Optimization methods
Lecture 1: Introduction. Convex sets. Convex functions

Alexandr Katrutsa

Modern State of Artificial Intelligence Masters Program
Moscow Institute of Physics and Technology

What is this course about?

Basic theory

- Convex sets and convex functions
- Optimality conditions
- Introduction to duality

Numerical methods

- First order methods and their accelerated versions
- Quasi-Newton methods
- Introduction to stochastic gradient methods

The place of this course in the program

- When you train some neural network, you solve some optimization problem
- Possible issues in this process will be discussed in the course
- How to solve these issues we will also discuss

Schedule and deadlines

- Lectures and webinars are once a week

Schedule and deadlines

- Lectures and webinars are once a week
- Home assignments are after every lecture

Schedule and deadlines

- Lectures and webinars are once a week
- Home assignments are after every lecture
- Grading policy will be announced during webinar

Schedule and deadlines

- Lectures and webinars are once a week
- Home assignments are after every lecture
- Grading policy will be announced during webinar
- Lecture slides are here:
  https://github.com/girafe-ai/msai-optimization

References

- S. Boyd and L. Vandenberghe *Convex Optimization*
  https://web.stanford.edu/~boyd/cvxbook/
- J. Nocedal, S. J. Wright *Numerical Optimization*
- I. Goodfellow et al *Deep learning book*

General methodology

Main steps for exploiting optimization methods in solving real-world problems:

1. Define objective function

General methodology

Main steps for exploiting optimization methods in solving real-world problems:

1. Define objective function
2. Define feasible set

General methodology

Main steps for exploiting optimization methods in solving real-world problems:

1. Define objective function
2. Define feasible set
3. Optimization problem statement and its analysis

General methodology

Main steps for exploiting optimization methods in solving real-world problems:

1. Define objective function
2. Define feasible set
3. Optimization problem statement and its analysis
4. Selection of the best algorithm for the stated problem

General methodology

Main steps for exploiting optimization methods in solving real-world problems:

1. Define objective function
2. Define feasible set
3. Optimization problem statement and its analysis
4. Selection of the best algorithm for the stated problem
5. Algorithm implementation and verification of its correctness

General problem statement

$$
\begin{aligned}
&\min_{\mathbf{x}\in\mathcal{X}} f_0(\mathbf{x}) \\
\text{s.t. } & f_i(\mathbf{x}) = 0,\ i = 1,\dots,p \\
& f_j(\mathbf{x}) \le 0,\ j = p+1,\dots,m,
\end{aligned}
$$

General problem statement

$$\begin{aligned}
&\min_{\mathbf{x} \in \mathcal{X}} f_0(\mathbf{x}) \\
\text{s.t. } & f_i(\mathbf{x}) = 0, \ i = 1, \dots, p \\
& f_j(\mathbf{x}) \leq 0, \ j = p+1, \dots, m,
\end{aligned}$$

- $\mathbf{x} \in \mathbb{R}^n$ — target vector

General problem statement

$$
\begin{aligned}
&\min_{\mathbf{x} \in \mathcal{X}} f_0(\mathbf{x}) \\
\text{s.t. } & f_i(\mathbf{x}) = 0,\ i = 1, \dots, p \\
& f_j(\mathbf{x}) \leq 0,\ j = p+1, \dots, m,
\end{aligned}
$$

- $\mathbf{x} \in \mathbb{R}^n$ — target vector
- $f_0(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ — objective function

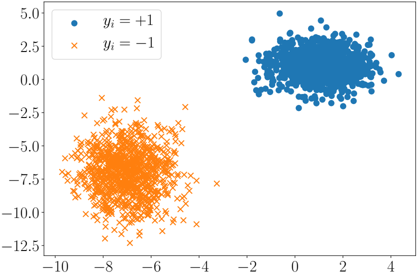
General problem statement

$$
\begin{aligned}
& \min_{\mathbf{x} \in \mathcal{X}} f_0(\mathbf{x}) \\
\text{s.t. } & f_i(\mathbf{x}) = 0, \; i = 1, \dots, p \\
& f_j(\mathbf{x}) \leq 0, \; j = p+1, \dots, m,
\end{aligned}
$$

- $\mathbf{x} \in \mathbb{R}^n$ — target vector
- $f_0(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ — objective function
- $f_k(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ — constraint functions

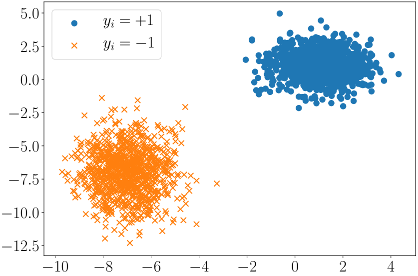
Linear classification problem

- Given dataset: $(\mathbf{x}_i, y_i)$, $\mathbf{x}_i \in \mathbb{R}^n,\ y_i = \{+1, -1\},\ i = 1, \dots, m$



Linear classification problem

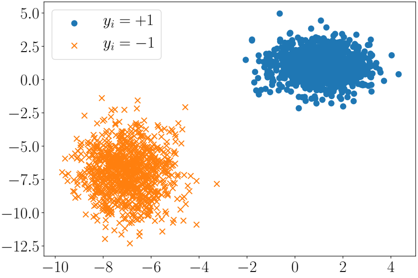
- Given dataset: $(\mathbf{x}_i, y_i)$, $\mathbf{x}_i \in \mathbb{R}^n,\ y_i = \{+1, -1\},\ i = 1, \dots, m$



- Linear classifier $\hat{y} = \mathrm{sign}(\mathbf{w}^\top \mathbf{x} + b)$

Linear classification problem

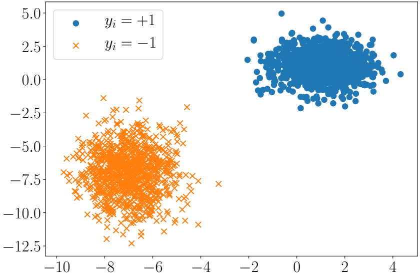
- Given dataset: $(\mathbf{x}_i, y_i)$, $\mathbf{x}_i \in \mathbb{R}^n,\ y_i = \{+1, -1\},\ i = 1, \dots, m$



- Linear classifier $\hat{y} = \operatorname{sign}(\mathbf{w}^\top \mathbf{x} + b)$
- $\begin{cases} \mathbf{w}^\top \mathbf{x}_i + b > 1, & y_i = +1 \\ \mathbf{w}^\top \mathbf{x}_i + b < -1, & y_i = -1 \end{cases}$

Linear classification problem

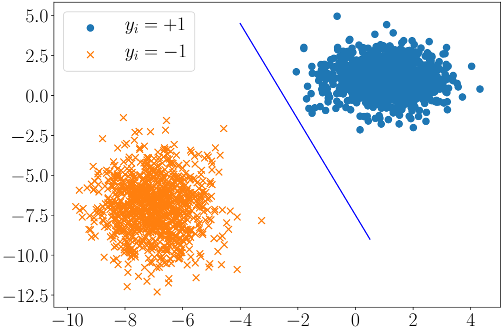
- Given dataset: $(\mathbf{x}_i, y_i)$, $\mathbf{x}_i \in \mathbb{R}^n,\ y_i = \{+1, -1\},\ i = 1, \dots, m$



