A hybrid control framework for fast methods under invexity: Non-Zeno trajectories with exponential rate

A. Sharifi Kolarijani, P. Mohajerin Esfahani and T. Keviczky

Abstract— **In this paper, we propose a framework to design a class of fast gradient-based methods in continuous-time that, in comparison with the existing literature including Nesterov's fast-gradient method, features a state-dependent, time-invariant damping term that acts as a feedback control input. The proposed design scheme allows for a user-defined, exponential rate of convergence for a class of nonconvex, unconstrained optimization problems in which the objective function satisfies the so-called Polyak–Łojasiewicz inequality. Formulating the optimization algorithm as a hybrid control system, a state-feedback input is synthesized such that a desired rate of convergence is guaranteed. Furthermore, we establish that the solution trajectories of the hybrid control system are Zeno-free.**

I. INTRODUCTION

Gradient-based optimization methods have received an increased level of interest from a wide range of communities recently because of their beneficiary properties such as ease of implementation. Originating from the *dynamical system* viewpoint to optimization algorithms suggested by Polyak in [22], a heavy-ball moving in a potential field, the *damped* 2nd-order ordinary differential equation (ODE)

$$\ddot{X}(t) + \gamma(t)\dot{X}(t) + \nabla f\big(X(t)\big) = 0 \tag{1}$$

has found numerous applications in design and analysis of a class of optimization algorithms. This class of algorithms is called *momentum-based* algorithms in the literature, where the function $\gamma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$ denotes the damping term. It has been shown that a higher rate of convergence can be achieved compared to the 1st-order ODE $\dot{X}(t) = -\nabla f\big(X(t)\big)$, i.e., the *gradient system*. This point of view has been employed and extended to various settings, see e.g., [1], [2].

Seemingly founded on a different basis, i.e., the notion of *estimate sequences*, Nesterov proposed his celebrated *accelerated algorithm* in [16] which has also been extended to other settings such as [18], [19]. Nonetheless, the performance of the approach remained mysterious despite many studies to describe the underlying principles of the accelerated methodology, such as [4], [6].

In this regard, the authors in [23] "suprisingly" discovered that Nesterov's method is a particular discretization of (1) with $\gamma(t) = \frac{3}{t}$. In other words, they have been able to show that acceleration-based methods and momentum-based methods are in fact the same in nature and originated from the ODE (1). We shall call the algorithms based on the 2nd-order ODE (1) *fast algorithms* in the rest of the paper.

Followed by the observation made in [23], the application of 2nd-order ODE (1) with a dynamical system viewpoint has become one of the prominent tools to design and analyze fast optimization methods. The ODE (1) is generalized into non-Euclidean settings and into higher order methods using the Bregman Lagrangian in [24]. Following [24], a "rate-matching" Lyapunov function is proposed in [25] with its monotonicity property established for both continuous and discrete dynamics. In the context of the dynamical system viewpoint, a control-oriented framework has been introduced to design and analyze optimization methods by [15]. The authors in [15] use the concept of integral quadratic constraints (IQC's) from the robust control literature to design iterative algorithms under the strong convexity assumption. The strong convexity assumption in [15] is relaxed to weaker assumptions, such as the quasiconvexity assumption, in [7] using an IQC-based approach. Utilizing dissipativity theory along with the IQC-based framework, in order to provide rate analyses, a framework to construct Lyapunov functions is proposed in [11].

In what follows, we mention some of the characteristic features of fast methods mainly in the continuous-time case:

- **Faster convergence rate:** Under convexity assumption, it has been shown that fast methods guarantee the convergence rate of $\mathcal{O}(\frac{1}{t^2})$ in continuous-time (and $\mathcal{O}(\frac{1}{k^2})$ in discrete-time where k is iteration index) whereas the gradient systems guarantee the convergence rate $\mathcal{O}(\frac{1}{t})$ (and the corresponding gradient descent method guarantees the convergence rate $\mathcal{O}(\frac{1}{k})$);
- **Non-monotonicity of fast methods:** Although fast methods guarantee an order of magnitude increase in the guaranteed convergence rate compared to the gradient-based methods, they suffer from a non-monotonic behavior. To avoid such an undesired behavior certain restarting schemes are proposed in the literature, such as the ones in [17], [21] in discrete-time and [23] in continuous-time.
- **Regularity for exponential convergence:** Under strong convexity assumption and generally using restarting schemes, it is shown that an exponential rate of convergence can be achieved, see e.g., [23], [25] in continuous-time fast methods (it is worth mentioning that the authors in [24] show an exponential convergence under uniform convexity assumption). In the discrete-time case, we refer the interested reader to [21] and the references therein for more details.

