arXiv:2104.10113v2 [math.OC] 30 Sep 2021

Exponentially Converging Distributed Gradient Descent with Intermittent Communication via Hybrid Methods

Katherine R. Hendrickson[a], Dawn M. Hustig-Schultz[b], Matthew T. Hale[a], and Ricardo G. Sanfelice[b]

Abstract

We present a hybrid systems framework for multi-agent optimization in which agents execute computations in continuous time and communicate in discrete time. The optimization algorithm is a hybrid version of parallelized coordinate descent. Agents implement a sample-and-hold strategy in which gradients are computed at communication times and held constant during flows between communications. Completeness of maximal solutions under these hybrid dynamics is established. Under assumptions of smoothness and strong convexity, we show that this system exponentially converges to the minimizer of an objective function. Simulation results illustrate this convergence rate.

I. Introduction

Convex optimization problems arise in many areas of engineering, including machine learning [1], communications [2], robotics [3], and others. Fundamentally, regardless of the application area, the goal is to design an algorithm that will converge to a minimum of an objective function, possibly under some constraints. Recently, there has been increased interest in studying optimization algorithms in continuous time using tools from dynamical systems to establish convergence to minimizers; see [4]–[6].

In this paper, we develop a hybrid optimization algorithm for the analysis of multi-agent systems with continuous-time updates and intermittent discrete-time communication events. This is motivated by two factors. First, we wish to leverage the large collection of tools from dynamical systems to analyze multi-agent optimization. Second, there exist many multi-agent controllers that operate in continuous time to minimize some objective function, e.g., in consensus [7] and coverage control [8], and our analyses will apply to such systems. However, while individual agents' computations occur in continuous time, communication between them inherently occur in discrete time

[a]K. Hendrickson and M. Hale are with the Department of Mechanical and Aerospace Engineering, University of Florida. Emails: {kat.hendrickson,matthewhale}@ufl.edu. Research partially supported by AFOSR under Grant no. FA9550-19-1-0169, by ONR under Grant no. N00014-19-1-2543, and by a task order contract from the Air Force Research Laboratory through Eglin AFB.

[b]D. Hustig-Schultz and R. G. Sanfelice are with the Department of Computer Engineering, University of California, Santa Cruz. Emails: {dhustigs,ricardo}@ucsc.edu. Research partially supported by NSF Grants no. ECS-1710621, CNS-1544396, and CNS-2039054, by AFOSR under Grants no. FA9550-19-1-0053, FA9550-19-1-0169, and FA9550-20-1-0238, by the Army Research Office under Grant no. W911NF-20-1-0253, and by CITRIS and the Banatao Institute at the University of California.

because communicated information arrives at isolated time instants. This mixture of continuous- and discrete-time elements naturally leads us to a hybrid system model.

The algorithm we propose is essentially a hybrid version of parallelized block coordinate descent [9], in which each agent updates only a small subset of all decision variables in continuous time, and agents communicate these updates to others in discrete time. In the proposed model, communication between agents occurs when a decreasing timer reaches zero, at which point the timer is reset to some value within a specified range. Agents use a sample-and-hold strategy in which gradients are computed at the communication times and then held constant and used continuously until the next communication event. This approach is inspired by recent work [10] that has successfully applied it to synchronization problems. We consider objective functions that satisfy typical, mild assumptions for distributed optimization, namely strongly convex objective functions with Lipschitz gradients.

We leverage the theory of hybrid systems to prove that the proposed hybrid algorithm has several desirable properties. First, we define a hybrid system model for this algorithm and show that, under these hybrid dynamics, every maximal solution is complete, with domain allowing arbitrarily large ordinary time. As a result, there are no theoretical obstructions to running this algorithm for arbitrarily long periods of time. Second, we use Lyapunov analysis to show that, even under intermittent information sharing, the hybrid optimization algorithm exponentially converges to the minimizer of an objective function. Furthermore, we derive an explicit convergence rate in terms of system parameters.

The developments in this paper can be regarded as continuous-time counterparts to "classical" discrete-time algorithms in multi-agent optimization [9]. Related research in multi-agent continuous-time optimization includes [5], [11], [12], though those works all use a consensus-based update law that executes computations and communications both in continuous time. However, we avoid continuous-time communications to account for cases in which they are not possible or simply undesirable, e.g., over long distances or when power is limited.