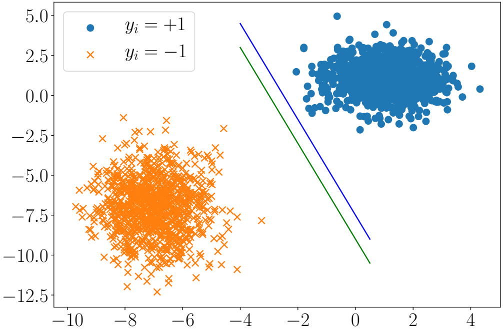
- Linear classifier $\hat{y} = \mathrm{sign}(\mathbf{w}^\top \mathbf{x} + b)$
- $\begin{cases} \mathbf{w}^\top \mathbf{x}_i + b > 1, & y_i = +1 \\ \mathbf{w}^\top \mathbf{x}_i + b < -1, & y_i = -1 \end{cases}$
- $y_i(\mathbf{w}^\top \mathbf{x}_i + b) > 1$

Possible separating hypelplanes

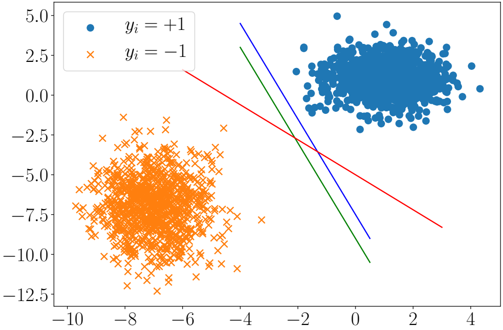
Possible separating hypelplanes



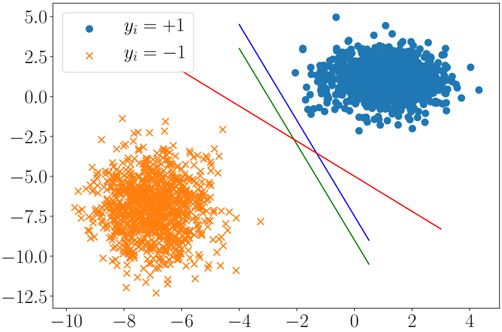
Possible separating hypelplanes



Possible separating hypelplanes



Possible separating hypelplanes



Q: How to define the separating hyperplane uniquely?

Margin maximization

Margin maximization

- For the support samples of every class the following holds

$$\begin{cases} \mathbf{w}^\top \mathbf{x}_k + b = 1, & y_k = +1 \\ \mathbf{w}^\top \mathbf{x}_j + b = -1, & y_j = -1 \end{cases}$$

Margin maximization

- For the support samples of every class the following holds

$$\begin{cases} \mathbf{w}^\top \mathbf{x}_k + b = 1, & y_k = +1 \\ \mathbf{w}^\top \mathbf{x}_j + b = -1, & y_j = -1 \end{cases}$$

- Distance between parallel hyperplanes $\mathbf{w}^\top \mathbf{x} + b = c_1$ and $\mathbf{w}^\top \mathbf{x} + b = c_2$:

$$d = \frac{|c_1 - c_2|}{\|\mathbf{w}\|_2} = \frac{2}{\|\mathbf{w}\|_2}$$

Margin maximization

- For the support samples of every class the following holds

$$\begin{cases} \mathbf{w}^\top \mathbf{x}_k + b = 1, & y_k = +1 \\ \mathbf{w}^\top \mathbf{x}_j + b = -1, & y_j = -1 \end{cases}$$

- Distance between parallel hyperplanes $\mathbf{w}^\top \mathbf{x} + b = c_1$ and $\mathbf{w}^\top \mathbf{x} + b = c_2$:

$$d = \frac{|c_1 - c_2|}{\|\mathbf{w}\|_2} = \frac{2}{\|\mathbf{w}\|_2}$$

- We want to maximize this distance or margin between two classes

Margin maximization

- For the support samples of every class the following holds

$$\begin{cases} \mathbf{w}^\top \mathbf{x}_k + b = 1, & y_k = +1 \\ \mathbf{w}^\top \mathbf{x}_j + b = -1, & y_j = -1 \end{cases}$$

- Distance between parallel hyperplanes $\mathbf{w}^\top \mathbf{x} + b = c_1$ and $\mathbf{w}^\top \mathbf{x} + b = c_2$:

$$d = \frac{|c_1 - c_2|}{\|\mathbf{w}\|_2} = \frac{2}{\|\mathbf{w}\|_2}$$

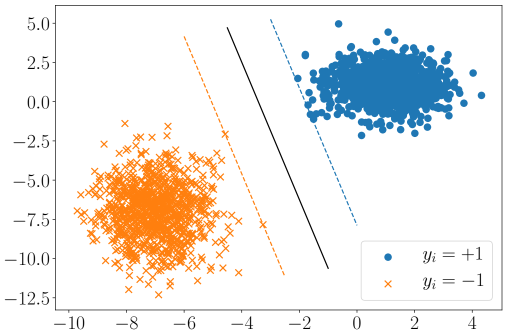
- We want to maximize this distance or margin between two classes

The final optimization problem

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|_2^2$$

$$\text{s.t. } y_i(\mathbf{w}^\top \mathbf{x}_i + b) > 1, \; i = 1, \ldots, m$$

Optimal separating hyperplane



What solutions are possible?

What solutions are possible?

Definition

A point $\mathbf{x}^*$ is called a point of ***global*** *minimum, if $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x}$ from the feasible set.*

What solutions are possible?

Definition

A point $\mathbf{x}^*$ *is called a point of* ***global*** *minimum, if* $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ *for all* $\mathbf{x}$ *from the feasible set.*

Definition

A point $\mathbf{x}^*$ *is called a point of* ***local*** *minimum, if* $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ *for all* $\mathbf{x}$ *from some neighborhood of point* $\mathbf{x}^*$ *and from feasible set.*

What solutions are possible?

Definition

*A point $\mathbf{x}^*$ is called a point of **global** minimum, if $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x}$ from the feasible set.*

Definition

*A point $\mathbf{x}^*$ is called a point of **local** minimum, if $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x}$ from some neighborhood of point $\mathbf{x}^*$ and from feasible set.*

Another form of problem statement

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathcal{X}} f_0(\mathbf{x})$$

$$\text{s.t. } f_i(\mathbf{x}) = 0,\ i = 1, \ldots, p$$

$$f_j(\mathbf{x}) \leq 0,\ j = p+1, \ldots, m,$$

How to solve such problems?

In general case:

- Very hard to solve
- randomized algorithms give a trade-off between running time and robustness of approximate solution

How to solve such problems?

In general case:

- Very hard to solve
- randomized algorithms give a trade-off between running time and robustness of approximate solution

However, some classes of optimization problems can be solved very efficiently

How to solve such problems?

In general case:

- Very hard to solve
- randomized algorithms give a trade-off between running time and robustness of approximate solution

However, some classes of optimization problems can be solved very efficiently

- Linear programming
- Linear least-squares problems
- Low-rank approximation problem
- Convex optimization

Convex optimization problem

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n}\ & f_0(\mathbf{x}) \\ \text{s.t.}\ & f_i(\mathbf{x}) \le 0,\ i = 1,\dots,m \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

Convex optimization problem

$$\begin{aligned} &\min_{\mathbf{x}\in\mathbb{R}^n} f_0(\mathbf{x}) \\ \text{s.t. } &f_i(\mathbf{x}) \leq 0,\ i = 1, \dots, m \\ &\mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

- f_0, f_i — convex functions:

$$f(\alpha \mathbf{x}_1 + \beta \mathbf{x}_2) \leq \alpha f(\mathbf{x}_1) + \beta f(\mathbf{x}_2),$$

where $\alpha, \beta \geq 0$ and $\alpha + \beta = 1$.

