

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

Dawn M. Hustig-Schultz

dhustigs@ucsc.edu

Ricardo G. Sanfelice

ricardo@ucsc.edu

October 10, 2023

Technical Report
Hybrid Systems Laboratory
Department of Computer Engineering
University of California, Santa Cruz

Technical Report No. TR-HSL-04-2021

Available at <https://hybrid.soe.ucsc.edu/biblio>

Readers of this material have the responsibility to inform all of the authors promptly if they wish to reuse, modify, correct, publish, or distribute any portion of this report.

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 and demonstrate the robustness of the uniting algorithm.

1 Introduction

1.1 Background and Motivation

There has been growing interest in analyzing accelerated gradient methods from a dynamical systems perspective [1], which permits the use of well established analysis tools, such as Lyapunov theory, to study convergence and stability properties of accelerated algorithms [2], [3], [4], [5], [6], [7]. The *heavy ball method* is an accelerated gradient method that guarantees convergence to the minimizer of a convex function L [8], and that achieves a faster convergence rate than classical gradient descent by adding a “velocity” term to the gradient. 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 [3], [2]. In [9] and [10] 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 [9] shows convergence with rate $\frac{1}{k}$ when L is convex. It is shown in [11] 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 [12] 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. For the case when L is strongly convex, and inspired by the heavy ball algorithm, two algorithms with a resettable velocity term are proposed in [13] and shown to guarantee exponential convergence. In [14], however, it was demonstrated 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. Global asymptotic stability of the minimizer, which is the property that all solutions that start close to the minimizer stay close, and solutions from all initial conditions converge to the minimizer, is demonstrated in [15], [16], when L is convex and smooth. The work in [2] provides several Lyapunov functions to establish global asymptotic stability of the minimizer and convergence rates for the heavy ball method, both when L is strongly convex and when L is convex.

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 [1], is

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

where $M > 0$ is the Lipschitz constant of the gradient 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 [1], the convergence rate of Nesterov's method is characterized as $\frac{1}{(t+2)^2}$ for (2) (for $t \geq 1$), when $\zeta = 1$, and when the minimizer is the origin, at which L is zero.

Exponential convergence of the discrete-time analogue of Nesterov's method is established for strongly convex L with Lipschitz continuous ∇L in [10] and [11]. The earliest dynamical system characterization for Nesterov's algorithm was proposed in [4], including a variation for higher friction, and the proposed characterizations were shown to have a convergence rate of $\frac{1}{t^2}$. In [5], the analysis of the dynamical system in [4] is extended to include optimization of objective functions L with non-Euclidean geometries, and this dynamical system is combined with mirror descent to design an acceler-

ated mirror descent ODE, with a convergence rate of $\frac{1}{t^2}$. In [6], a dynamical system, consisting of an Euler-Lagrange equation, is derived for Nesterov’s algorithm via a Bregman Lagrangian, with an exponential rate of convergence under ideal scaling and a rate of convergence of $\frac{1}{t^p}$ with $p \geq 2$ for a polynomial subclass of such a dynamical system.

In [6] an exponential rate of convergence for such a system under ideal scaling is provided, and, for a polynomial class of dynamical systems, a convergence rate of $\frac{1}{t^p}$ with $p \geq 2$ is shown. In [17] and [18], two hybrid algorithms based on the ODE in [4] are presented: one with a state-dependent, time-invariant damping input and another with an input that controls the magnitude of the gradient term. The algorithms require the objective function to satisfy the Polyak-Łojasiewicz inequality, which includes a subclass of nonconvex functions in which all stationary points are global minimizers. The authors in [19] propose two hybrid reset algorithms based on the ODE in [4], HAND-1 and HAND-2, which yield an exponential convergence rate for strongly convex L and a rate of $\frac{1}{t^2}$ for convex L , with the latter rate only assured until the first reset.

While the results in [1], [4], [5], [6], [10], 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 neighborhood of the minimizer in time that is uniform on the set of initial conditions [19], [20], [21]. After finite time, the error of such solutions becomes smaller than a given threshold [22]. Due to such a guarantee for solutions, UGAS is typically useful for certifying robustness to small perturbations in time-varying dynamical and hybrid systems [23], [22]. Remarkably, the algorithm with resets in the velocity term proposed in [13] can be shown to induce UGAS of the minimizer (with zero velocity term) and reduced oscillations, for the particular case when L is strongly convex. The algorithm with resets in [21] can be shown to induce UGAS of the minimizer 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 [19], via a counterexample, Nesterov-like algorithms do not necessarily assure UGAS of the minimizer when L is convex. In response to this, [19] proposes the HAND-1 and HAND-2 reset algorithms, based on the ODE in [4], and prove UGAS of the minimizer 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

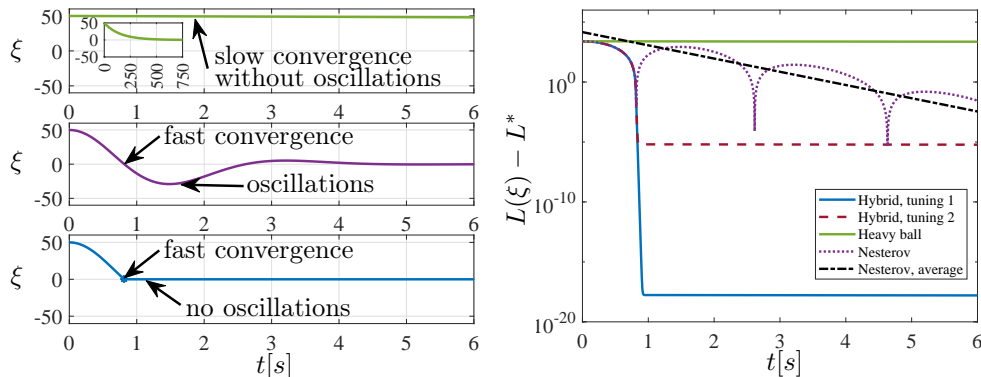


Figure 1: Comparison of the performance of heavy ball, with large λ , Nesterov’s method, and the proposed logic-based algorithm. The objective function is $L(\xi) = \xi^2$. Top left: the heavy ball algorithm, with large λ , converges very slowly. Top inset: zoomed out view of heavy ball. Middle left: Nesterov’s method converges quickly, but with oscillations. Bottom left: our proposed logic-based algorithm yields fast convergence, with no oscillations. Right: comparison of the value of $L(\xi) - L^*$ (in log scale) versus time for each algorithm. Different tunings of the logic-based algorithm’s parameters leads to modifications of the solution’s profile.

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 the minimizer), when the objective function 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, for instance, depends highly on the choice of λ and γ . In particular, for a fixed value of γ , the choice of the friction parameter λ significantly affects the asymptotic behavior of the solutions to (1). For rather simple choices of the function L , the literature on this method indicates that large values of λ are seen to give rise to slowly converging solutions resembling solutions yielded by steepest descent [3]. The top plot¹ on the left in Figure 1 demonstrates such behavior. In contrast, smaller values of λ give rise to fast solutions with oscillations getting wilder as λ decreases [3]. Nesterov’s method converges

¹Code at github.com/HybridSystemsLab/UnitingMotivation

quickly but also suffers from oscillations [4], as the coefficient of the velocity term starts small and tends toward zero (but being always positive) as t tends to infinity. The oscillatory behavior of Nesterov’s method, with $\zeta = 2$, is shown in the middle plot on the left in Figure 1.

Due to its implications on robustness, we are particularly interested in an algorithm that assures UGAS of the minimizer of L with a rate of convergence that holds for all time, and without the undesired oscillations. As pointed out in Section 1.1, these properties are not guaranteed by Nesterov’s method. The behavior shown in the top and middle plots in Figure 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 the minimizer. More precisely, without knowledge of the location of the minimizer, it selects Nesterov’s method to converge quickly to nearby the minimizer and, once solutions reach a neighborhood of the minimizer, switches to the heavy ball method with large λ to avoid oscillations. Such logic-based algorithms, or *uniting algorithms*, were first proposed in [24] and [25]. General uniting algorithms, with examples, are discussed in [23] and [22]. We use the hybrid systems framework for our proposed algorithm, as hybrid systems utilize hysteresis to avoid chattering at the switching boundary; see [13], [22], [23], [21]. An example solution to our proposed logic-based algorithm, shown in the bottom plot on the left of Figure 1, demonstrates the improvement obtained by using Nesterov’s method globally and the heavy ball method locally, under relatively mild assumptions on the objective function 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 the minimizer:* In Section 2 we propose a uniting algorithm that solves optimization problems of the form $\min_{\xi \in \mathbb{R}^n} L(\xi)$ with accelerated gradient methods. Designed using hybrid system tools, the algorithm unites Nesterov’s method in (3) globally and the heavy ball method in (1) with large λ locally to guarantee fast convergence with UGAS of the minimizer ξ^* of a convex objective function L ; see Sections 2 and 4. The establishment of UGAS solves the difficult problem of achieving such a property for Nesterov-like

algorithms [19], [26]. The algorithm we propose exploits measurements of ∇L and requires no knowledge of $L^* := L(\xi^*)$ or ξ^* . In practice, such measurements of ∇L are typically approximated from measurements of L . The algorithm, however, does not require measurements of the Hessian of L .

- 2) *Well-posedness and existence of solutions:* Nesterov’s method can suffer from error accumulation, due to its velocity term [27]. To overcome this issue, in Section 2 we prove well-posedness and existence of solutions for the proposed hybrid closed-loop algorithm. Hybrid systems that are *well-posed* are defined to be those hybrid systems, vaguely speaking, for which graphical limits of graphically convergent sequences of solutions, with no perturbations and with vanishing perturbations, respectively, are still solutions [23, Chapter 6]. It is important for our algorithm to be well-posed as we want to ensure robustness to small noise in measurements of the gradient of L .
- 3) *Robustness to small perturbations:* Due to the well-posedness of the proposed hybrid uniting algorithm, we show that the established UGAS property is robust to small perturbations in measurements of the gradient of L [23, Theorem 7.21]. We illustrate this robustness in Section 3 via numerical simulations that include small noise in measurements of the gradient.
- 4) *A (hybrid) convergence rate preserving the rates of Nesterov’s method and heavy ball:* In Sections 2 and 4 we show that our uniting algorithm attains a rate of $\frac{1}{(t+2)^2}$ for the global algorithm and $\exp(-(1-m)\psi t)$, where $m \in (0, 1)$ and $\alpha > 0$ are such that $\psi := \frac{m\alpha\gamma}{\lambda} > 0$ and² $\nu := \psi(\psi - \lambda) < 0$, for the local algorithm. The latter rate holds under the mild assumption on L of quadratic growth away from the minimizer. As mentioned in Section 1.1, Nesterov-like algorithms do not necessarily assure UGAS of the minimizer. The HAND-1 algorithm for convex L , proposed in [19], provides UGAS via a hybrid restarting mechanism that yields a convergence rate $\frac{1}{t^2}$. However, this convergence rate holds only until the first reset. The algorithm we propose not only renders the minimizer UGAS, but also has a (hybrid) convergence rate that preserves the rates of the individual optimization algorithms for all (hybrid) time. Moreover, the global rate

²Although the constant ν does not appear in the rate for the local algorithm, such a constant is used in the forthcoming Proposition 4.3 to derive this rate.

of our algorithm is commensurate with that of HAND-1. In Figure 2 and Section 3, our uniting algorithm is shown via numerical simulations³ to have improved performance over the HAND-1 algorithm in [19].

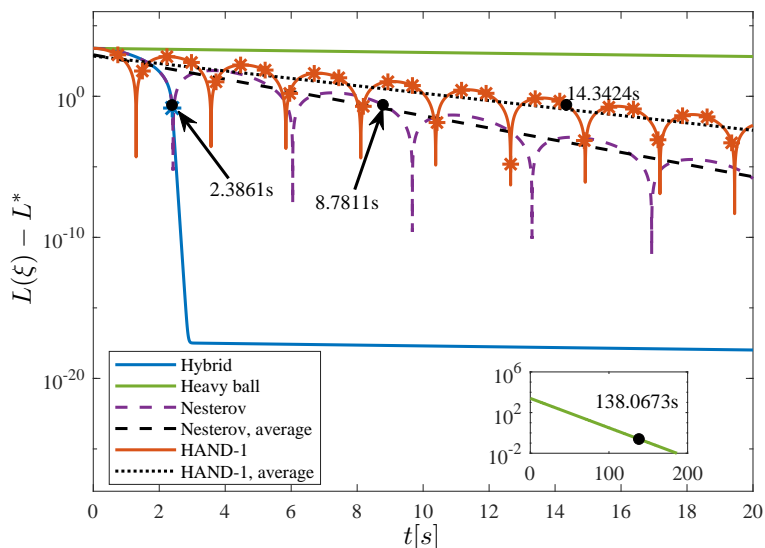


Figure 2: A comparison of the evolution of L over time for Nesterov’s method in (3), heavy ball, HAND-1 from [19], and our proposed uniting algorithm, for a function $L(\xi) := \xi^2$, with a single minimizer at $\xi^* = 0$. Nesterov’s method, shown in purple, settles to within 1% of ξ^* in about 8.8 seconds. The heavy ball algorithm, shown in green, settles to within 1% of ξ^* in about 138.1 seconds. HAND-1, shown in orange, settles to within 1% of ξ^* in about 14.3 seconds. The hybrid closed-loop system \mathcal{H} , shown in blue, settles to within 1% of z_1^* in about 2.4 seconds. As opposed to Figure 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 than in Figure 1.

- 5) *Extension of the results on Nesterov’s method in [1]*: In the process, in Section 4, we extend the properties and convergence results for Nesterov’s method in [1]. In particular, while the convergence rate results in [1] assume that $L(\xi_1^*) = 0$ at $\xi^* = 0$, and $\zeta = 1$ for (2), here we prove UGAS of the minimizer, with a convergence rate of $\frac{1}{(t+2)^2}$, for cost functions with a minimum value that is not necessarily zero, which holds for a generic

³Code at [gitHub.com/HybridSystemsLab/UnitingTradeoff](https://github.com/HybridSystemsLab/UnitingTradeoff)

parameter $\zeta > 0$. We achieve the relaxation on ζ by moving it into the numerator of the coefficient of the gradient, effectively decoupling ζ and M , namely,

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

Such a modification leads to faster convergence as ζ increases, and slower convergence as $\zeta \rightarrow 0$.

Preliminary work in [28] proposed an algorithm uniting Nesterov's method globally and heavy ball locally for \mathcal{C}^2 , strongly convex objective functions L , and included different results and examples that reflect such conditions, with proofs omitted due to space considerations. The uniting algorithm proposed in this paper relaxes the conditions in [28] to \mathcal{C}^1 , convex L with a unique minimizer. Such a relaxation is reflected in the results, examples, and proofs presented here.

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 vectors $v \in \mathbb{R}^n$ and $w \in \mathbb{R}^n$, $|v| = \sqrt{v^\top v}$ denotes the Euclidean vector norm of v , and $\langle v, w \rangle = v^\top w$ the inner product of v and w . 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} and the set of interior points of S is denoted $\text{int}(S)$. Given a set $S \subset \mathbb{R}^n \times \mathbb{R}^m$, the projection of S onto \mathbb{R}^n is defined as $\Pi(S) := \{x \in \mathbb{R}^n : \exists y \text{ such that } (x, y) \in S\}$. The distance of a point $x \in \mathbb{R}^n$ to a set $S \in \mathbb{R}^n$ is defined by $|x|_S = \inf_{y \in S} |y - x|$. Given a set-valued mapping $M : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$, the domain of M is the set $\text{dom}M = \{x \in \mathbb{R}^m : M(x) \neq \emptyset\}$, and the range of M is the set $\text{rge} M = \{y \in \mathbb{R}^n : \exists x \in \mathbb{R}^m \text{ such that } y \in M(x)\}$. A function $\alpha : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is a class- \mathcal{K}_∞ function, also written $\alpha \in \mathcal{K}_\infty$, if α is zero at zero, continuous, strictly increasing, and unbounded. A function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is a class- \mathcal{KL} function, also written $\beta \in \mathcal{KL}$, if it is nondecreasing in its first argument, nonincreasing in its second argument, $\lim_{r \rightarrow 0^+} \beta(r, s) = 0$ for each $s \in \mathbb{R}_{\geq 0}$, and $\lim_{s \rightarrow \infty} \beta(r, s) = 0$ for each $r \in \mathbb{R}_{\geq 0}$.

2 Uniting Optimization Algorithm

2.1 Problem Statement

As illustrated in Figure 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 [14] 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 (★): 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 (★). We interpret the ODEs in (1) and (3) as control systems consisting of a plant and a control algorithm [29] [22]. 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

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

The functions \bar{d} and $\bar{\beta}$ are defined expressly as in (5) for ease of analysis in the forthcoming Propositions 4.4-4.6. 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 [1]. 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

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

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

where the algorithm defined by κ_1 plays the role of the global algorithm in uniting control (see, e.g., [22, 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

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

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 [23, Definition 2.2]

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

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 by⁴ $(t, j) \in \mathbb{R}_{\geq 0} \times \mathbb{N}$, 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 $x := (z, q, \tau) \in \mathbb{R}^{2n} \times Q \times \mathbb{R}_{\geq 0}$ as⁵

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

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

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

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

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

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; see Figure 3. 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

⁴The variable t is the amount of time that has passed and j is the number of jumps that have occurred.

⁵Although h_0 in (7) does not depend on τ , to simplify notation we keep τ as an argument in the general function h_q .

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 complete algorithm, defined in (9)-(10), is summarized in Algorithm 1.

Algorithm 1 Uniting algorithm

- 1: Set $q(0,0)$ to 0, $\tau(0,0)$ to 0, and set $z(0,0)$ as an initial condition with an arbitrary value.
 - 2: **while** true **do**
 - 3: **if** $z \in \mathcal{T}_{0,1}$, $q = 0$, and $\tau = 0$ **then**
 - 4: Reset q to 1.
 - 5: **else if** $z \in \mathcal{T}_{1,0}$ and $q = 1$ **then**
 - 6: Reset q to 0 and τ to 0.
 - 7: **else if** $z \in \mathcal{U}_0$, $q = 0$, and $\tau = 0$ **then**
 - 8: Assign u to $\kappa_0(h_0(z))$ and update z , q , and τ according to (9a).
 - 9: **else if** $z \in \overline{\mathbb{R}^{2n}} \setminus \mathcal{T}_{1,0}$ and $q = 1$ **then**
 - 10: Assign u to $\kappa_1(h_1(z, \tau), \tau)$ and update z , q , and τ according to (9a).
 - 11: **end if**
 - 12: **end while**
-

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 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.

