

# On the Instability of Nesterov's ODE Under Non-Conservative Vector Fields

Daniel E. Ochoa<sup>✉</sup>, Member, IEEE, Mahmoud Abdelgalil<sup>✉</sup>, and Jorge I. Poveda<sup>✉</sup>

**Abstract**—We study the *instability* of Nesterov's ODE in non-conservative settings, where the driving term is not necessarily the gradient of a potential function. While convergence properties under Nesterov's ODE are well-characterized for settings with gradient-based driving terms, we show that the presence of *arbitrarily small* non-conservative terms can lead to instability. To resolve the instability issue, we study a regularization mechanism based on restarting. For this mechanism, we establish novel explicit bounds on the resetting period that ensure the decrease of a suitable Lyapunov function, thereby guaranteeing stability and “accelerated” convergence rates under suitable smoothness and monotonicity properties on the driving term. Numerical simulations support our results.

**Index Terms**—Averaging analysis, non-conservative systems, hybrid dynamical systems.

## I. INTRODUCTION

THE NESTEROV Accelerated Gradient method has been a cornerstone of optimization since its introduction in [1]. Its continuous-time analog, introduced in [2] and termed the *Nesterov's Ordinary Differential Equation* (ODE), has emerged as a powerful tool to study optimization algorithms using a continuous-time dynamical systems point of view [3]. Nesterov's ODE is defined by the equation

$$\ddot{x} + \frac{3}{\tau} \dot{x} + \mathcal{G}(x) = 0, \quad \dot{\tau} = \eta, \quad (1)$$

where  $x \in \mathbb{R}^n$ ,  $\tau \in [T_0, \infty)$ , with  $T_0 > 0$ ,  $\mathcal{G} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , and  $\eta \in (0, 1]$ . When  $\mathcal{G}(x) = \nabla J(x)$  for some convex potential function  $J$ , Nesterov's ODE achieves accelerated convergence to the minimum of  $J$  with rate  $O(1/t^2)$  [2].

The remarkable success of Nesterov's method in optimization motivates the exploration of its potential extensions to *non-conservative settings*, where the vector field  $\mathcal{G}$  is not necessarily the gradient of a potential function.

Received 25 July 2025; revised 7 October 2025; accepted 31 October 2025. Date of publication 17 November 2025; date of current version 3 December 2025. This work was supported in part by NSF ECCS CAREER under Grant 2305756; in part by CMMI under Grant 2228791; and in part by AFOSR under Grant FA9550-22-1-0211. Recommended by Senior Editor G. Giordano. (Corresponding author: Daniel E. Ochoa.)

Daniel E. Ochoa is with the Department of Electrical and Computer Engineering, University of California at Santa Cruz, Santa Cruz, CA 95064 USA (e-mail: dochoatamayo@ucsc.edu).

Mahmoud Abdelgalil and Jorge I. Poveda are with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093 USA.

Digital Object Identifier 10.1109/LCSYS.2025.3633320

Such scenarios naturally arise in areas such as game theory and consensus-based distributed optimization [4]. In game-theoretic contexts, it is well known that for a class of games known as *potential games*, convergence to equilibria can be achieved via gradient-like dynamics. Similarly, in multi-agent systems with undirected communication graphs, consensus dynamics are often interpreted as gradient flows. These observations raise an important question: can dynamical systems of the form (1) also provide benefits in settings where the game is not potential or the communication graph is directed?

Extending Nesterov's acceleration to such domains, however, presents several challenges. First, the lack of a potential function prevents one from formulating the ODE within an optimization framework and leveraging the body of existing stability results developed in that context. Second, while in optimization problems Nesterov's ODE achieves accelerated convergence through dynamic damping of the form  $\frac{3}{\tau} \dot{x}$ , this same term can have adverse effects in non-conservative systems. For instance, our previous studies on accelerated Nash equilibrium seeking [5] and distributed concurrent learning [6] revealed that this damping mechanism fails to mitigate the destabilizing influence of the non-conservative component of  $\mathcal{G}$  in (1), often leading to instability. Although similar behaviors have been observed numerically even in optimization problems [7], a theoretical explanation for this *instability phenomenon* has remained elusive.

The first contribution of this letter is to provide a theoretical explanation for the emergence of instability in Nesterov's ODE when the term  $\mathcal{G}$  is non-conservative. Specifically, by leveraging the Helmholtz decomposition theorem [8] we decompose  $\mathcal{G}$  into its conservative and non-conservative components, and employ the variation-of-constants formula [9, Proposition 9.6] to obtain a representation suitable for averaging analysis. We then study the average system using Floquet theory [10, Sec. 19] to identify subclasses of vector fields  $\mathcal{G}$  for which Nesterov's ODE *fails* to stabilize the set  $\mathcal{A} := \{x : \mathcal{G}(x) = 0\}$ . We further show that, in specific scenarios, this instability persists even when the non-conservative component of  $\mathcal{G}$  is arbitrarily small compared to the conservative component, thus providing new insights into adversarial perturbations that can destabilize such accelerated flows in traditional optimization settings. Our instability analysis focuses on linear non-conservative vector fields whose conservative components are strongly monotone.

The second main contribution is the development and analysis of a hybrid dynamical system that stabilizes  $\mathcal{A}$  by combining two mechanisms: continuous-time flows governed by Nesterov-like dynamics, and discrete-time resets triggered whenever  $\tau \in \mathbb{R}_{\geq 0}$  reaches the upper bound of a compact interval. Unlike the instability analysis, which focuses on linear vector fields, this hybrid framework encompasses a broader class of nonlinear vector fields satisfying suitable monotonicity and Lipschitz conditions. Extending the results in [5], we derive quasi-optimal reset conditions and establish improved convergence rate guarantees.

The rest of this letter is organized as follows. Section II presents the preliminaries. Section III prepares Nesterov's ODE for averaging, and Section IV presents its instability properties. Section V presents a hybrid mechanism to recover stability. Section VI presents a numerical example. Section VII ends with conclusions.

