



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

1

Convex Sets

2

Operations that Preserve Convexity

3

Separating and Supporting Hyperplanes



1

Convex Sets

2

Operations that Preserve Convexity

3

Separating and Supporting Hyperplanes

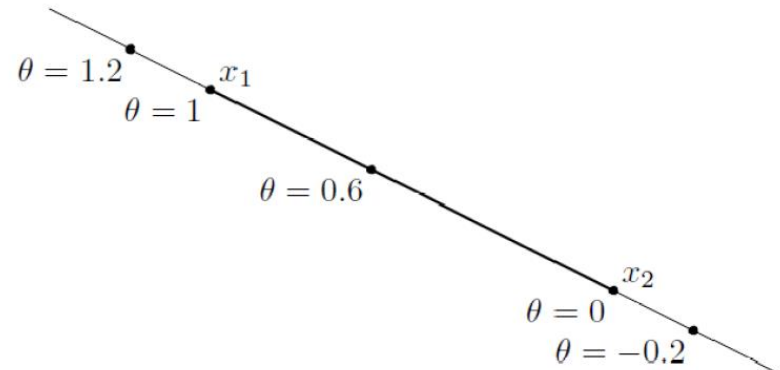


Affine Set



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

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



- 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)

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

Affine Set



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

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

- 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

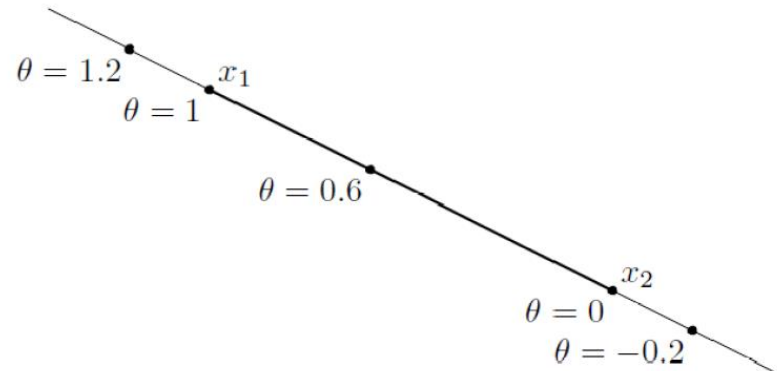
Convex Combination



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

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

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



- affine combination**

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

- convex combination**

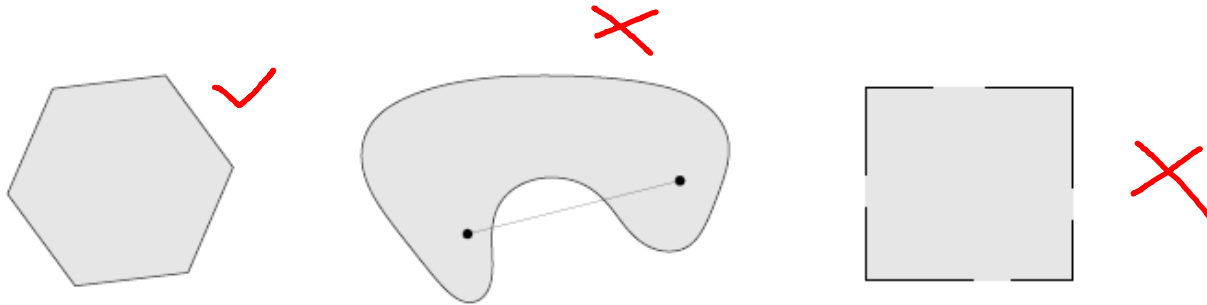
$$\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k \quad \text{with} \quad \theta_1 + \theta_2 + \cdots + \theta_k = 1 \quad \text{and} \quad \theta_i \geq 0, \forall i$$

Convex Set

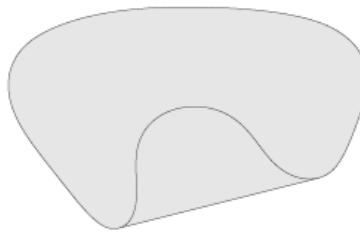


- a **convex set** C contains convex combinations of points in the set.

- Example:



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



$$S \subseteq \text{conv } S$$

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

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.

