Uniting Nesterov and Heavy Ball Methods for Uniform Global Asymptotic Stability of the Set of Minimizers ⋆

Dawn M. Hustig-Schultz [a], Ricardo G. Sanfelice [a]

[a] *Department of Electrical and Computer Engineering, University of California, Santa Cruz.*

Abstract

We propose a hybrid control algorithm that guarantees fast convergence and uniform global asymptotic stability of the unique minimizer of a $\mathcal{C}^1$, convex objective function. The algorithm, developed using hybrid system tools, employs a uniting control strategy, in which Nesterov's accelerated gradient descent is used "globally" and the heavy ball method is used "locally," relative to the minimizer. Without knowledge of its location, the proposed hybrid control strategy switches between these accelerated methods to ensure convergence to the minimizer without oscillations, with a (hybrid) convergence rate that preserves the convergence rates of the individual optimization algorithms. We analyze key properties of the resulting closed-loop system including existence of solutions, uniform global asymptotic stability, and convergence rate. Additionally, stability properties of Nesterov's method are analyzed, and extensions on convergence rate results in the existing literature are presented. Numerical results validate the findings.

Key words: Hybrid systems; dynamical systems; optimization; asymptotic stability; convergence rate.

1 Introduction

1.1 Background and Motivation

We propose an algorithm that solves optimization problems of the form $\min_{\xi\in\mathbb{R}^n} L(\xi)$ with accelerated gradient methods. The *heavy ball method* is an accelerated gradient method that guarantees convergence to the minimizer ξ^* of a convex function L [1], and that achieves a faster convergence rate than classical gradient descent by adding a "velocity" term to ∇L. The dynamical system characterization for this method is

$$\ddot{\xi} + \lambda\dot{\xi} + \gamma\nabla L(\xi) = 0 \tag{1}$$

where λ and γ are positive tunable parameters that represent friction and gravity, respectively; see [2], [3]. In [4] and [5] it is shown that the discrete-time version of the heavy ball method converges exponentially when L is strongly convex with a Lipschitz continuous gradient, and [4] shows convergence with rate $\frac{1}{k}$ when L is convex. It is shown in [6] that for strongly convex L with Lipschitz continuous ∇L global convergence of the discrete-time heavy ball method can only be guaranteed for condition numbers of about 18 or less, and it is found in [7] that the exact condition number of $9+5\sqrt{14} \approx 17.94$ denotes such a boundary between global convergence and non-convergence, for such objective functions. When L is strongly convex, and inspired by the heavy ball algorithm, two algorithms with a resettable velocity term are proposed in [8] and shown to guarantee exponential convergence. In [9], however, it was shown that the heavy ball algorithm converges exponentially for convex L when L also has the property of quadratic growth away from ξ^*. Global asymptotic stability of ξ^*, which is the property that all solutions that start close to ξ^* stay close, and solutions from all initial conditions converge to ξ^*, is demonstrated in [10], for convex and smooth L.

Another powerful accelerated method is *Nesterov's accelerated gradient descent*. One characterization of the dynamical system for Nesterov's method, for convex L, proposed in [11], is

⋆ This paper was not presented at any IFAC meeting. Corresponding author D. M. Hustig-Schultz.

This research has been partially supported by the National Science Foundation under Grant no. ECS-1710621, Grant no. CNS-2039054, and Grant no. CNS-2111688, by the Air Force Office of Scientific Research under Grant nos. FA9550-19-1-0169, FA9550-20-1-0238, and FA9550-23-1-0145, by the Air Force Research Laboratory under Grant nos. FA8651-22-1-0017 and FA8651-23-1-0004, and by the Army Research Office under Grant no. W911NF-20-1-0253.

Email addresses: dhustigs@ucsc.edu (Dawn M. Hustig-Schultz), ricardo@ucsc.edu (Ricardo G. Sanfelice).

$$\ddot{\xi} + 2\bar{d}(t)\dot{\xi} + \frac{1}{M\zeta^2}\nabla L(\xi + \bar{\beta}(t)\dot{\xi}) = 0, \qquad (2)$$

where $M > 0$ is the Lipschitz constant of ∇L and where the constant $\zeta > 0$ rescales time in solutions to (2). The dynamical system in (2) resembles the model of a mass-spring-damper, with a curvature-dependent damping term where the total damping is a linear combination of $\bar{d}(t)$ and $\bar{\beta}(t)$. In [11], the convergence rate of Nesterov's method is characterized as $\frac{1}{(t+2)^2}$ for (2) (for $t \geq 1$), when $\zeta = 1$, $\xi^* = 0$, and $L(\xi^*) = 0$. The stability properties of (2) are not revealed in [11], however.

While the results in [5], [6], and [11] characterize the convergence properties of Nesterov's method (or a variation of) the stability properties of the method are not revealed. A particularly useful property for optimization algorithms, called *uniform global asymptotic stability* (UGAS), requires that solutions reach a neigborhood of ξ^* in time that is uniform on the set of initial conditions [12], [13], [14]. After finite time, the error of such solutions becomes smaller than a given threshold [15]. Due to such a guarantee, UGAS is typically useful for certifying robustness to small perturbations in time-varying dynamical and hybrid systems [16], [15]. Remarkably, the algorithm with resets in the velocity term proposed in [8] can be shown to induce UGAS of ξ^* (with zero velocity term) and reduced oscillations, for the particular case when L is strongly convex. The algorithm with resets in [14] can be shown to induce UGAS of ξ^* when L is invex, has an exponential convergence rate when L satisfies the Polyak-Łojasiewicz inequality, and uniform global exponential stability (UGES) when L is strongly convex. Unfortunately, as shown in [12], via a counterexample, Nesterov-like algorithms do not necessarily assure UGAS of ξ^* when L is convex. In response to this, [12] proposes the HAND-1 and HAND-2 reset algorithms, and prove UGAS of ξ^* for both algorithms. The exponential convergence rate of HAND-2, however, only applies to strongly convex L, and the convergence rate of $\frac{1}{t^2}$ for HAND-1, for convex L, only holds up until the first reset.

The work in this paper is motivated by the lack of an accelerated gradient algorithm assuring UGAS, with a convergence rate that holds for all time and that resembles that of Nesterov's method (at least far from ξ^*), when L is convex. However, attaining such a rate is expected to lead to oscillations, which are typically seen in accelerated gradient methods. The performance of the heavy ball method depends highly on the choice of λ. For rather simple choices of L, large values of λ give rise to slowly converging solutions [2]. The top plot [1] in Fig. 1 demonstrates such behavior. In contrast, smaller values of λ give rise to fast solutions with oscillations getting wilder as λ decreases [2]. Nesterov's method converges

[1] Code at
`gitHub.com/HybridSystemsLab/UnitingMotivation`

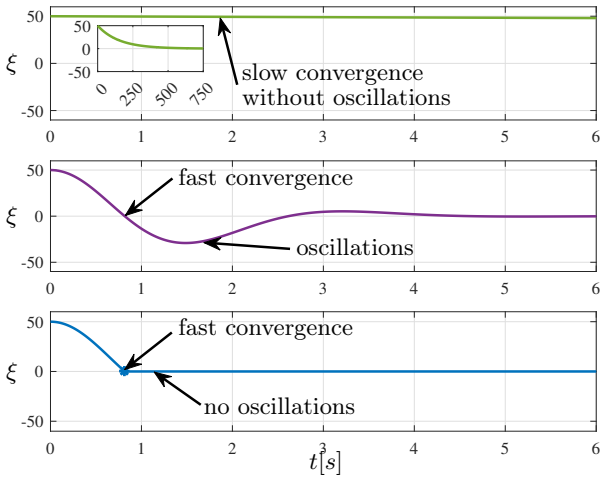

Fig. 1. Comparison of the performance of heavy ball, with large λ, Nesterov's method, and the proposed logic-based algorithm. The objective function is $L(\xi) = \xi^2$. Top: the heavy ball algorithm, with large λ, converges very slowly. Top inset: zoomed out view of heavy ball. Second from top: Nesterov's method converges quickly, but with oscillations. Third from top: our proposed logic-based algorithm yields fast convergence, with no oscillations.

quickly but also suffers from oscillations [11], as the coefficient of the velocity term starts small and tends toward zero (but being always positive) as t tends to infinity. Such behavior of Nesterov's method, with $\zeta = 2$, is shown in the second plot from the top in Fig. 1.

Due to its implications on robustness, we are particularly interested in an algorithm that assures UGAS of ξ^* with a rate of convergence that holds for all time, and without oscillations. As pointed out in Section 1.1, these properties are not guaranteed by Nesterov's method. The behavior shown in the first and second plots in Fig. 1 motivates the logic-based algorithm proposed in this paper. The proposed algorithm exploits the main features of heavy ball and Nesterov's method to achieve fast convergence and UGAS of ξ^*. More precisely, without knowledge of the location of ξ^*, it selects Nesterov's method to converge quickly to nearby ξ^* and, once solutions reach a neighborhood of ξ^*, switches to the heavy ball method with large λ to avoid oscillations. Such logic-based algorithms, or *uniting algorithms*, were first proposed in [17] and [18]. General uniting algorithms, with examples, are discussed in [16] and [15]. We use the hybrid systems framework for our proposed algorithm, as hybrid systems utilize hysteresis to avoid chattering at the switching boundary; see [8], [15], [16], [14]. An example solution to our proposed algorithm, shown in the third plot from the top in Fig. 1, demonstrates the improvement obtained, under relatively mild assumptions on L. The proposed algorithm guarantees UGAS and a (hybrid) convergence rate that holds for all $t \geq 0$.

