Linear Algebra 2: Lecture 11

Irena Penev

Summer 2023

So far, we have worked with vector spaces over arbitrary fields $\mathbb{F}$. In this lecture, we impose some additional structure on vector spaces, namely the "scalar product" and the "norm." A scalar product is a way of multiplying two vectors and obtaining a scalar. A norm is a way of measuring distance of a vector from the origin, or alternatively, measuring the length of a vector. As a trade-off for imposing this additional structure, we restrict ourselves to vector spaces over only two fields: $\mathbb{R}$ and $\mathbb{C}$. The theory that we develop in this lecture would not work for vector spaces over general fields $\mathbb{F}$.

1 The scalar product

1.1 The scalar product for vector spaces over $\mathbb{R}$

A *scalar product* (also called *inner product*) in a vector space V over the field $\mathbb{R}$ is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ that satisfies the following axioms:

r.1. for all $\mathbf{x} \in V$, $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$, and equality holds if and only if $\mathbf{x} = \mathbf{0}$;

r.2. for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$;

r.3. for all $\mathbf{x}, \mathbf{y} \in V$ and $\alpha \in \mathbb{R}$, $\langle \alpha\mathbf{x}, \mathbf{y} \rangle = \alpha\langle \mathbf{x}, \mathbf{y} \rangle$;

r.4. for all $\mathbf{x}, \mathbf{y} \in V$, $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$.

The name "scalar product" comes from the fact that we multiply two vectors and obtain a scalar as a result.

Note that axioms r.2 and r.3 from the definition above guarantee that the scalar product in a vector space V over $\mathbb{R}$ is linear in the first variable (when we keep the second variable fixed). But in fact, axioms r.2, r.3, and r.4 guarantee that it is linear in the second variable as well (when we keep the first variable fixed). Indeed:

r.2'. for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, we have that

$$\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle,$$

since $\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle \overset{r.4}{=} \langle \mathbf{y} + \mathbf{z}, \mathbf{x} \rangle \overset{r.2}{=} \langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{z}, \mathbf{x} \rangle \overset{r.4}{=} \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$;

r.3'. for all $\mathbf{x}, \mathbf{y} \in V$ and $\alpha \in \mathbb{R}$, we have that

$$\langle \mathbf{x}, \alpha\mathbf{y} \rangle \;=\; \alpha\langle \mathbf{x}, \mathbf{y} \rangle,$$

since $\langle \mathbf{x}, \alpha\mathbf{y} \rangle \overset{r.4}{=} \langle \alpha\mathbf{y}, \mathbf{x} \rangle \overset{r.3}{=} \alpha\langle \mathbf{y}, \mathbf{x} \rangle \overset{r.4}{=} \alpha\langle \mathbf{x}, \mathbf{y} \rangle$.

Perhaps the best known example of a scalar product is the "standard scalar product" (sometimes also called the "dot product") in $\mathbb{R}^n$. Given vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{R}^n$, the *standard scalar product* of $\mathbf{x}$ and $\mathbf{y}$, denoted by $\mathbf{x} \cdot \mathbf{y}$, is given by

$$\mathbf{x} \cdot \mathbf{y} \;:=\; \sum_{i=1}^{n} x_i y_i$$

For example, for vectors $\begin{bmatrix} 1 & -2 & 5 \end{bmatrix}^T, \begin{bmatrix} -3 & 2 & 1 \end{bmatrix}^T \in \mathbb{R}^3$, we have

$$\begin{bmatrix} 1 \\ -2 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix} \;=\; 1 \cdot (-3) + (-2) \cdot 2 + 5 \cdot 1 \;=\; -2.$$

Note that for vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{R}^n$, we have that

$$\mathbf{x}^T\mathbf{y} \;=\; \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \;=\; \begin{bmatrix} \sum_{i=1}^{n} x_i y_i \end{bmatrix} \;=\; \begin{bmatrix} \mathbf{x} \cdot \mathbf{y} \end{bmatrix}$$

So, if we identify 1×1 matrices with scalars, then we simply get that

$$\mathbf{x} \cdot \mathbf{y} \;=\; \mathbf{x}^T\mathbf{y}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Proposition 1.1. *The standard scalar product in $\mathbb{R}^n$ is a scalar product.*

Proof. We need to check that the standard scalar product satisfies the axioms from the definition of a scalar product.

r.1. For a vector $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ in $\mathbb{R}^n$, we have that

$$\mathbf{x} \cdot \mathbf{x} \;=\; \sum_{i=1}^{n} x_i^2 \;\overset{(*)}{\geq}\; 0,$$

and (*) is an equality if and only if $x_1 = \dots = x_n = 0$, i.e. if and only if $\mathbf{x} = \mathbf{0}$.

r.2. For vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$, $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$, and $\mathbf{z} = \begin{bmatrix} z_1 & \dots & z_n \end{bmatrix}^T$ in $\mathbb{R}^n$, we have that

$$\begin{aligned}(\mathbf{x}+\mathbf{y})\cdot\mathbf{z} &= \sum_{i=1}^{n}(x_i+y_i)z_i \\ &= \Big(\sum_{i=1}^{n} x_i z_i\Big) + \Big(\sum_{i=1}^{n} y_i z_i\Big) \\ &= \mathbf{x}\cdot\mathbf{z} + \mathbf{y}\cdot\mathbf{z}.\end{aligned}$$

r.3. For vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{R}^n$ and a scalar $\alpha \in \mathbb{R}$, we have that

$$(\alpha\mathbf{x})\cdot\mathbf{y} \;=\; \sum_{i=1}^{n}(\alpha x_i)y_i \;=\; \alpha\sum_{i=1}^{n} x_i y_i \;=\; \alpha(\mathbf{x}\cdot\mathbf{y}).$$

r.4. For vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{R}^n$, we have that

$$\mathbf{x}\cdot\mathbf{y} \;=\; \sum_{i=1}^{n} x_i y_i \;=\; \sum_{i=1}^{n} y_i x_i \;=\; \mathbf{y}\cdot\mathbf{x}.$$

