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 a natural choice for 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

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

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 non negative 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.

1 Solutions

Solution Exercice 1. :

1. Using (Smoothness),

$$\begin{aligned} f(x_{t+1}) &\leq f(x_t) + \langle \nabla f(x_t), x_{t+1} - x_t \rangle + \frac{L}{2}\|x_{t+1} - x_t\|^2 \\ &= f(x_t) + \langle \nabla f(x_t), -\gamma \nabla f(x_t) \rangle + \frac{L}{2}\| - \gamma \nabla f(x_t)\|^2 \\ &= f(x_t) - \gamma \left(1 - \frac{L\gamma}{2}\right) \|\nabla f(x_t)\|^2 \end{aligned}$$

2. Observing that $\gamma \left(1 - \frac{L\gamma}{2}\right) = \frac{1}{2L} - \frac{L}{2}(\gamma - \frac{1}{L})^2 \leq \frac{1}{2L}$, we have

$$\begin{aligned} f(x_{t+1}) &\leq f(x_t) - \frac{1}{2L}\|\nabla f(x_t)\|^2 + \frac{L}{2}\left(\gamma - \frac{1}{L}\right)^2 \|\nabla f(x_t)\|^2 \\ &\leq f(x_t) - \frac{1}{2L}\|\nabla f(x_t)\|^2, \end{aligned}$$

   reached for $\gamma = \frac{1}{L}$. Therefore this choice of step-size γ is optimal.

3. Using (Convexity), we have

$$f(x_t) - f_\star \leq \langle \nabla f(x_t), x_t - x_\star \rangle . \tag{8}$$

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

$$V_t = C_t(f(x_t) - f_\star) + \frac{L}{2}\|x_t - x_\star\|^2. \tag{9}$$

   Assuming V_t is non increasing,

$$C_t(f(x_t) - f_\star) \leq V_t \leq \cdots \leq V_t = \frac{L}{2}\|x_0 - x_\star\|^2. \tag{10}$$

   We conclude $f(x_t) - f_\star \leq \frac{L}{2}\frac{\|x_0 - x_\star\|^2}{C_t}$. Therefore, we want to have V_t non increasing for the largest possible C_t.

5. We want V_t non increasing. We then compute

$$\begin{aligned} V_{t+1} - V_t &= C_{t+1}(f(x_{t+1}) - f_\star) - C_t(f(x_t) - f_\star) + \frac{L}{2}\left(\|x_{t+1} - x_\star\|^2 - \|x_t - x_\star\|^2\right) \\ &= C_{t+1}(f(x_{t+1}) - f(x_t)) + (C_{t+1} - C_t)(f(x_t) - f_\star) + \frac{L}{2}\left(\|x_{t+1} - x_\star\|^2 - \|x_t - x_\star\|^2\right) \\ &\leq\ -\frac{C_{t+1}}{2L}\|\nabla f(x_t)\|^2 + (C_{t+1} - C_t)\langle \nabla f(x_t), x_t - x_\star \rangle - \langle \nabla f(x_t), x_t - x_\star \rangle + \frac{1}{2L}\|\nabla f(x_t)\|^2 \\ &\leq \frac{1 - C_{t+1}}{2L}\|\nabla f(x_t)\|^2 + (C_{t+1} - C_t - 1)\langle \nabla f(x_t), x_t - x_\star \rangle \end{aligned} \tag{11}$$

   Therefore, we verify $V_{t+1} - V_t \leq 0$ for sequences $(C_t)_{t\in\mathbb{N}}$ such that for all $t \geq 0$, $1 - C_{t+1} \leq 0$ and $C_{t+1} - C_t - 1 \leq 0$. Note the first inequality is equivalent to $\forall t \geq 1,\ C_t \geq 1$, while the second one leads to the growth constraint $\forall t \geq 0,\ C_{t+1} \leq C_t + 1$. Therefore, the largest such sequence verifies $\forall t \geq 0,\ C_t = t$.
   We conclude that $V_t = t(f(x_t) - f_\star) + \frac{L}{2}\|x_t - x_\star\|^2$ is non increasing and finally $\forall t \in \mathbb{N},\ f(x_t) - f_\star \leq \frac{L}{2}\frac{\|x_0 - x_\star\|^2}{t}$.