1.2 Contributions

The main contributions of this paper are as follows.

1) *A uniting algorithm for fast convergence and UGAS of ξ^*.*
2) *Well-posedness, existence of solutions, and robustness to small perturbations in measurements of ∇L.* Nesterov's method can suffer from error accumulation, due to its velocity term [19]. To overcome this issue, in Section 2 we prove well-posedness and existence of solutions for the proposed hybrid closed-loop algorithm. Due to such well-posedness, the established UGAS property is robust to small perturbations in measurements of ∇L [16, Theorem 7.21].

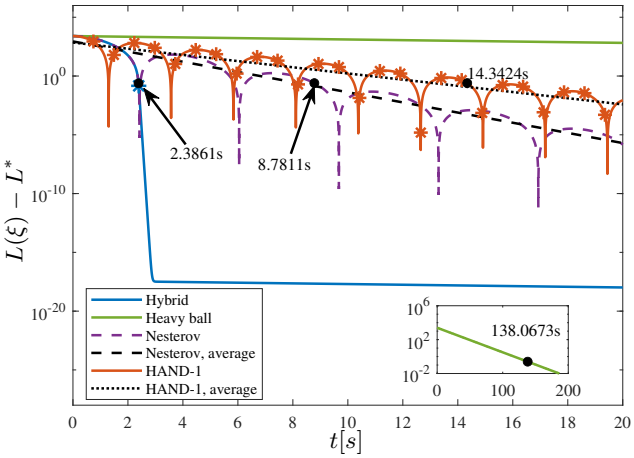

Fig. 2. A comparison of the evolution of L over time for Nesterov's method in (3), heavy ball, HAND-1 from [12], and our proposed uniting algorithm, for a function $L(\xi) := \xi^2$, with a single minimizer at $\xi^* = 0$. As opposed to Fig. 1, which uses $\zeta = 2$ for $\mathcal{H}_1$, this example uses $\zeta = 1$, which results in slower convergence of solutions to $\mathcal{H}$ and $\mathcal{H}_1$.

3) *A (hybrid) convergence rate preserving the rates of Nesterov's method and heavy ball.* Numerical simulations [2] in Fig. 2 and Section 3 show improved performance of our uniting algorithm over HAND-1.
4) *Extension of the results on Nesterov's method in [11].* We achieve such an extension by moving ζ into the numerator of the coefficient of ∇L, effectively decoupling ζ and M, namely,

$$\ddot{\xi} + 2\bar{d}(t)\dot{\xi} + \frac{\zeta^2}{M}\nabla L(\xi + \bar{\beta}(t)\dot{\xi}) = 0. \qquad (3)$$

While preliminary work in [20] proposed an algorithm uniting Nesterov's method globally and heavy ball locally for $\mathcal{C}^2$, strongly convex L, with different results and examples, the uniting algorithm proposed in this paper relaxes the conditions in [20] to $\mathcal{C}^1$, convex L with a unique minimizer. A technical report version of this paper, with more details [21], is available online.

[2] Code at `gitHub.com/HybridSystemsLab/UnitingTradeoff`

1.3 Notation

The sets of real, positive real, and natural numbers are denoted by $\mathbb{R}$, $\mathbb{R}_{>0}$, and $\mathbb{N}$, respectively. The closed unit ball, of appropriate dimension, in the Euclidean norm is denoted as $\mathbb{B}$. The set $\mathcal{C}^n$ represents the family of n-th continuously differentiable functions. For a vector $v \in \mathbb{R}^n$, $|v| = \sqrt{v^\top v}$ denotes the Euclidean vector norm of v. For any $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, $(x, y) := [x^\top, y^\top]^\top$. The closure of a set S is denoted $\overline{S}$. Given a set $S \subset \mathbb{R}^n \times \mathbb{R}^m$, the projection of S onto $\mathbb{R}^n$ is defined as $\Pi(S) := \{x \in \mathbb{R}^n : \exists y \text{ such that } (x, y) \in S\}$. Given a set-valued mapping $M : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$, the domain of M is the set $\mathrm{dom} M = \{x \in \mathbb{R}^m : M(x) \neq \emptyset\}$.

2 Uniting Optimization Algorithm

2.1 Problem Statement

As illustrated in Fig. 1, the performance of Nesterov's accelerated gradient descent commonly suffers from oscillations near the minimizer. This is also the case for the heavy ball method when $\lambda > 0$ is small. However, when λ is large, the heavy ball method converges slowly, albeit without oscillations. In Section 1 we discussed how Nesterov's algorithm guarantees a rate of $\frac{1}{(t+2)^2}$ for convex L. We also discussed how the heavy ball algorithm guarantees a rate of $\frac{1}{t}$ for convex L, although it was demonstrated in [9] that the heavy ball algorithm converges exponentially for convex L when such an objective function also has the property of quadratic growth away from its minimizer. We desire to attain the rate $\frac{1}{(t+2)^2}$ globally and an exponential rate locally, while avoiding oscillations via the heavy ball algorithm with large λ. We state the problem to solve as follows:

Problem 1 *Given a scalar, real-valued, continuously differentiable, and convex objective function L with a unique minimizer, design an optimization algorithm that, without knowing the function L or the location of its minimizer, has the minimizer UGAS, with a convergence rate of $\frac{1}{(t+2)^2}$ globally and an exponential convergence rate locally, and with robustness to arbitrarily small noise in measurements of ∇L.*

2.2 Modeling

In this section, we present an algorithm that solves Problem 1. We interpret the ODEs in (1) and (3) as control systems consisting of a plant and a control algorithm [22] [15]. Defining z_1 as ξ and z_2 as $\dot{\xi}$, the plant associated to these ODEs is given by the double integrator

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} z_2 \\ u \end{bmatrix} =: F_P(z, u) \quad (z, u) \in \mathbb{R}^{2n} \times \mathbb{R}^n =: C_P \quad (4)$$

With this model, the optimization algorithms that we consider assign u to a function of the state that involves the cost function, and such a function of the state may be time dependent. The control algorithm leading to (1) assigns u to $-\lambda z_2 - \gamma \nabla L(z_1)$ where $\gamma > 0$ and $\lambda > 0$, and the control algorithm leading to (3) assigns u to $-2\bar{d}(t)z_2 - \frac{\zeta^2}{M}\nabla L(z_1 + \bar{\beta}(t)z_2)$ where $\zeta > 0$, $M > 0$ is the Lipschitz constant for ∇L, and where $\bar{d}(t)$ and $\bar{\beta}(t)$ are defined, for all $t \geq 0$, as

$$\bar{d}(t) := \frac{3}{2(t+2)}, \quad \bar{\beta}(t) := \frac{t-1}{t+2}. \tag{5}$$

The functions $\bar{d}$ and $\bar{\beta}$ are defined expressly as in (5) for ease of analysis in the forthcoming Propositions 13-15. Such a time-varying definition satisfies the linear combination of the damping terms mentioned below (2). While constant terms can be used for (2), when L is strongly convex, constant damping terms are not adequate for convex L; see [11]. The proposed logic-based algorithm "unites" the two optimization algorithms modeled by κ_q, where the logic variable $q \in Q := \{0, 1\}$ indicates which algorithm is currently being used. The local and global algorithms, respectively, are defined as

$$\kappa_0(h_0(z)) = -\lambda z_2 - \gamma \nabla L(z_1) \tag{6a}$$

$$\kappa_1(h_1(z,t),t) = -2\bar{d}(t)z_2 - \frac{\zeta^2}{M}\nabla L(z_1 + \bar{\beta}(t)z_2) \tag{6b}$$

where the algorithm defined by κ_1 plays the role of the global algorithm in uniting control (see, e.g., [15, Chapter 4]), while the algorithm defined by κ_0 plays the role of the local algorithm. The outputs h_0 corresponding to the output for the heavy ball algorithm and h_1 corresponding to the output for Nesterov's algorithm are defined as

$$h_0(z) := \begin{bmatrix} z_2 \\ \nabla L(z_1) \end{bmatrix}, h_1(z,t) := \begin{bmatrix} z_2 \\ \nabla L(z_1 + \bar{\beta}(t)z_2) \end{bmatrix}. \tag{7}$$

Namely, the algorithm exploits measurements of ∇L, which in practice are typically approximated using measurements of L. The parameters $\lambda > 0$ and $\gamma > 0$ should be designed to achieve convergence without oscillations nearby the minimizer.

