An Introduction to Non-smooth Optimization

Lecture 01 - Mathematical Background

Jingwei LIANG

Institute of Natural Sciences, Shanghai Jiao Tong University

Email: optimization.sjtu@gmail.com

Office: Room 355, No. 6 Science Building

Outline

1. Vector spaces
2. Convex sets
3. Non-expansive operators
4. Fejér monotonicity
5. Convex functions
6. Differentiability
7. Convex minimization problem

Vector

Let $\mathbb{R}^n$ be the n-dimensional *real vector space*, a column vector of $\mathbb{R}^n$ is denoted by $\boldsymbol{a} \in \mathbb{R}^n$, with

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

- The number a_i is called the i'th element/component of the vector $\boldsymbol{a}$.

NB: By default we refer vector as column vector.

Matrix

A matrix with m rows and n columns is called an $m \times n$ matrix and denoted by

$$\boldsymbol{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix} \in \mathbb{R}^{m \times n}.$$

- The identity matrix of size n is a diagonal matrix

$$\boldsymbol{Id}_n = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}_{n \times n}.$$

Definition - Vector inner product

Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, their inner product or dot product returns a scalar

$$\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle = \sum_{i=1}^{n} x_i y_i.$$

- Alternative notation

$$\boldsymbol{x}^T \boldsymbol{y}.$$

- Given any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, their distance is

$$\|\boldsymbol{x} - \boldsymbol{y}\| = \sqrt{\langle \boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{x} - \boldsymbol{y} \rangle}.$$

Vector norm

Definition - Vector p-norm

Let $\boldsymbol{x} \in \mathbb{R}^n$ be a vector and $p \geq 1$, then the p-norm (also called ℓ_p-norm) of $\boldsymbol{x}$ is defined by

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}.$$

Definition - Vector p-norm

Let $\boldsymbol{x} \in \mathbb{R}^n$ be a vector and $p \geq 1$, then the p-norm (also called ℓ_p-norm) of $\boldsymbol{x}$ is defined by

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}.$$

A norm must satisfies

- Positivity: $\|\boldsymbol{x}\|_p \geq 0$, $\|\boldsymbol{x}\|_p = 0$ if and only if $\boldsymbol{x} = \mathbf{0}$.
- Homogeneity: $\|r\boldsymbol{x}\|_p = |r|\|\boldsymbol{x}\|_p, \ r \in \mathbb{R}$.
- Triangle inequality: $\|\boldsymbol{x}+\boldsymbol{y}\|_p \leq \|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p$.

Vector norm

Example - ℓ_2-norm (Euclidean norm)

Let $p = 2$ we obtain the Euclidean norm of $\boldsymbol{x}$

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

- $\|\boldsymbol{x}\|$ without subscript 2 is also used to denote ℓ_2-norm.

Example - ℓ_2-norm (Euclidean norm)

Let $p = 2$ we obtain the Euclidean norm of $\boldsymbol{x}$

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

- $\|\boldsymbol{x}\|$ without subscript 2 is also used to denote ℓ_2-norm.

Example - ℓ_1-norm

Let $p = 1$ we obtain the ℓ_1-norm of $\boldsymbol{x}$

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

Example - ℓ_2-norm (Euclidean norm)

Let $p = 2$ we obtain the Euclidean norm of $\boldsymbol{x}$

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

- $\|\boldsymbol{x}\|$ without subscript 2 is also used to denote ℓ_2-norm.

Example - ℓ_1-norm

Let $p = 1$ we obtain the ℓ_1-norm of $\boldsymbol{x}$

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

Example - ℓ_∞-norm

The infinity norm of $\boldsymbol{x}$ is defined by

$$\|\boldsymbol{x}\|_\infty = \max_{i=1,\dots,n} |x_i|.$$

Definition - Vector inner product

Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, their inner product or dot product returns a scalar

$$\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle = \sum_{i=1}^{n} x_i y_i.$$

Theorem - Cauchy-Schwarz inequality

For any two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^n$, the Cauchy-Schwarz inequality

$$|\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|$$

holds. Furthermore, equality holds if and only if $\boldsymbol{x} = \alpha \boldsymbol{y}$ for some $\alpha \in \mathbb{R}$.

Definition - Dual norm

Let $\|\cdot\|$ be a norm defined on $\mathbb{R}^n$, the associated *dual norm*, denoted by $\|\cdot\|_*$ is defined as

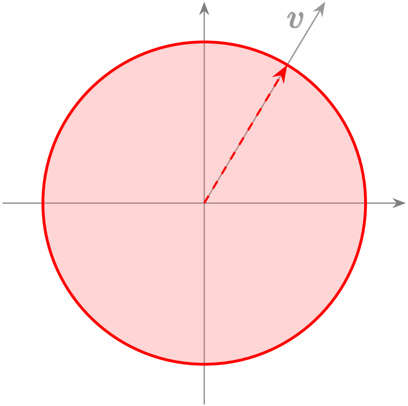
$$\|\boldsymbol{v}\|_* = \sup\left\{\langle \boldsymbol{v} \mid \boldsymbol{x}\rangle \;:\; \|\boldsymbol{x}\| \leq 1\right\}.$$

Dual norm

Definition - Dual norm

Let $\|\cdot\|$ be a norm defined on $\mathbb{R}^n$, the associated *dual norm*, denoted by $\|\cdot\|_*$ is defined as

$$\|\boldsymbol{v}\|_* = \sup\left\{\langle \boldsymbol{v} \mid \boldsymbol{x}\rangle \;:\; \|\boldsymbol{x}\| \leq 1\right\}.$$



The dual of the **Euclidean norm** is the **Euclidean norm**

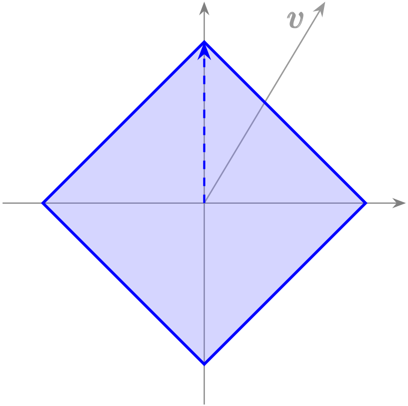
$$\sup\left\{\langle \boldsymbol{v} \mid \boldsymbol{x}\rangle \;:\; \|\boldsymbol{x}\|_2 \leq 1\right\} = \|\boldsymbol{v}\|_2.$$

Dual norm

Definition - Dual norm

Let $\|\cdot\|$ be a norm defined on $\mathbb{R}^n$, the associated *dual norm*, denoted by $\|\cdot\|_*$ is defined as

$$\|\mathbf{v}\|_* = \sup\left\{\langle \mathbf{v} \mid \mathbf{x}\rangle \ : \ \|\mathbf{x}\| \le 1\right\}.$$



