Convex Analysis for Optimization

Olga Kuryatnikova (Erasmus University Rotterdam)
kuryatnikova@ese.eur.nl

September-October 2024
Lecture 3

Course plan

- Week 1: Introduction to convexity
- Week 2: More on convex sets
- Week 3: Dual view of convex sets + more on convex functions
- Week 4: Dual view of convex functions
- Week 5: Duality and optimization
- Week 6: Introduction to algorithms, descend methods
- Week 7: Proximal methods, projected gradients
- Weeks 8 - 9: Fix point approach, averaged operators

Special cones

Convex cones based on a given set:

- Conic hull (smallest convex cone containing a given set)
- Recession cone (determines directions of unboundedness of a set)
- Polar cone (dual description of a set)
- Dual cone (dual description of a set)
- Normal cone (dual descriptions, optimality conditions)

Polar and dual cones of set $S \subseteq \mathbb{R}^n$

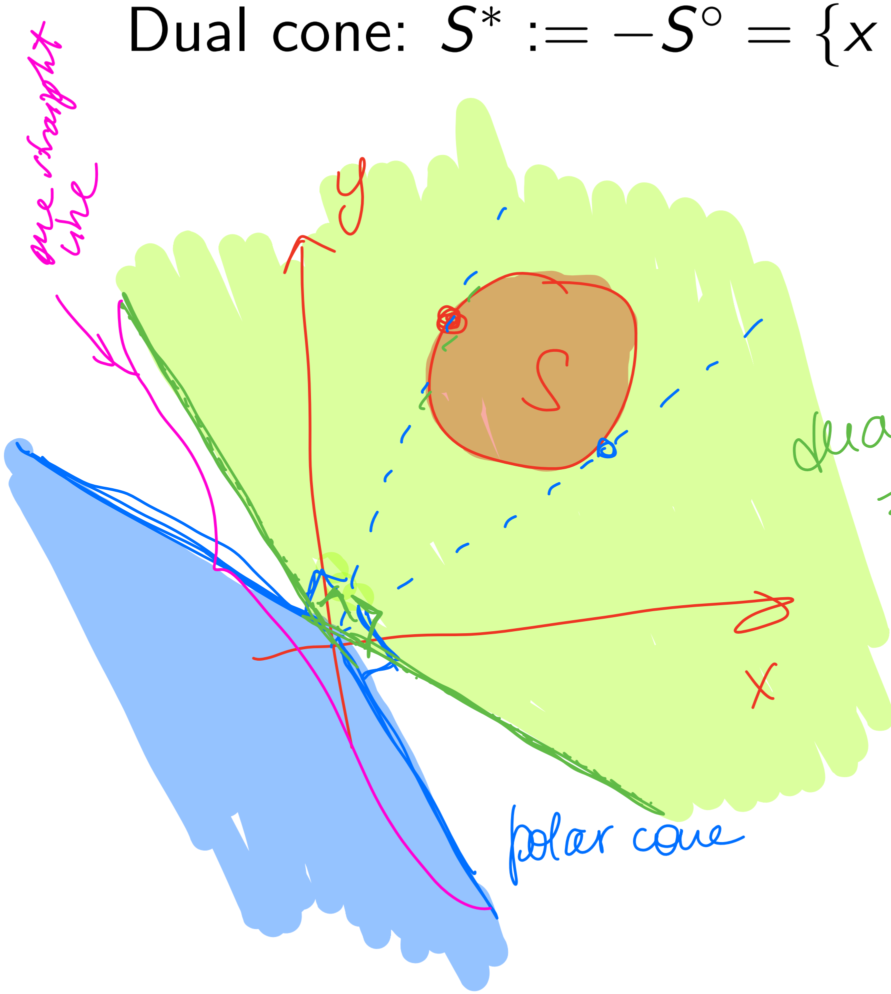
$\cos(\angle(x,y)) \le 0 \Rightarrow \angle(x,y) \in [\frac{\pi}{2}, \frac{3\pi}{2}]$

$= [90^\circ, 180^\circ]$

Polar cone: $S^\circ := \{x \in \mathbb{R}^n : x^\top y \le 0 \ \forall y \in S\}$

Dual cone: $S^* := -S^\circ = \{x \in \mathbb{R}^n : x^\top y \ge 0 \ \forall y \in S\}$

$\cos(\angle(x,y)) \ge 0 \Rightarrow \angle(x,y) \in [-\frac{\pi}{2}, \frac{\pi}{2}]$



dual cone = − polar cone

cones consisting of directions used to construct hyperplanes separating convex sets

Polar and dual cone properties

Those can be proven by definitions

For a non-empty set S (so, no convexity or closedness assumed):

$*$

$$S^{\circ} = \text{cl}(S)^{\circ} = \text{conv}(S)^{\circ} = \text{cone}(S)^{\circ}$$

Important result for duality, see Proposition 2.2.1 in Textbook for the proof.

Polar Cone Thm: $(S^{\circ})^{\circ} = \text{cl}(\text{conv}(S))$ for a non-empty cone S

All above also holds for the dual cone (i.e., if we replace $\circ$ with $*$).