## II. PRELIMINARIES

We use  $(u, v) = [u^\top, v^\top]^\top$  to denote the concatenation of  $u, v \in \mathbb{R}^n$ . The Euclidean inner product is denoted by  $\langle u, v \rangle$  for all  $u, v \in \mathbb{R}^n$ . We let  $|z| := \sqrt{\langle z, z \rangle}$  for each  $z \in \mathbb{R}^n$ , and use  $|z|_C := \min_{s \in C} |z - s|$  to refer to the minimum distance between  $z$  and a closed set  $C$ . Given a matrix  $A \in \mathbb{R}^{m \times n}$ , the induced 2-norm is defined by  $\|A\| := \sup_{|x|=1} |Ax|$ . The set of eigenvalues of a matrix  $A \in \mathbb{R}^{n \times n}$ , counted with multiplicities, is denoted by  $\text{spec}(A)$ . We use  $A_{ij} \in \mathbb{R}$  for the  $i$ -th entry of the  $j$ -th column of  $A \in \mathbb{R}^{n \times n}$ . Given  $\{a_k\}_{k=1}^n \subset \mathbb{R}^n$ ,  $\text{diag}\{a_k\}_{k=1}^n \in \mathbb{R}^{n \times n}$  denotes the matrix whose  $i$ -th diagonal entry equals  $a_i$  for all  $i \in \{1, 2, \dots, n\}$ , and whose off-diagonal entries are zero. A set-valued map  $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  assigns to each point  $x \in \mathbb{R}^n$  a set  $F(x) \subset \mathbb{R}^n$ . Given  $f : \mathcal{X} \rightarrow \mathcal{Y}$  and  $g : \mathcal{Y} \rightarrow \mathcal{Z}$ , we use  $g \circ f$  to denote their composition. The *flow*  $\Phi_s^f$  along a vector field  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  assigns to each point  $(s, x_0)$  the value  $\Phi_s^f(x_0) = \chi(s) \in \mathbb{R}^n$ , where  $\chi$  is the unique solution to  $\dot{x} = f(x)$  satisfying  $\chi(0) = x_0$ . Given a diffeomorphism  $\Psi : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , the *pullback* of  $f$  by  $\Psi$ , denoted  $\Psi^*f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , is defined by  $(\Psi^*f)(x) = (\frac{\partial \Psi^{-1}}{\partial x} \circ f \circ \Psi)(x)$  for all  $x \in \mathbb{R}^n$ , where  $\frac{\partial \Psi^{-1}}{\partial x}(y)$  denotes the Jacobian matrix of  $\Psi^{-1}$  evaluated at  $y \in \mathbb{R}^n$ . Given a smooth function  $J : \mathbb{R}^n \rightarrow \mathbb{R}$  we define its gradient by  $\nabla J(x) = (\frac{\partial J}{\partial x_1}(x), \dots, \frac{\partial J}{\partial x_n}(x))$ . Given a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , we define its divergence  $\text{div}f(x) := \sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x)$ , and let the rotation operator  $\overline{\text{rot}}f : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$  be given by  $(\overline{\text{rot}}f(x))_{ij} := \frac{\partial f_i(x)}{\partial x_j} - \frac{\partial f_j(x)}{\partial x_i}$  for each  $x \in \mathbb{R}^n$ . Given a matrix-valued function  $K : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ , we define the rotation operator as  $\text{rot}K(x) = \sum_{j=1}^n (\frac{\partial K_{ij}}{\partial x_1}(x), \dots, \frac{\partial K_{nj}}{\partial x_n}(x))$ , where  $K_{ij} : \mathbb{R}^n \rightarrow \mathbb{R}$  corresponds to the function  $x \mapsto (K(x))_{ij}$ .

To obtain our results, we use the following theorem relating the flows of two vector fields  $g, f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ :

**Theorem 1 (Variation-of-Constants Formula [9, Proposition 9.6]):** Let  $g$  be  $r$ -th and  $f$  be  $(r+1)$ -th continuously differentiable time-dependent vector fields. Let  $x_0 \in \mathbb{R}^n$ ,  $T > 0$ ,  $\tau \in [0, T]$ , and define  $\tilde{g}(\tau, x) = ((\Phi_\tau^f)^*g_\tau)(x)$ , where  $g_\tau(x) := g(x, \tau)$ , and  $(\Phi_\tau^f)^*g_\tau$  denotes the pullback of  $g_\tau$  along the flow of  $f$ . Then, for all  $\tau \in [0, T]$  where the flows  $\Phi_\tau^{f+g}$ ,  $\Phi_\tau^f$ , and  $\Phi_\tau^g$  exist, it follows that  $\Phi_\tau^{f+g}(x_0) = (\Phi_\tau^f \circ \Phi_\tau^g)(x_0)$ .

**Hybrid Dynamical Systems:** To formulate mechanisms that recover the stability of Nesterov's ODE, we use hybrid dynamical systems (HDS) with state  $x \in \mathbb{R}^n$  and dynamics

$$\mathcal{H} : \begin{cases} x \in C & \dot{x} \in F(x) \\ x \in D & x^+ \in G(x) \end{cases}, \quad (2)$$

where  $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  is the flow map,  $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  is the jump map,  $C \subset \mathbb{R}^n$  is the flow set, and  $D \subset \mathbb{R}^n$  is the jump set. We use the tuple  $\mathcal{H} = (C, F, D, G)$  to refer to the data of the HDS. Solutions to HDS are parameterized by continuous-time  $t \in \mathbb{R}_{\geq 0}$  and discrete-time  $j \in \mathbb{Z}_{\geq 0}$  and thus evolve on hybrid time domains. For a formal definition of solutions to HDS we refer the reader to [11, Sec. 2]. Standard continuous-time systems described by the dynamics  $\dot{x} = f(x)$  with  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , can be cast as HDS by setting  $C = \mathbb{R}^n$ ,  $D = \emptyset$ , and  $F(x) = \{f(x)\}$ .

**Definition 1 (Stability/Instability Notions):** A compact set  $\mathcal{A} \subset \mathbb{R}^n$  is said to be uniformly globally exponentially stable (UGES) for the system  $\mathcal{H}$  in (2) if every maximal solution to  $\mathcal{H}$  is complete, and there exist  $c_1, c_2 > 0$  such that  $|x(t, j)|\mathcal{A} \leq c_1|x(0, 0)|\mathcal{A}e^{-c_2(t+j)}$  for all  $(t, j) \in \text{dom } x$  and each solution  $x$  to  $\mathcal{H}$ . The set  $\mathcal{A}$  is unstable for  $\mathcal{H}$  if there exists  $\varepsilon > 0$  such that for each  $\delta > 0$  there exist a solution  $x$ , with  $|x(0, 0)|\mathcal{A} < \delta$ , and  $(T, J) \in \text{dom } x$  satisfying  $|x(T, J)|\mathcal{A} > \varepsilon$ .

## III. NESTEROV'S ODE: ANALYSIS VIA STANDARD AVERAGING

In this section, we prepare Nesterov's ODE for analysis via averaging. First, we transform (1) to the  $\tau$ -timescale by noting that every solution  $(x, \tau)$  to (1) satisfies  $\tau(t) = \eta t + T_0$ , with  $T_0 > 0$ , for all  $t \in \text{dom}(x, \tau)$ . This timescale change generates the following system with state  $(x, \tau) \in \mathbb{R}^n \times [T_0, \infty)$  and dynamics

$$\frac{d^2x}{d\tau^2} + \frac{3}{\tau} \frac{dx}{d\tau} + \gamma \mathcal{G}(x) = 0, \quad \frac{d\tau}{d\tau} = 1, \quad (3)$$

where  $\gamma := \eta^2$ . To analyze the effect of the vector field  $\mathcal{G}$  on the dynamics (3), we use the following generalization of the Helmholtz decomposition theorem [8, Th. 7.2]:

**Lemma 1 (Helmholtz Decomposition):** Suppose  $\mathcal{G}$  is analytic in  $\mathbb{R}^n$ . Then, there exist functions  $P : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ ,  $J : \mathbb{R}^n \rightarrow \mathbb{R}$ , and  $K : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$  such that  $\mathcal{G}(x) = \nabla J(x) + \text{rot}K(x)$  for all  $x \in \mathbb{R}^n$ , where  $J(x) := \sum_{i=1}^n P_{ii}(x)$  and  $K(x) := P(x) - P(x)^\top$  satisfy  $\text{div}(\text{rot}K(x)) = 0$  and  $\text{rot}(\nabla J) = 0$  for all  $x \in \mathbb{R}^n$ .