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

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

$$s \in C$$

- 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 $s_1, s_2, \dots \in C$, its limit $s = \lim_{N \rightarrow \infty} s_N \in C$

- 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

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

if the series converges, then

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

- more general, for $p(x): \mathbf{R}^n \rightarrow \mathbf{R}$, there is $p(x) \geq 0, \forall x \in C$ and $\int_C p(x) dx = 1$,

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

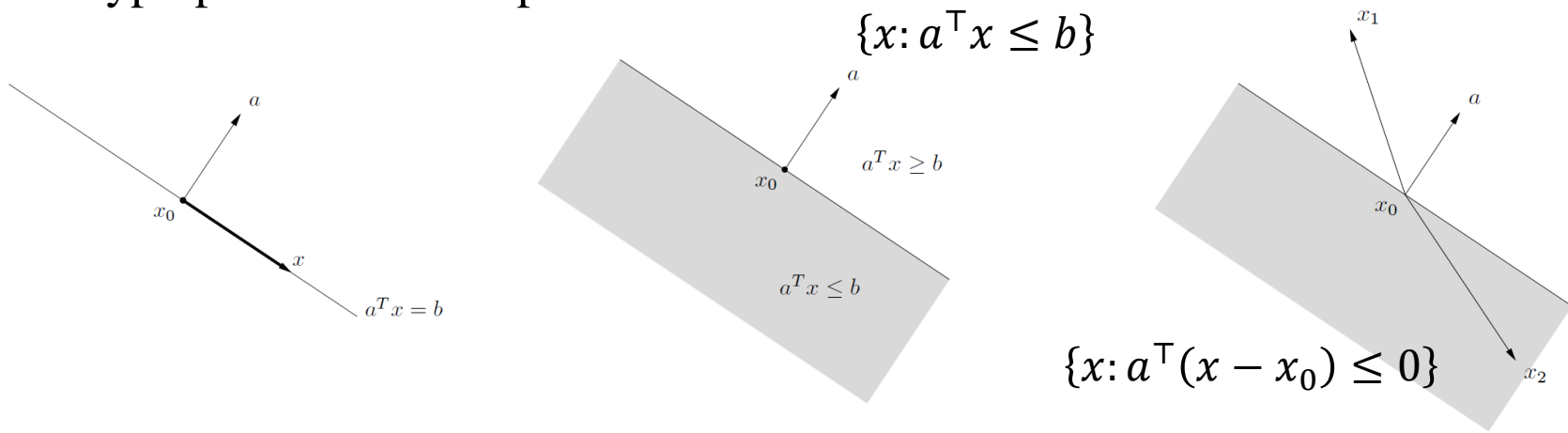
$$\int_C p(x) x dx \in C$$

“probability”

Convex Set: Examples



- hyperplanes and halfspaces



Convex Set: Examples



- polyhedra

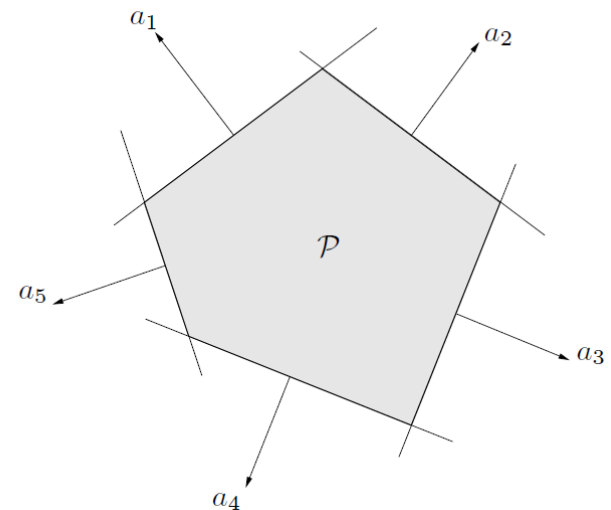
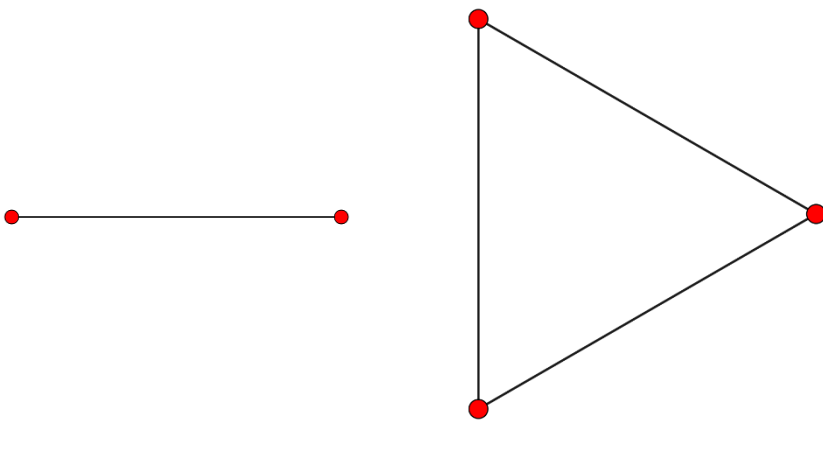