We use the hybrid systems framework to design our algorithm. A hybrid system $\mathcal{H}$ has data (C, F, D, G) and is defined as [16, Definition 2.2]

$$\mathcal{H} = \begin{cases} \dot{x} \in F(x) & x \in C \\ x^+ \in G(x) & x \in D \end{cases} \tag{8}$$

where $x \in \mathbb{R}^n$ is the system state, $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is the flow map, $C \subset \mathbb{R}^n$ is the flow set, $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is the jump map, and $D \subset \mathbb{R}^n$ is the jump set. Since the ODE in (3) is time varying, and since solutions to hybrid systems are parameterized by [3] $(t, j) \in \mathbb{R}_{\geq 0} \times \mathbb{N}$, we employ the state τ to capture ordinary time as a state variable, in this way, leading to a time-invariant hybrid system. To encapsulate the plant, static state-feedback laws, and the time-varying nature of the ODE in (3), we define a hybrid closed-loop system $\mathcal{H}$ with state $x := (z, q, \tau) \in \mathbb{R}^{2n} \times Q \times \mathbb{R}_{\geq 0}$ as

$$\left.\begin{aligned} \dot{z} &= \begin{bmatrix} z_2 \\ \kappa_q(h_q(z,\tau),\tau) \end{bmatrix} \\ \dot{q} &= 0 \\ \dot{\tau} &= q \end{aligned}\right\} =: F(x) \qquad x \in C := C_0 \cup C_1 \tag{9a}$$

$$\left.\begin{aligned} z^+ &= \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \\ q^+ &= 1 - q \\ \tau^+ &= 0 \end{aligned}\right\} =: G(x) \qquad x \in D := D_0 \cup D_1 \tag{9b}$$

The sets C_0, C_1, D_0, and D_1 are defined as

$$C_0 := \mathcal{U}_0 \times \{0\} \times \{0\}, \; C_1 := \overline{\mathbb{R}^{2n} \setminus \mathcal{T}_{1,0}} \times \{1\} \times \mathbb{R}_{\geq 0} \tag{10a}$$

$$D_0 := \mathcal{T}_{0,1} \times \{0\} \times \{0\}, \; D_1 := \mathcal{T}_{1,0} \times \{1\} \times \mathbb{R}_{\geq 0}. \tag{10b}$$

The sets $\mathcal{U}_0$, $\mathcal{T}_{1,0}$, and $\mathcal{T}_{0,1}$ are precisely defined in Section 2.3, using Lyapunov functions defined therein, but the idea behind their construction is as follows. The switch between κ_0 and κ_1 is governed by a *supervisory algorithm* implementing switching logic. The supervisor selects between these two optimization algorithms, based on the output of the plant in (7) and the optimization algorithm currently applied. When $z \in \mathcal{U}_0$, $q = 0$, and $\tau = 0$ (i.e., $x \in C_0$), due to the design of $\mathcal{U}_0$ in Section 2.3.1, then the state z is near the minimizer, which is denoted z_1^*, and the supervisor allows flows of (9) using κ_0 and $\dot{\tau} = q = 0$ to avoid oscillations. Conversely, when $z \in \overline{\mathbb{R}^{2n} \setminus \mathcal{T}_{1,0}}$ and $q = 1$ (i.e., $x \in C_1$), due to the design of $\mathcal{T}_{1,0}$ in Section 2.3.2, then the state z is far from the minimizer and the supervisor allows flows of (9) using κ_1 and $\dot{\tau} = q = 1$ to converge quickly to the neighborhood of the minimizer. When $z \in \mathcal{T}_{1,0}$ and $q = 1$ (i.e., $x \in D_1$), then this indicates that the state z is near the minimizer, and the supervisor assigns u to κ_0, resets q to 0, and resets τ to 0. Conversely, when $z \in \mathcal{T}_{0,1}$, $q = 0$, and $\tau = 0$ (i.e., $x \in D_0$), due to the design of $\mathcal{T}_{0,1}$ in Section 2.3.3, then this indicates that the state z is far from the minimizer and the supervisor assigns u to κ_1 and resets q to 1.

The reason that the state τ in (9) changes at the rate q during flows and is reset to 0 at jumps is that when the

[3] The variable t is the amount of time that has passed and j is the number of jumps that have occurred.

state x is in C_1, then $\dot{\tau} = q = 1$, which implies that τ behaves as ordinary time, so it is used to represent time in the time-varying algorithm κ_1. On the other hand, when the state x is in C_0, then $\dot{\tau} = q = 0$ causes the state τ to stay at zero, which is an appropriate value for τ as it is not required by the time-invariant algorithm κ_0. Such an evolution ensures that the set to asymptotically stabilize is compact.

We denote the closed-loop system resulting from κ_0 as $\mathcal{H}_0$, which is given by

$$\dot{z} = \begin{bmatrix} z_2 \\ \kappa_0(h_0(z)) \end{bmatrix} \qquad z \in \mathbb{R}^{2n} \tag{11}$$

and we denote the closed-loop system resulting from κ_1 as $\mathcal{H}_1$, which is given by

$$\dot{z} = \begin{bmatrix} z_2 \\ \kappa_1(h_1(z,\tau),\tau) \end{bmatrix},\ \dot{\tau} = 1 \qquad (z,\tau) \in \mathbb{R}^{2n} \times \mathbb{R}_{\geq 0}. \tag{12}$$

2.3 Design of the Hybrid Algorithm

In order for the supervisor to determine when the state component z_1 is close to the minimizer of L, denoted z_1^*, without knowledge of z_1^* or $L^* := L(z_1)$, we impose the following assumptions on L.

Assumption 2 *The function L is $\mathcal{C}^1$, convex[4], and has a single minimizer z_1^*.*

Assumption 3 (Quadratic growth of L) *The function L has quadratic growth away from its minimizer z_1^*; i.e., there exists $\alpha > 0$ such that $L(z_1) - L^* \geq \alpha |z_1 - z_1^*|^2$ for all $z_1 \in \mathbb{R}^n$.*

To make the switch back to κ_1, we impose the following assumption on L.

Assumption 4 (Lipschitz Continuity of ∇L) *The function ∇L is Lipschitz continuous with constant $M > 0$, namely, $|\nabla L(w_1) - \nabla L(u_1)| \leq M |w_1 - u_1|$ for all $w_1, u_1 \in \mathbb{R}^n$.*

Under Assumptions 2 and 3, the following lemma, used in some of the results to follow, relates the size of the gradient at a point to the distance from the point to z_1^*. Its proof is in [21].

Lemma 5 (Suboptimality) *Let L satisfy Assumptions 2 and 3, and let $\alpha > 0$ come from Assumption 3. For some $\varepsilon > 0$, if $z_1 \in \mathbb{R}^n$ is such that $|\nabla L(z_1)| \leq \varepsilon\alpha$, then $|z_1 - z_1^*| \leq \varepsilon$.*

[4] A function $L : \mathbb{R}^n \to \mathbb{R}$ is convex if, for all $u_1, w_1 \in \mathbb{R}^n$, $L(u_1) \geq L(w_1) + \langle \nabla L(w_1), u_1 - w_1 \rangle$ [23].

The suboptimality condition from Lemma 5 is typically used as a stopping condition for optimization, as it indicates that the argument of L is close enough to the minimizer z_1^* [23]. We exploit Lemma 5 to determine when the state component z_1 of the hybrid closed-loop system $\mathcal{H}$ is close enough to the minimizer z_1^* so as to switch to the local optimization algorithm, κ_0, in this way activating $\mathcal{H}_0$.

2.3.1 Design of the Set $\mathcal{U}_0$

The objective is to design $\mathcal{U}_0$ such that when $z \in \mathcal{U}_0$, $q = 0$, and $\tau = 0$, the state component z_1 is near z_1^* and the uniting algorithm allows flows of (9) with κ_0 and $q = 0$. For such a design, we use Assumptions 2 and 3 and the Lyapunov function

$$V_0(z) := \gamma \left(L(z_1) - L^* \right) + \frac{1}{2} |z_2|^2 \tag{13}$$

defined for each $z \in \mathbb{R}^{2n}$, where $\gamma > 0$. The choice of V_0 in (13) is used in the proof of the forthcoming Proposition 10 to establish UGAS of the minimizer for $\mathcal{H}_0$ in (11). Given $\varepsilon_0 > 0$, $c_0 > 0$, and $\gamma > 0$ from κ_0 in (6a), let $\alpha > 0$ come from Assumption 3 such that

$$\tilde{c}_0 := \varepsilon_0 \alpha > 0, \quad d_0 := c_0 - \gamma \left(\frac{\tilde{c}_0^2}{\alpha} \right) > 0. \tag{14}$$

Then, V_0 in (13) can be upper bounded as follows: due to L being $\mathcal{C}^1$, convex, and having a single minimizer z_1^* by Assumption 2, and due to L having quadratic growth away from z_1^* by Assumption 3, when $|\nabla L(z_1)| \leq \tilde{c}_0$, the suboptimality condition in Lemma 5 implies $|z_1 - z_1^*| \leq \frac{\tilde{c}_0}{\alpha}$, from where we get

$$V_0(z) \leq \gamma \left(\frac{\tilde{c}_0^2}{\alpha} \right) + \frac{1}{2} |z_2|^2 \tag{15}$$

Then, by defining the set $\mathcal{U}_0$ as

$$\mathcal{U}_0 := \left\{ z \in \mathbb{R}^{2n} : |\nabla L(z_1)| \leq \tilde{c}_0, \frac{1}{2} |z_2|^2 \leq d_0 \right\}, \tag{16}$$

every $z \in \mathcal{U}_0$ belongs to the c_0-sublevel set of V_0. In fact, using the conditions in (14) and (15), we have that for each $z \in \mathcal{U}_0$, $V_0(z) \leq \gamma \left(\frac{\tilde{c}_0^2}{\alpha} \right) + \frac{1}{2} |z_2|^2 \leq c_0$. Since κ_0 in (6a) is such that the set $\{z_1^*\} \times \{0\}$ is globally asymptotically stable for the closed-loop system resulting from controlling (4) by κ_0, as we show in the forthcoming Proposition 10, the set $\mathcal{U}_0$ is contained in the basin of attraction induced by κ_0.