The most similar works are [10], [13], which also study continuous-time optimization with discrete-time communication. However, those works also use consensus-based optimization algorithms in which each agent updates all decision variables. In contrast, we consider agents with a common objective function and require that each agent update only a small subset of decision variables. This has the advantage that an individual agent's computational burden can be small, even when solving high-dimensional problems.

The rest of the paper is organized as follows. Section II includes our problem statement, assumptions, and algorithm. Section III provides background on hybrid systems. We present our hybrid system model in Section IV and establish the existence of complete solutions. Section V proves that the hybrid multi-agent update law exponentially converges to the minimizer of an objective function. We include numerical results as validation in Section VI.

II. Problem Statement and Algorithm Overview

In this section, we state the class of problems that we consider and give an overview of the proposed hybrid optimization algorithm.

A. Problem Formulation

We consider a group of N agents jointly solving an optimization problem of the following form:

Problem 1. *Given an objective function $L : \mathbb{R}^n \to \mathbb{R}$,*

$$\text{minimize } L(x), \qquad x \in \mathbb{R}^n$$

using N distributed agents while requiring that (i) only one agent updates any entry of the decision variable x, and (ii) agents require only intermittent information sharing from others.

Each agent executes computations locally and then shares the results of those computations. Criterion (i) is there for scalability, only a single agent will update each decision variable. This reduces the computation load on agents and removes duplicated efforts. Criterion (ii) ensures that the algorithm performs even in environments where communications may be limited. In many practical settings, we expect bandwidth to be limited and/or agents to have limited onboard power available, which means communications should not be constant.

We assume the following about the objective function L.

Assumption 1. *The function L is twice continuously differentiable, β-strongly convex for some $\beta > 0$, and K-smooth (namely, ∇L is K-Lipschitz).* $\triangle$

Assumption 1 allows a large number of convex problems to be considered, such as strongly convex quadratic programs. It is a standard assumption in multi-agent optimization [9]. It implies that $K \geq \beta$.

We solve Problem 1 by applying gradient descent in continuous time using data received intermittently in discrete time. The proposed hybrid optimization algorithm uses jumps to characterize the discrete-time communication events and flows to represent the continuous-time dynamics. Analogously to past research that has developed distributed versions of the discrete-time gradient descent law, our update law during flows is based on the following (centralized) first-order dynamical system:

$$\dot{x} + \nabla L(x) = 0. \tag{1}$$

This is motivated by the use of gradient-based controllers in multi-agent systems, e.g., in consensus [7], as well as the simplicity of distributing gradient-based updates and the robustness to asynchrony that results from doing so [9]. Next, we distribute this across a team of agents.

B. Algorithmic Framework

We seek to distribute (1) across a team of agents in accordance with the parallelization requirement in Problem 1. We consider N agents indexed over $i \in [N] := \{1, \ldots, N\}$ and divide $x \in \mathbb{R}^n$ into N blocks. Then agent i is responsible for updating and communicating values of the i-th block, $x_i \in \mathbb{R}^{n_i}$, where $n_i \in \mathbb{N}$ and $\sum_{i \in [N]} n_i = n$. Thus, the variable x may be written as the vertical concatenation of all agents' blocks. Each agent performs gradient descent on their own block during flows but does not update any others.

Agents' updates occur in continuous time while communication of these updates occurs in discrete time. Communications are coordinated using a decreasing timer, τ, that is shared by all agents. When the timer reaches zero, all agents communicate their current values to all of the other agents and the timer resets to a value within a specified interval $[\tau_{min}, \tau_{max}]$. We assume that communicated data are received at the same time they are sent. These communicated blocks are gathered into the vector $\eta \in \mathbb{R}^n$ with the current value of x_i being assigned to η_i at

communication events. The value of η is used in each agent's continuous-time computations in a sample-and-hold manner between communication events. That is, each agent uses the previously communicated data in their updates rather than the continuously evolving values of the other agents. Formally, we write $\nabla_i L = \frac{\partial L}{\partial x_i}$, and during flows agent i executes

$$\dot{x}_i = -\nabla_i L(\eta).$$

This sample-and-hold method is common in the literature [10] and is used to demonstrate the feasibility of the hybrid approach in multi-agent optimization.

The complete algorithm is summarized in Algorithm 1.