Convex optimization problem

$$\begin{aligned} &\min_{\mathbf{x}\in\mathbb{R}^n} f_0(\mathbf{x}) \\ \text{s.t. } &f_i(\mathbf{x}) \leq 0,\ i = 1, \dots, m \\ &\mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

- f_0, f_i — convex functions:

  $$f(\alpha\mathbf{x}_1 + \beta\mathbf{x}_2) \leq \alpha f(\mathbf{x}_1) + \beta f(\mathbf{x}_2),$$

  where $\alpha, \beta \geq 0$ and $\alpha + \beta = 1$.
- no analytical solution

Convex optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} f_0(\mathbf{x})$$
$$\text{s.t. } f_i(\mathbf{x}) \leq 0,\ i = 1, \dots, m$$
$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

- f_0, f_i — convex functions:

$$f(\alpha\mathbf{x}_1 + \beta\mathbf{x}_2) \leq \alpha f(\mathbf{x}_1) + \beta f(\mathbf{x}_2),$$

  where $\alpha, \beta \geq 0$ and $\alpha + \beta = 1$.
- no analytical solution
- efficient algorithms

Convex optimization problem

$$
\begin{aligned}
&\min_{\mathbf{x} \in \mathbb{R}^n} f_0(\mathbf{x}) \\
\text{s.t. } & f_i(\mathbf{x}) \le 0, \; i = 1, \ldots, m \\
& \mathbf{A}\mathbf{x} = \mathbf{b}
\end{aligned}
$$

- f_0, f_i — convex functions:

  $$f(\alpha \mathbf{x}_1 + \beta \mathbf{x}_2) \le \alpha f(\mathbf{x}_1) + \beta f(\mathbf{x}_2),$$

  where $\alpha, \beta \ge 0$ and $\alpha + \beta = 1$.
- no analytical solution
- efficient algorithms
- special modeling helps to convert such problems to some standard form

Why convexity is so important?

Ralph Tyrrell Rockafellar (born 1935)

The great watershed in optimization is not between linearity and non-linearity, but convexity and non-convexity.

Why convexity is so important?

Ralph Tyrrell Rockafellar (born 1935)

The great watershed in optimization is not between linearity and non-linearity, but convexity and non-convexity.

- Local minimum is also global minimum

Why convexity is so important?

Ralph Tyrrell Rockafellar (born 1935)

The great watershed in optimization is not between linearity and non-linearity, but convexity and non-convexity.

- Local minimum is also global minimum
- Necessary optimality condition is also sufficient

Convex sets

Definition

A set $\mathcal{X} \subseteq \mathbb{R}^n$ is convex if for all $\alpha \in [0, 1]$ and for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ the following holds

$$\alpha \mathbf{x} + (1 - \alpha)\mathbf{y} \in \mathcal{X}.$$

Convex sets

Definition

A set $\mathcal{X} \subseteq \mathbb{R}^n$ is convex if for all $\alpha \in [0,1]$ and for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ the following holds

$$\alpha \mathbf{x} + (1-\alpha)\mathbf{y} \in \mathcal{X}.$$

Examples

- Polyhedron

Convex sets

Definition

A set $\mathcal{X} \subseteq \mathbb{R}^n$ *is convex if for all* $\alpha \in [0,1]$ *and for any* $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ *the following holds*

$$\alpha \mathbf{x} + (1-\alpha)\mathbf{y} \in \mathcal{X}.$$

Examples

- Polyhedron
- Hyperplanes

Convex sets

Definition

A set $\mathcal{X} \subseteq \mathbb{R}^n$ *is convex if for all* $\alpha \in [0, 1]$ *and for any* $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ *the following holds*

$$\alpha \mathbf{x} + (1 - \alpha)\mathbf{y} \in \mathcal{X}.$$

Examples

- Polyhedron
- Hyperplanes
- Balls in *any proper* norm and ellipsoids

Convex sets

Definition

A set $\mathcal{X} \subseteq \mathbb{R}^n$ is convex if for all $\alpha \in [0,1]$ and for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ the following holds

$$\alpha \mathbf{x} + (1-\alpha)\mathbf{y} \in \mathcal{X}.$$

Examples

- Polyhedron
- Hyperplanes
- Balls in *any proper* norm and ellipsoids
- Set of symmetric and non-negative definite matrices

Intersection of convex sets

Theorem

Intersection of finite or infinite number of convex sets $\mathcal{X}_i$ is a convex set:

$$\mathcal{X} = \bigcap_{i \in \mathcal{I}} \mathcal{X}_i.$$

Intersection of convex sets

Theorem

Intersection of finite or infinite number of convex sets $\mathcal{X}_i$ is a convex set:

$$\mathcal{X} = \bigcap_{i \in \mathcal{I}} \mathcal{X}_i.$$

Proof

- Consider $\mathbf{x}, \mathbf{y} \in \mathcal{X} \to \mathbf{x}, \mathbf{y} \in \mathcal{X}_i, \forall i \in \mathcal{I}$

Intersection of convex sets

Theorem

Intersection of finite or infinite number of convex sets $\mathcal{X}_i$ is a convex set:

$$\mathcal{X} = \bigcap_{i \in \mathcal{I}} \mathcal{X}_i.$$

Proof

- Consider $\mathbf{x}, \mathbf{y} \in \mathcal{X} \to \mathbf{x}, \mathbf{y} \in \mathcal{X}_i, \forall i \in \mathcal{I}$
- Consider point $\mathbf{z} = \alpha\mathbf{x} + (1-\alpha)\mathbf{y}$, $\alpha \in [0, 1]$

Intersection of convex sets

Theorem

Intersection of finite or infinite number of convex sets $\mathcal{X}_i$ is a convex set:

$$\mathcal{X} = \bigcap_{i \in \mathcal{I}} \mathcal{X}_i.$$

Proof

- Consider $\mathbf{x}, \mathbf{y} \in \mathcal{X} \to \mathbf{x}, \mathbf{y} \in \mathcal{X}_i, \forall i \in \mathcal{I}$
- Consider point $\mathbf{z} = \alpha \mathbf{x} + (1 - \alpha)\mathbf{y}$, $\alpha \in [0, 1]$
- Since $\mathcal{X}_i$ is convex for all $i \in \mathcal{I}$, $\mathbf{z} \in \mathcal{X}_i, \ \forall i \in \mathcal{I}$

Intersection of convex sets

Theorem

Intersection of finite or infinite number of convex sets $\mathcal{X}_i$ is a convex set:

$$\mathcal{X} = \bigcap_{i \in \mathcal{I}} \mathcal{X}_i.$$

Proof

- Consider $\mathbf{x}, \mathbf{y} \in \mathcal{X} \to \mathbf{x}, \mathbf{y} \in \mathcal{X}_i, \forall i \in \mathcal{I}$
- Consider point $\mathbf{z} = \alpha \mathbf{x} + (1 - \alpha)\mathbf{y}$, $\alpha \in [0, 1]$
- Since $\mathcal{X}_i$ is convex for all $i \in \mathcal{I}$, $\mathbf{z} \in \mathcal{X}_i, \ \forall i \in \mathcal{I}$
- Therefore, $\mathbf{z} \in \mathcal{X}$ and $\mathcal{X}$ is convex set

Affine map of convex set

Theorem

If the domain of any affine map is convex, then the image of this map is also convex.

