Accelerated Adaptive Concurrent Learning]Accelerated Concurrent Learning Algorithms via Data-Driven Hybrid Dynamics and Nonsmooth ODEs

(Extended Version with Proofs) **Robust Optimization over Networks Using Distributed Restarting of Accelerated Dynamics**

Technical Report

Daniel E. Ochoa, Jorge I. Poveda, César A. Uribe, Nicanor Quijano

University of Colorado **Boulder**

[

Daniel E. Ochoa, Jorge I. Poveda, César A. Uribe, Nicanor Quijano

Abstract

We present a new accelerated distributed algorithm for the robust solution of convex optimization problems over networks. We propose a novel distributed restarting mechanism for accelerated optimization dynamics with individual asynchronous time-varying coefficients. Graph-dependent restarting conditions are derived to establish suitable stability, convergence, and robustness properties for problems characterized by strongly convex smooth and non-smooth primal functions. Since the algorithm combines continuous-time dynamics and discrete-time dynamics, we model the complete system as a hybrid dynamical system. Numerical results illustrate our results.

I. Introduction

We study the accelerated, efficient and robust solution of accelerated distributed optimization problems over network systems characterized by connected and undirected graphs $\mathcal{G} := (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \ldots, n\}$ is the set of nodes, and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set of edges. We consider the setting where each node i has a local function $f_i : \mathbb{R}^p :\rightarrow \mathbb{R}$, and the network cooperates to find a common point $z^* \in \mathbb{R}^p$ that minimizes a global function defined as the summation of the local costs. This distributed optimization problem can be written as

$$\min_{z_1, z_2, \ldots, z_n \in \mathbb{R}^p} \quad \sum_{i=1}^{n} f_i(z_i), \quad \text{s.t.} \quad z_i = z_j, \quad \forall\ i, j \in \mathcal{V}, \tag{1}$$

which is also known in the literature as the *consensus-optimization* problem [1], and which has been shown to be relevant for several engineering applications in areas such as power systems, transportation systems, water distribution systems, and distributed network control, see [2] and references therein.

Discrete-time and continuous-time approaches to solve problem (1) have been extensively studied using gradient descent and Newton-based dynamics in [3], [4], primal-dual dynamics [5], and projected dynamics [6], to name just a few. However, a persistent challenge in the solution of problem (1) is to achieve fast rates of convergence without sacrificing essential robustness properties of the algorithms. As recently shown in [7], [8], this task is not trivial given that certain classes of accelerated continuous-time algorithms, such as Nesterov's ODE [9]–[11], can be destabilized under arbitrarily small disturbances on the states or gradients. Since these disturbances are unavoidable in practice, there is an urgent need for the development of robust, accelerated and distributed algorithms for the solution of problem (1).

In the literature of accelerated centralized optimization, one of the approaches that has received significant attention during the last years is the incorporation of restarting techniques. As a mater of fact, as shown in [9], [12], [13], [14], and [15], accelerated algorithms with restarting techniques can achieve exponential convergence rates in strongly convex optimization problems without having perfect knowledge of the condition number of the cost function. Moreover, restarting can also be used to induce suitable robustness properties in the Nesterov's ODE, provided the combination of the continuous-time dynamics and the discrete-time dynamics is carefully carried out [7]. While these ideas have been explored and validated in centralized optimization problems, as

D. E. Ochoa and J. I. Poveda are with the Department of Electrical, Computer and Energy Eng. at the University of Colorado, Boulder. Email: { *daniel.ochoa, jorge.poveda*}*@colorado.edu*. C. Uribe is with the LIDS lab at the Massachusetts Institute of Technology. Email: *cauribe@mit.edu*. N. Quijano is with the Department of Electronics Eng. at the University of Los Andes, Colombia. Email: *nquijano@uniandes.edu.co*.

Work supported in part by CU Boulder ASIRT Seed Grant, NSF grant CNS-1947613, and the the Yahoo! Faculty Engagement Program.

mentioned in [16], it remains an open question whether or not similar techniques could be pursued for distributed optimization problems of the form (1). As we will show in this paper, the answer to this question turns out to be positive.

The main contribution of this paper is the formulation and analysis of the first *robust and* distributed *restarting-based accelerated dynamics for the solution of network optimization problems of the form* (1). Since our restarting dynamics combine continuous-time dynamics and discrete-time dynamics, they are modeled as set-valued hybrid dynamical systems [17], for which stability, convergence, and robustness properties can be established using Lyapunov functions and the hybrid invariance principle. The construction of this hybrid system is not trivial due to the distributed nature of the system, which allows for multiple discrete-time updates in the network happening simultaneously in the standard time domain. In contrast to existing results that use projections or primal-dual approaches, we follow a complete dual approach that allows us to recast problem (1) as an unconstrained optimization problem with a suitable Laplacian-dependent structure on the dynamics of the momentum variables [16]. This reformulation, allows us to establish sufficient graph-dependent restarting conditions for the solution of the primal problem. To the knowledge of the authors, these are the first restarting results developed for accelerated distributed optimization algorithms.

The rest of this paper is organized as follows. Section II presents some preliminaries on hybrid dynamical systems. Section III presents the algorithm and the convergence results followed by Section IV that presents some numerical examples. Section V presents the analysis, and finally Section VI adds some concluding remarks.

Notation: We define $\mathbf{c}_n \in \mathbb{R}^n$ as the vector with all entries equal to $c \in \mathbb{R}$, and we use $|\cdot|$ as the Euclidean norm. We use $|x|_{\mathcal{A}} := \min_{y\in\mathcal{A}} |x-y|$ to denote the distance of a vector $x \in \mathbb{R}^n$ with respect to a compact set $\mathcal{A}$. A function $\phi : \mathbb{R}^n \to \mathbb{R}$ is radially unbounded if $\phi(x) \to \infty$ as $|x| \to \infty$, and it is said to be of class $\mathcal{C}^k$ if its k^{th} derivative is continuous. We use $\mathbf{D}(x) \in \mathbb{R}^{n\times n}$ to denote the diagonal matrix with diagonal given by the vector $x \in \mathbb{R}^n$. We use $I_n \in \mathbb{R}^{n\times n}$ to denote the identity matrix, and $S^n := S \times S \times \ldots \times S$ to denote the nth-Cartesian product of the set S.

II. Preliminaries

In this paper, we analyze optimization algorithms modeled as HDS [17], of the form

$$p \in C, \qquad \dot{p} = F(p) \tag{2a}$$

$$p \in D, \qquad p^+ \in G(p), \tag{2b}$$

where $p \in \mathbb{R}^n$ is the state, $F : \mathbb{R}^n \to \mathbb{R}^n$ is called the flow map, and $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is called the jump map. The sets C and D, called the flow set and the jump set, respectively, characterize the points in $\mathbb{R}^n$ where the system can flow or jump via equations (2a) or (2b), respectively. The *data* of the HDS is defined as the tuple $\mathcal{H} := \{C, F, D, G\}$. Solutions to HDS of the form (2) are defined on hybrid time domains, i.e., they are parameterized by both a continuous-time index $t \in \mathbb{R}_{\geq 0}$, and a discrete-time index $j \in \mathbb{Z}_{\geq 0}$. Consequently, the notation $\dot{p}$ in (2a) represents the derivative of p with respect to time t, i.e., $\frac{dp(t,j)}{dt}$; and p^+ in (2b) represents the value of p after an instantaneous jump, i.e., $p(t, j+1)$. For a precise definition of hybrid time domains and solutions to HDS of the form (2) we refer the reader to [17, Ch.2]. A HDS $\mathcal{H}$ is said to be well-posed if C and D are closed sets, $C \subset \text{dom}(F)$ and $D \subset \text{dom}(G)$, F is continuous in C, and G is outer-semicontinuous [17, Def. 5.9] and locally bounded [17, Def. 5.14] relative to D. A compact set $\mathcal{A} \subset \mathbb{R}^n$ is said to be uniformly globally asymptotically stable (UGAS) for a HDS $\mathcal{H}$ if there exists a class $\mathcal{KL}$ function [17, Def. 3.38] β such that every solution of $\mathcal{H}$ satisfies $|p(t,j)|_{\mathcal{A}} \leq \beta(|p(0,0)|_{\mathcal{A}}, t+j)$, for all $(t,j) \in \text{dom}(p)$. Note that the UGAS property is stronger than standard convergence notions used in optimization. In particular, UGAS implies *uniform* global stability *and uniform* global attractivity. For compact sets $\mathcal{A}$ and well-posed HDS, UGAS can be used to additionally certify desirable robustness properties.

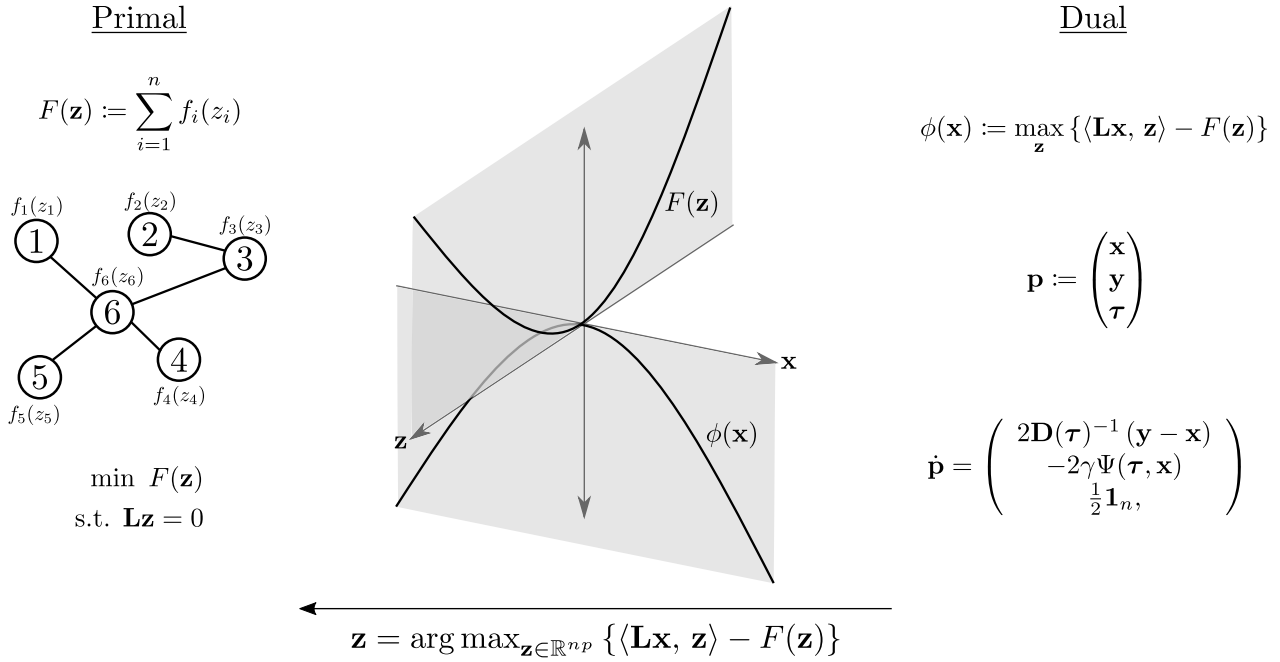

Fig. 1: We design an algorithm to solve the consensus-optimization problem from a complete dual perspective, by constructing dynamics for the *dual* variable $\mathbf{x} \in \mathbb{R}^{np}$. The proposed accelerated dynamics make use of a momentum variable $\mathbf{y} \in \mathbb{R}^{np}$, and timers τ_i for each one of the agents. In order to obtain the value of the *primal* variable $\mathbf{z}$ we make use of the relation $\mathbf{z} = \arg\max_{\mathbf{z}\in\mathbb{R}^{np}}\{\langle \mathbf{L}\mathbf{x},\, \mathbf{z}\rangle - F(\mathbf{z})\}$, where $\mathbf{L} = \mathcal{L} \otimes I_p \in \mathbb{R}^{np} \times \mathbb{R}^{np}$ and $\mathcal{L}$ is the Laplacian matrix of the communication graph $\mathcal{G}$.