The authors are with the Delft Center for Systems and Control, TU Delft, The Netherlands ({a.sharifikolarijani, p.mohajerinesfahani,t.keviczky}@tudelft.nl).

Research hypothesis: Based upon the above discussion, one may raise the following question: *treating the damping term $\gamma(t)$ as a state-dependent input $u(x)$, is it possible to propose a framework to synthesize this input such that the properties of the underlying optimization method are improved?*

Our methodology (Synthesis of damping term): In this paper, we provide a *control-oriented* framework to achieve an exponential rate of convergence $\mathcal{O}(e^{-\alpha t})$ for an unconstrained, smooth optimization problem in the suboptimality measure $f\big(X(t)\big) - f^*$, *given* a positive scalar α. Motivated by the above hypothesis, we first amend the dynamical system (1) by substituting the *time-dependent* damping coefficient $\gamma(t)$ by a *state-dependent* feedback input in the form of $u\big(X(t), \dot{X}(t)\big)$. Inspired by restarting schemes, we next extend the class of dynamical systems represented by ODE's to the class of hybrid control systems. The amended 2nd-order ODE becomes the *continuous flow* in the hybrid formulation. We then construct a state-dependent feedback law such that the convergence rate of $\mathcal{O}(e^{-\alpha t})$ is guaranteed. The *flow set* of the hybrid system is specified based on an admissible control input range $[u_{\min}, u_{\max}]$. Finally, the *jump map* of the hybrid control system is defined through the mapping $(X^\top, -\beta\nabla^\top f(X))^\top$ such that the jump map's range is a subset of the flow set. This work is in continuation of the author's previous work [14] where a particular focus is given to analyzing the zeno-free feature of the resulting controlled hybrid systems. For the sake of brevity, we exclude the technical proofs and refer to [13], [14] in which the detailed proofs are provided. A summary of our main contributions in the context of continuous-time fast methods follows: (i) Guaranteeing a desired exponential convergence, we provide a hybrid control framework to synthesize the damping term $\gamma(t)$ as a state-dependent feedback input $u(X, \dot{X})$ unlike the common time-dependent damping term in the literature (Theorem 3.1); (ii) Our framework requires the objective function to satisfy the Polyak–Łojasiewicz (PL) inequality that is a weaker regularity assumption compared to the one mentioned in the literature (i.e., strong convexity) (Assumption (A2)); and (iii) Under the additional assumption that the Hessian of the objective function is Lipschitz (Assumption (A3)), we establish that the solution trajectories to the hybrid control system are Zeno-free.

Notations: The sets $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$ denote the n-dimensional Euclidean space and the space of $m\times n$ dimensional matrices with real entries, respectively. For a matrix $M \in \mathbb{R}^{m\times n}$, $M^\top$ is the transpose of M, $M \succ 0$ ($\prec 0$) refers to M positive (negative) definite, $M \succeq 0$ ($\preceq 0$) refers to M positive (negative) semi-definite, and $\lambda_{\max}(M)$ denotes the maximum eigenvalue of M. The $n \times n$ identity matrix is denoted by I_n. For a vector $v \in \mathbb{R}^n$ and $i \in \{1, \cdots, n\}$, v_i represents the i-th entry of v and $\|v\| := \sqrt{\Sigma_{i=1}^n v_i^2}$ is the Euclidean 2-norm of v. For two vectors $x, y \in \mathbb{R}^n$, $\langle x, y\rangle := x^\top y$ denotes the Euclidean inner product. For a matrix M, $\|M\| := \sqrt{\lambda_{\max}(A^\top A)}$ is the induced 2-norm. Given the set $S \subseteq \mathbb{R}^n$, ∂S and $\mathrm{int}(S)$ represent the boundary and the interior of S, respectively.