2.3.2 Design of the Set $\mathcal{T}_{1,0}$

The objective is to design $\mathcal{T}_{1,0}$ such that when $z \in \mathcal{T}_{1,0}$ and $q = 1$, the state component z_1 is near z_1^* and the

supervisor resets q to 0, resets τ to 0, and assigns u to $\kappa_0(h_0(z))$. For such a design, we use Assumptions 2 and 3 and the Lyapunov function

$$V_1(z,\tau) := \frac{1}{2}\left|\bar{a}(\tau)\,(z_1 - z_1^*) + z_2\right|^2 + \frac{\zeta^2}{M}(L(z_1) - L^*) \quad (17)$$

defined for each $z \in \mathbb{R}^{2n}$ and each $\tau \geq 0$, where $\zeta > 0$, $M > 0$ is the Lipschitz constant of ∇L, and the function $\bar{a}$ is defined as

$$\bar{a}(\tau) := \frac{2}{\tau + 2}. \quad (18)$$

The choice of V_1 in (17) comes from [11], and is used to establish UGAS of the minimizer for $\mathcal{H}_1$ in (12) and the convergence rate $\frac{1}{(t+2)^2}$. In this same proof, the specific choice of $\bar{a}$ in (18), which comes from [11], is also used to show decrease of V_1. More details on how (17) and (18) are used in our analysis can be found in Proposition 13. Given $c_{1,0} \in (0, c_0)$ and $\varepsilon_{1,0} \in (0, \varepsilon_0)$, where $c_0 > 0$ and $\varepsilon_0 > 0$ come from Section 2.3.1, let $\tilde{c}_0$ and d_0 be given in (14), and let $\alpha > 0$ come from Assumption 3 such that

$$\tilde{c}_{1,0} := \varepsilon_{1,0}\alpha \in (0, \tilde{c}_0) \quad (19a)$$

$$d_{1,0} := c_{1,0} - \left(\frac{\tilde{c}_{1,0}}{\alpha}\right)^2 - \frac{\zeta^2}{M}\left(\frac{\tilde{c}_{1,0}^2}{\alpha}\right) \in (0, d_0) \quad (19b)$$

where $\zeta > 0$ comes from (3). Note that $\bar{a}$, defined via (18), which is in V_1, equals 1 when $\tau = 0$ and monotonically decreases toward zero (but being always positive) as τ tends to ∞. Namely, $\bar{a}$ is upper bounded by 1. Then, with V_1 given in (17), due to L being $\mathcal{C}^1$, convex, and having a single minimizer z_1^* by Assumption 2, and due to L having quadratic growth away from z_1^* by Assumption 3, when $|\nabla L(z_1)| \leq \tilde{c}_{1,0}$, the suboptimality condition in Lemma 5 implies $|z_1 - z_1^*| \leq \frac{\tilde{c}_{1,0}}{\alpha}$, from where we get

$$V_1(z,\tau) \leq \left(\frac{\tilde{c}_{1,0}}{\alpha}\right)^2 + |z_2|^2 + \frac{\zeta^2}{M}\left(\frac{\tilde{c}_{1,0}^2}{\alpha}\right). \quad (20)$$

Then, by defining $\mathcal{T}_{1,0}$ as

$$\mathcal{T}_{1,0} := \left\{z \in \mathbb{R}^{2n} : |\nabla L(z_1)| \leq \tilde{c}_{1,0}, |z_2|^2 \leq d_{1,0}\right\} \quad (21)$$

which, by construction, is contained in the interior of $\mathcal{U}_0$ defined in (16), every $z \in \mathcal{T}_{1,0}$ belongs to the $c_{1,0}$-sublevel set of V_1. In fact, using the conditions in (19) and (20), we have for each $z \in \mathcal{T}_{1,0}$, $V_1(z,\tau) \leq \left(\frac{\tilde{c}_{1,0}}{\alpha}\right)^2 + |z_2|^2 + \frac{\zeta^2}{M}\left(\frac{\tilde{c}_{1,0}^2}{\alpha}\right) \leq c_{1,0}$. The constants $\tilde{c}_0$, $\tilde{c}_{1,0}$, d_0, and $d_{1,0}$ in (14) and (19) comprise the hysteresis necessary to avoid chattering at the switching boundary. The idea behind these hysteresis boundaries is as follows. When $z \in \mathcal{U}_0$ and $q = 1$, we have that $z \in \overline{\mathbb{R}^{2n} \setminus \mathcal{T}_{1,0}}$, and it is not yet time to switch to κ_0 but to continue to flow using κ_1. But once $z \in \mathcal{T}_{1,0}$ then z is close enough to $\{z_1^*\} \times \{0\}$, and the supervisor switches to κ_0. Note that $\mathcal{T}_{0,1} \cap \mathcal{T}_{1,0} = \emptyset$. Fig. 3 illustrates the hysteresis mechanism in the design of $\mathcal{U}_0$ and $\mathcal{T}_{1,0}$.

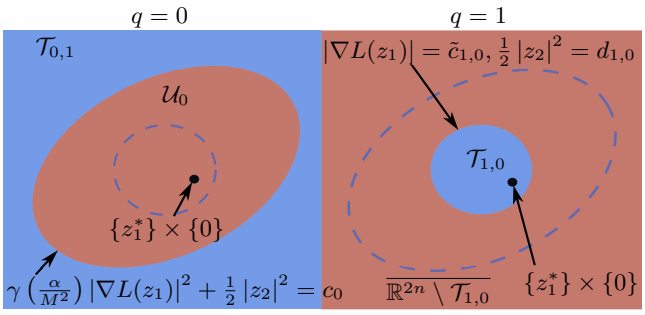

Fig. 3. An illustration of hysteresis in the design of the sets $\mathcal{U}_0$, $\mathcal{T}_{1,0}$, and $\mathcal{T}_{0,1}$ on $\mathbb{R}^{2n}$, via the constants $\tilde{c}_{1,0} \in (0, \tilde{c}_0)$, $d_{1,0} \in (0, d_0)$, and $c_0 > 0$. Left: due to the design of $\mathcal{U}_0$ in (16), every $z \in \mathcal{U}_0$ belongs to the c_0-sublevel set of the Lyapunov function V_0, where V_0 is defined via (13). Hence, the same value of $c_0 > 0$ is also used to define $\mathcal{T}_{0,1}$ as the closed complement of a sublevel set of V_0 with level equal to c_0. Right: the constants $\tilde{c}_{1,0} \in (0, \tilde{c}_0)$ and $d_{1,0} \in (0, d_0)$, defined via (19), are chosen such that the set $\mathcal{T}_{1,0}$ in (21) is contained in the interior of $\mathcal{U}_0$.

2.3.3 Design of the Set $\mathcal{T}_{0,1}$

The objective is to design $\mathcal{T}_{0,1}$ such that when $z \in \mathcal{T}_{0,1}$, $q = 0$, and $\tau = 0$, the state component z_1 is far from z_1^* and the supervisor resets q to 1 and assigns u to $\kappa_1(h_1(z,\tau),\tau)$ so that κ_1 steers z_1 back to nearby z_1^*. Given $c_0 > 0$, let $\alpha > 0$ come from Assumption 3, and let $M > 0$ come from Assumption 4. Then, using Assumption 4 with $u_1 = z_1^*$ and $w_1 = z_1$ yields $|\nabla L(z_1)| \leq M|z_1 - z_1^*|$ for all $z_1 \in \mathbb{R}^n$. Since L has quadratic growth away from z_1^* by Assumption 3, then dividing both sides of $|\nabla L(z_1)| \leq M|z_1 - z_1^*|$ by M and substituting into Assumption 3 leads to $L(z_1) - L^* \geq \frac{\alpha}{M^2}|\nabla L(z_1)|^2$, where $\alpha > 0$ comes from Assumption 3. Then, V_0 in (13) is lower bounded as follows: for each $z \in \mathbb{R}^{2n}$, $V_0(z) = \gamma(L(z_1) - L^*) + \frac{1}{2}|z_2|^2 \geq \gamma\left(\frac{\alpha}{M^2}\right)|\nabla L(z_1)|^2 + \frac{1}{2}|z_2|^2$. Using such a lower bound and the same $c_0 > 0$ as in Section 2.3.1, we define the set

$$\mathcal{T}_{0,1} := \left\{z \in \mathbb{R}^{2n} : \gamma\left(\frac{\alpha}{M^2}\right)|\nabla L(z_1)|^2 + \frac{1}{2}|z_2|^2 \geq c_0\right\}. \quad (22)$$

The set in (22) defines the (closed) complement of a sublevel set of the Lyapunov function V_0 in (13) with level equal to c_0. The constant c_0 is also a part of the hysteresis mechanism, as shown in Fig. 3. When $z \in \mathcal{U}_0$, $q = 0$, and $\tau = 0$, then the supervisor does not need to switch to κ_1, as the state component z is close enough to the minimizer to keep using κ_0. But if $z \in \mathcal{T}_{0,1}$ while $q = 0$ and $\tau = 0$, then z is far enough from the minimizer, and the supervisor then switches to κ_1.

While the constants $\tilde{c}_0$, $\tilde{c}_{1,0}$, d_0, $d_{1,0}$, and the set $\mathcal{T}_{0,1}$ in (22) depend on the constants $M > 0$ and $\alpha > 0$ which

characterize the objective function L, as long as M and α are positive, the uniform asymptotic stability property established in the forthcoming Theorem 8 still holds. As long as $M > 0$ and $\alpha > 0$ belong to a known set, the parameters $\tilde{c}_0$, $\tilde{c}_{1,0}$, d_0, and $d_{1,0}$ can still be tuned, treating such tuning as a worst-case tuning problem.

