2020 American Control Conference
Denver, CO, USA, July 1-3, 2020

The Heavy-Ball ODE with Time-Varying Damping: Persistence of Excitation and Uniform Asymptotic Stability

Jorge I. Poveda and Andrew R. Teel

***Abstract*— We study the uniform asymptotic stability properties of the heavy-ball optimization dynamics with general time-varying damping. Unlike existing results in the literature, which have focused mainly on standard convergence results, we study a stronger limiting notion called uniform asymptotic stability, which is instrumental for the design of feedback-based algorithms. Given that recent results in the literature have shown that a class of heavy-ball optimization dynamics with vanishing damping fails to satisfy this limiting notion, we study sufficient and necessary conditions on the time-varying coefficients such that uniform asymptotic stability for the set of minimizers of the cost function is achieved. Our main results show that such conditions are related to the notion of persistence of excitation, which is commonly used in adaptive control and system identification. Moreover, we show that the persistence of excitation condition is not necessary for a class of high-resolution accelerated optimization dynamics with Hessian-driven damping. Our results are established by using a nested Matrosov theorem that has not been used before in the analysis of accelerated optimization algorithms.**

I. INTRODUCTION

Recent years have seen a renewed interest in the analysis and design of efficient optimization algorithms using tools from dynamical systems theory, see for instance [1], [2], [3]. For instance, in [4] the authors linked Nesterov's optimization algorithm [5], given by the time-varying recursion

$$x_1^+ = x_2 - s\nabla\phi(x_2) \tag{1a}$$

$$x_2^+ = x_1^+ + \frac{j-1}{j+2}\Big(x_1^+ - x_1\Big), \tag{1b}$$

with $j \in \mathbb{Z}_{\geq 0}$, to the second order differential equation (ODE) given by

$$\ddot{z} + a(t)\dot{z} + \nabla\phi(z) = 0, \qquad t \geq t_0. \tag{2}$$

In particular, when $a(t) = 3/t$ the authors in [4] derived the time-varying ODE (2) by using the relation $t \approx k\sqrt{s}$ and taking the limit of the step size s to zero in (1). Moreover, in this case by suitable choices of energy functions it has been shown that any trajectory of (2) with $\dot{x}(0) = 0$ minimizes a smooth convex function ϕ at least at a rate of $\mathcal{O}(1/t^2)$, see [4], [6]. General dynamical systems of the form (2) have been extensively studied in the literature, e.g., [7], [8]. Several results have provided sufficient conditions on $a(\cdot)$ to guarantee that the state $z(t)$ converges to the set of minimizers of ϕ as $t \to \infty$. Similarly, unperturbed and perturbed heavy-ball dynamics with vanishing damping were studied in [9] and [10] under certain integrable conditions on the perturbation. Convergence properties for the continuous-time Nesterov's ODE and the heavy ball algorithm were also recently studied in [11]. Algorithms that incorporate continuous-time and discrete-time dynamics have been presented recently in [12], [13], [14], [15], [17], and [16].

While most of the results in the literature have focused on studying the convergence properties of equation (2), few works have studied its stability and robustness properties. For instance, it was recently shown in [13] that for a class of smooth convex cost functions the convergence properties of system (2) with $a(t) = c/t$ have no margins of robustness against arbitrarily small additive disturbances for any $c > 0$, which is problematic in feedback-based algorithms for real-time optimization, adaptive control, hybrid control, etc. Motivated by this fundamental limitation, in this paper we study sufficient and necessary conditions for uniform global asymptotic stability (UGAS) in a class of time-varying continuous-time optimization algorithms that cover systems of the form (2), and which can be written as

$$\dot{x} = -\Big(A(t) + kB(t,x)\Big)\frac{\partial H(x)}{\partial x}, \quad k \geq 0 \tag{3}$$

where the mapping $H : \mathbb{R}^n \to \mathbb{R}$ is a Hamiltonian function, $B : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \to \mathbb{R}^{n\times n}$, and $A : \mathbb{R}_{\geq 0} \to \mathbb{R}^{n\times n}$ is a positive semidefinite matrix. For the case when $k = 0$, system (3) was called a *degenerate gradient flow* (DGF) in [18]. These dynamics model Nesterov's ODE (2) under a particular choice of Hamiltonian and matrix A. On the other hand, when $k \neq 0$, system (3) can model a class of "high-resolution" optimization ODEs recently studied in [19], [20].

Based on these observations, the main contributions of this paper are the following: (a) We show that when $k = 0$ and $t \mapsto a(t)$ satisfies a persistence of excitation condition, the DGF (3) corresponding to the time-varying heavy-ball renders UGAS the set of minimizers of any cost function that is continuously differentiable, radially unbounded, and convex or gradient-dominated; (b) We establish that persistence of excitation of $a(\cdot)$ is not only a sufficient condition for UGAS, but also a *necessary* condition. Given that vanishing functions such as $1/t$ or $1/(t+1)$ are not persistently exciting, this result provides an alternative proof of the fact that Nesterov's ODE lacks uniform asymptotic stability properties, an observation made in [13] using Artstein's idea of limiting equations [21]; (c) We show that when $k \neq 0$

Jorge I. Poveda is with the Department of Electrical, Computer, and Energy Engineering, University of Colorado, Boulder, CO, 80309. Email:{jorge.poveda@colorado.edu.}

Andrew R. Teel is with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106-9560. Email:{teel@ucsb.edu}.

Research supported in part by CU Boulder Autonomous Systems Seed Grant No. 11005946, by AFOSR grant FA9550-18-1-0246, and by NSF grant ECCS-1508757.

and system (3) models a family of "high-resolution" ODEs also studied in [19] and [20], persistence of excitation is not needed in order to render UGAS the set of minimizers of the cost function; (d) Finally, we link the UGAS properties of system (2) with well-known robustness results developed in the control's literature, e.g., [22]. Our sufficiency results are established by using a nested Matrosov's theorem for time-varying dynamical systems; see [23] and [24] for related analytical tools. To our knowledge, this is the first paper that uses Matrosov's theorem to analyze accelerated optimization algorithms.