This proves that the standard scalar product in $\mathbb{R}^n$ really is a scalar product. □

A similar type of scalar product can be defined for matrices. Indeed, for matrices $A = [a_{i,j}]_{n\times m}$ and $B = [b_{i,j}]_{n\times m}$ in $\mathbb{R}^{n\times m}$, we can define $\langle A, B\rangle = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}b_{ij}$. It is easy to verify that this really is a scalar product in $\mathbb{R}^{n\times m}$ (the proof is similar to that of Proposition 1.1).

Here is another example, for those of you who have studied calculus. (If you have not studied calculus, you can ignore Proposition 1.2 below.)

Proposition 1.2. *Let $a, b \in \mathbb{R}$ be such that $a < b$, and let $\mathcal{C}_{[a,b]}$ be the vector space (over $\mathbb{R}$) of all continuous functions from the closed interval $[a,b]$ to $\mathbb{R}$.*[1] *Then the function $\langle\cdot,\cdot\rangle : \mathcal{C}_{[a,b]} \times \mathcal{C}_{[a,b]} \to \mathbb{R}$ defined by $\langle f, g\rangle := \int_a^b f(x)g(x)dx$ for all $f, g \in \mathcal{C}_{[a,b]}$ is a scalar product.*

[1] Recall from calculus that all such functions are integrable.

Proof. We must verify that the four axioms from the definition of a scalar product are satisfied. We first prove that axioms r.2, r.3, and r.4 are satisfied, and then we prove that axiom r.1 is satisfied (our proof of r.1 relies on r.3 and r.4, which why we need to prove them first).

r.2. For $f_1, f_2, f_3 \in \mathcal{C}_{[a,b]}$, we have that

$$\begin{aligned} \langle f_1 + f_2, f_3 \rangle &= \int_a^b \Big(f_1(x) + f_2(x) \Big) f_3(x) dx \\ &= \int_a^b \Big(f_1(x) f_3(x) + f_2(x) f_3(x) \Big) dx \\ &= \int_a^b f_1(x) f_3(x) dx + \int_a^b f_2(x) f_3(x) dx \\ &= \langle f_1, f_3 \rangle + \langle f_2, f_3 \rangle. \end{aligned}$$

r.3. For $f_1, f_2 \in \mathcal{C}_{[a,b]}$ and $\alpha \in \mathbb{R}$, we have that

$$\begin{aligned} \langle \alpha f_1, f_2 \rangle &= \int_a^b \Big(\alpha f_1(x) \Big) f_2(x) dx \\ &= \alpha \int_a^b f_1(x) f_2(x) dx \\ &= \alpha \langle f_1, f_2 \rangle. \end{aligned}$$

r.4. For $f_1, f_2 \in \mathcal{C}_{[a,b]}$, we have that

$$\langle f_1, f_2 \rangle \quad = \quad \int_a^b f_1(x) f_2(x) dx \quad = \quad \int_a^b f_2(x) f_1(x) dx \quad = \quad \langle f_2, f_1 \rangle.$$

r.1. Let $f \in \mathcal{C}_{[a,b]}$. Then

$$\langle f, f \rangle \quad = \quad \int_a^b f(x)^2 dx \quad \overset{(*)}{\geq} \quad 0,$$

where (*) follows from the fact that $f(x)^2 \geq 0$ for all $x \in [a,b]$. If $f(x) = 0$ for all $x \in [a,b]$, then obviously, $\langle f, f \rangle = 0$. Suppose now that there exists some $x_0 \in [a,b]$ such that $f(x_0) \neq 0$. We must show that $\langle f, f \rangle > 0$.

Suppose first that $f(x_0) > 0$. Set $m = \frac{f(x_0)}{2}$. (Clearly, $m > 0$.) Then since f is continuous on $[a,b]$, there exist $a_0, b_0 \in \mathbb{R}$ such that $a \leq a_0 \leq x_0 \leq b_0 \leq b$

and $a_0 < b_0$, and such that for all $x \in [a_0, b_0]$, we have that $f(x) \geq m$.[2] We now compute:

$$\begin{aligned} \langle f, f \rangle &= \int_a^b f(x)^2 dx \\ &= \int_a^{a_0} f(x)^2 dx + \int_{a_0}^{b_0} f(x)^2 dx + \int_{b_0}^{b} f(x)^2 dx \\ &\overset{(*)}{\geq} \int_{a_0}^{b_0} f(x)^2 dx \\ &\overset{(**)}{\geq} m^2 \\ &> 0, \end{aligned}$$

where (*) follows from the fact that $f(x)^2 \geq 0$ for all $x \in [a, a_0] \cup [b_0, b]$, and (**) follows from the fact that $f(x)^2 \geq m^2$ for all $x \in [a_0, b_0]$.