To ensure uniqueness of solutions to (3) and guarantee that the flow along the corresponding vector field is well-defined, we impose the following conditions on the elements of the decomposition in Lemma 1.

**Assumption 1:**

- (i) **Strong Monotonicity:**  $\exists \kappa_J > 0$  such that  $\langle \nabla J(x_1) - \nabla J(x_2), x_1 - x_2 \rangle \geq \kappa_J |x_1 - x_2|^2$  for all  $x_1, x_2 \in \mathbb{R}^n$ . Additionally, we have that  $\langle \text{rot}K(x_1) - \text{rot}K(x_2), x_1 - x_2 \rangle \geq 0$  for all  $x_1, x_2 \in \mathbb{R}^n$ .
- (ii) **Lipschitz continuity:**  $\exists \ell_J, \ell_R > 0$  such that  $|\nabla J(x_1) - \nabla J(x_2)| \leq \ell_J |x_1 - x_2|$  and  $|\text{rot}K(x_1) - \text{rot}K(x_2)| \leq \ell_K |x_1 - x_2| \forall x_1, x_2 \in \mathbb{R}^n$ .
- (iii) **Scaling relationship:**  $\ell_K = \alpha \sqrt{\ell_J}$  with  $\alpha \in (0, 1]$ .

Assumption 1-(i) ensures that the scalar function  $J$  is strongly convex, which, via [2, Th. 3], guarantees accelerated

convergence of Nesterov's ODE to the unique minimizer of  $J$  when the non-conservative term  $\text{rot}K$  is not present. The next example presents a class of vector fields that satisfies these conditions.

*Example 1:* Given  $Q \in \mathbb{R}^{n \times n}$ , the vector field  $\mathcal{G}(x) = Qx$  is analytic and satisfies Assumption 1. The Helmholtz decomposition in Lemma 1 can be explicitly constructed by letting  $J(x) = x^\top Q_s x$ ,  $(K(x))_{ij} = \frac{1}{2}(Q_a)_{ij}|x|^2$  for all  $x \in \mathbb{R}^n$  and  $i, j \in \{1, 2, \dots, n\}$ , where  $Q_s := \frac{1}{2}(Q + Q^\top)$  and  $Q_a := \frac{1}{2}(Q - Q^\top)$ .

Using Lemma 1 and Assumption 1, by letting  $s := \tau/\varepsilon$ , with  $\varepsilon := (\ell_J)^{-\frac{1}{2}}$ , and letting  $y := (y_1, y_2) = (x, \frac{dx}{ds}) \in \mathbb{R}^{2n}$ , from (3) we obtain that  $d\tau/ds = \varepsilon$  and

$$\begin{aligned} \frac{dy}{ds} &= \begin{pmatrix} y_2 \\ -\nabla \hat{J}(y_1) \end{pmatrix} + \varepsilon \begin{pmatrix} 0 \\ -\frac{3}{\varepsilon s + T_0} y_2 - \text{rot} \hat{K}(y_1) \end{pmatrix} \\ &=: h(y) + \varepsilon u(y, \varepsilon s), \end{aligned} \quad (4)$$

for all  $y \in \mathbb{R}^n$ , where  $\hat{J}(x) := \frac{\gamma}{\ell_J} J(x)$  and  $\hat{K}(x) := \frac{\alpha\gamma}{\ell_K} K(x)$ . By using the pullback of  $u_{\varepsilon s}(y) := u(y, \varepsilon s)$  via the flow along  $h$ , we obtain a system with state  $(z, \tau) \in \mathbb{R}^{2n} \times [T_0, \infty)$  and dynamics given by

$$\frac{dz}{ds} = \varepsilon \left( \left( \Phi_\tau^h \right)^* u_{\varepsilon s} \right) (z) =: \varepsilon v(z, \varepsilon s, s), \quad \frac{d\tau}{ds} = \varepsilon. \quad (5)$$

By Theorem 1, a solution  $y$  to system (4) and a solution  $z$  to system (5) with  $y(0) = z(0)$ , satisfy  $y(s) = \Phi_s^h(z(s))$  for all  $s \in \text{dom } y$ .

*Remark 1:* The system in (5) admits a series representation that expands the pullback in integrals of iterated Lie brackets between  $u_{\varepsilon s}$  and  $h$ . While closed forms are generally not available, simplifications emerge when certain Lie brackets vanish; an approach used for Euler-Lagrange systems in [12]. Although Nesterov's ODE has an Euler-Lagrange formulation that might benefit from similar simplifications, our instability analysis studies the linear case of Example 1, leaving the nonlinear case for future work.

#### IV. INSTABILITY UNDER LINEAR MAPPINGS

In this section, we study the setting where  $\mathcal{G}(x) = Qx$  for some  $Q \in \mathbb{R}^{n \times n}$  and all  $x \in \mathbb{R}^n$ . In this case, the Helmholtz decomposition yields  $J(x) = \frac{1}{2}x^\top Q_s x$  and  $\text{rot } K(x) = Q_a x$ , where  $Q_s$  and  $Q_a$  are the symmetric and skew-symmetric parts of  $Q$ , respectively. Therefore, the ODE in (4) becomes

$$\frac{dy}{ds} = \underbrace{\begin{pmatrix} 0 & I \\ -\hat{Q}_s & 0 \end{pmatrix}}_{=: A} y + \varepsilon \underbrace{\begin{pmatrix} 0 & 0 \\ -\hat{Q}_a - \frac{3}{\varepsilon s + T_0} I & 0 \end{pmatrix}}_{=: B(\varepsilon s)} y, \quad (6)$$

with  $d\tau/ds = \varepsilon$ , and where  $\hat{Q}_s = \frac{Q_s}{\|Q_s\|}$ ,  $\hat{Q}_a = \alpha \frac{Q_a}{\|Q_a\|}$ , with  $\alpha \in (0, 1]$ . Similarly, the dynamical system in (5) reduces to

$$\frac{dz}{ds} = \varepsilon \Lambda(\tau, s) z := \varepsilon \left( e^{-As} B(\tau) e^{As} \right) z, \quad \frac{d\tau}{ds} = \varepsilon. \quad (7)$$

Before presenting the first main result of this letter, we introduce a lemma that, under suitable conditions on  $Q_s$ , shows that the flow of the vector field  $y \mapsto Ay$  is periodic.

*Lemma 2:* Let  $\mathcal{G}(x) = Qx$  where  $Q \in \mathbb{R}^{n \times n}$  and suppose  $\mathcal{G}$  satisfies Assumption 1. Suppose there exists  $\mu \in \mathbb{R}$ ,  $\mu \neq 0$ ,

such that for each  $\lambda \in \text{spec}(Q_s)$  there exists  $c_\lambda \in \mathbb{Q}_{>0}$ <sup>1</sup> satisfying  $\lambda = (c_\lambda \mu)^2$ . Then, there exists  $\mathcal{T} > 0$  such that every solution  $\psi$  to

$$\frac{d\psi}{ds} = A\psi, \quad (8)$$

where  $A$  is defined in (6), is periodic with period  $\mathcal{T}$ .

