Nonlinear Optimization

Part I
Formulation of convex optimization problems

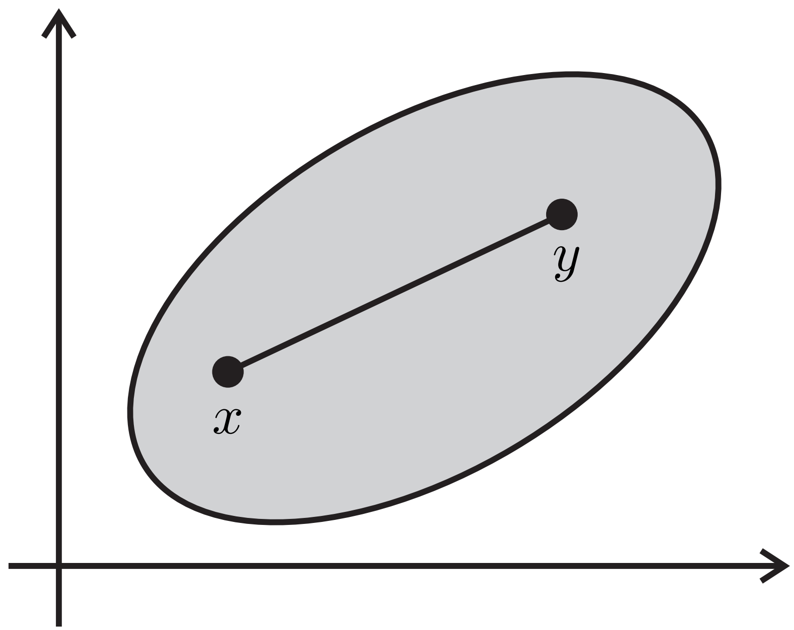
Instituto Superior Técnico – Carnegie Mellon University

jxavier@isr.ist.utl.pt

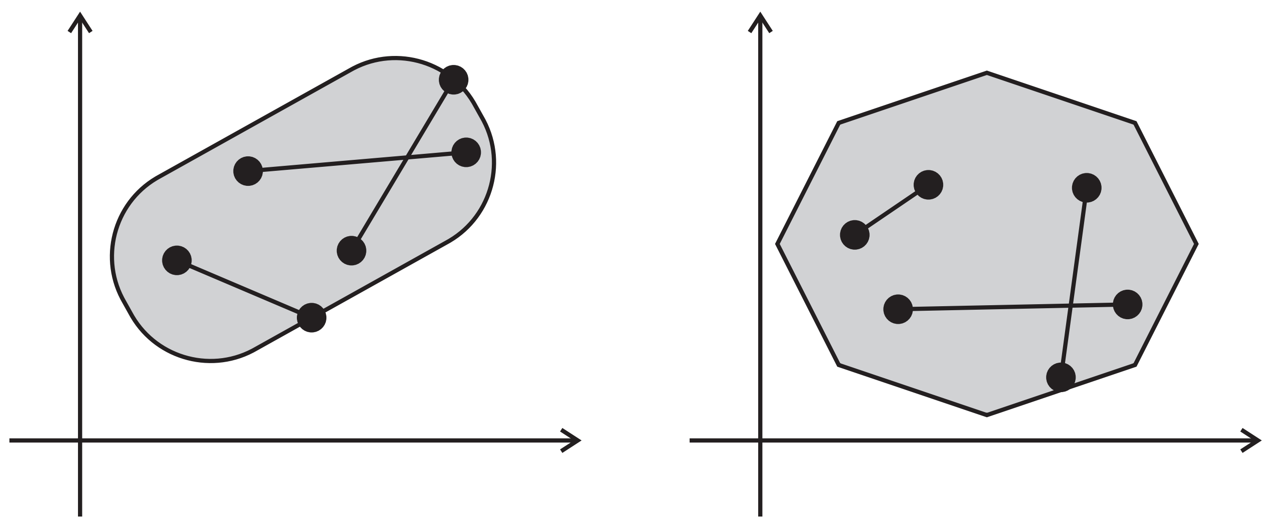
Convex sets

Definition (Convex set) Let V be a vector space over $\mathbb{R}$ (usually, $V = \mathbb{R}^n$, $V = \mathbb{R}^{n \times m}$, $V = \mathsf{S}^n$).

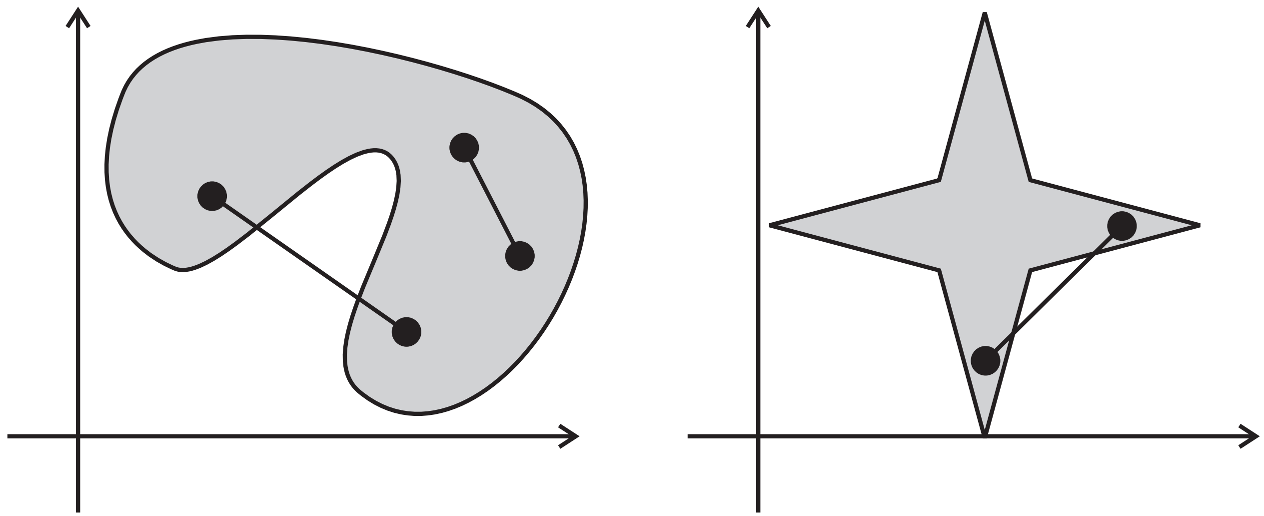
A set $S \subset V$ is convex if $x, y \in S \Rightarrow [x, y] \subset S$



Convex sets:



Nonconvex sets:



How do we recognize convex sets ?

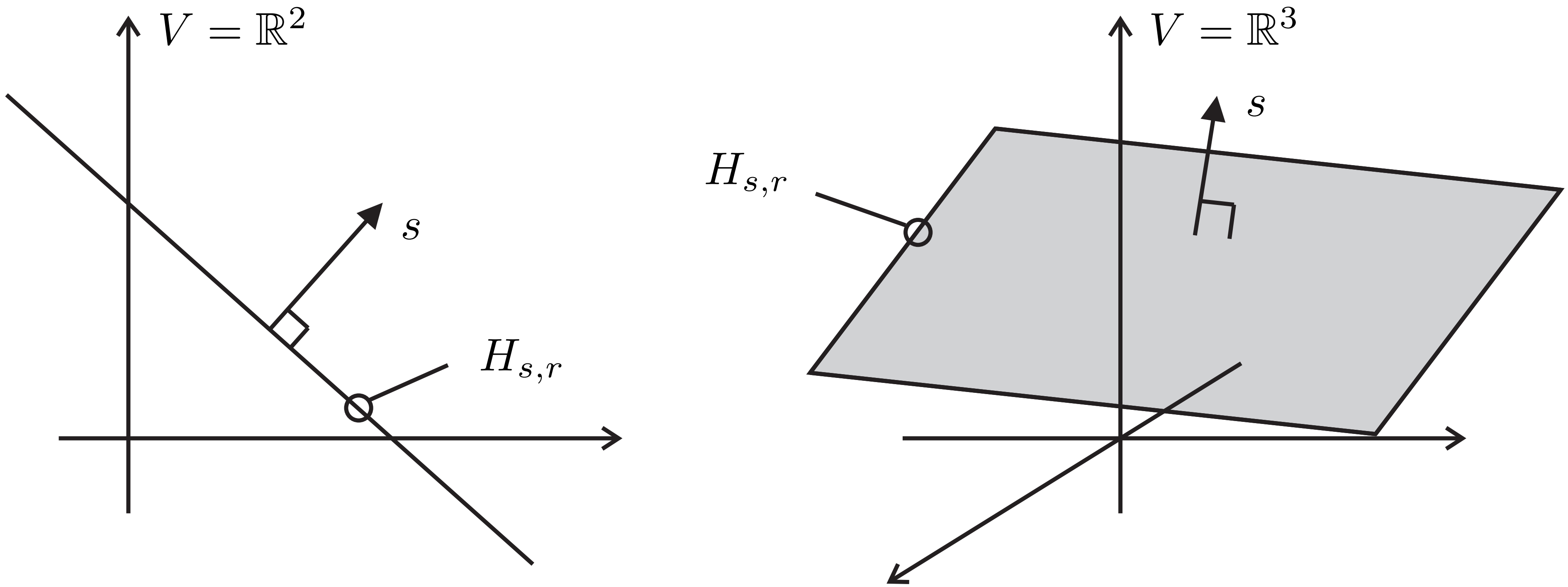
Vocabulary of simple ones

$$+$$

Apply convexity-preserving operations

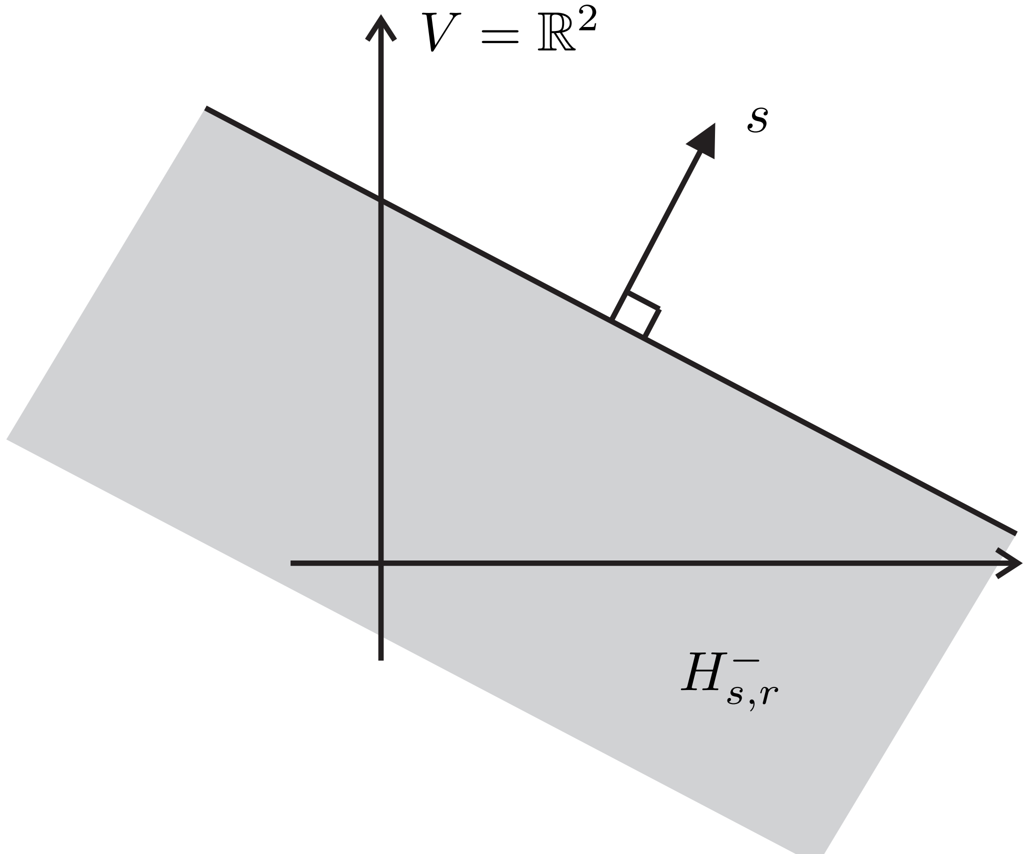
Example (Hyperplanes): for $s \in V - \{0\}$ and $r \in \mathbb{R}$

$$H_{s,r} := \{x \in V \;:\; \langle s, x \rangle = r\}$$



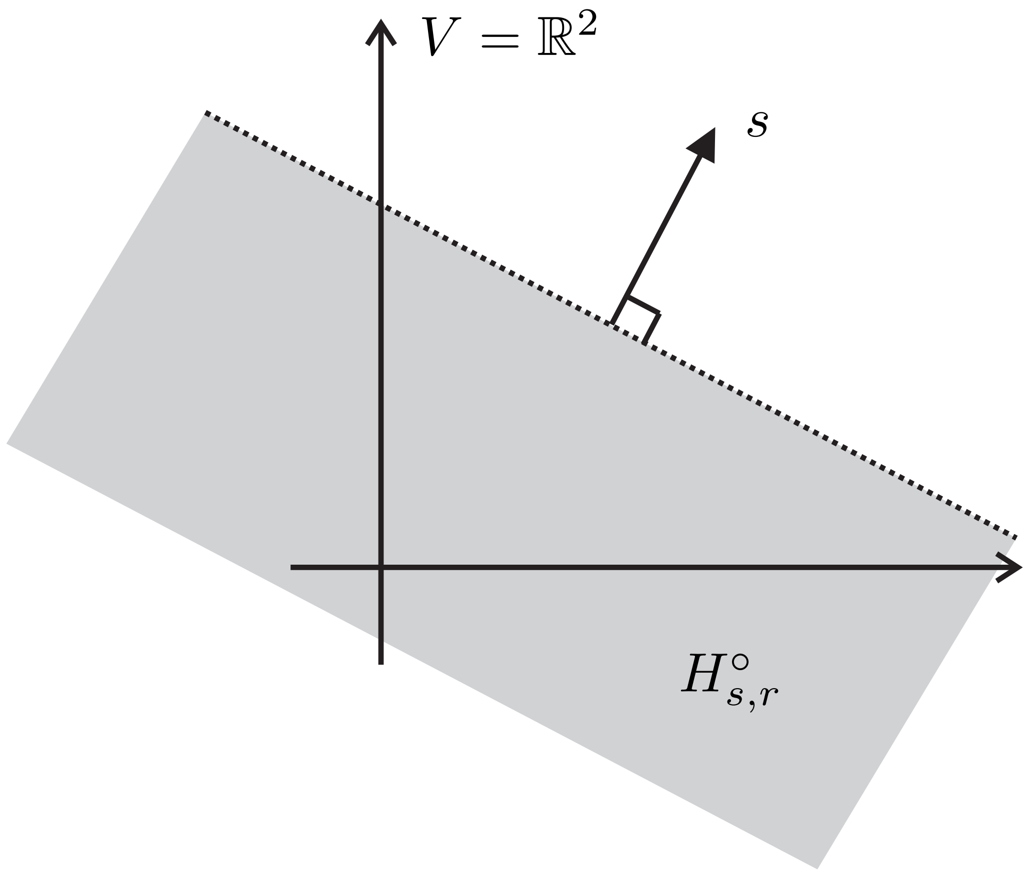
Example (Closed half-spaces): for $s \in V - \{0\}$ and $r \in \mathbb{R}$

$$H^-_{s,r} := \{x \in V \; : \; \langle s, x \rangle \leq r\}$$



Example (Open half-spaces): for $s \in V - \{0\}$ and $r \in \mathbb{R}$

$$H^{\circ}_{s,r} := \{x \in V \,:\, \langle s, x \rangle < r\}$$



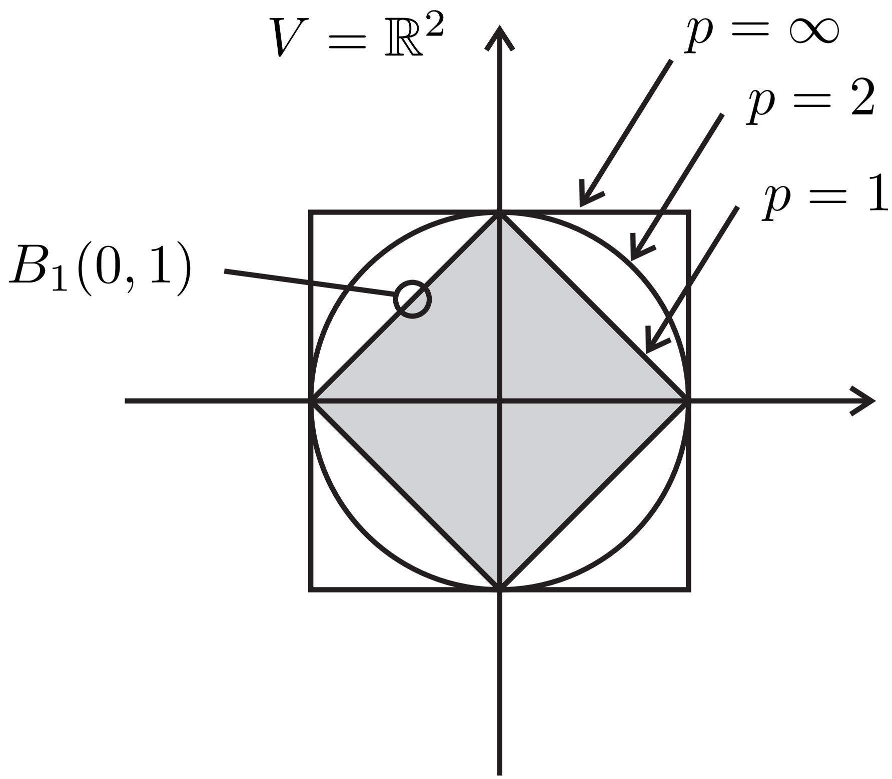
Example (norm balls):

$$B(c, R) = \{x \in V \ : \ \|x - c\| \leq R\}$$

where $\|\cdot\|$ denotes a norm on V

Special case: ℓ_p norms in $V = \mathbb{R}^n$

$$\|x\|_p = \begin{cases} \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} & \text{if } 1 \le p < \infty \\ \max_{i=1,\dots,n} |x_i| & \text{if } p = \infty \end{cases}$$



Special case: $\sigma_{\max}$ norm in $V = \mathbb{R}^{n\times m}$

SVD: for $A \in \mathbb{R}^{n\times m}$ $(n \geq m)$, there exist $U : n \times m$ $(U^\top U = I_m)$, $V : m \times m$ $(V^\top V = I_m)$ and

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_m \end{bmatrix} : m \times m$$

with $\sigma_{\max} = \sigma_1 \geq \sigma_2 \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_m = 0$ $(r = \text{rank}\, A)$ such that

$$A = U \Sigma V^\top$$

We also have

$$\sigma_{\max} = \sup\left\{u^\top Av \,:\, \|u\| = 1, \|v\| = 1\right\}$$

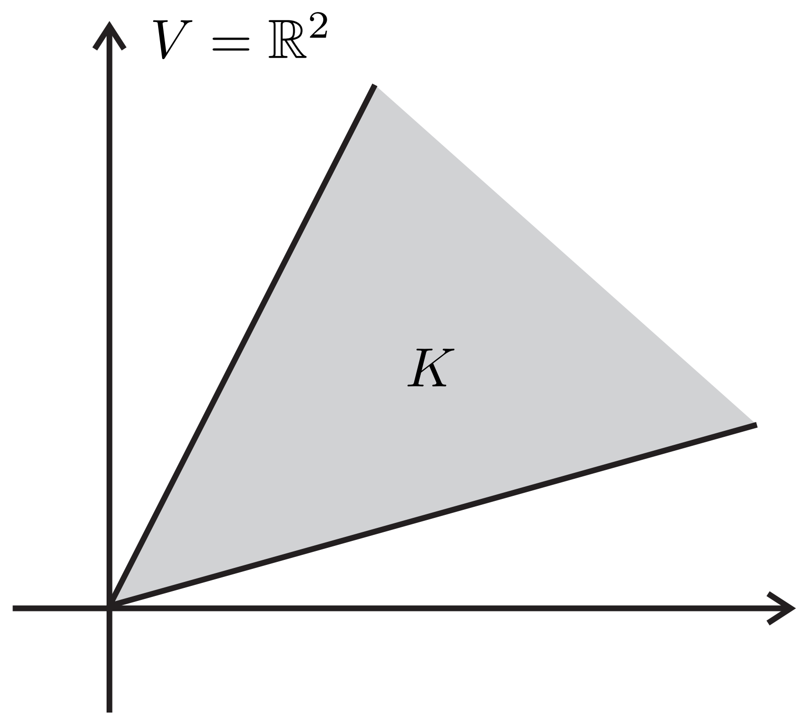
and

$$\|Ax\| \leq \sigma_{\max}(A)\,\|x\| \qquad \text{for all } x$$

Definition (Convex cone) Let V be a vector space.