II. Preliminaries

In this section, we first introduce the notion of hybrid control system that is adapted from [8]. We then present the class of optimization problems considered in this study followed by the formal description of the problem statement.

Definition 2.1 (Hybrid control system): A time-invariant hybrid control system $\mathcal{H}$ comprises a controlled ODE and a jump (or a reset) rule introduced as:

$$\begin{cases} \dot{x} &= F\big(x, u(x)\big), \qquad x \in \mathcal{C} \\ x^+ &= G(x), \qquad \text{otherwise}, \end{cases} \tag{$\mathcal{H}$}$$

where $x^+ \in \mathbb{R}^n$ is the state of the hybrid system after a jump, the function $u : \mathbb{R}^n \to \mathbb{R}^m$ denotes a feedback signal, the function $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is the flow map, the set $\mathcal{C} \subseteq \mathbb{R}^n$ is the flow set, and the function $G : \partial\mathcal{C} \to \mathrm{int}(\mathcal{C})$ represents the jump map.
The concept of *Zeno behavior* in hybrid dynamical systems refers to the phenomenon that an infinite number of jumps occur in a bounded time interval. By contrast, we call a solution trajectory Zeno-free if the number of jumps within any finite time interval is bounded. A sufficient condition to fulfill the Zeno-free feature is to provide a lower bound for the time interval between any consecutive jumps. It is worth noting that there exist solution concepts in the literature that accept Zeno behaviors, see for example [3], [9].

Consider the class of unconstrained optimization problems

$$f^* := \min_{X\in\mathbb{R}^n} f(X), \tag{2}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is an objective function. We now state the main problem addressed in this paper.

Problem 2.2: Consider the unconstrained optimization problem (2) where the objective function f is twice differentiable. Given a positive scalar α, design a fast gradient-based method in the form of a hybrid control system ($\mathcal{H}$) with α-exponential convergence rate, i.e. for any initial condition $X(0)$ and any $t \geq 0$ we have

$$f\big(X(t)\big) - f^* \leq e^{-\alpha t}\Big(f\big(X(0)\big) - f^*\Big),$$

where $\{X(t)\}_{t\geq 0}$ denotes the solution trajectory to the system ($\mathcal{H}$).

Assumption 2.3 (Regularity assumptions): We stipulate that the objective function $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable and fulfills the following properties.

- (Bounded Hessian) The Hessian of function f, denoted by $\nabla^2 f(x)$, is uniformly bounded, i.e.,

$$-\ell_f I_n \preceq \nabla^2 f(x) \preceq L_f I_n, \tag{A1}$$

  where ℓ_f and L_f are non-negative constants.
- (Gradient dominated) The function f satisfies the Polyak-Łojasiewicz inequality with a positive constant μ_f, i.e., for every x in $\mathbb{R}^n$ the inequality

$$\frac{1}{2}\big\|\nabla f(x)\big\|^2 \geq \mu_f\big(f(x) - f^*\big), \tag{A2}$$

  holds, where f^* is the minimum of f on $\mathbb{R}^n$.

- (Lipschitz Hessian) The Hessian of the function f is Lipschitz, i.e., for every x, y in $\mathbb{R}^n$ the inequality

$$\big\|\nabla^2 f(x) - \nabla^2 f(y)\big\| \leq H_f \|x - y\|, \tag{A3}$$

holds, where H_f is a positive constant.

Remark 2.4 (Lipschitz gradient): Since the function f is twice differentiable, Assumption (A1) implies that the function ∇f is also Lipschitz with a positive constant L_f, i.e., for every x, y in $\mathbb{R}^n$ we have

$$\big\|\nabla f(x) - \nabla f(y)\big\| \leq L_f \|x - y\|. \tag{3}$$

We record some key features of functions that satisfy (A2).

