Convex Sets and Functions

1 Convex sets

Of fundamental importance to our course will be the notion of convexity, applied to sets and functions. In particular, we will see that optimization algorithms will have global convergence guarantees when the minimization objective is a convex function, and the constraints form a convex set. We start with the definition of a convex set. Intuitively, we think of an object in the Euclidean space as convex if it has no holes or dents: for example, a circle is convex, while a crescent is not.

Definition. *A set $C \subseteq \mathbb{R}^n$ is* convex *if for every $\mathbf{u}, \mathbf{v} \in C$, and every $\lambda \in [0, 1]$, $(1-\lambda)\mathbf{u} + \lambda\mathbf{v} \in A$.*

The combination $(1-\lambda)\mathbf{u}+\lambda\mathbf{v}$ with $\lambda \in [0, 1]$ is called a *convex combination* of $\mathbf{u}$ and $\mathbf{v}$ (more generally, a convex combination of m vectors $\{\mathbf{u}_1, \dots, \mathbf{u}_m \in \mathbb{R}^n\}$ is the linear combination $\lambda_1\mathbf{u}_1 + \cdots + \lambda_m\mathbf{u}_m$ with non-negative weights summing up to 1: $\lambda_i \geq 0$, $\lambda_1 + \cdots + \lambda_m = 1$). In these terms, we can say that a convex set is the closure of all convex combinations of its points. Geometrically, the set is convex if the straight line segments connecting every pair of points lie within the set[1].

The following properties of convex sets can be shown easily:

Property. *Let $C_1, \dots, C_m$ be convex sets in $\mathbb{R}^n$. Then, the following sets are convex:*

1. $\cap_i C_i$ *(intersection of convex sets is convex)*
2. $\sum_i C_i = \{\mathbf{u}_1 + \cdots + \mathbf{u}_m : \mathbf{u}_1 \in C_1, \dots, \mathbf{u}_m \in C_m\}$ *(sum of convex sets is convex)*
3. $\mathbf{A}C + \mathbf{b} = \{\mathbf{A}\mathbf{u} + \mathbf{b} : \mathbf{u} \in C\}$ *for convex* $C \subseteq \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m\times n}$, *and* $\mathbf{b} \in \mathbb{R}^m$ *is convex (affine transformation of a convex set is convex)*

As an example, we will prove the first property leaving the rest to an exercise. Let $\mathbf{u}, \mathbf{v} \in C_1 \cap \cdots \cap C_m$, and let us fix some $\lambda \in [0, 1]$. From convexity of each of the sets, we have for each $i = 1, \dots, m$, $(1-\lambda)\mathbf{u} + \lambda\mathbf{v} \in C_i$. Consequently, $(1-\lambda)\mathbf{u} + \lambda\mathbf{v} \in C_1 \cap \cdots \cap C_m$. Since the statement is true for every two vectors $\mathbf{u}, \mathbf{v}$ in the intersection and for every $\lambda \in [0, 1]$, the set $C_1 \cap \cdots \cap C_m$ is convex.

[1]With this interpretation, the notion of convexity can be generalized to non-Euclidean (curved) domains, where shortest paths are no longer “straight”.

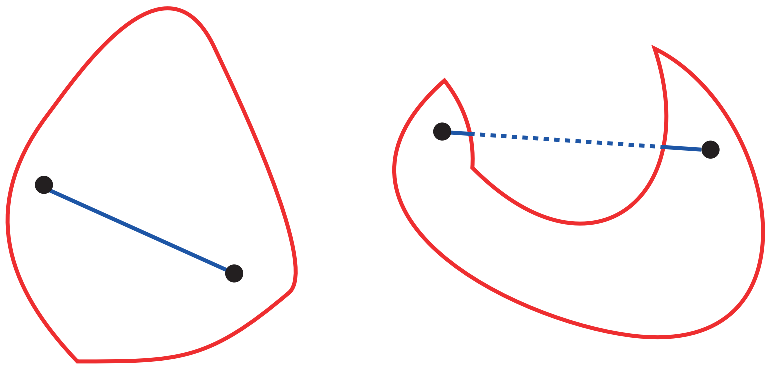

Figure 1: Example of a convex (left) and non-convex (right) sets in $\mathbb{R}^2$. A convex set contains the straight line segments connecting every pair of points in the set.

Exercise 1. *Prove the rest of the properties.*

Given a (non necessarily convex) set A, we can "convexify" it by adding to it all the convex combinations of the points. This leads to the following definition:

Definition. *The* convex hull *of a set $A \subseteq \mathbb{R}^n$, denoted as* $\text{conv}(\mathbf{A})$, *is the set of all convex combinations of the points in A.*

For example, the convex hull of a collection of points $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$, is

$$\text{conv}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\} = \left\{ \sum_{i=1}^{m} \lambda_i \mathbf{a}_i : \lambda_i \geq 0, \sum_{i=1}^{m} \lambda_i = 1 \right\}.$$

Sets of this form are usually called *convex polytopes* and are a multidimensional generalization of the notions of convex polygons (in $\mathbb{R}^2$) and polyhedrons (in $\mathbb{R}^3$). We will encounter convex polytopes in linear programming.