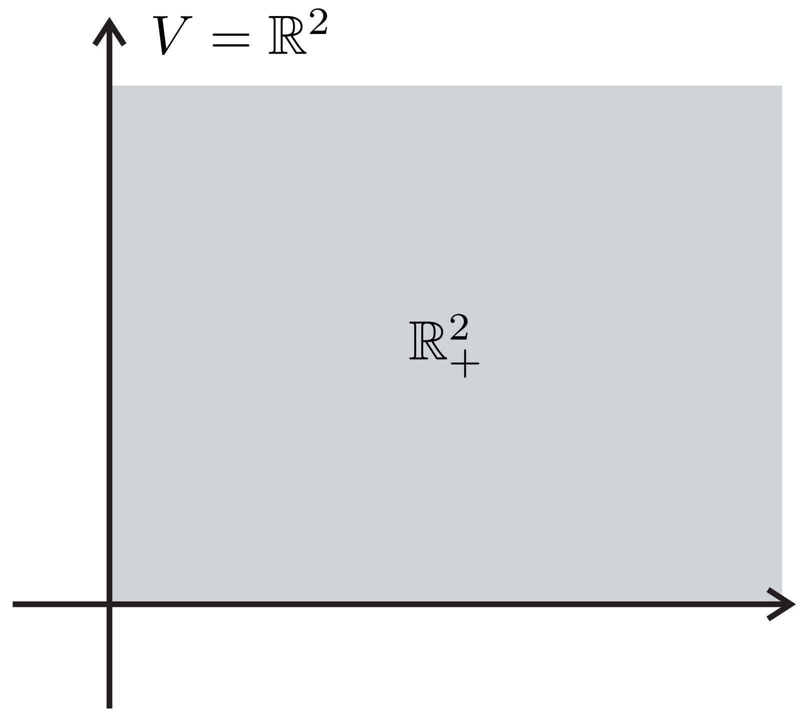
A set $K \subset V$ is a convex cone if K is convex and

$$\mathbb{R}_+ K = \{\alpha x \,:\, \alpha \geq 0, x \in K\} \subset K$$



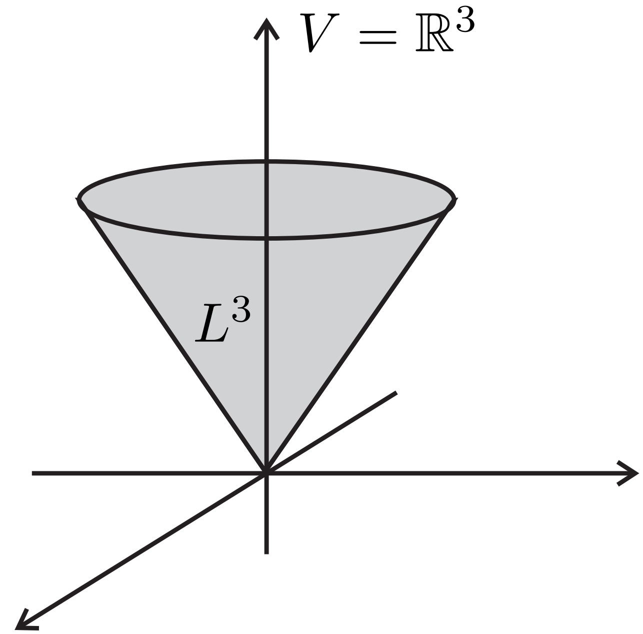
Example (Nonnegative orthant): with $V = \mathbb{R}^n$

$$\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \ : \ x \geq 0\}$$



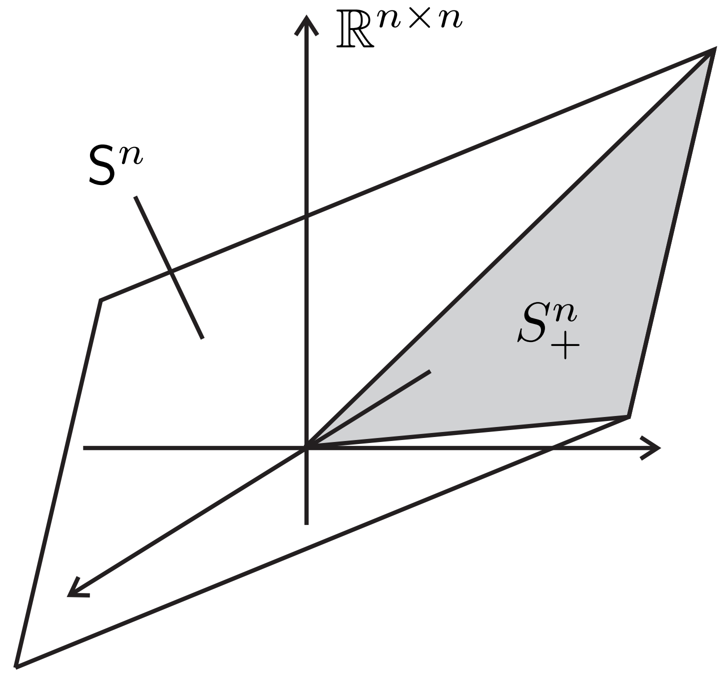
Example (2nd order cone): with $V = \mathbb{R}^{n+1}$

$$L^{n+1} = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \,:\, \|x\| \leq t\}$$



Example (Positive semidefinite cone): with $V = \mathsf{S}^n$

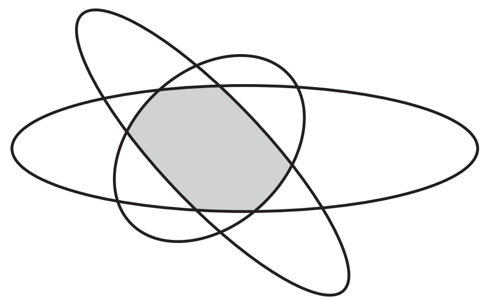
$$\mathsf{S}^n_+ = \{X \in \mathsf{S}^n \;:\; X \succeq 0\}$$



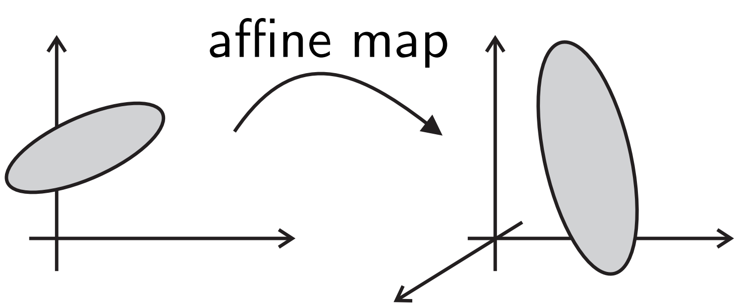
Note that $X \in S^n_+$ iff $\lambda(X) := (\lambda_{\min}(X), \lambda_2(X), \ldots, \lambda_{\max}(X)) \in \mathbb{R}^n_+$

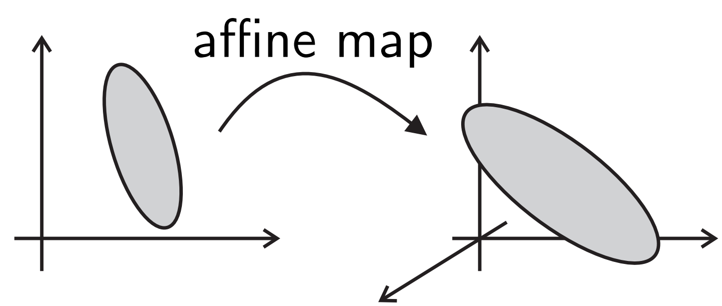
Convexity is preserved by:

- intersections



- affine images or inverse-images





Proposition Let $\{S_j : j \in J\}$ be a collection of convex subsets of the vector space V. Then,

$$\bigcap_{j \in J} S_j \text{ is convex.}$$

Note: the index set J might be uncountable

Proposition Let V, W be vector spaces and $A : V \to W$ be affine.

If $S \subset V$ is convex, then $A(S)$ is convex.

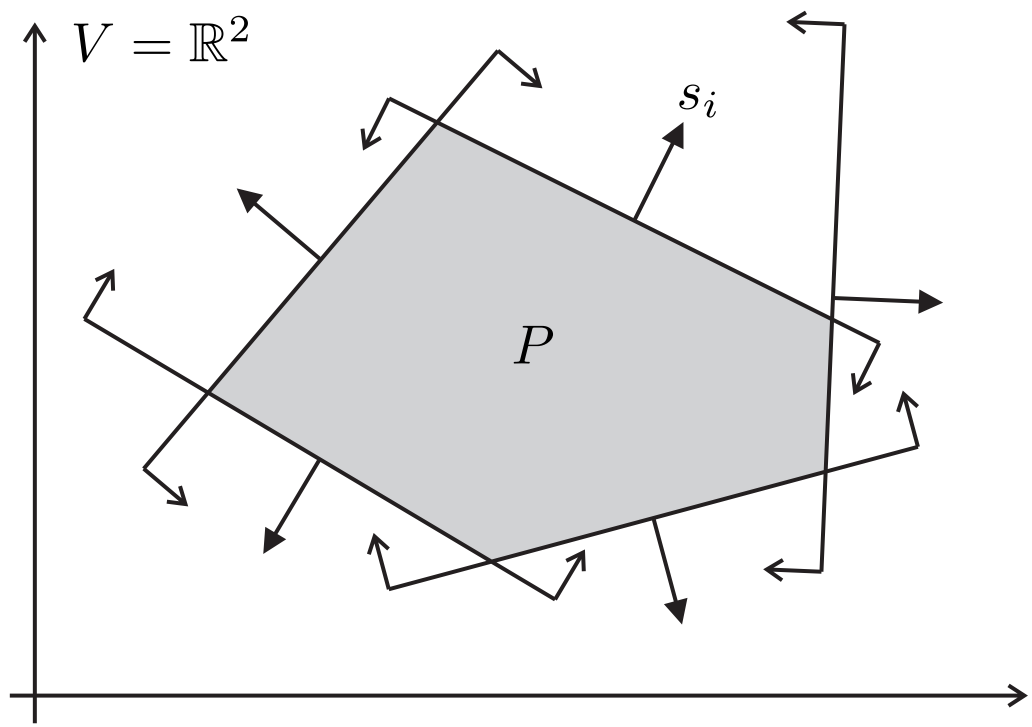
If $S \subset W$ is convex, then $A^{-1}(S)$ is convex

More examples of convex sets:

- Polyhedrons
- Polynomial separators for pattern recognition
- Sets of pre-filter signals
- Ellipsoids
- Linear Matrix Inequalities (LMIs)
- $V = \mathsf{S}^n$ and $S = \{X \in \mathsf{S}^n \,:\, \lambda_{\max}(X) \leq 1\}$
- Contraction matrices

Example (Polyhedrons):

$$P = \{x \in V \;:\; \langle s_i, x\rangle \leq r_i \;:\; i = 1, 2, \ldots, m\} = \bigcap_{i=1}^{m} H^{-}_{s_i, r_i}$$



Example (Polynomials separators): consider two sets in $\mathbb{R}^n$

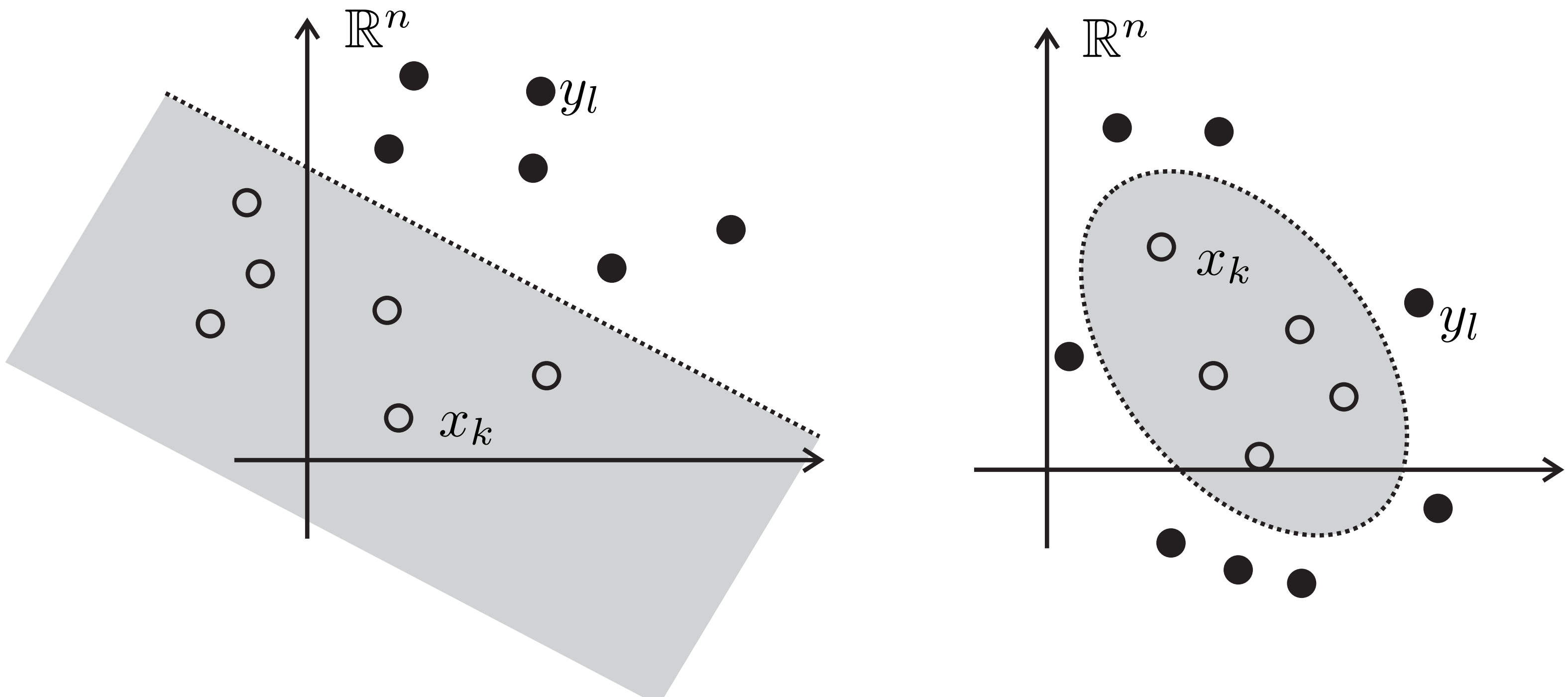
$$\mathcal{X} = \{x_1, x_2, \ldots, x_K\} \qquad \mathcal{Y} = \{y_1, y_2, \ldots, y_K\}$$

A separator is a function $f : \mathbb{R}^n \to \mathbb{R}$ such that

$$\begin{cases} f(x_k) > 0 \text{ for all } x_k \in \mathcal{X} \\ f(y_l) < 0 \text{ for all } y_l \in \mathcal{Y} \end{cases}$$

Some classes of separators:

- (linear) $f_{s,r}(x) = \langle s, x \rangle + r$
- (quadratic) $f_{A,s,r}(x) = \langle Ax, x \rangle + \langle s, x \rangle + r$



$\mathbb{R}^n$
y_l
x_k
$\mathbb{R}^n$
x_k
y_l

The set of linear separators

$$S = \{(s, r) \in \mathbb{R}^n \times \mathbb{R} \, : \, f_{s,r} \text{ is a separator}\}$$

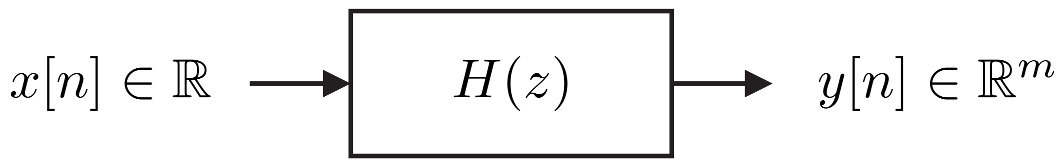
is convex

The set of quadratic separators

$$S = \{(A, s, r) \in \mathsf{S}^n \times \mathbb{R}^n \times \mathbb{R} \, : \, f_{A,s,r} \text{ is a separator}\}$$

is convex

Example (Pre-filter signals): a single-input multiple-output channel



with finite impulse-response (FIR)

$$H(z) = h_0 + h_1 z^{-1} + h_2 z^{-2} + \cdots + h_d z^{-d}$$

Input-output equation:

$$y[n] = h_0 x[n] + h_1 x[n-1] + h_2 x[n-2] + \cdots + h_d x[n-d]$$

Stacking N output samples (zero initial conditions):

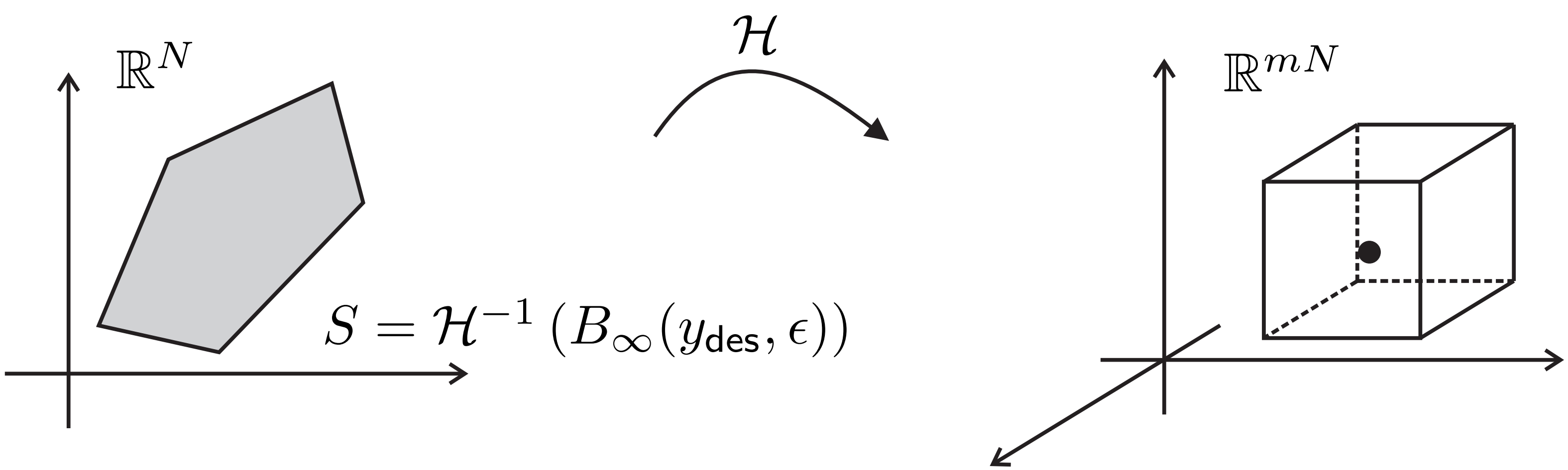
$$\underbrace{\begin{bmatrix} y[1] \\ y[2] \\ \vdots \\ y[N] \end{bmatrix}}_{y \in \mathbb{R}^{mN}} = \underbrace{\begin{bmatrix} h_0 & 0 & \cdots & 0 \\ h_1 & h_0 & \ddots & \vdots \\ \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & h_1 & h_0 \end{bmatrix}}_{\mathcal{H} \in \mathbb{R}^{mN \times N}} \underbrace{\begin{bmatrix} x[1] \\ x[2] \\ \vdots \\ x[N] \end{bmatrix}}_{x \in \mathbb{R}^{N}}$$

A desired output signal $y_{\mathsf{des}} \in \mathbb{R}^{mN}$ and tolerance $\epsilon > 0$ are given

Set of inputs fulfilling the specs

$$S = \{x \in \mathbb{R}^N \ : \ \|\mathcal{H}x - y_{\mathsf{des}}\|_\infty \leq \epsilon\}$$

