CSED490Y: Optimization for Machine Learning

Week 03-1: Convex optimization

Namhoon Lee

POSTECH

Spring 2022

Admin

The course registration period is now over.

For anyone who missed the previous lectures including ✓ newly registered:

- Please make sure you understand the course logistics; read "s01-1.pdf" carefully.
- We have covered the basic materials in the last few weeks. opt. math, ML

Admin

The course registration period is now over.

For anyone who missed the previous lectures including newly registered:

- Please make sure you understand the course logistics; read "s01-1.pdf" carefully.
- We have covered the basic materials in the last few weeks.

A few updates:

- We will switch back to offline classes from Week08 (Monday 11 April).
- The midterm exam will be taken on Week10 (Monday 25 April).

swap w8 and w10

Admin

Start working on the project!

Team up (Due: 11:59PM on Monday 14 March):

- Form a group of up to 3 members.
- Email TA about your group members by the due date.
- You may use the discussion board on PLMS to find your teammates.

Admin

Start working on the project!

Team up (Due: 11:59PM on Monday 14 March):

- Form a group of up to 3 members.
- Email TA about your group members by the due date.
- You may use the discussion board on PLMS to find your teammates.

What topics to work on?

- "Empirical study of *any* optimization and/for machine learning method."
- Examples and references: EPFL CS439, POSTECH CSED499, etc.
- Avoid self-plagiarism: you are not allowed to reuse work that you have already done (*e.g.*, previous research work, project, etc.).

Admin

W12 W9 W4

Example topics (from EPFL CS439):

- Local minima for deep learning: Can you find differences between the ‘shape’ of local minima that SGD finds, depending on different step-sizes or mini-batch size, vs e.g. AdaGrad or full gradient descent?
- Same thing for GANs? Or for matrix factorizations? ML
- Along a training trajectory for a deep net, does the behaviour of each step (update) resemble the convex case, is it different early or late in training?
- How do different optimization variants affect generalization (test error)?
- Second-order methods: Do (Quasi-)Newton methods go to differently shaped local minima in neural networks? Or: Is the secant method a viable alternative training method?
- ...

W8

Admin

Example topics (from POSTECH CSED499; 2021-1, 2021-2):

- Prior robust training on crowdsourcing with neural enhanced belief propagation
- CF-Layer: Boosting performance in noisy federated learning
- Achieving adversarial robustness via network pruning
- Hyperparameter optimization by unsupervised learning
- ...

Admin

Important:

- Make sure to check important dates for project.
- This course does <u>not require any assignments</u> other than project.
- If there is anything you are not sure of about the project, come talk to me.

Convex optimization

Consider an optimization problem:

error, risk, cost, loss

$x \in \mathbb{R}^n$

$$\boxed{\min_{x \in \mathbb{C}} \; f(x)}$$

s.t.

which reads as "minimize a function f subject to x being in the set $\mathbb{C}$".

Convex optimization

Consider an optimization problem:

$$\boxed{\min_{x \in \mathbb{C}} f(x)}$$

which reads as "minimize a function f subject to x being in the set $\mathbb{C}$".

We call the above a *convex* optimization problem if:

- The set $\mathbb{C}$ is a convex set.
- The function f is a convex function.

Convex optimization

Consider an optimization problem:

$$\min_{x \in \mathbb{C}} \; f(x)$$

which reads as “minimize a function f subject to x being in the set $\mathbb{C}$”.

We call the above a *convex* optimization problem if:

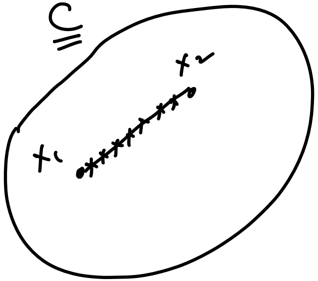
- The set $\mathbb{C}$ is a convex set.
- The function f is a convex function.