II. PRELIMINARIES

A. Notation

Given a compact set $\mathcal{A} \subset \mathbb{R}^n$ and a vector $z \in \mathbb{R}^n$, we define $|z|_{\mathcal{A}} := \min_{y\in\mathcal{A}} |z - y|$, and we use $|\cdot|$ to denote the standard Euclidean norm. We use $\mathbb{B}$ to denote a closed unit ball of appropriate dimension, $\rho\mathbb{B}$ to denote a closed ball of radius $\rho > 0$, and $\mathcal{A} + \rho\mathbb{B}$ to denote the union of all sets obtained by taking a closed ball of radius ρ around each point in the set $\mathcal{A}$. Throughout the paper, for two vectors $u, v \in \mathbb{R}^n$, we use $\langle u, v\rangle := u^T v$. In addition, for ease of notation we write (u, v) for $(u^T, v^T)^T$. A function $\phi : \mathbb{R}^n \to \mathbb{R}$ is said to be: a) C^k if its k-th derivative is continuous; b) radially unbounded if $\phi(z) \to \infty$ whenever $|z| \to \infty$, c) invex if every critical point is a global minimizer [25], i.e., $\{z \in \mathbb{R}^n : \nabla\phi(z) = 0\} = \arg\ \min_{z\in\mathbb{R}^n} \phi(z)$. A function $\alpha : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is said to be of class $\mathcal{K}_\infty$ if it is zero at zero, continuous, strictly increasing, and unbounded. A function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ is said to be of class $\mathcal{KL}$ if it is nondecreasing in its first argument, non-increasing in its second argument, $\lim_{r\to 0^+} \beta(r, s) = 0$ for each $s \in \mathbb{R}_{\geq 0}$, and $\lim_{s\to\infty} \beta(r, s) = 0$ for each $r \in \mathbb{R}_{\geq 0}$.

B. Dynamical Systems: Stability and Robustness

In this paper, we consider dynamical systems of the form

$$\dot{x} = f(t, x), \qquad t \geq t_0, \quad x_0 := x(t_0) \in \mathbb{R}^n, \tag{4}$$

where $f : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \to \mathbb{R}^n$ is measurable in t and locally Lipschitz in x. For these systems we are interested in establishing *uniform global asymptotic stability properties*, which are characterized by the following definitions.

Definition 1: A compact set $\mathcal{A} \subset \mathbb{R}^n$ is said to be uniformly globally stable (UGS) if there exists $\alpha \in \mathcal{K}_\infty$ such that, for each $(t_0, x_0) \in \mathbb{R}_{\geq 0} \times \mathbb{R}^n$ each solution of (4) satisfies $|x(t)|_{\mathcal{A}} \leq \alpha(|x_0|_{\mathcal{A}})$, for all $t \geq t_0$. □

Definition 2: A compact set $\mathcal{A} \subset \mathbb{R}^n$ is said to be uniformly globally attractive (UGA) if for each $r, \delta > 0$ there exists $T > 0$ such that each solution of (4) satisfies $|x_0|_{\mathcal{A}} \leq r \implies |x(t)|_{\mathcal{A}} \leq \delta$, for all $t \geq t_0 + T$. □

Definition 3: For system (4) a compact set $\mathcal{A} \subset \mathbb{R}^n$ is said to be uniformly globally asymptotically stable (UGAS) if it is UGS and UGA. □

The UGAS property can be equivalently characterized using $\mathcal{KL}$ functions: A compact set $\mathcal{A}$ is UGAS for system (4) if there exists a $\beta \in \mathcal{KL}$ such that every solution of (4) satisfies the bound

$$|x(t)|_{\mathcal{A}} \leq \beta(|x_0|_{\mathcal{A}}, t - t_0), \tag{5}$$

for all $t \geq t_0$. A traditional approach to establish UGAS in systems of the form (4) is to use suitable Lyapunov functions [22].

Lemma 1: Let $V : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \to \mathbb{R}$ be a C^1 function that satisfies the following conditions:

$$\alpha_1(|x|_{\mathcal{A}}) \leq V(t, x) \leq \alpha_2(|x|_{\mathcal{A}}), \tag{6a}$$

$$\dot{V}(t, x) := \frac{\partial V}{\partial t} + \frac{\partial V}{\partial x} f(t, x) \leq -\alpha_3(|x|_{\mathcal{A}}), \tag{6b}$$

for all $t \geq t_0 \geq 0$ and for all $x \in \mathbb{R}^n$, where $\alpha_1, \alpha_2 \in \mathcal{K}_\infty$, and $\alpha_3 : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is continuous and positive definite. Then, system (4) renders the set $\mathcal{A}$ UGAS. □

An appealing property of smooth dynamical systems that satisfy a bound of the form (5) is that, even if a Lyapunov function satisfying (6) is not readily available, desirable robustness properties can still be directly obtained via converse Lyapunov theorems. The following lemma is a consequence of [22, Thm. 4.16] and calculations like in the proof of [22, Lemma 9.3].

Lemma 2: Consider system (4) with f being C^2, and $\partial f/\partial x$ bounded on compact sets, uniformly in t. Suppose that (4) renders a nonempty compact set $\mathcal{A} \subset \mathbb{R}^n$ UGAS. Then, for each pair $\Delta > \nu > 0$ there exists $\varepsilon^* > 0$ such that for all measurable functions $e : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ satisfying $|e(t)| \leq \varepsilon^*$, all solutions with $x_0 \in \mathcal{A} + \Delta\mathbb{B}$ of the perturbed system

$$\dot{x} = f(t, x + e) + e,$$

satisfy the bound $|x(t)|_{\mathcal{A}} \leq \beta(|x_0|_{\mathcal{A}}, t - t_0) + \nu$, for all $t \geq t_0$. □

Lemma 2 establishes a structural robustness property with respect to sufficiently small bounded additive perturbations of arbitrary frequency, such as noisy measurements, unmodeled dynamics, etc. This property is relevant for applications where *feedback-based* algorithms are needed.