III. Accelerated Dynamics with Distributed Restarting

To solve problem (1), let $\mathbf{z} := [z_1^\top, z_2^\top, \dots, z_n^\top]^\top$ be the concatenation of the local decision variables of the nodes of the network $\mathcal{G}$. Define the global cost function $F(\mathbf{z}) := \sum_{i=1}^{n} f_i(z_i)$, and let $\mathbf{L} := \mathcal{L} \otimes I_p \in \mathbb{R}^{np} \times \mathbb{R}^{np}$, where $\mathcal{L}$ is the Laplacian matrix of the graph $\mathcal{G}$. We will make the following assumption on Problem (1):

Assumption 3.1

The local cost functions f_i are $\mathcal{C}^k$ and μ_i-strongly convex, i.e., there exists $\mu_i > 0$ such that for any $x, y \in \mathbb{R}^n$, $f_i(y) \geq f_i(x) + \nabla f_i(x)^\top (y - x) + \frac{\mu_i}{2}|y - x|^2$. The graph $\mathcal{G}$ is undirected, connected, and time-invariant. □

By Assumption 3.1, the extended Laplacian matrix $\mathbf{L}$ satisfies $\mathbf{L} = \mathbf{L}^\top$, $\ker(\mathbf{L}) = \text{span}(\mathbf{1}_{np})$, and $\ker(\mathbf{L})^\perp = \left\{\mathbf{y} \in \mathbb{R}^{np} : \mathbf{1}_{np}^\top \mathbf{y} = 0\right\}$. Thus, we can write Problem (1) as

$$\min_{\mathbf{z}\in\mathbb{R}^{np}} F(\mathbf{z}), \quad \text{s.t.} \quad \mathbf{L}\mathbf{z} = \mathbf{0}_{np}, \tag{3}$$

where F is also $\bar{\mu}$-strongly convex with $\bar{\mu} := \min_{i\in\mathcal{V}} \mu_i$. When the local gradients ∇f_i are also globally L_i-Lipschitz (a condition that we do not necessarily assume), the global gradient ∇F is globally $\bar{L}$-Lipschitz with $\bar{L} = \max_{i\in\mathcal{V}} L_i$.

To solve problem (3), we consider its dual problem:

$$\min_{\mathbf{x}\in\mathbb{R}^{np}} \phi(\mathbf{x}), \quad \text{with} \quad \phi(\mathbf{x}) := \max_{\mathbf{z}\in\mathbb{R}^{np}} \{\langle \mathbf{L}\mathbf{x},\, \mathbf{z}\rangle - F(\mathbf{z})\}, \tag{4}$$

which, as shown in [16], has zero duality gap under Assumption 3.1. By defining the mapping $h : \mathbb{R}^{np} \to \mathbb{R}^{np}$ as $h(\mathbf{u}) := \arg\max_{\mathbf{z}\in\mathbb{R}^{np}} \{\langle \mathbf{u}, \mathbf{z}\rangle - F(\mathbf{z})\}$, the gradient of the dual function ϕ can be computed as follows

$$\nabla\phi(\mathbf{x}) = \mathbf{L}h\big(\mathbf{L}\mathbf{x}\big). \tag{5}$$

Let $\mathcal{L}^2 := \mathcal{L}^\top\mathcal{L}$, and denote as $\lambda^+_{\min}(\mathcal{L}^2)$ the smallest positive eigenvalue of $\mathcal{L}^2$, and $\lambda_{\max}(\mathcal{L}^2)$ as its largest eigenvalue. The next lemma follows directly from the results in [16].

Lemma 3.1

The function ϕ is convex. Moreover, if F is $\bar{\mu}$-strongly convex and $\mathcal{C}^k$ at the point $\mathbf{z} := h(\mathbf{L}\mathbf{x})$, then ϕ is also $\mathcal{C}^k$ at $\mathbf{x}$, and $\nabla\phi$ is globally ℓ_ϕ-Lipschitz with $\ell_\phi = \lambda_{\max}(\mathcal{L}^2)/\bar{\mu}$. If, additionally, ∇F is globally $\bar{L}$-Lipschitz, then ϕ is also μ_ϕ-strongly convex on $\ker(\mathbf{L})^\perp$, with $\mu_\phi = \lambda^+_{\min}(\mathcal{L}^2)/\bar{L}$. □

Let $\mathcal{A}_\phi \subset \mathbb{R}^{np}$ be the set of solutions of Problem (4) subject to $\mathbf{x} \in \ker(\mathbf{L})^\perp$. By Lemma 3.1, if ∇F is not globally Lipschitz, the set $\mathcal{A}_\phi$ may not necessarily be bounded. Therefore, we will make the following technical assumption on ϕ.

Assumption 3.2

The level sets of the dual function ϕ with domain restricted to $\ker(\mathbf{L})^\perp$ are bounded. □

Finally, when ∇F is also globally $\bar{L}$-Lipschitz, we define the condition numbers of $\mathcal{L}^2$, and F, respectively, as $\kappa_{\mathcal{L}^2} := \lambda_{\max}(\mathcal{L}^2)/\lambda^+_{\min}(\mathcal{L}^2)$, and $\kappa_F := \bar{L}/\bar{\mu}$. These condition numbers will play an important role in the linear convergence properties of our algorithms.

A. Hybrid Dynamics with Distributed Restarting

We solve Problem (4) by considering a class of algorithms termed Hybrid Accelerated Restarting Distributed Dynamics (HARDD). Each node $i \in \mathcal{V}$ is endowed with three local states (x_i, y_i, τ_i), where $x_i, y_i \in \mathbb{R}^p$, and $\tau_i \in \mathbb{R}$ is a local timer. The overall network has states $\mathbf{x} := [x_1^\top, x_2^\top, \dots, x_n^\top] \in \mathbb{R}^{np}$, $\mathbf{y} := [y_1^\top, y_2^\top, \dots, y_n^\top]^\top \in \mathbb{R}^{np}$, and $\boldsymbol{\tau} := [\tau_1, \tau_2, \dots, \tau_n]^\top \in \mathbb{R}^n$. Using $\mathbf{p} := [\mathbf{x}^\top, \mathbf{y}^\top, \boldsymbol{\tau}^\top]^\top$, the continuous-time dynamics of the algorithms are given by:

$$\dot{\mathbf{p}} = F_A(\mathbf{p}) := \begin{pmatrix} 2\mathbf{D}(\boldsymbol{\tau}\otimes\mathbf{1}_p)^{-1}(\mathbf{y}-\mathbf{x}) \\ -2\gamma\Psi(\boldsymbol{\tau},\mathbf{x}) \\ \frac{1}{2}\mathbf{1}_n \end{pmatrix}, \tag{6}$$

where $\gamma \in \mathbb{R}_{>0}$ is a tunable gain, and $\Psi : \mathbb{R}^n \times \mathbb{R}^{np} \to \mathbb{R}^{np}$ is a mapping to be defined below. The dynamics (6) are allowed to evolve in the flow set:

$$\mathbf{p} \in C_A := \mathbb{R}^{np} \times \ker(\mathbf{L})^\perp \times [T_r, T_r + \Delta T]^n, \tag{7}$$

where $T_r > 0$ and $\Delta T > 0$ are tunable parameters. The discrete-time dynamics of the algorithms describe the restarting mechanism. They are parameterized by a constant $q \in \{0, 1\}$, a set-valued mapping $\mathcal{T} : \mathbb{R}^{2np+n} \rightrightarrows \mathbb{R}^{np+n}$, and a set $D_\tau \subset \mathbb{R}^n$. These dynamics are modeled by the following difference inclusion:

$$\mathbf{p}^+ \in G_A(\mathbf{p}) := \begin{pmatrix} \{(1-q)\mathbf{x}+q\mathbf{s}\} \\ \{\mathbf{y}\} \\ \{\mathbf{g}\} \end{pmatrix}, \ (\mathbf{s},\mathbf{g}) \in \mathcal{T}(\boldsymbol{p}) \Bigg), \tag{8}$$

which is allowed to evolve in the jump set:

$$\mathbf{p} \in D_A := \mathbb{R}^{np} \times \ker(\mathbf{L})^\perp \times D_\tau. \tag{9}$$

The constant q characterizes two different restarting algorithms: If ∇F is globally Lipschitz, then $q = 1$; otherwise, $q = 0$. The mapping $\Psi : \mathbb{R}^n \times \mathbb{R}^{np} \to \mathbb{R}^{np}$ in (6) is defined as

$$\Psi(\boldsymbol{\tau}, \mathbf{x}) := \mathbf{L}\mathbf{D}(\boldsymbol{\tau} \otimes \mathbf{1}_p) h\big(\mathbf{L}\mathbf{x}\big), \tag{10}$$

which preserves the sparsity properties of the graph because the computation $h(\mathbf{L}\mathbf{x})$ can be carried out locally by each node, and the matrix $\mathbf{D}(\boldsymbol{\tau})$ is diagonal. Thus, the complete vector field F_A in (6) can be computed in a distributed way.

The construction of the mapping F_A in (6) is motivated in part by Nesterov's ODE studied in [9] and [18] for the solution of *centralized* optimization problems, where Ψ is usually taken as the gradient of the cost function. Previous applications of these dynamics to distributed optimization problems (with no restarting) relied on assuming a centralized scalar timer τ that grows unbounded and coordinates the overall system, see [18]. However, in general, these types of dynamics may not guarantee UGAS of the set $\mathcal{A}_\phi$ due to their use of vanishing damping terms [7, Ex. 1]. Moreover, when each agent has its own timer τ_i, and Ψ is taken as the gradient of the cost function, the dynamics (6) do not necessarily render forward invariant the set $\ker(\mathbf{L})^\perp$ for the states $\mathbf{x}$ or $\mathbf{y}$. This issue prevents the direct application of typical restarting mechanisms that persistently reset to zero the momentum state [7], [12], [14]. Furthermore, in multi-agent systems (MAS) the restarting mechanism must be implemented in a distributed but coordinated way, since otherwise instability may emerge, see Figure 3. Also, note that Ψ, defined in (10), is different from $\nabla\phi$ because in general the product $\mathbf{L}\mathbf{D}(\boldsymbol{\tau})$ does not commute.

Next, we describe how to construct the mapping $\mathcal{T}$ and the set D_τ that characterize the restarting rule in (8)-(9).