Remark 2.5 (PL functions and invexity): The PL inequality implies the invexity of a function that is first introduced by [10]. The notion of invexity can be viewed as a generalization of the notion of convexity.

Remark 2.6 (Non-uniqueness of stationary points): The PL inequality requires the (not necessarily singular) set of stationary points of a function (i.e., $\{x : \nabla f(x) = 0\}$) to be global minimizers [5].

Example 1 (Popular PL functions [12]): Compositions of a strongly convex function and a linear function satisfy the PL inequality which include as an example least squares problems, i.e., $f(x) = \|Ax - b\|^2$. Any strictly convex function over a compact set satisfies the PL inequality. As such, the log-loss objective function in logistic regression, i.e., $f(x) = \Sigma_{i=1}^n \log\big(1 + \exp(b_i a_i^\top x)\big)$, is locally invex.

III. Main Results

We now present the main results of this work along with several remarks highlighting their implications. In what follows we use the notation $x := (x_1, x_2)$ such that the variables x_1 and x_2 represent the dynamics X and $\dot{X}$, respectively.

In the first step we provide a type of parameterization for the hybrid system ($\mathcal{H}$). Given a positive scalar α, the proposed parameterization denoted by $u_\alpha(x)$ enables achieving the rate of convergence $\mathcal{O}(e^{-\alpha t})$ in the suboptimality measure $f\big(X(t)\big) - f^*$. Motivated by the dynamics of fast gradient methods [23], we start with a 2nd-order ODE as the continuous evolution (or the flow map) $F : \mathbb{R}^{2n} \times \mathbb{R} \to \mathbb{R}^{2n}$ defined as

$$F\big(x, u_\alpha(x)\big) = \begin{pmatrix} x_2 \\ -\nabla f(x_1) \end{pmatrix} + \begin{pmatrix} 0 \\ -x_2 \end{pmatrix} u_\alpha(x). \tag{4a}$$

The feedback law $u_\alpha : \mathbb{R}^{2n} \to \mathbb{R}$ is given by

$$u_\alpha(x) = \alpha + \frac{\|\nabla f(x_1)\|^2 - \langle \nabla^2 f(x_1) x_2, x_2 \rangle}{\langle \nabla f(x_1), -x_2 \rangle}. \tag{4b}$$

The main feature of the proposed control structure is to ensure achieving an α-exponential convergence rate, see [14, Subsection 4.1] for more details. Notice that the state-dependent feedback input $u_\alpha(x)$ has replaced the time-dependent damping term $\gamma(t)$ in the dynamics (1). In the next step, we consider an admissible interval $[u_{\min}\ u_{\max}]$ to characterize a candidate flow set $\mathcal{C} \subset \mathbb{R}^{2n}$, i.e.,

$$\mathcal{C} = \big\{x \in \mathbb{R}^{2n} :\ u_\alpha(x) \in [u_{\min}, u_{\max}]\big\}, \tag{4c}$$

where $u_{\min}, u_{\max}$ represent the range of acceptable control values. Notice that the flow set $\mathcal{C}$ is the domain in which the hybrid system ($\mathcal{H}$) can evolve continuously. Furthermore, observe that when $\langle \nabla f(x_1), -x_2 \rangle = 0$ (e.g., at an optimal state x_1^* where $\nabla f(x_1^*) = 0$), the input (4b) is not well-defined. Nonetheless, the way the flow set $\mathcal{C}$ is defined in (4c) prevents the occurrence of such a case. We finally introduce the jump map $G : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ parameterized by a constant β

$$G(x) = \begin{pmatrix} x_1 \\ -\beta \nabla f(x_1) \end{pmatrix}. \tag{4d}$$

The parameter β ensures that the range space of the jump map G is a strict subset of $\mathrm{int}(\mathcal{C})$. By construction, one can inspect that any neighborhood of the optimizer x_1^* has a non-empty intersection with the flow set $\mathcal{C}$. That is, there always exist paths in the set $\mathcal{C}$ that allow the continuous evolution of the Hybrid system to approach arbitrarily close to the optimizer.

