Chapter 3: Convex Sets

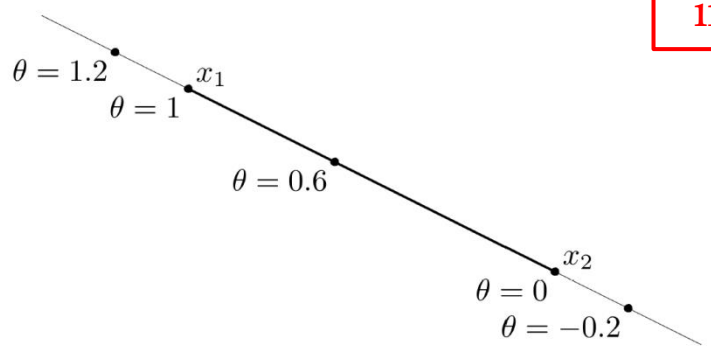
- Affine and convex sets
- Important examples
- Operators that preserve convexity
- Separating and supporting hyperplanes

Convex sets

- Affine sets

Lines and line segments: $\mathbf{y} = \theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2$

also defined for infinite-dim spaces



A set is called an *affine subspace* iff it contains all the lines passing through any two points.

If C is an affine subspace and $\mathbf{x}_0 \in C$, then the set

$$V = C - \mathbf{x}_0 = \{\mathbf{x} - \mathbf{x}_0 | \ \mathbf{x} \in C\}$$

is a linear subspace.

We define $\dim C = \dim V$.

Example: Solution set of linear equations.

Convex sets

- Affine sets

The set of all affine combinations of points in some set C is called the *affine hull* of C, and denoted aff C:

affine combination

$$\text{aff}C = \{\boxed{\theta_1\mathbf{x}_1 + ... + \theta_k\mathbf{x}_k |\ \mathbf{x}_1, ..., \mathbf{x}_k \in C, \theta_1 + ... + \theta_k = 1}\}.$$

The affine hull is the smallest affine set that contains C.

We define the *affine dimension* of a set C as the dimension of its affine hull.

Example: unit circle in $\mathbb{R}^2$

If aff$C$ is not the whole space, we define the *relative interior* of the set C, denoted ri$C$, as its interior relative to aff$C$:

$$\text{ri}C = \{\mathbf{x} \in C |\ B(\mathbf{x}, r) \cap \text{aff}C \subseteq C \text{ for some } r > 0\},$$

We can then define the *relative boundary* of a set C as $\bar{C} \setminus \text{ri}C$.

Example: unit square in $\mathbb{R}^3$

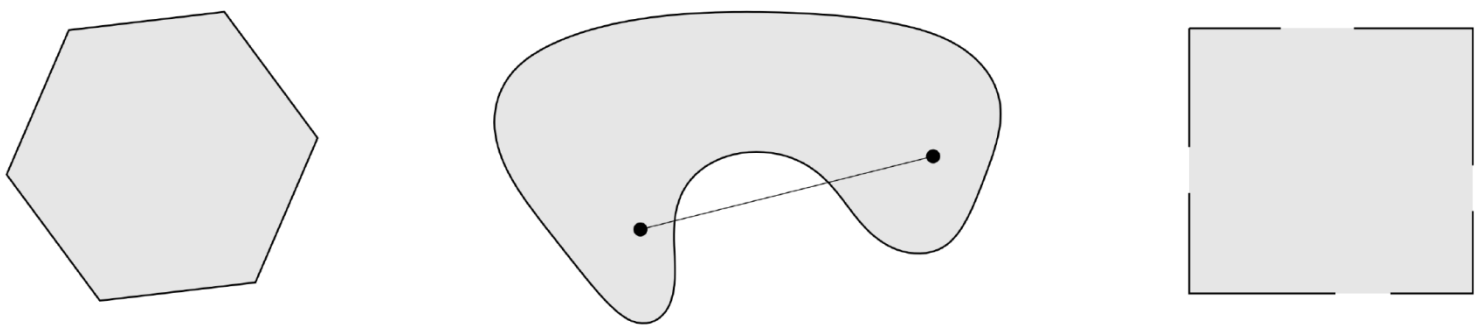
Convex sets

- Convex sets

A set C is *convex* if the line segment between any two points in C lies in C, i.e., if for any $\mathbf{x}_1, \mathbf{x}_2 \in C$ and any θ with $0 \leq \theta \leq 1$, we have

$$\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in C.$$

Every two points can see each other.



Convexity vs. closedness

Examples: $\emptyset$, $\{\mathbf{x}_0\}$, $\mathbb{R}^n$, and affine sets

Convex sets

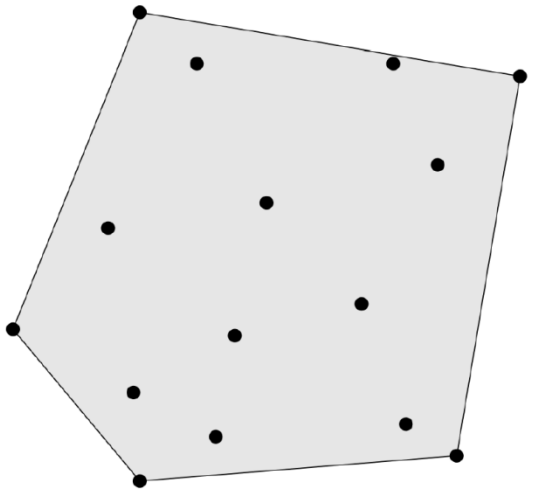
- Convex hull

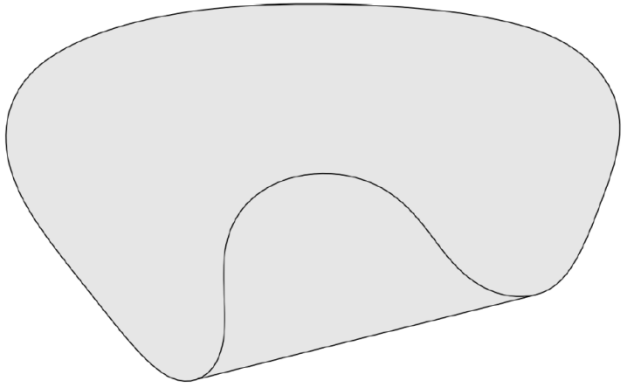
The *convex hull* of a set C, denoted conv$C$, is the set of all convex combinations of points in C:

Convex combination

$$\mathrm{conv}C = \{\boxed{\theta_1\mathbf{x}_1 + ... + \theta_k\mathbf{x}_k | \mathbf{x}_i \in C, \theta_i \geq 0, i = 1, ..., k, \theta_1 + ... + \theta_k = 1}\}.$$

conv$C$ is always convex. It is the smallest convex set that contains C.





Example: $\mathrm{conv}\{\mathbf{e}_i\mathbf{e}_j^T, i = 1, \cdots, m, j = 1, \cdots, n\}$, $\mathrm{conv}\{\mathbf{u}\mathbf{u}^T | \|\mathbf{u}\| = 1\}$, $\mathrm{conv}\{\mathbf{u}\mathbf{v}^T | \|\mathbf{u}\| = 1, \|\mathbf{v}\| = 1\}$

Convex sets

- General convex combination

Suppose $\theta_1, \theta_2, \ldots$ satisfy

$$\theta_i \geq 0, \quad i = 1, 2, \ldots, \quad \sum_{i=1}^{\infty} \theta_i = 1,$$

and $\mathbf{x}_1, \mathbf{x}_2, \ldots \in C$, where $C \subseteq \mathbb{R}^n$ is convex. Then

$$\sum_{i=1}^{\infty} \theta_i \mathbf{x}_i \in C,$$

if the series converges. More generally, suppose $p : \mathbb{R}^n \to \mathbb{R}$ satisfies $p(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in C$ and $\int_C p(\mathbf{x}) d\mathbf{x} = 1$, where $C \subseteq \mathbb{R}^n$ is convex. Then

$$\int_C p(\mathbf{x}) \mathbf{x} d\mathbf{x} \in C, \qquad \boxed{\mathbb{E}\mathbf{x} \in C}$$

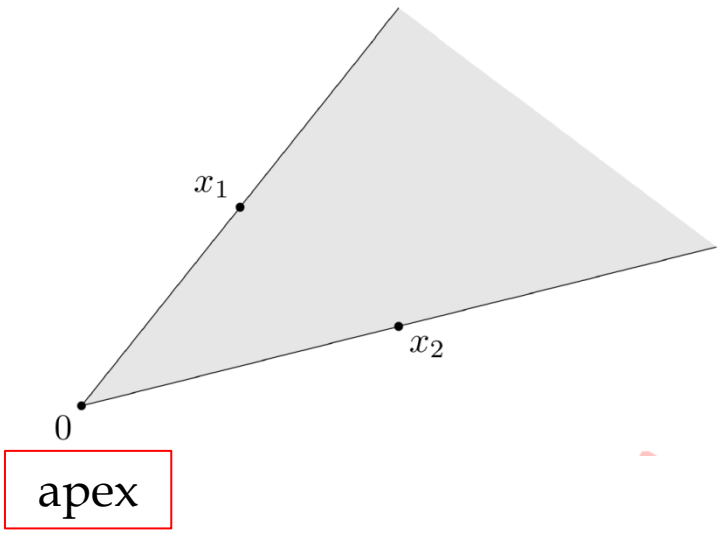
if the integral exists.

Convex sets

- Cones

A set C is called a *cone* if for every $\mathbf{x} \in C$ and $\theta > 0$ we have $\theta\mathbf{x} \in C$.
A set C is a *convex cone* if it is convex and a cone, which means that for any $\mathbf{x}_1, \mathbf{x}_2 \in C$ and $\theta_1, \theta_2 > 0$, we have

$$\theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2 \in C.$$



Convex sets

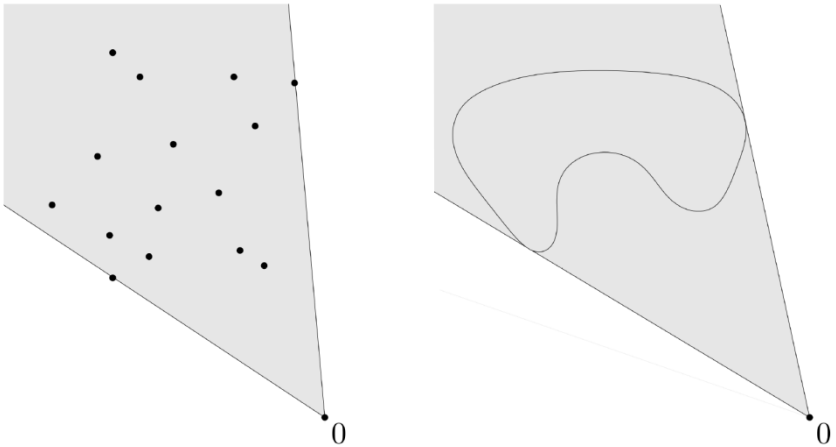
- Conic hull

The *conic hull* of a set C is the set of all conic combinations of points in C, i.e.,

Conic combination

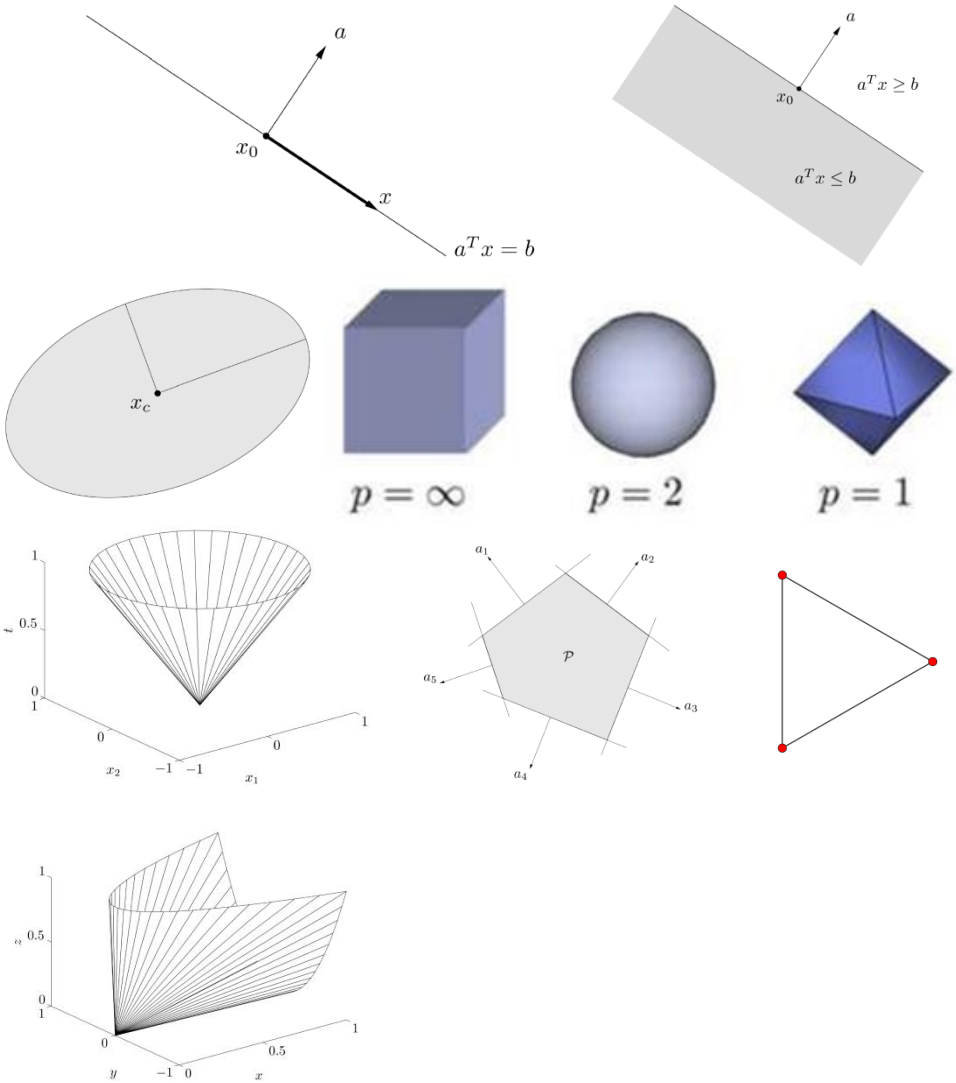
$$\{\boxed{\theta_1 \mathbf{x}_1 + \ldots + \theta_k \mathbf{x}_k \mid \mathbf{x}_i \in C, \theta_i \geq 0, i = 1, \ldots, k}\}.$$

It is the smallest convex cone that contains C.



Important examples

- line, line segment, ray, subspace
- hyperplanes and halfspaces
- Euclidean balls and ellipsoids
- norm balls and norm cones
- polyhedra, nonnegative orthant
- simplexes
- positive semidefinite cone



Operations that preserve convexity

- Intersection

If S_α is convex for every $\alpha \in \mathcal{A}$, then $\cap_{\alpha\in\mathcal{A}} S_\alpha$ is convex.

