Historical notes and errata to "Unifying Nesterov's accelerated gradient methods for convex and strongly convex objective functions" by Jungbin Kim and Insoon Yang

Jungbin Kim

March 10, 2026

The codes used to generate the figures in the paper [2] are available at `https://github.com/jungbinkim1/Unified-NAG`.

1 Historical notes

- The paper [2] is not credited with the first unified version of Nesterov acceleration because a unified algorithm was already presented in Nesterov's original work [4, Eq. 2.2.19]. Many results in Sections 3 and 4 of [2] amount to a refinement and a generalization of existing results and ideas scattered throughout the literature.

- The unification of non-strongly and strongly convex cases is a known theme in the literature, extending to many algorithms beyond just Nesterov acceleration. While it is hard to pinpoint a specific reference for this theme, a notable work is [6] (indeed one of the starting points for our research), which introduces ITEM as an algorithm that "unifies" OGM [1] and TMM [7].

- How much novelty does Unified AGM [2, §4.2] have? Let's first look at Nesterov's method for the non-strongly convex case ($\mu = 0$):

  $$x_{k+1} = y_k - \frac{1}{L}\nabla f(y_k), \qquad y_{k+1} = x_{k+1} + \frac{t_k - 1}{t_{k+1}}(x_{k+1} - x_k).$$

  One can set t_k using the inductive rule $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$. This rule makes it hard to see how the algorithm actually works, but we can still prove the $O(1/k^2)$ convergence rate using the fact $t_k \geq \frac{k+2}{2}$. A simpler and perhaps more popular choice is $t_k = \frac{k+2}{2}$. This choice is much more intuitive: you can directly see how the momentum coefficient changes as k increases, the $O(1/k^2)$ rate becomes obvious after a quick look at the proof, and the connection to the ODE model $\ddot{X} + \frac{3}{t}\dot{X} + \nabla f(X) = 0$ [5] is much more straightforward.

  Similarly, Unified AGM can be regarded as a more intuitive version of Nesterov's original unified algorithm [4, Eq. 2.2.19]. Since the momentum coefficient involves hyperbolic functions instead of messy inductive rules, the convergence rate and the connection to the ODE model are much easier to see.

- The credit for introducing the first unified continuous-time model for Nesterov acceleration belongs to [3]. Again, the results in [2, §§3.1, .4.1] are a refinement (see [2, Appendix A.3]).[1]
- As far as I know, the other results in [2] are truly original.
- The results in Section 6 of [2] were developed while engaging in discussions with other researchers.

2 Errata

- Page 2, three lines below Eq. 1: "Then, f is $(L + \frac{2\lambda}{m})$-smooth and $\frac{2\lambda}{m}$-strongly convex."
- Page 5, second paragraph: "where $\mathbf{t}_k := \iota\sqrt{s}k$ and $Z(t) = X(t) + \frac{t}{2}\mathrm{tanhc}(\frac{\sqrt{\mu}}{2}t)\dot{X}(t)$."
- Page 6, paragraph before Section 5.1: "when $p = 2$ and $h(x) = \frac{1}{2}\|x\|^2$."
- Page 9, paragraph before Section 8: "We set the parameters as $s = 10^{-3}$ (stepsize), $p = 3$ (order), $N = \sqrt{2}$, $M = 1/3$ (input constants), and $n = 50$."

References

[1] Donghwan Kim and Jeffrey A Fessler. Optimized first-order methods for smooth convex minimization. *Mathematical Programming*, 159(1):81–107, 2016.

[2] Jungbin Kim and Insoon Yang. Unifying nesterov's accelerated gradient methods for convex and strongly convex objective functions. In *International Conference on Machine Learning*, pages 16897–16954. PMLR, 2023.

[3] Hao Luo and Long Chen. From differential equation solvers to accelerated first-order methods for convex optimization. *Mathematical Programming*, pages 1–47, 2021.

[4] Yurii Nesterov. *Lectures on convex optimization*, volume 137. Springer, 2018.

[5] Weijie Su, Stephen Boyd, and Emmanuel J Candes. A differential equation for modeling nesterov's accelerated gradient method: Theory and insights. *Journal of Machine Learning Research*, 17:1–43, 2016.

[6] Adrien Taylor and Yoel Drori. An optimal gradient method for smooth strongly convex minimization. *Mathematical Programming*, pages 1–38, 2022.

[7] Bryan Van Scoy, Randy A Freeman, and Kevin M Lynch. The fastest known globally convergent first-order method for minimizing strongly convex functions. *IEEE Control Systems Letters*, 2(1):49–54, 2017.

[1] If the research process is of interest to you, I may add that I discovered the paper [3] when the writing of [2] was almost finished.