Supplementary material for "A Lyapunov analysis for accelerated gradient methods: from deterministic to stochastic case"

A Abstract Lyapunov analysis

A.1 Proof of Lemma 2.6

By (8) and using (FEP) and (7), we have

$$\begin{aligned}
&E(t_{k+1}, z_{k+1}) - E(t_k, z_k)\\
&\quad\leq \partial_t E(t_k, z_k)(t_{k+1} - t_k)\\
&\quad+ \langle \nabla E(t_k, z_k), z_{k+1} - z_k\rangle + \frac{L_E}{2}|z_{k+1} - z_k|^2\\
&\quad\leq h_k\Big(\partial_t E(t_k, z_k)\\
&\quad+ \langle \nabla E(t_k, z_k), g(t_k, z_k, \nabla f(z_k))\rangle\Big)\\
&\quad+ h_k\langle \nabla E(t_k, z_k), g_2(t_k, z_k)e_k\rangle\\
&\quad+ \frac{L_E h_k^2}{2}|g(t_k, z_k, \nabla f(z_k) + e_k)|^2\\
&\quad\leq -h_k r_E E(t_k, z_k)\\
&\quad- h_k a_E |g(t_k, z_k, \nabla f(z_k))|^2\\
&\quad+ h_k\langle \nabla E(t_k, z_k), g_2(t_k, z_k)e_k\rangle\\
&\quad+ \frac{L_E h_k^2}{2}|g(t_k, z_k, \nabla f(z_k) + e_k)|^2\\
&\quad\leq -h_k r_E E(t_k, z_k)\\
&\quad- h_k\left(a_E - \frac{L_E h_k}{2}\right)|g(t_k, z_k, \nabla f(z_k))|^2\\
&\quad+ h_k\langle \nabla E(t_k, z_k), g_2(t_k, z_k)e_k\rangle\\
&\quad+ L_E h_k^2\langle g(t_k, z_k, \nabla f(z_k)), g_2(t_k, z_k)e_k\rangle\\
&\quad+ \frac{L_E h_k^2}{2}|g_2(t_k, z_k)e_k|^2,
\end{aligned}$$

which concludes the proof.

A.2 Proof of Proposition 2.7

The proof of Proposition 2.7 can be done by induction and is an adaptation of the one of Oberman and Prazeres [2019]. Indeed, the initialization, $k = 0$, of E_k is trivial and for all $k \geq 1$, from (12), we have

$$\begin{aligned}
&\mathbb{E}[E(t_{k+1}, z_{k+1})]\\
&\qquad\leq (1 - h_k r_E)E(t_k, z_k) + \frac{h_k^2 L_E g_2(t_k, z_k)^2\sigma^2}{2}
\end{aligned}$$

and by definition of h_k, α, and using the induction assumption,

$$\begin{aligned}
\mathbb{E}[E(t_{k+1}, z_{k+1})] &\leq \left(1 - \frac{2}{k + \alpha^{-1}E_0^{-1}}\right)\frac{1}{\alpha(k + \alpha^{-1}E_0^{-1})}\\
&\quad+ \frac{1}{\alpha(k + \alpha^{-1}E_0^{-1})^2}\\
&\leq \frac{1}{\alpha(k + \alpha^{-1}E_0^{-1})}\\
&\quad- \frac{1}{\alpha(k + \alpha^{-1}E_0^{-1})^2}\\
&\leq \frac{1}{\alpha(k + 1 + \alpha^{-1}E_0^{-1})},
\end{aligned}$$

which concludes the proof.

A.3 Proof of Proposition 2.8

First, since $t_k = \sum_{i=1}^k \frac{c}{i^\alpha}$ ($t_0 = h_0 = 0$), we need $\alpha < 1$. Summing (14) over from 0 to $k - 1$, we obrain

$$\mathbb{E}[E(t_k, x_k)] \leq E_0 + \frac{\sigma^2}{2}\sum_{i=1}^{k-1} h_i^2(a_1 + a_2 t_i + a_3 t_i^2).$$

Now we want to prove that $\mathbb{E}[E(t_k, x_k)]$ is bounded.

