Optimization and Computational Linear Algebra for Data Science
Lecture 4: Norm and inner product

Léo Miolane · leo.miolane@gmail.com

September 21, 2020

Warning: *This material is not meant to be lecture notes. It only gathers the main concepts and results from the lecture, without any additional explanation, motivation, examples, figures...*

1 Norm

Definition 1.1 (*Norm*)

Let V be a vector space. A norm $\|\cdot\|$ on V is a function from V to $\mathbb{R}_{\geq 0}$ that verifies the following points:

(i) Homogeneity*: $\|\alpha v\| = |\alpha| \times \|v\|$ for all $\alpha \in \mathbb{R}$ and $v \in V$.*

(ii) Triangular inequality*: $\|u + v\| \leq \|u\| + \|v\|$ for all $u, v \in V$.*

(iii) Positive definiteness*: if $\|v\| = 0$ for some $v \in V$, then $v = 0$.*

Example 1.1. One can consider various norms over $\mathbb{R}^n$:

- The Euclidean norm $\|x\|_2 \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^n x_i^2}$.
- The ℓ_1 norm $\|x\|_1 \stackrel{\text{def}}{=} \sum_{i=1}^n |x_i|$.
- More generally, given $p \geq 1$, the ℓ_p-norm $\|x\|_p \stackrel{\text{def}}{=} \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$.
- The infinity-norm $\|x\|_\infty \stackrel{\text{def}}{=} \max(|x_1|, \dots, |x_n|)$.

2 Inner product

Definition 2.1 (*Inner product*)

Let V be a vector space. An inner product on V is a function $\langle \cdot, \cdot \rangle$ from $V \times V$ to $\mathbb{R}$ that verifies the following points:

(i) Symmetry*: $\langle u, v \rangle = \langle v, u \rangle$ for all $u, v \in V$.*

(ii) Linearity*: $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ and $\langle \alpha v, w \rangle = \alpha \langle v, w \rangle$ for all $u, v, w \in V$ and $\alpha \in \mathbb{R}$.*

(iii) Positive definiteness*: $\langle v, v \rangle \geq 0$ with equality if and only if $v = 0$.*

Example 2.1.

- For $V = \mathbb{R}^n$, the Euclidean inner product $\langle x, y \rangle = \sum_{i=1}^n x_i y_i = x^\mathsf{T} y$ is an inner product.

- If V is the set of all continuous functions on $[0,1]$, then $\langle f, g\rangle = \int_0^1 f(t)g(t)dt$ is an inner product.

- If V is the set of all random variables on a probability space Ω that have a finite variance, then $\langle X, Y\rangle = \mathbb{E}[XY]$ is an inner product on V.

Proposition 2.1 (***Norm induced by an inner product***)

If $\langle\cdot,\cdot\rangle$ is an inner product on V then $\|v\| \stackrel{\text{def}}{=} \sqrt{\langle v, v\rangle}$ is a norm on V. We say that the norm $\|\cdot\|$ is induced by the inner product $\langle\cdot,\cdot\rangle$.

Remark 2.1. *The Euclidean norm $\|\cdot\|_2$ on $V = \mathbb{R}^n$ is induced by the Euclidean inner product $x \cdot y = \sum_{i=1}^n x_i y_i$. Indeed, for $x \in \mathbb{R}^n$,*

$$\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2} = \sqrt{x \cdot x}.$$

Exercise 2.1. *Let $\langle\cdot,\cdot\rangle$ be an inner product on $\mathbb{R}^n$, and let $\|\cdot\|$ be the induced norm by $\langle\cdot,\cdot\rangle$.*

(**a**) *Show that for all $x, y \in \mathbb{R}^n$ we have*

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

(**b**) *Deduce from the previous question that the ℓ_1 norm $\|\cdot\|_1$ and the infinity norm $\|\cdot\|_\infty$ are **not** induced by an inner product.*

Theorem 2.1 (***Cauchy-Schwarz inequality***)

Let $\|\cdot\|$ be the norm induced by the inner product $\langle\cdot,\cdot\rangle$ on the vector space V. Then for all $x, y \in V$:

$$|\langle x, y\rangle| \leq \|x\|\,\|y\|. \tag{1}$$

Moreover, there is equality in (1) if and only if x and y are linearly dependent, i.e. $x = \alpha y$ or $y = \alpha x$ for some $\alpha \in \mathbb{R}$.

