TD 1 - Differentials

Exercise 1

$$f\colon x \longmapsto \|x\|^2$$
$$\mathbb{R}^n \mapsto \mathbb{R}$$

Using the definition of the $\|\cdot\|$ for any a, $\Delta x \in \mathbb{R}^n$

$$f(a+\Delta x) = \|a+\Delta x\|^2 = \|a\|^2 + \|\Delta x\|^2 + 2\langle a, \Delta x\rangle.$$

Under the assumption that Δx is infinitely small it can be rewritten as

$$f(a+\Delta x) = \|a\|^2 + \underbrace{\|\Delta x\|^2}_{o(\Delta x)} + 2\langle a, \Delta x\rangle = f(a) + \underbrace{2\langle a, \Delta x\rangle}_{df(a)(\Delta x)} + o(\Delta x).$$

Let us now write an explicit formula for the gradient

$$df(a)(\Delta x) = \langle \nabla f(a), \Delta x\rangle \Leftrightarrow \nabla f(a) = 2\left(a_1, a_2, \ldots, a_n\right)^\top = 2a.$$

Using partial derivatives of f the gradient can be written as

$$\nabla f(a) = \left(\frac{\partial f(a)}{\partial x_1}, \frac{\partial f(a)}{\partial x_2}, \ldots, \frac{\partial f(a)}{\partial x_n}\right)^\top = 2\left(a_1, a_2, \ldots, a_n\right)^\top = 2a.$$

Exercise 2

a)

For any $x, h \in \mathbb{R}^n$, we have :

$$\begin{aligned} f(x+h) = \|A(x+h)-b\|^2 &= \|Ax-b\|^2 + 2\langle Ax-b, Ah\rangle + \|Ah\|^2 \\ &= f(x) + \langle 2A^T(Ax-b), h\rangle + \|Ah\|^2 \end{aligned}$$

and since A is symmetric and $\|Ah\|^2 \leq (|||A|||\ h)^2 = o(h)$, we obtain

$$\nabla f(x) = 2A(Ax-b)$$

b)

We apply theorem of differentiation of composition : if $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}$ are both differentiable on $\mathbb{R}^n$, then $f \circ g$ is differentiable on $\mathbb{R}^n$ and for any $x \in \mathbb{R}^n$

$$\nabla(f \circ g)(x) = \underset{g}{\mathrm{Jac}}(x)^T \nabla f(g(x))$$

Applying this formula with $f = \|.\|^2$ and $g = G$, we obtain along with exercise 1 :

$$\nabla(\|.\|^2 \circ G)(x) = \left(\underset{G}{\mathrm{Jac}}(x)^T \nabla(\|.\|^2)(G(x))\right) = 2\,\underset{G}{\mathrm{Jac}}(x)^T G(x)$$

Exercise 3

a)

Gradient: by symmetry of A, for any $x, h \in \mathbb{R}^n$, we have :

$$\begin{aligned} f(x+h) &= (x+h)^T A(x+h) + p^T(x+h) + c = f(x) + 2(Ax)^T h + p^T h + h^T A h \\ &= f(x) + \langle 2Ax + p, h\rangle + h^t A h \end{aligned}$$

but since $h^t A h = o(\|h\|)$, we get $\nabla f(x) = 2Ax + p$.

Hessian: by definition of the Hessian, $\nabla^2 f(x) := \text{Jac}_{\nabla f}(x) = 2A$

b)

Gradient: if we denote by $G : \mathbb{R}^n \to \mathbb{R}^m$ the function such that for any $i \in [\![1, m]\!]$, and $x \in \mathbb{R}^n$, $G(x)_i = g_i(x)$, we have $\sum_{i=1}^m g_i(x)^2 = \|G(x)\|^2$. Hence, by exercise 2, question g, we have :

$$\nabla g(x) = 2 \operatorname*{Jac}_G(x)^T G(x)$$

Hessian: for any $p, q \in [\![1, n]\!]$, and $x \in \mathbb{R}^n$ we have

$$\begin{aligned}
\frac{\partial g}{\partial x_p}(x) &= 2 \sum_{i=1}^m \frac{\partial g_i}{\partial x_p}(x) \\
\frac{\partial^2 g}{\partial x_p \partial x_q}(x) &= 2 \sum_{i=1}^m \frac{\partial^2 g_i}{\partial x_p \partial x_q}(x) + \frac{\partial g_i}{\partial x_p}(x) \frac{\partial g_i}{\partial x_q}(x) \\
&= 2 \sum_{i=1}^m \nabla^2(g_i)(x)_{p,q} + (\nabla g_i)(x) \nabla g_i)(x)^T)_{p,q}
\end{aligned}$$

which leads to $\nabla^2 g = 2 \sum_{i=1}^m \nabla^2(g_i) + \nabla g_i(x) \nabla g_i(x)^T$