In many cases, the main difficulty in establishing UGAS via Lyapunov functions is to find a suitable function $\alpha_3(\cdot)$, independent of t, that satisfies condition (6b). It is well known that for *time-invariant* dynamical systems the condition $\dot{V}(x) \leq 0$ for all $x \in \mathbb{R}^n$, together with knowing that no solution x with an unbounded time domain keeps $t \mapsto V(x(t))$ equal to a nonzero constant, can be used to establish UGAS via the Krasovskii-LaSalle's invariance principle. However, for time-varying systems the application of the invariance principle is in general more difficult. However, in this case, Matrosov's theorem can be used to asses the stability properties of system (4).

C. Nested Matrosov Theorem and Persistence of Excitation

The key idea behind Matrosov's theorem (see [23], [24]) is to combine a Lyapunov function that establishes uniform global stability, with auxiliary functions that can be used to established uniform global attractivity.

Lemma 3 (Matrosov Theorem): Let $\mathcal{A} \subset \mathbb{R}^n$ be a compact set, and consider the time-varying system (4). Then, under the following assumptions the set $\mathcal{A}$ is UGAS:

1) The set $\mathcal{A}$ is uniformly globally stable.
2) There exists integers $j,\ m > 0$, and for each $\Delta > 0$ there exist a number $r > 0$ and:
   - Locally Lipschitz continuous functions $V_i : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \to \mathbb{R}$, $i \in \{1, 2, \ldots, j\}$.
   - A continuous function $\xi : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \to \mathbb{R}^m$.
   - Continuous functions $Y_i : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $i \in \{1, 2, \ldots, j\}$.

   such that, for almost all $(t, x) \in \mathbb{R}_{\geq 0} \times (\mathcal{A} + \Delta\mathbb{B})$,
$$\max\{|V_i(t,x)|, |\xi(t,x)|\} \leq \rho,$$
$$\dot{V}_i(t,x) \leq Y_i(x, \xi(t,x)).$$
3) For each integer $k \in \{1, 2, \ldots, j\}$ we have that
$$\Big\{(z,\psi) \in (\mathcal{A} + \Delta\mathbb{B}) \times \Delta\mathbb{B},\ Y_i(z,\psi) = 0,$$
$$\forall\ i \in \{1, 2, \ldots, k-1\}\Big\} \implies Y_k(z,\psi) \leq 0.$$
4) We have that
$$\Big\{(z,\psi) \in (\mathcal{A} + \Delta\mathbb{B}) \times \Delta\mathbb{B},\ Y_i(z,\psi) = 0,$$
$$\forall\ i \in \{1, 2, \ldots, j\}\Big\} \implies z \in \mathcal{A}. \qquad \square$$

Matrosov's theorem is sometimes used in combination with a class of persistence of excitation conditions that are used in system identification and adaptive control. We formalize this idea with the following definition:

Definition 4: A function $\lambda : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is said to be *persistently exciting* (PE) if $\exists\ T, \mu > 0$ such that
$$\int_t^{t+T} \lambda(\tau) d\tau \geq \mu, \tag{7}$$
for all $t \geq 0$. $\square$

The following Lemma will be instrumental for our results.

Lemma 4: Let $\lambda : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a bounded and globally Lipschitz function. If λ is PE, then the function $t \mapsto \lambda(t)^p$ is also PE for each $p > 0$. $\square$

III. MAIN RESULTS

Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a C^1 function satisfying $\phi^* := \min_{z \in \mathbb{R}^n} \phi(z) > -\infty$, and consider system (3) with state $x = (x_1, x_2)$, and Hamiltonian function
$$H(x) = \phi(x_1) - \phi^* + 0.5|x_2|^2. \tag{8}$$
When the matrices $A(t)$ and $B(t,x)$ are selected as
$$A(t) = \begin{bmatrix} 0 & -I \\ I & a(t)I \end{bmatrix},\ B(t,x) = \begin{bmatrix} 0 & 0 \\ b(t)I & \nabla^2\phi(x_1) \end{bmatrix}, \tag{9}$$
system (3) reduces to
$$\dot{x}_1 = \frac{\partial H(x)}{\partial x_2},\quad \dot{x}_2 = -u_1(t,x)\frac{\partial H(x)}{\partial x_2} - u_2(t)\frac{\partial H(x)}{\partial x_1}, \tag{10}$$
where $u_1(t,x) := a(t)I + k\nabla^2\phi(x_1)$ and $u_2(t) := 1 + kb(t)$. By using $x_1 = z$ and $x_2 = \dot{z}$, system (10) corresponds to the second order "high-resolution" ODE [19], [20]:
$$\ddot{z} + a(t)\dot{z} + \nabla\phi(z) = -k\Big(\nabla^2\phi(z)\dot{z} + b(t)\nabla\phi(z)\Big),$$
which was recently studied in [19], [20]. For the particular case when $k = 0$, the dynamics (10) further reduce to the heavy-ball dynamics with time-varying damping (2).

In this paper we are interested in finding conditions on the time-varying functions $a(\cdot)$ and $b(\cdot)$, as well as the cost function $\phi(\cdot)$, such that the set
$$\mathcal{A} := \mathcal{A}_\phi \times \{0\},\quad \text{where}\quad \mathcal{A}_\phi := \arg \min_{x_1 \in \mathbb{R}^n} \phi(x_1), \tag{11}$$
is UGAS for system (3). We start by characterizing a class of sufficient conditions on the signal $t \mapsto a(t)$ that achieves this goal for the case when $k = 0$.