2.4 Design of the Parameter λ

The heavy ball parameter $\lambda > 0$ should be made large enough to avoid oscillations near the minimizer, as stated in Sections 1.1, 1.2, and 2.1. To gain some intuition on how to tune λ, consider the quadratic objective function $L(z_1) = \frac{1}{2}a_1 z_1^2$, $a_1 > 0$, which was analyzed in detail in [2]. For such a case, solutions to the heavy ball algorithm are overdamped (i.e., converge slowly with no oscillations) when $\lambda > 2\sqrt{a_1}$, critically damped (i.e., the fastest convergence possible with no oscillations) when $\lambda = 2\sqrt{a_1}$, and underdamped (fast convergence with oscillations) when $\lambda < 2\sqrt{a_1}$. Therefore, setting $\lambda \geq 2\sqrt{a_1}$ gives the desired behavior of solutions to $\mathcal{H}_0$, for such an objective function. More generally, setting λ sufficiently large to avoid oscillations suffices, in practice. Numerically, λ can be tuned as follows. Choose an arbitrarily large value of λ. If there is still oscillations or overshoot locally, despite the switch from κ_1 to κ_0 being made near the minimizer, then gradually increase λ until the oscillations and overshoot disappear. See Example 9 where λ was tuned in such a way.

2.5 Well-posedness of the hybrid closed-loop system $\mathcal{H}$

When L satisfies Assumptions 2, 3, and 4, the hybrid closed-loop system $\mathcal{H}$ in (9) satisfies the hybrid basic conditions in [16, Assumption 6.5]. The satisfaction of such conditions is demonstrated in the following lemma. Its proof is in [21].

Lemma 6 (Well-posedness of $\mathcal{H}$) *Let the function L satisfy Assumptions 2, 3, and 4. Let the sets $\mathcal{U}_0$, $\mathcal{T}_{1,0}$, and $\mathcal{T}_{0,1}$ be defined via (21), and (22), respectively. Let the functions $\bar{d}$ and $\bar{\beta}$ be defined as in (5). Let κ_0 and κ_1 be defined via (6). Then, the hybrid closed-loop system $\mathcal{H}$ in (9) satisfies the hybrid basic conditions.*

In Theorem 8 we show that $\mathcal{H}$ has a compact pre-asymptotically stable set. In light of this property, Lemma 6 is key as it leads to pre-asymptotic stability that is robust to small perturbations [16, Theorem 7.21]. In the case of gradient-based algorithms, for instance, such perturbations can take the form of small noise in measurements of the gradient.

Under Assumptions 2, 3, and 4, every maximal solution to $\mathcal{H}$ is complete; see [21, Section 2.6] for such a result.

2.6 Main Result

The result in this section depends on the notion of UGAS, which is defined in [15] and [16] as follows.

Definition 7 (UGAS) *Given a hybrid closed-loop system $\mathcal{H}$ as in (8), a nonempty set $\mathcal{A} \subset \mathbb{R}^n$ is said to be uniformly globally stable for $\mathcal{H}$ if there exists a class-$\mathcal{K}_\infty$ function α such that any solution x to $\mathcal{H}$ satisfies $|x(t,j)|_{\mathcal{A}} \leq \alpha(|x(0,0)|_{\mathcal{A}})$ for all $(t,j) \in \operatorname{dom} x$; uniformly globally pre-attractive (UGpA) for $\mathcal{H}$ if for each $\varepsilon > 0$ and $\delta > 0$ there exists $T > 0$ such that, for any solution x to $\mathcal{H}$ with $|x(0,0)|_{\mathcal{A}} \leq \delta$, $(t,j) \in \operatorname{dom} x$ and $t + j \leq T$ imply $|x(t,j)|_{\mathcal{A}} \leq \varepsilon$; and uniformly globally pre-asymptotically stable (UGpAS) for $\mathcal{H}$ if it is both uniformly globally stable and uniformly globally pre-attractive.*

When every maximal solution is complete, then the prefix "pre" is dropped to obtain UGA and UGAS. In this section, we present a result that establishes UGAS of the set

$$\begin{aligned}\mathcal{A} :=& \left\{ z \in \mathbb{R}^{2n} \ : \ \nabla L(z_1) = z_2 = 0 \right\} \times \{0\} \times \{0\} \\ =& \{z_1^*\} \times \{0\} \times \{0\} \times \{0\} \end{aligned} \tag{23}$$

and a hybrid convergence rate that, globally, is equal to $\frac{1}{(t+2)^2}$ while locally, is exponential, for the hybrid closed loop algorithm $\mathcal{H}$ in (9) and (10). Recall that the state $x := (z, q, \tau) \in \mathbb{R}^{2n} \times Q \times \mathbb{R}_{\geq 0}$. In light of this, the first component of $\mathcal{A}$, namely, $\{z_1^*\}$, is the minimizer of L. The second component of $\mathcal{A}$, namely, $\{0\}$, reflects the fact that we need the velocity state z_2 to equal zero in $\mathcal{A}$ so that solutions are not pushed out of such a set. The third component in $\mathcal{A}$, namely, $\{0\}$, is due to the logic state ending with the value $q = 0$, namely using κ_0 as the state z reaches the set of minimizers of L. The last component in $\mathcal{A}$ is due to τ being set to, and then staying at, zero when the supervisor switches to κ_0.

Theorem 8 (UGAS of $\mathcal{A}$ for $\mathcal{H}$) *Let the function L satisfy Assumptions 2, 3, and 4. Let $\zeta > 0$, $\lambda > 0$, $\gamma > 0$, $c_{1,0} \in (0, c_0)$, and $\varepsilon_{1,0} \in (0, \varepsilon_0)$ be given. Let $\alpha > 0$ be generated by Assumption 3, and let $M > 0$ be generated by Assumption 4. Let $\tilde{c}_{1,0} \in (0, \tilde{c}_0)$ and $d_{1,0} \in (0, d_0)$ be defined via (14) and (19). Let the sets $\mathcal{U}_0$, $\mathcal{T}_{1,0}$, and $\mathcal{T}_{0,1}$ be defined via (21), and (22), respectively. Let the functions $\bar{d}$ and $\bar{\beta}$ be defined as in (5), and let κ_0 and κ_1 be defined via (6). Then, the set $\mathcal{A}$, defined via (23), is UGAS for $\mathcal{H}$ given in (9)-(10). Furthermore, each maximal solution $(t,j) \mapsto x(t,j) = (z(t,j), q(t,j), \tau(t,j))$ of the hybrid closed-loop algorithm $\mathcal{H}$ starting from C_1 with $\tau(0,0) = 0$ satisfies the following:*

1) The domain $\operatorname{dom} x$ of the solution x is of the form[5] *$\cup_{j=0}^{1}(I^j \times \{j\})$, with I^0 of the form $[t_0, t_1]$ and with I^1*

[5] We define the interval $I^j := \{t : (t,j) \in \operatorname{dom} x\}$.

of the form $[t_1, \infty)$ *for some* $t_1 \geq 0$ *defining the time of the first jump. In other words, the system experiences at most one jump;*

2) *For each* $t \in I^0$ *such that*[6] $t \geq 0$

$$L(z_1(t,0)) - L^* \leq \frac{4cM}{\zeta^2(t+2)^2}\left(|z_1(0,0) - z_1^*|^2 + |z_2(0,0)|^2\right) \quad (24)$$

where $L^* = L(z^*)$ *and* $c := \left(1+\zeta^2\right)\exp\left(\sqrt{\frac{13}{4} + \frac{\zeta^4}{M}}\right)$. *Namely,* $L(z_1(t,0)) - L^*$ *is* $\mathcal{O}\left(\frac{4cM}{\zeta^2(t+2)^2}\right)$;

3) *For each* $t \in I^1$, $L(z_1(t,1)) - L^*$ *is* $\mathcal{O}\left(\exp\left(-(1-m)\psi t\right)\right)$, *where* $m \in (0,1)$ *is such that* $\psi := \frac{m\alpha\gamma}{\lambda} > 0$ *and* $\nu := \psi(\psi - \lambda) < 0$.

As will be shown in the forthcoming proof of Theorem 8 in Section 4, solutions starting from C_1 jump no more than once. The UGAS of the hybrid closed-loop algorithm $\mathcal{H}$ in Theorem 8 is proved as follows. First, in the forthcoming Proposition 10, we establish UGAS of the set $\{z_1^*\} \times \{0\}$ for the closed-loop algorithm $\mathcal{H}_0$ in (11) via Lyapunov theory and the application of an invariance principle. Then, in the forthcoming Proposition 15, we prove UGAS of the set $\{z_1^*\} \times \{0\} \times \mathbb{R}_{\geq 0}$ for the closed-loop algorithm $\mathcal{H}_1$ in (12) via Lyapunov theory and a comparison principle. Then, UGAS of $\mathcal{A}$ for $\mathcal{H}$ and item 1) in Theorem 8 follow from a proof-by-contradiction employing the UGAS of $\mathcal{H}_0$, the UGAS of $\mathcal{H}_1$, and the construction of the sets $\mathcal{U}_0$, $\mathcal{T}_{1,0}$, and $\mathcal{T}_{0,1}$. The hybrid convergence rate of the closed-loop algorithm $\mathcal{H}$ in items 2) and 3) of Theorem 8 is proved in the forthcoming Propositions 12, 13, and 14.