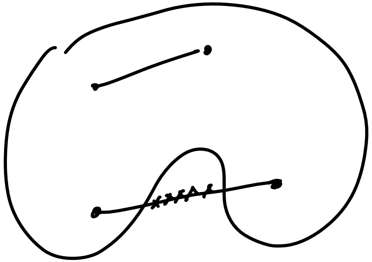
$\circledast$ Key property: “All (local) minima are (global) minima.”

Convex set

x_1 x_2

A set $\mathbb{C}$ is convex if the line segment between any two points in $\mathbb{C}$ also lies in $\mathbb{C}$.



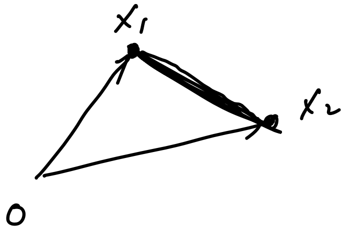


Convex set

Line segment between x_1 and x_2: all points

$$x = \theta x_1 + (1-\theta)x_2$$

with $0 \leq \theta \leq 1$.



$\theta = 0 \rightarrow x = x_2$

$\theta = 1 \rightarrow x = x_1$

$0 \leq \theta \leq 1$

Convex set

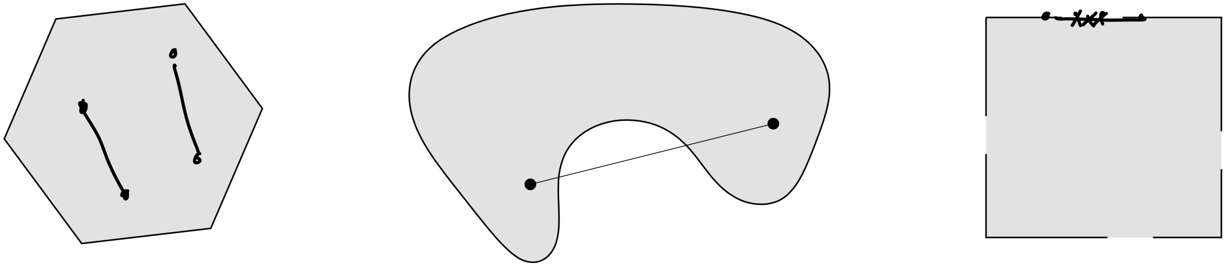
Line segment between x_1 and x_2: all points

$$x = \theta x_1 + (1-\theta) x_2$$

with $0 \leq \theta \leq 1$.

Convex set: a set $\mathbb{C}$ is convex if, for any $x_1, x_2 \in \mathbb{C}$ and any $0 \leq \theta \leq 1$, it contains the line segment between x_1 and x_2 in $\mathbb{C}$

$$* \quad \boxed{\underline{\theta x_1 + (1-\theta) x_2} \in \underline{\mathbb{C}}\ .}$$



C NC NC

Convex set

Convex combination of $x_1, \ldots, x_k$: any point x of the form

$$x = \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_k x_k$$

with $\theta_1 + \theta_2 + \ldots + \theta_k = 1$ and $\theta_i \geq 0$.

$$\sum_i \theta_i = 1$$

Convex set

Convex combination of $x_1, \dots, x_k$: any point x of the form

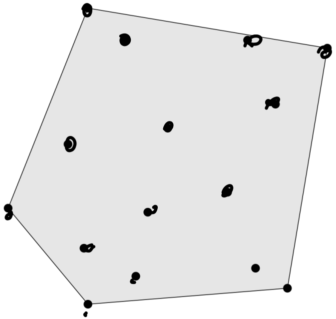
$$x = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_k x_k$$

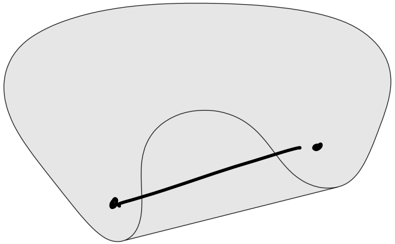
with $\theta_1 + \theta_2 + \dots + \theta_k = 1$ and $\theta_i \geq 0$.

