MATH 323

Linear Algebra

Lecture 22:
Orthogonal sets.
The Gram-Schmidt orthogonalization process.
Norm on a vector space.

Orthogonal sets

Let $\langle \cdot, \cdot \rangle$ denote the scalar product in $\mathbb{R}^n$.

Definition. Nonzero vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ form an **orthogonal set** if they are orthogonal to each other: $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ for $i \neq j$.

If, in addition, all vectors are of unit length, $\|\mathbf{v}_i\| = 1$, then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is called an **orthonormal set**.

Example. The standard basis $\mathbf{e}_1 = (1, 0, 0, \ldots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \ldots, 0)$, $\ldots$, $\mathbf{e}_n = (0, 0, 0, \ldots, 1)$. It is an orthonormal set.

Orthogonality $\implies$ linear independence

Theorem Suppose $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are nonzero vectors that form an orthogonal set. Then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are linearly independent.

Proof: Suppose $t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_k\mathbf{v}_k = \mathbf{0}$ for some $t_1, t_2, \ldots, t_k \in \mathbb{R}$. Our task is to show that $t_1 = t_2 = \cdots = t_k = 0$.

For any index i, $1 \leq i \leq k$ we have

$$\langle t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_k\mathbf{v}_k, \mathbf{v}_i \rangle = \langle \mathbf{0}, \mathbf{v}_i \rangle = 0$$

$$\implies t_1\langle \mathbf{v}_1, \mathbf{v}_i \rangle + t_2\langle \mathbf{v}_2, \mathbf{v}_i \rangle + \cdots + t_k\langle \mathbf{v}_k, \mathbf{v}_i \rangle = 0.$$

By orthogonality, $t_j\langle \mathbf{v}_j, \mathbf{v}_i \rangle = 0$ for all $j \neq i$. Then $t_i\langle \mathbf{v}_i, \mathbf{v}_i \rangle = 0$ as well, which implies $t_i = 0$.

Orthonormal bases

Suppose $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is an orthonormal basis for $\mathbb{R}^n$ (i.e., it is a basis and an orthonormal set).

Theorem Let $\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$ and $\mathbf{y} = y_1\mathbf{v}_1 + y_2\mathbf{v}_2 + \cdots + y_n\mathbf{v}_n$, where $x_i, y_j \in \mathbb{R}$. Then

(i) $\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_1 + x_2y_2 + \cdots + x_ny_n$,

(ii) $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$.

Proof: (ii) follows from (i) when $\mathbf{y} = \mathbf{x}$.

$$\langle \mathbf{x}, \mathbf{y} \rangle = \left\langle \sum_{i=1}^n x_i\mathbf{v}_i, \sum_{j=1}^n y_j\mathbf{v}_j \right\rangle = \sum_{i=1}^n x_i \left\langle \mathbf{v}_i, \sum_{j=1}^n y_j\mathbf{v}_j \right\rangle$$

$$= \sum_{i=1}^n \sum_{j=1}^n x_i y_j \langle \mathbf{v}_i, \mathbf{v}_j \rangle = \sum_{i=1}^n x_i y_i.$$

Suppose V is a subspace of $\mathbb{R}^n$. Let $\mathbf{p}$ be the orthogonal projection of a vector $\mathbf{x} \in \mathbb{R}^n$ onto V.

If V is a one-dimensional subspace spanned by a vector $\mathbf{v}$ then $\mathbf{p} = \dfrac{\langle \mathbf{x}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} \mathbf{v}$.

