Introduction to Optimization

Lecture 02: Calculus in $\mathbb{R}^N$.

Real vectors and their norms

$\mathbb{R}^N$ is the (real) vector space of N-tuples of real numbers (columns)

$$x \in \mathbb{R}^N \qquad \iff \qquad x = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}, \quad x_1, \ldots, x_N \in \mathbb{R}.$$

Real vectors and their norms

$\mathbb{R}^N$ is the (real) vector space of N-tuples of real numbers (columns)

$$x \in \mathbb{R}^N \qquad \Longleftrightarrow \qquad x = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}, \quad x_1, \dots, x_N \in \mathbb{R}.$$

The norm of $x \in \mathbb{R}^N$ is $\|x\| = \sqrt{x_1^2 + \cdots + x_N^2}$.

Real vectors and their norms

$\mathbb{R}^N$ is the (real) vector space of N-tuples of real numbers (columns)

$$x \in \mathbb{R}^N \qquad \Longleftrightarrow \qquad x = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}, \quad x_1, \ldots, x_N \in \mathbb{R}.$$

The norm of $x \in \mathbb{R}^N$ is $\|x\| = \sqrt{x_1^2 + \cdots + x_N^2}$.

Properties

- $\|x\| > 0$ for all $x \neq 0$ and $\|0\| = 0$.
- $\|\alpha x\| = |\alpha| \, \|x\|$ for all $x \in \mathbb{R}^N$ and $\alpha \in \mathbb{R}$.
- $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathbb{R}^N$.

Distances and balls

The distance between $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^N$ is $\mathrm{dist}(x, y) = \|x - y\|$.

Distances and balls

The distance between $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^N$ is $\mathrm{dist}(x, y) = \|x - y\|$.

Properties

- $\mathrm{dist}(x, y) > 0$ for all $x \neq y$ and $\mathrm{dist}(x, x) = 0$.

Distances and balls

The distance between $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^N$ is $\text{dist}(x, y) = \|x - y\|$.

Properties

- $\text{dist}(x, y) > 0$ for all $x \neq y$ and $\text{dist}(x, x) = 0$.
- $\text{dist}(x, y) = \text{dist}(y, x)$.

Distances and balls

The distance between $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^N$ is $\operatorname{dist}(x,y) = \|x - y\|$.

Properties

- $\operatorname{dist}(x,y) > 0$ for all $x \neq y$ and $\operatorname{dist}(x,x) = 0$.
- $\operatorname{dist}(x,y) = \operatorname{dist}(y,x)$.
- $\operatorname{dist}(x,y) \leq \operatorname{dist}(x,z) + \operatorname{dist}(z,y)$.

Distances and balls

The distance between $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^N$ is $\mathrm{dist}(x,y) = \|x-y\|$.

Properties

- $\mathrm{dist}(x,y) > 0$ for all $x \neq y$ and $\mathrm{dist}(x,x) = 0$.
- $\mathrm{dist}(x,y) = \mathrm{dist}(y,x)$.
- $\mathrm{dist}(x,y) \leq \mathrm{dist}(x,z) + \mathrm{dist}(z,y)$.

The open ball centered at $x \in \mathbb{R}^N$ with radius $r > 0$ is

$$B(x;r) = \{y \in \mathbb{R}^N \ : \ \mathrm{dist}(x,y) < r\}.$$

Distances and balls

The distance between $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^N$ is $\mathrm{dist}(x, y) = \|x - y\|$.

Properties

- $\mathrm{dist}(x, y) > 0$ for all $x \neq y$ and $\mathrm{dist}(x, x) = 0$.
- $\mathrm{dist}(x, y) = \mathrm{dist}(y, x)$.
- $\mathrm{dist}(x, y) \leq \mathrm{dist}(x, z) + \mathrm{dist}(z, y)$.

The open ball centered at $x \in \mathbb{R}^N$ with radius $r > 0$ is

$$B(x; r) = \{y \in \mathbb{R}^N \; : \; \mathrm{dist}(x, y) < r\}.$$

The closed ball centered at $x \in \mathbb{R}^N$ with radius $r > 0$ is

$$\bar{B}(x; r) = \{y \in \mathbb{R}^N \; : \; \mathrm{dist}(x, y) \leq r\}.$$

Topology I

A subset $A \subset \mathbb{R}^N$ is open if, for each $x \in A$, there is $r > 0$ such that $B(x; r) \subset A$.

Topology I

A subset $A \subset \mathbb{R}^N$ is open if, for each $x \in A$, there is $r > 0$ such that $B(x; r) \subset A$.

A subset $C \subset \mathbb{R}^N$ is closed if its complement is open.

Topology I

A subset $A \subset \mathbb{R}^N$ is open if, for each $x \in A$, there is $r > 0$ such that $B(x;r) \subset A$.

