Optimization in Machine Learning: Lecture 2

Convex Sets

by Xiaolin Huang xiaolinhuang@sjtu.edu.cn SEIEE 2-429

Institute of Image Processing and Pattern Recognition

http://www.pami.sjtu.edu.cn/

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

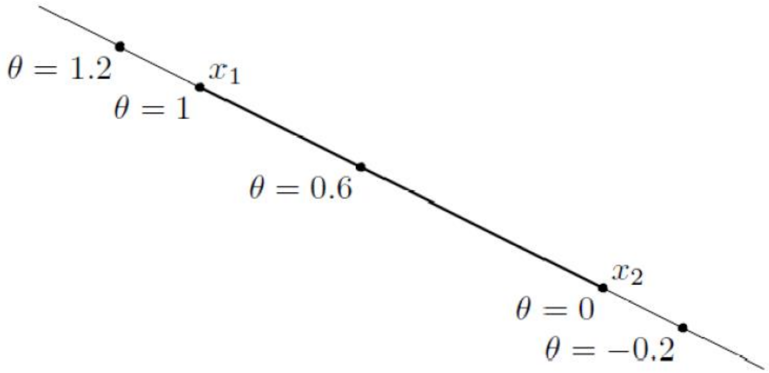
目录 Contents

目录 Contents

Affine Set

- all points through $x_1 \neq x_2$ could be expressed as a **line**

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



- **affine combination** of x_1 and x_2

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

- $C \subseteq \mathbf{R}^n$ is **affine** (an **affine set**) if and only if (iff)

$$\theta x_1 + (1 - \theta) x_2 \in C, \quad \forall \theta \in \mathbf{R}, \ \ \forall x_1, x_2 \in C$$

Affine Set

- more than two points: the **affine combination** of $x_1, x_2, \dots, x_k$

$$\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k \text{ with } \theta_1 + \theta_2 + \cdots + \theta_k = 1$$

- an affine set C contains affine combinations of any point in C.
- example:

subspace?

  - solution of linear equations is affine.
  - an affine set can be represented as solution of linear equations

"linear" and "affine"

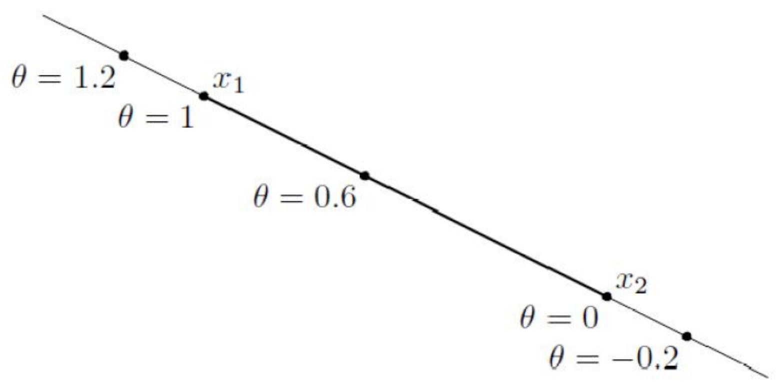
  - affine = linear + constant e.g., affine function = linear function + bias
  - prove: for an affine set C, for any $x_0 \in C$, $\{x - x_0 | x \in C\}$ is a (linear) subspace

Convex Combination

- all points through $x_1 \neq x_2$ could be expressed as a **line**

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

- further if $\theta \in [0,1]$, it becomes a **segment**



- **affine combination**

$$\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k \quad \text{with} \qquad \theta_1 + \theta_2 + \cdots + \theta_k = 1$$

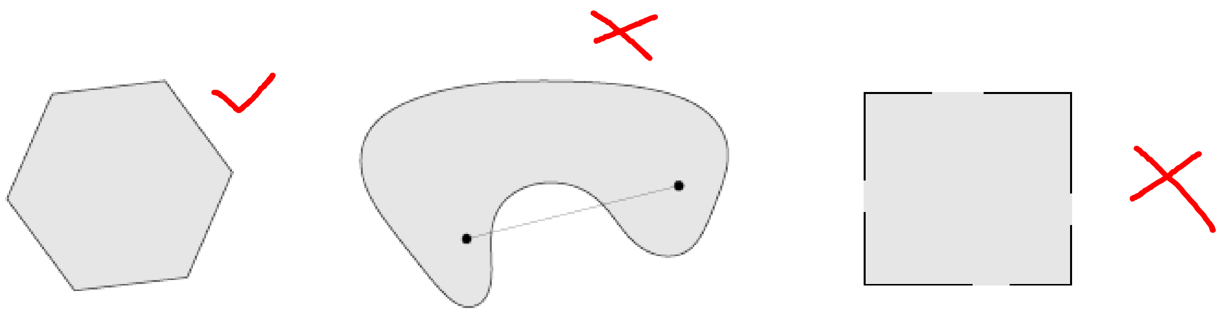
- **convex combination**

$$\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k \quad \text{with} \qquad \theta_1 + \theta_2 + \cdots + \theta_k = 1 \ \text{ and } \ \theta_i \geq 0, \forall i$$

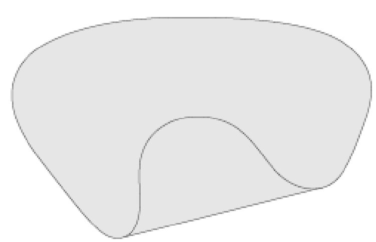
Convex Set

- a **convex set** C contains convex combinations of points in the set.

