MATH 311

Topics in Applied Mathematics I

Lecture 18: Orthogonal projection (continued). Least squares problems. Norm of a vector.

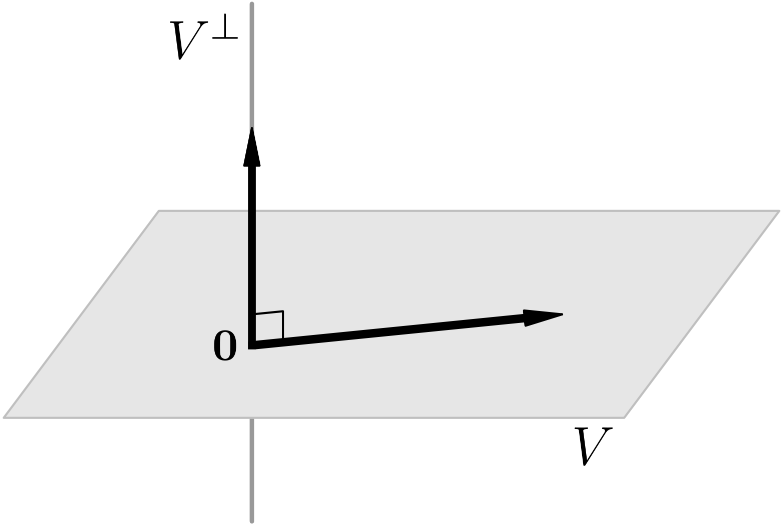
Orthogonal complement

Definition. Let $S \subset \mathbb{R}^n$. The **orthogonal complement** of S, denoted $S^\perp$, is the set of all vectors $\mathbf{x} \in \mathbb{R}^n$ that are orthogonal to S.

Theorem 1 (i) $S^\perp$ is a subspace of $\mathbb{R}^n$.
(ii) $(S^\perp)^\perp = \mathrm{Span}(S)$.

Theorem 2 If V is a subspace of $\mathbb{R}^n$, then
(i) $(V^\perp)^\perp = V$,
(ii) $V \cap V^\perp = \{\mathbf{0}\}$,
(iii) $\dim V + \dim V^\perp = n$.

Theorem 3 If V is the row space of a matrix, then $V^\perp$ is the nullspace of the same matrix.


$V^\perp$
$\mathbf{0}$
V

Orthogonal projection

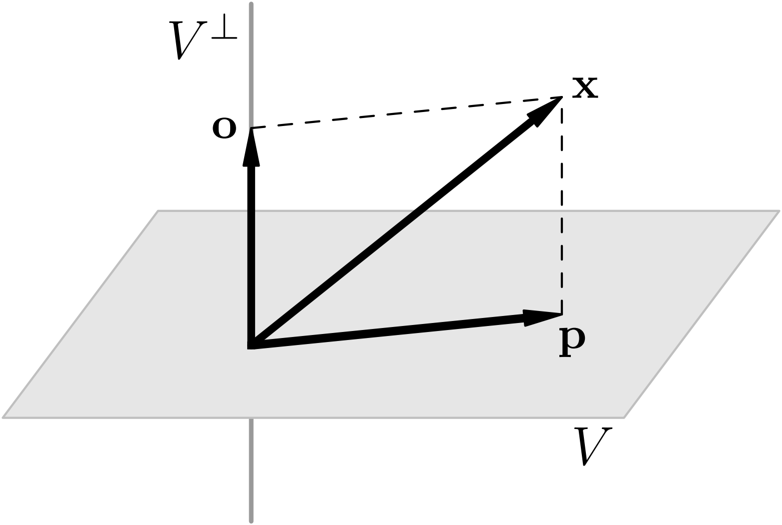
Theorem 1 Let V be a subspace of $\mathbb{R}^n$. Then any vector $\mathbf{x} \in \mathbb{R}^n$ is uniquely represented as $\mathbf{x} = \mathbf{p} + \mathbf{o}$, where $\mathbf{p} \in V$ and $\mathbf{o} \in V^{\perp}$.

In the above expansion, $\mathbf{p}$ is called the **orthogonal projection** of the vector $\mathbf{x}$ onto the subspace V.

