We recall the following definitions :

— A function f is said to be convex if and only if

$$\forall (x, y),\ f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle . \tag{Convexity}$$

— A function f is said to be L-smooth if and only if

$$\forall (x, z),\ f(x) \leq f(z) + \langle \nabla f(z), x - z \rangle + \frac{L}{2}\|x - z\|^2. \tag{Smoothness}$$

Exercice 1 (Simplest proof of GD). The Gradient descent method is defined by the update rule

$$x_{t+1} = x_t - \gamma \nabla f(x_t) \tag{GD}$$

where γ is called step-size.

Let f a L-smooth convex function and $(x_t)_{t\in\mathbb{N}}$ following this dynamic.

1. Prove that $f(x_{t+1}) \leq f(x_t) - \gamma \left(1 - \frac{L\gamma}{2}\right) \|\nabla f(x_t)\|^2$.
2. Given the previous inequality, provide the best choice for γ and prove the *Descent Lemma*
$$f(x_{t+1}) \leq f(x_t) - \frac{1}{2L}\|\nabla f(x_t)\|^2. \tag{Descent Lemma}$$
3. Use (Convexity) to upper bound $f(x_t) - f_\star$.
4. Let $(C_t)_{t\in\mathbb{N}}$ a non decreasing sequence such that $C_0 = 0$, and for all $t \geq 0$, define V_t as
$$V_t = C_t(f(x_t) - f_\star) + \frac{L}{2}\|x_t - x_\star\|^2. \tag{1}$$
Assuming V_t is non increasing, what convergence guarantee do we obtain ? Do we want to maximize C_t or to minimize it ?
5. Using the previously obtained inequalities, provide the optimal C_t such that V_t is non increasing.
Note $C_{t+1}(f(x_{t+1}) - f_\star) - C_t(f(x_t) - f_\star) = C_{t+1}(f(x_{t+1}) - f(x_t)) + (C_{t+1} - C_t)(f(x_t) - f_\star)$.

bonus : Based on tightest obtained bound on $V_{t+1} - V_t$, modify slightly V_t to obtain a convergence guarantee on the smallest observed gradient norm.

Exercice 2 (Simplest proof of NAG). The Nesterov accelerated gradient method is defined by the update rule

$$y_t = x_t + \beta_t(x_t - x_{t-1})x_{t+1} = y_t - \frac{1}{L}\nabla f(y_t) \tag{NAG}$$

where $(\beta_t)_{t\in\mathbb{N}}$ is called momentum parameter.

Let f a L-smooth convex function and $(x_t, y_t)_{t\in\mathbb{N}}$ following this dynamic.

Let $(C_t)_{t\in\mathbb{N}}$ a non decreasing sequence such that $C_0 = 0$. We want to define V_t for all $t \geq 0$, as

$$V_t = C_t(f(x_t) - f_\star) + R_t \tag{2}$$

where $R_t \geq 0$, $R_0 = \frac{L}{2}\|x_0 - x_\star\|^2$ and such that V is non increasing. Therefore, we would have

$$C_t(f(x_t) - f_\star) \leq V_t \leq V_0 = R_0 = \frac{L}{2}\|x_0 - x_\star\|^2,$$

and finally $f(x_t) - f_\star \leq \frac{L}{2}\frac{\|x_0 - x_\star\|^2}{C_t}$.

1. Using (Descent Lemma) once and (Convexity) twice, upper bound $C_{t+1}(f(x_{t+1})-f_\star)-C_t(f(x_t)-f_\star)=C_{t+1}(f(x_{t+1})-f(y_t))+C_t(f(y_t)-f(x_t))+(C_{t+1}-C_t)(f(y_t)-f_\star))$.
2. Find non negative R_t as mentioned above so that V_t is non increasing. To simplify the notation, we will define $\lambda_t \triangleq \sqrt{C_t}$. Propose an appropriate choice for the sequences λ and β, and prove that $f(x_t)-f_\star \leq \frac{L}{2}\frac{\|x_0-x_\star\|^2}{\lambda_t^2}$.
3. Conclude that $f(x_t)-f_\star = O(1/t^2)$.

Exercice 3 (Cocoercivity). The goal of the exercise is to re-establish the cocoercivity inequality

$$\frac{1}{2L}\|\nabla f(z)-\nabla f(y)\|^2 \leq f(z)-f(y)+\langle \nabla f(y), y-z\rangle \tag{Coco}$$

characterizing smooth convex functions.

1. Show that f is convex and L-smooth if and only if, $\forall(x,y,z)$

$$f(y)+\langle\nabla f(y), x-y\rangle \leq f(z)+\langle \nabla f(z), x-z\rangle+\frac{L}{2}\|x-z\|^2. \tag{3}$$

2. Show that f is convex and L-smooth if and only if, $\forall(y,z)$

$$0 \leq f(z)-f(y)+\langle\nabla f(y), y-z\rangle-\frac{1}{2L}\|\nabla f(z)-\nabla f(y)\|^2. \tag{4}$$

i.e.

$$\frac{1}{2L}\|\nabla f(z)-\nabla f(y)\|^2 \leq f(z)-f(y)+\langle\nabla f(y), y-z\rangle \tag{5}$$

3. Show that f is convex and L-smooth if and only if, $\forall(y,z)$

$$\frac{1}{L}\|\nabla f(z)-\nabla f(y)\|^2 \leq \langle\nabla f(y)-\nabla f(z), y-z\rangle\,. \tag{6}$$

Hint : for the reverse direction, show that f satisfies for any $\eta,\theta\in\mathbb{R}^d$, $\langle\theta-\eta, \nabla f(\eta)-\nabla f(\theta)\rangle \geq 0$, iif f is convex. Hint : consider $g(t)=f(\theta+t(\eta-\theta))$. Show that $t\geq 0$, $g'(t)\geq g'(0)$ and prove $f(\eta)\geq f(\theta)+\langle\nabla f(\theta), \eta-\theta\rangle$.

4. Find a similar inequality characterizing smooth strongly convex functions.

Exercice 4. In exercise 2 we showed that considering

$$V_t \triangleq \lambda_t^2(f(x_t)-f_\star)+\frac{L}{2}\|\lambda_t(x_t-x_\star)+(1-\lambda_t)(x_{t-1}-x_\star)\|^2, \tag{7}$$

we have $V_{t+1}\leq V_t$ for any $t\geq 0$, with $\lambda_{t+1}^2-\lambda_{t+1}=\lambda_t^2$ ($\lambda_0=0$) and $\beta_t=\frac{\lambda_t-1}{\lambda_{t+1}}$ ($x_{-1}=x_0$).

To do so, we computed $V_{t+1}-V_t$ and used (Smoothness) and (Convexity) to upper bound $f(x_{t+1}-f(y_t))$, $f(y_t)-f(x_t)$ and $f(y_t)-f_\star$.

1. Rewrite the same proof by using (Coco) instead of (Smoothness) and (Convexity).
2. Since (Coco) is stronger, you obtained $V_{t+1}-V_t\leq\Delta_t$ for a specific Δ. Prove that the sequence $V_t+\sum_{s=0}^{t-1}\Delta_s$ is non increasing.
3. Conclude on a convergence guarantee of the smallest observed gradient norm.