The first result of this section introduces a mechanism to compute the hybrid system's parameters $u_{\min}$, $u_{\max}$, and β to achieve the desired exponential convergence rate $\mathcal{O}(e^{-\alpha t})$.

Theorem 3.1 (Continuous-time hybrid dynamics [14]): Consider a positive scalar α and a smooth function $f : \mathbb{R}^n \to \mathbb{R}$ satisfying Assumptions (A1) and (A2). Then, the trajectory of the continuous-time hybrid control system ($\mathcal{H}$) with the respective parameters (4) and starting from any initial condition $x(0)$ with $\nabla f\big(x_1(0)\big) \neq 0$ satisfies

$$f\big(x_1(t)\big) - f^* \leq e^{-\alpha t}\Big(f\big(x_1(0)\big) - f^*\Big), \quad \forall t \geq 0, \tag{5}$$

if the scalars $u_{\min}$, $u_{\max}$, and β are chosen such that

$$u_{\min} < \alpha + \beta^{-1} - L_f \beta, \tag{6a}$$

$$u_{\max} > \alpha + \beta^{-1} + \ell_f \beta, \tag{6b}$$

$$\alpha \leq 2\mu_f \beta. \tag{6c}$$

Proof: See the proof of [14, Theorem 3.1]. ■

Remark 3.2 (Weaker regularity than strong convexity): The PL inequality is a weaker requirement than the strong convexity, which is often assumed in similar contexts [23], [24], [25]. It is worth noting that such a condition has also been used in the context of 1st-order oracle, non-accelerated algorithms [12].

Remark 3.3 (Hybrid embedding of restarting): The hybrid framework intrinsically captures a restarting scheme through the jump map. The scheme is a weighted gradient where the weight factor β is essentially characterized by the given data α, μ_f, ℓ_f, and L_f. One may inspect that the constant β can be in fact introduced as a state-dependent weight factor to potentially improve the performance. Nonetheless, for the sake of simplicity of exposition, we do not pursue this level of generality in this paper.

Remark 3.4 (Fundamental limits on control input): In order to guarantee the rate of convergence of $\mathcal{O}\big(e^{-\alpha t}\big)$, Theorem 3.1 asserts the following theoretical limits on $u_{\min}$ and $u_{\max}$: (i) the upper-bound on the admissible input interval $u_{\max}$ is required to be larger than α, and (ii) the lower-bound on the admissible input interval $u_{\min}$ has to be *negative* if the geometrical property $\alpha \geq \frac{2\mu_f}{\sqrt{\max\{L_f - 2\mu_f, 0\}}}$ holds based on the given α. In other words, Theorem 3.1 essentially indicates a lack of geometrical richness of the function f. As a result, it is required to *inject energy* to the dynamical system through *negative* damping in order to achieve an exponential rate of convergence. It is also natural to expect that the practical rate of convergence increases as $u_{\min}$ decreases.

Remark 3.5 (2nd-order information): Although our proposed framework requires 2nd-order information, i.e., the Hessian $\nabla^2 f$, this requirement only appears in a mild form as an evaluation in the same spirit as the modified Newton step proposed in [20]. Furthermore, we emphasize that our results still hold true if one replaces $\nabla^2 f(x_1)$ with its upper-bound $L_f I_n$ following essentially the same analysis. For further details we refer the reader to [14, Subsection 4.1].

The next result establishes a key feature of the solution trajectories generated by the dynamics $(\mathcal{H})$ with the respective parameters (4), that the solution trajectories are indeed *Zeno*-free. The Zeno-free feature is particularly of interest in this context since it facilitates the development of discrete-time algorithms based on the continuous-time analysis performed in this paper.

