

### **Convex sets**





#### Affine set

Suppose  $x_1, x_2$  are two points in  $\mathbb{R}^{\ltimes}$ . Then the line passing through them is defined as follows:

$$x = \theta x_1 + (1 - \theta) x_2, \theta \in \mathbb{R}$$

The set A is called **affine** if for any  $x_1, x_2$  from A the line passing through them also lies in A, i.e.

$$\forall \theta \in \mathbb{R}, \forall x_1, x_2 \in A : \theta x_1 + (1 - \theta)x_2 \in A$$

- i Example
  - $\mathbb{R}^n$  is an affine set.

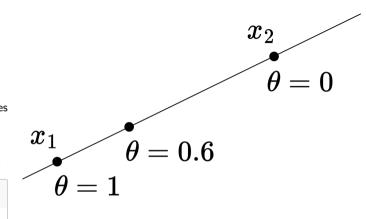


Figure 1: Illustration of a line between two vectors  $x_1$  and  $x_2$ 

y,z Convex sets

#### Affine set

Suppose  $x_1, x_2$  are two points in  $\mathbb{R}^{\times}$ . Then the line passing through them is defined as follows:

$$x = \theta x_1 + (1 - \theta) x_2, \theta \in \mathbb{R}$$

The set A is called **affine** if for any  $x_1, x_2$  from A the line passing through them also lies in A, i.e.

$$\forall \theta \in \mathbb{R}, \forall x_1, x_2 \in A : \theta x_1 + (1 - \theta)x_2 \in A$$

- i Example
  - $\mathbb{R}^n$  is an affine set.
  - The set of solutions  $\{x \mid \mathbf{A}x = \mathbf{b}\}$  is also an affine set.

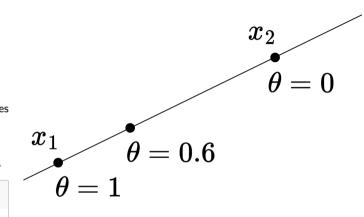


Figure 1: Illustration of a line between two vectors  $x_1$  and  $x_2$ 

Convex sets

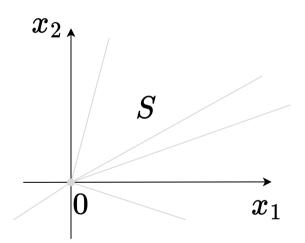
**⊕** 0 0

### Cone

A non-empty set S is called a cone, if:

$$\forall x \in S, \ \theta \ge 0 \ \rightarrow \ \theta x \in S$$

For any point in cone it also contains beam through this point.







The set S is called a convex cone, if:

$$\forall x_1, x_2 \in S, \ \theta_1, \theta_2 \ge 0 \quad \to \quad \theta_1 x_1 + \theta_2 x_2 \in S$$

Convex cone is just like cone, but it is also convex.



Convex cone: set that contains all conic combinations of points in the set

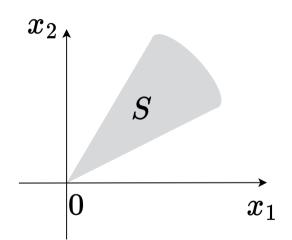
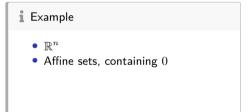


Figure 3: Illustration of a convex cone

The set S is called a convex cone, if:

$$\forall x_1, x_2 \in S, \ \theta_1, \theta_2 \ge 0 \quad \to \quad \theta_1 x_1 + \theta_2 x_2 \in S$$

Convex cone is just like cone, but it is also convex.



Convex cone: set that contains all conic combinations of points in the set

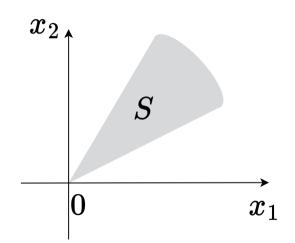


Figure 3: Illustration of a convex cone

The set S is called a convex cone, if:

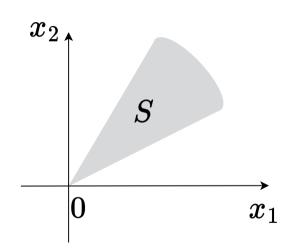
$$\forall x_1, x_2 \in S, \ \theta_1, \theta_2 \ge 0 \quad \to \quad \theta_1 x_1 + \theta_2 x_2 \in S$$

Convex cone is just like cone, but it is also convex.



- lacksquare
- Affine sets, containing 0
- Ray

Convex cone: set that contains all conic combinations of points in the set



Convex sets

The set S is called a convex cone, if:

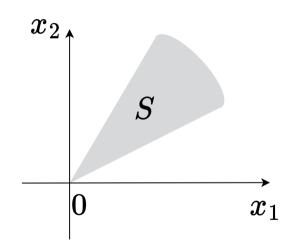
$$\forall x_1, x_2 \in S, \ \theta_1, \theta_2 \ge 0 \quad \to \quad \theta_1 x_1 + \theta_2 x_2 \in S$$

Convex cone is just like cone, but it is also convex.

# i Example

- lacksquare
- Affine sets, containing 0
- Ray
- S<sup>n</sup><sub>+</sub> the set of symmetric positive semi-definite matrices

Convex cone: set that contains all conic combinations of points in the set



Convex sets

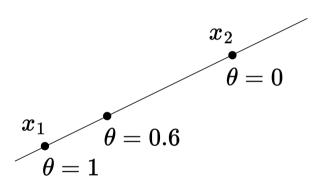
Figure 3: Illustration of a convex cone

### Line segment

Suppose  $x_1, x_2$  are two points in  $\mathbb{R}^n$ . Then the line segment between them is defined as follows:

$$x = \theta x_1 + (1 - \theta)x_2, \ \theta \in [0, 1]$$

Convex set contains line segment between any two points in the set.



 $f \to \min_{x \in \mathbb{R}^n}$ 

Convex sets

#### Convex set

The set S is called **convex** if for any  $x_1,x_2$  from S the line segment between them also lies in S, i.e.

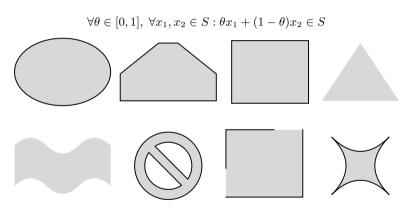


Figure 5: Top: examples of convex sets. Bottom: examples of non-convex sets.

### **i** Example

An empty set and a set from a single vector are convex by definition.

### i Example

Any affine set, a ray, a line segment - they all are convex sets.

Convex sets

### **Convex combination**

Let  $x_1, x_2, \ldots, x_k \in S$ , then the point  $\theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_k x_k$  is called the convex combination of points  $x_1, x_2, \ldots, x_k$  if  $\sum_{i=1}^k \theta_i = 1, \ \theta_i \geq 0$ .



#### Convex hull

The set of all convex combinations of points from S is called the convex hull of the set S.

$$\mathbf{conv}(S) = \left\{ \sum_{i=1}^k \theta_i x_i \mid x_i \in S, \sum_{i=1}^k \theta_i = 1, \ \theta_i \ge 0 \right\}$$

• The set  $\mathbf{conv}(S)$  is the smallest convex set containing S.

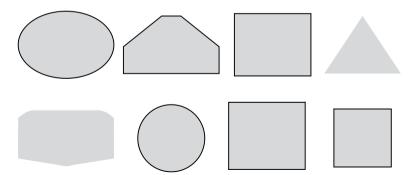


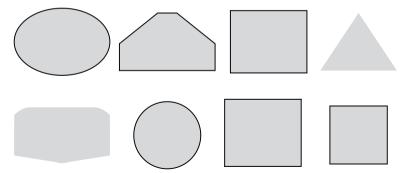
Figure 6: Top: convex hulls of the convex sets. Bottom: convex hull of the non-convex sets.  $\frac{1}{100}$ 

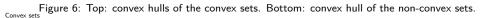
#### Convex hull

The set of all convex combinations of points from S is called the convex hull of the set S.

$$\mathbf{conv}(S) = \left\{ \sum_{i=1}^k \theta_i x_i \mid x_i \in S, \sum_{i=1}^k \theta_i = 1, \ \theta_i \ge 0 \right\}$$

- The set  $\mathbf{conv}(S)$  is the smallest convex set containing S.
- The set S is convex if and only if  $S = \mathbf{conv}(S)$ .







