

# Point Convergence of Nesterov's Accelerated Gradient Method: An AI-Assisted Proof

Uijeong Jang\*      Ernest K. Ryu\*

## Abstract

The Nesterov accelerated gradient method, introduced in 1983, has been a cornerstone of optimization theory and practice. Yet the question of its point convergence had remained open. In this work, we resolve this longstanding open problem in the affirmative. The discovery of the proof was heavily assisted by ChatGPT, a proprietary large language model, and we describe the process through which its assistance was elicited.

## 1 Introduction

Consider the optimization problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x),$$

where  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  is  $L$ -smooth and convex. The plain gradient descent method, whose origin dates back to Cauchy [7], is

$$x_{k+1} = x_k - \frac{1}{L} \nabla f(x_k)$$

for  $k = 0, 1, \dots$  with  $x_0 \in \mathbb{R}^n$ . Gradient descent is known to converge in the sense of  $x_k \rightarrow x_\infty \in \arg \min f$  and  $f(x_k) - \inf f \leq \mathcal{O}(1/k)$  [3]. However, the  $\mathcal{O}(1/k)$  rate is famously suboptimal and can be accelerated. Nesterov's 1983 seminal paper [9] presented the Nesterov accelerated gradient (NAG) method

$$\begin{aligned} x_{k+1} &= y_k - \frac{1}{L} \nabla f(y_k), \\ y_{k+1} &= x_{k+1} + \frac{t_k - 1}{t_{k+1}} (x_{k+1} - x_k) \end{aligned}$$

for  $k = 0, 1, \dots$  with  $x_0 = y_0 \in \mathbb{R}^n$ ,  $t_0 = 1$ , and  $t_{k+1}^2 - t_{k+1} \leq t_k^2$ . NAG is known to converge in the sense of  $f(x_k) - \inf f \leq \mathcal{O}(1/k^2)$ , an accelerated rate, when  $t_{k+1} = (1 + \sqrt{1 + 4t_k^2})/2$  or  $t_k = (k + 2)/2$  for  $k = 0, 1, \dots$ . However, whether

---

\*Department of Mathematics, University of California, Los Angeles.  
e-mail: [uijeongjang@math.ucla.edu](mailto:uijeongjang@math.ucla.edu) and [eryu@math.ucla.edu](mailto:eryu@math.ucla.edu)

NAG exhibits point convergence  $x_k \rightarrow x_\infty \in \arg \min f$  has remained an open problem.

In this work, we resolve this longstanding open problem in the affirmative by showing point convergence for NAG in the sense of

$$x_k \rightarrow x_\infty, \quad y_k \rightarrow x_\infty, \quad x_\infty \in \arg \min f.$$

The discovery of the proof was heavily assisted by ChatGPT, a proprietary large language model, and we describe the process through which its assistance was elicited.

This result, along with the use of AI in its discovery, was first announced by the authors on X.com, a social networking platform, where it received significant public attention. Given the level of interest, we are quickly releasing this archival record to formally document the proofs and facilitate their prompt dissemination. An updated version of this manuscript, offering a more detailed discussion of the historical context and relevant prior work, will be released in due course.

## 1.1 Notation

A function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  is convex if

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y), \quad \forall x, y \in \mathbb{R}^n, \theta \in [0, 1].$$

If  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  is differentiable and convex, then it satisfies the convexity inequality [11, Equation 2.1.2]

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \quad \forall x, y \in \mathbb{R}^n.$$

For  $L > 0$ , we say  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  is  $L$ -smooth if  $f$  is differentiable and

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad \forall x, y \in \mathbb{R}^n.$$

If  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  is  $L$ -smooth and convex, then it satisfies the cocoercivity inequality [11, Theorem 2.1.5]

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2, \quad \forall x, y \in \mathbb{R}^n.$$

Write  $\arg \min f$  to denote the set of minimizers of  $f$  and  $f_* = \inf_{x \in \mathbb{R}^n} f$  to denote the minimum value.

## 1.2 Concurrent work

The results presented in this manuscript were first announced by the authors on X.com through a series of daily posts from October 21 to October 24, 2025. Specifically, the convergence of the continuous-time dynamics for the case  $r = 3$  was announced on October 21, and the convergence of the discrete-time NAG method was announced on October 24.

On October 24, the authors of [6] contacted us, the authors of this work, with a manuscript also establishing point convergence of the discrete-time NAG. Their work builds upon the initial [X.com](#) post made on October 21 but was conducted independently of the October 24 post.

Beyond the analysis of NAG, the manuscript [6] contains additional results that do not overlap with the results of this work: their analysis extends to the infinite-dimensional Hilbert space setting, and they further argue that the FISTA method [4] also exhibits point convergence.

## 2 Point convergence in continuous time

Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a differentiable convex function with at least one minimizer. Consider the generalized Nesterov ODE

$$\ddot{X}(t) + \frac{r}{t}\dot{X}(t) + \nabla f(X(t)) = 0, \quad \text{for } t > 0$$

with  $r > 0$  and initial conditions  $X(0) = X_0$  and  $\dot{X}(0) = 0$ . The goal is to determine whether

$$X(t) \rightarrow X_\infty \in \arg \min f$$

holds, or whether there is an  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  such that  $\{X(t)\}_{t \geq 0}$  diverges. We take for granted the existence of a global solution to the ODE.

Prior work [2] has shown convergence for the overdamped regime  $r > 3$ . In the following, we establish convergence for the critical damping regime  $r = 3$ , which is the main case of interest, provide a partial convergence result for  $r \in (1, 3)$ , and demonstrate divergence for  $r \in (0, 1]$ .

### 2.1 Convergence for $r = 3$

For  $z \in \arg \min f$ , consider the following energy function:

$$\mathcal{E}_z(t) = t^2(f(X) - f_\star) + \frac{1}{2}\|t\dot{X} + 2(X - z)\|^2.$$

**Lemma 1.** *The energy function  $\mathcal{E}_z(t)$  converges to a finite limit and  $X(t)$  is bounded for all  $t$ .*

*Proof.* It is known that [13]

$$\frac{d}{dt}\mathcal{E}_z(t) = 2t\underbrace{(f(X) - f_\star - \langle \nabla f(X), X - z \rangle)}_{\leq 0, \text{ by convexity}} \leq 0$$

for  $t > 0$ . Therefore,  $\mathcal{E}_z(t)$  is nonincreasing and bounded below by zero. So, it converges to a finite limit.

