Chapter III: Convex Functions

1 Convex functions

Definition 1. Let $S \subseteq \mathbb{R}^n$. A function $f : S \to \mathbb{R}$ is *convex* if

- $\mathbf{dom}\, f = S$ is convex;
- $\forall \boldsymbol{x}, \boldsymbol{y} \in S,\ \forall \alpha \in [0,1],\ f((1-\alpha)\boldsymbol{x} + \alpha \boldsymbol{y}) \leq (1-\alpha) f(\boldsymbol{x}) + \alpha f(\boldsymbol{y})$.

Moreover, f is called *strictly* convex if the above inequality holds strictly for all $\boldsymbol{x} \neq \boldsymbol{y} \in S,\ \alpha \in (0,1)$.

The function f is (strictly) *concave* if $-f$ is (strictly) convex.

Note: If we define $f(\boldsymbol{x}) := +\infty$ whenever $\boldsymbol{x} \notin \mathbf{dom}\, f$, then f is convex iff the inequality

$$f((1-\alpha)\boldsymbol{x} + \alpha \boldsymbol{y}) \leq (1-\alpha) f(\boldsymbol{x}) + \alpha f(\boldsymbol{y}) \tag{1}$$

holds for all $\alpha \in [0,1]$ and for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$: The inequality trivially holds when one of $\boldsymbol{x}, \boldsymbol{y}$ is outside $\mathbf{dom}\, f$, and the fact that the inequality holds for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{dom}\, f$ implies the convexity of $\mathbf{dom}\, f$. Hence, we can always assume that f takes values in $\mathbb{R} \cup \{\infty\}$, and that $\mathbf{dom}\, f = \{\boldsymbol{x} : f(\boldsymbol{x}) < \infty\}$. (Or, in problems involving concavity, that f takes values in $\mathbb{R} \cup \{-\infty\}$, with $\mathbf{dom}\, f = \{\boldsymbol{x} : f(\boldsymbol{x}) > -\infty\}$.)

The next result shows that a function is convex iff its restriction to any line is convex. This can be very useful when we want to prove that a function is convex, since it reduces to show that some functions of one variable are convex.

Proposition 1. *Let $f : \mathbf{dom}\, f \to \mathbb{R}$ be a function with $\mathbf{dom}\, f \subseteq \mathbb{R}^n$. Then, f is convex if and only if its restriction to any line is convex. More precisely, if for all $\boldsymbol{x}_0 \in \mathbf{dom}\, f, \boldsymbol{u} \in \mathbb{R}^n$, the function*

$$g : t \mapsto f(\boldsymbol{x}_0 + t\boldsymbol{u})$$

is convex over $\mathbf{dom}\, g := \{t \in \mathbb{R} : \boldsymbol{x}_0 + t\boldsymbol{u} \in \mathbf{dom}\, f\}$.

Proof. If f is convex, then $\mathbf{dom}\, f$ is convex, so we can see that $\mathbf{dom}\, g$ is an interval (i.e., a convex subset of $\mathbb{R}$), and the convexity inequality trivially holds for g. Conversely, assume that g is convex for any $\boldsymbol{x}_0 \in \mathbb{R}^n$ and $\boldsymbol{u} \in \mathbb{R}^n$. Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{dom}\, f$ and $\alpha \in [0,1]$. We can set $\boldsymbol{x}_0 = \boldsymbol{x}, \boldsymbol{u} = \boldsymbol{y} - \boldsymbol{x}$, so that $g(0) = f(\boldsymbol{x})$, $g(1) = f(\boldsymbol{y})$, and $\boldsymbol{x}_0 + \alpha \boldsymbol{u} = (1-\alpha)\boldsymbol{x} + \alpha \boldsymbol{y}$. The convexity of $\mathbf{dom}\, g$ shows that g is well defined over the whole interval $[0,1]$, so $(1-\alpha)\boldsymbol{x} + \alpha \boldsymbol{y} \in \mathbf{dom}\, f$; this already shows that $\mathbf{dom}\, f$ is convex. Moreover, we have $g(\alpha) = g((1-\alpha)\cdot 0 + \alpha \cdot 1) \leq (1-\alpha) g(0) + \alpha g(1)$, which shows that the inequality (1) holds. □