Theorem 3.6 (Zeno-free hybrid trajectories [13]): Consider a smooth function $f : \mathbb{R}^n \to \mathbb{R}$ satisfying Assumption 2.3, and the corresponding hybrid control system $(\mathcal{H})$ with the respective parameters (4) satisfying (6). Given the initial condition $\big(x_1(0), -\beta\nabla f\big(x_1(0)\big)\big)$ with $\nabla f\big(x_1(0)\big) \neq 0$, the time between two consecutive jumps of the solution trajectory, denoted by τ, satisfies for any scalar $r > 1$

$$\tau \geq \log\left(\max\left\{\frac{a_1}{a_2 + a_3\left\|\nabla f\big(x_1(0)\big)\right\|} + 1, r^{1/\delta}\right\}\right), \quad (7)$$

where the constants involved are defined as

$$C := \frac{(u_{\max} - \alpha) + \sqrt{(u_{\max} - \alpha)^2 + 4L_f}}{2}, \quad (8a)$$

$$\delta := C + \max\{u_{\max}, -u_{\min}\}, \quad (8b)$$

$$\mathcal{L}_f := \max\{\ell_f, L_f\}, \quad (8c)$$

$$a_1 := \min\{u_{\max} - (\alpha + \beta^{-1} + \ell_f\beta), \quad (8d)$$

$$(\alpha + \beta^{-1} - L_f\beta) - u_{\min}\}, \quad (8e)$$

$$a_2 := rL_f\delta^{-1}(r\beta C + 1) + \beta^{-1} + (r^2 + r + 1)\beta\mathcal{L}_f, \quad (8f)$$

$$a_3 := r^3\beta^2 H_f\delta^{-1}. \quad (8g)$$

Consequently, the solution trajectories are Zeno-free.

Proof: See the proof of [13, Theorem 3.2]. ■

Remark 3.7 (Non-uniform inter-jumps): Notice that Theorem 3.6 suggests a lower bound for the inter-jump interval τ that depends on $\|\nabla f\big(x_1\big)\|$. In light of the fact that the solution trajectories converge to the optimal solutions, and as such $\nabla f\big(x_1\big)$ tends to zero, one can expect that the frequency at which the jumps occur reduces as the hybrid control system evolves in time. We refer to Section IV for a numerical example illustrating this phenomenon.

The main novelty of the proposed approach is to allude to a more general framework (i.e., the hybrid formulation) to capture the dynamics of fast methods, even the ones with restarting schemes. Although we have focused on the parameterization (4) for the hybrid dynamics $(\mathcal{H})$ by following a trajectory-based analysis, it is not difficult to see that other fast methods, whether continuous- or discrete-time, with either time- or state-dependent restarting schemes, can be brought to a hybrid formulation. More importantly, the authors believe that the fast methods in the literature that are already accompanied with Lyapunov functions are extremely suitable for this framework. As a result, one is then enabled to employ other specifically-tailored tools for hybrid systems in control theory and can hopefully provide a more unified form of analysis and interpretation for such methods.

In regard to the practical aspect of the user-defined rate α, one should bear in mind that although the continuous-time analysis in this paper does not warrant any restriction on α, a large α does, however, push the inter-jump interval τ toward zero, see Theorem 3.6. Thus, the impact of a large α is the possibility of more frequent jumps as the dynamics evolve in time. Observe that the continuous flow is the part of the hybrid dynamics that is the main source of speeding up the rate of convergence in the suboptimality measure. The jump map is simply a weighted gradient descent. Suppose that one seeks to discretize the continuous hybrid dynamics with a large α to attain a discrete-time hybrid control system (i.e., an iterative optimization algorithm). A large α may seem attractive in a continuous-time setting but this choice may lead to (i) a very small step size for the time discretization or (ii) frequent activation of the jump map. As a result, one should take into account other considerations in choosing α. We refer the reader to [14, Theorem 3.7] for a more detailed discussion.

IV. Numerical Example

In what follows, we present an academic example that by itself does not warrant the effectiveness of our proposed method (and in general gradient-based methods) but it serves as a comparison. Consider the problem (2) with the function $f(x_1) = x_1^\top Q x_1$ where $Q = \text{diag}\{0.1, 0.2, \cdots, 0.5\}$. It is straightforward to show that $L_f = 2\lambda_{max}(Q) = 1$, $\mu_f = 2\lambda_{\min}(Q) = 0.2$, and $\ell_f = 0$ (since f is convex). The hybrid system $(\mathcal{H})$ with the parameters (4) is simulated for two sets of parameters given in Table I, namely HD1 and HD2, which are depicted by the solid blue and dashed black lines in the figures, respectively. The results are also compared to Nesterov's accelerated dynamics (1) with $\gamma(t) = \frac{3}{t}$ employing the "speed restarting" proposed by [23], which we call NSR and it is depicted by the dash-dotted brown line in the figures.

