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

Peiyuan Zhang, Antonio Orvieto, Hadi Daneshmand,
Thomas Hofmann, Roy Smith

ETH Zurich

April 13, 2021

Acceleration in first-order optimization

We consider the unconstrained smooth minimization problem

$$\min_{x \in \mathbb{R}^d} f(x).$$

Suppose f is L-smooth and μ-strongly convex, i.e.

$$\mu I \preceq \nabla^2 f(x) \preceq LI,$$

acceleration$\rightarrow$ iteration $\{x_k\}_{k=1}^{\infty}$, has following convergence rate

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

In contrast, gradient descent (GD) converges in a slow rate as

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

Puzzling mechanism of acceleration

Most famous accelerated first-order gradient methods:

- Heavy-ball (HB) — Polyak, 1964

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

- Nesterov's accelerated gradient (NAG) — Nesterov, 1983

$$x_{k+1} = x_k - s\nabla f(x_k) + \beta(x_k - x_{k-1}) + \underbrace{\beta(\nabla f(x_k) - \nabla f(x_{k-1}))}_{\text{Gradient correction}}.$$

- HB $\rightarrow$ local acceleration (around solution),
- NAG $\rightarrow$ global acceleration.

Mechanism behind Nesterov's acceleration has puzzled people in decades.

Acceleration from continuous perspective

- The ODE limit of GD is the gradient flow (GF), i.e.

$$\dot{X} = -\nabla f(X), \tag{GF}$$

  which converges in rate $O(e^{-\mu t})$.

- For accelerated optimizers, Su et al. (2015) studied second order damping ODE

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

  converging in rate $O(e^{-\sqrt{\mu} t})$, as limit of both HB and NAG. The model does not reflect the difference between two iters.

- To overcome this, Shi et al. (2019) introduced

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

  for NAG to capture the gradient correction.

Euler numerical integrators

Euler integrators are crucial when deriving discrete iterations from continuous dynamics. For a first order ODE with two variables $(X, V) := (X, \dot{X})$

$$\begin{cases} \dot{X} = g(X, V) \\ \dot{V} = h(X, V), \end{cases}$$

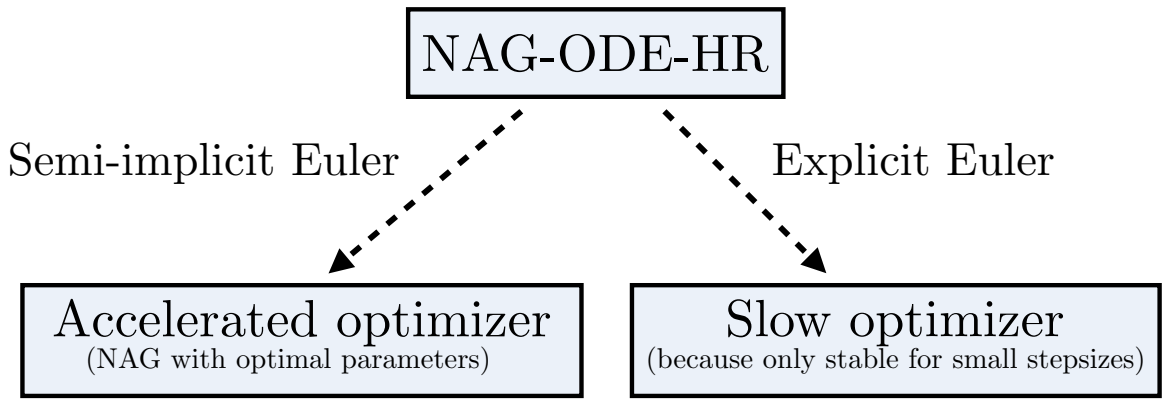
Euler integrator with step-size $\sqrt{s}$ has several variants of different properties:

Explicit:
$$\begin{cases} x_{k+1} - x_k = \sqrt{s} g(x_k, v_k) \\ v_{k+1} - v_k = \sqrt{s} h(x_k, v_k) \end{cases} \tag{EE}$$