Theorem 1: Consider the gradient flow (3) with Hamiltonian given by (8), and matrix A given by (9). Let $k = 0$ and suppose that the following assumptions hold:

1) The function $a(\cdot)$ is globally Lipschitz, and there exists $\gamma_1 > 0$ such that $a(t) \in [0, \gamma_1]$ for all $t \geq 0$.
2) The function $a(\cdot)$ is persistently exciting.
3) The cost function $\phi(\cdot)$ is radially unbounded, C^2, and satisfies at least one of the following conditions:
   a) $\phi(\cdot)$ is convex.
   b) $\phi(\cdot)$ is gradient-dominated, i.e., $\exists\ d > 0$ such that it satisfies the Polyak-Lojasiewicz (PL) inequality: $\phi(x_1) - \phi^* \leq \frac{1}{2d}|\nabla\phi(x_1)|^2$, for all $x_1 \in \mathbb{R}^n$.

Then, the set $\mathcal{A}$ is UGAS. $\square$

The PL inequality implies that ϕ is an invex function [25]. However, in general, the PL inequality does not imply convexity. For instance, the function $\phi(x_1) = x_1^2 + 3\sin^2(x_1)$ is invex but not convex, and it satisfies the PL inequality with $d = 1/32$; see [26, pp. 4].

Remark 1: Any positive constant or continuous periodic function $a : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ that is not identically zero satisfies the PE condition (7). However, condition (7) can also be satisfied by functions with arbitrarily large null sets, such as $a(t) = \max\left\{0, \frac{t}{1+t}\sin^3\left(\frac{t}{r}\right)\right\}$, which can have arbitrarily large null sets by taking r arbitrarily large [27, Ch. 6.3]. $\square$

The assumptions of Theorem 1 are satisfied by a variety of time-varying functions. However, condition (7) rules out any time-varying function $t \mapsto a(t)$ that converges to 0 as $t \to \infty$, including the function $a(t) = \frac{c}{t}$, for $t \geq t_0 > 0$, with $c > 0$, used in Nesterov's ODE [4].

The following theorem, which corresponds to the second main result of this paper, establishes that PE of $t \mapsto a(t)$ is indeed a necessary condition in order to obtain UGAS of $\mathcal{A}$.

Theorem 2: Consider the gradient flow (3) with Hamiltonian (8) and matrix A given by (9). Let $k = 0$ and suppose that the following holds:

1) The function $a : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is measurable.
2) The function ϕ is C^1 and radially unbounded.

Then, if the set $\mathcal{A} \times \{0\}$ is UGAS, the function $t \mapsto a(t)$ is persistently exciting. □

While Theorems 1 and 2 assume radial unboundedness of ϕ, it is possible to relax this assumption if one is interested in establishing only local uniform asymptotic stability results.

Remark 2: The necessity result established in Theorem 2 implies that UGAS of $\mathcal{A}$ cannot be achieved by Nesterov's ODE with $a(t) = c/t$, for any $c > 0$, thus providing an alternative and more general proof to the observation made in [13, Prop. 1]. □

Our last result pertains to the UGAS properties of system (3) for the case when $k > 0$. The following theorem is the third main result of this paper.

Theorem 3: Consider the gradient flow (3) with Hamiltonian given by (8), and matrices A and B given by (9). Let $k > 0$ and suppose that the following assumptions hold:

1) The function $\phi(\cdot)$ is C^2, radially unbounded, and convex.
2) The function $b : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is absolutely continuous, and there exists $\gamma_2, \gamma_3, \gamma_4 > 0$ such that
   - $a(t) \in [0, \gamma_2]$ and $b(t) \in [0, \gamma_3]$ for all $t \geq 0$.
   - $\dot{b}(t) \in [-\gamma_4, 0]$ for almost all $t \geq 0$.

Then, the set $\mathcal{A}$ is UGAS. □

An interesting conclusion of Theorem 3 is that whenever the term B in system (3) is "activated", the PE condition on $a(\cdot)$ is not needed any more in order to establish UGAS of $\mathcal{A}$. For example, when $\nabla\phi(\cdot)$ is L-globally Lipschitz, and one selects $k = 1/\sqrt{L}$, $b(t) = 1/t$, and $a(t) = 3b(t)$, the gradient flow (3) describes the dynamics

$$\ddot{z} + \frac{3}{t}\dot{z} + \nabla\phi(z) = -\frac{1}{\sqrt{L}}\left(\nabla^2\phi(z)\dot{z} + \frac{1}{t}\nabla\phi(z)\right), \quad (12)$$

which has been studied in [19], [20] as a better continuous-time approximation of Nesterov's algorithm (1). In this case, since $\dot{b} = -\frac{1}{t^2} < 0$, by fixing $t_0 > 0$, all our assumptions hold for $t \geq t_0$. If, for example, $\phi \in C^3$, by Lemma 2 the dynamics (12) are structurally robust.

IV. PROOFS

In this section, we present all the main proofs of the paper.

A. Proof of Lemma 4

For $p \in (0, 1)$, we note that $\lambda^p(t) \geq \lambda(t)/M^{1-p}$ where $|\lambda(t)| \leq M$ for all $t \geq 0$. Thus, it is immediate that λ^p is PE for $p \in (0, 1)$.