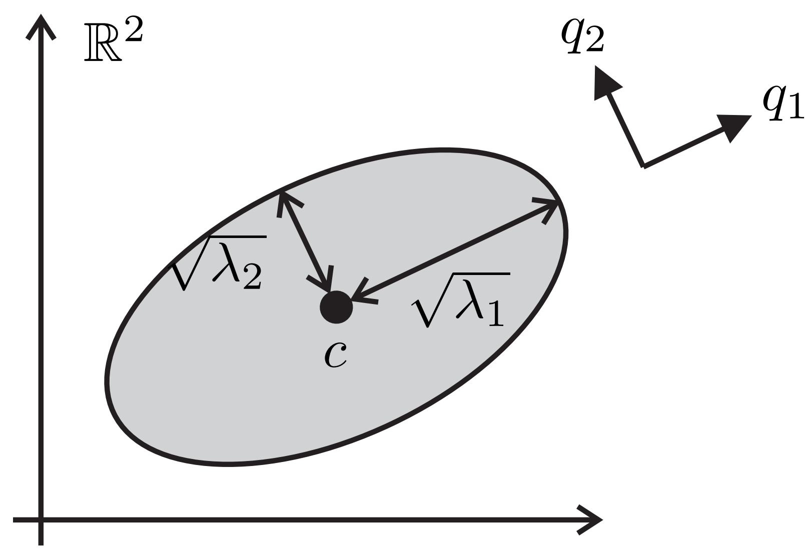
is convex



Example (Ellipsoids):

$$E(A, c) = \{x \in \mathbb{R}^n \,:\, (x-c)^\top A^{-1}(x-c) \le 1\} \qquad (A \succ 0)$$

$$A = Q\Lambda Q^\top \quad Q = \begin{bmatrix} q_1 & q_2 \end{bmatrix} \quad \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



Special case $A = R^2 I_n$ correspond to the ball $B(c, R)$

The function

$$\|\cdot\|_{A^{-1}} : \mathbb{R}^n \to \mathbb{R} \qquad x \mapsto \|x\|_{A^{-1}} = \sqrt{x^\top A^{-1} x}$$

is a norm (note that $\|x\|_{A^{-1}} = \left\|A^{-1/2} x\right\|$)

The ellipsoid

$$E(A, c) = \{x \in \mathbb{R}^n : \|x - c\|_{A^{-1}} \leq 1\}$$

is convex (norm ball)

Example (Linear Matrix Inequalities):

$$S = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \, : \, A_0 + x_1 A_1 + x_2 A_2 + \cdots + x_n A_n \succeq 0\}$$

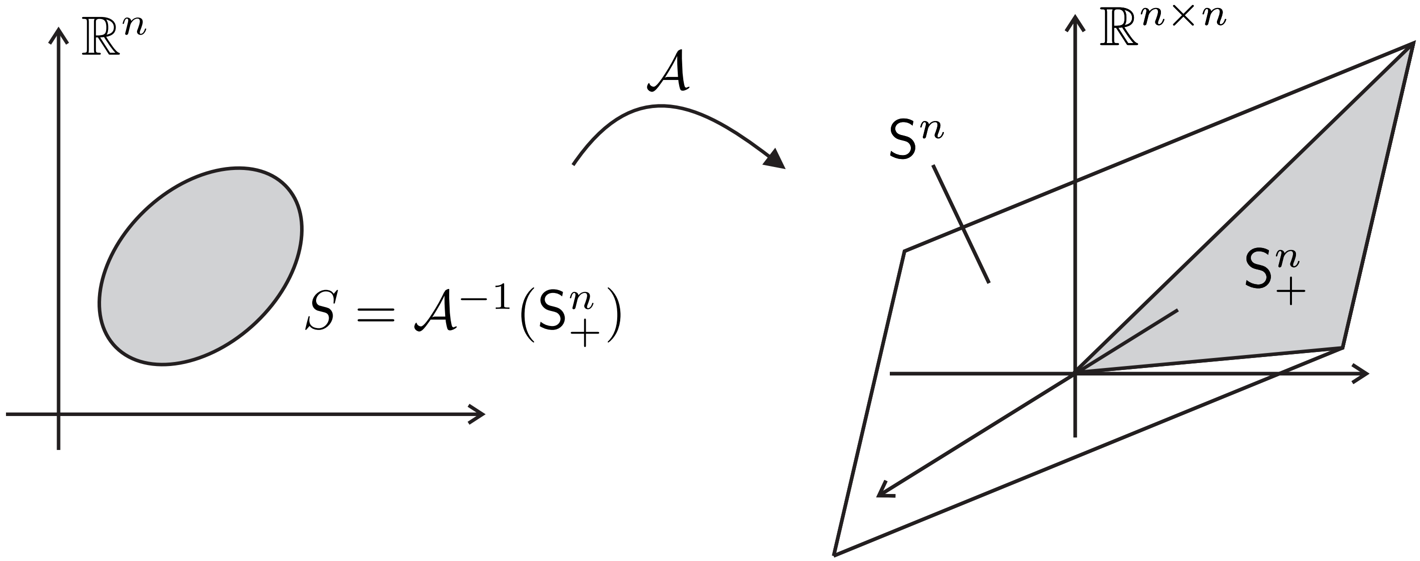
where $A_0, A_1, \ldots, A_n$ are given $m \times m$ symmetric matrices

The set S is convex because

$$S = \mathcal{A}^{-1}\left(\mathsf{S}^n_+\right)$$

where

$$\mathcal{A} : \mathbb{R}^n \to \mathsf{S}^n \qquad x \mapsto \mathcal{A}(x) = A_0 + x_1 A_1 + x_2 A_2 + \cdots + x_n A_n$$



Example: $V = \mathsf{S}^n$ and $S = \{X \in \mathsf{S}^n \,:\, \lambda_{\max}(X) \le 1\}$

The set S is convex:

$$S = \bigcap_{u \in \mathbb{R}^n \,:\, \|u\|=1} \underbrace{\{X \in \mathsf{S}^n \,:\, u^\top X u \le 1\}}_{S_u \;:\,\text{convex (closed half-space)}}$$

Example (Contraction matrices): $S = \{X \in \mathbb{R}^{n \times m} \,:\, \sigma_{\max}(X) \leq 1\}$

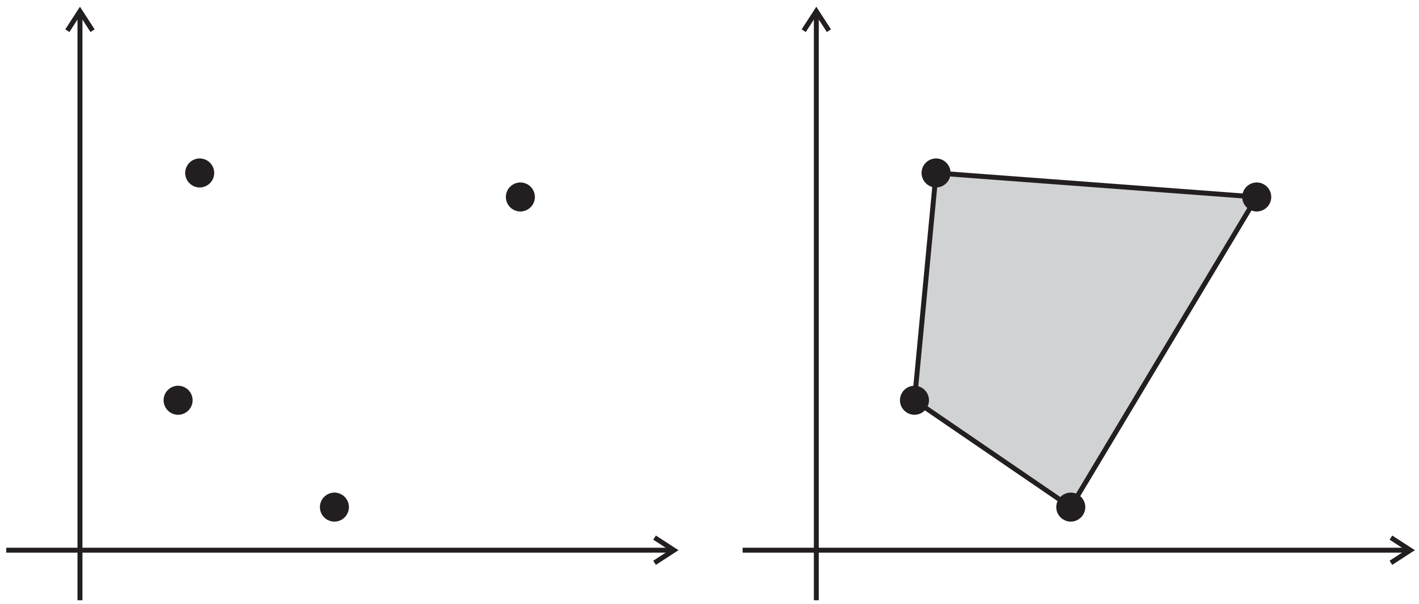
The set S is convex:

$$S = \bigcap_{u \in \mathbb{R}^n \,:\, \|u\|=1, v \in \mathbb{R}^m \,:\, \|v\|=1} \underbrace{\left\{X \in \mathbb{R}^{n \times m} \,:\, u^\top X v \leq 1\right\}}_{S_{u,v} \ \text{: convex (closed half-space)}}$$

Definition (Convex hull) The convex hull of a set S, written co S, is the smallest convex set containing S:

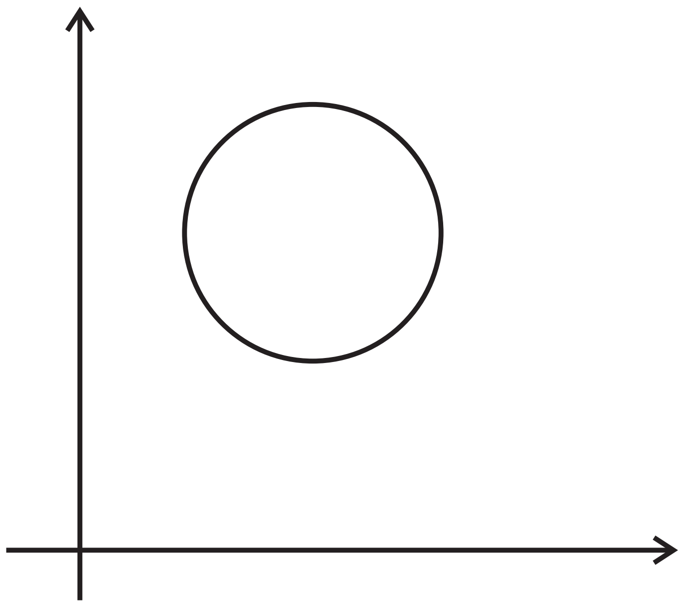
$$\mathsf{co}\, S = \left\{ \sum_{i=1}^{k} \alpha_i x_i \;:\; x_i \in S,\, \alpha_i \geq 0,\, \sum_{i=1}^{k} \alpha_i = 1, k = 1, 2, \ldots \right\}$$

Example:

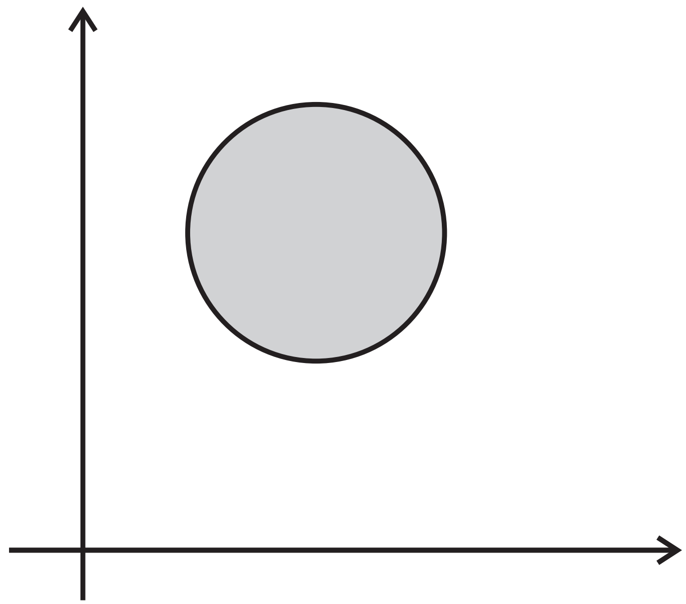


S $\quad$ $\operatorname{co} S$

Example:

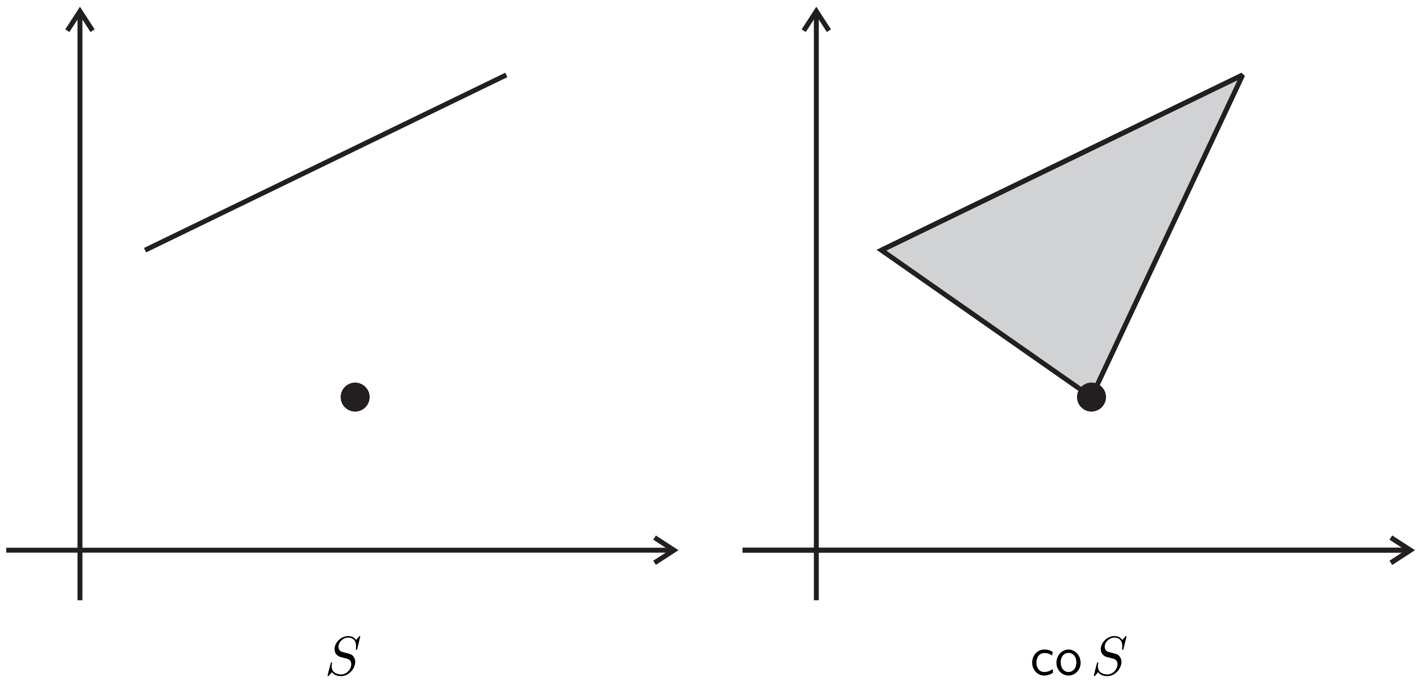


S



co S

Example:



Example (ℓ_p unit-sphere):

$$S = \{x \in \mathbb{R}^n \,:\, \|x\|_p = 1\} \quad \Rightarrow \quad \mathsf{co}\, S = \{x \in \mathbb{R}^n \,:\, \|x\|_p \le 1\}$$

Example (binary vectors):

$$S = \{(\pm 1, \pm 1, \ldots, \pm 1)\} \subset \mathbb{R}^n \quad \Rightarrow \quad \mathsf{co}\, S = B_\infty(0,1)$$

Example (unit vectors):

$$S = \{(0, \ldots, 1, \ldots, 0)\} \subset \mathbb{R}^n \quad \Rightarrow \quad \mathsf{co}\, S = \{x \in \mathbb{R}^n \,:\, x \ge 0, 1^\top x = 1\}$$

co S is called the unit simplex

Example (Stiefel matrices):

$$S = \{X \in \mathbb{R}^{m \times n} \,:\, X^\top X = I_n\} \quad \Rightarrow \quad \mathsf{co}\, S = \{X \,:\, \sigma_{\max}(X) \leq 1\}$$

Example (permutation matrices):

$$S = \{X \,:\, X_{ij} \in \{0,1\}, \sum_{i=1}^{n} X_{ij} = 1, \sum_{j=1} X_{ij} = 1\} \subset \mathbb{R}^{n \times n} \quad \Rightarrow$$

$$\mathsf{co}\, S = \{X \,:\, X \geq 0, \sum_{i=1}^{n} X_{ij} = 1, \sum_{j=1} X_{ij} = 1\}$$

co S is the set of doubly-stochastic matrices (this is Birkhoff's theorem)

Theorem (Projection onto closed convex sets) Let V be a vector space with inner-product $\langle \cdot, \cdot \rangle$ and induced norm $\|x\| = \sqrt{\langle x, x \rangle}$. Let $S \subset V$ be a non-empty closed convex set and $x \in V$. Then,

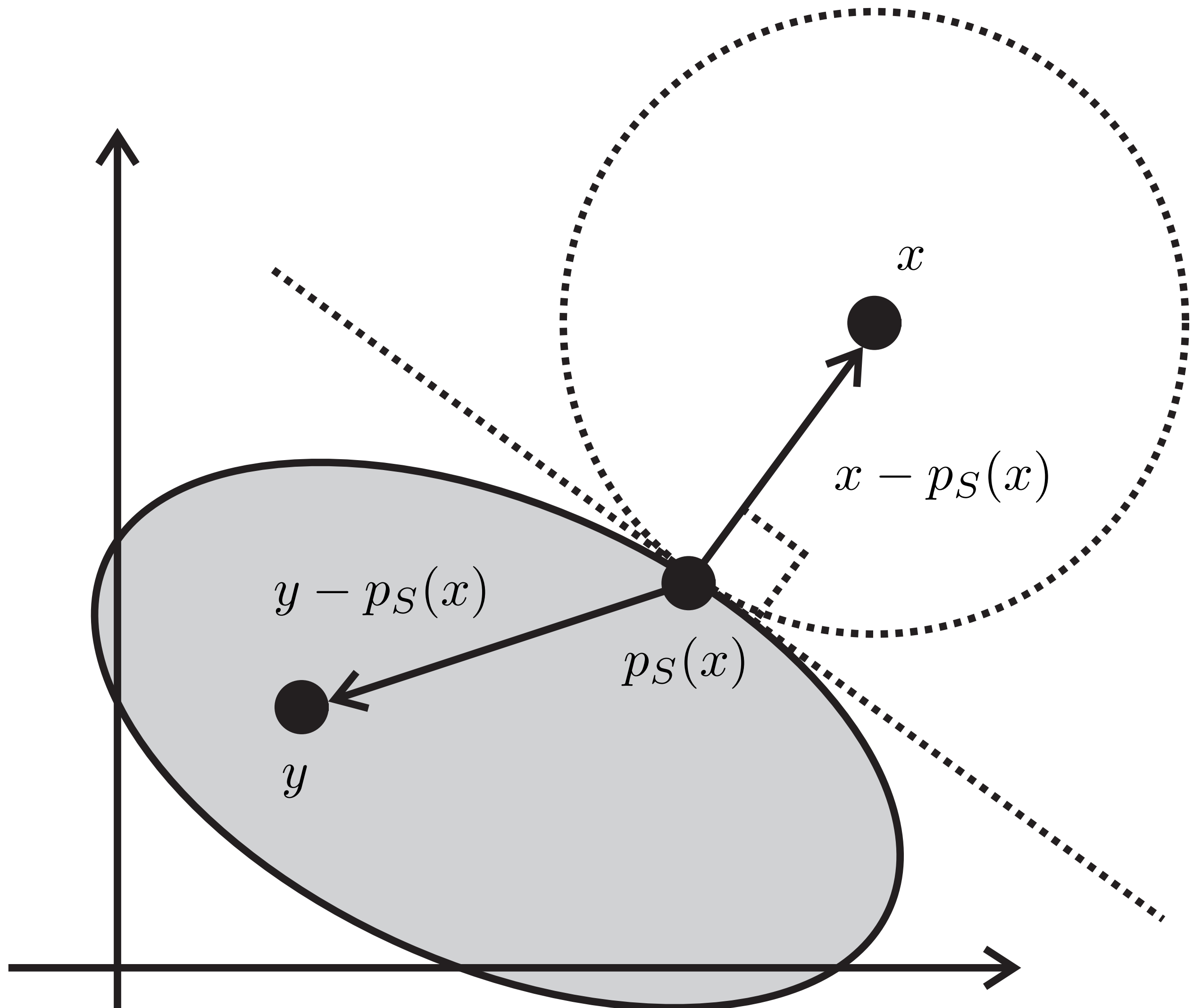
- **(Existence and uniqueness)** There is an unique $\widehat{x} \in S$ such that

  $$\|x - \widehat{x}\| \leq \|x - y\| \text{ for all } y \in S$$

  $\widehat{x}$ is called the projection of x onto S and is denoted by $p_S(x)$

- **(Characterization)** A point $\widehat{x} \in S$ is the projection $p_S(x)$ if and only if

  $$\langle x - \widehat{x}, y - \widehat{x} \rangle \leq 0 \text{ for all } y \in S.$$



x
$x - p_S(x)$
$y - p_S(x)$
$p_S(x)$
y

Inner-product induced norm is necessary:

$$S = \{(u, v) \,:\, v \leq 0\} \quad x = (0, 1) \quad \|\cdot\| = \|\cdot\|_\infty$$