If V admits an orthogonal basis $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ then

$$\mathbf{p} = \frac{\langle \mathbf{x}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 + \frac{\langle \mathbf{x}, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2 + \cdots + \frac{\langle \mathbf{x}, \mathbf{v}_k \rangle}{\langle \mathbf{v}_k, \mathbf{v}_k \rangle} \mathbf{v}_k.$$

Indeed, $\langle \mathbf{p}, \mathbf{v}_i \rangle = \displaystyle\sum_{j=1}^{k} \frac{\langle \mathbf{x}, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle} \langle \mathbf{v}_j, \mathbf{v}_i \rangle = \frac{\langle \mathbf{x}, \mathbf{v}_i \rangle}{\langle \mathbf{v}_i, \mathbf{v}_i \rangle} \langle \mathbf{v}_i, \mathbf{v}_i \rangle = \langle \mathbf{x}, \mathbf{v}_i \rangle$

$\implies \langle \mathbf{x} - \mathbf{p}, \mathbf{v}_i \rangle = 0 \implies \mathbf{x} - \mathbf{p} \perp \mathbf{v}_i \implies \mathbf{x} - \mathbf{p} \perp V.$

Coordinates relative to an orthogonal basis

Theorem If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is an orthogonal basis for $\mathbb{R}^n$, then

$$\mathbf{x} = \frac{\langle \mathbf{x}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 + \frac{\langle \mathbf{x}, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{x}, \mathbf{v}_n \rangle}{\langle \mathbf{v}_n, \mathbf{v}_n \rangle}\mathbf{v}_n$$

for any vector $\mathbf{x} \in \mathbb{R}^n$.

Corollary If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is an orthonormal basis for $\mathbb{R}^n$, then

$$\mathbf{x} = \langle \mathbf{x}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{x}, \mathbf{v}_2 \rangle \mathbf{v}_2 + \cdots + \langle \mathbf{x}, \mathbf{v}_n \rangle \mathbf{v}_n$$

for any vector $\mathbf{x} \in \mathbb{R}^n$.

The Gram-Schmidt orthogonalization process

Let V be a subspace of $\mathbb{R}^n$. Suppose $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ is a basis for V. Let

$$\mathbf{v}_1 = \mathbf{x}_1,$$

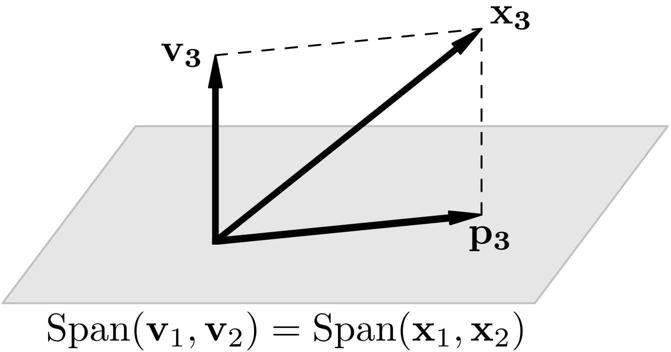
$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1,$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\langle \mathbf{x}_3, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 - \frac{\langle \mathbf{x}_3, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2,$$

. . . .

$$\mathbf{v}_k = \mathbf{x}_k - \frac{\langle \mathbf{x}_k, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 - \cdots - \frac{\langle \mathbf{x}_k, \mathbf{v}_{k-1} \rangle}{\langle \mathbf{v}_{k-1}, \mathbf{v}_{k-1} \rangle} \mathbf{v}_{k-1}.$$

Then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is an orthogonal basis for V.


$\mathbf{x_3}$
$\mathbf{v_3}$
$\mathbf{p_3}$
$\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2) = \mathrm{Span}(\mathbf{x}_1, \mathbf{x}_2)$

Any basis		Orthogonal basis
$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$	$\longrightarrow$	$\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$

Properties of the Gram-Schmidt process:

- $\mathbf{v}_j = \mathbf{x}_j - (\alpha_1 \mathbf{x}_1 + \dots + \alpha_{j-1} \mathbf{x}_{j-1})$, $1 \le j \le k$;
- the span of $\mathbf{v}_1, \dots, \mathbf{v}_j$ is the same as the span of $\mathbf{x}_1, \dots, \mathbf{x}_j$;
- $\mathbf{v}_j$ is orthogonal to $\mathbf{x}_1, \dots, \mathbf{x}_{j-1}$;
- $\mathbf{v}_j = \mathbf{x}_j - \mathbf{p}_j$, where $\mathbf{p}_j$ is the orthogonal projection of the vector $\mathbf{x}_j$ on the subspace spanned by $\mathbf{x}_1, \dots, \mathbf{x}_{j-1}$;
- $\|\mathbf{v}_j\|$ is the distance from $\mathbf{x}_j$ to the subspace spanned by $\mathbf{x}_1, \dots, \mathbf{x}_{j-1}$.

