Convex Optimization in Machine Learning and Computational Finance Lecture 2: The SDP Cone, Separation theorems

Dr. Michel Baes, Pr. Patrick Cheridito

RiskLab / ETH Zürich

Quick recall of last week's lecture

- Convexity (usually) makes a problem solvable in practice
- Convexity is sometimes difficult to spot

- Topological concepts (relint)
- Affine/convex/conic sets
- Examples of affine/convex/conic sets:
  balls, (and thus ellipsoids), hyperplanes,
  half-spaces, polyhedron, k-simplex
- Construction of convex sets from convex sets:
  intersection, Minkowski sum, light cone

A brief reminder on square matrices

Let $A \in \mathbb{R}^{n \times n}$. The identity matrix is denoted by I.
det(A) is the determinant of A.

The *eigenvalues* of A are roots of $p_A(t) := \det(tI - A)$.
Note: p_A is a real polynomial of degree n.
Hence the Fundamental Theorem of Algebra asserts that p_A has n complex roots $\lambda_1, \ldots, \lambda_n$.
As $[\lambda_i I - A]$ is degenerated, $K_i = \ker(\lambda_i I - A) \neq \{0\}$.
We call every $x_i \in K_i$ an *eigenvector* of A w.r.t. to λ_i:

$$Ax_i = \lambda_i x_i.$$

The *trace* of A is $\mathrm{Tr}(A) := \sum_i \lambda_i = \sum_i [A]_{ii}$, always in $\mathbb{R}$.
Exercise: prove the red assertions

A brief recall on symmetric matrices 1

Let $\mathbb{S}^n := \{X \in \mathbb{R}^{n\times n} : X = X^T\}$
be the set of symmetric $n \times n$ matrices.

> We consider in this slide each matrix as a point of the space $\mathbb{R}^{n\times n}$.

- $\mathbb{S}^n$ is a subspace of $\mathbb{R}^{n\times n}$ of dimension $\frac{n(n+1)}{2}$.
- Basis of $\mathbb{S}^2$:

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\}.$$

- **A scalar product for $\mathbb{S}^n$:** let $\langle X, Y\rangle_F := \mathrm{Tr}(XY)$.
  Then $\langle X, X\rangle_F > 0$ for $X \neq 0$ and $\langle \cdot, \cdot\rangle_F$ is bilinear.
  We call it the *Frobenius scalar product*
- We have $\langle X, uu^T\rangle_F = u^T X u$.

A brief recall on symmetric matrices 2

Spectral Theorem:
The eigenvalues $\lambda_i(X)$ of every $X \in \mathbb{S}^n$ are real.
It is possible to construct an *orthonormal basis* of real eigenvectors of X, that is:

$$||v_i||_2 = 1, \quad v_i^T v_j = 0 \text{ if } i \neq j, \quad Xv_i = \lambda_i(X)v_i.$$

Proof: Denoting by $\bar{z}$ the complex conjugate of $z \in \mathbb{C}$ and letting $\langle a, b\rangle := \sum_j \bar{a}_j b_j$ for $a, b \in \mathbb{C}^n$, we consider v_i an eigenvector w.r.t. $\lambda_i(X)$ such that $\langle v_i, v_i\rangle = 1$. We have:

$$\lambda_i(X) = \langle v_i, Xv_i\rangle = \overline{\langle Xv_i, v_i\rangle} = \overline{\lambda_i(X)},$$

thus $\lambda_i(X)$ is real, and we can take v_i real. Assume now that $\lambda_i(X) \neq \lambda_j(X)$. Then

$$\lambda_j(X)v_i^T v_j - \lambda_i(X)v_j^T v_i = v_i^T X v_j - v_j^T X v_i = 0,$$

and $v_j^T v_i = 0$. Therefore, the eigenspaces $\{v : Xv = \lambda_i(X)v\}$ are mutually orthogonal. It remains to take an orthonormal basis for each of these eigenspaces.

A brief recall on symmetric matrices 2

Spectral Theorem:

The eigenvalues $\lambda_i(X)$ of every $X \in \mathbb{S}^n$ are real.
It is possible to construct an *orthonormal basis* of eigenvectors of X, that is:

$$||v_i||_2 = 1, \quad v_i^T v_j = 0 \text{ if } i \neq j, \quad X v_i = \lambda_i(X) v_i.$$

Two useful reformulations:

1. There exists an orthogonal matrix V and a diagonal matrix Λ such that $X = V \Lambda V^T$.
$V = [v_1, \ldots, v_n]$, $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$.

