Revisiting the Role of Euler Numerical Integration on Acceleration and Stability in Convex Optimization

Peiyuan Zhang[1], Antonio Orvieto[1], Hadi Daneshmand[1,2], Roy Smith[1], Thomas Hofmann[1]

1: ETH Zurich, Switzerland 2: INRIA Paris, France

Main results

- A novel ODE model helps proving acceleration for multiple momentum methods;
- Both explicit and semi-implicit can generate accelerated methods;
- There is no direct relation between numerical stability and convergence speed.

Puzzling mechanism of acceleration in first-order gradient method

- Assume $\mu I \preceq \nabla^2 f \preceq LI$, *acceleration* indicates an algorithm, with iteration $\{x_k\}_{k=0}^{\infty}$, admits following convergence rate:

$$f(x_k) - f(x_0) \leq C(1-\sqrt{\mu/L})^k.$$

In contrast, gradient descent converges in a slow rate of $C(1-\mu/L)^k$.

- Most famous accelerated methods include Nesterov's accelerated gradient (NAG) and heavy-ball (HB)

$$x_{k+1} = x_k + \beta(x_k - x_{k-1}) - s\nabla f(x_k) \qquad \text{(HB)}$$

$$x_{k+1} = x_k + \beta(x_k - x_{k-1}) - s\nabla f(x_k) - \beta s(\nabla f(x_k) - \nabla f(x_{k-1})) \qquad \text{(NAG)}$$

- While NAG achieves global acceleration, HB is only guaranteed to accelerate locally. A question arises:

What is the mechanism behind this phenomena?

Continuous ODE for acceleration

- Recent research tackle this by studying the continuous limiting ODE of accelerated optimizers.
- Su et al. [1] studied second order damping ODE

$$\ddot{X} + 2\sqrt{\mu}\dot{X} + \nabla f(X) = 0 \qquad \text{(NAG-ODE)}$$

as continuous limit of both HB and NAG. This model does not reflect the difference between two iterations.

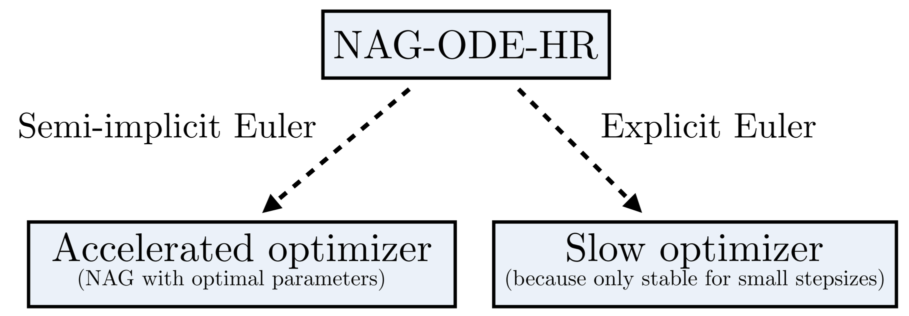
- To overcome this, Shi et al. [2] introduced

$$\ddot{X} + (2\sqrt{\mu} + \sqrt{s}\nabla^2 f(X))\dot{X} + C\nabla f(X) = 0 \qquad \text{(NAG-HR-ODE)}$$

for NAG to capture the difference between NAG and HB.

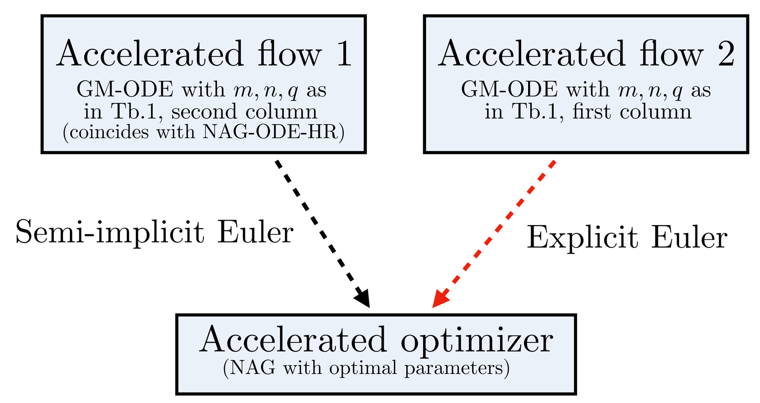
Old hypothesis: stable numerical integrator leads to acceleration

- Previous works like [3] suggest a close relationship between stability of numerical integrators and the acceleration of discrete iterations.



Our hypothesis: no direct relation between stability and acceleration

- Instead, this paper argues the choice of numerical integrator is not the sole decisive factor in achieving acceleration.



- To this end, a general ODE parameterized by (m, n, q) is proposed

$$\begin{cases} \dot{x} = -m\nabla f(X) - nV \\ \dot{V} = \nabla f(X) - qV. \end{cases} \qquad \text{(GM-ODE)}$$

Existing continuous models can be recasted as special case of GM-ODE.

