Tutorial 1 : Introduction to optimization and refresher course

Exercise 1 (Differentiation of some functions)**.** Compute the gradient and Hessian of the following functions:

1. $f : \mathbb{R}^d \to \mathbb{R} : x \mapsto \|\mathbf{A}x - b\|_2^2$ ($\mathbf{A}$ and b are constant matrix and vector).
2. $f : \mathbb{R}^d \to \mathbb{R} : x \mapsto x^\top \mathbf{A}x - b^\top x + c$ ($\mathbf{A}, b, c$ are constant matrix, vector and scalar).
3. $f : \mathbb{R}^d \to \mathbb{R} : x \mapsto \|x\|_2^a$ (where $a > 2$).
4. $g : \mathbb{R} \to \mathbb{R} : t \mapsto f(x + t(y - x))$ (x, y are two fixed vectors, f is a fixed C^2 function). Express the gradient and the Hessian matrix of g by those of f.

Exercise 2 (Differentiable but not C^1)**.** Consider the function:

$$f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right), & x \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

1. Is f differentiable?
2. Is f continuously differentiable?
3. Based on this function, can you construct a function f such that f is twice differentiable but not C^2?

Exercise 3 (Necessary conditions of optimal solution revisited)**.** If f is only differentiable and not C^1, is it still necessary that $\nabla f(x^\star) = 0$ for any local solution $x^\star$?

Exercise 4 (Properties of derivatives and gradient)**.** Given two differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$, we have:

$$\begin{aligned} \nabla(f + g)(x) &= \nabla f(x) + \nabla g(x) \\ \nabla(\alpha f)(x) &= \alpha \nabla f(x), \forall \alpha > 0 \\ \nabla(f \cdot g)(x) &= g(x)\nabla f(x) + f(x)\nabla f(x) \\ \nabla\left(\frac{f}{g}\right) &= \frac{g(x)\nabla f(x) - f(x)\nabla g(x)}{g(x)^2}, \quad \text{assuming that } g(x) > 0. \end{aligned} \tag{1}$$

Exercise 5 (Chain rule)**.** Given two differentiable functions $f : \mathbb{R}^k \to \mathbb{R}^\ell$ and $g : \mathbb{R}^d \to \mathbb{R}^k$, prove that the composition $f \circ g : \mathbb{R}^d \to \mathbb{R}^\ell$ is also differentiable and its Jacobian matrix is given by:

$$J_{f \circ g}(x) = J_f(g(x))J_g(x).$$

Exercise 6 (Two Taylor formulations)**.** Given a C^1 (resp. C^2) function $f : \mathbb{R}^d \to \mathbb{R}$, we have:

$$f(y) = f(x) + \int_0^1 \nabla f(x + t(y-x))^\top (y-x)\, dt \qquad , \forall x, y \in \mathbb{R}^d$$

$$\text{(resp.)} f(y) = f(x) + (y-x)^\top \nabla f(x) + \frac{1}{2}(y-x)^\top \nabla^2 f(x)(y-x) + R_2(x-y) \quad , \forall x, y \in \mathbb{R}^d, \tag{2}$$

where $R_2(x-y)$ is a reminder satisfying $\lim_{y \to x} \frac{R_2(x-y)}{\|y-x\|^2} = 0$.
Hint: you might need to use the fundamental theorem of calculus, i.e., if $f : \mathbb{R} \to \mathbb{R}$ is differentiable, we have:

$$f(b) = f(a) + \int_a^b f'(t)dt.$$