2. $X = \sum_{i=1}^n \lambda_i P_i$, where P_i is a *projector of rank one*, i.e. $P_i^2 = P_i$, and $\text{rank}(P_i) = 1$. $P_i = v_i v_i^T$.

There is also convexity in symmetric matrices

The *Positive Semidefinite Cone* $\mathbb{S}^n_+$,
or *Semidefinite Programming Cone (SDP Cone)* is:

$\{X \in \mathbb{S}^n : u^T X u \geq 0 \, \forall u \in \mathbb{R}^n\} = \cap_{u \in \mathbb{R}^n} \{X \in \mathbb{S}^n : \langle X, uu^T \rangle_F \geq 0\}$,

and is closed and convex, as intersection of halfspaces.

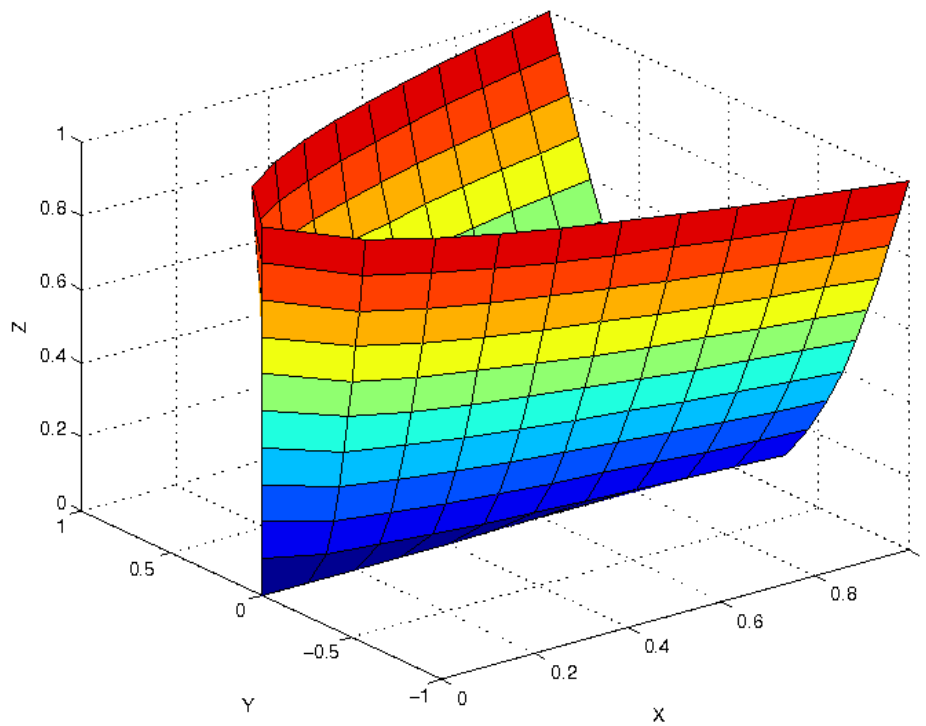
The interior $\mathbb{S}^n_{++}$ of $\mathbb{S}^n_+$ is:

$$\{X \in \mathbb{S}^n : u^T X u > 0 \, \forall u \in \mathbb{R}^n \backslash \{0\}\}.$$

Note: Note that $\lambda_{\min}(X) \geq 0$ iff $X \in \mathbb{S}^n_+$.
and $\lambda_{\min}(X) > 0$ iff $X \in \mathbb{S}^n_{++}$.

Illustration: the cone $\mathbb{S}^2_+$

$$\begin{pmatrix} x & y \\ y & z \end{pmatrix} \in \mathbb{S}^2_+ \quad \text{iff} \quad x, z \geq 0, xz \geq y^2.$$



There are many applications using the SDP cone. We will meet a handful of them later.

Convex functions: a very first glimpse

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$.

