An International Journal of Optimization and Control: Theories & Applications
ISSN: 2146-0957 eISSN: 2146-5703
Vol.16, No.1, pp.232-245 (2026)
https://doi.org/10.36922/IJOCTA025390163

RESEARCH ARTICLE

# Beyond Nesterov: Dynamical systems perspective with time-dependent inertia and conformable Bregman flows

Osama F. Abdel Aal [1†], Necdet Sinan Ozbek [1,2†], Jairo Viola [1†], and YangQuan Chen [1†*]

[1]Department of Mechanical and Aerospace Engineering, School of Engineering, University of California Merced, California, United States of America
[2]Department of Electrical and Electronic Engineering, Faculty of Engineering, Adana Alparslan Turkes Science and Technology University, Adana, Turkiye
*oabdelaal@ucmerced.edu, nozbek@atu.edu.tr, jviola@ucmerced.edu, ychen53@ucmerced.edu*

ARTICLE INFO

Article History:
Received: September 25, 2025
Revised: November 6, 2025
Accepted: November 10, 2025
Published Online: December 16, 2025

Keywords:
Accelerated Gradient Methods
Continuous optimization
Bregman lagrangian
Calculus of variation
Gradient algorithms
Conformable derivatives

AMS Classification 2010:
*26A33; 34A08; 35H15; 34K50 47H10; 60H10*

ABSTRACT

We present a Lyapunov-based framework for analyzing continuous-time accelerated optimization dynamics with time-dependent inertia and damping. By explicitly designing Lyapunov functions that account for varying inertia, we rigorously characterize convergence rates of the objective function, achieving exponential or polynomial acceleration beyond the classical $\mathcal{O}(1/t^2)$, even in the absence of strong convexity. Building on this foundation, we introduce a variational extension using conformable (fractional) derivatives in the Lagrangian formulation, replacing the classical velocity term with a time-weighted fractional velocity. This approach systematically modulates the system's effective inertia and damping, providing a principled mechanism to balance acceleration and stability, reduce oscillations, and interpolate smoothly between strongly damped gradient flows and momentum-driven dynamics. The resulting framework unifies Lyapunov analysis and fractional variational modeling, offering flexible, theoretically grounded design principles for fast and stable accelerated optimization.



## 1. Introduction

The shift toward analyzing gradient methods within a continuous-time framework offers a powerful lens for understanding optimization algorithms and their convergence behavior. In this setting, gradient flows are described by differential equations that naturally capture the underlying dynamics of optimization processes. This perspective establishes a robust analytical foundation for exploring the principles that govern these algorithms. By abstracting away the discretization artifacts typical of iterative methods, continuous-time models provide an effective means of examining stability, convergence rates, and critical phenomena in gradient-based techniques.

The work of by Su et al.[1] marked a significant advancement in interpreting optimization algorithms through this lens. A particularly important contribution of continuous-time models lies in their ability to describe accelerated methods, such as Nesterov's accelerated gradient (NAG) descent. Continuous analogs—including the Bregman Lagrangian and Polyak's heavy-ball method—expose the geometric structure and intricate dynamics of these techniques. Building on this, Su et al.[1] paved the way for subsequent research aimed at developing unified frameworks that encompass a wide range

*Corresponding Author

of algorithms.[bib2,bib3] Among these, the Bregman–Lagrangian approach[3] stands out for employing the Euler–Lagrange formulation to derive a general equation unifying multiple methods. Other works have focused on sharpening the characterization of convergence dynamics for accelerated optimization algorithms, leading to more precise descriptions of their time evolution.[4-8]

Mathematically, the shift toward continuous-time formulations provides greater flexibility in the analysis and design of optimization algorithms, machine learning (ML), and deep learning (DL), owing to the rich theory of ordinary differential equations (ODEs) and their widespread use in physical modeling. This perspective enables the incorporation of constraints, structural modifications, adaptation mechanisms, and redesign of the underlying dynamical systems. For instance, Weinan[9] and Le et al.[10] demonstrated that deep neural networks (DNNs) can be interpreted as discretizations of continuous dynamical systems. From this viewpoint, new possibilities emerge: adaptive step sizes can be associated with adaptively selected layers, higher-order or implicit discretization schemes can be employed, and advanced numerical techniques can be integrated into DNN training. Moreover, continuous-time analysis sheds light on the convergence of adaptive algorithms to critical points in both convex and nonconvex landscapes, as well as the conditions necessary to avoid saddle points and local maxima.[11,12]

Given the importance of continuous-time modeling, a substantial body of work has focused on developing more accurate formulations that capture the true dynamics of optimization methods while also designing iterative algorithms that preserve these dynamics. Symplectic discretization has recently emerged as a promising approach in this direction.[4,13,14]

Beyond providing analytical insights, these frameworks foster the development of novel optimization strategies by incorporating tools from dynamical systems theory. Notable examples include the use of Lyapunov functions and invariants to study stability and convergence,[15,16] as well as the synthesis of accelerated algorithms through state-dependent dynamics.[17-19]. Additionally, time-dependent and time-scaling techniques have been proposed to further enhance algorithmic performance.[20-23]

In control theory, Lyapunov analysis plays a fundamental role, offering a powerful framework for establishing the stability of dynamical systems by examining how a chosen Lyapunov candidate decreases along system trajectories. Within the context of optimization and gradient descent algorithms, the incorporation of suitable Lyapunov-like functions enables one to bound errors and guarantee convergence of the optimization trajectory to a local or global minimum, provided certain conditions—such as convexity or strong convexity—are satisfied.

Beyond stability analysis, Lyapunov theory introduces a valuable dimension to the design of optimization methods. By constructing appropriate time-dependent Lyapunov-like energy functions, it becomes possible not only to analyze convergence rates but also to redesign algorithms to achieve improved performance. This perspective has led to the development of enhanced convergence guarantees and accelerated schemes.[24-28]

In recent years, the study of convergence behavior and stability conditions in optimization and ML algorithms has become increasingly critical, as these factors are central to ensuring performance quality, reliability, and robustness. Drawing on fundamental stability notions from control theory—such as Lyapunov stability and input-to-state stability—offers a powerful framework for both analyzing and designing such algorithms.[29-30]

From the perspective of optimization, control-theoretic tools provide insights into key properties, including convergence rates, long-term dynamics, and robustness to uncertainties in state variables or gradient evaluations. In the ML setting, incorporating control theory has proven particularly useful in addressing challenges in neural network training and reinforcement learning, where it helps uncover the dynamic nature of learning algorithms and their sensitivity to fluctuating hyperparameters or system noise. Moreover, assessing algorithmic performance requires understanding its behavior around critical points of the loss function. Stability, in this context, is tied to whether the algorithm converges toward minima and how the surrounding geometry influences its trajectory.

Variational principles and generalized calculus have long provided powerful tools for modeling and controlling complex dynamical systems,[31] with fractional calculus in particular proving effective in capturing memory and hereditary effects.[32] Building on these insights, this work introduces a fractional-order differential equation as a continuous-time analogue of the gradient descent optimization algorithm. The formulation arises from a fractional extension of the Euler–Lagrange equation, with the objective of investigating whether the dynamics induced by fractional operators can achieve performance beyond Nesterov's accelerated scheme by leveraging their inherent flexibility.[?]

This study proposes two continuous-time dynamical frameworks that extend NAG method to achieve faster and more stable convergence. The first approach introduces a time-dependent inertia term into the ODE formulation of Nesterov's method and leverages Lyapunov theory to systematically design the inertia–time weighting and vanishing damping functions required for convergence rates faster than the classical $\mathcal{O}(1/t^2)$. This Lyapunov-based formulation provides an elegant and rigorous mechanism for constructing higher-order acceleration in optimization dynamics. The second approach employs the conformable derivative within a Bregman–Lagrangian framework, yielding a generalized Euler–Lagrange equation that naturally incorporates time-weighted dynamics. This conformable framework offers a systematic way to tune damping, allowing an explicit trade-off between convergence speed and stability of the optimization trajectory. Together, these results establish a unified foundation for designing next-generation accelerated algorithms that outperform classical schemes in both rate and robustness.

The remainder of the paper is structured as follows. Section 2 introduces the required definitions and preliminaries. Section 3 reviews the differential equation used for modeling Nesterov's accelerated gradient method (NAG). Section 4 introduces a modified NAG method as an approach toward higher rates of convergence. In Section 5, we present the conformable Bregman flows as a consistent method for deriving a time-weighted inertia term in the NAG model by extending the Euler–Lagrange framework with the conformable derivative. Section 6 presents numerical experiments on benchmark optimization functions, comparing the proposed approach against standard and accelerated gradient methods. Finally, Section 7 concludes the paper.