$$\{x: a_i^\top x \leq b_i, c_j^\top x = d_j\}$$

- 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

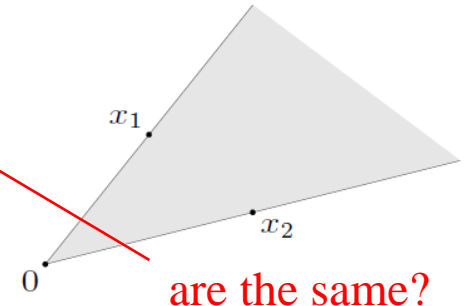


- conic set/cone C : $\theta x \in C, \forall \theta \geq 0, \forall x \in C$

- 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



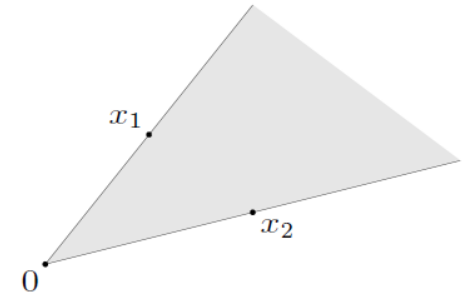
are the same?

Convex Set: Examples



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

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

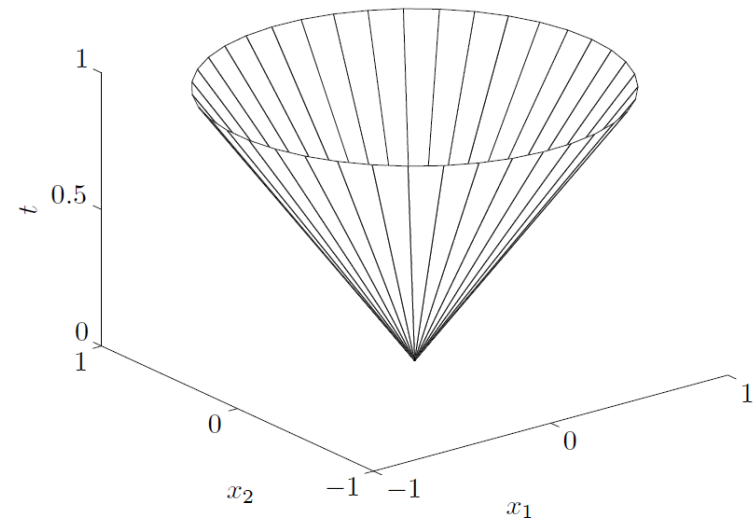


- convex cone**: a set contains conic combinations of points in the set
- conic hull
- second-order cone

