Convex Functions (I)

Lecture 4, Convex Optimization

National Taiwan University

March 17, 2023

Table of contents

Sublevel sets

Sublevel Sets

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

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

Sublevel sets of a convex function are convex

If f is a convex function, then for any $\alpha \in \mathbf{R}$, the α-sublevel set, C_α, is convex.

- The converse is not true. A function can have all its sublevel sets convex, but not be a convex function. (e.g., $f(x) = -e^x$.)
- If f is concave, then its α-**superlevel** set, given by $\{x \in \mathbf{dom}\ f | f(x) \geq \alpha\}$, is a convex set.

Sublevel sets – Example

Example

The geometric and arithmetic means of $x \in \mathbf{R}^n_+$ are

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

respectively. Suppose $0 \leq \beta \leq 1$, then the set

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

is convex since it is the 0-superlevel set of the concave function $G(x) - \beta A(x)$.

- It is also a convex cone.

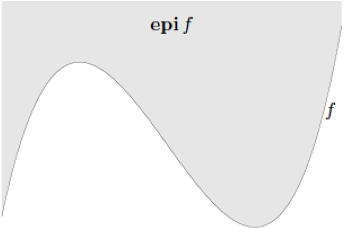
Epigraph

Graph

The **graph** of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as $\{(x, f(x)) \mid x \in \mathbf{dom}\ f\}$, a subset of $\mathbf{R}^{n+1}$.

Epigraph

The **epigraph** of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as **epi** $f = \{(x,t) | x \in \mathbf{dom}\ f, f(x) \le t\}$, which is a subset of $\mathbf{R}^{n+1}$.



Epigraph

Graph

The **graph** of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as $\{(x, f(x)) \mid x \in \mathbf{dom}\ f\}$, a subset of $\mathbf{R}^{n+1}$.

Epigraph

The **epigraph** of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as **epi** $f = \{(x,t)|x \in \mathbf{dom}\ f, f(x) \leq t\}$, which is a subset of $\mathbf{R}^{n+1}$.

The epigraph of convex functions

A function is convex if and only if its epigraph is a convex set.

The hypograph of concave functions

A function is concave if and only if its **hypograph**, defined as **hypo** $f = \{(x,t) \mid x \in \mathbf{dom}\ f, f(x) \geq t\}$, is a convex set.

Matrix fractional function

- The function $f : \mathbf{R}^n \times \mathbf{S}^n \to \mathbf{R}$, defined as

$$f(x, Y) = x^T Y^{-1} x,$$

is called a **matrix fractional function**, and is convex on $\mathbf{dom}\ f = \mathbf{R}^n \times \mathbf{S}^n_{++}$.

- Proof:

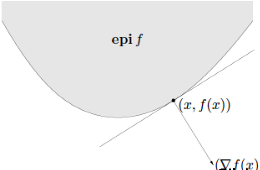
$$\begin{aligned} \textbf{epi}\ f &= \{(x, Y, t) \mid Y \succ 0, x^T Y^{-1} x \le t\} \\ &= \left\{(x, Y, t) \mid Y \succ 0, \begin{bmatrix} Y & x \\ x^T & t \end{bmatrix} \succeq 0\right\} \end{aligned}$$

is a convex set.

Epigraph and first-order condition for convexity

- If $(y, t) \in \textbf{epi } f$, then
$$t \geq f(y) \geq f(x) + \nabla f(x)^T (y - x),$$
implying
$$\begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0.$$

- This means that the hyperplane defined by $(\nabla f(x), -1)$ is a supporting hyperplane to **epi** f at the boundary point $(x, f(x))$.



Jensen's Inequality

- The basic inequality for convex functions

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y),$$

  is called **Jensen's inequality**.