Boundedness of the trajectory  $\{X(t)\}_{t \geq 0}$  is also known from [2], but we provide a proof here so that we can discuss its extension to the  $r \in (1, 3)$  case later in Section 2.2. Let  $W(t) := t^2(X(t) - z)$ . Then,

$$\dot{W} = t^2\dot{X} + 2t(X - z)$$

and thus

$$\mathcal{E}_z(t) = t^2(f(X) - f_\star) + \frac{1}{2} \left\| \frac{1}{t} \dot{W} \right\|^2 \leq \mathcal{E}_z(0).$$

So,  $\frac{1}{t} \dot{W}$  is bounded, say  $\sup_t \left\| \frac{1}{t} \dot{W} \right\| \leq M$ . Then we have

$$\|X - z\| = \frac{1}{t^2} \|W\| \leq \frac{1}{t^2} \int_0^t \|\dot{W}\| ds \leq \frac{1}{t^2} \int_0^t s M ds = \frac{M}{2}.$$

□

**Theorem 2.** *The solution  $X(t)$  of the Nesterov ODE converges to a minimizer. That is,*

$$X(t) \rightarrow X_\infty \in \arg \min f.$$

*Proof.* Since  $\{X(t)\}_{t \geq 0} \subset \mathbb{R}^n$  is bounded, the dynamics have at least one cluster point. Since  $f(X(t)) \rightarrow f_\star$  and  $f$  is continuous, all cluster points attain the function value  $f_\star$  and therefore are minimizers. If there is only one cluster point,  $X(t)$  converges to a limit, and we are done.

Therefore, assume  $z_1$  and  $z_2$  are cluster points of  $X(t)$ , so  $z_1, z_2 \in \arg \min f$ . We will now show  $z_1 = z_2$ . Let

$$h_i(t) := \|X(t) - z_i\|^2 \text{ for } i = 1, 2, \quad H(t) := \mathcal{E}_{z_1}(t) - \mathcal{E}_{z_2}(t).$$

Expanding the square in  $\mathcal{E}_z$  gives

$$\mathcal{E}_z(t) = t^2(f(X) - f_\star) + \frac{1}{2} t^2 \|\dot{X}\|^2 + 2h_z(t) + t h'_z(t).$$

The subtraction  $\mathcal{E}_{z_1} - \mathcal{E}_{z_2}$  cancels the  $t^2(f(X) - f_\star)$  and  $\frac{1}{2} t^2 \|\dot{X}\|^2$  terms and yields the linear ODE:

$$t(h_1 - h_2)' + 2(h_1 - h_2) = H(t). \tag{1}$$

Since both  $\mathcal{E}_{z_1}(t)$  and  $\mathcal{E}_{z_2}(t)$  converge to finite limits,

$$H(t) \rightarrow H(\infty) < \infty.$$

Multiply (1) by the integrating factor  $t$  to obtain

$$\frac{d}{dt}(t^2(h_1(t) - h_2(t))) = t H(t).$$

Then

$$(t^2(h_1(t) - h_2(t))) - (t_0^2(h_1(t_0) - h_2(t_0))) = \int_0^t s H(s) ds,$$

which implies

$$h_1(t) - h_2(t) = \frac{t_0^2}{t^2}(h_1(t_0) - h_2(t_0)) + \frac{1}{t^2} \int_0^t s H(s) ds.$$

By taking  $t \rightarrow \infty$ , the first term on the right-hand side vanishes and the second terms converges to  $\frac{H(\infty)}{2}$  by L'Hôpital's rule. Thus,  $h_1(t) - h_2(t)$  converges. Evaluating  $h_1 - h_2$  along two subsequences  $\{t_k\}_{k \geq 0}$  and  $\{s_k\}_{k \geq 0}$  such that  $X(t_k) \rightarrow z_1$  and  $X(s_k) \rightarrow z_2$  respectively, we get

$$\lim_{k \rightarrow \infty} (h_1(t_k) - h_2(t_k)) = \|z_1 - z_1\|^2 - \|z_1 - z_2\|^2 = -\|z_1 - z_2\|^2$$

and

$$\lim_{k \rightarrow \infty} (h_1(s_k) - h_2(s_k)) = \|z_2 - z_1\|^2 - \|z_2 - z_2\|^2 = \|z_2 - z_1\|^2$$

The two values must coincide since  $h_1 - h_2$  has a limit, therefore  $z_1 = z_2$ .  $\square$

## 2.2 Analysis for $r \in (1, 3)$

For the cases  $r < 3$ , we consider the ODE

$$\ddot{X}(t) + \frac{r}{t} \dot{X}(t) + \nabla f(X(t)) = 0, \quad \text{for } t > t_0$$

with initial conditions  $X(t_0)$  and  $\dot{X}(t_0)$  specified at time  $t_0 > 0$ . We introduce the initial time  $t_0$  to avoid technical complications associated with the singularity at  $t = 0$ .

For  $z \in \arg \min f$ , consider the following energy functions:

$$\begin{aligned} \mathcal{E}_z(t) &= t^{r-3} \left( t^2(f(X) - f_\star) + \frac{1}{2} \|t\dot{X} + (r-1)(X-z)\|^2 \right) \\ \mathcal{F}_z(t) &= t^{\frac{2r}{3}} (f(X) - f_\star) + \frac{r(3-r)}{9} t^{\frac{2r}{3}-2} \|X-z\|^2 + \frac{1}{2} t^{\frac{2r}{3}-2} \left\| t\dot{X} + \frac{2r}{3}(X-z) \right\|^2. \end{aligned}$$

**Lemma 3.** *Let  $r \in (1, 3)$ . The energy functions  $\mathcal{E}_z(t)$  and  $\mathcal{F}_z(t)$  are nonincreasing, hence converging to a finite limit. Also,  $f(X) \rightarrow f_\star$ , specifically with the rate*

$$f(X) - f_\star \leq \frac{\mathcal{F}_z(t_0)}{t^{\frac{2r}{3}}}$$

for  $t_0 > 0$ . Further,

- If  $\arg \min f$  is bounded, then  $\{X(t)\}_{t \geq t_0}$  is bounded.
- If  $\arg \min f$  is unbounded, then

$$\|X(t) - z\| \leq \min \left\{ \frac{3}{\sqrt{r(3-r)}} \sqrt{\mathcal{F}_z(t_0)} t^{\frac{3-r}{3}}, \frac{2\sqrt{2}}{r+1} \sqrt{\mathcal{E}_z(t_0)} t^{\frac{3-r}{2}} + \mathcal{O}\left(\frac{1}{t^{r-1}}\right) \right\}.$$

*Remark.* Unlike in the  $r = 3$  case, we do not have boundedness of  $\{X(t)\}_{t \geq t_0}$  in general. We leave this as an open problem for now.