B. Distributed and Coordinated Restarting Rule

For each node $j \in \mathcal{V}$ we define a set-valued mapping $\mathcal{R}_j : \mathbb{R} \rightrightarrows \mathbb{R}$ that is non-empty on $[T_r, T_r + \Delta T]$, given by

$$\mathcal{R}_j(\tau_j) := \begin{cases} \{T_r{+}\Delta T\} & \text{if } \tau_j \in (r_j, T_r{+}\Delta T] \\ \{T_r, T_r{+}\Delta T\} & \text{if } \tau_j{=}r_j \\ \{T_r\} & \text{if } \tau_j \in [T_r, r_j) \end{cases}, \tag{11}$$

where $r_j \in (T_r, T_r + \Delta T/n)$ is a tunable parameter of each node. The restarting mechanism works as follows: Whenever a timer τ_i of some node $i \in \mathcal{V}$ satisfies $\tau_i = T_r + \Delta T$, the timer τ_i is reset back to T_r, and the state x_i is instantaneously reset as $x_i^+ = (1 - q)x_i + qy_i$. Simultaneously, neighbors of agent i update their timers following the rule $\tau_j^+ \in \mathcal{R}_j(\tau_j)$. This distributed reset rule can generate different sequences of updates whenever more than one timer satisfies the jump condition $\tau_i = T_r + \Delta T$. To capture every possible sequence of such resets, let $\mathbf{g} := [g_1, g_2, \ldots, g_n]^\top \in \mathbb{R}^n$ and $\mathbf{s} := [s_1, s_2, \ldots, s_n]^\top \in \mathbb{R}^n$, and consider the set-valued mapping

$$\begin{aligned} G^0(\boldsymbol{p}) := \Big\{ (\mathbf{s}^\top, \mathbf{g}^\top)^\top : s_i = y_i,\ s_j = x_j\ \forall\ j \neq i,\ g_i = T_r, \\ g_j \in \mathcal{R}_j(\tau_j) \text{ if } (i,j) \in \mathcal{E},\ g_j = \tau_j \text{ if } (i,j) \notin \mathcal{E} \Big\}, \end{aligned}$$

which is defined to be non-empty only when $\tau_i = T_r + \Delta T$ and $\tau_j \in [T_r, T_r + \Delta T)$ for all $j \neq i$. The jump rule $\mathcal{T}$ in (8) is then defined as the outer-semi-continuous (osc) hull [19, pp.155] of G^0, i.e., $\mathcal{T}(\boldsymbol{p}) := \overline{G^0(\boldsymbol{p})}$. This construction preserves the sparsity properties of the graph $\mathcal{G}$. Finally, we define the set D_τ that triggers the resets as follows:

$$D_\tau := \left\{ \boldsymbol{\tau} \in [T_r, T_r{+}\Delta T]^n : \max_{i \in \mathcal{V}}\ \tau_i{=}T_r{+}\Delta T \right\}. \tag{12}$$

Note that the structures of $\mathcal{T}$ and D_τ imply that whenever two or more timers reach the value $T_r + \Delta T$, their resets will occur *sequentially* rather than in parallel. This behavior is induced on purpose by using the osc hull of G^0 to generate $\mathcal{T}$. As explained in [20], this construction guarantees well-posedness of multi-agent coordinated HDS.

C. Main Result

We are now ready to present the main result of this paper, which establishes tuning guidelines for the HDS $\mathcal{H}_A = \{C_A, F_A, D_A, G_A\}$, expressed in terms of the parameters of the primal Problem (1). In particular, we consider the following conditions on the tunable parameters $(\gamma, r_i, T_r, \Delta T)$:

(C.1) $$0 < T_r < \frac{1}{2q}\sqrt{\frac{\bar{\mu}}{\gamma \lambda_{\max}(\mathcal{L}^2)}},$$

(C.2) $$\Delta T > \left((2\kappa_F \kappa_{\mathcal{L}^2})^{\frac{q}{2}} - 1\right) T_r,$$

(C.3) $$T_r < r_i < T_r + \frac{\Delta T}{n}.$$

The stability properties of the HARDD dynamics will be stated with respect to the compact set $\mathcal{A} := \mathcal{A}_{x,y} \times \mathcal{A}_\tau$, where the sets $\mathcal{A}_{x,y}$ and $\mathcal{A}_\tau$ are defined as follows:

$$\mathcal{A}_{x,y} := \{\mathbf{x}, \mathbf{y} \in \mathbb{R}^{np} : \mathbf{y} = \mathbf{x},\ \mathbf{x} \in \mathcal{A}_\phi\},$$
$$\mathcal{A}_\tau := [T_r, T_r + \Delta T] \cdot \mathbf{1}_n \cup \{T_r, T_r + \Delta T\}^n.$$

Note that the set $\mathcal{A}_\tau$ describes a "synchronization" condition. We will also use $F^* := F(\mathbf{z}^*)$ to denote the optimal value of Problem (3), where $\mathbf{z}^* := \mathbf{1}_n \otimes z^* \in \mathbb{R}^{np}$, and $z^* \in \mathbb{R}^p$ is the unique solution of the primal Problem (1).

Theorem 3.2

Suppose that Assumptions 3.1 and 3.2 hold, and let the parameters $(\gamma, r_i, T_r, \Delta T)$ satisfy **(C.1)**-**(C.3)** for all $i \in \mathcal{V}$. Then, the following properties hold:

- **(P.1)** Every solution $\mathbf{p}$ of $\mathcal{H}_A$ has an unbounded time domain and it is uniformly non-Zeno, i.e., there exists at most n jumps in any time interval of length $2\Delta T$.
- **(P.2)** Every solution $\mathbf{p}$ of $\mathcal{H}_A$ satisfiies $|\boldsymbol{\tau}(t,j)|_{\mathcal{A}_\tau} = 0$ for all $(t,j) \in \text{dom}(\mathbf{p})$ such that $t + j \geq n + 2\Delta T$.
- **(P.3)** The compact set $\mathcal{A}$ is UGAS.

Moreover, for each compact set of initial conditions $K_0 \subset C_A \cup D_A$ the following acceleration properties hold with respect to the primal variable $\mathbf{z} = h(\mathbf{L}\mathbf{x})$:

- **(P4)** If $q = 1$, there exists $c_0, \lambda_0 > 0$ such that $F(\mathbf{z}(t,j)) - F^* \leq c_0 e^{-\lambda_0 (t+j)}$, for all $(t,j) \in \text{dom}(\mathbf{p})$.
- **(P5)** If $q = 0$, $\boldsymbol{\tau}(0,0) \in \mathcal{A}_\tau$, and $\mathbf{x}(0,0) \in \ker(\mathbf{L})^\perp$, then $F(\mathbf{z}(t,j)) - F^* \leq \frac{c_j}{\tau_i(t,j)}$, for all $i \in \mathcal{V}$, and all $(t,j) \in \text{dom}(\mathbf{p})$, where $\{c_j\}_{j=0}^\infty$ is a monotonically decreasing sequence of positive numbers.

To the knowledge of the authors, Theorem 3.2 provides the first network-dependent restarting conditions in the literature of accelerated distributed optimization over networks. Conditions **(C.1)**-**(C.3)** show the dependence of the restarting value T_r and the restarting frequency ΔT on the condition numbers of the primal global cost function and the graph $\mathcal{G}$, as well as the gain γ and the number of agents n. Note that these conditions can always be satisfied by taking ΔT sufficiently large, and T_r sufficiently small. However, as stated in property **(P.2)**, the larger ΔT is selected, the longer it will take the network to synchronize the timers. Property **(P.5)**

recovers the result of [18] established with a centralized timer, and Property **(P.4)** establishes linear convergence when the gradient of the primal function is globally Lipschitz. Note that when $q = 0$, conditions (C.1) and (C.2) reduce to $0 < T_r < \infty$ and $\Delta T > 0$.

Since the HARDD algorithms are modeled by a well-posed HDS $\mathcal{H}_A$, property **(P.3)** will be preserved, in a semi-global practical way, under *arbitrarily* small (possibly time-varying) perturbations on the states and dynamics of the algorithm. This property, which follows by [17, Lem. 7.20], is fundamental for feedback control applications where measurement disturbances are unavoidable.

Corollary 3.3

Let $e : \mathbb{R}_{\geq 0} \to \mathbb{R}^{2np+n}$ be a measurable function satisfying $\sup_{t\geq 0} |e(t)| \leq \bar{e}$, with $\bar{e} > 0$. Then, under conditions **(C.1)**, **(C.2)** and **(C.3)**, the dynamics

$$\mathbf{p} + e \in C_A, \quad \dot{\mathbf{p}} \in F_A(\mathbf{p} + e) + e, \tag{13a}$$

$$\mathbf{p} + e \in D_A, \quad \mathbf{p}^+ \in G_A(\mathbf{p} + e) + e, \tag{13b}$$

render the set $\mathcal{A}$ semi-globally practically asymptotically stable as $\bar{e} \to 0^+$.

Remark 3.1

Consider the jump map

$$\mathbf{p}^+ \in G_A(\mathbf{p}) := \begin{pmatrix} \{\mathbf{x}\} \\ \{\mathbf{s}\} \\ \{\mathbf{g}\} \end{pmatrix}, \ (\mathbf{s}, \mathbf{g}) \in \tilde{\mathcal{T}}(\boldsymbol{p}) \Bigg), \tag{14}$$

where $\tilde{\mathcal{T}}(\mathbf{p}) \coloneqq \overline{G^1(\mathbf{p})}$ and

$$G^1(\mathbf{p}) := \Big\{(\mathbf{s}^\top, \mathbf{g}^\top)^\top : s_i = x_i, \ s_j = y_j \ \forall \ j \neq i, \ g_i = T_r, \\ g_j \in \mathcal{R}_j(\tau_j) \text{ if } (i,j) \in \mathcal{E}, \ g_j = \tau_j \text{ if } (i,j) \notin \mathcal{E}\Big\},$$

which is inspired by the jump rule used in [7] for the centralized optimization of strongly convex functions. Additionally, let $\tau(0,0) \in \mathcal{A}_\tau$, $\mathbf{x}(0,0) = \mathbf{0}$, $\gamma = \frac{\bar{\mu}}{\bar{L}\mu_\phi}$, $T_r \approx 0$ and $\Delta T = e\sqrt{\frac{\bar{L}}{\bar{\mu}}}$. Then, by using the jump map (14) instead of (8) in the HARDD algorithm, the convergence in time of the sub-optimality measure $F(\mathbf{z}) - F(\mathbf{z}^*)$ is of order $\mathcal{O}\left(\sqrt{\frac{\bar{L}}{\bar{\mu}}} \log\left(\sqrt{\frac{\bar{L}}{\bar{\mu}}} \frac{1}{\epsilon}\right)\right)$ for any precision $\epsilon > 0$. This is the optimal linear convergence rate of the Nesterov algorithm, with which we recover the result of [12] for the classic discrete-time Nesterov dynamics.

IV. DISCRETIZATION AND NUMERICAL EXAMPLES

In this section, we apply the HARDD algorithm with $q = 1$ to solve a distributed linear regression problem over a network. The nodes aim to solve the optimization problem: $\min_{\mathbf{z}\in\mathbb{R}^{np}} \frac{1}{2nl} \sum_{i=1}^{n} \|H_i z_i - b_i\|$, s.t. $\mathbf{L}\mathbf{z} = \mathbf{0}_{np}$, where $H_i \in \mathbb{R}^{l\times p}$ and $b_i \in \mathbb{R}^l$ contain the data, l is the number of data points available per node and p is the dimension of such data points. We implement the HARDD algorithm by discretizing the flows using a Runge-Kutta method of 4-th order (RK4) with step-size $dt = 1 \times 10^{-3}$.