Definition 2. The $\alpha-$*sublevel* set of a real-valued function f is

$$C_\alpha(f) = \{\boldsymbol{x} \in \mathbf{dom}\, f : \; f(\boldsymbol{x}) \leq \alpha\}.$$

The $\alpha-$superlevel set of a real-valued function f is

$$C^\alpha(f) = \{\boldsymbol{x} \in \mathbf{dom}\, f : \; f(\boldsymbol{x}) \geq \alpha\}.$$

The *epigraph* of a function $f : \mathbf{dom}\, f \to \mathbb{R}$, with $\mathbf{dom}\, f \subseteq \mathbb{R}^n$, is the set

$$\mathbf{epi}\, f = \{(\boldsymbol{x}, t) \in \mathbb{R}^n \times \mathbb{R} : \; f(\boldsymbol{x}) \leq t\} \subseteq \mathbb{R}^{n+1}.$$

The *hypograph* of a function $f : \mathbf{dom}\, f \to \mathbb{R}$, with $\mathbf{dom}\, f \subseteq \mathbb{R}^n$, is the set

$$\mathbf{hypo}\, f = \{(\boldsymbol{x}, t) \in \mathbb{R}^n \times \mathbb{R} : \; f(\boldsymbol{x}) \geq t\} \subseteq \mathbb{R}^{n+1}.$$

The following proposition relates the convexity of f to the convexity of its sublevel sets and epigraph.

Proposition 2. *If f is convex (concave), then its α-sublevel sets (superlevel sets) are convex. The function f is convex (concave) iff its epigraph (hypograph) is convex.*

Example:
The converse of the first implication in Proposition 2 is not true. Indeed, consider the function of one variable $f(x) = 1 - e^{-x^2}$. Each sublevel set of this function is an interval, hence convex, but the function f is not convex.
Another example is the function of two variables $f(x, y) = xy$ over $\mathbf{dom}\, f := \mathbb{R}^2_+$. The superlevel sets of f are of the form $C^\alpha(f) = \mathbf{cl}\,\{x, y > 0 : \; y \geq \alpha/x\}$, which are convex sets, although f is not jointly concave in x and y: Indeed, $f(x, x) = x^2$, so f is not concave on the line $y = x$. #1

Remark: A function whose all sublevel (superlevel) sets are convex is called *quasi-convex* (*quasi-concave*).

2 Jensen's inequality

It is possible to extend the convexity inequality (1) to convex combinations of more than two points, or even to an infinite number of points (integrals). The proof works by induction:

Theorem 3. *Jensen's inequality Let f be convex and $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \in \mathbf{dom}\, f$. Then, for all $\boldsymbol{\lambda} \in \mathbb{R}^n_+$ with $\mathbf{1}^T\boldsymbol{\lambda} = 1$,*

$$f\left(\sum_{i=1}^n \lambda_i \boldsymbol{x}_i\right) \leq \sum_{i=1}^n \lambda_i f(\boldsymbol{x}_i).$$

More generally, let $\boldsymbol{X}$ be an integrable random variable with support in $\mathbf{dom}\, f$, i.e., $\mathbb{P}[\boldsymbol{X} \in \mathbf{dom}\, f] = 1$. Then,

$$f(\mathbb{E}[\boldsymbol{X}]) \leq \mathbb{E}[f(\boldsymbol{X})].$$

3 First and second order convexity conditions

Theorem 4 (First order conditions)**.** *Let f be differentiable at all points of its domain (in particular, this implies that $\mathbf{dom}\, f$ is open), and assume that $\mathbf{dom}\, f \subseteq \mathbb{R}^n$ is convex. Then, f is convex iff*

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

This theorem is important because it shows that when f is convex, we can obtain bounds on $f(\boldsymbol{y})$ by using only "the local information available at $\boldsymbol{x}$" (the value of f and its gradient).