#### Minkowski addition

The Minkowski sum of two sets of vectors  $S_1$  and  $S_2$  in Euclidean space is formed by adding each vector in  $S_1$  to each vector in  $S_2$ .

$$S_1 + S_2 = \{ \mathbf{s_1} + \mathbf{s_2} \mid \mathbf{s_1} \in S_1, \ \mathbf{s_2} \in S_2 \}$$

Similarly, one can define a linear combination of the sets.

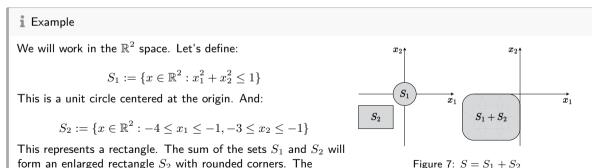


### Minkowski addition

The Minkowski sum of two sets of vectors  $S_1$  and  $S_2$  in Euclidean space is formed by adding each vector in  $S_1$  to each vector in  $S_2$ .

$$S_1 + S_2 = \{ \mathbf{s_1} + \mathbf{s_2} \mid \mathbf{s_1} \in S_1, \ \mathbf{s_2} \in S_2 \}$$

Similarly, one can define a linear combination of the sets.



resulting set will be convex.

## **Finding convexity**

In practice, it is very important to understand whether a specific set is convex or not. Two approaches are used for this depending on the context.

By definition.



## **Finding convexity**

In practice, it is very important to understand whether a specific set is convex or not. Two approaches are used for this depending on the context.

- By definition.
- Show that S is derived from simple convex sets using operations that preserve convexity.



# Finding convexity by definition

$$x_1, x_2 \in S, \ 0 \le \theta \le 1 \quad \rightarrow \quad \theta x_1 + (1 - \theta)x_2 \in S$$

i Example

Prove, that ball in  $\mathbb{R}^n$  (i.e. the following set  $\{\mathbf{x} \mid ||\mathbf{x} - \mathbf{x}_c|| \leq r\}$ ) - is convex.





Which of the sets are convex:

• Stripe,  $\{x \in \mathbb{R}^n \mid \alpha \leq a^\top x \leq \beta\}$ 



- Stripe,  $\{x \in \mathbb{R}^n \mid \alpha \leq a^\top x \leq \beta\}$
- Rectangle,  $\{x \in \mathbb{R}^n \mid \alpha_i \leq x_i \leq \beta_i, i = \overline{1, n}\}$



- Stripe,  $\{x \in \mathbb{R}^n \mid \alpha \leq a^{\top}x \leq \beta\}$
- Rectangle,  $\{x \in \mathbb{R}^n \mid \alpha_i \leq x_i \leq \beta_i, i = \overline{1,n}\}$
- Kleen,  $\{x \in \mathbb{R}^n \mid a_1^\top x \leq \overline{b}_1, a_2^\top x \leq b_2\}$



- Stripe,  $\{x \in \mathbb{R}^n \mid \alpha \leq a^\top x \leq \beta\}$
- Rectangle,  $\{x \in \mathbb{R}^n \mid \alpha_i \leq x_i \leq \beta_i, i = \overline{1, n}\}$
- Kleen,  $\{x \in \mathbb{R}^n \mid a_1^{\top} x \leq b_1, a_2^{\top} x \leq b_2\}$
- A set of points closer to a given point than a given set that does not contain a point,  $\{x \in \mathbb{R}^n \mid ||x x_0||_2 \leq ||x y||_2, \forall y \in S \subseteq \mathbb{R}^n\}$



#### Which of the sets are convex:

- Stripe,  $\{x \in \mathbb{R}^n \mid \alpha \leq a^{\top}x \leq \beta\}$
- Rectangle,  $\{x \in \mathbb{R}^n \mid \alpha_i \leq x_i \leq \beta_i, i = \overline{1,n}\}$
- Kleen,  $\{x \in \mathbb{R}^n \mid a_1^{\top} x \leq b_1, a_2^{\top} x \leq b_2\}$
- A set of points closer to a given point than a given set that does not contain a point,  $\{x \in \mathbb{R}^n \mid ||x x_0||_2 < ||x y||_2, \forall y \in S \subset \mathbb{R}^n\}$
- A set of points, which are closer to one set than another,  $\{x \in \mathbb{R}^n \mid \mathbf{dist}(x,S) \leq \mathbf{dist}(x,T), S, T \subseteq \mathbb{R}^n\}$



Convex sets

#### Which of the sets are convex:

- Stripe,  $\{x \in \mathbb{R}^n \mid \alpha \leq a^{\top}x \leq \beta\}$
- Rectangle,  $\{x \in \mathbb{R}^n \mid \alpha_i \leq x_i \leq \beta_i, i = \overline{1,n}\}$
- Kleen,  $\{x \in \mathbb{R}^n \mid a_1^{\top} x \leq b_1, a_2^{\top} x \leq b_2\}$
- A set of points closer to a given point than a given set that does not contain a point,  $\{x \in \mathbb{R}^n \mid ||x x_0||_2 \leq ||x y||_2, \forall y \in S \subseteq \mathbb{R}^n\}$
- A set of points, which are closer to one set than another,  $\{x \in \mathbb{R}^n \mid \mathbf{dist}(x,S) \leq \mathbf{dist}(x,T), S, T \subseteq \mathbb{R}^n\}$
- A set of points,  $\{x \in \mathbb{R}^n \mid x + X \subseteq S\}$ , where  $S \subseteq \mathbb{R}^n$  is convex and  $X \subseteq \mathbb{R}^n$  is arbitrary.

Convex sets

- Stripe,  $\{x \in \mathbb{R}^n \mid \alpha \leq a^{\top}x \leq \beta\}$
- Rectangle,  $\{x \in \mathbb{R}^n \mid \alpha_i < x_i < \beta_i, i = \overline{1, n}\}$
- Kleen,  $\{x \in \mathbb{R}^n \mid a_1^\top x < b_1, a_2^\top x < b_2\}$
- A set of points closer to a given point than a given set that does not contain a point.
  - $\{x \in \mathbb{R}^n \mid ||x x_0||_2 < ||x y||_2, \forall y \in S \subset \mathbb{R}^n\}$
- A set of points, which are closer to one set than another,  $\{x \in \mathbb{R}^n \mid \mathbf{dist}(x,S) \leq \mathbf{dist}(x,T), S, T \subseteq \mathbb{R}^n\}$
- A set of points,  $\{x \in \mathbb{R}^n \mid x + X \subseteq S\}$ , where  $S \subseteq \mathbb{R}^n$  is convex and  $X \subseteq \mathbb{R}^n$  is arbitrary.
- A set of points whose distance to a given point does not exceed a certain part of the distance to another given point is  $\{x \in \mathbb{R}^n \mid ||x - a||_2 < \theta ||x - b||_2, a, b \in \mathbb{R}^n, 0 < 1\}$



# Operations, that preserve convexity

The linear combination of convex sets is convex Let there be 2 convex sets  $S_x, S_y$ , let the set

$$S = \{s \mid s = c_1 x + c_2 y, \ x \in S_x, \ y \in S_y, \ c_1, c_2 \in \mathbb{R}\}$$

Take two points from  $S: s_1 = c_1x_1 + c_2y_1, s_2 = c_1x_2 + c_2y_2$  and prove that the segment between them  $\theta s_1 + (1-\theta)s_2, \theta \in [0,1]$  also belongs to S

$$\theta s_1 + (1 - \theta)s_2$$

$$\theta(c_1x_1+c_2y_1)+(1-\theta)(c_1x_2+c_2y_2)$$

$$c_1(\theta x_1 + (1 - \theta)x_2) + c_2(\theta y_1 + (1 - \theta)y_2)$$

$$c_1x + c_2y \in S$$

### The intersection of any (!) number of convex sets is convex

If the desired intersection is empty or contains one point, the property is proved by definition. Otherwise, take 2 points and a segment between them. These points must lie in all intersecting sets, and since they are all convex, the segment between them lies in all sets and, therefore, in their intersection.