As shown in [21], this discretization method preserves the main convergence properties of well-posed hybrid dynamics, provided the step size is sufficiently small. Indeed, as shown in [18], RK4 can also preserve

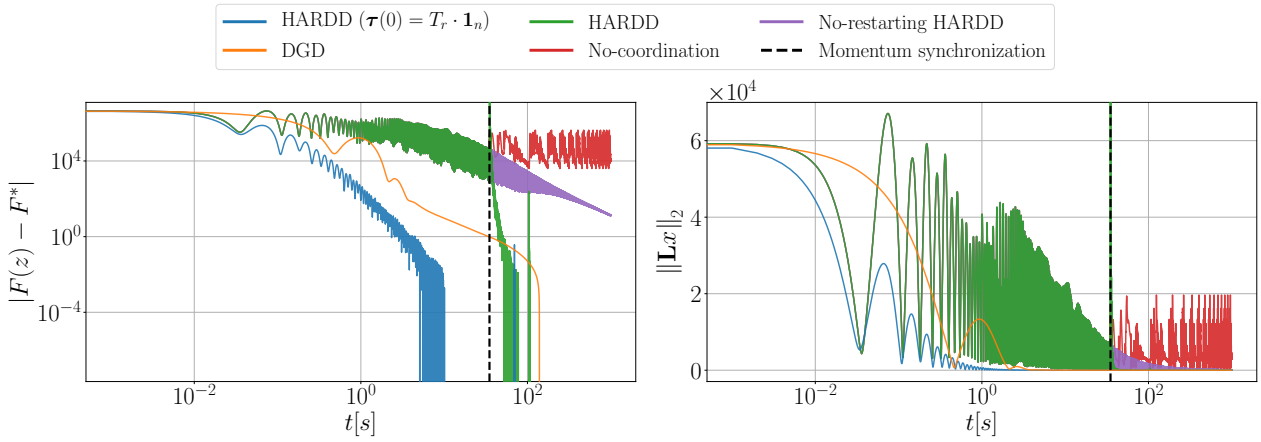

Fig. 2: (Left) Evolution in time of the sub-optimality measure, and (Right) evolution in time of the consensus distance for different algorithms and/or initializations.

acceleration. Figure 2 presents the numerical results with $\gamma = 1/4$, $\Delta T = 35$, $n = 5$, $p = 8$, and $l = 10$ on a ring graph. The data is generated by sampling from a normal distribution with mean 1.5 and standard deviation 3[1]. In the figure, the blue trajectory corresponds to the HARDD dynamics with synchronous initialization, i.e., $\tau_i = T_r = 0.1$ for all $i \in \mathcal{V}$. As expected, this initialization exhibits the best behavior. On the other hand, the green line indicates the behavior of the HARDD dynamics with asynchronous initialization of the timers. As shown in the plots, after the synchronization event $|\tau|_{\mathcal{A}_\tau} = 0$ occurs, the sub-optimality measure decreases rapidly. For the sake of comparison, we also show the solutions obtained without restarting mechanism (purple line), with individual restarting and no coordination between nodes (red line), and using the standard *synchronous* distributed gradient descent [3]. Figure 3 illustrates the importance of coordination in the restarting mechanism; in this case without coordination the dual variables do not converge. The purple trajectory (no restarting) is slow compared to the green trajectory, and arbitrarily small disturbances can destabilize this algorithm, as will be explored in Section IV-A. Finally, as shown in Figure 2, the red trajectories (restarting with no coordination) exhibit the worst behavior.

A. *Robustness*

In this section we compare the HARDD algorithm with and without restarting in order to stress the importance that this mechanism has in the robustness properties of the algorithm.

To do so, we first note that for the dynamics with no restarting there are no convergence guarantees. However, even if these dynamics do converge, they will suffer from the same limitation of Nesterov's ODE due to the unbounded growth of the timers τ_i, i.e., they will have zero margins of robustness with respect to arbitrarily small perturbations.

In order to explicitly show this behavior, we consider the consensus optimization problem with cost function

$$F(z) = \sum_{i=1}^{2} f_i(z_i) = \frac{1}{2} z^\top z, \qquad z \in \mathbb{R}^{2\times 2},$$

[1]The code used to generate the figures in this section can be found in https://github.com/deot95/HARDD

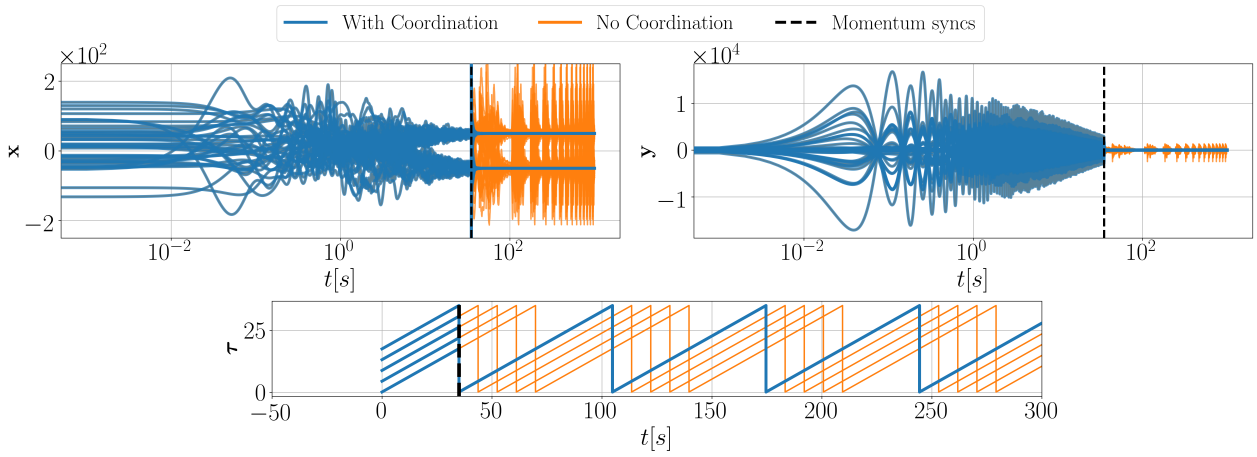

Fig. 3: Evolution in time of the dual-variables for the HARDD algorithm with and without coordination.

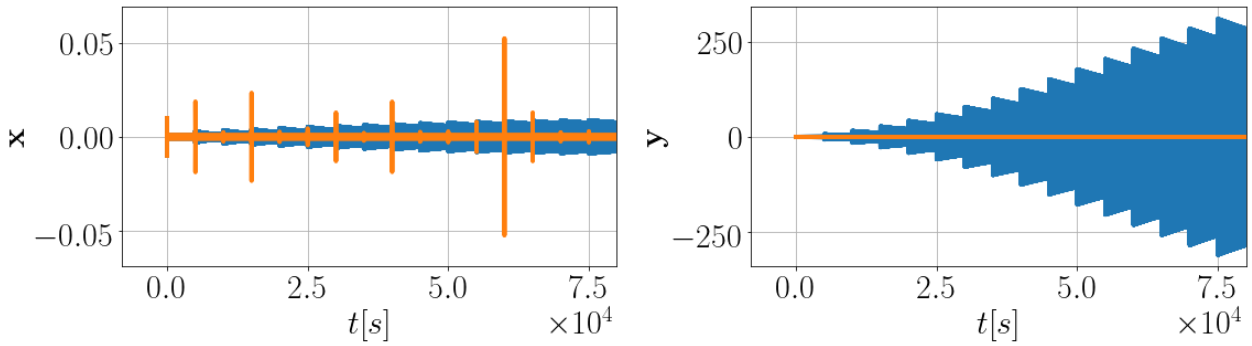

Fig. 4: (Blue) Instability of the optimization dynamics under disturbance $\varepsilon(t)$ when no restarting is implemented, i.e., $\Delta T = \infty$. As observed, the trajectories of the system diverge. (Orange) Robust asymptotic stability under the same disturbance $\varepsilon(t)$ with the HARDD algorithm and $\Delta T = 35$.

and $n = 2,\ p = 2$. For this case, the associated dual optimization problem is given by $\phi(x) = \frac{1}{2}x^\top \mathbf{L}^2 x$, where $\mathbf{L} \in \mathbb{R}^{4\times 4}$ is the communication matrix of the 2-agent system. With this in mind, we simulate the HARDD algorithm with ε-perturbed measurements of the state, i.e.,

$$\Psi(\tau, x) = \mathbf{LD}(\boldsymbol{\tau} \otimes \mathbf{1}_p)h\Big(\mathbf{L}\big(x + \varepsilon(t)\big)\Big), \tag{15}$$

where $\sup_{t\geq 0}|\varepsilon(t)| = 0.001$ and $\varepsilon(t) \in \ker(\mathbf{L})^\perp$ for all $t \geq 0$. In particular, we consider the following disturbance:

$$\varepsilon(t) = 0.001\eta(t)[1, 1, -1, -1]^\top, \tag{16}$$

where $\eta : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is a square periodic signal with unitary amplitude and period equal to 1×10^4. For the case when $\Delta T = \infty$ (the same setup that generates the purple line in Figure 2) this arbitrarily small disturbance induces the instability shown below in Figure 4-(a). On the other hand, when the restarting is activated, we obtain the robust stable behavior shown in Figure 4-(b). In the next section, we present the analysis of the hybrid algorithm.

V. Analysis

The convergence and stability analysis of the HDS $\mathcal{H}_A$ starts with the following lemma. The proof is almost identical to the proofs of [20, Prop. 1] and [22, Thm. 1], and it is presented here only for the sake of completness.