Convexity is necessary:

$$S = [-2, -1] \cup [1, 2] \quad x = 0$$

Closedness is necessary:

$$S =]1, 2] \quad x = 0$$

Examples of projections:

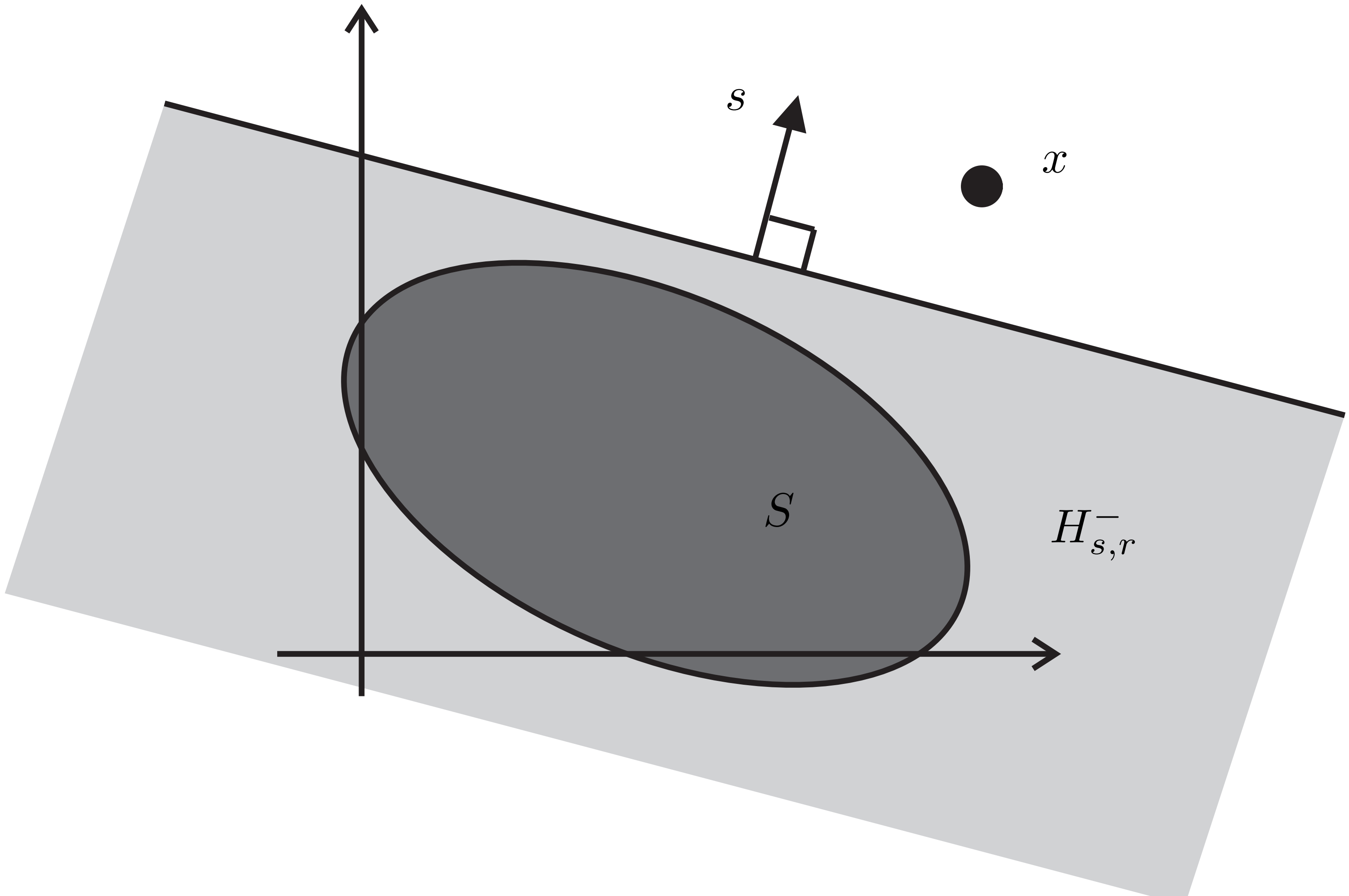
- $V = \mathbb{R}^n$, $S = \mathbb{R}^n_+$, $p_S(x) = x^+$
- $V = \mathsf{S}^n$, $S = \mathsf{S}^n_+$, $p_S(X) = X^+$

Corollary (Hyperplane separation) Let V be vector space with inner-product $\langle \cdot, \cdot \rangle$. Let $S \subset V$ be a non-empty closed convex set and $x \notin S$. Then, there exists an hyperplane $H_{s,r}$ $(s \in V - \{0\}, r \in \mathbb{R})$ that strictly separates x from S:

$$S \subset H_{s,r}^{-} \text{ and } x \notin H_{s,r}^{-}.$$

That is,

$$\langle s, y \rangle \leq r \text{ for all } y \in S \quad \text{and} \quad \langle s, x \rangle > r.$$



s
x
S
$H^{-}_{s,r}$

Proposition. Let $A \in \mathbb{R}^{n \times m}$. The set

$$A\mathbb{R}^m_+ = \{Ax \,:\, x \geq 0\} \subset \mathbb{R}^n$$

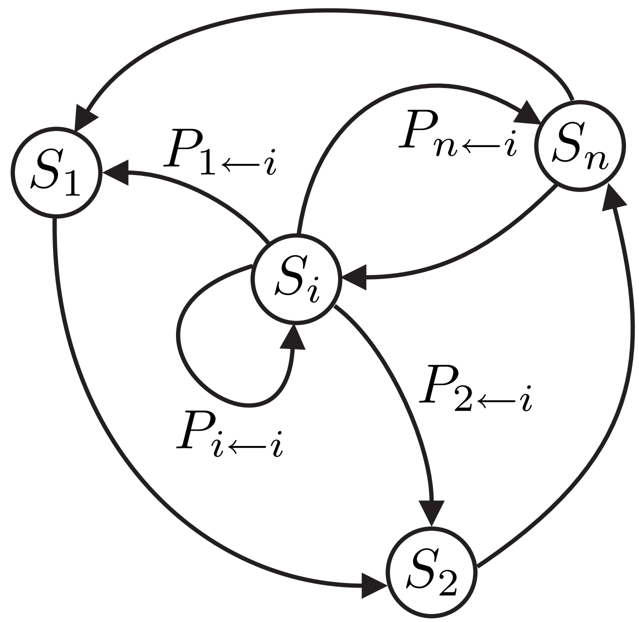
is closed

Lemma (Farkas). Let $A \in \mathbb{R}^{m \times n}$ and $c \in \mathbb{R}^n$. One and only one of the following sets is non-empty:

$$S_1 = \{x \in \mathbb{R}^n \,:\, Ax \leq 0,\, c^\top x > 0\} \qquad S_2 = \{y \in \mathbb{R}^m \,:\, c = A^\top y,\, y \geq 0\}$$

Application: stationary distributions for Markov chains

- n states: $\mathcal{S} = \{S_1, S_2, \dots, S_n\}$
- $X_t \in \mathcal{S}$ $(t = 0, 1, 2, \dots)$ state at time t
- $P_{j \leftarrow i} = \text{Prob}(X_{t+1} = S_j \mid X_t = S_i)$



- $p(t) := \begin{bmatrix} \text{Prob}(X_t = S_1) & \text{Prob}(X_t = S_2) & \cdots & \text{Prob}(X_t = S_n) \end{bmatrix}^\top$

$p(t)$ is a probability mass function (pmf):

$$p(t) \geq 0 \quad 1^\top p(t) = 1$$

for all t

Dynamics of $p(t)$:

$$p(t+1) = \underbrace{\begin{bmatrix} P_{1\leftarrow 1} & P_{1\leftarrow 2} & \cdots & P_{1\leftarrow n} \\ P_{2\leftarrow 1} & P_{2\leftarrow 2} & \cdots & P_{2\leftarrow n} \\ \vdots & \vdots & \cdots & \vdots \\ P_{n\leftarrow 1} & P_{n\leftarrow 2} & \cdots & P_{n\leftarrow n} \end{bmatrix}}_{P\,:\,\text{Transition matrix}} p(t)$$

A pmf $p^\star$ is said to be stationary if $p^\star = Pp^\star$

Result: any Markov chain has a stationary distribution

Definition (Proper cone) A convex cone K is proper if it is:

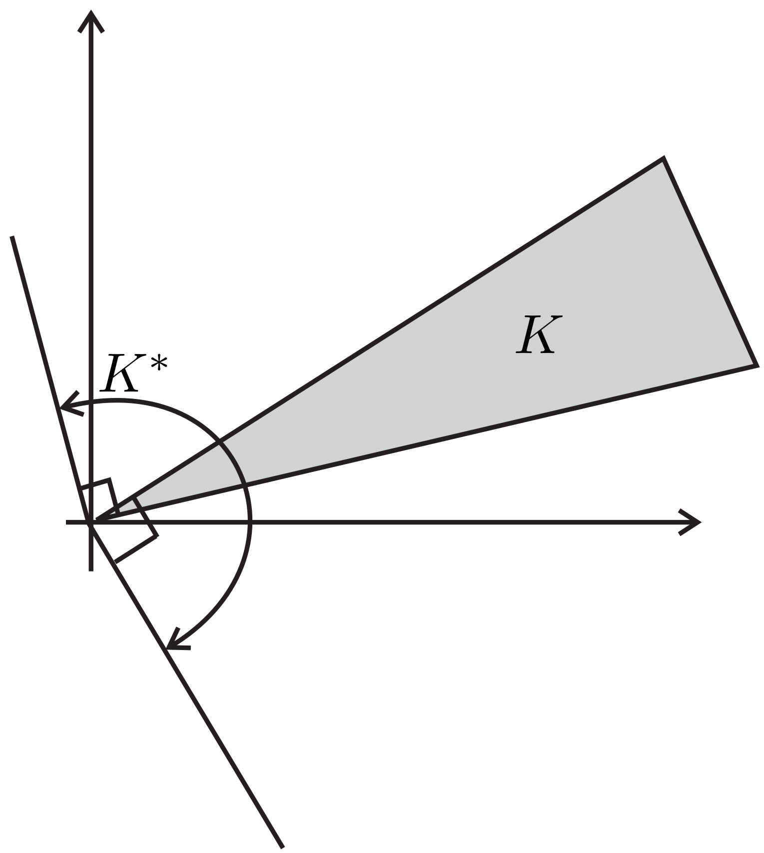
- closed
- solid ($\text{int}\, K \neq \emptyset$)
- pointed ($x, -x \in K \Leftrightarrow x = 0$)

Examples:

- non-negative orthant $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \,:\, x \geq 0\}$
- 2nd order cone $L^{n+1} = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \,:\, \|x\| \leq t\}$
- positive semidefinite cone $\mathsf{S}^n_+ = \{X \in \mathsf{S}^n \,:\, X \succeq 0\}$

Definition (Dual cone) Let K be a convex cone in the vector space V with inner-product $\langle \cdot, \cdot \rangle$. The dual cone of K is

$$K^* = \{s \in V \ : \ \langle s, x \rangle \geq 0 \text{ for all } x \in K\}$$



Dual cones are important for constructing dual programs

Examples:

- $K = L$ (linear subspace)

  $K^* = L^\perp$

- $K = \text{cone}\,(a_1, \ldots, a_m) = \{\sum_{i=1}^m \mu_i a_i \;:\; \mu_i \geq 0\}$

  $K^* = \{s \;:\; \langle s, a_i \rangle \geq 0\}$

- $K = \{x \;:\; \langle a_i, x \rangle \geq 0 \text{ for } i = 1, \ldots, m\}$

  $K^* = \text{cone}\,(a_1, \ldots, a_m)$

- $K = \{x \in \mathbb{R}^n \;:\; x_1 \geq \cdots \geq x_n \geq 0\}$

  $K^* = \{s \in \mathbb{R}^n \;:\; \sum_{j=1}^i s_j \geq 0, \; \forall_i\}$

Definition (Self-dual cones) A convex cone K is said to be self-dual if $K^* = K$

Examples of self-dual cones:

- non-negative orthant $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \,:\, x \geq 0\}$
- 2nd order cone $L^{n+1} = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} \,:\, \|x\| \leq t\}$
- positive semidefinite cone $\mathsf{S}^n_+ = \{X \in \mathsf{S}^n \,:\, X \succeq 0\}$

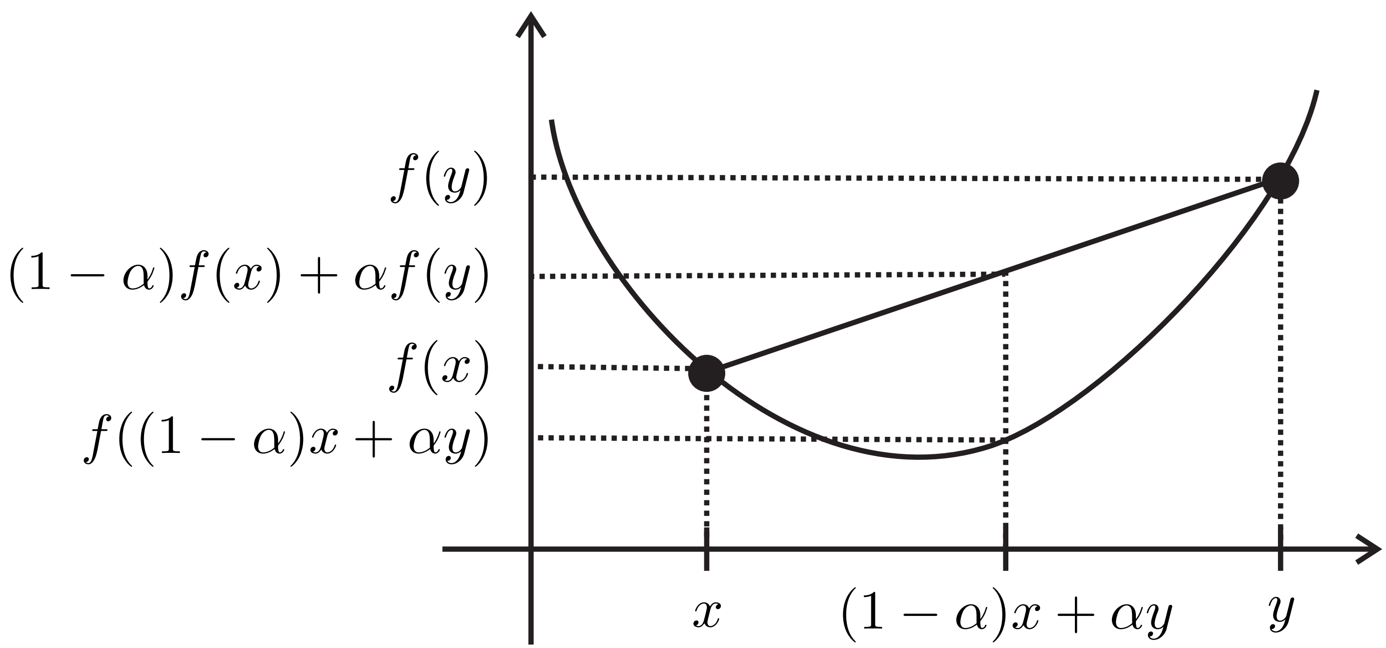
Convex functions

Definition (Convex function) Let V be a vector space over $\mathbb{R}$.

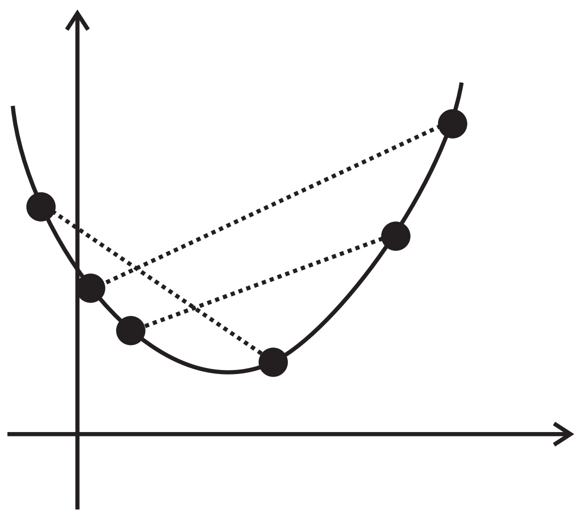
$f : S \to \mathbb{R}$ is convex if $S \subset V$ is convex and

$$f((1-\alpha)x + \alpha y) \leq (1-\alpha)f(x) + \alpha f(y)$$

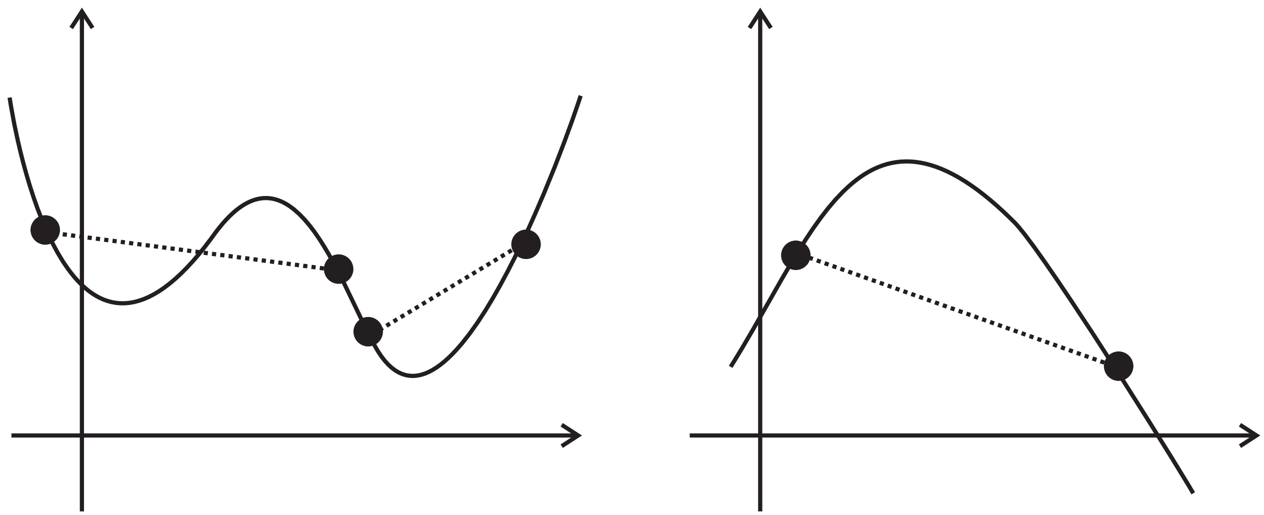
for all $x, y \in S$ and $\alpha \in [0, 1]$



Convex functions:



Nonconvex functions:



Proposition (Convexity is a $1D$ property) $f : S \subset V \to \mathbb{R}$ is convex iff for all $x \in S$ and $d \in V$

$$\phi_{x,d} : I_{x,d} := \{t \in \mathbb{R} : x + td \in S\} \subset \mathbb{R} \to \mathbb{R} \quad \phi(t) = f(x + td)$$

is convex

How do we recognize convex functions ?

Vocabulary of simple ones (definition, differentiability, epigraph)

$+$

Apply convexity-preserving operations

Example (Affine functions):

$$f(x) = \langle s, x \rangle + r$$

Example (norms):

$$f(x) = \|x\|$$

Special cases: ℓ_p norms, Frobenius norm, spectral norm, ...

Proposition (Convexity through differentiability) Let V be a vector space over $\mathbb{R}$ and $S \subset V$ an open convex set.

