Resource-Aware Discretization of Accelerated Optimization Flows: the Heavy-Ball Dynamics Case

Miguel Vaquero Pol Mestres Jorge Cortés

***Abstract*—This paper proposes a methodology for discretizing accelerated optimization flows while retaining their convergence properties. Inspired by the success of resource-aware control in developing efficient closed-loop feedback implementations on digital systems, we view the last sampled state of the system as the resource to be aware of. We illustrate our design methodology for discretization on a newly introduced continuous-time dynamics, the heavy-ball dynamics with displaced gradient. Our algorithm design employs techniques from resource-aware control that, in the present context, have interesting parallelisms with the discrete-time implementation of optimization algorithms. These include derivative- and performance-based triggers to monitor the evolution of the Lyapunov function as a way of determining the stepsize, exploiting sampled information to enhance performance, and employing high-order holds using more accurate integrators of the original dynamics. Our approach gives rise to variable-stepsize discrete-time algorithms that retain by design the monotonically decreasing properties of the Lyapunov certificate of the continuous-time heavy-ball dynamics with displaced gradient.**

I. Introduction

A recent body of research seeks to understand the acceleration phenomena of first-order discrete optimization methods by means of models that evolve in continuous time. Roughly speaking, the idea is to study the behavior of ordinary differential equations (ODEs) which arise as continuous limits of discrete-time accelerated algorithms. The basic premise is that the availability of the powerful tools of the continuous realm, such as differential calculus, Lie derivatives, and Lyapunov stability theory, can be then brought to bear to analyze and explain the accelerated behavior of these flows, in turn providing insight into their discrete counterparts and possibly guiding the synthesis of novel discrete algorithms. Realizing this requires solving the question of how to discretize the continuous flows while retaining their accelerated convergence properties. In fact, the discretization of accelerated continuous-time flows has proven to be challenging, where retaining acceleration seems to depend largely on the particular ODE and the discretization method employed. This paper addresses this challenge for the continuous-time heavy-ball dynamics by taking advantage of the resource-aware control paradigm to develop a principled approach to the discretization of accelerated optimization flows.

Literature Review: The acceleration phenomenon goes back to the seminal paper [1] introducing the so-called heavy-ball method, which employed momentum terms to speed up the convergence of the classical gradient descent method. The heavy-ball method achieves optimal convergence rate in a neighborhood of the minimizer for arbitrary convex functions and global optimal convergence rate for quadratic objective functions. Later on, the work [2] proposed the Nesterov's accelerated gradient method and, employing the technique of estimating sequences, showed that it converges globally with optimal convergence rate for convex and strongly-convex smooth functions. The algebraic nature of the technique of estimating sequences does not fully explain the mechanisms behind the acceleration phenomenon, and this has motivated many approaches in the literature to provide fundamental understanding and insights. These include coupling dynamics [3], dissipativity theory [4], integral quadratic constraints [5], [6], and geometric arguments [7].

Of specific relevance to this paper is a recent line of research initiated by [8] that seeks to understand the acceleration phenomenon in first-order optimization methods by means of models that evolve in continuous time. [8] introduced a second-order ODE as the continuous limit of Nesterov's accelerated gradient method and characterized its accelerated convergence properties using Lyapunov stability analysis. The ODE approach to acceleration now includes the use of Hamiltonian dynamical systems [9], [10], inertial systems with Hessian-driven damping [11], and high-resolution ODEs [12], [13]. This body of research is also reminiscent of the classical dynamical systems approach to algorithms in optimization, see [14], [15]. The question of how to discretize the continuous flows while maintaining their accelerated convergence rates has also attracted significant attention, motivated by the ultimate goal of fully understanding the acceleration phenomenon and taking advantage of it to design better optimization algorithms. Interestingly, discretizations of these ODEs do not necessarily lead to acceleration [16]. In fact, explicit discretization schemes, like forward Euler, can even become numerically unstable after a few iterations [17]. Most of the discretization approaches found in the literature are based on the study of well-known integrators, including symplectic integrators [9], [18], Runge-Kutta integrators [19] or modifications of Nesterov's *three sequences* [17], [18], [20]. Our previous work [21] instead developed a variable-stepsize discretization using zero-order holds and state-triggers based on the derivative of the Lyapunov function of the original continuous flow. Here, we provide a comprehensive approach based on powerful tools from resource-aware control, including performance-based triggering and state holds that more effectively use sampled information. We apply them to the heavy-ball dynamics with displaced gradient, a new dynamics also introduced here that has accelerated convergence rate. Other recent approaches to the acceleration phenomena and the synthesis of optimization algorithms using control-

This work was supported by NSF Award ECCS-1917177.

MV is with the School of Human Sciences and Technology, IE University, mvaquero@faculty.ie.edu. PM and JC are with the Department of Mechanical and Aerospace Engineering, University of California, San Diego, {pmestre,cortes}@ucsd.edu.

theoretic notions and techniques include [22], which employs hybrid systems to design a continuous-time dynamics with a feedback regulator of the viscosity of the heavy-ball ODE to guarantee arbitrarily fast exponential convergence, and [23], which introduced an algorithm which alternates between two (one fast when far from the minimizer but unstable, and another slower but stable around the minimizer) continuous heavy-ball dynamics.

Statement of Contributions: This paper proposes a resource-aware control framework to the discretization of accelerated optimization flows that takes advantage of their Lyapunov certificates guaranteeing asymptotic convergence. Our presentation illustrates the application of this approach in the case of the continuous-time heavy-ball dynamics. We rely on the key observation that resource-aware control provides a principled way to go from continuous-time control design to real-time implementation with stability and performance guarantees by opportunistically prescribing when certain resource should be employed. The resource to be aware of is the last sampled state of the system, and hence what we seek to maximize is the stepsize of the resulting discrete-time algorithm. We consider objective functions that are strongly convex, and continuously differentiable with Lipschitz gradients.

Our first contribution is the introduction of a second-order differential equation which we term heavy-ball dynamics with displaced gradient. This dynamics generalizes the continuous-time heavy-ball dynamics analyzed in the literature by evaluating the gradient of the objective function taking into account the second-order nature of the flow. We establish that the proposed dynamics retains the same convergence properties as the original one while providing additional flexibility for design in the form of a parameter that can be tuned according to the designer's criteria.

Our second contribution is the synthesis of criteria that determine the variable stepsize of the discrete-time implementation of the heavy-ball dynamics with displaced gradient. We refer to these criteria as event- or self-triggered, depending on whether the stepsize is implicitly or explicitly defined. We employ derivative- and performance-based triggering to ensure the algorithm retains the decrease of the Lyapunov function of the continuous flow. In doing so, we face the challenge that the evaluation of this function requires knowledge of the unknown optimizer of the objective function. To circumvent this hurdle, we derive bounds on the evolution of the Lyapunov function that can be evaluated without knowledge of the optimizer and enable the construction of computable surrogates. These bounds critically rely on the characterization of the optimizer as a critical point of the objective function. We characterize the asymptotic convergence properties of the resulting discrete-time algorithms, establishing the existence of a minimum inter-event time and exponential performance guarantees with regards to the decrease of the objective function. Notice that the existence of continuous flows is key for the application of resource-aware related techniques, stressing the importance of studying continuous-time dynamics in optimization.

Our last two contributions provide ways of exploiting the sampled information to enhance the algorithm performance. Our third contribution is an implementation of the algorithms that adaptively adjusts the value of the gradient displacement parameter depending on the region of the space to which the state belongs. Our fourth and last contribution builds on the fact that the continuous-time heavy-ball dynamics can be decomposed as the sum of a second-order linear dynamics with a nonlinear forcing term corresponding to the gradient of the objective function. Building on this observation, we design a hold for the resource-aware implementation that uses the samples only on the nonlinear term, and integrates exactly the resulting linear system with constant forcing, resulting in a more accurate approximation of the evolution of the continuous flow. We establish the existence of a minimum inter-event time and characterize the performance with regards to the objective function of the resulting high-order-hold algorithm. Finally, we illustrate the proposed optimization algorithms in simulation, comparing them against the heavy-ball and Nesterov's accelerated gradient methods and showing superior performance to other discretization methods proposed in the literature.

We conclude by noting that the proposed methodology for discretization based on resource-aware control is applicable to other accelerated optimization flows. Of course, each dynamics, along with their corresponding Lyapunov functions, are different in each case, and this makes it necessary to carefully work out the proper mathematical bounds to obtain the desired computable surrogates for each specific dynamics. Nevertheless, the general framework described here, together with the specific instantiation in the case of the heavy-ball method, offers a promising roadmap to tackle the discretization of other accelerated optimization flows.

II. Preliminaries

This section presents basic notation and preliminaries.

A. Notation

We denote by $\mathbb{R}$ and $\mathbb{R}_{>0}$ the sets of real and positive real numbers, resp. All vectors are column vectors. We denote their scalar product by $\langle \cdot, \cdot \rangle$. We use $\|\cdot\|$ to denote the 2-norm in Euclidean space. Given $\mu \in \mathbb{R}_{>0}$, a continuously differentiable function f is μ-strongly convex if $f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle + \frac{\mu}{2} \|x - y\|^2$ for x, $y \in \mathbb{R}^n$. Given $L \in \mathbb{R}_{>0}$ and a function $f : X \to Y$ between two normed spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, f is L-Lipschitz if $\|f(x) - f(x')\|_Y \leq L \|x - x'\|_X$ for x, $x' \in X$. The functions we consider here are continuously differentiable, μ-strongly convex and have L-Lipschitz continuous gradient. We refer to the set of functions with all these properties by $\mathcal{S}^1_{\mu,L}(\mathbb{R}^n)$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is positive definite relative to x_* if $f(x_*) = 0$ and $f(x) > 0$ for $x \in \mathbb{R}^n \setminus \{x_*\}$.

B. Resource-Aware Control

Our work builds on ideas from resource-aware control to develop discretizations of continuous-time accelerated flows. Here, we provide a brief exposition of its basic elements and refer to [24], [25] for further details.

Given a controlled dynamical system $\dot{p} = X(p, u)$, with $p \in \mathbb{R}^n$ and $u \in \mathbb{R}^m$, assume we are given a stabilizing continuous

state-feedback $\mathsf{k} : \mathbb{R}^n \to \mathbb{R}^m$ so that the closed-loop system $\dot{p} = X(p, \mathsf{k}(p))$ has p_* as a globally asymptotically stable equilibrium point. Assume also that a Lyapunov function $V : \mathbb{R}^n \to \mathbb{R}$ is available as a certificate of the globally stabilizing nature of the controller. Here, we assume this takes the form

$$\dot{V} = \langle \nabla V(p), X(p, \mathsf{k}(p)) \rangle \leq -\frac{\sqrt{\mu}}{4} V(p), \tag{1}$$

for all $p \in \mathbb{R}^n$. Although exponential decay of V along the system trajectories is not necessary, we restrict our attention to this case as it arises naturally in our treatment.

Suppose we are given the task of implementing the controller signal over a digital platform, meaning that the actuator cannot be continuously updated as prescribed by the specification $u = \mathsf{k}(p)$. In such case, one is forced to discretize the control action along the execution of the dynamics, while making sure that stability is still preserved. A simple-to-implement approach is to update the control action *periodically*, i.e., fix $h > 0$, sample the state as $\{p(kh)\}_{k=0}^{\infty}$ and implement

$$\dot{p}(t) = X(p(t), \mathsf{k}(p(kh))), \quad t \in [kh, (k+1)h].$$

This approach requires h to be small enough to ensure that V remains a Lyapunov function and, consequently, the system remains stable. By contrast, in *resource-aware control*, one employs the information generated by the system along its trajectory to update the control action in an opportunistic fashion. Specifically, we seek to determine in a state-dependent fashion a sequence of times $\{t_k\}_{k=0}^{\infty}$, not necessarily uniformly spaced, such that p_* remains a globally asymptotically stable equilibrium for the system

$$\dot{p}(t) = X(p(t), \mathsf{k}(p(t_k))), \quad t \in [t_k, t_{k+1}]. \tag{2}$$

The main idea to accomplish this is to let the state sampling be guided by the principle of maintaining the same type of exponential decay (1) along the new dynamics. To do this, one defines triggers to ensure that this decay is never violated by prescribing a new state sampling. Formally, one sets $t_0 = 0$ and $t_{k+1} = t_k + \mathrm{step}(p(t_k))$, where the stepsize is defined by

$$\mathrm{step}(\hat{p}) = \min\{t > 0 \mid b(\hat{p}, t) = 0\}. \tag{3}$$

Notice that the function b represents a property that is satisfied as long as $b \leq 0$ along the dynamics, which motivates the definition (3) as the "last instant" where the desired property is guaranteed to hold, and the necessity to re-evaluate the sampled state when b vanishes. We refer to the criteria as *event-triggering* or *self-triggering* depending on whether the evaluation of the function b requires monitoring of the state p along the trajectory of (2) (ET) or just knowledge of its initial condition $\hat{p}$ (ST). The computational complexity of the triggering criteria depends on the specific form of the function b: for instance, if the solutions to the equation $b(\hat{p}, t) = 0$ in the variable t can be expressed explicitly as a function of $\hat{p}$, then the computational complexity is minimal, as determining the stepsize just consists of evaluating the corresponding expression of the solution. In general, the event-triggering approach has a higher computational complexity than the self-triggering one. The more stringent requirements to implement event-triggering lead to larger stepsizes versus the more conservative ones characteristic of self-triggering. In order for the state sampling to be implementable in practice, the inter-event times $\{t_{k+1} - t_k\}_{k=0}^{\infty}$ must be uniformly lower bounded by a positive minimum inter-event time, abbreviated MIET. In particular, the existence of a MIET rules out the existence of Zeno behavior, i.e., the possibility of an infinite number of triggers in a finite amount of time.

Depending on how the evolution of the function V is examined, we describe two types of triggering conditions. In both cases, for a given $\hat{p} \in \mathbb{R}^n$, we let $p(t; \hat{p})$ denote the solution of $\dot{p}(t) = X(p(t), \mathsf{k}(\hat{p}))$ with initial condition $p(0) = \hat{p}$:

Derivative-based trigger: In this case, b^{d} is defined as an upper bound of the expression $\frac{d}{dt} V(p(t; \hat{p})) + \frac{\sqrt{\mu}}{4} V(p(t; \hat{p}))$. This definition ensures that (1) is maintained along (2);

Performance-based trigger: In this case, b^{p} is defined as an upper bound of the expression $V(p(t; \hat{p})) - e^{-\frac{\sqrt{\mu}}{4} t} V(\hat{p})$. Note that this definition ensures that the integral version of (1) is maintained along (2).

In general, the performance-based trigger gives rise to stepsizes that are at least as large as the ones determined by the derivative-based approach, cf. [26]. This is because the latter prescribes an update as soon as the exponential decay is about to be violated, and therefore, does not take into account the fact that the Lyapunov function might have been decreasing at a faster rate since the last update. Instead, the performance-based approach reasons over the *accumulated decay* of the Lyapunov function since the last update, potentially yielding longer inter-sampling times.

A final point worth mentioning is that, in the event-triggered control literature, the notion of *resource* to be aware of can be many different things, beyond the actuator described above, including the sensor, sensor-controller communication, communication with other agents, etc. This richness opens the way to explore more elaborate uses of the sampled information beyond the zero-order hold in (2), something that we also leverage later in our presentation.