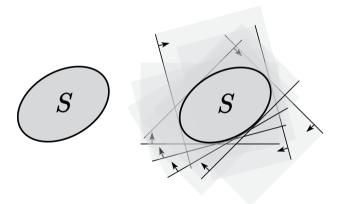


Figure 8: Intersection of halfplanes



# The image of the convex set under affine mapping is convex

$$S \subseteq \mathbb{R}^n$$
 convex  $\rightarrow f(S) = \{f(x) \mid x \in S\}$  convex  $(f(x) = \mathbf{A}x + \mathbf{b})$ 

Examples of affine functions: extension, projection, transposition, set of solutions of linear matrix inequality  $\{x \mid x_1A_1 + \ldots + x_mA_m \leq B\}$ . Here  $A_i, B \in \mathbf{S}^p$  are symmetric matrices  $p \times p$ .

Note also that the prototype of the convex set under affine mapping is also convex.

$$S \subseteq \mathbb{R}^m \text{ convex } \to f^{-1}(S) = \{x \in \mathbb{R}^n \mid f(x) \in S\} \text{ convex } (f(x) = \mathbf{A}x + \mathbf{b})$$

## **Example**

Let  $x \in \mathbb{R}$  is a random variable with a given probability distribution of  $\mathbb{P}(x = a_i) = p_i$ , where  $i = 1, \dots, n$ , and  $a_1 < \dots < a_n$ . It is said that the probability vector of outcomes of  $p \in \mathbb{R}^n$  belongs to the probabilistic simplex, i.e.

$$P = \{p \mid \mathbf{1}^T p = 1, p \succeq 0\} = \{p \mid p_1 + \ldots + p_n = 1, p_i \geq 0\}.$$

Determine if the following sets of p are convex:

• 
$$\mathbb{P}(x > \alpha) < \beta$$

## Example

Let  $x \in \mathbb{R}$  is a random variable with a given probability distribution of  $\mathbb{P}(x = a_i) = p_i$ , where  $i = 1, \dots, n$ , and  $a_1 < \ldots < a_n$ . It is said that the probability vector of outcomes of  $p \in \mathbb{R}^n$  belongs to the probabilistic simplex, i.e.

$$P = \{p \mid \mathbf{1}^T p = 1, p \succeq 0\} = \{p \mid p_1 + \ldots + p_n = 1, p_i \geq 0\}.$$

Determine if the following sets of p are convex:

- $\mathbb{P}(x > \alpha) \leq \beta$
- $\mathbb{E}|x^{201}| < \alpha \mathbb{E}|x|$



# Example

Let  $x \in \mathbb{R}$  is a random variable with a given probability distribution of  $\mathbb{P}(x = a_i) = p_i$ , where  $i = 1, \dots, n$ , and  $a_1 < \ldots < a_n$ . It is said that the probability vector of outcomes of  $p \in \mathbb{R}^n$  belongs to the probabilistic simplex, i.e.

$$P = \{p \mid \mathbf{1}^T p = 1, p \succeq 0\} = \{p \mid p_1 + \ldots + p_n = 1, p_i \geq 0\}.$$

Determine if the following sets of p are convex:

- $\mathbb{P}(x > \alpha) < \beta$
- $\mathbb{E}|x^{201}| < \alpha \mathbb{E}|x|$
- $\mathbb{E}|x^2| > \alpha \mathbb{V}x > \alpha$

### **Convex functions**





#### Jensen's inequality The function f(x), which is

defined on the convex set  $S \subseteq \mathbb{R}^n$ , is called **convex** on S, if:

$$B \subseteq \mathbb{R}$$
 , is called **collect** off  $B$ , ii.

for any  $x_1, x_2 \in S$  and  $0 \le \lambda \le 1$ . If the above inequality holds as strict inequality  $x_1 \neq x_2$  and

 $0 < \lambda < 1$ , then the function is called strictly convex on S.

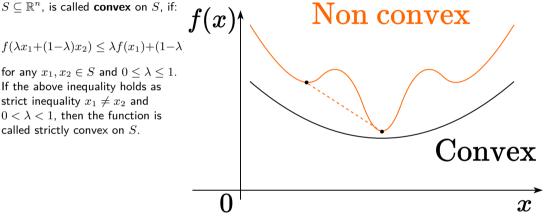


Figure 9: Difference between convex and non-convex function



# Jensen's inequality

#### i Theorem

Let f(x) be a convex function on a convex set  $X \subseteq \mathbb{R}^n$  and let  $x_i \in X, 1 \le i \le m$ , be arbitrary points from X. Then

$$f\left(\sum_{i=1}^{m} \lambda_i x_i\right) \le \sum_{i=1}^{m} \lambda_i f(x_i)$$

for any  $\lambda = [\lambda_1, \dots, \lambda_m] \in \Delta_m$  - probability simplex.

#### **Proof**

1. First, note that the point  $\sum_{i=1}^{m} \lambda_i x_i$  as a convex combination of points from the convex set X belongs to X.

Convex functions

# Jensen's inequality

#### i Theorem

Let f(x) be a convex function on a convex set  $X\subseteq\mathbb{R}^n$  and let  $x_i\in X, 1\leq i\leq m$ , be arbitrary points from X. Then

$$f\left(\sum_{i=1}^{m} \lambda_i x_i\right) \le \sum_{i=1}^{m} \lambda_i f(x_i)$$

for any  $\lambda = [\lambda_1, \dots, \lambda_m] \in \Delta_m$  - probability simplex.

#### **Proof**

- 1. First, note that the point  $\sum_{i=1}^{m} \lambda_i x_i$  as a convex combination of points from the convex set X belongs to X.
- 2. We will prove this by induction. For m=1, the statement is obviously true, and for m=2, it follows from the definition of a convex function.

େ ଚ

#### Jensen's inequality

3. Assume it is true for all m up to m=k, and we will prove it for m=k+1. Let  $\lambda\in\Delta k+1$  and

$$x = \sum_{i=1}^{k+1} \lambda_i x_i = \sum_{i=1}^k \lambda_i x_i + \lambda_{k+1} x_{k+1}.$$

Assuming  $0 < \lambda_{k+1} < 1$ , as otherwise, it reduces to previously considered cases, we have

$$x = \lambda_{k+1} x_{k+1} + (1 - \lambda_{k+1}) \bar{x},$$

where 
$$\bar{x} = \sum_{i=1}^k \gamma_i x_i$$
 and  $\gamma_i = \frac{\lambda_i}{1 - \lambda_{k+1}} \ge 0, 1 \le i \le k$ .

#### Jensen's inequality

3. Assume it is true for all m up to m=k, and we will prove it for m=k+1. Let  $\lambda\in\Delta k+1$  and

$$x = \sum_{i=1}^{k+1} \lambda_i x_i = \sum_{i=1}^{k} \lambda_i x_i + \lambda_{k+1} x_{k+1}.$$

Assuming  $0 < \lambda_{k+1} < 1$ , as otherwise, it reduces to previously considered cases, we have

$$x = \lambda_{k+1} x_{k+1} + (1 - \lambda_{k+1}) \bar{x},$$

where  $\bar{x} = \sum_{i=1}^k \gamma_i x_i$  and  $\gamma_i = \frac{\lambda_i}{1 - \lambda_{k+1}} \ge 0, 1 \le i \le k$ .

4. Since  $\lambda \in \Delta_{k+1}$ , then  $\gamma = [\gamma_1, \dots, \gamma_k] \in \Delta_k$ . Therefore  $\bar{x} \in X$  and by the convexity of f(x) and the induction hypothesis:

$$f\left(\sum_{i=1}^{k+1} \lambda_i x_i\right) = f\left(\lambda_{k+1} x_{k+1} + (1 - \lambda_{k+1})\bar{x}\right) \le \lambda_{k+1} f(x_{k+1}) + (1 - \lambda_{k+1}) f(\bar{x}) \le \sum_{i=1}^{k+1} \lambda_i f(x_i)$$

Thus, initial inequality is satisfied for m=k+1 as well.