Fig. 1 shows that the trajectories generated by the dynamics $(\mathcal{H})$ with the respective parameters HD1 or HD2 possess faster convergence behavior compared to $e^{-\alpha t}\big(f\big(x_1(0)\big) - f^*\big)$ (depicted by the solid yellow line) validating the claim

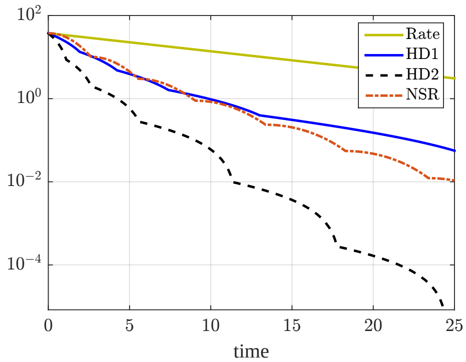

Fig. 1. Rate bound $e^{-\alpha t}\big(f\big(x_1(0)\big) - f^*\big)$, suboptimality decay $f\big(x_1(t)\big) - f^*$ for the trajectories of the hybrid systems (HD1 and HD2), and the speed restarting Nesterov's accelerated scheme (NSR) [23].

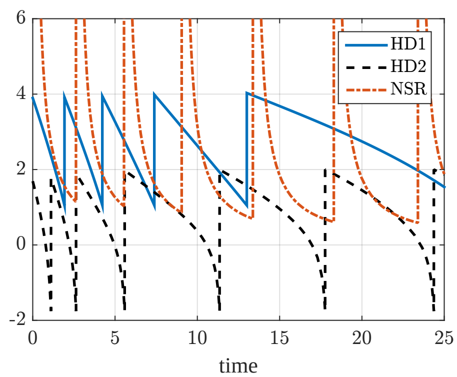

Fig. 2. Control inputs $u_\alpha\big(x(t)\big)$ of the hybrid systems (HD1 and HD2), and the damping coefficient $\gamma(t)$ for the speed restarting Nesterov's accelerated scheme (NSR) [23].

made in Theorem 3.1. Fig. 2 depicts the variation of damping term in the three cases NSR, HD1, and HD2. Three observations are due here. First, an exponential rate of convergence can be attained via a bounded interval for the damping term; the general form of damping term considered in the literature is $\frac{q}{t}$ where q is a positive scalar. Second, our framework allows the damping term to admit negative values in certain time intervals during which the convergence rate increases. Third, as the dynamics ($\mathcal{H}$) progress in time, the inter-jump intervals increase validating the assertion made in Remark 3.7.

V. Conclusions and Future Works

A system-theoretic framework was introduced to construct a class of optimization methods for unconstrained, smooth optimization problems satisfying the Polyak–Łojasiewicz inequality in continuous time. Our methodology is trajectory-based and enables achieving a desired, exponential rate of convergence. There are obvious directions that can be pursued to extend the results in this work, such as constraint satisfaction, parallelization, distributed implementation. However, we believe that in order to claim practical benefits of the presented results, for example in large scale problems, the next necessary step is to provide a discretization method for the continuous-time hybrid system proposed here, that is the subject of our current studies.

TABLE I

Hybrid dynamical system parameters.

	α	β	$u_{\min}$	$u_{\max}$
HD1	0.1	0.25	1.05	24.6
HD2	0.1	0.5	-1.755	12.6

References

[1] F. Alvarez, *On the minimizing property of a second order dissipative system in Hilbert spaces*, SIAM Journal on Control and Optimization, 38 (2000), pp. 1102–1119.

[2] H. Attouch and J. Bolte, *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*, Mathematical Programming, 116 (2009), pp. 5–16.