Proof

- Let $\mathcal{X}$ be a convex set and $\mathbf{x}, \mathbf{y} \in \mathcal{X}$

Affine map of convex set

Theorem

If the domain of any affine map is convex, then the image of this map is also convex.

Proof

- Let $\mathcal{X}$ be a convex set and $\mathbf{x}, \mathbf{y} \in \mathcal{X}$
- Let f be an affine map: $f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$

Affine map of convex set

Theorem

If the domain of any affine map is convex, then the image of this map is also convex.

Proof

- Let $\mathcal{X}$ be a convex set and $\mathbf{x}, \mathbf{y} \in \mathcal{X}$
- Let f be an affine map: $f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$
- Show that $\alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y}) \in f(\mathcal{X})$, where $\alpha \in [0, 1]$

Affine map of convex set

Theorem

If the domain of any affine map is convex, then the image of this map is also convex.

Proof

- Let $\mathcal{X}$ be a convex set and $\mathbf{x}, \mathbf{y} \in \mathcal{X}$
- Let f be an affine map: $f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$
- Show that $\alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y}) \in f(\mathcal{X})$, where $\alpha \in [0, 1]$
- Indeed,

$$\alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y}) = \alpha(\mathbf{A}\mathbf{x} + \mathbf{b}) + (1-\alpha)(\mathbf{A}\mathbf{y} + \mathbf{b}) =$$
$$\mathbf{A}(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) + \mathbf{b} = \mathbf{A}\mathbf{z} + \mathbf{b} = f(\mathbf{z}),$$

where $\mathbf{z} = \alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in \mathcal{X}$

Arithmetic operations under convex sets

Theorem

Minkowski sum of two convex sets is convex set.

Arithmetic operations under convex sets

Theorem

Minkowski sum of two convex sets is convex set.

Proof

- Let $\mathcal{X}_1, \mathcal{X}_2$ be convex sets. Consider $\mathcal{X} = \mathcal{X}_1 + \mathcal{X}_2 = \{\mathbf{x}_1 + \mathbf{x}_2 \mid \mathbf{x}_1 \in \mathcal{X}_1,\ \mathbf{x}_2 \in \mathcal{X}_2\}$

Arithmetic operations under convex sets

Theorem

Minkowski sum of two convex sets is convex set.

Proof

- Let $\mathcal{X}_1, \mathcal{X}_2$ be convex sets. Consider $\mathcal{X} = \mathcal{X}_1 + \mathcal{X}_2 = \{\mathbf{x}_1 + \mathbf{x}_2 \mid \mathbf{x}_1 \in \mathcal{X}_1,\ \mathbf{x}_2 \in \mathcal{X}_2\}$
- Let $\hat{\mathbf{x}} = \hat{\mathbf{x}}_1 + \hat{\mathbf{x}}_2$ and $\tilde{\mathbf{x}} = \tilde{\mathbf{x}}_1 + \tilde{\mathbf{x}}_2$ belong to $\mathcal{X}$. Show that $\alpha\hat{\mathbf{x}} + (1-\alpha)\tilde{\mathbf{x}} \in \mathcal{X}$

Arithmetic operations under convex sets

Theorem

Minkowski sum of two convex sets is convex set.

Proof

- Let $\mathcal{X}_1, \mathcal{X}_2$ be convex sets. Consider $\mathcal{X} = \mathcal{X}_1 + \mathcal{X}_2 = \{\mathbf{x}_1 + \mathbf{x}_2 \mid \mathbf{x}_1 \in \mathcal{X}_1,\ \mathbf{x}_2 \in \mathcal{X}_2\}$
- Let $\hat{\mathbf{x}} = \hat{\mathbf{x}}_1 + \hat{\mathbf{x}}_2$ and $\tilde{\mathbf{x}} = \tilde{\mathbf{x}}_1 + \tilde{\mathbf{x}}_2$ belong to $\mathcal{X}$. Show that $\alpha\hat{\mathbf{x}} + (1-\alpha)\tilde{\mathbf{x}} \in \mathcal{X}$
- Indeed,

  $\alpha\hat{\mathbf{x}}+(1-\alpha)\tilde{\mathbf{x}} = [\alpha\hat{\mathbf{x}}_1+(1-\alpha)\tilde{\mathbf{x}}_1]+[\alpha\hat{\mathbf{x}}_2+(1-\alpha)\tilde{\mathbf{x}}_2] = \mathbf{y}_1+\mathbf{y}_2,$

  where $\mathbf{y}_1 \in C_1$ and $\mathbf{y}_2 \in C_2$ since sets C_1, C_2 are convex.

Arithmetic operations under convex sets

Theorem

Minkowski sum of two convex sets is convex set.

Proof

- Let $\mathcal{X}_1, \mathcal{X}_2$ be convex sets. Consider $\mathcal{X} = \mathcal{X}_1 + \mathcal{X}_2 = \{\mathbf{x}_1 + \mathbf{x}_2 \mid \mathbf{x}_1 \in \mathcal{X}_1,\ \mathbf{x}_2 \in \mathcal{X}_2\}$
- Let $\hat{\mathbf{x}} = \hat{\mathbf{x}}_1 + \hat{\mathbf{x}}_2$ and $\tilde{\mathbf{x}} = \tilde{\mathbf{x}}_1 + \tilde{\mathbf{x}}_2$ belong to $\mathcal{X}$. Show that $\alpha\hat{\mathbf{x}} + (1-\alpha)\tilde{\mathbf{x}} \in \mathcal{X}$
- Indeed,

  $$\alpha\hat{\mathbf{x}}+(1-\alpha)\tilde{\mathbf{x}} = [\alpha\hat{\mathbf{x}}_1+(1-\alpha)\tilde{\mathbf{x}}_1]+[\alpha\hat{\mathbf{x}}_2+(1-\alpha)\tilde{\mathbf{x}}_2] = \mathbf{y}_1+\mathbf{y}_2,$$

  where $\mathbf{y}_1 \in C_1$ and $\mathbf{y}_2 \in C_2$ since sets C_1, C_2 are convex.

Corollary

Linear combination of convex sets is convex set

Arithmetic operations under convex sets

Theorem

Minkowski sum of two convex sets is convex set.