The dual of the ℓ_1**-norm** is the ℓ_∞**-norm**

$$\sup\left\{\langle \mathbf{v} \mid \mathbf{x}\rangle \ : \ \|\mathbf{x}\|_1 \le 1\right\} = \|\mathbf{v}\|_\infty.$$

Recall that in $\mathbb{R}^2$

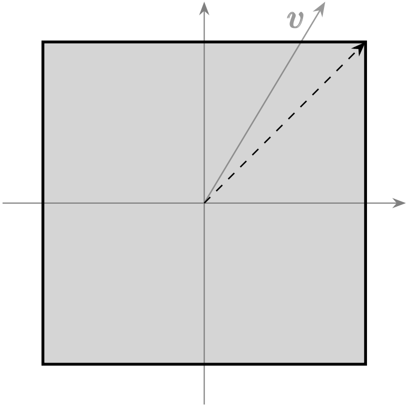
$$\langle \mathbf{v} \mid \mathbf{x}\rangle = v_1x_1 + v_2x_2.$$

Dual norm

Definition - Dual norm

Let $\|\cdot\|$ be a norm defined on $\mathbb{R}^n$, the associated *dual norm*, denoted by $\|\cdot\|_*$ is defined as

$$\|\boldsymbol{v}\|_* = \sup\left\{\langle \boldsymbol{v} \mid \boldsymbol{x}\rangle \ : \ \|\boldsymbol{x}\| \leq 1\right\}.$$



The dual of the ℓ_∞**-norm** is the ℓ_1**-norm**

$$\sup\left\{\langle \boldsymbol{v} \mid \boldsymbol{x}\rangle \ : \ \|\boldsymbol{x}\|_\infty \leq 1\right\} = \|\boldsymbol{v}\|_1.$$

Recall that in $\mathbb{R}^2$

$$\langle \boldsymbol{v} \mid \boldsymbol{x}\rangle = v_1x_1 + v_2x_2.$$

Dual norm

Proposition - Dual norm

Given $p, q \geq 1$, ℓ_p-norm and ℓ_q-norm are dual of each other if

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Dual norm

Proposition - Dual norm

Given $p, q \geq 1$, ℓ_p-norm and ℓ_q-norm are dual of each other if

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Theorem - generalized Cauchy-Schwarz inequality

Given any nonzero $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^n$, there holds

$$\langle \mathbf{v} \mid \mathbf{x}/\|\mathbf{x}\| \rangle \leq \sup \left\{ \langle \mathbf{v} \mid \mathbf{y} \rangle \; : \; \|\mathbf{y}\| \leq 1 \right\} = \|\mathbf{v}\|_* \quad \Longrightarrow \quad \langle \mathbf{v} \mid \mathbf{x} \rangle \leq \|\mathbf{v}\|_* \|\mathbf{x}\|$$

which holds for all $\mathbf{v}$ and $\mathbf{x}$.

- The inequality is **tight** in the sense that, for any $\mathbf{x}$ there exists a $\mathbf{v}$ such that the equality holds, and vice versa.

Convex sets

Definition, closedness

Definition - Convex set

A subset S of $\mathbb{R}^n$ is convex if for any $\boldsymbol{x}, \boldsymbol{y} \in S$ and $\lambda \in [0,1]$, there holds

$$\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in S.$$

$\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}$ is called the **convex combination** of $\boldsymbol{x}$ and $\boldsymbol{y}$.

Convex sets

Definition - Convex set

A subset S of $\mathbb{R}^n$ is convex if for any $\boldsymbol{x}, \boldsymbol{y} \in S$ and $\lambda \in [0, 1]$, there holds

$$\lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y} \in S.$$

$\lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y}$ is called the **convex combination** of $\boldsymbol{x}$ and $\boldsymbol{y}$.

Example - Hyper plane and half space

Given $\boldsymbol{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$,

- Hyper plane

$$H \stackrel{\text{def}}{=} \{\boldsymbol{x} \, : \, \boldsymbol{a}^T \boldsymbol{x} = b\}.$$

- Half space

$$H \stackrel{\text{def}}{=} \{\boldsymbol{x} \, : \, \boldsymbol{a}^T \boldsymbol{x} \leq b\}.$$

Proposition - Some properties

- Let S be a convex set, then $\beta S = \{\beta \mathbf{x} \ : \ \mathbf{x} \in S\}$ is convex.
- Let $S_i, i = 1, 2, \ldots, m$ be a family of convex sets, then
$$\bigcap_{i=1,2,\ldots,m} S_i$$
is convex.
- Let S_1, S_2 be two convex sets, then
$$S_1 + S_2 \quad \text{and} \quad S_1 - S_2$$
are convex.

Definition - Interior point

An element $\boldsymbol{x} \in S \subset \mathbb{R}^n$ is called an *interior point* of S if there $\exists \epsilon > 0$ for which

$$\left\{\boldsymbol{y} \ : \ \|\boldsymbol{y} - \boldsymbol{x}\| \leq \epsilon\right\} \subset S.$$

- The interior of S, *i.e.* $\mathrm{int}(S)$, denotes the set of all interior points of S.

Definition - Interior point

An element $\boldsymbol{x} \in S \subset \mathbb{R}^n$ is called an *interior point* of S if there $\exists \epsilon > 0$ for which

$$\left\{\boldsymbol{y} \ : \ \|\boldsymbol{y} - \boldsymbol{x}\| \leq \epsilon\right\} \subset S.$$

- The interior of S, *i.e.* $\operatorname{int}(S)$, denotes the set of all interior points of S.

A set S is **open** if $\operatorname{int}(S) = S$, it is **closed** if

$$\mathbb{R}^n \setminus S = \left\{\boldsymbol{x} \in \mathbb{R}^n \ : \ \boldsymbol{x} \notin S\right\}$$

is open.

Open and closed sets

Definition - Interior point

An element $\boldsymbol{x} \in S \subset \mathbb{R}^n$ is called an *interior point* of S if there $\exists \epsilon > 0$ for which

$$\left\{\boldsymbol{y} \ : \ \|\boldsymbol{y} - \boldsymbol{x}\| \leq \epsilon\right\} \subset S.$$

- The interior of S, *i.e.* $\mathrm{int}(S)$, denotes the set of all interior points of S.

A set S is **open** if $\mathrm{int}(S) = S$, it is **closed** if

$$\mathbb{R}^n \setminus S = \left\{\boldsymbol{x} \in \mathbb{R}^n \ : \ \boldsymbol{x} \notin S\right\}$$

is open.