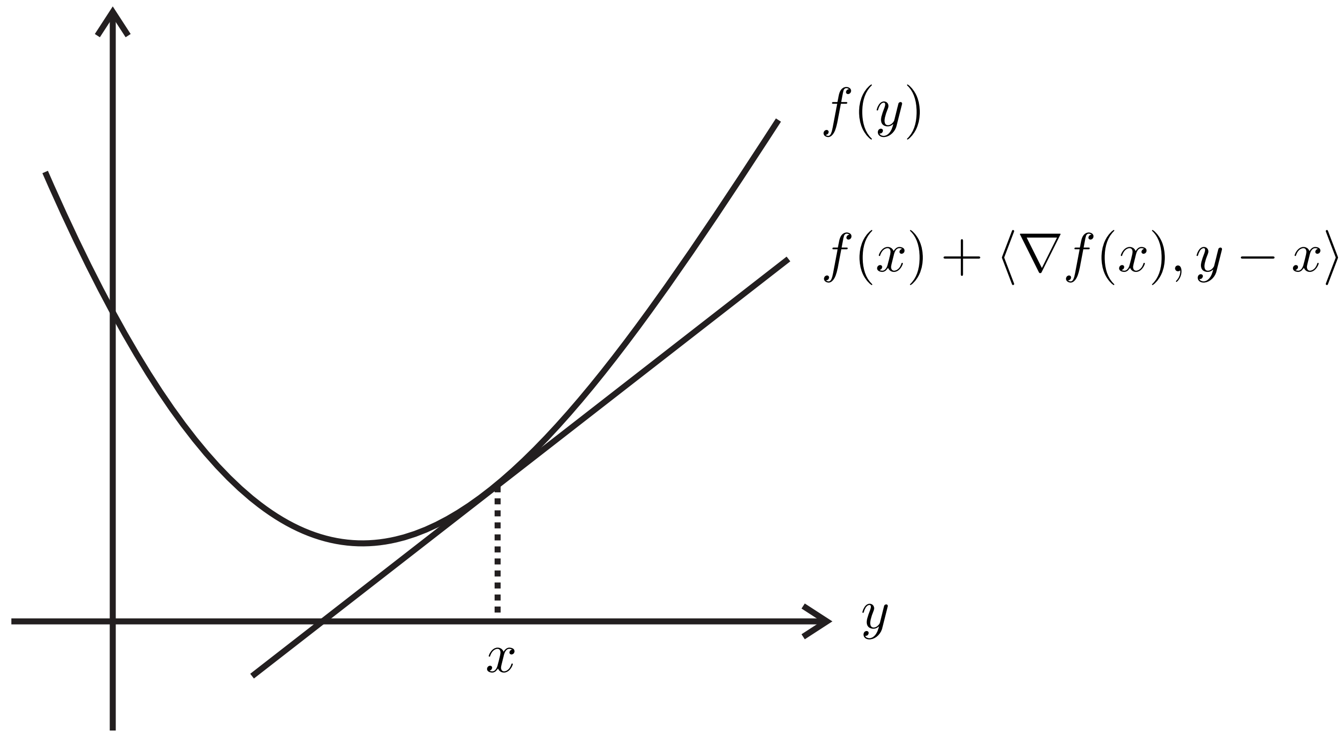
(1st order criterion) $f : S \to \mathbb{R}$ is convex iff

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \quad \text{for all } x, y \in S$$

(2nd order criterion) $f : S \to \mathbb{R}$ is convex iff

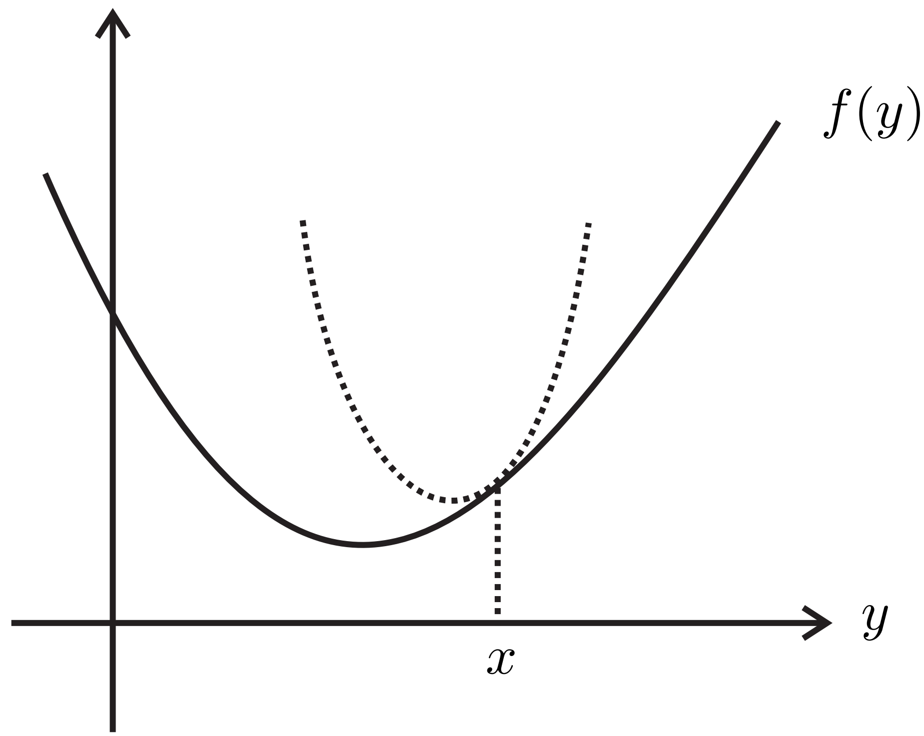
$$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in S$$

Geometrical interpretation of 1st order criterion:



Local information $f(x), \nabla f(x)$ yields a **global** bound

Geometrical interpretation of 2nd order criterion:



The graph of f is "pointing upwards"

Examples:

- $f : \mathbb{R}_{++} \to \mathbb{R}$

$$f(x) = \frac{1}{x}$$

- $f : \mathbb{R}_{++} \to \mathbb{R}$

$$f(x) = -\log(x)$$

- $f : \mathbb{R}_{+} \to \mathbb{R}$

$$f(x) = x\log(x) \qquad (0\log(0) := 0)$$

Examples:

- $f : \mathbb{R}^n \to \mathbb{R}$ (A symmetric)

$$f(x) = x^\top Ax + b^\top x + c$$

is convex iff $A \succeq 0$

- $f : \mathbb{R}^n \to \mathbb{R}$

$$f(x_1, \ldots, x_n) = \frac{1}{\gamma} \log\left(e^{\gamma x_1} + e^{\gamma x_2} + \cdots + e^{\gamma x_n}\right) \quad (\gamma > 0)$$

smooth approximation for non-smooth function $\max\{x_1, x_2, \ldots, x_n\}$

- $f : \mathbb{R}^n_{++} \to \mathbb{R}$

$$f(x_1, \ldots, x_n) = (x_1 \cdots x_n)^{1/n}$$

is concave

Fact: for any $A \succ 0$ and symmetric B, there exists S such that

$$A = SS^\top \quad B = S\Lambda S^\top$$

where Λ is diagonal (contains the eigenvalues of $A^{-1/2}BA^{-1/2}$)

- $f : \mathsf{S}^n_{++} \to \mathbb{R}$

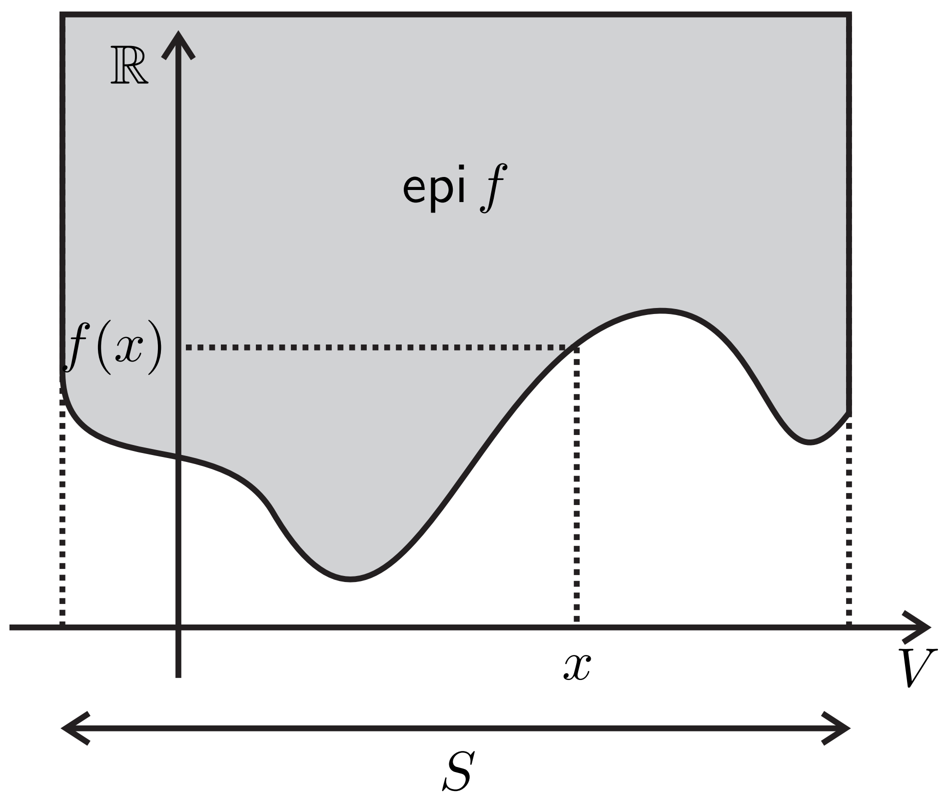
$$f(X) = \mathsf{tr}\left(X^{-1}\right)$$

- $f : \mathsf{S}^n_{++} \to \mathbb{R}$

$$f(X) = -\log\det(X)$$

Definition (Epigraph) Let V be a vector space. The epigraph of $f : S \subset V \to \mathbb{R}$ is the set

$$\text{epi}\, f = \{(x,t) : f(x) \le t, x \in S\} \subset V \times \mathbb{R}$$



Proposition (Convexity through epigraph)

$$f \text{ is a convex function} \quad \Leftrightarrow \quad \text{epi}\, f \text{ is a convex set}$$

Example: $\lambda_{\max} : \mathsf{S}^n \to \mathbb{R}$

Operations that preserve convexity:

- $f_1, \ldots, f_n$ are cvx

$$\oplus f_1 + \oplus f_2 + \cdots + \oplus f_n \text{ is cvx}$$

- f is cvx

$$(f \circ \text{affine}) \text{ is cvx}$$

- f_j $(j \in J)$ are cvx

$$\sup_{j \in J} f_j \text{ is cvx}$$

Proposition (Operations preserving convexity) Let V, W be vector spaces.

- if $f_j : S_j \subset V \to \mathbb{R}$ is convex for $j = 1, \ldots, m$ and $\alpha_j \geq 0$ then

$$\sum_{j=1}^{m} \alpha_j f_j : \bigcap_{j=1}^{m} S_j \subset V \to \mathbb{R} \text{ is convex}$$

- if $A : V \to W$ is an affine map and $f : S \subset W \to \mathbb{R}$ is convex, then

$$f \circ A : A^{-1}(W) \to \mathbb{R} \text{ is convex}$$

- if $f_j : S_j \subset V \to \mathbb{R}$ is convex for $j \in J$ then

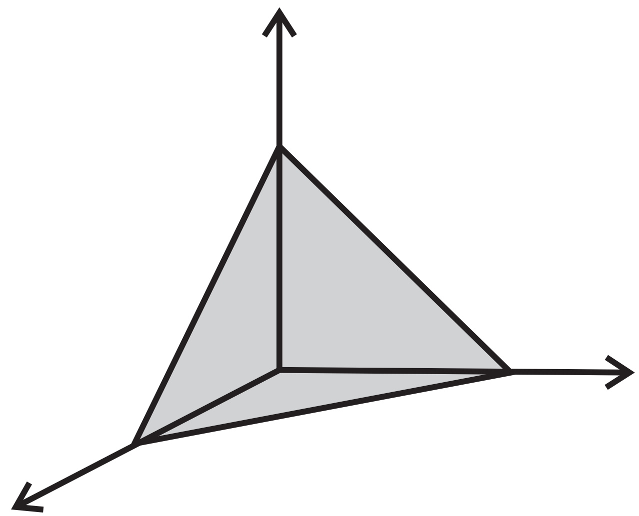
$$\sup_{j \in J} f_j : S \to \mathbb{R} \text{ is convex}$$

where $S := \{x \in \bigcap_{j \in J} S_j : \sup_{j \in J} f_j(x) < +\infty\}$

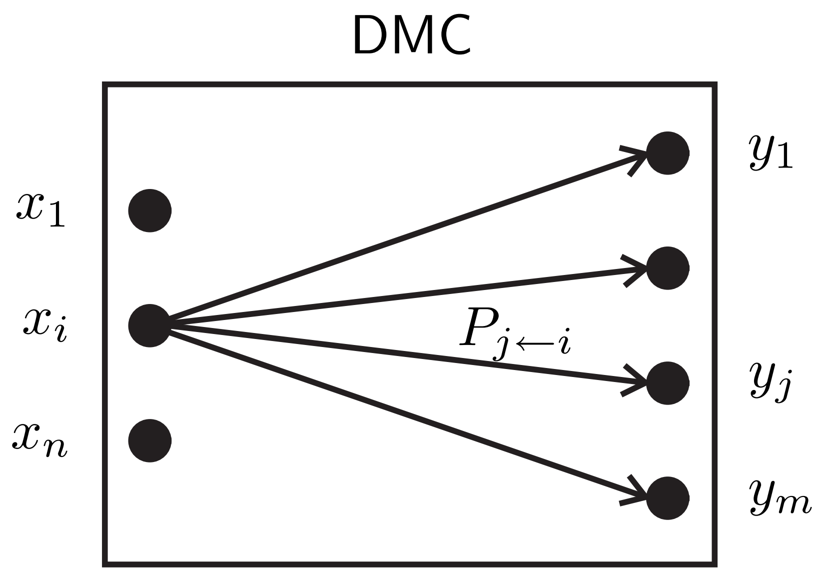
Example (entropy): $H : \Delta_n \to \mathbb{R}$

$$H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log(p_i)$$

is concave



Example (mutual information): a discrete memoryless channel (DMC)



$$P_{j \leftarrow i} := \text{Prob}(Y = y_j \mid X = x_i)$$

Input X has a probability mass function (pmf) $p_X \in \Delta_n$

Output Y has a pmf $p_Y \in \Delta_m$ given by $p_Y = Pp_X$

$$P := \underbrace{\begin{bmatrix} P_{1\leftarrow 1} & P_{1\leftarrow 2} & \cdots & P_{1\leftarrow n} \\ P_{2\leftarrow 1} & P_{2\leftarrow 2} & \cdots & P_{2\leftarrow n} \\ \vdots & \vdots & \cdots & \vdots \\ P_{m\leftarrow 1} & P_{m\leftarrow 2} & \cdots & P_{m\leftarrow n} \end{bmatrix}}_{\text{DMC transition matrix}}$$

Mutual information between X and Y is

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - \sum_{i=1}^{n} H(Y|X = x_i)\, p_X(i)$$

$I \, : \, \Delta_n \to \mathbb{R}$ is a concave function of p_X

Example (log-barrier for polyhedrons): $f : P \to \mathbb{R}$

$$f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^\top x)$$

is convex where $P = \{x \in \mathbb{R}^n : a_i^\top x < b_i \text{ for } i = 1, \ldots, m\}$

Example (sum of k largest eigenvalues): $f : \mathsf{S}^n \to \mathbb{R}$

$$f(X) = \lambda_1(X) + \lambda_2(X) + \cdots + \lambda_k(X)$$

is convex

Convexity through composition rules:

$$g \,:\, \mathbb{R}^n \to \mathbb{R}^m,\; g = (g_1, \ldots, g_m) \quad f \,:\, \mathbb{R}^m \to \mathbb{R}$$

$$\left.\begin{array}{c} f \text{ convex and } \nearrow \text{ in each variable} \\ g_i \text{ convex} \end{array}\right\} \Rightarrow (f \circ g) \text{ convex}$$

$$\left.\begin{array}{c} f \text{ convex and } \searrow \text{ in each variable} \\ g_i \text{ concave} \end{array}\right\} \Rightarrow (f \circ g) \text{ convex}$$

Proof (smooth case):

$$\nabla^2 (f \circ g)(x) = Dg(x) \nabla^2 f(g(x)) Dg(x)^\top + \sum_{i=1}^{m} \frac{\partial f}{\partial x_i}(g(x)) \nabla^2 g_i(x)$$

$$Dg(x) := [\, \nabla g_1(x)\, \nabla g_2(x)\, \cdots\, \nabla g_m(x)\,]$$

Examples:

- $f : \mathbb{R}^n \to \mathbb{R}$

$$f(x) = \log\left(e^{\|x\|^2 - 1} + e^{\max\{a_1^\top x + b_1, a_2^\top x + b_2\}}\right)$$

- $f : \mathbb{R}^n \to \mathbb{R}$

$$f(x) = g_{[1]}(x) + g_{[2]}(x) + \cdots + g_{[k]}(x)$$

with $g_j : \mathbb{R}^n \to \mathbb{R}$ convex

Proposition (Sublevel sets of convex functions are convex) Let V be a vector space and $f : S \subset V \to \mathbb{R}$ be convex. For any $r \in \mathbb{R}$ the sublevel set

$$S_r(f) = \{x \in S \, : \, f(x) \leq r\}$$

is convex

Example: the set

$$\{X \in \mathsf{S}^n_{++} \, : \, \mathsf{tr}\left(X^{-1}\right) \leq 5\}$$

is convex

Definition (Convexity w.r.t a generalized inequality) Let V, W be a vector spaces, $S \subset V$ a convex set and $K \subset W$ a proper cone. The map $F : S \subset V \to W$ is said to be K-convex if

$$F((1-\alpha)x + \alpha y) \preceq_K (1-\alpha)F(x) + \alpha F(y)$$

for all $x, y \in S$ and $\alpha \in [0, 1]$

Examples:

- $F : \mathbb{R}^n \to \mathbb{R}^m$ $F = (F_1, \ldots, F_m)$ is $\mathbb{R}^m_+$-convex iff all F_j are convex
- $F : \mathbb{R}^{n\times m} \to \mathsf{S}^m$
$$F(X) = X^\top A X$$
with $A \succeq 0$ is S^m_+-convex

Convex optimization problems

General optimization problem:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_j(x) \le 0 \quad j = 1, \ldots, m \\ & h_i(x) = 0 \quad i = 1, \ldots, p \end{aligned}$$

- f is the objective or cost function
- x is the optimization variable

If $m = p = 0$ the problem is unconstrained

A point x is feasible if $g_j(x) \le 0, h_i(x) = 0, f(x) < +\infty$

The optimal value of the problem is

$$f^\star := \inf\{f(x) \,:\, g_j(x) \le 0 \text{ for all } j, h_i(x) = 0 \text{ for all } i\}$$

If $f^\star = -\infty$ we say the problem is unbounded

If problem is infeasible we set $f^\star = +\infty$

Examples:

- $f^\star = 0$

$$\begin{array}{ll} \text{minimize} & x \\ \text{subject to} & x \geq 0 \end{array}$$

- $f^\star = -\infty$ (unbounded problem)

$$\begin{array}{ll} \text{minimize} & x \\ \text{subject to} & x \leq 0 \end{array}$$

- $f^\star = +\infty$ (infeasible problem)

$$\begin{array}{ll} \text{minimize} & x \\ \text{subject to} & x \leq 2 \\ & x \geq 3 \end{array}$$

A (global) solution is a feasible point $x^\star$ satisfying $f(x^\star) = f^\star$

There might be no solutions (problem is unsolvable). Examples:

- $f^\star = -\infty$

$$\begin{array}{ll} \text{minimize} & x \\ \text{subject to} & x \leq 0 \end{array}$$

- $f^\star = 0$ is not achieved

$$\text{minimize} \quad e^x$$

A local solution is a feasible point $x^\star$ which solves

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & g_j(x) \leq 0 & j = 1, \ldots, m \\ & h_i(x) = 0 & i = 1, \ldots, p \\ & \|x - x^\star\| \leq R & \end{array}$$

for some $R > 0$

Convex optimization problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_j(x) \leq 0 \quad j = 1, \ldots, m \\ & h_i(x) = 0 \quad i = 1, \ldots, p \end{array}$$

- f is convex
- g_j are convex
- h_i are affine

The set of feasible points is convex

Any local solution is a global solution

Linear optimization problem:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_j(x) \leq 0 \quad j = 1, \ldots, m \\ & h_i(x) = 0 \quad i = 1, \ldots, p \end{aligned}$$