*Proof.* These energy functions were originally introduced and analyzed in [1, 2]. It was also shown that

$$\frac{d}{dt} \mathcal{E}_z(t) = (r-1)t^{r-2} \underbrace{(f(X) - f_\star - \langle \nabla f(X), X - z \rangle)}_{\leq 0, \text{ by convexity}} + \underbrace{\left(\frac{r-3}{2}\right)t^{r-4}\|Y\|^2}_{\leq 0} \leq 0$$

and

$$\begin{aligned} t^{3-\frac{2r}{3}} \frac{d}{dt} \mathcal{F}_z(t) &= \frac{2r}{3} t^2 \underbrace{(f(X) - f_\star - \langle \nabla f(X), X - z \rangle)}_{\leq 0, \text{ by convexity}} \\ &\quad + \frac{r(3-r)}{9} \left[ \left(\frac{2r}{3} - 2\right)\|X-z\|^2 + 2\left(\langle X-z, V \rangle - \frac{2r}{3}\|X-z\|^2\right) \right] \\ &\quad + \frac{1}{2} \left(\frac{2r}{3} - 2\right)\|V\|^2 + \left(1 - \frac{r}{3}\right)\left(\|V\|^2 - \frac{2r}{3}\langle V, X-z \rangle\right) \\ &\leq -\frac{2r(3-r)(3+r)}{27}\|X-z\|^2 \leq 0. \end{aligned}$$

where  $Y := t\dot{X} + (r-1)(X-z)$  and  $V := t\dot{X} + \frac{2r}{3}(X-z)$ . Hence  $\mathcal{E}_z(t)$  and  $\mathcal{F}_z(t)$  are nonincreasing and converge to a finite limit. Therefore, we get

$$t^{\frac{2r}{3}}(f(X) - f_\star) \leq \mathcal{F}_z(t) \leq \mathcal{F}_z(t_0)$$

and hence the  $\mathcal{O}(t^{-\frac{2r}{3}})$  convergence rate.

If  $\arg \min f$  is bounded, then there exists  $\varepsilon_0 > 0$  such that the sublevel set  $E := \{x \mid f(x) \leq f_\star + \varepsilon_0\}$  is bounded. If not, there exists  $\|x_k\| \rightarrow \infty$  with  $f(x_k) \leq f_\star + \frac{1}{k}$  and  $x_k \neq z$ . Then set

$$d_k := \frac{x_k - z}{\|x_k - z\|}.$$

By compactness of the unit sphere, up to a subsequence we may assume  $d_k \rightarrow d$  for some  $\|d\| = 1$ , and note that  $\|x_k - z\| \rightarrow \infty$ . Fix any  $t > 0$  and define  $y_k := z + td_k$ . For all sufficiently large  $k$ , we have  $t \leq \|x_k - z\|$ , so  $y_k$  lies on the segment  $[z, x_k]$ . By convexity of  $f$ ,

$$f(y_k) \leq \left(1 - \frac{t}{\|x_k - z\|}\right)f(z) + \frac{t}{\|x_k - z\|}f(x_k) \leq f_\star + \frac{t}{\|x_k - z\|} \cdot \frac{1}{k}.$$

Since  $\|x_k - z\| \rightarrow \infty$ , the right-hand side tends to  $f_\star$ . So we get

$$f(z + td) \leq \liminf_{k \rightarrow \infty} f(y_k) \leq f_\star.$$

Hence  $f(z + td) = f_\star$  for all  $t > 0$ , which shows that  $\arg \min f$  is unbounded, contradicting the hypothesis. Now since  $f(X(t))$  converges to  $f_\star$ ,  $\{X(t)\}_{t \geq t_0}$  must enter  $E$  for  $t \geq T$ . Hence  $\{X(t)\}_{t \geq t_0}$  is bounded.

If  $\arg \min f$  is unbounded, let  $W(t) := t^{r-1}(X(t) - z)$ . Then,

$$\dot{W} = t^{r-1}\dot{X} + (r-1)t^{r-2}(X - z)$$

and thus

$$\mathcal{E}_z(t) = t^{r-3} \left( t^2(f(X) - f_\star) + \frac{1}{2} \left\| \frac{1}{t^{r-2}} \dot{W} \right\|^2 \right) \leq \mathcal{E}_z(t_0).$$

So,  $\left\| \frac{1}{t^{r-2}} \dot{W} \right\| \leq 2\sqrt{\mathcal{E}_z(t_0)} t^{\frac{r-1}{2}}$ . Then we have

$$\begin{aligned} \|X(t) - z\| &= \frac{1}{t^{r-1}} \|W(t)\| \\ &\leq \frac{1}{t^{r-1}} \left( \|W(t_0)\| + \int_{t_0}^t \|\dot{W}(s)\| ds \right) \\ &\leq \frac{t_0^{r-1} \|X(t_0) - z\|}{t^{r-1}} + \frac{\sqrt{2\mathcal{E}_z(t_0)}}{t^{r-1}} \int_{t_0}^t s^{\frac{r-1}{2}} ds \\ &= \frac{t_0^{r-1} \|X(t_0) - z\|}{t^{r-1}} + \frac{2\sqrt{2\mathcal{E}_z(t_0)}}{r+1} t^{\frac{3-r}{2}} + \mathcal{O}\left(\frac{1}{t^{r-1}}\right) \\ &= \frac{2\sqrt{2}}{r+1} \sqrt{\mathcal{E}_z(t_0)} t^{\frac{3-r}{2}} + \mathcal{O}\left(\frac{1}{t^{r-1}}\right). \end{aligned}$$

Meanwhile, we also have

$$\frac{r(3-r)}{9} t^{\frac{2r}{3}-2} \|X - z\|^2 \leq \mathcal{F}_z(t) \leq \mathcal{F}_z(t_0).$$

Thus,

$$\|X(t) - z\| \leq \frac{3}{\sqrt{r(3-r)}} \sqrt{\mathcal{F}_z(t_0)} t^{\frac{3-r}{3}}.$$

Putting altogether,

$$\|X(t) - z\| \leq \min \left\{ \frac{3}{\sqrt{r(3-r)}} \sqrt{\mathcal{F}_z(t_0)} t^{\frac{3-r}{3}}, \frac{2\sqrt{2}}{r+1} \sqrt{\mathcal{E}_z(t_0)} t^{\frac{3-r}{2}} + \mathcal{O}\left(\frac{1}{t^{r-1}}\right) \right\}.$$

□

**Theorem 4** (Uniqueness of the limit for  $r \in (1, 3)$ ). *Let  $r \in (1, 3)$ . Assume  $\arg \min f$  is bounded. Then, the solution  $X(t)$  of the generalized Nesterov ODE converges to a minimizer. That is,*

$$X(t) \rightarrow X_\infty \in \arg \min f.$$

*Proof.* Since  $\{X(t)\}_{t \geq t_0} \subset \mathbb{R}^n$  is bounded, the dynamics have at least one cluster point. Since  $f(X(t)) \rightarrow f_\star$  and  $f$  is continuous, all cluster points attain the function value  $f_\star$  and therefore are minimizers. If there is only one cluster point,  $X(t)$  converges to a limit, and we are done.