The **closure** and **boundary** of S are defined as

$$\mathrm{cl}(S) = \mathbb{R}^n \setminus \mathrm{int}(\mathbb{R}^n \setminus S) \quad \text{and} \quad \mathrm{bd}(S) = \mathrm{cl}(S) \setminus \mathrm{int}(S).$$

Non-expansive operators

Non-expansiveness and fixed-point

Non-expansive operator

Definition - Non-expansive operator

Let S be a non-empty subset of $\mathbb{R}^n$ and let $\mathcal{F} : S \to \mathbb{R}^n$. Then $\mathcal{F}$ is *non-expansive* if it is Lipschitz continuous with constant 1, *i.e.*

$$(\forall \mathbf{x}, \mathbf{y} \in S) \quad \|\mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|.$$

Non-expansive operator

Definition - Non-expansive operator

Let S be a non-empty subset of $\mathbb{R}^n$ and let $\mathcal{F} : S \to \mathbb{R}^n$. Then $\mathcal{F}$ is *non-expansive* if it is Lipschitz continuous with constant 1, *i.e.*

$$(\forall \mathbf{x}, \mathbf{y} \in S) \quad \|\mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|.$$

Definition - Firmly non-expansive operator

$\mathcal{F}$ is *firmly non-expansive* if

$$(\forall \mathbf{x}, \mathbf{y} \in S) \quad \|\mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{y})\|^2 + \|(\mathbf{Id} - \mathcal{F})(\mathbf{x}) - (\mathbf{Id} - \mathcal{F})(\mathbf{y})\|^2 \leq \|\mathbf{x} - \mathbf{y}\|^2.$$

The following are equivalent

- $\mathcal{F}$ is firmly non-expansive.
- $\mathbf{Id} - \mathcal{F}$ is firmly non-expansive.
- $2\mathcal{F} - \mathbf{Id}$ is non-expansive.

Non-expansive operator

Definition - Non-expansive operator

Let S be a non-empty subset of $\mathbb{R}^n$ and let $\mathcal{F} : S \to \mathbb{R}^n$. Then $\mathcal{F}$ is *non-expansive* if it is Lipschitz continuous with constant 1, *i.e.*

$$(\forall \mathbf{x}, \mathbf{y} \in S) \quad \|\mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|.$$

Definition - Averaged non-expansiveness

Let S be a non-empty subset of $\mathbb{R}^n$ and let $\mathcal{F} : S \to \mathbb{R}^n$. Then $\mathcal{F}$ is α*-averaged non-expansive* if there exist $\alpha \in]0, 1[$ and a non-expansive operator $\mathcal{R}$ such that

$$\mathcal{F} = (1 - \alpha)\mathbf{Id} + \alpha\mathcal{R}.$$

Fixed-points of non-expansive operator

Definition - Non-expansive operator

Let S be a non-empty convex subset of $\mathbb{R}^n$ and $\mathcal{F} : S \to \mathbb{R}^n$ be a non-expansive operator, the set of fixed points of $\mathcal{F}$, denoted by $\mathrm{fix}(\mathcal{F})$, is defined by

$$\mathrm{fix}(\mathcal{F}) \stackrel{\mathrm{def}}{=} \left\{\boldsymbol{x} \in S \ : \ \boldsymbol{x} = \mathcal{F}(\boldsymbol{x})\right\}.$$

Fixed-points of non-expansive operator

Definition - Non-expansive operator

Let S be a non-empty convex subset of $\mathbb{R}^n$ and $\mathcal{F} : S \to \mathbb{R}^n$ be a non-expansive operator, the set of fixed points of $\mathcal{F}$, denoted by $\mathrm{fix}(\mathcal{F})$, is defined by

$$\mathrm{fix}(\mathcal{F}) \stackrel{\mathrm{def}}{=} \left\{ \boldsymbol{x} \in S \ : \ \boldsymbol{x} = \mathcal{F}(\boldsymbol{x}) \right\}.$$

Proposition - Convexity

Let S be a non-empty closed convex subset of $\mathbb{R}^n$ and let $\mathcal{F} : S \to \mathbb{R}^n$ be non-expansive, then the set of fixed points $\mathrm{fix}(\mathcal{F})$ is *closed and convex*.

Fixed-points of non-expansive operator

Definition - Non-expansive operator

Let S be a non-empty convex subset of $\mathbb{R}^n$ and $\mathcal{F} : S \to \mathbb{R}^n$ be a non-expansive operator, the set of fixed points of $\mathcal{F}$, denoted by $\mathrm{fix}(\mathcal{F})$, is defined by

$$\mathrm{fix}(\mathcal{F}) \stackrel{\text{def}}{=} \big\{ \boldsymbol{x} \in S \; : \; \boldsymbol{x} = \mathcal{F}(\boldsymbol{x}) \big\}.$$

Proposition - Convexity

Let S be a non-empty closed convex subset of $\mathbb{R}^n$ and let $\mathcal{F} : S \to \mathbb{R}^n$ be non-expansive, then the set of fixed points $\mathrm{fix}(\mathcal{F})$ is *closed and convex*.

Theorem - Browder-Göhde-Kirk

Let S be a non-empty bounded closed convex subset of $\mathbb{R}^n$ and $\mathcal{F} : S \to S$ be a non-expansive operator. Then

$$\mathrm{fix}(\mathcal{F}) \neq \emptyset.$$

Fejér monotonicity

Fejér monotonicity, fixed-point iteration

Sequence and limits

A number $x^\star \in \mathbb{R}$ is called the limit of the sequence $\{x^{(k)}\}_{k\in\mathbb{N}}$ if for any positive $\epsilon > 0$ there exists a number $\bar{k} > 0$ such that for all $k \geq \bar{k}$, there holds

$$|x^{(k)} - x^\star| < \epsilon.$$

That is, $x^{(k)} \in [x^\star - \epsilon, x^\star + \epsilon]$ for all $k \geq \bar{k}$. In this case, we write

$$x^\star = \lim_{k\to+\infty} x^{(k)}$$

or

$$x^{(k)} \to x^\star.$$

- A sequence that has a limit is called a convergent sequence.
- Extension to sequences in $\mathbb{R}^n$.

Sequence and limits

Limit of convergent sequence A convergent sequence has only one limit.

Sequence and limits

Limit of convergent sequence A convergent sequence has only one limit.

Boundedness and convergence Every convergent sequence is bounded.

Sequence and limits

Limit of convergent sequence A convergent sequence has only one limit.

Boundedness and convergence Every convergent sequence is bounded.

Monotonicity and convergence Every monotone bounded sequence in $\mathbb{R}$ is convergent.

Limit of convergent sequence A convergent sequence has only one limit.

Boundedness and convergence Every convergent sequence is bounded.

Monotonicity and convergence Every monotone bounded sequence in $\mathbb{R}$ is convergent.

Subsequence and convergence Any subsequence of a convergent sequence is convergent.

Sequence and limits

Limit of convergent sequence A convergent sequence has only one limit.

Boundedness and convergence Every convergent sequence is bounded.

Monotonicity and convergence Every monotone bounded sequence in $\mathbb{R}$ is convergent.