Normalization

Let V be a subspace of $\mathbb{R}^n$. Suppose $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is an orthogonal basis for V.

Let $\mathbf{w}_1 = \dfrac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$, $\mathbf{w}_2 = \dfrac{\mathbf{v}_2}{\|\mathbf{v}_2\|}$,..., $\mathbf{w}_k = \dfrac{\mathbf{v}_k}{\|\mathbf{v}_k\|}$.

Then $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$ is an orthonormal basis for V.

Theorem Any non-trivial subspace of $\mathbb{R}^n$ admits an orthonormal basis.

Orthogonalization / Normalization

An alternative form of the Gram-Schmidt process combines orthogonalization with normalization.

Suppose $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ is a basis for a subspace $V \subset \mathbb{R}^n$. Let

$$\mathbf{v}_1 = \mathbf{x}_1, \quad \mathbf{w}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|},$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \langle \mathbf{x}_2, \mathbf{w}_1 \rangle \mathbf{w}_1, \quad \mathbf{w}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|},$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \langle \mathbf{x}_3, \mathbf{w}_1 \rangle \mathbf{w}_1 - \langle \mathbf{x}_3, \mathbf{w}_2 \rangle \mathbf{w}_2, \quad \mathbf{w}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|},$$

....

$$\mathbf{v}_k = \mathbf{x}_k - \langle \mathbf{x}_k, \mathbf{w}_1 \rangle \mathbf{w}_1 - \cdots - \langle \mathbf{x}_k, \mathbf{w}_{k-1} \rangle \mathbf{w}_{k-1},$$
$$\mathbf{w}_k = \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}.$$

Then $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$ is an orthonormal basis for V.

Problem. Let Π be the plane spanned by vectors $\mathbf{x}_1 = (1,1,0)$ and $\mathbf{x}_2 = (0,1,1)$.

(i) Find the orthogonal projection of the vector $\mathbf{y} = (4,0,-1)$ onto the plane Π.

(ii) Find the distance from $\mathbf{y}$ to Π.

First we apply the Gram-Schmidt process to the basis $\mathbf{x}_1, \mathbf{x}_2$:

$$\mathbf{v}_1 = \mathbf{x}_1 = (1,1,0),$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 = (0,1,1) - \frac{1}{2}(1,1,0) = (-1/2, 1/2, 1).$$

Now that $\mathbf{v}_1, \mathbf{v}_2$ is an orthogonal basis for Π, the orthogonal projection of $\mathbf{y}$ onto Π is

$$\mathbf{p} = \frac{\langle \mathbf{y}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 + \frac{\langle \mathbf{y}, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 = \frac{4}{2}(1,1,0) + \frac{-3}{3/2}(-1/2, 1/2, 1)$$

$$= (2,2,0) + (1,-1,-2) = (3,1,-2).$$

The distance from $\mathbf{y}$ to Π is $\|\mathbf{y} - \mathbf{p}\| = \|(1,-1,1)\| = \sqrt{3}$.

Problem. Find the distance from the point $\mathbf{y} = (0,0,0,1)$ to the subspace $V \subset \mathbb{R}^4$ spanned by vectors $\mathbf{x}_1 = (1,-1,1,-1)$, $\mathbf{x}_2 = (1,1,3,-1)$, and $\mathbf{x}_3 = (-3,7,1,3)$.