- Jensen's inequality can be extended to more than two points: If f is convex, $x_1, \cdots, x_k \in \textbf{dom } f$, and $\theta_1, \cdots, \theta_k \geq 0$ with $\theta_1 + \cdots + \theta_k = 1$, then

$$f(\theta_1 x_1 + \cdots + \theta_k x_k) \leq \theta_1 f(x_1) + \cdots + \theta_k f(x_k).$$

Jensen's Inequality

- Extension to infinite sum:

$$f\left(\int_S p(x)x dx\right) \leq \int_S f(x)p(x)dx,$$

  with $p(x) \geq 0$ on S, $\int_S p(x)dx = 1, S \subseteq \mathbf{dom}\ f$.
- If x is a random variable such that Prob $(x \in \mathbf{dom}\ f) = 1$, then

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

- Suppose $x \in \mathbf{dom}\ f \subseteq \mathbf{R}^n$, $z \in \mathbf{R}^n, \mathbf{E}(z) = 0$, and assume $\Pr(x + z \in \mathbf{dom}\ f) = 1$. Then we have

$$\mathbf{E}f(x+z) \geq f(x).$$

Inequalities

- Many famous inequalities can be derived by applying Jensen's inequality to some convex functions.
- The arithmetic-geometric mean inequality: $(a+b)/2 \geq \sqrt{ab}$.
- Noting that $-\log x$ is convex, and letting $\theta = 1/2$, we obtain

$$-\log \frac{a+b}{2} \leq \frac{-\log a - \log b}{2},$$

  implying the AM-GM inequality: $\sqrt{ab} \leq \dfrac{a+b}{2}$.
- Further, by taking

$$a = \frac{x_i^2}{\sum_{j=1}^n x_j^2}, b = \frac{y_i^2}{\sum_{j=1}^n y_j^2},$$

  and summing over i, we get the Cauchy's inequality

$$\left(\sum_{j=1}^n x_j y_j\right)^2 \leq \left(\sum_{j=1}^n |x_j||y_j|\right)^2 \leq \left(\sum_{j=1}^n x_j^2\right)\left(\sum_{j=1}^n y_j^2\right).$$

Hölder's Inequality

- Apply the Jensen's inequality to the function $-\log x$ again, with an arbitrary θ, $0 < \theta < 1$, we get an inequality more general than the AM-GM inequality:
$$a^\theta b^{1-\theta} \leq \theta a + (1-\theta)b.$$
- If we take $\theta = 1/p$, where $p > 1$. Let $q = 1/(1-\theta)$, then $q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$.
- By taking
$$a = \frac{|x_i|^p}{\sum_{j=1}^n |x_j|^p}, b = \frac{|y_i|^q}{\sum_{j=1}^n |y_j|^q},$$
and summing over i, we obtain
$$\sum_{j=1}^n |x_j||y_j| \leq \left(\sum_{j=1}^n |x_j|^p\right)^{1/p} \left(\sum_{j=1}^n |y_j|^q\right)^{1/q},$$
which implies the Hölder's inequality
$$\sum_{j=1}^n x_j y_j \leq \left(\sum_{j=1}^n |x_j|^p\right)^{1/p} \left(\sum_{j=1}^n |y_j|^q\right)^{1/q}.$$

Minkowski Inequality

- Using Hölder's inequality, and $\frac{1}{p}+\frac{1}{q}=1$, we have the following:

$$\begin{aligned}
||x+y||_p^p &= \sum_{i=1}^{n} |x_i+y_i|^p \\
&\leq \sum_i |x_i||x_i+y_i|^{p-1} + \sum_i |y_i||x_i+y_i|^{p-1} \\
\text{(by Hölder's ineq.)} \quad &\leq \left(\sum_i |x_i|^p\right)^{1/p} \left(\sum_i |x_i+y_i|^p\right)^{1/q} \\
&\quad + \left(\sum_i |y_i|^p\right)^{1/p} \left(\sum_i |x_i+y_i|^p\right)^{1/q} \\
&= \left(||x||_p + ||y||_p\right)\left(||x+y||_p\right)^{p-1}
\end{aligned}$$

- Therefore, we obtain that $||x+y||_p \leq ||x||_p + ||y||_p$, with $p>1$.
- Note that the inequality above also holds for $p=1$. So, we obtain the Minkowski inequality:

$$||x+y||_p \leq ||x||_p + ||y||_p,$$

with $p \geq 1$, which essentially says that $\forall p \geq 1$, ℓ_p-norm is a norm.