Theorem 5 (Second order conditions)**.** *Let f be twice differentiable at all points of its domain (in particular, this implies that $\mathbf{dom}\, f$ is open), and assume that $\mathbf{dom}\, f \subseteq \mathbb{R}^n$ is convex. Then, f is convex iff the matrix $\nabla^2 f(\boldsymbol{x})$ is positive semidefinite, for all $\boldsymbol{x} \in \mathbf{dom}\, f$.*

If $\nabla^2 f(\boldsymbol{x}) \succ 0$ (that is, the matrix $\nabla^2 f(\boldsymbol{x})$ is positive definite) for all points $\boldsymbol{x}$ of its domain, then f is strictly convex, but the converse is not necessarily true.

We skip the proofs of the theorems of this section. One possibility is to prove them for functions of a single variable first, and then the general result follows by applying Proposition 1. For a counter-example showing that strictly convex functions do not necessarily satisfy $\nabla^2 f(\boldsymbol{x}) \succ 0$, consider the strictly convex function $f : x \mapsto x^4$, whose second derivative is $f''(x) = 12x^2$ and vanishes at $x = 0$.

4 Examples of convex functions

- Every affine function is both convex and concave;
- $x \mapsto e^{ax}$ is convex on $\mathbb{R}$, for all $a \in \mathbb{R}$;
- $x \mapsto x^a$ is convex on $\mathbb{R}_+$, for all $a \geq 1$; concave on $\mathbb{R}_+$ for all $a \in (0, 1]$; convex on $\mathbb{R}_{++}$ for all $a \leq 0$.
- $x \mapsto \log(x)$ is concave over $\mathbb{R}_{++}$.
- $x \mapsto x \log(x)$ is convex over $\mathbb{R}_+$ (with $0 \log 0 := 0$).
- $\boldsymbol{x} \mapsto \|\boldsymbol{x}\|$ is convex over $\mathbb{R}^n$ (for *ANY* norm !); The squared norm $\boldsymbol{x} \mapsto \|\boldsymbol{x}\|^2$ is also convex;
- $\boldsymbol{x} \mapsto \max(x_1, \ldots, x_n)$ is convex over $\mathbb{R}^n$;
- The log-sum-exp function $\boldsymbol{x} \mapsto \log\left(e^{x_1} + \ldots + e^{x_n}\right)$ is convex over $\mathbb{R}^n$. (This function is often used as a smooth approximation of the above max −function);
- The quadratic function $\boldsymbol{x} \mapsto \boldsymbol{x}^T Q \boldsymbol{x} + \boldsymbol{a}^T \boldsymbol{x} + b$ is convex over $\mathbb{R}^n$ if and only if Q is positive semidefinite;
- The geometric mean $\boldsymbol{x} \mapsto \left(\prod_{i=1}^n x_i\right)^{1/n}$ is concave over $\mathbb{R}^n_+$.
- The log-det function $X \mapsto \log \det X$ is concave over $\mathbb{S}^n_{++}$.

It is easy to show the convexity (concavity) of the functions of one variable in the example. For all other functions, we will prove their convexity or concavity later in this course. For now we only show the concavity of the logdet function. By restriction to lines, it suffices to show that the function $g : t \mapsto \log\det(Z + tV)$ is concave for any $Z \succ 0$ and any $V \in \mathbb{S}^n$. We have:

$$\begin{aligned} g(t) = \log\det(Z + tV) &= \log\det\left(Z^{1/2}(I + tZ^{-1/2}VZ^{-1/2})Z^{1/2}\right) \\ &= \log\det Z + \log\det(I + tZ^{-1/2}VZ^{-1/2}) \\ &= \log\det Z + \sum_{i=1}^{n} \log(1 + t\lambda_i). \end{aligned}$$

Then, we take the second derivative of g: $g''(t) = -\sum_i \frac{\lambda_i^2}{(1+t\lambda_i)^2} \leq 0$, so g is concave.

5 Operations that preserve convexity

We have learned some kind of systematic way to establish the convexity of a function f: prove that its Hessian matrix $\nabla^2 f$ is positive semidefinite. In the practice however, this is often not the easiest way to proceed. Instead, we can show that f can be obtained by combining some *blocks* that are already known to be convex (or concave).