• 
$$f(x) = x^p, p > 1, x \in \mathbb{R}_+$$

- $f(x) = x^p, p > 1, x \in \mathbb{R}_+$
- $f(x) = ||x||^p, \ p > 1, x \in \mathbb{R}^n$



- $f(x) = x^p, p > 1, x \in \mathbb{R}_+$
- $f(x) = ||x||^p, \ p > 1, x \in \mathbb{R}^n$
- $f(x) = e^{cx}, c \in \mathbb{R}, x \in \mathbb{R}$



- $f(x) = x^p, p > 1, x \in \mathbb{R}_+$
- $f(x) = ||x||^p, \ p > 1, x \in \mathbb{R}^n$
- $f(x) = e^{cx}, c \in \mathbb{R}, x \in \mathbb{R}$
- $f(x) = -\ln x, \ x \in \mathbb{R}_{++}$



- $f(x) = x^p, p > 1, x \in \mathbb{R}_+$
- $f(x) = ||x||^p, \ p > 1, x \in \mathbb{R}^n$
- $f(x) = e^{cx}, c \in \mathbb{R}, x \in \mathbb{R}$
- $f(x) = -\ln x, \ x \in \mathbb{R}_{++}$
- $f(x) = x \ln x, \ x \in \mathbb{R}_{++}$

େ ଚେ 🕈

- $f(x) = x^p, p > 1, x \in \mathbb{R}_+$
- $f(x) = ||x||^p, \ p > 1, x \in \mathbb{R}^n$
- $f(x) = e^{cx}, c \in \mathbb{R}, x \in \mathbb{R}$
- $f(x) = -\ln x, \ x \in \mathbb{R}_{++}$
- $f(x) = x \ln x, \ x \in \mathbb{R}_{++}$
- The sum of the largest k coordinates  $f(x) = x_{(1)} + \ldots + x_{(k)}, \ x \in \mathbb{R}^n$



 $f \to \min_{x,y,z}$  Convex functions

- $f(x) = x^p, p > 1, x \in \mathbb{R}_+$
- $f(x) = ||x||^p, \ p > 1, x \in \mathbb{R}^n$
- $f(x) = e^{cx}, c \in \mathbb{R}, x \in \mathbb{R}$
- $f(x) = -\ln x, \ x \in \mathbb{R}_{++}$
- $f(x) = x \ln x, \ x \in \mathbb{R}_{++}$
- The sum of the largest k coordinates  $f(x) = x_{(1)} + \ldots + x_{(k)}, \ x \in \mathbb{R}^n$
- $f(X) = \lambda_{max}(X), X = X^T$

- $f(x) = x^p, p > 1, x \in \mathbb{R}_+$
- $f(x) = ||x||^p, \ p > 1, x \in \mathbb{R}^n$
- $f(x) = e^{cx}, c \in \mathbb{R}, x \in \mathbb{R}$
- $f(x) = -\ln x, \ x \in \mathbb{R}_{++}$
- $f(x) = x \ln x, \ x \in \mathbb{R}_{++}$
- The sum of the largest k coordinates  $f(x) = x_{(1)} + \ldots + x_{(k)}, \ x \in \mathbb{R}^n$
- $f(X) = \lambda_{max}(X), X = X^T$
- $f(X) = -\log \det X, X \in S_{++}^n$

 $f \to \min_{x,y,z}$  Convex functions

#### **Epigraph**

For the function f(x), defined on  $S \subseteq \mathbb{R}^n$ , the following set:

epi 
$$f = \{[x,\mu] \in S imes \mathbb{R} : f(x) \leq \mu\}$$

is called **epigraph** of the function f(x).

 $\mathring{\mathbb{I}}$  Convexity of the epigraph is the convexity of the function

For a function f(x), defined on a convex set X, to be convex on X, it is necessary and sufficient that the epigraph of f is a convex set.

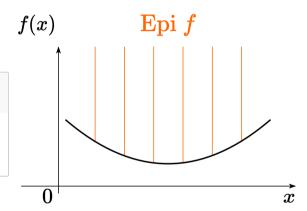


Figure 10: Epigraph of a function

Convex functions

### Convexity of the epigraph is the convexity of the function

1. **Necessity**: Assume f(x) is convex on X. Take any two arbitrary points  $[x_1, \mu_1] \in \text{epi} f$  and  $[x_2, \mu_2] \in \text{epi} f$ . Also take  $0 \le \lambda \le 1$  and denote  $x_{\lambda} = \lambda x_1 + (1 - \lambda)x_2$ ,  $\mu_{\lambda} = \lambda \mu_1 + (1 - \lambda)\mu_2$ . Then,

$$\lambda \begin{bmatrix} x_1 \\ \mu_1 \end{bmatrix} + (1 - \lambda) \begin{bmatrix} x_2 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} x_\lambda \\ \mu_\lambda \end{bmatrix}.$$

From the convexity of the set X, it follows that  $x_{\lambda} \in X$ . Moreover, since f(x) is a convex function,

$$f(x_{\lambda}) \le \lambda f(x_1) + (1 - \lambda)f(x_2) \le \lambda \mu_1 + (1 - \lambda)\mu_2 = \mu_{\lambda}$$

Inequality above indicates that  $\begin{bmatrix} x_{\lambda} \\ u_{\lambda} \end{bmatrix} \in \mathrm{epi}f.$  Thus, the epigraph of f is a convex set.

#### Convexity of the epigraph is the convexity of the function

1. **Necessity**: Assume f(x) is convex on X. Take any two arbitrary points  $[x_1, \mu_1] \in \text{epi} f$  and  $[x_2, \mu_2] \in \text{epi} f$ .

Also take  $0 \le \lambda \le 1$  and denote  $x_{\lambda} = \lambda x_1 + (1 - \lambda)x_2$ ,  $\mu_{\lambda} = \lambda \mu_1 + (1 - \lambda)\mu_2$ . Then,

$$\lambda \begin{bmatrix} x_1 \\ \mu_1 \end{bmatrix} + (1 - \lambda) \begin{bmatrix} x_2 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} x_\lambda \\ \mu_\lambda \end{bmatrix}.$$

From the convexity of the set X, it follows that  $x_{\lambda} \in X$ . Moreover, since f(x) is a convex function,

$$f(x_1) < \lambda f(x_1) + (1 - \lambda) f(x_2) < \lambda \mu_1 + (1 - \lambda) \mu_2 = \mu_1$$

Inequality above indicates that  $\begin{bmatrix} x_\lambda \\ \mu_\lambda \end{bmatrix} \in \mathrm{epi} f.$  Thus, the epigraph of f is a convex set.

2. **Sufficiency**: Assume the epigraph of f, epif, is a convex set. Then, from the membership of the points  $[x_1, \mu_1]$ and  $[x_2, \mu_2]$  in the epigraph of f, it follows that

$$egin{bmatrix} x_\lambda \ \mu_\lambda \end{bmatrix} = \lambda egin{bmatrix} x_1 \ \mu_1 \end{bmatrix} + (1-\lambda) egin{bmatrix} x_2 \ \mu_2 \end{bmatrix} \in \operatorname{\sf epi} f$$

particularly when  $\mu_1=f(x_1)$  and  $\mu_2=f(x_2)$ . Hence we arrive at the inequality

for any  $0 \le \lambda \le 1$ , i.e.,  $f(x_\lambda) \le \mu_\lambda = \lambda \mu_1 + (1 - \lambda)\mu_2$ . But this is true for all  $\mu_1 \ge f(x_1)$  and  $\mu_2 \ge f(x_2)$ ,

## Sublevel set

For the function f(x), defined on  $S \subseteq \mathbb{R}^n$ , the following set:

$$\mathcal{L}_{eta}=\{x\in S: f(x)\leq eta\}$$
 s called **sublevel set** or Lebesgue set of the function

is called **sublevel set** or Lebesgue set of the function f(x).

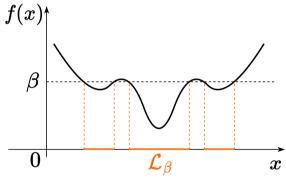


Figure 11: Sublevel set of a function with respect to level  $\beta$ 

#### Connection with epigraph

The function is convex if and only if its epigraph is a convex set.

#### i Example

Let a norm  $\|\cdot\|$  be defined in the space U. Consider the set:

$$K := \{(x, t) \in U \times \mathbb{R}^+ : ||x|| \le t\}$$

which represents the epigraph of the function  $x\mapsto \|x\|$ . This set is called the cone norm. According to the statement above, the set K is convex.

In the case where  $U=\mathbb{R}^n$  and  $||x||=||x||_2$  (Euclidean norm), the abstract set K transitions into the set:

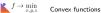
$$\{(x,t) \in \mathbb{R}^n \times \mathbb{R}^+ : ||x||_2 \le t\}$$



#### Connection with sublevel set

If f(x) - is a convex function defined on the convex set  $S \subseteq \mathbb{R}^n$ , then for any  $\beta$  sublevel set  $\mathcal{L}_{\beta}$  is convex.