Basic Operations that Preserve Convexity

- If f is convex and $\alpha \geq 0$, then αf is also convex.
- If both f_1 and f_2 are convex, then $f_1 + f_2$ is also convex.
- More generally, if $f_1, ..., f_n$ are convex functions, then any of their "conic combinations",

$$f = w_1 f_1 + \cdots + w_n f_n,$$

 is also convex (with $w_1, ..., w_n \geq 0$). This is also called the **nonnegative weighted sum**.
- Extension: if $f(x, y)$ is convex in x for any $y \in \mathcal{A}$, and $w(y) \geq 0$ for any $y \in \mathcal{A}$, then the function

$$g(x) = \int_{\mathcal{A}} w(y) f(x, y) dy$$

 is convex in x.

Basic Operations that Preserve Convexity

- Suppose $f : \mathbf{R}^n \to \mathbf{R}$, $A \in \mathbf{R}^{n \times m}$, and $b \in \mathbf{R}^n$. Define $g : \mathbf{R}^m \to \mathbf{R}$ by
$$g(x) = f(Ax + b)$$
with $\mathbf{dom}\ g = \left\{ x \,\middle|\, Ax + b \in \mathbf{dom}\ f \right\}$.
- If f is convex, then g is also convex.
- If f is concave, so is g.

Pointwise maximum (1/2)

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

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

  with $\mathbf{dom}\ f = \mathbf{dom}\ f_1 \cap \mathbf{dom}\ f_2$, is also convex.
- Proof:

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \\ &\leq \\ &= \\ &\leq \\ &= \theta f(x) + (1-\theta) f(y). \end{aligned}$$

Pointwise maximum (2/2)

- It can be easily extended: if $f_1, ..., f_m$ are convex, then their pointwise maximum

$$f(x) = \max \{f_1(x), ..., f_m(x)\},$$

is also convex.

Pointwise maximum – Examples

Piecewise-linear functions

A **piecewise-linear** function $f(x) = \max\left\{a_1^T x + b_1, \dots, a_L^T x + b_L\right\}$ is convex, since the affine functions $a_i^T x + b_i$ are all convex.

Sum of r largest components

For $x \in \mathbf{R}^n$, we denote by $x_{[i]}$ the ith largest component of x, i.e.,

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

are the components of x sorted in nonincreasing order $\left(\left\{x_{[1]}, \dots, x_{[n]}\right\} = \{x_1, \dots, x_n\}\right)$. Then the function $f(x) = \sum_{i=1}^{r} x_{[i]}$ is convex.

- Note that, as a generalization, the function $f(x) = \sum_{i=1}^{r} w_i x_{[i]}$ is also convex as long as $w_1 \geq w_2 \geq \dots \geq w_r \geq 0$.

Pointwise supremum

- If for each $y \in \mathcal{A}$, $f(x, y)$ is convex in x, then the function g, defined as
$$g(x) = \sup_{y \in \mathcal{A}} f(x, y)$$
is convex in x. Here
$$\mathbf{dom}\ g = \left\{ x \,\middle|\, (x, y) \in \mathbf{dom}\ f\ \forall y \in \mathcal{A},\ \ \sup_{y \in \mathcal{A}} f(x, y) < \infty \right\}.$$
- Similarly, the pointwise infimum of a set of concave functions is a concave function.

Recall: the **supremum** and **infimum** of a set $\mathcal{A}$ are defined as

$$\sup \mathcal{A} = \min \{y \mid y \geq x, \forall x \in \mathcal{A}\}$$ (i.e., the minimum upper bound of $\mathcal{A}$)

and

$$\inf \mathcal{A} = \max \{y \mid y \leq x, \forall x \in \mathcal{A}\}$$ (i.e., the maximum lower bound of $\mathcal{A}$),

respectively.

Pointwise supremum

- In terms of epigraphs, the pointwise supremum of functions corresponds to the intersection of epigraphs: if

$$g(x) = \sup_{y \in \mathcal{A}} f(x, y),$$

then we have

$$\textbf{epi } g = \bigcap_{y \in \mathcal{A}} \textbf{epi } f(\cdot, y).$$

- Thus, the result follows from the fact that the intersection of a family of convex sets is convex.