- f is affine
- g_j are affine
- h_i are affine

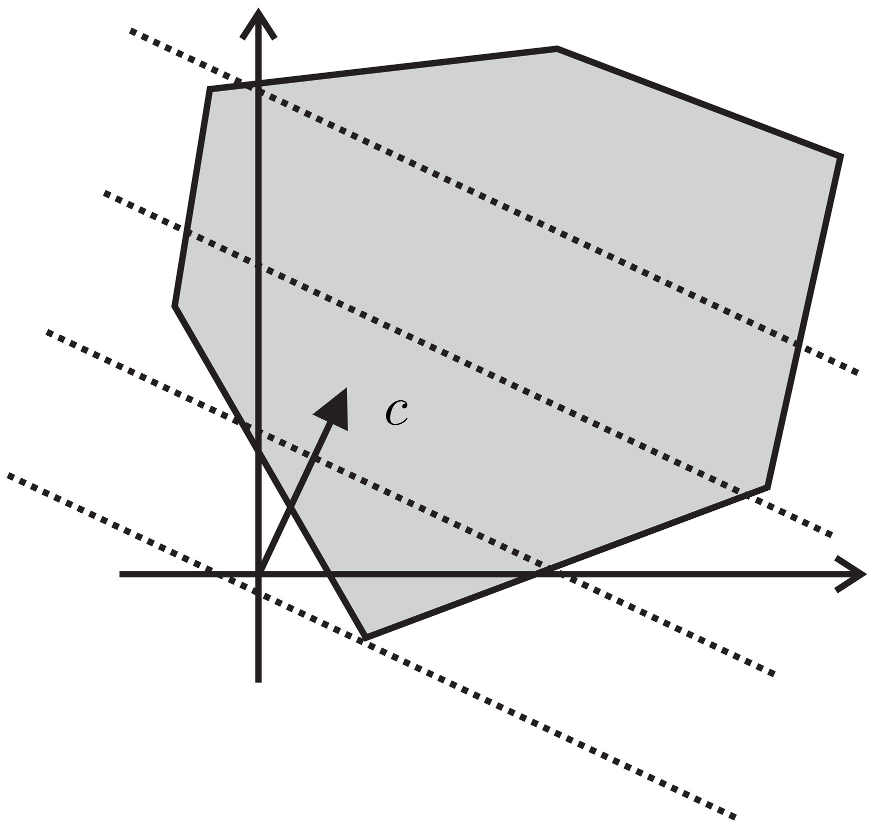
Linear optimization problem (inequality form):

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & Ax \leq b \end{array}$$

Linear optimization problem (standard form):

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

Geometrical interpretation:



Examples:

- Air traffic control
- Pattern separation: linear, quadratic, ... separators
- Chebyshev center of a polyhedron
- Problems involving the ℓ_∞ norm
- Problems involving the ℓ_1 norm
- Testing sphericity of a constellation
- Boolean relaxation

Air traffic example:

- n airplanes must land in the order $i = 1, 2, \ldots, n$
- t_i=time of arrival of plane i
- ith airplane must land in given time interval $[m_i, M_i]$

Goal: design $t_1, t_2, \ldots, t_n$ as to maximize $\min\{t_{i+1} - t_i \ : \ i\}$

Initial formulation:

$$\begin{aligned} \text{maximize} \quad & \underbrace{\min\{t_2 - t_1, t_3 - t_2, \dots, t_n - t_{n-1}\}}_{f(t_1, \dots, t_n)} \\ \text{subject to} \quad & t_i \leq t_{i+1} \\ & m_i \leq t_i \leq M_i \end{aligned}$$

Optimization variable: $(t_1, t_2, \dots, t_n)$

Epigraph form:

$$\begin{array}{ll} \text{maximize} & s \\ \text{subject to} & s \leq \min\{t_2 - t_1, t_3 - t_2, \ldots, t_n - t_{n-1}\} \\ & t_i \leq t_{i+1} \\ & m_i \leq t_i \leq M_i \end{array}$$

Optimization variable: $(t_1, t_2, \ldots, t_n, s)$

$$
\begin{array}{ll}
\text{maximize} & s \\
\text{subject to} & s \le t_{i+1} - t_i \\
& t_i \le t_{i+1} \\
& m_i \le t_i \le M_i
\end{array}
$$

Optimization variable: $(t_1, t_2, \ldots, t_n, s)$

Inequality form:

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & Ax \leq b \end{array}$$

$x \; = (t_1, \ldots, t_n, s)^\top$

$c = -(0, \ldots, 0, 1)^\top$

$$A = \begin{bmatrix} D & 1_{n-1} \\ D & 0_{n-1} \\ -I_n & 0_n \\ I_n & 0_n \end{bmatrix} \quad b = \begin{bmatrix} 0_{n-1} \\ 0_{n-1} \\ -m \\ M \end{bmatrix}$$

where

$$D = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -1 \end{bmatrix} : n \times (n-1) \quad m = \begin{bmatrix} -m_1 \\ -m_2 \\ \vdots \\ -m_n \end{bmatrix} \quad M = \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_n \end{bmatrix}$$

Pattern separation example:

- constellations $\mathcal{X} = \{x_1, \ldots, x_K\}$ and $\mathcal{Y} = \{y_1, \ldots, y_L\}$ in $\mathbb{R}^n$
- find a strict linear separator: $(s, r, \delta) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}_{++}$ such that

$$\begin{cases} s^\top x_k \geq r + \delta & \text{for all } k = 1, \ldots, K \\ s^\top y_l \leq r - \delta & \text{for all } l = 1, \ldots, L \end{cases}$$

System is homogeneous in (s, r, δ): normalize $\delta = 1$

Find $(s, r) \in \mathbb{R}^n \times \mathbb{R}$ such that

$$\begin{cases} s^\top x_k \geq r + 1 & \text{for all } k = 1, \ldots, K \\ s^\top y_l \leq r - 1 & \text{for all } l = 1, \ldots, L \end{cases}$$

Infeasible system if constellations overlap

Way out: allow model violations but minimize them

Problem formulation: find $(s, r, u, v) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^K \times \mathbb{R}^L$ such that

$$\begin{cases} s^\top x_k \geq r + 1 - u_k & \text{for all } k = 1, \ldots, K \\ s^\top y_l \leq r - 1 + v_l & \text{for all } l = 1, \ldots, L \\ u_k \geq 0 & \text{for all } k = 1, \ldots, K \\ v_l \geq 0 & \text{for all } l = 1, \ldots, L \end{cases}$$

and the average of model violations u and v is minimized

Linear program:

$$
\begin{array}{lll}
\text{minimize} & \frac{1}{K}1_K^\top u + \frac{1}{L}1_L^\top v & \\
\text{subject to} & s^\top x_k \geq r + 1 - u_k & k = 1, \ldots, K \\
& s^\top y_l \leq r - 1 + v_l & l = 1, \ldots, L \\
& u_k \geq 0 & k = 1, \ldots, K \\
& v_l \geq 0 & l = 1, \ldots, L
\end{array}
$$

Optimization variable: $(s, r, u, v) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^K \times \mathbb{R}^L$

Inequality form:

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & Ax \leq b \end{array}$$

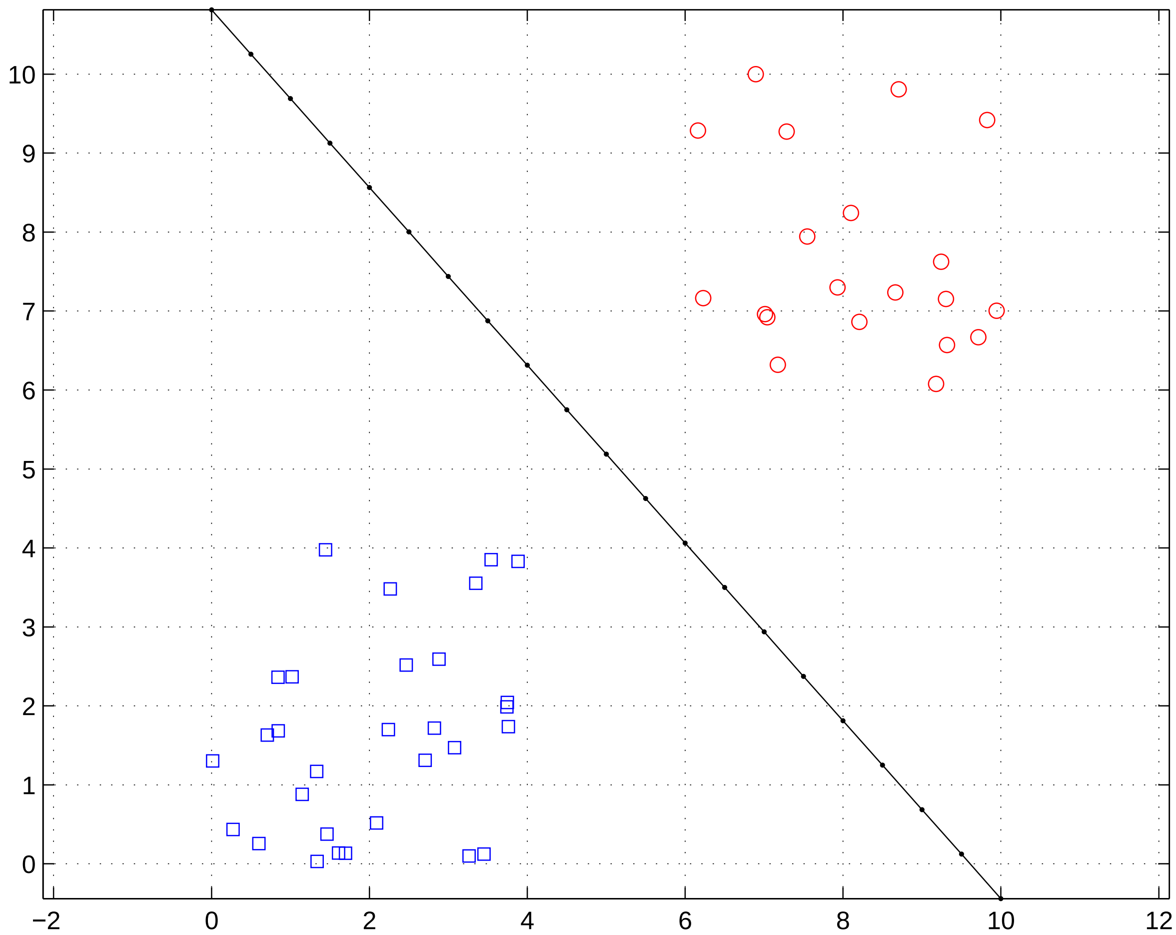
$x\,(:= (s, r, u, v)) \in \mathbb{R}^{n+1+K+L}$

$c = (0_n, 0, \frac{1}{K} 1_K, \frac{1}{L} 1_L)$

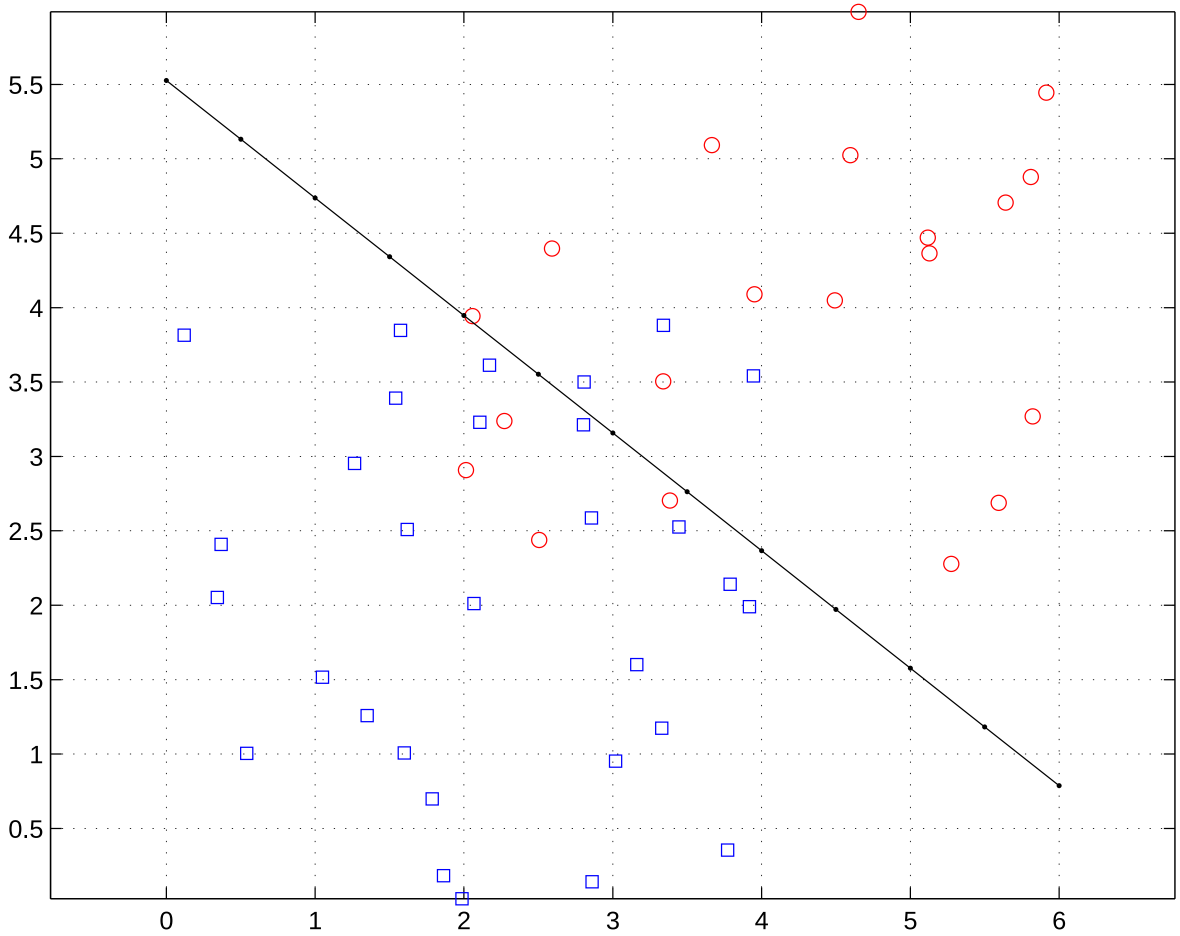
$$A = \begin{bmatrix} -X^\top & 1_K & -I_K & 0_{K\times L} \\ Y^\top & -1_L & 0_{L\times K} & -I_L \\ 0_{K\times n} & 0_K & -I_K & 0_{K\times L} \\ 0_{L\times n} & 0_L & 0_{L\times K} & -I_L \end{bmatrix} \quad b = \begin{bmatrix} -1_K \\ -1_L \\ 0_K \\ 0_L \end{bmatrix}$$

where $X = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \end{bmatrix}$ $\quad Y = \begin{bmatrix} y_1 & y_2 & \cdots & y_L \end{bmatrix}$

Example: constellations do not overlap



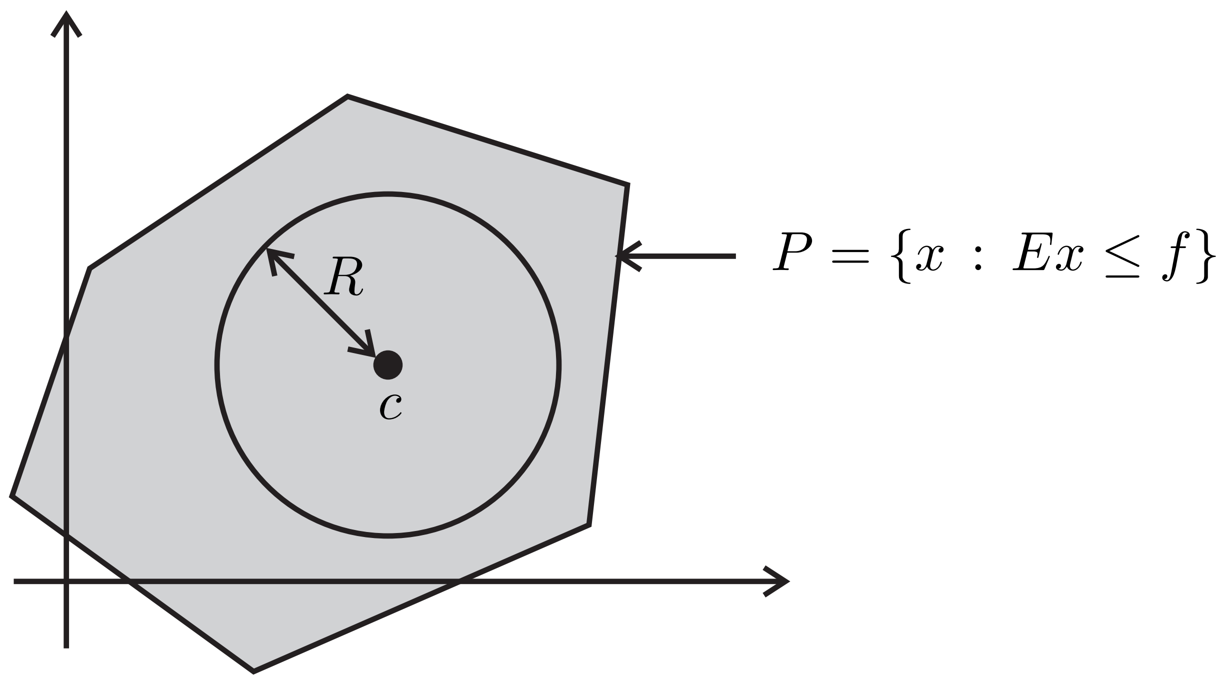
Example: constellations overlap



```
K = 30; L = 20; margin = 2; % choose margin = -2 for overlap
X = 4*rand(2,K); Y = 4+margin+4*rand(2,L);
figure(1); clf; plot(X(1,:),X(2,:),'bs'); grid on;
axis('equal'); hold on; plot(Y(1,:),Y(2,:),'ro');
f = [ zeros(2,1) ; 0 ; (1/K)*ones(K,1) ; (1/L)*ones(L,1) ];
A = [ -X' ones(K,1) -eye(K) zeros(K,L) ;
Y' -ones(L,1) zeros(L,K) -eye(L) ;
zeros(K,2) zeros(K,1) -eye(K) zeros(K,L) ;
zeros(L,2) zeros(L,1) zeros(L,K) -eye(L) ];
b = [ -ones(K,1) ; -ones(L,1) ; zeros(K,1); zeros(L,1) ];
x = linprog(f,A,b);
asol = x(1:2,:); bsol = x(3);
t = 0:0.5:10;
y = (bsol-asol(1)*t)/asol(2); plot(t,y,'k.-');
```

Chebyshev center example:

- find largest ball contained in a polyhedron



- application: largest robot arm that can operate in a polyhedral room

Initial formulation:

$$\begin{array}{ll} \text{maximize} & R \\ \text{subject to} & B(c, r) \subset P \end{array}$$

Optimization variable: $(c, R) \in \mathbb{R}^n \times \mathbb{R}$

Equivalent formulation:

$$\begin{aligned} \text{maximize} \quad & R \\ \text{subject to} \quad & B(c, R) \subset H^{-}_{e_1, f_1} \\ & \vdots \\ & B(c, R) \subset H^{-}_{e_m, f_m} \end{aligned}$$

Optimization variable: $(c, R) \in \mathbb{R}^n \times \mathbb{R}$

Key-fact:

$$B(c, R) \subset H_{e,f}^{-} = \{x : e^\top x \leq f\} \quad \Leftrightarrow \quad e^\top c + R\,\|e\| \leq f$$

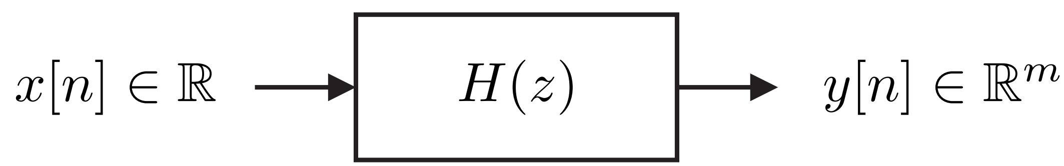
Linear program:

$$\begin{array}{ll} \text{maximize} & R \\ \text{subject to} & e_1^\top c + \|e_1\|\,R \leq f_1 \\ & \vdots \\ & e_m^\top c + \|e_m\|\,R \leq f_m \end{array}$$

Optimization variable: $(c, R) \in \mathbb{R}^n \times \mathbb{R}$

Example involving the ℓ_∞ norm:

- SIMO channel is given



- stacked data model: $y = \mathcal{H}x$ (recall page 25)
- input constraints: $Cx \leq d$

  e.g. $|x[n]| \leq P$, $|x[n] - x[n-1]| \leq S$, $|x[n] - x[m]| \leq V$, ...