## 2. Preliminaries

In this section, we recall several fundamental notions that will be employed throughout the paper, including the Bregman Lagrangian, convexity, and the conformable derivative.

### 2.1. The Bregman divergence

**Definition 1.** *Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a strictly convex and continuously differentiable function, often referred to as a* distance-generating function.[34] *The associated* Bregman divergence *between two points $x, y \in \mathbb{R}^n$ is defined as:*

$$R_\phi(x, y) = \phi(x) - \phi(y) - \langle \nabla\phi(y), x - y\rangle \quad (1)$$

**Remark 1.** *$R_\phi(p, q)$ is a non-negative operator when $\phi(x)$ is convex.*[34]

**Remark 2.** *Unlike a metric, $R_\phi$ is generally not symmetric and does not satisfy the triangle inequality. Nevertheless, it serves as a powerful measure of discrepancy between points, tightly linked to the geometry induced by $\phi$. For instance, when $\phi(x) = \frac{1}{2}\|x\|_2^2$, the Bregman divergence reduces to the squared Euclidean distance.*[35]

Bregman divergences play a central role in modern optimization, particularly in the design of first-order methods.[36] They provide the foundation for *mirror descent* algorithms,[37-39] where updates are performed relative to a non-Euclidean geometry tailored to the problem structure. They also arise in *Bregman proximal methods*, variational inequalities, and information geometry, where specific choices of $\phi$ (e.g., negative entropy) yield divergences such as the Kullback–Leibler divergence. In accelerated methods and variational formulations, Bregman divergences naturally appear as Lagrangian potentials that encode problem-dependent geometries, enabling algorithms to exploit curvature beyond the Euclidean setting.[3,36]

### 2.2. Convexity

**Definition 2.** *A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be convex if its domain is a convex set and if, for all $x_1, x_2 \in dom(f)$ and for all $\theta \in [0, 1]$, the following inequality holds:*[40]

$$f(\theta x_1 + (1 - \theta)\mathbf{x}_2) \leq \theta f(x_1) + (1 - \theta)f(x_2) \quad (2)$$

**Definition 3.** *A function $f : \mathbb{R}^n \to \mathbb{R}$ is called* convex *if its domain is convex and $\forall x_1, x_2 \in$* dom$(f)$*:*[40]

$$f(x_2) \geq f(x_1) + \langle \nabla f(x_1), x_2 - x_1\rangle, \quad (3)$$

**Remark 3.** *This first-order characterization states that the graph of a convex function always lies above its supporting hyperplanes. Convexity ensures that any local minimum of $f$ is also a global minimum, a property that underpins much of modern optimization theory.*

### 2.3. The Conformable derivative

The *conformable derivative* offers a local approach to fractional differentiation that is simpler in structure than classical fractional derivatives, yet preserves many of their key analytical features. It is particularly useful in modeling systems where memory effects or non-integer order dynamics are present, but where a simpler, more tractable derivative is desired.

For a function $f : [0, \infty) \to \mathbb{R}$ and order $\nu \in (0, 1]$, the conformable derivative provides a smooth interpolation between the classical derivative and fractional derivatives. Intuitively, it modifies the increment in the standard difference quotient by a factor that depends on the current point, which allows it to capture scaling properties inherent in fractional dynamics.

**Definition 4.** *Let $f : [0, \infty) \to \mathbb{R}$ be a function. The conformable $\nu$-th derivative at a point $t > 0$ is defined by the limit:*[41]

$$T_\nu(f)(t) = \lim_{\epsilon \to 0} \frac{f(t + \epsilon t^{1-\nu}) - f(t)}{\epsilon} \tag{4}$$

*This definition generalizes the classical derivative, as for $\nu = 1$, it reduces to the ordinary derivative of $f$ at $t$.*

**Lemma 1.** *If $f$ is differentiable at $t$, the conformable derivative can be rewritten in a form that directly involves the standard derivative:*[41]

$$T_\nu(f)(t) = t^{1-\nu} \frac{\mathrm{d}f}{\mathrm{d}t}(t) \tag{5}$$

***Proof.*** For completeness and to make the article self-contained, we briefly outline the proof (See Ref.[41]) Introducing $h = \epsilon t^{1-\nu}$, which implies $\epsilon = t^{\nu-1} h$, we acquired:

$$\begin{aligned} T_\nu(f)(t) &= \lim_{\epsilon \to 0} \frac{f(t + \epsilon t^{1-\nu}) - f(t)}{\epsilon} \\ &= \lim_{h \to 0} \frac{f(t+h) - f(t)}{h\, t^{\nu-1}} \\ &= t^{1-\nu} \lim_{h \to 0} \frac{f(t+h) - f(t)}{h} \\ &= t^{1-\nu} \frac{\mathrm{d}f}{\mathrm{d}t}(t), \end{aligned} \tag{6}$$

which completes the proof. □

**Remark 4.** *This formulation allows a straightforward extension of classical calculus rules. For instance, the chain rule for the conformable derivative takes the form:*[42]

$$\mathcal{T}_\nu(f \circ g)(t) = t^{1-\nu} \frac{\mathrm{d}f}{\mathrm{d}g} g'(t) = \frac{\mathrm{d}f}{\mathrm{d}g} \mathcal{T}_\nu(g(t)) \tag{7}$$

*which provides a systematic method for computing the derivative of composed functions in the conformable fractional calculus framework.*

Overall, the conformable derivative serves as a bridge between classical and fractional calculus, offering a flexible tool for analyzing systems that exhibit both local and scaling behaviors.

## 3. A Differential eEquation for modeling Nesterov's Accelerated Gradient method

Nesterov's seminal work[43] introduced an accelerated gradient scheme defined by the coupled recursions:

$$\begin{aligned} \chi_m &= y_{m-1} - h \nabla \Phi(y_m), \\ y_m &= \chi_m + \frac{m-1}{m+2} \big(\chi_m - \chi_{m-1}\big) \end{aligned} \tag{8}$$

where $h > 0$ denotes the step size. For the specific choice $h = 1/L_c$, with $L_c$ denoting the Lipschitz constant of $\nabla\Phi$, the method achieves the optimal convergence rate in the first-order oracle model.[44-46] The acceleration effect arises from the inclusion of the momentum term $\chi_m - \chi_{m-1}$, combined with the carefully tuned coefficient $\frac{m-1}{m+2} = 1 - \frac{3}{m}$.

Su et al.[1] demonstrated that the discrete scheme Equation (8) admits a continuous-time approximation governed by the second-order differential equation:

$$\ddot{X}(t) + \frac{3}{t} \dot{X}(t) + \nabla \Phi(X(t)) = 0 \tag{9}$$

with the correspondence between the discrete step size and continuous time given by $h = m\sqrt{t}$.

The recursion in Equation (8) may also be rewritten as:

$$\frac{\chi_{m+1} - \chi_m}{\sqrt{h}} = \frac{m-1}{m+2} \frac{\chi_m - \chi_{m-1}}{\sqrt{h}} - \sqrt{h}\, \nabla \Phi(y_m) \tag{10}$$

By introducing the interpolation $X(t) = \chi_{t/\sqrt{h}} = \chi_m$, with $X(t + \sqrt{h}) = \chi_{m+1}$, and applying Taylor expansions, one obtains:

$$\begin{aligned} \frac{\chi_{m+1} - \chi_m}{\sqrt{h}} &= \dot{X}(t) + \tfrac{1}{2} \ddot{X}(t) \sqrt{h} + O(\sqrt{h}), \\ \frac{\chi_m - \chi_{m-1}}{\sqrt{h}} &= \dot{X}(t) - \tfrac{1}{2} \ddot{X}(t) \sqrt{h} + O(\sqrt{h}), \\ \sqrt{h}\, \nabla \Phi(y_m) &= \sqrt{h}\, \nabla \Phi(X(t)) + O(\sqrt{h}) \end{aligned} \tag{11}$$

Substituting these expansions into Equation (10) yields the continuous formulation:

$$\begin{aligned} &\dot{X}(t) + \tfrac{1}{2} \ddot{X}(t) \sqrt{h} \\ &= \Big(1 - \tfrac{3\sqrt{h}}{t}\Big) \Big(\dot{X}(t) - \tfrac{1}{2} \ddot{X}(t) \sqrt{h}\Big) - \sqrt{h}\, \nabla \Phi(X(t)) \end{aligned} \tag{12}$$