Therefore, assume  $z_1$  and  $z_2$  are cluster points of  $X(t)$ , so  $z_1, z_2 \in \arg \min f$ . We will now show  $z_1 = z_2$ . Let

$$h_i(t) := \|X(t) - z_i\|^2 \text{ for } i = 1, 2, \quad H(t) := \mathcal{E}_{z_1}(t) - \mathcal{E}_{z_2}(t).$$

Expanding the square in  $\mathcal{E}_z$  gives

$$\mathcal{E}_z(t) = t^{r-3} \left[ t^2 (f(X) - f_\star) + \frac{1}{2} t^2 \|\dot{X}\|^2 + \frac{(r-1)^2}{2} h_z(t) + \frac{(r-1)}{2} t h'_z(t) \right].$$

The subtraction  $\mathcal{E}_{z_1} - \mathcal{E}_{z_2}$  cancels the  $t^{r-1}(f(X) - f_\star)$  and  $\frac{1}{2}t^{r-1}\|\dot{X}\|^2$  terms and yields the linear ODE:

$$t(h_1 - h_2)' + (r-1)(h_1 - h_2) = \frac{2}{(r-1)t^{r-3}} H(t). \quad (2)$$

Multiply (2) by the integrating factor  $t^{r-2}$  to obtain

$$\frac{d}{dt}(t^{r-1}(h_1(t) - h_2(t))) = \frac{2}{r-1} t H(t).$$

Since both  $\mathcal{E}_{z_1}(t)$  and  $\mathcal{E}_{z_2}(t)$  converge to finite limits,

$$H(t) \rightarrow H(\infty) < \infty.$$

Then

$$(t^{r-1}(h_1(t) - h_2(t))) - (t_0^{r-1}(h_1(t_0) - h_2(t_0))) = \frac{2}{r-1} \int_{t_0}^t s H(s) ds,$$

which implies

$$h_1(t) - h_2(t) = \frac{t_0^{r-1}}{t^{r-1}} (h_1(t_0) - h_2(t_0)) + \frac{2}{(r-1)t^{r-1}} \int_{t_0}^t s H(s) ds.$$

By taking  $t \rightarrow \infty$ , the first term on the right-hand side vanishes. For the second term,

$$\begin{aligned} \frac{2}{(r-1)t^{r-1}} \int_{t_0}^t s H(s) ds &= \frac{2}{(r-1)t^{r-1}} \int_{t_0}^t s(H(\infty) + H(s) - H(\infty)) ds \\ &= \frac{2}{(r-1)t^{r-1}} \int_{t_0}^t s H(\infty) ds + \frac{2}{(r-1)t^{r-1}} \int_{t_0}^t s(H(s) - H(\infty)) ds \\ &= \frac{2H(\infty)}{(r-1)t^{r-1}} \cdot \frac{t^2 - t_0^2}{2} + \frac{2}{(r-1)t^{r-1}} o(t^2) \\ &= \frac{H(\infty)}{(r-1)} t^{3-r} + o(t^{3-r}) \end{aligned}$$

If  $H(\infty) \neq 0$ , then it blows up when  $t \rightarrow \infty$ , which contradicts the boundedness of  $h_1 - h_2$ . Thus,  $H(\infty) = 0$  and  $h_1(t) - h_2(t) \rightarrow 0$ . Evaluating  $h_1 - h_2$  along

two subsequences  $\{t_k\}_{k \geq 0}$  and  $\{s_k\}_{k \geq 0}$  such that  $X(t_k) \rightarrow z_1$  and  $X(s_k) \rightarrow z_2$  respectively, we get

$$\lim_{k \rightarrow \infty} (h_1(t_k) - h_2(t_k)) = \|z_1 - z_1\|^2 - \|z_1 - z_2\|^2 = -\|z_1 - z_2\|^2$$

and

$$\lim_{k \rightarrow \infty} (h_1(s_k) - h_2(s_k)) = \|z_2 - z_1\|^2 - \|z_2 - z_2\|^2 = \|z_2 - z_1\|^2$$

The two values must coincide since  $h_1 - h_2$  has a limit, therefore  $z_1 = z_2$ .  $\square$

### 2.3 Divergence for $r \in (0, 1]$

In the following, we present a counterexample showing that point convergence does not hold in general for  $r \in (0, 1]$ .

**Theorem 5.** *Let  $f: \mathbb{R} \rightarrow \mathbb{R}$  be given by*

$$f(x) = \begin{cases} \frac{1}{2}(x-1)^2 & x > 1 \\ 0 & |x| \leq 1 \\ \frac{1}{2}(x+1)^2 & x < -1. \end{cases}$$

and consider

$$\ddot{X}(t) + \frac{r}{t}\dot{X}(t) + f'(X(t)) = 0, \quad \text{for } t > t_0 > 0$$

with  $r \in (0, 1]$  and initial conditions  $X(t_0) = 2$  and  $\dot{X}(t_0) = 0$ . Then,  $X(t)$  hits  $\pm 1$  infinitely many times and therefore diverges.

*Proof.* The equation is piecewise linear and can be written in three parts as follows: If  $x(t) > 1$ , let  $y(t) = x(t) - 1$ . Then

$$\ddot{y} + \frac{r}{t}\dot{y} + y = 0. \quad (\text{R})$$

If  $|x(t)| \leq 1$ , we have

$$\ddot{x} + \frac{r}{t}\dot{x} = 0. \quad (\text{C})$$

If  $x(t) < -1$ , let  $z(t) = x(t) + 1$ . Then

$$\ddot{z} + \frac{r}{t}\dot{z} + z = 0, \quad (\text{L})$$

We will show that the trajectory hits  $x = 1$  and  $x = -1$  infinitely many times, and hence cannot converge.

Consider equation (R) (the same argument applies to (L)). Let  $y(t) = t^{-r/2}u(t)$ ; a short calculation yields

$$\ddot{u} + \left(1 + \frac{r(2-r)}{4t^2}\right)u = 0. \quad (\star)$$

Because  $r \in (0, 1]$ , the coefficient in  $(\star)$  satisfies

$$1 + \frac{r(2-r)}{4t^2} \geq 1 \quad \text{for all } t > 0.$$

By the Sturm comparison theorem with  $w = \sin t$  satisfying  $\ddot{w} + w = 0$ , every nontrivial solution  $u$  of  $(\star)$  (and hence every nontrivial  $y$  of  $(R)$ ) has a zero in every interval of length at most  $\pi$ .

In particular, starting outside with  $y(t_0) > 0$  (or equivalently  $x(t_0) > 1$ ), there exists a first time  $t_1 > 0$  such that  $y(t_1) = 0$ ; that is,  $x(t_1) = 1$ . If  $y(t_1) = \dot{y}(t_1) = 0$ , then by the uniqueness theorem for linear ODEs, the only solution to  $(R)$  with this data is  $y \equiv 0$ . This contradicts  $y(t_0) = 1$ . Thus,  $\dot{x}(t_1) \neq 0$  and the trajectory enters the region  $|x| < 1$  with strictly negative velocity. Exactly the same reasoning applies to the left region (L): whenever the trajectory is outside on the left, it re-enters  $[-1, 1]$  at  $x = -1$  with nonzero velocity.