3 Numerical Example

Example 9 *In this example, to show the effectiveness of the uniting algorithm, we compare the hybrid closed-loop algorithm* $\mathcal{H}$, *defined via* (9) *and* (10), *with the individual closed-loop optimization algorithms* $\mathcal{H}_0$ *and* $\mathcal{H}_1$ *and with the HAND-1 algorithm from [12] which, in [12], is designed and analyzed for convex functions* L *satisfying Assumptions 2 and 4. The bound for HAND-1 is* $L(z_1(t,0)) - L^* \leq \frac{B}{t^2}$ *for all* $(t,j) \in \mathrm{dom}(z,\tau)$ *such that* $j = 0$, $z_1(0,0) = z_2(0,0)$, $\tau(0,0) = T_{\min}$, $z_1(0,0) \in K_0 := \{z_1^*\} + r\mathbb{B}$, *where* $B := \frac{r^2}{2c_1} + T_{\min}^2\left(L(z_1(0,0)) - L^*\right) > 0$, $r \in \mathbb{R}_{>0}$, $c_1 > 0$. *Such a rate is only guaranteed until the first jump.*

Next, we compare $\mathcal{H}_0$, $\mathcal{H}_1$, $\mathcal{H}$, *and HAND-1 in simulation. To compare these algorithms, the choice of objective function, parameter values, and initial conditions are as follows. We use the objective function* $L(z_1) := z_1^2$, *the gradient of which is Lipschitz continuous with* $M = 2$, *and which has a single minimizer at* $z_1^* = 0$. *This choice of objective function is made so that we can easily tune* λ, *as described in Section 2.4. We arbitrarily*[7] *chose the heavy ball parameter value* $\gamma = \frac{2}{3}$ *and we tuned* λ *to* 200 *by choosing a value arbitrarily larger than* $2\sqrt{a_1}$, *where* a_1 *comes from Section 2.4, and gradually increasing it until there is no overshoot in the hybrid algorithm. In Figure 2, we stated that choosing* $\zeta = 2$ *leads to faster convergence, for Nesterov's method in* (3) *and* $\mathcal{H}$, *than choosing* $\zeta = 1$. *In general, convergence for such algorithms is faster as* ζ *increases, and slower as* ζ *tends to zero. Given* $\zeta = 2$ *for Nesterov's algorithm and* $\mathcal{H}$, *the HAND-1 parameters* $c_1 = 0.5$ *and* $T_{\min} = \frac{1+\sqrt{7}}{2}$ *are chosen such that the resulting gain coefficients for* z_1 *and* z_2 *are the same for both* $\mathcal{H}$ *and HAND-1, so that these algorithms are compared on equal footing*[8]. *The remaining HAND-1 parameters,* r *and* δ_{med}, *have different values depending on the initial conditions* $z_1(0,0) = z_2(0,0)$, *listed in [21, Table 2], which leads to different values of* T_{med} *and* $T_{\max}$, *for each solution. Such values are chosen such that* $T_{med} \geq \sqrt{\frac{B}{\delta_{med}}} + T_{\min} > 0$. *Additionally, we choose* $T_{\max} = T_{med} + 1$. *The parameter values for the uniting algorithm are* $\varepsilon_0 = 10$, $\varepsilon_{1,0} = 5$, *and* $\alpha = 1$. *The remaining parameter values* c_0 *and* $c_{1,0}$ *are different depending on the initial condition* $z_1(0,0)$ *and are listed in [21, Table 2], which leads to different values of* d_0, *calculated via* (14), *and* $d_{1,0}$ *calculated via* (19). *These values are chosen for proper tuning of the algorithm, in order to get nice performance, and for exploiting the properties of Nesterov's method as long as we want. Initial conditions for all solutions to* $\mathcal{H}$ *are* $z_2(0,0) = 0$, $q(0,0) = 1$, *and* $\tau(0,0) = 0$, *with values of* $z_1(0,0)$ *listed in [21, Table 2]. Initial conditions for all solutions to HAND-1 are* $\tau(0,0) = T_{\min}$, *with values of* $z_1(0,0) = z_2(0,0)$ *listed in [21, Table 2].*

The time that it takes for each algorithm to settle within 1% *of* z_1^*, *averaged over solutions starting from ten different values*[9] *of* $z_1(0,0)$, *are as follows:* 0.811 *seconds for* $\mathcal{H}$, 690.759 *seconds for* $\mathcal{H}_0$, 4.409 *seconds for* $\mathcal{H}_1$, *and* 8.649 *seconds for HAND-1. Using the formula*

$$\left(\frac{(\textit{Time of } \mathcal{H}_0, \mathcal{H}_1, \textit{ or HAND-1}) - \textit{Time of } \mathcal{H}}{\textit{Time of } \mathcal{H}_0, \mathcal{H}_1, \textit{ or HAND-1}}\right) \times 100\%. \quad (25)$$

the average percent improvement of $\mathcal{H}$ *over each of the other algorithms is* 99.9% *over* $\mathcal{H}_0$, 81.6% *over* $\mathcal{H}_1$, *and* 90.6% *over HAND-1.*

[6] Note that at each $t \in I^0$, $q(t,0) = 1$, and at each $t \in I^1$, $q(t,1) = 0$.

[7] Although the choice of γ is arbitrary, we have found in general that choosing $\gamma \in (0,1)$ works well, in practice.

[8] Although there exist parameter values for which HAND-1 has faster, oscillation-free performance, due to the way $\mathcal{H}$ and HAND-1 relate to each other, they are compared fairly for a particular set of parameters.

[9] Code at `gitHub.com/HybridSystemsLab/UnitingDifferentICs`

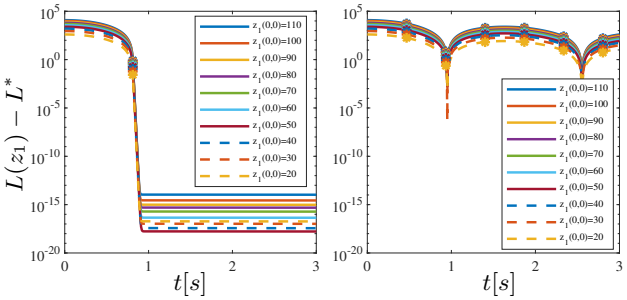

Fig. 4. The evolution of L over time, from different initial conditions, for $\mathcal{H}$ (left) and HAND-1 (right). All solutions are for the objective function $L(z_1) := z_1^2$, and the parameters used for HAND-1 and $\mathcal{H}$ are listed in [21, Table 2], with different values of c_0 and $c_{1,0}$ for each solution of $\mathcal{H}$, leading to different values of d_0 calculated via (14) and $d_{1,0}$ calculated via (19), and different values of r and δ_{med} for each solution of HAND-1, leading to different values of T_{med} and $T_{\max}$. Jumps are marked with asterisks.

Fig. 4 compares different solutions for $\mathcal{H}$ and HAND-1, from different values of $z_1(0,0)$, for the objective function $L(z_1) := z_1^2$. [21, Table 2] lists the times for which each solution settles to within 1% of z_1^* for both $\mathcal{H}$ and HAND-1, and shows the percent improvement of $\mathcal{H}$ over HAND-1. As can be seen in Fig. 4 and in [21, Table 2], the percent improvement of $\mathcal{H}$ over HAND-1 for all solutions is 90.6%, which shows consistency in the performance of $\mathcal{H}$ versus HAND-1.

For an illustration of the robustness of the UGAS property in Theorem 8 to perturbations, see [21, Example 4.1]. For a comparison between $\mathcal{H}$, HAND-1, and the hybrid Hamiltonian algorithm (HHA) in [14], see [21, Example 4.3]. For an illustration of the trade-off between speed of convergence and the resulting values of parameters for the uniting algorithm $\mathcal{H}$, for different values of $\zeta > 0$, see [21, Example 4.3].

4 Proof of Theorem 8

This section provides a proof of Theorem 8 from Section 2.6. Section 4.1 establishes UGAS of $\{z_1^*\} \times \{0\}$ and an exponential convergence rate for $\mathcal{H}_0$. Section 4.2 establishes UGAS of $\{z_1^*\} \times \{0\} \times \mathbb{R}_{\geq 0}$ and a convergence rate $\frac{1}{(t+2)^2}$ for $\mathcal{H}_1$. Section 4.3 uses the properties in Sections 4.1 and 4.2 to establish UGAS of $\mathcal{A}$, defined via (23), for $\mathcal{H}$. Finally, Section 4.4 proves the convergence rate of $\mathcal{H}$ using the convergence rates of the individual closed-loop algorithms $\mathcal{H}_0$ and $\mathcal{H}_1$ established in Sections 4.1 and 4.2, respectively.

4.1 Properties of $\mathcal{H}_0$

The following result establishes that the closed-loop algorithm $\mathcal{H}_0$ has the set $\{z_1^*\} \times \{0\}$ UGAS. To prove it, we use an invariance principle. Its proof is in [21].

Proposition 10 (UGAS of $\{z_1^*\} \times \{0\}$ for $\mathcal{H}_0$) *Let L satisfy Assumptions 2, 3, and 4. For each $\lambda > 0$ and $\gamma > 0$, the set $\{z_1^*\} \times \{0\}$ is UGAS for the closed-loop algorithm $\mathcal{H}_0$ in (11).*

Next, we establish the convergence rate of the closed-loop algorithm $\mathcal{H}_0$. To do so, we use the following Lyapunov function, proposed in [9, Lemma 4.2]:

$$V(z) := \gamma \left(L(z_1) - L^*\right) + \frac{1}{2} \left|\psi(z_1 - z_1^*) + z_2\right|^2 + \frac{\nu}{2} \left|z_1 - z_1^*\right|^2 \tag{26}$$

where, given $\lambda > 0$, $\psi > 0$ is chosen such that $\nu := \psi(\psi - \lambda) < 0$. When L satisfies Assumption 2, the following lemma, which is a version of [9, Lemma 4.2] tailored for the unperturbed heavy ball algorithm in (11), gives an upper bound on the change of V in (26). Its proof is in [21].