Subsequence and convergence Any subsequence of a convergent sequence is convergent.

Bolzano-Weierstrass Any bounded sequence has a convergent subsequence.

Definition - Fejér monotonicity

Let S be a non-empty subset of $\mathbb{R}^n$ and let $\{\mathbf{x}^{(k)}\}_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^n$. Then $\{\mathbf{x}^{(k)}\}_{k\in\mathbb{N}}$ is *Fejér monotone* with respect to S if

$$(\forall \mathbf{x} \in S)(k \in \mathbb{N}) \quad \|\mathbf{x}^{(k+1)} - \mathbf{x}\| \le \|\mathbf{x}^{(k)} - \mathbf{x}\|.$$

Definition - Fejér monotonicity

Let S be a non-empty subset of $\mathbb{R}^n$ and let $\{\boldsymbol{x}^{(k)}\}_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^n$. Then $\{\boldsymbol{x}^{(k)}\}_{k\in\mathbb{N}}$ is *Fejér monotone* with respect to S if

$$(\forall \boldsymbol{x} \in S)(k \in \mathbb{N}) \quad \|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}\| \leq \|\boldsymbol{x}^{(k)} - \boldsymbol{x}\|.$$

Theorem - Fejér monotonicity and convergence

Let S be a nonempty subset of $\mathbb{R}^n$ and let $\{\boldsymbol{x}^{(k)}\}_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^n$. Suppose that $\{\boldsymbol{x}^{(k)}\}_{k\in\mathbb{N}}$ is Fejér monotone with respect to S, then

- $\{\boldsymbol{x}^{(k)}\}_{k\in\mathbb{N}}$ is bounded. For every $\boldsymbol{x} \in S$, $\{\|\boldsymbol{x}^{(k)} - \boldsymbol{x}\|\}_{k\in\mathbb{N}}$ converges.

If every sequential cluster point of $\{\boldsymbol{x}^{(k)}\}_{k\in\mathbb{N}}$ belongs to S, then

- $\{\boldsymbol{x}^{(k)}\}_{k\in\mathbb{N}}$ converges to a point in S.

Definition - Fixed-point iteration

Let S be a nonempty closed convex subset of $\mathbb{R}^n$, let operator $\mathcal{F} : S \to S$ be non-expansive such that $\mathrm{fix}(\mathcal{F}) \neq \emptyset$. Let $\boldsymbol{x}^{(0)} \in S$, and set

$$(\forall k \in \mathbb{N}) \quad \boldsymbol{x}^{(k+1)} = \mathcal{F}(\boldsymbol{x}^{(k)}).$$

Suppose that $\boldsymbol{x}^{(k)} - \mathcal{F}(\boldsymbol{x}^{(k)}) \to \mathbf{0}$, then

- $\{\boldsymbol{x}^{(k)}\}_{k\in\mathbb{N}}$ converges to a point in $\mathrm{fix}(\mathcal{F})$.

- Only non-expansiveness does not guarantee convergence.

Theorem - Groetsch

Let S be a nonempty closed convex subset of $\mathbb{R}^n$, let operator $\mathcal{F} : S \to S$ be non-expansive such that $\mathrm{fix}(\mathcal{F}) \neq \emptyset$. Let $\{\lambda_k\}_{k\in\mathbb{N}}$ be a sequence in $[0,1]$ such that $\sum_k \lambda_k(1-\lambda_k) = +\infty$, and let $\boldsymbol{x}^{(0)} \in S$. Set

$$(\forall k \in \mathbb{N}) \quad \boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda_k\big(\mathcal{F}(\boldsymbol{x}^{(k)}) - \boldsymbol{x}^{(k)}\big).$$

Then the following hold

- $\{\boldsymbol{x}^{(k)}\}_{k\in\mathbb{N}}$ is Fejér monotone with respect to $\mathrm{fix}(\mathcal{F})$.
- $\{\mathcal{F}(\boldsymbol{x}^{(k)}) - \boldsymbol{x}^{(k)}\}_{k\in\mathbb{N}}$ converges to $\mathbf{0}$.
- $\{\boldsymbol{x}^{(k)}\}_{k\in\mathbb{N}}$ converges to a point in $\mathrm{fix}(\mathcal{F})$.

- When $\mathcal{F}$ is α-averaged non-expansive, then for $\{\lambda_k\}_{k\in\mathbb{N}}$, the condition changes to $\lambda_k \in [0, 1/\alpha]$ and

$$\sum_k \lambda_k \left(\frac{1}{\alpha} - \lambda_k\right) = +\infty.$$

Convex functions

Convex functions

Functions

Let $S \subset \mathbb{R}^n$, a function F is a mapping from S to $[-\infty, +\infty]$, *i.e.*

$$F \ : \ S \to [-\infty, +\infty].$$

- The *domain* of F is

$$\mathrm{dom}(F) \stackrel{\text{def}}{=} \Big\{ \boldsymbol{x} \in S \ : \ F(\boldsymbol{x}) < +\infty \Big\}.$$

- The *graph* of F is

$$\mathrm{gra}(F) \stackrel{\text{def}}{=} \Big\{ (\boldsymbol{x}, v) \in S \times \mathbb{R} \ : \ F(\boldsymbol{x}) = v \Big\}.$$

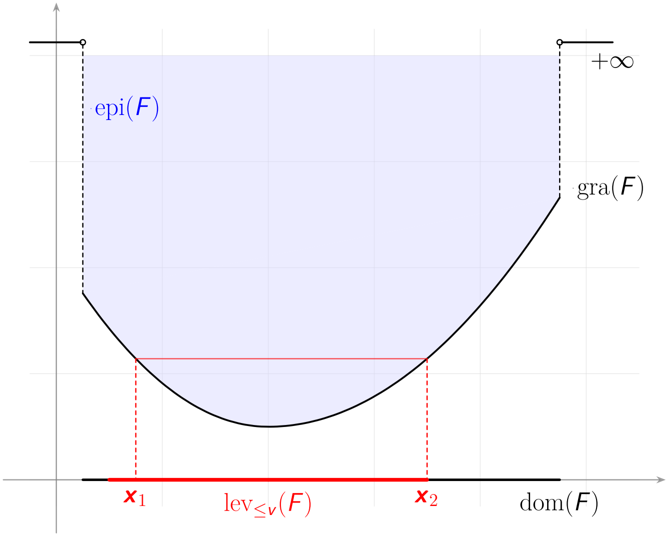
- The *epi graph* of F is

$$\mathrm{epi}(F) \stackrel{\text{def}}{=} \Big\{ (\boldsymbol{x}, v) \in S \times \mathbb{R} \ : \ F(\boldsymbol{x}) \leq v \Big\}.$$

- The *sub-level set* of F is

$$\mathrm{lev}_{\leq v}(F) \stackrel{\text{def}}{=} \Big\{ \boldsymbol{x} \in S \ : \ F(\boldsymbol{x}) \leq v \Big\}.$$

Functions



Closed function

Definition - Extended real line function

An *extended real-valued function* is a function defined over the entire underlying space that can take any real value, as well as the infinite values $-\infty$ and $+\infty$.