*Proof:* By the definition of  $A$  and the determinant formula for block matrices,  $\det(\lambda I - A) = \det(\lambda^2 I + \hat{Q}_s)$ . Thus, the eigenvalues of  $A$  are  $\pm i\omega_j$ , where  $\omega_j^2 = \lambda_j$  for  $\lambda_j \in \text{spec}(\hat{Q}_s)$ . Since  $J(x) = \frac{1}{2}x^\top Q_s x$  is strongly convex,  $\hat{Q}_s$  is positive definite, so all  $\omega_j \in \mathbb{R}_{>0}$ . Since  $\hat{Q}_s$  is positive definite, there exists an orthogonal matrix  $U$  that diagonalizes  $\hat{Q}_s$ . In the coordinates defined by  $U$ , the system (8) decouples into planar rotation blocks of the form  $\dot{\psi}_j = \begin{pmatrix} 0 & -\omega_j \\ \omega_j & 0 \end{pmatrix} \psi_j$ . Thus, each solution  $\psi$  to (8) is a superposition of periodic signals with frequencies  $\{\omega_j\}_{j=1}^n$ . By the assumption on  $Q_s$ , we have that  $\lambda_j = (c_j \mu)^2$  for  $c_j \in \mathbb{Q}$ ,  $c_j > 0$ . Then, writing  $c_j = a_j/b_j$  with  $a_j, b_j \in \mathbb{N}_{>0}$ , for each  $j \in \{1, 2, \dots, n\}$  there exists  $k_j \in \mathbb{Z}$  such that  $\omega_j = k_j \omega_0$  where  $\omega_0 := \mu/L$ , with  $L$  being the least common multiple of  $\{b_1, b_2, \dots, b_n\}$ . This result ensures periodicity of each solution  $\psi$ , with period  $\mathcal{T} = 2\pi/\omega_0$ . ■

The result of Lemma 2 enables the use of techniques for averaging of periodic systems with slow-time dependence (see [13, Sec. 3.3]), and leads to our first main result.

*Theorem 2:* Let  $\mathcal{G}(x) = Qx$ ,  $Q \in \mathbb{R}^{n \times n}$ , and assume  $\mathcal{G}$  satisfies Assumption 1. Suppose that: i) all off-diagonal entries of  $Q_a$  are nonzero, ii)  $Q_s$  satisfies the assumptions in Lemma 2, and iii) precisely one eigenvalue of  $Q_s$  has algebraic multiplicity greater than one. Then, for each  $\eta \in (0, 1]$ , the origin is unstable for the  $x$ -dynamics in Nesterov's ODE (1).

*Proof:* *Step 1 (Definition of the Average System):* By Lemma 2, there exists  $\mathcal{T} > 0$  such that  $\Lambda(\tau, s)$  is periodic in  $s$  with period  $\mathcal{T}$ . Indeed, since solutions to (8) are periodic, the matrix exponential  $e^{As}$  is periodic in  $s$ . Because  $B(\tau)$  is independent of  $s$ , the matrix  $\Lambda(\tau, s) = e^{-As} B(\tau) e^{As}$  is also periodic with period  $\mathcal{T}$ . By averaging (7) over  $s$ , we obtain the following slow time-varying<sup>2</sup> average system:

$$\frac{d\zeta}{ds} = \varepsilon \left( \bar{B}_1 + \frac{3}{\tau + T_0} \bar{B}_2 \right) \zeta, \quad \frac{d\tau}{ds} = \varepsilon, \quad (9)$$

where  $\bar{B}_k := \frac{1}{\mathcal{T}} \int_0^{\mathcal{T}} e^{-As} B_k e^{As} ds$ , for  $k \in \{1, 2\}$ , and  $B_1 := \begin{pmatrix} 0 & 0 \\ -\hat{Q}_a & 0 \end{pmatrix}$  and  $B_2 := \begin{pmatrix} 0 & 0 \\ 0 & -I \end{pmatrix}$ .

*Step 2 (Instability of the Average System):* We analyze the term  $\bar{B}_1$  in (9) which, as  $s \rightarrow \infty$ , dominates the stability properties of (9). First, since  $\hat{Q}_s$  is positive definite, by the spectral theorem [14, 2.5.6], there exists an orthonormal matrix  $P$  such that  $P^\top \hat{Q}_s P = \text{diag}\{q_k\}_{k=1}^n =: \Delta_s$ ,  $q_k > 0$ , where  $\{q_k\}_{k=1}^n = \text{spec}(\hat{Q}_s)$ , and where  $q_k > 0$ . Let  $\hat{P} := I_2 \otimes P$ , where  $\otimes$  denotes

<sup>1</sup> $\mathbb{Q}_{>0}$  denotes the set of positive rational numbers.

<sup>2</sup>The system is called slow time-varying because the resulting vector field depends on  $\tau$ , which evolves slowly since  $\frac{d\tau}{ds} = \varepsilon \ll 1$ .

the Kronecker product, and define  $\Delta := \hat{P}^\top A \hat{P} = \begin{pmatrix} 0 & I \\ -\Delta_s & 0 \end{pmatrix}$ . Then, we have that

$$\hat{P}^\top e^{\pm A\tau} \hat{P} = e^{\Delta\tau} = \begin{pmatrix} \mathcal{C}(s) & \pm \mathcal{S}_2(s) \\ \mp \mathcal{S}_1(s) & \mathcal{C}(s) \end{pmatrix}, \quad (10)$$

with  $\mathcal{C}(s) := \text{diag}\{\cos(\lambda_k s)\}_{k=1}^n$ ,  $\mathcal{S}_1(s) := \text{diag}\{\lambda_k \sin(\lambda_k s)\}_{k=1}^n$ ,  $\mathcal{S}_2(s) := \text{diag}\{\lambda_k^{-1} \sin(\lambda_k s)\}_{k=1}^n$ , and  $\lambda_k := \sqrt{q_k}$  for all  $k = \{1, 2, \dots, n\}$ . Using these definitions, we analyze the term  $\hat{P}^\top \bar{B}_1 \hat{P} = \frac{1}{T} \int_0^T e^{-\Delta s} \hat{P}^\top B_1 \hat{P} e^{\Delta s} ds$ . From (10), by letting  $\tilde{Q}_a := P^\top \hat{Q}_a P$ , we have that

$$e^{-\Delta s} \hat{P}^\top B_1 \hat{P} e^{\Delta s} = \begin{pmatrix} \mathcal{S}_2(s) \tilde{Q}_a \mathcal{C}(s) & \mathcal{S}_2(s) \tilde{Q}_a \mathcal{S}_2(s) \\ -\mathcal{C}(s) \tilde{Q}_a \mathcal{C}(s) & -\mathcal{C}(s) \tilde{Q}_a \mathcal{S}_2(s) \end{pmatrix}.$$

Then, from the definitions of  $\mathcal{C}(s)$  and  $\mathcal{S}_2(s)$ :

$$\begin{aligned} \left[ \mathcal{S}_2(s) \tilde{Q}_a \mathcal{C}(s) \right]_{ij} &= \lambda_i^{-1} \sin(\lambda_i s) \tilde{Q}_{a,ij} \cos(\lambda_j s) \\ \left[ \mathcal{S}_2(s) \tilde{Q}_a \mathcal{S}_2(s) \right]_{ij} &= \lambda_i^{-2} \sin(\lambda_i s) \tilde{Q}_{a,ij} \sin(\lambda_j s) \\ \left[ \mathcal{C}(s) \tilde{Q}_a \mathcal{C}(s) \right]_{ij} &= \cos(\lambda_i s) \tilde{Q}_{a,ij} \cos(\lambda_j s) \\ \left[ \mathcal{C}(s) \tilde{Q}_a \mathcal{S}_2(s) \right]_{ij} &= \lambda_j^{-1} \cos(\lambda_i s) \tilde{Q}_{a,ij} \sin(\lambda_j s). \end{aligned}$$

Since the period  $T$  can be written as the least common multiple of  $\{\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_n}\}$ , it follows that the terms  $\mathcal{C}(s) \tilde{Q}_a \mathcal{S}_2(s)$  and  $\mathcal{C}(s) \tilde{Q}_a \mathcal{C}(s)$  vanish under the averaging operation.