Lemma 5.1

Consider the hybrid dynamical system

$$\boldsymbol{\tau} \in [T_r,\ T_r + \Delta T]^n\,, \qquad \dot{\boldsymbol{\tau}} = \frac{1}{2}\mathbf{1}_n \tag{17a}$$

$$\boldsymbol{\tau} \in D_\tau, \qquad\qquad \boldsymbol{\tau}^+ \in \mathcal{T}(\boldsymbol{\tau}), \tag{17b}$$

where $\mathcal{T}$ is the osc hull of the set-valued mapping

$$G^0(\boldsymbol{\tau}) := \{\mathbf{g} : g_i = T_r, g_j \in \mathcal{R}_j(\tau_j) \text{ if } (i,j) \in \mathcal{E},\ g_j = \tau_j \text{ if } (i,j) \notin \mathcal{E}\}\,.$$

If r_i satisfies **(C3)**, then every solution is complete and uniformly non-Zeno, the set $\mathcal{A}_\tau$ is UGAS, and every solution satisfies $|\boldsymbol{\tau}(t,j)|_{\mathcal{A}_\tau} = 0$, for all $(t,j) \in \mathrm{dom}(\boldsymbol{\tau})$ such that $t + j \geq n + 2\Delta T$. □

Proof: Absence of finite escape times follows by compactness of the flow set and jump set. Being uniformly non-Zeno follows by the fact that after n jumps the system is necessarily synchronized and the timers satisfy $\boldsymbol{\tau} \in [T_r, T_r + \Delta T]^n \backslash D_\tau$, which implies that the system has to flow. Since the intervals of flow have a maximum duration of ΔT, it follows that there can be at most n consecutive jumps in any interval of length $2\Delta T$. This also implies completness of solutions. To show UGAS of $\mathcal{A}_\tau$ we define a Lyapunov function $V : [T_r, T_r + \Delta T]^n \to \mathbb{R}_{\geq 0}$ to be the the infimum of the lengths of all arcs that touch all timers (see Figure 5 for an illustration), where the points T_r and $T_r + \Delta T$ in the interval $[T_r, T_r + \Delta T]$ are identified to be the same to form a circle. Since all the timers have the same frequency, during the flows the Lyapunov function does not change, i.e., $\dot{V}(\boldsymbol{\tau}) = 0$. Moreover, during jumps the Lyapunov function cannot increase its value since jumps only happen whenever one or more timers satisfy the condition $\tau_i = T_r + \Delta T$, which either leaves the timers in the same position of the circle, or forces some of the timers to go to $T_r + \Delta T$. In both cases, $V(\boldsymbol{\tau}^+)$ does not increase. To show that V converges to zero in a fixed-time, we note that for any initial condition $\tau(0,0) \in [T_r, T_r + \Delta T]^n$ the Lyapunov function V always satisfies $V(\tau) \leq \Delta T\left(1 - \frac{1}{n}\right)$. Since all timers have the same frequency, and since $r_i \in (T_r, T_r + \frac{\Delta T}{n})$ for all i, there will exist a time $0 \leq t < 2\Delta T$ and some $j \in \{0, 1, \ldots, n\}$ such that $\tau_i(t,j) > r_i$ for all $i \in \mathcal{V}$. From this point, since the graph is connected and undirected, any jump induced by an agent j satisfying $\tau_j = T_r + \Delta T$ will be followed by at most $n-1$ jumps after which all timers will be synchronized at the position $\tau_i = T_r$, which implies $V(\boldsymbol{\tau}) = 0$. From this point, the system remains synchronized. Since no complete solution keeps V equal to a non-zero constant, UGAS of the set $\mathcal{A}_\tau$ follows now directly by the Hybrid Invariance Principle [17, Thm. 8.8]. ■

A. Proof of Theorem 3.2

We divide the proof in seven main steps. Refer to Figure 6 for an overview and visualization of the main aspects of the proof.

Step 1: Absence of Finite Escape Times

First, note that the function F_A is continuous in C_A. Also, by item (b) in Lemma 3.1, the gradient $\nabla\phi$ is

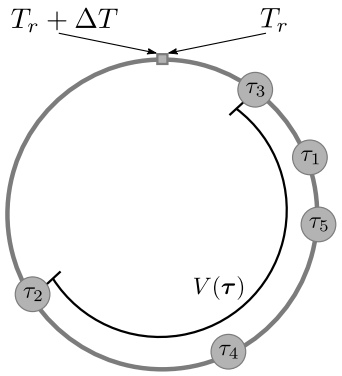

Fig. 5: Representation of the Lyapunov function used in proof of Lemma 5.1

globally Lipschitz. It then follows that since $T_r \leq \tau_i \leq T_r + \Delta T$ for all i, we have:

$$|\dot{\mathbf{y}}| \leq 2\gamma(T_r + \Delta T)|\mathbf{L}\mathbf{z}^*(\mathbf{L}\mathbf{x})| \leq 2\gamma(T_r + \Delta T)\ell_\phi|\mathbf{x} - \mathbf{s}^*|,$$
$$|\dot{\mathbf{x}}| \leq 2|\mathbf{D}(\boldsymbol{\tau} \otimes \mathbf{1}_p)^{-1}(\mathbf{y} - \mathbf{x})| \leq \frac{2}{T_r}|\mathbf{y} - \mathbf{x}|,$$

for any $\mathbf{s}^* \in \mathcal{A}_\phi$. Combining these inequalities, and the Comparison Lemma [23] we obtain that the flows (6) always generate bounded signals $(\mathbf{x}, \mathbf{y})$, which rules out finite escape times.

Step 2: Completeness of Solutions
We show that solutions cannot stop due to flows or jumps leaving $C_A \cup D_A$. By Lemma 5.1, the dynamics of the timers always generate complete solutions. On the other hand, by the properties of the Laplacian L, we have that $\mathbf{1}_{np}^\top \dot{\mathbf{y}} = 0$. Thus, the state $\mathbf{y}$ always remains in $\ker(\mathbf{L})^\perp$, which is unbounded. Since the state $\mathbf{x}$ evolves in $\mathbb{R}^{np}$, every solution of the HDS is complete and the hybrid time domains of the solutions are generated by the hybrid time domains of the HDS (17). This establishes properties **(P.1)** and **(P.2).**

Step 3: Fixed-Time Synchronization of Restarting Mechanisms
Let $k > 0$ and define the compact set $K := \mathcal{A}_\phi + k\mathbb{B}$. Let us restrict the data of the original HDS $\mathcal{H} := \{C_A, F_A, D_A, G_A\}$ by intersecting with K the $(\mathbf{x}, \mathbf{y})$-components of the flow set and the jump set. The resulting HDS has data $\mathcal{H}_K := \{C_{A,K}, F_A, D_{A,K}, G_A\}$, where

$$C_{A,K} := K \times (\ker(\mathbf{L})^\perp \cap K) \times [T_r, T_r + \Delta T]^n \tag{18a}$$
$$D_{A,K} := K \times (\ker(\mathbf{L})^\perp \cap K) \times D_\tau \tag{18b}$$

Since the dynamics of the timers $\boldsymbol{\tau}$ are independent of $\mathbf{x}$ and $\mathbf{y}$, by the definition of UGAS and by Lemma 5.1 the restricted HDS $\mathcal{H}_K$ renders UGAS the compact set

$$\mathcal{A}_{K,s} := K \times (\ker(\mathbf{L})^\perp \cap K) \times \mathcal{A}_\tau. \tag{19}$$

Step 4: Asymptotic Stability of Feasible Set

Let us now further restrict the flow and the jump sets of the HDS $\mathcal{H}_K$ with the set $\mathcal{A}_{K,s}$. We denote this new restricted HDS as $\mathcal{H}_{K,s} := \{C_{A,K,s}, F_A, D_{A,K,s}, G_A\}$, where

$$C_{A,K,s} = K \times (\ker(\mathbf{L})^\perp \cap K) \times ([T_r, T_r + \Delta T]^n \cap \mathcal{A}_\tau),$$
$$D_{A,K,s} = K \times (\ker(\mathbf{L})^\perp \cap K) \times (D_\tau \cap \mathcal{A}_\tau)$$

In this HDS, during flows the timers satisfy $\boldsymbol{\tau} = \alpha \mathbf{1}_n$ where $\alpha \in [T_r, T_r + \Delta T]$.

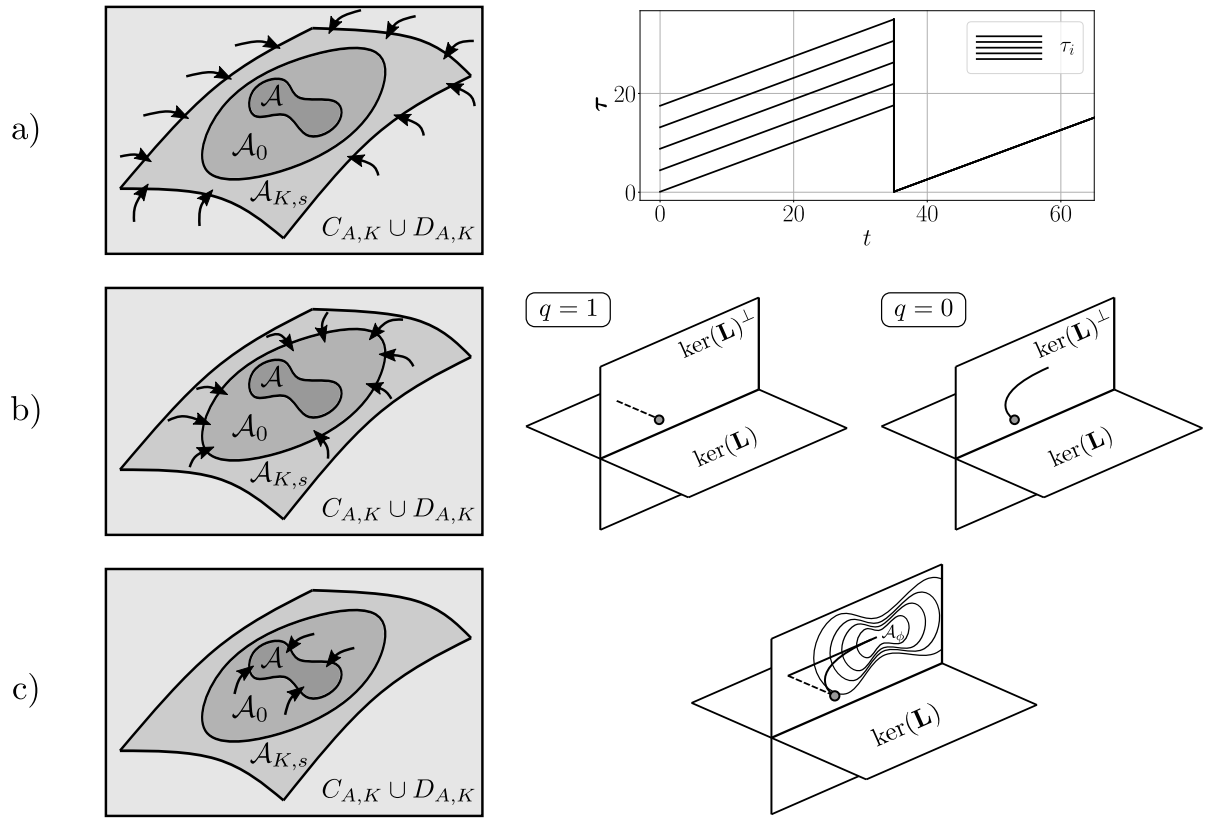

Fig. 6: In order to prove UGAS of the set $\mathcal{A}$: a) First, UGAS of the set $A_{K,s}$ from $C_{A,K} \cup D_{A,K}$ is proven by the fixed-time synchronization of the timers through the used restarting mechanism.
b) Second, and when $\nabla F(\mathbf{z})$ is globally Lipschitz ($q = 1$), the set $\ker(\mathbf{L})^\perp$ is rendered UGAS for $\mathbf{x}$ by the jump map $G_A(\mathbf{p})$ in (8). When $q = 0$, UGAS is achieved through the flow map (6). This fact, in conjunction with strong forward invariance of $\ker(\mathbf{L})^\perp$ and A_τ, for $\mathbf{y}$ and $\boldsymbol{\tau}$ respectively, render the set $\mathcal{A}_0$ UGAS from $A_{K,s}$.
c) Third, using Lyapunov analysis with the Lyapunov function in (20) and the formulated HDS with flow and jump sets (6),(8), UGAS of $\mathcal{A}$ from $\mathcal{A}_0$ is guaranteed.
Last, by the nested application of the Hybrid Reduction Principle [17, Cor. 7.24] UGAS of $\mathcal{A}$ is guaranteed.