Semi-implicit:
$$\begin{cases} x_{k+1} - x_k = \sqrt{s} g(x_k, v_k) \\ v_{k+1} - v_k = \sqrt{s} h(x_{k+1}, v_{k+1}). \end{cases} \tag{SIE}$$

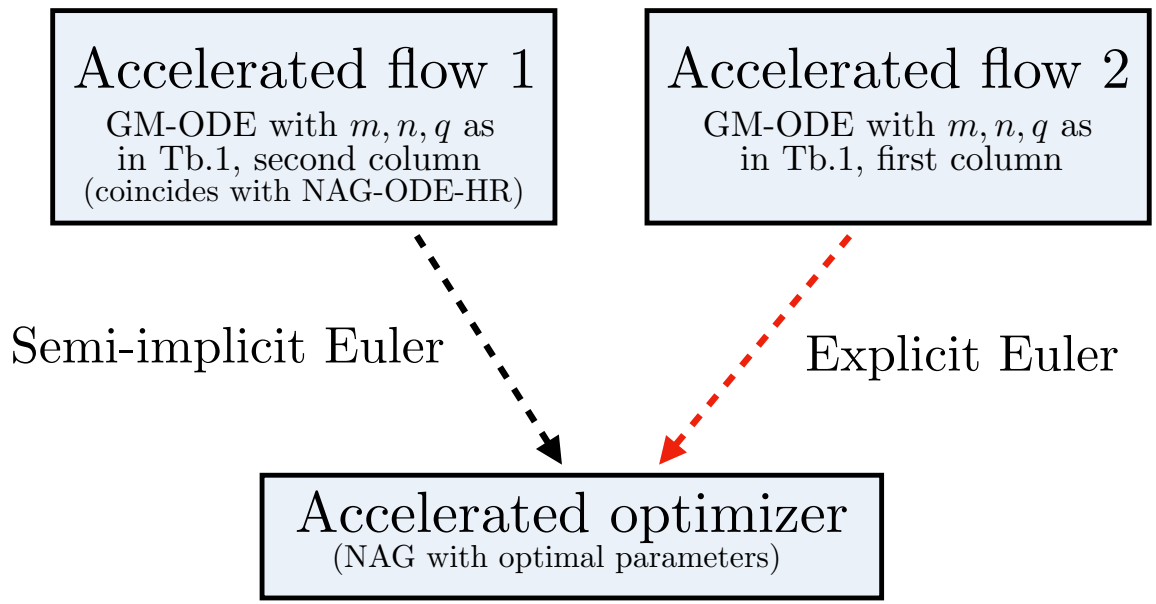
Acceleration as symplectic integration

Shi et al. (2019), Muehlebach & Jordan (2019) and França et al. (2020) suggest close relationship between stability (symplecticity) of numerical integrators and acceleration.



Our hypothesis

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



A novel model for momentum methods

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

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

The generality allows us to examine the relationship between numerical integrator and acceleration more rigorously.

	m	n	q
Gradient Flow	1	0	any
NAG-ODE	0	1	$2\sqrt{\mu}$
NAG-HR-ODE	$\sqrt{s}$	1	$2\sqrt{\mu}$

Table 1: Existing continuous model as special case of the novel GM-ODE.

Stability of continuous dynamics

With following Lyapunov function

$$L(x, v) = (qm + n)\big(f(x) - f(x^*)\big) + \frac{1}{4}\|q(x - x^*) - nv\|^2 + \frac{n(qm + n)}{4}\|v\|^2,$$

we prove that the trajectory (X, V) of GM-ODE asymptotically converges to the stable point $(x^*, 0)$ as

$$L(X(t), V(t)) \leq e^{-\gamma_1 t} L(X(0), V(0))$$

where $\gamma_1 := \min\left(\frac{\mu(n + qm)}{2q}, \frac{q}{2}\right)$.

Equivalence between integrators

Equivalence between semi-implicit and explicit Euler is proved by reparameterization.