III. Problem Statement

We aim to design discretization procedures of accelerated continuous flows that solve unconstrained optimization problems. Given a function $f : \mathbb{R}^n \to \mathbb{R}$, we deal with problems of the form

$$\min_{x \in \mathbb{R}^n} f(x).$$

Our motivation here is to show that principled approaches to discretization can retain the accelerated convergence properties of continuous-time dynamics, fill the gap between the continuous and discrete viewpoints on optimization algorithms, and lead to the construction of new ones. Throughout the paper, we focus on the continuous-time version of the celebrated heavy-ball method [1]. Assume f belongs to $\mathcal{S}^1_{\mu,L}(\mathbb{R}^n)$ and let x_* be its unique minimizer. The heavy-ball method is known to have an optimal convergence rate in a neighborhood of the minimizer. For its continuous-time counterpart, consider the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2022.3171307, IEEE Transactions on Automatic Control

following family of second-order equations parametrized by the variable $s \in \mathbb{R}_{>0}$, proposed in [12],

$$\begin{bmatrix}\dot{x}\\ \dot{v}\end{bmatrix} = \begin{bmatrix} v \\ -2\sqrt{\mu}v - (1+\sqrt{\mu s})\nabla f(x))\end{bmatrix}, \tag{4a}$$

$$x(0) = x_0, \quad v(0) = -\frac{2\sqrt{s}\nabla f(x_0)}{1+\sqrt{\mu s}}. \tag{4b}$$

We refer to this dynamics as $X_{\rm hb}$. The following result characterizes the convergence properties of (4) to $p_* = [x_*, 0]^T$.

Theorem III.1 ([12]). *Let $V : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be*

$$V(x,v) = (1+\sqrt{\mu s})(f(x) - f(x_*)) + \frac{1}{4}\left\|v\right\|^2 + \frac{1}{4}\left\|v + 2\sqrt{\mu}(x - x_*)\right\|^2, \tag{5}$$

which is positive definite relative to $[x_*, 0]^T$. Then $\dot{V} \leq -\frac{\sqrt{\mu}}{4}V$ along the dynamics (4) *and, as a consequence, $p_* = [x_*, 0]^T$ is globally asymptotically stable. Moreover, for $s \leq 1/L$, the exponential decrease of V implies*

$$f(x(t)) - f(x_*) \leq \frac{7\left\|x(0) - x_*\right\|^2}{2s} e^{-\frac{\sqrt{\mu}}{4}t}. \tag{6}$$

Theorem III.1, along with analogous results [12] for the Nesterov's accelerated gradient descent, serves as an inspiration to build Lyapunov functions that help to explain the accelerated convergence rate of the discrete-time methods.

The problem we seek to solve is establishing a way to discretize the continuous flow while retaining its accelerated convergence properties. Inspired by the success of resource-aware control in developing efficient closed-loop feedback implementations on digital systems, here we present a discretization approach to accelerated optimization flows using resource-aware control. At the basis of the approach taken here is the observation that the convergence rate (6) of the continuous flow is a direct consequence of the Lyapunov nature of the function (5). In fact, the integration of $\dot{V} \leq -\frac{\sqrt{\mu}}{4}V$ along the system trajectories yields

$$V(x(t), v(t)) \leq e^{-\frac{\sqrt{\mu}}{4}t} V(x(0), v(0)).$$

Since $f(x(t)) - f(x_*) \leq V(x(t), v(t))$, we deduce

$$f(x(t)) - f(x_*) \leq e^{-\frac{\sqrt{\mu}}{4}t} V(x(0), v(0)) = \mathcal{O}(e^{-\frac{\sqrt{\mu}}{4}t}).$$

The characterization of the convergence rate via the decay of the Lyapunov function is indeed common among accelerated optimization flows. This observation motivates the resource-aware approach to discretization pursued here, where the resource that we aim to use efficiently is the sampling of the state itself. By doing so, the ultimate goal is to give rise to large stepsizes that take maximum advantage of the decay of the Lyapunov function (and consequently of the accelerated nature) of the continuous-time dynamics in the resulting discrete-time implementation.

IV. Resource-Aware Discretization of Continuous-Time Heavy-Ball Dynamics

In this section we propose a discretization of accelerated optimization flows using state-dependent triggering and analyze the properties of the resulting discrete-time algorithm. For convenience, we use the shorthand notation $p = [x, v]^T$. In following with the exposition in Section II-B, we start by considering the zero-order hold implementation $\dot{p} = X_{\rm hb}(\hat{p})$, $p(0) = \hat{p}$ of the heavy-ball dynamics (4),

$$\dot{x} = \hat{v}, \tag{7a}$$

$$\dot{v} = -2\sqrt{\mu}\hat{v} - (1+\sqrt{\mu s})\nabla f(\hat{x}). \tag{7b}$$

Note that the solution trajectory takes the form $p(t) = \hat{p} + tX_{\rm hb}(\hat{p})$, which in discrete-time terminology corresponds to a forward-Euler discretization of (4). Component-wise, we have

$$x(t) = \hat{x} + t\hat{v},$$
$$v(t) = \hat{v} - t\big(2\sqrt{\mu}\hat{v} + (1+\sqrt{\mu s})\nabla f(\hat{x})\big).$$

As we pointed out in Section II-B, the use of sampled information opens the way to more elaborated constructions than the zero-order hold in (7). As an example, given the second-order nature of the heavy-ball dynamics, it would seem reasonable to leverage the (position, velocity) nature of the pair $(\hat{x}, \hat{v})$ (meaning that, at position $\hat{x}$, the system is moving with velocity $\hat{v}$) in approximating the gradient term ∇f by employing the modified zero-order hold:

$$\dot{x} = \hat{v}, \tag{8a}$$

$$\dot{v} = -2\sqrt{\mu}\hat{v} - (1+\sqrt{\mu s})\nabla f(\hat{x} + a\hat{v}), \tag{8b}$$

where $a \geq 0$. Note that the trajectory of (8) corresponds to the forward-Euler discretization of the continuous-time dynamics

$$\begin{bmatrix}\dot{x}\\ \dot{v}\end{bmatrix} = \begin{bmatrix} v \\ -2\sqrt{\mu}v - (1+\sqrt{\mu s})\nabla f(x + av))\end{bmatrix}, \tag{9}$$

We refer to this as the *heavy-ball dynamics with displaced gradient* and denote it by $X^a_{\rm hb}$. Note that (8) and (9) with $a = 0$ recover (7) and (4), respectively, so the presence of the parameter a provides additional richness in the type of dynamics considered, which as we show later, has important implications in providing flexibility for the design of discrete-time algorithms. In order to pursue the resource-aware approach laid out in Section II-B with the modified zero-order hold in (8), we first need to characterize the asymptotic convergence properties of the heavy-ball dynamics with displaced gradient, which we tackle next.

Remark IV.1. (Connection between the use of sampled information and high-resolution-ODEs). A number of works [27], [28], [29] have explored formulations of Nesterov's accelerated that employ displaced-gradient-like terms similar to the one used above. Here, we make this connection explicit. Given Nesterov's algorithm

$$y_{k+1} = x_k - s\nabla f(x_k)$$
$$x_{k+1} = y_{k+1} + \frac{1-\sqrt{\mu s}}{1+\sqrt{\mu s}}(y_{k+1} - y_k)$$

the work [12] obtains the following limiting high-resolution ODE

$$\ddot{x} + 2\sqrt{\mu}\dot{x} + \sqrt{s}\nabla^2 f(x)\dot{x} + (1+\sqrt{\mu s})\nabla f(x) = 0. \quad (10)$$

Interestingly, considering instead the evolution of the y-variable and applying similar arguments to the ones in [12], one instead obtains

$$\ddot{y} + 2\sqrt{\mu}\dot{y} + (1+\sqrt{\mu s})\nabla f\big(y + \frac{\sqrt{s}}{1+\sqrt{\mu s}}\dot{y}\big) = 0, \quad (11)$$

which corresponds to the continuous heavy-ball dynamics in (4) evaluated with a displaced gradient, i.e., (9). Even further, if we Taylor expand the last term in (11) as

$$\nabla f(y + \frac{\sqrt{s}}{1+\sqrt{\mu s}}\dot{y}) = \nabla f(y) + \nabla^2 f(y)\frac{\sqrt{s}}{1+\sqrt{\mu s}}\dot{y} + \mathcal{O}(s)$$

and disregard the $\mathcal{O}(s)$ term, we recover (10). This shows that (11) is just (10) with extra higher-order terms in s, and provides evidence of the role of gradient displacement in enlarging the modeling capabilities of high-resolution ODEs. •

A. Asymptotic Convergence of Heavy-Ball Dynamics with Displaced Gradient

In this section, we study the asymptotic convergence of heavy-ball dynamics with displaced gradient. Interestingly, for a sufficiently small, this dynamics enjoys the same convergence properties as the dynamics (4), as the following result shows.

Theorem IV.2. (Global asymptotic stability of heavy-ball dynamics with displaced gradient). *Let $\beta_1, \dots, \beta_4 > 0$ be*

$$\beta_1 = \sqrt{\mu_s}\mu, \quad \beta_2 = \frac{\sqrt{\mu_s}L}{\sqrt{\mu}},$$
$$\beta_3 = \frac{13\sqrt{\mu}}{16}, \quad \beta_4 = \frac{4\mu^2\sqrt{s} + 3L\sqrt{\mu}\sqrt{\mu_s}}{8L^2},$$

where, for brevity, $\sqrt{\mu_s} = 1 + \sqrt{\mu s}$, and define

$$a_1^* = \frac{2}{\beta_2^2}\Big(\beta_1\beta_4 + \sqrt{\beta_2^2\beta_3\beta_4 + \beta_1^2\beta_4^2}\Big). \quad (12)$$

Then, for $0 \le a \le a_1^*$, $\dot{V} \le -\frac{\sqrt{\mu}}{4}V$ along the dynamics (9) and, as a consequence, $p_* = [x_*, 0]^T$ is globally asymptotically stable. Moreover, for $s \le 1/L$, the exponential decrease of V implies (6) holds along the trajectories of $X_{\rm hb}^a$.

Proof. Note that

$$\begin{aligned}
&\langle \nabla V(p), X_{\rm hb}^a(p)\rangle + \frac{\sqrt{\mu}}{4}V(p) = \\
&= (1+\sqrt{\mu s})\langle \nabla f(x), v\rangle - \sqrt{\mu}\,\|v\|^2 - \sqrt{\mu_s}\langle \nabla f(x+av), v\rangle \\
&\quad - \sqrt{\mu}\sqrt{\mu_s}\langle \nabla f(x+av), x - x_*\rangle + \frac{\sqrt{\mu}}{4}V(x,v) \\
&= \underbrace{-\sqrt{\mu}\,\|v\|^2 - \sqrt{\mu}\sqrt{\mu_s}\langle \nabla f(x), x - x_*\rangle + \frac{\sqrt{\mu}}{4}V(x,v)}_{\text{Term I}} \\
&\quad \underbrace{-\sqrt{\mu_s}\langle \nabla f(x+av) - \nabla f(x), v\rangle}_{\text{Term II}} \\
&\quad \underbrace{-\sqrt{\mu}\sqrt{\mu_s}\langle \nabla f(x+av) - \nabla f(x), x - x_*\rangle}_{\text{Term III}},
\end{aligned}$$

where in the second equality, we have added and subtracted $\sqrt{\mu}\sqrt{\mu_s}\langle \nabla f(x), x - x_*\rangle$. Observe that "Term I" corresponds to $\langle \nabla V(p), X_{\rm hb}(p)\rangle + \frac{\sqrt{\mu}}{4}V(p)$ and is therefore negative by Theorem III.1. From [21], this term can be bounded as

$$\begin{aligned}
\text{Term I} \le\ & \frac{-13\sqrt{\mu}}{16}\|v\|^2 \\
& + \Big(\frac{4\mu^2\sqrt{s} + 3L\sqrt{\mu}\sqrt{\mu_s}}{8L^2}\Big)\|\nabla f(x)\|^2.
\end{aligned}$$

Let us study the other two terms. By strong convexity, we have $-\langle \nabla f(x+av) - \nabla f(x), v\rangle \le -a\mu\|v\|^2$, and therefore

$$\text{Term II} \le -a\sqrt{\mu_s}\mu\|v\|^2 \le 0.$$

Regarding Term III, one can use the L-Lipschitzness of ∇f and strong convexity to obtain

$$\text{Term III} \le \frac{a}{\mu}\sqrt{\mu}\sqrt{\mu_s}L\|v\|\,\|\nabla f(x)\|.$$

Now, using the notation in the statement, we can write

$$\begin{aligned}
&\langle \nabla V(p), X_{\rm hb}^a(p)\rangle + \frac{\sqrt{\mu}}{4}V(p) \quad (13)\\
&\le a\big(-\beta_1\|v\|^2 + \beta_2\|v\|\,\|\nabla f(x)\|\big) - \beta_3\|v\|^2 - \beta_4\|\nabla f(x)\|^2.
\end{aligned}$$

If $-\beta_1\|v\|^2 + \beta_2\|v\|\,\|\nabla f(x)\| \le 0$, then the RHS of (13) is negative for any $a \ge 0$. If $-\beta_1\|v\|^2 + \beta_2\|v\|\,\|\nabla f(x)\| > 0$, the RHS of (13) is negative if and only if

$$a \le \frac{\beta_3\|v\|^2 + \beta_4\|\nabla f(x)\|^2}{-\beta_1\|v\|^2 + \beta_2\|v\|\,\|\nabla f(x)\|}.$$

The RHS of this equation corresponds to $g(\|\nabla f(x)\|/\|\nabla v\|)$, with the function g defined in (A.3). From Lemma A.1, as long as $-\beta_1\|v\|^2 + \beta_2\|v\|\,\|\nabla f(x)\| > 0$, this function is lower bounded by

$$a_1^* = \frac{\beta_3 + \beta_4(z_{\rm root}^+)^2}{-\beta_1 + \beta_2 z_{\rm root}^+} > 0,$$

where $z_{\rm root}^+$ is defined in (A.4). This exactly corresponds to (12), concluding the result. □

Remark IV.3. (Adaptive displacement along the trajectories of heavy-ball dynamics with displaced gradient). From the proof of Theorem IV.2, one can observe that if (x, v) is such that $\underline{n} \le \|\nabla f(x)\| < \overline{n}$ and $\underline{m} \le \|v\| < \overline{m}$, for $\underline{n}, \overline{n}, \underline{m}, \overline{m} \in \mathbb{R}_{>0}$, then one can upper bound the LHS of (13) by

$$a(-\beta_1\underline{m}^2 + \beta_2\overline{m}\,\overline{n}) - \beta_3\underline{m}^2 - \beta_4\underline{n}^2.$$

If $-\beta_1\underline{m}^2 + \beta_2\overline{m}\,\overline{n} \le 0$, any $a \ge 0$ makes this expression negative. If instead $-\beta_1\underline{m}^2 + \beta_2\overline{m}\,\overline{n} \ge 0$, then a must satisfy

$$a \le \Big|\frac{\beta_3\underline{m}^2 + \beta_4\underline{n}^2}{-\beta_1\underline{m}^2 + \beta_2\overline{m}\,\overline{n}}\Big|. \quad (14)$$

This argument shows that over the region $R = \{(x,v) \mid \underline{n} \le \|\nabla f(x)\| < \overline{n}$ and $\underline{m} \le \|v\| < \overline{m}\}$, any $a \ge 0$ satisfying (14) ensures that $\dot{V} \le -\frac{\sqrt{\mu}}{4}V$, and hence the desired exponential decrease of the Lyapunov function. This observation opens

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2022.3171307, IEEE Transactions on Automatic Control

the way to modify the value of the parameter a adaptively along the execution of the heavy-ball dynamics with displaced gradient, depending on the region of state space visited by its trajectories. •

B. Triggered Design of Variable-Stepsize Algorithms

In this section we propose a discretization of the continuous heavy-ball dynamics based on resource-aware control. To do so, we employ the approaches to trigger design described in Section II-B on the dynamics $X^a_{\rm hb}$, whose forward-Euler discretization corresponds to the modified zero-order hold (8) of the heavy-ball dynamics.

Our starting point is the characterization of the asymptotic convergence properties of $X^a_{\rm hb}$ developed in Section IV-A. The trigger design necessitates of bounding the evolution of the Lyapunov function V in (5) for the continuous-time heavy-ball dynamics with displaced gradient along its zero-order hold implementation. However, this task presents the challenge that the definition of V involves the minimizer x_* of the optimization problem itself, which is unknown (in fact, finding it is the ultimate objective of the discrete-time algorithm we seek to design). In order to synthesize computable triggers, this raises the issue of bounding the evolution of V as accurately as possible while avoiding any requirement on the knowledge of x_*. We address this point by computing a surrogate of $\dot V + \frac{\sqrt{\mu}}{4}V$ which upper bounds it and enforcing the latter to be negative throughout the dynamics. The following result specifies the surrogate.