Closed function

Definition - Extended real line function

An *extended real-valued function* is a function defined over the entire underlying space that can take any real value, as well as the infinite values $-\infty$ and $+\infty$.

Example - Indicator function

Let $S \subset \mathbb{R}^n$ be a set, the *indicator function* of S is an extended real-valued function given by

$$\iota_S(\boldsymbol{x}) = \begin{cases} 0 : & \boldsymbol{x} \in S, \\ +\infty : & \boldsymbol{x} \notin S. \end{cases}$$

Closed function

Definition - Extended real line function

An *extended real-valued function* is a function defined over the entire underlying space that can take any real value, as well as the infinite values $-\infty$ and $+\infty$.

Example - Indicator function

Let $S \subset \mathbb{R}^n$ be a set, the *indicator function* of S is an extended real-valued function given by

$$\iota_S(\boldsymbol{x}) = \begin{cases} 0 : & \boldsymbol{x} \in S, \\ +\infty : & \boldsymbol{x} \notin S. \end{cases}$$

Definition - Closed function

A function $F : \mathbb{R}^n \to [-\infty, +\infty]$ is *closed* if

- *epi-graph* is closed.
- *sub-level set* is closed.

Convex functions

Definition - Convex function

Let $S \subset \mathbb{R}^n$ be a non-empty convex set, a function $F : S \to \mathbb{R}$ is said to be **convex** if for any $\boldsymbol{x}, \boldsymbol{y} \in S$ and any $\lambda \in (0, 1)$, there holds

$$F\big(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}\big) \le \lambda F(\boldsymbol{x}) + (1-\lambda)F(\boldsymbol{y}).$$

- If $-F$ is convex, then F is said to be **concave**.

Convex functions

Definition - Convex function

Let $S \subset \mathbb{R}^n$ be a non-empty convex set, a function $F : S \to \mathbb{R}$ is said to be **convex** if for any $\boldsymbol{x}, \boldsymbol{y} \in S$ and any $\lambda \in (0,1)$, there holds

$$F\big(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}\big) \leq \lambda F(\mathbf{x}) + (1-\lambda)F(\mathbf{y}).$$

- If $-F$ is convex, then F is said to be **concave**.

Example - Examples on $\mathbb{R}$

- Absolute value function $F(x) = |x|$ is closed and convex.
- The function $F(x) = -\log(x)$ is closed and convex.

Convex functions

Definition - Convex function

Let $S \subset \mathbb{R}^n$ be a non-empty convex set, a function $F : S \to \mathbb{R}$ is said to be **convex** if for any $\boldsymbol{x}, \boldsymbol{y} \in S$ and any $\lambda \in (0, 1)$, there holds

$$F\big(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}\big) \leq \lambda F(\boldsymbol{x}) + (1-\lambda)F(\boldsymbol{y}).$$

- If $-F$ is convex, then F is said to be **concave**.

Definition - Strong convexity

Function $F : \mathbb{R}^n \to \mathbb{R}$ is strongly convex if $\mathrm{dom}(F)$ is convex, there exists $\alpha > 0$ such that

$$F(\boldsymbol{x}) - \frac{\alpha}{2}\|\boldsymbol{x}\|^2$$

is convex.

Convex functions

Let $F : S \to \mathbb{R}$ be a convex function and $\beta > 0$, then βF is convex.

Convex functions

Let $F : S \to \mathbb{R}$ be a convex function and $\beta > 0$, then βF is convex.

Let $F_1, F_2 : S \to \mathbb{R}$ be convex functions, then $F_1 + F_2$ is convex.

- Sum of finitely many convex functions $\sum_{i=1}^{k} F_i$...

Convex functions

Let $F : S \to \mathbb{R}$ be a convex function and $\beta > 0$, then βF is convex.

Let $F_1, F_2 : S \to \mathbb{R}$ be convex functions, then $F_1 + F_2$ is convex.

- Sum of finitely many convex functions $\sum_{i=1}^{k} F_i$...

Let $F : S \to \mathbb{R}$ be a convex function and $(\alpha_i)_{i=1}^{r} \in]0, 1[$ such that $\sum_i \alpha_i = 1$, then $F(\sum_i \alpha_i \mathbf{x}_i) \leq \sum_i \alpha_i F(\mathbf{x}_i)$.

Let $F : S \to \mathbb{R}$ be a convex function and $\beta > 0$, then βF is convex.

Let $F_1, F_2 : S \to \mathbb{R}$ be convex functions, then $F_1 + F_2$ is convex.

- Sum of finitely many convex functions $\sum_{i=1}^{k} F_i$...

Let $F : S \to \mathbb{R}$ be a convex function and $(\alpha_i)_{i=1}^{r} \in]0,1[$ such that $\sum_i \alpha_i = 1$, then $F(\sum_i \alpha_i \mathbf{x}_i) \leq \sum_i \alpha_i F(\mathbf{x}_i)$.

Let $S \subset \mathbb{R}$ be a non-empty convex set and $F : S \to \mathbb{R}$ a convex function, then F is continuous along the interior of S.

Convex functions

Let $F : S \to \mathbb{R}$ be a convex function and $\beta > 0$, then βF is convex.

Let $F_1, F_2 : S \to \mathbb{R}$ be convex functions, then $F_1 + F_2$ is convex.

- Sum of finitely many convex functions $\sum_{i=1}^{k} F_i$...

Let $F : S \to \mathbb{R}$ be a convex function and $(\alpha_i)_{i=1}^{r} \in]0, 1[$ such that $\sum_i \alpha_i = 1$, then $F(\sum_i \alpha_i \mathbf{x}_i) \leq \sum_i \alpha_i F(\mathbf{x}_i)$.

Let $S \subset \mathbb{R}$ be a non-empty convex set and $F : S \to \mathbb{R}$ a convex function, then F is continuous along the interior of S.

Let $F : S \to \mathbb{R}$ be a convex function and any $\alpha \in \mathbb{R}$, then the sub-level set is convex.

Convex functions

Let $F : S \to \mathbb{R}$ be a convex function and $\beta > 0$, then βF is convex.

Let $F_1, F_2 : S \to \mathbb{R}$ be convex functions, then $F_1 + F_2$ is convex.

- Sum of finitely many convex functions $\sum_{i=1}^{k} F_i$...

Let $F : S \to \mathbb{R}$ be a convex function and $(\alpha_i)_{i=1}^{r} \in]0,1[$ such that $\sum_i \alpha_i = 1$, then $F(\sum_i \alpha_i \mathbf{x}_i) \leq \sum_i \alpha_i F(\mathbf{x}_i)$.