Pointwise supremum – Examples (1/3)

Support function of a set

Let $C \subseteq \mathbf{R}^n$ with $C \neq \varnothing$. The **support function** S_C associated with the set C, defined as

$$S_C(x) = \sup \left\{ x^T y \mid y \in C \right\},$$

with $\mathbf{dom}\, S_C = \left\{ x \mid \sup_{y \in C} x^T y < \infty \right\}$, is convex.

Pointwise supremum – Examples (2/3)

Distance to farthest point of a set

Let $C \subseteq \mathbf{R}^n$. The distance (in any norm) to the farthest point of C,

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

is convex.

Pointwise supremum – Examples (3/3)

Maximum eigenvalue of a symmetric matrix

The function $f(X) = \lambda_{max}(X)$, with $\mathbf{dom}\ f = \mathbf{S}^m$, is convex.

Proof:

$$f(X) = \sup \left\{ y^T X y \mid ||y||_2 = 1 \right\}.$$

Norm of a matrix

The function $f(X) = ||X||_2$ with $\mathbf{dom}\ f = \mathbf{R}^{p \times q}$, where $|| \cdot ||_2$ denotes the spectral norm or maximum singular value, is convex.

Proof:

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

is the pointwise supremum of a family of linear functions of X.

Convexity of composition of functions

Convexity of composition of functions

Let $h : \mathbf{R} \to \mathbf{R}$, and $g : \mathbf{R} \to \mathbf{R}$ and $f = h \circ g : \mathbf{R} \to \mathbf{R}$, $f(x) = h(g(x))$. Let $\mathbf{dom}\ f = \mathbf{dom}\ g = \mathbf{dom}\ h = \mathbf{R}$ and f, g, h be differentiable. Then,

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

Proof (for the case where h and g are both twice differentiable):

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x).$$

Examples – Convexity of composition of functions

- 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 $\{x \mid g(x) < 0\}$.

A generalization

Convexity of composition of functions

Let $h : \mathbf{R} \to \mathbf{R}$, and $g : \mathbf{R}^n \to \mathbf{R}$ and $f = h \circ g : \mathbf{R}^n \to \mathbf{R}$, $f(x) = h(g(x))$. Let $\mathbf{dom}\ f = \mathbf{dom}\ g = \mathbf{R}^n, \mathbf{dom}\ h = \mathbf{R}$, and f, g, h be differentiable. Then,

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

Proof idea: convexity is determined by the behavior of a function on arbitrary lines that intersect its domain.

Vector composition – A further generalization

Vector Composition

Suppose $f(x) = h(g(x)) = h(g_1(x), ..., g_k(x))$, with $h : \mathbf{R}^k \to \mathbf{R}, g_i : \mathbf{R}^n \to \mathbf{R}, i = 1, ..., k$. Then,

- f is convex if h is convex, h is n.d. in each argument, and g_i are convex,
- f is convex if h is convex, h is n.i. in each argument, and g_i are concave,
- f is concave if h is concave, h is n.d. in each argument, and g_i are concave.
- f is concave if h is concave, h is n.i. in each argument, and g_i are convex.

Proof: In the case h and g are twice differentaible, w.l.o.g., we can assume $n = 1$.

$$f''(x) = g'(x)^T \nabla^2 h(g(x)) g'(x) + \nabla h(g(x))^T g''(x),$$

Vector composition examples

- Let $h(z) = z_{[1]} + \ldots + z_{[r]}$, the sum of the r largest components of $z \in \mathbf{R}^k$. Then h is convex and nondecreasing in each argument.
- Suppose $g_1, \ldots, g_k$ are convex functions on $\mathbf{R}^n$. Then the composition function $f = h \circ g$, i.e., the pointwise sum of the r largest g_i's, is convex.
- The function $h(z) = \log(\sum_{i=1}^k e^{z_i})$ is convex and nondecreasing in each argument, so $\log(\sum_{i=1}^k e^{g_i})$ is convex whenever g_i are.
- For $0 < p \leq 1$, the function $h(z) = (\sum_{i=1}^k z_i^p)^{1/p}$ on $\mathbf{R}_+^k$ is concave, and its extension (which has the value $-\infty$ for $z \not\succeq 0$) 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.