Examples: 1. $\mathbb{S}^n_+ = \bigcap_{\mathbf{z}\neq\mathbf{0}} \{\mathbf{X} \in \mathbb{S}^n |\ \mathbf{z}^T\mathbf{X}\mathbf{z} \geq 0\}$.

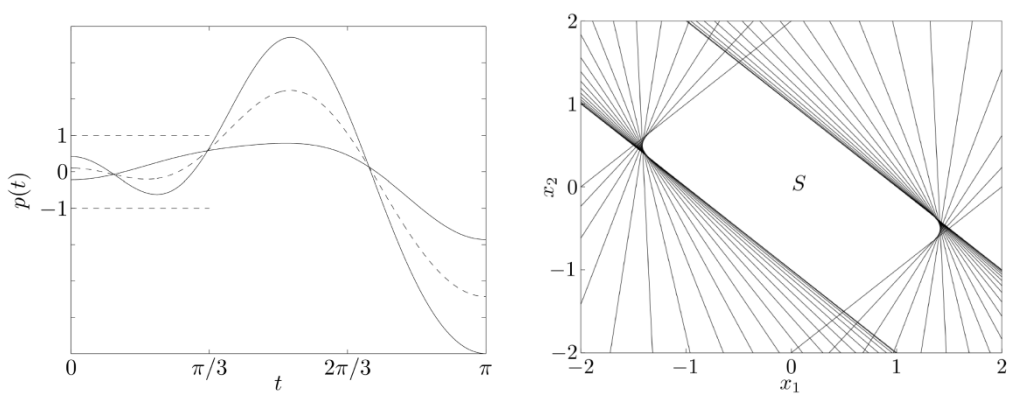
2. Consider the set

$$S = \{\mathbf{x} \in \mathbb{R}^m |\ |\ p(t)| \leq 1 \text{ for } |t| \leq \pi/3\},$$

where $p(t) = \sum_{k=1}^m x_k \cos kt$. The set S can be expressed as the intersection of an infinite number of slabs: $S = \cap_{|t|\leq\pi/3} S_t$, where

$$S_t = \{\mathbf{x} | -1 \leq (\cos t, ..., \cos mt)^T \mathbf{x} \leq 1\},$$

and so is convex.



Operations that preserve convexity

- Intersection

3. $\{\mathbf{X} | \|\mathbf{X}\|_* \leq 1\} = \bigcap_{\mathbf{U}^T\mathbf{U}=\mathbf{I}, \mathbf{V}^T\mathbf{V}=\mathbf{I}} \{\mathbf{X} | \langle \mathbf{U}\mathbf{V}^T, \mathbf{X} \rangle \leq 1\}.$
4. A closed convex set S is the intersection of all halfspaces that contain it:

$$S = \bigcap \{\mathcal{H} |\ \mathcal{H} \text{ halfspace}, S \subseteq \mathcal{H}\}.$$

Operations that preserve convexity

- Affine functions

Suppose $S \subseteq \mathbb{R}^n$ is convex and $f : \mathbb{R}^n \to \mathbb{R}^m$ is an *affine function*: $f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$. Then the *image* of S under f,

$$f(S) = \{f(\mathbf{x}) | \mathbf{x} \in S\},$$

is convex. Similarly, if $f : \mathbb{R}^k \to \mathbb{R}^n$ is an affine function, the *inverse image* of S under f,

$$f^{-1}(S) = \{\mathbf{x} | f(\mathbf{x}) \in S\},$$

is convex.

The *projection* of a convex set onto some of its coordinates is convex: if $S \subseteq \mathbb{R}^m \times \mathbb{R}^n$ is convex, then

$$T = \{\mathbf{x}_1 \in \mathbb{R}^m | (\mathbf{x}_1, \mathbf{x}_2) \in S \text{ for some } \mathbf{x}_2 \in \mathbb{R}^n\}$$

is convex.

Operations that preserve convexity

- Sum and Cartesian product

If S_1 and S_2 are convex, then their *sum* $S_1 + S_2 = \{\mathbf{x} + \mathbf{y} | \mathbf{x} \in S_1,\ \mathbf{y} \in S_2\}$ is convex, so is their *Cartesian product* $S_1 \times S_2 = \{(\mathbf{x}_1, \mathbf{x}_2) |\ \mathbf{x}_1 \in S_1, \mathbf{x}_2 \in S_2\}$.

Operations that preserve convexity

- Examples

1. The polyhedron $\{\mathbf{x}|\mathbf{A}\mathbf{x} \preceq \mathbf{b}, \mathbf{C}\mathbf{x} = \mathbf{d}\}$ can be expressed as the inverse image of the Cartesian product of the nonnegative orthant and the origin under the affine function $f(\mathbf{x}) = (\mathbf{b} - \mathbf{A}\mathbf{x},\ \mathbf{d} - \mathbf{C}\mathbf{x})$:

$$\{\mathbf{x}|\mathbf{A}\mathbf{x} \preceq \mathbf{b}, \mathbf{C}\mathbf{x} = \mathbf{d}\} = \{\mathbf{x}|f(\mathbf{x}) \in \mathbb{R}^m_+ \times \{0\}\}.$$

2. The condition

$$A(\mathbf{x}) = x_1\mathbf{A}_1 + ... + x_n\mathbf{A}_n \preceq \mathbf{B},$$

where $\mathbf{B}, \mathbf{A}_i \in \mathbb{S}^m$, is called a *linear matrix inequality (LMI)* in $\mathbf{x}$.

The solution set of a linear matrix inequality, $\{\mathbf{x}|A(\mathbf{x}) \preceq \mathbf{B}\}$, is convex. Indeed, it is the inverse image of the positive semidefinite cone under the affine function $f : \mathbb{R}^n \to \mathbb{S}^m$ given by $f(\mathbf{x}) = \mathbf{B} - A(\mathbf{x})$.

Operations that preserve convexity

- Examples

3. The set

$$\{\mathbf{x} | \mathbf{x}^T \mathbf{P} \mathbf{x} \leq (\mathbf{c}^T \mathbf{x})^2,\ \mathbf{c}^T \mathbf{x} \geq 0\}$$

where $\mathbf{P} \in \mathbb{S}_+^n$ and $\mathbf{c} \in \mathbb{R}^n$, is convex, since it is the inverse image of the second-order cone,

$$\{(\mathbf{z}, t) |\ \mathbf{z}^T \mathbf{z} \leq t^2, t \geq 0\},$$

under the affine function $f(\mathbf{x}) = (\mathbf{P}^{1/2}\mathbf{x},\ \mathbf{c}^T \mathbf{x})$.

4. The ellipsoid

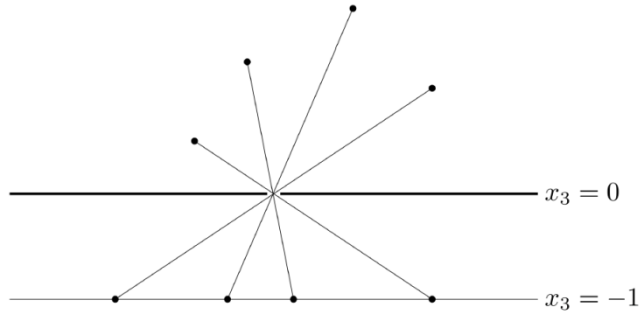
$$\epsilon = \{\mathbf{x} | (\mathbf{x} - \mathbf{x}_c)^T \mathbf{P}^{-1} (\mathbf{x} - \mathbf{x}_c) \leq 1\},$$

where $\mathbf{P} \in \mathbb{S}_{++}^n$, is the image of the unit Euclidean ball $\{\mathbf{u} |\ \|\mathbf{u}\|_2 \leq 1\}$ under the affine mapping $f(\mathbf{u}) = \mathbf{P}^{1/2}\mathbf{u} + \mathbf{x}_c$. (It is also the inverse image of the unit ball under the affine mapping $g(\mathbf{x}) = \mathbf{P}^{-1/2}(\mathbf{x} - \mathbf{x}_c)$.)

Operations that preserve convexity

- Perspective functions

We define the *perspective function* $P : \mathbb{R}^{n+1} \to \mathbb{R}^n$, with domain $\mathbf{dom}\ \ P = \mathbb{R}^n \times \mathbb{R}_{++}$, as $P(\mathbf{z}, t) = \mathbf{z}/t$.



The inverse image of a convex set under the perspective function is also convex: if $C \subseteq \mathbb{R}^n$ is convex, then

$$P^{-1}(C) = \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} | \mathbf{x}/t \in C, t > 0\}$$

is convex.

Question: If function f preserves convexity: if C_1 is convex then $f(C_1)$ is also convex, does f^{-1} also perserve convexity?

Operations that preserve convexity

- Linear-fractional functions

A linear-fractional function is formed by composing the perspective function with an affine function. Suppose $g : \mathbb{R}^n \to \mathbb{R}^{m+1}$ is affine, i.e.,

$$g(\mathbf{x}) = \begin{bmatrix} \mathbf{A} \\ \mathbf{c}^T \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{b} \\ d \end{bmatrix},$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{c} \in \mathbb{R}^n$, and $d \in \mathbb{R}$. The function $f : \mathbb{R}^n \to \mathbb{R}^m$ given by $f = P \circ g$, i.e.,

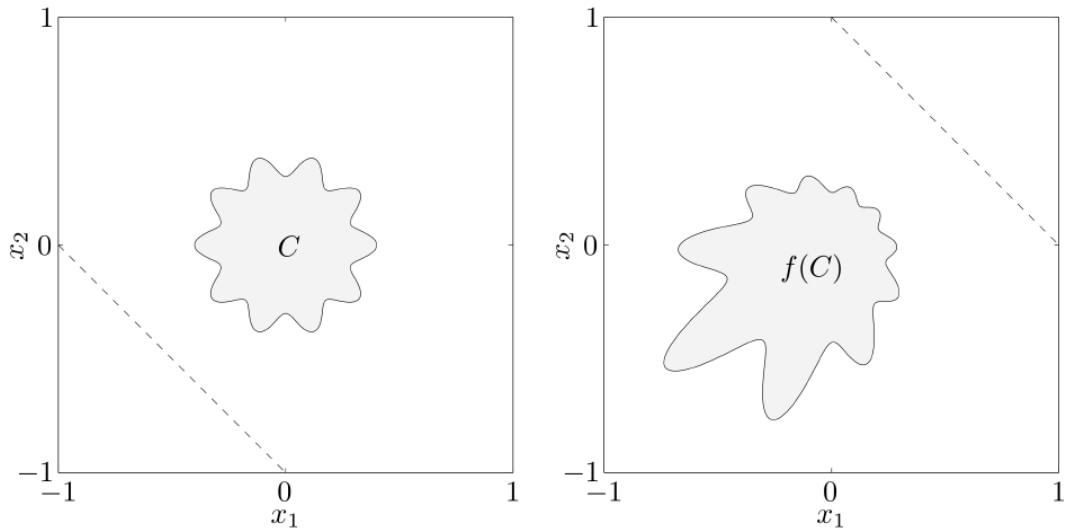
$$f(\mathbf{x}) = (\mathbf{A}\mathbf{x} + \mathbf{b})/(\mathbf{c}^T\mathbf{x} + d), \operatorname{dom} f = \{\mathbf{x} | \mathbf{c}^T\mathbf{x} + d > 0\},$$

is called a *linear-fractional* (or projective) function.

Operations that preserve convexity

- Linear-fractional functions

$$f(\mathbf{x}) = \frac{1}{\mathbf{x}_1 + \mathbf{x}_2 + 1}\mathbf{x}, \ \mathbf{dom}\ f = \{(x_1, x_2) | x_1 + x_2 + 1 > 0\}.$$



Conditional probabilities: Let $p_{ij} = \mathbb{P}(u = i, v = j)$. Then the conditional probability

$$f_{ij} = \frac{p_{ij}}{\sum_{k=1}^{n} p_{kj}}$$

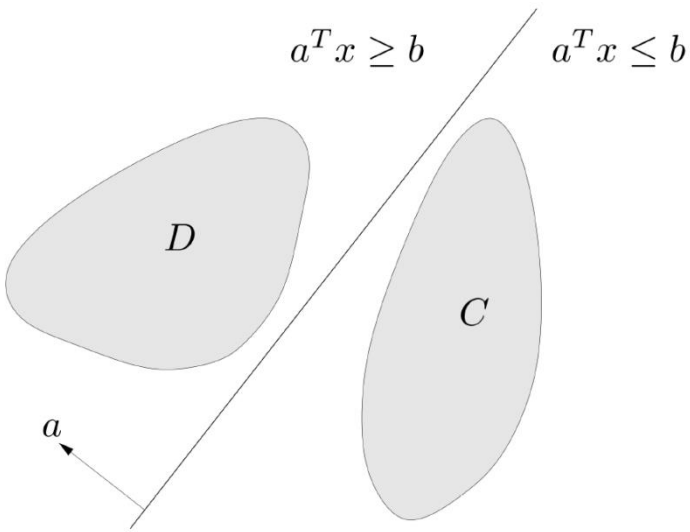
is obtained by a linear-fractional mapping from $\mathbf{p}$.

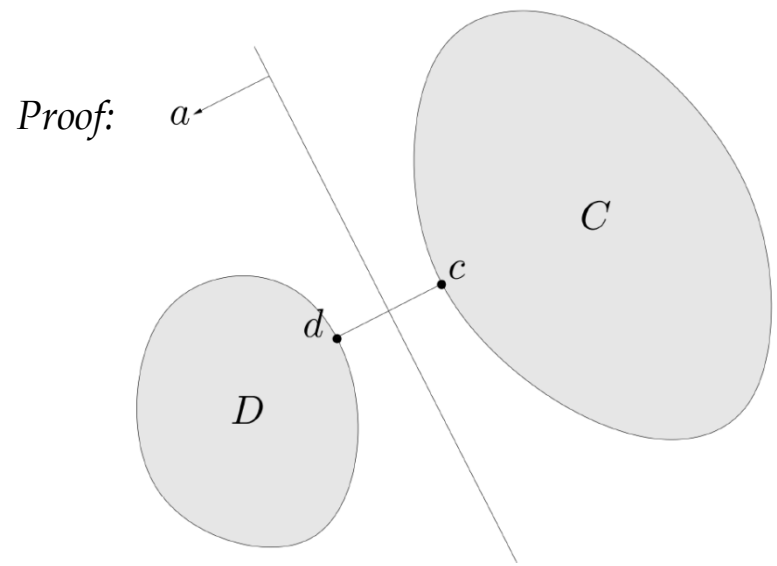
Separating and supporting hyperplanes

- Separating hyperplane theorem

Theorem 1. *Suppose C and D are two convex sets that do not intersect, i.e., $C \cap D = \emptyset$. Then there exist $\mathbf{a} \neq \mathbf{0}$ and b such that $\mathbf{a}^T\mathbf{x} \leq b$ for all $\mathbf{x} \in C$ and $\mathbf{a}^T\mathbf{x} \geq b$ for all $\mathbf{x} \in D$. In other words, the affine function $\mathbf{a}^T\mathbf{x} - b$ is nonpositive on C and nonnegative on D.*

The hyperplane $\{\mathbf{x} | \mathbf{a}^T\mathbf{x} = b\}$ is called a *separating hyperplane* for the sets C and D, or is said to separate the sets C and D.