If V is a line spanned by a vector $\mathbf{y}$ then $\mathbf{p} = \dfrac{\mathbf{x} \cdot \mathbf{y}}{\mathbf{y} \cdot \mathbf{y}} \mathbf{y}$.

Theorem 2 $\|\mathbf{x} - \mathbf{v}\| > \|\mathbf{x} - \mathbf{p}\|$ for any $\mathbf{v} \neq \mathbf{p}$ in V.

Thus $\|\mathbf{o}\| = \|\mathbf{x} - \mathbf{p}\| = \min_{\mathbf{v} \in V} \|\mathbf{x} - \mathbf{v}\|$ is the **distance** from the vector $\mathbf{x}$ to the subspace V.



$V^{\perp}$
$\mathbf{x}$
$\mathbf{o}$
$\mathbf{p}$
V

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

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

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

We have $\mathbf{x} = \mathbf{p} + \mathbf{o}$, where $\mathbf{p} \in \Pi$ and $\mathbf{o} \perp \Pi$. Then the orthogonal projection of $\mathbf{x}$ onto Π is $\mathbf{p}$ and the distance from $\mathbf{x}$ to Π is $\|\mathbf{o}\|$.

We have $\mathbf{p} = \alpha\mathbf{v}_1 + \beta\mathbf{v}_2$ for some $\alpha, \beta \in \mathbb{R}$. Then $\mathbf{o} = \mathbf{x} - \mathbf{p} = \mathbf{x} - \alpha\mathbf{v}_1 - \beta\mathbf{v}_2$.

$$\begin{cases} \mathbf{o} \cdot \mathbf{v}_1 = 0 \\ \mathbf{o} \cdot \mathbf{v}_2 = 0 \end{cases} \iff \begin{cases} \alpha(\mathbf{v}_1 \cdot \mathbf{v}_1) + \beta(\mathbf{v}_2 \cdot \mathbf{v}_1) = \mathbf{x} \cdot \mathbf{v}_1 \\ \alpha(\mathbf{v}_1 \cdot \mathbf{v}_2) + \beta(\mathbf{v}_2 \cdot \mathbf{v}_2) = \mathbf{x} \cdot \mathbf{v}_2 \end{cases}$$

$\mathbf{x} = (4, 0, -1), \;\; \mathbf{v}_1 = (1, 1, 0), \;\; \mathbf{v}_2 = (0, 1, 1)$

$$\begin{cases} \alpha(\mathbf{v}_1 \cdot \mathbf{v}_1) + \beta(\mathbf{v}_2 \cdot \mathbf{v}_1) = \mathbf{x} \cdot \mathbf{v}_1 \\ \alpha(\mathbf{v}_1 \cdot \mathbf{v}_2) + \beta(\mathbf{v}_2 \cdot \mathbf{v}_2) = \mathbf{x} \cdot \mathbf{v}_2 \end{cases}$$

$$\iff \begin{cases} 2\alpha + \beta = 4 \\ \alpha + 2\beta = -1 \end{cases} \iff \begin{cases} \alpha = 3 \\ \beta = -2 \end{cases}$$

$$\mathbf{p} = 3\mathbf{v}_1 - 2\mathbf{v}_2 = (3, 1, -2)$$
$$\mathbf{o} = \mathbf{x} - \mathbf{p} = (1, -1, 1)$$
$$\|\mathbf{o}\| = \sqrt{3}$$

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

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

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

Alternative solution: We have $\mathbf{x} = \mathbf{p} + \mathbf{o}$, where $\mathbf{p} \in \Pi$ and $\mathbf{o} \perp \Pi$. Then the orthogonal projection of $\mathbf{x}$ onto Π is $\mathbf{p}$ and the distance from $\mathbf{x}$ to Π is $\|\mathbf{o}\|$.

Notice that $\mathbf{o}$ is the orthogonal projection of $\mathbf{x}$ onto the orthogonal complement $\Pi^\perp$. In the previous lecture, we found that $\Pi^\perp$ is the line spanned by the vector $\mathbf{y} = (1, -1, 1)$. It follows that

$$\mathbf{o} = \frac{\mathbf{x} \cdot \mathbf{y}}{\mathbf{y} \cdot \mathbf{y}}\,\mathbf{y} = \frac{3}{3}\,(1, -1, 1) = (1, -1, 1).$$

Then $\mathbf{p} = \mathbf{x} - \mathbf{o} = (4, 0, -1) - (1, -1, 1) = (3, 1, -2)$ and $\|\mathbf{o}\| = \sqrt{3}$.