- Term $\frac{a_1\sigma^2}{2}\sum_{i=1}^{k-1} h_i^2$: By comparison series-integral, we have
$$\begin{aligned}
\sum_{i=1}^{k-1}\frac{c^2}{i^{2\alpha}} &\leq c^2\left(1 + \int_1^k \frac{1}{t^{2\alpha}}\,dt\right)\\
&\leq c^2\left(1 + \left[\frac{t^{1-2\alpha}}{1 - 2\alpha}\right]_1^k\right)\\
&\leq \frac{2\alpha c^2}{2\alpha - 1},
\end{aligned}$$
if $\alpha > \frac{1}{2}$.
Then,
$$\frac{a_1\sigma^2}{2}\sum_{i=1}^{k-1} h_i^2 \leq \frac{c^2\alpha a_1\sigma^2}{2\alpha - 1}.$$
- Term $\frac{a_2\sigma^2}{2}\sum_{i=1}^{k-1} h_i^2 t_i$: First, we have
$$t_i = \sum_{j=1}^i \frac{c}{j^\alpha} \leq \int_0^i \frac{c}{t^\alpha}\,dt = \frac{c i^{1-\alpha}}{1 - \alpha}.$$

Then,

$$\begin{aligned}\sum_{i=1}^{k-1} h_i^2 t_i &\leq \frac{c^3}{1-\alpha}\sum_{i=1}^{k-1}\frac{1}{i^{3\alpha-1}}\\ &\leq \frac{c^3}{1-\alpha}\left(1+\frac{1}{3\alpha-2}\right)\\ &= \frac{c^3(3\alpha-1)}{(1-\alpha)(3\alpha-2)},\end{aligned}$$

if $\alpha > \frac{2}{3}$. Now, if $\alpha > \frac{2}{3}$. Now, if $\alpha = \frac{2}{3}$,

$$\begin{aligned}\sum_{i=1}^{k-1} h_i^2 t_i &\leq 3c^3\sum_{i=1}^{k-1}\frac{1}{i}\\ &\leq \frac{c^3}{1-\alpha}\left(1+\log(k)\right).\end{aligned}$$

Then, we have

$$\frac{a_2\sigma^2}{2}\sum_{i=1}^{k-1} h_i^2 t_i \leq \begin{cases}\frac{c^3(3\alpha-1)a_2\sigma^2}{2(1-\alpha)(3\alpha-2)}, & \alpha > \frac{2}{3}\\ \frac{3c^3 a_2\sigma^2}{2}(1+\log(k)), & \alpha = \frac{2}{3}.\end{cases}$$

- Term $\frac{a_3\sigma^2}{2}\sum_{i=1}^{k-1} h_i^2 t_i^2$:

$$t_i^2 = \left(\sum_{j=1}^{i}\frac{c}{j^\alpha}\right)^2 \leq \frac{c^2 i^{2-2\alpha}}{(1-\alpha)^2}.$$

Then,

$$\begin{aligned}\sum_{i=1}^{k-1} h_i^2 t_i^2 &\leq \frac{c^4}{(1-\alpha)^2}\sum_{i=1}^{k-1}\frac{1}{i^{4\alpha-2}}\\ &\leq \frac{c^4}{(1-\alpha)^2}\left(1+\frac{1}{4\alpha-3}\right)\\ &= \frac{c^4(4\alpha-2)}{(1-\alpha)^2(4\alpha-3)},\end{aligned}$$

if $\alpha > \frac{3}{4}$, and, if $\alpha = \frac{3}{4}$,

$$\begin{aligned}\sum_{i=1}^{k-1} h_i^2 t_i^2 &\leq 16c^4\sum_{i=1}^{k-1}\frac{1}{i}\\ &\leq 16c^4\left(1+\log(k)\right).\end{aligned}$$

Then, we have

$$\frac{a_3\sigma^2}{2}\sum_{i=1}^{k-1} h_i^2 t_i^2 \leq \begin{cases}\frac{a_3 c^4(4\alpha-2)\sigma^2}{2(1-\alpha)^2(4\alpha-3)}, & \alpha > \frac{3}{4}\\ \frac{16c^4 a_3\sigma^2}{2}(1+\log(k)), & \alpha = \frac{3}{4}.\end{cases}$$