Proof. If $x = 0$ or $y = 0$ the result is obvious, we assume therefore to be in the case where $x \neq 0$ and $y \neq 0$. For $t \in \mathbb{R}$ we define the function $f(t) = \|tx - y\|^2$. Since the norm $\|\cdot\|$ is induced by the inner product $\langle\cdot,\cdot\rangle$ we have

$$f(t) = \langle tx - y, tx - y\rangle = t^2\|x\|^2 - 2t\langle x, y\rangle + \|y\|^2.$$

f is therefore a quadratic function of t. Notice that f is non-negative because $f(t) = \|tx-y\|^2 \geq 0$. This gives that its discriminant Δ is non-positive:

$$\Delta = (2\langle x, y\rangle)^2 - 4\|x\|^2\|y\|^2 \leq 0,$$

which proves (1). We have equality in (1) if and only if $\Delta = 0$ that is if and only if f admits a zero α, which is equivalent to $\alpha x - y = 0$, i.e. $y = \alpha x$. □

3 Orthogonality

In this section we consider an inner product $\langle\cdot,\cdot\rangle$ (that induces a norm $\|\cdot\|$) on a vector space V. For simplicity one may think of $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ to be the usual Euclidean dot product and norm on $V = \mathbb{R}^n$.

Definition 3.1 (*Orthogonality*)

- *We say that vectors x and y are* orthogonal *if $\langle x, y\rangle = 0$. We write then $x \perp y$.*
- *We say that a vector x is orthogonal to a set of vectors $A \subset V$ if x is orthogonal to all the vectors in A, i.e. $\forall y \in A,\ \langle x, y\rangle = 0$. We write then $x \perp A$.*
- *More generality we say that $A \subset V$ and $B \subset V$ are orthogonal if $\langle x, y\rangle = 0$ for all $x \in A$ and all $y \in B$. As before, we write $A \perp B$.*

Theorem 3.1 (*Pythagorean theorem*)

Let $x, y \in V$. Then

$$x \perp y \iff \|x+y\|^2 = \|x\|^2 + \|y\|^2.$$

Definition 3.2 (*Orthogonal and orthonormal families of vectors*)

Let $v_1, \dots, v_k$ be vectors of V. We say that the family of vectors $(v_1, \dots, v_k)$ is

- orthogonal *if the vectors $v_1, \dots, v_n$ are pairwise orthogonal, i.e. $\langle v_i, v_j\rangle = 0$ for all $i \neq j$.*
- orthonormal *if it is orthogonal and if all the v_i have unit norm: $\|v_1\| = \dots = \|v_k\| = 1$.*

Orthonormal basis are particularly convenient for computing coordinates of vectors:

Proposition 3.1

Assume that $\dim(V) = n$ and let $(v_1, \dots, v_n)$ be an **orthonormal** *basis of V. Then the coordinates of a vector $x \in V$ in the basis $(v_1, \dots, v_n)$ are $(\langle v_1, x\rangle, \dots, \langle v_n, x\rangle)$:*

$$x = \langle v_1, x\rangle v_1 + \dots + \langle v_n, x\rangle v_n.$$

Moreover, for all $y \in V$, we have $\langle x, y\rangle = \langle v_1, x\rangle\langle v_1, y\rangle + \dots + \langle v_n, x\rangle\langle v_n, y\rangle$. Taking $y = x$ leads to

$$\|x\| = \sqrt{\langle v_1, x\rangle^2 + \dots + \langle v_n, x\rangle^2}.$$

4 Orthogonal projection and distance to a subspace

We assume in this section that $V = \mathbb{R}^n$ and that $\langle \cdot, \cdot \rangle$, $\|\cdot\|$ are respectively the Euclidean dot product and Euclidean norm.