- design input x such that $\|\mathcal{H}x - y_{\mathsf{des}}\|_\infty$ is minimized

Initial formulation:

$$\begin{array}{ll} \text{minimize} & \|\mathcal{H}x - y_{\mathsf{des}}\|_\infty \\ \text{subject to} & Cx \leq d \end{array}$$

Optimization variable: $x \in \mathbb{R}^N$

Epigraph form:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \|\mathcal{H}x - y_{\mathsf{des}}\|_\infty \le t \\ & Cx \le d \end{array}$$

Optimization variable: $(x, t) \in \mathbb{R}^N \times \mathbb{R}$

Linear program:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & -t\mathbf{1} \leq \mathcal{H}x - y_{\text{des}} \leq t\mathbf{1} \\ & Cx \leq d \end{array}$$

Optimization variable: $(x, t) \in \mathbb{R}^N \times \mathbb{R}$

Example involving the ℓ_1 norm:

- “dictionary” $A : m \times n$ ($m \ll n$) and b are given
- find the sparsest solution of $Ax = b$

Initial formulation:

$$\begin{array}{ll} \text{minimize} & \mathbf{card}(x) \\ \text{subject to} & Ax = b \end{array}$$

Optimization variable: $x \in \mathbb{R}^n$

Convex approximation:

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = b \end{array}$$

Optimization variable: $x \in \mathbb{R}^n$

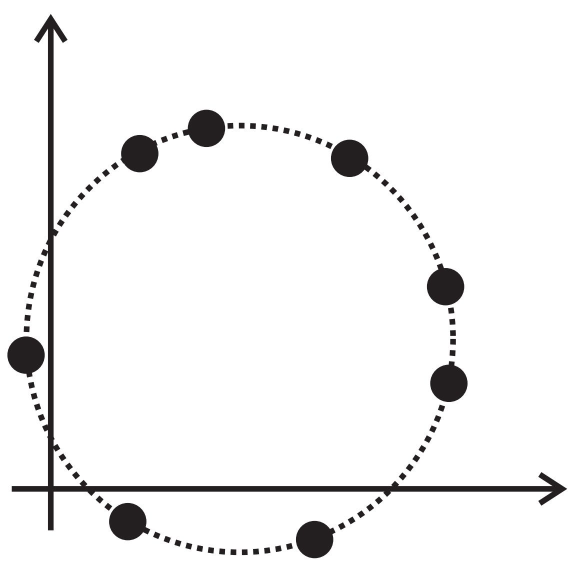
Linear program:

$$\begin{array}{ll} \text{minimize} & 1^\top y + 1^\top z \\ \text{subject to} & A(y - z) = b \\ & y \geq 0, z \geq 0 \end{array}$$

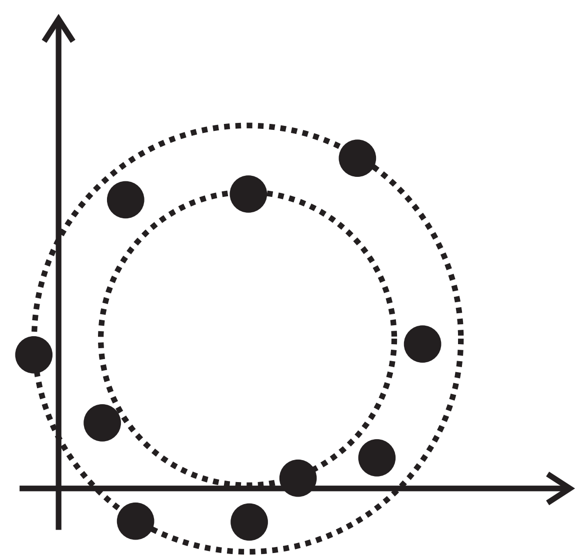
Optimization variable: $(y, z) \in \mathbb{R}^n \times \mathbb{R}^n$

Sphericity example:

- constellation $\mathcal{X} = \{x_1, x_2, \ldots, x_K\}$ is given
- how much spherical is the constellation ?



Spherical



Non spherical

Initial formulation:

$$\begin{array}{ll} \text{minimize} & R^2 - r^2 \\ \text{subject to} & r \le \|c - x_k\| \le R \quad k = 1, \ldots, K \end{array}$$

Optimization variable: $(c, r, R) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}$

$\mathcal{X}$ is spherical $\Leftrightarrow$ value of the problem is 0

Linear program:

$$\begin{array}{ll} \text{minimize} & \widetilde{R} - \widetilde{r} \\ \text{subject to} & \widetilde{r} \leq \|x_k\|^2 - 2x_k^\top c \leq \widetilde{R} \quad k = 1, \ldots, K \end{array}$$

Optimization variable: $(c, \widetilde{r}, \widetilde{R}) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}$

Boolean relaxation example:

- difficult optimization problem

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & Ax \leq b \\ & x_i \in \{0,1\} \quad i = 1,\ldots,n \end{array}$$

- relaxing $x_i \in \{0,1\}$ to $x_i \in [0,1]$ yields a linear program:

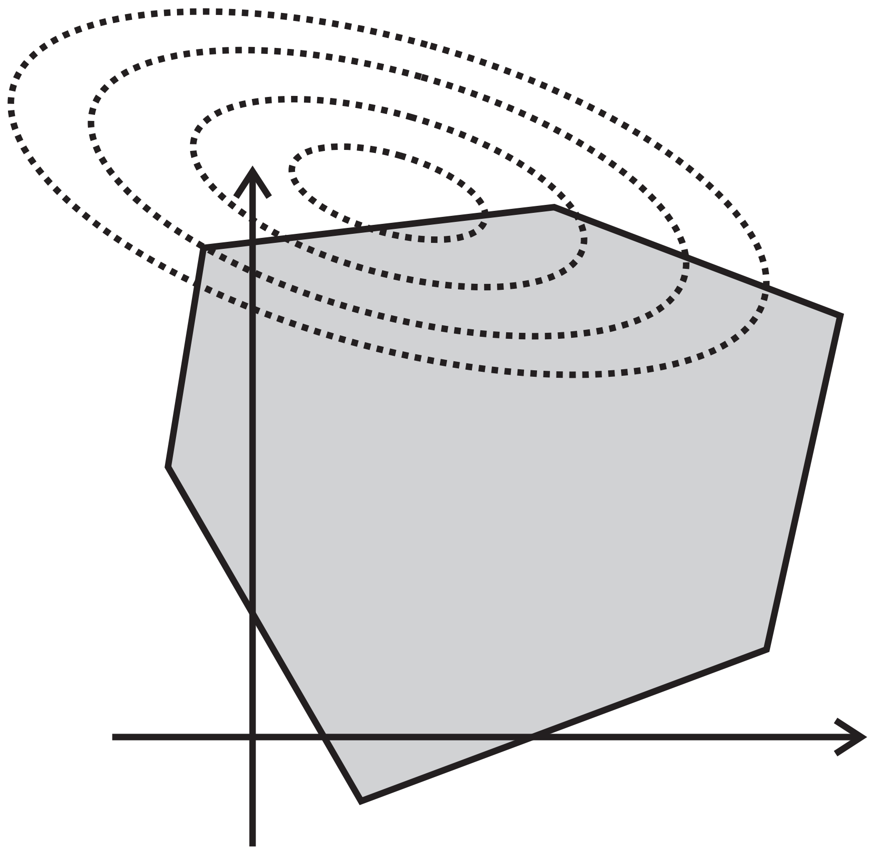
$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & Ax \leq b \\ & 0 \leq x_i \leq 1 \quad i = 1,\ldots,n \end{array}$$

Convex quadratic optimization problem:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_j(x) \leq 0 \quad j = 1, \ldots, m \\ & h_i(x) = 0 \quad i = 1, \ldots, p \end{aligned}$$

- f is convex and quadratic
- g_j are affine
- h_i are affine

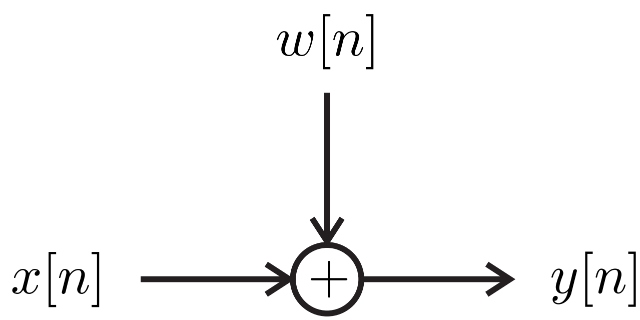
Geometrical interpretation:



Examples:

- ML estimation of constrained signals in additive Gaussian noise
- Portfolio optimization
- Basis pursuit denoising
- Support vector machines

ML estimation example:



$w[n]$ is Gaussian noise

Collect N samples:

$$\underbrace{\begin{bmatrix} y[1] \\ y[2] \\ \vdots \\ y[N] \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} x[1] \\ x[2] \\ \vdots \\ x[N] \end{bmatrix}}_{x} + \underbrace{\begin{bmatrix} w[1] \\ w[2] \\ \vdots \\ w[N] \end{bmatrix}}_{w}$$

Goal: estimate x given y

Signal x obeys polyhedral constraints $Cx \leq d$

Examples:

- known signal prefix

$$x[1] = 1, x[2] = 0, x[3] = -1$$

- power constraint

$$|x[n]| \leq P$$

- equal boundary values

$$x[1] = x[N]$$

- mean value is between -5 and $+5$

$$-5 \leq \frac{1}{N}\left(x[1] + x[2] + \cdots + x[N]\right) \leq 5$$

- rate constraint

$$|x[n] - x[n-1]| \leq R$$

Letting $w \sim \mathcal{N}(0, \Sigma)$ the pdf of y is

$$p_Y(y; x) = \frac{1}{(2\pi)^{N/2} \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(y-x)^\top \Sigma^{-1} (y-x)\right)$$

The ML estimate of x corresponds to the solution of

$$\begin{array}{ll} \text{maximize} & p_Y(y; x) \\ \text{subject to} & Cx \leq d \end{array}$$

Optimization variable: $x \in \mathbb{R}^N$

Convex quadratic program:

$$\begin{array}{ll} \text{minimize} & (x-y)^\top \Sigma^{-1} (x-y) \\ \text{subject to} & Cx \leq d \end{array}$$

Portfolio optimization example:

- T euros to invest among n assets
- r_i is the random rate of return of ith asset
- first 2 moments of the random vector $r = (r_1, r_2, \ldots, r_n)$ are known:

$$\mu = \mathsf{E}\{r\} \quad \Sigma = \mathsf{E}\left\{(r-\mu)(r-\mu)^\top\right\}$$

The return of an investment $x = (x_1, x_2, \ldots, x_n)$ is the random variable:

$$s(x) = r_1 x_1 + r_2 x_2 + \cdots + r_n x_n$$

First 2 moments of $s(x)$ are:

$$\mathsf{E}\{s(x)\} = \mu^\top x \quad \mathsf{var}(s(x)) = \mathsf{E}\{(s(x) - \mathsf{E}\{s(x)\})^2\} = x^\top \Sigma x$$

Goal: find the portfolio x with minimum variance guaranteeing a given mean return s_0

Optimization problem:

$$\begin{array}{ll} \text{minimize} & \mathsf{var}(s(x)) \\ \text{subject to} & x \geq 0 \\ & 1_n^\top x = T \\ & \mathsf{E}\{s(x)\} = s_0 \end{array}$$

Convex quadratic program:

$$
\begin{aligned}
\text{minimize} \quad & x^\top \Sigma x \\
\text{subject to} \quad & x \geq 0 \\
& 1_n^\top x = T \\
& \mu^\top x = s_0
\end{aligned}
$$

Goal: find the portfolio maximizing the expected mean return with a risk penalization

Optimization problem:

$$\begin{array}{ll} \text{maximize} & \mathsf{E}\{s(x)\} - \beta\,\mathsf{var}(s(x)) \\ \text{subject to} & x \geq 0 \\ & 1_n^\top x = T \end{array}$$

$\beta \geq 0$ is the risk-aversion parameter

Convex quadratic program:

$$\begin{array}{ll} \text{minimize} & -\mu^\top x + \beta x^\top \Sigma x \\ \text{subject to} & x \geq 0 \\ & 1_n^\top x = T \end{array}$$

Basis pursuit denoising example:

- "dictionary" $A : m \times n$ ($m \ll n$) and $b \in \mathbb{R}^m$ are given
- find x such that $Ax \simeq b$ and x is sparse

Heuristic to find sparse solutions:

$$\text{minimize} \quad \|Ax - b\|^2 + \beta \|x\|_1$$

$\beta \geq 0$ trades off data fidelity and sparsity

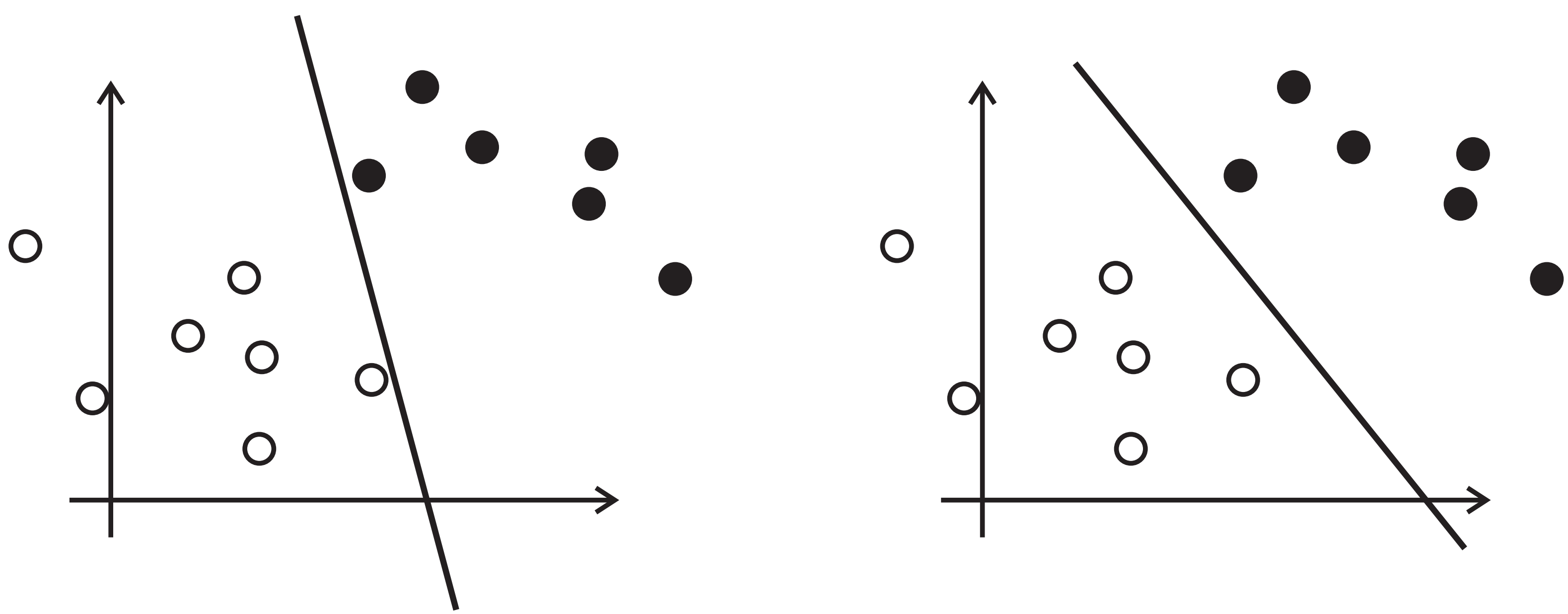
Convex quadratic program:

$$\begin{aligned} \text{minimize} \quad & \|A(y-z)-b\|^2 + \beta\left(1_n^\top y + 1_n^\top z\right) \\ \text{subject to} \quad & y \geq 0 \\ & z \geq 0 \end{aligned}$$

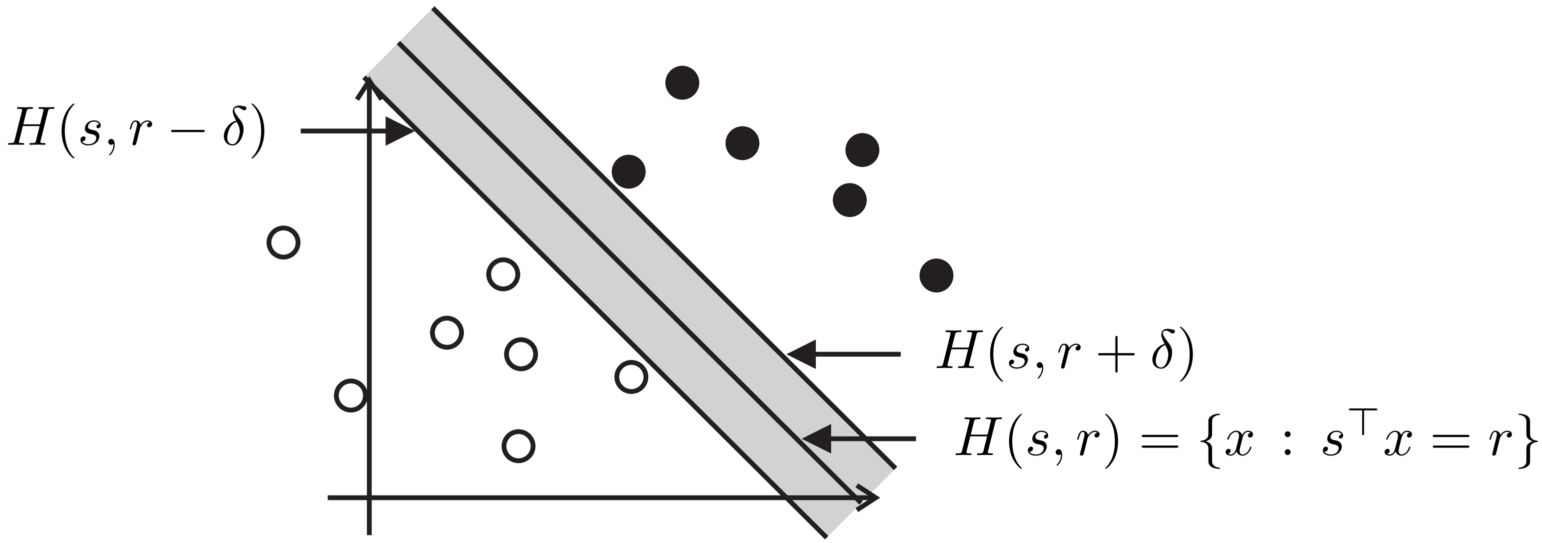
Optimization variable: $(y, z) \in \mathbb{R}^n \times \mathbb{R}^n$

Support vector machine example:

- constellations $\mathcal{X} = \{x_1, \ldots, x_K\}$ and $\mathcal{Y} = \{y_1, \ldots, y_L\}$ are given
- find the "best" linear separator



Optimization variable: $(s, r, \delta) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}_{++}$



Constraints:

- $\mathcal{X} \subset H(s, r + \delta)^+ = \{x : s^\top x \geq r + \delta\}$
- $\mathcal{Y} \subset H(s, r - \delta)^- = \{x : s^\top x \leq r - \delta\}$

Initial formulation:

$$\begin{array}{ll} \text{maximize} & \mathsf{dist}\left(H(s, r-\delta), H(s, r+\delta)\right) \\ \text{subject to} & \mathcal{X} \subset H(s, r+\delta)^{+} \\ & \mathcal{Y} \subset H(s, r-\delta)^{-} \\ & \delta > 0 \end{array}$$

Optimization variable: $(s, r, \delta) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}$

Optimization problem:

$$
\begin{array}{lll}
\text{maximize} & 2\dfrac{\delta}{\|s\|} & \\
\text{subject to} & s^\top x_k \geq r + \delta & k = 1, \ldots, K \\
& s^\top y_l \leq r - \delta & l = 1, \ldots, L \\
& \delta > 0 &
\end{array}
$$

If (s, r, δ) is feasible so is $\alpha(s, r, \delta)$ ($\alpha > 0$) with same cost value

Normalize $\delta = 1$:

$$\begin{array}{ll} \text{maximize} & 2\dfrac{1}{\|s\|} \\ \text{subject to} & s^\top x_k \geq r + 1 \quad k = 1, \ldots, K \\ & s^\top y_l \leq r - 1 \quad l = 1, \ldots, L \end{array}$$

Optimization variable: $(s, r) \in \mathbb{R}^n \times \mathbb{R}$

Convex quadratic program:

$$\begin{array}{ll} \text{minimize} & \|s\|^2 \\ \text{subject to} & s^\top x_k \geq r + 1 \quad k = 1, \ldots, K \\ & s^\top y_l \leq r - 1 \quad l = 1, \ldots, L \end{array}$$

Convex conic optimization problem:

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & \mathcal{A}_i(x) \in K_i \quad i = 1, \ldots, p \end{array}$$

- $\mathcal{A}_i$ are affine maps
- K_i are convex cones

Example: linear program

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & Ax \geq b \end{array}$$

is the conic optimization problem

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & \mathcal{A}(x) \in K \end{array}$$

- $\mathcal{A}(x) = Ax - b$
- $K = \mathbb{R}^m_{++}$

Second-order cone program (SOCP):

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & \mathcal{A}_i(x) \in K_i \quad i = 1, \ldots, p \end{array}$$

- $\mathcal{A}_i$ are affine maps
- K_i are second-order cones

Equivalently:

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & \|A_i x + b_i\| \le c_i^\top x + d_i \quad i = 1, \ldots, p \end{array}$$

Examples:

- ℓ_∞ approximation with complex data
- Geometric approximation
- Hyperbolic constraints
- Robust beamforming

Given $A \in \mathbb{C}^{m \times n}$ and $b \in \mathbb{C}^m$

$$\text{minimize} \quad \|Az - b\|_\infty$$

Optimization variable: $z \in \mathbb{C}^n$

For $w = (w_1, w_2, \ldots, w_n) \in \mathbb{C}^n$, $\|w\|_\infty = \max\{|w_i| \,:\, i\}$

Epigraph reformulation:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & |a_i^\top z - b_i| \leq t \quad i = 1, \ldots, m \end{array}$$

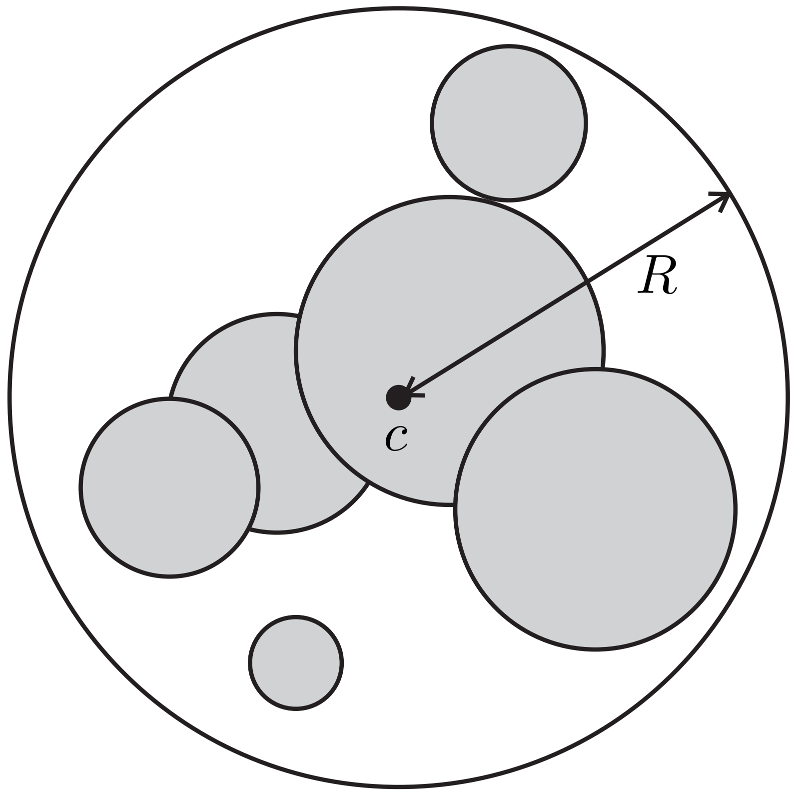
Optimization variable: $(t, z) \in \mathbb{R} \times \mathbb{C}^n$

Writing $z := x + jy$, yields the SOCP

$$
\begin{array}{ll}
\text{minimize} & t \\
\text{subject to} & \left\| \begin{bmatrix} (\operatorname{Re} a_i)^\top & -(\operatorname{Im} a_i)^\top \\ (\operatorname{Im} a_i)^\top & (\operatorname{Re} a_i)^\top \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \operatorname{Re} b_i \\ \operatorname{Im} b_i \end{bmatrix} \right\| \leq t \quad i = 1, \ldots, m
\end{array}
$$

Optimization variable: $(t, x, y) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n$

Find the smallest ball $B(c, R)$ containing n given ones $B(c_i, R_i)$



Initial formulation:

$$\begin{array}{ll} \text{minimize} & R \\ \text{subject to} & B(c_i, R_i) \subset B(c, R) \quad i = 1, \ldots, n \end{array}$$

Optimization variable: $(c, R) \in \mathbb{R} \times \mathbb{R}^n$

SOCP:

$$\begin{array}{ll} \text{minimize} & R \\ \text{subject to} & \|c - c_i\| + R_i \leq R \quad i = 1, \ldots, n \end{array}$$

Optimization variable: $(c, R) \in \mathbb{R} \times \mathbb{R}^n$

Fact:

$$\begin{cases} \|w\|^2 \leq yz \\ y \geq 0 \\ z \geq 0 \end{cases} \quad \Leftrightarrow \quad \left\| \begin{bmatrix} 2w \\ y - z \end{bmatrix} \right\| \leq y + z$$

Application:

$$\text{minimize} \quad \|Ax - b\|^4 + \|Cx - d\|^2$$

Optimization variable: $x \in \mathbb{R}^n$

Reformulation:

$$\begin{array}{ll} \text{minimize} & s+t \\ \text{subject to} & \|Ax-b\|^4 \le s \\ & \|Cx-d\|^2 \le t \end{array}$$

Optimization variable: $(x, s, t) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}$

$$
\begin{array}{ll}
\text{minimize} & s + t \\
\text{subject to} & \|Ax - b\|^4 \leq s \\
& \left\| \begin{bmatrix} 2(Cx - d) \\ t - 1 \end{bmatrix} \right\| \leq t + 1
\end{array}
$$

Optimization variable: $(x, s, t) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}$

$$
\begin{array}{ll}
\text{minimize} & s + t \\
\text{subject to} & \|Ax - b\|^2 \le u \\
& u^2 \le s \\
& \left\| \begin{bmatrix} 2(Cx - d) \\ t - 1 \end{bmatrix} \right\| \le t + 1
\end{array}
$$

Optimization variable: $(x, s, t, u) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}$

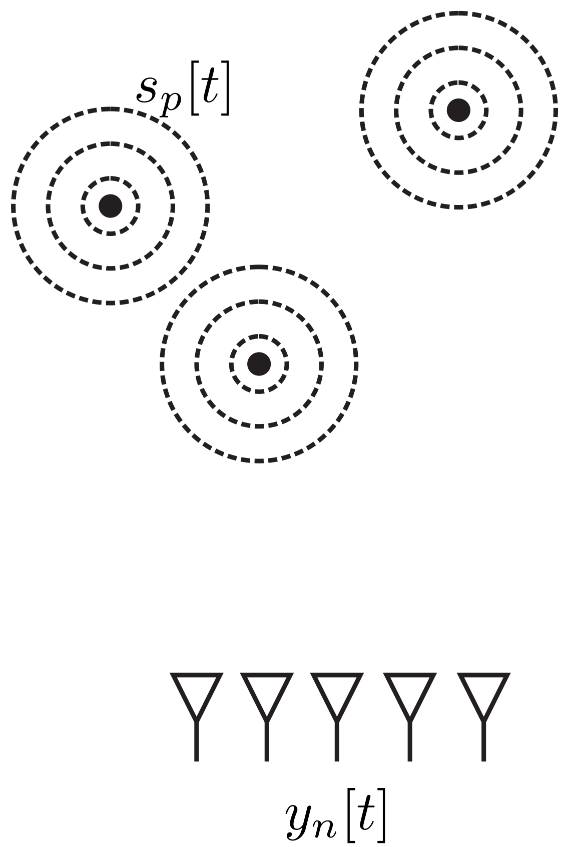
SOCP:

$$\begin{array}{ll} \text{minimize} & s+t \\ \text{subject to} & \left\| \begin{bmatrix} 2(Ax-b) \\ u-1 \end{bmatrix} \right\| \leq u+1 \\ & \left\| \begin{bmatrix} 2u \\ s-1 \end{bmatrix} \right\| \leq s+1 \\ & \left\| \begin{bmatrix} 2(Cx-d) \\ t-1 \end{bmatrix} \right\| \leq t+1 \end{array}$$

Optimization variable: $(x, s, t, u) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}$

Robust beamforming:

- N antennas at the base station
- P users



Baseband data model:

$$y[t] = \sum_{p=1}^{P} h_p s_p[t] + n[t]$$

- $y[t] \in \mathbb{C}^N$ is array snapshot at time t
- $h_p \in \mathbb{C}^N$ is pth channel vector
- $s_p[t] \in \mathbb{C}$ is symbol transmitted by pth user at time t
- $n[t] \in \mathbb{C}^N$ is additive noise

Goal: given $y[t]$ estimate $s_1[t]$

Assumption: h_1 is known and sources and noise are uncorrelated

Linear receiver:

$$\widehat{s}_1[t] = w^H y[t]$$

Minimum variance distortionless receiver (MVDR):

$$w_{\mathrm{MVDR}} = \begin{array}{ll} \arg\min & w^H R w \\ \text{subject to} & w^H h_1 = 1 \end{array}$$

$R = \mathrm{E}\left\{y[t]y[t]^H\right\} \simeq \frac{1}{T}\sum_{t=1}^{T} y[t]y[t]^H$

Robust formulation:

$$\begin{array}{ll} \text{minimize} & w^H R w \\ \text{subject to} & |w^H h| \geq 1 \text{ for all } h \in B(h_1, \epsilon) \end{array}$$

Optimization variable: $w \in \mathbb{C}^N$

Equivalent formulation:

$$\begin{array}{ll} \text{minimize} & w^H R w \\ \text{subject to} & |w^H h_1| - \epsilon \|w\| \geq 1 \end{array}$$

If w is feasible, so is $e^{j\theta} w$ with same cost

$$
\begin{array}{ll}
\text{minimize} & w^H R w \\
\text{subject to} & |w^H h_1| - \epsilon \|w\| \geq 1 \\
& \mathrm{Re}\left\{w^H h_1\right\} \geq 0 \\
& \mathrm{Im}\left\{w^H h_1\right\} = 0
\end{array}
$$

Writing $w = x + jy$ leads to SOCP:

$$\begin{array}{ll}\text{minimize} & t \\ \text{subject to} & \left\| A \begin{bmatrix} x \\ y \end{bmatrix} \right\| \leq t \\ & \left\| \epsilon \begin{bmatrix} x \\ y \end{bmatrix} \right\| \leq \begin{bmatrix} (\operatorname{Re} h_1)^\top & (\operatorname{Im} h_1)^\top \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} - 1 \\ & \begin{bmatrix} (\operatorname{Im} h_1)^\top & -(\operatorname{Re} h_1)^\top \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0 \end{array}$$

A is the square-root of

$$\begin{bmatrix} \operatorname{Re} R & \operatorname{Im} R \\ (\operatorname{Im} R)^\top & \operatorname{Re} R \end{bmatrix}$$

Semi-definite program (SDP):

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & \mathcal{A}_i(x) \in K_i \quad i = 1, \ldots, p \end{array}$$

- $\mathcal{A}_i$ are affine maps
- K_i are semi-definite matrix cones

Equivalently:

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & A_0 + x_1 A_1 + x_2 A_2 + \cdots x_n A_n \succeq 0 \end{array}$$

Examples:

- Eigenvalue optimization
- Manifold learning
- MAXCUT

Eigenvalue optimization:

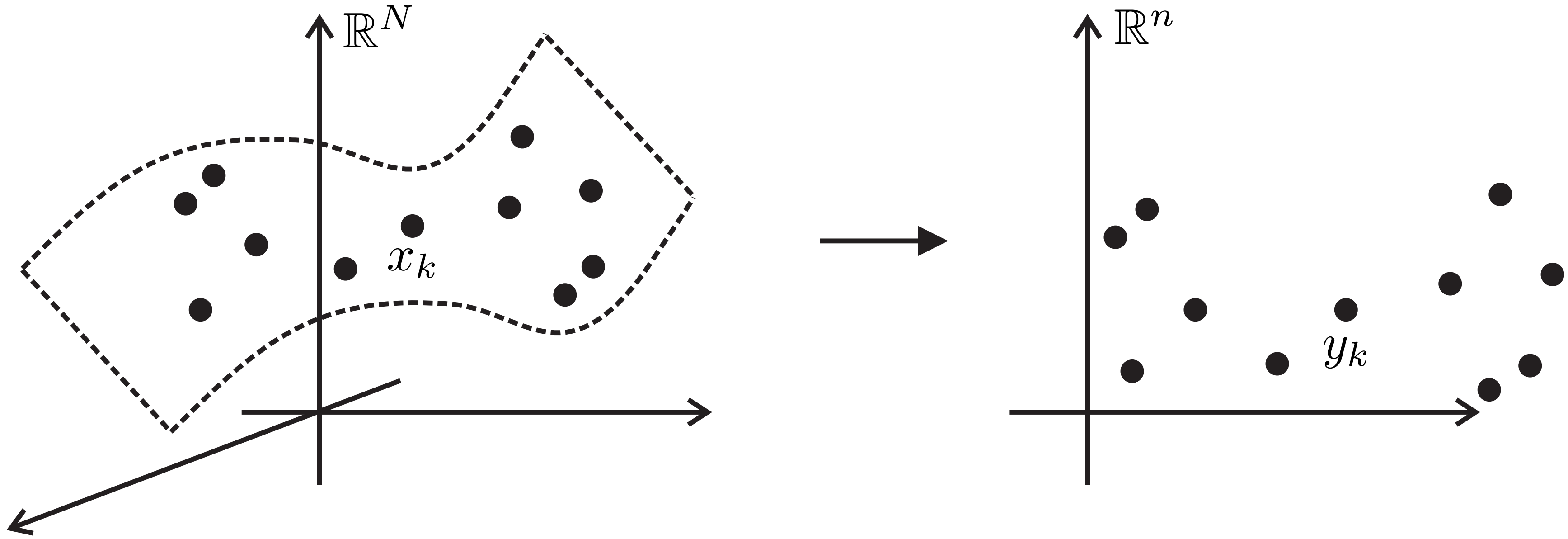
$$\text{minimize} \quad \lambda_{\max}\left(A_0 + x_1 A_1 + x_2 A_2 + \cdots + x_n A_n\right)$$

where $A_i \in \mathbf{S}^m$ are given

Equivalent SDP:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & A_0 + x_1 A_1 + x_2 A_2 + \cdots + x_n A_n \preceq t I_m \end{array}$$

Manifold learning:



Approach: "open" the manifold (maximum variance unfolding)

$$
\begin{array}{ll}
\text{maximize} & \sum_{k=1}^{K} \|y_k\|^2 \\
\text{subject to} & \frac{1}{K}\sum_{k=1}^{K} y_k = 0 \\
& \|y_k - y_l\|^2 = \|x_k - x_l\|^2 \quad \text{for all } (k,l) \in \mathcal{N}
\end{array}
$$

Optimization variable: $(y_1, y_2, \ldots, y_K) \in \mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n$

Equivalent formulation:

$$
\begin{array}{ll}
\text{maximize} & \operatorname{tr}\left(Y^\top Y\right) \\
\text{subject to} & 1^\top Y^\top Y 1 = 0 \\
& e_k^\top Y^\top Y e_k + e_l^\top Y^\top Y e_l - 2 e_k^\top Y^\top Y e_l = \|x_k - x_l\|^2 \quad \text{for all } (k,l) \in \mathcal{N}
\end{array}
$$

Optimization variable: $Y \in \mathbb{R}^{n \times K}$

$$
\begin{array}{ll}
\text{maximize} & \operatorname{tr}(G) \\
\text{subject to} & 1^\top G 1 = 0 \\
& e_k^\top G e_k + e_l^\top G e_l - 2 e_k^\top G e_l = \|x_k - x_l\|^2 \quad \text{for all } (k,l) \in \mathcal{N} \\
& G = Y^\top Y
\end{array}
$$

Optimization variable: $(G, Y) \in \mathbb{R}^{K \times K} \times \mathbb{R}^{n \times K}$

We can drop the variable Y

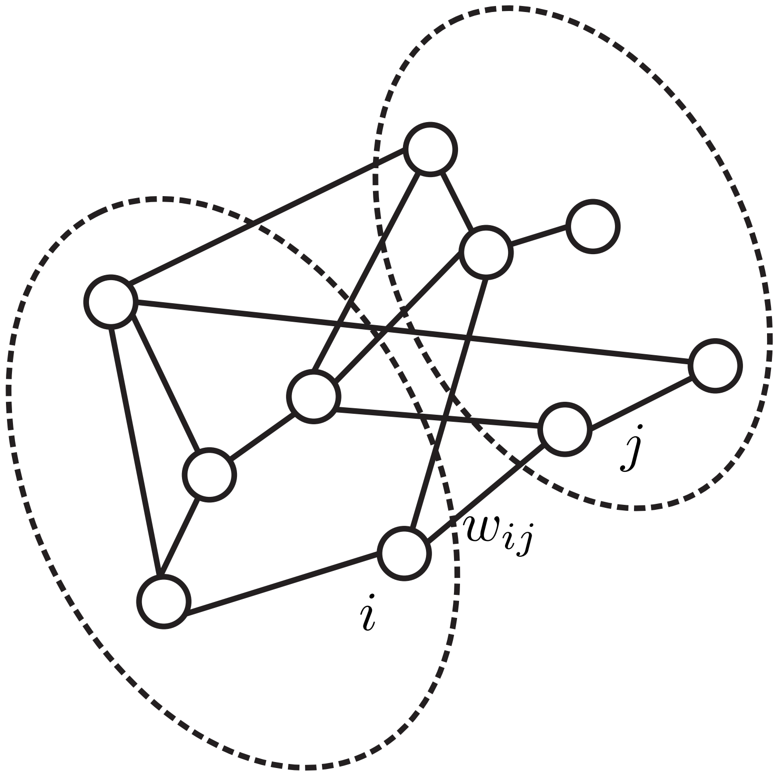
$$
\begin{array}{ll}
\text{maximize} & \operatorname{tr}(G) \\
\text{subject to} & 1^\top G 1 = 0 \\
& e_k^\top G e_k + e_l^\top G e_l - 2 e_k^\top G e_l = \|x_k - x_l\|^2 \quad \text{for all } (k,l) \in \mathcal{N} \\
& G \succeq 0 \\
& \operatorname{rank}(G) \leq n
\end{array}
$$

Optimization variable: $G \in \mathbb{R}^{K \times K}$

Removing the rank constraint yields the SDP:

$$\begin{array}{ll} \text{maximize} & \operatorname{tr}(G) \\ \text{subject to} & 1^\top G 1 = 0 \\ & e_k^\top G e_k + e_l^\top G e_l - 2 e_k^\top G e_l = \|x_k - x_l\|^2 \quad \text{for all } (k,l) \in \mathcal{N} \\ & G \succeq 0 \end{array}$$

MAXCUT problem



- $w_{ij} \geq 0$ is weight of edge (i, j)
- Value of a cut $=$ sum of weights of broken edges
- Find the maximum cut

$$\text{MAXCUT} = \begin{array}{ll} \max & \phi(x) \\ \text{subject to} & x_i^2 = 1 \quad i = 1, 2, \ldots, n \end{array}$$

where

$$\phi(x) = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} \frac{1 - x_i x_j}{2}$$

Optimization variable: $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$

$$
\begin{aligned}
\text{MAXCUT} = \quad & \max && \tfrac{1}{2}\textstyle\sum_{i,j=1}^{n} w_{ij}\,\tfrac{1-X_{ij}}{2} && \\
& \text{subject to} && X_{ii} = 1 && i = 1, 2, \ldots, n \\
& && X = xx^{\top} &&
\end{aligned}
$$

Optimization variable: $(X, x) \in \mathbb{R}^{n \times n} \times \mathbb{R}^n$

$$\text{MAXCUT} = \begin{array}{lll} \text{max} & \frac{1}{4}1_n^\top W 1_n - \frac{1}{4}\mathrm{tr}(WX) & \\ \text{subject to} & X_{ii} = 1 & i = 1, 2, \dots, n \\ & X = xx^\top & \end{array}$$

Optimization variable: $(X, x) \in \mathbb{R}^{n\times n} \times \mathbb{R}^n$

We can drop the variable x

$$\begin{array}{llll} \text{MAXCUT} = & \text{max} & \frac{1}{4} 1_n^\top W 1_n - \frac{1}{4} \mathrm{tr}(WX) & \\ & \text{subject to} & X_{ii} = 1 & i = 1, 2, \ldots, n \\ & & X \succeq 0 & \\ & & \text{rank}(X) = 1 & \end{array}$$

Optimization variable: $X \in \mathbb{R}^{n \times n}$

Removing the rank constraint yields the SDP:

$$
\begin{array}{llll}
\text{SDP} = & \max & \frac{1}{4} 1_n^\top W 1_n - \frac{1}{4} \mathrm{tr}(WX) & \\
& \text{subject to} & X_{ii} = 1 & i = 1, 2, \ldots, n \\
& & X \succeq 0 &
\end{array}
$$

Let $X^\star$ solve the SDP and $z \sim \mathcal{N}(0, X^\star)$

There holds

$$\mathsf{SDP} \geq \mathsf{MAXCUT} \geq \mathrm{E}\left\{\phi(\mathsf{sgn}(z))\right\} \geq 0.87\,\mathsf{SDP}$$