Convex hull can be equivalently defined as

1. The (unique) minimal convex set containing A.

2. The intersection of all convex sets containing A.

Exercise 2. *Prove that the above two definitions of the convex hull are equivalent.*

2 Convex functions

The notion of a convex set leads to the definition of a convex function.

Definition. *A function* $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ *with a convex domain* C *is* convex *if for every* $\mathbf{x} \neq \mathbf{y}$ *and* $0 < \lambda < 1$, $f((1-\lambda)\mathbf{x} + \lambda\mathbf{y}) \leq (1-\lambda)f(\mathbf{x}) + \lambda f(\mathbf{y})$. *The function is* strictly convex *if the latter inequality is sharp.*

A (strictly) *concave* function is a function f such that $-f$ is (strictly) convex. Note: there is no such a thing as a "concave set".

Geometrically, the convex combination $(1-\lambda)\mathbf{x} + \lambda\mathbf{y}$ of the arguments in $\mathbb{R}^n$ is a point on a line segment connecting $\mathbf{x}$ and $\mathbf{y}$, while the convex combination $(1-\lambda)f(\mathbf{x}) + \lambda f(\mathbf{y})$ of the values in $\mathbb{R}$ is a point on the line segment connecting $f(\mathbf{x})$ and $f(\mathbf{y})$. The inequality can be therefore interpreted by saying that the graph of a convex function lies below the straight line segment connecting any pair of points on it. This interpretation allows an alternative and somewhat more elegant definition of convex functions entirely through convex sets. We first define the *epigraph* of a function as the set of all points lying above the graph of f,

$$\mathrm{epi}(f) = \{(\mathbf{x}, y) : \mathbf{x} \in C, y \geq f(\mathbf{x})\}.$$

Note that the epigraph is a set in $\mathbb{R}^{n+1}$ (i.e., has one more dimension; more formally, it is a set in $C \times \mathbb{R}$). With this notion, we can give the following alternative definition of a convex function:

Property. f *is a convex function function iff* $\mathrm{epi}(f)$ *is a convex set.*

Exercise 3. *Prove equivalence of the two definitions.*

Note that in our first definition of a convex function, the convexity of the domain is essential (it is automatically included in the second definition, as an epigraph over a non-convex domain cannot be convex – show it!). Forgetting about this condition is a common mistake. However, it is frequently inconvenient to mention the function domain all the times, especially when working with compositions of multiple functions. It is therefore convenient to extend the function to entire $\mathbb{R}^n$ in such a way that a convex function remains convex, while a non-convex function remains non-convex. We define the *extended real line* as $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. Given a function $f : C \in \mathbb{R}^n \to \mathbb{R}$, we define its extension $\overline{f} : \mathbb{R}^n \to \overline{\mathbb{R}}$ as

$$\overline{f}(\mathbf{x}) = \begin{cases} f(\mathbf{x}) & : \mathbf{x} \in C \\ \infty & : \mathbf{x} \notin C. \end{cases}$$

Substituting $\overline{f}$ to the inequalities defining convexity, it is easy to verify their validity. With some abuse of notation, we will continue denoting the function range of $\mathbb{R}$, sometimes intending $\overline{\mathbb{R}}$.

Exercise 4. *Show that* $\overline{f}$ *is convex iff* f *is convex.*

3 Properties of convex functions

The following properties of convex functions are easy to show invoking the definition:

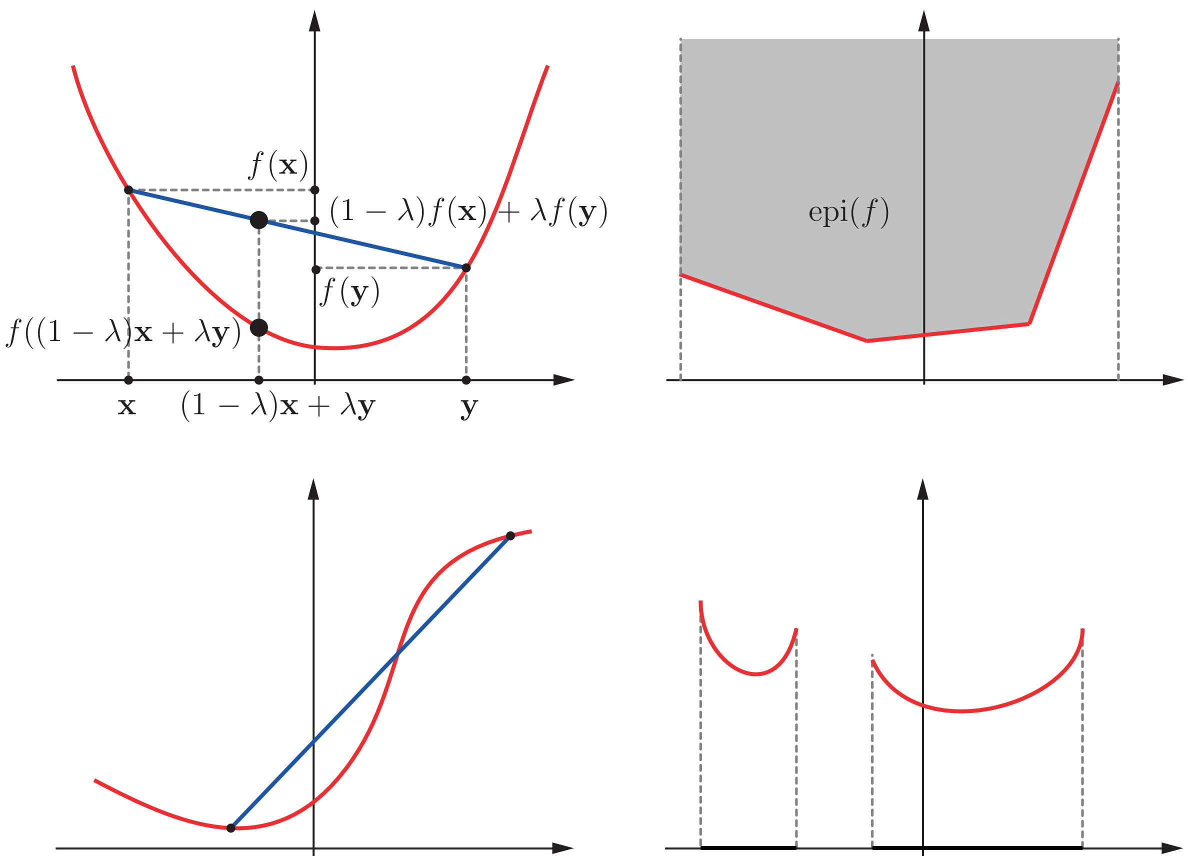

Figure 2: Example of convex (first row) and non-convex (second row) functions. The graph of a convex function must lie below the straight line segment connecting any two points on it. Alternatively the epigraph (the shaded region above the graph) has to be a convex set.

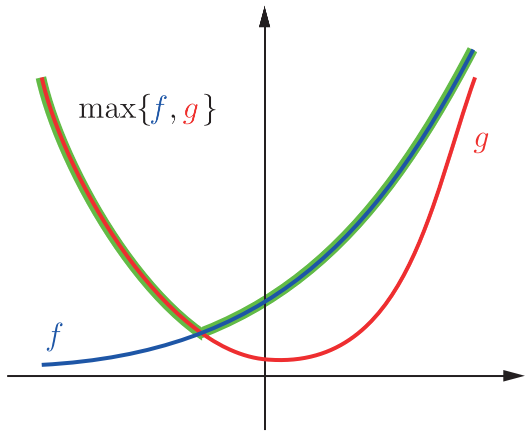

Figure 3: Point-wise maximum of convex functions is a convex function

Property. *The following holds for convex* $f, g : \mathbb{R}^n \to \mathbb{R}$*:*

1. f *is convex iff* $\varphi(\alpha) = f(\mathbf{x} + \alpha\mathbf{r})$ *is convex for every* $\mathbf{x}, \mathbf{r} \in \mathbb{R}^n$ *(convexity is equivalent to convexity on all one-dimensional sections)*

2. $\alpha f + \beta g$ *is convex for every* $\alpha, \beta \geq 0$ *(a non-negative linear combination of convex functions is convex)*

3. $\max\{f(\mathbf{x}), g(\mathbf{x})\}$ *is convex (point-wise maximum of convex functions is convex, as illustrated in Figure 3)*

4. *If* $h : \mathbb{R} \to \mathbb{R}$ *is convex and monotonically increasing, then* $h(f(\mathbf{x}))$ *is convex (transformation of a convex function by a convex increasing function is convex)*

5. *The* level sets *of* f, $\{\mathbf{x} : f(\mathbf{x}) \leq a\}$ *are convex sets for every* a.

Note that the non-negativity of the combination in Property 2 as well as the monotonicity of h in Property 4 are required to exclude results such as $-f$, which is clearly non-convex. The property of convex level sets will be of significant importance in our course.