Seen in LP Duality: the dual of the dual is the primal.

In LP we optimize over $Ax - b \geq 0 \Leftrightarrow Ax - b \in \mathbb{R}^m_+$, so we optimize over the cone $\mathbb{R}^m_+$. The dual of this cone is $\mathbb{R}^m_+$ as well, check that. ($\mathbb{R}^m_+$ is self-dual)

Normal cone

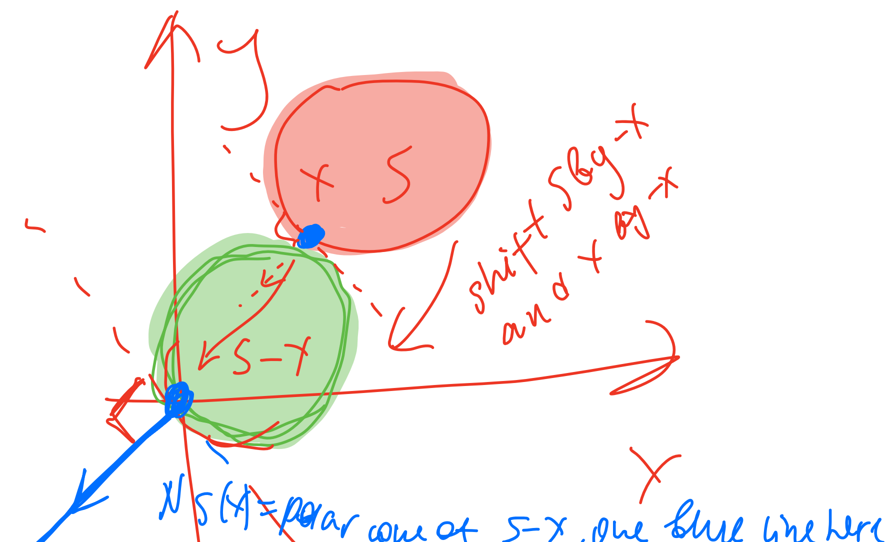
Def: the normal cone of $S \subseteq \mathbb{R}^n$ in $x \in S$ is

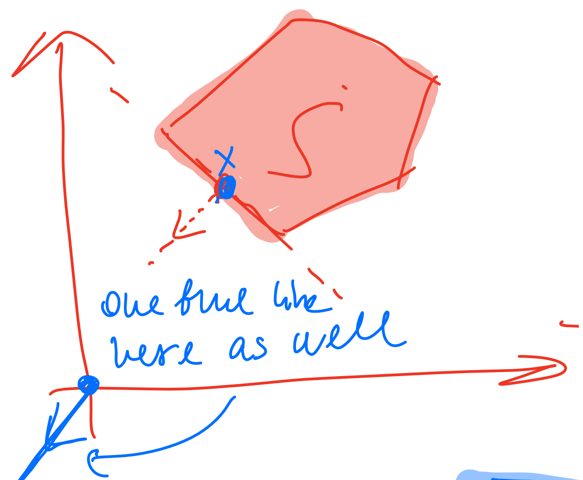
$$N_S(x) := \{y \in \mathbb{R}^n : y^\top (z - x) \leq 0 \ \forall z \in S\}.$$

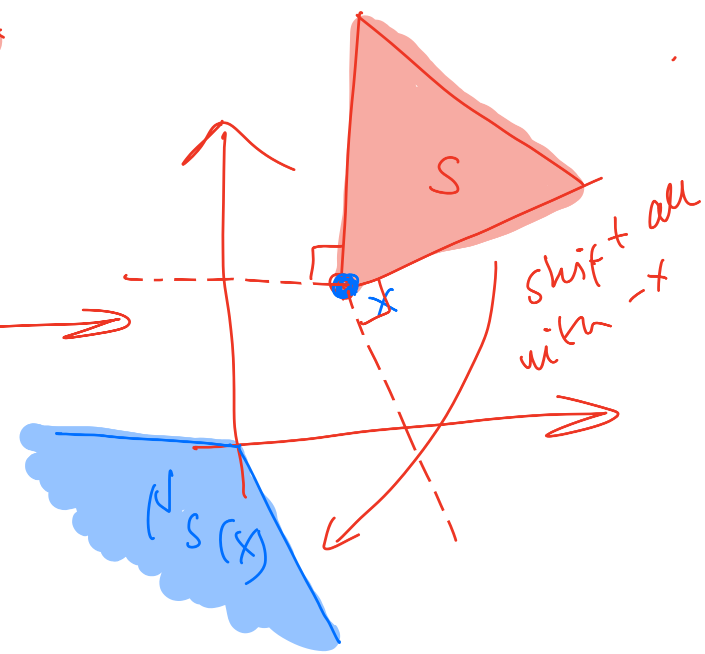
follows from separation theorems next

That is, the normal cone of S in x is $(S - x)^\circ$.