Let $S \subset \mathbb{R}$ be a non-empty convex set and $F : S \to \mathbb{R}$ a convex function, then F is continuous along the interior of S.

Let $F : S \to \mathbb{R}$ be a convex function and any $\alpha \in \mathbb{R}$, then the sub-level set is convex.

Definition - $\Gamma_0(\mathbb{R}^n)$

The set of all *proper*, closed and convex functions on $\mathbb{R}^n$ is denoted as $\Gamma_0(\mathbb{R}^n)$.

Differentiability

Gradient, sub-differential

Definition - Directional derivative

Let S be a nonempty subset of $\mathbb{R}^n$, $F : \mathbb{R}^n \to \mathbb{R}$, and $\boldsymbol{x} \in \mathrm{dom}(F)$. The *directional derivative* of F at $\boldsymbol{x}$ in the direction $\boldsymbol{y}$ is

$$\nabla_{\boldsymbol{y}} F(\boldsymbol{x}) = \lim_{\alpha \downarrow 0} \frac{F(\boldsymbol{x} + \alpha \boldsymbol{y}) - F(\boldsymbol{x})}{\alpha},$$

provided that the limits exists.

Gradient

Definition - Directional derivative

Let S be a nonempty subset of $\mathbb{R}^n$, $F : \mathbb{R}^n \to \mathbb{R}$, and $\boldsymbol{x} \in \mathrm{dom}(F)$. The *directional derivative* of F at $\boldsymbol{x}$ in the direction $\boldsymbol{y}$ is

$$\nabla_{\boldsymbol{y}} F(\boldsymbol{x}) = \lim_{\alpha \downarrow 0} \frac{F(\boldsymbol{x} + \alpha \boldsymbol{y}) - F(\boldsymbol{x})}{\alpha},$$

provided that the limits exists.

Definition - Gradient

Let S be a subset of $\mathbb{R}^n$, $F : S \to \mathbb{R}$, and suppose that F is differentiable at $\boldsymbol{x} \in S$. Then, there exists a unique vector $\nabla F(\boldsymbol{x}) \in \mathbb{R}^n$ such that such

$$(\forall \boldsymbol{y} \in \mathbb{R}^n \ \text{ with } \ \|\boldsymbol{y}\| = 1) \quad \nabla_{\boldsymbol{y}} F(\boldsymbol{x}) = \langle \boldsymbol{y} \mid \nabla F(\boldsymbol{x}) \rangle.$$

The (Gâteaux) gradient of F at $\boldsymbol{x} \in S \subset \mathrm{dom}(F)$ is an n-dimensional vector

$$\nabla F(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial F(\boldsymbol{x})}{\partial x_1} \\ \frac{\partial F(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \frac{\partial F(\boldsymbol{x})}{\partial x_n} \end{bmatrix} \in \mathbb{R}^n,$$

where the *partial derivative* is defined by

$$\frac{\partial F(\boldsymbol{x})}{\partial x_i} = \lim_{\alpha \downarrow 0} \frac{F(\boldsymbol{x} + \alpha \boldsymbol{e}_i) - F(\boldsymbol{x})}{\alpha}.$$

Gradient

The (Gâteaux) gradient of F at $\mathbf{x} \in S \subset \mathrm{dom}(F)$ is an n-dimensional vector

$$\nabla F(\mathbf{x}) = \begin{bmatrix} \frac{\partial F(\mathbf{x})}{\partial x_1} \\ \frac{\partial F(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial F(\mathbf{x})}{\partial x_n} \end{bmatrix} \in \mathbb{R}^n,$$

where the *partial derivative* is defined by

$$\frac{\partial F(\mathbf{x})}{\partial x_i} = \lim_{\alpha \downarrow 0} \frac{F(\mathbf{x} + \alpha \mathbf{e}_i) - F(\mathbf{x})}{\alpha}.$$

Proposition - Characterization of convexity

Let $S \subset \mathbb{R}^n$ be an open set and $F : S \to \mathbb{R}$ be convex and smooth differentiable, then

- $F(\mathbf{y}) \geq F(\mathbf{x}) + \langle \nabla F(\mathbf{x}) \mid \mathbf{y} - \mathbf{x} \rangle$.
- $\langle \mathbf{y} - \mathbf{x} \mid \nabla F(\mathbf{y}) - \nabla F(\mathbf{x}) \rangle \geq 0$.

Subdifferentiability

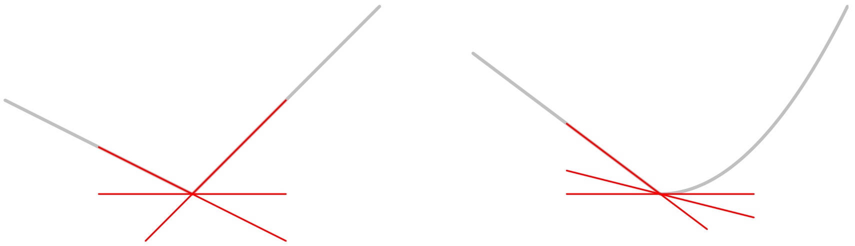
Definition - Subdifferential

Let $R : \mathbb{R}^n \to]-\infty, +\infty]$ be proper convex. The *subdifferential* of R at is the set-valued operator

$$\partial R : \mathbb{R}^n \rightrightarrows \mathbb{R}^n : \boldsymbol{x} \to \left\{ \boldsymbol{v} \in \mathbb{R}^n \ : \ (\forall \boldsymbol{y} \in \mathbb{R}^n) \ \langle \boldsymbol{y} - \boldsymbol{x} \mid \boldsymbol{v} \rangle + R(\boldsymbol{x}) \leq R(\boldsymbol{y}) \right\}.$$

Let $\boldsymbol{x} \in \mathbb{R}^n$, then R is *subdifferentiable* at $\boldsymbol{x}$ if $\partial R(\boldsymbol{x}) \neq \emptyset$.

- The elements of $\partial R(\boldsymbol{x})$ are the *subgradients* of R at $\boldsymbol{x}$.



Definition - Subdifferential

Let $R : \mathbb{R}^n \to]-\infty, +\infty]$ be proper convex. The *subdifferential* of R at is the set-valued operator

$$\partial R : \mathbb{R}^n \rightrightarrows \mathbb{R}^n : \mathbf{x} \to \left\{ \mathbf{v} \in \mathbb{R}^n \ : \ (\forall \mathbf{y} \in \mathbb{R}^n) \ \langle \mathbf{y} - \mathbf{x} \mid \mathbf{v} \rangle + R(\mathbf{x}) \leq R(\mathbf{y}) \right\}.$$

Let $\mathbf{x} \in \mathbb{R}^n$, then R is *subdifferentiable* at $\mathbf{x}$ if $\partial R(\mathbf{x}) \neq \emptyset$.

