02611 Optimization for Data Science (F25)

Convex sets and functions

Martin S. Andersen

Technical University of Denmark

Outline

Affine combinations and sets

An affine combination of k points $x_1, \ldots, x_k \in \mathbb{R}^n$ is a linear combination of the form

$$y = \sum_{i=1}^{k} \theta_i x_i, \quad \mathbb{1}^T \theta = 1.$$

A set is an affine set if it contains all affine combinations of its points; can always be written as

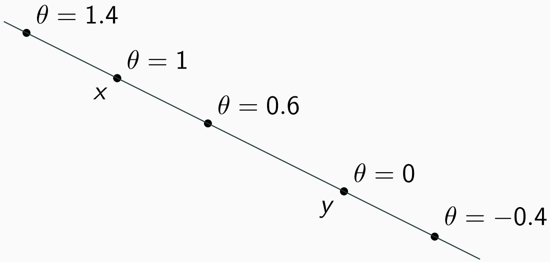
$$\mathcal{A} = \{x \in \mathbb{R}^n \,|\, Ax = b\}$$

The affine hull of a set $\mathcal{S} \subseteq \mathbb{R}^n$, denoted $\operatorname{aff} \mathcal{S}$, is the smallest affine set containing $\mathcal{S}$.

Affine combinations and sets (cont.)

Affine hull of two distinct points x and y

$$\mathrm{aff}\{x, y\} = \{\theta x + (1-\theta)y \mid \theta \in \mathbb{R}\}$$

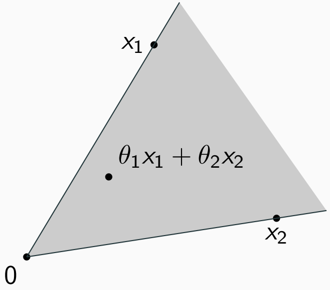


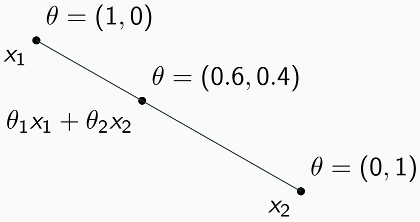
Conic and convex combinations

A linear combination of $x_1, \dots, x_k \in \mathbb{R}^n$,

$$y = \sum_{i=1}^{k} \theta_i x_i,$$

- is a conic combination if $\theta \in \mathbb{R}^k_+$
- is a convex combination if $\theta \in \Delta^k = \{\theta \in \mathbb{R}^k_+ \mid \mathbb{1}^T \theta = 1\}$



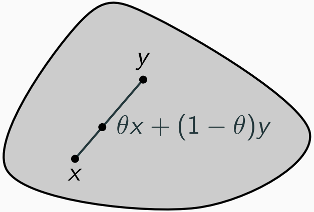


Convex sets

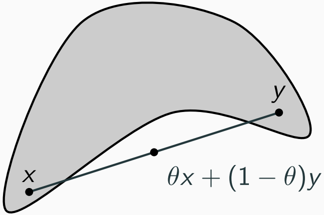
A set $C \subseteq \mathbb{R}^n$ is convex iff for all $x, y \in C$,

$$\theta x + (1-\theta)y \in C, \quad \forall \theta \in [0,1],$$

i.e., C contains all line segments connecting points in C.



Convex set



Non-convex set

Convex sets (cont.)

The dimension of a convex set $C \subseteq \mathbb{R}^n$ is the dimension of its affine hull,

$$\dim C = \dim(\operatorname{aff} C).$$

The relative interior of C is the interior of C within $\operatorname{aff} C$,

$$\operatorname{relint} C = \{x \in C \mid \exists\, \epsilon > 0 \text{ such that } B_2(x, \epsilon) \cap \operatorname{aff} C \subseteq C\},$$

where $B_2(x, \epsilon)$ is the Euclidean ball centered at x with radius ϵ.