- Example:



- **convex hull**: a set of all convex combinations of points in S



$S \subseteq \textbf{conv}\, S$

$\textbf{conv}\, S \subseteq \{C : S \subseteq C \text{ and } C \text{ is convex}\}$

Convex Set: Go to Infinity

- convex combination of infinite number of points in a convex set C

$$\sum_{i=1}^{\infty} \theta_i x_i = \theta_1 x_1 + \theta_2 x_2 + \cdots \quad \text{with} \quad \sum_{i=1}^{\infty} \theta_i = 1 \quad \text{and} \quad \theta_i \geq 0, \forall i$$

if the series converges, then

$$s = \sum_{i=1}^{\infty} \theta_i x_i \in C$$

how to prove

- Proof.

let $s_N = \sum_{i=1}^{N} \theta_i x_i \Big/ \sum_{i=1}^{N} \theta_i$, then $s_N \in C$

obviously, $\lim_{N \to \infty} s_N = s$

$$s \in C$$

- it is true only when C is closed

Closed and Open Set

- **interior point** of a set C

  $x \in C$ and $\exists \varepsilon > 0$ such that $\{y: \|x - y\| \leq \varepsilon\} \subseteq C$

  $\theta_i \geq 0, \forall i$

- a set C is **open** iff any point in C is an interior point

- a set C is **closed** iff its complementary set is open

  do we have sets which are both closed and open?

  - corollary: any convergent sequence $s_1, s_2, \ldots \in C$, its limit $s = \lim_{N\to\infty} s_N \in C$

- in an optimization problem, we generally consider about closed set
  - if the set is open, optimal solutions on the boundary are inaccessible
  - the gap of the optima for an open set and its **closure** is negligible
  - optimization problems always consider about "$\geq$" "$\leq$" rather than "$>$""$<$"

Convex Set: Go to Infinity

- convex combination of infinite number of points in a convex set C

$$\sum_{i=1}^{\infty} \theta_i x_i = \theta_1 x_1 + \theta_2 x_2 + \cdots \quad \text{with} \quad \sum_{i=1}^{\infty} \theta_1 = 1 \quad \text{and} \quad \theta_i \geq 0, \forall i$$

if the series converges, then

$$\sum_{i=1}^{\infty} \theta_i x_i \in C$$

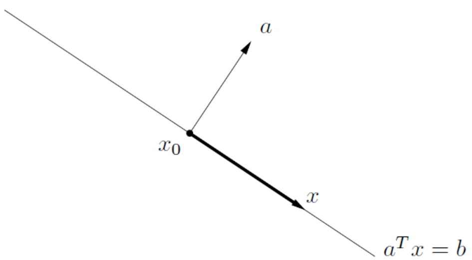
- more general, for $p(x)\colon \mathbf{R}^n \to \mathbf{R}$, there is $\underbrace{p(x) \geq 0, \forall x \in C \text{ and } \int_C \ p(x) dx = 1,}_{\text{“probability”}}$