Therefore,  $\hat{P}^\top \bar{B}_1 \hat{P} = \frac{1}{T} \begin{pmatrix} 0 & \bar{Q}_a^{(2)} \\ \bar{Q}_a^{(1)} & 0 \end{pmatrix}$ , where  $(\bar{Q}_a^{(1)})_{ij} := -(\tilde{Q}_a)_{ij}$ ,  $(\bar{Q}_a^{(2)})_{ij} := (\tilde{Q}_a)_{ij}/q_i$ , and

$$(\bar{Q}_a)_{ij} := \begin{cases} 0 & \text{if } i = j \vee q_i \neq q_j, \\ \tilde{Q}_{a,ij} & \text{if } q_i = q_j \end{cases}, \quad (11)$$

where we used the fact that  $\tilde{Q}_a^\top = -\tilde{Q}_a$  which implies that  $\tilde{Q}_{a,ii} = 0$  for all  $i \in \{1, 2, \dots, n\}$ .

Given that  $Q_s$  has exactly one degenerate eigenvalue, there exists  $q > 0$  such that  $q_i = q$  for all indices  $i$  in some subset  $\mathcal{I} \subset \{1, 2, \dots, n\}$ . Also, for all  $(j, k) \in \mathcal{J} := \{1, 2, \dots, n\} \setminus \mathcal{I}$ , we have  $q_j \neq q_k \neq q$ . Thus, from (11), both  $\bar{Q}_a^{(1)}$  and  $\bar{Q}_a^{(2)}$  have zeros in their  $j^{\text{th}}$  row and  $j^{\text{th}}$  column for all  $j \in \mathcal{J}$ . Hence, given  $\lambda \in \mathbb{C}$ , we have that  $\det(\lambda I - \hat{P}^\top \bar{B}_1 \hat{P}) = \lambda^{|\mathcal{J}|} \det(\lambda^2 I + \frac{1}{2q} \bar{Q}_{\mathcal{J}}^2)$ , where  $|\mathcal{J}|$  denotes the cardinality of  $\mathcal{J}$ , and where  $\bar{Q}_{\mathcal{J}}$  is obtained from  $\tilde{Q}_a$  by removing its  $j^{\text{th}}$  row and column for every  $j \in \mathcal{J}$ . Therefore,

$$\{\lambda = \pm \sqrt{\frac{-\mu^2}{2q}} : \mu \in \text{spec}(\tilde{Q}_{\mathcal{J}})\} \subset \text{spec}(\bar{B}_1), \quad (12)$$

where we used the fact that the spectrum of a matrix is invariant under similarity transformations.

Now, since  $\tilde{Q}_a$  is skew-symmetric,  $\tilde{Q}_{\mathcal{J}}$  is also skew-symmetric by construction. Additionally, given that  $Q_{a,ij} \neq 0$  for all  $i, j \in \{1, 2, \dots, n\}$  by assumption, it follows that there exists a set  $\{\omega_k\}_{k=1}^m \subset \mathbb{R}_{>0}$ , with  $m = \lfloor |\mathcal{J}|/2 \rfloor$ , such that  $\{\pm i \omega_k\}_{k=1}^m \subset \text{spec}(\tilde{Q}_{\mathcal{J}})$ . Thus, letting  $\mu = i\omega_k$  in (12) yields  $\lambda = \pm \omega_k / \sqrt{2q}$  and  $\lambda \in \text{spec}(\bar{B}_1)$ , giving a positive real eigenvalue for  $\bar{B}_1$ . Hence, the origin is unstable under the  $\phi$  dynamics of (9).

*Step 3 (Instability of the Original System):* Instability of system (7) follows from Step 2 by applying the Floquet

Theorem in [10, Sec. 2]. Also, by Theorem 1 we note that  $|y(s)| = |\Phi_s^h(z)| = |e^{As} z(s)| \geq \sigma_{\min}(e^{As}) |z(s)|$  for all  $s \in \mathbb{R}_{\geq 0}$ , where  $\sigma_{\min}(\Lambda)$  denotes the minimum singular value of  $\Lambda \in \mathbb{R}^{n \times n}$ . Since  $\sigma_{\min}(e^{As}) > 0$  by the same reasoning used in the proof of Lemma 2, the instability of system (6) is implied by the instability of system (7). ■

*Remark 2:* Theorem 2 provides a theoretical explanation for the instability observed numerically in [5] whenever non-conservative terms appear in the vector field  $\mathcal{G}$  in (1). Such terms arise when  $\mathcal{G}$  is the pseudo-gradient of a non-potential game, or in consensus-based dynamics when the communication graph is directed [6]. Note that the instability emerges regardless of the size of  $Q_a$ . In this sense, Theorem 2 also provides a procedure for the synthesis of arbitrarily small (compared to the conservative part of  $\mathcal{G}$ ), adversarial, state-dependent perturbations able to destabilize (1) under conservative mappings of the form  $\mathcal{G} = \nabla J$ , by adding to  $\nabla J$  a perturbation of the form  $\text{rot } K$ .

## V. PRECLUDING INSTABILITY VIA RESTARTING

To address the instability of Nesterov's ODE under non-conservative mappings, we regularize the dynamics using resets that restart the momentum. Similar approaches have been studied in [15] under conservative maps, in [5] for Nash equilibrium-seeking problems in monotone games, and in [6] for consensus-based concurrent learning in directed-graphs. Compared to these works, we establish improved convergence rates and tighter resetting conditions for a subclass of vector fields satisfying Assumption 1-(i)(ii), that includes nonlinear vector fields beyond the linear setting studied in Section IV.

Starting from Nesterov's ODE in (1), we let  $q := x$ ,  $p := \dot{x}$ , and embed the dynamics in the HDS  $\mathcal{H} = (C, f, D, g)$  with state  $\chi = (q, p, \tau) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_{\geq 0}$ , and data

$$\begin{aligned} f(\chi) &:= \begin{pmatrix} p \\ -\frac{3}{\tau}p - \mathcal{G}(q) \\ \eta \end{pmatrix}, \quad g(\chi) := \begin{pmatrix} q \\ 0 \\ T_0 \end{pmatrix}, \\ C &:= \{(q, p, \tau) : q \in \mathbb{R}^n, p \in \mathbb{R}^n, \tau \in [T_0, T]\}, \\ D &:= \{(q, p, \tau) : q \in \mathbb{R}^n, p \in \mathbb{R}^n, \tau = T\}, \end{aligned}$$

where  $\eta \in (0, 1]$  and  $0 < T_0 < T$  are tunable parameters. The HDS periodically resets  $\tau$  to  $T_0$  when it reaches  $T$ , preventing the damping term  $-3p/\tau$  from vanishing, and  $p$  to 0 to ensure the strict decrease of a suitable Lyapunov function across jumps. For this HDS, we analyze the stability of the compact set  $\mathcal{A} := \{x^*\} \times \{0\} \times [T_0, T]$ , where  $x^* \in \mathbb{R}^n$  is the unique point<sup>3</sup> that satisfies  $\mathcal{G}(x^*) = 0$ .