Proposition IV.4. (Upper bound for derivative-based triggering with zero-order hold). *Let $a \geq 0$ and define*

$$\begin{aligned} b^{\rm d}_{\rm ET}(\hat p, t; a) &= A_{\rm ET}(\hat p, t; a) + B_{\rm ET}(\hat p, t; a) + C_{\rm ET}(\hat p; a),\\ b^{\rm d}_{\rm ST}(\hat p, t; a) &= B^q_{\rm ST}(\hat p; a)t^2 + (A_{\rm ST}(\hat p; a) + B^l_{\rm ST}(\hat p; a))t\\ &\quad + C_{\rm ST}(\hat p; a), \end{aligned}$$

where

$$\begin{aligned} A_{\rm ET}(\hat p, t; a) &= 2\mu t\|\hat v\|^2 + \sqrt{\mu_s}\big(\langle \nabla f(\hat x + t\hat v) - \nabla f(\hat x), \hat v\rangle\\ &\quad + 2t\sqrt{\mu}\langle \nabla f(\hat x + a\hat v), \hat v\rangle + t\sqrt{\mu_s}\|\nabla f(\hat x + a\hat v)\|^2\big),\\ B_{\rm ET}(\hat p, t; a) &= \frac{\sqrt{\mu}t^2}{16}\|2\sqrt{\mu}\hat v + \sqrt{\mu_s}\nabla f(\hat x + a\hat v)\|^2\\ &\quad - \frac{t\mu}{4}\|\hat v\|^2 + \frac{\sqrt{\mu}\sqrt{\mu_s}}{4}\big(f(\hat x + t\hat v) - f(\hat x) +\\ &\quad - t\langle \hat v, \nabla f(\hat x + a\hat v)\rangle + \frac{t^2\sqrt{\mu_s}}{4}\|\nabla f(\hat x + a\hat v)\|^2\\ &\quad - \frac{t\sqrt{\mu}}{L}\|\nabla f(\hat x + a\hat v)\|^2 + t\sqrt{\mu}\langle a\hat v, \nabla f(\hat x + a\hat v)\rangle\big),\\ C_{\rm ET}(\hat p; a) &= -\frac{13\sqrt{\mu}}{16}\|\hat v\|^2 - \frac{\mu^2\sqrt{s}}{2}\frac{\|\nabla f(\hat x)\|^2}{L^2}\\ &\quad + \sqrt{\mu_s}\big(\frac{-3\sqrt{\mu}}{8L}\|\nabla f(\hat x)\|^2\\ &\quad + \sqrt{\mu}(f(\hat x) - f(\hat x + a\hat v)) + \sqrt{\mu}\|\nabla f(\hat x)\|\,\|a\hat v\|\\ &\quad - \frac{\mu^{3/2}}{2}\|a\hat v\|^2 - \langle \nabla f(\hat x + a\hat v) - \nabla f(\hat x), \hat v\rangle\\ &\quad + \sqrt{\mu}\langle \nabla f(\hat x + a\hat v), a\hat v\rangle\big),\\ A_{\rm ST}(\hat p; a) &= 2\mu\|\hat v\|^2 + \sqrt{\mu_s}\big(L\|\hat v\|^2 + 2\sqrt{\mu}\langle \nabla f(\hat x + a\hat v), \hat v\rangle\\ &\quad + \sqrt{\mu_s}\|\nabla f(\hat x + a\hat v)\|^2\big),\\ B^l_{\rm ST}(\hat p; a) &= \frac{\sqrt{\mu}}{4}\big(-\sqrt{\mu}\|\hat v\|^2 + \sqrt{\mu_s}(\langle \nabla f(\hat x)\\ &\quad - \nabla f(\hat x + a\hat v), \hat v\rangle - \frac{\sqrt{\mu}}{L}\|\nabla f(\hat x + a\hat v)\|^2\\ &\quad + \sqrt{\mu}\langle a\hat v, \nabla f(\hat x + a\hat v)\rangle)\big),\\ B^q_{\rm ST}(\hat p; a) &= \frac{\sqrt{\mu}}{16}\|2\sqrt{\mu}\hat v + \sqrt{\mu_s}\nabla f(\hat x + a\hat v)\|^2\\ &\quad + \frac{\sqrt{\mu}\sqrt{\mu_s}}{4}\big(\frac{L}{2}\|\hat v\|^2 + \frac{\sqrt{\mu_s}}{4}\|\nabla f(\hat x + a\hat v)\|^2\big),\\ C_{\rm ST}(\hat p; a) &= C_{\rm ET}(\hat p; a). \end{aligned}$$

Let $t \mapsto p(t) = \hat p + tX^a_{\rm hb}(\hat p)$ be the trajectory of the zero-order hold dynamics $\dot p = X^a_{\rm hb}(\hat p)$, $p(0) = \hat p$. Then, for $t \geq 0$,

$$\frac{d}{dt}V(p(t)) + \frac{\sqrt{\mu}}{4}V(p(t)) \leq b^{\rm d}_{\rm ET}(\hat p, t; a) \leq b^{\rm d}_{\rm ST}(\hat p, t; a).$$

The proof of this result is presented in Appendix A and critically relies on the characterization of the optimizer as a critical point of the objective function. The importance of Proposition IV.4 stems from the fact that the triggering conditions defined by $b^{\rm d}_{\#}$, $\# \in \{\rm ET, ST\}$, can be evaluated without knowledge of the optimizer x_*. We build on this result next to establish an upper bound for the performance-based triggering condition.

Proposition IV.5. (Upper bound for performance-based triggering with zero-order hold). *Let $a \geq 0$ and*

$$b^{\rm p}_{\#}(\hat p, t; a) = \int_0^t e^{\frac{\sqrt{\mu}}{4}\zeta} b^{\rm d}_{\#}(\hat p, \zeta; a)d\zeta,$$

for $\# \in \{\rm ET, ST\}$. Let $t \mapsto p(t) = \hat p + tX^a_{\rm hb}(\hat p)$ be the trajectory of the zero-order hold dynamics $\dot p = X^a_{\rm hb}(\hat p)$, $p(0) = \hat p$. Then, for $t \geq 0$,

$$V(p(t)) - e^{-\frac{\sqrt{\mu}}{4}t}V(\hat p) \leq e^{-\frac{\sqrt{\mu}}{4}t} b^{\rm p}_{\rm ET}(\hat p, t; a) \leq e^{-\frac{\sqrt{\mu}}{4}t} b^{\rm p}_{\rm ST}(\hat p, t; a).$$

Proof. We rewrite $V(p(t)) - e^{-\frac{\sqrt{\mu}}{4}t}V(\hat p) = e^{-\frac{\sqrt{\mu}}{4}t}(e^{\frac{\sqrt{\mu}}{4}t}V(p(t)) - V(\hat p))$, and note that

$$\begin{aligned} &e^{\frac{\sqrt{\mu}}{4}t}V(p(t)) - V(\hat p)\\ &\quad = \int_0^t \frac{d}{d\zeta}\big(e^{\frac{\sqrt{\mu}}{4}\zeta}V(p(\zeta)) - V(\hat p)\big)d\zeta\\ &\quad = \int_0^t e^{\frac{\sqrt{\mu}}{4}\zeta}\Big(\frac{d}{d\zeta}V(p(\zeta)) + \frac{\sqrt{\mu}}{4}V(p(\zeta))\Big)d\zeta. \end{aligned}$$

Note that the integrand corresponds to the derivative-based criterion bounded in Proposition IV.4. Therefore,

$$\begin{aligned} e^{\frac{\sqrt{\mu}}{4}t}V(p(t)) - V(\hat p) &\leq \int_0^t e^{\frac{\sqrt{\mu}}{4}\zeta} b^{\rm d}_{\rm ET}(\hat p, \zeta; a)d\zeta\\ &= b^{\rm p}_{\rm ET}(\hat p, t; a) \leq b^{\rm p}_{\rm ST}(\hat p, t; a) \end{aligned}$$

for $t \geq 0$, and the result follows. □

Propositions IV.4 and IV.5 provide us with the tools to determine the stepsize according to the derivative- and performance-based triggering criteria, respectively. For convenience, and following the notation in (3), we define the stepsizes

$$\text{step}^{\rm d}_{\#}(\hat p; a) = \min\{t > 0 \mid b^{\rm d}_{\#}(\hat p, t; a) = 0\}, \tag{15a}$$

$$\text{step}^{\text{p}}_{\#}(\hat{p}; a) = \min\{t > 0 \mid b^{\text{p}}_{\#}(\hat{p}, t; a) = 0\}, \tag{15b}$$

for $\# \in \{\text{ET}, \text{ST}\}$. Observe that, as long as $\hat{p} \neq p_* = [x_*, 0]^T$ and $0 \leq a \leq a_1^*$, we have $C_{\#}(\hat{p}; a) < 0$ for $\# \in \{\text{ST}, \text{ET}\}$ and, as a consequence, $b^{\text{d}}_{\#}(\hat{p}, 0; a) < 0$. The ET/ST terminology is justified by the following observation: in the ET case, the equation defining the stepsize is in general implicit in t, which in general requires a dedicated zero-finding routine to determine the stepsize. Instead, in the ST case, the equation defining the stepsize is explicit in t, and the stepsize can be readily determined by evaluating the expression. As a consequence, the ET implementation has a higher computational complexity than the ST one.

Equipped with this notation, we define the variable-stepsize algorithm described in Algorithm 1, which consists of following the dynamics (8) until the exponential decay of the Lyapunov function is violated as estimated by the derivative-based ($\diamond = \text{d}$) or the performance-based ($\diamond = \text{p}$) triggering condition. When this happens, the algorithm re-samples the state before continue flowing along (8).

Algorithm 1: Displaced-Gradient Algorithm

Design Choices: $\diamond \in \{\text{d}, \text{p}\}$, $\# \in \{\text{ET}, \text{ST}\}$
Initialization: Initial point (p_0), objective function (f), tolerance (ϵ), $a \geq 0$, $k = 0$
while $\|\nabla f(x_k)\| \geq \epsilon$ **do**
  Compute stepsize $\Delta_k = \text{step}^{\diamond}_{\#}(p_k; a)$
  Compute next iterate $p_{k+1} = p_k + \Delta_k X^a_{\text{hb}}(p_k)$
  Set $k = k + 1$
end

C. Convergence Analysis of Displaced-Gradient Algorithm

Here we study the convergence properties of the derivative- and performance-based implementations of the Displaced-Gradient Algorithm. In each case, we show that algorithm is implementable (i.e., it admits a MIET) and inherits the convergence rate from the continuous-time dynamics. The following result makes this precise in the case of the derivative-based implementation of Algorithm 1.

Theorem IV.6. (Convergence of derivative-based implementation of Displaced-Gradient Algorithm). *Let* $\hat{\beta}_1, \dots, \hat{\beta}_5 > 0$ *be*

$$\hat{\beta}_1 = \sqrt{\mu_s}(\frac{3\sqrt{\mu}}{2} + L), \qquad \hat{\beta}_2 = \sqrt{\mu}\sqrt{\mu_s}\frac{3}{2},$$
$$\hat{\beta}_3 = \frac{13\sqrt{\mu}}{16}, \qquad \hat{\beta}_4 = \frac{4\mu^2\sqrt{s} + 3L\sqrt{\mu}\sqrt{\mu_s}}{8L^2},$$
$$\hat{\beta}_5 = \sqrt{\mu_s}(\frac{5\sqrt{\mu}L}{2} - \frac{\mu^{3/2}}{2}),$$

and define

$$a_2^* = \alpha \min\Big\{\frac{-\hat{\beta}_1 + \sqrt{\hat{\beta}_1^2 + 4\hat{\beta}_5\hat{\beta}_3}}{2\hat{\beta}_5}, \frac{\hat{\beta}_4}{\hat{\beta}_2}\Big\}, \tag{16}$$

with $0 < \alpha < 1$. *Then, for* $0 \leq a \leq a_2^*$, $\diamond = \text{d}$, *and* $\# \in \{\text{ET}, \text{ST}\}$, *the variable-stepsize strategy in Algorithm 1 has the following properties*

(i) *the stepsize is uniformly lower bounded by the positive constant* $\text{MIET}(a)$, *where*

$$\text{MIET}(a) = -\nu + \sqrt{\nu^2 + \eta}, \tag{17}$$

$\eta = \min\{\eta_1, \eta_2\}$, $\nu = \max\{\nu_1, \nu_2\}$, *and*

$$\eta_1 = \frac{8a\sqrt{\mu_s}\left(a(\mu - 5L) - \frac{2L}{\sqrt{\mu}} - 3\right) + 13}{2\sqrt{\mu_s}L\left(3a^2\sqrt{\mu_s}L + 1\right) + 8\mu},$$
$$\eta_2 = -\frac{3\sqrt{\mu_s}\sqrt{\mu}L(4aL - 1) - 4\mu^2\sqrt{s}}{3\mu_s\sqrt{\mu}L^2},$$
$$\nu_1 = \frac{\mu\left(2a^3\sqrt{\mu_s}L^2 + a\sqrt{\mu_s} + 16\right)}{2\sqrt{\mu}\left(\sqrt{\mu_s}L\left(3a^2\sqrt{\mu_s}L + 1\right) + 4\mu\right)} + \frac{8\sqrt{\mu_s}L\left(2a^2\sqrt{\mu_s}L + 1\right)}{2\sqrt{\mu}\left(\sqrt{\mu_s}L\left(3a^2\sqrt{\mu_s}L + 1\right) + 4\mu\right)} + \frac{\sqrt{\mu_s}(aL(8aL + 1) + 4)}{\sqrt{\mu_s}L\left(3a^2\sqrt{\mu_s}L + 1\right) + 4\mu},$$
$$\nu_2 = \frac{a\mu + 8\sqrt{\mu_s} + 8\sqrt{\mu}}{3\sqrt{\mu_s}\sqrt{\mu}};$$

(ii) $\frac{d}{dt}V(p_k + tX^a_{\text{hb}}(p_k)) \leq -\frac{\sqrt{\mu}}{4}V(p_k + tX^a_{\text{hb}}(p_k))$ *for all* $t \in [0, \Delta_k]$ *and* $k \in \{0\} \cup \mathbb{N}$.

As a consequence, $f(x_{k+1}) - f(x_*) = \mathcal{O}(e^{-\frac{\sqrt{\mu}}{4}\sum_{i=0}^{k}\Delta_i})$.