While  $|x(t)| \leq 1$ , the trajectory follows (C):

$$\ddot{x} + \frac{r}{t}\dot{x} = 0.$$

Integrating once gives

$$\dot{x}(t) = \dot{x}(t_1) \left(\frac{t_1}{t}\right)^r.$$

Because the integral

$$\int_{t_1}^{\infty} \left(\frac{t_1}{t}\right)^r dt$$

diverges for  $r \leq 1$ , the trajectory traverses the entire flat region in finite time: there exists  $t_2 > t_1$  such that

$$x(t_2) = -1.$$

Moreover,

$$-2 = x(t_2) - x(t_1) = |\dot{x}(t_1)| t_1^r \int_{t_1}^{t_2} s^{-r} ds.$$

Since  $s^{-r} \leq t_1^{-r}$  for all  $s \geq t_1$ , it follows that

$$2 \leq |\dot{x}(t_1)| t_1^r \int_{t_1}^{t_2} t_1^{-r} ds = |\dot{x}(t_1)| (t_2 - t_1),$$

and therefore

$$t_2 - t_1 \geq \frac{2}{|\dot{x}(t_1)|}.$$

Also, from  $\dot{x}(t) = \dot{x}(t_1) \left(\frac{t_1}{t}\right)^r$ ,  $\dot{x}(t_2)$  has the same sign as  $\dot{x}(t_1)$ . Hence, the trajectory exits the flat region at the opposite endpoint with nonzero velocity. At  $t = t_2$ , we are at  $x = -1$  with  $\dot{x}(t_1) < 0$ . Hence immediately  $x(t) < -1$  for

$t > t_2$ , the equation (L) governs the motion. By the same reasoning applied to find  $t_1$ , there exists a next time  $t_3 > t_2$  such that

$$x(t_3) = -1 \quad \text{and} \quad \dot{x}(t_3) > 0,$$

corresponding to re-entry into the flat region  $|x| \leq 1$ . Repeating this reasoning inductively, we obtain an increasing sequence

$$t_0 < t_1 < t_2 < t_3 < t_4 < \dots$$

with

$$x(t_{4k-3}) = x(t_{4k}) = 1, \quad x(t_{4k-2}) = x(t_{4k-1}) = -1, \quad k = 1, 2, \dots$$

and  $\dot{x}(t_i) \neq 0$  for all  $i$ . Now, we claim that  $t_k \rightarrow \infty$ . Define the oscillator energy

$$E(t) := \frac{1}{2}\dot{x}(t)^2 + f(x(t))$$

which satisfies

$$\frac{d}{dt}E(t) = \dot{x}(t)\ddot{x}(t) + f'(x(t))\dot{x}(t) = -\frac{r}{t}\dot{x}(t)^2 \leq 0,$$

so  $E(t)$  is nonincreasing. Hence

$$\frac{1}{2}\dot{x}(t_{2k-1})^2 = E(t_{2k-1}) \leq E(t_1) = \frac{1}{2}\dot{x}(t_1)^2.$$

for  $k \geq 1$ . Consequently, for all  $k$ ,

$$t_{2k} - t_{2k-1} \geq \frac{2}{|\dot{x}(t_{2k-1})|} \geq \frac{2}{|\dot{x}(t_1)|} > 0.$$

So,

$$t_0 < t_1 < t_2 < t_3 < t_4 < \dots \rightarrow \infty.$$

Therefore, the trajectory alternates indefinitely between the endpoints  $x = \pm 1$ , and thus  $x(t)$  does not converge as  $t \rightarrow \infty$ . This completes the construction of a counterexample for every  $r \in (0, 1]$ .  $\square$

### 3 Point convergence in discrete time

In this section, we establish point convergence of the following discrete-time methods: Nesterov's 1983 NAG method [9] and Kim and Fessler's 2016 optimized gradient method [8].

### 3.1 Nesterov accelerated gradient

Consider the Nesterov accelerated gradient algorithm (NAG)

$$\begin{aligned}x_{k+1} &= y_k - \frac{1}{L} \nabla f(y_k), \\y_{k+1} &= x_{k+1} + \frac{t_k - 1}{t_{k+1}}(x_{k+1} - x_k)\end{aligned}$$

for  $k = 0, 1, \dots$  with  $x_0 = y_0 \in \mathbb{R}^n$  and a nonnegative sequence  $\{t_k\}_{k=0,1,\dots}$  satisfying  $t_0 = 1$ ,  $t_{k+1}^2 - t_{k+1} \leq t_k^2$  for  $k = 0, 1, \dots$ , and  $t_k \rightarrow \infty$  as  $k \rightarrow \infty$ . For notational convenience, also define  $t_{-1} = 0$ .

**Lemma 6** ([10, Lemma 1]). *The following is equivalent to NAG in the sense that it produces the same sequences  $\{x_k\}_{k \geq 0}$  and  $\{y_k\}_{k \geq 0}$ :*

$$\begin{aligned}x_{k+1} &= y_k - \frac{1}{L} \nabla f(y_k) \\z_{k+1} &= z_k - \frac{t_k}{L} \nabla f(y_k) \\y_{k+1} &= \left(1 - \frac{1}{t_{k+1}}\right) x_{k+1} + \frac{1}{t_{k+1}} z_{k+1}\end{aligned}$$

for  $k = 0, 1, \dots$ , where  $x_0 = y_0 = z_0 \in \mathbb{R}^n$ .

**Lemma 7** ([9]). *For  $x_\star \in \arg \min f$ , consider the following energy function for  $k = 0, 1, 2, \dots$ :*

$$\mathcal{E}_k(x_\star) = t_{k-1}^2 (f(x_k) - f_\star) + \frac{L}{2} \|z_k - x_\star\|^2$$

where  $z_{-1} = z_0$ . Assume  $t_{k+1}^2 - t_{k+1} \leq t_k^2$ , then  $\mathcal{E}_{k+1}(x_\star) \leq \mathcal{E}_k(x_\star)$  for all  $k \geq 0$ , hence  $\{\mathcal{E}_k(x_\star)\}_{k \geq 0}$  converges to a finite limit.