bonus : Rewriting (11) with $C_t = t$ gives

$$V_{t+1} - V_t \leq -\frac{t}{2L}\|\nabla f(x_t)\|^2.$$

Let's define $V_t' \triangleq V_t + \sum_{s=0}^{t-1} \frac{s}{2L}\|\nabla f(x_s)\|^2$.
We verify V' is non increasing since

$$V_{t+1}' - V_t' = V_{t+1} - V_t + \frac{t}{2L}\|\nabla f(x_t)\|^2 \leq 0.$$

And we conclude with

$$\frac{t(t-1)}{4L} \min_{0\leq s\leq t-1} \|\nabla f(x_s)\|^2 \leq \sum_{s=0}^{t-1} \frac{s}{2L}\|\nabla f(x_s)\|^2 \leq V_t' \leq V_0' = \frac{L}{2}\|x_0 - x_\star\|^2,$$

that $\min_{0\leq s\leq t-1} \|\nabla f(x_s)\|^2 \leq \frac{2L^2}{t(t-1)}\|x_0 - x_\star\|^2$.

Solution Exercice 2. :

1.

$$\begin{aligned}
&C_{t+1}(f(x_{t+1}) - f_\star) - C_t(f(x_t) - f_\star) \\
&\quad = C_{t+1}(f(x_{t+1}) - f(y_t)) + C_t(f(y_t) - f(x_t)) + (C_{t+1} - C_t)(f(y_t) - f_\star)) \\
&\quad \leq -\frac{C_{t+1}}{2L}\|\nabla f(y_t)\|^2 + C_t \langle \nabla f(y_t), y_t - x_t\rangle + (C_{t+1} - C_t)\langle \nabla f(y_t), y_t - x_\star\rangle \\
&\quad \leq -\frac{C_{t+1}}{2L}\|\nabla f(y_t)\|^2 + \langle \nabla f(y_t), C_{t+1}(y_t - x_\star) - C_t(x_t - x_\star)\rangle. \qquad (12)
\end{aligned}$$

2. Let's write (12)'s RHS as a difference of 2 squares.

$$\begin{aligned}
&-\frac{C_{t+1}}{2L}\|\nabla f(y_t)\|^2 + \langle \nabla f(y_t), C_{t+1}(y_t - x_\star) - C_t(x_t - x_\star)\rangle \\
&\quad = -\frac{LC_{t+1}}{2}\left[\left\|\frac{1}{L}\nabla f(y_t)\right\|^2 - 2\left\langle \frac{1}{L}\nabla f(y_t), (y_t - x_\star) - \frac{C_t}{C_{t+1}}(x_t - x_\star)\right\rangle\right] \\
&\quad = -\frac{LC_{t+1}}{2}\left[\left\|(y_t - x_\star) - \frac{C_t}{C_{t+1}}(x_t - x_\star) - \frac{1}{L}\nabla f(y_t)\right\|^2 - \left\|(y_t - x_\star) - \frac{C_t}{C_{t+1}}(x_t - x_\star)\right\|^2\right].
\end{aligned}$$

We now summarize writing everything in terms of the sequences x and λ.

$$\begin{aligned}
&C_{t+1}(f(x_{t+1}) - f_\star) - C_t(f(x_t) - f_\star) \\
&\quad \leq -\frac{L\lambda_{t+1}^2}{2}\left[\left\|(x_{t+1} - x_\star) - \frac{\lambda_t^2}{\lambda_{t+1}^2}(x_t - x_\star)\right\|^2 - \left\|(x_t - x_\star) + \beta_t(x_t - x_{t-1}) - \frac{\lambda_t^2}{\lambda_{t+1}^2}(x_t - x_\star)\right\|^2\right] \\
&\quad = -\frac{L}{2}\left[\left\|\lambda_{t+1}(x_{t+1} - x_\star) - \frac{\lambda_t^2}{\lambda_{t+1}}(x_t - x_\star)\right\|^2 - \left\|\left(\lambda_{t+1} - \frac{\lambda_t^2}{\lambda_{t+1}}\right)(x_t - x_\star) + \beta_t\lambda_{t+1}(x_t - x_{t-1})\right\|^2\right].
\end{aligned}$$