The convex hull of a set $A \subseteq \mathbb{R}^n$ is the smallest convex set that contains A,

$$\operatorname{conv} A = \cap\{S \subseteq \mathbb{R}^n \mid S \text{ is convex and } A \subseteq S\}.$$

Convexity-preserving operations

- intersection
- affine transformation (image and preimage)
- perspective transformation

Intersection

The intersection of convex sets is convex: if C_τ is convex for all $\tau \in T$, then

$$C = \bigcap_{\tau \in T} C_\tau$$

is convex.

Proof follows from the definition of convexity:

$$x, y \in C \implies x, y \in C_\tau, \ \forall \tau \in T$$

and hence for all $x, y \in C$ and $\theta \in [0, 1]$,

$$\theta x + (1 - \theta) y \in C_\tau, \ \forall \tau \in T \implies \theta x + (1 - \theta) y \in C$$

Affine transformation

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be an affine function defined as $f(x) = Ax + b$ with $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

The image of a convex set $C \subseteq \mathbb{R}^n$ under f

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

is convex.

The preimage of a convex set $C \subseteq \mathbb{R}^m$ under f

$$f^{-1}(C) = \{x \in \mathbb{R}^n \,|\, f(x) \in C\}$$

is convex.

Perspective transformation

The perspective function $P\colon \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}^n$ is defined as

$$P(u, s) = \frac{u}{s}.$$

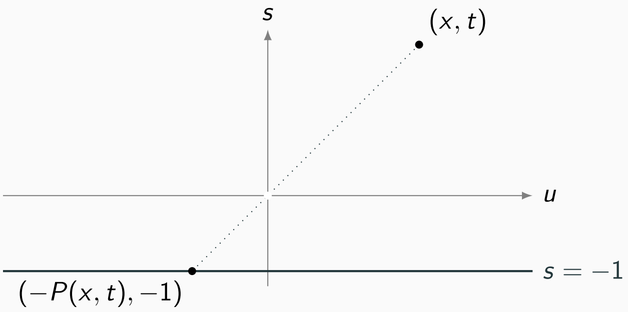
The perspective of a set $C \subseteq \operatorname{dom} P$ is the image of C under P

$$P(C) = \{u/s \,|\, (u, s) \in C\}.$$

If C is convex, then $P(C)$ is convex.

Perspective transformation (cont.)

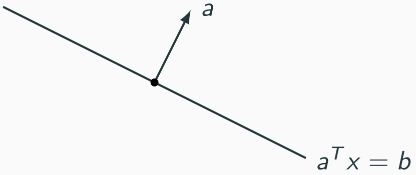
Interpretation: pinhole camera

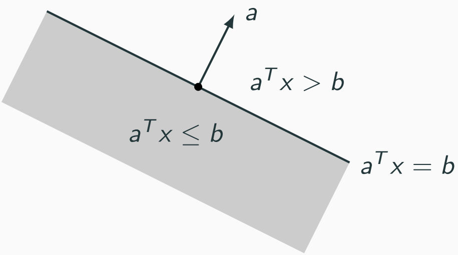


Hyperplanes and halfspaces

Given $a \in \mathbb{R}^n$ $(a \neq 0)$ and $b \in \mathbb{R}$

- the set $\mathcal{H} = \{x \in \mathbb{R}^n \mid a^T x = b\}$ is called a hyperplane
- the set $\{x \in \mathbb{R}^n \mid a^T x \leq b\}$ is called a (closed) halfspace

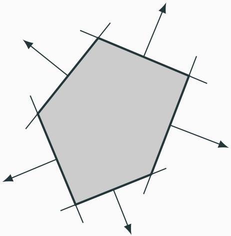


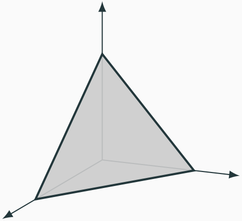


Polyhedral set

A polyhedral set is the intersection of a finite number of halfspaces

$$\mathcal{P} = \{x \in \mathbb{R}^n \mid a_i^T x \leq b_i, \ \ i = 1, \ldots, m\}$$