Lemma 5.2

For the HDS $\mathcal{H}_{K,s}$ the compact set $\mathcal{A}_0 = (\ker(\mathbf{L})^\perp \cap K) \times (\ker(\mathbf{L})^\perp \cap K) \times \mathcal{A}_\tau$ is UGAS. Moreover, if $q = 0$, then every complete solution satisfies $\mathbf{x}(t, j) \in \ker(\mathbf{L})^\perp$ for all $(t, j) \in \mathrm{dom}(\mathbf{p})$ such that $t + j \geq 2n + 4\Delta T$.

Proof: We show that $\mathcal{A}_0$ is strongly forward invariant and globally uniformly attractive, which implies UGAS via [17, Prop. 7.5]. Let $\mathbf{p}(0, 0) \in \mathcal{A}_0$. By Lemma 5.1 the set $\mathcal{A}_\tau$ is strongly forward invariant under the dynamics of $\boldsymbol{\tau}$. By Step 2, the set $\ker(\mathbf{L})^\perp$ is strongly forward invariant for the state $\mathbf{y}$. Moreover, by definition of the flow set $C_{A,K,s}$, during flows $\boldsymbol{\tau}(t, j) = \alpha \mathbf{1}_n$ with $\alpha \in [T_r, T_r + \Delta T]$. Thus, $\mathbf{1}_{np}^\top \dot{\mathbf{x}} = 0$ which implies that $\mathbf{1}_{np}^\top \mathbf{x}(t, 0) = 0$ for all $(t, 0) \in \mathrm{dom}(\mathbf{p})$ since $\mathbf{x}(0, 0) \in \ker(\mathbf{L})^\perp$. Because $\mathbf{x}^+ = \mathbf{x}$ $(q = 0)$ or $\mathbf{x}^+ = \mathbf{y}$ $(q = 1)$,

the state $\mathbf{x}$ remains in $\ker(\mathbf{L})^\perp$ also during jumps. To show that $\mathcal{A}_0$ is uniformly attractive it suffices to show that every complete trajectory of $\mathbf{x}$ converges to $\ker(\mathbf{L})^\perp$. Let $q = 0$ and consider the auxiliary variable $\tilde{\mathbf{x}} = \mathbf{1}_{np}^\top \mathbf{x}$, with dynamics $\dot{\tilde{\mathbf{x}}} = 2\mathbf{1}_{np}^\top \mathbf{D}(\boldsymbol{\tau} \otimes \mathbf{1}_p)^{-1}(\mathbf{y} - \mathbf{x}) = \frac{2}{\alpha}\mathbf{1}_{np}^\top(\mathbf{y} - \mathbf{x})$, where we used again the fact that $\boldsymbol{\tau} = \alpha \mathbf{1}_n$ during flows. Since $\mathbf{1}_{np}^\top \mathbf{y} = 0$ always hold, we obtain $\dot{\tilde{\mathbf{x}}} = -2\frac{1}{\alpha}\tilde{\mathbf{x}}$, which implies that $\tilde{\mathbf{x}}$ converges exponentially fast to zero, i.e., $\mathbf{x}(t, j)$ converges to $\ker(\mathbf{L})^\perp \cap K$ exponentially fast during flows. When $q = 1$, convergence of $\mathbf{x}$ to $\ker(\mathbf{L})^\perp$ happens in finite time after the n consecutive jumps induced by the synchronized timers $\boldsymbol{\tau}$, and the fact that $x_i = y_i$ after each jump. ■

Step 5: Asymptotic Stability of the Optimal Set

Having established UGAS of $\mathcal{A}_0$ for the HDS $\mathcal{H}_{K,s}$, we now proceed to further restrict the data of $\mathcal{H}_{K,s}$ using the set $\mathcal{A}_0$. In particular, we consider a HDS $\mathcal{H}_{K,s,0} := \{C_{A,K,s,0}, F_A, D_{A,K,s,0}, G_A\}$ with flow and jump sets

$$
\begin{aligned}
C_{A,K,s,0} &:= (\ker(\mathbf{L})^\perp \cap K) \times (\ker(\mathbf{L})^\perp \cap K) \times \big([T_r, T_r + \Delta T]^n \cap \mathcal{A}_\tau\big), \\
D_{A,K,s,0} &:= (\ker(\mathbf{L})^\perp \cap K) \times (\ker(\mathbf{L})^\perp \cap K) \times \big(D_\tau \cap \mathcal{A}_\tau\big).
\end{aligned}
$$

For this restricted HDS, the following two lemmas establish UGAS of the set $\mathcal{A}$, and suitable acceleration properties.

Lemma 5.3

Let $q = 0$. Then, the HDS $\mathcal{H}_{K,s,0}$ renders UGAS the set $\mathcal{A}$, and the dual function is minimized at a rate of $\mathcal{O}(1/\tau^2)$ during flows.

Proof: Consider the Lyapunov function

$$
V(\mathbf{p}) := \frac{1}{4}|\mathbf{x} - \mathbf{y}|^2 + \frac{1}{4}|\mathbf{y}|^2_{\mathcal{A}_\phi} + \gamma \frac{\boldsymbol{\tau}^\top \boldsymbol{\tau}}{n}(\phi(\mathbf{x}) - \phi^*) \tag{20}
$$

where $\phi^* = \phi(\mathcal{A}_\phi)$. This is a modified version of the "centralized" Lyapunov function considered in [7]. By construction, this function is positive definite with respect to the compact set $\mathcal{A}$ in $C_{A,K,s,0} \cup D_{A,K,s,0}$, and also radially unbounded due to Assumption 3.2. Since during flows $\boldsymbol{\tau}(t) = \alpha(t)\mathbf{1}_n$, with $\dot{\alpha}(t) = 0.5$, the gradient of V satisfies

$$
\begin{aligned}
\nabla V &= \begin{pmatrix} \frac{1}{2}(\mathbf{x} - \mathbf{y}) + \gamma \frac{\boldsymbol{\tau}^\top \boldsymbol{\tau}}{n} \nabla \phi(\mathbf{x}) \\ \frac{1}{2}(\mathbf{y} - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{y}^*) \\ \gamma \frac{2\boldsymbol{\tau}}{n}(\phi(\mathbf{x}) - \phi^*) \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{2}(\mathbf{x} - \mathbf{y}) + \gamma \alpha^2 \nabla \phi(\mathbf{x}) \\ \mathbf{y} - \frac{1}{2}(\mathbf{x} + \mathbf{y}^*) \\ 2\gamma \frac{\alpha}{n}(\phi(\mathbf{x}) - \phi^*)\mathbf{1}_n \end{pmatrix}
\end{aligned} \tag{21}
$$

where $\mathbf{y}^*$ is the Euclidean projection of $\mathbf{y}$ in $\mathcal{A}_\phi$. Therefore, the derivative of V along the trajectories of the system satisfies

$$
\begin{aligned}
\dot{V} &= \langle \nabla V, F_A \rangle \\
&= -\frac{|\mathbf{y} - \mathbf{x}|^2}{\alpha} + 2\gamma\alpha(\mathbf{y} - \mathbf{x})^\top \nabla \phi(\mathbf{x}) - 2\gamma\alpha \mathbf{y}^\top \nabla \phi(\mathbf{x}) \\
&\qquad + \gamma\alpha(\mathbf{x} + \mathbf{y}^*)^\top \nabla \phi(\mathbf{x}) + \gamma\alpha(\phi(\mathbf{x}) - \phi^*) \\
&= -\frac{|\mathbf{y} - \mathbf{x}|^2}{\alpha} - \gamma\alpha \Big((\mathbf{x} - \mathbf{y}^*)^\top \nabla \phi(\mathbf{x}) - (\phi(\mathbf{x}) - \phi^*) \Big).
\end{aligned} \tag{22}
$$

Since, by assumption ϕ is convex, we have that $(\mathbf{x}-\mathbf{y}^*)^\top\nabla\phi(\mathbf{x})-(\phi(\mathbf{x})-\phi^*)\geq 0$. Moreover, due to the fact that $\alpha\in[T_r,T_r+\Delta T]$, and by Assumption 3.2 from (22) we obtain that $\dot{V}<0$ for all $\mathbf{p}\in C_{A,K,s,0}\backslash\mathcal{A}$. On the other hand, jumps occur whenever $\tau_1=\tau_2=\ldots=\tau_n=T_r+\Delta T$. This condition will trigger n consecutive jumps, after which the system will flow again. Thus, we evaluate the Lyapunov function after the n jumps and obtain:

$$V(\mathbf{p}^{+^n})=\frac{1}{4}|\mathbf{x}-\mathbf{y}|^2+\frac{1}{4}|\mathbf{y}|^2_{\mathcal{A}_\phi}+\gamma T_r^2(\phi(\mathbf{x})-\phi^*).$$

Therefore, the difference $\Delta(\mathbf{p}^{+^n})V:=V(\mathbf{p}^{+^n})-V(\mathbf{p})$ satisfies

$$\begin{aligned}\Delta V(\mathbf{p}^{+^n})&=\frac{1}{4}|\mathbf{x}-\mathbf{y}|^2+\frac{1}{4}|\mathbf{y}|^2_{\mathcal{A}_\phi}+\gamma T_r^2(\phi(\mathbf{x})-\phi^*)\\&\qquad-\frac{1}{4}|\mathbf{x}-\mathbf{y}|^2-\frac{1}{4}|\mathbf{y}|^2_{\mathcal{A}_\phi}-\gamma(T_r+\Delta T)^2(\phi(\mathbf{x})-\phi^*)\\&=-\gamma(\phi(\mathbf{x})-\phi^*)\left(\Delta T^2+2T_r\Delta T\right)\\&\leq 0,\end{aligned}$$

for all $\mathbf{p}\in D_{A,K,s,0}$. Therefore, the Lyapunov function V has strict decrease during flows and does not increase during jumps. Since at most n jumps can happen between each interval of flow of duration $2\Delta T$, UGAS follows directly by the invariance principle [17, Thm. 8.8].

To obtain the convergence bound for the dual function, note that $\dot{V}\leq 0$ implies $V(\mathbf{p}(t,j))\leq V(\mathbf{p}(s,j))$ for all $(t,j),(s,j)\in\text{dom}(\mathbf{p})$ with $t\geq s$. In turn, this inequality implies that

$$\phi(\mathbf{x})-\phi^*\leq\frac{nV(s_j,j)}{\gamma\boldsymbol{\tau}^\top\boldsymbol{\tau}}=\frac{c_j}{\tau_i^2},\tag{23}$$

during flows, where $s_j=\inf\{t\geq 0:(t,j)\in\text{dom}(\mathbf{p})\}$, and $c_j:=V(s_j,j)/\gamma$. ■

Lemma 5.4

Let $q=1$. Then, the HDS $\mathcal{H}_{K,s,0}$ renders uniformly globally exponentially stable (UGES) the set $\mathcal{A}$, and ϕ is minimized at an exponential rate.