We can then conclude

$$C_{t+1}(f(x_{t+1}) - f_\star) + R_{t+1} \leq C_t(f(x_t) - f_\star) + R_t, \qquad (13)$$

with

$$R_{t+1} = \frac{L}{2}\left\|\lambda_{t+1}(x_{t+1}-x_\star) - \frac{\lambda_t^2}{\lambda_{t+1}}(x_t - x_\star)\right\|^2,$$
$$R_t = \frac{L}{2}\left\|\left(\lambda_{t+1} - \frac{\lambda_t^2}{\lambda_{t+1}}\right)(x_t - x_\star) + \beta_t\lambda_{t+1}(x_t - x_{t-1})\right\|^2.$$

Therefore, we need to have for all $t \geq 1$:

$$\lambda_t(x_t - x_\star) - \frac{\lambda_{t-1}^2}{\lambda_t}(x_{t-1} - x_\star) = \left(\lambda_{t+1} - \frac{\lambda_t^2}{\lambda_{t+1}} + \beta_t\lambda_{t+1}\right)(x_t - x_\star) - \beta_t\lambda_{t+1}(x_{t-1} - x_\star),$$

i.e.

$$\lambda_t = \left(\lambda_{t+1} - \frac{\lambda_t^2}{\lambda_{t+1}} + \beta_t\lambda_{t+1}\right),$$
$$\frac{\lambda_{t-1}^2}{\lambda_t} = \beta_t\lambda_{t+1}.$$

From the first line we get $\boxed{\beta_t = \frac{\lambda_t}{\lambda_{t+1}} + \frac{\lambda_t^2}{\lambda_{t+1}^2} - 1}$ and injecting it in the second line we obtain the recursion $\boxed{\frac{\lambda_{t-1}^2}{\lambda_t^2} = 1 + \frac{\lambda_t}{\lambda_{t+1}} - \frac{\lambda_{t+1}}{\lambda_t}}$.

The latter can be rearranged by substituting 1 on both sides, giving $\frac{\lambda_{t-1}^2 - \lambda_t^2}{\lambda_t^2} = \frac{\lambda_t^2 - \lambda_{t+1}^2}{\lambda_t\lambda_{t+1}}$.
We conclude that $\frac{\lambda_t^2 - \lambda_{t+1}^2}{\lambda_{t+1}} = \frac{\lambda_{t-1}^2 - \lambda_t^2}{\lambda_t} = \frac{\lambda_0^2 - \lambda_1^2}{\lambda_1} = -\lambda_1$ since $\lambda_0 = 0$.
This can also be written as $\boxed{\left(\frac{\lambda_{t+1}}{\lambda_1}\right)^2 - \frac{\lambda_{t+1}}{\lambda_1} = \left(\frac{\lambda_t}{\lambda_1}\right)^2}$ and $\boxed{\beta_t = \frac{\lambda_t/\lambda_1 - 1}{\lambda_{t+1}/\lambda_1}}$.
We want to maximize λ_1 such that $V_t \leq \frac{L}{2}\|x_0 - x_\star\|^2$.
We have :

$$\begin{aligned}
V_t &\leq V_{t-1} \\
&\leq V_1 \\
&= \lambda_1^2(f(x_1) - f_\star) + \frac{L}{2}\|x_0 - x_\star + \lambda_1(x_1 - x_0)\|^2 \\
&\leq \lambda_1^2\left(f(x_0) - f_\star - \frac{1}{2L}\|\nabla f(x_0)\|^2\right) + \frac{L}{2}\|x_0 - x_\star - \frac{\lambda_1}{L}\nabla f(x_0)\|^2 \\
&= \lambda_1^2(f(x_0) - f_\star) + \frac{L}{2}\|x_0 - x_\star\|^2 - \lambda_1\langle\nabla f(x_0), x_0 - x_\star\rangle \\
&= (\lambda_1^2 - \lambda_1)\langle\nabla f(x_0), x_0 - x_\star\rangle + \frac{L}{2}\|x_0 - x_\star\|^2
\end{aligned}$$
$$\leq \frac{L}{2}\|x_0 - x_\star\|^2$$

is valid for any $\lambda_1 \leq 1$.