$N_S(x) = \{0\}$ if $x \in \mathrm{int}(S)$; $N_S(x)$ contains at least one half-line otherwise.







Occurs in duality, optimality conditions.

Hyperplanes

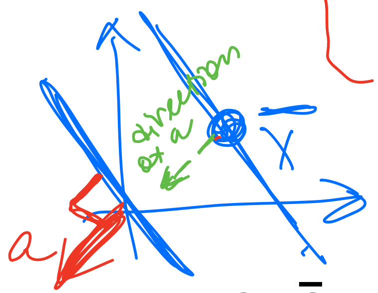
Recall: hyperplane for some $0 \neq a \in \mathbb{R}^n,\ b \in \mathbb{R}$

$$H := \{x \in \mathbb{R}^n : a^\top x = b\} = \bar{x} + \{x \in \mathbb{R}^n : a^\top x = 0\} \text{ for some } \bar{x} \in H$$

$$a^\top x = b,\ a^\top \bar{x} = b \iff \begin{cases} a^\top (x - \bar{x}) = 0, \\ \bar{x} \in H \end{cases} \implies H = \bar{x} + \{x : a^\top x = 0\}$$

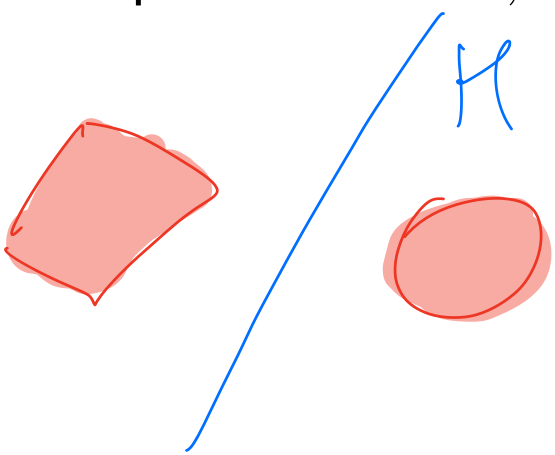
H goes through $\bar{x}$

H is normal to a



$H : a^\top (x - \bar{x}) = 0$

Def: H separates sets $S, \bar{S} \subseteq \mathbb{R}^n$ if $a^\top y \leq b \leq a^\top x,\ \forall y \in \bar{S}\ \forall x \in S$

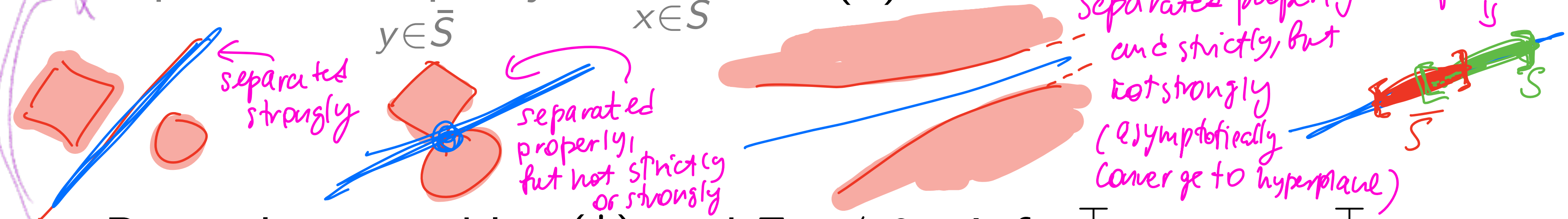


We can always take $a = -a$, and it is still normal to H, so the above inequalities could be reversed, both is fine, it's a matter of choice.

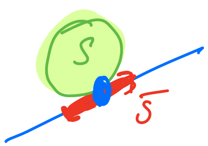
Set separation by hyperplane

$S, \bar{S} \subseteq \mathbb{R}^n$ separable: $\exists a \neq 0, b : a^\top y \leq b \leq a^\top x,\ \forall y \in \bar{S}\ \forall x \in S$.

Equiv: $\exists a : \sup_{y \in \bar{S}} a^\top y \leq \inf_{x \in S} a^\top x$ (*)

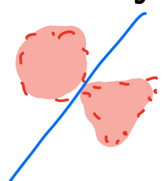


Properly separable: (*) and $\exists a \neq 0 : \inf_{y \in \bar{S}} a^\top y < \sup_{x \in S} a^\top x$

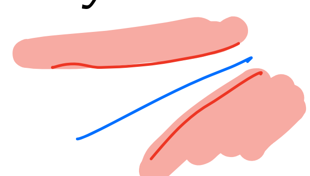


these are also properly separated as only one lies fully inside the hyperplane (the strict inequality above tells us there is a point in $\bar{S}$ or S where the inequality holds strictly)

Strictly separable: $\exists a \neq 0, b : a^\top y < b < a^\top x,\ \forall y \in \bar{S}\ \forall x \in S$



strictly separated (sets are open)