Suppose now that $f(x_0) < 0$. Then $-f(x_0) > 0$. So, by an argument completely analogous to the above (applied to $-f$), we have that $\langle -f, -f \rangle > 0$. We now use axioms r.3 and r.4 (which we have already verified) to obtain the following:

$$\langle -f, -f \rangle \overset{r.3}{=} -\langle f, -f \rangle \overset{r.4}{=} -\langle -f, f \rangle \overset{r.3}{=} \langle f, f \rangle.$$

Since $\langle -f, -f \rangle > 0$, it follows that $\langle f, f \rangle > 0$. □

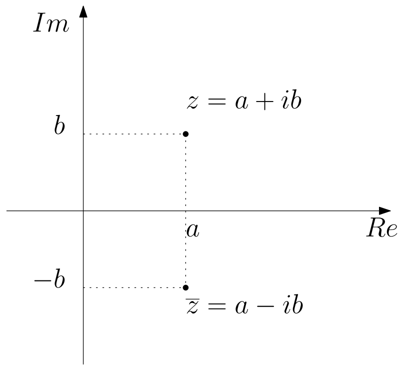
1.2 The scalar product for vector spaces over $\mathbb{C}$

Given a complex number $z = a + ib$ (where $a, b \in \mathbb{R}$),

- the *real part* of z is the real number $\text{Re}(z) = a$;
- the *imaginary part* of z is the real number $\text{Im}(z) = b$;
- the *absolute value* of z is the real number $|z| = \sqrt{a^2 + b^2}$;

[2] Essentially, this is because (by the continuity of f), we have that for $x \in [a, b]$, if $x \approx x_0$, then $f(x) \approx f(x_0)$. So, since $f(x_0) > m$, there exists some (sufficiently small) interval $[a_0, b_0]$ (with $a_0 < b_0$) such that $x_0 \in [a_0, b_0]$ and such that for all $x \in [a_0, b_0]$, we have that $f(x) \geq m$. Here is a formal proof, if you'd like one. Let $\varepsilon := m = \frac{f(x_0)}{2}$. By the continuity of f, there exists some $\delta > 0$ such that for all $x \in [a, b]$, if $|x - x_0| < \delta$, then $|f(x) - f(x_0)| < \varepsilon$. Now, let $a_0 := \max\{a, x_0 - \frac{\delta}{2}\}$ and $b_0 := \min\{b, x_0 + \frac{\delta}{2}\}$. Then $a \leq a_0 \leq x_0 \leq b_0 \leq b$ and $a_0 < b_0$. Moreover, by construction, for all $x \in [a_0, b_0]$, we have that $|x - x_0| \leq \frac{\delta}{2} < \delta$, and consequently, $|f(x) - f(x_0)| < \varepsilon$, i.e. $f(x_0) - \varepsilon < f(x) < f(x_0) + \varepsilon$. Since $f(x_0) = 2m > 0$ and $\varepsilon = m$, we deduce that for all $x \in [a_0, b_0]$, we have that $m < f(x) < 3m$, and in particular, $f(x) \geq m$.

- the *complex conjugate* of z is the complex number $\overline{z} = a - ib$ (see the picture below).



Obviously, for any complex number z, we have that $\overline{\overline{z}} = z$. We also have the following remark.

Remark 1.3. *For a complex number $z = a + ib$ (where $a, b \in \mathbb{R}$), we have that*

$$z\overline{z} \quad = \quad (a + ib)(a - ib) \quad = \quad a^2 - i^2b^2 \quad = \quad a^2 + b^2 \quad = \quad |z|^2;$$

in particular, $z\overline{z} \geq 0$, and equality holds if and only if $z = 0$.

The following proposition gives some familiar properties of the complex conjugate.

Proposition 1.4. *For all $z_1, z_2 \in \mathbb{C}$, the following hold:*

1. $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$;
2. $\overline{z_1 - z_2} = \overline{z_1} - \overline{z_2}$;
3. $\overline{z_1 z_2} = \overline{z_1}\;\overline{z_2}$;
4. *if $z_2 \neq 0$, then $\overline{z_1/z_2} = \overline{z_1}/\overline{z_2}$.*

Moreover, for all $z \in \mathbb{C}$ and non-negative integers n, we have that

5. $\overline{z^n} = (\overline{z})^n$.

Proof. High school math (or any introductory text on complex numbers). □

A *scalar product* (also called *inner product*) in a vector space V over the field $\mathbb{C}$ is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ that satisfies the following axioms:

c.1. for all $\mathbf{x} \in V$, $\langle \mathbf{x}, \mathbf{x} \rangle$ is a real number, $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$, and equality holds if and only if $\mathbf{x} = \mathbf{0}$;

c.2. for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$;

c.3. for all $\mathbf{x}, \mathbf{y} \in V$ and $\alpha \in \mathbb{C}$, $\langle \alpha\mathbf{x}, \mathbf{y} \rangle = \alpha\langle \mathbf{x}, \mathbf{y} \rangle$;

c.4. for all $\mathbf{x}, \mathbf{y} \in V$, $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$.

Note that axioms c.2 and c.3 from the definition above guarantee that the scalar product in a vector space V over $\mathbb{C}$ is linear in the first variable (when we keep the second variable fixed). Unlike in the real case, it is **not** linear in the second variable (when we keep the first variable fixed). We do, however, have the following:

c.2'. for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, we have that

$$\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle,$$

since $\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle \overset{c.4}{=} \overline{\langle \mathbf{y} + \mathbf{z}, \mathbf{x} \rangle} \overset{c.2}{=} \overline{\langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{z}, \mathbf{x} \rangle} = \overline{\langle \mathbf{y}, \mathbf{x} \rangle} + \overline{\langle \mathbf{z}, \mathbf{x} \rangle} \overset{c.4}{=} \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$;

c.3'. for all $\mathbf{x}, \mathbf{y} \in V$ and $\alpha \in \mathbb{C}$, we have that

$$\langle \mathbf{x}, \alpha\mathbf{y} \rangle = \overline{\alpha}\langle \mathbf{x}, \mathbf{y} \rangle,$$

since $\langle \mathbf{x}, \alpha\mathbf{y} \rangle \overset{c.4}{=} \overline{\langle \alpha\mathbf{y}, \mathbf{x} \rangle} = \overline{\alpha}\overline{\langle \mathbf{y}, \mathbf{x} \rangle} \overset{c.4}{=} \overline{\alpha}\langle \mathbf{x}, \mathbf{y} \rangle$.