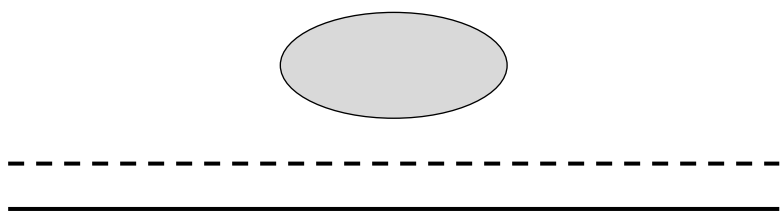
Separating and supporting hyperplanes

- Separating hyperplane theorem: example

Suppose C is convex and D is affine, i.e., $D = \{\mathbf{F}\mathbf{u} + \mathbf{g} |\ \mathbf{u} \in \mathbb{R}^m\}$, where $\mathbf{F} \in \mathbb{R}^{n \times m}$. Suppose C and D are disjoint, so by the separating hyperplane theorem there are $\mathbf{a} \neq 0$ and b such that $\mathbf{a}^T\mathbf{x} \leq b$ for all $\mathbf{x} \in C$ and $\mathbf{a}^T\mathbf{x} \geq b$ for all $\mathbf{x} \in D$.

Now $\mathbf{a}^T\mathbf{x} \geq b$ for all $\mathbf{x} \in D$ means $\mathbf{a}^T\mathbf{F}\mathbf{u} \geq b - \mathbf{a}^T\mathbf{g}$ for all $\mathbf{u} \in \mathbb{R}^m$. But a linear function is bounded below on $\mathbb{R}^m$ only when it is zero, so we conclude $\mathbf{a}^T\mathbf{F} = \mathbf{0}$ (and hence, $b \leq \mathbf{a}^T\mathbf{g}$).

Thus we conclude that there exists $\mathbf{a} \neq 0$ such that $\mathbf{F}^T\mathbf{a} = \mathbf{0}$ and $\mathbf{a}^T\mathbf{x} \leq \mathbf{a}^T\mathbf{g}$ for all $\mathbf{x} \in C$.

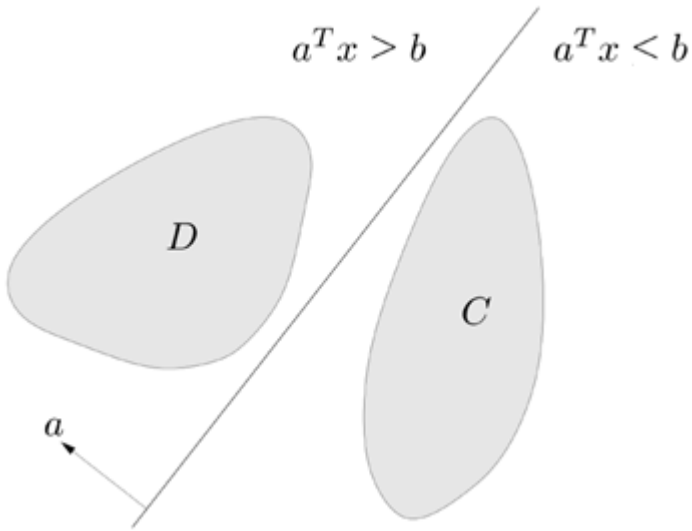


When is the separation hyperplane unique?

Separating and supporting hyperplanes

- Strict separation

If the separating hyperplane satisfies the stronger condition that $\mathbf{a}^T\mathbf{x} < b$ for all $\mathbf{x} \in C$ and $\mathbf{a}^T\mathbf{x} > b$ for all $\mathbf{x} \in D$, then the sets C and D are called *strictly separated*.



Disjoint convex sets need not be strictly separable by a hyperplane (even when the sets are closed)

Example: a point and a closed convex set; a closed convex set is the intersection of all halfspaces that contain it

Separating and supporting hyperplanes

- Converse separating hyperplane theorems

Theorem 1. *Any two convex sets C and D, at least one of which is open, are disjoint if and only if there exists a separating hyperplane.*

Example: (Theorem of alternatives for strict linear inequalities) We derive the necessary and sufficient conditions for solvability of a system of strict linear inequalities $\boxed{\mathbf{A}\mathbf{x} < \mathbf{b}.}$

pair of alternatives

These inequalities are infeasible if and only if the (convex) sets

$$C = \{\mathbf{b} - \mathbf{A}\mathbf{x} | \mathbf{x} \in \mathbb{R}^n\}, \quad D = \mathbb{R}^m_{++} = \{\mathbf{y} \in \mathbb{R}^m | \mathbf{y} \succ \mathbf{0}\}$$

do not intersect. The set D is open, C is an affine set. Hence by the above theorem, C and D are disjoint iff there exists a separating hyperplane, i.e., a nonzero $\boldsymbol{\lambda} \in \mathbb{R}^m$ and $\mu \in \mathbb{R}$ such that $\boldsymbol{\lambda}^T\mathbf{y} \leq \mu$ on C and $\boldsymbol{\lambda}^T\mathbf{y} \geq \mu$ on D.

$\mu \leq 0$ and $\boldsymbol{\lambda} \geq \mathbf{0}$, $\boldsymbol{\lambda} \neq \mathbf{0}$. $\Rightarrow$ $\boxed{\exists \boldsymbol{\lambda} \text{ s.t. } \boldsymbol{\lambda} \neq \mathbf{0},\ \boldsymbol{\lambda} \geq \mathbf{0},\ \mathbf{A}^T\boldsymbol{\lambda} = \mathbf{0},\ \boldsymbol{\lambda}^T\mathbf{b} \leq 0.}$

Separating and supporting hyperplanes

- Converse separating hyperplane theorems

Theorem 1 (Theorem of the Alternative (Fakas' Lemma)). *For* $\mathbf{A} \in \mathbb{R}^{m \times n}$ *and* $\mathbf{b} \in \mathbb{R}^m$ *the following are strong alternatives:*

1. $\exists \mathbf{x} \in \mathbb{R}^n_+$ *such that* $\mathbf{A}\mathbf{x} = \mathbf{b}$,
2. $\exists \mathbf{y} \in \mathbb{R}^m$ *such that* $\mathbf{A}^T\mathbf{y} \geq \mathbf{0}$ *and* $\mathbf{b}^T\mathbf{y} < 0$.

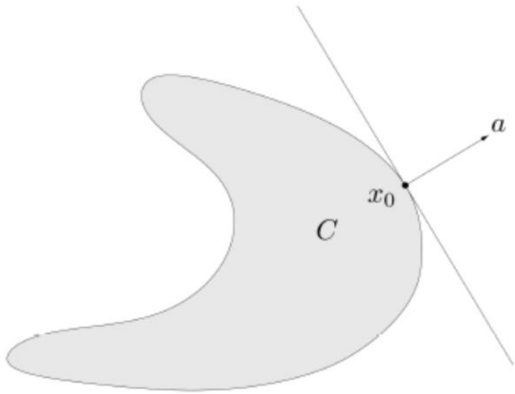
Proof. 1) $\Longrightarrow \neg 2)$: For $\mathbf{x} \in \mathbb{R}^n_+$ with $\mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{y} \in \mathbb{R}^m$ with $\mathbf{A}^T\mathbf{y} \geq 0$ we have $\mathbf{b}^T\mathbf{y} = \mathbf{x}^T\mathbf{A}^T\mathbf{y} \geq 0$.

$\neg 1) \Longrightarrow 2)$: $C := cone(\mathbf{A})$ is a closed convex cone which does not contain the vector $\mathbf{b}$: by the Separating Hyperplane Theorem there exists a $\mathbf{y} \in \mathbb{R}^m$ with $\langle \mathbf{y}, \mathbf{x} \rangle \geq 0 > \langle \mathbf{y}, \mathbf{b} \rangle$ for all $\mathbf{x} \in C$, in particular $\mathbf{A}_i^T\mathbf{y} = \langle \mathbf{y}, \mathbf{A}_i \rangle \geq 0$, $\forall i$, that is, $\mathbf{A}^T\mathbf{y} \geq \mathbf{0}$.

Separating and supporting hyperplanes

- Supporting hyperplanes

Suppose $C \subseteq \mathbb{R}^n$, and $\mathbf{x}_0$ is a point in its boundary ∂C. If $\mathbf{a} \neq \mathbf{0}$ satisfies $\mathbf{a}^T\mathbf{x} \leq \mathbf{a}^T\mathbf{x}_0$ for all $\mathbf{x} \in C$, then the hyperplane $\{\mathbf{x} |\ \mathbf{a}^T\mathbf{x} = \mathbf{a}^T\mathbf{x}_0\}$ is called a *supporting hyperplane* to C at the point $\mathbf{x}_0$.



Theorem 1 (Supporting Hyperplane Theorem)**.** *For any nonempty convex set C, and any $\mathbf{x}_0 \in \partial C$, there exists a supporting hyperplane to C at $\mathbf{x}_0$.*

Proof: Two cases: $C^\circ \neq \emptyset$ and $C^\circ = \emptyset$.

Chapter 4: Convex Functions

- Basic properties and examples
- Operations that preserve convexity
- The conjugate function
- Envelope function and Proximal mapping

Basic properties and examples

A function $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* if $\boxed{\mathrm{dom}\, f \text{ is a convex set}}$ and if for all $\mathbf{x}$, $\mathbf{y} \in \mathrm{dom}\, f$, and θ with $0 \le \theta \le 1$, we have

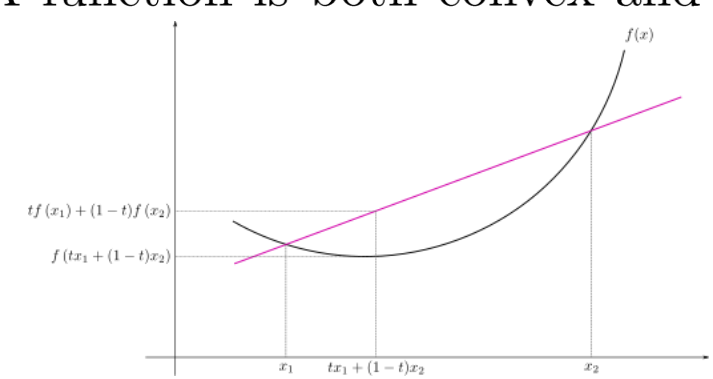
$$f\left(\boxed{\theta\mathbf{x} + (1-\theta)\mathbf{y}}\right) \le \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}). \tag{1}$$

A function f is *strictly convex* if strict inequality holds in (1) whenever $\mathbf{x} \neq \mathbf{y}$ and $0 < \theta < 1$.

A function f is *strongly convex* if

$$f\left(\theta\mathbf{x} + (1-\theta)\mathbf{y}\right) \le \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}) - \frac{\theta(1-\theta)\overset{\boxed{\textit{modulus}}}{\boxed{\mu}}}{2}\|\mathbf{y}-\mathbf{x}\|^2, \quad \forall \theta \in [0,1]. \tag{2}$$

f is *concave*, *strictly concave*, *strongly concave* if $-f$ is convex, strictly convex, strongly convex. A function is both convex and concave iff it is an affine function.



Basic properties and examples

A convex function is continuous on the relative interior of its domain; it can have discontinuities only on its relative boundary.

Theorem 1 (Rademacher's Theorem)**.** *A convex function is differentiable almost everywhere on the relative interior of its domain.*

- Extended-value extensions

If f is convex we define its *extended-value extension* $\tilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ by

$$\tilde{f}(\mathbf{x}) = \begin{cases} f(\mathbf{x}), & \mathbf{x} \in \operatorname{dom} f \\ \infty, & \mathbf{x} \notin \operatorname{dom} f. \end{cases}$$

We will use the same symbol to denote a convex function and its extension.

Example: Indicator function of a convex set

$$\begin{aligned} &\min_{\mathbf{x}} \ f(\mathbf{x}), \\ &s.t. \ \mathbf{x} \in \mathcal{C}. \end{aligned} \quad \Longleftrightarrow \quad \min_{\mathbf{x}} \ f(\mathbf{x}) + \tilde{I}_{\mathcal{C}}(\mathbf{x}).$$

Basic properties and examples

- Proper function

f is called *proper* if $f(\mathbf{x}) < \infty$ for at least one $\mathbf{x} \in \mathcal{X}$ and $f(\mathbf{x}) > -\infty$ for all $\mathbf{x} \in \mathcal{X}$, and we say that f is *improper* if it is not proper. In words, a function is proper iff its epigraph is nonempty and does not contain a vertical line.

Basic properties and examples

- First-order conditions

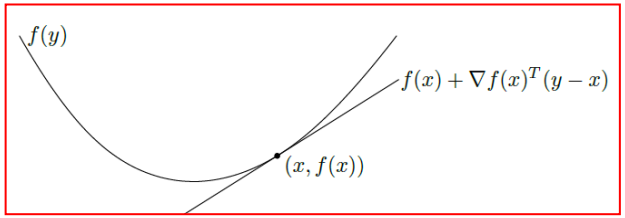
Suppose f is differentiable. Then f is convex iff $\operatorname{dom} f$ is convex and

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \tag{1}$$

holds for all $\mathbf{x}$, $\mathbf{y} \in \operatorname{dom} f$.

Proof. If f is convex, then $f((1-\alpha)\mathbf{x} + \alpha\mathbf{y}) \leq (1-\alpha)f(\mathbf{x}) + \alpha f(\mathbf{y})$, which can be rewritten as

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \frac{f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\alpha},$$



Letting $\alpha \to 0^+$, we have (1). If (1) holds, we have

$$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq f(\mathbf{x}) - (1-\alpha)\langle \nabla f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle,$$
$$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq f(\mathbf{y}) + \alpha\langle \nabla f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle.$$

Multiplying the first inequality with α and the second with $(1-\alpha)$ and adding them together, we obtain $f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y})$.

Basic properties and examples

If $\nabla f(\mathbf{x}) = \mathbf{0}$, then for all $\mathbf{y} \in \operatorname{dom} f$, $f(\mathbf{y}) \geq f(\mathbf{x})$, *i.e.*, $\mathbf{x}$ is a global minimizer of f.

Strictly convex:

$$f(\mathbf{y}) > f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle, \quad \text{if } \mathbf{y} \neq \mathbf{x}. \tag{1}$$

Proof. $f(\mathbf{y}) > f(\mathbf{x}) + \frac{f(\mathbf{x}+\alpha(\mathbf{y}-\mathbf{x}))-f(\mathbf{x})}{\alpha}, \quad \forall \alpha \in (0,1).$

For all $\alpha \in (0,1)$ by the convexity we have $f(\mathbf{x}+\alpha(\mathbf{y}-\mathbf{x}))-f(\mathbf{x}) \geq \alpha\langle \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle$. Thus $\frac{f(\mathbf{x}+\alpha(\mathbf{y}-\mathbf{x}))-f(\mathbf{x})}{\alpha} \geq \langle \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle$, $\forall \alpha \in (0,1)$. So (1) holds.

Strongly convex: $f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{\mu}{2}\|\mathbf{y} - \mathbf{x}\|^2$.