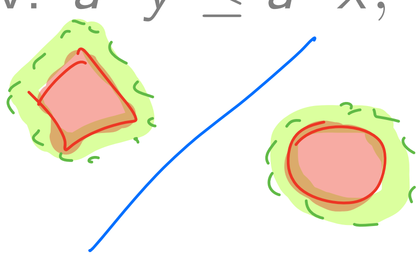


strictly separated (asymptotically converge to hyperplane)

Strongly separable: $\exists a \neq 0 : \sup_{y \in \bar{S}} a^\top y < \inf_{x \in S} a^\top x$

Equiv: $a^\top y \leq a^\top x,\ \forall y \in \bar{S} + B(0, \varepsilon)\ \forall x \in S + B(0, \varepsilon)$ for some $\varepsilon > 0$

open ball with center 0, radius ε



strongly separated = separated even if we expand them a bit in all directions.

implies

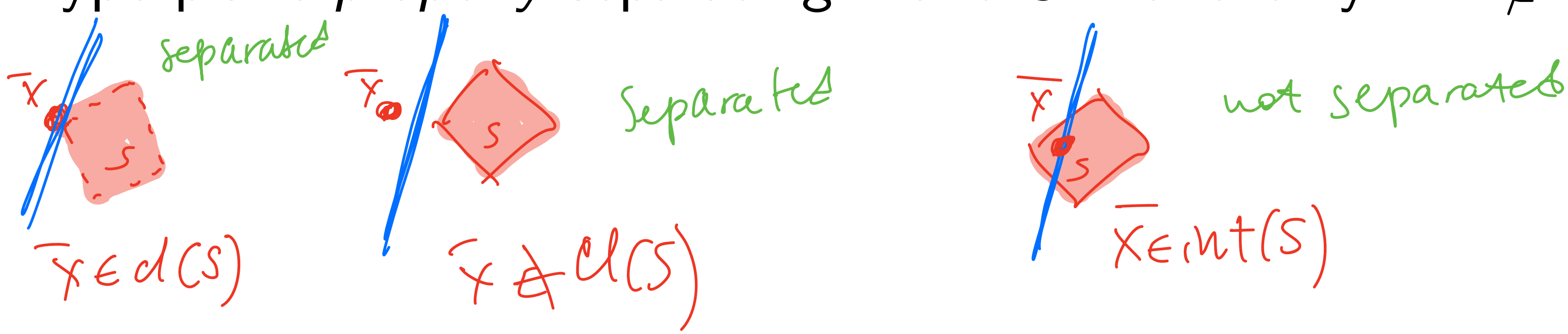
implies

implies

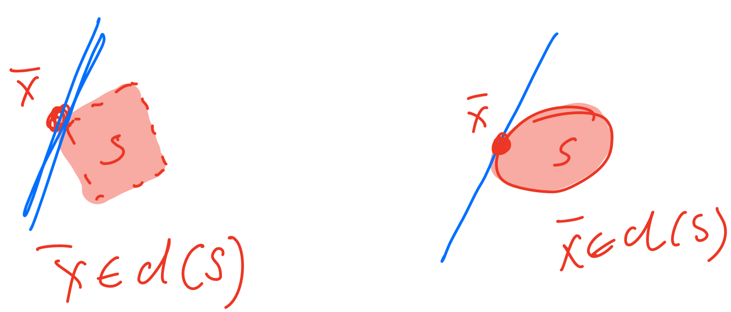
implies

Proper Separation Theorem

Thm: Let $S \subseteq \mathbb{R}^n$ be nonempty and convex, and let $\bar{x} \in \mathbb{R}^n$. There is a hyperplane *properly* separating $\bar{x}$ and S if and only if $x \notin \mathrm{ri}(S)$.

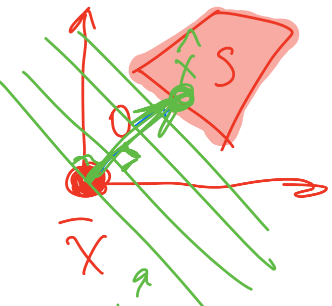


Def: hyperplane $H := \bar{x} + \{x \in \mathbb{R}^n : a^\top x = 0\}$ supports set S in point $\bar{x}$ if $\bar{x} \in \mathrm{cl}(S)$ and $\inf_{x \in S} a^\top x = a^\top \bar{x}$ ($i.e.$, $a^\top \bar{x} \leq a^\top x \; \forall x \in S$).



Both are supporting

Proof of Proper Separation Theorem



separating hyperplanes with normal $\hat{x}$.

Assume $\bar{x}=0$: we can shift $S \to S-\bar{x}$, $\bar{x}\to\bar{x}-\bar{x}$ and then shift back.

① Argue that $0 \in ri(S) \Rightarrow$ cannot properly separate 0 from S (either $S=\{0\}$ or by line segm. principle and prolongation lem. $0\in ri(S)$, $y\in ri(S) \Rightarrow \alpha y, -\alpha y \in S$ for some $\alpha\in(0,1)$).