In general, a family of Nesterov-type accelerated schemes can be described by the continuous model:

$$\ddot{X}(t) + \frac{r}{t}\dot{X}(t) + \nabla\Phi(X(t)) = 0, \tag{13}$$

where $r > 0$ is a damping parameter. For comparison, the standard gradient descent method:

$$\chi_{m+1} = \chi_m - \mu\nabla\Phi(\chi_m), \tag{14}$$

corresponds in the continuous limit to the first-order system:

$$\dot{X}(t) + \nabla\Phi(X(t)) = 0. \tag{15}$$

## 4. A modified Nestrov model for Accelerated Gradient methods: Toward higher rates

### 4.1. Problem formulation

To formulate the problem setting, consider the unconstrained optimization problem:

$$\min_{x\in\mathbb{R}^n} \quad f(x), \tag{16}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is the cost function. The following assumptions are introduced:

**Assumption 1.** *The function $f$ is assumed to attain its minimum value $f^* = f(x^*) = \min f(x) > -\infty$ at $x^* \in \mathbb{R}^n$, i.e., $f^* > -\infty$*

**Assumption 2.** *The objective function $f : \mathbb{R}^n \to \mathbb{R}$ is convex; that is, for all $x, y \in \mathbb{R}^n$,*

$$f(y) \geq f(x) + \langle \nabla f(x),\, y - x\rangle, \quad \forall x, y \in \mathrm{dom}(f) \tag{17}$$

**Remark 5.** *This convexity assumption ensures that any local minimum of $f$ is also a global minimum, thereby allowing the use of convex optimization methods to efficiently find the optimal solution.*

We investigated the convergence properties of continuous-time accelerated optimization models under general conditions on the damping and time-scaling coefficients. Our analysis was grounded in a newly formulated Lyapunov framework, which enables a rigorous characterization of the rate of convergence of the objective function values. In particular, we focused on a modified continuous-time model incorporating a time-dependent inertia term. The introduction of this inertia term effectively enhanced the system dynamics, enabling exponential convergence of the objective function values and allowing the Lyapunov design to achieve polynomial rates surpassing the classical $\mathcal{O}(1/t^2)$ associated with NAG method,[44] even in the absence of strong convexity assumptions. By explicitly tailoring the Lyapunov function to account for the time-varying inertia, the proposed approach ensures stability while attaining these accelerated convergence rates.

The dynamics of the ODE in Equation (13) can be interpreted as a continuous-time dissipative system, in which the inertia is initially unity, the damping coefficient diminishes over time, and the gradient functions as a nonlinear restoring force. To attain accelerated convergence, we extended this framework by introducing a time-dependent mass (inertia) term into the dynamical equation, thereby modulating the system's effective inertia throughout its evolution.

$$m(t)\ddot{X} + \mu(t)\dot{X} + \nabla f(X) = 0, \tag{18}$$

where the coefficients $m, \mu : [t_0, +\infty) \to \mathbb{R}_{\geq 0}$ represent the inertia and viscous damping coefficients, respectively, and are continuously differentiable time functions.

**Remark 6.** *The well-posedness of the corresponding Cauchy problem associated with Equation (18), given the initial conditions $x(t_0) = x_0 \in \mathbb{R}^n, \dot{x}(t_0) = \dot{x}_0 \in \mathbb{R}^n$, is ensured. Provided that $\nabla f$ is Lipschitz continuous over bounded domains and the coefficients are continuously differentiable, the local existence of the solution follows directly from the nonautonomous form of the Cauchy–Lipschitz theorem.then analyzed*

We then analyzed the convergence behavior of the proposed dynamics using Lyapunov theory. The goal was to construct a time-varying energy function whose decay rate guarantees a desired convergence speed.

### 4.2. Convergence rate: A lyapunov approach

In this section, Lyapunov theory is used to derive new results by establishing a set of inequalities whose solutions guarantee a desired rate of convergence. This approach not only enhances the theoretical understanding of accelerated optimization but also guides the development of more powerful and faster algorithms. With suitably designed Lyapunov functions, one can obtain explicit convergence rates and provide theoretical guarantees for the proposed modifications to NAG.

Define the following energy-like function:

$$V = c(t)[f - f^*] + \frac{1}{2}\|\sigma(t)[x - x^*] + \eta(t)\dot{x}\|^2, \tag{19}$$

where $c(t) : [t_0, +\infty) \to \mathbb{R}_{\geq 0}$ is an increasing time function, and $\sigma(t), \eta(t)$ are time-scaling factors. We aimed for $f - f^*$ to have a rate of convergence of $\mathcal{O}(1/c(t))$ to ensure an upper-bounded energy function. We differentiated the Lyapunov function with respect to time:

$$\begin{aligned}\dot{V} &= c(t)\dot{x}^T\nabla f(x) + \dot{c}(t)[f - f^*] \\ &+ \sigma(t)\dot{\sigma}(t)\|x - x^*\|^2 - \frac{\eta(t)}{m(t)}\dot{x}^T\nabla f(x) \\ &+ \sigma(t)\left[\sigma(t) + \dot{\eta}(t) - \eta(t)\frac{\mu(t)}{m(t)}\right]\dot{x}^T(x - x^*) \\ &+ \dot{\sigma}(t)\eta(t)\dot{x}^T(x - x^*) - \sigma(t)\frac{\eta(t)}{m(t)}\nabla f(x)^T(x - x^*) \\ &+ \eta(t)\left[\sigma(t) + \dot{\eta}(t) - \eta(t)\frac{\mu(t)}{m(t)}\right]\|\dot{x}\|^2.\end{aligned} \tag{20}$$

Applying the convexity property ((Equation 17) and setting all mixed terms in (Equation 20) to zero, one obtains:

$$\dot{V} \leq \left[\dot{c}(t) - \sigma\frac{\eta(t)}{m(t)}\right](f - f^*), \tag{21}$$

subject to:

$$\begin{aligned}&\sigma(t) + \dot{\eta}(t) - \eta(t)\frac{\mu(t)}{m(t)} = 0, \\ &\dot{c}(t) - \sigma\frac{\eta(t)}{m(t)} \leq 0, \\ &c(t) = \frac{\eta(t)^2}{m(t)}, \\ &\dot{\sigma}(t) = 0\end{aligned} \tag{22}$$

**Theorem 1.** *If the function $f$ satisfies Assumptions 1 and 2, given the inequalities in (Equation 22), the convergence rate of the algorithm given in (Equation 18) is:*

$$f(x(t)) - \min_{x\in\mathbb{R}^n} f(x(t)) = \mathcal{O}\left(\frac{1}{c(t)}\right) \tag{23}$$

***Proof.*** Note that $\dot{\sigma}(t) = 0$ implies that $\sigma$ is constant, while the second and third equations can be combined into $\dot{c} \leq \sigma\frac{c(t)}{\eta(t)}$. Thus, under the derived conditions in (Equation 22), the energy-like function $V(\cdot)$ is nonnegative and nonincreasing. Therefore, for all $t \geq t_0$, $V(t) \leq V(t_0)$, which implies:

$$c(t)[f - f^*] \leq V(t_0) \implies f - f^* = \mathcal{O}\left(\frac{1}{c(t)}\right) \tag{24}$$

□

The derived inequalities provide a foundation for designing time-dependent parameters that achieve faster convergence, as discussed next.

### 4.3. Higher convergence rates by design

One of the most significant parameters influencing the behavior of accelerated gradient methods is the inertia term, which determines the influence of past gradients on future updates. Classical NAG formulations employ a fixed momentum coefficient, which is not necessarily ideal for all problem structures. However, recent studies reveal that using adaptive momentum schemes can improve both stability conditions and the convergence rate.

Introducing a time-weighted inertia into the dynamics of NAG model for convex optimization allows one to obtain higher convergence rates than the classical $\mathcal{O}(1/t^2)$. Time-weighted inertia dynamically adjusts the acceleration constant based on the optimization trajectory.

