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 shownthat $\mathcal{T}(\Delta\mathbf{X}) = -\mathbf{X}^{-1}\Delta\mathbf{X}\mathbf{X}^{-1}$.

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

Basic properties and examples

- Sublevels

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

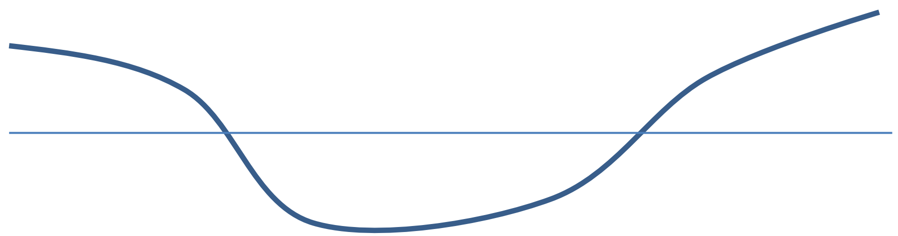
$$C_\alpha = \{\mathbf{x} \in \operatorname{dom} f \mid f(\mathbf{x}) \leq \alpha\}.$$

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-superlevel set of the function $G(\mathbf{x}) - \alpha A(\mathbf{x})$, which is concave. In fact, the set is positively homogeneous, so it is a convex cone.

Basic properties and examples

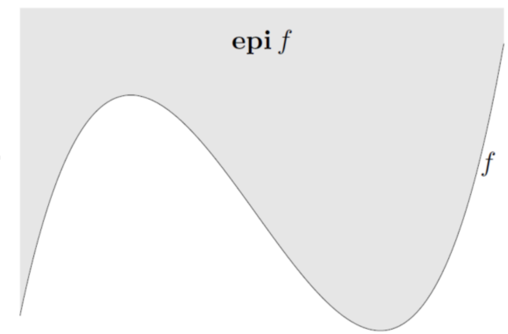
- Epigraph

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

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

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:

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

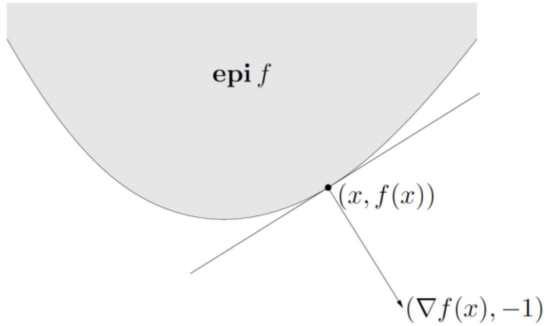
The last condition is a linear matrix inequality in $(\mathbf{x}, \mathbf{Y}, t)$, and therefore $\operatorname{epi} f$ is convex.

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

- 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

- Jensen's inequality and extensions

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

$$\Updownarrow$$

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

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

$$\Updownarrow$$

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

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

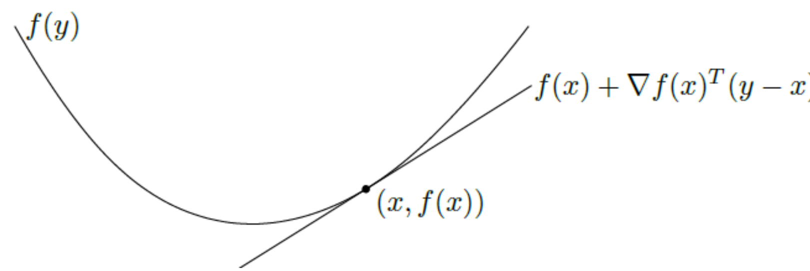
- Bregman distance

Given a differentiable strictly convex function $f : C \to \mathbb{R}$, where $C \subset \mathbb{R}^n$ is a convex set, the Bregman distance is defined as:

$$B_f(\mathbf{y}, \mathbf{x}) = f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle. \tag{1}$$

It it clear that $B_f(\mathbf{y}, \mathbf{x}) \geq 0$ for all $\mathbf{x}, \mathbf{y} \in C$ due to the convexity of f. However, the Bregman distance may not be symmetric: $B_f(\mathbf{y}, \mathbf{x}) \neq B_f(\mathbf{x}, \mathbf{y})$.

Examples:



- $f(\mathbf{x}) = \|\mathbf{x}\|^2$.
- $f(\mathbf{x}) = \mathbf{x}^T \mathbf{P} \mathbf{x}$.
- $f(\mathbf{x}) = \sum_i x_i \log x_i$.

Basic properties and examples

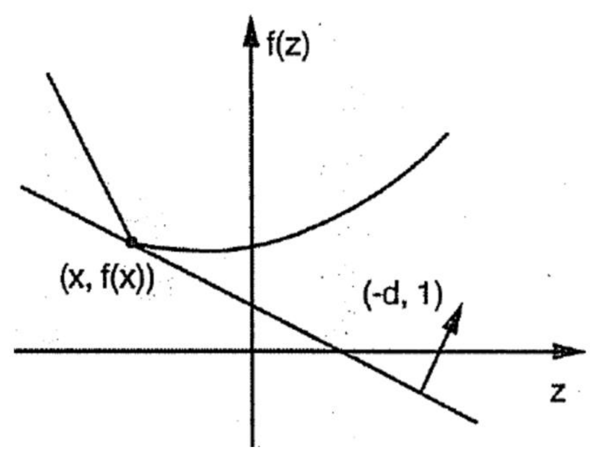
- Subgradient

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

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?



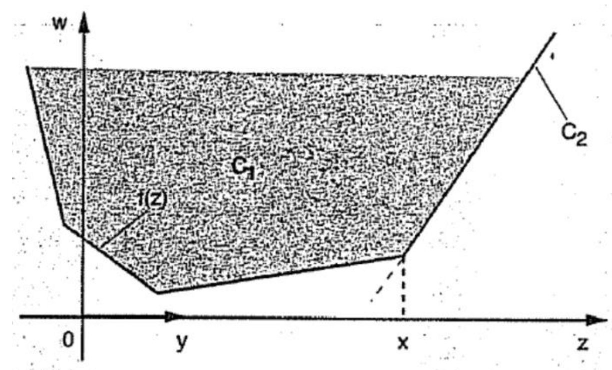
Basic properties and examples

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

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

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

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

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

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.