Proof

- Let $\mathcal{X}_1, \mathcal{X}_2$ be convex sets. Consider $\mathcal{X} = \mathcal{X}_1 + \mathcal{X}_2 = \{\mathbf{x}_1 + \mathbf{x}_2 \mid \mathbf{x}_1 \in \mathcal{X}_1,\ \mathbf{x}_2 \in \mathcal{X}_2\}$
- Let $\hat{\mathbf{x}} = \hat{\mathbf{x}}_1 + \hat{\mathbf{x}}_2$ and $\tilde{\mathbf{x}} = \tilde{\mathbf{x}}_1 + \tilde{\mathbf{x}}_2$ belong to $\mathcal{X}$. Show that $\alpha\hat{\mathbf{x}} + (1-\alpha)\tilde{\mathbf{x}} \in \mathcal{X}$
- Indeed,
  $$\alpha\hat{\mathbf{x}}+(1-\alpha)\tilde{\mathbf{x}} = [\alpha\hat{\mathbf{x}}_1+(1-\alpha)\tilde{\mathbf{x}}_1]+[\alpha\hat{\mathbf{x}}_2+(1-\alpha)\tilde{\mathbf{x}}_2] = \mathbf{y}_1+\mathbf{y}_2,$$
  where $\mathbf{y}_1 \in C_1$ and $\mathbf{y}_2 \in C_2$ since sets C_1, C_2 are convex.

Corollary

Linear combination of convex sets is convex set

Exercise

Proof that Cartesian product of convex sets is convex

Cones

Definition
A set $\mathcal{K}$ is a cone if for any $\mathbf{x} \in \mathcal{K}$ and arbitrary number $\theta \geq 0$ we have $\theta \mathbf{x} \in \mathcal{K}$.

Definition
*A set $\mathcal{K}$ is called **convex** cone if for any points $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{K}$ and any numbers $\theta_1 \geq 0,\ \theta_2 \geq 0$ we have $\theta_1 \mathbf{x}_1 + \theta_2 \mathbf{x}_2 \in \mathcal{K}$.*

Cones

Definition

A set $\mathcal{K}$ is a cone if for any $\mathbf{x} \in \mathcal{K}$ and arbitrary number $\theta \geq 0$ we have $\theta \mathbf{x} \in \mathcal{K}$.

Definition

A set $\mathcal{K}$ is called ***convex*** *cone if for any points $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{K}$ and any numbers $\theta_1 \geq 0,\ \theta_2 \geq 0$ we have $\theta_1 \mathbf{x}_1 + \theta_2 \mathbf{x}_2 \in \mathcal{K}$.*

Important cones

Cones

Definition

A set $\mathcal{K}$ is a cone if for any $\mathbf{x} \in \mathcal{K}$ and arbitrary number $\theta \geq 0$ we have $\theta \mathbf{x} \in \mathcal{K}$.

Definition

*A set $\mathcal{K}$ is called **convex** cone if for any points $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{K}$ and any numbers $\theta_1 \geq 0,\ \theta_2 \geq 0$ we have $\theta_1 \mathbf{x}_1 + \theta_2 \mathbf{x}_2 \in \mathcal{K}$.*

Important cones

- Nonnegative orthant $\mathbb{R}^n_+ = \{\mathbf{x} \in \mathbb{R}^n \mid x_i \geq 0,\ i = 1, \dots, n\} \rightarrow$ Linear programming (LP)

Cones

Definition

A set $\mathcal{K}$ is a cone if for any $\mathbf{x} \in \mathcal{K}$ and arbitrary number $\theta \geq 0$ we have $\theta\mathbf{x} \in \mathcal{K}$.

Definition

A set $\mathcal{K}$ is called ***convex*** *cone if for any points $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{K}$ and any numbers $\theta_1 \geq 0,\ \theta_2 \geq 0$ we have $\theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2 \in \mathcal{K}$.*

Important cones

- Nonnegative orthant $\mathbb{R}^n_+ = \{\mathbf{x} \in \mathbb{R}^n \mid x_i \geq 0,\ i = 1, \ldots, n\} \rightarrow$ Linear programming (LP)
- Second-order cone $\{(\mathbf{x}, t) \in \mathbb{R}^{n+1} \mid \|\mathbf{x}\|_2 \leq t\} \rightarrow$ Second-order cone programming (SOCP)

Cones

Definition

A set $\mathcal{K}$ is a cone if for any $\mathbf{x} \in \mathcal{K}$ and arbitrary number $\theta \geq 0$ we have $\theta\mathbf{x} \in \mathcal{K}$.

Definition

*A set $\mathcal{K}$ is called **convex** cone if for any points $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{K}$ and any numbers $\theta_1 \geq 0,\ \theta_2 \geq 0$ we have $\theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2 \in \mathcal{K}$.*

Important cones

- Nonnegative orthant $\mathbb{R}^n_+ = \{\mathbf{x} \in \mathbb{R}^n \mid x_i \geq 0,\ i = 1, \dots, n\} \rightarrow$ Linear programming (LP)
- Second-order cone $\{(\mathbf{x}, t) \in \mathbb{R}^{n+1} \mid \|\mathbf{x}\|_2 \leq t\} \rightarrow$ Second-order cone programming (SOCP)
- Symmetric positive semi-definite matrices $\mathbf{S}^n_+ \rightarrow$ Semidefinite programming (SDP)

Convex hull

Definition

Convex hull of the set $\mathcal{G}$ is called such set $\operatorname{conv}(\mathcal{G})$ that

- *it is an intersection of all convex sets containing $\mathcal{G}$*
- *it is a set of all convex combinations of points from $\mathcal{G}$*

$$\operatorname{conv}(\mathcal{G}) = \left\{ \sum_{i=1}^{k} \theta_i \mathbf{x}_i \mid \mathbf{x}_i \in \mathcal{G}, \sum_{i=1}^{k} \theta_i = 1, \theta_i \geq 0 \right\}$$

- *it is a minimal convex set containing $\mathcal{G}$*

How to use convex hulls

- Assume that you face with optimization problem with **non-convex** feasible set

How to use convex hulls

- Assume that you face with optimization problem with **non-convex** feasible set
- You can convexify feasible set with its convex hull

How to use convex hulls

- Assume that you face with optimization problem with **non-convex** feasible set
- You can convexify feasible set with its convex hull
- Solve the problem in the new feasible set

How to use convex hulls

- Assume that you face with optimization problem with **non-convex** feasible set
- You can convexify feasible set with its convex hull
- Solve the problem in the new feasible set
- Recover approximate solution of the original problem from the solution of the problem with convex feasible set

Convex function

Definition

Function $f : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ *is called convex (strictly convex), if* $\mathcal{X}$ *is convex set and* $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$ *and* $\alpha \in [0, 1]$ $(\alpha \in (0, 1))$ *we have:*

$$f(\alpha \mathbf{x}_1 + (1 - \alpha)\mathbf{x}_2) \leq \ (<) \ \alpha f(\mathbf{x}_1) + (1 - \alpha) f(\mathbf{x}_2)$$

Convex function

Definition

Function $f : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ *is called convex (strictly convex), if* $\mathcal{X}$ *is convex set and* $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$ *and* $\alpha \in [0,1]$ $(\alpha \in (0,1))$ *we have:*

$$f(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leq \ (<) \ \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2)$$

Definition

Function f *is concave, if function* $-f$ *is convex.*

Convex function

Definition

Function $f : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ is called convex (strictly convex), if $\mathcal{X}$ is convex set and $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$ and $\alpha \in [0,1]$ $(\alpha \in (0,1))$ we have:

$$f(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leq \; (<) \; \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2)$$

Definition

Function f is concave, if function $-f$ is convex.

