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.

Solution for Exercise 1. We have:

1.
$$\begin{aligned} f(x) &= (\mathbf{A}x - b)^\top(\mathbf{A}x - b) = x^\top \mathbf{A}^\top \mathbf{A}x - 2b^\top \mathbf{A}x + \|b\|_2^2, \\ \nabla f(x) &= 2\mathbf{A}^\top \mathbf{A}x - 2\mathbf{A}^\top b, \\ \nabla^2 f(x) &= 2\mathbf{A}^\top \mathbf{A}. \end{aligned}$$

2.
$$\begin{aligned} \nabla f(x) &= 2\mathbf{A}x - b, \\ \nabla^2 f(x) &= 2\mathbf{A}. \end{aligned}$$

3.
$$\begin{aligned} f(x) &= \left(\sum_{i=1}^d x_i^2\right)^{\frac{a}{2}}, \\ \nabla f(x) &= \frac{a}{2}\left(\sum_{i=1}^d x_i^2\right)^{\frac{a}{2}-1} x, \\ \nabla^2 f(x) &= \frac{a}{2}\left(\left(\frac{a}{2} - 1\right) xx^\top + \left(\sum_{i=1}^d x_i^2\right)^{\frac{a}{2}-1} \mathbf{I}\right). \end{aligned}$$

4.
$$\begin{aligned} \nabla g(t) &= \nabla f(x + t(y - x))^\top (y - x), \\ \nabla^2 g(t) &= (y - x)^\top \nabla^2 f(x + t(y - x))(y - x). \end{aligned}$$

$\square$

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 continuously differentiable but not C^2?

Solution for Exercise 2. We have:

1. Yes, f is differentiable because:

$$f'(x) = \begin{cases} 2x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right), & \text{if } x \neq 0 \\ 0, & \text{if } x = 0 \end{cases}.$$

2. f is not continuously differentiable because f' is discontinuous at $x = 0$.
3. Take the integral of $f(x)$.

$\square$

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 $x^\star$?

Solution for Exercise 3. Yes, the same proof still applied because we do not use the continuity of the derivatives. $\square$

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:

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

Solution for Exercise 4. The first two claims are clear from definition. We will only deal with the third and the fourth claims.

1. We have:

$$\begin{aligned} &\lim_{d\to 0} \frac{f(x+d)g(x+d) - f(x)g(x)}{\|d\|} \\ &= \lim_{d\to 0} \frac{f(x+d)g(x+d) - f(x)g(x+d)}{\|d\|} + \lim_{d\to 0} \frac{g(x)f(x+d) - g(x)f(x)}{\|d\|} \\ &= \lim_{d\to 0} g(x+d) \lim_{d\to 0} \frac{f(x+d) - f(x)}{\|d\|} + fg(x) \lim_{d\to 0} \frac{f(x+d) - f(x)}{d} \\ &= f(x)\langle \nabla g(x), d\rangle + g(x)\langle \nabla f(x), d\rangle = \langle g(x)\nabla f(x) + f(x)\nabla g(x), d\rangle. \end{aligned}$$

2. It is sufficient to compute the gradient of the function $h(x) = \frac{1}{g(x)}$ and apply the third claim. Since $h(x)g(x) = 1$ a constant, $\nabla(g(x)h(x)) = 0$.

$$0 = \nabla(g \cdot h)(x) = h(x)\nabla g(x) + g(x)\nabla h(x).$$

Therefore,

$$\nabla h(x) = -\frac{1}{g(x)^2}\nabla g(x).$$

□

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:

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

Solution to Exercise 5. By the definition of Jacobian matrix, we have:

$$\begin{aligned} f \circ g(x + d) &= f(g(x) + J_g(x)d + R_1(d)) \quad \text{where } \lim_{d \to 0} \frac{\|R_1(d)\|}{\|d\|} = 0 \\ &= f(g(x)) + J_f(g(x))J_g(x)d + J_f(g(x))R_1(d) + R_2(d) \quad \text{where } \lim_{d \to 0} \frac{\|R_2(d)\|}{\|d\|} = 0 \end{aligned}$$

Therefore,

$$\lim_{d \to 0} \frac{\|f \circ g(x + d) - f \circ g(x) - J_f(g(x))J_g(x)d\|}{\|d\|} = 0.$$

or equivalently, $J_{f \circ g}(x) = J_f(g(x))J_g(x)$. □

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

$$\begin{aligned} f(y) &= f(x) + \int_0^1 \nabla f(x + t(y - x))^\top (y - x)\, dt \quad &, \forall x, y \in \mathbb{R}^d \\ \text{(resp.)} f(y) &= f(x) + (y - x)^\top \nabla f(x) + \frac{1}{2}(y - x)^\top \nabla^2 f(x)(y - x) + R_2(x - y) \quad &, \forall x, y \in \mathbb{R}^d, \end{aligned} \tag{2}$$

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:

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

Solution of Exercise 6. We have:

1. Consider the function $g(t) = f(x + t(y - x))$. We have:

$$\begin{aligned} f(y) = g(1) &= g(0) + \int_0^1 g'(t)dt \\ &= f(x) + \int_0^1 \nabla f(x + t(y - x))^\top (y - x)dt \end{aligned}$$

2. Consider the same function $g(t)$, we have:

$$\begin{aligned}
f(y) = g(1) &= g(0) + \int_0^1 g'(t)dt \\
&= g(0) - \big[(1-t)g'(t)\big]_0^1 + \int_0^1 (1-t)g''(t)dt \\
&= f(x) + \nabla f(x)^\top (y-x) + \int_0^1 (1-t)(y-x)^\top \nabla^2 f(x+t(y-x))(y-x)dt \\
&= f(x) + \nabla f(x)^\top (y-x) + \frac{1}{2}(y-x)^\top \nabla^2 f(x)(y-x) + R_2(y-x)
\end{aligned}$$

where R_2 is given by:

$$R_2(y-x) = \int_0^1 (y-x)^\top (\nabla^2 f(x+t(y-x)) - \nabla^2 f(x))(y-x)dt$$

Finally, we remark that:

$$\begin{aligned}
|R_2(y-x)| &= \left| \int_0^1 (y-x)^\top (\nabla^2 f(x+t(y-x)) - \nabla^2 f(x))(y-x)dt \right| \\
&\leq \int_0^1 \|(y-x)^\top (\nabla^2 f(x+t(y-x)) - \nabla^2 f(x))(y-x)\| dt \\
&\leq \int_0^1 \|(y-x)\|^2 \|\nabla^2 f(x+t(y-x)) - \nabla^2 f(x)\| dt \\
&= \|y-x\|^2 \int_0^1 \|\nabla^2 f(x+t(y-x)) - \nabla^2 f(x)\| dt = o(\|y-x\|^2),
\end{aligned}$$

because f is C^2. □