of $\mathbf{V}$ (the **right-singular vectors** of $\mathbf{A}$) are eigenvectors of $\mathbf{A}^\top\mathbf{A}$, and the columns of $\mathbf{U}$ (the **left-singular vectors** of $\mathbf{A}$) are eigenvectors of $\mathbf{A}\mathbf{A}^\top$.

SVD — Geometric intuition

- Think of $\mathbf{A}$ as a three-step transform:

$$\mathbf{A} = \underbrace{\mathbf{U}}_{\text{rotate/reflect in } \mathbb{R}^m} \underbrace{\boldsymbol{\Sigma}}_{\text{axis-wise stretch}} \underbrace{\mathbf{V}^\top}_{\text{rotate/reflect in } \mathbb{R}^n}.$$

- The unit sphere in $\mathbb{R}^n$ is mapped by $\mathbf{V}^\top$ to itself (rotation/reflection), then by $\boldsymbol{\Sigma}$ to an axis-aligned ellipsoid (semi-axes σ_i), then by $\mathbf{U}$ to a rotated ellipsoid in $\mathbb{R}^m$.
- Right singular vectors v_i give the principal input directions, left singular vectors u_i give the principal output directions; σ_i are the semi-axis lengths (gains).
- For $k < r$, truncating after k terms keeps the k *most amplifying* directions.

SVD — Proof sketch (via spectral theorem)

Idea. Use that $\mathbf{A}^\top\mathbf{A}$ is symmetric positive semidefinite.

1. $\mathbf{A}^\top\mathbf{A}$ is symmetric $\Rightarrow$ by the spectral theorem, there exists an orthonormal basis $\{v_1, \dots, v_n\}$ of eigenvectors with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. Set $\sigma_i = \sqrt{\lambda_i}$ and $\mathbf{V} = [v_1 \ \dots \ v_n]$.
2. Define $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \dots, \sigma_{\min(m,n)})$ (embedded in $\mathbb{R}^{m\times n}$ by padding zeros).
3. For each i with $\sigma_i > 0$, set
$$u_i = \frac{\mathbf{A}v_i}{\sigma_i} \in \mathbb{R}^m.$$
Then $\|u_i\|_2 = 1$ and $\langle u_i, u_j\rangle = \delta_{ij}$ for $\sigma_i, \sigma_j > 0$. Complete to an orthonormal basis $\{u_1, \dots, u_m\}$ of $\mathbb{R}^m$ and set $\mathbf{U} = [u_1 \ \dots \ u_m]$.
4. By construction, $\mathbf{A}v_i = \sigma_i u_i$ for all i with $\sigma_i > 0$, and $\mathbf{A}v_i = \mathbf{0}$ when $\sigma_i = 0$. Hence $\mathbf{AV} = \mathbf{U\Sigma}$, i.e.
$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top.$$

Applications of the SVD

- **Low-rank approximation.** The SVD provides the best low-rank approximations to a matrix in both the Frobenius norm and the 2-norm.
- **Pseudoinverse.** The SVD provides a way to compute the Moore-Penrose pseudoinverse of a matrix.
- **Principal component analysis (PCA).** PCA is a technique for dimensionality reduction that uses the SVD to find a low-dimensional representation of data that captures as much variance as possible.

Next class

Class 6 : Calculus

Exam : 24th of October, 2h. From 2pm to 4pm.

Warning

Check the timetable! Room is different from the usual one. (D101)