So now, we have two cases:

- Case $a_1, a_2, b_1 > 0$, $a_3 = b_2 = 0$: In that case, we have shown that

$$\begin{aligned}&\mathbb{E}[E(t_k, x_k)]\\ &\leq E_0 + \begin{cases}\frac{c^2\alpha a_1\sigma^2}{2\alpha-1} + \frac{c^3(3\alpha-1)a_2\sigma^2}{2(1-\alpha)(3\alpha-2)}, & \alpha > \frac{2}{3}\\ 2c^2 a_1\sigma^2 + \frac{3c^3 a_2\sigma^2}{2}(1+\log(k)), & \alpha = \frac{2}{3}.\end{cases}\end{aligned}$$

In addition, by (15),

$$\mathbb{E}[E(t_k, x_k)] \geq b_1 t_k(\mathbb{E}[f(x_k)] - f^*),$$

combine with

$$t_k \geq \frac{c}{1-\alpha}(k^{1-\alpha} - 1),$$

we obtain

$$\begin{aligned}&\mathbb{E}[f(x_k)] - f^*\\ &\leq \begin{cases}\frac{\frac{1-\alpha}{c}E_0 + \left(\frac{a_1 c(1-\alpha)\alpha}{2\alpha-1} + \frac{a_2 c^2(3\alpha-1)}{2(3\alpha-2)}\right)\sigma^2}{b_1(k^{1-\alpha}-1)}, & \alpha \in \left(\frac{2}{3}, 1\right)\\ \frac{\frac{1}{3c}E_0 + \left(\frac{2a_1 c}{3} + \frac{a_2 c^2}{2}(1+\log(k))\right)\sigma^2}{b_1(k^{1/3}-1)}, & \alpha = \frac{2}{3}.\end{cases}\end{aligned}$$

- Case $a_1 = a_2 = b_1 = 0$, $a_3 > 0, b_2 > 0$: In that case, we have shown that

$$\begin{aligned}&\mathbb{E}[E(t_k, x_k)]\\ &\leq E_0 + \begin{cases}\frac{a_3 c^4(4\alpha-2)\sigma^2}{2(1-\alpha)^2(4\alpha-3)}, & \alpha > \frac{3}{4}\\ \frac{16c^4 a_3\sigma^2}{2}(1+\log(k)), & \alpha = \frac{3}{4}.\end{cases}\end{aligned}$$

By (15),

$$\mathbb{E}[E(t_k, x_k)] \geq b_2 t_k^2(\mathbb{E}[f(x_k)] - f^*),$$

combine with

$$t_k^2 \geq \frac{c^2}{(1-\alpha)^2}(k^{1-\alpha} - 1)^2,$$

we obtain

$$\begin{aligned}&\mathbb{E}[f(x_k)] - f^*\\ &\leq \begin{cases}\frac{\frac{(1-\alpha)^2}{c^2}E_0 + \frac{a_3 c^2(4\alpha-2)\sigma^2}{2(4\alpha-3)}}{b_2(k^{1-\alpha}-1)^2}, & \alpha \in \left(\frac{3}{4}, 1\right)\\ \frac{\frac{1}{16c^2}E_0 + \frac{a_3 c^2\sigma^2}{2}(1+\log(k))}{b_2(k^{1/4}-1)^2}, & \alpha = \frac{3}{4},\end{cases}\end{aligned}$$

which concludes the proof.

B Gradient descent

B.1 Proof of Proposition 3.1

- In the convex case, we first start to look for (7):

$$\begin{aligned}&\partial_t E^c(t,z) - \nabla E^c(t,z)\nabla f(z)\\ &\qquad = f(z) - f^* - \langle t\nabla f(z) + z - x^*, \nabla f(z)\rangle\\ &\qquad = f(z) - f^* - \langle z - x^*, \nabla f(z)\rangle - t|\nabla f(z)|^2\\ &\qquad \leq -t|\nabla f(z)|^2,\end{aligned}$$

by convexity, which gives $r_{E^c} = 0$ and $a_{E^c} = t$. Now, by 1-convexity of the quadratic term and L-smoothness of f,