*Theorem 3:* Let  $\mathcal{G}$  satisfy Assumption 1-(i),(ii). Suppose that  $\eta \in (0, 1)$ , and  $\underline{T} < T \leq \bar{T}$ , where  $\underline{T}^2 := T_0^2 + 4\eta^2/\kappa_J$  and  $\bar{T} := 2 \min\{3(1-\eta), \kappa_J \eta\}/\ell_K$ . Then, the set  $\mathcal{A}$  is UGES for  $\mathcal{H}$ . Moreover, there exists  $\rho > 0$  such that, for each compact set  $K_0 \subset C \cup D$ , there exist  $M_J, M_G > 0$  for which each solution  $\chi = (q, p, \tau)$  with  $\chi(0, 0) \in K_0$  satisfies

$$\tilde{J}(q(t, j)) \leq \frac{M_J T^2 e^{-\rho j}}{\tau^2(t, j)}, \quad |\mathcal{G}(q(t, j))|^2 \leq \frac{M_G T^2 e^{-\rho j}}{\tau^2(t, j)}$$

$\forall (t, j) \in \text{dom } \chi$ , where  $\tilde{J}(q) := J(q) - J(x^*) \quad \forall q \in \mathbb{R}^n$ .

<sup>3</sup>Guaranteed to exist due to the strong monotonicity properties of  $\mathcal{G}$ .

*Proof:* For each  $\chi \in C \cup D$ , consider the Lyapunov function

$$V(\chi) = a \left| q + \frac{\tau}{b} p - x^* \right|^2 + c \tau^2 |p|^2 + \delta \tau^2 (J(q) - J(x^*)),$$

where  $a = 2\eta b/T^2$ ,  $b := 3 - \eta$ ,  $c := \frac{3a(1-\eta)}{2\eta b^2}$ , and  $\delta := \frac{a}{\eta b} = \frac{2}{T^2}$ . Since  $\tau \in [T_0, T]$ , and  $J$  is strongly convex by Assumption 1, we obtain  $\underline{c} |\chi|_{\mathcal{A}}^2 \leq V(\chi) \leq \bar{c} |\chi|_{\mathcal{A}}^2$  where  $\underline{c} := T_0^2 \min\{c, \frac{\delta \kappa_J}{2}\}$ ,  $\bar{c} := \max\{a + \frac{aT}{b} + \frac{\delta T^2 \ell_J}{2}, mT^2 + \frac{aT}{b}\}$ , and  $m := \frac{a}{b^2} + c = \delta/2$ .

Computing  $\dot{V}(\chi) = \langle \nabla V(\chi), f(\chi) \rangle$ , and using the Helmholtz decomposition  $\mathcal{G}(q) = \nabla J(q) + \text{rot } K(q)$ , yields

$$\begin{aligned} \dot{V}(\chi) &= -2c\tau(3-\eta)|p|^2 \\ &\quad - \frac{2a\tau}{b} (\langle q - x^*, \nabla J(q) \rangle - (J(q) - J(x^*))) \\ &\quad - \frac{2a\tau}{b} \langle q - x^*, \text{rot } K(q) \rangle - \delta \tau^2 \langle p, \text{rot } K(q) \rangle \quad \forall \chi \in C. \end{aligned}$$

By Assumption 1,  $\langle \nabla J(q), q - x^* \rangle - (J(q) - J(x^*)) \geq \frac{\kappa_J}{2} |q - x^*|^2$ ,  $\langle q - x^*, \text{rot } K(q) \rangle \geq 0$ , and  $|\text{rot } K(q)| \leq \ell_K |q - x^*|$  for all  $q \in \mathbb{R}^n$ . Using these facts, together with the Cauchy-Schwarz and Young's inequalities, we obtain that

$$\begin{aligned} \dot{V}(\chi) &\leq -\tau \lambda \left( |p|^2 + |q - x^*|^2 \right) + \frac{\delta T \tau \ell_K}{2} \left( |p|^2 + |q - x^*|^2 \right) \\ &= -\lambda \tau \left( 1 - \frac{T}{\bar{T}} \right) |\chi|_{\mathcal{A}}^2 \leq -\mu V(\chi) \quad \forall \chi \in C, \end{aligned}$$

where  $\lambda := \min\{2c(3-\eta), \frac{a\kappa_J}{b}\}$  and  $\mu := \lambda(\bar{T} - T)T_0/\bar{T}c$ .

On the other hand, for  $\chi \in D$ ,  $V(g(\chi)) - V(\chi) = -mT^2 |p|^2 - \frac{2aT}{b} \langle q - x^*, p \rangle - \delta(T^2 - T_0^2)(J(q) - J(x^*))$ . Defining  $\Gamma := \sqrt{(T^2 - T_0^2)\kappa_J}$ , and applying Cauchy-Schwarz together with Young's inequality with parameter  $\theta = T/\Gamma$ , and using the strong convexity of  $J$ , by letting  $\Delta V(\chi) := V(g(\chi)) - V(\chi)$ , we obtain that

$$\begin{aligned} \Delta V(\chi) &\leq -\left( \frac{aT^2}{2\eta b} - \frac{aT^2}{b\Gamma} \right) |p|^2 - \left( \frac{a\Gamma^2}{2\eta b} - \frac{a\Gamma}{b} \right) |q - x^*|^2 \\ &= -\nu_1 |p|^2 - \nu_2 |q - x^*|^2 \leq -\frac{\nu}{\bar{c}} V(\chi) \end{aligned}$$

for all  $\chi \in D$ , where  $\nu_1 := 1 - \frac{2\eta}{\Gamma}$ ,  $\nu_2 := \frac{\Gamma(\Gamma-2\eta)}{T^2}$ , and  $\nu := \min\{\nu_1, \nu_2\}$ . Since  $T > \bar{T}$  implies  $\Gamma > 2\eta$ ,  $\nu \in (0, 1)$ .