A subset $C \subset \mathbb{R}^N$ is closed if its complement is open.

Example

Open balls are open sets. Closed balls are closed sets.

Topology I

A subset $A \subset \mathbb{R}^N$ is open if, for each $x \in A$, there is $r > 0$ such that $B(x; r) \subset A$.

A subset $C \subset \mathbb{R}^N$ is closed if its complement is open.

Example

Open balls are open sets. Closed balls are closed sets.

Proposition

If a sequence in a closed set is convergent, its limit has to be in the set.

Topology II

A subset $B \subset \mathbb{R}^N$ is bounded if it is contained in a ball.

Topology II

A subset $B \subset \mathbb{R}^N$ is bounded if it is contained in a ball.

Proposition

Every bounded sequence in $\mathbb{R}^N$ has a convergent subsequence.

Topology II

A subset $B \subset \mathbb{R}^N$ is bounded if it is contained in a ball.

Proposition

Every bounded sequence in $\mathbb{R}^N$ has a convergent subsequence.

Finally, a subset $K \subset \mathbb{R}^N$ is compact if it is both closed and bounded.

Topology II

A subset $B \subset \mathbb{R}^N$ is bounded if it is contained in a ball.

Proposition

Every bounded sequence in $\mathbb{R}^N$ has a convergent subsequence.

Finally, a subset $K \subset \mathbb{R}^N$ is compact if it is both closed and bounded.

Proposition

Every sequence in a compact subset of $\mathbb{R}^N$ has a convergent subsequence. The limits of all convergent subsequences must lie in the set.

Dot product

The dot product (or also inner product) of $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^N$ is

$$x \cdot y = x_1 y_1 + \cdots + x_N y_N.$$

Dot product

The dot product (or also inner product) of $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^N$ is

$$x \cdot y = x_1 y_1 + \cdots + x_N y_N.$$

Properties

- $x \cdot x = \|x\|^2$ for all $x \in \mathbb{R}^N$.

Dot product

The dot product (or also inner product) of $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^N$ is

$$x \cdot y = x_1 y_1 + \cdots + x_N y_N.$$

Properties

- $x \cdot x = \|x\|^2$ for all $x \in \mathbb{R}^N$.
- $x \cdot y = y \cdot x$ for all $x, y \in \mathbb{R}^N$.

Dot product

The dot product (or also inner product) of $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^N$ is

$$x \cdot y = x_1 y_1 + \cdots + x_N y_N.$$

Properties

- $x \cdot x = \|x\|^2$ for all $x \in \mathbb{R}^N$.
- $x \cdot y = y \cdot x$ for all $x, y \in \mathbb{R}^N$.
- $(\alpha x + z) \cdot y = \alpha(x \cdot y) + z \cdot y$ for all $x, y, z \in \mathbb{R}^N$ and $\alpha \in \mathbb{R}$.

Dot product

The dot product (or also inner product) of $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^N$ is

$$x \cdot y = x_1 y_1 + \cdots + x_N y_N.$$

Properties

- $x \cdot x = \|x\|^2$ for all $x \in \mathbb{R}^N$.
- $x \cdot y = y \cdot x$ for all $x, y \in \mathbb{R}^N$.
- $(\alpha x + z) \cdot y = \alpha(x \cdot y) + z \cdot y$ for all $x, y, z \in \mathbb{R}^N$ and $\alpha \in \mathbb{R}$.

Other notations: $x \cdot y = \langle x, y \rangle = \langle x \mid y \rangle = x^T y$ (product of matrices).

Perpendicularity and parallelism

Proposition

$|x \cdot y| \leq \|x\| \, \|y\|$ *for all* $x, y \in \mathbb{R}^N$.

Perpendicularity and parallelism

Proposition

$|x \cdot y| \leq \|x\| \, \|y\|$ *for all* $x, y \in \mathbb{R}^N$.

The vectors $x, y \in \mathbb{R}^N \setminus \{0\}$ are perpendicular or orthogonal if $x \cdot y = 0$. In that case, we write $x \perp y$.

Perpendicularity and parallelism

Proposition

$|x \cdot y| \leq \|x\| \, \|y\|$ *for all* $x, y \in \mathbb{R}^N$.

The vectors $x, y \in \mathbb{R}^N \setminus \{0\}$ are perpendicular or orthogonal if $x \cdot y = 0$. In that case, we write $x \perp y$.

The vectors $x, y \in \mathbb{R}^N \setminus \{0\}$ are parallel if there is $\alpha \in \mathbb{R}$ such that $x = \alpha y$. We write $x \parallel y$.

Perpendicularity and parallelism

Proposition

$|x \cdot y| \leq \|x\| \, \|y\|$ *for all* $x, y \in \mathbb{R}^N$.

The vectors $x, y \in \mathbb{R}^N \setminus \{0\}$ are perpendicular or orthogonal if $x \cdot y = 0$. In that case, we write $x \perp y$.