Figure 3 shows the feedback diagram of this hybrid closed-loop system \mathcal{H} . We denote the closed-loop system resulting from κ_0 as \mathcal{H}_0 , which is given by

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

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

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

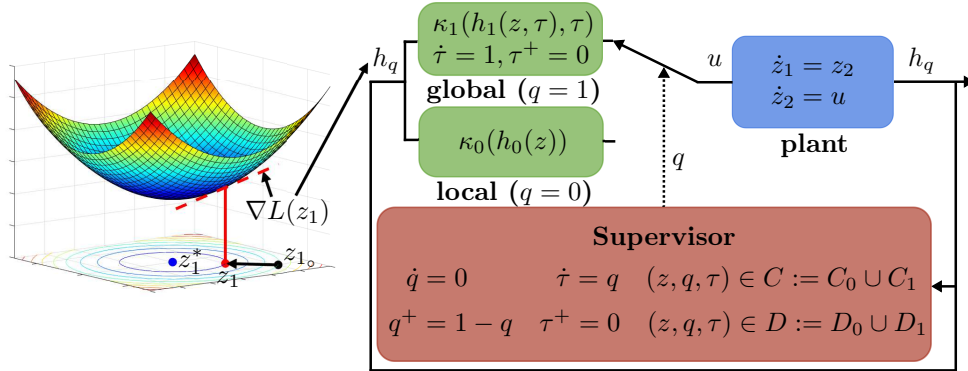


Figure 3: Feedback diagram of the hybrid closed-loop system \mathcal{H} (on the right), in (9), uniting global and local optimization algorithms. An example optimization problem to solve is shown on the left and, for this example optimization problem, measurements of the gradient are used for the input of κ_q .

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 2.1 *The function L is \mathcal{C}^1 , convex⁶, and has a single minimizer z_1^* .*

Assumption 2.2 (Quadratic growth of L) *The function L has quadratic growth away from its minimizer z_1^* ; i.e., there exists $\alpha > 0$ such that*

$$L(z_1) - L^* \geq \alpha |z_1 - z_1^*|^2 \quad \forall z_1 \in \mathbb{R}^n. \quad (13)$$

Remark 2.3 *Assumption 2.1, which is a common assumption used in the analysis of optimization algorithms [30] [31], ensures that the objective function is continuously differentiable, which is necessary for well-posedness of*

⁶A function $L : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if, for all $u_1, w_1 \in \mathbb{R}^n$, $L(u_1) \geq L(w_1) + \langle \nabla L(w_1), u_1 - w_1 \rangle$ [30].

\mathcal{H} , as was explained in Section 1.2. Additionally, the convex property and the restriction that L has a single minimizer z_1^* in Assumption 2.1 rules out the possibility of the objective function having a continuum of minimizers or multiple isolated minimizers. Assumption 2.2, which is used for the construction of \mathcal{U}_0 , $\mathcal{T}_{1,0}$, and $\mathcal{T}_{0,1}$, is employed as a means of determining when the state z is near the minimizer of L , via measurements of the gradient. Such an assumption is also commonly used in the analysis of convex optimization algorithms; see, e.g., [32], [33].

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

Assumption 2.4 (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 $w_1, u_1 \in \mathbb{R}^n$.

Remark 2.5 *Assumption 2.4 is used in the forthcoming construction of $\mathcal{T}_{0,1}$. Additionally, Assumption 2.4 is commonly used in nonlinear analysis to ensure that the differential equations of the individual optimization algorithms, for example, those in (3) and (1), do not have solutions that escape in finite time, which is used to guarantee existence and completeness of maximal solutions to \mathcal{H}_q [34, Theorem 3.2].*

Under Assumptions 2.1 and 2.2, 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^* .

Lemma 2.6 (Suboptimality): *Let L satisfy Assumptions 2.1 and 2.2, and let $\alpha > 0$ come from Assumption 2.2. For some $\varepsilon > 0$, if $z_1 \in \mathbb{R}^n$ is such that $|\nabla L(z_1)| \leq \varepsilon\alpha$, then $|z_1 - z_1^*| \leq \varepsilon$.*

Proof. Combining Assumption 2.1 and (13) from Assumption 2.2 with $u_1 = z_1^*$ and $w_1 = z_1$ yields

$$\alpha |z_1 - z_1^*|^2 \leq |L(z_1) - L^*| \leq |\langle \nabla L(z_1), z_1^* - z_1 \rangle| \leq |\nabla L(z_1)| |z_1 - z_1^*| \quad (14)$$

where the first inequality holds since $L(z_1) \geq L^*$. Then,

$$|z_1 - z_1^*| \leq \frac{1}{\alpha} |\nabla L(z_1)|. \quad (15)$$

From (15), we can deduce that $|\nabla L(z_1)| \leq \varepsilon\alpha$ implies $|z_1 - z_1^*| \leq \frac{1}{\alpha}(\varepsilon\alpha) = \varepsilon$. \square

The suboptimality condition from Lemma 2.6 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^* [30]. We exploit Lemma 2.6 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 ; see Figure 3.

2.3.1 Design of the Set \mathcal{U}_0

Recall from lines 7-8 of Algorithm 1 that 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 2.1 and 2.2 and the Lyapunov function

$$V_0(z) := \gamma(L(z_1) - L^*) + \frac{1}{2}|z_2|^2 \quad (16)$$

defined for each $z \in \mathbb{R}^{2n}$, where $\gamma > 0$. The choice of V_0 in (16) leads to \dot{V}_0 showing decrease of V_0 for each $z \in \mathbb{R}^{2n}$, in the proof of the forthcoming Proposition 4.1. Such a property of V_0 is needed 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 2.2 such that

$$\tilde{c}_0 := \varepsilon_0\alpha > 0, \quad d_0 := c_0 - \gamma\left(\frac{\tilde{c}_0^2}{\alpha}\right) > 0. \quad (17)$$

Then, V_0 in (16) can be upper bounded, using Assumption 2.1 as done to arrive to (14), as follows: for each $z \in \mathbb{R}^{2n}$

$$V_0(z) = \gamma(L(z_1) - L^*) + \frac{1}{2}|z_2|^2 \leq \gamma|\nabla L(z_1)||z_1 - z_1^*| + \frac{1}{2}|z_2|^2. \quad (18)$$

Then, due to L being \mathcal{C}^1 , convex, and having a single minimizer z_1^* by Assumption 2.1, and due to L having quadratic growth away from z_1^* by Assumption 2.2, when $|\nabla L(z_1)| \leq \tilde{c}_0$, the suboptimality condition in Lemma 2.6 implies $|z_1 - z_1^*| \leq \frac{\tilde{c}_0}{\alpha}$, from where we get

$$V_0(z) \leq \gamma\left(\frac{\tilde{c}_0^2}{\alpha}\right) + \frac{1}{2}|z_2|^2 \quad (19)$$

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

$$\mathcal{U}_0 := \left\{ z \in \mathbb{R}^{2n} : |\nabla L(z_1)| \leq \tilde{c}_0, \frac{1}{2} |z_2|^2 \leq d_0 \right\}, \quad (20)$$

every $z \in \mathcal{U}_0$ belongs to the c_0 -sublevel set of V_0 . In fact, using the conditions in (17) and (19), we have that for each $z \in \mathcal{U}_0$,

$$V_0(z) \leq \gamma \left(\frac{\tilde{c}_0^2}{\alpha} \right) + \frac{1}{2} |z_2|^2 \leq c_0. \quad (21)$$

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 4.1, the set \mathcal{U}_0 is contained in the basin of attraction induced by κ_0 .

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

Recall from lines 5-6 of Algorithm 1 that 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 2.1 and 2.2 and the Lyapunov function

$$V_1(z, \tau) := \frac{1}{2} |\bar{a}(\tau) (z_1 - z_1^*) + z_2|^2 + \frac{\zeta^2}{M} (L(z_1) - L^*) \quad (22)$$

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

$$\bar{a}(\tau) := \frac{2}{\tau + 2}. \quad (23)$$

The choice of V_1 in (22) comes from [1], and such a choice leads to \dot{V}_1 showing decrease of V_1 , for each $z \in \mathbb{R}^{2n}$ and each $\tau \in \mathbb{R}_{\geq 0}$, in the proof of the forthcoming Proposition 4.4. Such a property of V_1 is needed 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 (23), which comes from [1], leads to the elimination of the cross term $\langle z_1 - z_1^*, z_2 \rangle$ – which has indeterminate sign – in the upper bound on \dot{V}_1 . In other words, without the definitions of V_1 in (22) and \bar{a} in (23), decrease of V_1 and, consequently, UGAS of the minimizer

for \mathcal{H}_1 and the convergence rate $\frac{1}{(t+2)^2}$ cannot be established. More details on how (22) and (23) are used in our analysis can be found in Appendix C. 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 (17), and let $\alpha > 0$ come from Assumption 2.2 such that

$$\tilde{c}_{1,0} := \varepsilon_{1,0}\alpha \in (0, \tilde{c}_0) \quad (24a)$$

$$d_{1,0} := c_{1,0} - \left(\frac{\tilde{c}_{1,0}}{\alpha}\right)^2 - \frac{\zeta^2}{M} \left(\frac{\tilde{c}_{1,0}^2}{\alpha}\right) \in (0, d_0) \quad (24b)$$

where $\zeta > 0$ comes from (3). Note that \bar{a} , defined via (23), 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 (22) and using Assumption 2.1 with $u_1 = z_1^*$ and $w_1 = z_1$,

$$V_1(z, \tau) \leq |z_1 - z_1^*|^2 + |z_2|^2 + \frac{\zeta^2}{M} |\nabla L(z_1)| |z_1 - z_1^*|. \quad (25)$$

Then, due to L being \mathcal{C}^1 , convex, and having a single minimizer z_1^* by Assumption 2.1, and due to L having quadratic growth away from z_1^* by Assumption 2.2, when $|\nabla L(z_1)| \leq \tilde{c}_{1,0}$, the suboptimality condition in Lemma 2.6 implies $|z_1 - z_1^*| \leq \frac{\tilde{c}_{1,0}}{\alpha}$, from where we get

$$V_1(z, \tau) \leq \left(\frac{\tilde{c}_{1,0}}{\alpha}\right)^2 + |z_2|^2 + \frac{\zeta^2}{M} \left(\frac{\tilde{c}_{1,0}^2}{\alpha}\right). \quad (26)$$

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

$$\mathcal{T}_{1,0} := \{z \in \mathbb{R}^{2n} : |\nabla L(z_1)| \leq \tilde{c}_{1,0}, |z_2|^2 \leq d_{1,0}\} \quad (27)$$

which, by construction, is contained in the interior of \mathcal{U}_0 defined in (20), every $z \in \mathcal{T}_{1,0}$ belongs to the $c_{1,0}$ -sublevel set of V_1 . In fact, using the conditions in (24) and (26), we have for each $z \in \mathcal{T}_{1,0}$,

$$V_1(z, \tau) \leq \left(\frac{\tilde{c}_{1,0}}{\alpha}\right)^2 + |z_2|^2 + \frac{\zeta^2}{M} \left(\frac{\tilde{c}_{1,0}^2}{\alpha}\right) \leq c_{1,0}. \quad (28)$$

The constants \tilde{c}_0 , $\tilde{c}_{1,0}$, d_0 , and $d_{1,0}$ in (17) and (24) 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 $\mathcal{T}_{0,1} \cap \mathcal{T}_{1,0} = \emptyset$. Figure 4 illustrates the hysteresis mechanism in the design of \mathcal{U}_0 and $\mathcal{T}_{1,0}$.

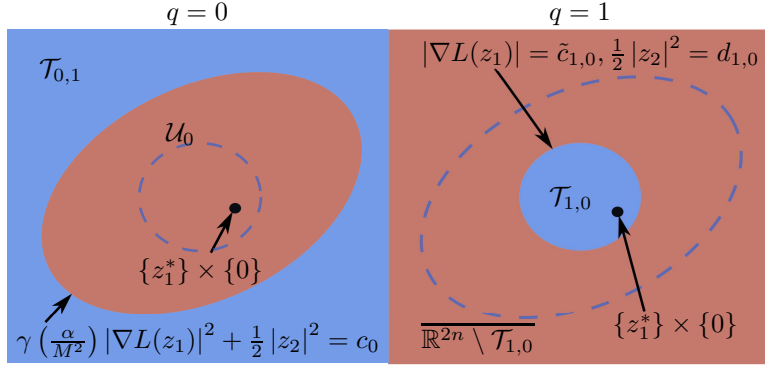


Figure 4: An illustration of hysteresis in the design of the sets \mathcal{U}_0 , $\mathcal{T}_{1,0}$, and $\mathcal{T}_{0,1}$ on \mathbb{R}^{2n} , via the constants $\tilde{c}_{1,0} \in (0, \tilde{c}_0)$, $d_{1,0} \in (0, d_0)$, and $c_0 > 0$. Left: due to the design of \mathcal{U}_0 in (20), every $z \in \mathcal{U}_0$ belongs to the c_0 -sublevel set of the Lyapunov function V_0 , where V_0 is defined via (16). 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 (24), are chosen such that the set $\mathcal{T}_{1,0}$ in (27) is contained in the interior of \mathcal{U}_0 .

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

Recall from lines 3-4 of Algorithm 1 that 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 2.2, and let $M > 0$ come from Assumption 2.4. Then, using Assumption 2.4 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 2.2, then dividing both sides of $|\nabla L(z_1)| \leq M |z_1 - z_1^*|$ by M and substituting into (13) leads to $L(z_1) - L^* \geq \frac{\alpha}{M^2} |\nabla L(z_1)|^2$, where $\alpha > 0$ comes from Assumption 2.2.

Then, V_0 in (16) is lower bounded as follows: for each $z \in \mathbb{R}^{2n}$,

$$V_0(z) = \gamma (L(z_1) - L^*) + \frac{1}{2} |z_2|^2 \geq \gamma \left(\frac{\alpha}{M^2} \right) |\nabla L(z_1)|^2 + \frac{1}{2} |z_2|^2. \quad (29)$$

Using the right-hand side of (29) and the same $c_0 > 0$ as in Section 2.3.1, we define the set

$$\mathcal{T}_{0,1} := \left\{ z \in \mathbb{R}^{2n} : \gamma \left(\frac{\alpha}{M^2} \right) |\nabla L(z_1)|^2 + \frac{1}{2} |z_2|^2 \geq c_0 \right\}. \quad (30)$$

The set in (30) defines the (closed) complement of a sublevel set of the Lyapunov function V_0 in (16) with level equal to c_0 . The constant c_0 is also a part of the hysteresis mechanism, as shown in Figure 4. 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 (30) 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 2.11 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 [3]. 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 Examples 3.1, 3.2, and 3.7 where λ was tuned in such a way.

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

When L satisfies Assumptions 2.1, 2.2, and 2.4, the hybrid closed-loop system \mathcal{H} in (9) satisfies the hybrid basic conditions from [23] and [22], defined as follows.

Definition 2.7 (Hybrid basic conditions) *A hybrid system \mathcal{H} is said to satisfy the hybrid basic conditions if its data (C, F, D, G) is such that*

- (A1) C and D are closed subsets of \mathbb{R}^n ;
- (A2) $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is outer semicontinuous and locally bounded relative to C , $C \subset \text{dom } F$, and $F(x)$ is convex for every $x \in C$;
- (A3) $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is outer semicontinuous and locally bounded relative to D , and $D \subset \text{dom } G$.

The hybrid basic conditions are defined in such a manner as to ensure that a hybrid system satisfying such conditions does not have any discontinuities. In particular, for a discontinuous system, if the state starts close to one of the points of a discontinuity and small measurement noise is present, the solution remains nearby such a point, even when the noise is arbitrarily small. The limit of such a solution as the noise goes to zero is a solution to a differential inclusion (or a difference inclusion, or a hybrid inclusion) which is a *Krasovskii regularization* of the discontinuous system. Such a solution, when the right-hand side of the Krasovskii regularization is bounded, is also a *Hermes solution*, and represents an equilibrium point of the discontinuous system, from which the state cannot converge to the set of interest. Conversely, when a hybrid system satisfies the hybrid basic conditions of closed sets C and D and outer semicontinuous maps F and G , this ensures that any existing stability properties of such a hybrid system are robust to small perturbations. See [23] and [22] for more details. The satisfaction of such conditions is demonstrated in the following lemma. A hybrid closed-loop system \mathcal{H} that satisfies the hybrid basic conditions is said to be well-posed in the sense that the limit of a graphically convergent sequence of solutions to \mathcal{H} having a mild boundedness property is also a solution to \mathcal{H} [23].

Lemma 2.8 (*Well-posedness of \mathcal{H}*): *Let the function L satisfy Assumptions 2.1, 2.2, and 2.4. Let the sets \mathcal{U}_0 , $\mathcal{T}_{1,0}$, and $\mathcal{T}_{0,1}$ be defined via (27), and (30), 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.*

Proof. See Section A. □

In Theorem 2.11 we show that \mathcal{H} has a compact pre-asymptotically stable set. In light of this property, Lemma 2.8 is key as it leads to pre-asymptotic stability that is robust to small perturbations [23, 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.

2.6 Existence of solutions to \mathcal{H}

Under Assumptions 2.1, 2.2, and 2.4, every maximal solution to \mathcal{H} is complete⁷ and bounded, as stated in the following lemma. Such a property is useful since it guarantees that nontrivial solutions to \mathcal{H} exist from each initial point in $C \cup D$, and that such solutions do not escape $C \cup D$. When every maximal solution is complete, then uniform global pre-asymptotic stability⁸ of the set \mathcal{A} becomes UGAS. The following lemma also states that $\Pi(C_0) \cup \Pi(D_0) = \mathbb{R}^{2n}$ and $\Pi(C_1) \cup \Pi(D_1) = \mathbb{R}^{2n}$. Such a property ensures that nontrivial solutions to \mathcal{H} , which exist from each initial point in $C \cup D$, also exist from any initial point in $\mathbb{R}^{2n} \times Q \times \mathbb{R}_{\geq 0}$.