where C is a convex set, if the following integral exists, then

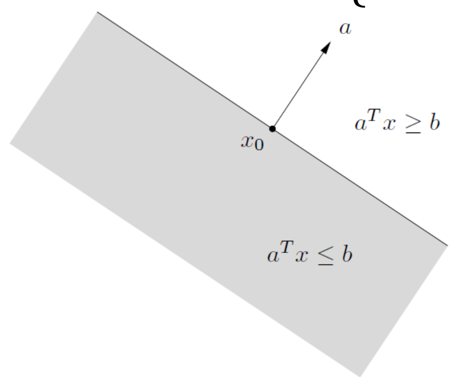
$$\int_C \ p(x) x \, dx \in C$$

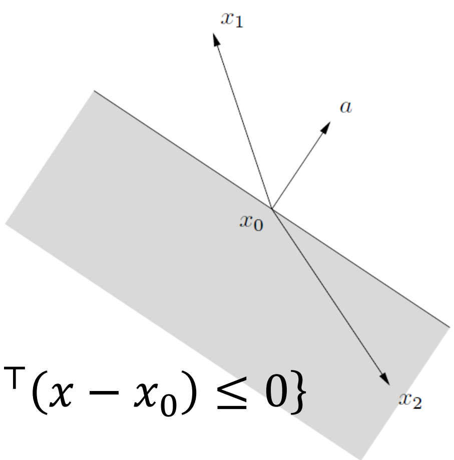
Convex Set: Examples

- hyperplanes and halfspaces



$\{x: a^\top x \leq b\}$



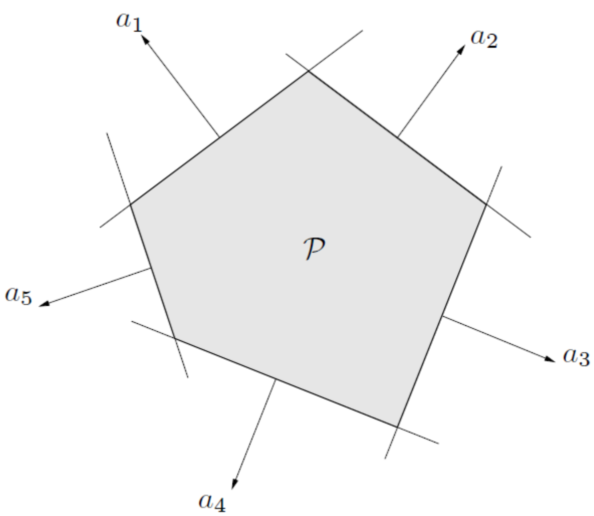


$\{x: a^\top (x - x_0) \leq 0\}$

Convex Set: Examples

- polyhedra

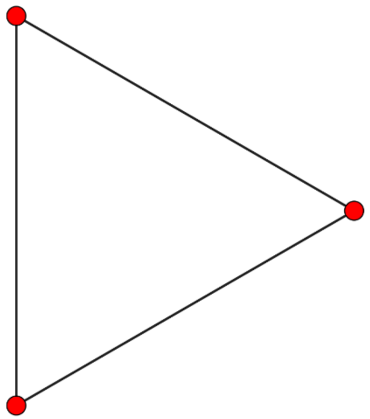
$$\{x : a_i^\top x \le b_i, c_j^\top x = d_j\}$$

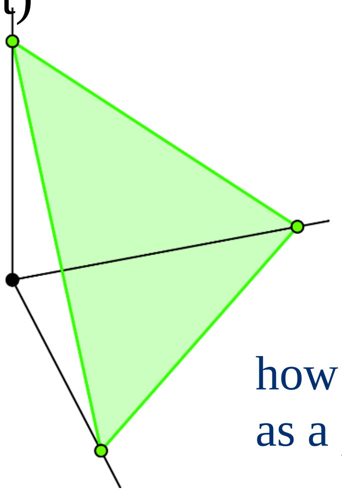


- simplexes

  a convex hull of affinely independent points

  (i.e., the difference is linearly independent)





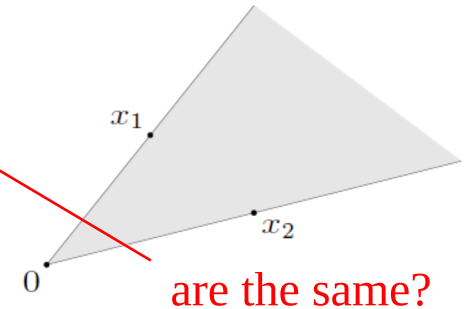
how to represent a simplex as a polyhedra

Convex Set: Examples

- conic set/cone C: $\theta x \in C, \forall \theta \geq 0, \forall x \in C$
- conic combination/nonnegative combination

$$\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k \quad \text{with} \quad \theta_i \geq 0$$

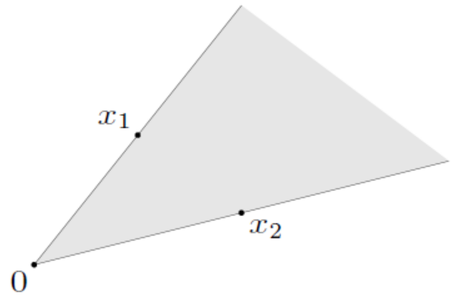
- a set contains conic combinations of points in the set



are the same?

Convex Set: Examples

- conic set/cone C: $\theta x \in C, \forall \theta \geq 0, \forall x \in C$



- conic combination/nonnegative combination

$$\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k \quad \text{with} \quad \theta_i \geq 0$$

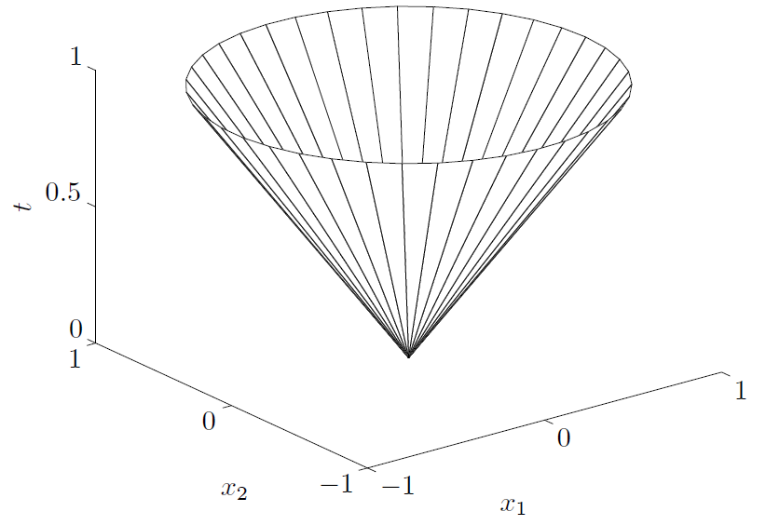
- **convex cone**: a set contains conic combinations of points in the set