$$\begin{aligned}
E^c(t_{k+1}, z_{k+1}) - E^c(t_k, z_k) \\
&\le t_{k+1}(f(z_k) - f^* + \langle \nabla f(z_k), z_{k+1} - z_k\rangle \\
&+ \frac{L}{2}|z_{k+1} - z_k|^2) - t_k(f(z_k) - f^*) \\
&+ \frac{1}{2}|z_{k+1} - x^*|^2 - \frac{1}{2}|z_k - x^*|^2 \\
&\le \langle t_k \nabla f(z_k) + z_k - x^*, z_{k+1} - z_k\rangle \\
&+ \frac{Lt_{k+1} + 1}{2}|z_{k+1} - z_k|^2 \\
&+ (t_{k+1} - t_k)(f(z_k) - f^*) \\
&+ (t_{k+1} - t_k)\langle \nabla f(z_k), z_{k+1} - z_k\rangle \\
&\le (t_{k+1} - t_k)\partial_t E^c(t_k, z_k) \\
&+ \langle \nabla E^c(t_k, z_k), z_{k+1} - z_k\rangle \\
&+ \frac{Lt_{k+1} + 1}{2}|z_{k+1} - z_k|^2,
\end{aligned}$$

since, by (16),

$$(t_{k+1} - t_k)\langle \nabla f(z_k), z_{k+1} - z_k\rangle \le 0.$$

Then $L_{E^c} = Lt_{k+1} + 1$.

- In the strongly convex case,

$$\begin{aligned}
\partial_t E^{sc}(z) - \nabla E^{sc}(z)\nabla f(z) \\
&= -\langle \nabla f(z) + \mu(z - x^*), \nabla f(z)\rangle \\
&= -\mu\langle z - x^*, \nabla f(z)\rangle - |\nabla f(z)|^2 \\
&\le -\mu\left(f(z) - f^* - \frac{\mu}{2}|z - x^*|^2\right) \\
&- |\nabla f(z)|^2,
\end{aligned}$$

by strong convexity and then $r_{E^{sc}} = \mu$ and $a_{E^{sc}} = 1$. Concerning (8), since E^{sc} is time independent, (8) is equivalent to L-smoothness condition which gives $L_{E^{sc}} = L + \mu$.

C Accelerated rate: convex case

C.1 ODE and derivation of Nesterov's method

Derivation of (H-ODE) Solve for v in the first line of (1st-ODE)

$$v = \frac{t}{2}(\dot{x} + \frac{1}{\sqrt{L}}\nabla f(x)) + x$$

differentiate to obtain

$$\dot{v} = \frac{1}{2}(\dot{x} + \frac{1}{\sqrt{L}}\nabla f(x)) + \frac{t}{2}(\ddot{x} + \frac{1}{\sqrt{L}}D^2 f(x)\cdot\dot{x}) + \dot{x}.$$

Insert into the second line of (1st-ODE)

$$\frac{1}{2}(\dot{x} + \frac{1}{\sqrt{L}}\nabla f(x)) + \frac{t}{2}(\ddot{x} + \frac{1}{\sqrt{L}}D^2 f(x)\cdot\dot{x}) + \dot{x} = -\frac{t}{2}\nabla f(x).$$

Simplify to obtain (H-ODE).

Proof of Proposition 4.2 The system (FE-C) with a constant time step $h_k = \frac{1}{\sqrt{L}}$ gives $t_k = h(k+2) = \frac{k+2}{\sqrt{L}}$ and

$$\begin{cases} x_{k+1} - x_k &= \frac{2}{k+2}(v_k - x_k) - \frac{1}{L}\nabla f(y_k) \\ v_{k+1} - v_k &= -\frac{k+2}{2L}\nabla f(y_k) \end{cases}$$

Eliminate the variable v_k using the definition of y_k in (FE-C) to obtain (C-Nest).