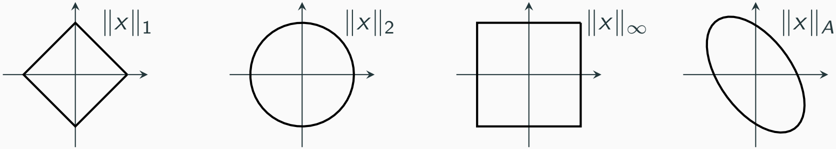
Norm balls and ellipsoids

Given a norm $\|\cdot\|$ on $\mathbb{R}^n$, the set

$$B(c, r) = \{x \in \mathbb{R}^n \mid \|x - c\| \leq r\}$$

is a norm ball centered at c with radius $r > 0$.

An ellipsoid is a norm ball induced by a quadratic norm $\|\cdot\|_A$ for some $A \in \mathbb{S}^n_{++}$.



Convex cones

A set $K \subseteq \mathbb{R}^n$ is a cone if $x \in K \implies tx \in K$ for all $t \geq 0$.

- K is pointed if $K \cap (-K) = \{0\}$
- K is full-dimensional if aff $K = \mathbb{R}^n$ (or equivalently, int $K \neq \emptyset$)
- a proper cone is convex, closed, pointed, and full-dimensional
- a polyhedral cone is the intersection of a finite number of halfspaces

Example $\mathbb{S}^n_+ = \{A \in \mathbb{S}^n \mid x^T A x \geq 0 \; \forall x \in \mathbb{R}^n\}$ is a proper cone in $\mathbb{S}^n$

$$A \in \mathbb{S}^n_+ \implies tA \in \mathbb{S}^n_+ \; \forall\, t \geq 0 \qquad \text{and} \qquad \mathbb{S}^n_+ = \bigcap_{x \in \mathbb{R}^n} \{A \in \mathbb{S}^n \mid \langle A, xx^T \rangle \geq 0\}$$

i.e., $\mathbb{S}^n_+$ is the intersection of infinitely many closed halfspaces

Norm cones

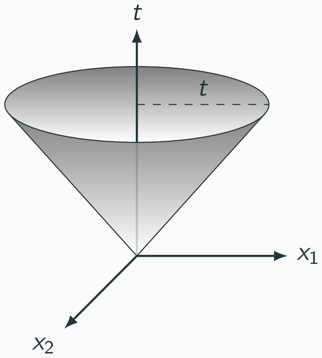
The norm cone associated with a norm $\|\cdot\|$ on $\mathbb{R}^{n-1}$ is the set

$$K = \{(x, t) \in \mathbb{R}^{n-1} \times \mathbb{R} \mid \|x\| \leq t\}$$

Example

Second-order cone (Lorentz cone)

$$\mathbb{Q}^n = \{(x, t) \in \mathbb{R}^{n-1} \times \mathbb{R} \mid \|x\|_2 \leq t\}$$

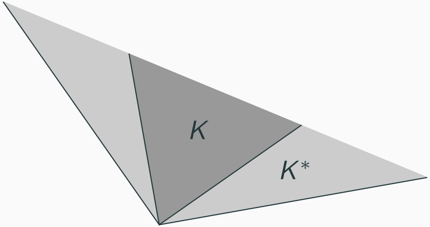


Dual cone

The dual cone of a convex cone $K \subseteq \mathbb{R}^n$ is the set

$$K^* = \{y \in \mathbb{R}^n \,|\, y^T x \geq 0 \; \forall\, x \in K\}$$

- K^* is always convex and closed
- $K^{**} = K$ if K is a proper cone
- K is self-dual if $K^* = K$
- $-(K^*)$ is called the polar cone of K