Vector composition examples

- Suppose $p \geq 1$, and $g_1, \ldots, g_k$ are convex and nonnegative. Then the function $(\sum_{i=1}^k g_i(x)^p)^{1/p}$ is convex.
  - Proof idea: The ℓ_p-norm is convex, and is nondecreasing in each argument if the considered domain is $\mathbf{dom}\ ||\cdot||_p = \mathbf{R}_+^k$.
- The geometric mean $h(z) = (\prod_{i=1}^k z_i)^{1/k}$ on $\mathbf{R}_+^k$ 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,

$$\left(\prod_{i=1}^k g_i\right)^{1/k}.$$

Minimization (1/2)

Minimization and convexity

If f is convex in (x, y), and C is a convex nonempty set, then

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

is convex in x, provided $g(x) > -\infty$ for some x [a], with

$$\mathbf{dom}\ g = \{x \mid (x, y) \in \mathbf{dom}\ f,\ \exists y \in C\}.$$

[a] Since $g(x)$ will be shown to be convex, this actually implies $g(x) > -\infty$ for all x.

Minimization (2/2)

Proof of the convexity of $g(x)$:

- Let $x_1, x_2 \in \textbf{dom}\ g$ and $\epsilon > 0$. Then $\exists y_1, y_2 \in C$ such that $f(x_i, y_i) \leq g(x_i) + \epsilon$ for $i = 1, 2$.
- For any $\theta, 0 \leq \theta \leq 1$, we have

$$\begin{aligned} g(\theta x_1 + (1-\theta)x_2) &= \inf_{y \in C} f(\theta x_1 + (1-\theta)x_2, y) \\ &\leq f(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2) \\ &\leq \theta f(x_1, y_1) + (1-\theta) f(x_2, y_2) \\ &\leq \theta g(x_1) + (1-\theta) g(x_2) + \epsilon \end{aligned}$$

- Since this holds for any $\epsilon > 0$, we conclude that $g(\theta x_1 + (1-\theta)x_2) \leq \theta g(x_1) + (1-\theta) g(x_2)$ for all $x_1, x_2 \in \textbf{dom}\ g$ and for any $\theta \in [0, 1]$.

Example – Distance to a set

- The distance of a point x to a set $S \subseteq \mathbf{R}^n$, in the norm $||\cdot||$, is defined as
$$\textbf{dist}\ (x, S) = \inf_{y \in S} ||x - y||.$$
- The function $||x - y||$ is convex in (x, y), so if the set S is convex, the distance function **dist** (x, S) is a convex function of x.

Example

- Suppose h is convex. Then the function g defined as

$$g(x) = \inf \{h(y) \mid Ay = x\} \quad ^1$$

  is convex.

- Proof: We define f by [2]

$$f(x, y) = \begin{cases} h(y) & \text{if } Ay = x \\ \infty & \text{otherwise} \end{cases}$$

  which is convex in (x, y). Then g is the minimum of f over y, and hence is convex. (It is not hard to show directly that g is convex.)

---

[1] In fact, it can be shown that $g(x) = \min \{h(y) \mid Ay = x\}$.

[2] Note that $\mathbf{dom}\ f = \{(x, y) \mid Ay = x\}$ is convex.

Conjugate functions

Conjugate functions

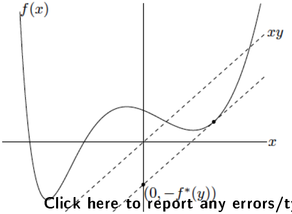
Let $f : \mathbf{R}^n \to \mathbf{R}$. The function $f^* : \mathbf{R}^n \to \mathbf{R}$, defined as

$$f^*(y) = \sup_{x \in \mathbf{dom}\ f} \left(y^T x - f(x) \right),$$

is called the **conjugate** of the function f. The domain of f^* is

$$\mathbf{dom}\ f^* = \left\{ y \in \mathbf{R}^n \mid \exists z \in \mathbf{R} \text{ s.t. } \forall x \in \mathbf{dom}\ f,\ y^T x - f(x) < z \right\}.$$