Exercise 5. *Prove the above properties.*

4 Jensen's inequality

One of the most important inequalities associated with convex functions is Jensen's inequality:

Theorem (Jensen's inequality)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function, $\mathbf{x}_1, \ldots \mathbf{x}_m \in \mathbb{R}^m$, and $\lambda_1, \ldots, \lambda_m \geq 0$ such that $\lambda_1 + \cdots + \lambda_m = 1$. Then,*

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

Jensen's inequality is a generalization of the one we used in the definition of a convex set. Geometrically, it says that the graph of a convex function lies below the polytope formed by any m points on it. The inequality can be further generalized to infinite sums and even integrals.

Exercise 6. *Use Jensen's inequality to show the celebrate relation between the arithmetic and the geometric mean,*

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geq \left(\prod_{i=1}^{n} x_i\right)^{1/n}.$$

Hint: the function $-\log x$ is convex.

5 Derivatives of convex functions

Differentiable convex functions manifest important local properties (i.e., properties expressed in terms of their derivatives). We start with the one-dimensional case, stating the following well-known property:

Property. *A $\mathcal{C}^1$ function $f : \mathbb{R} \to \mathbb{R}$ is (strictly) convex iff f' is monotonically non-decreasing (increasing).*

Exercise 7. *Prove the above property.*
Hint: use the mean value theorem, stating that for a $\mathcal{C}^1$ function on an interval $[x_1, x_2]$ there exists $x_3 \in (x_1, x_2)$ such that

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = f'(x_3).$$

A direct consequence of the monotonicity of the first-order derivative is non-negativity of the second-order derivative. Formally, if a convex f is $\mathcal{C}^2$, then $f'' \geq 0$. It can actually be shown that the statement holds in both direction, i.e., $f \in \mathcal{C}^2$ is convex iff $f'' \geq 0$. This is not true for strict convexity: while $f'' > 0$ implies strict convexity of f, a strictly convex function may have a second-order derivative that is not everywhere strictly positive (for example, $f(x) = x^4$ is strictly positive, but its second order derivative $f''(x) = 12x^2$ vanishes at $x = 0$).

Another direct consequence can be seen from the firs-order Taylor expansion. By Taylor's theorem, for every choice of x, there exists some y such that

$$f(x + d) = f(x) + f'(x)d + \frac{1}{2}f''(y)d^2$$

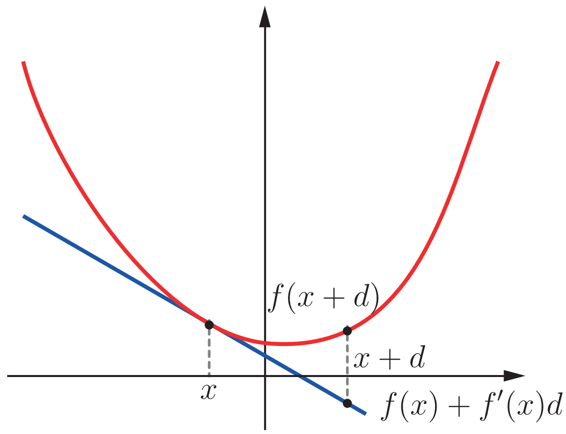

Figure 4: A convex function lies above its tangents.

(this second-order term is usually called the remainder of the Taylor series). However, since $f'' \geq 0$, we have

$$f(x+d) \geq f(x) + f'(x)d.$$

Geometrically, the inequality means that the graph of f lies above all its tangents (Figure 4). Despite we used the fact that f is $\mathcal{C}^2$, this result is more general and holds for any $\mathcal{C}^1$ function. The statement also holds in the converse direction, i.e., a $\mathcal{C}^1$ f is convex iff its graph lies above all its tangents.

The generalization of these properties to the multivariate case is relatively straightforward, as is summarized below:

Property (Gradient inequality). *Let $f \in \mathcal{C}^1$ be convex. Then, $f(\mathbf{x}+\mathbf{d}) \geq f(\mathbf{x}) + \mathbf{d}^{\mathrm{T}}\nabla f(\mathbf{x})$.*

Geometrically, this means that the graph of a convex function lies above all tangent planes.

The non-negativity of the second-order derivative is generalized as

Property (Gradient inequality). *A $\mathcal{C}^2$ function $f : \mathbb{R}^n \to \mathbb{R}$ is (strictly) convex iff it has a positive semi-definite (positive definite) Hessian.*

To prove the above, recall that f is convex iff for every $\mathbf{x}, \mathbf{r} \in \mathbb{R}^n$, the function $\varphi(\alpha) = f(\mathbf{x} + \alpha\mathbf{r})$ is convex. The latter holds iff

$$0 \leq \varphi''(\alpha) = \mathbf{r}^{\mathrm{T}}\nabla^2 f(\mathbf{x})\mathbf{r},$$

which in turn holds iff $\nabla^2 f(\mathbf{x}) \succeq 0$.