Generalized inequalities

A proper cone $K \subseteq \mathbb{R}^n$ defines a relation $\succeq_K$ and a generalized inequality defined as

$$x \succeq_K y \iff x - y \in K \quad \text{and} \quad x \succ_K y \iff x - y \in \operatorname{int} K$$

- $K = \mathbb{R}_+$ yields the usual inequality $x \geq y$ for $x, y \in \mathbb{R}$ ($\geq$ is a total order on $\mathbb{R}$)
- $K = \mathbb{R}^n_+$ yields the componentwise inequality: $x \succeq_K y \iff x_i \geq y_i$ for all $i \in \mathbb{N}_n$
- $K = \mathbb{S}^n_+$ yields the Loewner order: $A \succeq_K B \iff A - B \in \mathbb{S}^n_+$

The relation $\succeq_K$ generally defines a partial order on $\mathbb{R}^n$ with the following properties:

1. $x \succeq_K x$ for all $x \in \mathbb{R}^n$ (reflexivity)
2. if $x \succeq_K y$ and $y \succeq_K x$, then $x = y$ (antisymmetry)
3. if $x \succeq_K y$ and $y \succeq_K z$, then $x \succeq_K z$ (transitivity)

Generalized inequalities (cont.)

The strict relation $\succ_K$ defines a strict partial order on $\mathbb{R}^n$ with the following properties:

1. there is no $x \in \mathbb{R}^n$ such that $x \succ_K x$ (irreflexivity)
2. if $x \succ_K y$, then $y \not\succ_K x$ (asymmetry)
3. if $x \succ_K y$ and $y \succ_K z$, then $x \succ_K z$ (transitivity)

Transitivity implies that

$$a \succeq_K b, \; c \succeq_K d \implies a + c \succeq_K b + d$$
$$a \succ_K b, \; c \succ_K d \implies a + c \succ_K b + d$$

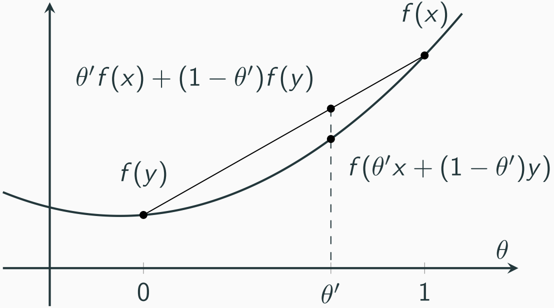
Example Define $K = \mathbb{R}^2_+$, $x = (3, 1)$, and $y = (0, 2)$.

$$x \succ_K 0, \quad y \succeq_K 0, \quad x \not\succeq_K y, \quad y \not\succeq_K x$$

Convex functions

$f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex if its domain $\operatorname{dom} f$ is a convex set and for all $x, y \in \operatorname{dom} f$,

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y), \quad \forall\, \theta \in [0,1].$$

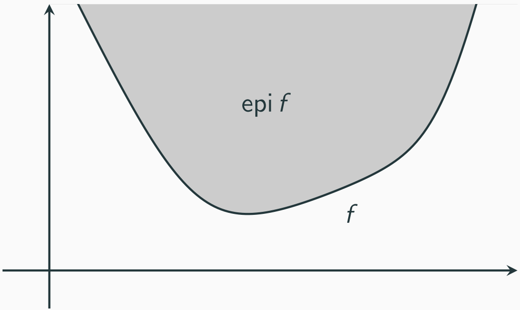


f is concave if $-f$ is convex

Convex functions: epigraph characterization

$f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex if and only if its epigraph is a convex set, *i.e.*, for all $x, y \in \operatorname{dom} f$

$$\theta \begin{bmatrix} x \\ f(x) \end{bmatrix} + (1-\theta) \begin{bmatrix} y \\ f(y) \end{bmatrix} \in \operatorname{epi} f, \quad \forall\, \theta \in [0,1]$$



Strict and strong convexity

$f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is

- strictly convex if and only if for all $x, y \in \operatorname{dom} f$ with $x \neq y$