- *Nonnegative scaling*: if f is convex, then αf is convex for all $\alpha \geq 0$;
- *Sum*: if f_1 and f_2 are convex, then $f_1 + f_2$ is convex ;
- *Composition with affine maping*: if f is convex, then $\boldsymbol{x} \mapsto f(A\boldsymbol{x} + b)$ is convex;
- *Pointwise maximum*: if $f_1, \ldots, f_n$ are convex, then

  $$\boldsymbol{x} \mapsto \max(f_1(\boldsymbol{x}), \ldots, f_n(\boldsymbol{x}))$$

  is convex over $\mathbf{dom}\, f = \bigcap_{i=1}^n \mathbf{dom}\, f_i$. This extends to *infinite supremums*: let $f : X \times Y \to \mathbb{R}$, and assume that $\boldsymbol{x} \mapsto f(\boldsymbol{x}, \boldsymbol{y})$ is convex over X, for all $\boldsymbol{y} \in Y$. Then,

  $$\boldsymbol{x} \mapsto \sup_{\boldsymbol{y} \in Y} f(\boldsymbol{x}, \boldsymbol{y})$$

  is convex (over its natural domain, which consists of all $\boldsymbol{x}$'s such that the supremum is $< \infty$).
- *Minimization*: if $f : \mathbb{R}^{n+m} \to \mathbb{R}$ is convex on $\mathbf{dom}\, f$ (i.e., $f(\boldsymbol{x}, \boldsymbol{y})$ is *jointly* convex in $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^m$). Then,

  $$g : \boldsymbol{x} \mapsto \inf_{\boldsymbol{y} \in \mathbb{R}^m} f(\boldsymbol{x}, \boldsymbol{y})$$

  is convex over its natural domain, $\mathbf{dom}\, g = \{\boldsymbol{x} \in \mathbb{R}^n |\ \exists \boldsymbol{y} \in \mathbb{R}^m : (\boldsymbol{x}, \boldsymbol{y}) \in \mathbf{dom}\, f\}$;
- *Perspective of a function*: If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then

  $$g : (\boldsymbol{x}, t) \mapsto t\, f\left(\frac{\boldsymbol{x}}{t}\right)$$

  is convex over $\mathbf{dom}\, g = \mathbf{dom}\, f \times \mathbb{R}_{++}$ (Note that g is a function of $n + 1$ variables);
- *Composition rules*: Let $h : \mathbb{R}^k \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^k$, and define $f = h \circ g$.
  - (i) If h is convex, h is nondecreasing in each argument, and g_i is convex ($\forall i \in [k]$), then f is convex.
  - (ii) If h is convex, h is nonincreasing in each argument, and g_i is concave ($\forall i \in [k]$), then f is convex.
  - (iii) If h is concave, h is nondecreasing in each argument, and g_i is concave ($\forall i \in [k]$), then f is concave.
  - (iv) If h is concave, h is nonincreasing in each argument, and g_i is convex ($\forall i \in [k]$), then f is concave.

To recall these rules, we can simply study the case when $k = n = 1$ and h, g are twice differentiable:

$$f'(x) = g'(x)h'(g(x));$$

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

Then, the case (i) follows from $g'' \geq 0, h' \geq 0, h'' \geq 0 \implies f'' \geq 0$.

Example:

- $f(\boldsymbol{x}) = -\sum_{i=1}^m \log(b_i - \boldsymbol{a}_i^T \boldsymbol{x})$ is convex over the polyhedron $\mathbf{dom}\, f := \{\boldsymbol{x} : A\boldsymbol{x} \leq \boldsymbol{b}\}$, where $A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m]^T \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. Indeed, $-\log$ is convex, so $\boldsymbol{x} \mapsto -\log(b_i - \boldsymbol{a}_i^T \boldsymbol{x})$ is convex (composition with affine mapping), and finally f is a sum of convex functions, hence it is convex.
- $f(X) = \lambda_{\max}(X)$ is convex over $\mathbb{S}^n$. Indeed, we have the variational characterization

$$\lambda_{\max}(X) = \sup_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^T X \boldsymbol{v},$$

