

# Convexity. Strong convexity.

## Seminar

Optimization for ML. Faculty of Computer Science. HSE University

## Line Segment

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

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



Figure 1: Illustration of a line segment between points  $x_1, x_2$

# 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.

$$\forall \theta \in [0, 1], \forall x_1, x_2 \in S : \theta x_1 + (1 - \theta)x_2 \in S$$

## i Example

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

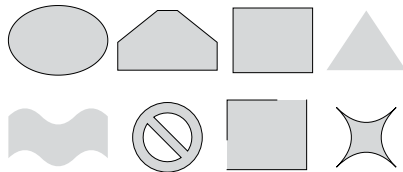


Figure 2: Top: examples of convex sets. Bottom: examples of non-convex sets.

# Problem 1

## Question

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

## Problem 2

### Question

Is stripe -  $\{x \in \mathbb{R}^n \mid \alpha \leq a^\top x \leq \beta\}$  - convex?

## Problem 3

### Question

Let  $S$  be such that  $\forall x, y \in S \rightarrow \frac{1}{2}(x + y) \in S$ . Is this set convex?

## Problem 4

### Question

The set  $S = \{x \mid x + S_2 \subseteq S_1\}$ , where  $S_1, S_2 \subseteq \mathbb{R}^n$  with  $S_1$  convex. Is this set convex?

# Convex Function

The function  $f(x)$ , which is defined on the convex set  $S \subseteq \mathbb{R}^n$ , is called **convex** on  $S$ , if:

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$$

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

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 3: Difference between convex and non-convex function



# 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  $\mu > 0$ .

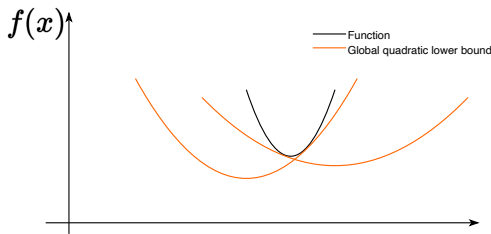


Figure 4: Strongly convex function is greater or equal than Taylor quadratic approximation at any point

## 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:

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



## 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:

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

# Motivational Experiment with JAX

Why convexity and strong convexity is important? Check the simple code snippet.

## Problem 5

### Question

Show, that  $f(x) = \|x\|$  is convex on  $\mathbb{R}^n$ .

### Question

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

## Problem 6

### Question

Show, that if  $f(x)$  is convex on  $\mathbb{R}^n$ , then  $\exp(f(x))$  is convex on  $\mathbb{R}^n$ .

## Problem 7

### Question

If  $f(x)$  is convex nonnegative function and  $p \geq 1$ . Show that  $g(x) = f(x)^p$  is convex.

## Problem 8

### i Question

Show that, if  $f(x)$  is concave positive function over convex  $S$ , then  $g(x) = \frac{1}{f(x)}$  is convex.

### i Question

Show, that the following function is convex on the set of all positive denominators

$$f(x) = \frac{1}{x_1 - \frac{1}{x_2 - \frac{1}{x_3 - \frac{1}{\dots}}}}, x \in \mathbb{R}^n$$



## Problem 9

### Question

Let  $S = \{x \in \mathbb{R}^n \mid x \succ 0, \|x\|_\infty \leq M\}$ . Show that  $f(x) = \sum_{i=1}^n x_i \log x_i$  is  $\frac{1}{M}$ -strongly convex.

# Polyak-Lojasiewicz (PL) Condition

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

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

The example of a function, that satisfies the PL-condition, but is not convex.

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

Example of PL non-convex function  Open in Colab.

# Logistic regression

**i** Given

Data:  $X \in \mathbb{R}^{m \times n}, y \in \{0, 1\}^n$ .

# Logistic regression

**i** Given

Data:  $X \in \mathbb{R}^{m \times n}, y \in \{0, 1\}^n$ .

**!** To find

Find function, that translates object  $x$  to probability  $p(y = 1|x)$ :

$$p : \mathbb{R}^m \rightarrow (0, 1), p(x) \equiv \sigma(x^T w) = \frac{1}{1 + \exp(-x^T w)}$$

# Logistic regression

## Given

Data:  $X \in \mathbb{R}^{m \times n}, y \in \{0, 1\}^n$ .

## To find

Find function, that translates object  $x$  to probability  $p(y = 1|x)$ :  
 $p: \mathbb{R}^m \rightarrow (0, 1), p(x) \equiv \sigma(x^T w) = \frac{1}{1 + \exp(-x^T w)}$

## Criterion

Binary cross-entropy (logistic loss):  
 $L(p, X, y) = - \sum_{i=1}^n y_i \log p(X_i) + (1 - y_i) \log (1 - p(X_i)),$   
that is minimized with respect to  $w$ .

# Logistic regression

## i Given

Data:  $X \in \mathbb{R}^{m \times n}, y \in \{0, 1\}^n$ .