There is also a complex version of the standard scalar product. Given vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{C}^n$, the *standard scalar product* of $\mathbf{x}$ and $\mathbf{y}$, denoted by $\mathbf{x} \cdot \mathbf{y}$, is given by

$$\mathbf{x} \cdot \mathbf{y} := \sum_{i=1}^{n} x_i \overline{y_i}.$$

For example, for vectors $\begin{bmatrix} 1 - 2i & -2 + i \end{bmatrix}^T$ and $\begin{bmatrix} 2 + i & 1 + 3i \end{bmatrix}^T$ in $\mathbb{C}^2$, we have

$$\begin{aligned} \begin{bmatrix} 1 - 2i \\ -2 + i \end{bmatrix} \cdot \begin{bmatrix} 2 + i \\ 1 + 3i \end{bmatrix} &= (1 - 2i)\overline{(2 + i)} + (-2 + i)\overline{(1 + 3i)} \\ &= (1 - 2i)(2 - i) + (-2 + i)(1 - 3i) \\ &= 1 + 2i. \end{aligned}$$

Proposition 1.5. *The standard scalar product in $\mathbb{C}^n$ is a scalar product.*

Proof. We need to check that the standard scalar product satisfies the axioms from the definition of a scalar product.

c.1. For a vector $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}$ in $\mathbb{C}^n$, we have that

$$\mathbf{x} \cdot \mathbf{x} \quad = \quad \sum_{i=1}^{n} x_i \overline{x_i} \quad \overset{(*)}{\geq} 0,$$

where (*) follows from Remark 1.3. Moreover, by Remark 1.3, the inequality (*) is an equality if and only if $x_1 = \dots = x_n = 0$, i.e. if and only if $\mathbf{x} = \mathbf{0}$.

c.2 For vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$, $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$, and $\mathbf{z} = \begin{bmatrix} z_1 & \dots & z_n \end{bmatrix}^T$ in $\mathbb{C}^n$, we have that

$$\begin{aligned} (\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} &= \sum_{i=1}^{n} (x_i + y_i) \overline{z_i} \\ &= \Big(\sum_{i=1}^{n} x_i \overline{z_i} \Big) + \Big(\sum_{i=1}^{n} y_i \overline{z_i} \Big) \\ &= \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}. \end{aligned}$$

c.3. For vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{C}^n$ and a scalar $\alpha \in \mathbb{C}$, we have that

$$(\alpha \mathbf{x}) \cdot \mathbf{y} \quad = \quad \sum_{i=1}^{n} (\alpha x_i) \overline{y_i} \quad = \quad \alpha \sum_{i=1}^{n} x_i \overline{y_i} \quad = \quad \alpha (\mathbf{x} \cdot \mathbf{y}).$$

c.4. For vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{C}^n$, we have that

$$\mathbf{x} \cdot \mathbf{y} \quad = \quad \sum_{i=1}^{n} x_i \overline{y_i} \quad = \quad \sum_{i=1}^{n} \overline{\overline{x_i} y_i} \quad = \quad \overline{\sum_{i=1}^{n} \overline{x_i} y_i} \quad = \quad \overline{\sum_{i=1}^{n} y_i \overline{x_i}} \quad = \quad \overline{\mathbf{y} \cdot \mathbf{x}}.$$

This proves that the standard scalar product in $\mathbb{C}^n$ really is a scalar product. □

2 The norm

In this section, we introduce the notion of a "norm" $||\cdot||$ in a vector space V over $\mathbb{R}$ or $\mathbb{C}$. The idea is that for a vector $\mathbf{x} \in V$, $||\mathbf{x}||$ is the distance from $\mathbf{x}$ to the origin, or alternatively, the length of the vector $\mathbf{x}$; $||\mathbf{x}||$ is always supposed to be a non-negative real number (even if V is a vector space over $\mathbb{C}$). For vectors $\mathbf{x}, \mathbf{y} \in V$, $||\mathbf{x} - \mathbf{y}||$ is supposed to be the distance between $\mathbf{x}$ and $\mathbf{y}$. Distance can be defined in a a variety of ways. We first study norms induced by a scalar product (see subsection 2.1). The definition of a norm in general is given in subsection 2.2, and some additional examples of norms are given in subsection 2.3.

2.1 The norm induced by a scalar product

Given a scalar product $\langle\cdot,\cdot\rangle$ in a vector space V over $\mathbb{R}$ or $\mathbb{C}$, we define the *norm in V induced by* $\langle\cdot,\cdot\rangle$ to be the function $||\cdot|| : V \to \mathbb{R}$ given by

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

for all $\mathbf{x} \in V$. In view of r.1 and c.1, for all $\mathbf{x} \in V$, we have that $||\mathbf{x}||$ is a non-negative **real** number,[3] and moreover, $||\mathbf{x}|| = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

Proposition 2.1. *Let V be a vector space over $\mathbb{R}$ or $\mathbb{C}$, let $\langle\cdot,\cdot\rangle$ be a scalar product in V, and let $||\cdot||$ be the norm in V induced by $\langle\cdot,\cdot\rangle$. Then for all vectors $\mathbf{x} \in V$ and scalars α,[4] we have that*

$$||\alpha\mathbf{x}|| \;=\; |\alpha|\;||\mathbf{x}||.$$

Proof. If the vector space V is over $\mathbb{R}$, then for all vectors $\mathbf{x} \in V$ and scalars $\alpha \in \mathbb{R}$, we have that

$$\begin{aligned}
||\alpha\mathbf{x}|| &= \sqrt{\langle \alpha\mathbf{x}, \alpha\mathbf{x}\rangle} \\
&= \sqrt{\alpha^2\langle \mathbf{x}, \mathbf{x}\rangle} && \text{by r.3 and r.3'} \\
&= |\alpha|\sqrt{\langle \mathbf{x}, \mathbf{x}\rangle} \\
&= |\alpha|\;||\mathbf{x}||.
\end{aligned}$$