Examples of convex functions

- x^p for $x \geq 0$ and $p \geq 1$
- $x \log x$, where $x > 0$
- $\max\{x_1, \ldots, x_n\}$
- $\|\mathbf{x}\|$
- $\log \left(\sum\limits_{i=1}^{n} e^{x_i} \right)$
- $-\log \det \mathbf{X}$ for $\mathbf{X} \in \mathbf{S}^n_{++}$

Epigraph and convexity

Definition

A set $\mathrm{epi}\ f = \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} \mid t \geq f(\mathbf{x})\}$ *is called epigraph of* f.

Epigraph and convexity

Definition
A set $\operatorname{epi} f = \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} \mid t \geq f(\mathbf{x})\}$ *is called epigraph of* f.

Theorem
Function f *is convex* $\Leftrightarrow \operatorname{epi} f$ *is convex set.*

Epigraph and convexity

Definition

A set $\operatorname{epi} f = \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} \mid t \geq f(\mathbf{x})\}$ *is called epigraph of* f.

Theorem

Function f *is convex* $\Leftrightarrow \operatorname{epi} f$ *is convex set.*

Proof

1. Let f be a convex function

Epigraph and convexity

Definition

A set $\operatorname{epi} f = \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} \mid t \geq f(\mathbf{x})\}$ *is called epigraph of* f*.*

Theorem

Function f *is convex* $\Leftrightarrow \operatorname{epi} f$ *is convex set.*

Proof

1. Let f be a convex function
    - Consider any two points from epigraph $(\mathbf{x}_1, t_1)$ and $(\mathbf{x}_2, t_2)$, where $t_1 \geq f(\mathbf{x}_1)$ and $t_2 \geq f(\mathbf{x}_2)$

Epigraph and convexity

Definition
A set $\text{epi } f = \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} \mid t \geq f(\mathbf{x})\}$ *is called epigraph of* f.

Theorem
Function f *is convex* $\Leftrightarrow$ $\text{epi } f$ *is convex set.*

Proof
1. Let f be a convex function
   - Consider any two points from epigraph $(\mathbf{x}_1, t_1)$ and $(\mathbf{x}_2, t_2)$, where $t_1 \geq f(\mathbf{x}_1)$ and $t_2 \geq f(\mathbf{x}_2)$
   - Check that the point $(\alpha \mathbf{x}_1 + (1 - \alpha)\mathbf{x}_2, \alpha t_1 + (1 - \alpha) t_2)$ also belongs to epigraph

Epigraph and convexity

Definition

A set $\operatorname{epi} f = \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} \mid t \geq f(\mathbf{x})\}$ *is called epigraph of* f*.*

Theorem

Function f *is convex* $\Leftrightarrow \operatorname{epi} f$ *is convex set.*

Proof

1. Let f be a convex function
   - Consider any two points from epigraph $(\mathbf{x}_1, t_1)$ and $(\mathbf{x}_2, t_2)$, where $t_1 \geq f(\mathbf{x}_1)$ and $t_2 \geq f(\mathbf{x}_2)$
   - Check that the point $(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2, \alpha t_1 + (1-\alpha)t_2)$ also belongs to epigraph
   - From the convexity of f follows
     $\alpha t_1 + (1-\alpha)t_2 \geq \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2) \geq f(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2)$.

Epigraph and convexity

Definition

A set $\operatorname{epi} f = \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} \mid t \geq f(\mathbf{x})\}$ *is called epigraph of* f.

Theorem

Function f *is convex* $\Leftrightarrow$ $\operatorname{epi} f$ *is convex set.*

Proof

1. Let f be a convex function
   - Consider any two points from epigraph $(\mathbf{x}_1, t_1)$ and $(\mathbf{x}_2, t_2)$, where $t_1 \geq f(\mathbf{x}_1)$ and $t_2 \geq f(\mathbf{x}_2)$
   - Check that the point $(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2, \alpha t_1 + (1-\alpha)t_2)$ also belongs to epigraph
   - From the convexity of f follows
     $\alpha t_1 + (1-\alpha)t_2 \geq \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2) \geq f(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2)$.
2. Let epigraph $\operatorname{epi} f$ is convex set

Epigraph and convexity

Definition

A set $\operatorname{epi} f = \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} \mid t \geq f(\mathbf{x})\}$ *is called epigraph of* f*.*

Theorem

Function f *is convex* $\Leftrightarrow \operatorname{epi} f$ *is convex set.*

Proof

1. Let f be a convex function
   - Consider any two points from epigraph $(\mathbf{x}_1, t_1)$ and $(\mathbf{x}_2, t_2)$, where $t_1 \geq f(\mathbf{x}_1)$ and $t_2 \geq f(\mathbf{x}_2)$
   - Check that the point $(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2, \alpha t_1 + (1-\alpha)t_2)$ also belongs to epigraph
   - From the convexity of f follows
     $\alpha t_1 + (1-\alpha)t_2 \geq \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2) \geq f(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2)$.
2. Let epigraph $\operatorname{epi} f$ is convex set
   - $(\mathbf{x}_1, f(\mathbf{x}_1))$ and $(\mathbf{x}_2, f(\mathbf{x}_2)) \in \operatorname{epi} f$, then
     $(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2, \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2)) \in \operatorname{epi} f$

Epigraph and convexity

Definition

A set $\mathrm{epi}\ f = \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} \mid t \geq f(\mathbf{x})\}$ *is called epigraph of* f.

Theorem

Function f *is convex* $\Leftrightarrow$ $\mathrm{epi}\ f$ *is convex set.*

Proof

1. Let f be a convex function
   - Consider any two points from epigraph $(\mathbf{x}_1, t_1)$ and $(\mathbf{x}_2, t_2)$, where $t_1 \geq f(\mathbf{x}_1)$ and $t_2 \geq f(\mathbf{x}_2)$
   - Check that the point $(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2, \alpha t_1 + (1-\alpha)t_2)$ also belongs to epigraph
   - From the convexity of f follows $\alpha t_1 + (1-\alpha)t_2 \geq \alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2) \geq f(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2)$.
2. Let epigraph $\mathrm{epi}\ f$ is convex set
   - $(\mathbf{x}_1, f(\mathbf{x}_1))$ and $(\mathbf{x}_2, f(\mathbf{x}_2)) \in \mathrm{epi}\ f$, then $(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2, \alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2)) \in \mathrm{epi}\ f$
   - From the definition of epigraph follows convexity of f

Strongly convex function

Definition

*Function $f : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ is called **strongly** convex with constant $m > 0$, if $\mathcal{X}$ is convex set and $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$ и $\alpha \in [0, 1]$ we have:*

$$f(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \le \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2) - \frac{m}{2}\alpha(1-\alpha)\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2$$

Strongly convex function

Definition

*Function $f : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ is called **strongly** convex with constant $m > 0$, if $\mathcal{X}$ is convex set and $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$ и $\alpha \in [0,1]$ we have:*

$$f(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leq \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2) - \frac{m}{2}\alpha(1-\alpha)\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2$$

- Convexity $\supset$ strict convexity $\supset$ strong convexity

Strongly convex function

Definition

Function $f : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ is called ***strongly*** *convex with constant $m > 0$, if $\mathcal{X}$ is convex set and $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$ и $\alpha \in [0, 1]$ we have:*
$$f(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leq \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2) - \frac{m}{2}\alpha(1-\alpha)\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2$$

- Convexity $\supset$ strict convexity $\supset$ strong convexity
- Theoretical analysis of methods in the case of strongly convex functions significantly differs from the one for convex functions

Gradient and hessian: preliminaries

Consider $f : \mathbb{R}^n \to \mathbb{R}$

- Directional derivative

$$f'_{\mathbf{d}}(\mathbf{x}) = \lim_{\alpha \to 0} \frac{f(\mathbf{x} + \alpha \mathbf{d}) - f(\mathbf{x})}{\alpha}$$

- Gradient $f'(\mathbf{x})$ is a vector such that $[f'(\mathbf{x})]_i = \frac{\partial f}{\partial x_i}$
- Hessian is a square matrix $f''(\mathbf{x})$ such that $[f''(\mathbf{x})]_{ij} = \frac{\partial f}{\partial x_i x_j}$

Differential criteria of convexity

We consider convex function as strongly convex function with $m = 0$.