Definition 4.1 (*Orthogonal projection and distance to a subspace*)

Let S be a subspace of $\mathbb{R}^n$. The orthogonal projection *of a vector x onto S is defined as the vector $P_S(x)$ is S that minimizes the distance to x:*

$$P_S(x) \overset{\text{def}}{=} \underset{y \in S}{\arg\min}\, \|x - y\|.$$

The distance of x to the subspace S is then defined as

$$d(x, S) \overset{\text{def}}{=} \min_{y \in S} \|x - y\| = \|x - P_S(x)\|.$$

Proposition 4.1

*Let S be a subspace of $\mathbb{R}^n$ and let $(v_1, \ldots, v_k)$ be an **orthonormal basis** of S. Then for all $x \in \mathbb{R}^n$,*

$$P_S(x) = \langle v_1, x \rangle v_1 + \cdots + \langle v_k, x \rangle v_k.$$

In other words, if we let

$$V = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_k \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{n \times k},$$

then P_S is a linear transformation whose matrix is VV^T:

$$\forall x \in \mathbb{R}^n, \quad P_S(x) = VV^\mathsf{T} x.$$

Proof. Let us add vectors $v_{k+1}, \ldots, v_n$ to the basis $(v_1, \ldots, v_k)$ to obtain an orthonormal basis of $\mathbb{R}^n$. (This is made possible by the Gram-Schmidt orthonormalization principle that we will see in the next lecture.) Let $\alpha_1 = \langle x, v_1 \rangle, \ldots, \alpha_n = \langle x, v_n \rangle$ be the coordinates of x in the basis $(v_1, \ldots, v_n)$. Let $y \in S$, and let $\beta_1, \ldots, \beta_k$ be its coordinates in the basis $(v_1, \ldots, v_k)$. By Proposition 3.1:

$$\|x - y\|^2 = \sum_{i=1}^{k} (\alpha_i - \beta_i)^2 + \sum_{i=k+1}^{n} \alpha_i^2.$$

Minimizing this quantity over $y \in S$ is equivalent to minimizing it over the coordinates $\beta_1, \ldots, \beta_k$ of y. The minimum is uniquely achieved for $\beta_i = \alpha_i$ for all i, hence

$$P_S(x) \stackrel{\text{def}}{=} \operatorname*{arg\,min}_{y \in S} \|x - y\| = \alpha_1 v_1 + \cdots + \alpha_k v_k = \langle v_1, x \rangle v_1 + \cdots + \langle v_k, x \rangle v_k.$$

The second part of the proposition is a rewritting of this last equation, obtained by noticing that

$$V^\mathsf{T} x = \begin{pmatrix} - \, v_1 \, - \\ \vdots \\ - \, v_k \, - \end{pmatrix} x = \begin{pmatrix} \langle v_1, x \rangle \\ \vdots \\ \langle v_k, x \rangle \end{pmatrix}.$$

$\square$

Corollary 4.1

For all $x \in \mathbb{R}^n$,

- *$x - P_S(x)$ is orthogonal to S.*
- *$\|P_S(x)\| \leq \|x\|$.*

Definition 4.2 (*Orthogonal complement*)

Let S be a subspace of $\mathbb{R}^n$. The orthogonal complement of S is defined by

$$S^\perp \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \,|\, x \perp S\} = \{x \in \mathbb{R}^n \,|\, \forall y \in S, \, \langle x, y \rangle = 0\}.$$

Proposition 4.2

Let S be a subspace of $\mathbb{R}^n$. Then $S^\perp$ is also a subspace of $\mathbb{R}^n$ with dimension

$$\dim(S^\perp) = n - \dim(S).$$