Lemma 11 *Let L satisfy Assumption 2, and let $\lambda > 0$ and $\gamma > 0$, which come from $\mathcal{H}_0$ in (11), be given. For each $\psi > 0$ such that $\nu := \psi(\psi - \lambda) < 0$, the following bound is satisfied for each $z \in \mathbb{R}^{2n}$: $\dot{V}(z) \leq -\psi\left(a(z_1) + 2\nu c(z_1)\right) + 2(\psi - \lambda) b(z)$, where V is defined in (26), $a(z_1) := \gamma\left(L(z_1) - L^*\right)$, $b(z) := \frac{1}{2}\left|\psi(z_1 - z_1^*) + z_2\right|^2$, and $c(z_1) := \frac{1}{2}\left|z_1 - z_1^*\right|^2$.*

We employ Lemma 5.2 to show that when L satisfies Assumptions 2 and 3, the convergence rate of the closed-loop algorithm $\mathcal{H}_0$ in (11) is exponential. This is supported by the following proposition, which is a version of [9, Theorem 3.2] tailored for the unperturbed heavy ball algorithm $\mathcal{H}_0$ in (11). Its proof is in [21].

Proposition 12 (Convergence rate for $\mathcal{H}_0$) *Let L satisfy Assumptions 2 and 3, let $\alpha > 0$ come from Assumption 3, and let $\lambda > 0$ and $\gamma > 0$ come from $\mathcal{H}_0$ in (11). For each $m \in (0,1)$ such that $\psi := \frac{m\alpha\gamma}{\lambda} > 0$ and $\nu := \psi(\psi - \lambda) < 0$, each maximal solution $t \mapsto z(t)$ to the closed-loop algorithm $\mathcal{H}_0$ satisfies*

$$L(z_1(t)) - L^* = \mathcal{O}\left(\exp\left(-(1-m)\psi t\right)\right) \tag{27}$$

for all $t \in \operatorname{dom} z$ $(= \mathbb{R}_{\geq 0})$.

4.2 Properties of $\mathcal{H}_1$

When L satisfies Assumptions 2 and 4, then we can derive an upper bound, for all $t \geq 0$, on V_1 in (17) along solutions to $\mathcal{H}_1$. To derive such a bound, we extend [11, Proposition 3.2] to functions L with generic L^*, z_1^*, and $\zeta > 0$, in the following proposition. Its proof is in [21].

Proposition 13 *Let L satisfy Assumptions 2 and 4. Then, each maximal solution $t \mapsto (z(t), \tau(t))$ to the closed-loop algorithm $\mathcal{H}_1$ in (12) with $\tau(0) = 0$ satisfies $V_1(z(t), t) \leq \frac{4}{(t+2)^2} V_1(z(0), 0)$ for all $t \geq 0$, where V_1 is defined via (17).*

The following proposition establishes that the closed-loop algorithm $\mathcal{H}_1$ has a convergence rate $\frac{1}{(t+2)^2}$ for all $t \geq 0$. To prove it, we use Proposition 13. Its proof is in [21].

Proposition 14 (Convergence rate for $\mathcal{H}_1$) *Let L satisfy Assumptions 2 and 4. Let $\zeta > 0$ and $M > 0$ come from Assumption 4. Then, for each maximal solution $t \mapsto (z(t), \tau(t))$ to the closed-loop algorithm $\mathcal{H}_1$ in (12) with $\tau(0) = 0$, the following holds:*

$$\begin{aligned} &\frac{\zeta^2}{M}(L(z_1(t)) - L^*) \qquad (28)\\ &\leq V_1(z(t), t) \leq \frac{4c}{(t+2)^2}\left(|z_1(0) - z_1^*|^2 + |z_2(0)|^2\right) \end{aligned}$$

for all $t \geq 0$, where $c := \left(1 + \zeta^2\right) \exp\left(\sqrt{\frac{13}{4} + \frac{\zeta^4}{M}}\right)$.

The following proposition establishes that the closed-loop system $\mathcal{H}_1$ in (12) has the set

$$\mathcal{A}_1 := \{z_1^*\} \times \{0\} \times \mathbb{R}_{\geq 0} \qquad (29)$$

UGAS. To prove it, we use Proposition 14 and [16, Theorem 3.18]. Its proof is in [21].

Proposition 15 (UGAS of $\mathcal{A}_1$ in (29) for $\mathcal{H}_1$) *Let L satisfy Assumptions 2 and 4. Let $\zeta > 0$ and let $M > 0$ come from Assumption 4. Then, the set $\mathcal{A}_1$ in (29) is UGAS for $\mathcal{H}_1$.*

4.3 Uniform Global Asymptotic Stability of $\mathcal{A}$ for $\mathcal{H}$

The hybrid closed-loop algorithm $\mathcal{H}$ satisfies the hybrid basic conditions by Lemma 6, satisfying the first assumption of [21, Theorem A.3]. Furthermore, $\Pi(C_0) \cup \Pi(D_0) = \mathbb{R}^{2n}$, $\Pi(C_1) \cup \Pi(D_1) = \mathbb{R}^{2n}$, and each maximal solution $(t, j) \mapsto x(t, j) = (z(t, j), q(t, j), \tau(t, j))$ to $\mathcal{H}$ in (9)-(10) is complete and bounded by [21, Proposition 2.9]. Since by Assumption 2, L has a unique minimizer z_1^*, then $\mathcal{A}$, defined via (23), is compact by construction, and $\mathcal{U} = \mathbb{R}^{2n} \times Q \times \mathbb{R}_{\geq 0}$ contains a nonzero open neighborhood of $\mathcal{A}$, satisfying the second assumption of [21, Theorem A.3].

To prove attractivity [10] of $\mathcal{A}$, we proceed by contradiction. Suppose there exists a complete solution x to $\mathcal{H}$ such that $\lim_{t+j\to\infty} |x(t, j)|_{\mathcal{A}} \neq 0$. Since [21, Proposition 2.9] guarantees completeness of maximal solutions, we have the following cases:

a) There exists $(t', j') \in \operatorname{dom} x$ such that $x(t, j) \in C_1 \setminus D_1$ for all $(t, j) \in \operatorname{dom} x, t + j \geq t' + j'$;
b) There exists $(t', j') \in \operatorname{dom} x$ such that $x(t, j) \in C_0 \setminus (\mathcal{A} \cup D_0)$ for all $(t, j) \in \operatorname{dom} x, t + j \geq t' + j'$;
c) There exists $(t', j') \in \operatorname{dom} x$ such that $x(t, j) \in D$ for all $(t, j) \in \operatorname{dom} x, t + j \geq t' + j'$.

Case a) contradicts the fact that, by Proposition 15, the set $\mathcal{A}_1$, defined via (29), is UGAS for $\mathcal{H}_1$. Such UGAS of $\mathcal{A}_1$ implies there exist $\tilde{c}_1 \in (0, \tilde{c}_{1,0})$ and $d_1 \in (0, d_{1,0})$ such that the state z reaches $(\{z_1^*\} + \tilde{c}_1\mathbb{B}) \times (\{0\} + d_1\mathbb{B}) \subset \mathcal{T}_{1,0}$ at some finite flow time $t \geq 0$ or as $t \to \infty$. In turn, due to the construction of C_1 and D_1 in (10), with $\mathcal{T}_{1,0}$ defined via (21), the solution x must reach D_1 at some $(t, j) \in \operatorname{dom} x, t + j \geq t' + j'$. Therefore, case a) does not happen.

Case b) contradicts the fact that, by Proposition 10, $\{z_1^*\} \times \{0\}$ is UGAS for $\mathcal{H}_0$. In fact, $\lim_{t+j\to\infty} |x(t, j)|_{\mathcal{A}} = 0$, and since $\mathcal{A} \subset C_0$, case b) does not happen.

Case c) contradicts the fact that, due to the construction of $\mathcal{T}_{1,0}$ in (21) and $\mathcal{T}_{0,1}$ in (22), we have $G(D) \cap D := ((\mathcal{T}_{0,1} \times \{1\} \times \{0\}) \cup (\mathcal{T}_{1,0} \times \{0\} \times \{0\}))$ $\cap ((\mathcal{T}_{0,1} \times \{0\} \times \{0\}) \cup (\mathcal{T}_{1,0} \times \{1\} \times \mathbb{R}_{\geq 0})) = \emptyset$ where D is defined in (10). Such an equality holds since $\mathcal{T}_{1,0} \cap \mathcal{T}_{0,1} = \emptyset$; see the end of Section 2.3.2. Therefore, case c) does not happen.

Therefore, cases a)-c) do not happen, and each maximal and complete solution $x = (z, q, \tau)$ to $\mathcal{H}$ with $\tau(0, 0) = 0$ converges to $\mathcal{A}$. Consequently, by the construction of C and D in (10), the UGAS of $\mathcal{A}_1$ (defined via (29)) for $\mathcal{H}_1$ established in Proposition 15, the UGAS of $\{z_1^*\} \times \{0\}$ for $\mathcal{H}_0$ established in Proposition 10, and since each maximal solution to $\mathcal{H}$ is complete by [21, Proposition 2.9], the set $\mathcal{A}$ is UGAS for $\mathcal{H}$.