- The elements of $\partial R(\mathbf{x})$ are the *subgradients* of R at $\mathbf{x}$.

Example - Indicator function

Let S be a non-empty convex subset of $\mathbb{R}^n$. Then

$$\partial \iota_S(\mathbf{x}) = \mathcal{N}_S(\mathbf{x}) = \begin{cases} \{\mathbf{v} \in \mathbb{R}^n \ : \ \sup\langle \mathbf{v} \mid S - \mathbf{x} \rangle \leq 0\} : & \mathbf{x} \in S, \\ \emptyset : & \text{o.w.} \end{cases}$$

Definition - Subdifferential

Let $R : \mathbb{R}^n \to]-\infty, +\infty]$ be proper convex. The *subdifferential* of R at is the set-valued operator

$$\partial R : \mathbb{R}^n \rightrightarrows \mathbb{R}^n : \boldsymbol{x} \to \left\{ \boldsymbol{v} \in \mathbb{R}^n \ : \ (\forall \boldsymbol{y} \in \mathbb{R}^n) \ \langle \boldsymbol{y} - \boldsymbol{x} \mid \boldsymbol{v} \rangle + R(\boldsymbol{x}) \leq R(\boldsymbol{y}) \right\}.$$

Let $\boldsymbol{x} \in \mathbb{R}^n$, then R is *subdifferentiable* at $\boldsymbol{x}$ if $\partial R(\boldsymbol{x}) \neq \emptyset$.

- The elements of $\partial R(\boldsymbol{x})$ are the *subgradients* of R at $\boldsymbol{x}$.

Proposition - Convexity of subdifferential

Let $R : \mathbb{R}^n \to]-\infty, +\infty]$ be proper convex and $\boldsymbol{x} \in \mathrm{dom}(R)$. Then

- $\mathrm{dom}(\partial R) \subset \mathrm{dom}(R)$.
- $\partial R(\boldsymbol{x})$ is closed and convex.

Convex minimization problem

Minimizers, and characterizations

Definition - Infimum and minimum

Let $\Phi : S \to [-\infty, +\infty]$ and let C be a subset of S.

- The infimum of Φ over C is $\inf \Phi(C)$; it is also denoted by

$$\inf_{\boldsymbol{x} \in C} \Phi(\boldsymbol{x}).$$

- Φ achieves its infimum over C if there exists $\boldsymbol{y} \in C$ such that

$$\Phi(\boldsymbol{y}) = \inf \Phi(C).$$

In this case, we write

$$\Phi(\boldsymbol{y}) = \min \Phi(C) \quad \text{or} \quad \Phi(\boldsymbol{y}) = \min_{\boldsymbol{x} \in C} \Phi(\boldsymbol{x})$$

and call $\min \Phi(C)$ the minimum of Φ over C.

Global and local minimum

Definition - Global and local minimum

Let $\Phi : \mathbb{R}^n \to]-\infty, +\infty]$ be proper, and let $\boldsymbol{x} \in \mathbb{R}^n$

- $\boldsymbol{x}$ is a (global) minimizer of Φ if

$$\Phi(\boldsymbol{x}) = \inf \ \Phi(\mathbb{R}^n)$$

  and $\Phi(\boldsymbol{x}) = \min \ \Phi(\mathbb{R}^n) \in \mathbb{R}$.
- The set of minimizers of Φ is denoted by $\mathrm{Argmin}(\Phi)$.
- If $\mathrm{Argmin}(\Phi)$ is a singleton, its unique element is denoted by $\underset{\boldsymbol{x} \in \mathbb{R}^n}{\mathrm{argmin}} \, \Phi(\boldsymbol{x})$.

Let S be a subset of $\mathbb{R}^n$ such that $S \cap \mathrm{dom}(\Phi) \neq \emptyset$

- A minimizer of Φ over S is a minimizer of $\Phi + \iota_S$.
- If $\exists \rho > 0$ such that $\boldsymbol{x}$ is a minimizer of Φ over $\mathcal{B}(\boldsymbol{x}; \rho)$, then $\boldsymbol{x}$ is a local minimizer of Φ.

Global and local minimum

Definition - Global and local minimum

Let $\Phi : \mathbb{R}^n \to]-\infty, +\infty]$ be proper, and let $\boldsymbol{x} \in \mathbb{R}^n$

- $\boldsymbol{x}$ is a (global) minimizer of Φ if
$$\Phi(\boldsymbol{x}) = \inf \ \Phi(\mathbb{R}^n)$$
and $\Phi(\boldsymbol{x}) = \min \ \Phi(\mathbb{R}^n) \in \mathbb{R}$.
- The set of minimizers of Φ is denoted by $\mathrm{Argmin}(\Phi)$.
- If $\mathrm{Argmin}(\Phi)$ is a singleton, its unique element is denoted by $\underset{\boldsymbol{x} \in \mathbb{R}^n}{\mathrm{argmin}} \, \Phi(\boldsymbol{x})$.

Let S be a subset of $\mathbb{R}^n$ such that $S \cap \mathrm{dom}(\Phi) \neq \emptyset$

- A minimizer of Φ over S is a minimizer of $\Phi + \iota_S$.
- If $\exists \rho > 0$ such that $\boldsymbol{x}$ is a minimizer of Φ over $\mathcal{B}(\boldsymbol{x}; \rho)$, then $\boldsymbol{x}$ is a local minimizer of Φ.

Theorem - Convexity and local minimizer

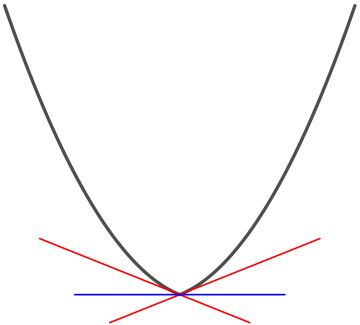
Let $\Phi : \mathbb{R}^n \to]-\infty, +\infty]$ be proper convex. Then every local minimizer of Φ is a minimizer.