$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta)f(y), \quad \forall\, \theta \in [0,1]$$

- strongly convex with parameter $\mu > 0$ if and only if for all $x, y \in \operatorname{dom} f$

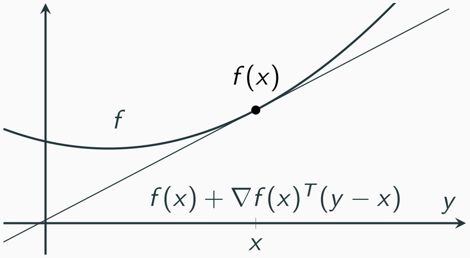
$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) - \frac{\theta(1-\theta)\mu}{2}\|x-y\|_2^2, \quad \forall\, \theta \in [0,1]$$

  equivalently, f is μ-strongly convex if and only if $g(x) = f(x) - (\mu/2)\|x\|_2^2$ is convex

First-order condition for convexity

Suppose $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is continuously differentiable and $\operatorname{dom} f$ is open and convex. Then f is convex if and only if

$$f(y) \geq f(x) + \nabla f(x)^T (y - x), \quad \forall x, y \in \operatorname{dom} f$$



- first-order Taylor expansion of f provides a global affine lower bound
- x is a global minimizer of f if and only if $\nabla f(x) = 0$

First-order condition for convexity (cont.)

Strict convexity f is strictly convex if and only if for all $x, y \in \operatorname{dom} f$ with $x \neq y$

$$f(y) > f(x) + \nabla f(x)^T (y - x)$$

Strong convexity f is μ-strongly convex if and only if for all $x, y \in \operatorname{dom} f$

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{\mu}{2} \|y - x\|_2^2$$

- a stationary point of f is a unique global minimizer of f if f is strictly or strongly convex
- strong convexity implies that sublevel sets of f are bounded

Second-order condition for convexity

A twice continuously differentiable function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex on $\operatorname{dom} f$ if and only if

$$\nabla^2 f(x) \succeq 0, \quad \forall\, x \in \operatorname{dom} f$$

- $\nabla^2 f(x) \succ 0$ for all $x \in \operatorname{dom} f$ implies that f is strictly convex (the converse is not true)
- $\nabla^2 f(x) \succeq \mu I$ for all $x \in \operatorname{dom} f$ and for some $\mu > 0$ implies that f is μ-strongly convex

Convexity-preserving operations

- scaling, sums, and integrals
- pointwise maximum and supremum
- affine transformation
- perspective transformation
- partial infimum
- square of nonnegative convex function

Nonnegative scaling if f is convex, then so is αf for $\alpha \geq 0$

Sum if f and g are proper convex functions, then $f + g$ is proper convex

Integral if $f : \mathbb{R}^n \times \mathbb{R}^p \to \overline{\mathbb{R}}$ and $f(x, y)$ is proper convex in x for each $y \in \mathbb{R}^p$, then

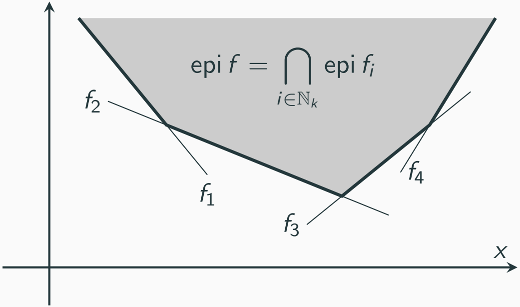
$$h(x) = \int_{\mathbb{R}^p} f(x, y)\, dy$$

is proper convex

Pointwise maximum and supremum

The pointwise maximum of a family of convex functions $f_i \colon \mathbb{R}^n \to \overline{\mathbb{R}}$ for $i \in \mathbb{N}_k$ is itself convex.

$$f(x) = \max_{i \in \mathbb{N}_k} f_i(x)$$



Pointwise maximum and supremum (cont.)