The vectors $x, y \in \mathbb{R}^N \setminus \{0\}$ are parallel if there is $\alpha \in \mathbb{R}$ such that $x = \alpha y$. We write $x \parallel y$.

Exercise

Show that $x \parallel y$ if, and only if, $|x \cdot y| = \|x\| \, \|y\|$.

Angles and triangles

The angle θ between $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^N$ is

$$\cos^{-1}\left(\frac{x \cdot y}{\|x\|\,\|y\|}\right).$$

Angles and triangles

The angle θ between $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^N$ is

$$\cos^{-1}\left(\frac{x \cdot y}{\|x\|\,\|y\|}\right).$$

Law of cosines

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2 - 2\|x\|\,\|y\|\cos(\theta).$$

Angles and triangles

The angle θ between $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^N$ is

$$\cos^{-1}\left(\frac{x \cdot y}{\|x\|\,\|y\|}\right).$$

Law of cosines

$\|x+y\|^2 = \|x\|^2 + \|y\|^2 - 2\|x\|\,\|y\|\cos(\theta).$

Pythagoras's Theorem

$x \perp y$ if, and only if, $\|x+y\|^2 = \|x\|^2 + \|y\|^2$.

Convex sets

A subset $\subset \mathbb{R}^N C$ is convex if $\lambda x + (1-\lambda)y \in C$ whenever $x, y \in C$ and $\lambda \in [0,1]$.

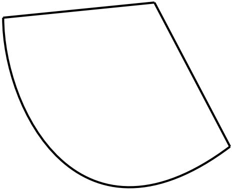
Convex sets

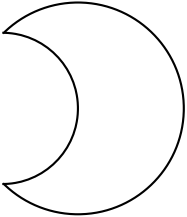
A subset $\subset \mathbb{R}^N C$ is convex if $\lambda x + (1-\lambda)y \in C$ whenever $x, y \in C$ and $\lambda \in [0,1]$. In other words, if the segment joining any two points of C also belongs to C.

Convex sets

A subset $\subset \mathbb{R}^N C$ is convex if $\lambda x + (1-\lambda)y \in C$ whenever $x, y \in C$ and $\lambda \in [0,1]$. In other words, if the segment joining any two points of C also belongs to C.

This set is convex





This one is not

Projection

Theorem

Let $C \subset \mathbb{R}^N$ be nonempty, closed and convex. For each $x \in \mathbb{R}^N$, there is a unique point $\hat{x} \in C$ such that

$$\operatorname{dist}(x, \hat{x}) = \min\{\operatorname{dist}(x, y) \; : \; y \in C\}.$$

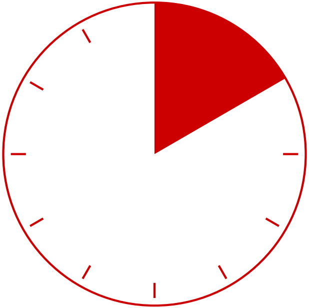
Moreover, $\hat{x}$ is the only point in C that satisfies the inequality

$$(x - \hat{x}) \cdot (y - \hat{x}) \leq 0$$

for all $y \in C$.

The point $\hat{x}$ is the projection of x onto C, and is denoted by $P_C(x)$.

Break



Differentiability and gradient

Let $A \subset \mathbb{R}^N$ be nonempty and open. A function $f : A \to \mathbb{R}$ is differentiable at $x \in A$ (in the sense of Gâteaux) if the directional derivative

$$f'(x; h) = \lim_{t \to 0} \frac{f(x + th) - f(x)}{t}$$

exists for all $h \in \mathbb{R}^N$, and there is $g \in \mathbb{R}^N$ such that

$$g \cdot h = f'(x; h)$$

for all $h \in \mathbb{R}^N$. If this case, the gradient of f at x is $\nabla f(x) = g$.

Differentiability and gradient

Let $A \subset \mathbb{R}^N$ be nonempty and open. A function $f : A \to \mathbb{R}$ is differentiable at $x \in A$ (in the sense of Gâteaux) if the directional derivative

$$f'(x; h) = \lim_{t \to 0} \frac{f(x + th) - f(x)}{t}$$

exists for all $h \in \mathbb{R}^N$, and there is $g \in \mathbb{R}^N$ such that

$$g \cdot h = f'(x; h)$$

for all $h \in \mathbb{R}^N$. If this case, the gradient of f at x is $\nabla f(x) = g$.

As usual, f is differentiable on A if it is so at every point of A.