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



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



Convex Set: Examples

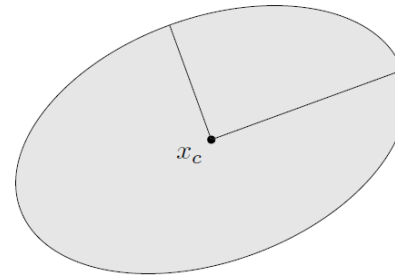
- Euclid balls and ellipsoids

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

P is symmetric and positive definite

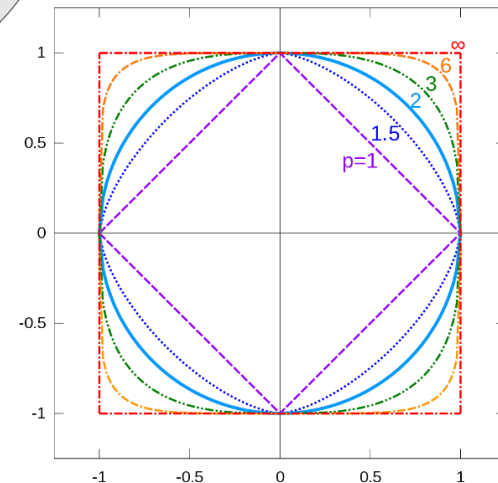
$$P = P^\top$$

$$P \succ 0$$



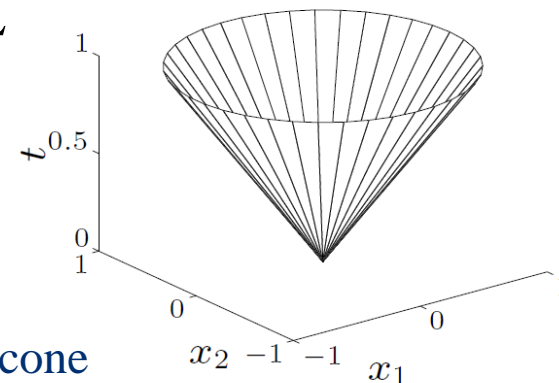
- norm ball $\{x: \|x - x_c\| \leq t\}$

$$\{x: \|x - x_c\|_1 \leq 1\} \quad \{x: \|x - x_c\|_2 \leq 1\} \quad \{x: \|x - x_c\|_\infty \leq 1\}$$



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

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



second-order cone

Convex Set: Examples



- positive semidefinite matrices

Do you remember PSD?

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

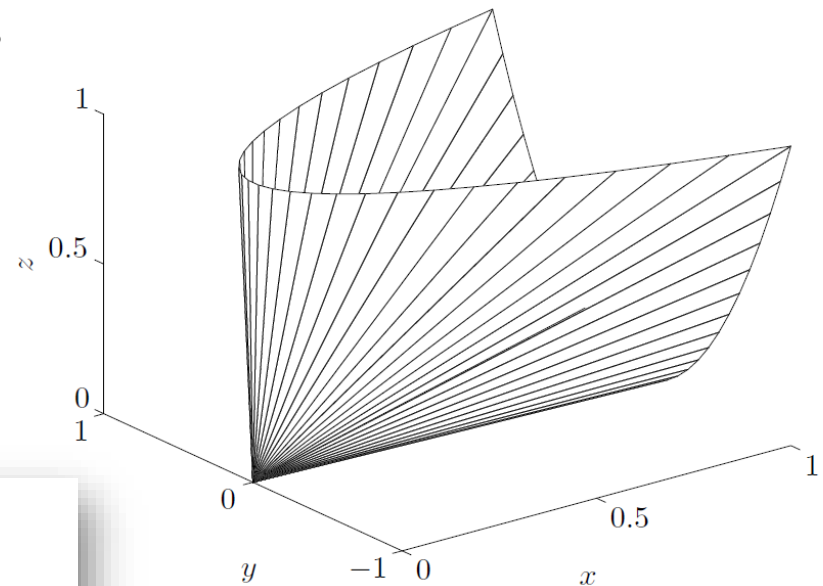
$$\mathbf{S}_+^n = \{X \in \mathbf{S}^n : X \succeq 0\}$$

- 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 $\mathbf{S}^n = \{X \in \mathbf{R}^{n \times n}: X^\top = X\}$
- 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 $K = \mathbf{R}_+^n$

$$x \preceq_{\mathbf{R}_+^n} y \iff y - x \in \mathbf{R}_+^n \iff x_i \leq y_i$$

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

$$X \preceq_{\mathbf{S}_+^n} Y \iff Y - X \in \mathbf{S}_+^n$$

Generalized Inequality

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

$$x \preceq_K y \iff y - x \in K$$

$$x \prec_K y \iff y - x \in \text{int } K$$

1

Convex Sets

2

Operations that Preserve Convexity

3

Separating and Supporting Hyperplanes



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

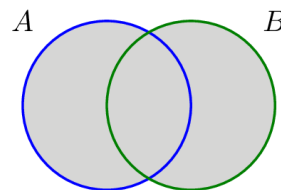
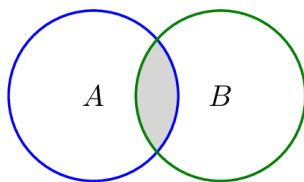
Intersection