Proof. Regarding fact (i), we prove the result for the ST-case, as the ET-case follows from $\text{step}^{\text{d}}_{\text{ET}}(\hat{p}; a) \geq \text{step}^{\text{d}}_{\text{ST}}(\hat{p}; a)$. We start by upper bounding $C_{\text{ST}}(\hat{p}; a)$ by a negative quadratic function of $\|\hat{v}\|$ and $\|\nabla f(\hat{x})\|$ as follows,

$$\begin{aligned} C_{\text{ST}}(\hat{p}; a) &= -\frac{13\sqrt{\mu}}{16}\|\hat{v}\|^2 + \sqrt{\mu_s}\frac{-3\sqrt{\mu}}{8L}\|\nabla f(\hat{x})\|^2 \\ &- \frac{\mu^2\sqrt{s}}{2L^2}\|\nabla f(\hat{x})\|^2 + \sqrt{\mu_s}\big(\sqrt{\mu}\underbrace{(f(\hat{x}) - f(\hat{x} + a\hat{v}))}_{\text{(a)}} \\ &+ \sqrt{\mu}\underbrace{\|\nabla f(\hat{x})\|\,\|a\hat{v}\|}_{\text{(b)}} - \frac{\mu^{3/2}}{2}\|a\hat{v}\|^2 \\ &+ \underbrace{\langle\nabla f(\hat{x}) - \nabla f(\hat{x} + a\hat{v}), \hat{v}\rangle}_{\text{(c)}} + \sqrt{\mu}\underbrace{\langle\nabla f(\hat{x} + a\hat{v}), a\hat{v}\rangle}_{\text{(d)}}\big). \end{aligned}$$

Using the L-Lipschitzness of the gradient and Young's inequality, we can easily upper bound

$$\begin{aligned} \text{(a)} &\leq \underbrace{\langle\nabla f(\hat{x} + a\hat{v}), -a\hat{v}\rangle + \frac{L}{2}a^2\|\hat{v}\|^2}_{\text{Using (A.1c)}} \\ &= \langle\nabla f(\hat{x} + a\hat{v}) - \nabla f(\hat{x}), -a\hat{v}\rangle + \frac{L}{2}a^2\|\hat{v}\|^2 \\ &\quad + \langle\nabla f(\hat{x}), -a\hat{v}\rangle \\ &\leq La^2\|\hat{v}\|^2 + \frac{L}{2}a^2\|\hat{v}\|^2 + a\big(\frac{\|\nabla f(\hat{x})\|^2}{2} + \frac{\|\hat{v}\|^2}{2}\big) \\ &= \frac{3La^2 + a}{2}\|\hat{v}\|^2 + \frac{a}{2}\|\nabla f(\hat{x})\|^2, \\ \text{(b)} &\leq a\big(\frac{\|\nabla f(\hat{x})\|^2}{2} + \frac{\|\hat{v}\|^2}{2}\big), \\ \text{(c)} &\leq La\|\hat{v}\|^2, \\ \text{(d)} &= \langle\nabla f(\hat{x} + a\hat{v}) - \nabla f(\hat{x}) + \nabla f(\hat{x}), a\hat{v}\rangle \end{aligned}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2022.3171307, IEEE Transactions on Automatic Control

$$\leq La^2 \|\hat{v}\|^2 + \langle \nabla f(\hat{x}), a\hat{v}\rangle$$
$$= \frac{2La^2+a}{2}\|\hat{v}\|^2 + \frac{a}{2}\|\nabla f(\hat{z})\|^2.$$

Note that, with the definition of the constants $\hat{\beta}_1, \dots, \hat{\beta}_5 > 0$ in the statement, we can write

$$C_{\mathrm{ST}}(\hat{p};a) \leq a\hat{\beta}_1\|\hat{v}\|^2 + a^2\hat{\beta}_5\|\hat{v}\|^2 + a\hat{\beta}_2\|\nabla f(\hat{x})\|^2 - \hat{\beta}_3\|\hat{v}\|^2 - \hat{\beta}_4\|\nabla f(\hat{x})\|^2.$$

Therefore, for $a \in [0, a_2^*]$, we have

$$a\hat{\beta}_1 + a^2\hat{\beta}_5 - \hat{\beta}_3 \leq a_2^*\hat{\beta}_1 + (a_2^*)^2\hat{\beta}_5 - \hat{\beta}_3 = -\gamma_1 < 0$$
$$a\hat{\beta}_2 - \hat{\beta}_4 \leq a_2^*\hat{\beta}_2 - \hat{\beta}_4 = -\gamma_2 < 0,$$

and hence $C_{\mathrm{ST}}(\hat{p};a) \leq -\gamma_1\|\hat{v}\|^2 - \gamma_2\|\nabla f(\hat{x})\|^2$. Similarly, introducing

$$\gamma_3 = 2a^2\mu_s L^2 + 2a^2\sqrt{\mu_s}\sqrt{\mu}L^2 + \sqrt{\mu_s}\sqrt{\mu} + \sqrt{\mu_s}L + 2\mu,$$
$$\gamma_4 = 2\mu_s + 2\sqrt{\mu_s}\sqrt{\mu},\ \gamma_5 = \frac{1}{8}a\sqrt{\mu_s}\left(2a^2\mu L^2 + \mu + 2\sqrt{\mu}L\right),$$
$$\gamma_6 = \frac{a\mu\sqrt{\mu_s}}{4},\ \gamma_7 = \frac{3}{8}a^2\mu_s\sqrt{\mu}L^2 + \frac{1}{8}\sqrt{\mu_s}\sqrt{\mu}L + \frac{\mu^{3/2}}{2},$$
$$\gamma_8 = \frac{3\mu_s\sqrt{\mu}}{8},$$

one can show that

$$A_{\mathrm{ST}}(\hat{p};a) \leq \hat{A}_{\mathrm{ST}}(\hat{p};a) = \gamma_3\|\hat{v}\|^2 + \gamma_4\|\nabla f(\hat{x})\|^2,$$
$$B^l_{\mathrm{ST}}(\hat{p};a) \leq \hat{B}^l_{\mathrm{ST}}(\hat{p};a) = \gamma_5\|\hat{v}\|^2 + \gamma_6\|\nabla f(\hat{x})\|^2,$$
$$B^q_{\mathrm{ST}}(\hat{p};a) \leq \hat{B}^q_{\mathrm{ST}}(\hat{p};a) = \gamma_7\|\hat{v}\|^2 + \gamma_8\|\nabla f(\hat{x})\|^2.$$

Thus, from (15a), we have

$$\mathrm{step}^{\mathrm{d}}_{\mathrm{ST}}(\hat{p};a) \geq \frac{-(\hat{A}_{\mathrm{ST}}(\hat{p};a) + \hat{B}^l_{\mathrm{ST}}(\hat{p};a))}{2\hat{B}^q_{\mathrm{ST}}(\hat{p};a)} + \sqrt{\left(\frac{\hat{A}_{\mathrm{ST}}(\hat{p};a) + \hat{B}^l_{\mathrm{ST}}(\hat{p};a)}{2\hat{B}^q_{\mathrm{ST}}(\hat{p};a)}\right)^2 - \frac{C_{\mathrm{ST}}(\hat{p};a)}{\hat{B}^q_{\mathrm{ST}}(\hat{p};a)}}. \tag{18}$$

Using now [21, supplementary material, Lemma 1], we deduce

$$\eta \leq \frac{-C_{\mathrm{ST}}(\hat{p};a)}{\hat{B}^q_{\mathrm{ST}}(\hat{p};a)}, \quad \frac{\hat{A}_{\mathrm{ST}}(\hat{p};a) + \hat{B}^l_{\mathrm{ST}}(\hat{p};a)}{2\hat{B}^q_{\mathrm{ST}}(\hat{p};a)} \leq \nu,$$

where

$$\eta = \min\{\frac{\gamma_1}{\gamma_7}, \frac{\gamma_2}{\gamma_8}\}, \quad \nu = \max\{\frac{\gamma_3+\gamma_5}{2\gamma_7}, \frac{\gamma_4+\gamma_6}{2\gamma_8}\}.$$

With these elements in place and referring to (18), we have

$$\mathrm{step}^{\mathrm{d}}_{\mathrm{ST}}(\hat{p};a) \geq \frac{-(\hat{A}_{\mathrm{ST}}(\hat{p};a) + \hat{B}^l_{\mathrm{ST}}(\hat{p};a))}{2\hat{B}^q_{\mathrm{ST}}(\hat{p};a)} + \sqrt{\left(\frac{\hat{A}_{\mathrm{ST}}(\hat{p};a) + \hat{B}^l_{\mathrm{ST}}(\hat{p};a)}{2\hat{B}^q_{\mathrm{ST}}(\hat{p};a)}\right)^2 + \eta}.$$

We observe now that $z \mapsto g(z) = -z + \sqrt{z^2+\eta}$ is monotonically decreasing and lower bounded. So, if z is upper bounded, then $g(z)$ is lower bounded by a positive constant. Taking $z = \frac{(\hat{A}_{\mathrm{ST}}(\hat{p};a)+\hat{B}^l_{\mathrm{ST}}(\hat{p};a))}{2\hat{B}^q_{\mathrm{ST}}(\hat{p};a)} \leq \nu$ gives the bound of the stepsize. Finally, the algorithm design together with Proposition IV.4 ensure fact (ii) throughout its evolution. □

It is worth noticing that the derivative-based implementation of the Displaced-Gradient Algorithm generalizes the algorithm proposed in our previous work [21] (in fact, the strategy proposed there corresponds to the choice $a = 0$). The next result characterizes the convergence properties of the performance-based implementation of Algorithm 1.

Theorem IV.7. (Convergence of performance-based implementation of Displaced-Gradient Algorithm). *For $0 \leq a \leq a_2^*$, $\diamond = \mathrm{p}$, and $\# \in \{\mathrm{ET}, \mathrm{ST}\}$, the variable-stepsize strategy in Algorithm 1 has the following properties*

(i) *the stepsize is uniformly lower bounded by the positive constant* $\mathrm{MIET}(a)$*;*

(ii) $V(p_k + tX^a_{\mathrm{hb}}(p_k)) \leq e^{-\frac{\sqrt{\mu}}{4}t}V(p_k)$ *for all* $t \in [0, \Delta_k]$ *and* $k \in \{0\} \cup \mathbb{N}$.

As a consequence, $f(x_{k+1}) - f(x_*) = \mathcal{O}(e^{-\frac{\sqrt{\mu}}{4}\sum_{i=0}^{k}\Delta_i})$.

Proof. To show (i), notice that it is sufficient to prove that $\mathrm{step}^{\mathrm{p}}_{\mathrm{ST}}$ is uniformly lower bounded away from zero. This is because of the definition of stepsize in (15b) and the fact that $b^{\mathrm{p}}_{\mathrm{ET}}(\hat{p},t;a) \leq b^{\mathrm{p}}_{\mathrm{ST}}(\hat{p},t;a)$ for all $\hat{p}$ and all t. For an arbitrary fixed $\hat{p}$, note that $t \mapsto b^{\mathrm{d}}_{\mathrm{ST}}(\hat{p},t;a)$ is strictly negative in the interval $[0, \mathrm{step}^{\mathrm{d}}_{\mathrm{ST}}(p;a))$ given the definition of stepsize in (15a). Consequently, the function $t \mapsto b^{\mathrm{p}}_{\mathrm{ST}}(\hat{p},t;a) = \int_0^t e^{\frac{\sqrt{\mu}}{4}\zeta}b^{\mathrm{d}}_{\mathrm{ST}}(\hat{p};\zeta,a)d\zeta$ is strictly negative over $(0, \mathrm{step}^{\mathrm{d}}_{\mathrm{ST}}(\hat{p};a))$. From the definition of $\mathrm{step}^{\mathrm{p}}_{\mathrm{ST}}$, it then follows that $\mathrm{step}^{\mathrm{p}}_{\mathrm{ST}}(\hat{p};a) \geq \mathrm{step}^{\mathrm{d}}_{\mathrm{ST}}(\hat{p};a)$. The result now follows by noting that $\mathrm{step}^{\mathrm{d}}_{\mathrm{ST}}$ is uniformly lower bounded away from zero by a positive constant, cf. Theorem IV.6(i).

To show (ii), we recall that $\Delta_k = \mathrm{step}^{\mathrm{p}}_{\#}(p_k;a)$ for $\# \in \{\mathrm{ET}, \mathrm{ST}\}$ and use Proposition IV.5 for $\hat{p} = p_k$ to obtain, for all $t \in [0, \Delta_k]$,

$$V(p(t)) - e^{-\frac{\sqrt{\mu}}{4}t}V(p_k) \leq e^{-\frac{\sqrt{\mu}}{4}t}b^{\mathrm{p}}_{\#}(p_k,t;a) \leq e^{-\frac{\sqrt{\mu}}{4}t}b^{\mathrm{p}}_{\#}(p_k,\Delta_k;a) = 0,$$

as claimed. □

The proof of Theorem IV.7 brings up an interesting geometric interpretation of the relationship between the stepsizes determined according to the derivative- and performance-based approaches. In fact, since

$$\frac{d}{dt}b^{\mathrm{p}}_{\#}(\hat{p},t;a) = e^{\frac{\sqrt{\mu}}{4}t}b^{\mathrm{d}}_{\#}(\hat{p},t;a),$$

we observe that $\mathrm{step}^{\mathrm{d}}_{\#}(\hat{p};a)$ is precisely the (positive) critical point of $t \mapsto b^{\mathrm{p}}_{\#}(\hat{p},t;a)$. Therefore, $\mathrm{step}^{\mathrm{p}}_{\mathrm{ST}}(\hat{p};a)$ is the smallest nonzero root of $t \mapsto b^{\mathrm{p}}_{\#}(\hat{p},t;a)$, whereas $\mathrm{step}^{\mathrm{d}}_{\mathrm{ST}}(\hat{p};a)$ is the time where $t \mapsto b^{\mathrm{p}}_{\#}(\hat{p},t;a)$ achieves its smallest value, and consequently is furthest away from zero. This confirms the fact that the performance-based approach obtains larger stepsizes than the derivative-based approach.

V. Exploiting Sampled Information to Enhance Algorithm Performance

Here we describe two different refinements of the implementations proposed in Section IV to further enhance their

performance. Both of them are based on further exploiting the sampled information about the system. The first refinement, cf. Section V-A, looks at the possibility of adapting the value of the gradient displacement as the algorithm is executed. The second refinement, cf. Section V-B, develops a high-order hold that more accurately approximates the evolution of the continuous-time heavy-ball dynamics with displaced gradient.

A. Adaptive Gradient Displacement

The derivative- and performance-based triggered implementations in Section IV-B both employ a constant value of the parameter a. Here, motivated by the observation made in Remark IV.3, we develop triggered implementations that adaptively adjust the value of the gradient displacement depending on the region of the space to which the state belongs. Rather than relying on the condition (14), which would require partitioning the state space based on bounds on $\nabla f(x)$ and v, we seek to compute on the fly a value of the parameter a that ensures the exponential decrease of the Lyapunov function at the current state. Formally, the strategy is stated in Algorithm 2.