Example:
$f : \mathbf{R}^1 \to \mathbf{R}$, $f^* : \mathbf{R}^1 \to \mathbf{R}$



Example – Revenue and Profit Functions

- Let $r = (r_1, ..., r_n)$ denote the vector of resource quantities consumed, $S(r)$ denote the sales revenue derived from the product produced, $p = (p_1, ..., p_n)$ denote the vector of unit prices of resources.
- Then the profit is

$$S(r) - p^T r.$$

- Given the price vector p, the maximum profit is given by

$$M(p) = \sup_r \left(S(r) - p^T r \right),$$

or

$$M(p) = (-S)^*(-p).$$

Conjugate functions

- A conjugate function

$$f^*(y) = \sup_{x \in \mathbf{dom}\ f} \left(y^T x - f(x)\right)$$

  is always convex.
  - $\because$ it is the pointwise supremum of a family of convex (indeed, affine) functions of y.
  - This is true whether or not f is convex.
  - Note that when f is convex, the subscript $x \in \mathbf{dom}\ f$ is not necessary since $y^T x - f(x) = -\infty$ for $x \notin \mathbf{dom}\ f$.

Conjugate Functions – Examples for $f : \mathbf{R} \to \mathbf{R}$

- Affine function. $f(x) = ax + b$. The function $yx - ax - b$ is bounded if and only if $y = a$. Therefore $\mathbf{dom}\ f^* = \{a\}$, and $f^*(a) = -b$.
- Negative logarithm. $f(x) = -\log x$, with $\mathbf{dom}\ f = \mathbf{R}_{++}$. The function $xy + \log x$ is unbounded above if $y \geq 0$ and reaches its maximum at $x = -1/y$ otherwise. Therefore, $\mathbf{dom}\ f^* = \{y \mid y < 0\} = -\mathbf{R}_{++}$ and $f^*(y) = -\log(-y) - 1$ for $y < 0$.
- Exponential. $f(x) = e^x$. The function $xy - e^x$ is unbounded above if $y < 0$. It can be shown that $\mathbf{dom}\ f^* = \mathbf{R}_+$ and

$$f^*(y) = \begin{cases} y \log y - y, & y > 0 \\ 0, & y = 0 \end{cases}.$$

Conjugate Functions – Examples for $f : \mathbf{R} \to \mathbf{R}$ (1/2)

- Negative entropy. $f(x) = x \log x$, with $\mathbf{dom}\ f = \mathbf{R}_+$ (and $f(0) = 0$). The function $xy - x \log x$ is bounded above on $\mathbf{R}_+$ for all y, hence $\mathbf{dom}\ f^* = \mathbf{R}$. It attains its maximum at $x = e^{y-1}$, and substituting we find $f^*(y) = e^{y-1}$.
- Inverse. $f(x) = 1/x$ on $\mathbf{R}_{++}$. 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 $x = (-y)^{-1/2}$. Therefore we have $f^*(y) = -2(-y)^{1/2}$, with $\mathbf{dom}\ f^* = -\mathbf{R}_+$.

Conjugate Functions – Examples for $f : \mathbf{R}^n \to \mathbf{R}$ (2/2)

- Strictly convex quadratic function. Consider $f(x) = \frac{1}{2}x^T Qx$, with $Q \in \mathbf{S}^n_{++}$. The function $y^T x - \frac{1}{2}x^T Qx$ is bounded above as a function of x for all y. It attains its maximum at $x = Q^{-1}y$, so
$$f^*(y) = \frac{1}{2}y^T Q^{-1}y.$$
- Log-sum-exp function. Consider
$$f(x) = \log\left(\sum_{i=1}^n e^{x_i}\right).$$
Then, $f^*(y) = \sum_{i=1}^n y_i \log y_i$ with
$$\mathbf{dom}\ f^* = \left\{y \mid \mathbf{1}^T y = 1, y \succeq 0\right\}.$$

Indicator function

- Let I_S be the indicator function of a (not necessarily convex) set $S \subseteq \mathbf{R}^n$, i.e., $I_S(x) = 0$ on $\mathbf{dom}\ I_S = S$.
- Its conjugate is
$$I_S^*(y) = \sup_{x \in S} y^T x,$$
which is the support function of the set S.