- 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 $f: \mathbf{R}^n \rightarrow \mathbf{R}^m$ 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 :

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

- 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 $\{x: Ax \preceq b, c_j^\top x = d_j\}$ is the inverse image of $\mathbf{R}_+^m \times \{0\}$ under

$$f(x) = (b - Ax, d - Cx): \{x: Ax \preceq b, c_j^\top x = d_j\} = \{x: f(x) \in \mathbf{R}_+^m \times \{0\}\}$$

- solution of LMI $\{x: A(x) = x_1 A_1 + x_2 A_2 + \cdots + x_n A_n \preceq B\}$ 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 $P: \mathbf{R}^{n+1} \rightarrow \mathbf{R}^n$

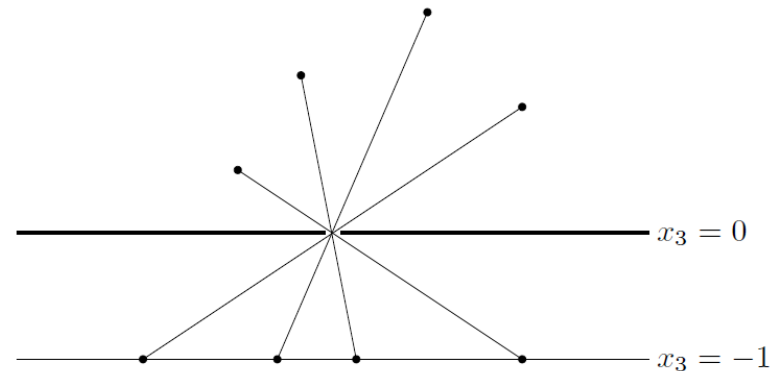
$$P(x, t) = x/t$$

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

- linear-fractional functions

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

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



Observe a convex object through a pin-hole

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

1

Convex Sets

2

Operations that Preserve Convexity

3

Separating and Supporting Hyperplanes



Separating Hyperplane Theorem

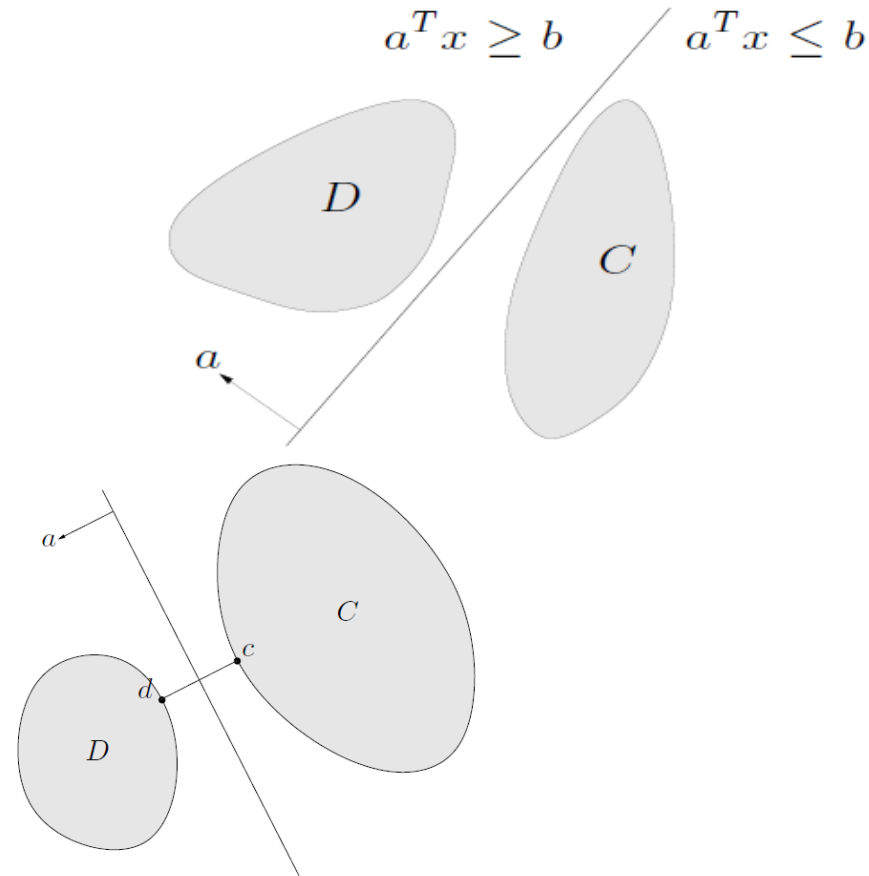


- if C and D are disjoint convex sets, i.e., $C \cap D = \emptyset$, then there exists at least one **separating hyperplane** such that