Algorithm 2: Adaptive Displaced-Gradient Algorithm

Design Choices: $\diamond \in \{\mathrm{d}, \mathrm{p}\}$, $\# \in \{\mathrm{ET}, \mathrm{ST}\}$
Initialization: Initial point (p_0), objective function (f), tolerance (ϵ), increase rate ($r_i > 1$), decrease rate ($0 < r_d < 1$), stepsize lower bound (τ), $a \geq 0$, $k = 0$
while $\|\nabla f(x_k)\| \geq \epsilon$ **do**
  increase = True
  exit = False
  while exit = False **do**
    while $C_{\#}(p_k; a) \geq 0$ **do**
      $a = ar_d$
      increase = False
    end
    if $\mathrm{step}^{\diamond}_{\#}(p_k; a) \geq \tau$ **then**
      exit = True
    else
      $a = ar_d$
      increase = False
  end
  Compute stepsize $\Delta_k = \mathrm{step}^{\diamond}_{\#}(p_k; a)$
  Compute next iterate $p_{k+1} = p_k + \Delta_k X^a_{\mathrm{hb}}(p_k)$
  Set $k = k + 1$
  if increase = True **then**
    $a = ar_i$
end

Proposition V.1. (Convergence of Adaptive Displaced-Gradient Algorithm). *For $\diamond \in \{\mathrm{d}, \mathrm{p}\}$, $\# \in \{\mathrm{ET}, \mathrm{ST}\}$, and $\tau \leq \min_{a \in [0, a_2^*]} \mathrm{MIET}(a)$, the variable-stepsize strategy in Algorithm 2 has the following properties:*

(i) *it is executable (i.e., at each iteration, the parameter a is determined in a finite number of steps);*
(ii) *the stepsize is uniformly lower bounded by τ;*
(iii) *it satisfies $f(x_{k+1}) - f(x_*) = \mathcal{O}(e^{-\frac{\sqrt{\mu}}{4}\sum_{i=0}^{k}\Delta_i})$, for $k \in \{0\} \cup \mathbb{N}$.*

Proof. Notice first that the function $a \mapsto \mathrm{MIET}(a) > 0$ defined in (17) is continuous and therefore attains its minimum over a compact set. At each iteration, Algorithm 2 first ensures that $C_{\#}(\hat{p}; a) < 0$, decreasing a if this is not the case. We know this process is guaranteed as soon as $a < a_2^*$ (cf. proof of Theorem IV.6) and hence only takes a finite number of steps. Once $C_{\#}(\hat{p}; a) < 0$, the stepsize could be computed to guarantee the desired decrease of the Lyapunov function V. The algorithm next checks if the stepsize is lower bounded by τ. If that is not the case, then the algorithm reduces a and re-checks if $C_{\#}(\hat{p}; a) < 0$. With this process and in a finite number of steps, the algorithm eventually either computes a stepsize lower bounded by τ with $a > a_2^*$ or a decreases enough to make $a \leq a_2^*$, for which we know that the stepsize is already lower bounded by τ. These arguments establish facts (i) and (ii) at the same time. Finally, fact (iii) is a consequence of the prescribed decreased of the Lyapunov function along the algorithm execution. □

B. Discretization via High-Order Hold

The modified zero-order hold based on employing displaced gradients developed in Section IV is an example of the possibilities enabled by more elaborate uses of sampled information. In this section, we propose another such use based on the observation that the continuous-time heavy-ball dynamics can be decomposed as the sum of a linear term and a nonlinear term. Specifically, we have

$$X^a_{\mathrm{hb}}(p) = \begin{bmatrix} v \\ -2\sqrt{\mu}v \end{bmatrix} + \begin{bmatrix} 0 \\ -\sqrt{\mu_s}\nabla f(x + av) \end{bmatrix}.$$

Note that the first term in this decomposition is linear, whereas the other one contains the potentially nonlinear gradient term that complicates finding a closed-form solution. Keeping this in mind when considering a discrete-time implementation, it would seem reasonable to perform a zero-order hold only on the nonlinear term while exactly integrating the resulting differential equation. Formally, a zero-order hold at $\hat{p} = [\hat{x}, \hat{v}]$ of the nonlinear term above yields a system of the form

$$\begin{bmatrix} \dot{x} \\ \dot{v} \end{bmatrix} = A \begin{bmatrix} x \\ v \end{bmatrix} + b, \tag{19}$$

with $p(0) = \hat{p}$, and where

$$A = \begin{bmatrix} 0 & 1 \\ 0 & -2\sqrt{\mu} \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ -\sqrt{\mu_s}\nabla f(\hat{x} + a\hat{v}) \end{bmatrix}.$$

Equation (19) is an in-homogeneous linear dynamical system, which is integrable by the method of variation of constants [30]. Its solution is given by $p(t) = e^{At}\big(\int_0^t e^{-A\zeta} b d\zeta + p(0)\big)$, or equivalently,

$$x(t) = \hat{x} - \frac{\sqrt{\mu_s}\nabla f(\hat{x} + a\hat{v})t}{2\sqrt{\mu}} + (1 - e^{-2\sqrt{\mu}t})\frac{\sqrt{\mu_s}\nabla f(\hat{x} + a\hat{v}) + 2\sqrt{\mu}\hat{v}}{4\mu}, \tag{20a}$$

$$v(t) = e^{-2\sqrt{\mu}t}\hat{v} + (e^{-2\sqrt{\mu}t} - 1)\frac{\sqrt{\mu_s}\nabla f(\hat{x} + a\hat{v})}{2\sqrt{\mu}}. \tag{20b}$$

We refer to this trajectory as a *high-order-hold integrator*. In order to develop a discrete-time algorithm based on this type of integrator, the next result provides a bound of the evolution of the Lyapunov function V along the high-order-hold integrator trajectories. The proof is presented in Appendix A.

Proposition V.2. (Upper bound for derivative-based triggering with high-order hold). *Let $a \geq 0$ and define*

$$\begin{aligned}\mathfrak{b}^{\mathrm{d}}_{\mathrm{ET}}(\hat{p},t;a) &= \mathfrak{A}_{\mathrm{ET}}(\hat{p},t;a) + \mathfrak{B}_{\mathrm{ET}}(\hat{p},t;a)\\ &\quad + \mathfrak{C}_{\mathrm{ET}}(\hat{p};a) + \mathfrak{D}_{\mathrm{ET}}(\hat{p},t;a),\\ \mathfrak{b}^{\mathrm{d}}_{\mathrm{ST}}(\hat{p},t;a) &= (\mathfrak{A}^{q}_{\mathrm{ST}}(\hat{p};a) + \mathfrak{B}^{q}_{\mathrm{ST}}(\hat{p};a))t^2 + (\mathfrak{A}^{l}_{\mathrm{ST}}(\hat{p};a)\\ &\quad + \mathfrak{B}^{l}_{\mathrm{ST}}(\hat{p};a) + \mathfrak{D}_{\mathrm{ST}}(\hat{p};a))t + \mathfrak{C}_{\mathrm{ST}}(\hat{p};a),\end{aligned}$$

where

$$\begin{aligned}\mathfrak{A}_{\mathrm{ET}}(\hat{p},t;a) &= \sqrt{\mu_s}(\langle \nabla f(x(t)) - \nabla f(\hat{x}), v(t)\rangle\\ &\quad - \langle v(t)-\hat{v}, \nabla f(\hat{x}+a\hat{v})\rangle\\ &\quad - \sqrt{\mu}\langle x(t)-\hat{x}, \nabla f(\hat{x}+a\hat{v})\rangle)\\ &\quad - \sqrt{\mu}\langle v(t)-\hat{v}, v(t)\rangle,\\ \mathfrak{B}_{\mathrm{ET}}(\hat{p},t;a) &= \frac{\sqrt{\mu}}{4}\big(\sqrt{\mu_s}(f(x(t)) - f(\hat{x}))\\ &\quad - \sqrt{\mu}\sqrt{\mu_s}t\frac{\|\nabla f(\hat{x}+a\hat{v})\|^2}{L}\\ &\quad + \sqrt{\mu}\sqrt{\mu_s}t\langle \nabla f(\hat{x}+a\hat{v}), a\hat{v}\rangle + \frac{1}{4}(\|v(t)\|^2 - \|\hat{v}\|^2)\\ &\quad + \frac{1}{4}\|v(t)-\hat{v}+2\sqrt{\mu}(x(t)-\hat{x})\|^2\\ &\quad + \frac{1}{2}\langle v(t)-\hat{v}+2\sqrt{\mu}(x(t)-\hat{x}), \hat{v}\rangle\big),\\ \mathfrak{C}_{\mathrm{ET}}(\hat{p};a) &= C_{\mathrm{ET}}(\hat{p};a),\\ \mathfrak{D}_{\mathrm{ET}}(\hat{p},t;a) &= \sqrt{\mu_s}\langle \nabla f(\hat{x}), v(t)-\hat{v}\rangle\\ &\quad - \sqrt{\mu}\langle \hat{v}, v(t)-\hat{v}\rangle,\end{aligned}$$

and

$$\begin{aligned}\mathfrak{A}^{l}_{\mathrm{ST}}(\hat{p};a) &= \|2\sqrt{\mu}\hat{v} + \sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\|\Big(\sqrt{\mu}\|\hat{v}\|\\ &\quad + \frac{L\sqrt{\mu_s}}{2\sqrt{\mu}}\|\hat{v}\| + \frac{3\sqrt{\mu_s}}{2}\|\nabla f(\hat{x}+a\hat{v})\|\Big)\\ &\quad + \frac{\mu_s}{2}\|\nabla f(\hat{x}+a\hat{v})\|\Big(\frac{L}{\sqrt{\mu}}\|\hat{v}\| + \|\nabla f(\hat{x}+a\hat{v})\|\Big),\\ \mathfrak{A}^{q}_{\mathrm{ST}}(\hat{p};a) &= \|2\sqrt{\mu}\hat{v} + \sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\|\\ &\quad \cdot \Big(\big(\frac{L\sqrt{\mu_s}}{2\sqrt{\mu}} + \sqrt{\mu}\big)\|2\sqrt{\mu}\hat{v} + \sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\|\\ &\quad + \frac{L\mu_s}{2\sqrt{\mu}}\|\nabla f(\hat{x}+a\hat{v})\|\Big),\\ \mathfrak{B}^{l}_{\mathrm{ST}}(\hat{p};a) &= \frac{\sqrt{\mu}\sqrt{\mu_s}}{4}\Big(\frac{\sqrt{\mu_s}}{2\sqrt{\mu}}\|\nabla f(\hat{x}+a\hat{v})\|\,\|\nabla f(\hat{x})\|\\ &\quad + \frac{1}{2}\|2\sqrt{\mu}\hat{v}+\sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\|\big(\frac{\|\nabla f(\hat{x})\|}{\sqrt{\mu}} + \frac{\|\hat{v}\|}{\sqrt{\mu_s}}\big)\\ &\quad - \sqrt{\mu}\frac{\|\nabla f(\hat{x}+a\hat{v})\|^2}{L} + (a\sqrt{\mu} - \frac{1}{2})\langle \nabla f(\hat{x}+a\hat{v}), \hat{v}\rangle\Big),\\ \mathfrak{B}^{q}_{\mathrm{ST}}(\hat{p};a) &= \frac{10\mu^2 + L^2\sqrt{\mu_s}}{32\mu^{3/2}}\\ &\quad \cdot \|2\sqrt{\mu}\hat{v} + \sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\|^2\\ &\quad + \frac{\mu_s\left(4\mu^2 + L^2\sqrt{\mu_s}\right)}{32\mu^{3/2}}\|\nabla f(\hat{x}+a\hat{v})\|^2\\ &\quad + \frac{\sqrt{\mu_s}\left(4\mu^2 + L^2\sqrt{\mu_s}\right)}{16\mu^{3/2}}\|2\sqrt{\mu}\hat{v} + \sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\|\\ &\quad \cdot \|\nabla f(\hat{x}+a\hat{v})\|),\\ \mathfrak{C}_{\mathrm{ST}}(\hat{p};a) &= C_{\mathrm{ST}}(\hat{p};a),\\ \mathfrak{D}_{\mathrm{ST}}(\hat{p};a) &= \|2\sqrt{\mu}\hat{v} + \sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\| \cdot\\ &\quad \Big(\sqrt{\mu_s}\|\nabla f(\hat{x})\| + \sqrt{\mu}\|\hat{v}\|\Big).\end{aligned}$$

Let $t \mapsto p(t)$ be the high-order-hold integrator trajectory (20) *from $p(0) = \hat{p}$. Then, for $t \geq 0$,*

$$\frac{d}{dt}V(p(t)) + \frac{\sqrt{\mu}}{4}V(p(t)) \leq \mathfrak{b}^{\mathrm{d}}_{\mathrm{ET}}(\hat{p},t;a) \leq \mathfrak{b}^{\mathrm{d}}_{\mathrm{ST}}(\hat{p},t;a).$$

Analogously to what we did in Section IV-B, we build on this result to establish an upper bound for the performance-based triggering condition with the high-order-hold integrator.

Proposition V.3. (Upper bound for performance-based triggering with high-order hold). *Let $0 \leq a$ and*

$$\mathfrak{b}^{\mathrm{p}}_{\#}(\hat{p},t;a) = \int_0^t e^{\frac{\sqrt{\mu}}{4}\zeta}\mathfrak{b}^{\mathrm{d}}_{\#}(\hat{p},\zeta;a)d\zeta, \tag{21}$$

for $\# \in \{\mathrm{ET},\mathrm{ST}\}$. Let $t \mapsto p(t)$ be the high-order-hold integrator trajectory (20) *from $p(0) = \hat{p}$. Then, for $t \geq 0$,*

$$V(p(t)) - e^{-\frac{\sqrt{\mu}}{4}t}V(\hat{p}) \leq e^{-\frac{\sqrt{\mu}}{4}t}\mathfrak{b}^{\mathrm{p}}_{\mathrm{ET}}(\hat{p},t;a) \leq e^{-\frac{\sqrt{\mu}}{4}t}\mathfrak{b}^{\mathrm{p}}_{\mathrm{ST}}(\hat{p},t;a).$$

Using Proposition V.2, the proof of this result is analogous to that of Proposition IV.5, and we omit it for space reasons. Propositions V.2 and V.3 are all we need to fully specify the variable-stepsize algorithm based on high-order-hold integrators. Formally, we set

$$\mathfrak{step}^{\diamond}_{\#}(\hat{p};a) = \min\{t > 0 \mid \mathfrak{b}^{\diamond}_{\#}(\hat{p},t;a) = 0\}, \tag{22}$$

for $\diamond \in \{\mathrm{d},\mathrm{p}\}$ and $\# \in \{\mathrm{ET},\mathrm{ST}\}$. With this in place, we design Algorithm 3, which is a higher-order counterpart to Algorithm 2, and whose convergence properties are characterized in the following result.

Proposition V.4. (Convergence of Adaptive High-Order-Hold Algorithm). *For $\diamond \in \{\mathrm{d},\mathrm{p}\}$, and $\# \in \{\mathrm{ET},\mathrm{ST}\}$, there exists $\mathrm{MIET}^{\diamond}$ such that for $\tau \leq \mathrm{MIET}^{\diamond}$, the variable-stepsize strategy in Algorithm 3 has the following properties:*

(i) *it is executable (i.e., at each iteration, the parameter a is determined in a finite number of steps);*
(ii) *the stepsize is uniformly lower bounded by τ;*
(iii) *it satisfies $f(x_{k+1}) - f(x_*) = \mathcal{O}(e^{-\frac{\sqrt{\mu}}{4}\sum_{i=0}^{k}\Delta_i})$, for $k \in \{0\} \cup \mathbb{N}$.*

We omit the proof of this result, which is analogous to that of Proposition V.1, with lengthier computations.

VI. Simulations

Here we illustrate the performance of the algorithms resulting from the proposed resource-aware discretization approach to accelerated optimization flows. Specifically, we simulate in two examples the performance-based implementation of the Displaced Gradient algorithm (denoted DG^{p}) and the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2022.3171307, IEEE Transactions on Automatic Control

Algorithm 3: Adaptive High-Order-Hold Algorithm

Design Choices: $\diamond \in \{\mathrm{d}, \mathrm{p}\}$, $\# \in \{\mathrm{ET}, \mathrm{ST}\}$
Initialization: Initial point (p_0), objective function (f), tolerance (ϵ), increase rate ($r_i > 1$), decrease rate ($0 < r_d < 1$), stepsize lower bound (τ), $a \geq 0$, $k = 0$
while $\|\nabla f(x_k)\| \geq \epsilon$ **do**
  increase = True
  exit = False
  while exit = False **do**
    while $\mathfrak{C}_{\#}(p_k; a) \geq 0$ **do**
      $a = ar_d$
      increase = False
    end
    if $\mathfrak{step}^{\diamond}_{\#}(p_k; a) \geq \tau$ **then**
      exit = True
    else
      $a = ar_d$
      increase = False
  end
  Compute stepsize $\Delta_k = \mathfrak{step}^{\diamond}_{\#}(p_k; a)$
  Compute next iterate p_{k+1} using (20)
  Set $k = k + 1$
  if increase = True **then**
    $a = ar_i$
end

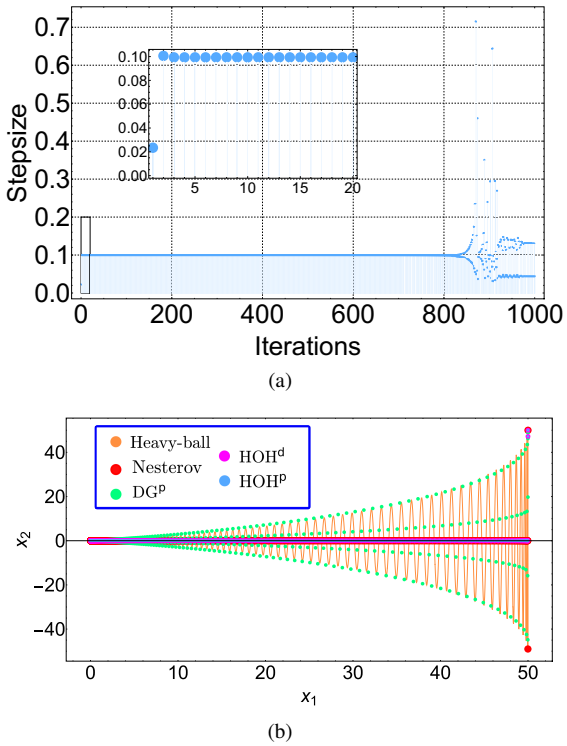


Fig. 1. Ill-conditioned quadratic objective function example. (a) Evolution of the stepsize along the execution of HOH$^{\mathrm{p}}$ during the first 1000 iterations. (b) State evolution along DG$^{\mathrm{p}}$, HOH$^{\mathrm{d}}$, HOH$^{\mathrm{p}}$, continuous heavy-ball dynamics, and Nesterov's method starting from $x = (50, 50)$ and $v = (-0.0023, -4.7139)$.

derivative- and performance-based implementations of the High-Order-Hold (HOH$^{\mathrm{d}}$ and HOH$^{\mathrm{p}}$ respectively) algorithms. We compare these algorithms against the Nesterov's accelerated gradient and the heavy-ball methods, as they exhibit similar or superior performance to the discretization approaches proposed in the literature, cf. Section I. Note that the latter are constant-stepsize methods, whereas the algorithms developed here are variable-stepsize ones, where the stepsize is computed online using state information. In fact, given our design procedure, DG$^{\mathrm{p}}$, HOH$^{\mathrm{d}}$, and HOH$^{\mathrm{p}}$ can be understood as different variable-stepsize implementations of the heavy-ball method. As the plots below show, the discretizations developed here retain the convergence rate of their continuous counterpart, with a performance regarding the objective function that is comparable or slightly better than state-of-the-art optimization algorithms.