Interestingly, the differential inequalities in Equation (22) can be solved for the scaling factor $\eta(t)$ and the inertia-weighting time function $m(t)$ to satisfy a predesigned convergence-rate function $c(t)$ such that:

$$f(x(t)) - \min_{x\in\mathbb{R}^n} f(x(t)) = \mathcal{O}\left(\frac{1}{c(t)}\right) \tag{25}$$

Consider a polynomial rate of convergence, we designed our convergence rate based on $c(t) = t^\rho$, where $\rho > 2$. Upon solving the inequalities in (Equation 22), one finds the required scaling factor such that $f - f^* = \mathcal{O}\left(\frac{1}{t^\rho}\right)$. By selecting $\eta(t) = \frac{\sigma}{\rho}t$, the time-dependent inertia term can be expressed as:

$$m(t) = \left(\frac{\sigma}{\rho}\right)^2 t^{2-\rho} \tag{26}$$

The damping term can then be designed as:

$$\mu(t) = m(t)\frac{\dot{\eta}(t) + \sigma}{\eta(t)} = \frac{\sigma^2(\rho + 1)}{\rho^2}t^{1-\rho}. \tag{27}$$

The combination of Equations (27) and (26) satisfies the coupled differential inequality in (Equation 22). In Nesterov's theory of accelerated methods, one key property for achieving fast convergence for a general convex function $f$ is that the dynamics involve an asymptotically vanishing damping coefficient, i.e., $\mathcal{D}(t) \to 0$ as $t \to \infty$, where $\mathcal{D}(t)$ is the damping term defined as:

$$\mathcal{D}(t) = \frac{\mu(t)}{m(t)} = \frac{\dot{\eta}(t) + \sigma}{\eta(t)}. \tag{28}$$

Now, to design an exponential rate of convergence, we set $c(t) = e^{\lambda t}$, where $\lambda > 0$, such that $f - f^* = \mathcal{O}(e^{-\lambda t})$. Upon choosing the factor $\eta(t) = rt$, the inertia scaling becomes:

$$m(t) = \frac{\eta(t)^2}{c(t)} = r^2t^2e^{-\lambda t} \tag{29}$$

The next section introduces a consistent way to modulate the inertia dynamics via the conformable Bregman flow, which modifies the Bregman Lagrangian with a time-weighted derivative of the state variable within the Euler–Lagrange variational framework, providing a systematic way to control the trade-off between fast convergence and the corresponding oscillatory behavior.

## 5. Time-weighted derivatives: A variational model of accelerated methods in optimization

In this section, we introduce a variational framework for accelerated optimization dynamics based on time-weighted (fractional) derivatives defined through the conformable derivative operator. By replacing the classical first-order velocity term in the Lagrangian formulation with a conformable fractional derivative, we obtained a generalized Euler–Lagrange system that naturally extends traditional continuous-time models of Nesterov-type acceleration. This fractional velocity term introduces an adaptive, time-dependent scaling of the damping coefficient, effectively modulating the system's inertia in a systematic and analytically consistent manner. As a result, the proposed model provides a principled mechanism to balance acceleration and stability, mitigating the oscillatory behaviors often observed in classical accelerated methods while preserving their fast convergence characteristics. Moreover, this formulation offers a unified variational perspective from which inertial effects can be tuned continuously via the fractional order, enabling a smooth interpolation between strongly damped gradient flows and momentum-driven dynamics.

### 5.1. Conformable bregman flows

The Bregman divergence employs a strictly convex function $\phi(x)$ to measure the distance between two points $p$ and $q$, defined as:[3]

$$R_\phi(p,q) = \phi(p) - \phi(q) - \langle \nabla\phi(q), p-q\rangle , \quad (30)$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product.

Building on this concept, Wibisono *et al.* [3] introduced the Bregman Lagrangian, a continuous-time variational principle that unifies a broad class of accelerated optimization methods. For time-dependent scaling $\alpha(t)$, damping $\beta(t)$, and weighting $\gamma(t)$, the Bregman Lagrangian associated with a convex objective $f$ is:[3]

$$\mathcal{L}(X,V,t) = e^{\alpha(t)+\gamma(t)}\Big[R_\phi(X+e^{-\alpha(t)}V, X) - e^{\beta(t)}f(X)\Big] \quad (31)$$

Here $X\in\mathbb{R}^n$, $V$ is the velocity, and $t$ is the continuous-time variable.

We extended the Bregman Lagrangian to a velocity term of fractional order $\nu$:

$$\mathcal{L}(X,\mathcal{T}_\nu(X),t) = e^{\alpha_t+\gamma_t}\Big[R_\phi\big(X+e^{-\alpha_t}\mathcal{T}_\nu(X), X\big) - e^{\beta_t}f(X)\Big], \quad (32)$$

where $\mathcal{T}_\nu(\cdot)$ denotes the conformable fractional derivative of order $\nu$.

Defining the function:

$$J(X) = \int_{\mathcal{T}} \mathcal{L}\big(X,\mathcal{T}_\nu(X),t\big)\,dt \quad (33)$$

we investigated the continuous model describing the Nesterov scheme (see Equation 10) using the fractional Euler–Lagrange equation:[42,48]

$$\frac{\partial\mathcal{L}}{\partial X} - \mathcal{T}_\nu\left(\frac{\partial\mathcal{L}}{\partial\mathcal{T}_\nu(X)}\right) = 0. \quad (34)$$

The partial derivatives are:

$$\frac{\partial\mathcal{L}}{\partial X} = e^{\gamma_t+\alpha_t}\Big[\nabla\Phi(X+e^{-\alpha_t}\mathcal{T}_\nu(X)) - \nabla\Phi(X) - e^{-\alpha_t}\mathcal{T}_\nu(X)\nabla^2\Phi(X) - e^{\beta_t}\nabla f(X)\Big], \quad (35)$$

and

$$\frac{\partial\mathcal{L}}{\partial\mathcal{T}_\nu(X)} = e^{\gamma_t}\big[\nabla\Phi(X+e^{-\alpha_t}\mathcal{T}_\nu(X)) - \nabla\Phi(X)\big] \quad (36)$$

Evaluating the second term in the Euler–Lagrange equation equation, Equation 36 gives:

$$\begin{aligned}
&\mathcal{T}_\nu\Big(\frac{\partial\mathcal{L}}{\partial\mathcal{T}_\nu(X)}\Big)\\
&= \mathcal{T}_\nu(e^{\gamma_t})\Big[\nabla\Phi(X+e^{-\alpha_t}\mathcal{T}_\nu(X)) - \nabla\Phi(X)\Big]\\
&+ e^{\gamma_t}\nabla^2\Phi(X+e^{-\alpha_t}\mathcal{T}_\nu(X))\Big\{\mathcal{T}_\nu(X)\\
&+ \mathcal{T}_\nu(e^{-\alpha_t})\,\mathcal{T}_\nu(X) + e^{-\alpha_t}\,\mathcal{T}_\nu\mathcal{T}_\nu(X)\Big\}\\
&- e^{\gamma_t}\nabla^2\Phi(X)\mathcal{T}_\nu(X)
\end{aligned} \quad (37)$$

Substituting Equations 35–37 yields:

$$\begin{aligned}
&\mathcal{T}_\nu\mathcal{T}_\nu(X) + e^{\alpha_t}\mathcal{T}_\nu(X)\big[\mathcal{I} - \mathcal{M}\nabla^2\Phi(X)\big]\\
&+ e^{\alpha_t}\mathcal{T}_\nu(X)\big[\mathcal{T}_\nu(e^{-\alpha_t}) - \mathcal{M}\nabla^2\Phi(X)\big]\\
&+\mathcal{M}e^{\alpha_t}\big(e^{\alpha_t} - e^{-\gamma_t}\mathcal{T}_\nu(e^{\gamma_t})\big)\big[\nabla\Phi(X+e^{-\alpha_t}\mathcal{T}_\nu(X))\\
&- \nabla\Phi(X)\big] + \mathcal{M}e^{\beta_t+2\alpha_t}\nabla f(X) = 0.
\end{aligned} \quad (38)$$

where:

$$\mathcal{M} = \big[\nabla^2\Phi(X+e^{-\alpha_t}\mathcal{T}_\nu(X))\big]^{-1} \quad (39)$$

In Euclidean space, $\nabla^2\Phi(X) = \mathcal{I}$, hence:

$$\begin{aligned}
&\mathcal{T}_\nu\mathcal{T}_\nu(X) + e^{\alpha_t}\mathcal{T}_\nu(X)\big[\mathcal{T}_\nu(e^{-\alpha_t}) + 1\big]\\
&+\mathcal{M}e^{\alpha_t}\big(e^{\alpha_t} - e^{-\gamma_t}\mathcal{T}_\nu(e^{\gamma_t})\big)\big[\nabla\Phi(X+e^{-\alpha_t}\mathcal{T}_\nu(X))\\
&- \nabla\Phi(X)\big] + \mathcal{M}e^{\beta_t+2\alpha_t}\nabla f(X) = 0.
\end{aligned} \quad (40)$$