First we apply the Gram-Schmidt process to vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ and obtain an orthogonal basis $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ for the subspace V. Next we compute the orthogonal projection $\mathbf{p}$ of the vector $\mathbf{y}$ onto V:

$$\mathbf{p} = \frac{\langle \mathbf{y}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 + \frac{\langle \mathbf{y}, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 + \frac{\langle \mathbf{y}, \mathbf{v}_3 \rangle}{\langle \mathbf{v}_3, \mathbf{v}_3 \rangle}\mathbf{v}_3.$$

Then the distance from $\mathbf{y}$ to V equals $\|\mathbf{y} - \mathbf{p}\|$.

Alternatively, we can apply the Gram-Schmidt process to vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{y}$. We should obtain an orthogonal system $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$. Then the desired distance will be $\|\mathbf{v}_4\|$.

$\mathbf{x}_1 = (1, -1, 1, -1),\ \mathbf{x}_2 = (1, 1, 3, -1),$
$\mathbf{x}_3 = (-3, 7, 1, 3),\ \mathbf{y} = (0, 0, 0, 1).$

$$\mathbf{v}_1 = \mathbf{x}_1 = (1, -1, 1, -1),$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 = (1, 1, 3, -1) - \frac{4}{4}(1, -1, 1, -1)$$

$$= (0, 2, 2, 0),$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\langle \mathbf{x}_3, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 - \frac{\langle \mathbf{x}_3, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2$$

$$= (-3, 7, 1, 3) - \frac{-12}{4}(1, -1, 1, -1) - \frac{16}{8}(0, 2, 2, 0)$$

$$= (0, 0, 0, 0).$$

The Gram-Schmidt process can be used to check linear independence of vectors! It failed because the vector $\mathbf{x}_3$ is a linear combination of $\mathbf{x}_1$ and $\mathbf{x}_2$. V is a plane, not a 3-dimensional subspace. To fix things, it is enough to drop $\mathbf{x}_3$, i.e., we should orthogonalize vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}$.

$$\tilde{\mathbf{v}}_3 = \mathbf{y} - \frac{\langle \mathbf{y}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 - \frac{\langle \mathbf{y}, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2$$

$$= (0,0,0,1) - \frac{-1}{4}(1,-1,1,-1) - \frac{0}{8}(0,2,2,0)$$

$$= (1/4, -1/4, 1/4, 3/4).$$

$$|\tilde{\mathbf{v}}_3| = \left|\left(\frac{1}{4}, -\frac{1}{4}, \frac{1}{4}, \frac{3}{4}\right)\right| = \frac{1}{4}\,|(1,-1,1,3)| = \frac{\sqrt{12}}{4} = \frac{\sqrt{3}}{2}.$$

Norm

The notion of *norm* generalizes the notion of length of a vector in $\mathbb{R}^n$.

Definition. Let V be a vector space. A function $\alpha : V \to \mathbb{R}$ is called a **norm** on V if it has the following properties:

(i) $\alpha(\mathbf{x}) \geq 0$, $\alpha(\mathbf{x}) = 0$ only for $\mathbf{x} = \mathbf{0}$ (positivity)
(ii) $\alpha(r\mathbf{x}) = |r|\,\alpha(\mathbf{x})$ for all $r \in \mathbb{R}$ (homogeneity)
(iii) $\alpha(\mathbf{x}+\mathbf{y}) \leq \alpha(\mathbf{x}) + \alpha(\mathbf{y})$ (triangle inequality)

Notation. The norm of a vector $\mathbf{x} \in V$ is usually denoted $\|\mathbf{x}\|$. Different norms on V are distinguished by subscripts, e.g., $\|\mathbf{x}\|_1$ and $\|\mathbf{x}\|_2$.

Examples. $V = \mathbb{R}^n$, $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$.

- $\|\mathbf{x}\|_\infty = \max(|x_1|, |x_2|, \dots, |x_n|)$.