  which is a pointwise supremum of linear functions of the form $X \mapsto \langle X, \boldsymbol{v}\boldsymbol{v}^T \rangle$;
- $f : (\boldsymbol{x}, t) \mapsto \frac{\|\boldsymbol{x}\|^2}{t}$ is convex over $\mathbb{R}^{n+1}$, because this is the perspective of the squared norm function: #2

$$f(\boldsymbol{x}, t) = t \|\frac{\boldsymbol{x}}{t}\|^2.$$

- The function $g : \boldsymbol{x} \mapsto \mathrm{dist}(\boldsymbol{x}, S) := \inf_{\boldsymbol{y} \in S} \|\boldsymbol{x} - \boldsymbol{y}\|$ is convex if S is convex. To see this, we first observe that $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \|\boldsymbol{x} - \boldsymbol{y}\|$ is *jointly convex* in $\boldsymbol{x}$ and $\boldsymbol{y}$, since it is the composition of the norm with the affine map $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{x} - \boldsymbol{y}$. Then, we can apply the *minimization rule* to the extended-value function

$$f : (\boldsymbol{x}, \boldsymbol{y}) \mapsto \begin{cases} \|\boldsymbol{x} - \boldsymbol{y}\| & \text{if } \boldsymbol{y} \in S; \\ +\infty & \text{otherwise,} \end{cases}$$

  which is convex if S is convex (or equivalently, we can redefine f as its restriction to $\mathbf{dom}\, f = \mathbb{R}^n \times S$, and we observe that S convex $\implies$ $\mathbf{dom}\, f$ convex).

6 Conjugate function

Definition 3. Let $f : \mathbb{R}^n \to \mathbb{R}$. The *conjugate function* of f, also known as *Fenchel conjugate*, is

$$f^* : \boldsymbol{y} \mapsto \sup_{\boldsymbol{x} \in \mathbf{dom}\, f} \langle \boldsymbol{y}, \boldsymbol{x} \rangle - f(\boldsymbol{x}).$$

f^* is implicitly defined with values in $\mathbb{R} \cup \{\infty\}$, so we have $\mathbf{dom}\, f^* := \{\boldsymbol{y} \in \mathbb{R}^n : f^*(\boldsymbol{y}) < \infty\}$.

The next proposition gives a few properties of conjugate functions:

Proposition 6. *Let $f : \mathbb{R}^n \to \mathbb{R}$. Then,*

- *f^* is convex (*even if f is not*).*
- *If f is convex and the epigraph of f is closed, then $f = f^{**}$.*
- Fenchel-Young inequality*: $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n,\ \langle \boldsymbol{x}, \boldsymbol{y} \rangle \leq f(\boldsymbol{x}) + f^*(\boldsymbol{y})$*
- *If f is differentiable, and $\boldsymbol{x}^*$ solves the equation $\boldsymbol{y} = \nabla f(\boldsymbol{x}^*)$, then $f^*(\boldsymbol{y}) = \boldsymbol{x}^{*T} \nabla f(\boldsymbol{x}^*) - f(\boldsymbol{x}^*)$.*

We finish this section with a few examples of conjugate functions:

Example:

- Let $f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x} + b$ be an affine function. The function $\langle \boldsymbol{y}, \boldsymbol{x} \rangle - f(\boldsymbol{x}) = \langle \boldsymbol{y} - \boldsymbol{a}, \boldsymbol{x} \rangle - \boldsymbol{b}$ is unbounded over $\mathbb{R}^n$, unless $\boldsymbol{y} = \boldsymbol{a}$. Hence,

$$\mathbf{dom}\, f^* = \{\boldsymbol{a}\}, \text{ with } f^*(\boldsymbol{a}) = -\boldsymbol{b}.$$

#3

- Let f be the strictly quadratic function $\boldsymbol{x} \mapsto \frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x}$, where $Q \succ 0$ (Q is positive definite). Then, $\boldsymbol{x} \mapsto \boldsymbol{x}^T\boldsymbol{y} - \frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x}$ is minimized over $\boldsymbol{x} \in \mathbb{R}^n$ for $\boldsymbol{x} = Q^{-1}\boldsymbol{y}$ ($\forall \boldsymbol{y} \in \mathbb{R}^n$). Hence,

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