The function f(x) defined on the convex set  $S \subseteq \mathbb{R}^n$  is closed if and only if for any  $\beta$  sublevel set  $\mathcal{L}_{\beta}$  is closed.



#### Reduction to a line

 $f:S\to\mathbb{R}$  is convex if and only if S is a convex set and the function g(t)=f(x+tv) defined on  $\{t\mid x+tv\in S\}$  is convex for any  $x\in S,v\in\mathbb{R}^n$ , which allows checking convexity of the scalar function to establish convexity of the vector function.

Convex functions

## Strong convexity criteria





## First-order differential criterion of convexity

The differentiable function f(x) defined on the convex set

$$S \subseteq \mathbb{R}^n$$
 is convex if and only if  $\forall x,y \in S$ :

$$f(y) \ge f(x) + \nabla f^{T}(x)(y - x)$$

Let  $y=x+\Delta x$ , then the criterion will become more tractable:

$$f(x + \Delta x) \ge f(x) + \nabla f^{T}(x) \Delta x$$

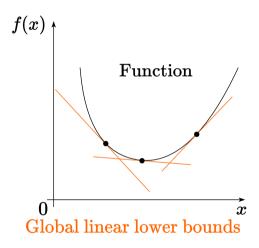


Figure 12: Convex function is greater or equal than Taylor linear approximation at any point

⊕ ი დ

## Second-order differential criterion of convexity

Twice differentiable function f(x) defined on the convex set  $S \subseteq \mathbb{R}^n$  is convex if and only if  $\forall x \in \mathbf{int}(S) \neq \emptyset$ :

$$\nabla^2 f(x) \succeq 0$$

In other words,  $\forall y \in \mathbb{R}^n$ :

Strong convexity criteria

$$\langle y, \nabla^2 f(x)y \rangle \ge 0$$

• Definition (Jensen's inequality)

- Definition (Jensen's inequality)
- Differential criteria of convexity

- Definition (Jensen's inequality)
- Differential criteria of convexity
- Operations, that preserve convexity



- Definition (Jensen's inequality)
- Differential criteria of convexity
- Operations, that preserve convexity
- Connection with epigraph

The function is convex if and only if its epigraph is a convex set.

- Definition (Jensen's inequality)
- Differential criteria of convexity
- Operations, that preserve convexity
- Connection with epigraph

The function is convex if and only if its epigraph is a convex set.

Connection with sublevel set

If f(x) - is a convex function defined on the convex set  $S \subseteq \mathbb{R}^n$ , then for any  $\beta$  sublevel set  $\mathcal{L}_{\beta}$  is convex.

The function f(x) defined on the convex set  $S \subseteq \mathbb{R}^n$  is closed if and only if for any  $\beta$  sublevel set  $\mathcal{L}_{\beta}$  is closed.

എ റ ഉ

- Definition (Jensen's inequality)
- Differential criteria of convexity
- Operations, that preserve convexity
- Connection with epigraph

The function is convex if and only if its epigraph is a convex set.

Connection with sublevel set

If f(x) - is a convex function defined on the convex set  $S \subseteq \mathbb{R}^n$ , then for any  $\beta$  sublevel set  $\mathcal{L}_{\beta}$  is convex.

The function f(x) defined on the convex set  $S \subseteq \mathbb{R}^n$  is closed if and only if for any  $\beta$  sublevel set  $\mathcal{L}_{\beta}$  is closed.

Reduction to a line

 $f:S \to \mathbb{R}$  is convex if and only if S is a convex set and the function g(t)=f(x+tv) defined on  $\{t\mid x+tv\in S\}$  is convex for any  $x\in S, v\in \mathbb{R}^n$ , which allows checking convexity of the scalar function to establish convexity of the vector function.

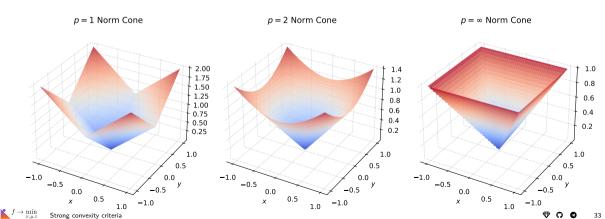


#### **Example:** norm cone

Let a norm  $\|\cdot\|$  be defined in the space U. Consider the set:

$$K := \{(x, t) \in U \times \mathbb{R}^+ : ||x|| \le t\}$$

which represents the epigraph of the function  $x\mapsto \|x\|$ . This set is called the cone norm. According to the statement above, the set K is convex.  $\P$ Code for the figures



# Strong convexity

f(x), defined on the convex set  $S \subseteq \mathbb{R}^n$ , is called  $\mu$ -strongly

 $f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda)f(x_2) - \frac{\mu}{2}\lambda(1-\lambda)\|x_1 - x_2\|^2$ 

for any  $x_1, x_2 \in S$  and  $0 \le \lambda \le 1$  for some  $\mu > 0$ .

convex (strongly convex) on S, if:

f(x)Function

Global quadratic lower bounds

Figure 14: Strongly convex function is greater or equal than Taylor quadratic approximation at any point

# First-order differential criterion of strong convexity

Differentiable f(x) defined on the convex set  $S \subseteq \mathbb{R}^n$  is  $\mu$ -strongly convex if and only if  $\forall x, y \in S$ :

$$f(y) \ge f(x) + \nabla f^{T}(x)(y - x) + \frac{\mu}{2} ||y - x||^{2}$$

# First-order differential criterion of strong convexity

Differentiable f(x) defined on the convex set  $S \subseteq \mathbb{R}^n$  is  $\mu$ -strongly convex if and only if  $\forall x, y \in S$ :

$$f(y) \ge f(x) + \nabla f^{T}(x)(y - x) + \frac{\mu}{2} ||y - x||^{2}$$

Let  $y = x + \Delta x$ , then the criterion will become more tractable:

$$f(x + \Delta x) \ge f(x) + \nabla f^{T}(x)\Delta x + \frac{\mu}{2} ||\Delta x||^{2}$$



# First-order differential criterion of strong convexity

Differentiable f(x) defined on the convex set  $S \subseteq \mathbb{R}^n$  is  $\mu$ -strongly convex if and only if  $\forall x, y \in S$ :

$$f(y) \ge f(x) + \nabla f^{T}(x)(y - x) + \frac{\mu}{2} ||y - x||^{2}$$

Let  $y = x + \Delta x$ , then the criterion will become more tractable:

$$f(x + \Delta x) \ge f(x) + \nabla f^{T}(x)\Delta x + \frac{\mu}{2} ||\Delta x||^{2}$$

#### i Theorem

Let f(x) be a differentiable function on a convex set  $X \subseteq \mathbb{R}^n$ . Then f(x) is strongly convex on X with a constant  $\mu > 0$  if and only if

$$f(x) - f(x_0) \ge \langle \nabla f(x_0), x - x_0 \rangle + \frac{\mu}{2} ||x - x_0||^2$$

for all  $x, x_0 \in X$ .

## Proof of first-order differential criterion of strong convexity

**Necessity**: Let  $0 < \lambda \le 1$ . According to the definition of a strongly convex function,

$$f(\lambda x + (1 - \lambda)x_0) \le \lambda f(x) + (1 - \lambda)f(x_0) - \frac{\mu}{2}\lambda(1 - \lambda)\|x - x_0\|^2$$

or equivalently,

$$f(x) - f(x_0) - \frac{\mu}{2} (1 - \lambda) ||x - x_0||^2 \ge \frac{1}{\lambda} [f(\lambda x + (1 - \lambda)x_0) - f(x_0)] =$$

$$= \frac{1}{\lambda} [f(x_0 + \lambda(x - x_0)) - f(x_0)] = \frac{1}{\lambda} [\lambda \langle \nabla f(x_0), x - x_0 \rangle + o(\lambda)] =$$

$$= \langle \nabla f(x_0), x - x_0 \rangle + \frac{o(\lambda)}{\lambda}.$$

Thus, taking the limit as  $\lambda \downarrow 0$ , we arrive at the initial statement.