The pointwise supremum of a family of uncountably many convex functions is itself convex.

Let $f : \mathbb{R}^n \times \mathbb{R}^p \to (-\infty, +\infty]$ and suppose $f(x, y)$ is convex in x for each $y \in \mathcal{Y}$. Then

$$h(x) = \sup_{y \in \mathcal{Y}} f(x, y)$$

is convex.

Affine transformation

Suppose $f\colon \mathbb{R}^m \to (-\infty, \infty]$ is convex and $g\colon \mathbb{R}^n \to (-\infty, +\infty]$ is defined as

$$g(x) = f(Ax + b), \quad A \in \mathbb{R}^{m \times n},\; b \in \mathbb{R}^m.$$

Then g is convex since $\operatorname{epi} g$ is the preimage of $\operatorname{epi} f$ under an affine transformation, *i.e.*,

$$\operatorname{epi} g = \{(x, t) \mid f(Ax + b) \leq t\} = \{(x, t) \mid (Ax + b, t) \in \operatorname{epi} f\}.$$

If f is twice continuously differentiable, then by convexity of f,

$$\nabla^2 g(x) = A^T \nabla^2 f(Ax + b) A \succeq 0.$$

Perspective transformation

The perspective of a function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is the function $P_f : \mathbb{R}^n \times \mathbb{R} \to (-\infty, +\infty]$ defined as

$$P_f(x, t) = \begin{cases} tf(x/t), & t > 0, \\ +\infty, & t \leq 0. \end{cases}$$

If f is proper convex, then P_f is convex.

Example Suppose $g(x, t) = x^T x/t$ with $\operatorname{dom} g = \mathbb{R}^n \times \mathbb{R}_{++}$.

We can write g as $g(x, t) = P_f(x, t)$ for $f(x) = x^T x$, and hence g is convex.

Partial infimum

The partial infimum of $f\colon \mathbb{R}^n \times \mathbb{R}^p \to (-\infty, +\infty]$ is a function $h\colon \mathbb{R}^n \to (-\infty, +\infty]$, defined as

$$h(x) = \inf_y f(x, y).$$

If f is convex, then h is convex.

Let $T\colon \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}$ be the linear transformation defined as $T(x, y, t) = (x, t)$. Then

$$T(\{(x, y, t) \mid f(x, y) < t\}) = \{(x, t) \mid h(x) < t\}.$$

Interpretation: the strict epigraph of h is the image of that of f under T.

Partial infimum (cont.)

Example Let $f : \mathbb{R}^n \to \mathbb{R}$ be defined as

$$f(x, y) = \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}^T \qquad \text{where} \qquad \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \in \mathbb{S}^n_{++}.$$

Convexity of f implies that the partial infimum

$$h(x) = \inf_y \; f(x, y) = x^T(A - BC^{-1}B^T)x$$

is convex, and hence the Schur complement $A - BC^{-1}B^T$ must be positive semidefinite.

Square of nonnegative convex function

Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be a nonnegative convex function. Then $g = f^2$ is convex.

By convexity of f, we have for all $x, y \in \mathbb{R}^n$ and $\theta \in [0, 1]$,

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y).$$

Squaring both sides yields

$$\begin{aligned} g(\theta x + (1-\theta)y) &\le \theta^2 f(x)^2 + (1-\theta)^2 f(y)^2 + 2\theta(1-\theta)f(x)f(y) \\ &= -\theta(1-\theta)(f(x) - f(y))^2 + \theta f(x)^2 + (1-\theta)f(y)^2 \\ &\le \theta f(x)^2 + (1-\theta)f(y)^2 \\ &= \theta g(x) + (1-\theta)g(y). \end{aligned}$$