Using the scaling condition and chain rule (Equation 7),

$$e^{\alpha_t} = e^{-\gamma_t}\mathcal{T}_\nu(e^{\gamma_t}) = t^{1-\nu}\dot{\gamma}_t, \tag{41}$$

Equation 40 becomes:

$$\begin{aligned}\mathcal{T}_\nu\mathcal{T}_\nu(X) + e^{\alpha_t}\mathcal{T}_\nu(X)\big[\mathcal{T}_\nu(e^{-\alpha_t}) + 1\big] \\ + \mathcal{M}e^{\beta_t+2\alpha_t}\nabla f(X) = 0\end{aligned} \tag{42}$$

Incorporating Equations 6 and 7, Equation 42 can be rewritten as:

$$\begin{aligned}(t^{1-\nu})^2\ddot{X} + t^{1-\nu}\dot{X}\big[(1-\nu)t^{-\nu} - \dot{\alpha}_t t^{1-\nu} + e^{\alpha_t}\big] \\ + \mathcal{M}e^{\beta_t+2\alpha_t}\nabla f(X) = 0\end{aligned} \tag{43}$$

$$\begin{aligned}t^{2-2\nu}\ddot{X} + \big[(1-\nu)t^{1-2\nu} - t^{2-2\nu}\dot{\alpha}_t + e^{\alpha_t}t^{1-\nu}\big]\dot{X} \\ + \mathcal{M}e^{\beta_t+2\alpha_t}\nabla f(X) = 0\end{aligned} \tag{44}$$

We considered a subfamily of Bregman Lagrangians parameterized by a positive scalar $\rho > 0$, with coefficients defined as $\alpha_t = \log\rho - \nu\log t$ for $\beta_t = \beta_0 + \nu\rho\log t$ for $\beta_{t0}$ and $\gamma_t = \gamma_0 + \rho\log t$ where $\rho \geq 2$ is a constant characterizing the rate of convergence. These time functions satisfy the ideal scaling conditions of Equations 41 and 50.

The resulting Euler–Lagrange Equation 44 takes the form:

$$t^{2-2\nu}\ddot{X} + (p+1)t^{1-2\nu}\dot{X} + p^2 t^{\nu(p-2)}\mathcal{M}\nabla f(X) = 0 \tag{45}$$

One can notice that for $\rho = 2, \nu = 1$, Equation 45 reduces to the classical NAG. This method provides a rigorous, systematic way to tune the damping coefficient, achieving a principled trade-off between convergence speed and stability of the optimization trajectory.

### 5.2. Convergence rate

To study convergence, we adopted a Lyapunov function approach. By defining the energy function:

$$\begin{aligned}\mathcal{F}_t =& D_\phi\big(X^*, X + e^{-\alpha_t}\mathcal{T}_\nu(X)\big) \\ &+ e^{\beta_t}\big(f(X) - f(X^*)\big)\end{aligned} \tag{46}$$

its conformable time derivative is:

$$\begin{aligned}\mathcal{T}_\nu(\mathcal{F}) = \mathcal{T}_\nu(\beta_t)e^{\beta_t}(f - f^*) \\ + e^{\beta_t}\big(\mathcal{T}_\nu(X)\nabla f(X) - \mathcal{T}_\nu(f^*)\big) \\ + \big\langle \mathcal{T}_\nu\big(\nabla\phi(X + e^{-\alpha_t}\mathcal{T}_\nu(X))\big), \\ X - X^* + e^{-\alpha_t}\mathcal{T}_\nu(X)\big\rangle\end{aligned} \tag{47}$$

The inner-product term satisfies the Euler–Lagrange condition (Equation 42), therefore, Equation 47 becomes:

$$\begin{aligned}\mathcal{T}_\nu(\mathcal{F}) = -e^{\alpha_t+\beta_t}R_f(X^*, X) \\ + \big(\mathcal{T}_\nu(\beta_t) - e^{\alpha_t}\big)e^{\beta_t}(f - f^*)\end{aligned} \tag{48}$$

Since $f$ is convex, $R_f(X^*, X) \geq 0$ the first term non-positive. Recalling $\mathcal{T}_\nu(h) = t^{1-\nu}\frac{dh}{dt}$, we have:

$$\begin{aligned}\frac{d\mathcal{F}}{dt} = -t^{\nu-1}e^{\alpha_t+\beta_t}R_f(X^*, X) \\ + \big(t^{1-\nu}\dot{\beta}_t - e^{\alpha_t}\big)t^{\nu-1}e^{\beta_t}(f - f^*)\end{aligned} \tag{49}$$

For the negativity of the second term, we introduced the second scaling condition:

$$\hat{\beta}_t = t^{1-\nu}\dot{\beta}_t \leq e^{\alpha_t}, \tag{50}$$

thats ensures $\frac{d\mathcal{F}}{dt} \leq 0$. Hence, the convergence rate of the modified variational model with conformable time-weighted velocity is:

$$f - f^* \leq \mathcal{O}\big(e^{-\beta_t}\big) \tag{51}$$

## 6. Results and discussion

We evaluated the effectiveness of the proposed Lyapunov-based accelerated optimization method on several benchmark functions, including the Sphere, Trid, and 10-dimensional Rosenbrock functions. All simulations were performed using pre-designed convergence rates dictated by the Lyapunov framework. Figures 1 and 2 present the trajectories and convergence behavior of the Sphere and Trid functions, respectively, with a designed polynomial convergence rate of $\mathcal{O}(1/t^3)$. The results show that the proposed method significantly accelerated convergence compared with the classical NAG approach. In particular, the Lyapunov-based design ensured a systematic reduction in the objective function gap while providing predictable and stable convergence patterns. Figure 3 illustrates the performance of the proposed method under an exponential convergence design, $\mathcal{O}(e^{-0.5t})$, demonstrating even faster attraction toward the minimizer with reduced oscillatory behavior compared to classical NAG. Figures (4 and 5) extend the analysis to the 10-dimensional nonconvex Rosenbrock function and the Rastrigin benchmark function, respectively, using a Lyapunov-designed convergence rate of $\mathcal{O}(1/t^4)$. The proposed approach maintained accelerated convergence while handling the nonconvexity of the landscape. The results highlight the method's ability to control oscillations while preserving rapid convergence, illustrating the trade-off inherent in accelerated optimization: higher convergence rates can induce overshooting and oscillatory behavior due to reduced damping, whereas slower convergence provides smoother trajectories but at the cost of speed. To address this trade-off, we introduced a fractional-order conformable derivative in the optimization

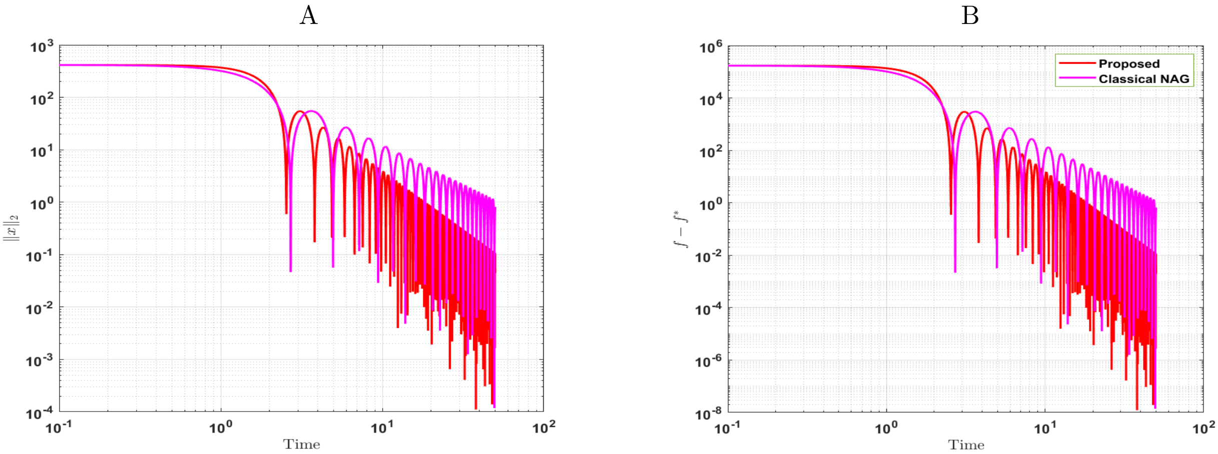


**Figure 1.** Comparison of the classical Nesterov's accelerated gradient (NAG) method and the proposed approach with a designed rate of $\mathcal{O}(1/t^3)$ on the Sum of Squares function. (A) Trajectory of the state norm x as function of time. (B) Convergence of the function value gap $f - f^*$