Lemma 1. (Informal) *Any semi-implicit discretization of GM-ODE with parameters* $(m_{SIE}, n_{SIE}, q_{SIE})$ *can be viewed as explicit Euler discretization of GM-ODE with parameters* (m_{EE}, n_{EE}, q_{EE}) *if following condition holds:*

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

$\Longrightarrow$ **energy-preservation/geometric properties of semi-implicit (symplectic) integration are not strictly necessary to achieve acceleration.**

Claimed in Shi et al. (2019), Muehlebach & Jordan (2019) and França et al. (2020)

Semi-implicit is unsurprisingly accelerated

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

Thm 2. (Informal) *Assume f is L-Lipschitz and μ-strongly convex. If some parameter condition is satisfied, the iteration $\{x_k\}_{k=0}^{\infty}$ of* ***semi-implicit Euler*** *of GM-ODE yields accelerated convergence*

$$f(x_k) - f(x^*) \leq C_1(1 - C_2\sqrt{\mu/L})^k.$$

$\Longrightarrow$ The generality of GM-ODE allows to prove acceleration of novel momentum methods like QHM (Ma et al., (2018)).

Acceleration of Explicit Euler

$\rightarrow$ By equivalence, explicit Euler is **also** proved to accelerate.

Lemma 2. (Informal) *Assume f is L-Lipschitz and μ-strongly convex. If some parameter condition is satisfied, the iteration $\{x_k\}_{k=0}^{\infty}$ of **explicit Euler** of GM-ODE yields accelerated convergence*

$$f(x_k) - f(x^*) \leq C_1(1 - C_2\sqrt{\mu/L})^k.$$

$\Longrightarrow$ The old notion that

Explicit Euler is inferior in acceleration due to its unstable nature

is rejected by the above results.

Simple experimental validation

This is also verified by empirical results.

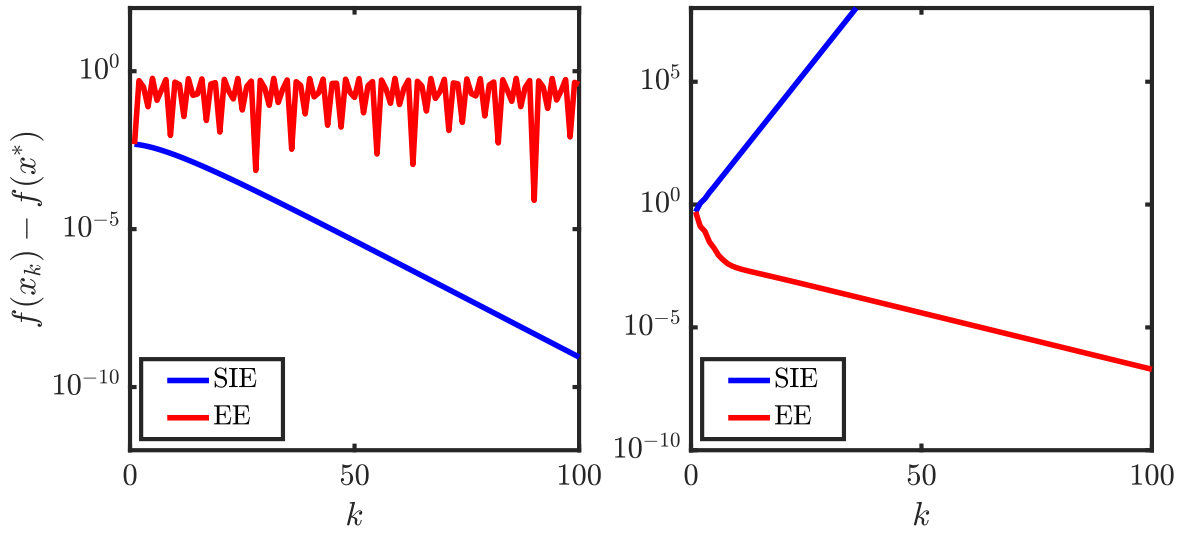


Figure 1: Quadratic loss: SIE vs. EE. To show SIE and EE are neither superior nor inferior to each other, in each subplot, we use the same parameters for both SIE and EE discretization.

$\Longrightarrow$ Stability and convergence is determined by joint choice of parameters and numerical integrator.

Conclusions

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

Thank you for your attention!