Optimization of Ill-Conditioned Quadratic Objective Function

Consider the optimization of the objective function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x) = 10^{-2}x_1^2 + 10^2 x_2^2$. Note that $\mu = 2 \cdot 10^{-2}$ and $L = 2 \cdot 10^2$. We use $s = \mu/(36L^2)$ and initialize the velocity according to (4b). For DG$^{\mathrm{p}}$, HOH$^{\mathrm{d}}$, and HOH$^{\mathrm{p}}$, we set $a = 0.1$ and implement the event-triggered approach (at each iteration, we employ a numerical zero-finding routine to explicitly determine the stepsizes $\mathrm{step}^{\mathrm{p}}_{\mathrm{ET}}$, $\mathfrak{step}^{\mathrm{d}}_{\mathrm{ET}}$, and $\mathfrak{step}^{\mathrm{p}}_{\mathrm{ET}}$, respectively).

Figure 1(a) illustrates how the stepsize of HOH$^{\mathrm{p}}$ changes during the first 1000 iterations. After the tuning of the stepsize during the first iterations, it becomes quite steady (likely due to the simplicity of quadratic functions) until the trajectory approaches the minimizer. After 5 iterations, the algorithm stepsize becomes almost equal to the optimal stepsize.

Figure 1(b) compares the performance of DG$^{\mathrm{p}}$, HOH$^{\mathrm{d}}$, and HOH$^{\mathrm{p}}$ against the continuous heavy-ball method and the discrete Nesterov method for strongly convex functions. To implement the continuous heavy-ball method, we directly obtained the solution of the ODE and plotted its trajectories. The DG$^{\mathrm{p}}$ algorithm takes large stepsizes following the evolution of the continuous heavy-ball along the straight lines $p(t) = p_k + tX^a_{\mathrm{hb}}(p_k)$. Meanwhile, the higher-order nature of the hold employed by HOH$^{\mathrm{d}}$ and HOH$^{\mathrm{p}}$ makes them able to leap over the oscillations, yielding a state evolution similar to Nesterov's method. Figure 2 shows a comparison of the evolution of the objective and Lyapunov functions (where the heavy-ball method implemented is the discrete version [1]). We use the stepsizes $\frac{1}{L}$ for the Nesterov method and $\frac{4}{(\sqrt{L}+\sqrt{\mu})^2}$ for the heavy-ball method, as commonly found in the literature [12]. We observe that after some initial iterations, HOH$^{\mathrm{p}}$ outperforms Nesterov's method. Eventually, also DG$^{\mathrm{p}}$ catches up to Nesterov's method.

Logarithmic Regression

Consider the optimization of the regularized logistic regression cost function $f : \mathbb{R}^4 \to \mathbb{R}$ defined by $f(x) = \sum_{i=1}^{10} \log(1+e^{-y_i\langle z_i, x\rangle}) + \frac{1}{2}\|x\|^2$, where the points $\{z_i\}_{i=1}^{10} \subset \mathbb{R}^4$ are generated randomly using a uniform distribution in the interval $[-5, 5]$, and the points $\{y_i\}_{i=1}^{10} \subset \{-1, 1\}$ are generated similarly with quantized values. This objective function is 1-strongly convex and one can also compute the value $L = 177.49$. We use $a = 0.025$ and $s = \mu/(36L^2)$, and initialize

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2022.3171307, IEEE Transactions on Automatic Control

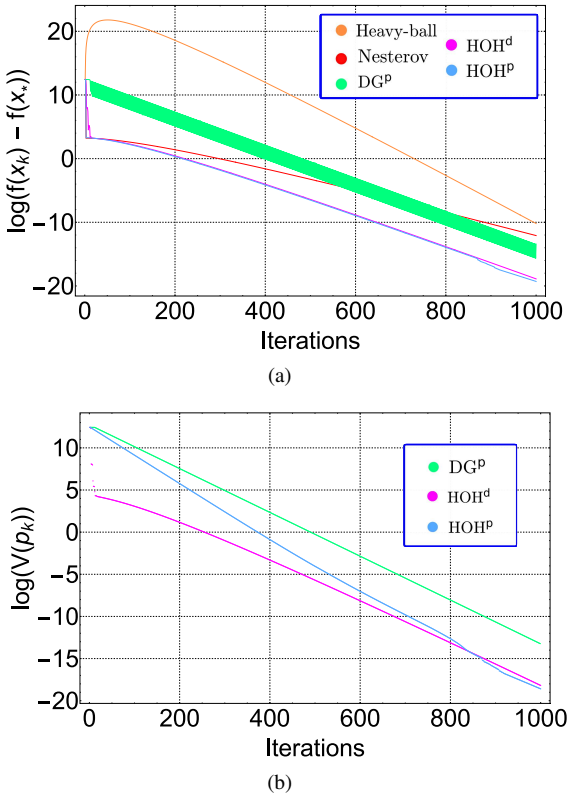


Fig. 2. Ill-conditioned quadratic objective function example. (a) Evolution of the logarithm of the objective function under DG^{p}, HOH^{d}, HOH^{p}, the heavy-ball method, and Nesterov's method starting from $x = (50, 50)$ and $v = (-0.0023, -4.7139)$. (b) Corresponding evolution of the logarithm of the Lyapunov function along DG^{p}, HOH^{d}, and HOH^{p}.

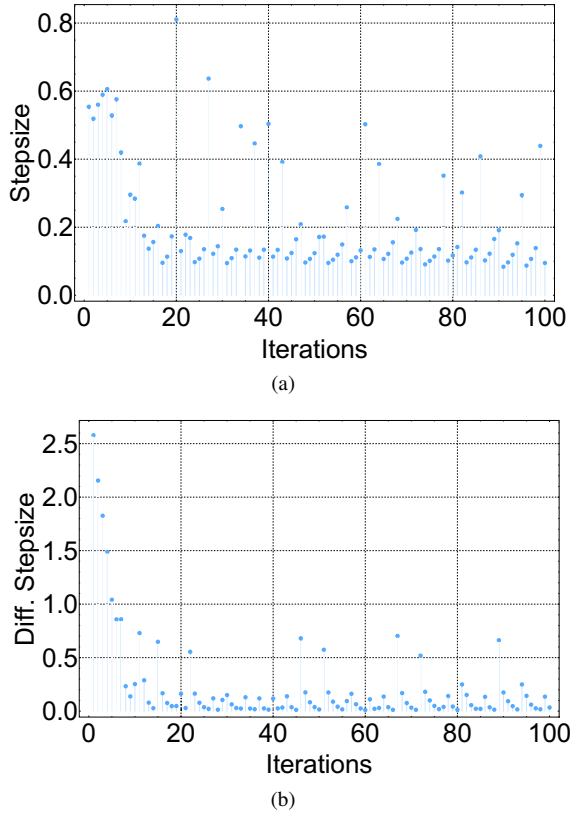


Fig. 3. Logarithmic regression example. (a) Evolution of the stepsize along the execution of HOH^{p} starting from $x = (50, 50, 50, 50)$ and $v = (-0.1026, -0.09265, -0.1078, -0.0899)$. Notice the complex pattern, with significant increases and oscillations along the trajectory. (b) Difference between the optimal stepsize (computed using the exact Lyapunov function, which assumes knowledge of the minimizer) and the stepsize of HOH^{p}. The largest difference is achieved at the beginning: after a few iterations, the difference decreases significantly, periodically becoming almost zero.

the velocity according to (4b). Figure 3(a) show the evolution of the stepsize along HOH^{p}, which changes as a function of the state looking to satisfy the desired decay of the Lyapunov function. Figure 3(b) shows the difference between the optimal stepsize, computed with complete knowledge of the Lyapunov function, and the stepsize computed using HOH^{p}. This plot is an illustration of the tightness of the upper bound for the expression $\dot{V} + \frac{\sqrt{\mu}}{4} V$ given by Proposition V.3. Figure 4 shows the evolution of the objective and Lyapunov functions. We observe how HOH^{d} and HOH^{p} outperform Nesterov's method, although eventually the heavy-ball algorithm performs the best. The Lyapunov function decreases at a much faster rate along HOH^{d} and HOH^{p} than along DG^{p}.

VII. Conclusions

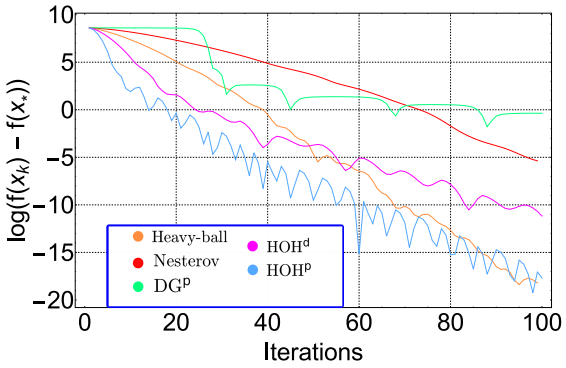
We have introduced a resource-aware control framework to the discretization of accelerated optimization flows that specifically takes advantage of their dynamical properties. We have exploited fundamental concepts from opportunistic state-triggering related to the various ways of encoding the notion of valid Lyapunov certificates, the use of sampled-data information, and the construction of state estimators and holders to synthesize variable-stepsize optimization algorithms that retain by design the convergence properties of the continuous-time heavy-ball dynamics with displaced gradient. The proposed methodology is general and applicable, with the appropriate derivations, to other accelerated optimization flows and in fact we expect this work will spur the development of other variable-stepsize implementations of accelerated optimization flows. We believe these results open the way to a number of exciting research directions. Among them, we highlight the characterization of how close the computed stepsize is from the stepsize that would be obtained using the original Lyapunov function, the design of adaptive learning schemes to refine the use of sampled data and optimize the algorithm performance with regards to the objective function, the use of tools and insights from hybrid systems for analysis and design, the incorporation of re-start schemes as triggering conditions to avoid overshooting and oscillations, the development of distributed implementations for network optimization problems, and the extension of the proposed design methodology to constrained optimization problems.

References

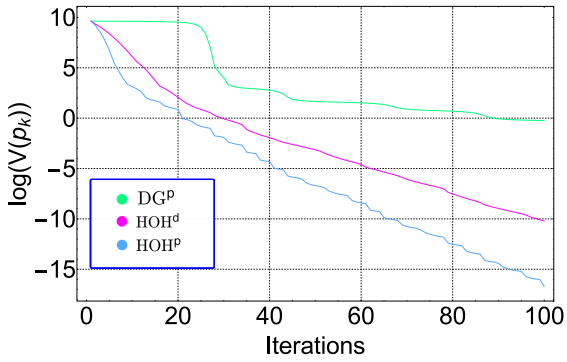
[1] B. T. Polyak, "Some methods of speeding up the convergence of iterative methods," *USSR Computational Mathematics and Mathematical Physics*, vol. 4, no. 5, pp. 1–17, 1964.

[2] Y. E. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," *Soviet Mathematics Doklady*, vol. 27, no. 2, pp. 372–376, 1983.

[3] Z. Allen-Zhu and L. Orecchia, "Linear Coupling: An Ultimate Unification of Gradient and Mirror Descent," in *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, Dagstuhl, Germany, 2017, pp. 1–22.



(a)



(b)

Fig. 4. Logarithmic regression example. (a) Evolution of the logarithm of the objective function under $\mathrm{DG^p}$, $\mathrm{HOH^d}$, $\mathrm{HOH^p}$, the heavy-ball method, and Nesterov's method starting from $x = (50, 50, 50, 50)$ and $v = (-0.1026, -0.09265, -0.1078, -0.0899)$. (b) Corresponding evolution of the logarithm of the Lyapunov function along $\mathrm{DG^p}$, $\mathrm{HOH^d}$, and $\mathrm{HOH^p}$.

[4] B. Hu and L. Lessard, "Dissipativity theory for Nesterov's accelerated method," in *International Conference on Machine Learning*, International Convention Centre, Sydney, Australia, August 2017, pp. 1549–1557.

[5] 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.

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

[7] S. Bubeck, Y. Lee, and M. Singh, "A geometric alternative to Nesterov's accelerated gradient descent," *arXiv preprint arXiv:1506.08187*, 2015.

[8] W. Su, S. Boyd, and E. J. Candès, "A differential equation for modeling Nesterov's accelerated gradient method: theory and insights," *Journal of Machine Learning Research*, vol. 17, pp. 1–43, 2016.

[9] M. Betancourt, M. Jordan, and A. C. Wilson, "On symplectic optimization," *arXiv preprint arXiv: 1802.03653*, 2018.

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

[11] H. Attouch, Z. Chbani, J. Fadili, and H. Riahi, "First-order optimization algorithms via inertial systems with Hessian driven damping," *arXiv preprint arXiv:1907.10536*, 2019.

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

[13] B. Sun, J. George, and S. Kia, "High-resolution modeling of the fastest first-order optimization method for strongly convex functions," *arXiv preprint arXiv:2008.11199*, 2020.

[14] R. W. Brockett, "Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems," *Linear Algebra and its Applications*, vol. 146, pp. 79–91, 1991.

[15] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*. Springer, 1994.

[16] B. Shi, S. S. Du, M. I. Jordan, and W. J. Su, "Acceleration via symplectic discretization of high-resolution differential equations," *arXiv preprint arXiv:1902.03694*, 2019.

[17] 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.

[18] A. Wilson, L. Mackey, and A. Wibisono, "Accelerating rescaled gradient descent: Fast optimization of smooth functions," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, Eds., vol. 31. Curran Associates, Inc., 2019, pp. 13 555–13 565.

[19] J. Zhang, A. Mokhtari, S. Sra, and A. Jadbabaie, "Direct Runge-Kutta discretization achieves acceleration," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31. Curran Associates, Inc., 2018, pp. 3900–3909.

[20] A. C. Wilson, B. Recht, and M. I. Jordan, "A Lyapunov analysis of momentum methods in optimization," *arXiv preprint arXiv:1611.02635*, 2018.

[21] M. Vaquero and J. Cortés, "Convergence-rate-matching discretization of accelerated optimization flows through opportunistic state-triggered control," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, vol. 32, pp. 9767–9776.