On the other hand, if the vector space V is over $\mathbb{C}$, then for all vectors $\mathbf{x} \in V$ and scalars $\alpha \in \mathbb{C}$, we have that

$$\begin{aligned}
||\alpha\mathbf{x}|| &= \sqrt{\langle \alpha\mathbf{x}, \alpha\mathbf{x}\rangle} \\
&= \sqrt{\alpha\overline{\alpha}\langle \mathbf{x}, \mathbf{x}\rangle} && \text{by c.3 and c.3'} \\
&= \sqrt{|\alpha|^2\langle \mathbf{x}, \mathbf{x}\rangle} && \text{by Remark 1.3} \\
&= |\alpha|\sqrt{\langle \mathbf{x}, \mathbf{x}\rangle} \\
&= |\alpha|\;||\mathbf{x}||.
\end{aligned}$$

□

Note that if $||\cdot||$ is the norm induced on $\mathbb{R}^n$ by the standard scalar product in $\mathbb{R}^n$, then for all $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$, we have that

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

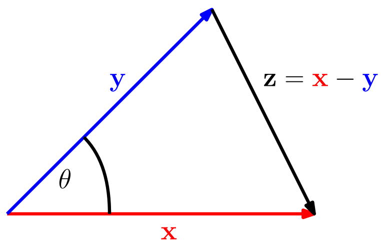
[3] This happens even if V is a vector space over $\mathbb{C}$.

[4] So, $\alpha \in \mathbb{R}$ or $\alpha \in \mathbb{C}$, depending on whether the vector space V is over $\mathbb{R}$ or $\mathbb{C}$.

So, we simply get the standard Euclidean length in $\mathbb{R}^n$. We note that if $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ are non-zero vectors in $\mathbb{R}^n$, then we have that

$$\mathbf{x} \cdot \mathbf{y} \quad = \quad ||\mathbf{x}||\, ||\mathbf{y}||\, \cos\theta,$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$. To see this, consider the triangle formed by $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z} := \mathbf{x} - \mathbf{y}$,[5] and let θ be the angle between $\mathbf{x}$ and $\mathbf{y}$ in this triangle.



We then compute

$$\begin{aligned} ||\mathbf{z}||^2 &= \mathbf{z} \cdot \mathbf{z} \\ &= (\mathbf{x} - \mathbf{y}) \cdot (\mathbf{x} - \mathbf{y}) \\ &= \underbrace{\mathbf{x} \cdot \mathbf{x}}_{=||\mathbf{x}||^2} - \mathbf{x} \cdot \mathbf{y} - \mathbf{y} \cdot \mathbf{x} + \underbrace{\mathbf{y} \cdot \mathbf{y}}_{=||\mathbf{y}||^2} \\ &= ||\mathbf{x}||^2 + ||\mathbf{y}||^2 - 2\mathbf{x} \cdot \mathbf{y} \end{aligned}$$

On the other hand, the Law of Cosines (for triangles) tells us that

$$||\mathbf{z}||^2 \quad = \quad ||\mathbf{x}||^2 + ||\mathbf{y}||^2 - 2||\mathbf{x}||\, ||\mathbf{y}||\, \cos\theta.$$

So, $||\mathbf{x}||^2 + ||\mathbf{y}||^2 - 2\mathbf{x} \cdot \mathbf{y} = ||\mathbf{z}||^2 = ||\mathbf{x}||^2 + ||\mathbf{y}||^2 - 2||\mathbf{x}||\, ||\mathbf{y}||\, \cos\theta$, and consequently,

$$\mathbf{x} \cdot \mathbf{y} \quad = \quad ||\mathbf{x}||\, ||\mathbf{y}||\, \cos\theta,$$

as we had claimed. Note that this means that non-zero vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are orthogonal (i.e. the angle between them is $90°$) if and only if $\mathbf{x} \cdot \mathbf{y} = 0$.[6]

Let us now return to norms defined by arbitrary scalar products. Given a vector space V over $\mathbb{R}$ or $\mathbb{C}$, and given a scalar product $\langle \cdot, \cdot \rangle$ on V, we say that vectors $\mathbf{x}$ and $\mathbf{y}$ in V are *orthogonal* (and we write $\mathbf{x} \perp \mathbf{y}$) if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. When our scalar product is the **standard** scalar product in $\mathbb{R}^n$, this corresponds

[5] This triangle may possibly be "degenerate" (i.e. one-dimensional). This happens if $\mathbf{x}$ and $\mathbf{y}$ are scalar multiples of each other.

[6] This is because for an angle θ, with $0° \leq \theta \leq 180°$, we have that $\cos\theta = 0$ if and only if $\theta = 90°$.

to our usual geometric interpretation, as we saw above. For norms induced by general scalar products, this is how we **define** orthogonality.[7] Note that it follows from r.4 and c.4 that $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ if and only if $\langle \mathbf{y}, \mathbf{x} \rangle = 0$. So, $\mathbf{x} \perp \mathbf{y}$ if and only if $\mathbf{y} \perp \mathbf{x}$, as one would expect. We also note that $\mathbf{0}$ is orthogonal to every vector in V. Indeed, for all $\mathbf{y} \in V$, we have that $\langle \mathbf{0}, \mathbf{y} \rangle = \langle 0\mathbf{0}, \mathbf{y} \rangle \stackrel{(*)}{=} 0\langle \mathbf{0}, \mathbf{y} \rangle = 0$, where (*) follows from r.3 or c.3.