Basic examples of convex and concave functions

- Linear and affine functions are both convex and concave.
- Absolute value: $f(x) = |x|$ is convex on $\mathbb{R}$.
- Powers: $f(x) = x^\alpha$ with $\operatorname{dom} f = \mathbb{R}_{++}$ is concave if $\alpha \in [0, 1]$ and convex if $\alpha \notin (0, 1)$.
- Powers of absolute value: $f(x) = |x|^\alpha$ is convex on $\mathbb{R}$ if $\alpha \geq 1$.
- Exponential function: $f(x) = \exp(x)$ is convex on $\mathbb{R}$.
- Logarithm: $f(x) = \ln(x)$ with $\operatorname{dom} f = \mathbb{R}_{++}$ is concave.
- Negative entropy: $f(x) = x \ln(x)$ with $\operatorname{dom} f = \mathbb{R}_+$ and $f(0) = 0$ is convex.
- Quadratic-over-linear: $f(x, y) = x^2/y$ with $\operatorname{dom} f = \mathbb{R} \times \mathbb{R}_{++}$ is convex.
- One-sided square: $f(x) = \max(0, x)^2$ is convex on $\mathbb{R}$.

Norms

All norms are convex functions, which is an immediate consequence of the triangle inequality.

Equivalently, if $f(x) = \|x\|$ is a norm, then

$$\operatorname{epi} f = \{(x, t) \mid \|x\| \leq t\}$$

is a norm cone, which is a convex set.

Indicator and support functions

The indicator function of a set $C \subseteq \mathbb{R}^n$ is the function $I_C \colon \mathbb{R}^n \to \{0, +\infty\}$ defined as

$$I_C(x) = \begin{cases} 0, & x \in C, \\ +\infty, & x \notin C. \end{cases}$$

This is convex function if and only if the set C is convex, and it is proper if $C \neq \emptyset$.

The support function of a nonempty set $C \subseteq \mathbb{R}^n$ is the function $S_C \colon \mathbb{R}^n \to \overline{\mathbb{R}}$ defined as

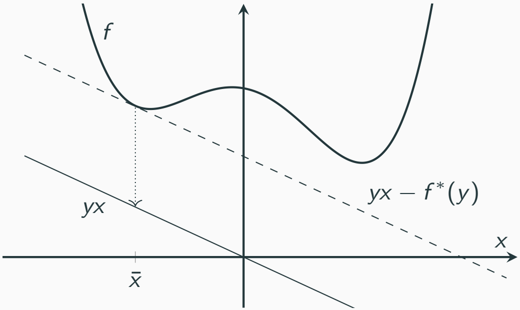
$$S_C(x) = \sup_{y \in C} x^T y.$$

This is always a convex function, even if C is not convex.

Conjugate function

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

$$f^*(y) = \sup_x \{y^T x - f(x)\}. \tag{1}$$



Properties of the conjugate function

- f^* is always convex, even if f is not convex.
- f^* is closed (epi f^* is the intersection of closed halfspaces).
- $f^*(0) = \sup_x \{0 - f(x)\} = -\inf_x f(x)$ is the negative of the infimum of f.
- Fenchel–Young inequality: if $f : \mathbb{R}^n \to (-\infty, +\infty]$ is proper, then

$$f(x) + f^*(y) \geq y^T x, \quad \forall x, y,$$

  with equality if the supremum of $y^T x - f(x)$ is attained at x.
- The biconjugate of f is $f^{**} = (f^*)^*$ and satisfies

$$f(x) \geq \sup_y \{x^T y - f^*(y)\} = f^{**}(x).$$

- f^{**} is the (lower) convex envelope of f
- Fenchel–Moreau theorem: if f is proper, then $f^{**} = f$ if and only if f is convex and closed.

Example: conjugate of strongly convex quadratic form

Let $f : \mathbb{R}^n \to \mathbb{R}$ be defined as

$$f(x) = \frac{1}{2} x^T P x, \qquad P \in \mathbb{S}^n_{++}.$$

The conjugate function is

$$f^*(y) = \sup_x \left\{ y^T x - \frac{1}{2} x^T P x \right\} = \frac{1}{2} y^T P^{-1} y.$$

The Fenchel–Young inequality yields

$$\frac{1}{2} x^T P x + \frac{1}{2} y^T P^{-1} y \geq y^T x$$

with equality iff $y = Px$.