C.2 Lyapunov analysis for (1st-ODE)

Proposition C.1. *Suppose f is convex and L-smooth. Let x, v be solutions of* (1st-ODE) *and let $E^{ac,c}$ be given by* (22). *Then $E^{ac,c}$ is a continuous Lyapunov function with $r_{E^{ac,c}} = 0$ and $a_{E^{ac,c}} = \frac{t^2}{\sqrt{L}}$ (with gap $t^2|\nabla f(x)|^2$) i.e.*

$$\frac{d}{dt}E^{ac,c}(t, x(t), v(t); 0) \le -\frac{t^2}{\sqrt{L}}|\nabla f(x)|^2.$$

Corollary C.2. *Let f be a convex and L-smooth function. Let $(x(t), v(t))$ be a solution to* (1st-ODE), *then for all $t > 0$,*

$$f(x(t)) - f^* \le \frac{2}{t^2}|v_0 - x^*|^2.$$

Proof of Proposition C.1. First, by definition of $E^{ac,c}$, we have

$$\begin{aligned}
\frac{d}{dt}E^{ac,c}(t, x(t), v(t)) &\le 2t(f(x) - f^*) + t^2\langle \nabla f(x), \dot{x}\rangle \\
&+ 4\langle v - x^*, \dot{v}\rangle \\
&\le 2t(f(x) - f^*) + 2t\langle \nabla f(x), v - x\rangle \\
&- \frac{t^2}{\sqrt{L}}|\nabla f(x)|^2 \\
&- 2t\langle v - x^*, \nabla f(x)\rangle \\
&\le 2t(f(x) - f^* - \langle x - x^*, \nabla f(x)\rangle) \\
&- \frac{t^2}{\sqrt{L}}|\nabla f(x)|^2.
\end{aligned}$$

The proof is concluded by convexity,

$$f(x) - f^* - \langle x - x^*, \nabla f(x)\rangle \le 0.$$

□

C.3 Classical inequality in the perturbed case

Before proving Proposition 4.4, using the convexity and the L-smoothness of f, we prove a gneralization of the classical inequality obtained in Attouch et al. [2016] or Su et al. [2014] in the case $e_k = 0$:

$$\begin{aligned} t_k^2(f(x_{k+1} - f^*) - t_{k-1}^2(f(x_k - f^*) &\leq -h^2(f(x_k) - f^*) \\ &+ 2ht_k\langle \nabla f(y_k), v_k - x^* \rangle \\ &- \left(\frac{1}{\sqrt{L}} - \frac{h}{2}\right) ht_k^2 |\nabla f(y_k)|^2 \\ &- \frac{ht_k^2}{\sqrt{L}}\langle \nabla f(y_k), e_k\rangle \\ &+ h^2 t_k^2 \left\langle \nabla f(y_k) + \frac{e_k}{2}, e_k \right\rangle. \end{aligned} \tag{30}$$

The proof is a perturbed version of the one in Beck and Teboulle [2009], Su et al. [2014], Attouch et al. [2016]. First we prove the following inequality:

Lemma C.3. *Assume f is a convex, L-soothness function. For all x, y, z, f satisfies*

$$f(z) \leq f(x) + \langle \nabla f(y), z - x\rangle + \frac{L}{2}|z - y|^2. \tag{31}$$

Proof. By L-smoothness,

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

and since f is convex,

$$f(y) - f(x) \leq \langle \nabla f(y), y - x\rangle.$$

We conclude the proof comibning these two inequalities. □

Now apply inequality (31) at $(x, y, z) = (x_k, y_k, x_{k+1})$:

$$\begin{aligned} &f(x_{k+1}) \\ &\quad \leq f(x_k) + \langle \nabla f(y_k), x_{k+1} - x_k\rangle + \frac{L}{2}|x_{k+1} - y_k|^2 \\ &\quad \leq f(x_k) + \frac{2h}{t_k}\langle \nabla f(y_k), v_k - x_k\rangle - \frac{h}{\sqrt{L}}|\nabla f(y_k)|^2 \\ &\quad - \frac{h}{\sqrt{L}}\langle \nabla f(y_k), e_k\rangle + \frac{h^2}{2}|\nabla f(y_k) + e_k|^2, \end{aligned}$$

and then

$$\begin{aligned} &f(x_{k+1}) \\ &\leq f(x_k) + \frac{2h}{t_k}\langle \nabla f(y_k), v_k - x_k\rangle - \left(\frac{1}{\sqrt{L}} - \frac{h}{2}\right) h|\nabla f(y_k)|^2 \\ &\quad - \frac{h}{\sqrt{L}}\langle \nabla f(y_k), e_k\rangle + h^2 \left\langle \nabla f(y_k) + \frac{e_k}{2}, e_k \right\rangle. \end{aligned} \tag{32}$$