Overdetermined system of linear equations:

$$\begin{cases} x + 2y = 3 \\ 3x + 2y = 5 \\ x + y = 2.09 \end{cases} \iff \begin{cases} x + 2y = 3 \\ -4y = -4 \\ -y = -0.91 \end{cases}$$

No solution: inconsistent system

Assume that a solution (x_0, y_0) does exist but the system is not quite accurate, namely, there may be some errors in the right-hand sides.

Problem. Find a good approximation of (x_0, y_0).

One approach is the **least squares fit**. Namely, we look for a pair (x, y) that minimizes the sum $(x + 2y - 3)^2 + (3x + 2y - 5)^2 + (x + y - 2.09)^2$.

Least squares solution

System of linear equations:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \iff A\mathbf{x} = \mathbf{b}$$

For any $\mathbf{x} \in \mathbb{R}^n$ define a **residual** $r(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$.

The **least squares solution x** to the system is the one that minimizes $\|r(\mathbf{x})\|$ (or, equivalently, $\|r(\mathbf{x})\|^2$).

$$\|r(\mathbf{x})\|^2 = \sum_{i=1}^{m} (a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n - b_i)^2$$

Let A be an $m \times n$ matrix and let $\mathbf{b} \in \mathbb{R}^m$.

Theorem A vector $\hat{\mathbf{x}}$ is a least squares solution of the system $A\mathbf{x} = \mathbf{b}$ if and only if it is a solution of the associated **normal system** $\boxed{A^T A\mathbf{x} = A^T\mathbf{b}.}$

Proof: $A\mathbf{x}$ is an arbitrary vector in $R(A)$, the column space of A. Hence the length of $r(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$ is minimal if $A\mathbf{x}$ is the orthogonal projection of $\mathbf{b}$ onto $R(A)$. That is, if $r(\mathbf{x})$ is orthogonal to $R(A)$.

We know that $\{\text{row space}\}^\perp = \{\text{nullspace}\}$ for any matrix. In particular, $R(A)^\perp = N(A^T)$, the nullspace of the transpose matrix of A. Thus $\hat{\mathbf{x}}$ is a least squares solution if and only if

$$A^T r(\hat{\mathbf{x}}) = \mathbf{0} \iff A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0} \iff A^T A\hat{\mathbf{x}} = A^T\mathbf{b}.$$

Corollary The normal system $A^T A\mathbf{x} = A^T\mathbf{b}$ is always consistent.

Problem. Find the least squares solution to

$$\begin{cases} x + 2y = 3 \\ 3x + 2y = 5 \\ x + y = 2.09 \end{cases}$$

$$\begin{pmatrix} 1 & 2 \\ 3 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \\ 2.09 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 3 & 1 \\ 2 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 3 & 1 \\ 2 & 2 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 5 \\ 2.09 \end{pmatrix}$$

$$\begin{pmatrix} 11 & 9 \\ 9 & 9 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 20.09 \\ 18.09 \end{pmatrix} \iff \begin{cases} x = 1 \\ y = 1.01 \end{cases}$$

Problem. Find the constant function that is the least squares fit to the following data

x	0	1	2	3
$f(x)$	1	0	1	2

$$f(x) = c \implies \begin{cases} c = 1 \\ c = 0 \\ c = 1 \\ c = 2 \end{cases} \implies \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} (c) = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$$(1,1,1,1) \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} (c) = (1,1,1,1) \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$c = \frac{1}{4}(1+0+1+2) = 1$ (mean arithmetic value)