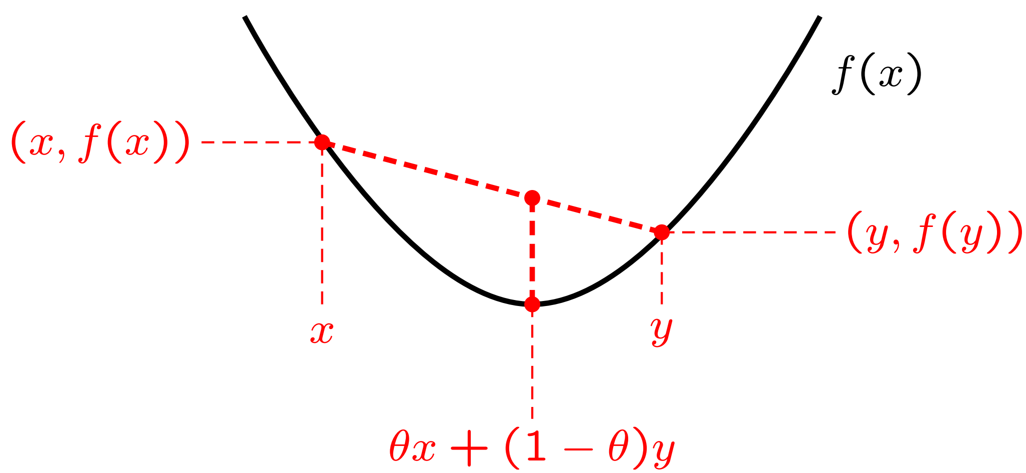
The *domain* of f is dom$f := \{x \in \mathbb{R}^n : f(x) < +\infty\}$.

[Also called "effective domain"]

The *epigraph* of f is epi$f := \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq t\}$.

f is *convex* if epi(f) is convex

$\Leftrightarrow$ dom$f$ convex and $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$
for all $x, y \in$ dom$f$, $0 \leq \theta \leq 1$.



Some convex functions constructed from convex sets

Let $Q \subseteq \mathbb{R}^n$ be a convex set.

1. The *characteristic function* of Q is:

$$\chi_Q(x) := \begin{cases} 0, & \text{if } x \in Q \\ +\infty, & \text{otherwise.} \end{cases}$$

2. Suppose that $0 \in Q$.
The *Minkowski function* of Q is:

$$\mu_Q(x) := \inf\{t > 0 : x \in tQ\}.$$

[Also called Minkowski functional]

Note: epiμ_Q is the light cone of Q, which is a convex cone.
μ_Q is *positively homogenous*: $\mu_Q(tx) = t\mu_Q(x)$
and *subadditive*: $\mu_Q(x+y) \leq \mu_Q(x) + \mu_Q(y)$
for $t > 0$, $x, y \in \mathbb{R}^n$.
$\mu_Q(x) < 1$ for $x \in \operatorname{int} Q$.

The supremum of convex functions is convex

Remember:

The intersection of convex sets is convex.

Note: Works (also) for uncountable intersections:
e.g. if U_α is convex $\forall \alpha \in \mathbb{R}$, then $\cap_{\alpha \in \mathbb{R}} U_\alpha$ is convex.

Let A be any set
and $(f_\alpha)_{\alpha \in A}$ be convex functions $\mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$.

We have:

$$\bigcap_{\alpha \in A} \text{epi}(f_\alpha) = \text{epi}\left(\sup_{\alpha \in A} f_\alpha\right)$$

Therefore $x \mapsto \sup_{\alpha \in A} f_\alpha(x)$ is convex.

Constructing convex functions from convex functions

Many other constructions preserve convexity

Let $f, g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex

- **Sum.** $f + g$ is convex
- **Positive scaling.** If $\alpha \geq 0$, then αf is convex
- **Composition.** Let $h : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$.
  If h is convex and increasing, then $x \mapsto h(f(x))$ is convex.
- **Infimum convolution.** The function

$$f \,\square\, g(x) := \inf\{f(y) + g(x - y) : y \in \mathbb{R}^n\})$$

is convex [Hint: $\text{epi}(f \,\square\, g) = \text{epi}(f) + \text{epi}(g)$]

