Optimization and Computational Linear Algebra for Data Science
Lecture 9: Convex functions

Léo Miolane · leo.miolane@gmail.com

November 8, 2019

Warning: *This material is not meant to be lecture notes. It only gathers the main concepts and results from the lecture, without any additional explanation, motivation, examples, figures...*

1 Convex sets

Definition 1.1 (*Convex set*)

A set $C \subset \mathbb{R}^n$ is convex *if for all $x, y \in C$ and all $\alpha \in [0, 1]$,*

$$\alpha x + (1-\alpha) y \in C.$$

Remark 1.1. *Subspaces of $\mathbb{R}^n$ are convex sets.*

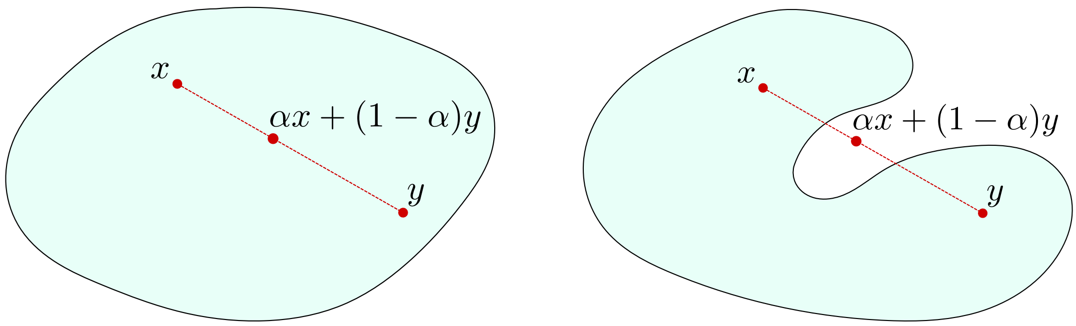

Figure 1: Left: a convex set. Right: a non-convex set.

Definition 1.2 (*Convex combination*)

We say that $y \in \mathbb{R}^n$ is a convex combination of $x_1, \dots, x_k \in \mathbb{R}^n$ if there exists $\alpha_1, \dots, \alpha_k \geq 0$ such that

$$y = \sum_{i=1}^k \alpha_i x_i \qquad \text{and} \qquad \sum_{i=1}^k \alpha_i = 1.$$

Proposition 1.1

If C is convex then all convex combination of elements of C remains in C.

2 Convex functions

Definition 2.1

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if for all $x, y \in \mathbb{R}^n$ and all $\alpha \in [0, 1]$,

$$f(\alpha x + (1 - \alpha)y) \le \alpha f(x) + (1 - \alpha)f(y). \tag{1}$$

We say that f is strictly convex *is there is strict inequality in (1) whenever $x \ne y$ and $\alpha \in (0, 1)$.*
A function f is concave (respectively strictly concave) if $-f$ is convex (respectively strictly convex).

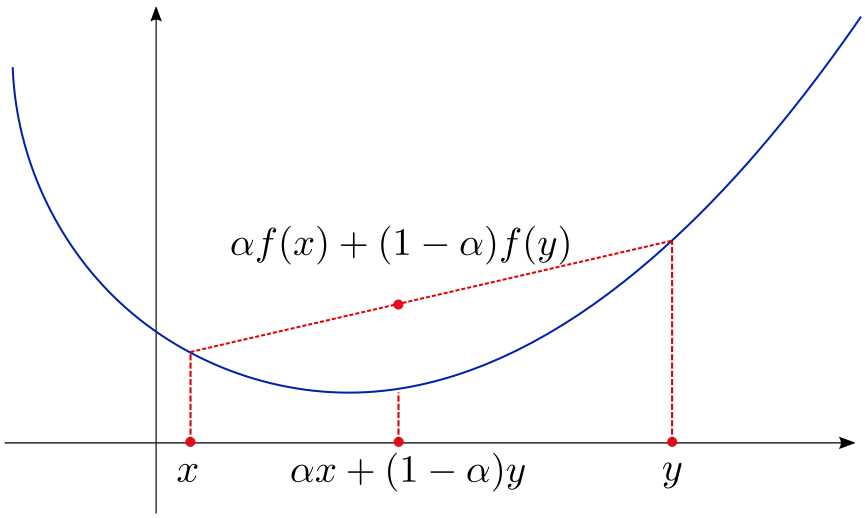

Figure 2: A convex function.

Notice that a linear function is also a convex function since it verifies (1) with equality, but is not strictly convex.