Algorithm 1: Distributed Hybrid Gradient Descent

```
Initialization: set x_o, η_o ∈ ℝ^n and τ_o ∈ [0, τ_max];
while τ ≥ 0 do
    ẋ_i = −∇_i L(η), for all i ∈ {1, ..., N};
    τ̇ = −1;
    if τ = 0 then
        reset η_i to x_i, for all i ∈ {1, ..., N};
        reset τ to a value in [τ_min, τ_max];
    end
end
```

The next section provides the tools that will be used to analyze Algorithm 1.

III. Hybrid System Preliminaries

In this section, we recount the background material necessary for the hybrid system modeling and analysis in the remainder of the paper.

A. Preliminaries on Hybrid Systems

For the purposes of this paper, a hybrid system $\mathcal{H}$ has data (C, f, D, G) that takes the general form

$$\mathcal{H} = \begin{cases} \dot{x} = f(x) & x \in C \\ x^+ \in G(x) & x \in D \end{cases}, \tag{2}$$

where $x \in \mathbb{R}^n$ is the system's state, and f defines the flow map and continuous dynamics for which C is the flow set. The set-valued jump map G captures the system's discrete behavior for the jump set D. More information on this definition and hybrid systems can be found in [14].

Definition 1 (Hybrid Basic Conditions, [14])**.** *A hybrid system $\mathcal{H}$ as in* (2) *with data (C, f, D, G) satisfies the* hybrid basic conditions *if*

- *C and D are closed subsets of $\mathbb{R}^n$;*

- *f is a continuous function from $\mathbb{R}^n \to \mathbb{R}^n$;*
- *$G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is outer semicontinuous and locally bounded relative to D, and $D \subset$ dom G.*

If a hybrid system meets the hybrid basic conditions, then we say that the system is *well-posed* (Theorem 6.30, [14]).

We denote solutions to $\mathcal{H}$ by ϕ, which we parameterize by $(t, j) \in \mathbb{R}_{\geq 0} \times \mathbb{N}$, where t denotes the ordinary (continuous) time, and j denotes the jump (discrete) time. Per Definition 2.3 in [14], dom $\phi \subset \mathbb{R}_{\geq 0} \times \mathbb{N}$ is a *hybrid time domain* if for all $(T, J) \in$ dom ϕ, the dom $\phi \cap ([0, T] \times \{0, 1, \ldots, J\}$ can be written as $\bigcup_{j=0}^{J-1}([t_j, t_{j+1}], j)$ for some finite sequence of times $0 = t_0 \leq t_1 \leq \cdots \leq t_J$. We say that a solution ϕ is *complete* if dom ϕ is unbounded. A solution ϕ to $\mathcal{H}$ is called maximal if it cannot be extended further.

IV. Hybrid System Model

In this section, we define a hybrid system model that encompasses all agents' current states and their most recently communicated state values. Towards defining this model, we first formally define the timer that governs communication events. This allows us to define the hybrid subsystems that are distributed across the agents. Building on this, we present a definition of the hybrid system modelling the N agents, their algorithm, and the mechanism governing the communication events. Finally, we show the existence of solutions and conclude that all maximal solutions are complete.

A. Mechanism Governing the Communication Events

We seek to account for intermittent communication events that occur only at some time instances t_j, for $j \in \mathbb{N}$, that are not known *a priori*. We assume that the sequence $\{t_j\}_{j=1}^{\infty}$ is strictly increasing and unbounded. Between consecutive time events, some amount of time elapses which we upper and lower bound with positive scalars τ_{min} and τ_{max}:

$$0 < \tau_{min} \leq t_{j+1} - t_j \leq \tau_{max} \quad \forall j \in \mathbb{N} \setminus \{0\}. \tag{3}$$

The upper bound τ_{max} prevents infinitely long communication delays and ensures convergence, while the lower bound τ_{min} rules out Zeno behavior.

To generate events at times t_j satisfying (3), let τ be the timer that governs when agents exchange data, where τ is defined by

$$\begin{aligned} \dot{\tau} &= -1 && \tau \in [0, \tau_{max}], \\ \tau^+ &\in [\tau_{min}, \tau_{max}] && \tau = 0, \end{aligned}$$

for $\tau_{min}, \tau_{max} \in \mathbb{R}_{>0}$. The timer τ steadily decreases until it reaches zero. At this point, it is reset to a value within $[\tau_{min}, \tau_{max}]$.

There is indeterminacy built into the timer in that the reset map is only confined to a compact interval, $[\tau_{min}, \tau_{max}]$, where τ_{min} and τ_{max} are both positive real numbers.