Proposition 2.9 (*Existence of solutions to \mathcal{H}*): *Let the function L satisfy Assumptions 2.1, 2.2, and 2.4. Let the sets \mathcal{U}_0 , $\mathcal{T}_{1,0}$, and $\mathcal{T}_{0,1}$ be defined via (27), and (30), 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, $\Pi(C_0) \cup \Pi(D_0) = \mathbb{R}^{2n}$, $\Pi(C_1) \cup \Pi(D_1) = \mathbb{R}^{2n}$, and each maximal solution $(t, j) \mapsto x(t, j) = (z(t, j), q(t, j), \tau(t, j))$ to \mathcal{H} in (9) is bounded and complete.*

Proof. See Section B. □

⁷A solution x to \mathcal{H} is called maximal if it cannot be extended further. A solution is called complete if its domain is unbounded.

⁸Uniform global pre-asymptotic stability indicates the possibility of a maximal solution that is not complete, even though it may be bounded.

2.7 Main Result

The result in this section depends on the notion of stability, uniform global stability, pre-attractivity, uniform global pre-attractivity, and uniform global pre-asymptotic stability (UGpAS) which are listed in the following definition, from [22] and [23].

Definition 2.10 (Stability and attractivity notions) *Given a hybrid closed-loop system \mathcal{H} as in (8), a nonempty set $\mathcal{A} \subset \mathbb{R}^n$ is said to be*

- Stable for \mathcal{H} if for each $\varepsilon > 0$ there exists $\delta > 0$ such that each solution x to \mathcal{H} with $|x(0, 0)|_{\mathcal{A}} \leq \delta$ satisfies $|x(t, j)|_{\mathcal{A}} \leq \varepsilon$ for all⁹ $(t, j) \in \text{dom } x$;
- Uniformly globally stable for \mathcal{H} if there exists a class- \mathcal{K}_∞ 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 $(t, j) \in \text{dom } x$;
- Pre-attractive for \mathcal{H} if there exists $\mu > 0$ such that every solution x to \mathcal{H} with $|x(0, 0)|_{\mathcal{A}} \leq \mu$ is such that $(t, j) \mapsto |x(t, j)|_{\mathcal{A}}$ is bounded and if x is complete then $\lim_{(t, j) \in \text{dom } x, t+j \rightarrow \infty} |x(t, j)|_{\mathcal{A}} = 0$;
- 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 $|x(0, 0)|_{\mathcal{A}} \leq \delta$, $(t, j) \in \text{dom } x$ and $t + j \leq T$ imply $|x(t, j)|_{\mathcal{A}} \leq \varepsilon$;
- Uniformly globally pre-asymptotically stable (UGpAS) for \mathcal{H} if it is both uniformly globally stable and uniformly globally pre-attractive.

In the notions involving convergence in Definition 2.10, when every maximal solution is complete, then the prefix “pre” is dropped to obtain attractivity, UGA, and UGAS. The prefix “pre” is in the notions involving convergence in Definition 2.10 to allow for maximal solutions that are not complete. When every maximal solution is complete, such a property guarantees that nontrivial solutions exist from each initial point in $C \cup D$ to the hybrid system resulting from using our proposed uniting algorithm.

As was mentioned in Section 1.1, establishing UGAS for Nesterov’s algorithm is a difficult problem to solve, due to its time-varying nature, as some

⁹The domain of x , namely, $\text{dom } x \subset \mathbb{R}_{\geq 0} \times \mathbb{N}$, is a hybrid time domain, which is a set such that for each $(T, J) \in \text{dom } x$, $\text{dom } x \cap ([0, T] \times \{0, 1, \dots, J\}) = \cup_{j=0}^J ([t_j, t_{j+1}], j)$ for a finite sequence of times $0 = t_0 \leq t_1 \leq t_2 \leq \dots \leq t_{J+1}$.

solutions converge in a non-uniform way. We show in this section that our proposed uniting algorithm overcomes such a difficulty.

In this section, we present a result that establishes UGAS of the set

$$\mathcal{A} := \{z \in \mathbb{R}^{2n} : \nabla L(z_1) = z_2 = 0\} \times \{0\} \times \{0\} = \{z_1^*\} \times \{0\} \times \{0\} \times \{0\} \quad (31)$$

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 2.11 (*UGAS of \mathcal{A} for \mathcal{H}*): *Let the function L satisfy Assumptions 2.1, 2.2, and 2.4. 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 2.2, and let $M > 0$ be generated by Assumption 2.4. Let $\tilde{c}_{1,0} \in (0, \tilde{c}_0)$ and $d_{1,0} \in (0, d_0)$ be defined via (17) and (24). Let the sets \mathcal{U}_0 , $\mathcal{T}_{1,0}$, and $\mathcal{T}_{0,1}$ be defined via (27), and (30), 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 (31), 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 $\text{dom } x$ of the solution x is of the form¹⁰ $\cup_{j=0}^1 (I^j \times \{j\})$, with I^0 of the form $[t_0, t_1]$ and with I^1 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 that¹¹ $t \geq 0$*

$$L(z_1(t, 0)) - L^* \leq \frac{4cM}{\zeta^2(t+2)^2} (|z_1(0, 0) - z_1^*|^2 + |z_2(0, 0)|^2) \quad (32)$$

¹⁰We define the interval $I^j := \{t : (t, j) \in \text{dom } x\}$.

¹¹Note that at each $t \in I^0$, $q(t, 0) = 1$, and at each $t \in I^1$, $q(t, 1) = 0$.

where $L^* = L(z^*)$ and $c := (1 + \zeta^2) \exp\left(\sqrt{\frac{13}{4} + \frac{\zeta^4}{M}}\right)$. 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 $\mathcal{O}(\exp(-(1-m)\psi t))$, 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 2.11 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 2.11 is proved as follows. First, in the forthcoming Proposition 4.1, 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 4.6, 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 2.11 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 2.11 is proved in the forthcoming Propositions 4.3, 4.4, and 4.5.

3 Numerical Examples

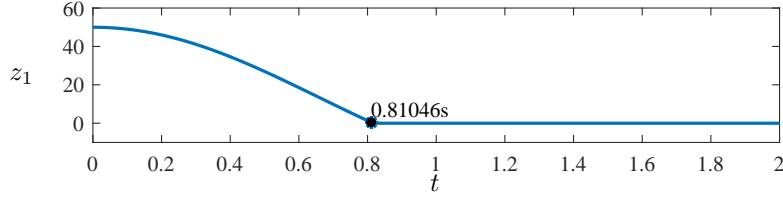
In this section, we present multiple numerical examples to illustrate the hybrid closed-loop algorithm in (9) and (10). Example 3.1 first illustrates the operation of the nominal hybrid closed-loop system \mathcal{H} , and then demonstrates the robustness of \mathcal{H} to different amounts of noise in measurements of ∇L . Example 3.2 compares solutions to the hybrid closed-loop algorithm in (9) and (10) with solutions to \mathcal{H}_0 , \mathcal{H}_1 , and HAND-1 from [19], with parameters chosen such that HAND-1 and \mathcal{H} are compared on equal footing. Example 3.2 then compares multiple solutions of \mathcal{H} , starting from different initial values of z_1 , to multiple solutions of HAND-1 from such initial values of z_1 , to show that \mathcal{H} has a consistent percentage of improvement over HAND-1 for different solutions. Example 3.3 compares solutions to the hybrid closed-loop algorithm in (9) and (10) with solutions to \mathcal{H}_0 , \mathcal{H}_1 , HAND-1 from [19], and the hybrid Hamiltonian algorithm (HHA) from [21], with parameters chosen such that HAND-1, HHA, and \mathcal{H} are compared on equal

footing. Example 3.7 illustrates the trade-off between speed of convergence and the resulting values of parameters for the uniting algorithm \mathcal{H} , for different tunings of $\zeta > 0$. As in Example 3.2, the parameter values for Example 3.7 are chosen such that HAND-1 and \mathcal{H} are compared on equal footing.

Example 3.1 *In this example, we simulate a solution to the nominal hybrid closed-loop system \mathcal{H} to illustrate how the uniting algorithm works. Then, we compare that same solution to solutions with different amounts of noise in measurements of ∇L . For both the nominal system and the perturbed system, 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 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. For Nesterov’s algorithm, we chose $\zeta = 2$. 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. The parameter values for the uniting algorithm are $c_0 = 7000$, $c_{1,0} \approx 6819.68$, $\varepsilon_0 = 10$, $\varepsilon_{1,0} = 5$, and $\alpha = 1$, which yield the values $\tilde{c}_0 = 10$, $\tilde{c}_{1,0} = 5$, $d_0 = 6933$, and $d_{1,0} = 6744$, which are calculated via (17) and (24). These values are chosen for proper tuning of the algorithm, in order to get nice performance, and the value of $c_{1,0}$ is chosen to exploit the properties of Nesterov’s method for a longer time, so that the nominal solution gets closer to the minimizer faster. Initial conditions for \mathcal{H} are $z_1(0,0) = 50$, $z_2(0,0) = 0$, $q(0,0) = 1$, and $\tau(0,0) = 0$. The plot on the top in Figure 5 shows the solution to the nominal hybrid closed-loop algorithm¹² \mathcal{H} , namely, the value of z_1 over time, with the time it takes for the solution to settle to within 1% of z_1^* marked with a black dot and labeled in seconds. The jump at which the switch from \mathcal{H}_1 to \mathcal{H}_0 occurs is labeled with an asterisk. The solution converges quickly, without oscillations near the minimizer.*

To show that the UGAS of \mathcal{A} , established in Theorem 2.11, is robust to small perturbations, due to the hybrid closed-loop system \mathcal{H} satisfying the hybrid basic conditions by Lemma 2.8. we simulate the hybrid algorithm, using the objective function, parameter values, and initial conditions listed in

¹²Code at github.com/HybridSystemsLab/UnitingRobustness



σ	$\limsup_{t+j \rightarrow \infty} z_1(t, j) - z_1^* $	$\limsup_{t+j \rightarrow \infty} L(z_1(t, j)) - L^* $
0.01	8.857×10^{-6}	7.844×10^{-11}
0.1	8.011×10^{-4}	6.418×10^{-7}
0.5	9.039×10^{-4}	8.171×10^{-7}
1	6.982×10^{-3}	4.875×10^{-5}
5	9.459×10^{-3}	8.947×10^{-5}
10	1.450×10^{-2}	2.103×10^{-4}
15	4.938×10^{-2}	2.438×10^{-3}
20	5.992×10^{-2}	3.591×10^{-3}
25	6.663×10^{-2}	4.439×10^{-3}

Figure 5: Top: The evolution over time of z_1 , for the nominal hybrid closed-loop algorithm \mathcal{H} , for a function $L(z_1) := z_1^2$ with a single minimizer at $z_1^* = 0$. The time at which the solution settles to within 1% of z_1^* is marked with a dot and labeled in seconds. The jump is labeled with an asterisk. Bottom: Simulation results for perturbed solutions using zero mean Gaussian noise, with each simulation using a different value of the standard deviation σ . Results listed are for a large value of $t + j$.

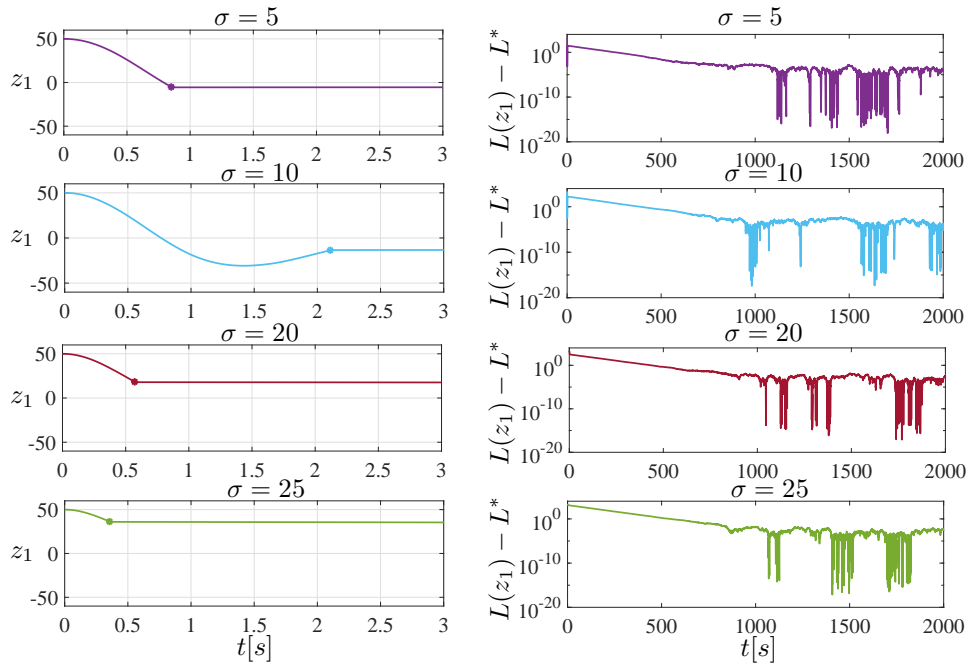


Figure 6: Simulation results for hybrid closed-loop algorithm \mathcal{H} , for a function $L(z_1) := z_1^2$ with a single minimizer at $z_1^* = 0$, with zero-mean Gaussian noise added to measurements of the gradient. Each subplot is labeled with the standard deviation used. Left subplots: the value of z_1 over time for each perturbed solution, with the jump in each solution labeled by an asterisk. Right subplots: the corresponding value of L over time for each perturbed solution.

the first paragraph of this example, with zero-mean Gaussian noise added to measurements of the gradient. Separate simulations were run for each of the following standard deviations: $\sigma \in \{0.01, 0.1, 0.5, 1, 5, 10, 15, 20, 25\}$. Figure 6 shows some of these perturbed solutions, with each subplot labeled with the corresponding standard deviation used¹³. The subplots on the left side of Figure 6 show the value of z_1 over time for different standard deviations, and the subplots on the right side of Figure 6 show the corresponding value of L over time for such standard deviations. Note that, while all perturbed solutions shown in Figure 6 get close to the minimizer quickly, such perturbed solutions do not get as close to the minimizer as the solution to the nominal algorithm does; see the plot on the top in Figure 5. Also note that as the standard deviation gets larger, the corresponding perturbed solution stays slightly farther away from the minimizer. The results for all standard deviations are listed in the table in Figure 5, showing the neighborhood of z_1^* that each solution settles to, for a large value of $t+j$, along with the corresponding value of L .

Example 3.2 *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 [19] which, in [19], is designed and analyzed for convex functions L satisfying Assumptions 2.1 and 2.4. First, we compare the convergence rates of \mathcal{H} and HAND-1 analytically. Using an alternate state space representation, namely, $z_1 := \xi$ and $z_2 := \xi + \frac{\tau}{2}\dot{\xi}$, the HAND-1 algorithm has state $(z, \tau) \in \mathbb{R}^{2n+1}$ and data (C, F, D, G)*

$$F(z, \tau) := \begin{bmatrix} \frac{z}{\tau}(z_2 - z_1) \\ -2c_1\tau\nabla L(z_1) \\ 1 \end{bmatrix} \quad (z, \tau) \in C, \quad G(z, \tau) := \begin{bmatrix} z \\ T_{\min} \end{bmatrix} \quad (z, \tau) \in D \quad (33)$$

where $c_1 > 0$ and the flow and jump sets are

$C := \{(z, \tau) \in \mathbb{R}^{2n+1} : \tau \in [T_{\min}, T_{\max}]\}$ and

$D := \{(z, \tau) \in \mathbb{R}^{2n+1} : \tau \in [T_{\text{med}}, T_{\max}]\}$, with $0 < T_{\min} < T_{\text{med}} < T_{\max} < \infty$,

and $T_{\text{med}} \geq \sqrt{\frac{B}{\delta_{\text{med}}}} + T_{\min} > 0$, $\delta_{\text{med}} > 0$. It is shown in [19] that each maximal

¹³Code found at same link as in Footnote 12.

solution $(t, j) \mapsto (z(t, j), \tau(t, j))$ to the HAND-1 algorithm satisfies

$$L(z_1(t, 0)) - L^* \leq \frac{B}{t^2} \quad (34)$$

for all $(t, j) \in \text{dom}(z, \tau)$ 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 $B := \frac{r^2}{2c_1} + T_{\min}^2 (L(z_1(0, 0)) - L^*) > 0$, $r \in \mathbb{R}_{>0}$, $c_1 > 0$.

For the hybrid closed-loop algorithm \mathcal{H} , the coefficient of the bound on \mathcal{H}_1 from (32), namely,

$$L(z_1(t, 0)) - L^* \leq \frac{4cM}{\zeta^2(t+2)^2} (|z_1(0, 0) - z_1^*|^2 + |z_2(0, 0)|^2) \quad (35)$$

for each $t \in I^0$, $t \geq 0$, at which $q(t, 0) = 1$, and for each $\zeta > 0$, and $M > 0$, is $\frac{4cM}{\zeta^2} (|z_1(0, 0) - z_1^*|^2 + |z_2(0, 0)|^2)$, where $c := (1 + \zeta^2) \exp\left(\sqrt{\frac{13}{4} + \frac{\zeta^4}{M}}\right)$. The coefficient of the bound in HAND-1 is $B := \frac{r^2}{2c_1} + T_{\min}^2 (L(z_1(0, 0)) - L^*)$. Since, as $t \rightarrow \infty$, $\frac{1}{(t+2)^2} \rightarrow \frac{1}{t^2}$, then, comparing the coefficients of the bounds, the bound in (35) is slightly better than (34) since $\frac{r^2}{2c_1}$ is very large for small t . Neglecting the $\frac{r^2}{2c_1}$ term, however, the bound on \mathcal{H}_1 (35) matches (34). The rate for HAND-1, nevertheless, is only guaranteed until the first jump. After this, there is no characterized bound for HAND-1. In contrast, \mathcal{H} has a characterized bound for the domain of every solution such that $t \geq 0$. Namely, it has rate $\frac{1}{(t+2)^2}$ until the state z is within a small neighborhood of the minimizer – where the rate then switches to $\exp(-(1-m)\psi t)$, where, given $\gamma > 0$ and $\lambda > 0$, $m \in (0, 1)$ is such that $\psi = \frac{m\alpha\gamma}{\lambda} > 0$ and $\nu = \psi(\psi - \lambda) < 0$.