The Pythagorean theorem. *Let V be a vector space over $\mathbb{R}$ or $\mathbb{C}$, let $\langle \cdot, \cdot \rangle$ be a scalar product in V, and let $||\cdot||$ be the norm in V induced by the scalar product $\langle \cdot, \cdot \rangle$. Let $\mathbf{x}$ and $\mathbf{y}$ be orthogonal vectors in V. Then*

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

Proof. Since $\mathbf{x} \perp \mathbf{y}$, we have that $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ and $\langle \mathbf{y}, \mathbf{x} \rangle = 0$. So,

$$\begin{aligned} ||\mathbf{x} + \mathbf{y}||^2 &= \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle \\ &= \underbrace{\langle \mathbf{x}, \mathbf{x} \rangle}_{=||\mathbf{x}||^2} + \underbrace{\langle \mathbf{x}, \mathbf{y} \rangle}_{=0} + \underbrace{\langle \mathbf{y}, \mathbf{x} \rangle}_{=0} + \underbrace{\langle \mathbf{y}, \mathbf{y} \rangle}_{=||\mathbf{y}||^2} \\ &= ||\mathbf{x}||^2 + ||\mathbf{y}||^2, \end{aligned}$$

which is what we needed to show. □

The Cauchy–Schwarz inequality. *Let $\langle \cdot, \cdot \rangle$ be a scalar product in a vector space V over $\mathbb{R}$ or $\mathbb{C}$, and let $||\cdot||$ be the norm in V induced by $\langle \cdot, \cdot \rangle$. Then*

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

for all $\mathbf{x}, \mathbf{y} \in V$.

Proof. Fix $\mathbf{x}, \mathbf{y} \in V$. We may assume that $\langle \mathbf{x}, \mathbf{y} \rangle \neq 0$, for otherwise, the result is immediate. Note that this implies that $\mathbf{x}, \mathbf{y} \neq \mathbf{0}$, and consequently, $||\mathbf{x}||, ||\mathbf{y}|| \neq 0$. We set

$$\mathbf{z} \quad := \quad \tfrac{\langle \mathbf{y}, \mathbf{y} \rangle}{\langle \mathbf{x}, \mathbf{y} \rangle}\mathbf{x} - \mathbf{y},$$

and we compute

$$\begin{aligned} \langle \mathbf{z}, \mathbf{y} \rangle &= \langle \tfrac{\langle \mathbf{y}, \mathbf{y} \rangle}{\langle \mathbf{x}, \mathbf{y} \rangle}\mathbf{x} - \mathbf{y}, \mathbf{y} \rangle \\ &\stackrel{(*)}{=} \tfrac{\langle \mathbf{y}, \mathbf{y} \rangle}{\langle \mathbf{x}, \mathbf{y} \rangle}\langle \mathbf{x}, \mathbf{y} \rangle - \langle \mathbf{y}, \mathbf{y} \rangle \\ &= 0, \end{aligned}$$

[7] For example, for the scalar product defined on $\mathcal{C}_{[-\pi,\pi]}$ in Proposition 1.2, we have that $\sin x \perp \cos x$, since $\langle \sin x, \cos x \rangle = \int\limits_{-\pi}^{\pi} \sin x \cos x dx = 0$.

where (*) follows from r.2 and r.3 if V is over $\mathbb{R}$, or from c.2 and c.3 if V is over $\mathbb{C}$. We have now shown that $\mathbf{z} \perp \mathbf{y}$, and so by the Pythagorean theorem, we have that

$$||\mathbf{z} + \mathbf{y}||^2 \quad = \quad ||\mathbf{z}||^2 + ||\mathbf{y}||^2.$$

But by construction, $\mathbf{z} + \mathbf{y} = \frac{\langle \mathbf{y}, \mathbf{y} \rangle}{\langle \mathbf{x}, \mathbf{y} \rangle}\mathbf{x}$, and consequently (by Proposition 2.1), we have that

$$\begin{aligned} ||\mathbf{z} + \mathbf{y}|| &= ||\tfrac{\langle \mathbf{y}, \mathbf{y} \rangle}{\langle \mathbf{x}, \mathbf{y} \rangle}\mathbf{x}|| \\ &= |\tfrac{\langle \mathbf{y}, \mathbf{y} \rangle}{\langle \mathbf{x}, \mathbf{y} \rangle}| \; ||\mathbf{x}|| \\ &= \tfrac{|\langle \mathbf{y}, \mathbf{y} \rangle|}{|\langle \mathbf{x}, \mathbf{y} \rangle|} \; ||\mathbf{x}|| \\ &= \tfrac{||\mathbf{y}||^2}{|\langle \mathbf{x}, \mathbf{y} \rangle|} \; ||\mathbf{x}||. \end{aligned}$$

So,

$$\tfrac{||\mathbf{y}||^4}{|\langle \mathbf{x}, \mathbf{y} \rangle|^2} \; ||\mathbf{x}||^2 \quad = \quad ||\mathbf{z} + \mathbf{y}||^2 \quad = \quad ||\mathbf{z}||^2 + ||\mathbf{y}||^2 \quad \geq \quad ||\mathbf{y}||^2,$$

which yields

$$\tfrac{||\mathbf{y}||^4}{|\langle \mathbf{x}, \mathbf{y} \rangle|^2} \; ||\mathbf{x}||^2 \quad \geq \quad ||\mathbf{y}||^2.$$