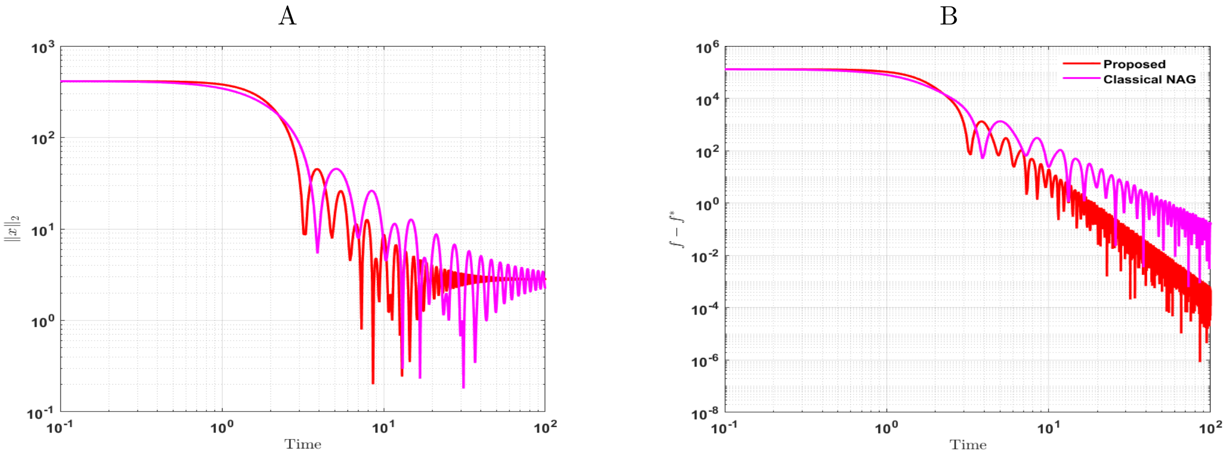


**Figure 2.** Comparison of the classical Nesterov's accelerated gradient (NAG) method and the proposed approach with a designed rate of $\mathcal{O}(1/t^3)$ on the Trid function. (A) Trajectory of the state norm x as function of time. (B) Convergence of the function value gap $f - f^*$

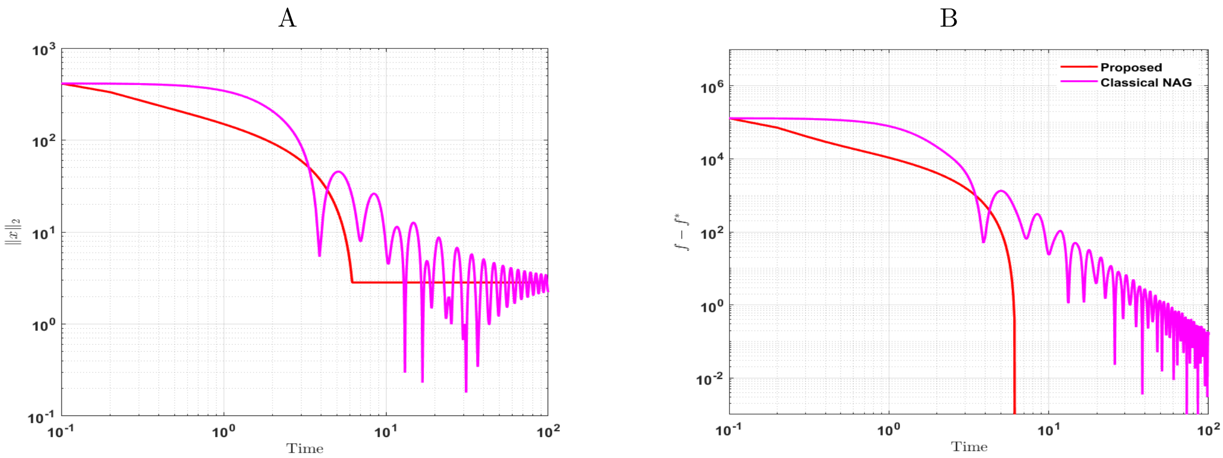


**Figure 3.** Comparison of the classical Nesterov's accelerated gradient (NAG) method and the proposed approach with a designed rate of $\mathcal{O}(e^{-0.5t})$ on the Trid function. (A) Trajectory of the state norm x as function of time. (B) Convergence of the function value gap $f - f^*$

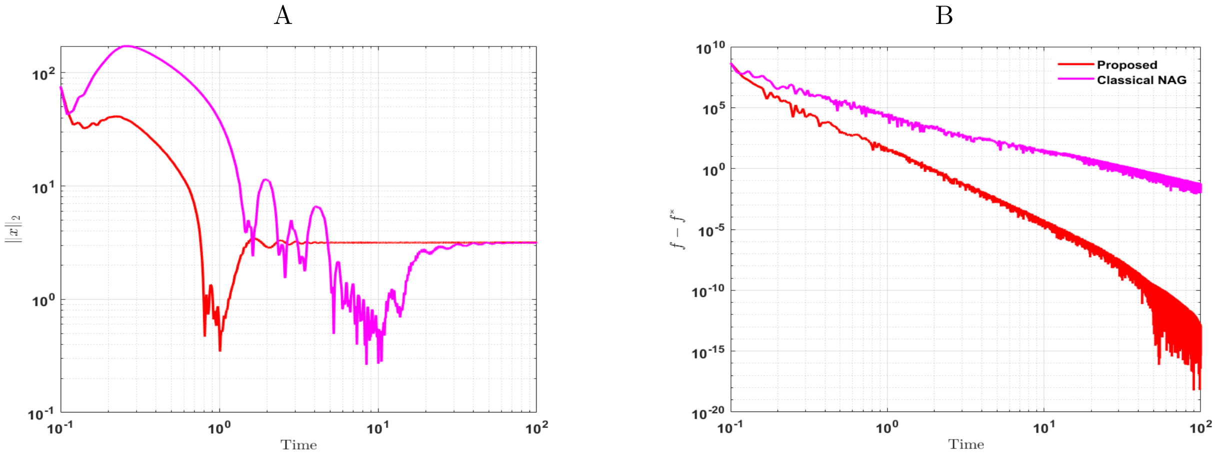


**Figure 4.** Comparison of the classical Nesterov's accelerated gradient (NAG) method and the proposed approach with a designed rate of $\mathcal{O}(1/t^4)$ on the 10-dimensional nonconvex Rosenbrock function. (A) Trajectory of the state norm x as function of time. (B) Convergence of the function value gap $f - f^*$

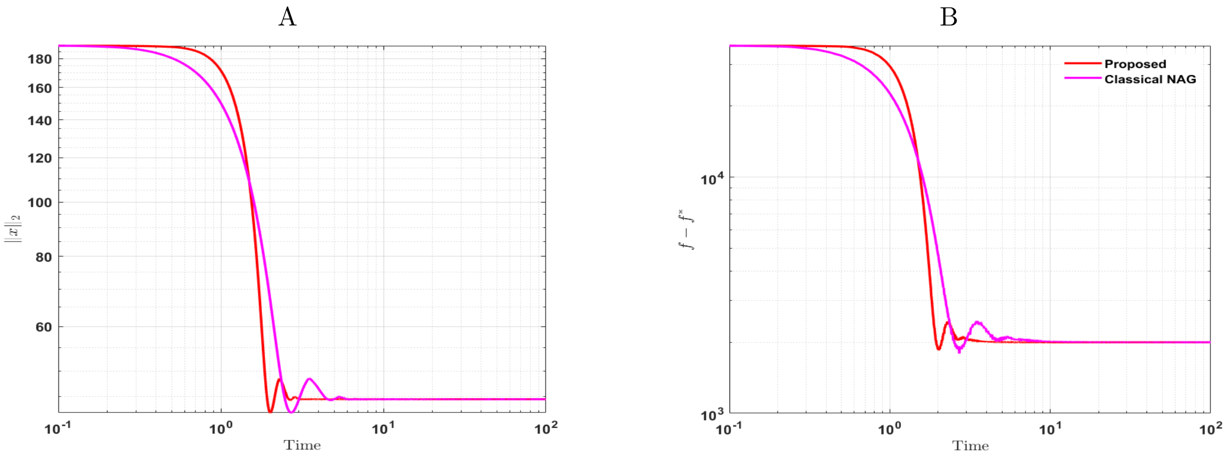


**Figure 5.** Comparison of the classical Nesterov's accelerated gradient (NAG) method and the proposed approach with a designed rate of $\mathcal{O}(1/t^4)$ on the 10-dimensional nonconvex Rastrigin function. (A) Trajectory of the state norm x as function of time. (B) Convergence of the function value gap $f - f^*$