**Lemma 8.** *The sequence  $\{x_k\}_{k \geq 0}$  produced by NAG is bounded.*

*Proof.* From Lemma 7,  $\{z_k\}_{k \geq 0}$  is bounded. From the equivalent form of NAG, we have

$$\begin{aligned}x_{k+1} &= y_k - \frac{1}{L} \nabla f(y_k) \\&= y_k + \frac{1}{t_k} (z_{k+1} - z_k) \\&= \left(1 - \frac{1}{t_k}\right) x_k + \frac{1}{t_k} z_k + \frac{1}{t_k} (z_{k+1} - z_k) \\&= \left(1 - \frac{1}{t_k}\right) x_k + \frac{1}{t_k} z_{k+1}\end{aligned}\tag{3}$$

for any  $x_\star \in \arg \min f$ . We have  $\|z_k\|$  bounded, say  $\sup_k \|z_k\| \leq M$ . This gives the bound

$$\|x_{k+1}\| \leq \left(1 - \frac{1}{t_k}\right) \|x_k\| + \frac{1}{t_k} \|z_{k+1}\| \leq \max(\|x_k\|, M).$$

for all  $k \geq 0$ . So, we get

$$\|x_k\| \leq \max(\|x_0\|, M) < \infty.$$

□

**Lemma 9** ([5, Lemma A.4]). *Let  $\{h_k\}_{k \geq 0} \subset \mathbb{R}$  and  $\{\varphi_k\}_{k \geq 0} \subset \mathbb{R}$  be sequences of real and positive numbers respectively, and such that  $\sum_{k=0}^{\infty} \frac{1}{\varphi_k} = \infty$ . Then if*

$$h_{k+1} + \varphi_k(h_{k+1} - h_k) \rightarrow c < \infty,$$

*then  $h_k \rightarrow c$ .*

**Theorem 10.** *The sequences  $\{x_k\}_{k \geq 0}$  and  $\{y_k\}_{k \geq 0}$  of NAG converge to the same minimizer. That is,*

$$x_k \rightarrow x_{\infty}, \quad y_k \rightarrow x_{\infty}, \quad x_{\infty} \in \arg \min f.$$

*Proof.* Since  $\{x_k\}_{k \geq 0} \subset \mathbb{R}^n$  is bounded, the dynamics have at least one cluster point. Also, together with

$$t_{k-1}^2(f(x_k) - f_{\star}) \leq \mathcal{E}_k(x_{\star}) \leq \mathcal{E}_{-1}(x_{\star}) = \frac{L}{2}\|x_0 - x_{\star}\|^2$$

and  $t_k \rightarrow \infty$ , we have  $f(x_k) \rightarrow f_{\star}$ . Since  $f$  is continuous, all cluster points attain the function value  $f_{\star}$  and therefore are minimizers. If there is only one cluster point,  $x_k$  converges to a limit, and we are done.

Therefore, assume  $x_{\star}$  and  $\tilde{x}_{\star}$  are cluster points of  $\{x_k\}_{k \geq 0}$ , so  $x_{\star}, \tilde{x}_{\star} \in \arg \min f$ . We will now show  $x_{\star} = \tilde{x}_{\star}$ . Let

$$h_k := \|x_k - x_{\star}\|^2 - \|x_k - \tilde{x}_{\star}\|^2, \quad H_k := \mathcal{E}_k(x_{\star}) - \mathcal{E}_k(\tilde{x}_{\star}).$$

Then,

$$\begin{aligned} \frac{2}{L}H_{k+1} &= -2\langle z_{k+1}, x_{\star} - \tilde{x}_{\star} \rangle + \|x_{\star}\|^2 - \|\tilde{x}_{\star}\|^2 \\ h_{k+1} &= -2\langle x_{k+1}, x_{\star} - \tilde{x}_{\star} \rangle + \|x_{\star}\|^2 - \|\tilde{x}_{\star}\|^2 \\ h_k &= -2\langle x_k, x_{\star} - \tilde{x}_{\star} \rangle + \|x_{\star}\|^2 - \|\tilde{x}_{\star}\|^2 \end{aligned}$$

We use the equality  $t_k x_{k+1} - (t_k - 1)x_k = z_{k+1}$  from (3) to get

$$t_k h_{k+1} - (t_k - 1)h_k = h_{k+1} + (t_k - 1)(h_{k+1} - h_k) = \frac{2}{L}H_{k+1}.$$

Note that  $t_k - 1 > 0$  for  $k \geq 2$ . Also, from  $t_{k+1} \leq \frac{1+\sqrt{1+4t_k^2}}{2}$  and  $\sqrt{1+4t_k^2} \leq 2t_k + 1$ , we have

$$t_{k+1} \leq t_k + 1$$

and therefore  $t_k \leq t_0 + k = k + 1$ . So,

$$\sum_{k=2}^{\infty} \frac{1}{t_k - 1} \geq \sum_{k=2}^{\infty} \frac{1}{k} = \infty.$$

Lastly,  $\frac{2}{L}H_{k+1}$  converges. Thus, by Lemma 9,  $h_k \rightarrow \frac{2}{L}H_\infty$ . Passing  $h_k$  through two subsequences  $\{x_{m_k}\}_{k \geq 0}$  and  $\{x_{n_k}\}_{k \geq 0}$  such that  $x_{m_k} \rightarrow x_*$  and  $x_{n_k} \rightarrow \tilde{x}_*$  respectively, we get

$$\lim_{k \rightarrow \infty} h_{m_k} = \|x_* - x_*\|^2 - \|x_* - \tilde{x}_*\|^2 = -\|x_* - \tilde{x}_*\|^2$$

and

$$\lim_{k \rightarrow \infty} h_{n_k} = \|\tilde{x}_* - x_*\|^2 - \|\tilde{x}_* - \tilde{x}_*\|^2 = \|\tilde{x}_* - x_*\|^2.$$

The two values must coincide, therefore  $x_* = \tilde{x}_*$  and  $\{x_k\}_{k \geq 0}$  converges to  $x_\infty \in \arg \min f$ . Convergence of  $\{y_k\}_{k \geq 0}$  then follows from  $t_k \rightarrow \infty$  and

$$y_{k+1} = \left(1 - \frac{1}{t_{k+1}}\right)x_{k+1} + \frac{1}{t_{k+1}}z_{k+1} \rightarrow x_\infty.$$

□

### 3.2 Optimized gradient method

Consider the optimized gradient method (OGM)

$$\begin{aligned} x_{k+1} &= y_k - \frac{1}{L}\nabla f(y_k), \\ y_{k+1} &= x_{k+1} + \frac{\theta_k - 1}{\theta_{k+1}}(x_{k+1} - x_k) + \frac{\theta_k}{\theta_{k+1}}(x_{k+1} - y_k) \end{aligned}$$

for  $k = 0, 1, \dots$  with  $x_0 = y_0 \in \mathbb{R}^n$ , and a nonnegative sequence  $\{\theta_k\}_{k=0,1,\dots}$  satisfying  $\theta_0 = 1$  and  $\theta_{k+1}^2 - \theta_{k+1} = \theta_k^2$  for  $k = 0, 1, \dots$ . For notational convenience, also define  $\theta_{-1} = 0$ .