The quadratic bounds on  $V$ , combined with the strict decrease of  $V$  during the flows and jumps of  $\mathcal{H}$  establish that  $\mathcal{A}$  is UGES for  $\mathcal{H}$  via [16, Th. 1]. To obtain the convergence bounds for  $\mathcal{G}$  and  $\tilde{J}$ , let  $K_0 \subset C \cup D$  be compact. For any solution  $\chi$  to  $\mathcal{H}$  with  $\chi(0, 0) \in K_0$ , define  $c_j := V(\chi(s_j, j))$  where  $s_j := \min\{t : (t, j) \in \text{dom } \chi\}$ . The flow and jump conditions give  $c_{j+1} \leq e^{-\rho} c_j$  where  $\rho := -\ln(1 - \frac{\nu}{\bar{c}}) + \mu(T - T_0) > 0$ . Thus,  $c_j \leq e^{-\rho j} V(\chi(0, 0))$ . From  $\delta \tau^2 (J(q) - J(x^*)) \leq V(\chi)$ , we obtain  $\tilde{J}(q(t, j)) \leq V(\chi(0, 0)) T^2 e^{-\rho j} / 2\tau^2(t, j)$  for all  $(t, j) \in \text{dom } \chi$ . Using strong convexity and Lipschitz continuity of  $\mathcal{G}$  gives  $|\mathcal{G}(q(t, j))|^2 \leq (\ell_J + \ell_K)^2 V(\chi(0, 0)) T^2 e^{-\rho j} \kappa_J \tau^2(t, j)$ . Letting  $M_J := \frac{1}{2} \max_{\chi \in K_0} V(\chi)$  and  $M_{\mathcal{G}} := \frac{2(\ell_J + \ell_K)^2}{\kappa_J} M_J$  completes the proof. Additional details are presented in the extended manuscript [17]. ■

*Corollary 1:* Under the assumptions of Theorem 3, as  $T_0 \rightarrow 0$ ,  $T \in [2\bar{T}, e\bar{T}]$  guarantees convergence of  $q(t, j)$  to  $x^*$  at a rate of order  $\mathcal{O}(e^{-\eta\sqrt{\kappa_J}t})$ .

*Proof:* For any solution  $\chi = (q, p, \tau)$  to  $\mathcal{H}$ , the state  $\tau$  resets every  $(T - T_0)/\eta$  units of time, giving  $j \leq \frac{\eta t}{T - T_0}$  for any  $(t, j) \in \text{dom } \chi$ . Additionally, from Theorem 3 and strong convexity, we obtain  $|q(t, j) - x^*|^2 \leq k_M \exp(-\alpha(T, T_0)t)$  where  $\alpha(T, T_0) = -\frac{\eta}{T - T_0} \ln(1 - \frac{\nu}{\bar{c}})$ .

As  $T_0 \rightarrow 0$ , we have  $\nu \rightarrow \min\{1, \kappa_J\}(1 - \frac{\bar{T}}{\bar{T}})$  where  $\bar{T} \rightarrow \frac{2\eta}{\sqrt{\kappa_J}}$ . Setting  $\xi = \frac{T}{\bar{T}}$  and  $\beta = \frac{\min\{1, \kappa_J\}}{\bar{c}}$ , we maximize  $f(\xi) = -\xi \ln(1 - \beta(1 - \xi))$ . The unique critical point satisfies  $\ln(1 - \beta(1 - \xi^*)) + \frac{\beta \xi^*}{1 - \beta(1 - \xi^*)} = 0$ , which can be verified to be a maximum by checking  $g(\xi) = \ln(1 - \beta(1 - \xi)) + \frac{\beta \xi}{1 - \beta(1 - \xi)}$  is strictly increasing with  $g(0) < 0$  and  $\lim_{\xi \rightarrow 1^-} g(\xi) > 0$ . For  $\beta = 1$ , we obtain  $\xi^* = e^{-1}$ , giving  $T^{\text{opt}} = e\bar{T}$ . For small  $\beta$ ,  $\xi^* \approx 1/2$ , giving  $T^{\text{opt}} \approx 2\bar{T}$ . In general,  $T^{\text{opt}} \in [2\bar{T}, e\bar{T}]$ .

At  $T = T^{\text{opt}}$ ,  $\alpha(T) = \frac{\sqrt{\kappa_J}}{2} \frac{\beta(\xi^*)^2}{1 - \beta(1 - \xi^*)}$ , yielding convergence rate  $\mathcal{O}(e^{-c\eta\sqrt{\kappa_J}t})$  for  $c > 0$ . Additional details are presented in the extended manuscript [17]. ■

*Remark 3:* Theorem 3 yields the *semi-acceleration* bound  $\mathcal{O}(1/\tau(t, j)^2)$  along flow intervals, matching the classical  $\mathcal{O}(1/t^2)$  rate of Nesterov-type ODEs with damping  $3/t$  for smooth convex objectives [2] during intervals of flow. In low-curvature regimes ( $\kappa_J < 1$ ), Corollary 1 improves upon gradient-like dynamics  $\dot{x} = -\mathcal{G}(x)$ , which attain  $\mathcal{O}(e^{-\kappa_J t})$  convergence rates under Assumption 1-(i),(ii), and refines [5, Lemma 5] by removing the factor  $\sigma := (\ell_J + \ell_K)/\kappa_J$  from the exponent in  $\mathcal{O}(e^{-(\sqrt{\kappa_J}/\sigma)t})$ . When  $\mathcal{G} = \nabla J$ , existing acceleration results achieve exponential convergence with exponent proportional to  $\sqrt{\kappa_J}$  [3, Th. 1]; our result attains the same qualitative  $\sqrt{\kappa_J}$  scaling without the gradient-structure assumption.

## VI. NUMERICAL SIMULATIONS

*Instability Example:* We consider a non-conservative driving term  $\mathcal{G}(x) = (Q_s + Q_a)x = \ell_J(\hat{Q}_s + \frac{1}{\sqrt{\ell_J}} \hat{Q}_a)x$  with  $\hat{Q}_s := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ ,  $\hat{Q}_a := \begin{pmatrix} 0 & 0.5 \\ -0.5 & 0 \end{pmatrix}$ , and  $\ell_J = 100$ , leading to  $\varepsilon = 0.1$ . With these choices, all the conditions of Theorem 2 are satisfied. We set  $\tau(0) = T_0 = 0.1$ ,  $y(0) = z(0) = (0.1, -0.1, 0, 0)$  for systems (6)-(7). We also simulate system (8) from  $\psi(0) = y(0)$  and the average system (9) from  $\zeta(0) = z(0)$ . The trajectories are shown in Figure 1. The left plot shows the periodic behavior of the linear system (8), where the  $\psi$ -components exhibit bounded oscillations consistent with the periodic flow predicted by Lemma 2. The middle plot validates the averaging analysis by showing how the solution  $z(\tau)$  of the pulled-back system (7) closely follows the solution  $\zeta(\tau)$  of the averaged system (9) on the timescale  $\mathcal{O}(1/\varepsilon)$ . The right plot confirms the instability predicted by Theorem 2.

*Precluding Instability via Restarting:* We simulate the restarting hybrid dynamics  $\mathcal{H}$  using the same vector field  $\mathcal{G}(x) = (Q_s + Q_a)x$ , which yields  $\ell_K = 5$ ,  $\ell_J = 100$ , and  $\kappa_J = 100$ . We set  $\eta = 1/2$ ,  $T_0 = 0.1$ , and  $T = 0.3844$ , which satisfy the condition  $0.14 = \bar{T} < T \leq \bar{T} = 0.6$  from Theorem 3. Using  $\chi(0, 0) = (10^4, -10^4, 10^4, -10^4, 0.1)$ , Figure 2 shows the evolution of  $|q(t) - x^*|$  for both Nesterov's ODE and the HDS  $\mathcal{H}$ . The red stars indicate the restarting events. The simulations confirm both the UGES property of Theorem 3 and the  $\mathcal{O}(e^{-\sqrt{\kappa_J}t})$  convergence rate of Corollary 1.