[3] J.-P. Aubin, J. Lygeros, M. Quincampoix, S. Sastry, and N. Seube, *Impulse differential inclusions: a viability approach to hybrid systems*, IEEE Transactions on Automatic Control, 47 (2002), pp. 2–20.

[4] S. Bubeck, Y. T. Lee, and M. Singh, *A geometric alternative to Nesterov's accelerated gradient descent*, arXiv preprint arXiv:1506.08187, (2015).

[5] B. D. Craven and B. M. Glover, *Invex functions and duality*, Journal of the Australian Mathematical Society, 39 (1985), pp. 1–20.

[6] Y. Drori and M. Teboulle, *Performance of first-order methods for smooth convex minimization: a novel approach*, Mathematical Programming, 145 (2014), pp. 451–482.

[7] M. Fazlyab, A. Ribeiro, M. Morari, and V. M. Preciado, *Analysis of optimization algorithms via integral quadratic constraints: Non-strongly convex problems*, arXiv preprint arXiv:1705.03615, (2017).

[8] R. Goebel, R. G. Sanfelice, and A. R. Teel, *Hybrid dynamical systems: modeling, stability, and robustness*, Princeton University Press, 2012.

[9] R. Goebel and A. R. Teel, *Solutions to hybrid inclusions via set and graphical convergence with stability theory applications*, Automatica, 42 (2006), pp. 573–587.

[10] M. A. Hanson, *On sufficiency of the Kuhn-Tucker conditions*, Journal of Mathematical Analysis and Applications, 80 (1981), pp. 545–550.

[11] B. Hu and L. Lessard, *Dissipativity theory for Nesterov's accelerated method*, arXiv preprint arXiv:1706.04381, (2017).

[12] H. Karimi, J. Nutini, and M. Schmidt, *Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition*, Springer International Publishing, 2016, pp. 795–811.

[13] A. S. Kolarijani, P. Mohajerin Esfahani, and T. Keviczky, *Continuous-time accelerated methods via a hybrid control lens*, arXiv preprint arXiv:1807.07805, (2018).

[14] ———, *Fast gradient-based methods with exponential rate: A hybrid control framework*, in International Conference on Machine Learning, 2018, pp. 2733–2741.

[15] L. Lessard, B. Recht, and A. Packard, *Analysis and design of optimization algorithms via integral quadratic constraints*, SIAM Journal on Optimization, 26 (2016), pp. 57–95.

[16] Y. Nesterov, *A method of solving a convex programming problem with convergence rate $\mathcal{O}(1/k^2)$*, in Soviet Mathematics Doklady, vol. 27, 1983, pp. 372–376.

[17] ———, *Introductory lectures on convex optimization: a basic course*, Springer Science and Business Media, 2004.

[18] ———, *Smooth minimization of non-smooth functions*, Mathematical Programming, 103 (2005), pp. 127–152.

[19] ———, *Gradient methods for minimizing composite functions*, Mathematical Programming, 140 (2013), pp. 125–161.

[20] Y. Nesterov and B. T. Polyak, *Cubic regularization of newton method and its global performance*, Mathematical Programming, 108 (2006), pp. 177–205.

[21] B. O'Donoghue and E. Candès, *Adaptive restart for accelerated gradient schemes*, Foundations of Computational Mathematics, 15 (2015), pp. 715–732.

[22] B. T. POLYAK, *Some methods of speeding up the convergence of iteration methods*, USSR Computational Mathematics and Mathematical Physics, 4 (1964), pp. 1–17.

[23] W. SU, S. BOYD, AND E. CANDÈS, *A differential equation for modeling Nesterov's accelerated gradient method: theory and insights*, Journal of Machine Learning Research, 17 (2016), pp. 1–43.

[24] A. WIBISONO, A. C. WILSON, AND M. I. JORDAN, *A variational perspective on accelerated methods in optimization*, Proceedings of the National Academy of Sciences, 113 (2016), pp. E7351–E7358.

[25] A. C. WILSON, B. RECHT, AND M. I. JORDAN, *A Lyapunov analysis of momentum methods in optimization*, arXiv preprint arXiv:1611.02635, (2016).