Exercise 4

a)

Let us define $g \colon t \mapsto \bar{x} + tu$, then $q = f \circ g$. Now let us use the chain rule for the derivative of composition of the functions.

$$dq_t(h) = \left(df_{g(t)} \circ \underbrace{dg_t}_{hu} \right)(h) = df_{\bar{x}+tu}(hu) = \nabla f(\bar{x} + tu)^\top hu.$$

which leads to

$$q'(t) = \nabla f(\bar{x} + tu)^\top u.$$

b)

For any $t, h \in \mathbb{R}$, h near to 0, we have :

$$\begin{aligned}
q'(t + h) &= u^T \nabla f(x + (t + h)u) = u^T \nabla f(x + tu + hu) \\
&= u^T \nabla f(x + tu + hu) = u^T (\nabla f(x + tu) + J_{\nabla f}(x + tu)hu + o(h)) \\
&= q'(t) + u^T J_{\nabla f}(x + tu)u\, h + o(h)
\end{aligned}$$

but by definition of the hessian, $J_{\nabla f}((x + tu)) = \nabla^2 f(x + tu)$. Therefore,

$$q''(t) = u^T \nabla^2 f(x + tu)u$$

c), d)

For the function q Taylor series in 0 are the following

$$q(t) = q(0) + tq'(0) + o(t)$$

for the first order approximation and

$$q(t) = q(0) + tq'(0) + \frac{t^2}{2} q''(0) + o(t^2)$$

for the second one. As far as $\bar{x}$ is a local minimum it means that for any t that is small enough and for any u

$$f(\bar{x}+tu) \geq f(\bar{x}) \Leftrightarrow q(t) \geq q(0).$$

Gradient: using the first order approximation of $q(t)$ for this inequality we have

$$q(0) + tq'(0) + o(t) \geq q(0) \Leftrightarrow q'(0) \geq 0.$$

Now using the result from a) we have

$$\forall\, u, \quad tq'(0) = tu^\top \nabla f(\bar{x}) \geq 0 \Rightarrow \nabla f(\bar{x}) = 0.$$

Hessian: using the second order approximation of $q(t)$ in 0 we have

$$q(0) + \cancelto{0}{tq'(0)} + \frac{t^2}{2} q''(0) + o(t^2) \geq q(0).$$

Using the result of b) we have

$$\forall\, u, \quad q''(0) = u^\top \nabla^2 f(\bar{x}) u \geq 0.$$

e)

Gradient: by the definition

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right) = 0 \Leftrightarrow \forall\, i \quad \frac{\partial f}{\partial x_i} = 0.$$

Hessian: let H be a symmetric matrix

$$H = \begin{bmatrix} \alpha & \gamma \\ \gamma & \beta \end{bmatrix}.$$

Then positive-semidefiniteness of H is on the one hand $\forall\, u,\ u^\top H u \geq 0$ and on the other hand $\forall\, i,\ \lambda_i \geq 0$, where $\{\lambda_i\}$ are eigenvalues of H. It is known that $\det H = \prod_i \lambda_i = \alpha\beta - \gamma^2$ and $\operatorname{tr} H = \sum_i \lambda_i = \alpha + \beta$. Then

$$\forall\, i,\ \lambda_i \geq 0 \Leftrightarrow \det H \geq 0 \,\&\, \operatorname{tr} H \geq 0.$$

Exercise 5

a)

On the one hand, $\forall x_0 \in X$, $y_0 \in Y$

$$\begin{cases} f(x_0) & \geq \inf_{x\in X} f(x) \\ g(y_0) & \geq \inf_{y \in Y} g(y) \end{cases}$$

and summing it up we have

$$\inf_{x\in X} f(x) + \inf_{y\in Y} g(y) \leq f(x_0) + g(y_0) \Leftrightarrow \inf_{x\in X} f(x) + \inf_{y\in Y} g(y) \leq \inf_{x\in X,\, y\in Y} \left(f(x) + g(y)\right).$$