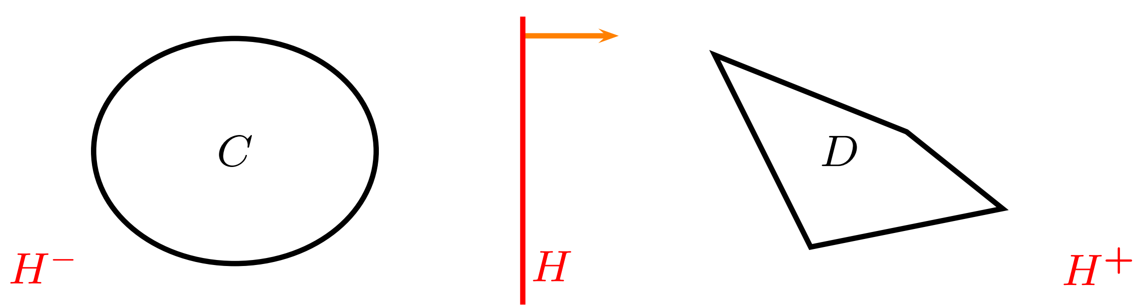
Separation theorems

General idea

Let $\langle \cdot, \cdot \rangle$ be a scalar product of $\mathbb{R}^n$.
Let $C, D \subseteq \mathbb{R}^n$ nonempty convex sets,
such that relint$(C) \cap$ relint$(D) = \phi$.

> Does there exist a hyperplane
> $H := \{x : \langle a, x \rangle = b\}$, $a \neq 0$,
> *separating* C from D?

i.e. letting $H^- := \{x : \langle a, x \rangle \leq b\}$, $H^+ := \{x : \langle a, x \rangle \geq b\}$,
then $C \subseteq H^-$ and $D \subseteq H^+$.

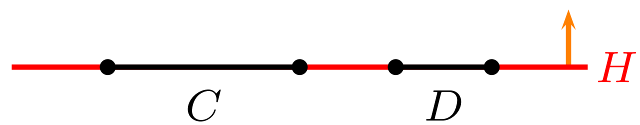


Different kinds of separation

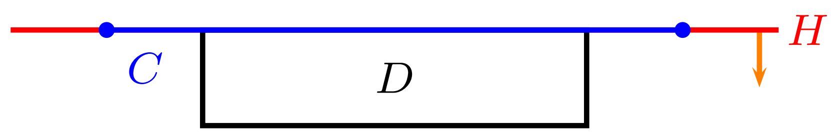
A separation is:

counterexample

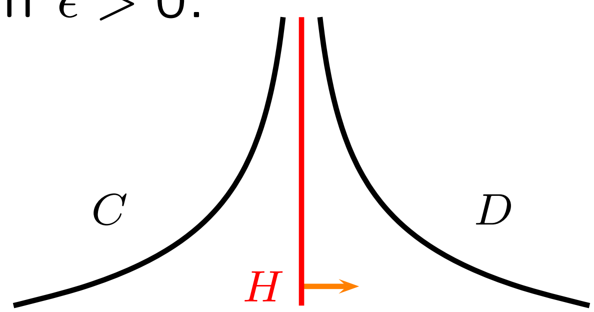
- *proper* if $H \not\supseteq (C \cup D)$



- *strict* if $C \subseteq \text{int}\, H^-, D \subseteq \text{int}\, H^+$



- *strong* if $C + B(0,\epsilon) \subseteq \textbf{int}(H^-)$ and $D + B(0,\epsilon) \subseteq \textbf{int}(H^+)$ for an $\epsilon > 0$.



An obvious rewriting

Does there exist a hyperplane
$H := \{x : \langle a, x\rangle = b\}$, $a \neq 0$,
separating C from D?

is equivalent to

Does there exist a linear function f
and a number α such that
$f(x) \leq \alpha$ for every $x \in C$
and $f(y) \geq \alpha$ for every $y \in D$?

Note: **Proper separation** iff equality does not hold $\forall x, y$.
Strict separation iff inequalities are strict.
Strong separation iff true for all $\alpha \in]\alpha_0 - \epsilon, \alpha_0 + \epsilon[$ for an $\epsilon > 0$.

Equivalent characterizations of separation

Does there exist a linear function f
and a number α such that
$f(x) \leq \alpha$ for every $x \in C$
and $f(x) \geq \alpha$ for every $x \in D$?