[22] A. S. Kolarijani, P. M. Esfahani, and T. Keviczky, "Fast gradient-based methods with exponential rate: a hybrid control framework," in *International Conference on Machine Learning*, July 2018, pp. 2728–2736.

[23] D. Hustig-Schultz and R. G. Sanfelice, "A robust hybrid Heavy-Ball algorithm for optimization with high performance," in *American Control Conference*, July 2019, pp. 151–156.

[24] W. P. M. H. Heemels, K. H. Johansson, and P. Tabuada, "An introduction to event-triggered and self-triggered control," in *IEEE Conf. on Decision and Control*, Maui, HI, 2012, pp. 3270–3285.

[25] C. Nowzari, E. Garcia, and J. Cortés, "Event-triggered control and communication of networked systems for multi-agent consensus," *Automatica*, vol. 105, pp. 1–27, 2019.

[26] P. Ong and J. Cortés, "Event-triggered control design with performance barrier," in *IEEE Conf. on Decision and Control*, Miami Beach, FL, Dec. 2018, pp. 951–956.

[27] M. Laborde and A. Oberman, "A Lyapunov analysis for accelerated gradient methods: From deterministic to stochastic case," in *AISTATS*, vol. 108, Online, 2020, pp. 602–612.

[28] M. Muehlebach and M. Jordan, "A dynamical systems perspective on Nesterov acceleration," in *International Conference on Machine Learning*, vol. 97, Long Beach, California, USA, 2019, pp. 4656–4662.

[29] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *International Conference on Machine Learning*, vol. 28, Atlanta, Georgia, USA, 2013, pp. 1139–1147.

[30] L. Perko, *Differential Equations and Dynamical Systems*, 3rd ed., ser. Texts in Applied Mathematics. New York: Springer, 2000, vol. 7.

[31] S. Lang, *Real and Functional Analysis*, 3rd ed. New York: Springer, 1993.

APPENDIX A

Throughout the appendix, we make use of a number of basic facts that we gather here for convenience,

$$f(x_*) - f(x) \leq -\frac{\|\nabla f(x)\|^2}{2L} \tag{A.1a}$$

$$\frac{\|\nabla f(x)\|}{L} \leq \|x - x_*\| \leq \frac{\|\nabla f(x)\|}{\mu} \tag{A.1b}$$

$$f(y) - f(x) - \langle \nabla f(x), y - x \rangle \leq \frac{L}{2}\|y - x\|^2 \tag{A.1c}$$

$$\frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2 \leq \langle \nabla f(x) - \nabla f(y), x - y \rangle \tag{A.1d}$$

$$f(y) - f(x) - \langle \nabla f(x), y - x \rangle \leq \frac{1}{2\mu}\|\nabla f(y) - \nabla f(x)\|^2 \tag{A.1e}$$

We also resort at various points to the expression of the gradient of V,

$$\nabla V(p) = \begin{bmatrix} \sqrt{\mu_s}\nabla f(x) + \sqrt{\mu}v + 2\mu(x - x_*) \\ v + \sqrt{\mu}(x - x_*) \end{bmatrix}. \tag{A.2}$$

The following result is used in the proof of Theorem IV.2.

Lemma A.1. *For $\beta_1, \dots, \beta_4 > 0$, the function*

$$g(z) = \frac{\beta_3 + \beta_4 z^2}{-\beta_1 + \beta_2 z} \tag{A.3}$$

is positively lower bounded on $(\beta_1/\beta_2, \infty)$.

Proof. The derivative of g is

$$g'(z) = \frac{-\beta_2\beta_3 + \beta_4 z(-2\beta_1 + \beta_2 z)}{(\beta_1 - \beta_2 z)^2}.$$

The solutions to $g'(z) = 0$ are then given by

$$z^{\pm}_{\rm root} = \frac{\beta_1\beta_4 \pm \sqrt{\beta_2^2\beta_3\beta_4 + \beta_1^2\beta_4^2}}{\beta_2\beta_4}. \tag{A.4}$$

Note that $z^-_{\rm root} < 0 < \beta_1/\beta_2 < z^+_{\rm root}$, g' is negative on $(z^-_{\rm root}, z^+_{\rm root})$, and positive on $(z^+_{\rm root}, \infty)$. Therefore the minimum value over $(\beta_1/\beta_2, \infty)$ is achieved at $z^+_{\rm root}$, and corresponds to $g(z^+_{\rm root}) > 0$. □

Proof of Proposition IV.4. We break out $\frac{d}{dt}V(p(t)) + \frac{\sqrt{\mu}}{4}V(p(t))$ as follows

$$\begin{aligned} &\frac{d}{dt}V(\hat{p} + tX^a_{\rm hb}(\hat{p})) + \frac{\sqrt{\mu}}{4}V(\hat{p} + tX^a_{\rm hb}(\hat{p})) = \\ &= \underbrace{\langle \nabla V(\hat{p}), X^a_{\rm hb}(\hat{p})\rangle + \frac{\sqrt{\mu}}{4}V(\hat{p})}_{\text{Term I + II + III}} \\ &\quad + \underbrace{\langle \nabla V(\hat{p} + tX^a_{\rm hb}(\hat{p})) - \nabla V(\hat{p}), X^a_{\rm hb}(\hat{p})\rangle}_{\text{Term IV + V}} \\ &\quad + \frac{\sqrt{\mu}}{4}\underbrace{(V(\hat{p} + tX^a_{\rm hb}(\hat{p})) - V(\hat{p}))}_{\text{Term VI}}, \end{aligned}$$

and bound each term separately.

Term I + II + III. From the definition (5) of V and the fact that $\|y_1 + y_2\|^2 \le 2\|y_1\|^2 + 2\|y_2\|^2$, we have

$$\begin{aligned} V(\hat{p}) &= \sqrt{\mu_s}(f(\hat{x}) - f(x_*)) + \frac{1}{4}\|\hat{v}\|^2 \\ &\quad + \frac{1}{4}\|\hat{v} + 2\sqrt{\mu}(\hat{x} - x_*)\|^2 \\ &\le \sqrt{\mu_s}(f(\hat{x}) - f(x_*)) \\ &\quad + \frac{1}{4}\|\hat{v}\|^2 + \frac{2}{4}\|\hat{v}\|^2 + \frac{2}{4}\|2\sqrt{\mu}(\hat{x} - x_*)\|^2 \\ &= \sqrt{\mu_s}(f(\hat{x}) - f(x_*)) + \frac{3}{4}\|\hat{v}\|^2 + 2\mu\|\hat{x} - x_*\|^2. \end{aligned}$$

Using this bound, we obtain

$$\begin{aligned} &\langle \nabla V(\hat{p}), X^a_{\rm hb}(\hat{p})\rangle + \frac{\sqrt{\mu}}{4}V(\hat{p}) \\ &\le -\sqrt{\mu}\|\hat{v}\|^2 + \frac{\sqrt{\mu}}{4}\sqrt{\mu_s}(f(\hat{x}) - f(x_*)) + \frac{3\sqrt{\mu}}{16}\|\hat{v}\|^2 \\ &\quad + \frac{\mu\sqrt{\mu}}{2}\|\hat{x} - x_*\|^2 + \sqrt{\mu_s}\langle \nabla f(\hat{x}) - \nabla f(\hat{x} + a\hat{v}), \hat{v}\rangle \\ &\quad - \sqrt{\mu}\sqrt{\mu_s}\langle \nabla f(\hat{x} + a\hat{v}), \hat{x} - x_*\rangle. \end{aligned}$$

Writing 0 as $0 = a\hat{v} - a\hat{v}$ and using strong convexity, we can upper bound $\langle \nabla f(\hat{x} + a\hat{v}), x_* - \hat{x}\rangle$ in the last summand by the expression

$$f(x_*) - f(\hat{x} + a\hat{v}) - \frac{\mu}{2}\|\hat{x} + a\hat{v} - x_*\|^2 + \langle \nabla f(\hat{x} + a\hat{v}), a\hat{v}\rangle.$$

Substituting this bound above and re-grouping terms,

$$\begin{aligned} &\langle \nabla V(\hat{p}), X^a_{\rm hb}(\hat{p})\rangle + \frac{\sqrt{\mu}}{4}V(\hat{p}) \le -\sqrt{\mu}\|\hat{v}\|^2 \\ &\quad + \underbrace{\sqrt{\mu}\sqrt{\mu_s}\Big(\frac{1}{4}(f(\hat{x}) - f(x_*)) + f(x_*) - f(\hat{x} + a\hat{v})\Big)}_{\text{(a)}} \\ &\quad + \frac{3\sqrt{\mu}}{16}\|\hat{v}\|^2 + \sqrt{\mu_s}\langle \nabla f(\hat{x}) - \nabla f(\hat{x} + a\hat{v}), \hat{v}\rangle \\ &\quad \underbrace{+\frac{\mu\sqrt{\mu}}{2}\|\hat{x} - x_*\|^2 + \sqrt{\mu}\sqrt{\mu_s}(-\frac{\mu}{2}\|\hat{x} + a\hat{v} - x_*\|^2)}_{\text{(b)}} \\ &\quad + \sqrt{\mu}\sqrt{\mu_s}\langle \nabla f(\hat{x} + a\hat{v}), a\hat{v}\rangle. \end{aligned}$$

Observe that

$$\begin{aligned} \text{(a)} &= \sqrt{\mu}\sqrt{\mu_s}\big(-\frac{3}{4}(f(\hat{x}) - f(x_*)) + f(\hat{x}) - f(\hat{x} + a\hat{v})\big), \\ \text{(b)} &\le -\frac{\mu^2\sqrt{s}}{2}\|\hat{x} - x_*\|^2 + \sqrt{\mu_s}\mu^{3/2}\|\hat{x} - x_*\|\,\|a\hat{v}\| \\ &\quad - \sqrt{\mu_s}\mu^{3/2}/2\,\|a\hat{v}\|^2, \end{aligned}$$

where, in the expression of (a), we have expressed 0 as $0 = 3/4(f(\hat{x}) - f(\hat{x}))$ and, in the expression of (b), we have expanded the square and used the Cauchy-Schwartz inequality [31]. Finally, resorting to (A.1), we obtain

$$\langle \nabla V(\hat{p}), X^a_{\rm hb}(\hat{p})\rangle + \frac{\sqrt{\mu}}{4}V(\hat{p}) \le C_{\rm ET}(\hat{p}; a) = C_{\rm ST}(\hat{p}; a).$$

- **Term IV + V**. Using (A.2) we have

$$\nabla V(\hat{p} + tX^a_{\rm hb}(\hat{p})) = \begin{bmatrix} \sqrt{\mu_s}\nabla f(\hat{x} + t\hat{v}) + \sqrt{\mu}\hat{v} - 2\mu t\hat{v} \\ -t\sqrt{\mu}\sqrt{\mu_s}\nabla f(\hat{x} + a\hat{v}) + 2\mu(\hat{x} + t\hat{v} - x_*) \\ \hat{v} - 2t\sqrt{\mu}\hat{v} - t\sqrt{\mu_s}\nabla f(\hat{x} + a\hat{v}) + \sqrt{\mu}(\hat{x} + t\hat{v} - x_*) \end{bmatrix}.$$

Therefore, $\nabla V(\hat{p} + tX^a_{\rm hb}(\hat{p})) - \nabla V(\hat{p})$ reads

$$\begin{bmatrix} \sqrt{\mu_s}(\nabla f(\hat{x} + t\hat{v}) - \nabla f(\hat{x})) - t\sqrt{\mu}\sqrt{\mu_s}\nabla f(\hat{x} + a\hat{v}) \\ -\sqrt{\mu}t\hat{v} - t\sqrt{\mu_s}\nabla f(\hat{x} + a\hat{v}) \end{bmatrix}$$

and hence

$$\begin{aligned} &\langle \nabla V(\hat{p} + tX^a_{\rm hb}(\hat{p})) - \nabla V(\hat{p}), X^a_{\rm hb}(\hat{p})\rangle \\ &= \sqrt{\mu_s}\langle \nabla f(\hat{x} + t\hat{v}) - \nabla f(\hat{x}), \hat{v}\rangle \\ &\quad + 2t\sqrt{\mu}\sqrt{\mu_s}\langle \nabla f(\hat{x} + a\hat{v}), \hat{v}\rangle + 2t\mu\|\hat{v}\|^2 \\ &\quad + t\mu_s\|\nabla f(\hat{x} + a\hat{v})\|^2. \end{aligned}$$

The RHS of the last expression is precisely $A_{\rm ET}(\hat{p}, t; a)$. Using the L-Lipschitzness of ∇f, one can see that $A_{\rm ET}(\hat{p}, t; a) \le A_{\rm ST}(p; a)t$.

- **Term VI**. From (5),

$$V(\hat{p} + tX^a_{\rm hb}(\hat{p})) - V(\hat{p}) = \sqrt{\mu_s}(f(\hat{x} + t\hat{v}) - f(x_*))$$

$$
\begin{aligned}
&+\frac{1}{4}\left\|\hat{v}-2t\sqrt{\mu}\hat{v}-t\sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\right\|^2\\
&+\frac{1}{4}\left\|\hat{v}-2t\sqrt{\mu}\hat{v}-t\sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\right.\\
&\left.+2\sqrt{\mu}(\hat{x}+t\hat{v}-x_*)\right\|^2-\sqrt{\mu_s}(f(\hat{x})-f(x_*))\\
&-\frac{1}{4}\left\|\hat{v}\right\|^2-\frac{1}{4}\left\|\hat{v}+2\sqrt{\mu}(\hat{x}-x_*)\right\|^2.
\end{aligned}
$$

Expanding the squares in the second and third summands, and simplifying, we obtain

$$
\begin{aligned}
V(\hat{p}+tX^a_{\rm hb}(\hat{p}))-V(\hat{p})&=\sqrt{\mu_s}(f(\hat{x}+t\hat{v})-f(\hat{x}))\\
&+\frac{1}{4}\left\|-2t\sqrt{\mu}\hat{v}-t\sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\right\|^2\\
&+\frac{1}{2}\langle\hat{v},-2t\sqrt{\mu}\hat{v}-t\sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\rangle\\
&+\frac{1}{4}\left\|-t\sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\right\|^2\\
&+\frac{1}{2}\langle\hat{v}+2\sqrt{\mu}(\hat{x}-x_*),-t(\sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\rangle\\
&=\sqrt{\mu_s}(f(\hat{x}+t\hat{v})-f(\hat{x}))\\
&+\frac{1}{4}\left\|-2t\sqrt{\mu}\hat{v}-t\sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\right\|^2\\
&-t\sqrt{\mu}\left\|\hat{v}\right\|^2-t\sqrt{\mu_s}\langle\hat{v},\nabla f(\hat{x}+a\hat{v})\rangle\\
&+\frac{1}{4}\left\|-t\sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\right\|^2\\
&+\langle\sqrt{\mu}(\hat{x}-x_*),-t\sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\rangle.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\langle x_*-\hat{x},\nabla f(\hat{x}+a\hat{v})\rangle\\
&=\langle x_*-\hat{x}-av,\nabla f(\hat{x}+a\hat{v})\rangle+\langle a\hat{v},\nabla f(\hat{x}+a\hat{v})\rangle\\
&\leq-\frac{\left\|\nabla f(\hat{x}+a\hat{v})\right\|^2}{L}+\langle a\hat{v},\nabla f(\hat{x}+a\hat{v})\rangle,
\end{aligned}
$$

where in the inequality we have used (A.1d) with $x=\hat{x}+a\hat{v}$ and $y=x_*$. Using this in the equation above, one identifies the expression of $B_{\rm ET}(p,t;a)$. Finally, applying (A.1c), one can show that $B_{\rm ET}(p,t;a)\leq B^l_{\rm ST}(p;a)t+B^q_{\rm ST}(p;a)t^2$, concluding the proof. □

Proof of Proposition V.2. For convenience, let

$$
X^{a,\hat{p}}_{\rm hb}(p)=\begin{bmatrix}v\\-2\sqrt{\mu}v-\sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\end{bmatrix},
$$

where $\hat{p}=[\hat{x},\hat{v}]$. We next provide a bound for the expression

$$
\begin{aligned}
\frac{d}{dt}V(p(t)))+\frac{\sqrt{\mu}}{4}V(p(t))&=\underbrace{\langle\nabla V(\hat{p}),X^{a,\hat{p}}_{\rm hb}(\hat{p})\rangle+\frac{\sqrt{\mu}}{4}V(\hat{p})}_{\text{Term I + II + III}}\\
&+\underbrace{\langle\nabla V(p(t))-\nabla V(\hat{p}),X^{a,\hat{p}}_{\rm hb}(p(t))\rangle}_{\text{Term IV}}\\
&+\underbrace{\langle\nabla V(\hat{p}),X^{a,\hat{p}}_{\rm hb}(p(t))-X^{a,\hat{p}}_{\rm hb}(\hat{p})\rangle}_{\text{Term V}}+\frac{\sqrt{\mu}}{4}\underbrace{(V(p(t))-V(\hat{p}))}_{\text{Term VI}}.
\end{aligned}
$$

Next, we bound each term separately.