Now consider $p > 1$. Suppose the PE condition holds for λ with $T > 0$ and $\mu > 0$. Then, for each $t \geq 0$, there exists $s \in [t, t + T]$ such that $\lambda(s) \geq \mu/T$. Let L denote the Lipschitz constant of λ and define $m := \mu/(2LT)$. It follows that $\lambda(r) \geq \frac{\mu}{2T}$, for all $r \in [s - m, s + m] \cap [t, t + T]$ and thus

$$\lambda^p(r) \geq \left(\frac{\mu}{2T}\right)^p, \quad \forall\, r \in [s - m, s + m] \cap [t, t + T]. \quad (13)$$

In turn, since $s \in [t, t + T]$, it follows that

$$\int_t^{t+T} \lambda^p(\tau)d\tau \geq m\left(\frac{\mu}{2T}\right)^p = \left(\frac{\mu}{2T}\right)^{p+1}\frac{1}{L}. \quad (14)$$

Thus, λ^p is PE. ■

B. Proof of Theorem 1

The proof is based on Matrosov's theorem. We divide the proof in three steps. In each step, we find an auxiliary function V_i in order to obtain the bounds of Lemma 3.

Step 1: Consider the Lyapunov function candidate

$$V_1(x) = \frac{|x_2|^2}{2} + \phi(x_1) - \phi^*, \quad (15)$$

which is positive definite with respect to $\mathcal{A}$ and radially unbounded. Define $\xi_1(t, x_2) := \sqrt{a(t)}|x_2|$, and note that if $a(\cdot)$ is uniformly bounded, then so it is $\sqrt{a(\cdot)}$. Therefore, on the compact set $\mathcal{A} + \Delta\mathbb{B} \subset \mathbb{R}^{2n}$ we know there exists $r_1 > 0$ such that for all $t \geq 0$ and all $x \in \mathcal{A} + \Delta\mathbb{B}$ we have:

$$\max\{V_1(x), \xi_1(x_2, t)\} \leq r_1.$$

Taking the derivative of V_1 with respect to time, we obtain:

$$\begin{aligned}\dot{V}_1(t, x) &= -a(t)|x_2|^2 - x_2^\top\nabla\phi(x_1) + \nabla\phi(x_1)^\top x_2\\ &= -a(t)|x_2|^2 \leq 0, \end{aligned}\quad (16)$$

Inequality (16) establishes uniform global stability (UGS) of the set $\mathcal{A}$.

Step 2: We now consider additional auxiliary differentiable functions that are bounded and have bounded derivatives on compact sets. From Matrosov's theorem, we know that any terms in the derivative of subsequent auxiliary functions that vanish with $\xi_1(x_2, t)$ are not problematic, even if they are positive. Therefore, we can consider the auxiliary function

$$V_2(t, x) = a(t)^{\frac{3}{2}}\nabla\phi(x_1)^\top x_2,$$

and we define $\xi_2(t, x_1) := a(t)^{\frac{3}{4}}|\nabla\phi(x_1)|$. Since V_2 and ξ_2 are continuous functions, and $a(t)$ is nonnegative and upper bounded, there exists $r_2 > 0$ such that for almost all $t \geq 0$ and all $x \in \mathcal{A} + \Delta\mathbb{B}$ we have $\max\{|V_2(t, x)|, \xi_2(x_1, t)\} \leq r_2$. Taking the derivative of V_2 along the solutions of the system, we obtain:

$$\begin{aligned}\dot{V}_2(t, x) &= a(t)^{\frac{3}{2}}x_2^\top\nabla^2\phi(x_1)x_2 - a(t)^{\frac{5}{2}}\nabla\phi(x_1)^\top x_2\\ &\quad - a(t)^{\frac{3}{2}}|\nabla\phi(x_1)|^2 + 1.5a(t)^{\frac{1}{2}}\nabla\phi(x_1)^\top x_2\dot{a}(t),\\ &\leq -a(t)^{\frac{3}{2}}|\nabla\phi(x_1)|^2 + \rho_r\xi_1(x_2, t),\end{aligned}$$

where $\rho_r > 0$ is a bound that holds on compact sets, and where we used the boundedness of $a(t)$ and $\dot{a}(t)$. Thus, $\dot{V}_2$ can be upper bounded as $\dot{V}_2(t, x) \leq -\xi_2(x_1, t)^2 + \rho_r\xi_1(x_2, t)$. Thus, we know that any terms in the derivative of subsequent auxiliary functions that vanish with $\xi_2(x_1, t)$ are not problematic, even if they are positive.

Step 3: Consider the next auxiliary function

$$V_3(t, x) := -V_1(x)\int_t^\infty e^{(t-\tau)}a(\tau)^{\frac{3}{2}}d\tau,$$

which for any $T > 0$ satisfies $V_3(t,x) \leq -e^{-T}V_1(x)\int_t^{t+T} a(\tau)^{\frac{3}{2}}d\tau$. Since Lemma 4 implies that $t \mapsto a(t)^{\frac{3}{2}}$ is PE, condition (7) gives the inequality $-\int_t^{t+T} a(\tau)^{\frac{3}{2}}d\tau \leq -\mu$. Thus, the auxiliary function $V_3(t,x)$ satisfies the bound $V_3(t,x) \leq -\mu\exp(-T)V_1(x)$. Moreover, since $a(\cdot)$ is bounded by γ_1, and $\int_t^{\infty} e^{t-\tau}d\tau = 1$, we have that $V_1(x)\int_t^{\infty} e^{(t-\tau)}a(\tau)^{\frac{3}{2}}d\tau \leq V_1(x)\gamma^{1.5}$. Thus, $V_3(t,x) \geq -V_1(x)\gamma^{1.5}$. Therefore, there exists $r_3 > 0$ such that for almost all $t \geq 0$ and all $x \in \mathbb{R}^n$ we have $|V_3(t,x)| \leq r_3$. Taking the derivative of V_3 with respect to x_2 and x_1 we obtain:

$$\frac{\partial V_3}{\partial x_2} = -\left(\int_t^{\infty} e^{(t-\tau)}a(\tau)^{\frac{3}{2}}d\tau\right)x_2$$
$$\frac{\partial V_3}{\partial x_1} = -\left(\int_t^{\infty} e^{(t-\tau)}a(\tau)^{\frac{3}{2}}d\tau\right)\nabla\phi(x_1).$$