Conjugate Functions – Examples for $f : \mathbf{R}^n \to \mathbf{R}$

- Norm. Let $||\cdot||$ be a norm on $\mathbf{R}^n$, with dual norm $||\cdot||_*$. We will show that the conjugate of $f(x) = ||x||$ is

$$f^*(y) = \begin{cases} 0, & ||y||_* \leq 1 \\ \infty, & \text{otherwise} \end{cases},$$

  i.e., the conjugate of a norm is the indicator function[3] of the dual norm unit ball.
- The definition of the dual norm of a given norm is defined in the following pages.

[3]The **indicator function** of a convex set can be defined to have a zero value within the set and infinity outside the set. It will occur again several times in the rest of the course.

Introduction to Dual Norms (1/3)

- Let $||\cdot||$ be a norm on $\mathbf{R}^n$. The associated **dual norm**, denoted $||\cdot||_*$, is defined as

$$||z||_* = \sup\left\{z^T x \mid ||x|| \leq 1\right\}.$$

- It can be shown that

$$||z||_* = \sup\left\{|z^T x| \mid ||x|| \leq 1\right\}$$

and

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

- A dual norm is also a norm.
  - *Hint*: $||u+v||_* = \sup\left\{(u+v)^T x \mid ||x|| \leq 1\right\}$

Introduction to Dual Norms (2/3)

- From the definition of dual norm we have the inequality

$$z^T x \leq ||x|| \; ||z||_*,$$

  for all x and z.
- The dual of the dual norm is the original norm: we have $||x||_{**} = ||x||$ for all x.
  - *Hint*: $||x||_{**} = \sup_{z \neq 0} \frac{x^T z}{||z||_*}$
- The dual of the Euclidean norm is the Euclidean norm, since $\sup \left\{ z^T x \mid ||x||_2 \leq 1 \right\} = ||z||_2$.
  - This follows from the Cauchy-Schwarz inequality.
  - For nonzero z, the value of x that maximizes $z^T x$ over $||x||_2 \leq 1$ is $z/||z||_2$.

Introduction to Dual Norms (3/3)

- The dual of the ℓ_∞-norm is the ℓ_1-norm:

$$\sup\left\{z^T x \;\middle|\; ||x||_\infty \leq 1\right\} = \sum_{i=1}^{n} |z_i| = ||z||_1.$$

- The dual of the ℓ_1-norm is the ℓ_∞-norm.
- More generally, the dual of the ℓ_p-norm is the ℓ_q-norm, where q satisfies

$$\frac{1}{p} + \frac{1}{q} = 1,$$

i.e., $q = p/(p-1)$.
    - *Hint*: Hölder's inequality: $u^T v \leq ||u||_p ||v||_q$.

Conjugate Functions – Examples for $f : \mathbf{R}^n \to \mathbf{R}$

- Come back to the example of the conjugate function of a norm. Let $||\cdot||$ be a norm on $\mathbf{R}^n$, with dual norm $||\cdot||_*$. We now show that the conjugate of $f(x) = ||x||$ is

$$f^*(y) = \begin{cases} 0, & ||y||_* \leq 1 \\ \infty, & \text{otherwise} \end{cases}.$$

- Proof: If $||y||_* > 1$, then by definition of the dual norm, there is a $z \in \mathbf{R}^n$ with $||z|| \leq 1$ and $y^T z > 1$. Taking $x = tz$ and letting $t \to \infty$, we have $y^T x - ||x|| = t(y^T z - ||z||) \to \infty$, which shows that $f^*(y) = \infty$.
- Conversely, if $||y||_* \leq 1$, then we have $y^T x \leq ||x|| ||y||_*$ for all x, which implies for all x, $y^T x - ||x|| \leq 0$. Therefore $x = 0$ is the value that maximizes $y^T x - ||x||$, with maximum value 0.

Conjugate Functions – Examples for $f : \mathbf{S}^n_{++} \to \mathbf{R}$

- Log-determinant. We consider $f(X) = \log \mathbf{det}\ X^{-1}$ on $\mathbf{S}^n_{++}$. The conjugate function is defined as

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

since $\mathbf{tr}\ (YX)$ is the standard inner product on $\mathbf{S}^n$. It can be shown that $\mathbf{dom}\ f^* = -\mathbf{S}^n_{++}$ and

$$f^*(Y) = \log \mathbf{det}\ (-Y)^{-1} - n.$$