$$a^\top x \leq b, \forall x \in C$$

$$a^\top x \geq b, \forall x \in D$$

- 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



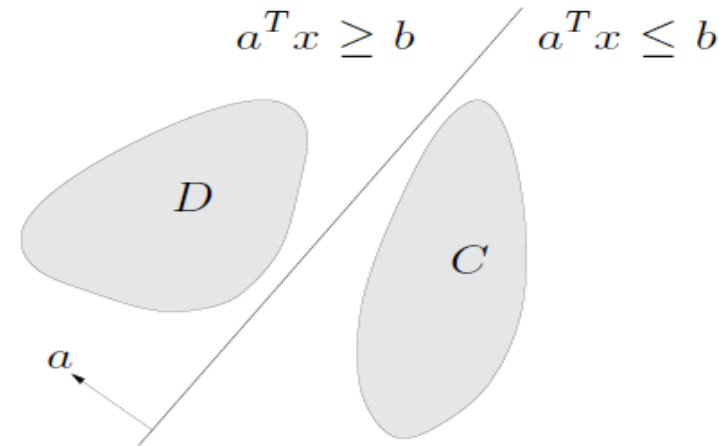
- if C and D are disjoint convex sets, i.e., $C \cap D = \emptyset$, then there exists at least one **separating hyperplane** such that

$$a^T x \leq b, \forall x \in C$$

$$a^T x \geq b, \forall x \in D$$

- 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



□ 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 $Ax < b$ with $A \in \mathbf{R}^{m \times n}$ iff the following system is feasible

$$\lambda \neq 0, \lambda \geq 0, A^T \lambda = 0, \lambda^T b \leq 0$$

- Proof sketch: $C = \{y: y = b - Ax, x \in \mathbf{R}^n\}$ $D = \{y \in \mathbf{R}^m: y > 0\}$

- C and D are disjoint, there is a separating hyperplane
- further $\lambda^T(b - Ax) \leq \mu, \forall x \Leftrightarrow A^T \lambda = 0, \lambda^T b \leq \mu$

$$\begin{aligned} \lambda^T y &\leq \mu, \forall y \in C \\ \lambda^T y &\geq \mu, \forall y \in D \end{aligned}$$

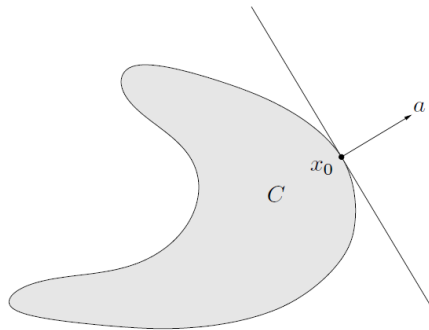
$$\lambda^T y \geq \mu, \forall y > 0 \Leftrightarrow \mu \leq 0, \lambda \geq 0, \lambda \neq 0,$$

- only one of the systems could be feasible


- $Ax < b$
- $\lambda \neq 0, \lambda \geq 0, A^T \lambda = 0, \lambda^T b \leq 0$