Since $\langle \mathbf{x}, \mathbf{y} \rangle$ and $||\mathbf{y}||$ are both non-zero, we have that $\frac{|\langle \mathbf{x}, \mathbf{y} \rangle|^2}{||\mathbf{y}||^2}$ is defined and positive. So, we may multiply both sides of the inequality above by $\frac{|\langle \mathbf{x}, \mathbf{y} \rangle|^2}{||\mathbf{y}||^2}$ to obtain

$$||\mathbf{x}||^2 ||\mathbf{y}||^2 \quad \geq \quad |\langle \mathbf{x}, \mathbf{y} \rangle|^2.$$

By taking the square root of both sides, we get

$$||\mathbf{x}|| \; ||\mathbf{y}|| \quad \geq \quad |\langle \mathbf{x}, \mathbf{y} \rangle|,$$

which is what we needed to show. □

Corollary 2.2. *For all* $x_1, \dots, x_n, y_1, \dots, y_n \in \mathbb{R}$*, we have that*

$$\Big(\sum_{i=1}^{n} x_i y_i \Big)^2 \quad \leq \quad \Big(\sum_{i=1}^{n} x_i^2 \Big) \Big(\sum_{i=1}^{n} y_i^2 \Big).$$

Proof. If we consider the standard scalar product in $\mathbb{R}^n$, the Cauchy-Schwarz inequality yields

$$|\sum_{i=1}^{n} x_i y_i| \quad \leq \quad \sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} y_i^2}.$$

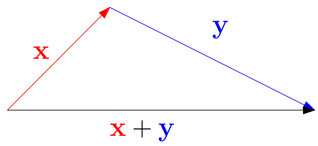
for all $x_1, \dots, x_n, y_1, \dots, y_n \in \mathbb{R}$. By squaring both sides, we obtain the desired inequality. □

As a further corollary of the Cauchy-Schwarz inequality, we obtain the following.

The triangle inequality. *Let $\langle \cdot, \cdot \rangle$ be a scalar product in a vector space V over $\mathbb{R}$ or $\mathbb{C}$, and let $|| \cdot ||$ be the norm in V induced by $\langle \cdot, \cdot \rangle$. Then*

$$||\mathbf{x}+\mathbf{y}|| \quad \leq \quad ||\mathbf{x}|| + ||\mathbf{y}||$$

for all $\mathbf{x}, \mathbf{y} \in V$.



Proof. We prove the result for the case when V is a vector space over $\mathbb{C}$. The real case is similar, but easier. We first remark that for all complex numbers $z = a + ib$ (where $a, b \in \mathbb{R}$), we have that

- $z + \overline{z} = 2a = 2\text{Re}(z)$;
- $\text{Re}(z) = a \leq |a| \leq \sqrt{a^2 + b^2} = |z|$.

Now, fix $\mathbf{x}, \mathbf{y} \in V$. Then we have the following:

$$\begin{aligned}
||\mathbf{x}+\mathbf{y}||^2 &= \langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y} \rangle \\
&= \underbrace{\langle \mathbf{x}, \mathbf{x} \rangle}_{=||\mathbf{x}||^2} + \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle + \underbrace{\langle \mathbf{y}, \mathbf{y} \rangle}_{=||\mathbf{y}||^2} && \text{by c.2 and c.2'} \\
&= ||\mathbf{x}||^2 + ||\mathbf{y}||^2 + \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle \\
&= ||\mathbf{x}||^2 + ||\mathbf{y}||^2 + \langle \mathbf{x}, \mathbf{y} \rangle + \overline{\langle \mathbf{x}, \mathbf{y} \rangle} && \text{by c.4} \\
&= ||\mathbf{x}||^2 + ||\mathbf{y}||^2 + 2\text{Re}\Big(\langle \mathbf{x}, \mathbf{y} \rangle\Big) \\
&\leq ||\mathbf{x}||^2 + ||\mathbf{y}||^2 + 2|\langle \mathbf{x}, \mathbf{y} \rangle| \\
&\leq ||\mathbf{x}||^2 + ||\mathbf{y}||^2 + 2||\mathbf{x}||\;||\mathbf{y}|| && \text{by the Cauchy-Scwarz inequality} \\
&= (||\mathbf{x}|| + ||\mathbf{y}||)^2.
\end{aligned}$$

By taking the square root of both sides, we obtain

$$||\mathbf{x}+\mathbf{y}|| \quad \leq \quad ||\mathbf{x}|| + ||\mathbf{y}||,$$

which is what we needed to show. □

2.2 The norm in general

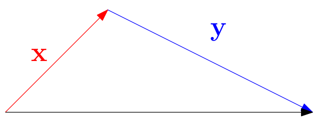
A *norm* in a vector space V over $\mathbb{R}$ or $\mathbb{C}$ is a function $||\cdot|| : V \to \mathbb{R}$ that satisfies the following axioms:

n.1. for all vectors $\mathbf{x} \in V$, we have that $||\mathbf{x}|| \geq 0$, and equality holds if and only if $\mathbf{x} = \mathbf{0}$;

n.2. for all vectors $\mathbf{x} \in V$ and scalars α,[8] we have that $||\alpha\mathbf{x}|| = |\alpha|\ ||\mathbf{x}||$;

n.3. for all vectors $\mathbf{x}, \mathbf{y} \in V$, we have that $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$.

As stated at the beginning of section 2, a norm in a vector space V over $\mathbb{R}$ or $\mathbb{C}$ gives a way of measuring the distance of a vector from the origin, or equivalently, measuring the length of a vector. The norm of a vector is always a non-negative real number (regardless of whether the vector space is over $\mathbb{R}$ or $\mathbb{C}$). We note that n.3 is referred to as the "triangle inequality." The idea is that vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{x} + \mathbf{y}$ form a triangle (see the picture below), and the length of the third side cannot be greater than the sum of lengths of the other two sides.



