MATH-329 Nonlinear optimization
Exercise session 3: Convexity

Instructor: Nicolas Boumal
TAs: Guifré Sánchez, Antoine Gonon

Document compiled on September 23, 2025

1. Strict and strong.

1. Give an example of a function $f\colon \mathbb{R} \to \mathbb{R}$ which is convex but not strictly convex.
2. Give an example of a function $f\colon \mathbb{R} \to \mathbb{R}$ which is strictly convex but not strongly convex.
3. Give an example of a function $f\colon \mathbb{R} \to \mathbb{R}$ which is strictly convex yet not bounded below.
4. Give an example of a function $f\colon \mathbb{R} \to \mathbb{R}$ which is strictly convex and bounded below yet does not have a minimum.

2. Quadratic functions. Let $\mathcal{E} = \mathbb{R}^n$ with the usual inner product and let $f(x) = \frac{1}{2}x^\top Ax + b^\top x + c$.

1. Show that f is convex if and only if $A \succeq 0$.
2. Show that f is strictly convex if and only if $A \succ 0$.
3. Show that f is μ-strongly convex if and only if $A \succeq \mu I$. What is the best choice of μ is terms of A?

3. Jensen's inequality. Let $\mathcal{E}$ be a linear space. Let $f\colon \mathcal{E} \to \mathbb{R}$ be a convex function. Show that for all $x_1, \ldots, x_n \in \mathcal{E}$ and any $\lambda_1, \ldots, \lambda_n \geq 0$ such that $\lambda_1 + \cdots + \lambda_n = 1$ we have

$$f(\lambda_1 x_1 + \cdots + \lambda_n x_n) \leq \lambda_1 f(x_1) + \cdots + \lambda_n f(x_n).$$

Hint: proceed by induction on n.

We call the quantity $\lambda_1 x_1 + \cdots + \lambda_n x_n$ a convex combination of the points $x_1, \ldots, x_n$. The result that you proved shows that if X is a discrete random variable taking values $x_1, \ldots, x_n \in \mathcal{E}$ with probabilities $p_1, \ldots, p_n$ respectively, then we have

$$f(\mathbf{E}[X]) \leq \mathbf{E}[f(X)],$$

where $\mathbf{E}$ denotes mathematical expectation. This inequality generalizes to any random variable X.

4. Log-sum-exp. *If you prefer you can consider the following exercise with $k = 2$.*

1. Show that the *log-sum-exp function* is convex from $\mathbb{R}^k$ to $\mathbb{R}$ ($t > 0$ is a fixed, real parameter):

$$f(x) = t \log\left(\sum_{i=1}^{k} e^{x_i/t}\right). \tag{1}$$

This function is often used in applications because it is a smooth approximation of the maximum function. Indeed:

2. With $\bar{x} = \max_i x_i$ show that

$$\bar{x} \leq f(x) = \bar{x} + t \log\left(\sum_{i=1}^{k} e^{\frac{x_i - \bar{x}}{t}}\right) \leq \bar{x} + t\log(k). \tag{2}$$

Thus, the smaller t is, the better the approximation. However:

3. From an optimization perspective (for example, if we plan to use gradient descent), can you see a reason why we should not take t too small?

Note: on a computer, it is necessary to use expression (2) rather than expression (1) to compute f (and its derivatives). Indeed, expression (1) can lead to overflow when t is small because it involves computing exponentials of possibly large numbers. In contrast, expression (2) only involves exponentials of nonpositive numbers. Still, even with expression (2), evaluating f and its derivatives can get tricky numerically when t is small.

5. Norms. Let $\mathcal{E}$ be a Euclidean space.

1. Show that any norm on $\mathcal{E}$ is convex.

2. Show that any squared norm on $\mathcal{E}$ is convex.

Interestingly, a norm is never differentiable at $x = 0$. Do you see why? However:

3. Let $\langle \cdot, \cdot \rangle$ be an inner product on $\mathcal{E}$ and $\| \cdot \|$ the associated norm (that is, $\|x\| = \sqrt{\langle x, x \rangle}$ for all $x \in \mathcal{E}$). Prove that the squared norm $x \mapsto \|x\|^2$ is differentiable.

A norm may not be differentiable if it is not derived from an inner product. Can you come up with an example?