We conclude that with $\lambda_{t+1}^2 - \lambda_{t+1} = \lambda_t^2$ and $\beta_t = \frac{\lambda_t - 1}{\lambda_{t+1}}$,

$$f(x_t) - f_\star \leq \frac{L}{2} \frac{\|x_0 - x_\star\|^2}{\lambda_t^2}. \tag{14}$$

3. $\lambda_{t+1}^2 - \lambda_{t+1} = \lambda_t^2$ can also be written as $\lambda_{t+1} = \frac{1}{2} + \sqrt{\lambda_t^2 + \frac{1}{4}}$.
   Then, $\lambda_{t+1} \geq \frac{1}{2} + \lambda_t$, hence $\lambda_t \geq \frac{t}{2}$.
   This is sufficient to conclude that $f(x_t) - f_\star \leq 2L\frac{\|x_0 - x_\star\|^2}{t^2}$.
   Moreover, since $\lambda_t \to 0$, we can show that $\lambda_{t+1} - \lambda_t \to \frac{1}{2}$, and then $\lambda_t \sim \frac{t}{2}$.

Solution Exercice 3. :

1. If f is said convex and L-smooth, then using (Convexity) and (Smoothness), we get $\forall (x, y, z)$:

$$f(y) + \langle \nabla f(y), x - y\rangle \leq \ f(x) \leq f(z) + \langle \nabla f(z), x - z\rangle + \frac{L}{2}\|x - z\|^2. \tag{15}$$

   and thus (3). Reciprocally, if $\forall (x, y, z)$

$$f(y) + \langle \nabla f(y), x - y\rangle \leq f(z) + \langle \nabla f(z), x - z\rangle + \frac{L}{2}\|x - z\|^2. \tag{16}$$

   then with $x = y$, we get, $\forall (x, z)$:

$$f(x) \leq f(z) + \langle \nabla f(z), x - z\rangle + \frac{L}{2}\|x - z\|^2. \tag{17}$$

   which is (Smoothness) and with $x = z$, we get, $\forall (x, y)$:

$$f(y) + \langle \nabla f(y), x - y\rangle \leq f(x). \tag{18}$$

   which is (Convexity).

2. By (3), f is convex and L-smooth if and only if, $\forall (x, y, z)$

$$0 \leq f(z) - f(y) + \langle \nabla f(y), y\rangle - \langle \nabla f(z), z\rangle + \langle \nabla f(z) - \nabla f(y), x\rangle + \frac{L}{2}\|x - z\|^2. \tag{19}$$

   thus if and only if, $\forall (y, z)$

$$0 \leq f(z) - f(y) + \langle \nabla f(y), y - z\rangle + \underbrace{\min_{x\in\mathbb{R}^d} \left(\langle \nabla f(z) - \nabla f(y), x - z\rangle + \frac{L}{2}\|x - z\|^2\right)}_{= -\frac{1}{2L}\|\nabla f(z) - \nabla f(y)\|^2}.$$

   which gives (4).

3. (5) summed with the same inequality with y, z permuted gives (6).
   For the other direction :
   (a) (6) implies that ∇f is Lipschitz by Cauchy Schwartz.
   (b) $g'(t) = \langle \nabla f(\theta + t(\eta - \theta)), \theta - \eta\rangle$ and thus for all $t > 0$, we have $g'(t) - g'(0) = \frac{1}{t}\langle \nabla f(\theta + t(\eta - \theta)) - \nabla f(\theta), t(\theta - \eta)\rangle \geq 0$. Writing $g(1) = g(0) + \int_0^1 g'(t)\mathrm{d}t \geq g(0) + g'(0)$ we get (Convexity)

Remark : another solution, is too show that $g(x) := \frac{L}{2}\|x\|^2 - f(x)$ is L-smooth and to conclude that f is thus convex.

To do so, we observe that :

$$\frac{1}{L}\|\nabla g(y) - \nabla g(z)\|^2 \leq \langle \nabla g(y) - \nabla g(z), y - z\rangle \tag{20}$$

$$\Leftrightarrow \quad \frac{1}{L}\|L(y-z) - (\nabla f(y) - \nabla f(z))\|^2 \leq \langle L(y-z) - (\nabla f(y) - \nabla f(z)), y - z\rangle . \tag{21}$$

$$\Leftrightarrow \quad \frac{1}{L}\|\nabla f(y) - \nabla f(z)\|^2 \leq \langle \nabla f(y) - \nabla f(z), y - z\rangle . \tag{22}$$

Thus (6) for f is equivalent to (6) for g! And as remarked above, (6) implies L-smoothness.