Next, we compare \mathcal{H}_0 , \mathcal{H}_1 , \mathcal{H} , and HAND-1 in simulation. To compare these algorithms, we use the same objective function L , heavy ball parameter values λ and γ , Lipschitz parameter M , Nesterov parameter ζ , and uniting algorithm parameter values c_0 , $c_{1,0}$, ε_0 , $\varepsilon_{1,0}$, α , \tilde{c}_0 , $\tilde{c}_{1,0}$, d_0 , and $d_{1,0}$ as in Example 3.1. 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 footing¹⁴. The remaining HAND-1

¹⁴Although 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.

parameters, r and δ_{med} , have different values depending on the initial conditions $z_1(0,0) = z_2(0,0)$, listed in 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 Table 2, which leads to different values of d_0 , calculated via (17), and $d_{1,0}$ calculated via (24). 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 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 Table 2.

Algorithm	Average time to converge (s)	Average % improvement
\mathcal{H}	0.811	–
\mathcal{H}_0	690.759	99.9
\mathcal{H}_1	4.409	81.6
HAND-1	8.649	90.6

Table 1: Average times for which \mathcal{H} , \mathcal{H}_0 , \mathcal{H}_1 , and HAND-1 settle to within 1% of z_1^* , and the average percent improvement of \mathcal{H} over each algorithm. Percent improvement is calculated via (36). The objective function used for this table is $L(z_1) := z_1^2$.

Table 1 shows the time that each algorithm takes to settle within 1% of z_1^* , averaged over solutions starting from ten different values¹⁵ of $z_1(0,0)$ (listed in the first column of Table 2), and the average percent improvement of \mathcal{H} over \mathcal{H}_0 , \mathcal{H}_1 , and HAND-1, which is calculated using the following formula

$$\left(\frac{(\text{Time of } \mathcal{H}_0, \mathcal{H}_1, \text{ or HAND-1}) - \text{Time of } \mathcal{H}}{\text{Time of } \mathcal{H}_0, \mathcal{H}_1, \text{ or HAND-1}} \right) \times 100\%. \quad (36)$$

As can be seen in Table 1, \mathcal{H} converges faster than the other algorithms, and

¹⁵Code at github.com/HybridSystemsLab/UnitingDifferentICs

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

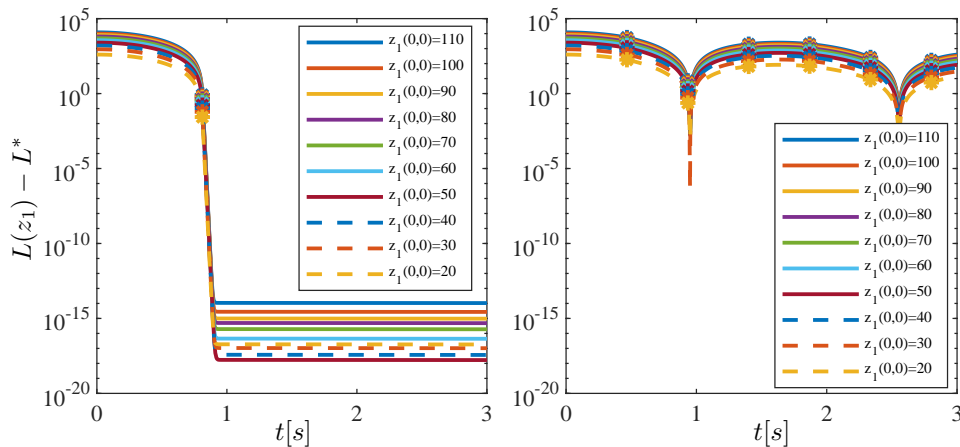


Figure 7: The evolution of L over time, from different initial conditions, for \mathcal{H} (left) and HAND-1 (right). All solutions are for the objective function $L(z_1) := z_1^2$, and the parameters used for HAND-1 and \mathcal{H} are listed in Table 2, with different values of c_0 and $c_{1,0}$ for each solution of \mathcal{H} , leading to different values of d_0 calculated via (17) and $d_{1,0}$ calculated via (24), 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.

Figure 7 compares different solutions for \mathcal{H} and HAND-1, from different values of $z_1(0,0)$, for the objective function $L(z_1) := z_1^2$. 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 Figure 7 and in 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.

The bound for HAND-1, shown in (34) and which holds only until the first reset, is only guaranteed when $z_1(0,0) = z_2(0,0)$. This leads to a required nonzero velocity for HAND-1 in most scenarios, which leads to overshoot. In contrast, \mathcal{H} has no such constraint on $z_2(0,0)$, which can be set to zero in all scenarios. The lack of such a constraint on the initial condition $z_2(0,0)$ for the hybrid closed-loop algorithm \mathcal{H} is essential to its improved performance over HAND-1, as the overshoot in solutions to HAND-1 due to $z_1(0,0) =$

$z_1(0, 0)$	c_0	$c_{1,0}$	r	δ_{med}	Time to converge (s)		% Improve- ment
					\mathcal{H}	HAND-1	
110	34000	32719.231	111	240700	0.811	8.649	90.6
100	28000	27053.704	101	199000	0.811	8.65	90.6
90	23000	21927.75	91	161300	0.811	8.648	90.6
80	18000	17341.37	81	127550	0.811	8.65	90.6
70	14000	13294.565	71	97700	0.811	8.649	90.6
60	10500	9787.333	61	71875	0.811	8.648	90.6
50	7000	6819.676	51	50000	0.810	8.65	90.6
40	5000	4391.593	41	32075	0.811	8.65	90.6
30	3000	2503.083	31	18110	0.811	8.648	90.6
20	2000	1154.148	21	8112	0.811	8.648	90.6

Table 2: Times for which \mathcal{H} and HAND-1 settle to within 1% of z_1^* , and percent improvement of \mathcal{H} over HAND-1, for solutions from different initial conditions, shown in Figure 7. The objective function used for this table is $L(z_1) := z_1^2$.

$z_2(0, 0)$ leads to a slower convergence time than for \mathcal{H} , as seen in Table 1. Moreover, as described previously in this example, no bound for HAND-1 is characterized after the first reset, whereas the (hybrid) convergence bound characterized for \mathcal{H} holds for the domain of every solution such that $t \geq 1$.

Example 3.3 *In this example, 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 , the HAND-1 algorithm from [19], and the hybrid Hamiltonian algorithm (HHA) from [21]. In [21], the HHA algorithm is designed and analyzed for objective functions L which satisfy the following assumptions:*

Assumption 3.4

- (L1) *The function L is \mathcal{C}^1 ;*
- (L2) *The function L has compact sublevel sets;*
- (L3) *∇L is Lipschitz continuous with constant $M \in (0, \bar{M}]$.*

First, we compare the convergence rates of \mathcal{H} and HHA analytically. Letting $H : \mathbb{R}^{2n} \rightarrow \mathbb{R}_{\geq 0}$ denote the separable Hamiltonian

$$H(z) := L(z_1^*) - L^* + \frac{1}{2} |z_2|^2 \quad (37)$$

and letting J and R_0 be defined as

$$J := \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}, R_0 := \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix} \quad (38)$$

the HHA algorithm has state $(z, \tau) \in \mathbb{R}^{2n+1}$ and data (C, F, D, G)

$$F(z, \tau) := \begin{bmatrix} J\nabla H(z) \\ 1 \end{bmatrix} = \begin{bmatrix} z_2 \\ -\nabla L(z) \\ 1 \end{bmatrix} \quad (z, \tau) \in C \quad (39a)$$

$$G(z, \tau) := \begin{bmatrix} R_0 z \\ 0 \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ 0 \end{bmatrix} \quad (z, \tau) \in D \quad (39b)$$

where the flow and jump sets are

$$C := C_0 \times [0, \bar{T}] \quad (40a)$$

$$D := (C_0 \times \{\bar{T}\}) \cup (D_0 \times [0, \bar{T}]) \quad (40b)$$

where $C_0 := \{z \in \mathbb{R}^{2n} : \langle \nabla L(z_1), z_2 \rangle \leq 0\}$, $D_0 := \{z \in \mathbb{R}^{2n} : \langle \nabla L(z_1), z_2 \rangle = 0, |z_2|^2 \geq |\nabla L(z_1)|^2 / \bar{M}\}$, and $\bar{T} \in (0, \infty)$ is the timeout parameter of the timer τ . The following assumption is imposed on HHA in (39)-(40):

Assumption 3.5 No solution to the flow dynamics starting from $(z_1(0, 0), 0, 0)$ with $\nabla L(z_1) \neq 0$ causes the timer to timeout, i.e., reaches $(C_0 \setminus D_0) \times \{\bar{T}\}$.

It is shown in [21] that Assumption 3.5 holds for quadratic functions L with $A = A^\top > 0$, when $\bar{T} := \frac{n\pi}{2\sqrt{\lambda_{\min}(A)}}$, where $\lambda_{\min}(A)$ denotes the minimum eigenvalue of A .

The forthcoming convergence rate results in [21] hold when L satisfies the following assumption:

Assumption 3.6 *The function L satisfies the Polyak-Łojasiewicz inequality with $\varsigma > 0$, namely,*

$$|\nabla L(z_1)|^2 \geq 2\varsigma (L(z_1) - L^*). \quad (41)$$

It is shown in [21] that when L satisfies Assumptions 3.4 and 3.6, and HHA in (39)-(40) satisfies Assumption 3.5, each maximal solution $(t, j) \mapsto (z(t, j), \tau(t, j))$ to HHA in (39)-(40) satisfies

$$L(z_1(t, j)) - L^* \leq (L(z_1(0, 0)) - L^*) \min \{1, \exp(-\Omega(t - \theta))\} \quad (42)$$

for each $(t, j) \in \text{dom}(z, \tau)$ such that $z_2(0, 0) = 0$ and $\tau(0, 0) = 0$, where

$$\Omega := \ln \left(1 + \frac{\varsigma}{M} \right) \bar{T}^{-1}, \quad \theta := \bar{T} \quad (43)$$

and where $\varsigma > 0$ comes from Assumption 3.6 and $M > 0$ comes from Assumption 3.4. Comparing the bound in (42) with the bound on \mathcal{H}_1 in (12), from (32), the HHA algorithm reaches the neighborhood of the minimizer faster than the hybrid closed-loop algorithm \mathcal{H} in (9)-(10). The assumptions imposed on L for HHA, however, are different from the assumptions on L for the hybrid closed-loop algorithm \mathcal{H} . While some assumptions on L are weaker for HHA – for instance, Assumption 3.6 – other assumptions are stronger, such as items (L2) and (L3) of Assumption 3.4. Moreover, Assumption 3.5 is imposed on HHA, which is not imposed on the hybrid closed-loop algorithm \mathcal{H} . Table 3 summarizes the assumptions and the convergence rate results for HHA, \mathcal{H} , and HAND-1.

Next, we compare \mathcal{H}_0 , \mathcal{H}_1 , \mathcal{H} , HAND-1, and HHA 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 chose the heavy ball parameter value $\gamma = \frac{2}{3}$ and we tuned λ to 40 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. Since the gain coefficient of z_1 for HHA is fixed at -1 , then to ensure that the gain coefficients of z_1 for \mathcal{H} and HHA are the same, the only value possible for ζ , for the objective function L chosen, is $\sqrt{2}$, for Nesterov’s algorithm. In Addition, since z_2 does not appear in the dynamics of \dot{z}_2 of HHA in (39),*

Algorithm	Assumptions	Convergence Rate
\mathcal{H}	Assumptions 2.1, 2.2, and 2.4	$\frac{1}{(t+2)^2}$ globally and $\exp(-(1-m)\psi t)$ locally, where $m \in (0, 1)$ s.t. $\psi := \frac{m\alpha\gamma}{\lambda} > 0$ and $\nu := \psi(\psi - \lambda) < 0$.
HAND-1	Assumptions 2.1 and 2.4	$\frac{1}{t^2}$ until the first reset (at $j = 1$).
HHA	Assumptions 3.4, 3.5, and 3.6	$\min\{1, \exp(-\Omega(t - \theta))\}$, where Ω and θ are defined via (43).

Table 3: A comparison of assumptions and convergence rates, for HAND-1 from [19], HHA from [21], and the hybrid closed-loop algorithm \mathcal{H} .

we could not compare gain coefficients for z_2 for both \mathcal{H} and HHA. For the remaining HHA parameters, we chose $\bar{M} = 2$, since this satisfies $\bar{M} \geq M$. To ensure Assumption 3.5 is satisfied, we chose $\bar{T} := \frac{\pi}{2\sqrt{1}}$. Given $\zeta = \sqrt{2}$, the HAND-1 parameters $c_1 = 0.25$ and $T_{\min} = \frac{1+\sqrt{13}}{2}$ 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. The remaining HAND-1 parameters, we chose $r = 51$ and $\delta_{med} = 8650$ such that $T_{med} \geq \sqrt{\frac{B}{\delta_{med}}} + T_{\min} > 0$, and we chose $T_{\max} = T_{med} + 1$ to ensure resets happen at the proper times. The uniting algorithm parameters are $c_0 = 7000$, $c_{1,0} \approx 1354.025$, $\varepsilon_0 = 10$, $\varepsilon_{1,0} = 5$, and $\alpha = 1$, which yield the values $\tilde{c}_0 = 10$, $\tilde{c}_{1,0} = 5$, $d_0 \approx 6933.3$, and $d_{1,0} \approx 1304.0$, which are calculated via (17) and (24). 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 \mathcal{H} are $z_1(0, 0) = 50$, $z_2(0, 0) = 0$, $q(0, 0) = 1$, and $\tau(0, 0) = 0$. Initial conditions for HAND-1 are $z_1(0, 0) = z_2(0, 0) = 50$ and $\tau(0, 0) = T_{\min}$. Initial conditions for HHA are $z_1(0, 0) = 50$, $z_2(0, 0) = 0$, and $\tau(0, 0) = 0$.

Figure 8 and Table 4 show the time that each algorithm takes to settle within¹⁶ 1% of z_1^* . Table 4 shows the percent improvement \mathcal{H} over \mathcal{H}_0 , \mathcal{H}_1 , and HAND-1, which is calculated using (36). Since HHA converges expo-

¹⁶Code at github.com/HybridSystemsLab/UnitingComparison

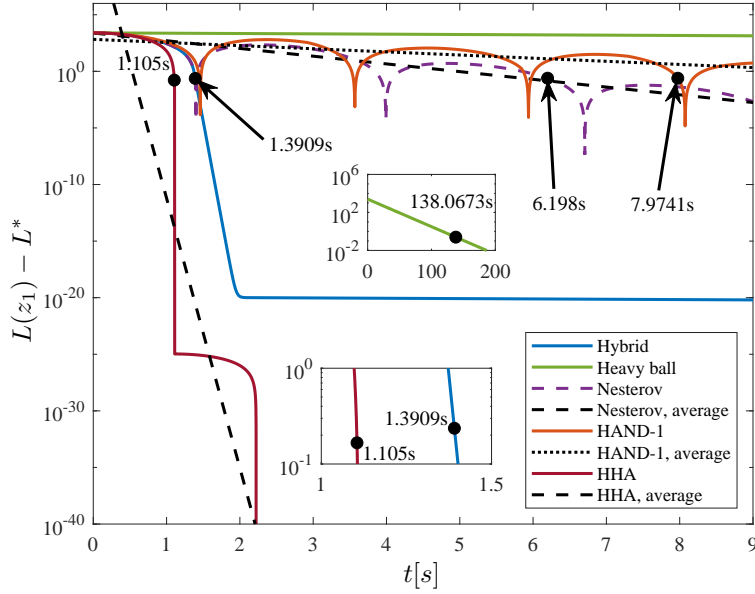


Figure 8: A comparison of the evolution of L over time for Nesterov’s method in (3), heavy ball, HAND-1 from [19], HHA from [21], and our proposed uniting algorithm, for a function $L(z_1) := z_1^2$, with a single minimizer at $z_1^* = 0$. Nesterov’s method, shown in purple, settles to within 1% of z_1^* in about 6.2 seconds. The heavy ball algorithm, shown in green, settles to within 1% of z_1^* in about 138.1 seconds. HAND-1, shown in orange, settles to within 1% of z_1^* in about 8.0 seconds. HHA, shown in red, settles to within 1% of z_1^* in about 1.1 seconds. The hybrid closed-loop system \mathcal{H} , shown in blue, settles to within 1% of z_1^* in about 1.4 seconds.

Algorithm	Time to converge (s)	% improvement
\mathcal{H}	1.390	–
\mathcal{H}_0	138.1	99.0
\mathcal{H}_1	6.191	77.5
HAND-1	7.974	82.6
HHA	1.105	–

Table 4: Average times for which \mathcal{H} , \mathcal{H}_0 , \mathcal{H}_1 , HAND-1, and HHA settle to within 1% of z_1^* , and the percent improvement of \mathcal{H} over each algorithm. Percent improvement is calculated via (36). The objective function used for this table is $L(z_1) := z_1^2$.

nentially, as shown in the bound in (42), such a rate explains why HHA is faster than \mathcal{H} in Figure 8 and Table 4. The hybrid closed-loop algorithm \mathcal{H} converges faster than \mathcal{H}_0 , \mathcal{H}_1 , and HAND-1, however.

Example 3.7 *This example explores the trade-off that results from using different values of $\zeta > 0$ for the uniting algorithm. Particularly, for $\zeta = 1$, we first compare the uniting algorithm in simulation with the individual optimization algorithms \mathcal{H}_0 , \mathcal{H}_1 , and the HAND-1 algorithm from [19], using the same objective function as in Example 3.2, and next we compare the resulting solutions with those in Table 1. Recall that the objective function in Example 3.2 is $L(z_1) := z_1^2$, the gradient of which is Lipschitz continuous with $M = 2$, and which has a single minimizer at $z_1^* = 0$. Since the gain coefficient of ∇L is proportional to ζ^2 , we choose different parameters for the HAND-1 algorithm for the simulation depicted in¹⁷ Figure 2, so that the gain coefficients of z_1 and z_2 are the same for HAND-1 and \mathcal{H} in this simulation. Namely, given $\zeta = 1$, for HAND-1 we choose $T_{\min} = 3$ and $c_1 = 0.25$. For the other HAND-1 parameters, we choose $r = 51$ and $\delta_{med} = 4010$ such that $T_{med} \geq \sqrt{\frac{B}{\delta_{med}}} + T_{\min} > 0$, and we again choose $T_{\max} = T_{med} + 1$ to ensure resets happen at the proper times. We arbitrarily choose $\gamma = \frac{2}{3}$, and we tuned λ to 40 by choosing a value arbitrarily larger than $2\sqrt{a_1}$ and gradually increasing until there was no overshoot in the hybrid algorithm.*

¹⁷Code found at same link as in Footnote 3

The uniting algorithm parameters are $c_0 = 320$, $c_{1,0} \approx 271.584$, $\varepsilon_0 = 10$, $\varepsilon_{1,0} = 5$, and $\alpha = 1$, which yield the values $\tilde{c}_0 = 10$, $\tilde{c}_{1,0} = 5$, $d_0 \approx 253.333$, and $d_{1,0} \approx 234.084$, which are calculated via (17) and (24). 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 \mathcal{H} are $z_1(0,0) = 50$, $z_2(0,0) = 0$, $q(0,0) = 1$, and $\tau(0,0) = 0$, and for HAND-1 are $z_1(0,0) = z_2(0,0) = 50$ and $\tau(0,0) = T_{\min}$.