Note: **Proper separation** iff

$$\sup\{f(x) : x \in C\} \leq \inf\{f(x) : x \in D\}$$

$$\text{and } \inf\{f(x) : x \in C\} < \sup\{f(x) : x \in D\}.$$

Strong separation iff

$$\sup\{f(x) : x \in C\} < \inf\{f(x) : x \in D\}$$

The Extension Theorem
of Hahn and Banach


The ingredients:

E: vector space

$p : E \to \mathbb{R}$: convex and pos. homogenous

$L \subseteq E$: linear subspace

$f : L \to \mathbb{R}$: linear function

$f(x) \leq p(x)$ for $x \in L$

Theorem: It is possible to extend f to E while preserving its dominance by p: there exists a linear function $\widehat{f} : E \to \mathbb{R}$ such that $\widehat{f}(x) = f(x)$ on L, and $\widehat{f}(x) \leq p(x)$ on E.

and its proof (in finite dimension)

The simplest proof for $\dim(E) = \infty$ requires Zorn's Lemma (transfinite induction), which is beyond the scope of this lecture.

Proof: Assume first that $E = \mathbb{R}^n$, and $\dim(L) = n - 1$.
There is an $e \in E$ such that $E = L + \mathbb{R}e$. Our candidate function $\hat{f}$ will have the form $\hat{f}(x + \alpha e) = f(x) + \alpha\gamma$ for all $x \in L$ and $\alpha \in \mathbb{R}$. We "just" need to find a $\gamma = \hat{f}(e) \in \mathbb{R}$ for which $f(x) + \alpha\gamma \leq p(x + \alpha e)$.
Let:

$$l := \sup_{y \in L} f(y) - p(y - e), \quad u := \inf_{x \in L} p(x + e) - f(x),$$

so that $-p(-e) \leq l \leq u \leq p(e)$ because for all $x, y \in L$:

$$p(x + e) - f(x) - f(y) + p(y - e) \geq p(x + y) - f(x) - f(y) \geq 0.$$

Taking $\gamma := u$, we have for $\alpha > 0$ and $x \in L$:

$$\alpha\gamma = \alpha u \leq \alpha[p(x/\alpha + e) - f(x/\alpha)] = p(x + \alpha e) - f(x),$$

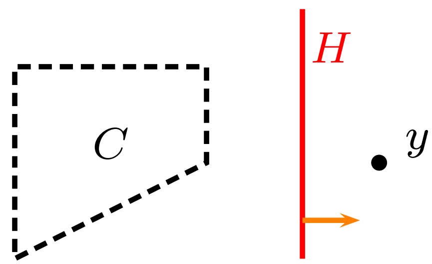
and for $\alpha < 0$ and $y \in L$:

$$|\alpha|\gamma \geq |\alpha| l \geq |\alpha|[f(y/|\alpha|) - p(y/|\alpha| - e)] = f(y) - p(y + \alpha e).$$

Iterating this process, we can extend f when $E = L + \mathbb{R}e_1 + \cdots + \mathbb{R}e_k$.

Separating one point from an open convex set

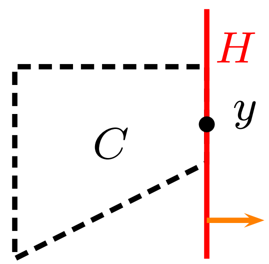
Let $C \subseteq \mathbb{R}^n$ nonempty open convex set,
and $D := \{y\}$, where $y \notin C$.
There exists a linear function $\widehat{f}$
such that $\widehat{f}(x) < \widehat{f}(y)$ for all $x \in C$.



Proof: Wlog, $0 \in C$, so that $y \neq 0$.
Let $p(x) := \mu_C(x)$, the Minkowski function of C,
$L := \mathbb{R}y$, and $f(ty) := t$ for $t \in \mathbb{R}$.
Then $\mu_C(y) \geq 1$ because $y \notin C$. Hence $\mu_C \geq f$ on L.
Note that $\mu_C(x) < 1$ for $x \in C$, because C is open.
By Hahn-Banach, there is a linear $\widehat{f}$ such that
$\widehat{f}(y) = f(y) = 1$ and $\widehat{f}(x) \leq \mu_C(x) < 1$ for $x \in C$.