Proof. Follow the proof of convexity.

Basic properties and examples

- Second-order conditions

Assume that f is twice differentiable. Then f is convex iff $\operatorname{dom} f$ is convex and its Hessian is positive semidefinite: for all $\mathbf{x} \in \operatorname{dom} f$,

$$\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}.$$

For a function on $\mathbb{R}$, this reduces to the simple condition $f''(x) \geq 0$ (and $\operatorname{dom} f$ convex, *i.e.*, an interval), which means that the derivative is nondecreasing. The condition $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ can be interpreted geometrically as the requirement that the graph of the function have positive (upward) curvature at $\mathbf{x}$.

Strictly convex: $\nabla^2 f(\mathbf{x}) \succ \mathbf{0}$.
Strongly convex: $\nabla^2 f(\mathbf{x}) \succeq \mu \mathbf{I}$.

Basic properties and examples

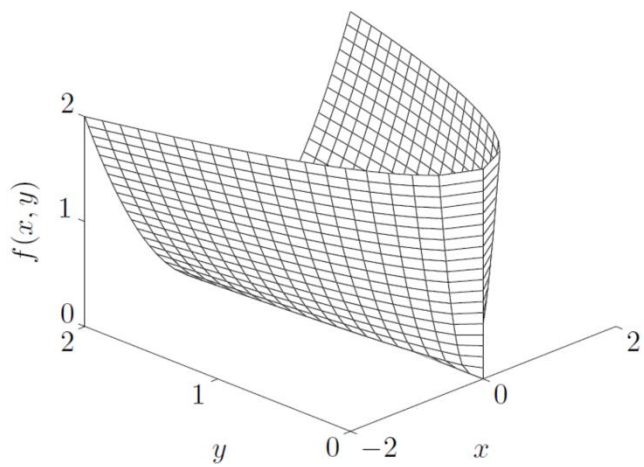
- Examples

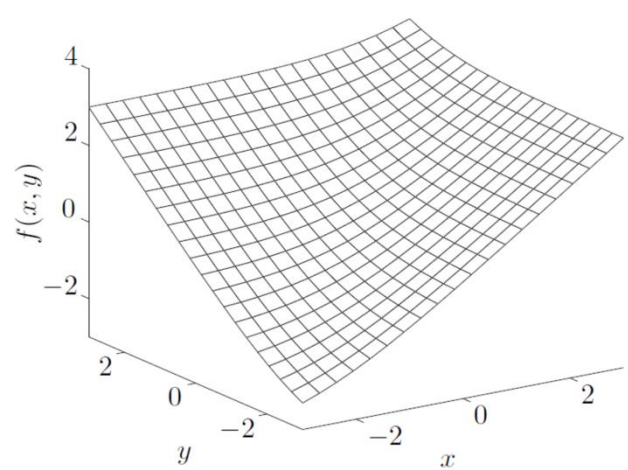
- *Exponential.* e^{ax} is convex on $\mathbb{R}$, for any $a \in \mathbb{R}$.
- *Powers.* x^a is convex on $\mathbb{R}_{++}$ when $a \geq 1$ or $a \leq 0$, and concave for $0 \leq a \leq 1$.
- *Powers of absolute value.* $|x|^p$, for $p \geq 1$, is convex on $\mathbb{R}$.
- *Logarithm.* $\log x$ is concave on $\mathbb{R}_{++}$.
- *Negative entropy.* $x \log x$ (either on $\mathbb{R}_{++}$, or on $\mathbb{R}_{+}$, defined as 0 for $x = 0$) is convex.

Basic properties and examples

- Examples

- *Norms.* Every norm on $\mathbb{R}^n$ is convex.
- *Max function.* $f(\mathbf{x}) = \max\{x_1, \ldots, x_n\}$ is convex on $\mathbb{R}^n$.
- $f(x, y) = x^2/y$ with $\operatorname{dom} f = \mathbb{R} \times \mathbb{R}_{++} = \{(x, y) \in \mathbb{R}^2 \mid y > 0\}$ is convex.
- *Log-sum-exp.* $f(\mathbf{x}) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex on $\mathbb{R}^n$.
- *Geometric mean.* $f(\mathbf{x}) = (\prod_{i=1}^n x_i)^{1/n}$ is concave on $\operatorname{dom} f = \mathbb{R}^n_{++}$.





Basic properties and examples

- Examples
- *Log-determinant.* $f(\mathbf{X}) = \log\det \mathbf{X}$ is concave on $\operatorname{dom} f = \mathbb{S}^n_{++}$.

The Hessian of f at $\mathbf{X}$ is a fourth-order tensor $\mathcal{T}$. We have shown that $\mathcal{T}(\Delta\mathbf{X}) = -\mathbf{X}^{-1}\Delta\mathbf{X}\mathbf{X}^{-1}$.

$$\langle \mathcal{T}(\Delta\mathbf{X}), \Delta\mathbf{X}\rangle = -\operatorname{tr}\left[(\mathbf{X}^{-1}\Delta\mathbf{X}\mathbf{X}^{-1})\Delta\mathbf{X}\right] = -\operatorname{tr}\left[\mathbf{X}^{-1}(\Delta\mathbf{X}\mathbf{X}^{-1}\Delta\mathbf{X})\right] \leq 0.$$

Basic properties and examples

- Sublevels

The α-*sublevel set* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

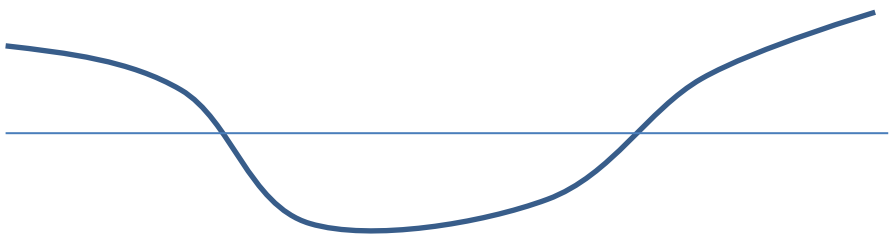
$$C_\alpha = \{\mathbf{x} \in \operatorname{dom} f \mid f(\mathbf{x}) \leq \alpha\}.$$

Sublevel sets of a convex function are convex, for any value of α.

The converse is not true: a function can have all its sublevel sets convex, but not be a convex function. Such functions are called *quasi-convex functions*.

Quasi-convex:

$$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq \max\{f(\mathbf{x}), f(\mathbf{y})\}, \quad \alpha \in [0,1].$$



Basic properties and examples

- Sublevels

Example: The geometric and arithmetic means of $\mathbf{x} \in \mathbb{R}^n_+$ are, respectively,

$$G(\mathbf{x}) = \left(\prod_{i=1}^{n} x_i\right)^{1/n}, \quad A(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

The arithmetic-geometric mean inequality states that $G(\mathbf{x}) \leq A(\mathbf{x})$.

Suppose $0 \leq \alpha \leq 1$, and consider the set

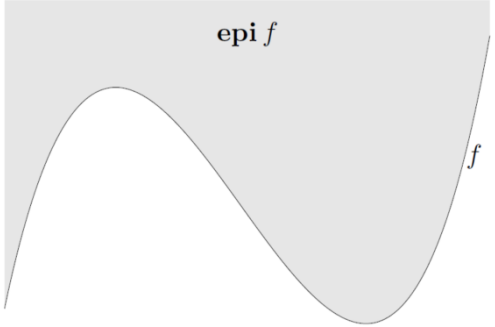
$$\{\mathbf{x} \in \mathbb{R}^n_+ \mid G(\mathbf{x}) \geq \alpha A(\mathbf{x})\},$$

i.e., the set of vectors with geometric mean at least as large as a factor α times the arithmetic mean. This set is convex, since it is the 0-sublevel set of the function $\alpha A(\mathbf{x}) - G(\mathbf{x})$, which is convex. In fact, the set is positively homogeneous, so it is a convex cone.

Basic properties and examples

- Epigraph

The *epigraph* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as



$$\operatorname{epi} f = \{(\mathbf{x}, t) \mid \mathbf{x} \in \operatorname{dom} f, f(\mathbf{x}) \leq t\},$$

which is a subset of $\mathbb{R}^{n+1}$.

A function is convex iff its epigraph is a convex set.

Example: $f(\mathbf{x}, \mathbf{Y}) = \mathbf{x}^T \mathbf{Y}^{-1} \mathbf{x}$ is convex on $\operatorname{dom} f = \mathbb{R}^n \times \mathbb{S}^n_{++}$.

By its epigraph:

$$\begin{aligned}\operatorname{epi} f &= \{(\mathbf{x}, \mathbf{Y}, t) \mid \mathbf{Y} \succ \mathbf{0},\ \mathbf{x}^T \mathbf{Y}^{-1} \mathbf{x} \leq t\} \\ &= \left\{ (\mathbf{x}, \mathbf{Y}, t) \;\middle|\; \begin{bmatrix} \mathbf{Y} & \mathbf{x} \\ \mathbf{x}^T & t \end{bmatrix} \succeq \mathbf{0},\ \mathbf{Y} \succ \mathbf{0} \right\}.\end{aligned}$$

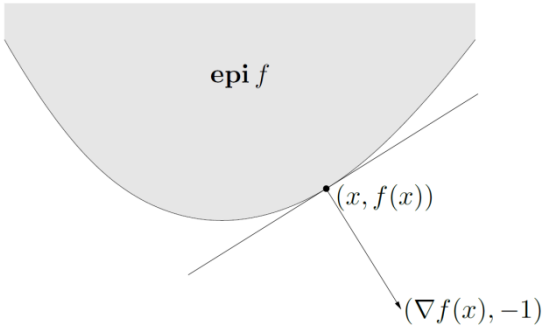
The last condition is a linear matrix inequality in $(\mathbf{x}, \mathbf{Y}, t)$, and therefore $\operatorname{epi} f$ is convex.

Basic properties and examples

- Epigraph

Many results for convex functions can be proved (or interpreted) geometrically using epigraphs, and applying results for convex sets. As an example, consider the first-order condition for convexity:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle .$$



If $(\mathbf{y}, t) \in \operatorname{epi} f$, then

$$t \geq f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle .$$

We can express this as:

$$(\mathbf{y}, t) \in \operatorname{epi} f \Longrightarrow \begin{bmatrix} \nabla f(\mathbf{x}) \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} \mathbf{y} \\ t \end{bmatrix} - \begin{bmatrix} \mathbf{x} \\ f(\mathbf{x}) \end{bmatrix} \right) \leq 0.$$

This means that the hyperplane defined by $(\nabla f(\mathbf{x}), -1)$ supports $\operatorname{epi} f$ at the boundary point $(\mathbf{x}, f(\mathbf{x}))$.

Basic properties and examples

- Jensen's inequality and extensions

$$f\left(\theta\mathbf{x}+(1-\theta)\mathbf{y}\right) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}).$$

$$\Updownarrow$$

$$f(\theta_1\mathbf{x}_1 + \cdots + \theta_k\mathbf{x}_k) \leq \theta_1 f(\mathbf{x}_1) + \cdots + \theta_k f(\mathbf{x}_k).$$

$$f\left(\int_S p(\mathbf{x})\mathbf{x}\,d\mathbf{x}\right) \leq \int_S f(\mathbf{x})p(\mathbf{x})\,d\mathbf{x}.$$

$$\Updownarrow$$

$$f(\mathbb{E}\,\mathbf{x}) \leq \mathbb{E}\,f(\mathbf{x}).$$

Basic properties and examples

- Inequality

Arithmetic-geometric mean inequality:

$$\sqrt{ab} \leq (a+b)/2.$$

Hölder's inequality: for $p, q > 1$, $1/p + 1/q = 1$, and $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n$,

$$\sum_{i=1}^{n} x_i y_i \leq \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \left(\sum_{i=1}^{n} |y_i|^q \right)^{1/q}.$$

Basic properties and examples

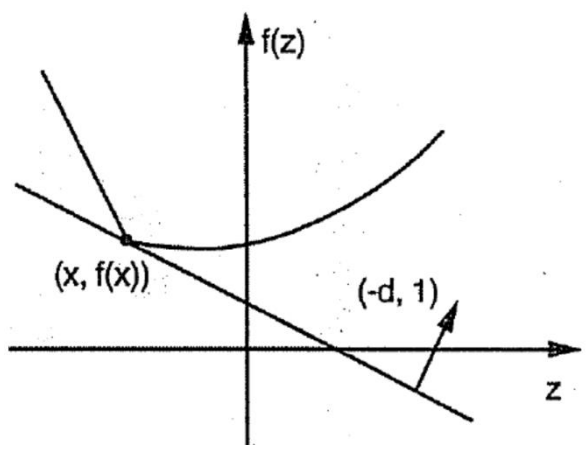
- Subgradient

$$\partial f(\mathbf{x}) = \{\mathbf{g} | f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle, \forall \mathbf{x}, \mathbf{y} \in \operatorname{dom} f\}.$$

Subgradient can be identified with a non-vertical supporting hyperplane to the epigraph of f at $(\mathbf{x}, f(\mathbf{x}))$.

Proposition 1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a proper convex function. The subgradient $\partial f(\mathbf{x})$ is nonempty, convex, and compact for all $\mathbf{x} \in (\operatorname{dom} f)^\circ$.*

$\partial f(\mathbf{x})$ may be empty when $\mathbf{x} \in \partial(\operatorname{dom} f)$. Example?



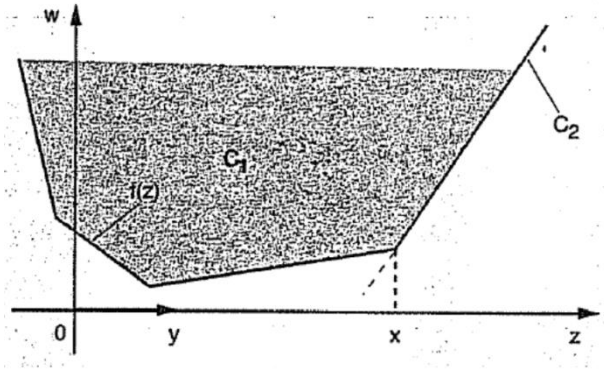
Basic properties and examples

- Subgradient

Proposition 1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. For every $\mathbf{x} \in \mathbb{R}^n$, we have*

$$f'(\mathbf{x}; \mathbf{y}) = \max_{\mathbf{g} \in \partial f(\mathbf{x})} \langle \mathbf{y}, \mathbf{g} \rangle, \quad \forall \mathbf{y} \in \mathbb{R}^n. \tag{1}$$

In particular, f is differentiable at $\mathbf{x}$ with gradient $\nabla f(\mathbf{x})$ iff it has $\nabla f(\mathbf{x})$ as its unique subgradient at $\mathbf{x}$.



Proof: Apply Separating Hyperplane Theorem to

$$C_1 = \{(\mathbf{z}, w) | f(\mathbf{z}) < w\},$$

and

$$C_2 = \{(\mathbf{z}, w) | \mathbf{z} = \mathbf{x} + \alpha \mathbf{y}, w = f(\mathbf{x}) + \alpha f'(\mathbf{x}; \mathbf{y}), \alpha \geq 0\}.$$

Basic properties and examples

- Subgradient

Example: $|x|$, $\max\{0, \frac{1}{2}(x^2 - 1)\}$, $I_{\mathcal{C}}(\mathbf{x})$.