- conic hull

- second-order cone

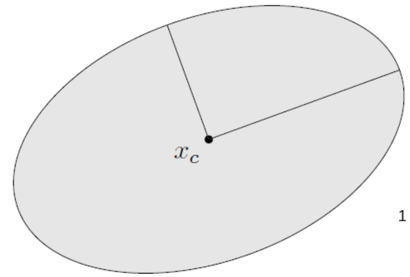
$$C = \{(x, t) \in R^{n+1} : \|x\|_2 \leq t\}$$

$$\begin{bmatrix} x \\ t \end{bmatrix}^\top \begin{bmatrix} I & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ t \end{bmatrix} \leq 0, t \geq 0$$



Convex Set: Examples

- Euclid balls and ellipsoids

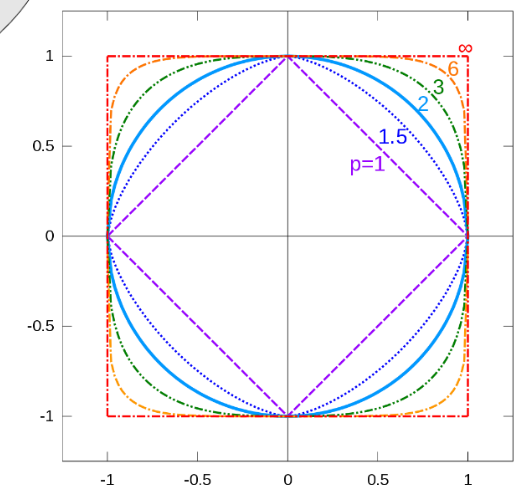


$$\{x: (x - x_c)^\top P^{-1}(x - x_c) \le 1\}$$

P is symmetric and positive definite

$$P = P^\top \qquad\qquad P \succ 0$$

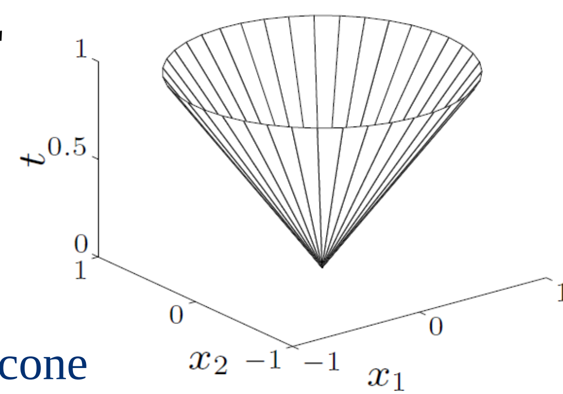
- norm ball $\{x: \|x - x_c\| \le t\}$



$$\{x: \|x - x_c\|_1 \le 1\} \qquad \{x: \|x - x_c\|_2 \le 1\} \qquad \{x: \|x - x_c\|_\infty \le 1\}$$

- norm cone $\{(x,t): \|x\| \le t\}$, which is very useful in ML

$$\min_x \ \|x\| + \lambda\|Ax - Y\| \quad \rightarrow \quad \min_{x,t} \ t + \lambda\,\|Ax - Y\|, \text{s.t.}\ \|x\| \le t$$



second-order cone

Convex Set: Examples

Do you remember PSD?

- positive semidefinite matrices
  - symmetric $n \times n$ matrices $\mathbf{S}^n = \{X \in \mathbf{R}^{n \times n} : X^\top = X\}$
  - symmetric positive semidefinite matrices

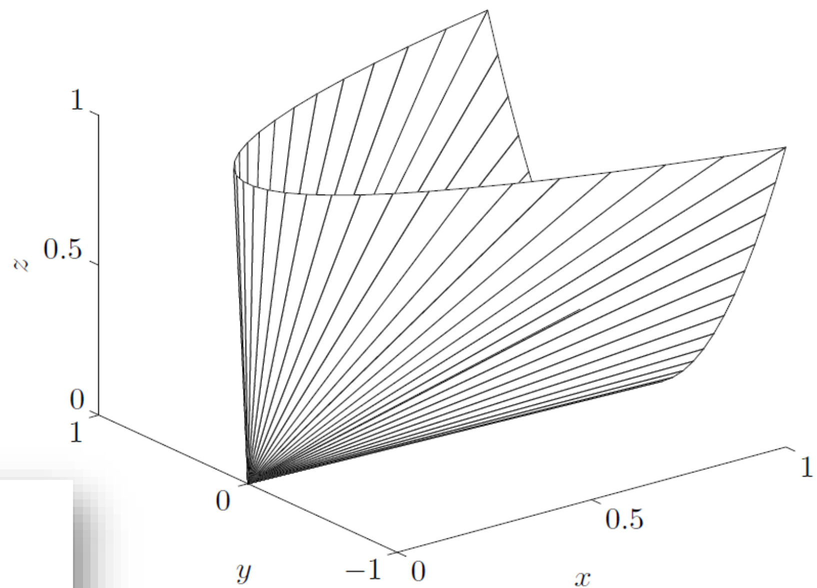
$$\mathbf{S}_+^n = \{X \in \mathbf{S}^n : X \boxed{\succcurlyeq} 0\}$$