**Lemma 11** ([8, Proposition 5]). *The following is equivalent to OGM in the sense that it produces the same sequences  $\{x_k\}_{k \geq 0}$  and  $\{y_k\}_{k \geq 0}$ :*

$$\begin{aligned} x_{k+1} &= y_k - \frac{1}{L}\nabla f(y_k) \\ z_{k+1} &= z_k - \frac{2\theta_k}{L}\nabla f(y_k) \\ y_{k+1} &= \left(1 - \frac{1}{\theta_{k+1}}\right)x_{k+1} + \frac{1}{\theta_{k+1}}z_{k+1} \end{aligned}$$

for  $k = 0, 1, \dots$ , where  $x_0 = y_0 = z_0 \in \mathbb{R}^n$ .

**Lemma 12** ([12, Theorem 1]). *For  $x_* \in \arg \min f$ , consider the following energy function for  $k \geq 0$ :*

$$\mathcal{E}_k(x_*) = 2\theta_k^2 \underbrace{\left(f(y_k) - f_* - \frac{1}{2L}\|\nabla f(y_k)\|^2\right)}_{\geq 0, \text{ by cocoercivity}} + \frac{L}{2}\|z_{k+1} - x_*\|^2$$

where  $y_{-1} = y_0$  and  $z_{-1} = z_0$ . Then  $\mathcal{E}_{k+1}(x_*) \leq \mathcal{E}_k(x_*)$  for all  $k = -1, 0, 1, 2, \dots$ , hence  $\{\mathcal{E}_k(x_*)\}_{k \geq -1}$  converges to a finite limit.

**Lemma 13.** *The sequences  $\{x_k\}_{k \geq 0}$  and  $\{y_k\}_{k \geq 0}$  produced by OGM are bounded.*

*Proof.* From the OGM updates, we have

$$x_{k+1} = y_k + \frac{z_{k+1} - z_k}{2\theta_k}, \quad y_{k+1} = \left(1 - \frac{1}{\theta_{k+1}}\right)x_{k+1} + \frac{1}{\theta_{k+1}}z_{k+1}.$$

By Lemma 12,  $\{z_k\}$  is bounded, i.e.,  $\sup_k \|z_k\| \leq M$ . Then,

$$\|x_{k+1}\| \leq \|y_k\| + \frac{\|z_{k+1}\| + \|z_k\|}{2\theta_k} \leq \|y_k\| + \frac{M}{\theta_k}, \quad (4)$$

$$\|y_{k+1}\| \leq \left(1 - \frac{1}{\theta_{k+1}}\right)\|x_{k+1}\| + \frac{M}{\theta_{k+1}}. \quad (5)$$

Substituting (4) into (5) gives

$$\|y_{k+1}\| \leq \left(1 - \frac{1}{\theta_{k+1}}\right)\|y_k\| + M\left(\frac{1}{\theta_{k+1}} + \frac{1}{\theta_k}\left(1 - \frac{1}{\theta_{k+1}}\right)\right).$$

Since  $\theta_{k+1} \leq \theta_k + 1$ ,

$$\begin{aligned} \frac{1}{\theta_{k+1}} + \frac{1}{\theta_k}\left(1 - \frac{1}{\theta_{k+1}}\right) &= \frac{1}{\theta_{k+1}} + \frac{1}{\theta_k} - \frac{1}{\theta_k\theta_{k+1}} \\ &= \frac{1}{\theta_{k+1}} + \frac{\theta_{k+1} - 1}{\theta_k\theta_{k+1}} \\ &\leq \frac{1}{\theta_{k+1}} + \frac{\theta_k}{\theta_k\theta_{k+1}} \\ &= \frac{2}{\theta_{k+1}} \end{aligned}$$

Hence,

$$\|y_{k+1}\| \leq \left(1 - \frac{1}{\theta_{k+1}}\right)\|y_k\| + \frac{2M}{\theta_{k+1}}. \quad (6)$$

Let  $B := \max\{\|x_0\|, 2M\}$ . We show by induction that  $\|y_k\| \leq B$  for all  $k$ . For  $k = 0$ , this holds by the definition of  $B$ . Assume  $\|y_k\| \leq B$ . Then from (6),

$$\|y_{k+1}\| \leq \left(1 - \frac{1}{\theta_{k+1}}\right)B + \frac{2M}{\theta_{k+1}} \leq \left(1 - \frac{1}{\theta_{k+1}}\right)B + \frac{B}{\theta_{k+1}} = B,$$

since  $2M \leq B$ . Thus, by induction,  $\{y_k\}_{k \geq 0}$  is bounded. Finally, by (4),

$$\|x_{k+1}\| \leq \|y_k\| + \frac{M}{\theta_k} \leq B + M,$$

showing that  $\{x_k\}_{k \geq 0}$  is also bounded.  $\square$

**Theorem 14.** *The sequences  $\{x_k\}_{k \geq 0}$  and  $\{y_k\}_{k \geq 0}$  of OGM converge to the same minimizer. That is,*

$$x_k \rightarrow x_\infty, \quad y_k \rightarrow x_\infty, \quad x_\infty \in \arg \min f.$$

*Proof.* Since  $\{x_k\}_{k \geq 0} \subset \mathbb{R}^n$  is bounded, the dynamics have at least one cluster point. Also,

$$\begin{aligned} 2\theta_{k-1}^2(f(x_k) - f_\star) &\leq 2\theta_{k-1}^2 \left( f(y_{k-1}) - f_\star - \frac{1}{2L} \|\nabla f(y_{k-1})\|^2 \right) \\ &\leq \mathcal{E}_{k-1}(x_\star) \leq \mathcal{E}_{-1}(x_\star) = \frac{L}{2} \|z_0 - x_\star\|^2. \end{aligned}$$

Together with  $\theta_k \rightarrow \infty$  implies  $f(x_k) \rightarrow f_\star$ . Hence, all cluster points attain the function value  $f_\star$  and therefore are minimizers. As in the NAG case, let  $x_\star, \tilde{x}_\star \in \arg \min f$  be two cluster points of  $\{x_k\}_{k \geq 0}$ . Let

$$h_k^x := \|x_k - x_\star\|^2 - \|x_k - \tilde{x}_\star\|^2, \quad h_k^y := \|y_k - x_\star\|^2 - \|y_k - \tilde{x}_\star\|^2,$$

and

$$H_k := \mathcal{E}_k(x_\star) - \mathcal{E}_k(\tilde{x}_\star).$$

From the definition of  $\mathcal{E}_k(x_\star)$ , the only dependence on  $x_\star$  is through  $\|z_k - x\|^2$ , so for any  $x_\star, \tilde{x}_\star \in \arg \min f$  we have

$$\frac{2}{L} H_{k+1} = -2\langle z_{k+1}, x_\star - \tilde{x}_\star \rangle + \|x_\star\|^2 - \|\tilde{x}_\star\|^2.$$