Basic properties and examples

- Subgradient

Proposition 1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function.*

(a) If $\mathcal{X}$ is a bounded set, then the set $\cup_{\mathbf{x}\in\mathcal{X}} \partial f(\mathbf{x})$ is bounded.

(b) If a sequence $\{\mathbf{x}_k\}$ converges to a vector $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{g}_k \in \partial f(\mathbf{x}_k)$ for all k, then the sequence $\{\mathbf{g}_k\}$ is bounded and each of its accumulation points is a subgradient of f at $\mathbf{x}$.

Proposition 2. *Let $f_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \cdots, m$, be convex functions and let $f = f_1 + \cdots + f_m$. Then*

$$\partial f(\mathbf{x}) = \partial f_1(\mathbf{x}) + \cdots + \partial f_m(\mathbf{x}).$$

Basic properties and examples

- Subgradient

Proposition 1 (Chain Rule). *(a) Let $f : \mathbb{R}^m \to \mathbb{R}$ be a convex function, and let $\mathbf{A}$ be an $m \times n$ matrix. Then the subgradient of the function F, defined by $F(\mathbf{x}) = f(\mathbf{A}\mathbf{x})$, is given by*

$$\partial F(\mathbf{x}) = \mathbf{A}^T \partial f(\mathbf{A}\mathbf{x}).$$

(b) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and let $h : \mathbb{R} \to \mathbb{R}$ be a differentiable scalar function. Then the function F, defined by $F(\mathbf{x}) = h(f(\mathbf{x}))$, is directionally differentiable at all $\mathbf{x}$, given by

$$F'(\mathbf{x}; \mathbf{y}) = h'(f(\mathbf{x}))f'(\mathbf{x}; \mathbf{y}), \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

Furthermore, if h is convex and monotonically nondecreasing, then F is convex and its subgradient is given by

$$\partial F(\mathbf{x}) = \partial h(f(\mathbf{x}))\partial f(\mathbf{x}) = \{g\mathbf{g} | g \in \partial h(f(\mathbf{x})), \mathbf{g} \in \partial f(\mathbf{x})\}, \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

Basic properties and examples

- Subgradient

Theorem 1 (Subgradient of norms)**.** *Let $\mathcal{H}$ be a real Hilbert space endowed with an inner product $\langle \cdot, \cdot \rangle$ and a norm $\|\cdot\|$. Then $\partial\|\mathbf{x}\| = \{\mathbf{y} | \langle \mathbf{y}, \mathbf{x} \rangle = \|\mathbf{x}\|$ and $\|\mathbf{y}\|^* \leq 1\}$, where $\|\cdot\|^*$ is the dual norm of $\|\cdot\|$.*

Proof. Let $S = \{\mathbf{y} | \langle \mathbf{y}, \mathbf{x} \rangle = \|\mathbf{x}\|$ and $\|\mathbf{y}\|^* \leq 1\}$.
For every $\mathbf{y} \in \partial\|\mathbf{x}\|$, we have

$$\|\mathbf{w} - \mathbf{x}\| \geq \|\mathbf{w}\| - \|\mathbf{x}\| \geq \langle \mathbf{y}, \mathbf{w} - \mathbf{x} \rangle, \quad \forall\ \mathbf{w} \in \mathcal{H}. \tag{1}$$

Choosing $\mathbf{w} = 0$ and $\mathbf{w} = 2\mathbf{x}$ for the second inequality above, which results from the convexity of norm $\|\cdot\|$, we can deduce that

$$\|\mathbf{x}\| = \langle \mathbf{y}, \mathbf{x} \rangle. \tag{2}$$

Basic properties and examples

- Subgradient

On the other hand, (1) gives

$$\|\mathbf{w}-\mathbf{x}\| \geq \langle \mathbf{y}, \mathbf{w}-\mathbf{x} \rangle, \quad \forall\, \mathbf{w} \in \mathcal{H}. \tag{3}$$

So

$$\left\langle \mathbf{y}, \frac{\mathbf{w}-\mathbf{x}}{\|\mathbf{w}-\mathbf{x}\|} \right\rangle \leq 1, \quad \forall\, \mathbf{w} \neq \mathbf{x}.$$

Therefore $\|\mathbf{y}\|^* \leq 1$. Thus $\partial\|\mathbf{x}\| \subset S$.

For every $\mathbf{y} \in S$, we have

$$\langle \mathbf{y}, \mathbf{w}-\mathbf{x} \rangle = \langle \mathbf{y}, \mathbf{w} \rangle - \langle \mathbf{y}, \mathbf{x} \rangle = \langle \mathbf{y}, \mathbf{w} \rangle - \|\mathbf{x}\| \leq \|\mathbf{y}\|^* \|\mathbf{w}\| - \|\mathbf{x}\| \leq \|\mathbf{w}\| - \|\mathbf{x}\|, \quad \forall\, \mathbf{w} \in \mathcal{H}, \tag{4}$$

where the second equality utilizes $\langle \mathbf{y}, \mathbf{x} \rangle = \|\mathbf{x}\|$ and the first inequality is by the definition of dual norm. Thus, $\mathbf{y} \in \partial\|\mathbf{x}\|$. So $S \subset \partial\|\mathbf{x}\|$.

Basic properties and examples

- Subgradient

Theorem 1 (Danskin's Theorem)**.** *Let $\mathcal{Z}$ be a compact subset of $\mathbb{R}^m$, and let $\phi : \mathbb{R}^n \times \mathcal{Z} \to \mathbb{R}$ be continuous and such that $\phi(\cdot, \mathbf{z}) : \mathbb{R}^n \to \mathbb{R}$ is convex for each $\mathbf{z} \in \mathcal{Z}$. Define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(\mathbf{x}) = \max_{\mathbf{z} \in \mathcal{Z}} \phi(\mathbf{x}, \mathbf{z})$ and*

$$\mathcal{Z}(\mathbf{x}) = \left\{ \bar{\mathbf{z}} \,\middle|\, \phi(\mathbf{x}, \bar{\mathbf{z}} = \max_{\mathbf{z} \in \mathcal{Z}} \phi(\mathbf{x}, \mathbf{z})) \right\}.$$

If $\phi(\cdot, \mathbf{z})$ is differentiable for all $\mathbf{z} \in \mathcal{Z}$ and $\nabla_x \phi(\mathbf{x}, \cdot)$ is continuous on $\mathcal{Z}$ for each $\mathbf{x}$, then

$$\partial f(\mathbf{x}) = conv\{\nabla_x \phi(\mathbf{x}, \mathbf{z}) | \mathbf{z} \in \mathcal{Z}(\mathbf{x})\}, \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

Basic properties and examples

- Subgradient

Example: $\partial\|\mathbf{X}\|_*$, $\partial\|\mathbf{X}\|_2$.

Operations that preserve convexity

- Nonnegative weighted sums

A nonnegative weighted sum of convex functions,

$$f = w_1 f_1 + \cdots + w_m f_m,$$

is convex.

Operations that preserve convexity

- Composition with an affine mapping

Suppose $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{A} \in \mathbb{R}^{n \times m}$, and $\mathbf{b} \in \mathbb{R}^n$. Define $g : \mathbb{R}^m \to \mathbb{R}$ by

$$g(\mathbf{x}) = f(\mathbf{A}\mathbf{x} + \mathbf{b}),$$

with $\operatorname{dom} g = \{\mathbf{x} \mid \mathbf{A}\mathbf{x} + \mathbf{b} \in \operatorname{dom} f\}$. Then if f is convex, so is g.

Operations that preserve convexity

- Composition with an affine mapping

Suppose $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{A} \in \mathbb{R}^{n \times m}$, and $\mathbf{b} \in \mathbb{R}^n$. Define $g : \mathbb{R}^m \to \mathbb{R}$ by

$$g(\mathbf{x}) = f(\mathbf{A}\mathbf{x} + \mathbf{b}),$$

with $\operatorname{dom} g = \{\mathbf{x} \mid \mathbf{A}\mathbf{x} + \mathbf{b} \in \operatorname{dom} f\}$. Then if f is convex, so is g.

Operations that preserve convexity

- Pointwise maximum and supremum

If f_1 and f_2 are convex functions then their *pointwise maximum* f, defined by

$$f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x})\},$$

with $\operatorname{dom} f = \operatorname{dom} f_1 \cap \operatorname{dom} f_2$, is also convex.
Example. 1 (Piecewise-linear functions): The function

$$f(\mathbf{x}) = \max\{\langle \mathbf{a}_1, \mathbf{x}\rangle + \mathbf{b}_1, \ldots, \langle \mathbf{a}_L, \mathbf{x}\rangle + \mathbf{b}_L\}$$

defines a piecewise-linear (or really, affine) function (with L or fewer regions). It is convex since it is the pointwise maximum of affine functions.

Operations that preserve convexity

- Pointwise maximum and supremum

2. (Sum of r largest components): For $\mathbf{x} \in \mathbb{R}^n$ we denote by $x_{[i]}$ the ith largest component of $\mathbf{x}$, *i.e.*,

$$x_{[1]} \geq x_{[2]} \geq \cdots \geq x_{[n]}$$

are the components of $\mathbf{x}$ sorted in nonincreasing order. Then the function

$$f(\mathbf{x}) = \sum_{i=1}^{r} x_{[i]},$$

i.e., the sum of the r largest elements of $\mathbf{x}$, is a convex function.

Operations that preserve convexity

- Pointwise maximum and supremum

3. (Support function of a set): Let $C \subseteq \mathbb{R}^n$, with $C \neq \emptyset$. The *support function* S_C associated with the set C is defined as

$$S_C(\mathbf{x}) = \sup\{\langle \mathbf{x}, \mathbf{y} \rangle \mid \mathbf{y} \in C\}$$

(and, naturally, $\operatorname{dom} S_C = \{\mathbf{x} \mid \sup_{\mathbf{y} \in C} \langle \mathbf{x}, \mathbf{y} \rangle < \infty\}$).

4. (Distance to farthest point of a set): Let $C \subseteq \mathbb{R}^n$. The distance (in any norm) to the farthest point of C,

$$f(\mathbf{x}) = \sup_{\mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\|,$$

is convex.

Operations that preserve convexity

- Pointwise maximum and supremum

5. (Least-squares cost as a function of weights): Let $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$. In a weighted least-squares problem we minimize the objective function $\sum_{i=1}^n w_i(\langle \mathbf{a}_i, \mathbf{x}\rangle - b_i)^2$ over $\mathbf{x} \in \mathbb{R}^m$. We refer to w_i as *weights*, and allow negative w_i (which opens the possibility that the objective function is unbounded below).

We define the (optimal) *weighted least-squares cost* as

$$g(\mathbf{w}) = \inf_{\mathbf{x}} \sum_{i=1}^n w_i \left(\langle \mathbf{a}_i, \mathbf{x}\rangle - b_i\right)^2,$$

with domain

$$\operatorname{dom} g = \left\{ \mathbf{w} \;\middle|\; \inf_{\mathbf{x}} \sum_{i=1}^n w_i \left(\langle \mathbf{a}_i, \mathbf{x}\rangle - b_i\right)^2 > -\infty \right\}.$$

Since g is the infimum of a family of linear functions of $\mathbf{w}$ (indexed by $\mathbf{x} \in \mathbb{R}^m$), it is a concave function of $\mathbf{w}$.

Operations that preserve convexity

- Pointwise maximum and supremum

7. (Norm of a matrix): Consider $f(\mathbf{X}) = \|\mathbf{X}\|_2$ with $\operatorname{dom} f = \mathbb{R}^{p \times q}$, where $\|\cdot\|_2$ denotes the spectral norm or maximum singular value. Convexity of f follows from

$$f(\mathbf{X}) = \sup\{\mathbf{u}^T \mathbf{X} \mathbf{v} \mid \|\mathbf{u}\|_2 = 1,\ \|\mathbf{v}\|_2 = 1\},$$

which shows it is the pointwise supremum of a family of linear functions of $\mathbf{X}$.

Operations that preserve convexity

- Composition – Scalar composition

We examine conditions on $h : \mathbb{R}^k \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^k$ that guarantee convexity or concavity of their composition $f = h \circ g : \mathbb{R}^n \to \mathbb{R}$, defined by

$$f(\mathbf{x}) = h(g(\mathbf{x})), \quad \operatorname{dom} f = \{\mathbf{x} \in \operatorname{dom} g \mid g(\mathbf{x}) \in \operatorname{dom} h\}.$$

$$n = 1: \quad f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x).$$

When $\operatorname{dom} h = \mathbb{R}$:

f is convex if h is **convex** and **nondecreasing**, and g is **convex**,
f is convex if h is **convex** and **nonincreasing**, and g is **concave**.

When $\operatorname{dom} h \neq \mathbb{R}$, change h to $\tilde{h}$!

Example: $g(x) = x^2$, with $\operatorname{dom} g = \mathbb{R}$, and $h(x) = 0$, with $\operatorname{dom} h = [1, 2]$. Here g is convex, and h is convex and nondecreasing. But the function $f = h \circ g$, given by

$$f(x) = 0, \quad \operatorname{dom} f = [-\sqrt{2}, -1] \cup [1, \sqrt{2}],$$

is not convex, since its domain is not convex.

Operations that preserve convexity

- Composition – Scalar composition

Examples:

- If g is convex then $\exp g(x)$ is convex.
- If g is concave and positive, then $\log g(x)$ is concave.
- If g is concave and positive, then $1/g(x)$ is convex.
- If g is convex and nonnegative and $p \geq 1$, then $g(x)^p$ is convex.
- If g is convex then $-\log(-g(x))$ is convex on $\{x \mid g(x) < 0\}$.

Operations that preserve convexity

- Composition – Vector composition

$$f(\mathbf{x}) = h(g(\mathbf{x})) = h(g_1(\mathbf{x}), \ldots, g_k(\mathbf{x})),$$

with $h : \mathbb{R}^k \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$. Again without loss of generality we can assume $n = 1$. As in the case $k = 1$, we start by assuming the functions are twice differentiable, with $\operatorname{dom} g = \mathbb{R}$ and $\boxed{\operatorname{dom} h = \mathbb{R}^k}$, in order to discover the composition rules. We have

$$f''(x) = (\mathbf{g}'(x))^T \nabla^2 h(\mathbf{g}(x)) \mathbf{g}'(x) + \nabla h(\mathbf{g}(x))^T \mathbf{g}''(x).$$

f is convex if $\tilde{h}$ is **convex** and **nondecreasing** in each argument, and g_i is **convex**,
f is convex if $\tilde{h}$ is **convex** and **nonincreasing** in each argument, and g_i is **concave**.

Operations that preserve convexity

- Composition – Vector composition

Examples:

- Let $h(\mathbf{z}) = z_{[1]} + \cdots + z_{[r]}$, the sum of the r largest components of $\mathbf{z} \in \mathbb{R}^k$. Then h is convex and nondecreasing in each argument. Suppose $g_1, \ldots, g_k$ are convex functions on $\mathbb{R}^n$. Then the composition function $f = h \circ \mathbf{g}$, *i.e.*, the pointwise sum of the r largest g_i's, is convex.