Differential criteria of convexity

We consider convex function as strongly convex function with $m = 0$.

Theorem (First order criterion)

Let function $f(\mathbf{x})$ is differentiable and its domain is a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then $f(\mathbf{x})$ is strongly convex with $m \geq 0$ iff

$$f(\mathbf{x}) - f(\mathbf{x}^*) \geq \left\langle f'(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \right\rangle + \frac{m}{2} \|\mathbf{x} - \mathbf{x}^*\|_2^2, \quad \forall \mathbf{x}, \mathbf{x}^* \in X.$$

Differential criteria of convexity

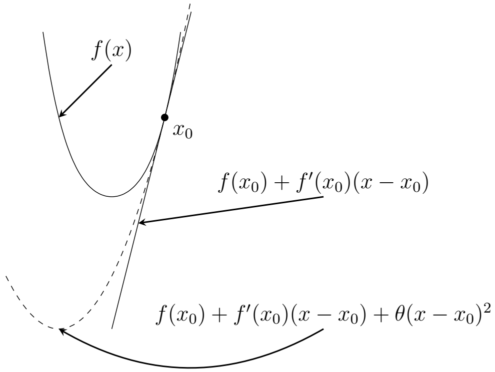
We consider convex function as strongly convex function with $m = 0$.

Theorem (First order criterion)

Let function $f(\mathbf{x})$ is differentiable and its domain is a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then $f(\mathbf{x})$ is strongly convex with $m \geq 0$ iff

$$f(\mathbf{x}) - f(\mathbf{x}^*) \geq \left\langle f'(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \right\rangle + \frac{m}{2} \|\mathbf{x} - \mathbf{x}^*\|_2^2, \quad \forall \mathbf{x}, \mathbf{x}^* \in X.$$

Illustration for the first order criterion



Theorem (Second order criterion)

Twice continuously differentiable function f is convex $\Leftrightarrow$
$f''(\mathbf{x}) \succeq m\mathbf{I}$

What function compositions preserve convexity?

- If $f(\mathbf{x})$ is convex, then $g(\mathbf{x}) = f(\mathbf{A}\mathbf{x} + \mathbf{b})$ is convex

What function compositions preserve convexity?

- If $f(\mathbf{x})$ is convex, then $g(\mathbf{x}) = f(\mathbf{A}\mathbf{x} + \mathbf{b})$ is convex
- If $f(\mathbf{x})$ is convex, then $g(t) = f(\mathbf{x} + t\mathbf{y})$ is convex

What function compositions preserve convexity?

- If $f(\mathbf{x})$ is convex, then $g(\mathbf{x}) = f(\mathbf{A}\mathbf{x} + \mathbf{b})$ is convex
- If $f(\mathbf{x})$ is convex, then $g(t) = f(\mathbf{x} + t\mathbf{y})$ is convex
- If $f_i(\mathbf{x})$ are convex, then $f(\mathbf{x}) = \max_{i=1,\ldots,m} f_i(\mathbf{x})$ is convex

What function compositions preserve convexity?

- If $f(\mathbf{x})$ is convex, then $g(\mathbf{x}) = f(\mathbf{A}\mathbf{x} + \mathbf{b})$ is convex
- If $f(\mathbf{x})$ is convex, then $g(t) = f(\mathbf{x} + t\mathbf{y})$ is convex
- If $f_i(\mathbf{x})$ are convex, then $f(\mathbf{x}) = \max\limits_{i=1,\ldots,m} f_i(\mathbf{x})$ is convex
- The sum of convex functions with non-negative coefficients is convex function

What function compositions preserve convexity?

- If $f(\mathbf{x})$ is convex, then $g(\mathbf{x}) = f(\mathbf{A}\mathbf{x} + \mathbf{b})$ is convex
- If $f(\mathbf{x})$ is convex, then $g(t) = f(\mathbf{x} + t\mathbf{y})$ is convex
- If $f_i(\mathbf{x})$ are convex, then $f(\mathbf{x}) = \max\limits_{i=1,\dots,m} f_i(\mathbf{x})$ is convex
- The sum of convex functions with non-negative coefficients is convex function
- Scalar composition $h(f(\mathbf{x}))$

Local minimum of convex function is also a global minimum

Theorem

If f is a convex function and $\mathbf{x}^*$ is a point of local minimum, the $\mathbf{x}^*$ is a point of global minimum.

Local minimum of convex function is also a global minimum

Theorem

If f is a convex function and $\mathbf{x}^*$ is a point of local minimum, the $\mathbf{x}^*$ is a point of global minimum.

Proof

- Assume that there exists a point $\mathbf{y}^*$ such that $\mathbf{y}^* \neq \mathbf{x}^*$ and $\mathbf{y}^*$ is a point of global minimum: $f(\mathbf{y}^*) < f(\mathbf{x}^*)$

Local minimum of convex function is also a global minimum

Theorem

If f is a convex function and $\mathbf{x}^*$ is a point of local minimum, the $\mathbf{x}^*$ is a point of global minimum.

Proof

- Assume that there exists a point $\mathbf{y}^*$ such that $\mathbf{y}^* \neq \mathbf{x}^*$ and $\mathbf{y}^*$ is a point of global minimum: $f(\mathbf{y}^*) < f(\mathbf{x}^*)$
- By definition of a point of local minimum: $f(\mathbf{x}^*) \leq f(\mathbf{x})$, where $\|\mathbf{x}^* - \mathbf{x}\|_2 \leq \delta$

Local minimum of convex function is also a global minimum

Theorem

If f is a convex function and $\mathbf{x}^*$ is a point of local minimum, the $\mathbf{x}^*$ is a point of global minimum.

Proof

- Assume that there exists a point $\mathbf{y}^*$ such that $\mathbf{y}^* \neq \mathbf{x}^*$ and $\mathbf{y}^*$ is a point of global minimum: $f(\mathbf{y}^*) < f(\mathbf{x}^*)$
- By definition of a point of local minimum: $f(\mathbf{x}^*) \leq f(\mathbf{x})$, where $\|\mathbf{x}^* - \mathbf{x}\|_2 \leq \delta$
- Choose sufficiently small $\alpha \in (0, 1)$ and consider a point $\mathbf{z} = (1 - \alpha)\mathbf{x}^* + \alpha\mathbf{y}^*$ such that $\|\mathbf{x}^* - \mathbf{z}\|_2 \leq \delta$

Local minimum of convex function is also a global minimum

Theorem

If f is a convex function and $\mathbf{x}^*$ is a point of local minimum, the $\mathbf{x}^*$ is a point of global minimum.