  - symmetric positive definite matrices

$$\mathbf{S}_{++}^n = \{X \in \mathbf{S}^n : X \succ 0\}$$

- the set $\mathbf{S}_+^n$ is a cone, e.g.,

$$X = \begin{bmatrix} x & y \\ y & z \end{bmatrix} \in \mathbf{S}_+^2 \iff x \geq 0, \quad z \geq 0, \quad xz \geq y^2.$$



Convex Set: Examples

- positive semidefinite matrices
  - symmetric $n \times n$ matrices $\quad \mathbf{S}^n = \{X \in \mathbf{R}^{n\times n}: X^\top = X\}$
  - symmetric positive semidefinite matrices

$$\mathbf{S}_+^n = \{X \in \mathbf{S}^n: X \succcurlyeq 0\}$$

  - symmetric positive definite matrices

$$\mathbf{S}_{++}^n = \{X \in \mathbf{S}^n: X \succ 0\}$$

- nonnegative orthant $K = \mathbf{R}_+^n$

$$x \preccurlyeq_{\mathbf{R}_+^n} y \Leftrightarrow y - x \in \mathbf{R}_+^n \Leftrightarrow x_i \le y_i$$

- PSD cone $K = \mathbf{S}_+^n$

$$X \preccurlyeq_{\mathbf{S}_+^n} Y \Leftrightarrow Y - X \in \mathbf{S}_+^n$$

Generalized Inequality

- K is a proper cone:
  - closed
  - solid: non-empty interior
  - pointed: contains no line
- K defines a generalized inequality

$$x \preccurlyeq_K y \Leftrightarrow y - x \in K$$

$$x \prec_K y \Leftrightarrow y - x \in \mathbf{int}\, K$$

目录 Contents

Prove and Establish Convex Sets

- by definition

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

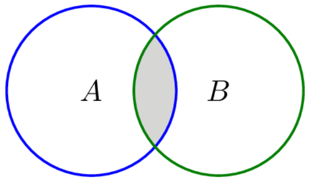
- from basic convex sets mentioned before, we can obtain convex sets by the following **operations that preserve convexity**:
    - intersection
    - affine functions
    - perspective function
    - linear-fractional functions

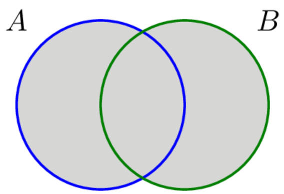
Intersection

- the intersection of (any number of) convex sets is convex
  - a polyhedra is the intersection of halfspaces and hyperplanes
  - semi-positive cone is convex:

$$\mathbf{S}_+^n = \{X \in \mathbf{S}^n : X \succcurlyeq 0\} = \bigcap_z \{X \in \mathbf{S}^n : z^\top X z \geq 0\}$$

- proof is quite trivial
- “And” is a good logic for optimization
- “Or” corresponds to union, which usually brings non-convexity





Affine Functions

- suppose $f: \mathbf{R}^n \to \mathbf{R}^m$ is affine, i.e., $f(x) = Ax + b$
  - the image of a convex set C under f is convex

$$f(C) = \{f(x): x \in C\}$$

  - the inverse image of a convex set C under f is convex :

$$f^{-1}(C) = \{x: f(x) \in C\}$$

- scaling, translation, projection
- if C_1 and C_2 are convex, so is their sum

$$C_1 + C_2 = \{x: x = x_1 + x_2, x_1 \in C_1, x_2 \in C_2\}$$

Affine Functions

- suppose $f: \mathbf{R}^n \rightarrow \mathbf{R}^m$ is affine, i.e., $f(x) = Ax + b$
  - the image of a convex set C under f is convex

$$f(C) = \{f(x): x \in C\}$$

  - the inverse image of a convex set C under f is convex :

$$f^{-1}(C) = \{x: f(x) \in C\}$$

- polyhedron $\{x: Ax \preccurlyeq b, c_j^\top x = d_j\}$ is the inverse image of $\mathbf{R}_+^m \times \{0\}$ under $f(x) = (b - Ax, d - Cx)$: $\{x: Ax \preccurlyeq b, c_j^\top x = d_j\} = \{x: f(x) \in \mathbf{R}_+^m \times \{0\}\}$