Numerical integrator: bridge between continuous and discrete domain

- GM-ODE produces a rich class of momentum methods through semi-implicit and explicit Euler discretization

$$\text{EE}: \begin{cases} x_{k+1} - x_k = -m\sqrt{s}\nabla f(x_k) - n\sqrt{s}v_k \\ v_{k+1} - v_k = \sqrt{s}\nabla f(x_k) - q\sqrt{s}v_k. \end{cases}$$

$$\text{SIE}: \begin{cases} x_{k+1} - x_k = -m\sqrt{s}\nabla f(x_k) - n\sqrt{s}v_k \\ v_{k+1} - v_k = \sqrt{s}\nabla f(x_{k+1}) - q\sqrt{s}v_k. \end{cases}$$

- The generality of GM-ODE allows us to examine the relationship between numerical integrator and acceleration more rigorously. As a result, for example, NAG can be expressed as both EE or SIE numerical integrators of GM-ODE with different parameters.

- We prove that, when applied to GM-ODE, semi-implicit and explicit express the same discrete iteration if following condition holds:

$$m_{\text{EE}} = m_{\text{SIE}} + \sqrt{s}n_{\text{SIE}}, \quad n_{\text{EE}} = (1 - q\sqrt{s})n_{\text{SIE}}.$$

Both semi-implicit and explicit Euler accelerate

- Not surprisingly, we reconfirm the acceleration of semi-implicit Euler.

Theorem 1. *Under proper condition, the iteration $\{x_k\}_{k=1}^{\infty}$ generated by semi-implicit Euler discretization of GM-ODE enjoys acceleration*

$$f(x_k) - f(x^*) \leq C(1-\sqrt{\mu/L})^k.$$

- By the above equivalence, acceleration is also possible from explicit Euler.

Lemma 2. *Under proper condition, the iteration $\{x_k\}_{k=1}^{\infty}$ generated by explicit Euler discretization of GM-ODE enjoys acceleration*

$$f(x_k) - f(x^*) \leq C(1-\sqrt{\mu/L})^k.$$

- Therefore, we conclude that there is no direct relationship between stability of numerical integrator and acceleration. Explicit Euler is not inferior than semi-implicit in achieving acceleration.

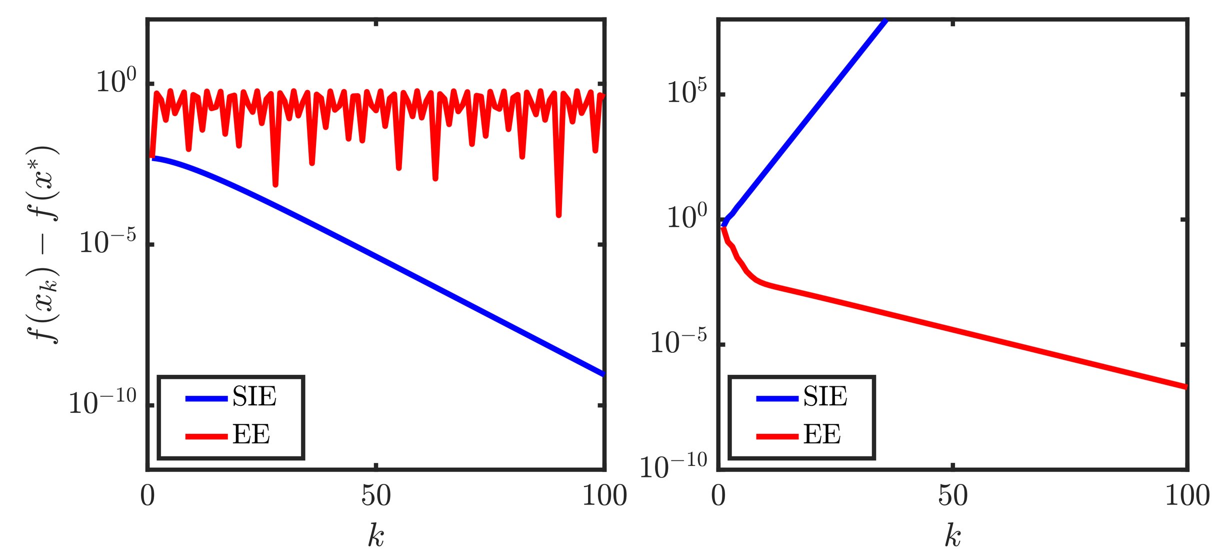


Figure 1: To show that SIE and EE are neither superior nor inferior to each other, in each subplot,we use the same parameters for both SIE and EE. This suggests the stability and convergence is determined by the joint choice of parameters and integrator together.

Reference

[1] Weijie Su, Stephen Boyd, and Emmanuel Candes. A differential equation for modeling Nesterov's accelerated gradient method. In NIPS 2014.

[2] Bin Shi, Simon S Du, Michael I Jordan, and Weijie Su. Understanding the acceleration phenomenon via high-resolution differential equations. As arXiv preprint.

[3] Bin Shi, Simon S Du, Weijie Su, and Michael I Jordan. Acceleration via symplectic discretization of high-resolution differential equations. In NIPS 2019.