② Now, let $0\notin ri(S)$. Simple case: $\overset{(=\bar{x})}{0} \notin cl(S) \Leftrightarrow \inf_{x\in S}\|x\|_2 > 0 \Rightarrow \exists\, \hat{x}\in cl(S)$ with minimal norm and $\hat{x}\neq 0$

Take any hyperplane with direction $\frac{\hat{x}}{\|\hat{x}\|_2} := a$.

We know: $a^T\bar{x} = a^T 0 = 0$. Show: $a^Ty \ge 0$ for all $y\in ri(S)$ $\Rightarrow$ $\inf_{y\in S} a^Ty \ge \sup_{x\in\{0\}} a^Tx = 0$,

follows from line segment principle

If $\dim(S)=0$, so $S=\{\hat{x}\}$, the result follows as a^Ty for $y\in S$ is $\|\hat{x}\|_2^2>0$.

If $\dim(S)>0$, use contradiction: show that if $a^Ty<0$, $\hat{x}$ does not have minimal norm.

Take $y\in ri(S)$, $y\neq\hat{x}$, $\forall\alpha\in[0,1)$ we have $z=\alpha\hat{x}+(1-\alpha)y\in S$

show: $z^Tz < \hat{x}^T\hat{x}$

$$z^Tz-\hat{x}^T\hat{x} = (\alpha^2-1)\hat{x}^T\hat{x} + \overbrace{2\alpha(1-\alpha)}^{1-\alpha^2-(1-\alpha)^2}\hat{x}^Ty + (1-\alpha)^2y^Ty$$

$$= (\alpha^2-1)(\hat{x}^T\hat{x}-\hat{x}^Ty) + (1-\alpha)^2(y^Ty-\hat{x}^Ty)$$

$$= (1-\alpha)\Big((1-\alpha)\underbrace{(y^Ty-\hat{x}^Ty)}_{=\varepsilon>0} - (1+\alpha)\underbrace{(\hat{x}^T\hat{x}-\hat{x}^Ty)}_{=\delta>0}\Big) < 0$$

for some α close to 1, check $\alpha=\frac{\varepsilon}{\delta+\varepsilon}$

$\varepsilon>0$, by Cauchy-Schwartz inequality and minimality of norm $\hat{x}$.

$\delta>0$, by assumption $\hat{x}^Ty<0$

When we shift back, the hyperplane equation changes from $a^Ty=0$ to $a^T(y-\bar{x})=0 \Rightarrow$ a belongs to $N_S(\bar{x})$

Harder case: $\overset{(=\bar{x})}{0} \in cl(S) \setminus ri(S)$. Then $\inf_{x \in S} x^T x = 0$, so we cannot take $\hat{x} = 0$ and $a = \frac{\hat{x}}{\|\hat{x}\|}$. Thus, we take any sequence $\tilde{x}_k \xrightarrow{k\to\infty} 0 = \bar{x}$, $\tilde{x}_k \notin cl(S)$. From the previous reasoning, we can separate $\tilde{x}_k$ from S by hyperplanes with normal directions $a_k = \frac{\hat{x}_k}{\|\hat{x}_k\|}$, $\hat{x}_k = \underset{x \in S - \{\tilde{x}_k\}}{\arg\inf} \|x\| \neq 0$.

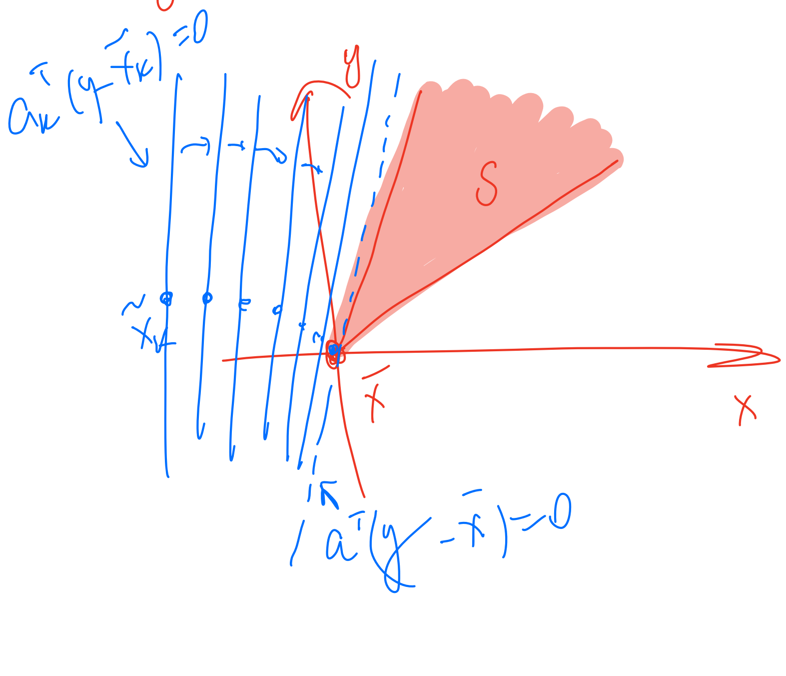
Then $a_k^T y \leq 0$ for all $y \in ri(S - \{\tilde{x}_k\})$, for all k

$\Rightarrow$ $a_k^T (y - \tilde{x}_k) \leq 0$ for all $y \in ri(S)$, for all k

$\Rightarrow$ $\lim_{k\to\infty} a_k^T (y - \tilde{x}_k) = a^T (y - \bar{x}) \leq 0$ for all $y \in ri(S)$

and thus for all $y \in cl(S)$ as those are limit points of $y \in ri(S)$ by the segment principle.