B. Hybrid Subsystems

Recall that agent i stores and updates its own state variable $x_i \in \mathbb{R}^{n_i}$, and the variable x is the vertical concatenation of all agents' states. Data from all agents are collectively stored in $\eta \in \mathbb{R}^n$ at communication events. We define the state of agent i's hybrid system as $\xi_i = (x_i, \eta, \tau)$, where x_i is agent i's state (the one it is responsible for updating), η is the memory state storing the states of the agents measured at communication events, and τ is defined as above. This leads to the hybrid subsystem given by

$$\dot{\xi}_i = \begin{bmatrix} -\nabla_i L(\eta) \\ 0 \\ -1 \end{bmatrix} \qquad \xi_i \in \mathbb{R}^{n_i} \times \mathbb{R}^n \times [0, \tau_{max}]$$

$$\xi_i^+ \in \begin{bmatrix} x_i \\ x \\ [\tau_{min}, \tau_{max}] \end{bmatrix} \qquad \xi_i \in \mathbb{R}^{n_i} \times \mathbb{R}^n \times \{0\},$$

where $x = (x_1^T, \ldots, x_N^T)^T$.

C. Combined Hybrid System

We are now ready to combine the distributed subsystems into one hybrid system for analysis. First, we define a variable $z = (z_1, z_2) \in \mathbb{R}^n \times \mathbb{R}^n$ such that

$$z_1 = \text{col}(x_1, \ldots, x_N)$$

$$z_2 = \eta,$$

where $\text{col}(x_1, \ldots, x_N) = (x_1^T, \ldots, x_N^T)^T$.

We define the state of the combined hybrid system as $\xi = (z_1, z_2, \tau) \in \mathcal{X}$, where z_1, z_2, and τ are defined as above, and $\mathcal{X} := \mathbb{R}^n \times \mathbb{R}^n \times [0, \tau_{max}]$. This leads to the combined hybrid system $\mathcal{H} = (C, f, D, G)$ given by

$$\dot{\xi} = \begin{bmatrix} -\nabla L(z_2) \\ 0 \\ -1 \end{bmatrix} := f(\xi) \qquad \xi \in C, \tag{4}$$

$$\xi^+ \in \begin{bmatrix} z_1 \\ z_1 \\ [\tau_{min}, \tau_{max}] \end{bmatrix} := G(\xi) \qquad \xi \in D, \tag{5}$$

where $C := \mathcal{X}$ and $D := \mathbb{R}^n \times \mathbb{R}^n \times \{0\}$.

D. Hybrid Basic Conditions

We now demonstrate that $\mathcal{H}$ meets the hybrid basic conditions and is well-posed.

Lemma 1. *Let L satisfy Assumption 1. Then, the hybrid system given by $\mathcal{H}$ with data (C, f, D, G) defined in* (4)–(5) *satisfies the hybrid basic conditions from Definition 1 and is nominally well-posed as a result.*

Proof: The sets C and D are closed subsets of $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$ by definition. Due to our assumption that ∇L is continuous, f is a continuous function from $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$ to $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$. By construction, G is outer semicontinuous and locally bounded relative to D. Finally, $D \subset \text{dom } G$ because dom G is $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$. ■

E. Existence of Solutions

In addition to being well-posed, there exists a nontrivial solution to $\mathcal{H}$ from each point in $C \cup D$, and all maximal solutions are complete and not Zeno under mild conditions on problem parameters. Complete solutions cannot be extended further and their domains are unbounded. Practically, this means that the proposed algorithm may run for an arbitrarily long period of time and does not reach a point where it can neither flow nor jump.

Lemma 2 (Existence of Solutions)**.** *Let Assumption 1 hold. Let τ_{min} and τ_{max} be such that $0 < \tau_{min} \leq \tau_{max} < \frac{\beta^2}{3K^3}$, where β is the strong convexity constant of L and K is the Lipschitz constant of ∇L. Then there exists a nontrivial solution to $\mathcal{H} = (C, f, D, G)$ from every initial point in $C \cup D$. Additionally, every maximal solution ϕ to the hybrid system $\mathcal{H}$ is complete and not Zeno.*

Proof: See the appendix. ■

V. Convergence Analysis

In this section, we define the set for solutions to converge to and present some useful properties of the hybrid system $\mathcal{H}$ in Lemmas 3 and 4. We then propose a Lyapunov function in Lemma 5. As an interim result, we show that for a solution $\phi = (\phi_{z_1}, \phi_{z_2}, \phi_\tau)$ to $\mathcal{H}$ in (4)–(5), if $\phi_{z_1}(0,0) = \phi_{z_2}(0,0)$, we are able to bound the distance from the minimizer of L for all $(t,j) \in \text{dom } \phi$. Finally, we present our main result, exponential convergence to the minimizer of L, in Theorem 1.