Separating one point from an open convex set

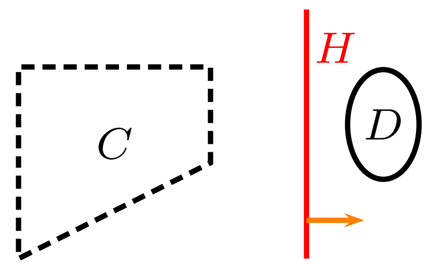
Let $C \subseteq \mathbb{R}^n$ nonempty open convex set,
and $D := \{y\}$, where $y \notin C$.
There exists a linear function $\hat{f}$
such that $\hat{f}(x) < \hat{f}(y)$ for all $x \in C$.



Proof: Wlog, $0 \in C$, so that $y \neq 0$.
Let $p(x) := \mu_C(x)$, the Minkowski function of C,
$L := \mathbb{R}y$, and $f(ty) := t$ for $t \in \mathbb{R}$.
Then $\mu_C(y) \geq 1$ because $y \notin C$. Hence $\mu_C \geq f$ on L.
Note that $\mu_C(x) < 1$ for $x \in C$, because C is open.
By Hahn-Banach, there is a linear $\hat{f}$ such that
$\hat{f}(y) = f(y) = 1$ and $\hat{f}(x) \leq \mu_C(x) < 1$ for $x \in C$.

Separating a convex set from an open convex set

Let $C \subseteq \mathbb{R}^n$ nonempty open convex set,
and $D \subseteq \mathbb{R}^n$ convex such that $C \cap D = \phi$.
There exists a linear function $\widehat{f}$
such that $\widehat{f}(x) < \widehat{f}(y)$ for all $x \in C$, $y \in D$.



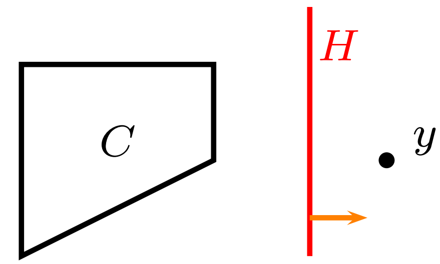
Proof: Let $U := C - D = \{x - y : x \in C, y \in D\}$.
Then $0 \notin U$. Also, U is open because C is.
U is convex as well (it's a Minkowski sum).
Separating U from the point $\{0\}$, we have:

$$\widehat{f}(x - y) < \widehat{f}(0) = 0 \;\; \forall\, x \in C, y \in D,$$

that is $\widehat{f}(x) < \widehat{f}(y)$.

Immediate extensions: Strong separation

Let $C \subseteq \mathbb{R}^n$ convex set with nonempty interior
and $D = \{y\}$ convex such that $y \notin \mathrm{cl}(C)$.
Then C and D can be strongly separated.



Proof: Exactly the same idea.
Here, $\mu_C(y) > 1$, and strong separation follows.

Corollary: Let $C \subseteq \mathbb{R}^n$ convex set with nonempty interior
and D convex such that $0 \notin \mathrm{cl}(C - D)$.
Then C and D can be strongly separated.

Strong separation also holds when C and D are closed,
C or D is compact and $C \cap D = \emptyset$ (see exercise)

"Immediate" extensions : Affine separation

Let $C \subseteq \mathbb{R}^n$ convex set with relint$(C) = C \neq \phi$
and D affine $C \cap D = \phi$.
There exists a hyperplane H of dimension $n-1$
such that $D \subseteq H$, and $C \subseteq \text{int}\, H^+$.

Note: Nothing new for $D = \{y\}$ and relint$(C) = \text{int}\,(C)$.

Proof (sketch): Assume $0 \in C$, let $x_0 \in D$, $V := D - \{x_0\}$, $V' := \text{aff}(C)$, and $E := V + V'$. Set $\widehat{C} := C + V$. Then $\text{int}_E(\widehat{C}) \neq \phi$ and $\widehat{C} \cap D = \phi$. Define the linear function $f(v + tx_0) := t$ for every $v \in V$, $t \in \mathbb{R}$, and $p(x) := \mu_{\widehat{C}}(x)$. Then

$$t = f(v + tx_0) = tf(v/t + x_0) \leq t\mu_{\widehat{C}}(v/t + x_0) = \mu_{\widehat{C}}(v + tx_0)$$