 $f \to \min_{x,y,z}$  Strong convexity criteria

#### Proof of first-order differential criterion of strong convexity **Sufficiency**: Assume the inequality in the theorem is satisfied for all $x, x_0 \in X$ . Take $x_0 = \lambda x_1 + (1 - \lambda)x_2$ , where

 $x_1, x_2 \in X$ ,  $0 \le \lambda \le 1$ . According to the inequality, the following inequalities hold:

$$f(x_1) - f(x_0) \ge \langle \nabla f(x_0), x_1 - x_0 \rangle + \frac{\mu}{2} ||x_1 - x_0||^2,$$

 $f(x_2) - f(x_0) \ge \langle \nabla f(x_0), x_2 - x_0 \rangle + \frac{\mu}{2} ||x_2 - x_0||^2.$ 

Multiplying the first inequality by  $\lambda$  and the second by  $1-\lambda$  and adding them, considering that

$$x_1 - x_0 = (1 -$$

 $x_1 - x_0 = (1 - \lambda)(x_1 - x_2), \quad x_2 - x_0 = \lambda(x_2 - x_1),$ 

and 
$$\lambda(1-\lambda)^2+\lambda^2(1-\lambda)=\lambda(1-\lambda)$$
, we get

 $f \to \min_{x,y,z}$  Strong convexity criteria

$$\lambda f(x_1) + (1 - \lambda)f(x_2) - f(x_0) - \frac{\mu}{2}\lambda(1 - \lambda)||x_1 - x_2||^2 \ge$$

$$\langle \nabla f(x_0), \lambda x_1 + (1 - \lambda)x_2 - x_0 \rangle = 0.$$

Thus, inequality from the definition of a strongly convex function is satisfied. It is important to mention, that  $\mu=0$ stands for the convex case and corresponding differential criterion.

## Second-order differential criterion of strong convexity

Twice differentiable function f(x) defined on the convex set  $S \subseteq \mathbb{R}^n$  is called  $\mu$ -strongly convex if and only if  $\forall x \in \mathbf{int}(S) \neq \emptyset$ :

$$\nabla^2 f(x) \succeq \mu I$$

In other words:

$$\langle y, \nabla^2 f(x)y \rangle \ge \mu \|y\|^2$$



## Second-order differential criterion of strong convexity

Twice differentiable function f(x) defined on the convex set  $S \subseteq \mathbb{R}^n$  is called  $\mu$ -strongly convex if and only if  $\forall x \in \mathbf{int}(S) \neq \emptyset$ :

$$\nabla^2 f(x) \succeq \mu I$$

In other words:

$$\langle y, \nabla^2 f(x)y \rangle \ge \mu \|y\|^2$$

#### i Theorem

Let  $X\subseteq\mathbb{R}^n$  be a convex set, with  $\mathrm{int}X\neq\emptyset$ . Furthermore, let f(x) be a twice continuously differentiable function on X. Then f(x) is strongly convex on X with a constant  $\mu>0$  if and only if

$$\langle y, \nabla^2 f(x)y \rangle \ge \mu \|y\|^2$$

for all  $x \in X$  and  $y \in \mathbb{R}^n$ .

## Proof of second-order differential criterion of strong convexity

The target inequality is trivial when  $y = \mathbf{0}_n$ , hence we assume  $y \neq \mathbf{0}_n$ .

**Necessity**: Assume initially that x is an interior point of X. Then  $x + \alpha y \in X$  for all  $y \in \mathbb{R}^n$  and sufficiently small  $\alpha$ . Since f(x) is twice differentiable,

$$f(x + \alpha y) = f(x) + \alpha \langle \nabla f(x), y \rangle + \frac{\alpha^2}{2} \langle y, \nabla^2 f(x) y \rangle + o(\alpha^2).$$

Based on the first order criterion of strong convexity, we have

$$\frac{\alpha^2}{2}\langle y, \nabla^2 f(x)y \rangle + o(\alpha^2) = f(x + \alpha y) - f(x) - \alpha \langle \nabla f(x), y \rangle \ge \frac{\mu}{2} \alpha^2 ||y||^2.$$

This inequality reduces to the target inequality after dividing both sides by  $\alpha^2$  and taking the limit as  $\alpha \downarrow 0$ .

If  $x \in X$  but  $x \notin \text{int} X$ , consider a sequence  $\{x_k\}$  such that  $x_k \in \text{int} X$  and  $x_k \to x$  as  $k \to \infty$ . Then, we arrive at the target inequality after taking the limit.

#### Proof of second-order differential criterion of strong convexity

**Sufficiency**: Using Taylor's formula with the Lagrange remainder and the target inequality, we obtain for  $x + y \in X$ :

$$f(x+y) - f(x) - \langle \nabla f(x), y \rangle = \frac{1}{2} \langle y, \nabla^2 f(x+\alpha y)y \rangle \ge \frac{\mu}{2} ||y||^2,$$

where  $0 \le \alpha \le 1$ . Therefore,

$$f(x+y) - f(x) \ge \langle \nabla f(x), y \rangle + \frac{\mu}{2} ||y||^2.$$

Consequently, by the first order criterion of strong convexity, the function f(x) is strongly convex with a constant  $\mu$ . It is important to mention, that  $\mu = 0$  stands for the convex case and corresponding differential criterion.

 $f \to \min_{x,y,z}$  Strong convexity criteria

#### **Convex and concave function**

Show, that  $f(x) = c^{\top}x + b$  is convex and concave.





### Simplest strongly convex function

Show, that  $f(x) = x^{\top} A x$ , where  $A \succeq 0$  - is convex on  $\mathbb{R}^n$ . Is it strongly convex?





#### **Convexity and continuity**

Let f(x) - be a convex function on a convex set  $S \subseteq \mathbb{R}^n$ . Then f(x) is continuous  $\forall x \in ri(S)$ .

#### i Proper convex function

Function  $f: \mathbb{R}^n \to \mathbb{R}$  is said to be **proper convex function** if it never takes on the value  $-\infty$  and not identically equal to  $\infty$ .

#### i Indicator function

$$\delta_S(x) = \begin{cases} \infty, & x \in S, \\ 0, & x \notin S, \end{cases}$$

is a proper convex function.



#### Convexity and continuity

Let f(x) - be a convex function on a convex set  $S \subseteq \mathbb{R}^n$ . Then f(x) is continuous  $\forall x \in \mathbf{ri}(S)$ .

#### i Proper convex function

Function  $f:\mathbb{R}^n\to\mathbb{R}$  is said to be **proper convex** function if it never takes on the value  $-\infty$  and not identically equal to  $\infty$ .

#### i Closed function

Function  $f:\mathbb{R}^n \to \mathbb{R}$  is said to be **closed** if for each  $\alpha \in \mathbb{R}$ , the sublevel set is a closed set. Equivalently, if the epigraph is closed, then the function f is closed.

#### i Indicator function

$$\delta_S(x) = \begin{cases} \infty, & x \in S, \\ 0, & x \notin S, \end{cases}$$
 is a proper convex function.

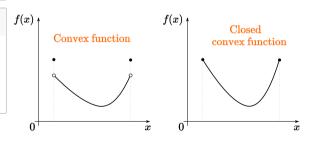


Figure 15: The concept of a closed function is introduced to avoid such breaches at the border.

#### Facts about convexity

• f(x) is called (strictly, strongly) concave, if the function -f(x) - is (strictly, strongly) convex.

#### Facts about convexity

- f(x) is called (strictly, strongly) concave, if the function -f(x) is (strictly, strongly) convex.
- Jensen's inequality for the convex functions:

$$f\left(\sum_{i=1}^{n} \alpha_i x_i\right) \le \sum_{i=1}^{n} \alpha_i f(x_i)$$

for  $\alpha_i \geq 0$ ;  $\sum_{i=1}^n \alpha_i = 1$  (probability simplex)

For the infinite dimension case:

$$f\left(\int\limits_{S} xp(x)dx\right) \le \int\limits_{S} f(x)p(x)dx$$

If the integrals exist and  $p(x) \ge 0$ ,  $\int p(x)dx = 1$ .