Convex hull of a set $\mathbb{C}$: set of all convex combinations of points in $\mathbb{S}$

$$\{\theta_1 x_1 + \dots + \theta_k x_k \mid x_i \in \mathbb{C},\ \theta_1 + \dots + \theta_k = 1,\ \theta_i \geq 0\}\ .$$

$\{x_1 \cdots x_k\}$





Convex set

Cone: if for every $x \in \mathbb{C}$ and $\theta \geq 0$ we have

$$\theta x \in \mathbb{C} .$$

Convex set

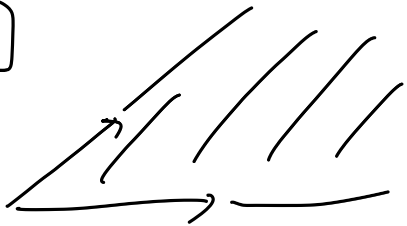
Cone: if for every $x \in \mathbb{C}$ and $\theta \geq 0$ we have

$$\theta x \in \mathbb{C} \ .$$

Conic (nonnegative) combination of x_1 and x_2: any point of the form

$$x = \boxed{\underline{\theta_1 x_1} + \theta_2 x_2}$$

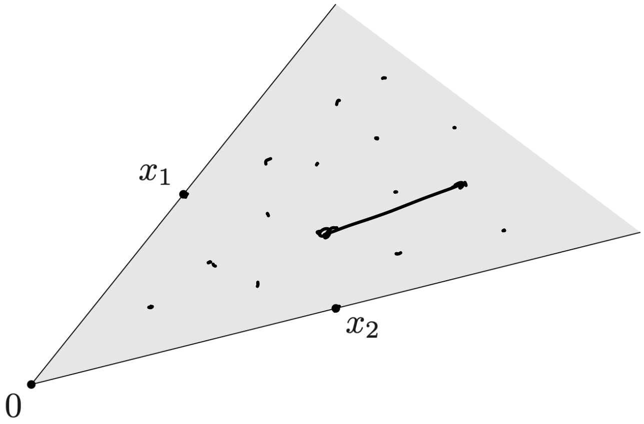
Cone



with $\underline{\theta_1 \geq 0}, \underline{\theta_2 \geq 0}$. $\left(\sum_i \theta_i = 1 \right)$

Convex cone: a set $\mathbb{C}$ is a convex cone if it is convex and a cone; for any $x_1, x_2 \in \mathbb{C}$ and $\theta_1, \theta_2 \geq 0$, we have

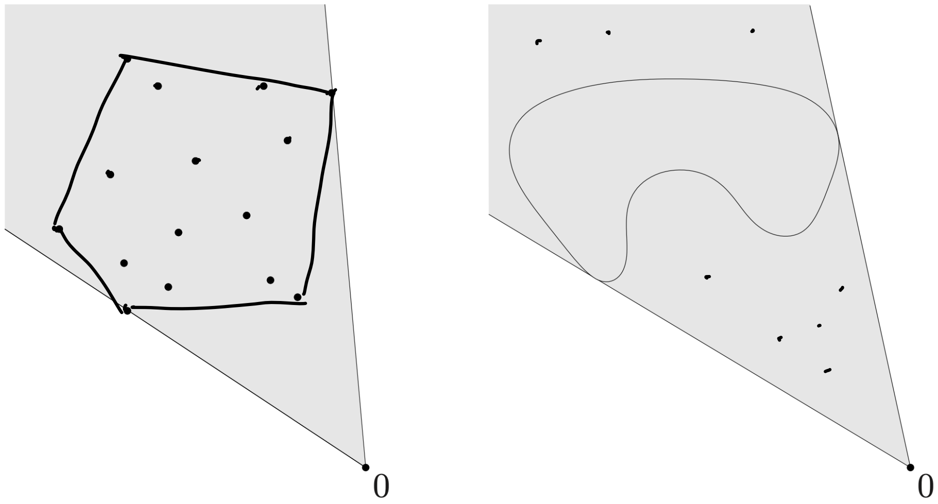
$$x = \boxed{\theta_1 x_1 + \theta_2 x_2} \in \mathbb{C} \; .$$