Thus,

$$\begin{aligned}\frac{\partial V_3}{\partial x_2}^{\top}\dot{x}_2 &= -\left(\int_t^{\infty} e^{(t-\tau)}a(\tau)^{\frac{3}{2}}d\tau\right)x_2^{\top}\dot{x}_2\\ &\leq a(t)|x_2|^2\rho + \left(\int_t^{\infty} e^{(t-\tau)}a(\tau)^{\frac{3}{2}}d\tau\right)x_2^{\top}\nabla\phi(x_1),\end{aligned}$$

where $\rho > 0$ is a bound for the integral that holds on compact sets. We also have:

$$\frac{\partial V_3}{\partial x_1}^{\top}\dot{x}_1 = -\left(\int_t^{\infty} e^{(t-\tau)}a(\tau)^{\frac{3}{2}}d\tau\right)\nabla\phi(x_1)^{\top}x_2,$$

and

$$\begin{aligned}\frac{\partial V_3}{\partial t} &= V_1(x)\left[e^{t-\tau}a(\tau)^{\frac{3}{2}}\Big|_{\tau=t} - \int_t^{\infty}\frac{\partial}{\partial t}\left[e^{t-\tau}a(\tau)^{\frac{3}{2}}\right]d\tau\right],\\ &= a(t)^{\frac{3}{2}}V_1(x) + V_3(t,x).\end{aligned}$$

Therefore, the derivative of V_3 satisfies

$$\begin{aligned}\dot{V}_3(t,x) \leq\ & V_3(t,x) + a(t)^{\frac{3}{2}}V_1(x) + a(t)|x_2|^2\rho\\ &+ \left(\int_t^{\infty} e^{(t-\tau)}a(\tau)^{\frac{3}{2}}d\tau\right)x_2^{\top}\nabla\phi(x_1)\\ &- \left(\int_t^{\infty} e^{(t-\tau)}a(\tau)^{\frac{3}{2}}d\tau\right)\nabla\phi(x_1)^{\top}x_2.\end{aligned}$$

Since the last two terms cancel each other, we obtain

$$\begin{aligned}\dot{V}_3(t,x) &= V_3(t,x) + a(t)^{\frac{3}{2}}V_1(x) + a(t)|x_2|^2\rho,\\ &\leq -\mu e^{-T}V_1(x) + V_1(x)a(t)^{\frac{3}{2}} + a(t)|x_2|^2\rho. \quad (17)\end{aligned}$$

Using the Polyak-Lojasiewicz inequality to bound the second term, we obtain:

$$\begin{aligned}\dot{V}_3(t,x) \leq\ & -\mu e^{-T}V_1(x) + \left(\frac{1}{2d}|\nabla\phi(x_1)|^2 + \frac{|x_2|^2}{2}\right)a(t)^{\frac{3}{2}}\\ &+ a(t)|x_2|^2\rho,\\ \leq\ & -\mu e^{-T}V_1(x_1) + \xi_2(x_1,t)\rho_r + \xi_1(x_2,t)\rho_r,\end{aligned} \quad (18)$$

where $\rho_r > 0$ is a bound that holds on compact sets. On the other hand, if ϕ is convex but does not satisfy the PL inequality, we can use convexity to bound $\dot{V}_3$ in (17) as

$$\begin{aligned}\dot{V}_3(t,x) \leq\ & -\mu e^{-T}V_1(x) + a(t)|x_2|^2\rho\\ &+ \left(|\nabla\phi(x_1)||x_1 - x^*| + \frac{|x_2|^2}{2}\right)a(t)^{\frac{3}{2}},\end{aligned}$$

for all $x^* \in \mathcal{A}_\phi$. Since $\mathcal{A}_\phi$ is compact, $\exists\ \tilde{\rho}_r > 0$ such that

$$\dot{V}_3(t,x) \leq -\mu e^{-T}V_1(x_1) + \xi_2(x_1,t)\tilde{\rho}_r + \xi_1(x_2,t)\rho_r. \quad (19)$$

Neglecting $\xi_1(t,x_1)$ and $\xi_2(t,x_2)$ in (18) or (19), for almost all $t \geq 0$ and all $x \in \mathbb{R}^{2n}$ we have $\dot{V}_3(t,x) \leq 0$, and by the definition of V_1, we have that the right-hand side of (19) is zero only at $\mathcal{A}$. Therefore, we can now apply Lemma 3 with the functions V_i, for $i \in \{1,2,3\}$, $\xi : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \to \mathbb{R}^2$, and functions Y_i given by, $Y_1(x,\xi(t,x)) := -\xi_1(t,x_2)^2$, $Y_2(x,\xi(t,x)) := -\xi_2(t,x_1)^2 + \rho_r\xi_1(t,x_2)$, $Y_3(x,\xi(t,x)) := -\mu e^{-T}V_1(x_1) + \xi_2(t,x_1)\rho_r + \xi_1(t,x_2)\rho_r$. This establishes the result. ■