If we apply (31) also at $(x, y, z) = (x^*, y_k, x_{k+1})$ we obtain

$$\begin{aligned} &f(x_{k+1}) \\ &\quad \leq f^* + \langle \nabla f(y_k), x_{k+1} - x^*\rangle + \frac{L}{2}|x_{k+1} - y_k|^2 \\ &\quad \leq f^* + \langle \nabla f(y_k), y_k - x^*\rangle - \frac{h}{\sqrt{L}}|\nabla f(y_k)|^2 \\ &\quad - \frac{h}{\sqrt{L}}\langle \nabla f(y_k), e_k\rangle + \frac{h^2}{2}|\nabla f(y_k) + e_k|^2 \end{aligned}$$

then,

$$\begin{aligned} &f(x_{k+1}) \\ &\leq f^* + \langle \nabla f(y_k), y_k - x^*\rangle - \left(\frac{1}{\sqrt{L}} - \frac{h}{2}\right) h|\nabla f(y_k)|^2 \\ &\quad - \frac{h}{\sqrt{L}}\langle \nabla f(y_k), e_k\rangle + h^2 \left\langle \nabla f(y_k) + \frac{e_k}{2}, e_k \right\rangle. \end{aligned} \tag{33}$$

Summing $\left(1 - \frac{2h}{t_k}\right)$(32) and $\frac{2h}{t_k}$(33), we have

$$\begin{aligned} f(x_{k+1}) - f^* \leq &\left(1 - \frac{2h}{t_k}\right)(f(x_k) - f^*) \\ &+ \frac{2h}{t_k}\langle \nabla f(y_k), v_k - x^*\rangle \\ &- \left(\frac{1}{\sqrt{L}} - \frac{h}{2}\right) h|\nabla f(y_k)|^2 \\ &- \frac{h}{\sqrt{L}}\langle \nabla f(y_k), e_k\rangle \\ &+ h^2 \left\langle \nabla f(y_k) + \frac{e_k}{2}, e_k \right\rangle. \end{aligned}$$

Then,

$$\begin{aligned} t_k^2(f(x_{k+1}) - f^*) &\leq (t_k - 2h)\, t_k(f(x_k) - f^*) \\ &+ 2ht_k\langle \nabla f(y_k), v_k - x^*\rangle \\ &- \left(\frac{1}{\sqrt{L}} - \frac{h}{2}\right) ht_k^2|\nabla f(y_k)|^2 \\ &- \frac{ht_k^2}{\sqrt{L}}\langle \nabla f(y_k), e_k\rangle \\ &+ h^2 t_k^2 \left\langle \nabla f(y_k) + \frac{e_k}{2}, e_k \right\rangle \\ &\leq \left(t_{k-1}^2 - h^2\right)(f(x_k) - f^*) \\ &+ 2ht_k\langle \nabla f(y_k), v_k - x^*\rangle \\ &- \left(\frac{1}{\sqrt{L}} - \frac{h}{2}\right) ht_k^2|\nabla f(y_k)|^2 \\ &- \frac{ht_k^2}{\sqrt{L}}\langle \nabla f(y_k), e_k\rangle \\ &+ h^2 t_k^2 \left\langle \nabla f(y_k) + \frac{e_k}{2}, e_k \right\rangle, \end{aligned}$$

which concludes the proof.