4. We can do the exact same derivation for strongly convex functions adding curvature to (Convexity). Another way it to notice that f is L-smooth and μ-strongly convex if and only if $f - \frac{\mu}{2}\|x - x_\star\|^2$ is $L - \mu$-smooth and convex. Applying (Coco) to $f - \frac{\mu}{2}\|x - x_\star\|^2$ therefore answers the question. And the obtained inequality is of course the same using the 2 approaches.

Solution Exercice 4. :

1. First we compute

$$\begin{aligned}
V_{t+1} - V_t \leq\ & \lambda_{t+1}^2(f(x_{t+1}) - f_\star) - \lambda_t^2(f(x_t) - f_\star) \\
& + \frac{L}{2}\|\lambda_{t+1}(x_{t+1} - x_\star) + (1-\lambda_{t+1})(x_t - x_\star)\|^2 - \frac{L}{2}\|\lambda_t(x_t - x_\star) + (1-\lambda_t)(x_{t-1} - x_\star)\|^2 \\
=\ & \lambda_{t+1}^2(f(x_{t+1}) - f(y_t)) + \lambda_{t+1}(f(y_t) - f_\star) + \lambda_t^2(f(y_t) - f(x_t)) \\
& + \frac{L}{2}\|\lambda_{t+1}(x_{t+1} - x_\star) + (1-\lambda_{t+1})(x_t - x_\star)\|^2 - \frac{L}{2}\|\lambda_t(x_t - x_\star) + (1-\lambda_t)(x_{t-1} - x_\star)\|^2 \\
\leq\ & -\frac{\lambda_{t+1}^2}{2L}\|\nabla f(y_t)\|^2 + \lambda_{t+1}\langle \nabla f(y_t), y_t - x_\star\rangle + \lambda_t^2\langle \nabla f(y_t), y_t - x_t\rangle \\
& + \frac{L}{2}\|\lambda_{t+1}(x_{t+1} - x_\star) + (1-\lambda_{t+1})(x_t - x_\star)\|^2 - \frac{L}{2}\|\lambda_t(x_t - x_\star) + (1-\lambda_t)(x_{t-1} - x_\star)\|^2
\end{aligned}$$

Then, by rearranging terms and using (NAG), we conclude

$$V_{t+1} - V_t \leq -\frac{1}{2L}\left[\lambda_{t+1}^2\|\nabla f(x_{t+1})\|^2 + \lambda_{t+1}\|\nabla f(y_t)\| + \lambda_t^2\|\nabla f(y_t) - \nabla f(x_t)\|^2\right].$$

Which is the wanted inequality $V_{t+1} - V_t \leq -\Delta_t$, with

$$\Delta_t = \frac{1}{2L}\left[\lambda_{t+1}^2\|\nabla f(x_{t+1})\|^2 + \lambda_{t+1}\|\nabla f(y_t)\| + \lambda_t^2\|\nabla f(y_t) - \nabla f(x_t)\|^2\right]. \tag{23}$$

2. We compute the difference

$$\left(V_{t+1} + \sum_{s=0}^{t}\Delta_s\right) - \left(V_t + \sum_{s=0}^{t-1}\Delta_s\right) = V_{t+1} - V_t + \Delta_t \leq 0.$$

3. We have that for any $t \geq 0$,

$$\frac{1}{2L}\sum_{s=0}^{t-1} \lambda_{s+1}^2 \|\nabla f(x_{s+1})\|^2 \leq V_t + \sum_{s=0}^{t-1} \Delta_s \leq V_0 = \frac{L}{2}\|x_0 - x_\star\|^2. \tag{24}$$

We conclude that

$$\min_{0\leq s\leq t-1} \|\nabla f(x_{s+1})\|^2 \leq \frac{L^2\|x_0 - x_\star\|^2}{\sum_{s=0}^{t-1} \lambda_{s+1}^2} = O\left(\frac{1}{t^3}\right). \tag{25}$$