- solution of LMI $\{x: A(x) = x_1 A_1 + x_2 A_2 + \cdots + x_n A_x \preccurlyeq B\}$ is convex, since

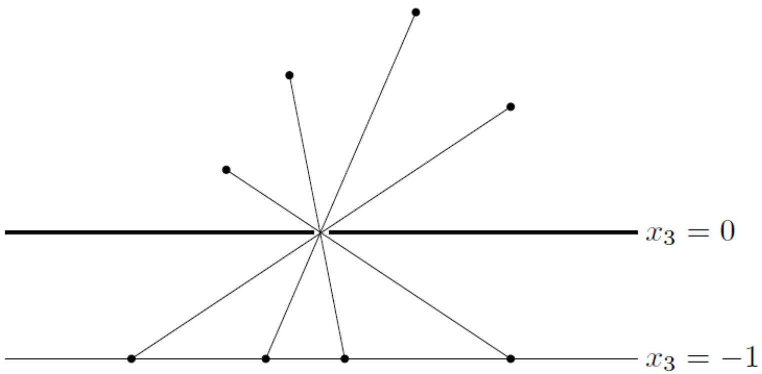
it is the inverse image of $\mathbf{S}_+^n$ in the affine function $f(x) = B - A(x)$.

Perspective and linear-fractional function

- perspective function P: $\mathbf{R}^{n+1} \to \mathbf{R}^n$

$$P(x,t) = x/t$$

$$\textbf{dom } P = \{(x,t) : t > 0\}$$



Observe a convex object though a pin-hole

- linear-fractional functions

$$f(x) = \frac{Ax + b}{c^\top x + d}$$

$$\textbf{dom } f = \{x | c^\top x + d > 0\}$$

  - it is a composition of linear function and perspective function
  - conditional probability: for random $u \in \{1, \ldots, n\}$ $v \in \{1, \ldots, m\}$ with probability $p_{ij} = \text{prob}(u = i, v = j)$, then the conditional property $f_{ij} = \text{prob}(u = i | v = j)$ is

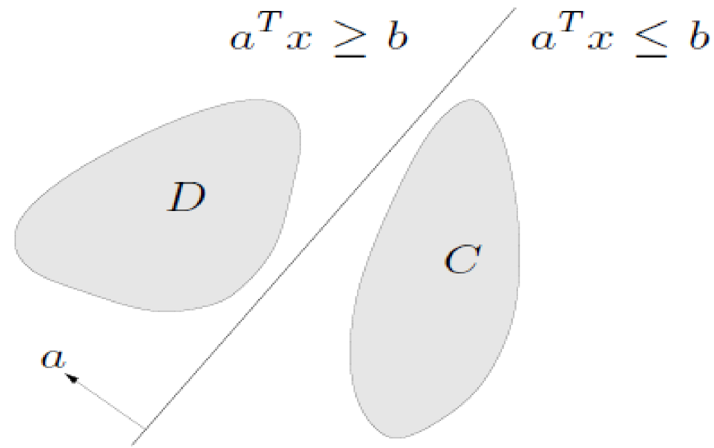
$$f_{ij} = p_{ij} / \Sigma_k p_{kj}$$

目录 Contents

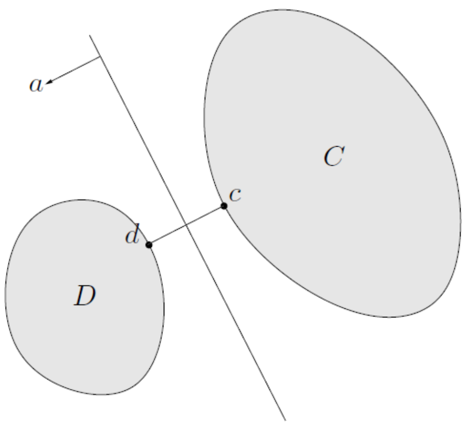
Separating Hyperplane Theorem

- if *C* and *D* are disjoint convex sets, i.e., $C \cap D = \emptyset$, then there exists at least one **separating hyperplane** such that

$$a^{\top}x \leq b, \forall x \in C$$
$$a^{\top}x \geq b, \forall x \in D$$



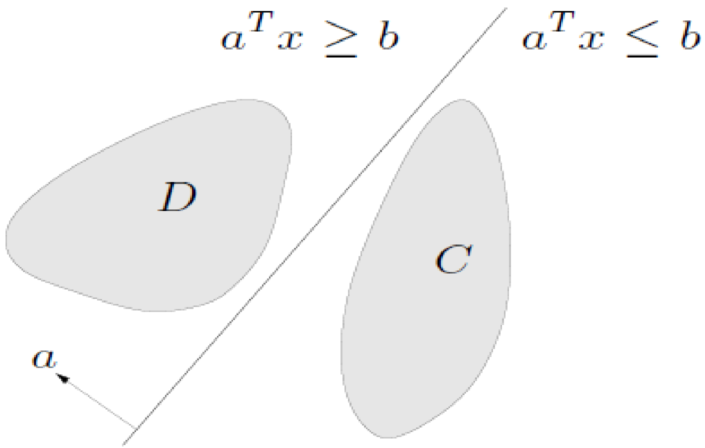
- Proof sketch for the case, the distance between *C* and *D* are positive
  - find the closest points
  - define the hyperplane by the middle point and the orthogonal direction