More about the gradient

Remark

Let f be differentiable at x, and let $g = \nabla f(x)$ be its gradient at that point. If e_i denotes the i-th canonical vector in $\mathbb{R}^N$, then

$$g_i = \nabla f(x) \cdot e_i = f'(x; e_i) = \lim_{t \to 0} \frac{f(x + te_i) - f(x)}{t} = \frac{\partial f}{\partial x_i}(x).$$

More about the gradient

Remark

Let f be differentiable at x, and let $g = \nabla f(x)$ be its gradient at that point. If e_i denotes the i-th canonical vector in $\mathbb{R}^N$, then

$$g_i = \nabla f(x) \cdot e_i = f'(x; e_i) = \lim_{t \to 0} \frac{f(x + te_i) - f(x)}{t} = \frac{\partial f}{\partial x_i}(x).$$

Example

Let us compute the gradient of the function $f : \mathbb{R}^N \to \mathbb{R}$, defined by

$$f(x) = \frac{1}{2} \|Ax - y\|^2,$$

where A is a real matrix of size $M \times N$ and $b \in \mathbb{R}^M$.

First order optimality condition

Theorem (Fermat's Rule)

Let $f : A \subset \mathbb{R}^N \to \mathbb{R}$ *and let* $\emptyset \neq C \subset A$ *be convex. If* $\hat{x} \in C$ *is such that* $f(\hat{x}) \leq f(y)$ *for all* $y \in C$, *and if* f *is differentiable at* $\hat{x}$, *then*

$$\nabla f(\hat{x}) \cdot (y - \hat{x}) \geq 0$$

for all $y \in C$.

First order optimality condition

Theorem (Fermat's Rule)

Let $f : A \subset \mathbb{R}^N \to \mathbb{R}$ and let $\emptyset \neq C \subset A$ be convex. If $\hat{x} \in C$ is such that $f(\hat{x}) \leq f(y)$ for all $y \in C$, and if f is differentiable at $\hat{x}$, then

$$\nabla f(\hat{x}) \cdot (y - \hat{x}) \geq 0$$

for all $y \in C$.

If, moreover, $\hat{x} \in \mathrm{int}(C)$, then $\nabla f(\hat{x}) = 0$.

L-smoothness

A differentiable function $f : A \subset \mathbb{R}^N \to \mathbb{R}$ is L-smooth, with $L > 0$, if

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$$

for all $x, y \in A$.

L-smoothness

A differentiable function $f : A \subset \mathbb{R}^N \to \mathbb{R}$ is L-smooth, with $L > 0$, if

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$$

for all $x, y \in A$.

Proposition (Descent Lemma)

If f is L-smooth and A is convex, then

$$|f(y) - f(x) - \nabla f(x) \cdot (y - x)| \leq \frac{L}{2}\|x - y\|^2$$

for all $x, y \in A$.

Second order conditions

Proposition (Taylor's Approximation)

Let $f : \mathrm{dom}(f) \to \mathbb{R}$ be twice differentiable at x. For each $d \in \mathbb{R}^N$,

$$\lim_{t \to 0} \frac{1}{t^2} \left| f(x+td) - f(x) - t\langle \nabla f(x), d\rangle - \frac{t^2}{2}\langle \nabla^2 f(x)d, d\rangle \right| = 0.$$

Second order conditions

Proposition (Taylor's Approximation)

Let $f : \mathrm{dom}(f) \to \mathbb{R}$ be twice differentiable at x. For each $d \in \mathbb{R}^N$,

$$\lim_{t \to 0} \frac{1}{t^2} \left| f(x+td) - f(x) - t\langle \nabla f(x), d\rangle - \frac{t^2}{2}\langle \nabla^2 f(x)d, d\rangle \right| = 0.$$

Theorem (Second order optimality conditions)

Let $f : \mathrm{dom}(f) \to \mathbb{R}$ be twice differentiable at $\hat{x}$.

i) *If $\hat{x}$ is a local minimizer of f, then $\nabla f(\hat{x}) = 0$ and $\nabla^2 f(\hat{x}) \geq 0$.*

Second order conditions

Proposition (Taylor's Approximation)

Let $f : \mathrm{dom}(f) \to \mathbb{R}$ *be twice differentiable at* x. *For each* $d \in \mathbb{R}^N$,

$$\lim_{t \to 0} \frac{1}{t^2} \left| f(x+td) - f(x) - t\langle \nabla f(x), d\rangle - \frac{t^2}{2}\langle \nabla^2 f(x)d, d\rangle \right| = 0.$$

Theorem (Second order optimality conditions)

Let $f : \mathrm{dom}(f) \to \mathbb{R}$ *be twice differentiable at* $\hat{x}$.

i) *If* $\hat{x}$ *is a local minimizer of* f, *then* $\nabla f(\hat{x}) = 0$ *and* $\nabla^2 f(\hat{x}) \geq 0$.

ii) *If* $\nabla f(\hat{x}) = 0$ *and* $\nabla^2 f(\hat{x}) > 0$, *then* $\hat{x}$ *is a strict local minimizer of* f.