Exercise 2.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ a convex function and $\alpha \in \mathbb{R}$. Show that the "α-sublevel set"*

$$C_\alpha = \{x \in \mathbb{R}^n \,|\, f(x) \le \alpha\}$$

is convex.

2.1 Convex function and differential

Proposition 2.1

A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for all $x, y \in \mathbb{R}^n$

$$f(y) \ge f(x) + \nabla f(x)^\mathsf{T}(y - x).$$

Corollary 2.1

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function and $x \in \mathbb{R}^n$. Then

$$x \text{ is a minimizer of } f \quad \Longleftrightarrow \quad \nabla f(x) = 0.$$

Proposition 2.2

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice-differentiable function. We denote by H_f the Hessian matrix of f. Then f is convex if and only if for all $x \in \mathbb{R}^n$, $H_f(x)$ is positive semi-definite.

When $f : \mathbb{R} \to \mathbb{R}$ is twice differentiable, we get that f is convex if and only if $f'' \geq 0$.

It can be complicated to check that a function f is convex using Proposition 2.2 when f is a function of multiple variables ($n \geq 2$). The next proposition shows that we can always reduce to the unidimensional case, by checking that the restriction of f on every line is convex:

Proposition 2.3

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if the function

$$\begin{array}{rccc} g: & \mathbb{R} & \to & \mathbb{R} \\ & t & \mapsto & f(x+tv) \end{array}$$

is convex for all $x, v \in \mathbb{R}^n$.

2.2 Jensen's inequality

Proposition 2.4 (*Jensen's inequality*)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then for all $x_1, \dots, x_k \in \mathbb{R}^n$ and all $\alpha_1, \dots, \alpha_k \geq 0$ such that $\sum_{i=1}^k \alpha_i = 1$ we have

$$f\left(\sum_{i=1}^k \alpha_i x_i\right) \leq \sum_{i=1}^k \alpha_i f(x_i).$$

More generally, if X is a random variable that takes value in $\mathbb{R}^n$ we have

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

Remark 2.1. *If f is concave then Proposition 2.4 holds, but with inequalities in the reverse order.*

Example 2.1 (Discrete entropy). Let Z be a random variable that take value in $\{1, \dots, k\}$ and write $p_i = \mathbb{P}(Z = i)$. The entropy of Z is defined as

$$H(Z) = -\sum_{i=1}^k p_i \log(p_i).$$

The entropy of Z is a measure of the uncertainty associated with Z. We apply Jensen's inequality to the concave function log:

$$H(Z) = \sum_{i=1}^k p_i \log(1/p_i) \leq \log\left(\sum_{i=1}^k p_i/p_i\right) = \log(k).$$

Notice that $H(Z) = \log(k)$ when Z is uniformly distributed over $\{1, \dots, k\}$, i.e. $\mathbb{P}(Z = i) = 1/k$ for all i. *Conclusion*: maximal entropy is achieved for the uniform distribution.

2.3 Operations that preserve convexity

Proposition 2.5 (***Non-negative linear combination of convex functions***)

Let $f_1, \dots, f_k$ be convex functions from $\mathbb{R}^n \to \mathbb{R}$ and let $\alpha_1, \dots, \alpha_k \geq 0$. Then the function f defined by

$$f(x) = \sum_{i=1}^{k} \alpha_i f_i(x)$$

is convex. In particular a sum of convex functions is convex.

Proposition 2.6 (***Supremum of convex functions***)

Let $(f_i)_{i \in S}$ is a family of convex functions from $\mathbb{R}^n \to \mathbb{R}$. Then the function

$$f(x) = \sup_{i \in S} f_i(x)$$

is convex. In particular, a supremum of affine functions is a convex function.

Proposition 2.7 (***Composition with affine function***)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function, $A \in \mathbb{R}^{n \times m}$ and $b \in \mathbb{R}^n$. Then the function $g : \mathbb{R}^m \to \mathbb{R}$ defined by

$$g(x) = f(Ax + b)$$

is convex.

Further reading

See [1] Chapters 2 and 3 for example of properties of convex sets/functions. See also `http://web.stanford.edu/class/ee364a/lectures.html` for nice lecture slides. The book [2] is a great reference for convex analysis, but is mathematically more involved.

References

[1] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, `https://web.stanford.edu/~boyd/cvxbook/`, 2004.

[2] R Tyrrell Rockafellar. *Convex analysis*, volume 28. Princeton university press, 1970.