## ! To find

Find function, that translates object  $x$  to probability  $p(y = 1|x)$ :  
 $p: \mathbb{R}^m \rightarrow (0, 1), p(x) \equiv \sigma(x^T w) = \frac{1}{1 + \exp(-x^T w)}$

## 💡 Criterion

Binary cross-entropy (logistic loss):

$$L(p, X, y) = - \sum_{i=1}^n y_i \log p(X_i) + (1 - y_i) \log (1 - p(X_i)),$$

that is minimized with respect to  $w$ .

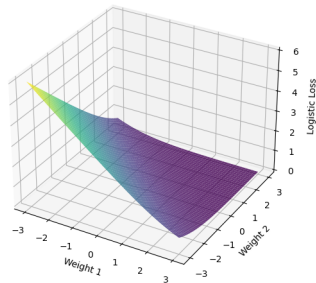


Figure 6: Logistic Loss in Parameter Space for  $x=(1,1), y=1$

We can make this problem  $\mu$ -strongly convex if we consider regularized logistic loss as criterion:  $L(p, X, y) + \frac{\mu}{2} \|w\|_2^2$ .

Check the  logistic regression experiments.

# Support Vector Machine (SVM)

**i** Given

Data:  $X \in \mathbb{R}^{m \times n}, y \in \{-1, 1\}^n$ .

# Support Vector Machine (SVM)

**i** Given

Data:  $X \in \mathbb{R}^{m \times n}, y \in \{-1, 1\}^n$ .

**!** To find

Find a hyperplane that maximizes the margin between two classes:


$$f : \mathbb{R}^m \rightarrow \{-1, 1\}, f(x) = \text{sign}(w^T x + b).$$



# Support Vector Machine (SVM)

 Given

Data:  $X \in \mathbb{R}^{m \times n}$ ,  $y \in \{-1, 1\}^n$ .

 To find

Find a hyperplane that maximizes the margin between two classes:


$$f : \mathbb{R}^m \rightarrow \{-1, 1\}, f(x) = \text{sign}(w^T x + b).$$

 Criterion

Hinge loss:

$L(w, X, y) = \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^n \max(0, 1 - y_i (X_i^T w + b))$ , that is minimized with respect to  $w$  and  $b$ .

This problem is strongly-convex due to squared Euclidean norm.

Check the  SVM experiments in the same notebook.

# Support Vector Machine (SVM)

**i** Given

Data:  $X \in \mathbb{R}^{m \times n}$ ,  $y \in \{-1, 1\}^n$ .

**!** To find

Find a hyperplane that maximizes the margin between two classes:

$$f: \mathbb{R}^m \rightarrow \{-1, 1\}, f(x) = \text{sign}(w^T x + b).$$

**💡** Criterion

Hinge loss:

$L(w, X, y) = \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^n \max(0, 1 - y_i(X_i^T w + b))$ , that is minimized with respect to  $w$  and  $b$ .

This problem is strongly-convex due to squared Euclidean norm.

Check the  SVM experiments in the same notebook.

 Practical examples

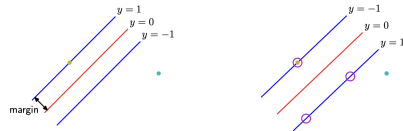


Figure 7: Support Vector Machine



Figure 8:  $L_2$ -Regularized Hinge Loss in Parameter Space for  $x=(1,1)$ ,  $y=1$

## Some other curious examples

- **Low-rank matrix approximation**

$$\min_X \|A - X\|_F^2 \text{ s.t. } \text{rank}(X) \leq k.$$

## Some other curious examples

- **Low-rank matrix approximation**

$$\min_X \|A - X\|_F^2 \text{ s.t. } \text{rank}(X) \leq k.$$

**i** Question

Is it convex?

## Some other curious examples

- **Low-rank matrix approximation**

$$\min_X \|A - X\|_F^2 \text{ s.t. } \text{rank}(X) \leq k.$$

**i** Question

Is it convex?

By Eckart-Young theorem this can be solved using SVD:  $X^* = U_k \Sigma_k V_k^T$ , where  $A = U \Sigma V^T$ .

## Some other curious examples

- **Low-rank matrix approximation**

$$\min_X \|A - X\|_F^2 \text{ s.t. } \text{rank}(X) \leq k.$$

**i** Question

Is it convex?

By Eckart-Young theorem this can be solved using SVD:  $X^* = U_k \Sigma_k V_k^T$ , where  $A = U \Sigma V^T$ .

- **Convex relaxation via nuclear norm**

$$\min_X \text{rank}(X), \text{ s.t. } X_{ij} = M_{ij}, (i, j) \in I.$$

## Some other curious examples

- **Low-rank matrix approximation**

$$\min_X \|A - X\|_F^2 \text{ s.t. } \text{rank}(X) \leq k.$$

**i** Question

Is it convex?

By Eckart-Young theorem this can be solved using SVD:  $X^* = U_k \Sigma_k V_k^T$ , where  $A = U \Sigma V^T$ .

- **Convex relaxation via nuclear norm**

$$\min_X \text{rank}(X), \text{ s.t. } X_{ij} = M_{ij}, (i, j) \in I.$$

NP-hard problem, but  $\|A\|_* = \text{trace}(\sqrt{A^T A}) = \sum_{i=1}^{\text{rank}(A)} \sigma_i(A)$  is a convex envelope of the matrix rank.