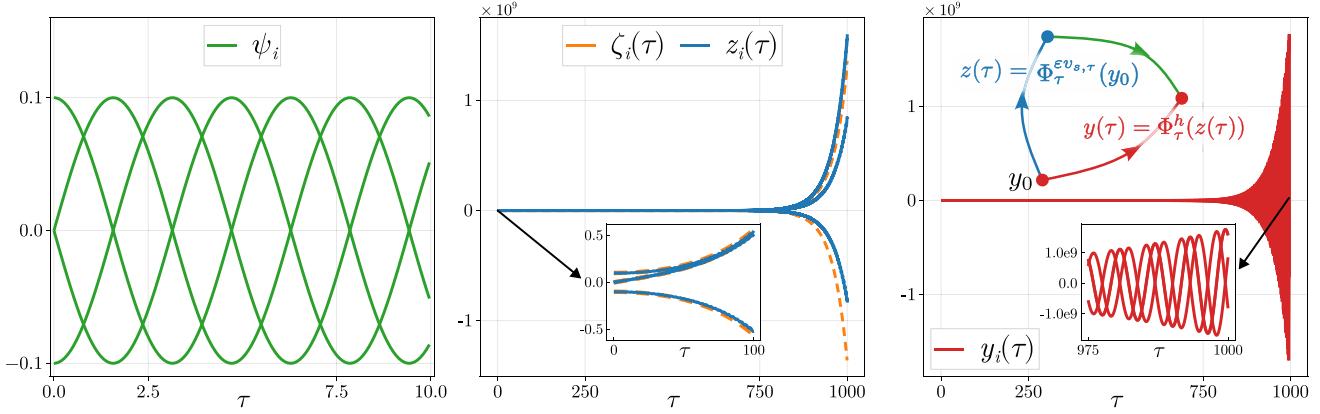


Fig. 1. Left: Components of  $\psi$  showing a solution of the periodic system (8). Middle: Components of  $z$  and  $\zeta$  showing the solutions of the pulled-back system (7) and the averaged system (9). Right: Components of  $y$  showing a solution of Nesterov's ODE in (6).

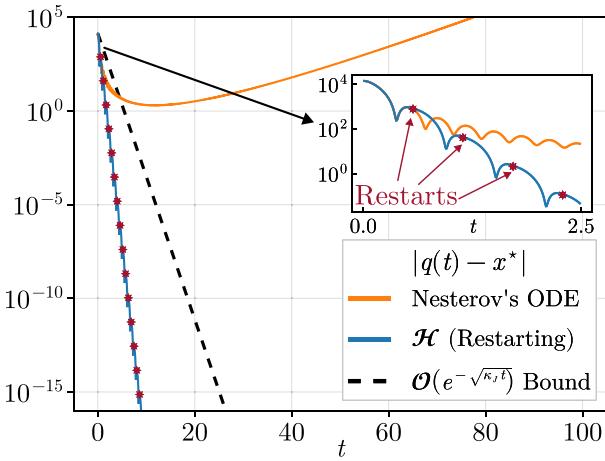


Fig. 2. Comparison of trajectories of  $|q(t) - x^*|$  under Nesterov's ODE and the restarting HDS  $\mathcal{H}$ .

## VII. CONCLUSION AND FUTURE DIRECTIONS

In this letter, we prove that for linear vector fields with nonzero skew-symmetric components, Nesterov's ODE can exhibit instability even under strong monotonicity conditions. To resolve this instability, we design a hybrid dynamical system that achieves robust stability while inducing convergence rates of order  $\mathcal{O}(e^{-\sqrt{\kappa_J}t})$  through periodic restarting mechanisms. Future work will focus on Lie bracket expansions of the pullback operator to analyze the instability of Nesterov's ODE in the case of nonlinear vector fields.

## REFERENCES

- [1] Y. Nesterov, "A method for solving the convex programming problem with convergence rate  $\mathcal{O}(\frac{1}{k^2})$ ," *Dokl. Akad. Nauk SSSR*, vol. 269, no. 3, pp. 543–547, 1983.
- [2] W. Su, S. Boyd, and E. J. Candès, "A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights," *J. Mach. Learn. Res.*, vol. 17, no. 153, pp. 1–43, 2016.
- [3] B. Shi, S. S. Du, M. I. Jordan, and W. J. Su, "Understanding the acceleration phenomenon via high-resolution differential equations," *Math. Program.*, vol. 195, no. 1, pp. 79–148, Sep. 2022.
- [4] B. Gharesifard and J. Cortés, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 781–786, Mar. 2014.
- [5] D. E. Ochoa and J. I. Poveda, "Momentum-based Nash set-seeking over networks via multitime scale hybrid dynamic inclusions," *IEEE Trans. Autom. Control*, vol. 69, no. 7, pp. 4245–4260, Jul. 2024.
- [6] D. E. Ochoa, M. U. Javed, X. Chen, and J. I. Poveda, "Decentralized concurrent learning with coordinated momentum and restart," *Syst. Control Lett.*, vol. 193, Nov. 2024, Art. no. 105931.
- [7] J. I. Poveda and N. Li, "Inducing uniform asymptotic stability in non-autonomous accelerated optimization dynamics via hybrid regularization," in *Proc. IEEE 58th Conf. Decis. Control (CDC)*, 2019, pp. 3000–3005.
- [8] E. Glötzl and O. Richters, "Helmholtz decomposition and potential functions for n-dimensional analytic vector fields," *J. Math. Anal. Appl.*, vol. 525, no. 2, Sep. 2023, Art. no. 127138.
- [9] F. Bullo and A. D. Lewis, *Geometric Control of Mechanical Systems: Modeling, Analysis, and Design for Simple Mechanical Control Systems* (Applied Mathematics), vol. 49. New York, NY, USA: Springer, 2005.
- [10] A. V. Sarychev, "Lie- and chronologico-algebraic tools for studying stability of time-varying systems," *Syst. Control Lett.*, vol. 43, no. 1, pp. 59–76, May 2001.
- [11] R. Goebel, R. G. Sanfelice, and A. R. Teel, *Hybrid Dynamical Systems: Modeling, Stability, and Robustness*. Princeton, NJ, USA: Princeton Univ. Press, 2012.
- [12] F. Bullo, "Averaging and vibrational control of mechanical systems," *SIAM J. Control Optim.*, vol. 41, no. 2, pp. 542–562, Jan. 2002.
- [13] J. A. Sanders, F. Verhulst, and J. Murdock, *Averaging Methods in Nonlinear Dynamical Systems*, vol. 59. New York, NY, USA: Springer, 2007.
- [14] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2017.
- [15] A. R. Teel, J. I. Poveda, and J. Le, "First-order optimization algorithms with resets and Hamiltonian flows," in *Proc. IEEE 58th Conf. Decis. Control (CDC)*, 2019, pp. 5838–5843.
- [16] A. R. Teel, F. Forni, and L. Zaccarian, "Lyapunov-based sufficient conditions for exponential stability in hybrid systems," *IEEE Trans. Autom. Control*, vol. 58, no. 6, pp. 1591–1596, Jun. 2013.
- [17] D. E. Ochoa, M. Abdelgalil, and J. I. Poveda, "On the instability of Nesterov's ODE under non-conservative vector fields," 2025, *arXiv:2501.13244*.