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 $S = \{\mathbf{y} | \langle \mathbf{y}, \mathbf{x}\rangle = \|\mathbf{x}\| \text{ and } \|\mathbf{y}\|^* \leq 1\}$.
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 $\mathrm{dom}\, g = \{\mathbf{x} \mid \mathbf{A}\mathbf{x} + \mathbf{b} \in \mathrm{dom}\, f\}$. 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

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

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

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

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

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

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

When $\mathbf{dom}\, h = \mathbb{R}$:

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

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

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

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 $\operatorname{dom} h = \mathbb{R}^k$, in order to discover the composition rules. We have

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

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 $\log(\sum_{i=1}^k e^{g_i(\mathbf{x})})$ 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 $f(\mathbf{x}) = (\sum_{i=1}^k g_i(\mathbf{x})^p)^{1/p}$ 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 $f(\mathbf{x}) = (\sum_{i=1}^k g_i(\mathbf{x})^p)^{1/p}$ 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 $\mathbf{dom}\, h = \mathbb{R}^k$, so $h = \tilde{h}$. This function is convex, and nondecreasing, so we conclude $h(\mathbf{g}(\mathbf{x}))$ is a convex function of $\mathbf{x}$. 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 $(\sum_{i=1}^k g_i(\mathbf{x})^p)^{1/p}$ 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

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

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.

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

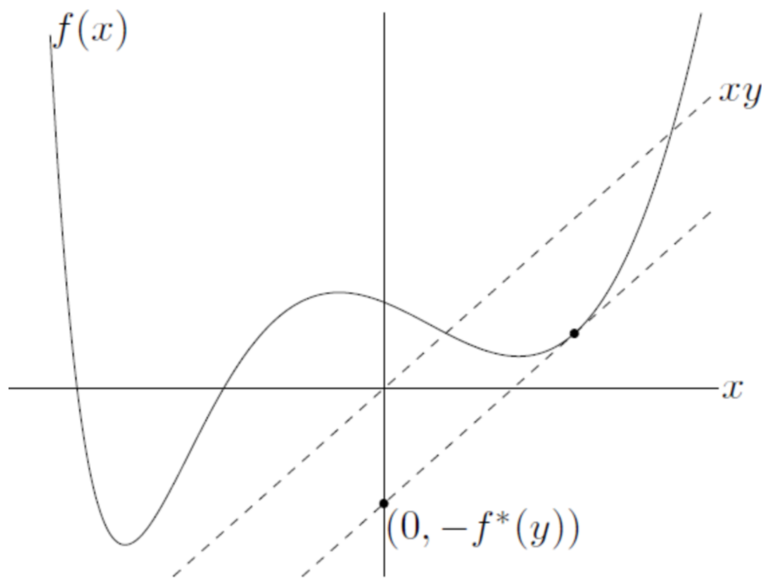
- Definition and examples

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

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

$$f^*(\mathbf{y}) = \sup_{\mathbf{x} \in \operatorname{dom} f} \left(\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x}) \right), \tag{1}$$

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 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 $\operatorname{dom} f^* = \{a\}$ and $f^*(a) = -b$.

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

- *Negative entropy.* $f(x) = x \log x$, with $\operatorname{dom} f = \mathbb{R}_+$. The function $xy - x \log x$ is bounded above on $\mathbb{R}_+$ for all y, hence $\operatorname{dom} f^* = \mathbb{R}$. 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 $\operatorname{dom} f^* = -\mathbb{R}_+$.

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* $\mathbb{S}^n_{++}$.

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

We first show that $\operatorname{tr}(\mathbf{Y}\mathbf{X}) + \log\det\mathbf{X}$ *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*

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

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

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

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

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

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*

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