- The function $h(\mathbf{z}) = \log(\sum_{i=1}^k e^{z_i})$ is convex and nondecreasing in each argument, so $\log(\sum_{i=1}^k e^{g_i(x)})$ is convex whenever g_i are.

- For $0 < p \leq 1$, the function $h(\mathbf{z}) = (\sum_{i=1}^k z_i^p)^{1/p}$ on $\mathbb{R}_+^k$ is concave, and its extension (which has the value $-\infty$ for $\mathbf{z} \not\succeq \mathbf{0}$ is nondecreasing in each component. So if g_i are concave and nonnegative, we conclude that $f(x) = (\sum_{i=1}^k g_i(x)^p)^{1/p}$ is concave.

Operations that preserve convexity

- Composition – Vector composition

Examples:

- Suppose $p \geq 1$, and $g_1, \ldots, g_k$ are convex and nonnegative. Then the function $f(x) = (\sum_{i=1}^k g_i(x)^p)^{1/p}$ is convex. To show this, we consider the function $h : \mathbb{R}^k \to \mathbb{R}$ defined as

$$h(\mathbf{z}) = \left(\sum_{i=1}^k \max\{z_i, 0\}^p \right)^{1/p},$$

with $\boxed{\operatorname{dom} h = \mathbb{R}^k}$, so $h = \tilde{h}$. This function is convex, and nondecreasing, so we conclude $h(\mathbf{g}(x))$ is a convex function of x. For $\mathbf{z} \succeq \mathbf{0}$, we have $h(\mathbf{z}) = (\sum_{i=1}^k z_i^p)^{1/p}$, so our conclusion is that $(\sum_{i=1}^k g_i(x)^p)^{1/p}$ is convex.

- The geometric mean $h(\mathbf{z}) = (\prod_{i=1}^k z_i)^{1/k}$ on $\mathbb{R}_+^k$ is concave and its extension is nondecreasing in each argument. It follows that if $g_1, \ldots, g_k$ are nonnegative concave functions, then so is their geometric mean, $(\prod_{i=1}^k g_i)^{1/k}$.

Operations that preserve convexity

- Minimization

If f is convex in $(\mathbf{x}, \mathbf{y})$, and C is a convex nonempty set, then the function

$$g(\mathbf{x}) = \inf_{\mathbf{y} \in C} f(\mathbf{x}, \mathbf{y}) \tag{1}$$

$$\operatorname{epi} g = \{(\mathbf{x}, t) \mid (\mathbf{x}, \mathbf{y}, t) \in \operatorname{epi} f \text{ for some } \mathbf{y} \in C\}.$$

is convex in $\mathbf{x}$, provided $g(\mathbf{x}) > -\infty$ for some $\mathbf{x}$ (which implies $g(\mathbf{x}) > -\infty$ for all $\mathbf{x}$). The domain of g is the projection of $\operatorname{dom} f$ on its $\mathbf{x}$-coordinates, *i.e.*,

$$\operatorname{dom} g = \{\mathbf{x} \mid (\mathbf{x}, \mathbf{y}) \in \operatorname{dom} f \text{ for some } \mathbf{y} \in C\}.$$

Examples:

- The distance of a point $\mathbf{x}$ to a set $S \subseteq \mathbb{R}^n$, in the norm $\|\cdot\|$, is defined as

$$\operatorname{dist}(\mathbf{x}, S) = \inf_{\mathbf{y} \in S} \|\mathbf{x} - \mathbf{y}\|.$$

- Suppose h is convex. Then the function g defined below is convex:

$$g(\mathbf{x}) = \inf\{h(\mathbf{y}) \mid \mathbf{A}\mathbf{y} = \mathbf{x}\}.$$

Operations that preserve convexity

- Perspective of a function

If $f : \mathbb{R}^n \to \mathbb{R}$, then the *perspective* of f is the function $g : \mathbb{R}^{n+1} \to \mathbb{R}$ defined by

$$g(\mathbf{x}, t) = tf(\mathbf{x}/t),$$

with domain

$$\operatorname{dom} g = \{(\mathbf{x}, t) \mid \mathbf{x}/t \in \operatorname{dom} f,\ t > 0\}.$$

The perspective operation preserves convexity: If f is a convex function, then so is its perspective function g.

$$\begin{aligned}(\mathbf{x}, t, s) \in \operatorname{epi} g &\iff tf(\mathbf{x}/t) \leq s \\ &\iff f(\mathbf{x}/t) \leq s/t \\ &\iff (\mathbf{x}/t, s/t) \in \operatorname{epi} f.\end{aligned}$$

Operations that preserve convexity

- Perspective of a function

Examples:

- The perspective of the convex function $f(\mathbf{x}) = \langle \mathbf{x}, \mathbf{x} \rangle$ on $\mathbb{R}^n$ is

$$g(\mathbf{x}, t) = t \langle \mathbf{x}/t, \mathbf{x}/t \rangle = \frac{\langle \mathbf{x}, \mathbf{x} \rangle}{t},$$

  which is convex in $(\mathbf{x}, t)$ for $t > 0$.

- Consider the convex function $f(x) = -\log x$ on $\mathbb{R}_{++}$. Its perspective is

$$g(x, t) = -t \log(x/t) = t \log(t/x) = t \log t - t \log x,$$

  and is convex on $\mathbb{R}^2_{++}$. The function g is called the *relative entropy* of t and x.

The conjugate function

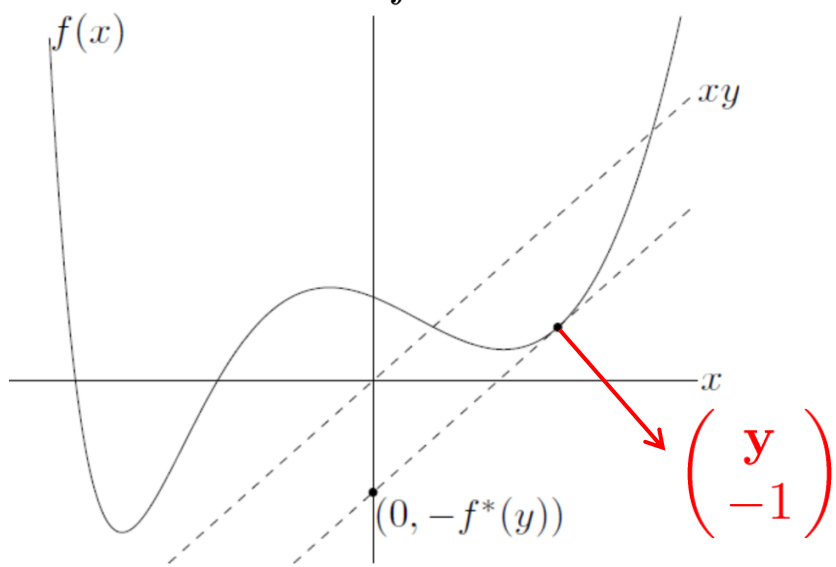
- Definition and examples

$\boxed{f^*(\mathbf{y}) \text{ is always convex!}}$

Let $f : \mathbb{R}^n \to \mathbb{R}$. The function

$$f^*(\mathbf{y}) = \sup_{\mathbf{x} \in \operatorname{dom} f} (\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x})) = \sup_{(\mathbf{x},t) \in \operatorname{epi} f} \begin{pmatrix} \mathbf{y} \\ -1 \end{pmatrix}^T \begin{pmatrix} \mathbf{x} \\ t \end{pmatrix}, \tag{1}$$

is called the *conjugate* of the function f. The domain of the conjugate function consists of $\mathbf{y} \in \mathbb{R}^n$ for which the supremum is finite, *i.e.*, for which the difference $\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x})$ is bounded above on $\operatorname{dom} f$.



The conjugate function

- Definition and examples
- *Affine function.* $f(x) = ax + b$. As a function of x, $yx - ax - b$ is bounded if and only if $y = a$. Therefore $\mathrm{dom}\, f^* = \{a\}$ and $f^*(a) = -b$.

- *Negative logarithm.* $f(x) = -\log x$, with $\mathrm{dom}\, f = \mathbb{R}_{++}$. The function $xy + \log x$ is unbounded above if $y \geq 0$ and reaches its maximum at $x = -1/y$ otherwise. Therefore, $\mathrm{dom}\, f^* = \{y \mid y < 0\} = -\mathbb{R}_{++}$ and $f^*(y) = -\log(-y) - 1$ for $y < 0$.

- *Exponential.* $f(x) = e^x$. $xy - e^x$ is unbounded if $y < 0$. For $y > 0$, $xy - e^x$ reaches its maximum at $x = \log y$, so we have $f^*(y) = y \log y - y$. For $y = 0$, $f^*(y) = \sup_x -e^x = 0$. So $\mathrm{dom}\, f^* = \mathbb{R}_+$ and $f^*(y) = y \log y - y$.

- *Negative entropy.* $f(x) = x \log x$, with $\mathrm{dom}\, f = \mathbb{R}_+$. The function $xy - x \log x$ is bounded above on $\mathbb{R}_+$ for all y, hence $\mathrm{dom}\, f^* = \mathbb{R}$. It attains its maximum at $x = e^{y-1}$, and substituting we find $f^*(y) = e^{y-1}$.

- *Inverse.* $f(x) = 1/x$ on $\mathbb{R}_{++}$. For $y > 0$, $yx - 1/x$ is unbounded above. For $y = 0$ this function has supremum 0; for $y < 0$ the supremum is attained at $x = (-y)^{-1/2}$. So $f^*(y) = -2(-y)^{1/2}$, with $\mathrm{dom}\, f^* = -\mathbb{R}_+$.

The conjugate function

- Definition and examples

Example 1 (Strictly convex quadratic function)**.** *Consider* $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x}$, *with* $\mathbf{Q} \in \mathbb{S}^n_{++}$. *The function* $\langle \mathbf{y}, \mathbf{x} \rangle - \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x}$ *is bounded above as a function of* $\mathbf{x}$ *for all* $\mathbf{y}$. *It attains its maximum at* $\mathbf{x} = \mathbf{Q}^{-1}\mathbf{y}$, *so*

$$f^*(\mathbf{y}) = \frac{1}{2}\mathbf{y}^T\mathbf{Q}^{-1}\mathbf{y}.$$

Example 2 (Indicator function)**.** *Let* I_S *be the indicator function of a (not necessarily convex) set* $S \subseteq \mathbb{R}^n$, i.e., $I_S(\mathbf{x}) = 0$ *on* $\operatorname{dom} I_S = S$. *Its conjugate is*

$$I_S^*(\mathbf{y}) = \sup_{\mathbf{x} \in S} \langle \mathbf{y}, \mathbf{x} \rangle ,$$

which is the support function of the set S.

The conjugate function

- Definition and examples

Example 3 (Log-determinant)**.** *We consider* $f(\mathbf{X}) = \log\det \mathbf{X}^{-1}$ *on* $\mathbb{S}^n_{++}$.

$$f^*(\mathbf{Y}) = \sup_{\mathbf{X}\succ\mathbf{0}} (\mathrm{tr}\,(\mathbf{YX}) + \log\det \mathbf{X}).$$

We first show that $\mathrm{tr}\,(\mathbf{YX}) + \log\det\mathbf{X}$ *is unbounded above unless* $\mathbf{Y} \prec \mathbf{0}$*. If* $\mathbf{Y} \not\prec \mathbf{0}$*, then* $\mathbf{Y}$ *has an eigenvector* $\mathbf{v}$*, with* $\|\mathbf{v}\|_2 = 1$*, and eigenvalue* $\lambda \geq 0$*. Taking* $\mathbf{X} = \mathbf{I} + t\mathbf{v}\mathbf{v}^T$ *we find that*

$$\begin{aligned}\mathrm{tr}\,(\mathbf{YX}) + \log\det\mathbf{X} &= \mathrm{tr}\,(\mathbf{Y}) + t\lambda + \log\det(\mathbf{I} + t\mathbf{v}\mathbf{v}^T)\\ &= \mathrm{tr}\,(\mathbf{Y}) + t\lambda + \log(1+t) \longrightarrow +\infty.\end{aligned}$$

Now consider the case $\mathbf{Y} \prec \mathbf{0}$*. The optimum* $\mathbf{X}$ *satisfies:*

$$\nabla_{\mathbf{X}}(\mathrm{tr}\,(\mathbf{YX}) + \log\det\mathbf{X}) = \mathbf{Y} + \mathbf{X}^{-1} = \mathbf{0},$$

which yields $\mathbf{X} = -\mathbf{Y}^{-1}$*. Therefore we have*

$$f^*(\mathbf{Y}) = \log\det(-\mathbf{Y})^{-1} - n, \quad \mathrm{dom}\, f^* = -\mathbb{S}^n_{++}.$$

The conjugate function

- Definition and examples

Example 4 (Log-sum-exp function)**.** *We consider* $f(\mathbf{x}) = \log(\sum_{i=1}^n e^{x_i})$. *By setting the gradient of* $\langle \mathbf{y}, \mathbf{x}\rangle - f(\mathbf{x})$ *w.r.t.* $\mathbf{x}$ *equal to zero, we obtain the condition*

$$y_i = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}, \quad i = 1, \ldots, n.$$

These equations are solvable for $\mathbf{x}$ *if and only if* $\mathbf{y} \succ \mathbf{0}$ *and* $\langle \mathbf{1}, \mathbf{y}\rangle = 1$. *By substituting the expression for* y_i *into* $\langle \mathbf{y}, \mathbf{x}\rangle - f(\mathbf{x})$ *we obtain* $f^*(\mathbf{y}) = \sum_{i=1}^n y_i \log y_i$. *This expression for* f^* *is still correct if some components of* $\mathbf{y}$ *are zero, as long as* $\mathbf{y} \succeq \mathbf{0}$ *and* $\langle \mathbf{1}, \mathbf{y}\rangle = 1$, *and we interpret* $0 \log 0$ *as* 0.