Positivity and homogeneity are obvious. Let $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$. Then $\mathbf{x} + \mathbf{y} = (x_1 + y_1, \dots, x_n + y_n)$.

$$|x_i + y_i| \le |x_i| + |y_i| \le \max_j |x_j| + \max_j |y_j|$$

$$\implies \max_j |x_j + y_j| \le \max_j |x_j| + \max_j |y_j|$$

$$\implies \|\mathbf{x} + \mathbf{y}\|_\infty \le \|\mathbf{x}\|_\infty + \|\mathbf{y}\|_\infty.$$

- $\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_n|$.

Positivity and homogeneity are obvious.
The triangle inequality: $|x_i + y_i| \le |x_i| + |y_i|$

$$\implies \sum_j |x_j + y_j| \le \sum_j |x_j| + \sum_j |y_j|$$

Examples. $V = \mathbb{R}^n$, $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.

- $\|\mathbf{x}\|_p = \big(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\big)^{1/p}$, $p > 0$.

Remark. $\|\mathbf{x}\|_2 =$ Euclidean length of $\mathbf{x}$.

Theorem $\|\mathbf{x}\|_p$ is a norm on $\mathbb{R}^n$ for any $p \geq 1$.

Positivity and homogeneity are still obvious (and hold for any $p > 0$). The triangle inequality for $p \geq 1$ is known as the **Minkowski inequality**:

$$\big(|x_1 + y_1|^p + |x_2 + y_2|^p + \cdots + |x_n + y_n|^p\big)^{1/p} \leq$$
$$\leq \big(|x_1|^p + \cdots + |x_n|^p\big)^{1/p} + \big(|y_1|^p + \cdots + |y_n|^p\big)^{1/p}.$$

Normed vector space

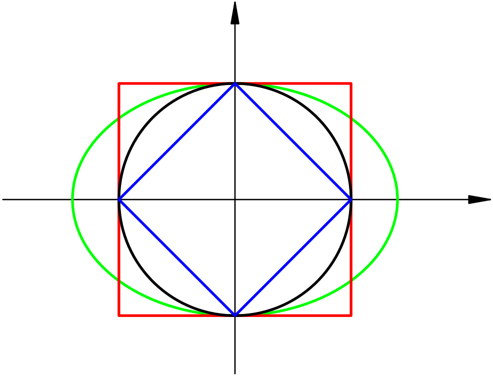
Definition. A **normed vector space** is a vector space endowed with a norm.

The norm defines a distance function on the normed vector space: $\operatorname{dist}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$.

Then we say that a vector $\mathbf{x}$ is a good *approximation* of a vector $\mathbf{x}_0$ if $\operatorname{dist}(\mathbf{x}, \mathbf{x}_0)$ is small.

Also, we say that a sequence $\mathbf{x}_1, \mathbf{x}_2, \dots$ *converges* to a vector $\mathbf{x}$ if $\operatorname{dist}(\mathbf{x}, \mathbf{x}_n) \to 0$ as $n \to \infty$.

Unit circle: $\|\mathbf{x}\| = 1$



$\|\mathbf{x}\| = (x_1^2 + x_2^2)^{1/2}$	black
$\|\mathbf{x}\| = \left(\frac{1}{2}x_1^2 + x_2^2\right)^{1/2}$	green
$\|\mathbf{x}\| = \|x_1\| + \|x_2\|$	blue
$\|\mathbf{x}\| = \max(\|x_1\|, \|x_2\|)$	red

Examples. $V = C[a,b]$, $f : [a,b] \to \mathbb{R}$.

- $\|f\|_\infty = \max_{a \le x \le b} |f(x)|$.
- $\|f\|_1 = \int_a^b |f(x)|\,dx$.
- $\|f\|_p = \left(\int_a^b |f(x)|^p\,dx\right)^{1/p}$, $p > 0$.

Theorem $\|f\|_p$ is a norm on $C[a,b]$ for any $p \ge 1$.