Separating Hyperplane Theorem

- if C and D are disjoint convex sets, i.e., $C \cap D = \emptyset$, then there exists at least one **separating hyperplane** such that

$$a^\top x \le b, \forall x \in C$$
$$a^\top x \ge b, \forall x \in D$$



- Proof sketch for the case, the distance between C and D are positive
  - find the closest points
  - define the hyperplane by the middle point and the orthogonal direction

- why we need distance being positive?
- is this assumption meaningful?
- how about non-convex sets?
- how about the converse theorem?
(if separating hyperplane, then disjoint)

Alternative Theorem

- There is no solution for a system of strict linear inequalities $Ax \prec b$ with $A \in R^{m\times n}$ iff the following system is feasible

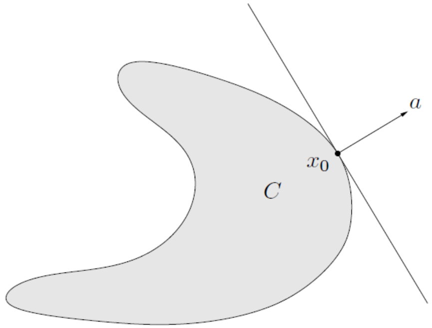
$$\lambda \neq 0, \lambda \succcurlyeq 0, A^{\top}\lambda = 0, \lambda^{\top} b \leq 0$$

- Proof sketch: $C = \{y: y = b - Ax, x \in \mathbf{R}^n\}$ $D = \{y \in \mathbf{R}^m: y \succ 0\}$
  - C and D are disjoint, there is a separating hyperplane $\lambda^{\top} y \leq \mu, \forall y \in C$ $\lambda^{\top} y \geq \mu, \forall y \in D$
  - further $\lambda^{\top}(b - Ax) \leq \mu, \forall x \Leftrightarrow A^{\top}\lambda = 0, \lambda^{\top} b \leq \mu$

    $\lambda^{\top} y \geq \mu, \forall y \succ 0 \Leftrightarrow \mu \leq 0, \lambda \succcurlyeq 0, \lambda \neq 0,$
  - only one of the systems could be feasible
    - $Ax \prec b$
    - $\lambda \neq 0, \lambda \succcurlyeq 0, A^{\top}\lambda = 0, \lambda^{\top} b \leq 0$

- slackness condition
- dual variable
- support vector

Supporting Hyperplane Theorem



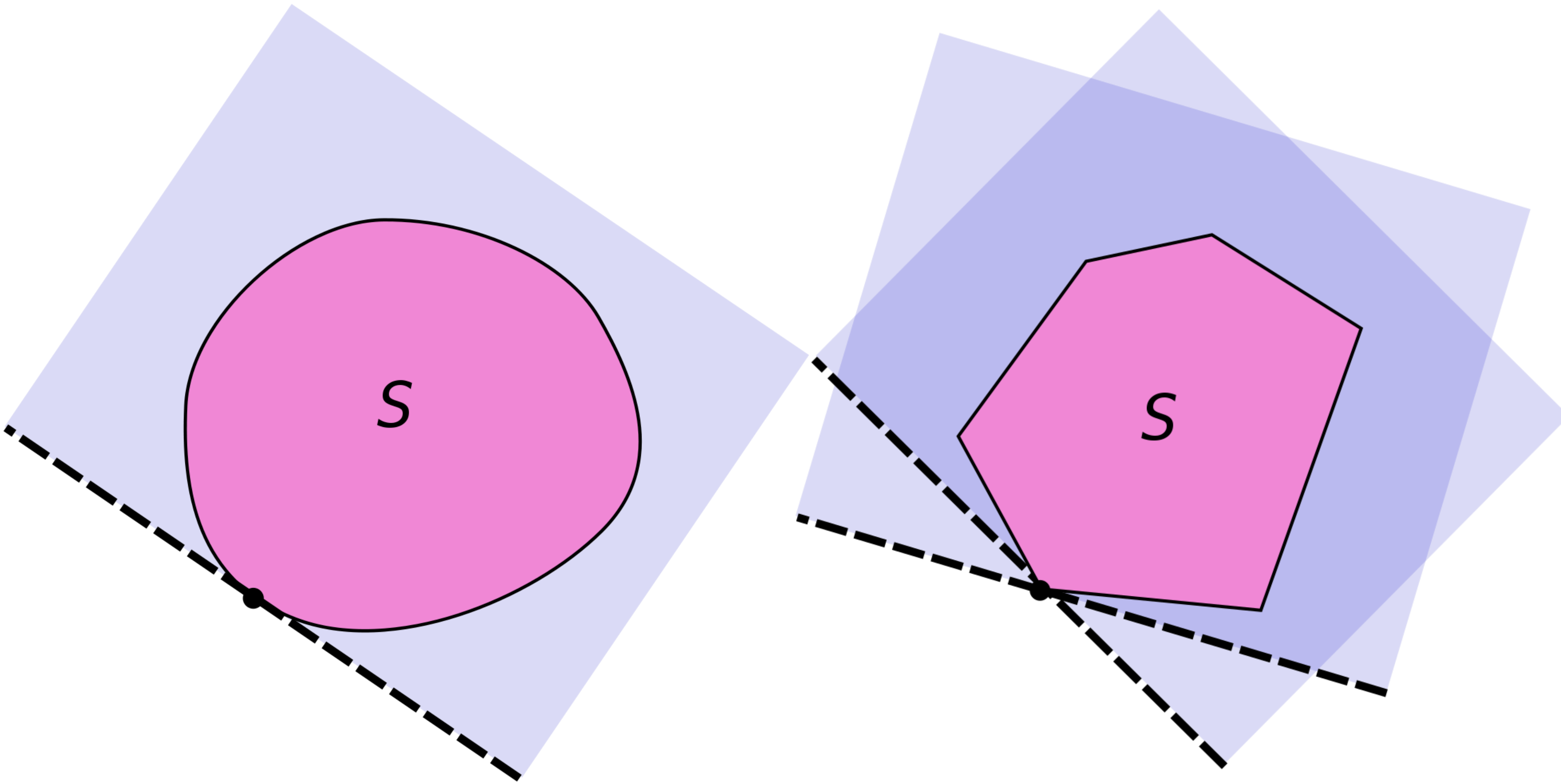
For a set $C \subseteq \mathbf{R}^n$ and a point in its boundary x_0, if $a \neq 0$ satisifes $a^\top x \leq a^\top x_0, \forall x \in C$, then we call the corresponding hyperplane a **supporting hyperplane** to C at x_0