First, we compare solutions to each algorithm within Figure 2 itself. Table 5 shows the time that each algorithm takes to settle within 1% of z_1^* , averaged over solutions starting from ten different values of $z_1(0,0)$ (listed in the first column of Table 2), and the percent improvement of \mathcal{H} over \mathcal{H}_0 , \mathcal{H}_1 , and HAND-1, which is calculated using (36). While the closed-loop algorithm \mathcal{H} still converges faster than all the other algorithms in Figure 2 and Table 5, the improvement over \mathcal{H}_0 , \mathcal{H}_1 , and HAND-1 is smaller than it is in Table 1.

Algorithm	Average time to converge (s)	Average % improvement
\mathcal{H}	2.387	–
\mathcal{H}_0	138.066	98.3
\mathcal{H}_1	8.782	72.8
HAND-1	14.343	83.4

Table 5: Times for which \mathcal{H} , \mathcal{H}_0 , \mathcal{H}_1 , and HAND-1 settle to within 1% of z_1^* , and percent improvement of \mathcal{H} over each algorithm, as shown in Figure 2. Percent improvement is calculated via (36). The objective function used for this table is $L(z_1) := z_1^2$.

Next, we compare solutions using $\zeta = 1$, in Figure 2, with solutions using $\zeta = 2$, in Table 1. Since $\zeta > 0$ scales time in solutions to (3), Then smaller values of ζ result in slower settling to within 1% of z_1^* for \mathcal{H}_1 with less frequent oscillations, as seen in Figure 2 with $\zeta = 1$ (about 8.8 seconds), while larger values of ζ result in settling to within 1% of z_1^* for \mathcal{H}_1 faster, with more frequent oscillations, as seen in Figure 1 and Table 1 with $\zeta = 2$ (about 4.5 seconds). For the uniting algorithm, this translates to faster settling to within 1% of z_1^* with $\zeta = 2$ (about 0.8 seconds), in Figure 1 and Table 1, compared with slower settling to within 1% of z_1^* with $\zeta = 1$ (about 2.4 seconds), in Figure 2, but with no oscillations, in both cases, due to the switch to \mathcal{H}_0 . In both Figure 2, and Table 1, the uniting algorithm converges more quickly

than the HAND-1 algorithm, when both algorithms are tuned to have the same gain coefficients for the z_1 and z_2 terms. Although larger ζ results in faster convergence, the trade-off is that even though the z_2 (velocity) term generally reduces quickly as it approaches the neighborhood of the minimizer for any size of ζ , the z_2 still ends up relatively larger near the minimizer than it is when ζ is smaller. The consequence is that, when ζ is larger, $d_{1,0}$ needs to be set much larger so that the uniting algorithm can still make the switch to \mathcal{H}_0 at the proper time. This also means that $c_{1,0}$ needs to be set much larger, due to the definition of $d_{1,0}$ in (24). Additionally, c_0 and d_0 , also need to be set larger to ensure the algorithm still has adequate hysteresis. Recall that, in Example 3.2, for $\zeta = 2$, we have the parameter values $c_0 = 7000$, $c_{1,0} \approx 6819.676$, $d_0 = 6933$, and $d_{1,0} = 6744$, which are quite large, while for the simulation shown in Figure 2 these same parameters have much smaller values, as listed in the second paragraph of this example.

4 Proof of Theorem 2.11

This section provides a proof of Theorem 2.11 from Section 2.7. The proof consists of the following steps.

- Section 4.1 establishes UGAS of $\{z_1^*\} \times \{0\}$, and an exponential convergence rate for the closed-loop algorithm \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 the closed-loop algorithm \mathcal{H}_1 ;
- Section 4.3 uses the properties in Sections 4.1 and 4.2 and a proof-by-contradiction to establish UGAS of \mathcal{A} , defined via (31), for \mathcal{H} ;
- 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 in (11) has the set $\{z_1^*\} \times \{0\}$ UGAS. To prove it, we use an invariance principle.

Proposition 4.1 (UGAS of $\{z_1^*\} \times \{0\}$ for \mathcal{H}_0) *Let L satisfy Assumptions 2.1, 2.2, and 2.4. 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).*

Proof. By Proposition 2.9, each maximal solution to the closed-loop algorithm \mathcal{H}_0 , defined via (11), is bounded, complete, and unique. Recall that, in the proof of Proposition 2.9, it was shown that V_0 in (16) satisfies (74) for all $z \in \mathbb{R}^{2n}$, since λ is positive. Therefore, by an application of Theorem D.3, since $\gamma > 0$ and $\lambda > 0$, the set $\{z_1^*\} \times \{0\}$ is stable for the closed-loop algorithm \mathcal{H}_0 . Since by Lemma 2.8 \mathcal{H}_0 satisfies the hybrid basic conditions, then, using the invariance principle in Theorem D.6, every maximal solution that is complete and bounded approaches the largest weakly invariant set for \mathcal{H}_0 in (11) that is contained in

$$\left\{ z \in \mathbb{R}^{2n} : \dot{V}_0(z) = 0 \right\} \cap \left\{ z \in \mathbb{R}^{2n} : V_0(z) = r \right\}, \quad r \geq 0. \quad (44)$$

Such a set is nonempty only when $r = 0$ and, precisely, is equal to $\{z_1^*\} \times \{0\}$. This property can be seen by noticing that $\left\{ z \in \mathbb{R}^{2n} : \dot{V}_0(z) = 0 \right\} = \left\{ z \in \mathbb{R}^{2n} : z_2 = 0 \right\}$, and that after setting z_2 to zero in (11) we obtain $\begin{bmatrix} \dot{z}_1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -\gamma \nabla L(z_1) \end{bmatrix}$. For any solution to this system, its z_1 component satisfies $0 = \gamma \nabla L(z_1)$, which, since $\gamma > 0$ and since $\nabla L(z_1) = 0$ only when z_1 is the minimizer of L , leads to $z_1 = z_1^*$. Then, the only maximal solution that starts and stays in (44) is the solution from $\{z_1^*\} \times \{0\}$, for which $r = 0$. Then, every bounded and complete solution to the closed-loop algorithm \mathcal{H}_0 converges to $\{z_1^*\} \times \{0\}$. The arguments above involving the Lyapunov theorem in Theorem D.3 and the invariance principle in Theorem D.6 yield global pre-asymptotic stability of $\{z_1^*\} \times \{0\}$ for \mathcal{H}_0 . Since by Proposition 2.9, each maximal solution to \mathcal{H}_0 is complete, then $\{z_1^*\} \times \{0\}$ is globally asymptotically stable for the closed-loop algorithm \mathcal{H}_0 . Since \mathcal{H}_0 satisfies the hybrid basic conditions by Lemma 2.8, then, by Theorem D.4, $\{z_1^*\} \times \{0\}$ is UGAS for \mathcal{H}_0 . \square

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 [14, Lemma 4.2], for \mathcal{H}_0 :

$$V(z) := \gamma(L(z_1) - L^*) + \frac{1}{2} |\psi(z_1 - z_1^*) + z_2|^2 + \frac{\nu}{2} |z_1 - z_1^*|^2 \quad (45)$$

where, given $\lambda > 0$, $\psi > 0$ is chosen such that $\nu := \psi(\psi - \lambda) < 0$. When L satisfies Assumption 2.1, the following lemma, which is a version of [14, Lemma 4.2] tailored for the unperturbed heavy ball algorithm in (11), gives an upper bound on the change of the Lyapunov function in (45).

Lemma 4.2 *Let L satisfy Assumption 2.1, 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 $z \in \mathbb{R}^{2n}$:*

$$\dot{V}(z) \leq -\psi(a(z_1) + 2\nu c(z_1)) + 2(\psi - \lambda)b(z) \quad (46)$$

where V is defined in (45), $a(z_1) := \gamma(L(z_1) - L^*)$, $b(z) := \frac{1}{2}|\psi(z_1 - z_1^*) + z_2|^2$, and $c(z_1) := \frac{1}{2}|z_1 - z_1^*|^2$.

Proof. Since L is \mathcal{C}^1 , convex, and has a single minimizer z_1^* , and since $\nabla V(z) = [\gamma \nabla L(z_1) + \psi(\psi(z_1 - z_1^*) + z_2) + \nu(z_1 - z_1^*) \quad \psi(z_1 - z_1^*) + z_2]$, then we evaluate the derivative of V , defined via (45), using the map $z \mapsto F_P(\kappa_0(h_0(z)))$, where F_P is defined via (4), κ_0 is defined in (6a), and h_0 is defined via (7). For each $z \in \mathbb{R}^{2n}$, we obtain

$$\begin{aligned} \dot{V}(z) &= \langle \nabla V(z), F_P(\kappa_0(h_0(z))) \rangle = \left\langle \nabla V(z), \begin{bmatrix} z_2 \\ \kappa_0(h_0(z)) \end{bmatrix} \right\rangle \quad (47) \\ &= \gamma \langle \nabla L(z_1), z_2 \rangle + \psi \langle z_2, \psi(z_1 - z_1^*) + z_2 \rangle + \nu \langle z_2, z_1 - z_1^* \rangle \\ &\quad - \lambda \langle z_2, \psi(z_1 - z_1^*) + z_2 \rangle - \gamma \langle \nabla L(z_1), \psi(z_1 - z_1^*) + z_2 \rangle \\ &= -\gamma\psi \langle \nabla L(z_1), z_1 - z_1^* \rangle + (\nu + \psi(\psi - \lambda)) \langle z_2, z_1 - z_1^* \rangle + (\psi - \lambda) |z_2|^2. \end{aligned}$$

Note that $|\psi(z_1 - z_1^*) + z_2|^2 = |z_2|^2 + 2\psi \langle z_2, z_1 - z_1^* \rangle + \psi^2 |z_1 - z_1^*|^2$, from where we obtain $|z_2|^2 = |\psi(z_1 - z_1^*) + z_2|^2 - 2\psi \langle z_2, z_1 - z_1^* \rangle - \psi^2 |z_1 - z_1^*|^2$. Substituting the expression for $|z_2|^2$ into (47), we arrive at, for all $z \in \mathbb{R}^{2n}$,

$$\begin{aligned} \dot{V}(z) &= -\gamma\psi \langle \nabla L(z_1), z_1 - z_1^* \rangle + (\psi - \lambda) |\psi(z_1 - z_1^*) + z_2|^2 \\ &\quad + (\nu - \psi(\psi - \lambda)) \langle z_2, z_1 - z_1^* \rangle - \psi^2(\psi - \lambda) |z_1 - z_1^*|^2 \\ &= -\gamma\psi \langle \nabla L(z_1), z_1 - z_1^* \rangle + 2(\psi - \lambda)b(z) - 2\psi\nu c(z_1) \quad (48) \end{aligned}$$

since $\nu = \psi(\psi - \lambda)$, where $b(z) = \frac{1}{2}|\psi(z_1 - z_1^*) + z_2|^2$ and $c(z_1) = \frac{1}{2}|z_1 - z_1^*|^2$. Since L is \mathcal{C}^1 , convex, and has a unique minimizer by Assumption 2.1, then using the definition of convexity in Footnote 6 with $u_1 = z_1^*$ and $w_1 = z_1$, we

get $-(L(z_1) - L^*) \geq -\langle \nabla L(z_1), z_1 - z_1^* \rangle$. Substituting it into (48) yields, for all $z \in \mathbb{R}^{2n}$, $\dot{V}(z) \leq -\psi a(z_1) + 2(\psi - \lambda)b(z) - 2\psi\nu c(z_1)$, where $a(z_1) = \gamma(L(z_1) - L^*)$, and (46) is satisfied. \square

We employ Lemma 5.2 to show that when L satisfies Assumptions 2.1 and 2.2, 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 [14, Theorem 3.2] tailored for the unperturbed heavy ball algorithm \mathcal{H}_0 in (11).

Proposition 4.3 (*Convergence rate for \mathcal{H}_0*) *Let L satisfy Assumptions 2.1 and 2.2, let $\alpha > 0$ come from (13), 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*

$$L(z_1(t)) - L^* = \mathcal{O}(\exp(-(1-m)\psi t)) \quad \forall t \in \text{dom } z (= \mathbb{R}_{\geq 0}). \quad (49)$$

Proof. By Lemma 4.2, the bound in (46) is satisfied for V in (45) for each $z \in \mathbb{R}^{2n}$ since, by Assumption 2.1, L is \mathcal{C}^1 , convex, and has a single minimizer z_1^* . Then, since $\psi = \frac{m\alpha\gamma}{\lambda} > 0$ is such that $\nu = \psi(\psi - \lambda) < 0$ and c is nonnegative, this leads to

$$V(z) = a(z_1) + b(z) + \nu c(z_1) \leq a(z_1) + b(z) \quad \forall z \in \mathbb{R}^{2n} \quad (50)$$

where a , b , and c are defined below (46). By Assumption 2.2, L has quadratic growth away from z_1^* . Therefore, we have, for all $z \in \mathbb{R}^{2n}$,

$$\begin{aligned} a(z_1) + 2\nu c(z_1) &= a(z_1) - 2|\nu|c(z_1) = \gamma(L(z_1) - L^*) - |\nu||z_1 - z_1^*|^2 \\ &\geq \gamma(L(z_1) - L^*) - \frac{|\nu|(L(z_1) - L^*)}{\alpha} = \left(1 - \frac{|\nu|}{\alpha\gamma}\right)a(z_1). \end{aligned} \quad (51)$$

Observe that, for each $m \in (0, 1)$ such that $\psi = \frac{m\alpha\gamma}{\lambda} > 0$ and $\nu = \psi(\psi - \lambda) < 0$, we have

$$|\nu| = \psi(\lambda - \psi) \leq \lambda\psi = m\alpha\gamma \quad (52)$$

It follows from (51) and (52) that

$$a(z_1) + 2\nu c(z_1) \geq (1-m)a(z_1) \quad (53)$$

for all $z \in \mathbb{R}^{2n}$. Noticing that from (50) we have $a(z_1) + \frac{1}{(1-m)}b(z) \geq a(z_1) + b(z) \geq V(z)$, substituting (53) into (46) we have

$$\dot{V}(z) \leq -(1-m)\psi a(z_1) + 2(\psi - \lambda)b(z)$$

$$\begin{aligned}
&\leq -(1-m)\psi a(z_1) + \psi b(z) + (\psi - 2\lambda)b(z) \\
&\leq -(1-m)\psi a(z_1) + \psi b(z) \\
&\leq -(1-m)\psi \left(a(z_1) + \frac{1}{(1-m)}b(z) \right) \\
&\leq -(1-m)\psi V(z)
\end{aligned} \tag{54}$$

for all $z \in \mathbb{R}^{2n}$. The third inequality comes from the fact that we choose $\psi = \frac{m\alpha\gamma}{\lambda} > 0$ such that $\psi - \lambda < 0$ and, consequently, $\psi - 2\lambda < 0$. Applying Grönwall's inequality to (54) shows that every maximal solution $t \mapsto z(t)$ to (1) satisfies $V(z(t)) \leq V(z(0)) \exp(-(1-m)\psi t)$ for all $t \in \text{dom } z (= \mathbb{R}_{\geq 0})$. Therefore, each maximal solution $t \mapsto z(t)$ to the closed-loop algorithm \mathcal{H}_0 in (11) satisfies (49) for all $t \in \text{dom } z (= \mathbb{R}_{\geq 0})$. \square

4.2 Properties of \mathcal{H}_1

When L satisfies Assumptions 2.1 and 2.4, then we can derive an upper bound, for all $t \geq 0$, on the Lyapunov function V_1 in (22) along solutions to \mathcal{H}_1 . To derive such a bound, we extend [1, Proposition 3.2], which assumes $L^* = 0$, $z_1^* = 0$, and $\zeta = 1$, to the general case of $L^* \in \mathbb{R}$, a single minimizer $z_1^* \in \mathbb{R}^n$, and $\zeta > 0$, in the following proposition.

Proposition 4.4 *Let L satisfy Assumptions 2.1 and 2.4. 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*

$$V_1(z(t), t) \leq \frac{4}{(t+2)^2} V_1(z(0), 0) \tag{55}$$

for all $t \geq 0$, where V_1 is defined via (22).

Proof. See Section C. \square

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 4.4. This proposition is a new result, which was not analyzed in [1].

Proposition 4.5 *(Convergence rate for \mathcal{H}_1) Let L satisfy Assumptions 2.1 and 2.4. Let $\zeta > 0$ and $M > 0$ come from Assumption 2.4. 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:

$$\begin{aligned} & \frac{\zeta^2}{M}(L(z_1(t)) - L^*) \\ & \leq V_1(z(t), t) \leq \frac{4c}{(t+2)^2} (|z_1(0) - z_1^*|^2 + |z_2(0)|^2) \end{aligned} \quad (56)$$

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

Proof. The proof consists of the following steps.

- 1) First, we use the definition of convexity in Footnote 6 and the Lipschitz continuity of ∇L in Assumption 2.4, to show that V_1 satisfies

$$V_1(z, \tau) \leq \alpha_2 |z|_{\mathcal{A}_2}^2 := (1 + \zeta^2) |z|_{\mathcal{A}_2}^2 \quad (57)$$

where $1 + \zeta^2 > 0$;

- 2) Then, we use the Lipschitz continuity of ∇L in Assumption 2.4 and the comparison principle to show that the bound in step 1) along $t \mapsto z(t)$ satisfies $V_1(z(t), t) \leq \alpha_2 \exp\left(2\left(\sqrt{\frac{13}{4} + \frac{\zeta^4}{M}}\right)t\right) (|z_1(0) - z_1^*|^2 + |z_2(0)|^2)$ for all $t \geq 0$;

- 3) Next, we show that at $t = 0$, $V_1(z(0), 0)$ is upper bounded by $c (|z_1(0) - z_1^*|^2 + |z_2(0)|^2)$, where $c = (1 + \zeta^2) \exp\left(\sqrt{\frac{13}{4} + \frac{\zeta^4}{M}}\right)$;

- 4) Finally, we combine the bound in 3) with (55) to get (56) for all $t \geq 0$.