Proof: By Lemma 3.1, the function ϕ is μ_ϕ-strongly convex in $\ker(\mathbf{L})^\perp$, and has a globally ℓ_ϕ-Lipschitz gradient. Thus, $\mathcal{A}_\phi=\{\mathbf{x}^*\}$. We consider again the same Lyapunov function $V(\mathbf{p})$, which now satisfies

$$\underline{c}|\mathbf{p}|^2_{\mathcal{A}}\leq V(\mathbf{p})\leq\overline{c}|\mathbf{p}|^2_{\mathcal{A}}$$

with $\underline{c}:=0.25\min\{1,2\gamma {T_r}^2\mu_\phi\}$ and $\overline{c}:=0.25\max\{3,6\gamma(T_r+\Delta T)^2\ell_\phi\}$. Since the continuous-time dynamics are the same of Lemma 5.3, we still have $\dot{V}\leq 0$. However, using strong convexity we have

$$(\phi(\mathbf{x})-\phi^*)-(\mathbf{x}-\mathbf{x}^*)^\top\nabla\phi(\mathbf{x})\leq-\frac{\mu_\phi}{2}|\mathbf{x}-\mathbf{x}^*|^2$$

to further obtain

$$\begin{aligned}\dot{V} &\leq -\frac{|\mathbf{y}-\mathbf{x}|^2}{\alpha} - \gamma\mu_\phi\frac{\alpha}{2}|\mathbf{x}-\mathbf{x}^*|^2 \\ &\leq -\frac{|\mathbf{y}-\mathbf{x}|^2}{T_r+\Delta T} - \gamma\mu_\phi\frac{T_r}{2}|\mathbf{x}-\mathbf{x}^*|^2 \\ &\leq \min\left\{\frac{1}{T_r+\Delta T},\ \frac{\gamma T_r\mu_\phi}{4}\right\}\left(-|\mathbf{y}-\mathbf{x}|^2 - |\mathbf{x}-\mathbf{x}^*|^2\right) - \gamma\mu_\phi\frac{T_r}{4}|\mathbf{x}-\mathbf{x}^*| \\ &\leq -\frac{1}{2}\min\left\{\frac{1}{T_r+\Delta T},\ \frac{\gamma T_r\mu_\phi}{4}\right\}\left(|\mathbf{y}-\mathbf{x}^*|^2 + |\mathbf{x}-\mathbf{x}^*|^2\right) \\ &\leq -\frac{\rho}{\overline{c}}V(\mathbf{p})\end{aligned}$$

for all $\mathbf{p} \in C_{A,K,s,0}$, where $\rho := 0.5\min\{\frac{1}{T_r+\Delta T}, 0.25\gamma T_r\mu_\phi\}$. On the other hand, after the n consecutive jumps triggered by the condition $\tau_1 = \tau_2 = \ldots = \tau_n = T_r + \Delta T$, the change in the Lyapunov function is

$$\begin{aligned}\Delta V(\mathbf{p}^{+^n}) = \frac{1}{4}\left|\mathbf{x}^{+^n} - \mathbf{y}^{+^n}\right|^2 + \frac{1}{4}\left|\mathbf{y}^{+^n} - \mathbf{x}^*\right|^2 + \gamma T_r^2\left(\phi\left(\mathbf{x}^{+^n}\right) - \phi^*\right) \\ - \frac{1}{4}|\mathbf{x}-\mathbf{y}|^2 - \frac{1}{4}|\mathbf{y}-\mathbf{x}^*|^2 - \gamma(T_r+\Delta T)^2(\phi(\mathbf{x}) - \phi^*) \\ = \gamma {T_r}^2(\phi(\mathbf{y}) - \phi^*) - \frac{1}{4}|\mathbf{x}-\mathbf{y}|^2 - \gamma(T_r+\Delta T)^2(\phi(\mathbf{x}) - \phi^*).\end{aligned}$$

Since ϕ is strongly convex in $\ker(\mathbf{L})^\perp$, and $\nabla\phi$ is globally Lipschitz, we have that

$$\phi(\mathbf{y}) - \phi^* \leq \frac{1}{2}\ell_\phi|\mathbf{y}-\mathbf{x}^*|^2,\ \forall\ \mathbf{y} \in \ker(\mathbf{L})^\perp, \tag{24}$$

and

$$\phi(\mathbf{x}) - \phi^* \geq \frac{1}{2}\mu_\phi|\mathbf{x}-\mathbf{x}^*|^2,\ \forall\ \mathbf{x} \in \ker(\mathbf{L})^\perp. \tag{25}$$

Using these two inequalities to further upper bound $\Delta V(\mathbf{p}^{+^n})$, we obtain:

$$\begin{aligned}\Delta V(\mathbf{p}^{+^n}) &\leq \gamma\ell_\phi\frac{T_r^2}{2}|\mathbf{y}-\mathbf{x}^*|^2 - \frac{1}{4}|\mathbf{x}-\mathbf{y}|^2 - \gamma\mu_\phi\frac{(T_r+\Delta T)^2}{2}|\mathbf{x}-\mathbf{x}^*|^2 \\ &\leq -T_1|\mathbf{x}-\mathbf{y}|^2 - T_2|\mathbf{x}-\mathbf{x}^*|^2 \\ &\leq -\underline{T}(|\mathbf{x}-\mathbf{y}|^2 + |\mathbf{x}-\mathbf{x}^*|^2),\end{aligned}$$

where $T_1 := \frac{1}{4} - \gamma\ell_\phi T_r^2$, $T_2 := \frac{1}{2}\gamma\mu_\phi(T_r+\Delta T)^2 - \gamma\ell_\phi T_r^2$ and $\underline{T} = \min\{T_1, T_2\}$. The constants T_1 and T_2 are positive provided the following holds: $0 < T_r < 1/(2\sqrt{\gamma\ell_\phi})$ and $(T_r+\Delta T)/T_r > \sqrt{2\kappa_\phi}$, which are precisely conditions **(C.1)** and **(C.2)**.

Finally, since V can be upper bounded as

$$\begin{aligned}V(\mathbf{p}) &\leq \overline{c}\left(|\mathbf{x}-\mathbf{x}^*|^2 + |\mathbf{y}-\mathbf{x}^*|^2\right) \\ &\leq \overline{c}\left(|\mathbf{x}-\mathbf{x}^*|^2 + 2|\mathbf{y}-\mathbf{x}|^2 + 2|\mathbf{x}-\mathbf{x}^*|^2\right) \\ &\leq 3\overline{c}\left(|\mathbf{x}-\mathbf{x}^*|^2 + |\mathbf{y}-\mathbf{x}|^2\right),\end{aligned}$$

we obtain

$$\Delta V(\mathbf{p}^{+^n}) \leq -\beta V(\mathbf{p})$$

where $\beta := \underline{T}/3\overline{c}$. These bounds, and the fact that the system is uniformly non-Zeno with n consecutive jumps followed by a constant interval of flow, establish UGES of $\mathcal{A}$ via [24, Thm. 1].

Step 6: Nested Application of the Reduction Principle
We now repeatedly apply the Hybrid Reduction Principle [17, Cor. 7.24] to establish UGAS for the original hybrid system $\mathcal{H}_A$. First, since the set $\mathcal{A}$ is UGAS for the HDS $\mathcal{H}_{K,s,0}$, and the set $\mathcal{A}_0$ is UGAS for the HDS $\mathcal{H}_{K,s}$, by the reduction principle, we obtain that the set $\mathcal{A}$ is UGAS for the HDS $\mathcal{H}_{K,s}$. Moreover, since the compact set $A_{K,s}$ is UGAS for the HDS $\mathcal{H}_K$, it follows again by the reduction principle that $\mathcal{H}_K$ renders UGAS the set $\mathcal{A}$. Finally, since by Step 1, there are no finite escape times in the original HDS $\mathcal{H}_A$, and since the compact set K was arbitrary, we have that the set $\mathcal{A}$ is indeed UGAS for the HDS $\mathcal{H}_A$ with no restriction. This establishes the stability result of property **(P.3)**. See Figure 6 for an illustration of the nested application of the reduction principle to guarantee asympotic stability of the set $\mathcal{A}$.

Step 7: Optimal Bounds for the Primal. Finally, we derive convergence bounds for the primal problem (1) based on the convergence bounds derived for the dual problem in Steps 1-6.

Let $q = 1$. Since the gradient of ϕ is globally Lipschitz, we have

$$|\nabla\phi(\mathbf{x})|^2 \leq \frac{\lambda_{\max}(\mathcal{L}^2)}{\overline{\mu}}(\phi(\mathbf{x}) - \phi^*) \tag{26}$$

Using (5),

$$|\mathbf{L}\mathbf{z}|^2 \leq \frac{\lambda_{\max}(\mathcal{L}^2)}{\overline{\mu}}(\phi(\mathbf{x}) - \phi^*) \tag{27}$$

Let us decompose $\mathbf{z} = \bar{\mathbf{z}} + \tilde{\mathbf{z}}$, where $\bar{\mathbf{z}} \in \ker(\mathbf{L})$ and $\tilde{\mathbf{z}} \in \ker(\mathbf{L})^\perp$. We then have

$$|\mathbf{L}\tilde{\mathbf{z}}|^2 \leq \frac{\lambda_{\max}(\mathcal{L}^2)}{\overline{\mu}}(\phi(\mathbf{x}) - \phi^*) \tag{28}$$

and since $\tilde{\mathbf{z}} \in \ker(\mathbf{L})^\perp$, which implies that $|\mathbf{L}\tilde{\mathbf{z}}|^2 \geq \lambda^+_{\min}(\mathcal{L}^2)|\tilde{z}|^2 s$, we get

$$|\tilde{\mathbf{z}}|^2 \leq \frac{\lambda_{\max}(\mathcal{L}^2)}{\overline{\mu}\lambda^+_{\min}(\mathcal{L}^2)}(\phi(\mathbf{x}) - \phi^*). \tag{29}$$

Using (5) and the definition of $\mathbf{z} = h(\cdot)$ in we get:

$$\langle \mathbf{L}\mathbf{x}, \mathbf{z}\rangle - F(\mathbf{z}) \geq \langle \mathbf{L}\mathbf{x}, \mathbf{z}^*\rangle - F(\mathbf{z}^*), \tag{30}$$

where $\mathbf{z}^*$ is the minimizer of the primal problem. Now, using the fact that $\mathbf{L}\mathbf{z}^* = 0$ we obtain:

$$\begin{aligned} F(\mathbf{z}) &\leq F(\mathbf{z}^*) + \langle \mathbf{L}\mathbf{x},\, \mathbf{z} - \mathbf{z}^*\rangle \\ &= F(\mathbf{z}^*) + \langle \mathbf{x},\, \mathbf{L}\mathbf{z}\rangle \\ &= F(\mathbf{z}^*) + \langle \mathbf{x},\, \mathbf{L}\tilde{\mathbf{z}}\rangle \\ &= F(\mathbf{z}^*) + \langle \mathbf{L}\mathbf{x},\, \tilde{\mathbf{z}}\rangle \\ &\leq F(\mathbf{z}^*) + |\langle \nabla F(\mathbf{z}), \tilde{\mathbf{z}}\rangle|, \end{aligned}$$

where in the last step we have used that $\mathbf{L}\mathbf{x} = \nabla F(\mathbf{z})$ by KKT conditions. Moreover, by the Cauchy-Schwartz inequality we get:

$$F(\mathbf{z}) - F(\mathbf{z}^*) \leq |\nabla F(\mathbf{z})||\tilde{\mathbf{z}}|. \tag{31}$$