Conic hull of a set $\mathbb{C}$: the set of all conic combinations of points in $\mathbb{C}$

$$\{\theta_1 x_1 + \ldots + \theta_k x_k \mid x_i \in \mathbb{C},\ \theta_i \geq 0\}\ .$$

$$\left(\sum_i \theta_i = 1\right)$$

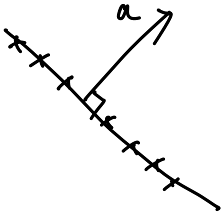


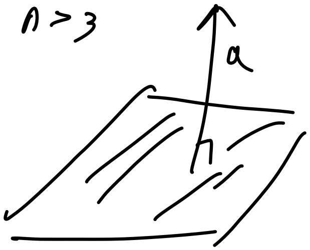
Hyperplane:

$$\mathbb{H} = \{x \in \mathbb{R}^n : a^\top x = b\}$$

where $a \neq 0$ and $b \in \mathbb{R}$.

n = 2 line.

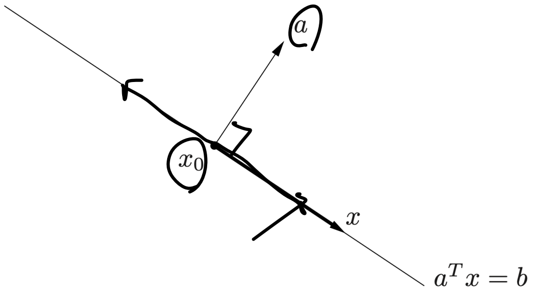




Hyperplane:

$$\mathbb{H} = \{x \in \mathbb{R}^n : a^\top x = b\}$$

where $a \neq 0$ and $b \in \mathbb{R}$.



$\underline{a}^T(x - x_0) = 0$

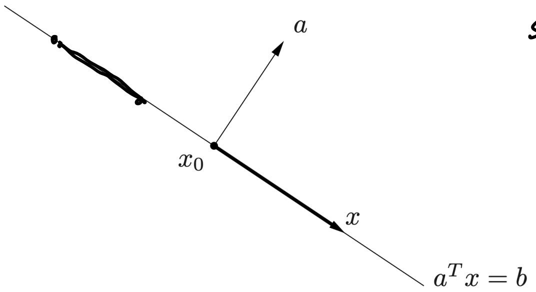
$b = a^T x_0$

Convex set

Hyperplane:

$$\mathbb{H} = \{x \in \mathbb{R}^n : a^\top x = b\}$$

where $a \neq 0$ and $b \in \mathbb{R}$.



$\boxed{x_1, x_2 \in H}$

show $\theta x_1 + (1-\theta) x_2 \in H$

$\Leftrightarrow a^T(\theta x_1 + (1-\theta)x_2) \overset{?}{=} b$

$= \theta a^T x_1 + (1-\theta) a^T x_2$

$= \theta b + (1-\theta) b$

$= b$

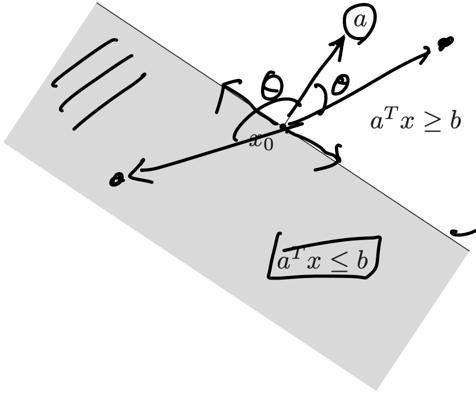
- Notice the set of points on the line is a convex set.
- You can also prove hyperplanes are convex.