Here we may assume that $\lim_{k\to\infty} a_k$ exists as $\|a_k\| = 1$, so we get a bounded sequence and can restrict it to some convergent subsequence.



Separating two sets

not always necessary

Let $S, \bar{S} \subseteq \mathbb{R}^n$ be nonempty, convex, and disjoint.

Thm (separation): There is a hyperplane separating S and $\bar{S}$. disjoint is sufficient but not necessary; could have $0 \in ri(S-\bar{S})$ or $S, \bar{S} \subseteq$ one hyperplane

Idea of proofs: separate $S-\bar{S}$ and 0, use the proof from page before (or any other proof of separating a set from a point)

Thm (proper s.): There is a hyperplane *properly* separating S and $\bar{S}$ if and only if $\mathrm{ri}(S) \cap \mathrm{ri}(\bar{S}) = \emptyset$. → disjoint is sufficient but not necessary, condition m is weaker.

Follows from proper separation theorem and $0 \notin ri(S-\bar{S}) = ri(S) - ri(\bar{S})$ iff $ri(S) \cap ri(\bar{S}) = \phi$

Thm (strict s.): There is a hyperplane *strictly* separating S and $\bar{S}$ if $S - \bar{S}$ is closed. → need disjoint

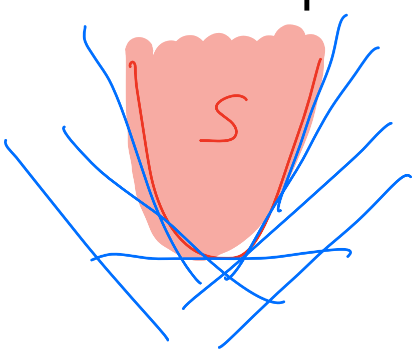
Same idea of proof as Proper sep. Thm, use disjointness and closedness of $S-\bar{S}$, so $0 \notin cl(S-\bar{S})$, so $\hat{x} = \arg\min_{x \in S-\bar{S}} \|x\| \neq 0$, take $a = \frac{\hat{x}}{\|\hat{x}\|}$, $b = \frac{\|\hat{x}\|}{2}$ $\Rightarrow \underbrace{a^T x \geq \frac{\hat{x}^T\hat{x}}{\|\hat{x}\|} > b \text{ for } x \in S-\bar{S} \text{ and } a^T x < b \text{ for } x = 0}_{\text{Exercise: check}}$

Thm (strong s.): There is a hyperplane *strongly* separating S and $\bar{S}$ if and only if $0 \notin \mathrm{cl}(S - \bar{S})$. → don't even need disjoint, it follows from the condition

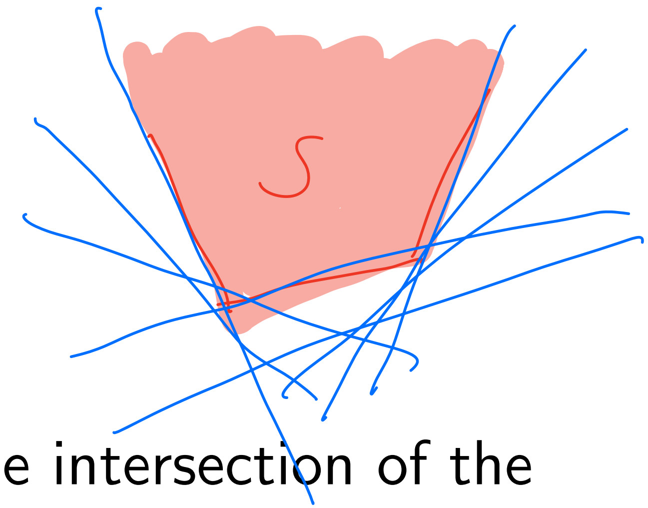
Same idea + need to introduce a small ball to show that the sets shifted by that ball can still be separated

Dual description of convex sets

Thm: For $S \subseteq \mathbb{R}^n$, the set cl(conv(S)) is the intersection of the closed halfspaces that contain S.



$$S = \bigcap_{y \notin \mathrm{int}(S)} \left[y + \{x : a_y^T x \geq 0\} \right]$$



Corollary: If S is closed and convex, S is the intersection of the closed halfspaces that contain it.

Proof idea: a halfspace is defined by a hyperplane, and $\overbrace{cl(conv(S))}^{\bar{S}}$ is closed and convex. If $x \in \bar{S} \Rightarrow$ all halfspace equations hold for x.

If $x \notin \bar{S}$, at least one half-space is defined by a hyperplane separating x from $\bar{S}$.