Combining (29) and (31), we get

$$F(\mathbf{z}) - F(\mathbf{z}^*) \leq |\nabla F(\mathbf{z})|\sqrt{\frac{\lambda_{\max}(\mathcal{L}^2)}{\overline{\mu}\lambda^+_{\min}(\mathcal{L}^2)}(\phi(\mathbf{x}) - \phi^*)}. \tag{32}$$

By UGAS, for each compact set of initial condition K_0 there exists $M > 0$ such that $|\nabla F(\mathbf{z})| < M$. By exponential stability [24] and property (24) we obtain:

$$\phi(\mathbf{x}) - \phi^* \leq \frac{1}{2}\ell_\phi c^2 |\mathbf{x}(0,0) - \mathbf{x}^*|^2 \exp(-2\lambda(t+j)), \tag{33}$$

where $c^2 \coloneqq \frac{\overline{c}}{\underline{c}}$ and $\lambda = \frac{\rho}{c}$. Combining this with (32) we finally obtain for all initial conditions $\mathbf{x}(0,0) \in K_0$ that:

$$F(\mathbf{z}) - F(\mathbf{z}^*) \leq cM\sqrt{\frac{\lambda_{\max}(\mathcal{L}^2)}{2\overline{\mu}\lambda^+_{\min}(\mathcal{L}^2)}}|\mathbf{x}(0,0) - \mathbf{x}^*| \exp(-\lambda(t+j)). \tag{34}$$

When $q = 0$ we can follow exactly the same steps, using the bound (23) instead of using (24). ■

B. Proof of Remark 3.1

Suppose that $\mathbf{x}_0 = 0$ and that $\tau_0 \in \mathcal{A}_\tau$. Then, we have that

$$\begin{aligned}
\mathbf{z}_0 &= \arg\max_{\mathbf{z}} \{\langle \mathbf{L}\mathbf{x}_0,\, \mathbf{z}\rangle - F(\mathbf{z})\} \\
&= \arg\max_{\mathbf{z}} \{-F(\mathbf{z})\} \\
&= \arg\min_{\mathbf{z}} F(\mathbf{z}) \\
&\implies F(0,0) = \min_{\mathbf{z}} F(\mathbf{z}).
\end{aligned}$$

Moreover, from the definition of ϕ, we have that $\phi(\mathbf{x}(0,0)) = -\min_{\mathbf{z}} F(\mathbf{z}) = -F(\mathbf{z}(0,0))$. Similarly, we obtain that $\phi^* = -F(\mathbf{z}^*) = -\min_{\mathbf{L}\mathbf{z}=0} F(\mathbf{z})$, which implies that

$$\phi(\mathbf{x}(0,0)) - \phi^* = F(\mathbf{z}^*) - F(\mathbf{z}(0,0)). \tag{35}$$

Now, using the jump map (14), and following the analysis done in [7] for the centralized optimization case we can obtain that

$$\phi\left(\mathbf{x}(t,j)\right) - \phi^* \leq k_1^j \left(\phi\left(\mathbf{x}(0,0)\right) - \phi^*\right), \tag{36}$$

where $k_1 \coloneqq \frac{(\gamma\mu_\phi)^{-1} + T_r^2}{\Delta T^2}$. Using (32) and (35) in (36), in addition to the fact that the solutions of the HDS remain bounded so that $|F(\mathbf{z})| < M_0$, with M_0 being a constant that depends on initial conditions, we obtain that

$$\begin{aligned}
F(\mathbf{z}(t,j)) - F(\mathbf{z}^*) &\leq k_1^{\frac{j}{2}} |\nabla F(\mathbf{z})| \sqrt{\frac{\lambda_{\max}(\mathcal{L}^2)}{\overline{\mu}\lambda^+_{\min}(\mathcal{L}^2)}(F(\mathbf{z}^*) - F(0,0))} \\
&\leq k_1^{\frac{j}{2}} |\nabla F(\mathbf{z})| \sqrt{\frac{\overline{L}}{2\overline{\mu}} \frac{\lambda_{\max}(\mathcal{L}^2)}{\lambda^+_{\min}(\mathcal{L}^2)}}|\mathbf{z}^* - \mathbf{z}(0,0)| \\
&\leq k_1^{\frac{j}{2}} M_0 \sqrt{\frac{\overline{L}}{2\overline{\mu}} \frac{\lambda_{\max}(\mathcal{L}^2)}{\lambda^+_{\min}(\mathcal{L}^2)}}|\mathbf{z}^* - \mathbf{z}(0,0)| \\
&\leq \alpha k_1^{\frac{j}{2}} |\mathbf{z}^* - \mathbf{z}(0,0)|
\end{aligned} \tag{37}$$

where $\alpha \coloneqq M_0\sqrt{\frac{\overline{L}}{2\overline{\mu}} \frac{\lambda_{\max}(\mathcal{L}^2)}{\lambda^+_{\min}(\mathcal{L}^2)}}$, and we have used the fact that $F(\mathbf{z}^*) - F(\mathbf{z}(0,0)) \leq \frac{\overline{L}}{2}|\mathbf{z}^* - \mathbf{z}(0,0)|^2$ due to strong convexity and global Lipschitz continuity of $\nabla F(\mathbf{z})$. By the periodicity of the solutions, and letting

$j = t/\Delta T$, we take the derivative of $k_1^j = k_1^{\frac{t}{\Delta T}}$ with respect to ΔT and equate to zero to obtain the optimal restarting rate

$$\Delta T = e\sqrt{\frac{1}{\gamma\mu_\phi} + T_r^2}. \tag{38}$$

By replacing (38) in (37) we obtain that

$$F(\mathbf{z}(t,j)) - F(\mathbf{z}^*) \leq e^{-j}\alpha|\mathbf{z}^* - \mathbf{z}(0,0)|. \tag{39}$$

For an arbitrary precision $\epsilon > 0$ we thus obtain that $F(\mathbf{z}(t,j)) - F(\mathbf{z}^*) < \epsilon$ whenever $j \geq \log\left(\frac{\alpha|\mathbf{z}^* - \mathbf{z}(0,0)|}{\epsilon}\right)$. By multiplying both sides by the ΔT of (38), we obtain that the precision is achieved for

$$t \geq e\sqrt{\frac{1}{\gamma\mu_\phi} + T_r^2}\log\left(\frac{\alpha|\mathbf{z}^* - \mathbf{z}(0,0)|}{\epsilon}\right). \tag{40}$$

Hence, letting $T_r \approx 0$ and $\gamma = \frac{\overline{\mu}}{\overline{L}\mu_\phi}$ the convergence time (40) reduces to

$$\begin{aligned} t &\geq e\sqrt{\frac{\overline{L}}{\overline{\mu}}}\log\left(\frac{\alpha|\mathbf{z}^* - \mathbf{z}(0,0)|}{\epsilon}\right) \\ &= e\sqrt{\frac{\overline{L}}{\overline{\mu}}}\log\left(M_0\sqrt{\frac{\overline{L}}{2\overline{\mu}}\frac{\lambda_{\max}(\mathcal{L}^2)}{\lambda_{\min}^+(\mathcal{L}^2)}}\frac{|\mathbf{z}^* - \mathbf{z}(0,0)|}{\epsilon}\right), \end{aligned}$$

which obtains the result.

VI. Conclusions

We introduced a hybrid accelerated optimization algorithm with distributed restarting. Unlike previous results in the literature, each node implements its momentum state with dynamic coefficients using a local restarting mechanism that shares information with neighboring agents. By using Lyapunov tools, and in order to guarantee suitable stability and robustness properties, sufficient periodic restarting conditions in terms of the parameters of the network and the primal cost function were established for the closed-loop system. Future research directions will study adaptive distributed restarting techniques and zero-order distributed algorithms. Moreover, we are currently studying the discretization of the flows prescribed for the algorithm using Runge-Kutta techniques, which are known to preserved acceleration in discrete-time under appropriate choice of the step size.

References

[1] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[2] D. S. Acharya and N. Nath, "Applications of multi agent systems in control engineering: A state of the art survey," *International Journal of Innovative Research in Advanced Engineering*, vol. 2, no. 4, pp. 113–128, 2015.

[3] B. Gharesifard and J. Cortes, "Distributed continuous-time convex optimization on weigh-balanced digraphs," *IEEE Transactions on Automatic Control*, vol. 59, pp. 781–786, 2014.

[4] A. Mokhtari, Q. Ling, and A. Ribeiro, "Network newton distributed optimization methods," *IEEE Transactions on Signal Processing*, vol. 65, pp. 146–161, 2017.

[5] J. Cortes and S. K. Niederlander, "Distributed coordination for nonsmooth convex optimization via saddle-point dynamics," *Journal of Nonlinear Science*, vol. 29, pp. 1247–1272, 2019.

[6] A. Nedic, A. Ozdaglar, and P. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Trans. Autom. Contr.*, vol. 55, pp. 922–938, 2010.

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

[8] J. I. Poveda and A. R. Teel, "The heavy-ball ode with time-varying damping: Persistence of excitation and uniform asymptotic stability," in *2020 American Control Conference (ACC)*, pp. 773–778, IEEE, 2020.

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

[10] A. C. Wilson, B. Recht, and M. I. Jordan, "A lyapunov analysis of momentum methods in optimization," *arXiv preprint arXiv:1611.02635v4*, 2016.

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

[12] O'Donoghue and E. J. Candes, "Adaptive restart for accelerated gradient schemes," *Foundations of Computational Mathematics*, vol. 15, no. 3, pp. 715–732, 2013.

[13] J. I. Poveda and N. Li, "Robust hybrid zero-order optimization algorithms with acceleration via averaging in continuous time," *arXiv:1909.00265*, 2019.

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

[15] D. M. Hustig-Schultz and R. Sanfelice, "A robust hybrid heavy ball algorithm for optimization with high performance," *American Control Conference*, pp. 151–156, 2019.

[16] C. A. Uribe, S. Lee, A. Gasnikov, and A. Nedić, "A dual approach for optimal algorithms in distributed optimization over networks," *Optimization Methods and Software*, vol. 0, no. 0, pp. 1–40, 2020.

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

[18] J. Zhang, C. A. Uribe, A. Mokhtari, and A. Jadbabaie, "Achieving acceleration in distributed optimization via direct discretization of the Heavy-Ball ODE," in *2019 American Control Conference (ACC)*, pp. 3408–3413, IEEE, 2019.

[19] R. T. Rockafellar and R. J. Wets, *Variational Analysis*. Springer, 1998.

[20] J. I. Poveda and A. R. Teel, "Hybrid mechanisms for robust synchronization and coordination of multi-agent networked sampled-data systems," *Automatica*, vol. 99, pp. 41–53, 2019.

[21] R. G. Sanfelice and A. R. Teel, "Dynamical properties of hybrid systems simulators," *Automatica*, vol. 46, pp. 239–248, 2010.

[22] A. R. Teel and J. I. Poveda, "A hybrid systems approach to global synchronization and coordination of multi-agent sampled-data systems," *In proc. of Analy. and Design of Hybrid Syst.*, pp. 123–128, 2015.

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

[24] A. R. Teel, F. Forni, and L. Zaccarian, "Lyapunov-based sufficient conditions for exponential stability in hybrid systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 6, pp. 1591–1596, 2013.