



**Strongly convex functions. Polyak -  
Lojasiewicz Condition. Conjugate sets**

**Daniil Merkulov**

Optimization methods. MIPT

## 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  $\forall x, y \in S$ :

$$f(y) \geq f(x) + \nabla f^T(x)(y - x)$$

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

$$f(x + \Delta x) \geq f(x) + \nabla f^T(x)\Delta x$$



Figure 1: Convex function is greater or equal than Taylor linear approximation at any point

## 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  $\forall x \in \text{int}(S) \neq \emptyset$ :

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

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

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

# Tools for discovering convexity

- Definition (Jensen's inequality)

# Tools for discovering convexity

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

# Tools for discovering convexity

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

# Tools for discovering 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.



# Tools for discovering 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.

- 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.

# Tools for discovering 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.

- **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**

$f : S \rightarrow \mathbb{R}$  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:

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

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.  Code for the figures

$p = 1$  Norm Cone



$p = 2$  Norm Cone



$p = \infty$  Norm Cone



## Strong convexity

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

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

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



Figure 3: 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  $\forall x, y \in S$ :

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

## 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  $\forall x, y \in S$ :

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

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

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

## 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  $\forall x, y \in S$ :

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

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

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

### 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

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

for all  $x, x_0 \in X$ .

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

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

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

or equivalently,

$$\begin{aligned} f(x) - f(x_0) - \frac{\mu}{2}(1 - \lambda)\|x - x_0\|^2 &\geq \frac{1}{\lambda}[f(\lambda x + (1 - \lambda)x_0) - f(x_0)] = \\ &= \frac{1}{\lambda}[f(x_0 + \lambda(x - x_0)) - f(x_0)] = \frac{1}{\lambda}[\lambda\langle \nabla f(x_0), x - x_0 \rangle + o(\lambda)] = \\ &= \langle \nabla f(x_0), x - x_0 \rangle + \frac{o(\lambda)}{\lambda}. \end{aligned}$$

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



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

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

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

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

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

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

$$\begin{aligned} \lambda f(x_1) + (1 - \lambda)f(x_2) - f(x_0) - \frac{\mu}{2} \lambda(1 - \lambda) \|x_1 - x_2\|^2 \geq \\ \langle \nabla f(x_0), \lambda x_1 + (1 - \lambda)x_2 - x_0 \rangle = 0. \end{aligned}$$

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  $\forall x \in \text{int}(S) \neq \emptyset$ :

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

In other words:

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

## 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  $\forall x \in \text{int}(S) \neq \emptyset$ :

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

In other words:

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

### **i** Theorem

Let  $X \subseteq \mathbb{R}^n$  be a convex set, with  $\text{int}X \neq \emptyset$ . 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

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

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

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

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  $x_k \rightarrow x$  as  $k \rightarrow \infty$ . 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  $x + y \in X$ :

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

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

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

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.

# Convex and concave function

## Example

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

# Simplest strongly convex function

## Example

Show, that  $f(x) = x^\top Ax$ , 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  $\forall x \in \text{ri}(S)$ .<sup>a</sup>

### i Proper convex function

Function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  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.

---

<sup>a</sup>Please, read here about difference between interior and relative interior.



# Convexity and continuity

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

## i Proper convex function

Function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  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.

<sup>a</sup>Please, read here about difference between interior and relative interior.

## i Closed function

Function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  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.



Figure 4: 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:

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

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

For the infinite dimension case:

$$f\left(\int_S x p(x) dx\right) \leq \int_S f(x) p(x) dx$$

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

## 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:

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

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

For the infinite dimension case:

$$f\left(\int_S x p(x) dx\right) \leq \int_S f(x) p(x) dx$$

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

- 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:

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

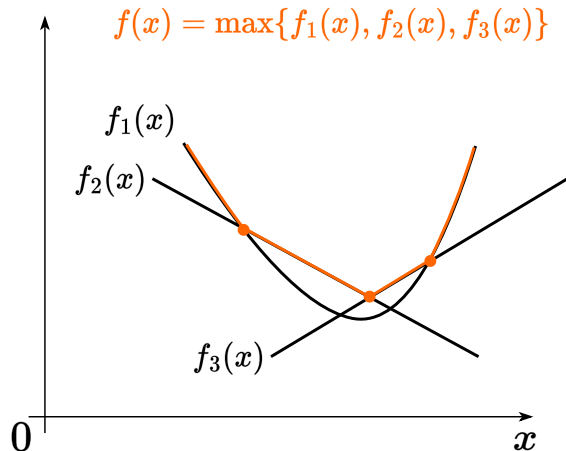


Figure 5: Pointwise maximum (supremum) of convex functions is convex

## Operations that preserve convexity

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



Figure 5: Pointwise maximum (supremum) of convex functions is convex

## Operations that preserve convexity

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

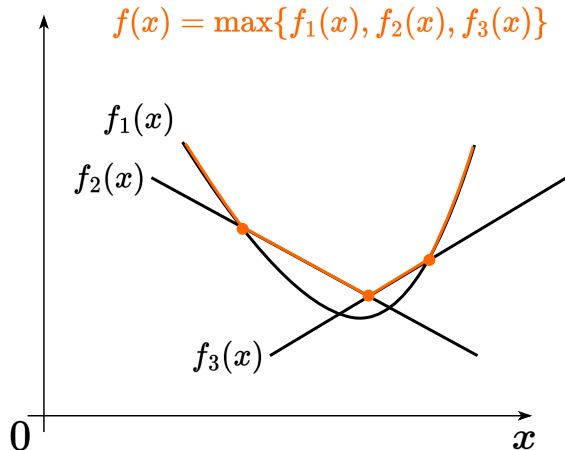


Figure 5: Pointwise maximum (supremum) of convex functions is convex

## Operations that preserve convexity

- Non-negative sum of the convex functions:  
 $\alpha f(x) + \beta g(x)$ ,  $(\alpha \geq 0, \beta \geq 0)$ .
- Composition with affine function  $f(Ax + b)$  is convex, if  $f(x)$  is convex.
- Pointwise maximum (supremum) of any number of functions: If  $f_1(x), \dots, f_m(x)$  are convex, then  $f(x) = \max\{f_1(x), \dots, f_m(x)\}$  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 5: Pointwise maximum (supremum) of convex functions is convex



## Operations that preserve convexity

- Non-negative sum of the convex functions:  
 $\alpha f(x) + \beta g(x)$ ,  $(\alpha \geq 0, \beta \geq 0)$ .
- Composition with affine function  $f(Ax + b)$  is convex, if  $f(x)$  is convex.
- Pointwise maximum (supremum) of any number of functions: If  $f_1(x), \dots, f_m(x)$  are convex, then  $f(x) = \max\{f_1(x), \dots, f_m(x)\}$  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 5: Pointwise maximum (supremum) of convex functions is convex

## Operations that preserve convexity

- Non-negative sum of the convex functions:  
 $\alpha f(x) + \beta g(x), (\alpha \geq 0, \beta \geq 0).$
- Composition with affine function  $f(Ax + b)$  is convex, if  $f(x)$  is convex.
- Pointwise maximum (supremum) of any number of functions: If  $f_1(x), \dots, f_m(x)$  are convex, then  $f(x) = \max\{f_1(x), \dots, f_m(x)\}$  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  $x/t \in S, t > 0$ .
- Let  $f_1 : S_1 \rightarrow \mathbb{R}$  and  $f_2 : S_2 \rightarrow \mathbb{R}$ , where  $\text{range}(f_1) \subseteq S_2$ . 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 5: Pointwise maximum (supremum) of convex functions is convex

# Maximum eigenvalue of a matrix is a convex function

## Example

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

## Other forms of convexity

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

## Other forms of convexity

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

## Other forms of convexity

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

## Other forms of convexity

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

## Other forms of convexity

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



## Other forms of convexity

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

## Other forms of convexity

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

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

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

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

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. [🔗Link 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  $\mu > 0$ ,

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

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. [🔗Link to the code](#)

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



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

Non-convex PL function



## Convexity in ML

# Linear Least Squares aka Linear Regression



Figure 8: 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:

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

For example, we might have a dataset of  $m$  users, each represented by  $n$  features. Each row  $x_i^\top$  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>1</sup>

1. Is this problem convex? Strongly convex?

# Linear Least Squares aka Linear Regression <sup>1</sup>

1. Is this problem convex? Strongly convex?
2. What do you think about convergence of Gradient Descent for this problem?

---

<sup>1</sup>Take a look at the  example of real-world data linear least squares problem



## $l_2$ -regularized Linear Least Squares

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

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

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

Take a look at the code

# Most important difference between convexity and strong convexity

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

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



Figure 9: Convex problem does not have convergence in domain

# Most important difference between convexity and strong convexity

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

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



Figure 10: But if you add even small amount of regularization, you will ensure convergence in domain

# Most important difference between convexity and strong convexity

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

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



Figure 11: Another way to ensure convergence in the previous problem is to switch the dimension values

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

Convex binary logistic regression.  $\mu=0$ .



Figure 12: Only small precision is achievable with sublinear convergence

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

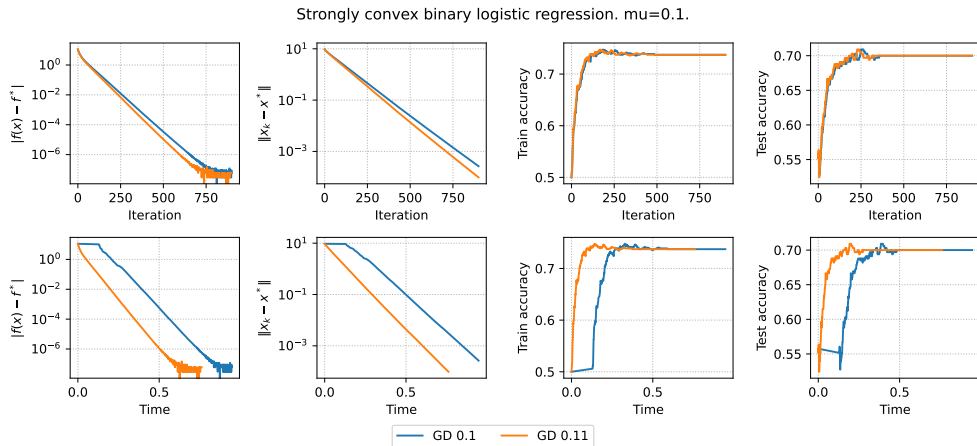


Figure 13: Strong convexity ensures linear convergence

# 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**:

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

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

$$f_0(x) \rightarrow \min_{x \in S}, \quad (\text{CP})$$

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

Figure 14: The idea behind the definition of a convex optimization problem