- slackness condition
- dual variable
- support vector

Supporting Hyperplane Theorem

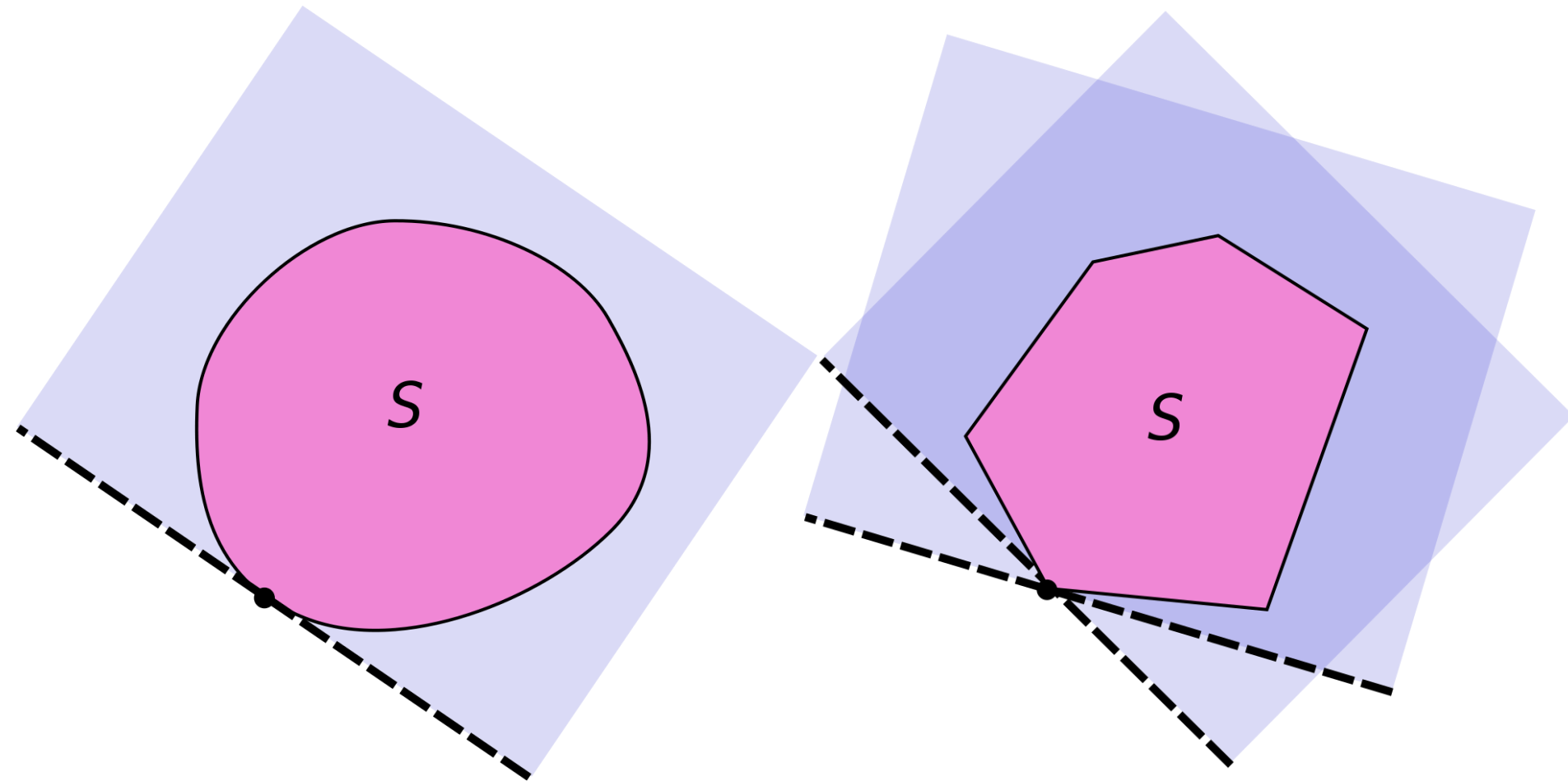


For a set $C \subseteq \mathbf{R}^n$ and a point in its boundary x_0 , if $a \neq 0$ satisfies $a^\top x \leq a^\top x_0, \forall x \in C$, then we call the corresponding hyperplane a **supporting hyperplane** to C at x_0

- **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



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.

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

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

$$s \in C$$

- it is true only when C is closed

Applications of Supporting Hyperplanes



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

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

if the series converges, then

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

- Proof.
 - 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