Proceeding with step 1), the Lyapunov function V_1 , defined via (22), can be upper bounded by a class- \mathcal{K}_∞ function, namely, defining the set

$$\mathcal{A}_2 := \{z_1^*\} \times \{0\} \quad (58)$$

then, V_1 satisfies

$$V_1(z, \tau) \leq \alpha_2 |z|_{\mathcal{A}_2}^2 \quad (59)$$

for all $(z, \tau) \in \mathbb{R}^{2n} \times \mathbb{R}_{\geq 0}$, and with α_2 derived as follows. Since \bar{a} , defined via (23), equals 1 at $\tau = 0$ and \bar{a} is monotonically decreasing toward zero

(but being always positive) as τ tends to ∞ , then \bar{a} is upper bounded by 1, and, consequently, the first term of V_1 can be upper bounded, for all $(z, \tau) \in \mathbb{R}^{2n} \times \mathbb{R}_{\geq 0}$, as follows:

$$\frac{1}{2} \left| \frac{2}{(\tau + 2)} (z_1 - z_1^*) + z_2 \right|^2 \leq |z_1 - z_1^*|^2 + |z_2|^2. \quad (60)$$

The second term of V_1 can be bounded as follows. Since by Assumption 2.1, L is \mathcal{C}^1 , convex, and has a single minimizer z_1^* , then, since $\nabla L(z_1^*) = 0$, we can upper bound $L(z_1) - L^*$ in the following manner, using the definition of convexity in Footnote 6 and the Lipschitz continuity of ∇L in Assumption 2.4, using $u_1 = z_1^*$ and $w_1 = z_1$: $|L(z_1) - L^*| \leq |\langle \nabla L(z_1), z_1^* - z_1 \rangle| \leq |\nabla L(z_1)| |z_1 - z_1^*| \leq M |z_1 - z_1^*|^2$, for all $z_1 \in \mathbb{R}^n$. Therefore, since $L(z_1) \geq L^*$, we can upper bound the second term of V_1 as follows:

$$\frac{\zeta^2}{M} (L(z_1) - L^*) \leq \zeta^2 |z_1 - z_1^*|^2 \leq \zeta^2 (|z_1 - z_1^*|^2 + |z_2|^2) \quad (61)$$

for all $z \in \mathbb{R}^{2n}$. Using (60) and (61) $V_1(z, \tau)$ is upper bounded as in (57) for each $z \in \mathbb{R}^{2n}$ and each $\tau \in \mathbb{R}_{\geq 0}$.

Next, for step 2), in order to apply the comparison principle, we define the system

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} z_2 \\ -2\bar{d}(t)z_2 - \frac{\zeta^2}{M} \nabla L(z_1 + \bar{\beta}(t)z_2) \end{bmatrix} =: f(z, t) \quad z \in \mathbb{R}^{2n}. \quad (62)$$

Since ∇L is Lipschitz continuous with constant $M > 0$ by Assumption 2.4, then using Assumption 2.4 with $w_1 = z_1 + \bar{\beta}(t)z_2$ and $u_1 = z_1^*$ yields, for each $z_1, z_2 \in \mathbb{R}^n$ and each $t \in \mathbb{R}_{\geq 0}$,

$$|\nabla L(z_1 + \bar{\beta}(t)z_2)| \leq M |z_1 - z_1^* + \bar{\beta}(t)z_2|. \quad (63)$$

Then, since $|\bar{d}(t)| \leq \frac{3}{4}$ and $|\bar{\beta}(t)| \leq 1$ for all $t \geq 0$, we have

$$\begin{aligned} |f(z, t)|^2 &= |z_2|^2 + \left| -2\bar{d}(t)z_2 - \frac{\zeta^2}{M} \nabla L(z_1 + \bar{\beta}(t)z_2) \right|^2 \\ &\leq |z_2|^2 + \frac{9}{4} |z_2|^2 + \frac{\zeta^4}{M^2} |\nabla L(z_1 + \bar{\beta}(t)z_2)|^2 \\ &\leq \frac{13}{4} |z_2|^2 + \frac{\zeta^4}{M} (|z_1 - z_1^*|^2 + |z_2|^2) \end{aligned}$$

$$\begin{aligned}
&= \frac{\zeta^4}{M} |z_1 - z_1^*|^2 + \left(\frac{13}{4} + \frac{\zeta^4}{M} \right) |z_2|^2 \\
&\leq \left(\frac{13}{4} + \frac{\zeta^4}{M} \right) |z|_{\mathcal{A}_2}^2
\end{aligned} \tag{64}$$

for all $z \in \mathbb{R}^{2n}$ and all $t \in \mathbb{R}_{\geq 0}$, where \mathcal{A}_2 is defined via (58). The second inequality in (64) comes from applying (63). The comparison principle [34, Lemma 3.4], leads to the following bound of the norm of the solution to (62):

$$|z(t)|_{\mathcal{A}_2} \leq \exp \left(\frac{1}{2} \sqrt{\frac{13}{4} + \frac{\zeta^4}{M}} t \right) |z(0)|_{\mathcal{A}_2} \tag{65}$$

for all $t \geq 0$. Then, (59) along $t \mapsto (z(t))$ reduces to, for all $t \geq 0$

$$\begin{aligned}
V_1(z(t), t) &\leq (1 + \zeta^2) |z(t)|_{\mathcal{A}_2}^2 \\
&\leq (1 + \zeta^2) \exp \left(\sqrt{\frac{13}{4} + \frac{\zeta^4}{M}} t \right) (|z_1(0) - z_1^*|^2 + |z_2(0)|^2).
\end{aligned} \tag{66}$$

In step 3), we evaluate this bound at $t = 0$. Finally, for step 4), taking $c = (1 + \zeta^2) \exp \left(\sqrt{\frac{13}{4} + \frac{\zeta^4}{M}} \right)$, combining (55) with 3) at $t = 0$ yields (56) for all $t \geq 0$. \square

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

$$\mathcal{A}_1 := \{z_1^*\} \times \{0\} \times \mathbb{R}_{\geq 0} \tag{67}$$

UGAS. To prove it, we use Proposition 4.5 and [23, Theorem 3.18]. This proposition is a new result, which was not analyzed in [1].

Proposition 4.6 (*UGAS of \mathcal{A}_1 in (67) for \mathcal{H}_1*) *Let L satisfy Assumptions 2.1 and 2.4. Let $\zeta > 0$ and let $M > 0$ come from Assumption 2.4. Then, the set \mathcal{A}_1 in (67) is UGAS for \mathcal{H}_1 .*

Proof. By Proposition 2.9, each maximal solution to \mathcal{H}_1 in (12) is complete and unique. Next, since L is \mathcal{C}^1 , convex, and has a unique minimizer by Assumption 2.1, then $\mathcal{A}_1 \subset \mathbb{R}^{2n} \times \mathbb{R}_{\geq 0}$, defined via (67), is closed by construction, satisfying the first assumption of [23, Theorem 3.18]. Then, since by Assumption 2.1, L is \mathcal{C}^1 , then V_1 in (22) is continuously differentiable

and therefore, since $\mathbb{R}^{2n} \times \mathbb{R}_{\geq 0} \subset \text{dom } V_1$, V_1 is a Lyapunov function candidate for \mathcal{H}_1 by [23, Definition 3.16], satisfying the second assumption of [23, Theorem 3.18].

Next, since the distance of the state τ to $\mathbb{R}_{\geq 0}$ is always zero, then we show that V_1 in (22) is radially unbounded in z , relative to \mathcal{A}_2 , defined via (58). Since L has quadratic growth away from z_1^* with constant $\alpha > 0$, by Assumption 2.2, and due to \bar{a} , defined via (23), equaling 1 at $\tau = 0$ and monotonically decreasing toward zero (but being always positive) as τ tends to ∞ , then we lower bound V_1 as follows:

$$\begin{aligned}
V_1(z, \tau) &= \frac{1}{2} |\bar{a}(\tau) (z_1 - z_1^*) + z_2|^2 + \frac{\zeta^2}{M} (L(z_1) - L^*) & (68) \\
&\geq \frac{1}{2} |\bar{a}(\tau) (z_1 - z_1^*) + z_2|^2 + \frac{\alpha \zeta^2}{M} |z_1 - z_1^*|^2 \\
&\geq \frac{\bar{a}^2(\tau)}{2} |z_1 - z_1^*|^2 + \bar{a}(\tau) \langle z_1 - z_1^*, z_2 \rangle + \frac{1}{2} |z_2|^2 + \frac{\alpha \zeta^2}{M} |z_1 - z_1^*|^2 \\
&\geq \left(\frac{\bar{a}^2(\tau)}{2} + \frac{\alpha \zeta^2}{M} \right) |z_1 - z_1^*|^2 + \frac{\bar{a}(\tau)}{2} \langle z_1 - z_1^*, z_2 \rangle \\
&\quad + \frac{\bar{a}(\tau)}{2} \langle z_1 - z_1^*, z_2 \rangle + \frac{1}{2} |z_2|^2 \\
&\geq [(z_1 - z_1^*)^\top \quad z_2^\top] P \begin{bmatrix} z_1 - z_1^* \\ z_2 \end{bmatrix}
\end{aligned}$$

for each $z \in \mathbb{R}^{2n}$ and each $\tau \in \mathbb{R}_{\geq 0}$, where

$$P := \begin{bmatrix} \left(\frac{\bar{a}^2(\tau)}{2} + \frac{\alpha \zeta^2}{M} \right) & \frac{\bar{a}(\tau)}{2} \\ \frac{\bar{a}(\tau)}{2} & \frac{1}{2} \end{bmatrix}. \quad (69)$$

Next, we show that P in (69) is positive definite, so that there exists α_1 such that¹⁸

$$\alpha_1 |z|_{\mathcal{A}_2}^2 \leq [(z_1 - z_1^*)^\top \quad z_2^\top] P \begin{bmatrix} z_1 - z_1^* \\ z_2 \end{bmatrix} \leq V_1(z, \tau) \quad (70)$$

for each $z \in \mathbb{R}^{2n}$ and each $\tau \in \mathbb{R}_{\geq 0}$. To that end, we show that the leading principal minors of P in (69) are strictly positive, as follows. Since $\bar{a}(\tau) \in$

¹⁸It was already shown that there exists α_2 such that the upper bound on V_1 in (57) holds.

$(0, 1]$ for each $\tau \in \mathbb{R}_{\geq 0}$, $\alpha > 0$ from Assumption 2.2, $\zeta > 0$, and $M > 0$ from Assumption 2.4, we have:

$$\left(\frac{\bar{a}^2(\tau)}{2} + \frac{\alpha\zeta^2}{M} \right) > 0 \quad (71a)$$

$$\det(P) = \frac{1}{2} \left(\frac{\bar{a}^2(\tau)}{2} + \frac{\alpha\zeta^2}{M} \right) - \left(\frac{\bar{a}(\tau)}{2} \right)^2 = \frac{\alpha\zeta^2}{2M} > 0. \quad (71b)$$

Therefore, since the leading principal minors of P are strictly positive, then P is positive definite. Hence, there exists α_1 such that (70) is true, and V_1 is radially unbounded in z , relative to \mathcal{A}_2 .

By Proposition 4.4, V_1 satisfies (89) for each $z \in \mathbb{R}^{2n}$ and $\tau \in \mathbb{R}_{\geq 0}$. Since L is \mathcal{C}^1 , convex, and has a unique minimizer by Assumption 2.1, then L is positive definite with respect to z_1^* and, consequently, V_1 is positive definite with respect to \mathcal{A}_1 in (67). Then, since $\bar{a}(\tau) \in (0, 1]$ for each $\tau \geq 0$, $\rho(|x|_{\mathcal{A}_1}) := \bar{a}(\tau)V_1(z, \tau)$ is positive definite with respect to \mathcal{A}_1 . Therefore, by an application of [23, Theorem 3.18], every complete solution to (12) converges to \mathcal{A}_1 in (67). The arguments above involving the Lyapunov theorem in [23, Theorem 3.18] yield UGpAS of \mathcal{A}_1 for \mathcal{H}_1 in (12). Since by Proposition 2.9, each maximal solution to \mathcal{H}_1 is complete, then \mathcal{A}_1 is UGAS for \mathcal{H}_1 . \square

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 2.8, satisfying the first assumption of Theorem D.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 Proposition 2.9. Since by Assumption 2.1, L has a unique minimizer z_1^* , then \mathcal{A} , defined via (31), 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 Theorem D.3.

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

- a) There exists $(t', j') \in \text{dom } x$ such that $x(t, j) \in C_1 \setminus D_1$ for all $(t, j) \in \text{dom } x, t + j \geq t' + j'$;

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

Case a) contradicts the fact that, by Proposition 4.6, the set \mathcal{A}_1 , defined via (67), is UGAS for \mathcal{H}_1 . Such UGAS of \mathcal{A}_1 , guaranteed by Proposition 4.6, 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 $t \rightarrow \infty$. In turn, due to the construction of C_1 and D_1 in (10), with $\mathcal{T}_{1,0}$ defined via (27), the solution x must reach D_1 at some $(t, j) \in \text{dom } x, t + j \geq t' + j'$. Therefore, case a) does not happen.

Case b) contradicts the fact that, by Proposition 4.1, $\{z_1^*\} \times \{0\}$ is UGAS for \mathcal{H}_0 . In fact, $\lim_{t+j \rightarrow \infty} |x(t, j)|_{\mathcal{A}} = 0$, 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 (27) and $\mathcal{T}_{0,1}$ in (30), 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 $G(D)$ is defined via (73) and 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 (67)) for \mathcal{H}_1 established in Proposition 4.6, the UGAS of $\{z_1^*\} \times \{0\}$ for \mathcal{H}_0 established in Proposition 4.1, and since each maximal solution to \mathcal{H} is complete by 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 (27) and $\mathcal{T}_{0,1}$

in (30) 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 4.1, 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 2.4, then by Proposition 4.5, since L satisfies Assumptions 2.1 and 2.4, each maximal solution $t \mapsto (z(t), \tau(t))$ to the closed-loop algorithm \mathcal{H}_1 with $\tau(0, 0) = 0$ satisfies (56), for all $t \geq 0$, where c is defined below (56). By Proposition 4.3, since L satisfies Assumptions 2.1 and 2.2, 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 (49) for all $t \in \text{dom } z (= \mathbb{R}_{\geq 0})$. 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 2.2, and $M > 0$ from Assumption 2.4, due to the construction of \mathcal{U}_0 , $\mathcal{T}_{1,0}$, and $\mathcal{T}_{0,1}$ in (20), (27), and (30), with $\tilde{c}_{1,0} \in (0, \tilde{c}_0)$ and $d_{1,0} \in (0, d_0)$ defined via (17) and (24), 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 (32) 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 2.11 for each $t \in I^1$ at which $q(t, 1)$ is equal to 0.

5 Extensions

Some possible extensions of Theorem 2.11, Proposition 4.1, Lemma 4.2, and Propositions 4.3, 4.4, 4.5, and 4.6 are as follows. A potential approach to discretizing the hybrid closed-loop algorithm \mathcal{H} in (9)-(10) can be found in

[35]. 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 2.11, Proposition 4.1, Lemma 4.2, and Propositions 4.3, 4.4, 4.5, and 4.6.

It is possible to extend Theorem 2.11, Proposition 4.1, Lemma 4.2, and Propositions 4.3, 4.4, 4.5, and 4.6 to include \mathcal{C}^1 , convex objective functions L with a compact and connected set of minimizers. Such an extension could be achieved via the use of Clarke’s generalized derivative (see [36]). Additionally, Clarke’s generalized derivative could be utilized to extend the analysis of the hybrid closed-loop algorithm to include nonsmooth convex objective functions L with a compact and connected set of minimizers.

6 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 [19]. 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 [1] 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.

Appendices

A Proof of Lemma 2.8

The objective function L is \mathcal{C}^1 , convex, and has a single minimizer by Assumption 2.1. Therefore, since ∇L is continuous, the following hold: the set \mathcal{U}_0 , defined via (20), is closed since it is a sublevel set of the continuous function V_0 ; due to \bar{a} in (23) being continuous, the set $\mathcal{T}_{1,0}$, defined via (27), is closed since it is a sublevel set of the continuous function V_1 ; the set $\mathcal{T}_{0,1}$, defined via (30), is closed since it is the closed complement of a set. Therefore, since the sets \mathcal{U}_0 , $\mathcal{T}_{1,0}$, and $\mathcal{T}_{0,1}$ are closed, then the sets D_0 , D_1 , C_0 , and C_1 are closed. Since C and D are both finite unions of finite and closed sets, then C and D are also closed.

Since \bar{d} and $\bar{\beta}$, defined via (5), are continuous, and since by Assumption 2.1, L is \mathcal{C}^1 , then h_q in (7) and κ_q in (6) are continuous. In turn, the map $z \mapsto F_P(z, \kappa_q(h_q(z, \tau), \tau))$ is also continuous since F_P in (4) is a \mathcal{C}^1 function of κ_q and h_q . Therefore, $x \mapsto F(x)$ is continuous. The map G satisfies (A3) by construction since it is continuous.

B Proof of Proposition 2.9

Since Assumptions 2.1, 2.2, and 2.4 hold, then \mathcal{H} satisfies the hybrid basic conditions by Lemma 2.8. With $\tilde{c}_0 > 0$ and $d_0 > 0$ defined via (17), since L is \mathcal{C}^1 , convex, has a single minimizer by Assumption 2.1, and has quadratic growth away from z_1^* by Assumption 2.2, from the arguments below (18), every $z \in \mathcal{U}_0$ belongs to the c_0 -sublevel set of V_0 ; recall that \mathcal{U}_0 is defined in (20) and that V_0 is defined via (16). Additionally, since by Assumption 2.2 L has quadratic growth away from z_1^* , then $\mathcal{T}_{0,1}$ in (30), defines the closed complement of a sublevel set of V_0 with level equal to c_0 . Therefore, due to the definitions of \mathcal{U}_0 in (20) and $\mathcal{T}_{0,1}$ in (30), $\Pi(C_0) \cup \Pi(D_0) = \mathbb{R}^{2n}$. Furthermore, since $\mathcal{T}_{1,0}$ is defined via (27), and since by the definitions of C_1 and D_1 in (10), C_1 is the closed complement of D_1 , then $\Pi(C_1) \cup \Pi(D_1) = \mathbb{R}^{2n}$.