To show that each maximal and complete solution x to $\mathcal{H}$ jumps no more than twice, we proceed by contradiction. Without loss of generality, suppose there exists a maximal and complete solution that jumps three times. We have the following possible cases: i) The solution first jumps at a point in D_0, then jumps at a point in D_1, and then jumps at a point in D_0; or ii) The solution first jumps at a point in D_1, then jumps at a point in D_0, and then jumps at a point in D_1. Case i) does not hold since, once the jump in D_1 occurs, the solution x is in $(\mathcal{T}_{1,0} \times \{0\} \times \{0\}) \subset C_0$. Due to the construction of $\mathcal{T}_{1,0}$ in (21) and $\mathcal{T}_{0,1}$ in (22) such that $\mathcal{T}_{1,0} \cap \mathcal{T}_{0,1} = \emptyset$, as described in the contradiction of case c) above, and due to the UGAS of $z_1^* \times \{0\}$ for $\mathcal{H}_0$ by Proposition 10, the solution x will never return to D_0. Therefore, case i) does not happen. Case ii) leads to a contradiction for the same reason, and in this case, once the first jump in D_1 occurs, no more jumps happen. Therefore, since cases i)-ii) do not happen, each maximal and complete solution x to $\mathcal{H}$ with $\tau(0, 0) = 0$ has no more than two jumps.

4.4 Convergence Rate of $\mathcal{H}$

Finally, we prove the hybrid convergence rate of $\mathcal{H}$. Letting $\zeta > 0$ and letting $M > 0$ come from Assumption

[10] For a definition of attractivity, see [15, Definition 3.1].

4, then by Proposition 14, since L satisfies Assumptions 2 and 4, each maximal solution $t \mapsto (z(t), \tau(t))$ to the closed-loop algorithm $\mathcal{H}_1$ with $\tau(0,0) = 0$ satisfies (28), for all $t \geq 0$, where c is defined below (28). By Proposition 12, since L satisfies Assumptions 2 and 3, then, given $\gamma > 0$ and $\lambda > 0$, for each $m \in (0,1)$ such that $\psi := \frac{m\alpha\gamma}{\lambda} > 0$ and $\nu := \psi(\psi - \lambda) < 0$, each maximal solution $t \mapsto z(t)$ to the closed-loop algorithm $\mathcal{H}_0$ satisfies (27) for all $t \in \operatorname{dom} z \ (= \mathbb{R}_{\geq 0})$. Since maximal solutions $(t,j) \mapsto x(t,j) = (z(t,j), q(t,j), \tau(t,j))$ to $\mathcal{H}$ starting from C_1 are guaranteed to jump no more than once, as implied by the contradiction in cases i)-ii) above, then the domain of each maximal solution x to $\mathcal{H}$ starting from C_1 is $\cup_{j=0}^{1}(I^j, j)$, with I^0 of the form $[t_0, t_1]$ and with I^1 of the form $[t_1, \infty)$. Therefore, given $\zeta > 0$, $\lambda > 0$, $\gamma > 0$, $c_{1,0} \in (0, c_0)$, $\varepsilon_{1,0} \in (0, \varepsilon_0)$, $\alpha > 0$ from Assumption 3, and $M > 0$ from Assumption 4, due to the construction of $\mathcal{U}_0$, $\mathcal{T}_{1,0}$, and $\mathcal{T}_{0,1}$ in (16), (21), and (22), with $\tilde{c}_{1,0} \in (0, \tilde{c}_0)$ and $d_{1,0} \in (0, d_0)$ defined via (14) and (19), and due to the individual convergence rates of $\mathcal{H}_1$ and $\mathcal{H}_0$, each maximal solution $(t,j) \mapsto x(t,j) = (z(t,j), q(t,j), \tau(t,j))$ to the hybrid closed-loop algorithm $\mathcal{H}$ that starts in C_1, such that $\tau(0,0) = 0$, satisfies (24) for each $t \in I^0$ at which $q(t,0)$ is equal to 1 and $t \geq 0$, and satisfies item 3) of Theorem 8 for each $t \in I^1$ at which $q(t,1)$ is equal to 0.

Remark 16 *Although it is outside the scope of this paper, a potential approach to discretizing the hybrid closed-loop algorithm $\mathcal{H}$ in* (9)-(10) *can be found in [24]. Such a discretization approach, which is designed for hybrid systems and which has assumptions that are satisfied by forward Euler and p-stage Runge-Kutta consistent methods, for example, would yield results similar to Theorem 8, Proposition 10, Lemma 11, and Propositions 12, 13, 14, and 15.*

5 Conclusion

We presented an algorithm, designed using hybrid system tools, that unites Nesterov's accelerated algorithm and the heavy ball algorithm to ensure fast convergence and UGAS of the unique minimizer for $\mathcal{C}^1$, convex objective functions L. The hybrid convergence rate is $\frac{1}{(t+2)^2}$ globally and exponential locally. In simulation, we showed performance improvement not only over the individual heavy ball and Nesterov algorithms, but also over the HAND-1 algorithm in [12]. In the process, we proved the existence of solutions for the individual heavy ball and Nesterov algorithms, and we extended the convergence rate results for Nesterov's algorithm in [11] to functions L with generic z_1^*, L^*, and $\zeta > 0$. Additionally, we established UGAS of the minimizer for Nesterov's algorithm, when L is $\mathcal{C}^1$, convex, and has a unique minimizer. Future work will extend the uniting algorithm to a general framework, allowing the local and global algorithms to be any accelerated gradient algorithm. We will also extend the uniting algorithm to learning applications.

References

[1] B. T. Polyak, "Some methods of speeding up the convergence of iteration methods," *USSR Computational Mathematics and Mathematical Physics*, vol. 4, no. 5, pp. 1–17, 1964.

[2] H. Attouch, X. Goudou, and P. Redont, "The heavy ball with friction method, I. the continuous dynamical system: global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system," *Communications in Contemporary Mathematics*, vol. 2, no. 01, pp. 1–34, 2000.

[3] B. Polyak and P. Shcherbakov, "Lyapunov functions: an optimization theory perspective," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 7456–7461, 2017.

[4] E. Ghadimi, H. R. Feyzmahdavian, and M. Johansson, "Global convergence of the heavy-ball method for convex optimization," in *14th IEEE European Control Conference*, 2015, pp. 310–315.

[5] M. Fazlyab, M. Morari, and V. M. Preciado, "Design of first-order optimization algorithms via sum-of-squares programming," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 4445–4452.

[6] 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.

[7] A. Badithela and P. Seiler, "Analysis of the heavy-ball algorithm using integral quadratic constraints," in *2019 American control conference (ACC)*. IEEE, 2019, pp. 4081–4085.

[8] J. H. Le and A. R. Teel, "Hybrid heavy-ball systems: reset methods for optimization with uncertainty," in *2021 American Control Conference (ACC)*. IEEE, 2021, pp. 2236–2241.

[9] O. Sebbouh, C. Dossal, and A. Rondepierre, "Convergence rates of damped inertial dynamics under geometric conditions and perturbations," *SIAM Journal on Optimization*, vol. 30, no. 3, pp. 1850–1877, 2020.

[10] S. Michalowsky and C. Ebenbauer, "Extremum control of linear systems based on output feedback," in *55th IEEE Conference on Decision and Control*, 2016, pp. 2963–2968.

[11] M. Muehlebach and M. Jordan, "A dynamical systems perspective on nesterov acceleration," in *International Conference on Machine Learning*. PMLR, 2019, pp. 4656–4662.

[12] J. I. Poveda and N. Li, "Inducing uniform asymptotic stability in non-autonomous accelerated optimization dynamics via hybrid regularization," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 3000–3005.

[13] ——, "Robust hybrid zero-order optimization algorithms with acceleration via averaging in time," *Automatica*, vol. 123, p. 109361, 2021.

[14] A. R. Teel, J. I. Poveda, and J. Le, "First-order optimization algorithms with resets and hamiltonian flows," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 5838–5843.

[15] R. G. Sanfelice, *Hybrid Feedback Control*. New Jersey: Princeton University Press, 2021.

[16] R. Goebel, R. G. Sanfelice, and A. R. Teel, *Hybrid Dynamical Systems: Modeling, Stability, and Robustness*. New Jersey: Princeton University Press, 2012.

[17] A. R. Teel and N. Kapoor, "Uniting local and global controllers," in *1997 European Control Conference (ECC)*. IEEE, 1997, pp. 3868–3873.

[18] A. R. Teel, O. E. Kaiser, and R. M. Murray, "Uniting local and global controllers for the caltech ducted fan," in *Proceedings of the 1997 American Control Conference (Cat. No. 97CH36041)*, vol. 3. IEEE, 1997, pp. 1539–1543.

[19] N. Flammarion and F. Bach, "From averaging to acceleration, there is only a step-size," in *Conference on Learning Theory*, 2015, pp. 658–695.

[20] D. M. Hustig-Schultz and R. G. Sanfelice, "Uniting nesterov's accelerated gradient descent and the heavy ball method for strongly convex functions with exponential convergence rate," in *2021 American Control Conference (ACC)*. IEEE, 2021, pp. 959–964.

[21] ——, "Uniting nesterov and heavy ball methods for uniform global asymptotic stability of the set of minimizers," *arXiv preprint*, 2022. [Online]. Available: https://arxiv.org/abs/2202.07739

[22] D. Hustig-Schultz and R. G. Sanfelice, "A robust hybrid heavy ball algorithm, for optimization with high performance," in *Proceedings of the American Control Conference*, July 2019, pp. 151–156.

[23] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[24] R. G. Sanfelice and A. R. Teel, "Dynamical properties of hybrid systems simulators," *Automatica*, vol. 46, no. 2, p. 239–248, 2010. [Online]. Available: https://hybrid.soe.ucsc.edu/files/preprints/40.pdf