C.4 Proof of Proposition 4.4

By defintion of v_{k+1}, we have

$$\begin{aligned}
2|v_{k+1} - x^*|^2 - 2|v_k - x^*|^2 \\
&= -2ht_k\langle v_k - x^*, \nabla f(y_k) + e_k\rangle \\
&+ \frac{h^2 t_k^2}{2}|\nabla f(y_k) + e_k|^2 \\
&= -2ht_k\langle v_k - x^*, \nabla f(y_k) \\
&+ \frac{h^2 t_k^2}{2}|\nabla f(y_k)|^2 \\
&- 2ht_k\langle v_k - x^*, e_k\rangle \\
&+ h^2 t_k^2 \left\langle \nabla f(y_k) + \frac{e_k}{2}, e_k \right\rangle.
\end{aligned}$$

Therefore, combining it with (30) from Appendix C.3, we obtain

$$\begin{aligned}
E_{k+1}^{ac,c} - E_k^{ac,c} \leq -h^2(f(x_k) - f^*) \\
- \left(\frac{1}{\sqrt{L}} - h\right) ht_k^2 |\nabla f(y_k)|^2 \\
- 2ht_k\langle v_k - x^* - \frac{t_k}{\sqrt{L}}\nabla f(y_k), e_k\rangle \\
+ 2h^2 t_k^2 \left\langle \nabla f(y_k) + \frac{e_k}{2}, e_k \right\rangle,
\end{aligned}$$

and (24) is proved.

D Accelerated rate: strongly convex case

D.1 ODE and derivation of Nesterov's method

Equivalence between (1st-ODE-SC) **and** (H-ODE-SC) Solve for v in the first line of (1st-ODE-SC)

$$v = \frac{1}{\sqrt{\mu}}(\dot{x} + \frac{1}{\sqrt{L}}\nabla f(x)) + x$$

differentiate to obtain

$$\dot{v} = \frac{1}{\sqrt{\mu}}(\ddot{x} + \frac{1}{\sqrt{L}}D^2 f(x)\cdot\dot{x}) + \dot{x}.$$

Insert into the second line of (1st-ODE-SC)

$$\frac{1}{\sqrt{\mu}}(\ddot{x} + \frac{1}{\sqrt{L}}D^2 f(x)\cdot\dot{x}) + \dot{x} = -\dot{x} - \left(\frac{1}{\sqrt{L}} + \frac{1}{\sqrt{\mu}}\right)\nabla f(x).$$

Simplify to obtain (H-ODE-SC).

Proof of Proposition 5.3 (FE-SC) with $h_k = 1/\sqrt{L}$ becomes

$$\begin{cases}
x_{k+1} - x_k &= \frac{\sqrt{C_f^{-1}}}{1+\sqrt{C_f^{-1}}}(v_k - x_k) - \frac{1}{L}\nabla f(y_k) \\
v_{k+1} - v_k &= \frac{\sqrt{C_f^{-1}}}{1+\sqrt{C_f^{-1}}}(x_k - v_k) - \frac{1}{\sqrt{L\mu}}\nabla f(y_k)
\end{cases}$$

Eliminate the variable v_k using the definition of y_k to obtain (SC-Nest).

D.2 Lyapunov analysis

n the next proposition, we show that $E^{ac,c}$ is a rate-generating Lyapunov function, in the sense of Definition 2.3, for system (1st-ODE-SC) and its explicit discretization (FE-SC).

Proposition D.1. *Suppose f is μ-strongly convex and L-smooth. Let (x, v) be a solution of* (1st-ODE-SC) *and (x_k, v_k) be a sequences generated by* (FE-SC)*. Let $E^{ac,sc}$ be given by* (27)*. Then $E^{ac,sc}$ is a continuous Lyapunov function with $r_{E^{ac,sc}} = \sqrt{\mu}$ and $a_{E^{ac,sc}} = \frac{1}{\sqrt{L}}$ i.e.*

$$\frac{d}{dt}E^{ac,sc}(x,v) \leq -\sqrt{\mu}E^{ac,sc}(x,v) - \frac{1}{\sqrt{L}}|\nabla f(x)|^2 - \frac{\mu\sqrt{\mu}}{2}|v - x|^2. \tag{34}$$

Then we retrieve the usual optimal rates in the continuous and discrete cases.

Corollary D.2. *Let f be a μ-strongly convex and L-smooth function. Let $(x(t), v(t))$ be a solution to* (1st-ODE)*, then for all $t > 0$,*

$$f(x(t)) - f^* + \frac{\mu}{2}|v(t) - x^*|^2 \leq \exp(-\sqrt{\mu}t)E^{ac,sc}(x_0, v_0).$$

The proof of Corollary D.2 results immedialtly from Proposition D.1 and then, we focus on the proof of (34) in the following.