- **Term I + II + III**. Since $X^{a,\hat{p}}_{\rm hb}(\hat{p})=X^a_{\rm hb}(\hat{p})$, this term is exactly the same as Term I + II + III in the proof of Proposition IV.4, and hence the bound obtained there is valid.

- **Term IV**. Using (A.2), we have

$$
\begin{aligned}
&\langle\nabla V(p(t))-\nabla V(\hat{p}),X^{a,\hat{p}}_{\rm hb}(p(t))\rangle\\
&=\sqrt{\mu_s}\langle\nabla f(x(t))-\nabla f(\hat{x}),v(t)\rangle\\
&\quad+\sqrt{\mu}\langle v(t)-\hat{v},v(t)\rangle+2\mu\langle x(t)-\hat{x},v(t)\rangle\\
&\quad-2\sqrt{\mu}\langle v(t)-\hat{v},v(t)\rangle-\sqrt{\mu_s}\langle v(t)-\hat{v},\nabla f(\hat{x}+a\hat{v})\rangle\\
&\quad-2\mu\langle x(t)-\hat{x},v(t)\rangle-\sqrt{\mu_s}\sqrt{\mu}\langle x(t)-\hat{x},\nabla f(\hat{x}+a\hat{v})\rangle\\
&=\sqrt{\mu_s}\langle\nabla f(x(t))-\nabla f(\hat{x}),v(t)\rangle-\sqrt{\mu}\langle v(t)-\hat{v},v(t)\rangle\\
&\quad-\sqrt{\mu_s}\langle v(t)-\hat{v},\nabla f(\hat{x}+a\hat{v})\rangle\\
&\quad-\sqrt{\mu_s}\sqrt{\mu}\langle x(t)-\hat{x},\nabla f(\hat{x}+a\hat{v})\rangle,
\end{aligned}
$$

from where we obtain Term IV $\leq\mathfrak{A}_{\rm ET}(\hat{p},t;a)$. Now, using $0=\hat{v}-\hat{v}$, the L-Lipschitzness of ∇f, and the Cauchy-Schwartz inequality, we have

$$
\begin{aligned}
|\mathfrak{A}_{\rm ET}(\hat{p},t;a)|\leq\sqrt{\mu_s}L\left\|x(t)-\hat{x}\right\|&\left(\left\|v(t)-\hat{v}\right\|+\left\|\hat{v}\right\|\right)\\
&+\sqrt{\mu}\left\|v(t)-\hat{v}\right\|^2+\sqrt{\mu}\left\|v(t)-\hat{v}\right\|\left\|\hat{v}\right\|\\
&+\sqrt{\mu_s}\left\|v(t)-\hat{v}\right\|\left\|\nabla f(\hat{x}+a\hat{v})\right\|\\
&+\sqrt{\mu_s}\sqrt{\mu}\left\|x(t)-\hat{x}\right\|\left\|\nabla f(\hat{x}+a\hat{v})\right\|.
\end{aligned}
$$

Using (20), the triangle inequality, and $1-e^{-2\sqrt{\mu}t}\leq2\sqrt{\mu}t$, we can write

$$
\begin{aligned}
\left\|x(t)-\hat{x}\right\|&\leq\frac{t}{2\sqrt{\mu}}\left\|2\sqrt{\mu}\hat{v}+\sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\right\|\\
&\quad+\frac{\sqrt{\mu_s}t}{2\sqrt{\mu}}\left\|\nabla f(\hat{x}+a\hat{v})\right\|,\qquad\text{(A.5a)}\\
\left\|v(t)-\hat{v}\right\|&\leq t\left\|2\sqrt{\mu}\hat{v}+\sqrt{\mu_s}\nabla f(\hat{x}+a\hat{v})\right\|.\qquad\text{(A.5b)}
\end{aligned}
$$

Substituting into the bound for $|\mathfrak{A}_{\rm ET}(\hat{p},t;a)|$ above, we obtain

$$
|\mathfrak{A}_{\rm ET}(\hat{p},t;a)|\leq\mathfrak{A}^q_{ST}(\hat{p};a)t^2+\mathfrak{A}^l_{\rm ST}(\hat{p};a)t
$$

as claimed.

- **Term V**. Using (A.2), we have

$$
\begin{aligned}
&\left\langle\nabla V(\hat{p}),X^{a,\hat{p}}_{\rm hb}(p(t))-X^{a,\hat{p}}_{\rm hb}(\hat{p})\right\rangle\\
&=\langle\begin{bmatrix}\sqrt{\mu_s}\nabla f(\hat{x})+\sqrt{\mu}\hat{v}+2\mu(\hat{x}-x_*)\\\hat{v}+\sqrt{\mu}(\hat{x}-x_*)\end{bmatrix},\\
&\quad\begin{bmatrix}v(t)-\hat{v}\\-2\sqrt{\mu}(v(t)-\hat{v})\end{bmatrix}\rangle\\
&=\sqrt{\mu_s}\langle\nabla f(\hat{x}),v(t)-\hat{v}\rangle+\sqrt{\mu}\langle\hat{v},v(t)-\hat{v}\rangle\\
&\quad+2\mu\langle\hat{x}-x_*,v(t)-\hat{v}\rangle-2\sqrt{\mu}\langle\hat{v},v(t)-\hat{v}\rangle\\
&\quad-2\mu\langle\hat{x}-x_*,v(t)-\hat{v}\rangle=\mathfrak{D}_{\rm ET}(\hat{p},t;a).
\end{aligned}
$$

Taking the absolute value and using the Cauchy-Schwartz inequality in conjunction with (A.5), we obtain the expression corresponding to $\mathfrak{D}_{\rm ST}$.

- **Term VI**. From (5),

$$
\begin{aligned}
V(p(t))-V(\hat{p})&=\sqrt{\mu_s}(f(x(t))-f(x_*))+\frac{1}{4}\left\|v(t)\right\|^2\\
&+\frac{1}{4}\left\|v(t)+2\sqrt{\mu}(x(t)-x_*)\right\|^2\\
&-\sqrt{\mu_s}(f(\hat{x})-f(x_*))-\frac{1}{4}\left\|\hat{v}\right\|^2\\
&-\frac{1}{4}\left\|\hat{v}+2\sqrt{\mu}(\hat{x}-x_*)\right\|^2.
\end{aligned}
$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2022.3171307, IEEE Transactions on Automatic Control

Expanding the third summand (using $x(t) = \hat{x}+(x(t)-\hat{x})$ and $v(t) = \hat{v}+(v(t)-\hat{v})$) as $\left\|\hat{v} + 2\sqrt{\mu}(\hat{x} - x_*)\right\|^2+2\langle\hat{v}+2\sqrt{\mu}(\hat{x}-x_*), v(t)-\hat{v}+2\sqrt{\mu}(x(t)-\hat{x})\rangle+\left\|v(t) - \hat{v} + 2\sqrt{\mu}(x(t) - \hat{x})\right\|^2$, we obtain after simplification

$$\begin{aligned} V(p(t)) - V(\hat{p}) &= \sqrt{\mu_s}(f(x(t)) - f(\hat{x})) \qquad \text{(A.6)}\\ &+ \frac{1}{4}(\|v(t)\|^2 - \|\hat{v}\|^2) + \frac{1}{4}\|v(t) - \hat{v} + 2\sqrt{\mu}(x(t) - \hat{x})\|^2\\ &+ \frac{1}{2}\langle v(t) - \hat{v} + 2\sqrt{\mu}(x(t) - \hat{x}), \hat{v} + 2\sqrt{\mu}(\hat{x} - x_*)\rangle. \end{aligned}$$

Using (20), we have

$$\begin{aligned} &\langle v(t) - \hat{v} + 2\sqrt{\mu}(x(t) - \hat{x}), 2\sqrt{\mu}(\hat{x} - x_*)\rangle\\ &= -2\sqrt{\mu}\sqrt{\mu_s}t\langle\nabla f(\hat{x} + a\hat{v}), \hat{x} - x_*\rangle\\ &= -2\sqrt{\mu}\sqrt{\mu_s}t\langle\nabla f(\hat{x} + a\hat{v}), \hat{x} + a\hat{v} - x_*\rangle\\ &\quad - 2\sqrt{\mu}\sqrt{\mu_s}t\langle\nabla f(\hat{x} + a\hat{v}), -a\hat{v}\rangle\\ &\leq -2\sqrt{\mu}\sqrt{\mu_s}t\frac{\|\nabla f(\hat{x} + a\hat{v})\|^2}{L}\\ &\quad + 2\sqrt{\mu}\sqrt{\mu_s}t\langle\nabla f(\hat{x} + a\hat{v}), a\hat{v}\rangle, \end{aligned}$$

where we have used (A.1d) to derive the inequality. Substituting this bound into (A.6), we obtain $\frac{\sqrt{\mu}}{4}(V(p(t)) - V(\hat{p})) \leq \mathfrak{B}_{\mathrm{ET}}(\hat{p}, t; a)$. To obtain the ST-expressions, we bound each remaining term separately as follows. Note that

$$\begin{aligned} &f(x(t)) - f(\hat{x}) \underbrace{\leq}_{\text{(A.1e)}} \langle\nabla f(\hat{x}), x(t) - \hat{x}\rangle + \frac{L^2}{2\mu}\|x(t) - \hat{x}\|^2\\ &\leq \|x(t) - \hat{x}\|\,\|\nabla f(\hat{x})\| + \frac{L^2}{2\mu}\|x(t) - \hat{x}\|^2\\ &\leq \frac{t}{2\sqrt{\mu}}\|2\sqrt{\mu}\hat{v} + \sqrt{\mu_s}\nabla f(\hat{x} + a\hat{v})\|\,\|\nabla f(\hat{x})\|\\ &\quad + \frac{\sqrt{\mu_s}t}{2\sqrt{\mu}}\|\nabla f(\hat{x} + a\hat{v})\|\,\|\nabla f(\hat{x})\|\\ &\quad + \frac{L^2}{2\mu}(\frac{t^2}{4\mu}\|2\sqrt{\mu}\hat{v} + \sqrt{\mu_s}\nabla f(\hat{x} + a\hat{v})\|^2\\ &\quad + \frac{\mu_s t^2}{4\mu}\|\nabla f(\hat{x} + a\hat{v})\|^2\\ &\quad + \frac{\sqrt{\mu_s}t^2}{2\mu}\|2\sqrt{\mu}\hat{v} + \sqrt{\mu_s}\nabla f(\hat{x} + a\hat{v})\|\,\|\nabla f(\hat{x} + a\hat{v})\|), \end{aligned}$$

where we have used (A.5a) to obtain the last inequality. Next,

$$\begin{aligned} &\|v(t)\|^2 - \|\hat{v}\|^2 = \|v(t) - \hat{v}\|^2 + 2\langle v(t) - \hat{v}, \hat{v}\rangle\\ &\leq t^2\|2\sqrt{\mu}\hat{v} + \sqrt{\mu_s}\nabla f(\hat{x} + a\hat{v})\|^2\\ &\quad + 2t\|2\sqrt{\mu}\hat{v} + \sqrt{\mu_s}\nabla f(\hat{x} + a\hat{v})\|\,\|\hat{v}\|, \end{aligned}$$

where we have used (A.5b) to obtain the last inequality. Using $\|y_1 + y_2\|^2 \leq 2\|y_1\|^2 + 2\|y_2\|^2$, we bound

$$\begin{aligned} &\|v(t)-\hat{v}+2\sqrt{\mu}(x(t)-\hat{x})\|^2 \leq 2\|v(t)-\hat{v}\|^2+8\mu\|x(t)-\hat{x}\|^2\\ &\leq 2t^2\|2\sqrt{\mu}\hat{v} + \sqrt{\mu_s}\nabla f(\hat{x} + a\hat{v})\|^2 + 4\sqrt{\mu}t\cdot\\ &\quad\cdot\left(\|2\sqrt{\mu}\hat{v} + \sqrt{\mu_s}\nabla f(\hat{x} + a\hat{v})\| + \sqrt{\mu_s}\|\nabla f(\hat{x} + a\hat{v})\|\right)^2, \end{aligned}$$

where we have used (A.5). Finally,

$$\langle v(t) - \hat{v} + 2\sqrt{\mu}(x(t) - \hat{x}), \hat{v}\rangle \leq -\sqrt{\mu_s}t\langle\nabla f(\hat{x} + a\hat{v}), \hat{v}\rangle.$$

Employing these bounds in the expression of $\mathfrak{B}_{\mathrm{ET}}$, we obtain $|\mathfrak{B}_{\mathrm{ET}}(\hat{p}, t; a)| \leq \mathfrak{B}^q_{ST}(\hat{p}; a)t^2 + \mathfrak{B}^l_{\mathrm{ST}}(\hat{p}; a)t$, as claimed. □

Miguel Vaquero was born in Galicia, Spain. He received his Licenciatura and Master's degree in mathematics from the Universidad de Santiago de Compostela, Spain and the Ph.D. degree in mathematics from Instituto de Ciencias Matemáticas (ICMAT), Spain in 2015. He was then a postdoctoral scholar working on the ERC project "Invariant Submanifolds in Dynamical Systems and PDE" also at ICMAT. From 2017 to 2020, he was a postdoctoral scholar in the Department of Mechanical and Aerospace Engineering of UC San Diego. Since January 2021, he has been an Assistant Professor in the School of Human Sciences and Technology at IE University, Madrid, Spain. His interests include optimization, dynamical systems, control theory, machine learning, and geometric mechanics.

Pol Mestres received a Bachelor's Degree in Mathematics and a Bachelor's Degree in Engineering Physics from the Universitat Politècnica de Catalunya, Barcelona, Spain, in 2020. He was a visiting scholar at University of California, San Diego (UCSD) from September 2019 to March 2020, and is now a PhD student in the Dynamics and Control Graduate Program at UCSD. His research interests include optimization, dynamical systems, control theory and spreading processes in networks.

Jorge Cortés (M'02, SM'06, F'14) received the Licenciatura degree in mathematics from Universidad de Zaragoza, Zaragoza, Spain, in 1997, and the Ph.D. degree in engineering mathematics from Universidad Carlos III de Madrid, Madrid, Spain, in 2001. He held postdoctoral positions with the University of Twente, Twente, The Netherlands, and the University of Illinois at Urbana-Champaign, Urbana, IL, USA. He was an Assistant Professor with the Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA, USA, from 2004 to 2007. He is currently a Professor in the Department of Mechanical and Aerospace Engineering, University of California, San Diego, CA, USA. He is a Fellow of IEEE and SIAM. At the IEEE Control Systems Society, he has been a Distinguished Lecturer (2010-2014) and an elected member (2018-2020) of its Board of Governors, and is currently its Director of Operations. His research interests include distributed control and optimization, network science, resource-aware control, nonsmooth analysis, reasoning and decision making under uncertainty, network neuroscience, and multi-agent coordination in robotic, power, and transportation networks.