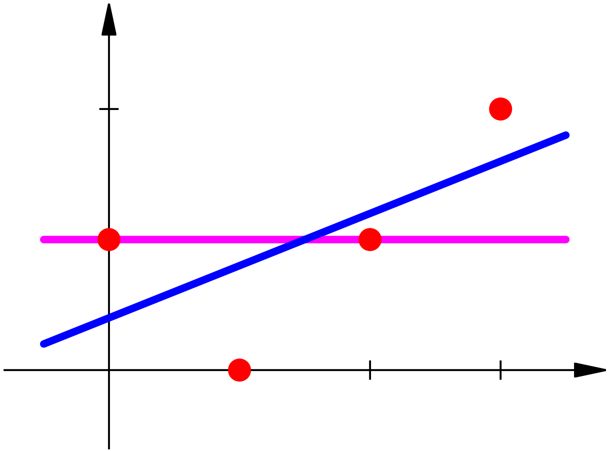
Problem. Find the linear polynomial that is the least squares fit to the following data

x	0	1	2	3
$f(x)$	1	0	1	2

$$f(x) = c_1 + c_2 x \implies \begin{cases} c_1 = 1 \\ c_1 + c_2 = 0 \\ c_1 + 2c_2 = 1 \\ c_1 + 3c_2 = 2 \end{cases} \implies \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 4 & 6 \\ 6 & 14 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 8 \end{pmatrix} \iff \begin{cases} c_1 = 0.4 \\ c_2 = 0.4 \end{cases}$$



Problem. Find the quadratic polynomial that is the least squares fit to the following data

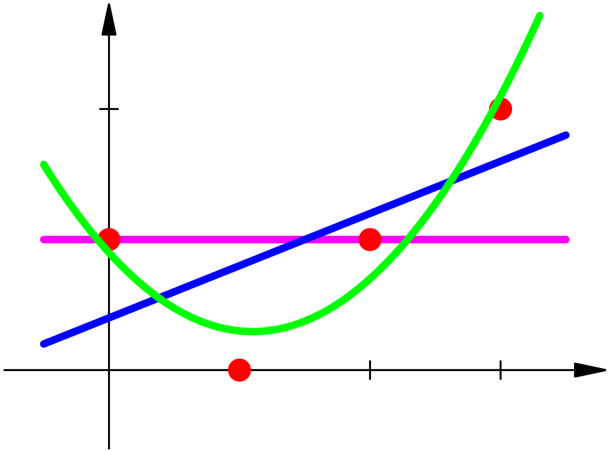
x	0	1	2	3
$f(x)$	1	0	1	2

$$f(x) = c_1 + c_2 x + c_3 x^2$$

$$\Longrightarrow \begin{cases} c_1 = 1 \\ c_1 + c_2 + c_3 = 0 \\ c_1 + 2c_2 + 4c_3 = 1 \\ c_1 + 3c_2 + 9c_3 = 2 \end{cases} \Longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 4 & 6 & 14 \\ 6 & 14 & 36 \\ 14 & 36 & 98 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 8 \\ 22 \end{pmatrix} \Longleftrightarrow \begin{cases} c_1 = 0.9 \\ c_2 = -1.1 \\ c_3 = 0.5 \end{cases}$$



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

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

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

$$|x_i + y_i| \le |x_i| + |y_i| \le \max_j |x_j| + \max_j |y_j|$$
$$\implies \max_j |x_j + y_j| \le \max_j |x_j| + \max_j |y_j|$$
$$\implies \|\mathbf{x} + \mathbf{y}\|_\infty \le \|\mathbf{x}\|_\infty + \|\mathbf{y}\|_\infty.$$

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

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

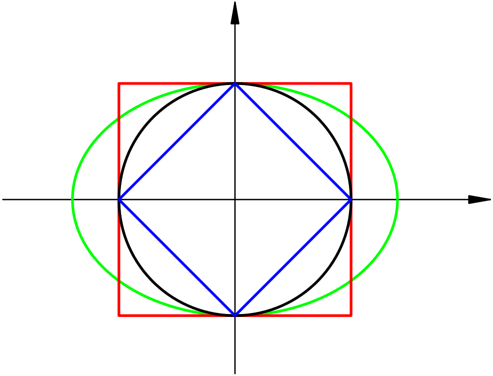
Definition. A **normed vector space** is a vector space endowed with a norm.

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

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

Also, we say that a sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots$ *converges* to a vector $\mathbf{x}$ if $\mathrm{dist}(\mathbf{x}, \mathbf{x}_n) \to 0$ 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 $p \ge 1$.