Convex set

Half space:

$a^T x = b$ — H

$$\mathbb{H}^+ = \{x \in \mathbb{R}^n : a^\top x \geq b\} \quad \text{or} \quad \mathbb{H}^- = \{x \in \mathbb{R}^n : a^\top x \leq b\}\ .$$



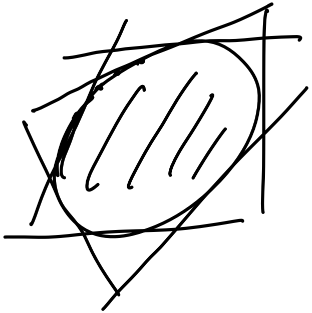
$a^T(x - x_0) = 0$

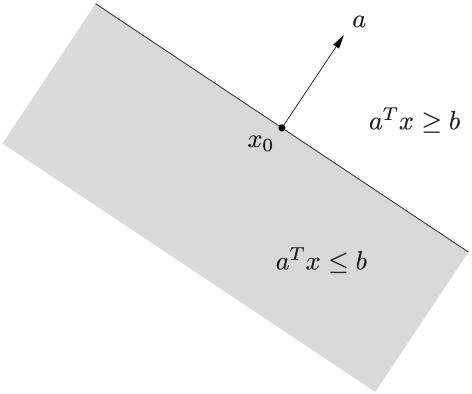
$a^T(x - x_0) \leq 0$

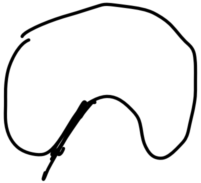
Convex set

Half space:

$$\mathbb{H}^{+} = \{x \in \mathbb{R}^n : a^\top x \geq b\} \quad \text{or} \quad \mathbb{H}^{-} = \{x \in \mathbb{R}^n : a^\top x \leq b\} \ .$$







- You can draw a convex set with half spaces; for a nonconvex set you can't.

Norm ball with center x_c and radius r:

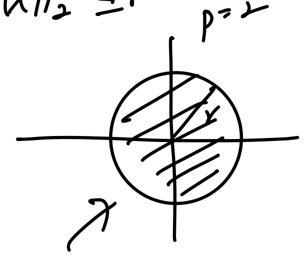
$$\{x : \|x - x_c\|_p \leq r\} .$$

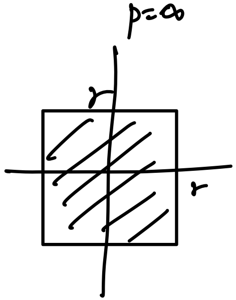
0

$\|x\|_2 \leq r$

$p=2$ Euclidean ball

$p=\infty$





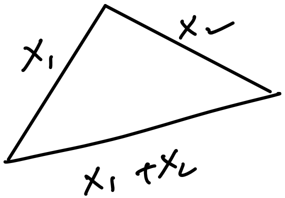
Norm ball with center x_c and radius r:

$$\{x : \|x - x_c\|_p \le r\} \ .$$

$x_1, x_2 \in S$

show $\boxed{\theta x_1 + (1-\theta) x_2} \in S$

Exercise. (triangle inequality)



$\|x_1 + x_2\| \le \|x_1 + x_2\|$

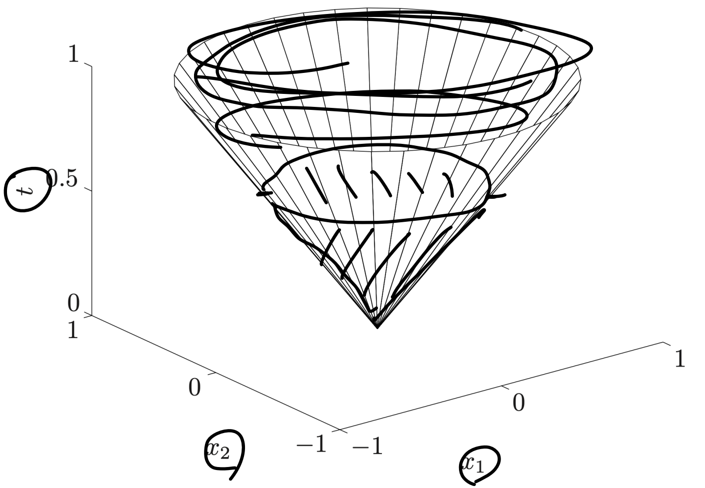
- Prove Euclidean balls are convex:

Norm cone:

$$\{(x, t) : \|x\|_p \leq t\}$$

$x \in \mathbb{R}^2$

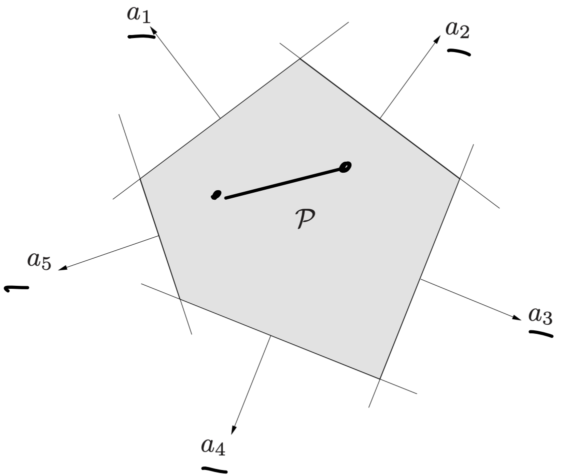
t

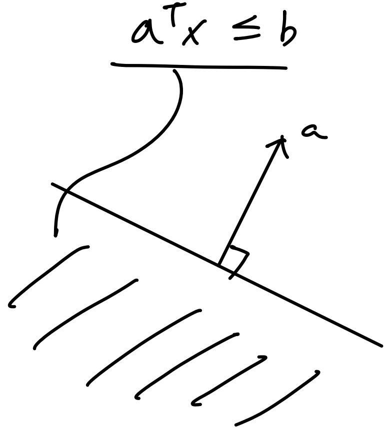


Polyhedron: the solution set of a finite number of linear equalities and inequalities

$$\mathbb{P} = \{x : \boxed{a_i^\top x \leq b_i}, i = 1, \ldots, m,\ c_j^\top x = d_j, j = 1, \ldots, p\} \ .$$

W01-2
opt.
Transportation



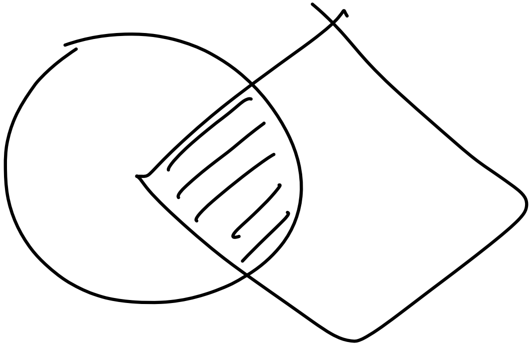


A polyhedron is the intersection of a finite number of halfspaces and hyperplanes.

Convex set

Intersections of convex sets are convex.

Let $C_i, i \in \mathbb{I}$ be convex sets, where $\mathbb{I}$ is a index set. Then $C = \cap_{i\in\mathbb{I}} C_i$ is a convex set.



Convex set

Intersections of convex sets are convex.

Let $C_i, i \in \mathbb{I}$ be convex sets, where $\mathbb{I}$ is a index set. Then $C = \cap_{i \in \mathbb{I}} C_i$ is a convex set.

obj : linear
constraint : linear

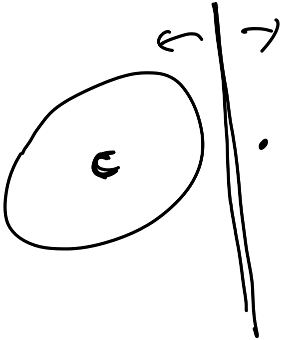
Example: linear program with linear inequalities constraints $Ax \leq b$.