Proof

- Assume that there exists a point $\mathbf{y}^*$ such that $\mathbf{y}^* \neq \mathbf{x}^*$ and $\mathbf{y}^*$ is a point of global minimum: $f(\mathbf{y}^*) < f(\mathbf{x}^*)$
- By definition of a point of local minimum: $f(\mathbf{x}^*) \leq f(\mathbf{x})$, where $\|\mathbf{x}^* - \mathbf{x}\|_2 \leq \delta$
- Choose sufficiently small $\alpha \in (0, 1)$ and consider a point $\mathbf{z} = (1-\alpha)\mathbf{x}^* + \alpha\mathbf{y}^*$ such that $\|\mathbf{x}^* - \mathbf{z}\|_2 \leq \delta$
- $f(\mathbf{x}^*) \leq f(\mathbf{z}) \leq \alpha f(\mathbf{y}^*) + (1-\alpha)f(\mathbf{x}^*) < f(\mathbf{x}^*)$

Local minimum of convex function is also a global minimum

Theorem

If f is a convex function and $\mathbf{x}^*$ is a point of local minimum, the $\mathbf{x}^*$ is a point of global minimum.

Proof

- Assume that there exists a point $\mathbf{y}^*$ such that $\mathbf{y}^* \neq \mathbf{x}^*$ and $\mathbf{y}^*$ is a point of global minimum: $f(\mathbf{y}^*) < f(\mathbf{x}^*)$
- By definition of a point of local minimum: $f(\mathbf{x}^*) \leq f(\mathbf{x})$, where $\|\mathbf{x}^* - \mathbf{x}\|_2 \leq \delta$
- Choose sufficiently small $\alpha \in (0, 1)$ and consider a point $\mathbf{z} = (1 - \alpha)\mathbf{x}^* + \alpha \mathbf{y}^*$ such that $\|\mathbf{x}^* - \mathbf{z}\|_2 \leq \delta$
- $f(\mathbf{x}^*) \leq f(\mathbf{z}) \leq \alpha f(\mathbf{y}^*) + (1 - \alpha) f(\mathbf{x}^*) < f(\mathbf{x}^*)$
- We get a contradiction, therefore assumption is incorrect and $\mathbf{x}^*$ is a point of global minimum

Jensen's inequality

Theorem

If function f is convex, then $f\left(\sum_{i=1}^{k} \alpha_i \mathbf{x}_i\right) \le \sum_{i=1}^{k} \alpha_i f(\mathbf{x}_i)$, *where* $\sum_{i=1}^{k} \alpha_i = 1,\ \alpha_i \ge 0.$

Jensen's inequality

Theorem

If function f is convex, then $f\left(\sum_{i=1}^{k} \alpha_i \mathbf{x}_i\right) \le \sum_{i=1}^{k} \alpha_i f(\mathbf{x}_i)$, *where* $\sum_{i=1}^{k} \alpha_i = 1,\ \alpha_i \ge 0.$

Proof by induction

- Base $k = 2$ holds according to the definition

Jensen's inequality

Theorem

If function f is convex, then $f\left(\sum_{i=1}^{k}\alpha_i\mathbf{x}_i\right) \leq \sum_{i=1}^{k}\alpha_i f(\mathbf{x}_i)$, *where* $\sum_{i=1}^{k}\alpha_i = 1,\ \alpha_i \geq 0.$

Proof by induction

- Base $k = 2$ holds according to the definition
- Assume the inequality holds for $k = m - 1$:
  $f\left(\sum_{i=1}^{m-1}\alpha\mathbf{x}_i\right) \leq \sum_{i=1}^{m-1}\alpha_i f(\mathbf{x}_i)$ and $\sum_{i=1}^{m-1}\alpha_i = 1$, $\alpha_i \geq 0$

Jensen's inequality

Theorem

If function f is convex, then $f\left(\sum_{i=1}^{k}\alpha_i\mathbf{x}_i\right) \leq \sum_{i=1}^{k}\alpha_i f(\mathbf{x}_i)$, *where* $\sum_{i=1}^{k}\alpha_i = 1,\ \alpha_i \geq 0.$

Proof by induction

- Base $k = 2$ holds according to the definition
- Assume the inequality holds for $k = m - 1$:
  $f\left(\sum_{i=1}^{m-1}\alpha\mathbf{x}_i\right) \leq \sum_{i=1}^{m-1}\alpha_i f(\mathbf{x}_i)$ and $\sum_{i=1}^{m-1}\alpha_i = 1$, $\alpha_i \geq 0$
- Consider $k = m$: $f\left(\sum_{i=1}^{m}\hat{\alpha}_i\mathbf{x}_i\right) = f\left(\sum_{i=1}^{m-1}\hat{\alpha}\mathbf{x}_i + \hat{\alpha}_m\mathbf{x}_m\right) =$
  $f\left((1-\hat{\alpha}_m)\sum_{i=1}^{m-1}\frac{\hat{\alpha}_i}{1-\hat{\alpha}_m}\mathbf{x}_i + \hat{\alpha}_m\mathbf{x}_m\right) \leq$
  $(1-\hat{\alpha}_m)f\left(\sum_{i=1}^{m-1}\frac{\hat{\alpha}_i}{1-\hat{\alpha}_m}\mathbf{x}_i\right) + \hat{\alpha}_m f(\mathbf{x}_m) \leq \sum_{i=1}^{k}\alpha_i f(\mathbf{x}_i)$

Corollaries and generalizations

- If we write Jensen's inequality for the function $-\log x$, we get inequality for geometric and arithmetic means

$$\frac{1}{m}\sum_{i=1}^{m} x_i \geq \sqrt[m]{x_1 \cdot \ldots \cdot x_m}$$

Corollaries and generalizations

- If we write Jensen's inequality for the function $-\log x$, we get inequality for geometric and arithmetic means

$$\frac{1}{m}\sum_{i=1}^{m} x_i \geq \sqrt[m]{x_1 \cdot \ldots \cdot x_m}$$

- Hölder's inequality

$$\sum_{i=1}^{n} x_i y_i \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |y_i|^q\right)^{1/q}, \quad \frac{1}{p} + \frac{1}{q} = 1$$

Corollaries and generalizations

- If we write Jensen's inequality for the function $-\log x$, we get inequality for geometric and arithmetic means

$$\frac{1}{m}\sum_{i=1}^{m} x_i \geq \sqrt[m]{x_1 \cdot \ldots \cdot x_m}$$

- Hölder's inequality

$$\sum_{i=1}^{n} x_i y_i \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |y_i|^q\right)^{1/q}, \quad \frac{1}{p}+\frac{1}{q}=1$$

- The generalization of Jensen's inequality gives the inequality for the convex function of the expected value

$$f(\mathbb{E}(\mathbf{x})) \leq \mathbb{E}(f(\mathbf{x}))$$

Summary on convex functions

- Convex, strictly convex and strongly convex functions

Summary on convex functions

- Convex, strictly convex and strongly convex functions
- Examples and how to verify convexity of function

Summary on convex functions

- Convex, strictly convex and strongly convex functions
- Examples and how to verify convexity of function
- Operations that preserve convexity

Summary on convex functions

- Convex, strictly convex and strongly convex functions
- Examples and how to verify convexity of function
- Operations that preserve convexity
- Jensen inequality