Example: conjugate of indicator function

Let I_C be the indicator function of a set $C \subseteq \mathbb{R}^n$. Then the conjugate function is

$$I_C^*(y) = \sup_x \{y^T x - I_C(x)\} = \sup_{x \in C} y^T x = S_C(y).$$

Special case C is a nonempty convex cone $K \subset \mathbb{R}^n$

$$S_K(y) = \sup_{x \in K} y^T x = I_{-(K^*)}(y)$$

i.e., $I_K^* = I_{-(K^*)}$ is the indicator function of the polar cone $-(K^*)$.

Dual norm

The dual norm of a norm $\|\cdot\|$ on $\mathbb{R}^n$ is defined as

$$\|y\|_* = \sup_{\|x\|\leq 1} y^T x = \sup_x \{y^T x - I_B(x)\} = I_B^*(y) = S_B(y)$$

where $B = \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$.

Alternatively, we can write

$$\|y\|_* = \sup_{x\neq 0} \frac{x^T y}{\|x\|}$$

which leads to a generalized Cauchy–Schwartz inequality

$$\|x\|\|y\|_* \geq |x^T y|, \qquad \forall x, y \in \mathbb{R}^n.$$

Dual of dual norm

The dual of the dual norm is $\|\cdot\|_{**} = \|\cdot\|$.

Proof sketch

- $\|y\|_* > 1 \iff \exists x \colon x^T y / \|x\| > 1 \implies \sup_x \{y^T x - \|x\|\} = +\infty$
- $\|y\|_* \le 1 \iff \nexists x \colon x^T y / \|x\| > 1 \implies \sup_x \{y^T x - \|x\|\} = 0$
- conjugate function of $\|x\|$ is the indicator function I_{B_*} of the dual norm ball

$$B_* = \{y \in \mathbb{R}^n \mid \|y\|_* \le 1\}$$

- $\|\cdot\|_{**} = I_{B_*}^* = \|\cdot\|$

Dual norm examples

- Euclidean norm on $\mathbb{R}^n$: $\|y\|_2 = \sqrt{x^T x}$

$$\|y\|_* = \sup_{\|x\|_2 \leq 1} y^T x = \|y\|_2$$

- ℓ_1 norm on $\mathbb{R}^n$: $\|y\|_1 = \sum_{i=1}^n |y_i|$

$$\|y\|_* = \sup_{\|x\|_1 \leq 1} y^T x = \|y\|_\infty$$

- ℓ_p norm $(p \geq 1)$ on $\mathbb{R}^n$: $\|y\|_p = \left(\sum_{i=1}^n |y_i|^p\right)^{1/p}$

$$\|y\|_* = \sup_{\|x\|_p \leq 1} y^T x = \|y\|_q, \qquad 1/p + 1/q = 1$$

- quadratic norm on $\mathbb{R}^n$: $\|y\|_A = \sqrt{y^T A y}$ with $A \in \mathbb{S}^n_{++}$

$$\|y\|_* = \sup_{\|x\|_A \leq 1} y^T x = \|y\|_{A^{-1}}$$

Dual norm examples (cont.)

- Frobenius norm on $\mathbb{R}^{m\times n}$: $\|Y\|_F = \sqrt{\langle Y, Y\rangle}$

$$\|Y\|_* = \sup_{\|X\|_F \le 1} \langle Y, X\rangle = \|Y\|_F$$

- Spectral norm on $\mathbb{R}^{m\times n}$: $\|Y\|_2 = \sigma_1(Y)$

$$\|Y\|_* = \sup_{\|X\|_2 \le 1} \langle Y, X\rangle = \sum_{i=1}^{\min(m,n)} \sigma_i(Y)$$

(follows from on Neumann's trace inequality and unitary invariance of the spectral norm)