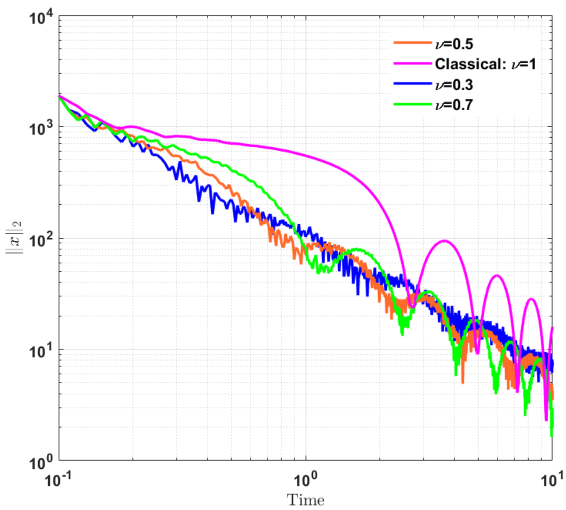


**Figure 6.** Effect of the fractional order $\nu$ on convergence behavior compared with classical Nesterov's method on the ill-conditioned 10-dimensional Ellipsoid function. Smaller $\nu$ values yield less oscillation and greater stability in the optimization trajectory.

dynamics, which modulates the effective damping in a systematic manner. Figure 6 shows simulations on the ill-conditioned 10-dimensional Ellipsoid function for various fractional orders $\nu \in (0, 1)$. As $\nu$ decreased, the trajectories exhibited reduced oscillations and enhanced stability while retaining fast convergence. This demonstrates that the fractional time-weighted term $t^{-\nu}$ acts as a tunable damping mechanism, allowing a principled balance between acceleration and stability.

Overall, the results confirm that the proposed Lyapunov-based design provides accelerated convergence across convex and nonconvex benchmarks while offering predictable behavior. Incorporating fractional derivatives further enhances stability, mitigating oscillations and overshooting in ill-conditioned problems, without sacrificing convergence speed. These findings highlight the flexibility and robustness of the proposed framework in controlling both convergence rate and stability in continuous-time accelerated optimization.

## 7. Conclusion

In this work, we proposed and analyzed two dynamical-system–based approaches for accelerating optimization beyond the classical NAG method. The first approach employed a Lyapunov-based design by introducing a time-dependent inertia term into the continuous-time ODE formulation. This allowed us to systematically construct new convergence rates that exceed those of the classical Nesterov scheme, as confirmed by our simulations on ill-conditioned convex and non-convex benchmark functions, including the 10-dimensional Ellipsoid and Rosenbrock functions.

The second approach leveraged a conformable derivative within a Bregman–Lagrangian framework, incorporating a time-weighted velocity term. This method provides a rigorous, systematic way to tune the damping coefficient, achieving a principled trade-off between convergence speed and stability of the optimization trajectory. Our numerical experiments show that, for fractional orders $\nu < 1$, the conformable dynamics significantly reduce oscillations while maintaining fast convergence, particularly on challenging landscapes such as the Rastrigin function, where the convergence rate approaches $1/t^4$.

Overall, these results highlight the power of variational and generalized-calculus frameworks in designing accelerated optimization methods. The Lyapunov-based approach offers a clear path to new, faster convergence rates, while the conformable-derivative approach enables fine control of the trade-off between speed and stability, suggesting a promising direction for the development of next-generation accelerated algorithms.

## Acknowledgments

None.

## Funding

None.

## Conflict of interest

YangQuan Chen is an Editorial Board Member of this journal, but was not in any way involved in the editorial and peer-review process conducted for this paper, directly or indirectly. Separately, other authors declared that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

## Author contributions

*Conceptualization:* Osama F. Abdel Aal, YangQuan Chen
*Formal analysis:* Osama F. Abdel Aal, YangQuan Chen
*Investigation:* Osama F. Abdel Aal, Necdet Sinan Ozbek, Jairo Viola
*Methodology:* Osama F. Abdel Aal, Necdet Sinan Ozbek, Jairo Viola
*Writing – original draft:* Osama F. Abdel Aal, Necdet Sinan Ozbek, Jairo Viola
*Writing–review & editing:* Osama F. Abdel Aal

## Availability of data

Not applicable.

## AI tools statement

All authors confirm that no AI tools were used in the preparation of this manuscript.

## References

1. W. Su, S. Boyd, and E. J. Cand‘es. A differential equation for modeling nesterov's accelerated gradient method: Theory and insights. *J Mach Learn Res.* 2016;17(153):1–43.
2. Park C, Cho Y, and Yang I. Generalized continuous-time models for nesterov's accelerated gradient methods. 2024.
3. Wibisono A, Wilson AC, and Jordan MI. A Variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences.* 2016;113(47):E7351–E7358.

4. Shi B, Du SS, Jordan MI and Su WJ. Understanding the Acceleration Phenomenon via High-resolution differential equations. *Mathematical Programming*. 2022;1–70.
5. Chen S, Shi B and Yuan Y-x, Revisiting the acceleration phenomenon via high resolution differential equations. 2022;arXiv preprint arXiv:2212.05700.
6. Muehlebach M and Jordan MI. Optimization with momentum: Dynamical, control-theoretic, and symplectic perspectives. *J Mach Learn Res.* 2021;22.
7. Chen S, Liu J, Wang P, Xu C, Cai S, and Chu J. Accelerated Optimization in Deep Learning with a Proportional-Integral-Derivative Controller. *Nature Communications.* 2024;15(1):10263.
8. Zhu TT, Hu R, and Fang YP. Fast convergence rates and trajectory convergence of a tikhonov regularized inertial primal–dual dynamical system with time scaling and vanishing damping. *Journal of Computational and Applied Mathematics*. 460:116394, 2025.
9. Weinan E. A proposal on machine learning via dynamical systems Communications in Mathematics and Statistics. 2017;5:1–11.
10. Le DM, Patil OS, Nino CF, and Dixon WE. Accelerated gradient approach for deep neural network-based adaptive control of unknown nonlinear systems. *IEEE Transactions on Neural Networks and Learning Systems.* 2024;36(4):62990-6313.
11. Da Silva AB and Gazeau M. A General System of Differential Equations To Model Firstorder Adaptive Algorithms. *J Mach Learn Res.* 2020;21.
12. Mao Z, Suzuki S, Nabae H, Miyagawa S, Suzumori K, and Maeda S. Machine Learning enhanced Soft Robotic System Inspired by Rectal Functions to Investigate Fecal Incontinence. Bio-Design and Manufacturing. 2025;8(3):482-494.
13. Muehlebach M and Jordan M. A Dynamical Systems Perspective on Nesterov Acceleration, in International Conference on Machine Learning. PMLR. 2019;4656–4662.
14. Shi B, Du SS, Su W, and Jordan MI. Acceleration via Symplectic Discretization of High resolution Differential Equations. *Advances in Neural Information Processing Systems.* 2019;32.
15. Moucer C, Taylor A and Bach F. A Systematic Approach To Lyapunov Analyses of Continuous-time Models in Convex Optimization. *SIAM Journal on Optimization.* 2023;33(3):1558–1586.
16. Sanz Serna JM, and Zygalakis KC. The Connections Between Lyapunov Functions For Some Optimization Algorithms and Differential Equations. *SIAM Journal on Numerical Analysis.* 2021;59(3):1542–1565.
17. Kolarijani AS, Esfahani PM, and Keviczky T. Continuous-time Accelerated Methods via A Hybrid Control Lens. *IEEE Transactions on Automatic Control.* 2020; 65(8):3425-3440.
18. Lin T, and Jordan MI. A control-theoretic perspective on optimal high-order optimization," *Mathematical Programming.* 2022;195:929–975.
19. Attouch H, Bot RI, and Csetnek ER. Fast Optimization via Inertial Dynamics With Closedloop Damping. *Journal of the European Mathematical Society.* 2022;25(5):1985–2056.
20. Adly S. and Attouch H. Accelerated Optimization Through Time-scale Analysis of Inertial Dynamics With Asymptotic Vanishing and Hessian-driven Dampings. *Optimization.* 2024;1–38.
21. Castera C, Attouch H, Fadili J and Ochs P. Continuous Newton-like Methods Featuring Inertia and Variable Mass. *SIAM Journal on Optimization.* 2024;34(1):251–277.
22. Attouch H, Chbani Z and Riahi H. Fast Proximal Methods via Time Scaling of Damped Inertial Dynamics. *SIAM Journal on Optimization.* 2019;29(3):2227–2256.
23. Cheng X, Liu J, and Shang Z. A class of generalized nesterov's accelerated gradient method from dynamical perspective. *arXiv preprint ar.* 2025;Xiv:2508.12816.
24. Laborde M and Oberman A. A Lyapunov Analysis For Accelerated Gradient Methods: From Deterministic To Stochastic Case," in Proceedings of The Twenty Third International Conference on Artificial Intelligence and Statistics (S. Chiappa and R. Calandra, eds.), vol. 108 of Proceedings of Machine Learning
25. Ross I. An Optimal Control Theory For Nonlinear Optimization. *Journal of Computational and Applied Mathematics.* 2019;354:39–51.
26. Wilson AC, Recht B, and Jordan MI. A Lyapunov Analysis of Accelerated Methods in Optimization. *Journal of Machine Learning Research.* 2021;22(113):1-34.
27. Verma VR, Nishad DK, Sharma V, Singh VK, Verma A, and Shah DR. Quantum Machine Learning for Lyapunov-stabilized Computation Offloading in Next-Generation MEC Networks. *Scientific Reports.* 2025;15:405.
28. Dobson P, Sanz-Serna JM and Zygalakis KC. On the connections between optimization algorithms, lyapunov functions, and differential equations: Theory and insights. *SIAM Journal on Optimization.* 2025;35(1):537–566.
29. Kravitz H, Dur´on C, Nieves B, and Brio M. Data-driven optimization and parameter estimation for a metric graph epidemic model with applications to covid-19 spread in poland: A real-world example of optimization for a challenging rosenbrock-type objective function. *An International Journal of Optimization and Control: Theories & Applications (IJOCTA).* 2025;16:750–778.
30. Hanachi SB, Sellami B and Belloufi M. Global convergence property with inexact line search for a new conjugate gradient method. *An International Journal of Optimization and Control: Theories & Applications (IJOCTA).* 2025;15:25-34.