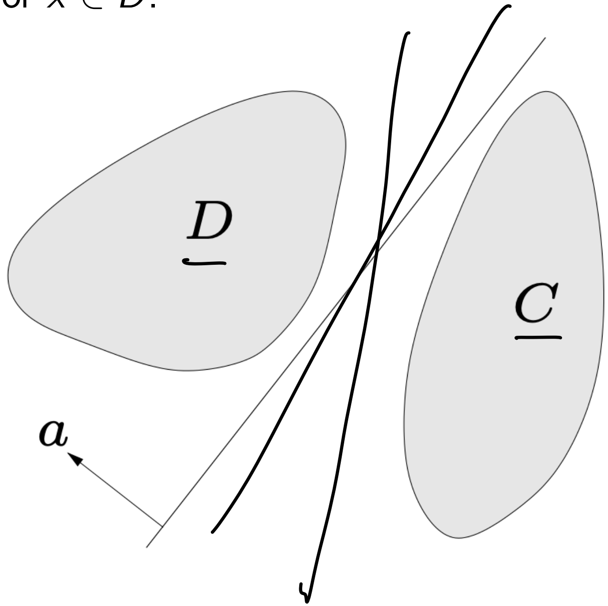
- Each constraint $a_i^\top x \leq b_i$ defines a half-space.
- Half-spaces are convex sets.
- So the set of x satisfying $Ax \leq b$ is the intersection of convex sets.

Convex set

Separating hyperplane theorem

If C and D are nonempty disjoint convex sets, there exist $a \neq 0$, b such that $a^\top x \leq b$ for $x \in C$ and $a^\top x \geq b$ for $x \in D$.

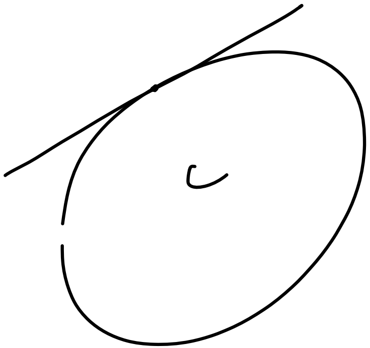


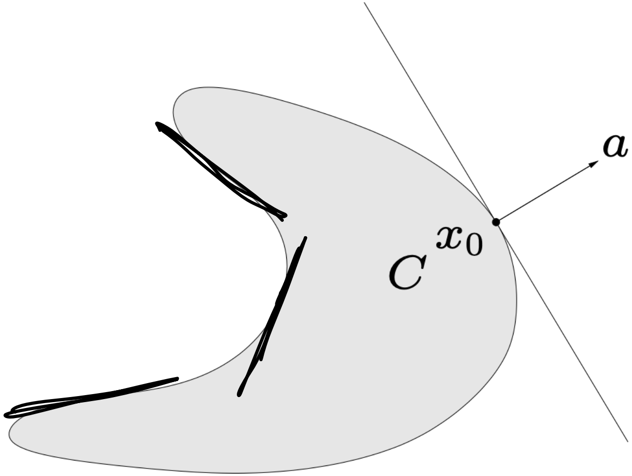


Convex set

Supporting hyperplane theorem

If C is convex, then there exist a supporting hyperplane at every boundary point of C.





Thank you

Any questions?

Credits

A lot of material in this course is borrowed or derived from the following:

- Numerical Optimization, Jorge Nocedal and Stephen J. Wright.
- Convex Optimization, Stephen Boyd and Lieven Vandenberghe.
- Convex Optimization, Ryan Tibshirani.
- Optimization for Machine Learning, Martin Jaggi and Nicolas Flammarion.
- Optimization Algorithms, Constantine Caramanis.
- Advanced Machine Learning, Mark Schmidt.