Due to the definitions of C_0 , D_0 , C_1 , and D_1 in (10), \mathcal{U}_0 in (20), $\mathcal{T}_{1,0}$ in (27), and $\mathcal{T}_{0,1}$ in (30), then $C \setminus D$ is equal to $\text{int}(C)$. Hence, for each point

$x \in C \setminus D$, the tangent cone to C at x is

$$T_C(x) := \begin{cases} \mathbb{R}^{2n} \times \{0\} \times \{0\} & \text{if } x \in C_0 \setminus D_0, \\ \mathbb{R}^{2n} \times \{1\} \times \mathbb{R}_{\geq 0} & \text{if } x \in C_1 \setminus D_1. \end{cases} \quad (72)$$

Therefore, $F(x) \cap T_C(x) \neq \emptyset$, satisfying (VC) of for each point $x \in C \setminus D$, and nontrivial solutions exist for every initial point in $(C_0 \cup C_1) \cup (D_0 \cup D_1)$, where $\Pi(C_0) \cup \Pi(D_0) = \mathbb{R}^{2n}$ and $\Pi(C_1) \cup \Pi(D_1) = \mathbb{R}^{2n}$. To prove that item (c) of Proposition D.1 does not hold, we need to show that $G(D) \subset C \cup D$. With D defined in (10), $G(D) = (\mathcal{T}_{0,1} \times \{1\} \times \{0\}) \cup (\mathcal{T}_{1,0} \times \{0\} \times \{0\})$.

$$G(D) = (\mathcal{T}_{0,1} \times \{1\} \times \{0\}) \cup (\mathcal{T}_{1,0} \times \{0\} \times \{0\}). \quad (73)$$

Notice that $\mathcal{T}_{1,0} \times \{0\} \times \{0\} \subset C_0$ and $\mathcal{T}_{0,1} \times \{1\} \times \{0\} \subset C_1$. Therefore, $G(D) \subset C$; hence $G(D) \subset C \cup D$. Therefore, item (c) of Proposition D.1 does not hold. Then it remains to prove that item (b) does not happen.

To this end, we show first that \mathcal{H}_0 , defined via (11), has no finite time escape¹⁹, and has unique and bounded solutions. Since L is \mathcal{C}^1 by Assumption 2.1, and ∇L is Lipschitz continuous by Assumption 2.4, then h_0 in (7) and κ_0 in (6a) are Lipschitz continuous, which, since F_P is a \mathcal{C}^1 function of h_0 and κ_0 , means the map $z \mapsto F_P(z, \kappa_0(h_0(z)))$ is also Lipschitz continuous. Therefore, by [34, Theorem 3.2], $\dot{z} = F_P(z, \kappa_0(h_0(z)))$ has no finite time escape and each maximal solution to \mathcal{H}_0 is unique. To show that each maximal solution to \mathcal{H}_0 is bounded, we use the Lyapunov function in (16), defined for each $z \in \mathbb{R}^{2n}$. Then, solutions to $\dot{z} = F_P(z, \kappa_0(h_0(z)))$ starting from any c_V -sublevel set $W := \{z \in \mathbb{R}^{2n} : V_0(z) \leq c_V\}$, $c_V \geq 0$, remains in such a set for all time since V_0 satisfies

$$\dot{V}_0(z) = \langle \nabla V_0(z), F_P(z, \kappa_0(h_0(z))) \rangle = -\lambda |z_2|^2 \leq 0 \quad (74)$$

for each $z \in \mathbb{R}^{2n}$, since λ is positive. Then, to show that V_0 is radially unbounded, we derive class- \mathcal{K}_∞ functions α_1 and α_2 such that²⁰, for all $z \in$

¹⁹*Finite escape time* describes when there exists a solution $t \mapsto x(t)$ to a continuous-time nonlinear system that satisfies $\lim_{t \nearrow t_e} |x(t)| = \infty$ for some finite time t_e .

²⁰Since L has quadratic growth away from z_1^* by Assumption 2.2, then the choice of α_1 comes from lower bounding $L(z_1) - L^*$ in V_0 via Assumption 2.2. The choice of α_2 comes from the following: since L is \mathcal{C}^1 , convex, and has a single minimizer by Assumption 2.1, then the expression $L(z_1) - L^*$ in V_0 is upper bounded using the definition of convexity in Footnote 6, by the same process that $L(z_1) - L^*$ is upper bounded in (14), to get (18). Then, $|\nabla L(z_1)|$ in (18) is upper bounded via Assumption 2.4 with $u_1 = z_1^*$ and $w_1 = z_1$.

\mathbb{R}^{2n} , with $z^* := (z_1^*, 0)$,

$$\begin{aligned} \alpha_1(|z - z^*|) &:= \min \left\{ \alpha\gamma, \frac{1}{2} \right\} |z - z^*|^2 \leq V_0(z) \\ &\leq \alpha_2(|z - z^*|) := \left(M\gamma + \frac{1}{2} \right) |z - z^*|^2. \end{aligned} \tag{75}$$

Since L has quadratic growth away from z_1^* with constant $\alpha > 0$ by Assumption 2.2, then the choice of α_1 comes from lower bounding V_0 as follows

$$\begin{aligned} V_0(z) &= \gamma(L(z_1) - L^*) + \frac{1}{2}|z_2|^2 \geq \alpha\gamma|z_1 - z_1^*|^2 + \frac{1}{2}|z_2|^2 \\ &\geq \min \left\{ \alpha\gamma, \frac{1}{2} \right\} |z - z^*|^2 = \alpha_1(|z - z^*|) \end{aligned} \tag{76}$$

for all $z \in \mathbb{R}^{2n}$. The choice of α_2 comes from the following. Since L is \mathcal{C}^1 , convex, has a single minimizer by Assumption 2.1, and since ∇L is Lipschitz continuous with constant $M > 0$ by Assumption 2.4, we upper bound V_0 in the following manner, using the definition of convexity in Footnote 6 to get (18), and then using the Lipschitz bound in Assumption 2.4 with $u_1 = z_1^*$ and $w_1 = z_1$ to upper bound (18), yielding

$$\begin{aligned} V_0(z) &= \gamma(L(z_1) - L^*) + \frac{1}{2}|z_2|^2 \leq \gamma|\nabla L(z_1)||z_1 - z_1^*| + \frac{1}{2}|z_2|^2 \\ &\leq M\gamma|z_1 - z_1^*|^2 + \frac{1}{2}|z_2|^2 \\ &\leq \left(M\gamma + \frac{1}{2} \right) |z - z^*|^2 = \alpha_2(|z_1 - z_1^*|) \end{aligned} \tag{77}$$

for all $z \in \mathbb{R}^{2n}$. Since (75) is satisfied for V_0 in (16) for all $z \in \mathbb{R}^{2n}$, then V_0 is radially unbounded (in z , relative to $\{z_1^*\} \times \{0\}$). Therefore, W is compact and, due to (74), forward invariant for \mathcal{H}_1 , that is, any nontrivial solution starting in the subset W is complete and stays in W . Therefore, each maximal solution to \mathcal{H}_0 , defined via (11), is bounded.

Next, we show that \mathcal{H}_1 in (12) has no finite time escape from $\mathbb{R}^{2n} \times \mathbb{R}_{\geq 0}$, and has unique solutions. Since \bar{d} and $\bar{\beta}$, defined via (5), are continuous, and

since by Assumption 2.1, L is \mathcal{C}^1 , then h_1 in (7) and κ_1 in (6) are also continuous. Furthermore, since by Assumption 2.4 ∇L is Lipschitz continuous, then h_1 in (7) and κ_1 in (6) are Lipschitz continuous which, in turn, means the map $z \mapsto F_P(z, \kappa_1(h_1(z, \tau), \tau))$ is Lipschitz continuous. Consequently, since the map $z \mapsto F_P(z, \kappa_1(h_1(z, \tau), \tau))$ is Lipschitz continuous and since the solution component τ of \mathcal{H}_1 increases linearly, then by [34, Theorem 3.2], \mathcal{H}_1 in (12) has no finite escape time from $\mathbb{R}^{2n} \times \mathbb{R}_{\geq 0}$ and each maximal solution to \mathcal{H}_0 is unique. Therefore, each maximal solution to \mathcal{H}_1 , defined via (12), is complete and unique.

Since \mathcal{H}_0 has no finite time escape from \mathbb{R}^{2n} and \mathcal{H}_1 has no finite time escape from $\mathbb{R}^{2n} \times \mathbb{R}_{\geq 0}$, then this means $\dot{x} = F(x)$ has no finite time escape from C for \mathcal{H} , as q does not change in C and as the state component τ is bounded in C , namely, the state component τ – which is always reset to 0 in D – increases linearly in C_1 and remains at 0 in C_0 . Therefore, there is no finite time escape from $C \cup D$, for solutions x to \mathcal{H} . Therefore, item (b) from Proposition D.1 does not hold.

C Proof of Proposition 4.4

The Lyapunov function V_1 , defined via (22), is positive definite with respect to \mathcal{A}_1 , defined via (67), since, by Assumption 2.1, L is \mathcal{C}^1 , convex, and has a unique minimizer z_1^* . Then, letting

$$\bar{v}_1(z, \tau) := z_1 + \bar{\beta}(\tau)z_2, \quad (78)$$

letting $\varphi(z, \tau) := \bar{a}(\tau) (\bar{a}(\tau) (z_1 - z_1^*) + z_2) + \frac{\zeta^2}{M} \nabla L(z_1)$, and since $\nabla V_1(z, \tau) = \left[\varphi(z, \tau) \quad (\bar{a}(\tau) (z_1 - z_1^*) + z_2) \quad \frac{d\bar{a}(\tau)}{d\tau} \langle z_1 - z_1^*, (\bar{a}(\tau) (z_1 - z_1^*) + z_2) \rangle \right]$, we evaluate the derivative of V_1 , using the map $z \mapsto F_P(z, \kappa_1(h_1(z, \tau), \tau))$, where F_P is defined in (4), κ_1 is defined via (6b), and h_1 is defined in (7), to yield

$$\begin{aligned} \dot{V}_1(z, \tau) &= \left\langle \nabla V_1(z, \tau), \begin{bmatrix} F_P(z, \kappa_1(h_1(z, \tau), \tau)) \\ 1 \end{bmatrix} \right\rangle \\ &= \left\langle \nabla V_1(z, \tau), \begin{bmatrix} \left[-2\bar{d}(\tau)z_2 - \frac{\zeta^2}{M} \nabla L(\bar{v}_1(z, \tau)) \right] \\ 1 \end{bmatrix} \right\rangle \\ &= \bar{a}(\tau) \langle \bar{a}(\tau) (z_1 - z_1^*) + z_2, z_2 \rangle + \frac{\zeta^2}{M} \langle z_2, \nabla L(z_1) \rangle - 2\bar{d}(\tau) |z_2|^2 \end{aligned}$$

$$\begin{aligned}
& -2\bar{d}(\tau)\bar{a}(\tau)\langle z_1 - z_1^*, z_2 \rangle - \frac{\bar{a}(\tau)\zeta^2}{M}\langle z_1 - z_1^*, \nabla L(\bar{v}_1(z, \tau)) \rangle \\
& - \frac{\zeta^2}{M}\langle z_2, \nabla L(\bar{v}_1(z, \tau)) \rangle + \bar{a}(\tau)\frac{d\bar{a}(\tau)}{d\tau}|z_1 - z_1^*|^2 + \frac{d\bar{a}(\tau)}{d\tau}\langle z_1 - z_1^*, z_2 \rangle \\
= & -\frac{\bar{a}(\tau)\zeta^2}{M}\langle z_1 - z_1^*, \nabla L(\bar{v}_1(z, \tau)) \rangle + \bar{a}(\tau)\frac{d\bar{a}(\tau)}{d\tau}|z_1 - z_1^*|^2 \\
& + (\bar{a}(\tau) - 2\bar{d}(\tau))|z_2|^2 + \left(\bar{a}^2(\tau) - 2\bar{d}(\tau)\bar{a}(\tau) + \frac{d\bar{a}(\tau)}{d\tau} \right) \langle z_1 - z_1^*, z_2 \rangle \\
& - \frac{\zeta^2}{M}\langle z_2, \nabla L(\bar{v}_1(z, \tau)) - \nabla L(z_1) \rangle \tag{79}
\end{aligned}$$

for all $(z, \tau) \in \mathbb{R}^{2n} \times \mathbb{R}_{\geq 0}$. Since L is \mathcal{C}^1 , convex, and has a unique minimizer by Assumption 2.1, then using the definition of convexity in Footnote 6 with $u_1 = z_1^*$ and $w_1 = \bar{v}_1(z, \tau)$, where \bar{v}_1 is defined via (78), we get

$$-\langle \bar{v}_1(z, \tau) - z_1^*, \nabla L(\bar{v}_1(z, \tau)) \rangle \leq -(L(\bar{v}_1(z, \tau)) - L^*) \tag{80}$$

for each $z \in \mathbb{R}^{2n}$ and $\tau \in \mathbb{R}_{\geq 0}$. Using the definition of convexity in Footnote 6 with $u_1 = \bar{v}_1(z, \tau)$, where \bar{v}_1 is defined via (78), and $w_1 = z_1$ yields

$$\langle \nabla L(z_1), \bar{\beta}(\tau)z_2 \rangle \leq L(\bar{v}_1(z, \tau)) - L(z_1) \tag{81}$$

for each $z \in \mathbb{R}^{2n}$ and $\tau \in \mathbb{R}_{\geq 0}$. Combining (80) and (81) yields $-\langle \bar{v}_1(z, \tau) - z_1^*, \nabla L(\bar{v}_1(z, \tau)) \rangle + \langle \nabla L(z_1), \bar{\beta}(\tau)z_2 \rangle \leq -L(\bar{v}_1(z, \tau)) + L(\bar{v}_1(z, \tau)) - L(z_1) + L^*$. Then, rearranging terms gives, for all $z \in \mathbb{R}^{2n}$ and $\tau \in \mathbb{R}_{\geq 0}$,

$$\begin{aligned}
& -\langle z_1 - z_1^*, \nabla L(\bar{v}_1(z, \tau)) \rangle \\
& \leq -(L(z_1) - L^*) + \langle \bar{\beta}(\tau)z_2, \nabla L(\bar{v}_1(z, \tau)) - \nabla L(z_1) \rangle. \tag{82}
\end{aligned}$$

Substituting the bound in (82) into (79) yields

$$\begin{aligned}
\dot{V}_1(z, \tau) \leq & -\frac{\bar{a}(\tau)\zeta^2}{M} ((L(z_1) - L^*) - \langle \bar{\beta}(\tau)z_2, \nabla L(\bar{v}_1(z, \tau)) - \nabla L(z_1) \rangle) \\
& + \bar{a}(\tau)\frac{d\bar{a}(\tau)}{d\tau}|z_1 - z_1^*|^2 + (\bar{a}(\tau) - 2\bar{d}(\tau))|z_2|^2 \\
& + \left(\bar{a}^2(\tau) - 2\bar{d}(\tau)\bar{a}(\tau) + \frac{d\bar{a}(\tau)}{d\tau} \right) \langle z_1 - z_1^*, z_2 \rangle \\
& - \frac{\zeta^2}{M}\langle z_2, \nabla L(\bar{v}_1(z, \tau)) - \nabla L(z_1) \rangle \tag{83}
\end{aligned}$$

for all $(z, \tau) \in \mathbb{R}^{2n} \times \mathbb{R}_{\geq 0}$. Then, noticing that $\frac{\bar{a}(\tau)}{2} |\bar{a}(\tau) (z_1 - z_1^*) + z_2|^2 = \frac{\bar{a}^3(\tau)}{2} |z_1 - z_1^*|^2 + \bar{a}^2(\tau) \langle z_1 - z_1^*, z_2 \rangle + \frac{\bar{a}(\tau)}{2} |z_2|^2$, adding it to and subtracting it from (83), and rearranging terms, yields

$$\begin{aligned}
\dot{V}_1(z, \tau) &\leq -\bar{a}(\tau)V_1(z, \tau) + \bar{a}(\tau) \frac{d\bar{a}(\tau)}{d\tau} |z_1 - z_1^*|^2 + (\bar{a}(\tau) - 2\bar{d}(\tau)) |z_2|^2 \\
&\quad + \left(\bar{a}^2(\tau) - 2\bar{d}(\tau)\bar{a}(\tau) + \frac{d\bar{a}(\tau)}{d\tau} \right) \langle z_1 - z_1^*, z_2 \rangle + \frac{\bar{a}^3(\tau)}{2} |z_1 - z_1^*|^2 \\
&\quad + \frac{\bar{a}(\tau)}{2} |z_2|^2 + \bar{a}^2(\tau) \langle z_1 - z_1^*, z_2 \rangle \\
&\quad - \frac{\zeta^2}{M} (1 - \bar{\beta}(\tau)\bar{a}(\tau)) \langle z_2, \nabla L(\bar{v}_1(z, \tau)) - \nabla L(z_1) \rangle \\
&\leq -\bar{a}(\tau)V_1(z, \tau) + \left(\frac{\bar{a}^3(\tau)}{2} + \bar{a}(\tau) \frac{d\bar{a}(\tau)}{d\tau} \right) |z_1 - z_1^*|^2 \\
&\quad + \left(\frac{3\bar{a}(\tau)}{2} - 2\bar{d}(\tau) \right) |z_2|^2 \\
&\quad + \left(2\bar{a}^2(\tau) - 2\bar{d}(\tau)\bar{a}(\tau) + \frac{d\bar{a}(\tau)}{d\tau} \right) \langle z_1 - z_1^*, z_2 \rangle \\
&\quad - \frac{\zeta^2}{M} (1 - \bar{\beta}(\tau)\bar{a}(\tau)) \langle z_2, \nabla L(\bar{v}_1(z, \tau)) - \nabla L(z_1) \rangle \tag{84}
\end{aligned}$$

for all $(z, \tau) \in \mathbb{R}^{2n} \times \mathbb{R}_{\geq 0}$. Due to the definitions of the functions \bar{a} and \bar{d} , in (23) and (5), respectively, the cross term $\langle z_1 - z_1^*, z_2 \rangle$ vanishes since $2\bar{a}^2(\tau) - 2\bar{d}(\tau)\bar{a}(\tau) + \frac{d\bar{a}(\tau)}{d\tau} = 2 \left(\frac{2}{\tau+2} \right)^2 - 2 \left(\frac{3}{2(\tau+2)} \right) \left(\frac{2}{\tau+2} \right) - \frac{2}{(\tau+2)^2} = 0$. Moreover, the definitions of the functions \bar{d} and \bar{a} lead to the $|z_1 - z_1^*|^2$ and $|z_2|^2$ terms in (84) vanishing due to $\frac{\bar{a}^3(\tau)}{2} + \bar{a}(\tau) \frac{d\bar{a}(\tau)}{d\tau} = \frac{\left(\frac{2}{\tau+2}\right)^3}{2} + \left(\frac{2}{\tau+2}\right) \left(-\frac{2}{(\tau+2)^2}\right) = 0$ and $\frac{3\bar{a}(\tau)}{2} - 2\bar{d}(\tau) = \frac{3\left(\frac{2}{\tau+2}\right)}{2} - 2 \left(\frac{3}{2(\tau+2)}\right) = 0$. The bound in (84) reduces to

$$\dot{V}_1(z, \tau) \leq -\bar{a}(\tau)V_1(z, \tau) - \frac{\zeta^2}{M} (1 - \bar{\beta}(\tau)\bar{a}(\tau)) \langle z_2, \nabla L(\bar{v}_1(z, \tau)) - \nabla L(z_1) \rangle \tag{85}$$

for all $(z, \tau) \in \mathbb{R}^{2n} \times \mathbb{R}_{\geq 0}$. By Assumption 2.1, L is \mathcal{C}^1 and convex. By [31, Theorem 2.1.3], a function L is \mathcal{C}^1 and convex if and only if, for each