In fact $\operatorname{dom} f^* = \{\mathbf{y} | \langle \mathbf{1}, \mathbf{y}\rangle = 1, \mathbf{y} \succeq \mathbf{0}\}$. *To show this, suppose that a component of* $\mathbf{y}$ *is negative, say,* $y_k < 0$. *Then we can show that* $\langle \mathbf{y}, \mathbf{x}\rangle - f(\mathbf{x})$ *is unbounded above by choosing* $x_k = -t$, *and* $x_i = 0$, $i \neq k$, *and letting* t *go to infinity.*

If $\mathbf{y} \succeq \mathbf{0}$ *but* $\langle \mathbf{1}, \mathbf{y}\rangle \neq 1$, *we choose* $\mathbf{x} = t\mathbf{1}$, *so that*

$$\langle \mathbf{y}, \mathbf{x}\rangle - f(\mathbf{x}) = t(\langle \mathbf{1}, \mathbf{y}\rangle - 1) - \log n \longrightarrow \infty.$$

The conjugate function

- Definition and examples

Example 5 (Norm)**.** *Let* $\|\cdot\|$ *be a norm on* $\mathbb{R}^n$*, with dual norm* $\|\cdot\|_*$*. The conjugate of* $f(\mathbf{x}) = \|\mathbf{x}\|$ *is*

$$f^*(\mathbf{y}) = I_{\mathcal{B}}(\mathbf{y}),$$

where $\mathcal{B} = \{\mathbf{y} | \|\mathbf{y}\|_* \leq 1\}$ *is the unit ball of the dual norm.*

Example 6 (Norm squared)**.** *Now consider the function* $f(\mathbf{x}) = (1/2)\|\mathbf{x}\|^2$*, where* $\|\cdot\|$ *is a norm, with dual norm* $\|\cdot\|_*$*. Its conjugate is* $f^*(\mathbf{y}) = (1/2)\|\mathbf{y}\|_*^2$*.*

Example 7 (Support function)**.** $f(\mathbf{x}) = S_{\mathcal{C}}(\mathbf{x}) = \sup_{\mathbf{y}\in\mathcal{C}} \langle \mathbf{x}, \mathbf{y} \rangle$*, where* $\mathcal{C}$ *is a closed convex set. Then its conjugate is the indicator function of* $\mathcal{C}$*:* $f^*(\mathbf{y}) = I_{\mathcal{C}}(\mathbf{y})$*.*

The conjugate function

- Basic properties

Fenchel's inequality:

$$f(\mathbf{x}) + f^*(\mathbf{y}) \geq \langle \mathbf{x}, \mathbf{y} \rangle, \quad \forall \mathbf{x}, \mathbf{y}.$$

For example, if $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2$ then $f^*(\mathbf{y}) = \frac{1}{2}\|\mathbf{y}\|_*^2$ and thus

$$\frac{1}{2}\|\mathbf{x}\|^2 + \frac{1}{2}\|\mathbf{y}\|_*^2 \geq \langle \mathbf{x}, \mathbf{y} \rangle,$$

which is generalization of $\frac{1}{2}\|\mathbf{x}\|_2^2 + \frac{1}{2}\|\mathbf{y}\|_2^2 \geq \langle \mathbf{x}, \mathbf{y} \rangle$.

When $f(\mathbf{x}) = (1/2)\mathbf{x}^T\mathbf{Q}\mathbf{x}$, where $\mathbf{Q} \in \mathbb{S}_{++}^n$, we obtain the inequality

$$\langle \mathbf{x}, \mathbf{y} \rangle \leq (1/2)\mathbf{x}^T\mathbf{Q}\mathbf{x} + (1/2)\mathbf{y}^T\mathbf{Q}^{-1}\mathbf{y}.$$

The conjugate function

- Basic properties

Conjugate of the conjugate

Theorem: If f is proper, convex and closed, then $f^{**} = f$.

Proof. It is easy to prove that $f^{**} \leq f$, i.e. epi$f \subseteq$ epif^{**}. Suppose $\exists(\mathbf{x}_0^T, \gamma)^T \notin$ epi$f$, where $\gamma \geq f^{**}(\mathbf{x}_0)$. Since f is proper, convex and closed, its epigraph is closed and does not include a vertical line. Thus there exists a hyperplane parameterized as $(\mathbf{w}^T, \zeta)^T$ $(\zeta \neq 0)$ and μ such that it strictly separates epi$f$ and the point $(\mathbf{x}_0^T, \gamma)^T$:

$$(\mathbf{w}^T, \zeta)(\mathbf{x}^T, t)^T < \mu < (\mathbf{w}^T, \zeta)(\mathbf{x}_0^T, \gamma)^T, \quad \forall \mathbf{x} \in \operatorname{dom} f, t \geq f(\mathbf{x}).$$

Since t can be arbitrarily large, ζ must be negative. W.l.o.g, we assume $\zeta = -1$. Then taking $t = f(\mathbf{x})$,

$$\mathbf{w}^T\mathbf{x} - f(\mathbf{x}) < \mu < \mathbf{w}^T\mathbf{x}_0 - \gamma \leq \mathbf{w}^T\mathbf{x}_0 - f^{**}(\mathbf{x}_0), \quad \forall \mathbf{x} \in \operatorname{dom} f.$$

Thus $f^*(\mathbf{w}) < \mathbf{w}^T\mathbf{x}_0 - f^{**}(\mathbf{x}_0)$, contradicting Fenchel's inequality.

The conjugate function

- Basic properties

Useful for finding a convex surrogate of a nonconvex function!

Conjugate of the conjugate

Theorem: For any f, f^{**} is the largest convex function not exceeding f, i.e., *convex envelop* of f.

Proof. We first have $f^{***}(\mathbf{y}) = f^*(\mathbf{y})$ because $f^*(\mathbf{y})$ is a propoer, convex and closed function. And it is easy to prove that if $f_1(\mathbf{x}) \leq f_2(\mathbf{x})$, then $f_1^*(\mathbf{y}) \geq f_2^*(\mathbf{y})$.

Suppose g is proper, convex and closed, and $f^{**}(\mathbf{x}) \leq g(\mathbf{x}) \leq f(\mathbf{x})$. Then by the above results,

$$f^*(\mathbf{y}) \leq g^*(\mathbf{y}) \leq f^{***}(\mathbf{y}) = f^*(\mathbf{y}).$$

Thus $f^*(\mathbf{y}) = g^*(\mathbf{y})$ and hence $f^{**} = g^{**} = g$.

The conjugate function

- Basic properties

Conjugate of the conjugate

Examples. 1. The convex envelope of ℓ_0-pseudo-norm on unit ℓ_1 ball is the ℓ_1-norm.

Proof. We first compute $f^*(\mathbf{y}) = \sup_{\|\mathbf{x}\|\leq 1}(\langle \mathbf{y}, \mathbf{x}\rangle - \|\mathbf{x}\|_0)$. When $\|\mathbf{x}\| \leq 1$,

$$f^*(\mathbf{y}) \leq \|\mathbf{y}\|_\infty \|\mathbf{x}\|_1 - \|\mathbf{x}\|_0 \leq \|\mathbf{y}\|_\infty - \|\mathbf{x}\|_0.$$

$$f^*(\mathbf{y}) = \begin{cases} \|\mathbf{y}\|_\infty - 1, & \text{if } \|\mathbf{y}\|_\infty \geq 1, \\ 0, & \text{if } \|\mathbf{y}\|_\infty < 1 \end{cases} = \max(\|\mathbf{y}\|_\infty - 1, 0).$$

Next, we compute $f^{**}(\mathbf{x}) = \sup_{\mathbf{y}}(\langle \mathbf{y}, \mathbf{x}\rangle - f^*(\mathbf{y}))$:

$$f^{**}(\mathbf{x}) \leq \|\mathbf{y}\|_\infty \|\mathbf{x}\|_1 - \max(\|\mathbf{y}\|_\infty - 1, 0).$$

$$f^{**}(\mathbf{x}) = \begin{cases} \|\mathbf{x}\|_1, & \text{if } \|\mathbf{x}\|_1 \leq 1, \\ +\infty, & \text{if } \|\mathbf{x}\|_1 > 1. \end{cases}$$

The conjugate function

- Basic properties

Conjugate of the conjugate

Examples. 2. The convex envelope of rank on unit 2-norm ball is the nuclear-norm.

Proof. 1) Compute rank*. Let $q = \min(m, n)$. By von Neumann's trace thm,

$$\text{rank}^*(\mathbf{Y}) = \sup_{\|\mathbf{X}\|_2 \le 1} \{\langle \mathbf{X}, \mathbf{Y} \rangle - \text{rank}(\mathbf{X})\} \le \sup_{\|\mathbf{X}\|_2 \le 1} \left\{ \sum_{i=1}^{q} \sigma_i(\mathbf{X})\sigma_i(\mathbf{Y}) - \text{rank}(\mathbf{X}) \right\}.$$

The equality can hold when $\mathbf{X}$ and $\mathbf{Y}$ have the same singular subspaces. Suppose $\text{rank}(\mathbf{X}) = r$, $0 \le r \le q$. Then

$$\text{rank}^*(\mathbf{Y}) = \sup_{\|\mathbf{X}\|_2 \le 1} \left\{ \sum_{i=1}^{r} \sigma_i(\mathbf{Y}) - r \right\} = \sup_{\|\mathbf{X}\|_2 \le 1} \left\{ \sum_{i=1}^{r} (\sigma_i(\mathbf{Y}) - 1) \right\}.$$

So r should be chosen such that $\sigma_r(\mathbf{Y}) > 1$ and $\sigma_{r+1}(\mathbf{Y}) \le 1$, and

$$\text{rank}^*(\mathbf{Y}) = \sum_{i=1}^{r} (\sigma_i(\mathbf{Y}) - 1).$$

The conjugate function

- Basic properties

2) Compute rank**. By von Neumann's trace thm. again,

$$\mathrm{rank}^{**}(\mathbf{X}) = \sup_{\mathbf{Y}}\{\langle \mathbf{X}, \mathbf{Y}\rangle - \mathrm{rank}^{*}(\mathbf{Y})\} \leq \sup_{\mathbf{Y}} \left\{ \sum_{i=1}^{q} \sigma_i(\mathbf{X})\sigma_i(\mathbf{Y}) - \sum_{i=1}^{r} (\sigma_i(\mathbf{Y}) - 1) \right\},$$

where r is such that $\sigma_r(\mathbf{Y}) > 1$ and $\sigma_{r+1}(\mathbf{Y}) \leq 1$. The equality can also hold.

If $\|\mathbf{X}\|_2 > 1$, we can choose $\sigma_1(\mathbf{Y})$ large enough so that $\mathrm{rank}^{**}(\mathbf{X}) \to +\infty$, because the coefficient of $\sigma_1(\mathbf{Y})$ is a positive value $\sigma_1(\mathbf{X}) - 1$. If $\|\mathbf{X}\|_2 \leq 1$,

$$\begin{aligned}
&\sum_{i=1}^{q} \sigma_i(\mathbf{X})\sigma_i(\mathbf{Y}) - \sum_{i=1}^{r} (\sigma_i(\mathbf{Y}) - 1) \\
&= \sum_{i=1}^{q} \sigma_i(\mathbf{X})\sigma_i(\mathbf{Y}) - \sum_{i=1}^{r} (\sigma_i(\mathbf{Y}) - 1) - \sum_{i=1}^{q} \sigma_i(\mathbf{X}) + \sum_{i=1}^{q} \sigma_i(\mathbf{X}) \\
&= \sum_{i=1}^{r} (\sigma_i(\mathbf{X}) - 1)(\sigma_i(\mathbf{Y}) - 1) + \sum_{i=r+1}^{q} \sigma_i(\mathbf{X})(\sigma_i(\mathbf{Y}) - 1) + \sum_{i=1}^{q} \sigma_i(\mathbf{X}) \\
&\leq \sum_{i=1}^{q} \sigma_i(\mathbf{X}) = \|\mathbf{X}\|_*.
\end{aligned}$$

The equality can hold. Thus $\mathrm{rank}^{**}(\mathbf{X}) = \|\mathbf{X}\|_*$ over the set $\{\mathbf{X} | \|\mathbf{X}\|_2 \leq 1\}$.

The conjugate function

- Basic properties

Differentiable functions

The conjugate of a differentiable function f is also called the *Legendre transform* of f.

Suppose f is convex and differentiable, with $\boxed{\mathrm{dom}\, f = \mathbb{R}^n.}$ Any maximizer $\mathbf{x}^*$ of $\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x})$ satisfies $\mathbf{y} = \nabla f(\mathbf{x}^*)$, and conversely, if $\mathbf{x}^*$ satisfies $\mathbf{y} = \nabla f(\mathbf{x}^*)$, then $\mathbf{x}^*$ maximizes $\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x})$. Therefore, if $\mathbf{y} = \nabla f(\mathbf{x}^*)$, we have

$$f^*(\mathbf{y}) = \mathbf{x}^{*T} \nabla f(\mathbf{x}^*) - f(\mathbf{x}^*).$$

This allows us to determine $f^*(\mathbf{y})$ for any $\mathbf{y}$ for which we can solve the gradient equation $\mathbf{y} = \nabla f(\mathbf{z})$ for $\mathbf{z}$.

We can express this another way. Let $\mathbf{z} \in \mathbb{R}^n$ be arbitrary and define $\mathbf{y} = \nabla f(\mathbf{z})$. Then we have

$$f^*(\mathbf{y}) = \mathbf{z}^T \nabla f(\mathbf{z}) - f(\mathbf{z}).$$

The conjugate function

- Basic properties

Useful for finding the gradient for the dual problem!

Proposition 5. *If f is convex and closed, then*

$$\mathbf{y} \in \partial f(\mathbf{x}) \Leftrightarrow \mathbf{x} \in \partial f^*(\mathbf{y}) \Leftrightarrow f(\mathbf{x}) + f^*(\mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle.$$

Corollary 6. *If f is strictly convex, then f^* is differentiable, with*

$$\nabla f^*(\mathbf{y}) = \operatorname*{argmax}_{\mathbf{x}} \left\{ \langle \mathbf{x}, \mathbf{y} \rangle - f(\mathbf{x}) \right\}.$$

The conjugate function

- Basic properties

Proposition 6. *If f is ρ-strongly convex, then f^* is ρ^{-1}-smooth:*

$$\|\nabla f^*(\mathbf{y}_1) - \nabla f^*(\mathbf{y}_2)\| \leq \rho^{-1}\|\mathbf{y}_1 - \mathbf{y}_2\|.$$

Proof:

$$\mathbf{x}_i = \operatorname*{argmax}_{\mathbf{x}} \langle \mathbf{x}, \mathbf{y}_i \rangle - f(\mathbf{x}) \Longrightarrow \mathbf{y}_i \in \partial f(\mathbf{x}_i), \mathbf{x}_i = \nabla f^*(\mathbf{y}_i), i = 1, 2.$$

$$f(\mathbf{x}_2) - f(\mathbf{x}_1) \geq \langle \mathbf{y}_1, \mathbf{x}_2 - \mathbf{x}_1 \rangle + \frac{\rho}{2}\|\mathbf{x}_1 - \mathbf{x}_2\|^2,$$
$$f(\mathbf{x}_1) - f(\mathbf{x}_2) \geq \langle \mathbf{y}_2, \mathbf{x}_1 - \mathbf{x}_2 \rangle + \frac{\rho}{2}\|\mathbf{x}_1 - \mathbf{x}_2\|^2.$$

$$\|\mathbf{y}_1 - \mathbf{y}_2\|\|\mathbf{x}_1 - \mathbf{x}_2\| \geq \langle \mathbf{y}_1 - \mathbf{y}_2, \mathbf{x}_1 - \mathbf{x}_2 \rangle \geq \rho\|\mathbf{x}_1 - \mathbf{x}_2\|^2.$$
$$\rho^{-1}\|\mathbf{y}_1 - \mathbf{y}_2\| \geq \|\mathbf{x}_1 - \mathbf{x}_2\| = \|\nabla f^*(\mathbf{y}_1) - \nabla f^*(\mathbf{y}_2)\|.$$

The conjugate function

- Basic properties

Scaling and composition with affine transformation

For $a > 0$ and $b \in \mathbb{R}$, the conjugate of $g(x) = af(x) + b$ is $g^*(y) = af^*(y/a) - b$.