- **Supporting Hyperplane Theorem**: for a convex set, there exists at least one supporting hyperplane at every boundary point.

  - how about non-convex sets?
  - can we have more?

- a convex set could be represented by (maybe infinite) linear inequalities
- (partial converse): if a set is close, has nonempty interior, and has a supporting hyperplane at every point in its boundary, then it is convex.

Supporting Hyperplane Theorem



https://en.wikipedia.org/

Applications of Supporting Hyperplanes

- convex combination of infinite number of points in a convex set C

$$\sum_{i=1}^{\infty} \theta_i x_i = \theta_1 x_1 + \theta_2 x_2 + \cdots \quad \text{with} \quad \sum_{i=1}^{\infty} \theta_i = 1 \quad \text{and} \quad \theta_i \geq 0, \forall i$$

if the series converges, then

$$s = \sum_{i=1}^{\infty} \theta_i x_i \in C$$

- Proof.

let $s_N = \sum_{i=1}^{N} \theta_i x_i \Big/ \sum_{i=1}^{N} \theta_i$, then $s_N \in C$

obviously, $\lim\limits_{N\to\infty} s_N = s$

$$s \in C$$

- it is true only when C is closed

Applications of Supporting Hyperplanes

- convex combination of infinite number of points in a convex set C

$$\sum_{i=1}^{\infty} \theta_i x_i = \theta_1 x_1 + \theta_2 x_2 + \cdots \quad \text{with} \quad \sum_{i=1}^{\infty} \theta_1 = 1 \quad \text{and} \quad \theta_i \geq 0, \forall i$$

if the series converges, then

$$s = \sum_{i=1}^{\infty} \theta_i x_i \in C$$

- Proof.
  - using supporting hyperplane theorem

if s is in the boundary of C, then $C \subseteq \{x: a_s^\top x \leq a_s^\top s\}$

all x_i are in the hyperplane, i.e., $a_s^\top x_i = a_s^\top s$, otherwise, there is a contradiction

$$a_s^\top s = \sum_{i=1}^{\infty} \theta_i a_s^\top x_i < \sum_{i=1}^{\infty} \theta_i a_s^\top s = a_s^\top s$$

then we can reduce the dimension by one. Repeat it until zero dimension.

Conclusion

- convex optimization is to minimize a convex function over a **convex set**
    - **convex combination**
    - **definition of convex sets**
    - **examples**
    - **operations that preserve convexity**
    - **separating hyperplane theorem**
    - **supporting hyperplane theorem**

Conclusion and Home Work

- convex optimization is to minimize a convex function over a **convex set**
  - **convex combination**
  - **definition of convex sets**
  - **examples**
  - **operations that preserve convexity**
  - **separating hyperplane theorem**
  - **supporting hyperplane theorem**
- **Excise 2.3:** for your understanding of convexity

  Excise 2.19: the use of inverse image to check convexity

  Excise 2.23: the supporting hyperplane representation for convex sets

THANKS