 $f \to \min_{x,y,z}$  Strong convexity criteria

#### Facts about convexity

- f(x) is called (strictly, strongly) concave, if the function -f(x) is (strictly, strongly) convex.
- Jensen's inequality for the convex functions:

$$f\left(\sum_{i=1}^{n} \alpha_i x_i\right) \le \sum_{i=1}^{n} \alpha_i f(x_i)$$

for  $\alpha_i \geq 0$ ;  $\sum_{i=1}^n \alpha_i = 1$  (probability simplex)

For the infinite dimension case:

$$f\left(\int\limits_{S} xp(x)dx\right) \le \int\limits_{S} f(x)p(x)dx$$

If the integrals exist and  $p(x) \geq 0$ ,  $\int p(x)dx = 1$ .

• If the function f(x) and the set S are convex, then any local minimum  $x^* = \arg\min_{x \in S} f(x)$  will be the global one. Strong convexity guarantees the uniqueness of the solution.

### Operations that preserve convexity • Non-negative sum of the convex functions:

$$\alpha f(x) + \beta g(x), (\alpha \ge 0, \beta \ge 0).$$

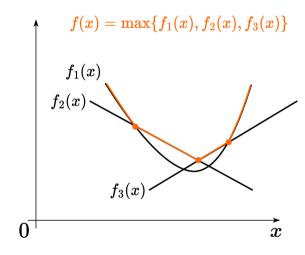


Figure 16: Pointwise maximum (supremum) of convex functions is convex

### Operations that preserve convexity Non-negative sum of the convex functions:

- Non-negative sum of the convex functions:  $\alpha f(x) + \beta g(x), (\alpha \ge 0, \beta \ge 0).$
- Composition with affine function f(Ax+b) is convex, if f(x) is convex.

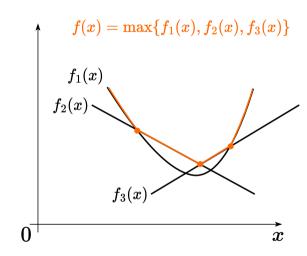


Figure 16: Pointwise maximum (supremum) of convex functions is convex

- Non-negative sum of the convex functions:  $\alpha f(x) + \beta g(x), (\alpha > 0, \beta > 0).$
- Composition with affine function f(Ax + b) is convex, if f(x) is convex.
- Pointwise maximum (supremum) of any number of functions: If  $f_1(x),\ldots,f_m(x)$  are convex, then  $f(x)=\max\{f_1(x),\ldots,f_m(x)\}$  is convex.

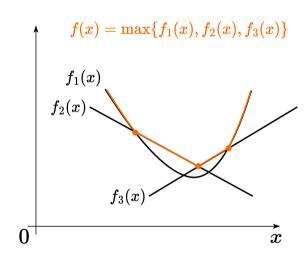


Figure 16: Pointwise maximum (supremum) of convex functions is convex



- Non-negative sum of the convex functions:  $\alpha f(x) + \beta g(x), (\alpha > 0, \beta > 0).$
- Composition with affine function f(Ax + b) is convex, if f(x) is convex.
- Pointwise maximum (supremum) of any number of functions: If  $f_1(x), \ldots, f_m(x)$  are convex, then  $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$  is convex.
- If f(x,y) is convex on x for any  $y \in Y$ :  $g(x) = \sup_{y \in Y} f(x,y)$  is convex.

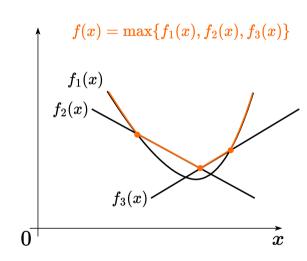


Figure 16: Pointwise maximum (supremum) of convex functions is convex

- Non-negative sum of the convex functions:  $\alpha f(x) + \beta q(x), (\alpha > 0, \beta > 0).$
- Composition with affine function f(Ax + b) is convex, if f(x) is convex.
- Pointwise maximum (supremum) of any number of functions: If  $f_1(x), \ldots, f_m(x)$  are convex, then  $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$  is convex.
- If f(x,y) is convex on x for any  $y \in Y$ :  $g(x) = \sup_{y \in Y} f(x,y)$  is convex.
- If f(x) is convex on S, then g(x,t)=tf(x/t) is convex with  $x/t\in S, t>0$ .

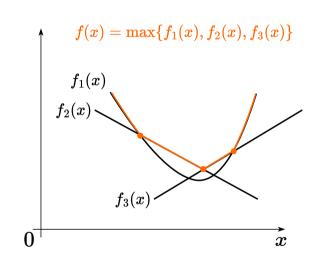


Figure 16: Pointwise maximum (supremum) of convex functions is convex

- Non-negative sum of the convex functions:  $\alpha f(x) + \beta g(x), (\alpha > 0, \beta > 0).$
- Composition with affine function f(Ax + b) is convex, if f(x) is convex.
- Pointwise maximum (supremum) of any number of functions: If  $f_1(x), \ldots, f_m(x)$  are convex, then  $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$  is convex.
- If f(x,y) is convex on x for any  $y \in Y$ :  $g(x) = \sup_{x \in Y} f(x,y)$  is convex.
- If f(x) is convex on S, then g(x,t)=tf(x/t) is convex with  $x/t \in S, t>0$ .
- Let  $f_1: S_1 \to \mathbb{R}$  and  $f_2: S_2 \to \mathbb{R}$ , where range $(f_1) \subseteq S_2$ . If  $f_1$  and  $f_2$  are convex, and  $f_2$  is increasing, then  $f_2 \circ f_1$  is convex on  $S_1$ .

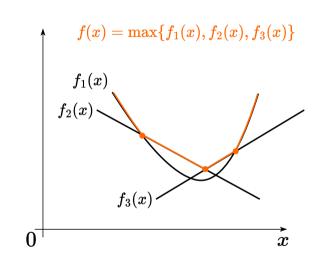


Figure 16: Pointwise maximum (supremum) of convex functions is convex

### Maximum eigenvalue of a matrix is a convex function

i Example

Show, that  $f(A) = \lambda_{max}(A)$  - is convex, if  $A \in S^n$ .



• Log-convexity:  $\log f$  is convex; Log convexity implies convexity.



- Log-convexity:  $\log f$  is convex; Log convexity implies convexity.
- Log-concavity:  $\log f$  concave; **not** closed under addition!



- ullet Log-convexity:  $\log f$  is convex; Log convexity implies convexity.
- $\bullet$  Log-concavity:  $\log f$  concave;  $\mathbf{not}$  closed under addition!
- Exponential convexity:  $[f(x_i+x_j)]\succeq 0$ , for  $x_1,\ldots,x_n$

େ ଚ

- ullet Log-convexity:  $\log f$  is convex; Log convexity implies convexity.
- $\bullet$  Log-concavity:  $\log f$  concave;  $\mathbf{not}$  closed under addition!
- Exponential convexity:  $[f(x_i + x_j)] \succeq 0$ , for  $x_1, \ldots, x_n$
- Operator convexity:  $f(\lambda X + (1 \lambda)Y)$

େ ଚ

- Log-convexity:  $\log f$  is convex; Log convexity implies convexity.
- $\bullet$  Log-concavity:  $\log f$  concave; not closed under addition!
- Exponential convexity:  $[f(x_i + x_j)] \succeq 0$ , for  $x_1, \ldots, x_n$
- Operator convexity:  $f(\lambda X + (1 \lambda)Y)$
- Quasiconvexity:  $f(\lambda x + (1 \lambda)y) \le \max\{f(x), f(y)\}$





- Log-convexity:  $\log f$  is convex; Log convexity implies convexity.
- ullet Log-concavity:  $\log f$  concave; **not** closed under addition!
- Exponential convexity:  $[f(x_i + x_j)] \succeq 0$ , for  $x_1, \ldots, x_n$
- Operator convexity:  $f(\lambda X + (1 \lambda)Y)$
- Quasiconvexity:  $f(\lambda x + (1 \lambda)y) \le \max\{f(x), f(y)\}$
- Pseudoconvexity:  $\langle \nabla f(y), x y \rangle \geq 0 \longrightarrow f(x) \geq f(y)$



- ullet Log-convexity:  $\log f$  is convex; Log convexity implies convexity.
- ullet Log-concavity:  $\log f$  concave; **not** closed under addition!
- Exponential convexity:  $[f(x_i + x_j)] \succeq 0$ , for  $x_1, \ldots, x_n$
- Operator convexity:  $f(\lambda X + (1 \lambda)Y)$
- Quasiconvexity:  $f(\lambda x + (1 \lambda)y) \le \max\{f(x), f(y)\}$
- Pseudoconvexity:  $\langle \nabla f(y), x y \rangle \ge 0 \longrightarrow f(x) \ge f(y)$
- Discrete convexity:  $f: \mathbb{Z}^n \to \mathbb{Z}$ ; "convexity + matroid theory."