31. Tran BK and Leok M. Variational principles for Hamiltonian systems. *arXiv preprint*. 2024;arXiv:2410.02960.
32. Cresson J. Fractional Embedding of Differential Operators and Lagrangian Systems," *Journal of Mathematical Physics*. 2007;48(3).
33. Aal OFA, Viola J, and Chen Y, Fractional Order Euler-Lagrange Model for Accelerated Gradient Methods. *IFAC-PapersOnLine*. 2024;58(12):466-471.
34. Pham T and Wagner H. Fast Kd-trees for the Kullback-Leibler Divergence and other Decomposable Bregman Divergences. 2025;arXiv preprint arXiv:2502.13425.
35. Banerjee A, Merugu S, Dhillon IS and Ghosh J. Clustering with Bregman Divergences. *Journal of Machine Learning Research*. 2005;6:1705–1749.
36. Cha J, Kim Y, Shin J, Cho J, Kim S, and Ryu J. Optimization of Bregman Variational Learning Dynamics. *arXiv preprint ar*. 2025;Xiv:2510.20227.
37. Yu D, Jiang W, Wan Y, and Zhang L. Mirror Descent under Generalized Smoothness. *arXiv preprint*. 2025;arXiv:2502.00753.
38. Hayashi M. Bregman-Divergence-based Arimoto-Blahut Algorithm. *IEEE Transactions on Information Theory*. 2025.
39. Gao X, Jiang Y, Cai X, and Wang K. An Accelerated Mirror Descent Algorithm for Constrained Nonconvex Problems. *Optimization*. 2025;1–26.
40. Boyd SP and Vandenberghe L. Convex optimization. *Cambridge University Press*. 2004.
41. Khalil R, Al Horani M, Yousef A and Sababheh M. A New Definition of Fractional Derivative. *Journal of Computational and Applied Mathematics*. 2014;26465–70.
42. Chung WS. Fractional Newton Mechanics with Conformable Fractional Derivative. *Computers & Mathematics with Applications*. 2015;69(4):322–330.
43. Nesterov YE. A Method For Solving The Convex Programming Problem With Convergence Rate. *in Dokl. Akad. Nauk Sssr*. 1983;269:543–547.
44. Nesterov Y. Introductory Lectures on Convex Optimization: A Basic Course, vol. 87. *Springer Science & Business Media*. 2003.
45. Nesterov Y. Nonsmooth Convex Optimization. *Lectures on Convex Optimization*. 2018;139–240.
46. Gaud N, Murthy PK, Hassan MMM, Ganguly A, Mali V, Randive MLB, and Singh A. Enhanced nirmal optimizer with damped nesterov acceleration: a comparative analysis. *arXiv preprint* 2025;arXiv:2508.16550.
47. Haraux A. Systemes Dynamiques Dissipatifs Et Applications. 1991.
48. Almeida R. Fractional Calculus of Variations for composed Functionals with Generalized Derivatives.

**Osama F. Abdel Aal** received the B.S. degree in Mechatronics Engineering in 2011 and the M.S. degree in Mechanical Engineering in 2017, both from the University of Jordan, Amman, Jordan. He is currently pursuing the Ph.D. degree in Mechanical Engineering at the University of California, Merced, USA. His research interests include nonlinear dynamics, mechatronics, optimization and control systems.
`https://orcid.org/0000-0002-0325-3201`

**Necdet Sinan Ozbek** received his Ph.D. in Electrical and Electronics Engineering from Çukurova University in 2018. He is currently an Associate Professor at Adana Alparslan Türkeş Science and Technology University in Türkiye. His research interests include control theory, unmanned aerial vehicles, fractional-order control, data-driven systems, and nonlinear dynamics. He has authored numerous scientific publications in highly prestigious international journals and contributed extensively to advanced control methodologies Dr. Özbek has held international research positions at the MESA Lab, University of California, Merced, and at the Laboratoire des Signaux et Systèmes (L2S), CentraleSupélec, Paris-Saclay. He serves as a reviewer for several top-tier journals in control, automation, and cyber-physical systems. His recent work focuses on finite-time and prescribed-time convergence in optimization, robust control of UAVs, and energy-aware cyber-physical system design.
`https://orcid.org/0000-0002-7184-9015`

**Jairo Viola** (Member, IEEE) received the bachelor's and master's degrees in electronic engineering from Pontifical Bolivarian University, Colombia, in 2013 and 2017, respectively, and the Ph.D. degree in mechanical engineering from the University of California at Merced in 2022. He currently serves as external research fellow of the Mechatronics, Embedded Systems and Automation Lab of the university of California, Merced. His research interests include advanced control systems, artificial intelligence, digital twin, self-optimizing control, machine learning, big data, edge computing, and applied fractional-order calculus.
`https://orcid.org/0000-0003-2739-5383`

**YangQuan Chen** earned his Ph.D. from Nanyang Technological University, Singapore, in 1998. He had been a faculty of Electrical Engineering at Utah State University (USU) from 2000-12. He joined the School of Engineering, University of California, Merced (UCM) in summer 2012 teaching "Mechatronics", "Engineering Service Learning", "Unmanned Aerial Systems" and "Dgital Twins" for undergraduates and "Fractional Order Mechanics", "Linear Multivariable Control", "Nonlinear Controls", "Advanced

Controls: Robustness and Optimality" for graduates. His research interests include mechatronics for sustainability, cognitive process control, small multi-UAV based cooperative multi-spectral "personal remote sensing", applied fractional calculus in controls, modeling and complex signal processing; distributed measurement and control of distributed parameter systems with mobile actuator and sensor networks. He is listed in Highly Cited Researchers by Clarivate Analytics from 2018 to 2021. He received Research of the Year awards from USU (12) and UCM (20).
https://orcid.org/0000-0002-7422-5988

An International Journal of Optimization and Control: Theories & Applications
(https://accscience.com/journal/ijocta)