Proof of (34). Using (1st-ODE-SC), we obtain

$$
\begin{aligned}
\frac{d}{dt}E^{ac,sc}(x,v) &= \langle \nabla f(x), \dot{x}\rangle \\
&+ \sqrt{\mu}\langle v - x^*, \dot{v}\rangle \\
&= \sqrt{\mu}\langle \nabla f(x), v - x\rangle \\
&- \frac{1}{\sqrt{L}}|\nabla f(x)|^2 - \mu\sqrt{\mu}\langle v - x^*, v - x\rangle \\
&- \sqrt{\mu}\langle \nabla f(x), v - x^*\rangle \\
&= -\sqrt{\mu}\langle \nabla f(x), x - x^*\rangle \\
&- \frac{1}{\sqrt{L}}|\nabla f(x)|^2 \\
&- \frac{\mu\sqrt{\mu}}{2}\left[|v - x^*|^2 + |v - x|^2 - |x - x^*|^2\right].
\end{aligned}
$$

By strong convexity, we have

$$
\begin{aligned}
\frac{d}{dt}E^{ac,sc}(x,v) &\le -\sqrt{\mu}\left(f(x) - f^* + \frac{\mu}{2}|x - x^*|^2\right) \\
&- \frac{1}{\sqrt{L}}|\nabla f(x)|^2 \\
&- \frac{\mu\sqrt{\mu}}{2}\left[|v - x^*|^2 + |v - x|^2 - |x - x^*|^2\right] \\
&\le -\sqrt{\mu}E^{ac,sc}(x,v) \\
&- \frac{1}{\sqrt{L}}|\nabla f(x)|^2 - \frac{\mu\sqrt{\mu}}{2}|v - x|^2.
\end{aligned}
$$

which establishes (34). □

D.3 Proof of Proposition 5.5

First, arguing as in Proposition D.1,

$$
\begin{aligned}
f(x_{k+1}) - f(x_k) &\le \langle \nabla f(y_k), y_k - x_k\rangle - \frac{\mu}{2}|y_k - x_k|^2 \\
&+ \left(\frac{h^2}{2} - \frac{h}{\sqrt{L}}\right)|\nabla f(y_k)|^2 \\
&- \frac{h}{\sqrt{L}}\langle \nabla f(y_k), e_k\rangle \\
&+ h^2\left\langle \nabla f(y_k) + \frac{e_k}{2}, e_k\right\rangle,
\end{aligned}
$$

and,

$$
\begin{aligned}
&\frac{\mu}{2}|v_{k+1} - x^*|^2 - \frac{\mu}{2}|v_k - x^*|^2 \\
&\le -h\sqrt{\mu}E_k^{ac,sc} \\
&+ (\sqrt{\mu} + Lh)\frac{\sqrt{\mu}}{2}|x_k - y_k|^2 + \frac{h^2}{2}|\nabla f(y_k)|^2 \\
&- h\sqrt{\mu}\langle v_k - x^* + x_k - y_k, e_k\rangle \\
&+ h^2\left\langle \nabla f(y_k) + \frac{e_k}{2}, e_k\right\rangle.
\end{aligned}
$$

Summing these two inequalities,

$$
\begin{aligned}
E_{k+1}^{ac,sc} - E_k^{ac,sc} &\le -h\sqrt{\mu}E_k^{ac,sc} + \left(h^2 - \frac{h}{\sqrt{L}}\right)|\nabla f(y_k)|^2 \\
&+ \left(\frac{h\sqrt{\mu}L}{2} - \frac{\sqrt{\mu}}{2h}\right)|x_k - y_k|^2 \\
&- h\sqrt{\mu}\langle x_k - y_k + v_k - x^*, e_k\rangle \\
&+ \frac{h}{\sqrt{L}}\langle \nabla f(y_k), e_k\rangle + 2h^2\left\langle \nabla f(y_k) + \frac{e_k}{2}, e_k\right\rangle.
\end{aligned}
$$