Fermat's rule

Theorem - Fermat's rule

Let $\Phi : \mathbb{R}^n \to]-\infty, +\infty]$ be proper. Then

$$\operatorname{Argmin}(\Phi) = \operatorname{zer}(\partial\Phi) \stackrel{\text{def}}{=} \left\{ \boldsymbol{x} \in \mathbb{R}^n \ : \ \mathbf{0} \in \partial\Phi(\boldsymbol{x}) \right\}.$$



Characterization of minimizers

Problem - A non-smooth problem

Let $F \in \Gamma_0(\mathbb{R}^n)$, $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ be non-zero bounded linear and $R \in \Gamma_0(\mathbb{R}^m)$

$$\min_{\boldsymbol{x}} \left\{ \Phi(\boldsymbol{x}) \stackrel{\text{def}}{=} F(\boldsymbol{x}) + R(\boldsymbol{K}\boldsymbol{x}) \right\},$$

Problem - A non-smooth problem

Let $F \in \Gamma_0(\mathbb{R}^n)$, $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ be non-zero bounded linear and $R \in \Gamma_0(\mathbb{R}^m)$

$$\min_{\boldsymbol{x}} \left\{ \Phi(\boldsymbol{x}) \stackrel{\text{def}}{=} F(\boldsymbol{x}) + R(\boldsymbol{K}\boldsymbol{x}) \right\},$$

- In general,

$$\partial \Phi = \partial \big(F + R \circ \boldsymbol{K} \big) \neq \partial F + \boldsymbol{K}^* \circ \partial R \circ \boldsymbol{K}.$$

Problem - A non-smooth problem

Let $F \in \Gamma_0(\mathbb{R}^n)$, $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ be non-zero bounded linear and $R \in \Gamma_0(\mathbb{R}^m)$

$$\min_{\boldsymbol{x}} \Big\{ \Phi(\boldsymbol{x}) \stackrel{\text{def}}{=} F(\boldsymbol{x}) + R(\boldsymbol{K}\boldsymbol{x}) \Big\},$$

- In general,

$$\partial\Phi = \partial\big(F + R \circ \boldsymbol{K}\big) \neq \partial F + \boldsymbol{K}^* \circ \partial R \circ \boldsymbol{K}.$$

- Suppose $\mathrm{dom}(R) \cap \boldsymbol{K}\mathrm{dom}(F) \neq \emptyset$,

$$\partial F + \boldsymbol{K}^* \circ \partial R \circ \boldsymbol{K} \subset \partial\big(F + R \circ \boldsymbol{K}\big).$$

Proposition - Characterization of minimizers

Let $F \in \Gamma_0(\mathbb{R}^n)$, $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ be non-zero bounded and $R \in \Gamma_0(\mathbb{R}^m)$. Then the following holds

- $\mathrm{zer}\big(\partial F + \boldsymbol{K}^* \circ \partial R \circ \boldsymbol{K}\big) \subset \mathrm{Argmin}(F + R \circ \boldsymbol{K})$.
- Suppose $\mathrm{Argmin}(F + R \circ \boldsymbol{K}) \neq \emptyset$ and
$$\mathrm{ri}\big(\mathrm{dom}(R)\big) \cap \mathrm{ri}\big(\boldsymbol{K}\mathrm{dom}(F)\big) \neq \emptyset.$$
Then
$$\mathrm{Argmin}(F + R \circ \boldsymbol{K}) = \mathrm{zer}\big(\partial F + \boldsymbol{K}^* \circ \partial R \circ \boldsymbol{K}\big) \neq \emptyset.$$

Let $\boldsymbol{x}^\star \in \mathrm{Argmin}(F + R \circ \boldsymbol{K})$, the corresponding optimality condition reads
$$\boldsymbol{0} \in \partial F(\boldsymbol{x}^\star) + \boldsymbol{K}^* \partial R(\boldsymbol{K}\boldsymbol{x}^\star).$$

Definition - Set-valued operator

An operator $\mathcal{A} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is set-valued if for every $\boldsymbol{x} \in \mathbb{R}^n$, $\mathcal{A}(\boldsymbol{x})$ is a subset of $\mathbb{R}^n$. Its graph is defined by

$$\mathrm{gra}(\mathcal{A}) = \Big\{(\boldsymbol{x}, \boldsymbol{u}) \in \mathbb{R}^n \times \mathbb{R}^n \ : \ \boldsymbol{u} \in \mathcal{A}(\boldsymbol{x})\Big\}.$$

Definition - Set-valued operator

An operator $\mathcal{A} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is set-valued if for every $\boldsymbol{x} \in \mathbb{R}^n$, $\mathcal{A}(\boldsymbol{x})$ is a subset of $\mathbb{R}^n$. Its graph is defined by

$$\mathrm{gra}(\mathcal{A}) = \left\{(\boldsymbol{x}, \boldsymbol{u}) \in \mathbb{R}^n \times \mathbb{R}^n \ : \ \boldsymbol{u} \in \mathcal{A}(\boldsymbol{x})\right\}.$$

Definition - Monotone operator

Let $\mathcal{A} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$. Then A is monotone if

$$\big(\forall (\boldsymbol{x}, \boldsymbol{u}) \in \mathrm{gra}(\mathcal{A})\big)\big((\boldsymbol{y}, \boldsymbol{v}) \in \mathrm{gra}(\mathcal{A})\big) \quad \langle \boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{u} - \boldsymbol{v} \rangle \geq 0.$$

It is moreover *maximally monotone* if $\mathrm{gra}(\mathcal{A})$ cannot be contained properly by the graph of another monotone operator $\mathcal{B}$.

Monotone operators

Definition - Monotone operator

Let $\mathcal{A} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$. Then A is monotone if

$$\big(\forall (\boldsymbol{x}, \boldsymbol{u}) \in \operatorname{gra}(\mathcal{A})\big)\big((\boldsymbol{y}, \boldsymbol{v}) \in \operatorname{gra}(\mathcal{A})\big) \quad \langle \boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{u} - \boldsymbol{v} \rangle \geq 0.$$

It is moreover *maximally monotone* if $\operatorname{gra}(\mathcal{A})$ cannot be contained properly by the graph of another monotone operator $\mathcal{B}$.

Theorem - Moreau

Let $R \in \Gamma_0(\mathbb{R}^n)$. Then ∂R is maximally monotone.

Definition - Monotone operator

Let $\mathcal{A} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$. Then A is monotone if

$$\big(\forall(\boldsymbol{x}, \boldsymbol{u}) \in \operatorname{gra}(\mathcal{A})\big)\big((\boldsymbol{y}, \boldsymbol{v}) \in \operatorname{gra}(\mathcal{A})\big) \quad \langle \boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{u} - \boldsymbol{v} \rangle \geq 0.$$

It is moreover *maximally monotone* if $\operatorname{gra}(\mathcal{A})$ cannot be contained properly by the graph of another monotone operator $\mathcal{B}$.

Theorem - Moreau

Let $R \in \Gamma_0(\mathbb{R}^n)$. Then ∂R is maximally monotone.

Proposition -

Let $\mathcal{A} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be maximally monotone and $\boldsymbol{x} \in \mathbb{R}^n$. Then $\mathcal{A}(\boldsymbol{x})$ is *closed and convex*.

References

- Amir Beck: First-order methods in optimization, Vol. 25. SIAM, 2017.
- Heinz H. Bauschke and Patrick L. Combettes: Convex analysis and monotone operator theory in Hilbert spaces, Vol. 408. New York: Springer, 2011.