Suppose $\mathbf{A} \in \mathbb{R}^{n \times n}$ is nonsingular and $\mathbf{b} \in \mathbb{R}^n$. Then the conjugate of $g(\mathbf{x}) = f(\mathbf{A}\mathbf{x} + \mathbf{b})$ is

$$g^*(\mathbf{y}) = f^*(\mathbf{A}^{-T}\mathbf{y}) - \mathbf{b}^T\mathbf{A}^{-T}\mathbf{y},$$

with $\operatorname{dom} g^* = \mathbf{A}^T \operatorname{dom} f^*$.

The conjugate function

- Basic properties

Separable functions

If $f(\mathbf{u}, \mathbf{v}) = f_1(\mathbf{u}) + f_2(\mathbf{v})$, where f_1 and f_2 are convex functions with conjugates f_1^* and f_2^*, respectively, then

$$f^*(\mathbf{w}, \mathbf{z}) = f_1^*(\mathbf{w}) + f_2^*(\mathbf{z}).$$

In other words, the conjugate of a *separable* convex function is the sum of the conjugates.

Envelope function and Proximal mapping

Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a closed proper function. For a scalar $c > 0$, define the corresponding *envelope function* $\mathrm{Env}_c f$ and the *proximal mapping* $\mathrm{Prox}_c\, f$ by

Involution: $(f \boxtimes g)(\mathbf{x}) = \inf_{\mathbf{y}+\mathbf{z}=\mathbf{x}} f(\mathbf{y}) + g(\mathbf{z})$.

$$\begin{aligned} \mathrm{Env}_c f(\mathbf{x}) &= \inf_{\mathbf{w}} \left\{ f(\mathbf{w}) + \frac{1}{2c}\|\mathbf{w}-\mathbf{x}\|^2 \right\}, \\ \mathrm{Prox}_c\, f(\mathbf{x}) &= \operatorname*{argmin}_{\mathbf{w}} \left\{ f(\mathbf{w}) + \frac{1}{2c}\|\mathbf{w}-\mathbf{x}\|^2 \right\}. \end{aligned} \tag{1}$$

Remark:

1. The envelope function $\mathrm{Env}_c f$ is an underestimate of the function f, i.e., $\mathrm{Env}_c f(\mathbf{x}) \le f(\mathbf{x})$, $\forall \mathbf{x} \in \mathbb{R}^n$. Furthermore, $\mathrm{Env}_c f$ is a real-valued continuous function, whereas f itself may only be extended real-valued and lower semi-continuous.

2. It is obvious that $\mathbf{u} = \mathrm{Prox}_c\, f(\mathbf{x}) \Leftrightarrow \mathbf{x} - \mathbf{u} \in \partial f(\mathbf{u})$.

Envelope function and Proximal mapping

Usefulness of proximal mapping in optimization

1. One of the two basic update schema

forward

backward

Let $T = I - \alpha \nabla f$ and $\tilde{T} = (I + \alpha \nabla f)^{-1}$. Then

$$\text{either } \boxed{\mathbf{x}_{k+1} = T(\mathbf{x}_k)} \text{ or } \boxed{\mathbf{x}_{k+1} = \tilde{T}(\mathbf{x}_k)}.$$

$$\boxed{\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha \nabla f(\mathbf{x}_k)} \qquad \boxed{\mathbf{x}_{k+1} = \operatorname*{argmin}_{\mathbf{x}} f(\mathbf{x}) + \frac{1}{2\alpha}\|\mathbf{x} - \mathbf{x}_k\|^2}$$

2. Sufficient descent and encourages convergence

$$\mathbf{x}_{k+1} = \operatorname*{argmin}_{\mathbf{x}} \left\{ f(\mathbf{x}) + \frac{1}{2\alpha}\|\mathbf{x} - \mathbf{x}_k\|^2 \right\}$$

$$\Rightarrow f(\mathbf{x}_{k+1}) + \frac{1}{2\alpha}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \leq f(\mathbf{x}_k)$$

$$\Rightarrow f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k) \leq -\frac{1}{2\alpha}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2$$

KL condition

$$\Rightarrow \sum_{k=0}^{+\infty} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 < +\infty \Rightarrow \sum_{k=0}^{+\infty} \|\mathbf{x}_{k+1} - \mathbf{x}_k\| < +\infty$$

Envelope function and Proximal mapping

Examples: $f(\mathbf{x}) = \chi_C(\mathbf{x})$

Projection onto halfspace: $\mathcal{C} = \{\mathbf{x} | \mathbf{a}^T\mathbf{x} \leq b\}$ $(\mathbf{a} \neq \mathbf{0})$:

$$P_{\mathcal{C}}(\mathbf{x}) = \begin{cases} \mathbf{x} + \frac{b - \mathbf{a}^T\mathbf{x}}{\|\mathbf{a}\|^2}\mathbf{a}, & \text{if } \mathbf{a}^T\mathbf{x} > b, \\ \mathbf{x}, & \text{if } \mathbf{a}^T\mathbf{x} \leq b. \end{cases}$$

Projection onto hyperbox $\mathcal{C} = \{\mathbf{x} | \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}\}$:

$$(P_{\mathcal{C}}(\mathbf{x}))_i = \begin{cases} l_i, & \text{if } x_i \leq l_i, \\ x_i, & \text{if } l_i \leq \mathbf{x}_i \leq u_i, \\ u_i, & \text{if } x_i \geq u_i. \end{cases}$$

Projection onto nonnegative orthant: $\mathcal{C} = \mathbb{R}^n_+$:

$$P_{\mathcal{C}}(\mathbf{x}) = \max(\mathbf{x}, \mathbf{0}).$$

Envelope function and Proximal mapping

Projection onto positive semi-definite cone: $\mathcal{C} = \mathbb{S}_+^n$:

$$P_{\mathcal{C}}(\mathbf{X}) = \sum_{i=1}^{n} \max(\lambda_i, 0)\mathbf{q}_i\mathbf{q}_i^T,$$

where $\sum_{i=1}^{n} \lambda_i \mathbf{q}_i \mathbf{q}_i^T$ is the eigenvalue decomposition of $\mathbf{X}$.
Projection onto unit Euclidean ball: $\mathcal{C} = \{\mathbf{x} | \|\mathbf{x}\| \leq 1\}$:

$$P_{\mathcal{C}}(\mathbf{x}) = \begin{cases} \frac{1}{\|\mathbf{x}\|}\mathbf{x}, & \text{if } \|\mathbf{x}\| > 1, \\ \mathbf{x}, & \text{if } \|\mathbf{x}\| \leq 1. \end{cases}$$

Projection onto unit matrix 2-norm ball: $\mathcal{C} = \{\mathbf{X} | \|\mathbf{X}\|_2 \leq 1\}$:

$$P_{\mathcal{C}}(\mathbf{X}) = \begin{cases} \mathbf{U}\,\mathrm{Diag}(\min(\boldsymbol{\sigma}, 1))\mathbf{V}^T, & \text{if } \|\mathbf{X}\|_2 > 1, \\ \mathbf{X}, & \text{if } \|\mathbf{X}\|_2 \leq 1. \end{cases}$$

where $\mathbf{U}\,\mathrm{Diag}(\boldsymbol{\sigma})\mathbf{V}^T$ is the SVD of $\mathbf{X}$.

Envelope function and Proximal mapping

Examples: $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^t\mathbf{x} + \mathbf{c}$, $\|\mathbf{x}\|_1$, $-\sum_{i=1}^{n} \log x_i$, $f(\mathbf{X}) = \|\mathbf{X}\|_*$

Theorem 1. *For each $\tau > 0$ and $\mathbf{Y} \in \mathbb{R}^{n_1 \times n_2}$,*

$$\mathcal{D}_\tau(\mathbf{Y}) = \operatorname*{argmin}_{\mathbf{X}} \left\{ \tau\|\mathbf{X}\|_* + \frac{1}{2}\|\mathbf{X} - \mathbf{Y}\|_F^2 \right\}, \tag{1}$$

where $\mathcal{D}_\tau(\mathbf{Y})$ is the singular value thresholding operator defined as

$$\mathcal{D}_\tau(\mathbf{Y}) = \mathbf{U}\,\mathrm{diag}(\{(\sigma_i - \tau)_+\})\mathbf{V}^T, \tag{2}$$

in which $\mathbf{U}\,\mathrm{diag}(\{\sigma_i\})\mathbf{V}^T$ is the SVD of $\mathbf{Y}$.

Envelope function and Proximal mapping

Separable functions

If $f(\mathbf{u}, \mathbf{v}) = f_1(\mathbf{u}) + f_2(\mathbf{v})$, where f_1 and f_2 are closed proper functions, then

$$\mathrm{Prox}_c f(\mathbf{u}, \mathbf{v}) = \mathrm{Prox}_c f_1(\mathbf{u}) + \mathrm{Prox}_c f_2(\mathbf{v}).$$

Scaling and composition with orthogonal transformation

For $a > 0$ and $b \in \mathbb{R}$, the conjugate of $g(x) = f(ax + b)$ is

$$\mathrm{Prox}_c g(x) = a^{-1}(\mathrm{Prox}_{ca^2} f(ax + b) - b).$$

Suppose $g(\mathbf{x}) = f(\mathbf{A}\mathbf{x} + \mathbf{b})$, where $\mathbf{A} \in \mathbb{R}^{n \times m}$ satisfies $\mathbf{A}\mathbf{A}^T = \lambda^{-1}\mathbf{I}$ $(\lambda > 0)$ and $\mathbf{b} \in \mathbb{R}^n$. Then

$$\mathrm{Prox}_c g(\mathbf{x}) = (\mathbf{I} - \lambda \mathbf{A}^T \mathbf{A})\mathbf{x} + \lambda \mathbf{A}^T (\mathrm{Prox}_{c\lambda^{-1}} f(\mathbf{A}\mathbf{x} + \mathbf{b}) - \mathbf{b}).$$

Envelope function and Proximal mapping

Proof. $\mathbf{w} = \mathrm{Prox}_c\, g(\mathbf{x})$ is the solution of the optimization problem:

$$\min_{\mathbf{w},\mathbf{z}} f(\mathbf{z}) + \frac{1}{2c}\|\mathbf{w} - \mathbf{x}\|^2, \quad s.t. \quad \mathbf{A}\mathbf{w} + \mathbf{b} = \mathbf{z}.$$

Eliminating $\mathbf{w}$ gives: $\mathbf{w} = \mathbf{x} + \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{z} - \mathbf{b} - \mathbf{A}\mathbf{x}) = (\mathbf{I} - \lambda\mathbf{A}^T\mathbf{A})\mathbf{x} + \lambda\mathbf{A}^T(\mathbf{z} - \mathbf{b})$. The optimal $\mathbf{z}$ is the minimizer of

$$f(\mathbf{z}) + \frac{\lambda^2}{2c}\|\mathbf{A}^T(\mathbf{z} - \mathbf{b} - \mathbf{A}\mathbf{x})\|^2 = f(\mathbf{z}) + \frac{\lambda}{2c}\|\mathbf{z} - \mathbf{b} - \mathbf{A}\mathbf{x}\|^2,$$

which is $\mathbf{z} = \mathrm{Prox}_{c\lambda^{-1}}\, f(\mathbf{A}\mathbf{x} + \mathbf{b})$.

Envelope function and Proximal mapping

Theorem 2. *For any function f, $\mathrm{Prox}_c f(\mathbf{x})$ is a monotonic function of $\mathbf{x}$ in the sense that:*

$$\langle \mathbf{x}_1 - \mathbf{x}_2, \mathbf{y}_1 - \mathbf{y}_2 \rangle \geq 0, \quad \forall \mathbf{y}_i \in Prox_c f(\mathbf{x}_i), i = 1, 2. \tag{1}$$

Proposition 1. *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a closed proper convex function and $c > 0$. The envelope function $Env_c f$ is convex and smooth and its gradient is given by*

$$\nabla Env_c f(\mathbf{x}) = \frac{1}{c}(\mathbf{x} - Prox_c f(\mathbf{x})). \tag{2}$$

The envelope function $\mathrm{Env}_c f$ is smooth, regardless of whether f is smooth.

Envelope function and Proximal mapping

Proposition 2. *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a closed proper convex function and $c > 0$. The proximal mapping $\operatorname{Prox}_c f$ is single-valued and is continuous: $\operatorname{Prox}_c f(x) \to \operatorname{Prox}_{c^*} f(\mathbf{x}^*)$ whenever $(\mathbf{x}, c) \to (\mathbf{x}^*, c^*)$, with $c^* > 0$.*

Theorem 3 (Moreau Decomposition)**.** $\mathbf{x} = \operatorname{Prox}_c f(\mathbf{x}) + c \operatorname{Prox}_{c^{-1}} f^*(c^{-1}\mathbf{x})$.

Proof. Let $\mathbf{u} = \operatorname{Prox}_c f(\mathbf{x})$. Then

$$\begin{aligned}\mathbf{u} = \operatorname{Prox}_c f(\mathbf{x}) &\iff -c^{-1}(\mathbf{u} - \mathbf{x}) \in \partial f(\mathbf{u}) \\ &\iff \mathbf{u} \in \partial f^*(-c^{-1}(\mathbf{u} - \mathbf{x}))\end{aligned}$$

Let $\mathbf{z} = -c^{-1}(\mathbf{u} - \mathbf{x})$. Then $\mathbf{u} = \mathbf{x} - c\mathbf{z}$ and thus

$$\begin{aligned}&\iff \mathbf{x} - c\mathbf{z} \in \partial f^*(\mathbf{z}) \\ &\iff \mathbf{0} \in \partial f^*(\mathbf{z}) + c(\mathbf{z} - c^{-1}\mathbf{x}) \\ &\iff \mathbf{z} = \operatorname{Prox}_{c^{-1}} f^*(c^{-1}\mathbf{x}).\end{aligned}$$

So $\mathbf{x} = \mathbf{u} + c\mathbf{z} = \operatorname{Prox}_c f(\mathbf{x}) + c \operatorname{Prox}_{c^{-1}} f^*(c^{-1}\mathbf{x})$.

Envelope function and Proximal mapping

Example: proximal mapping of a norm.
We know that if $f(\mathbf{x}) = \|\mathbf{x}\|$, then $f^*(\mathbf{y}) = I_{\mathcal{B}}(\mathbf{y})$, where $\mathcal{B}$ is the unit ball of the dual norm $\|\cdot\|_*$. Then by Moreay decomposition:

$$\begin{aligned}
\mathrm{Prox}_c f(\mathbf{x}) &= \mathbf{x} - c\mathrm{Prox}_{c^{-1}} f^*(\mathbf{x}/c) \\
&= \mathbf{x} - cP_{\mathcal{B}}(\mathbf{x}/c) \\
&= \mathbf{x} - P_{c\mathcal{B}}(\mathbf{x}).
\end{aligned}$$

Examples: $f(\mathbf{x}) = \|\mathbf{x}\|_1$, $\|\mathbf{x}\|_2$, $f(\mathbf{X}) = \|\mathbf{X}\|_2$, $\|\mathbf{X}\|_*$.