C. Proof of Theorem 2

Suppose $\mathcal{A} \times \{0\}$ is UGAS. Consider the function V_1 defined on (15). This function satisfies $\dot{V}_1(t,x) = -a(t)|x_2|^2$. Since V_1 is continuous and satisfies $V_1(x) = 0$ for all $x \in \mathcal{A}$, there exists a function $\alpha_2 \in \mathcal{K}_\infty$ such that $V_1(x) \leq \alpha_2(|x|_{\mathcal{A}})$. Moreover, by UGAS, there exists a function $\beta \in \mathcal{KL}$ such that $|x(t)|_{\mathcal{A}} \leq \beta(|x_0|_{\mathcal{A}}, t-t_0)$, for all $t \geq t_0 \geq 0$. Therefore,

$$V_1(x(t)) \leq \alpha_2(\beta(|x(0)|_{\mathcal{A}}, t-t_0)) := \tilde{\beta}(|x(0)|_{\mathcal{A}}, t-t_0),$$

where $\tilde{\beta} \in \mathcal{KL}$. Suppose by contradiction that $a(\cdot)$ is not PE. Then, for all $T, \mu > 0\ \exists\ t^* > 0$ such that

$$\int_{t^*}^{t^*+T} |a(\tau)|d\tau < \mu. \quad (20)$$

Let $x_0 \notin \mathcal{A}$ and consider a solution of (3) starting at (t^*, x_0), denoted by $x(t,t^*,x_0)$. Pick T sufficiently large, such that $\tilde{\beta}(|x_0|_{\mathcal{A}}, T) \leq 0.5V_1(x_0)$. Such T always exist because $\tilde{\beta} \in \mathcal{KL}$. Define μ such that $\mu \leq -\frac{\log(0.5)}{2}$, which satisfies $\mu > 0$. Using $\dot{V}_1$ and the definition of V_1 we get $\dot{V}_1(t,x) \geq -a(t)V_1(x)$. Using the Comparison Lemma from t^* to t we get $V_1(x(t,t^*,x_0)) \geq V_1(x_0)\exp\left(-2\int_{t^*}^{t} a(\tau)d\tau\right)$. Substituting t by $t^* + T$:

$$V_1(x(t^*+T), t^*, x_0) \geq V_1(x_0)\exp\left(-2\int_{t^*}^{t^*+T} a(\tau)d\tau\right).$$

Using the bound (20) and the monotonicity of the exponential function we obtain $V_1(x(t^*+T, t^*, x_0)) > V_1(x_0)\exp(-2\mu)$, and by the selection of μ we obtain

$$V_1(x(t^*+T, t^*, x_0)) > \frac{V_1(x_0)}{2}. \quad (21)$$

On the other hand, by the selection of T, we have that

$$V_1(x(t^*+T, t^*, x_0)) \leq \tilde{\beta}(|x_0|_{\mathcal{A}}, T) \leq \frac{V_1(x_0)}{2},$$

which contradicts (21). Thus, $t \mapsto a(t)$ must be PE. ■

D. Proof of Theorem 3

Consider the Lyapunov function candidate $V_1(t,x) := \frac{|x_2|^2}{2} + (1+kb(t))\,(\phi(x_1)-\phi^*)$, which satisfies $\dot{V}_1(t,x) = -a(t)|x_2|^2 - kx_2^\top \nabla^2\phi(x_1)x_2 + k\dot{b}(t)(\phi(x_1)-\phi^*) \leq 0$. Therefore, the set $\mathcal{A}$ is UGS. Since ϕ is convex we can consider the following two possible cases:

(a) Suppose first that there exists $\varepsilon > 0$ such that $\nabla^2\phi(x_1) \geq \varepsilon I$, e.g., ϕ is strongly convex. In this case, we can define $\xi_1(t,x_2) = \sqrt{a(t)}|x_2|$ to bound $\dot{V}_1$ as follows

$$\dot{V}_1(t,x) \leq -k\varepsilon|x_2|^2 - \xi_1(t,x_2)^2 =: Y_1(x,\xi(t,x)). \quad (22)$$

Consider the auxiliary function $V_2(x) := \nabla\phi(x_1)^\top x_2 + 0.5k|\nabla\phi(x_1)|^2$, which satisfies

$$\begin{aligned}\dot{V}_2(t,x) &\leq -|\nabla\phi(x_1)|^2 + \gamma^{0.5}\rho_r\xi_1(t,x_2) + \rho_r|x_2|^2\\ &=: Y_2(x,\xi(t,x)),\end{aligned}$$

where ρ_r is a bound for the gradient and the Hessian of ϕ that holds on compact sets. Since $Y_1(x,\xi)=0$ and $Y_2(x,\xi)=0$ imply $x \in \mathcal{A}$, we obtain UGAS of $\mathcal{A}$ by Lemma 3.

(b) Suppose now that $\nabla^2\phi(x_1) \succeq 0$. Then, $\dot{V}_1(t,x) \leq -\xi_1(t,x_2)^2 - kx_2^\top\nabla^2\phi(x_1)x_2 =: Y_1(x,\xi(t,x))$, and

$$\begin{aligned}\dot{V}_2(t,x) &\leq -|\nabla\phi(x_1)|^2 + \gamma^{0.5}\xi_1(t,x_2) + x_2^\top\nabla^2\phi(x_1)x_2\\ &=: Y_2(x,\xi(t,x)).\end{aligned}$$

Let $\xi_2(t,x_1) = |\nabla\phi(x_1)|$ and consider the auxiliary function $V_3(x) := -x_1^\top x_2$, which satisfies

$$\begin{aligned}\dot{V}_3(t,x) &\leq -|x_2|^2 + k|x_1||\nabla^2\phi(x_1)x_2| + \gamma^{0.5}\rho_r\xi_1(t,x)\\ &\quad + (1+k\gamma_3)\rho_r\xi_2(t,x) =: Y_3(x,\xi(t,x)),\end{aligned}$$

where ρ_r is a bound on compact sets for x_1. This establishes that the conditions $Y_1(x,\xi)=0$, $Y_2(x,\xi)=0$, $Y_3(x,\xi)=0$ imply that $x \in \mathcal{A}$ because $x_2^\top\nabla^2\phi(x_1)x_2 = 0$ implies $\nabla^2\phi(x_1)x_2 = 0$. By Lemma 3 the set $\mathcal{A}$ is UGAS. ■

V. CONCLUSIONS

This paper studied the UGAS properties of a class of time-varying gradient flows that have recently emerged in the literature of accelerated optimization algorithms. When the gradient flow is degenerate, we established persistence of excitation conditions that guarantee UGAS of the set of minimizers for a general class of cost functions. Moreover, we showed that the persistence of excitation condition is indeed a necessary condition in order to achieve UGAS of compact attractors with radially unbounded cost functions. For the case when the degenerate gradient flow is conditioned by an additional Hessian-dependent term, persistence of excitation turns out to be not necessary for UGAS.