On the other hand, by the definition $\forall\, \varepsilon > 0, \exists\, x_0 \in X, y_0 \in Y:$

$$\begin{cases} f(x_0) & < \inf_{x\in X} f(x) + \frac{\varepsilon}{2} \\ g(y_0) & < \inf_{y\in Y} g(x) + \frac{\varepsilon}{2} \end{cases}$$

Summing up these two inequalities we have

$$\inf_{x\in X} f(x) + \inf_{y\in Y} g(x) + \varepsilon > f(x_0) + g(y_0) > \inf_{x\in X,\, y\in Y} \left(f(x) + g(y)\right).$$

Finally, setting $\varepsilon \to 0$ we have

$$\inf_{x\in X} f(x) + \inf_{y\in Y} g(y) \geq \inf_{x\in X,\, y\in Y} \left(f(x)+g(y)\right).$$

If there are points $\bar{x}$ and $\bar{y}$ s.t. $f(\bar{x}) = \inf_{x\in X} f(x)$ and $g(\bar{y}) = \inf_{y\in Y} g(y)$ then

$$\inf_{x\in X} f(x) + \inf_{y\in Y} g(y) = f(\bar{x}) + g(\bar{y}) \geq \inf_{x\in X,\, y\in Y} \left(f(x)+g(y)\right),$$

using the equality proven above we have

$$f(\bar{x}) + g(\bar{y}) = \inf_{x\in X,\, y\in Y} \left(f(x)+g(y)\right),$$

that means that pair $(\bar{x}, \bar{y})$ minimizes $f + y$ on $X \times Y$.

b)

Let us rewrite the problem in separable view

$$\begin{cases} \min c^\top x \\ l_i \leq x_i \leq u_i \ \forall\, i \in \{1,\dots,n\} \\ x \in \mathbb{R}^n \end{cases} \Leftrightarrow \begin{cases} \min \sum\limits_{i=1}^n c_i x_i \\ l_i \leq x_i \leq u_i \ \forall\, i \in \{1,\dots,n\} \\ x \in \mathbb{R}^n \end{cases} \Leftrightarrow \begin{cases} \min \sum\limits_{i=1, c_i>0}^n c_i x_i + \sum\limits_{i=1, c_i<0}^n c_i x_i + \sum\limits_{i=1, c_i=0}^n c_i x_i \\ l_i \leq x_i \leq u_i \ \forall\, i \in \{1,\dots,n\} \\ x \in \mathbb{R}^n \end{cases}$$

Using the result from a) all problems for x_i may be solved independently of each other. Then explicit solution is following

$$\begin{cases} x_i \in [l_i, u_i] & \text{if } c_i = 0, \\ x_i = l_i & \text{if } c_i > 0, \\ x_i = u_i & \text{if } c_i < 0 \end{cases}$$

and the minimum will be equal to

$$\min = \sum_{i=1, c_i>0}^n c_i l_i + \sum_{i=1, c_i<0}^n c_i u_i$$

Exercise 6

a)

Let us define function $\varphi\colon t \mapsto f(y + t(x-y))$. Then, from the fundamental theorem of calculus it follows

$$f(x) - f(y) = \varphi(1) - \varphi(0) = \int_0^1 \varphi\prime(t)dt = \int_0^1 (x-y)^\top \nabla f(y + t(x-y))dt$$

b)

Using the result of a)

$$\begin{aligned} f(x) = f(y) + \int_0^1 (x-y)^\top \nabla f(y+t(x-y))dt &= f(y) + \int_0^1 (x-y)^\top \left[\nabla f(y+t(x-y)) + \nabla f(y) - \nabla f(y)\right] dt \\ &= f(y) + (x-y)^\top \nabla f(y) + \int_0^1 (x-y)^\top \underbrace{\left[\nabla f(y+t(x-y)) - \nabla f(y)\right]}_{\leq Lt\|x-y\|} dt \\ &\leq f(y) + (x-y)^\top \nabla f(y) + \|x-y\|^2 L \int_0^1 t dt = f(y) + (x-y)^\top \nabla f(y) + \frac{L}{2}\|x-y\|^2 L \end{aligned}$$

c)

Consider a function $f\colon x \mapsto \|x\|^2$. As we know from ex. 1 $\nabla f(a) = 2a$ that means that f has L-Lipschitz gradient with $L = 2$. It is easy to see, that

$$f(x) = \|x\|^2 = \|y + (x - y)\|^2 = \|y\|^2 + \underbrace{\|x - y\|^2}_{= \frac{L}{2}\|x-y\|^2} + \underbrace{2\langle y, x - y\rangle}_{(x-y)^\top \nabla f(y)} .$$