More info on convex functions

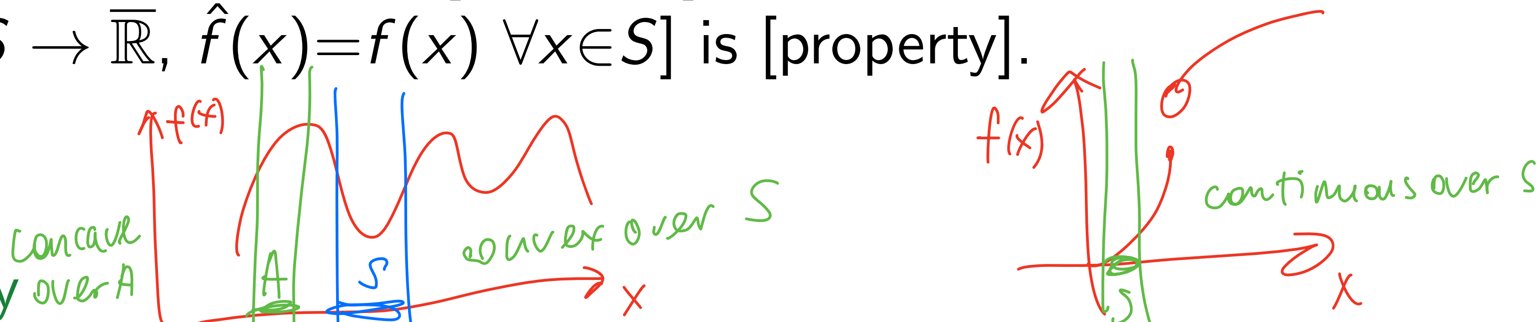
- Popular convex functions
- Convexity preserving operations on functions
- Continuity and closedness
- Differentiable convex functions (for next lecture)

Some definitions

Consider $f : \mathbb{R}^n \to (-\infty, \infty]$.

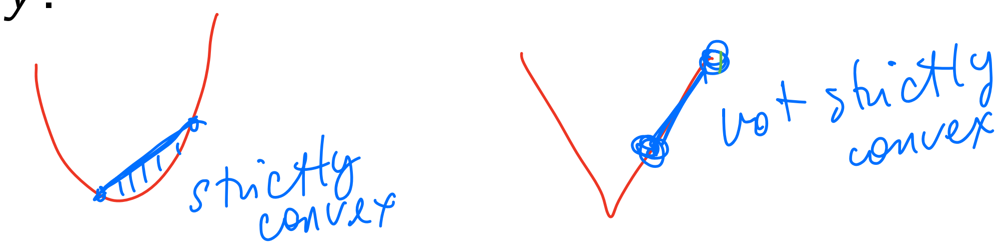
Convexity/continuity/some other property over a set

Def: Let $C \subseteq S \subseteq \mathbb{R}^n$, then f is [property] over S if its restriction to S defined by $[\hat{f} : S \to \overline{\mathbb{R}},\ \hat{f}(x) = f(x)\ \forall x \in S]$ is [property].



Strict convexity

Def: f is strictly convex if $f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y)$ for $\alpha \in (0,1)$, $x \neq y$.



Strong convexity

Def: f is strongly convex if $f - \sigma \|x\|_2^2$ is convex for some $\sigma > 0$.

"Very convex" i.e., at least as convex as Euclidean norm (2-norm).
An affine function is not strongly convex, for example.

Popular convex functions

- Affine functions: $a^\top x + b$ or $\sum_{i=1}^{n} A_{ij} x_{ij} + b$ if x is a matrix
- Norms: l_p norm for $p \geq 1$: $(\sum_{i=1}^{n} = |x|^{1/p})^p$, ∞-norm: $\max_{i=1}^{n} |x_i|$;
  spectral norm: $\sigma_{\max}(x) = (\lambda_{\max}(x^\top x))^{1/2}$ if x is a matrix
- Sums of squares of polynomials: $\sum_{j=1}^{m} (p_j(x))^2$
- Max: $\max_{i=1}^{n} x_i$
- Log-sum-exp: $\log(\sum_{i=1}^{n} \exp(x_i))$
- log-determinant: -log(det(x)) is convex on the set of positive definite matrices x

Convexity preserving operations on functions

Consider functions $\mathbb{R}^n \to (-\infty, \infty]$. The following is convex.

- Sum of convex functions, even of infinitely many functions
- For convex f: $f(Ax + b)$, $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^n$
- Conic combination of convex functions

frequently occuring case in optimization, proof by definition.

- Supremum: $\sup_{i \in I} f_i(x)$, $\sup_{y \in Y} f(x, y)$ if f is convex in x for all y

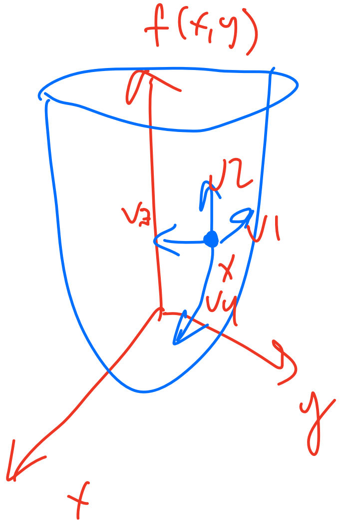
interesting to prove case, see Proposition 3.3.1 in textbook.

