Exercise 3

① Let $f_1, f_2, \dots, f_m$ be smooth with parameters $L_1, L_2, \dots, L_m$. Show that the function $f = \sum_{i=1}^{m} f_i f$ is smooth with parameter $\sum_{i=1}^{m} L_i$.

For every f_i, $i = 1, \dots, m$, the following holds:
(they are smooth with parameter L_i)

$$f_i(y) \leq f_i(x) + \langle Df_i(x), y-x \rangle + \frac{L_i}{2} \|y-x\|^2 \quad , \; i = 1, \dots, m$$

If we sum all of these statements, we get:

$$\sum_{i=1}^{n} f_i(y) \leq \sum_{i=1}^{m} f_i(x) + \underline{\langle \sum_{i=1}^{n} Df_i(x), y-x \rangle} + \frac{\sum_{i=1}^{n} L_i}{2} \underline{\|y-x\|^2}$$

$$\hookrightarrow \sum_{i=1}^{n} \langle Df_i(x), y-x \rangle = \langle \sum_{i=1}^{n} Df_i(x), y-x \rangle$$

$$\hookrightarrow \sum_{i=1}^{n} \frac{L_i}{2} \|y-x\|^2 = \frac{1}{2} \|y-x\|^2 \cdot \sum_{i=1}^{n} L_i$$

Because of $f = \sum_{i=1}^{m} f_i$ we have:

$$f(y) \leq f(x) + \langle f(x), y-x \rangle + \frac{\sum_{i=1}^{n} L_i}{2} \|y-x\|^2.$$

② Let f be smooth with parameter L and A a matrix. Show that $f \circ A$ is smooth with parameter $L\|A\|^2$.

f is L-smooth, we have:

$$\|Df(x) - Df(y)\| \leq L \|y-x\|$$

Let $g(x) = f(xA)$. We want to show that:

$$\|Dg(x) - Dg(y)\| \leq L\|A\|^2 \|y-x\|.$$

We start from f:

$$\begin{aligned} \|Df(xA) - Df(yA)\| &\leq L \|yA - xA\| \quad \longrightarrow \|\vec{\lambda} A\| \leq \|\vec{\lambda}\| \|A\| \\ &\leq L \|(y-x)A\| \\ &\leq L \|A\| \|y-x\| \\ &\leq L \|A\|^2 \|y-x\| \quad \text{because } \|A\| > 0 \text{ and we have } \leq \end{aligned}$$

We have shown that:

$$\|Dg(x) - Dg(y)\| \leq L\|A\|^2 \|y-x\|$$

Therefore $g(x)$ is $L\|A\|^2$ smooth.

④ Theoretical fixed point

- Rewrite $x_{k+1} = g(x_k)$, as a step of gradient descent.

$$x_{k+1} = x_k - \alpha f'(x_k) = g(x_k)$$

Derive f'.

$$f'(x_k) = \frac{-g(x_k) + x_k}{\alpha} \qquad f(x_k) = \frac{1}{\alpha} \int (x_k - g(x_k))\,dx_k = \frac{1}{2\alpha} x_k^2 - \frac{1}{\alpha} \int g(x_k)\,dx_k$$

- Give sufficient conditions on g to ensure convergence.

$$|g'(x)| \leq k, \quad k < 1 \quad \forall x \in (a, b)$$

In the case for $\log(1+x)$, if we look at the graph of its derivative, we can see that there is no k, because in -0.1 the gradient has value > 1.

The function $\log(2+x)$, we see that on the given interval has a value below 0,6.

* g has to be also continuous and differentiable on that interval, so it can achieve local convergence.

The optimal step size is:

$$\alpha^* = \frac{1}{L}$$ where L is the smoothness constant of f.