MATH-329 Nonlinear optimization
Exercise session 4: Hessians and Newton's method

Instructor: Nicolas Boumal
TAs: Antoine Gonon and Guifré Sánchez

Document compiled on September 24, 2025

Exercises marked with $(*)$ will be used in later exercises or in the homeworks: you might want to prioritize those.

1. Good and bad global behavior of Newton's method.

1. Consider the function $f(x) = \frac{1}{4}x^4 - x^2 + 2x + 1$. What is the behavior of Newton's method on f if the initial point is $x_0 = 0$? (Observe numerically first.)
2. Argue that $f(x) = \log(e^x + e^{-x})$ has a Lipschitz continuous gradient, a Lipschitz continuous Hessian, and is strictly convex (it helps to plot the function). What is the behavior of Newton's method on this function with $x_0 = 1$? And with $x_0 = 1.5$?
3. Consider the Rosenbrock function
$$f(x, y) = (a - x)^2 + b(y - x^2)^2$$
with $a = 1$ and $b = 100$. Run Newton's method with $x_0 = (-1.2, 1)$. Compare the performance with the one of gradient descent for this initial point.

2.$(*)$ Regression loss function. We let $y \in \mathbb{R}^m$ and $F\colon \mathbb{R}^n \to \mathbb{R}^m$ be a C^2 map. Consider the C^2 regression function $f\colon \mathbb{R}^n \to \mathbb{R}$ defined by
$$f(x) = \frac{1}{2}\|F(x) - y\|^2.$$
Find the gradient and the Hessian of f as a function of the derivatives of F.
Hint: There are two ways to see this. You can see F as m scalar functions that have a gradient and a Hessian and express everything as a function of this. Or you can define the Jacobian of F as $J(x) = \mathrm{D}F(x) \in \mathbb{R}^{m\times n}$. Then J is a function $\mathbb{R}^n \to \mathbb{R}^{m\times n}$ and for all $x, u \in \mathbb{R}^n$ we have $\mathrm{D}J(x)[u] \in \mathbb{R}^{m\times n}$.

3. Computing Hessians. For the following functions $f\colon \mathcal{E} \to \mathbb{R}$ give an expression for the gradient and the Hessian. Specifically, for the Hessian compute $\nabla^2 f(x)[u]$ for all $x, u \in \mathcal{E}$.

1. Given $a \in \mathbb{R}^n$, consider $f(x) = \frac{1}{2}(x^\top x + (a^\top x)^2)$ for all $x \in \mathbb{R}^n$.
2. $(*)$ Given $A \in \mathbb{R}^{m\times n}$, $M \in \mathbb{R}^{n\times n}$, consider the function $f\colon \mathbb{R}^{m\times n} \to \mathbb{R}$ defined by
$$f(X) = \frac{1}{2}\|A^\top X - M\|_\mathrm{F}^2,$$
where $\|\cdot\|_\mathrm{F}$ is the Frobenius norm.