$w_1, u_1 \in \mathbb{R}^n$,

$$\langle \nabla L(w_1) - \nabla L(u_1), w_1 - u_1 \rangle \geq 0. \quad (86)$$

Then, since $\bar{\beta}(\tau) \geq 0$ for all $t \geq 0$, using the bound in (86) with $w_1 = \bar{v}_1(z, \tau)$, where \bar{v}_1 is defined in (78), and $u_1 = z_1$, we get, for all $z \in \mathbb{R}^{2n}$ and $\tau \in \mathbb{R}_{\geq 0}$,

$$\begin{aligned} \langle \bar{v}_1(z, \tau) - z_1, \nabla L(\bar{v}_1(z, \tau)) - \nabla L(z_1) \rangle &= \\ \bar{\beta}(\tau) \langle z_2, \nabla L(\bar{v}_1(z, \tau)) - \nabla L(z_1) \rangle &\geq 0 \\ -\bar{\beta}(\tau) \langle z_2, \nabla L(\bar{v}_1(z, \tau)) - \nabla L(z_1) \rangle &\leq 0 \\ -\langle z_2, \nabla L(\bar{v}_1(z, \tau)) - \nabla L(z_1) \rangle &\leq 0 \end{aligned} \quad (87)$$

Therefore, since $1 - \bar{\beta}(\tau)\bar{a}(\tau) \geq 0$, due to \bar{a} , defined via (23), equaling 1 at $\tau = 0$ and monotonically decreasing toward zero (but being always positive) as τ tends to ∞ , and due to $\bar{\beta}$, defined via (5), equaling 0 at $\tau = 0$ and monotonically increasing to 1 as τ tends to ∞ , we use (87) to upper bound the last term of (85) as follows:

$$-\frac{\zeta^2}{M} (1 - \bar{\beta}(\tau)\bar{a}(\tau)) \langle z_2, \nabla L(\bar{v}_1(z, \tau)) - \nabla L(z_1) \rangle \leq 0 \quad (88)$$

This leads to, $z \in \mathbb{R}^{2n}$ and $\tau \in \mathbb{R}_{\geq 0}$,

$$\dot{V}_1(z, \tau) \leq -\bar{a}(\tau)V_1(z, \tau). \quad (89)$$

Applying Grönwall's Inequality to (89), namely,

$$\begin{aligned} V_1(z(t), t) &\leq V_1(z(0), 0) \exp\left(-\int_0^t \bar{a}(\tau) d\tau\right) \\ &= V_1(z(0), 0) \exp(-2 \ln(t+2) - 2 \ln(2)) \\ &= V_1(z(0), 0) \exp\left(-\ln\left(\frac{t+2}{2}\right)^2\right) \\ &= V_1(z(0), 0) \left(\frac{1}{\exp\left(\ln\left(\frac{t+2}{2}\right)^2\right)}\right) \\ &= \frac{4}{(t+2)^2} V_1(z(0), 0) \end{aligned}$$

shows that each maximal solution $t \mapsto (z(t), \tau(t))$ to the closed-loop algorithm \mathcal{H}_1 , such that $\tau(0) = 0$, satisfies (55), for all $t \geq 0$.

D General Results for Hybrid Systems

The following proposition, from [23], is used to prove the existence of solutions to the hybrid closed-loop system.

Proposition D.1 (*Basic existence of solutions*): *Let $\mathcal{H} = (C, F, D, G)$ satisfy Definition 2.7. Take an arbitrary $\xi \in C \cup D$. If $\xi \in D$ or*

(VC) there exists a neighborhood U of ξ such that for every $x \in U \cap C$,

$$F(x) \cap T_C(x) \neq \emptyset,$$

then there exists a nontrivial solution x to \mathcal{H} with $x(0, 0) = \xi$. If (VC) holds for every $\xi \in C \setminus D$, then there exists a nontrivial solution to \mathcal{H} from every initial point in $C \cup D$, and every²¹ $x \in \mathcal{S}_{\mathcal{H}}$ satisfies exactly one of the following conditions:

- (a) x is complete;
- (b) $\text{dom } x$ is bounded and the interval I^J , where $J = \sup_j \text{dom } x$, has nonempty interior and $t \mapsto x(t, J)$ is a maximal solution to $\dot{z} \in F(z)$, in fact $\lim_{t \rightarrow T} |x(t, J)| = \infty$, where $T = \sup_t \text{dom } x$;
- (c) $x(T, J) \notin C \cup D$, where $(T, J) = \sup \text{dom } x$.

Furthermore, if $G(D) \subset C \cup D$, then (c) above does not occur.

The following definition, from [22, Definition 3.17], describes the basic properties that a function must satisfy to serve as a Lyapunov function for the hybrid closed-loop algorithm \mathcal{H} .

Definition D.2 (Lyapunov function candidate) *The sets $\mathcal{U}, \mathcal{A} \subset \mathbb{R}^n$, and the function $V : \text{dom } V \rightarrow \mathbb{R}$ define a Lyapunov function candidate on \mathcal{U} with respect to \mathcal{A} for the hybrid closed-loop system $\mathcal{H} = (C, F, D, G)$ if the following conditions hold:*

1. $(\overline{C} \cup D \cup G(D)) \cup \mathcal{U} \subset \text{dom } V$;

²¹The set $\mathcal{S}_{\mathcal{H}}$ contains all maximal solutions to \mathcal{H} .

2. \mathcal{U} contains an open neighborhood of $\mathcal{A} \cap (C \cup D \cup G(D))$;
3. V is continuous on \mathcal{U} and locally Lipschitz on an open set containing $\overline{C} \cap \mathcal{U}$;
4. V is positive definite on $\overline{C} \cup D \cup G(D)$ with respect to \mathcal{A} .

The following theorem is used to prove the uniform global asymptotic stability of the hybrid closed-loop system, via Lyapunov stability and an invariance principle.

Theorem D.3 (*Hybrid Lyapunov theorem*): Given sets $\mathcal{U}, \mathcal{A} \subset \mathbb{R}^n$ and a function $V : \text{dom } V \rightarrow \mathbb{R}$ defining a Lyapunov candidate on \mathcal{U} with respect to \mathcal{A} for the closed-loop hybrid system $\mathcal{H} = (C, F, D, G)$, suppose

- \mathcal{H} satisfies the hybrid basic conditions;
- \mathcal{A} is compact and \mathcal{U} contains a nonzero open neighborhood of \mathcal{A} ;
- \dot{V} and ΔV satisfy

$$\dot{V}(x) = \max_{\xi \in F(x)} \langle \nabla V(x), \xi \rangle \leq 0 \quad \forall x \in C \cap \mathcal{U} \quad (90)$$

$$\Delta V(x) := \max_{\xi \in G(x)} V(\xi) - V(x) \leq 0 \quad \forall x \in D \cap \mathcal{U} \quad (91)$$

Then \mathcal{A} is stable. Furthermore, \mathcal{A} is attractive and, hence, pre-asymptotically stable if any of the following conditions hold:

1. Strict decrease during flows and jumps:

$$\dot{V}(x) < 0 \quad \forall x \in (C \cap \mathcal{U}) \setminus \mathcal{A} \quad (92)$$

$$\Delta V(x) < 0 \quad \forall x \in (D \cap \mathcal{U}) \setminus \mathcal{A} \quad (93)$$

2. Strict decrease during flows and no instantaneous Zeno:
 - (a) $\dot{V}(x) < 0$ for each $x \in (C \cap \mathcal{U}) \setminus \mathcal{A}$,
 - (b) any instantaneous Zeno solution x to \mathcal{H} where $\text{rge } x \subset \mathcal{U}$ converges to \mathcal{A} ;

3. Strict decrease during jumps and no complete continuous solution:

- (a) $\Delta V(x) < 0$ for each $x \in (D \cap \mathcal{U}) \setminus \mathcal{A}$,
- (b) any complete continuous solution x to \mathcal{H} where $\text{rge } x \subset \mathcal{U}$ converges to \mathcal{A} ;

4. *Weak decrease during flows and jumps:* for each $\chi \in \mathcal{U}$ with $r := V(\chi) > 0$ there is no complete solution x to \mathcal{H} , $x(0,0) = \chi$ such that

$$\text{rge } x \subset \{x : V(x) = r\} \cap \mathcal{U} \quad (94)$$

and the set \mathcal{U} is the subset of the basin of pre-attraction.

Observe that, if the set \mathcal{A} is pre-asymptotically stable via Theorem D.3 and the Lyapunov function V also has compact sublevel sets, namely, for each $c_V > 0$, $\{x : V(x) \leq c_V\}$ is compact, then the origin is *globally pre-asymptotically stable*.

The following result is used to show that, when a hybrid closed-loop algorithm \mathcal{H} has a set \mathcal{A} globally asymptotically stable, then when \mathcal{H} satisfies the hybrid basic conditions, the set \mathcal{A} is also uniformly globally asymptotically stable²² for \mathcal{H} .

Theorem D.4 (*Pre-asymptotic stability implies \mathcal{KL} pre-asymptotic stability*): Suppose that the hybrid closed-loop system \mathcal{H} satisfies the hybrid basic conditions and that a compact set \mathcal{A} is pre-asymptotically stable with basin of pre-attraction $\mathcal{B}_{\mathcal{A}}^p$. Then, $\mathcal{B}_{\mathcal{A}}^p$ is open and \mathcal{A} is \mathcal{KL} pre-asymptotically stable on $\mathcal{B}_{\mathcal{A}}^p$ for \mathcal{H} ; namely, there exists a function $\beta \in \mathcal{KL}$ such that

$$|x(0,0)|_{\mathcal{A}} \leq \beta(|x(0,0)|_{\mathcal{A}}, t+j) \quad \forall (t,j) \in \text{dom } x \quad (95)$$

for each $x \in \mathcal{S}_{\mathcal{H}}(\mathcal{B}_{\mathcal{A}}^p)$.

For Proposition 4.1 and Theorem D.6 we use the following definition of weak invariance, from [23].

Definition D.5 (Weak invariance) Given a hybrid system \mathcal{H} , a set $S \subset \mathbb{R}^n$ is said to be

- *weakly forward invariant* if for every $\xi \in S$ there exists at least one complete $x \in \mathcal{S}_{\mathcal{H}}(\xi)$ with $\text{rge } x \subset S$;

²²Uniform global asymptotic stability allows an equivalent characterization involving a class- \mathcal{KL} function [23].

- *weakly backward invariant if for every $\xi \in S$ and every $T > 0$, there exists at least one $x \in \mathcal{S}_{\mathcal{H}}(S)$ such that for some $(t^*, j^*) \in \text{dom } x$, $t^* + j^* \geq T$, it is the case that $x(t^*, j^*) = \xi$ and $x(t, j) \in S$ for all $(t, j) \in \text{dom } x$ with $t + j \leq t^* + j^*$;*
- *weakly invariant if it is both weakly forward invariant and weakly backward invariant.*

The following *hybrid invariance principle*, from [22, Theorem 3.23], is used to establish attractivity when only a “weak” Lyapunov function is available – meaning that the function does not strictly decrease along both flows and jumps of the hybrid system. It is also useful to check where particular solutions of interest converge to.

Theorem D.6 (*Hybrid Invariance Principle*): *Given a hybrid closed-loop system $\mathcal{H} = (C, F, D, G)$ with state $x \in \mathbb{R}^n$ satisfying the hybrid basic conditions, nonempty $\mathcal{U} \subset \mathbb{R}^n$, and a function $V : \text{dom } V \rightarrow \mathbb{R}$, suppose that D.2 is satisfied, and that (90) and (91) hold. With $X := C \cup D \cup G(D)$, we employ the following definitions:*

$$V^{-1}(r) := \{x \in X : V(x) = r\} \quad (96)$$

$$\dot{V}^{-1}(0) := \left\{x \in C : \dot{V}(x) = 0\right\} \quad (97)$$

$$\Delta V^{-1}(0) := \{x \in D : \Delta V(x) = 0\} \quad (98)$$

Let x be a precompact solution to \mathcal{H} with $\overline{\text{rge } x} \subset \mathcal{U}$. Then, for some $r \in V(\mathcal{U} \cap X)$, the following hold:

1. *The solution x converges to the largest weakly invariant set in*

$$V^{-1}(r) \cap \mathcal{U} \cap \left[\dot{V}^{-1}(0) \cup (\Delta V^{-1}(0) \cap G(\Delta V^{-1}(0))) \right]; \quad (99)$$

2. *The solution x converges to the largest weakly invariant set in*

$$V^{-1}(r) \cap \mathcal{U} \cap \Delta V^{-1}(0) \cap G(\Delta V^{-1}(0)) \quad (100)$$

if in addition the solution X is Zeno;

3. The solution x converges to the largest weakly invariant set in

$$V^{-1}(r) \cap \mathcal{U} \cap \dot{V}^{-1}(0) \quad (101)$$

if, in addition, the solution x is such that, for some $a > 0$ and some $J \in \mathbb{N}$, $t_{j+1} - t_j > a$ for all $j \geq J$; i.e., the given solution x is such that the elapsed time between consecutive jumps is eventually bounded below by a positive constant a .

References

- [1] M. Muehlebach and M. Jordan, “A dynamical systems perspective on nesterov acceleration,” in *International Conference on Machine Learning*. PMLR, 2019, pp. 4656–4662.
- [2] B. Polyak and P. Shcherbakov, “Lyapunov functions: an optimization theory perspective,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 7456–7461, 2017.
- [3] 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.
- [4] W. Su, S. Boyd, and E. J. Candès, “A differential equation for modeling nesterov’s accelerated gradient method: Theory and insights,” *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 5312–5354, 2016.
- [5] W. Krichene, A. Bayen, and P. Bartlett, “Accelerated mirror descent in continuous and discrete time,” *Advances in neural information processing systems*, vol. 28, pp. 2845–2853, 2015.
- [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] A. C. Wilson, B. Recht, and M. I. Jordan, “A lyapunov analysis of momentum methods in optimization,” Mar. 2016.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] S. Michalowsky and C. Ebenbauer, “The multidimensional n-th order heavy ball method and its application to extremum seeking,” in *53rd IEEE Conference on Decision and Control*, 2014, pp. 2660–2666.
- [16] —, “Extremum control of linear systems based on output feedback,” in *55th IEEE Conference on Decision and Control*, 2016, pp. 2963–2968.

- [17] A. S. Kolarijani, P. M. Esfahani, and T. Keviczky, “Continuous-time accelerated methods via a hybrid control lens,” *IEEE Transactions on Automatic Control*, pp. 3425–3440, Oct. 2019.
- [18] ———, “Fast gradient-based methods with exponential rate: A hybrid control framework,” in *International Conference on Machine Learning*, 2018, pp. 2728–2736.
- [19] 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.
- [20] ———, “Robust hybrid zero-order optimization algorithms with acceleration via averaging in time,” *Automatica*, vol. 123, p. 109361, 2021.
- [21] 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.
- [22] R. G. Sanfelice, *Hybrid Feedback Control*. New Jersey: Princeton University Press, 2021.
- [23] R. Goebel, R. G. Sanfelice, and A. R. Teel, *Hybrid Dynamical Systems: Modeling, Stability, and Robustness*. New Jersey: Princeton University Press, 2012.
- [24] A. R. Teel and N. Kapoor, “Uniting local and global controllers,” in *1997 European Control Conference (ECC)*. IEEE, 1997, pp. 3868–3873.
- [25] 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.
- [26] J. I. Poveda and A. R. Teel, “The heavy-ball ODE with time-varying damping: Persistence of excitation and uniform asymptotic stability,” in *2020 American Control Conference (ACC)*. IEEE, 2020, pp. 773–778.

- [27] N. Flammarion and F. Bach, “From averaging to acceleration, there is only a step-size,” in *Conference on Learning Theory*, 2015, pp. 658–695.
- [28] 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.
- [29] 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.
- [30] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [31] Y. Nesterov, “Introductory lectures on convex optimization, vol. 87,” 2004.
- [32] D. Drusvyatskiy and A. S. Lewis, “Error bounds, quadratic growth, and linear convergence of proximal methods,” *Mathematics of Operations Research*, 2018.
- [33] H. Karimi, J. Nutini, and M. Schmidt, “Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2016, pp. 795–811.
- [34] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, New Jersey: Prentice Hall, 2002.
- [35] 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>
- [36] F. H. Clarke, *Optimization and nonsmooth analysis*. SIAM, 1990.