when $t > 0$, as $v/t + x_0 \in D$. For $t \leq 0$, we also have $f(v + tx_0) \leq \mu_{\widehat{C}}(v+tx_0)$. By Hahn-Banach, we can extend f to $\hat{f}$ such that $\hat{f}(x) = 1$ when $x \in D$ and $\hat{f}(x) \leq \mu_{\widehat{C}}(x) < 1$ when $x \in \widehat{C}$. Now, $H := \{x : \hat{f}(x) = 1\}$ works if $E = \mathbb{R}^n$. Otherwise, let $\mathbb{R}^n = E + \sum_i \mathbb{R}e_i$, and set $g(x + \sum_i \lambda_i e_i) := \hat{f}(x)$. Then $H := \{x : g(x) = 1\}$ works.

A fundamental consequence of separation theorems

Theorem: Every closed convex set $C \subseteq \mathbb{R}^n$ is the intersection of the half-spaces containing C.

Proof: Let C' be that intersection.

- As C' is closed and convex, $C \subseteq C'$.
- Assume that $C' \backslash C \neq \phi$, and let $x \in C' \backslash C$. As $x \in \mathbb{R}^n \backslash C$, which is open, $\exists \epsilon > 0$ such that $B(x, \epsilon) \subseteq \mathbb{R}^n \backslash C$.
  We separate $B(x, \epsilon)$ from C: $\exists f : \mathbb{R}^n \to \mathbb{R}$, linear with

$$t := \inf\{f(z) : z \in C\} > \sup\{f(z) : z \in B(x, \epsilon)\} \geq f(x)$$

  Consider the half-space $H^+ := \{z \in \mathbb{R}^n : f(z) \geq t\}$, so that $C \subseteq H^+$. H^+ is in the collection of half-spaces defining C'. However $x \notin H^+$, a contradiction.

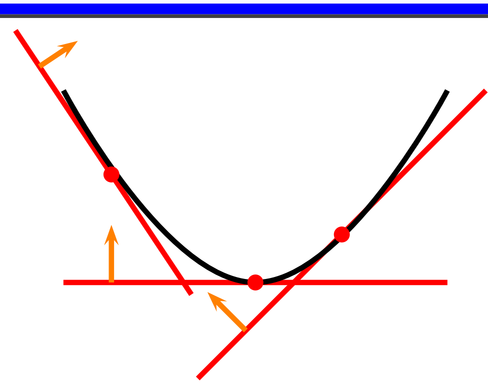
A fundamental consequence of separation theorems

Theorem: Every closed convex set $C \subseteq \mathbb{R}^n$
is the intersection of the half-spaces containing C.

▶ **Reformulation for convex functions:**

A *lower semicontinuous function*, or *LSC function*
is a function with a closed epigraph.

Every convex LSC function
is the supremum of some affine functions



For next week

- (Even) deeper consequences of Hahn-Banach:

Duality theory

- The Fundamental Theorem of Asset Pricing

Addendum: The Minkowski function of a convex set is convex

Proof 1: The light cone of a convex set is convex

Proof 2: Let $x, y \in Q \subseteq \mathbb{R}^n$, a convex set.

μ_Q is positively homogenous, so we just need to show

$$\mu_Q(x+y) \leq \mu_Q(x) + \mu_Q(y).$$

Define $\bar{x} = x/\mu_Q(x)$ and $\bar{y} := y/\mu_Q(y)$,
so that $\mu_Q(\bar{x}) = \mu_Q(\bar{y}) = 1$, thus $\bar{x}, \bar{y} \in \mathrm{cl}(Q)$.

Since Q is convex, $z = \lambda\bar{x} + (1-\lambda)\bar{y} \in \mathrm{cl}(Q)$
for $\lambda := \frac{\mu_Q(x)}{\mu_Q(x)+\mu_Q(y)}$.

Thus $\mu_Q(z) \leq 1$.

But $z = \frac{x+y}{\mu_Q(x)+\mu_Q(y)}$. So $\mu_Q(z) = \frac{\mu_Q(x+y)}{\mu_Q(x)+\mu_Q(y)}$,
and $\mu_Q(x+y) \leq \mu_Q(x) + \mu_Q(y)$.