- Partial inf: $\inf_{y \in Y} f(x, y)$ if Y is convex and f is convex in (x, y)
- Integral: $\int_{y \in Y} f(x, y) dy$ if f is convex in x for all y
- Composition: $f(g_1(x), \ldots, g_n(x))$ is convex if f is convex and for each $i = 1, \ldots, n$ at least one of three facts holds: [g_i convex, f non-decreasing in x_i]; [g_i concave, f non-increasing in x_i]; [g_i affine]

example $\underset{\text{decreasing}}{f(x) = -x}$, $\begin{cases} g(x) = x^2 \Rightarrow f(g(x)) = -x^2 \text{ is concave} \\ h(x) = -x^2 \Rightarrow f(h(x)) = x^2 \text{ is convex} \end{cases}$

Restricting a convex function to a line

Thm: $f: \mathbb{R}^n \to (-\infty, \infty]$ is convex if and only if its restriction to a line $g_{x,v}(t)$ is convex for any fixed x, v, where $g(t) := f(x + tv)$.



Can follow the function in any direction, and the resulting line is convex. Follows from the definition of convexity.

Types of continuity

Consider a function $f\colon S \to \overline{\mathbb{R}}$

Def: f is lower semicontinuous in x if $f(x) \leq \liminf_{y\to x} f(y), \forall (y) \subset S$.

Def: f is continuous in $x \in \mathrm{dom}(f)$ if $f(x) = \lim_{y\to x} f(y), \forall (y) \subset \mathrm{dom}(f)$

Def: f is Lipshitz-continuous with constant $L > 0$ if
$\|f(x) - f(y)\|_2 \leq L\|x - y\|_2$ for all $x, y \in \mathrm{dom}(f)$

Semicontinuity and closedness

Def: $f : S \to \overline{\mathbb{R}}$ is closed if its epigraph epi(f) is a closed set.

Thm: Function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is closed if and only if
$\iff$ f is lower-semicontinuous
$\iff$ level set $V_\gamma =: \{x \in \mathbb{R}^n : \gamma \geq f(x)\}$ is closed for any $\gamma \in \mathbb{R}$

Continuity and convexity

Thm: $f : S \to \overline{\mathbb{R}}$ proper and convex $\Rightarrow$ f continuous over $\mathrm{ri}(\mathrm{dom}(f))$.

Corollary: A convex function $\mathbb{R}^n \to \mathbb{R}$ is continuous.

If there is time: Farkas' lemma as an example of using polar cones

Farkas' lemma

Let $a_1, \ldots, a_m \in \mathbb{R}^n$. Then $\{x \in \mathbb{R}^n : a_j^\top x \geq 0 \ \forall j = 1, \ldots, m\}$ and cone$(a_1, \ldots, a_m)$ are closed convex cones dual to each other.

Note: Textbook uses $a_j^\top x \leq 0$, and so the cones become polar.

Interpretations of Farkas' lemma

Let $c, a_1, \ldots, a_m \in \mathbb{R}^n$. Then $c^\top x \geq 0$ for all $x \in S$, where

$$S := \{x \in \mathbb{R}^n : \underbrace{a_j^\top x \geq 0 \; \forall j = 1, \ldots, k, \;\; a_i^\top x =}_{\text{can write this as } A^\top x \in \mathbb{R}_+^k \times \{0\}^{m-k}} 0 \; \forall i = k+1, \ldots, m\}$$

if and only if

$$c = \sum_{j=1}^{k} a_j y_j + \sum_{i=k+1}^{m} a_i y_i \text{ for some } \underbrace{y \in \mathbb{R}^m, \; y_1, \ldots, y_k \geq 0}_{\text{can write this as } y \in \mathbb{R}_+^k \times \mathbb{R}^{m-k}}. \quad (\#)$$

Interpretation as a theorem of alternatives:

Either $c^\top x < 0$ or c has representation (#).

Interpretation as simplification of optimization:

instead of "for all $x \in S$" (hard constraint) get "$\exists\, y$ with repres. (#)" (easier constraint).

Generalized Farkas' lemma

Let $K \subseteq \mathbb{R}^m$ be a closed convex cone, $c \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times m}$. Let the cone $\{Ay : y \in K^*\}$ be closed. Then $x^\top c \geq 0$ for all $x \in S$, where

In the previous slide
$K^* = \mathbb{R}_+^k \times \mathbb{R}^{m-k}$,
$K = \mathbb{R}_+^k \times \{0\}^{m-k}$.

$$S := \{x \in \mathbb{R}^n : A^\top x \in K\}$$

if and only if

$$c = Ay \text{ for some } y \in K^*.$$