**♥** ೧ **0** 

# Polyak- Lojasiewicz condition. Linear convergence of gradient descent without convexity

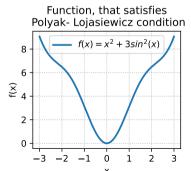
PL inequality holds if the following condition is satisfied for some  $\mu>0$ ,

$$\|\nabla f(x)\|^2 \ge 2\mu(f(x) - f^*) \forall x$$

It is interesting, that Gradient Descent converges linearly under this condition (weaker, then strong convexity).

The following functions satisfy the PL-condition, but are not convex. **\PL**ink to the code

$$f(x) = x^2 + 3\sin^2(x)$$



# Polyak- Lojasiewicz condition. Linear convergence of gradient descent without convexity

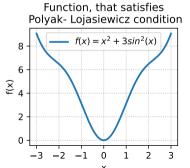
PL inequality holds if the following condition is satisfied for some  $\mu > 0$ ,

$$\|\nabla f(x)\|^2 \ge 2\mu(f(x) - f^*) \forall x$$

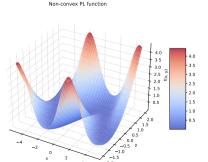
It is interesting, that Gradient Descent converges linearly under this condition (weaker, then strong convexity).

The following functions satisfy the PL-condition, but are not convex. **\display**Link to the code

$$f(x) = x^2 + 3\sin^2(x)$$



$$f(x,y) = \frac{(y - \sin x)^2}{2}$$



#### Convexity in ML



Convexity in ML



#### **Linear Least Squares aka Linear Regression**

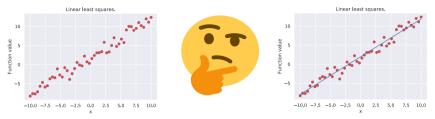


Figure 19: Illustration

In a least-squares, or linear regression, problem, we have measurements  $X \in \mathbb{R}^{m \times n}$  and  $y \in \mathbb{R}^m$  and seek a vector  $\theta \in \mathbb{R}^n$  such that  $X\theta$  is close to y. Closeness is defined as the sum of the squared differences:

$$\sum_{i=1}^{m} (x_i^{\top} \theta - y_i)^2 = ||X\theta - y||_2^2 \to \min_{\theta \in \mathbb{R}^n}$$

For example, we might have a dataset of m users, each represented by n features. Each row  $x_*^\top$  of X is the features for user i, while the corresponding entry  $y_i$  of y is the measurement we want to predict from  $x_i^{\top}$ , such as ad spending. The prediction is given by  $x_i^{\top} \theta$ .

#### **Linear Least Squares aka Linear Regression** <sup>2</sup>

1. Is this problem convex? Strongly convex?

#### **Linear Least Squares aka Linear Regression** <sup>2</sup>

1. Is this problem convex? Strongly convex?

Convexity in ML

2. What do you think about convergence of Gradient Descent for this problem?

<sup>&</sup>lt;sup>2</sup>Take a look at the **@**example of real-world data linear least squares problem

#### *l*<sub>2</sub>-regularized Linear Least Squares

In the underdetermined case, it is often desirable to restore strong convexity of the objective function by adding an  $l_2$ -penality, also known as Tikhonov regularization,  $l_2$ -regularization, or weight decay.

$$||X\theta - y||_2^2 + \frac{\mu}{2}||\theta||_2^2 \to \min_{\theta \in \mathbb{R}^n}$$

Note: With this modification the objective is  $\mu$ -strongly convex again.

Take a look at the \$\mathbb{e}\text{code}

#### Most important difference between convexity and strong convexity

$$f(x) = \frac{1}{2m} \|Ax - b\|_2^2 + \frac{\mu}{2} \|x\|_2^2 \to \min_{x \in \mathbb{R}^n}, \quad A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$$

Convex least squares regression. m=50. n=100. mu=0.

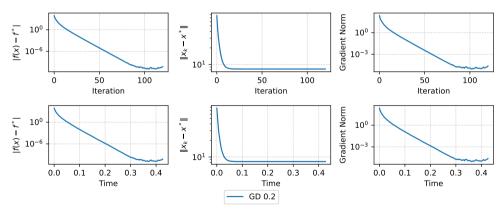


Figure 20: Convex problem does not have convergence in domain

**⊕** ດ **ø** 

#### Most important difference between convexity and strong convexity

$$f(x) = \frac{1}{2m} \|Ax - b\|_2^2 + \frac{\mu}{2} \|x\|_2^2 \to \min_{x \in \mathbb{R}^n}, \quad A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$$

Strongly convex least squares regression.  $m=50. \ n=100. \ mu=0.1.$ 

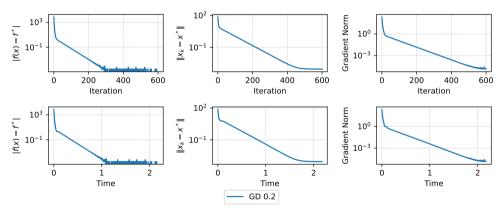


Figure 21: But if you add even small amount of regularization, you will ensure convergence in domain

**⊕ດ** ø

#### Most important difference between convexity and strong convexity

$$f(x) = \frac{1}{2m} \|Ax - b\|_2^2 + \frac{\mu}{2} \|x\|_2^2 \to \min_{x \in \mathbb{R}^n}, \quad A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$$

Strongly convex least squares regression. m=100. n=50. mu=0.

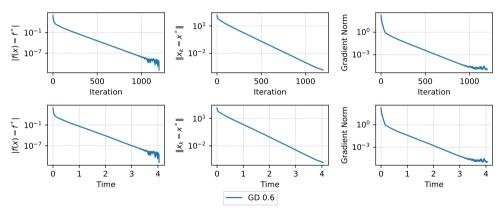


Figure 22: Another way to ensure convergence in the previous problem is to switch the dimension values

**⊕** ດ **ø** 

## You have to have strong convexity (or PL) to ensure convergence with a high precision

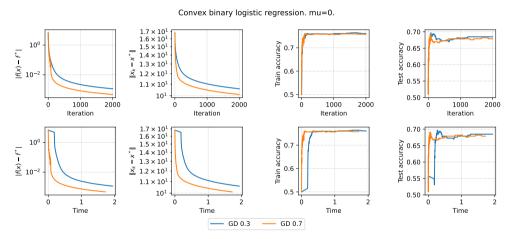


Figure 23: Only small precision is achievable with sublinear convergence

## You have to have strong convexity (or PL) to ensure convergence with a high precision

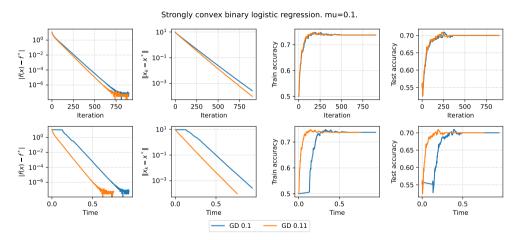


Figure 24: Strong convexity ensures linear convergence

⇔റ⊘

#### **Convex optimization problem**

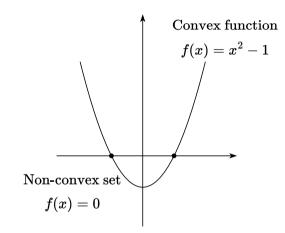


Figure 25: The idea behind the definition of a convex optimization problem

Note, that there is an agreement in notation of mathematical programming. The problems of the following type are called **Convex optimization problem**:

$$f_0(x) o \min_{x \in \mathbb{R}^n}$$
  
s.t.  $f_i(x) \le 0, \ i = 1, \dots, m$  (COP)  
 $Ax = b,$ 

where all the functions  $f_0(x), f_1(x), \ldots, f_m(x)$  are convex and all the equality constraints are affine. It sounds a bit strange, but not all convex problems are convex optimization problems.

$$f_0(x) \to \min_{x \in S},$$
 (CP)

where  $f_0(x)$  is a convex function, defined on the convex set S. The necessity of affine equality constraint is essential.

 $\underset{x,y,z}{\min}$  Convexity in ML