Using the identity

$$\theta_{k+1} y_{k+1} - (\theta_{k+1} - 1)x_{k+1} = z_{k+1},$$

and the relations

$$-2\langle x_{k+1}, x_\star - \tilde{x}_\star \rangle = h_{k+1}^x - \|x_\star\|^2 + \|\tilde{x}_\star\|^2, \quad -2\langle y_{k+1}, x_\star - \tilde{x}_\star \rangle = h_{k+1}^y - \|x_\star\|^2 + \|\tilde{x}_\star\|^2,$$

we obtain

$$\theta_{k+1} h_{k+1}^x - (\theta_{k+1} - 1)h_{k+1}^y = \frac{2}{L} H_{k+1}. \quad (7)$$

Furthermore, since

$$y_{k+1} = x_k + \frac{z_{k+1} - z_k}{2\theta_k},$$

we can similarly derive

$$h_{k+1}^y = h_k^x + \frac{1}{L\theta_k} (H_{k+1} - H_k). \quad (8)$$

Combining the (7) and (8) yields the recursion

$$\theta_{k+1} h_{k+1}^x = (\theta_{k+1} - 1)h_k^x + \left[ \frac{\theta_{k+1} - 1}{L\theta_k} (H_{k+1} - H_k) + \frac{2}{L} H_{k+1} \right].$$

Or equivalently,

$$h_{k+1}^x + (\theta_{k+1} - 1)(h_{k+1}^x - h_k^x) = \frac{1}{L} \left( \frac{\theta_{k+1} - 1}{\theta_k} (H_{k+1} - H_k) + 2H_{k+1} \right) \quad (9)$$

Note we have  $\theta_{k+1} - 1 > 0$  for  $k \geq 0$ . Also, as we already proved in the proof of Theorem 10, we have

$$\sum_{k=0}^{\infty} \frac{1}{\theta_{k+1} - 1} \geq \sum_{k=0}^{\infty} \frac{1}{k+2-1} = \infty.$$

So, to apply Lemma 9, it suffices to show the convergence of the right-hand side of (9). This follows from  $H_k \rightarrow H_\infty$  and

$$\frac{\theta_{k+1} - 1}{\theta_k} |H_{k+1} - H_k| \leq \frac{\theta_k + 1 - 1}{\theta_k} |H_{k+1} - H_k| \rightarrow 0.$$

Hence  $h_{k+1}^x$  converges and we use the subsequence argument to conclude  $x_k \rightarrow x_\infty$ . Convergence of  $\{y_k\}_{k \geq 0}$  then follows from

$$y_{k+1} = \left(1 - \frac{1}{\theta_{k+1}}\right) x_{k+1} + \frac{1}{\theta_{k+1}} z_{k+1} \rightarrow x_\infty.$$

□

## 4 Eliciting AI-assistance

The discovery of the proofs presented in this work was heavily assisted by ChatGPT, a proprietary large language model. We believe that the mathematical results presented in this work hold value independent of the method of their discovery. At the same time, we believe our discovery serves as a case study demonstrating how AI can be leveraged to accelerate the discovery of new mathematics. For this reason, we briefly describe the process through which ChatGPT's assistance was elicited.

The process began by prompting ChatGPT to solve the continuous-time problem. The model did not produce the correct answer in a single attempt; rather, the process was highly interactive. ChatGPT generated numerous arguments, approximately 80% of which were incorrect, but several ideas felt novel and worth further exploring. Whenever a new idea emerged, whether correct or only partially so, we distilled the key insight and prompted ChatGPT to develop it further.

The authors' contribution was to filter out incorrect arguments, consolidate a consistent set of valid facts, identify promising lines of reasoning, and determine when a particular approach had been fully explored. ChatGPT's contribution was to generate candidate arguments, substantially accelerate the exploration of potential avenues, particularly by quickly ruling out unproductive directions, and ultimately produce the final proof argument.

Once the proof for the continuous-time setting with  $r = 3$  was discovered, the extensions to  $r \in (1, 3)$  and the translation to the discrete-time Nesterov accelerated gradient method were relatively straightforward. The prompting technique was to provide the proof for the continuous-time case in LaTeX code,

presenting the desired extension as a theorem statement, and instructing ChatGPT to produce a proof of the new theorem using ideas from the supplied proof.

For the case  $r \in (1, 3)$  in continuous-time, however, we were not able to establish boundedness in full generality. We prompted ChatGPT to prove the boundedness of  $\{X(t)\}_{t \geq t_0}$ , suggesting several potentially promising approaches, but the model was unable to produce a valid proof. We then carefully re-examined the problem ourselves to ensure that no simple argument had been overlooked. After a reasonable amount of effort, we concurred with ChatGPT that the technique used for establishing boundedness for the case  $r = 3$  does not immediately extend to the case  $r < 3$ . We therefore leave this issue as an open problem.

## References

- [1] H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont. Fast convergence of inertial dynamics and algorithms with asymptotic vanishing damping. *Mathematical Programming*, 2019.
- [2] H. Attouch, Z. Chbani, and H. Riahi. Rate of convergence of the Nesterov accelerated gradient method in the subcritical case  $\alpha \leq 3$ . *ESAIM: Control, Optimisation and Calculus of Variations*, 25:2, 2019.
- [3] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. 2nd edition. Springer, 2017.
- [4] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.
- [5] R. I. Boț, E. Chencene, E. R. Csetnek, and D. A. Hulett. Accelerating diagonal methods for bilevel optimization: Unified convergence via continuous-time dynamics. *arXiv:2505.14389*, 2025.
- [6] R. I. Boț, J. Fadili, and D.-K. Nguyen. The iterates of Nesterov’s accelerated algorithm converge in the critical regimes. *arXiv preprint*, 2025.
- [7] A.-L. Cauchy. Méthode générale pour la résolution des systèmes d’équations simultanées. *Comptes Rendus Hebdomadaires des Séances de l’Académie des Sciences*, 25:536–538, 1847.
- [8] D. Kim and J. A. Fessler. Optimized first-order methods for smooth convex minimization. *Mathematical Programming*, 159(1):81–107, 2016.
- [9] Y. Nesterov. A method for solving the convex programming problem with convergence rate  $O(1/k^2)$ . *Soviet Mathematics Doklady*, 1983.
- [10] Y. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1):127–152, 2005.

- [11] Y. Nesterov. *Lectures on Convex Optimization*, volume 137. Springer, 2018.
- [12] C. Park, J. Park, and E. K. Ryu. Factor- $\sqrt{2}$  acceleration of accelerated gradient methods. *Applied Mathematics & Optimization*, 88(3):77, 2023.
- [13] W. Su, S. Boyd, and E. J. Candès. A differential equation for modeling nesterov’s accelerated gradient method: Theory and insights. *Neural Information Processing Systems*, 2014.