REFERENCES

[1] B. V. Scoy, R. A. Freeman, and K. M. Lynch, "The fastest known globally convergent first-order method for minimizing strongly convex functions," *IEEE Control Systems Letters*, vol. 2, no. 1, pp. 49–54, 2018.

[2] L. Lessard, B. Recht, and A. Packard, "Analysis and design of optimization algorithms via integral quadratic constraints," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 57–95, 2016.

[3] C. J. Maddison, D. Paulin, Y. W. Teh, B. O'Donoghue, and A. Doucet, "Hamiltonian descent methods," *arXiv:1809.05042*, 2018.

[4] W. Su, S. Boyd, and E. Candes, "A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights," *Journal of Machine Learning Research*, vol. 17, no. 153, pp. 1–43, 2016.

[5] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic course*. Boston, MA.: Kluwer Academic Publishers, 2004.

[6] A. Wibisono, A. C. Wilson, and M. I. Jordan, "A variational perspective on accelerated methods in optimization," *Proceedings of the National Academy of Sciences*, vol. 113, no. 47, pp. E7351–E7358, 2016.

[7] H. Attouch and F. Alvarez, *The heavy ball with friction dynamical system for convex constrained minimization problems*. Nguyen V.H., Strodiot JJ., Tossings P. (eds) Optimization. Lecture Notes in Economics and Mathematical Systems, 2000, vol. 481.

[8] A. Cabot, H. Engler, and S. Gadat, "On the long time behavior of second order differential equations with asymptotically small dissipation," *Transactions of the American Mathematical Society*, vol. 361, no. 11, pp. 5983–6017, 2009.

[9] H. Attouch, Z. Chbani, and H. Riahi, "Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$." *ESAIM: COCV*, vol. 25, no. 2, 2019.

[10] H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont, "Fast convergence of inertial dynamics and algorithms with asymptotic vanishing damping," *Math. Program*, vol. 168, no. 1, pp. 123–175, 2018.

[11] O. Sebbouh, C. Dossal, and A. Rondepierre, "Nesterov's acceleration and Polyak's heavy ball method in continuous time: convergence rate analysis under geometric conditions and perturbations," *hal-02173978*, 2019.

[12] D. M. Hustig-Schultz and R. G. Sanfelcie, "A robust hybrid heavy ball algorithm for optimization with high performance," *Proc. of American Contr. Conf.*, pp. 151–156, 2019.

[13] J. I. Poveda and N. Li, "Inducing uniform asymptotic stability in nonautonomous accelerated optimization dynamics via hybrid regularization," *58th IEEE Conference on Decision and Control*, pp. 3000–3005, 2019.

[14] D. Ochoa, J. I. Poveda, C. Uribe, and N. Quijano, "Robust resource allocation with momentum," *IEEE Conference on Decision and Control*, pp. 3954–3959, 2019.

[15] A. S. Kolarijani, P. M. Esfahani, and T. Keviczky, "Continuous-time accelerated methods via a hybrid control lens," *IEEE Transactions on Automatic and Control, DOI 10.1109/TAC.2019.2945287*, 2018.

[16] A. R. Teel, J. I. Poveda, and J. Le, "First-order optimization algorithms with resets and hamiltonian flows," *58th IEEE Conference on Decision and Control*, 2019.

[17] J. I. Poveda, M. Benosman, and A. R. Teel, "Hybrid online learning control in networked multi-agent systems: A survey," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 2, pp. 228–261, 2019.

[18] R. Brockett, "The rate of descent for degenerate gradient flows," *Proceedings of the 2000 MTNS, Perpignan, France.*, 2000.

[19] B. Shi, S. S. Du, M. I. Jordan, and W. J. Su, "Understanding the acceleration phenomenon via high-resolution differential equations," *arXiv preprint arXiv:1810.08907*, 2018.

[20] M. Laborde and A. Oberman, "A Lyapunov analysis for accelerated gradient methods: From deterministic to stochastic case," *arXiv:1908.07861*, 2019.

[21] Z. Artstein, "Uniform asymptotic stability via the limiting equations," *Journal of Differential Equations*, vol. 27, pp. 172–189, 1978.

[22] H. K. Khalil, *Nonlinear Systems*. Upper Saddle River, NJ: Prentice Hall, 2002.

[23] A. Loria, E. Panteley, D. Popović, and A. R. Teel, "A nested Matrosov theorem and persistency of excitation for uniform convergence in stable nonautonomous systems," *IEEE Transactions on Automatica Control*, vol. 50, no. 2, pp. 183–198, 2005.

[24] R. G. Sanfelice and A. R. Teel, "Asymptotic stability in hybrid systems via nested Matrosov functions," *IEEE Transactions on Automatic and Control*, vol. 54, no. 7, pp. 1569–1574, 2009.

[25] A. Ben-Israel and B. Mond, "What is invexity?" *J. Australia. Math. Soc. Ser. B*, pp. 1–9, 1986.

[26] H. Karimi, J. Nutini, and M. Schmidt, "Linear convergence of gradient and proximal-gradient methods under the Polyak-Lojasiewicz condition," *arXiv:1608.04636*, 2018.

[27] M. Malisoff and F. Mazenc, *Constructions of Strict Lyapunov Functions*. Springer, 2009.