It follows from the results of subsection 2.1 that any norm induced by a scalar product in a vector space V over $\mathbb{R}$ or $\mathbb{C}$ really is a norm, i.e. it is a function from V to $\mathbb{R}$ that satisfies axioms n.1, n.2, and n.3 above. The fact that axiom n.1 is satisfied is immediate from the construction of a norm induced by a scalar product, the fact that n.2 is satisfied follows from Proposition 2.1, and the fact that n.3 is satisfied follows from the triangle inequality proven in subsection 2.1.

2.3 Other examples of norms

For a positive integer p, we define the *p-norm*, denoted by $||\cdot||_p$, on $\mathbb{R}^n$ by setting

$$||\mathbf{x}||_p \quad := \quad \Big(\sum_{i=1}^{n} |x_i|^p \Big)^{\frac{1}{p}}$$

for all $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ in $\mathbb{R}^n$. We omit the proof of the fact that this really is a norm in $\mathbb{R}^n$. We do note, however, that for $p = 2$, we get

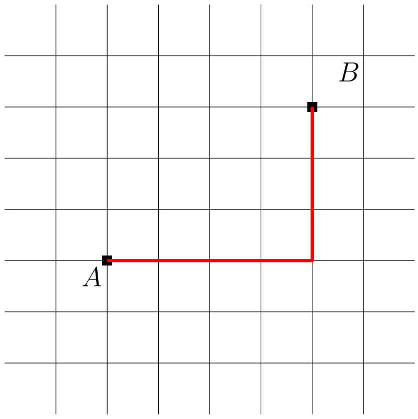
$$||\mathbf{x}||_2 \quad = \quad \sqrt{\sum_{i=1}^{n} x_i^2},$$

[8] So, $\alpha \in \mathbb{R}$ or $\alpha \in \mathbb{C}$, depending on whether V is a vector space over $\mathbb{R}$ or $\mathbb{C}$.

which is precisely the norm induced by the standard scalar product in $\mathbb{R}^n$, i.e. the standard Euclidean norm in $\mathbb{R}^n$. For $p = 1$, we get

$$||\mathbf{x}||_1 \;=\; \sum_{i=1}^{n} |x_i|.$$

We note that the $||\cdot||_1$ norm is sometimes called the "Manhattan norm." This is because streets and avenues in Manhattan form a perfect grid (more or less), and so $||\cdot||_1$ gives the actual walking distance between two places in Manhattan (see the picture below).



Another norm of interest is the so called "Chebyshev distance" on $\mathbb{R}^n$, denoted by $||\cdot||_\infty$. It is defined by

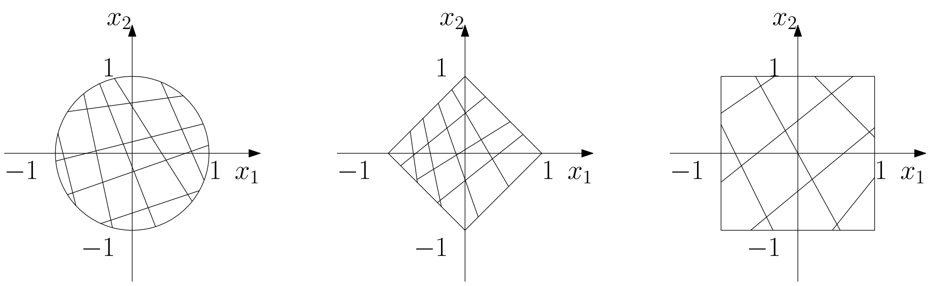
$$||\mathbf{x}||_\infty \;:=\; \max\{|x_1|, \ldots, |x_n|\}$$

for all vectors $\mathbf{x} = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix}^T$ in $\mathbb{R}^n$.

Given a norm $||\cdot||$ in a vector space V over $\mathbb{R}$ or $\mathbb{C}$, the *unit disk* is the set

$$\{\mathbf{x} \in V \mid ||\mathbf{x}|| \le 1\}.$$

For the purposes of gaining intuition, it may be useful to consider unit disks in $\mathbb{R}^2$ with respect to the norms $||\cdot||_2$, $||\cdot||_1$, and $||\cdot||_\infty$. These are represented in the picture below.



unit disk in $\mathbb{R}^2$ w.r.t. $||\cdot||_2$

unit disk in $\mathbb{R}^2$ w.r.t. $||\cdot||_1$

unit disk in $\mathbb{R}^2$ w.r.t. $||\cdot||_\infty$

Finally, if you have studied calculus,[9] recall that for $a, b \in \mathbb{R}$ such that $a < b$, $\mathcal{C}_{[a,b]}$ is the vector space (over $\mathbb{R}$) of all continuous functions from $[a, b]$ to $\mathbb{R}$. For a real number $p \geq 1$, we have the norm $|| \cdot ||_p$ on $\mathcal{C}_{[a,b]}$ given by

$$||f||_p \quad = \quad \left(\int_a^b |f(x)|^p \right)^{\frac{1}{p}}$$

for all $f \in \mathcal{C}_{[a,b]}$, and we also have the norm $|| \cdot ||_\infty$ on $\mathcal{C}_{[a,b]}$ given by

$$||f||_\infty \quad = \quad \max_{x \in [a,b]} |f(x)|$$

for all $f \in \mathcal{C}_{[a,b]}$. Once again, we omit the proof of the fact that $|| \cdot ||_p$ (for a real number $p \geq 1$) and $|| \cdot ||_\infty$ really are norms on $\mathcal{C}_{[a,b]}$.

[9]If you haven't, you may ignore this paragraph.