A. Convergence Set

Let $\mathbf{0}_n$ be the vector of zeros in $\mathbb{R}^n$; similarly, let $\mathbf{0}_{n_i}$ be the vector of zeros in $\mathbb{R}^{n_i}$. Convergence using gradient descent occurs when the gradient of L is $\mathbf{0}_n$. Given a complete solution $\phi = (\phi_{z_1}, \phi_{z_2}, \phi_\tau)$ to the hybrid system $\mathcal{H}$, we seek to assure that $\lim_{t+j\to\infty} \nabla_i L(\phi_{z_2}(t,j)) = \mathbf{0}_{n_i}$, for $i = 1, \ldots, N$. This is equivalent to a set convergence problem where the set to converge to for the hybrid system $\mathcal{H}$ is given by

$$\begin{aligned}\mathcal{A} &:= \{\xi = (z_1, z_2, \tau) \in \mathcal{X} : \nabla L(z_2) = \mathbf{0}_n, z_2 = z_1, \tau \in [0, \tau_{max}]\} \\ &= \{x^*\} \times \{x^*\} \times [0, \tau_{max}], \end{aligned} \tag{6}$$

where x^* is the unique fixed point of ∇L. Equivalence of the expression for $\mathcal{A}$ stems from Assumption 1: because L is strongly convex, it has a unique minimum (denoted by x^*) and this unique minimum is the unique stationary point of ∇L. Given a vector $\xi = (z_1, z_2, \tau) \in \mathcal{X}$, the squared distance from $\mathcal{A}$ is given by $|\xi|^2_{\mathcal{A}} := \|z - z^*\|^2 = \|z_1 - x^*\|^2 + \|z_2 - x^*\|^2$, where $\|\cdot\|$ denotes the Euclidean norm throughout this paper.

B. Useful Properties of $\mathcal{H}$

Combining gradient descent with a bound on $\tau_{\max}$ allows us to establish relationships that prove useful during Lyapunov analysis.

Lemma 3. *Let Assumption 1 hold. Consider the hybrid system given by $\mathcal{H}$ with data (C, f, D, G) defined in* (4)–(5). *Let τ_{min} and τ_{max} be such that $0 < \tau_{min} \leq \tau_{max} < \frac{\beta^2}{3K^3}$, where β is the strong convexity constant of L and K is the Lipschitz constant of ∇L. Denote the unique fixed point of ∇L by x^*. Pick a solution $\phi = (\phi_{z_1}, \phi_{z_2}, \phi_\tau)$ to $\mathcal{H}$ such that $\phi_{z_1}(0,0) = \phi_{z_2}(0,0)$. For each $I^j := \{t : (t,j) \in \text{dom } \phi\}$ with nonempty interior and with $t_{j+1} > t_j$ such that $[t_j, t_{j+1}] = I^j$, we have*

$$\phi_{z_1}(t,j) = \phi_{z_2}(t_j,j) - (t-t_j)\nabla L(\phi_{z_2}(t_j,j)) \tag{7}$$

$$\phi_{z_2}(t,j) = \phi_{z_2}(t_j,j), \tag{8}$$

for all $t \in (t_j, t_{j+1})$. Additionally, for all $(t,j) \in \text{dom } \phi$, the following are satisfied:

$$\|\phi_{z_1}(t,j) - x^*\|^2 \leq q(t,t_j)\|\phi_{z_2}(t_j,j) - x^*\|^2; \tag{9}$$

$$\|\phi_{z_1}(t,j) - \phi_{z_2}(t,j)\| \leq \tau_{max}\|\nabla L(\phi_{z_2}(t_j,j))\|; \tag{10}$$

$$\|\phi_{z_1}(t,j) - x^*\|^2 \geq B\|\phi_{z_2}(t_j,j) - x^*\|^2; \tag{11}$$

where $q(t,t_j) := (1 - 2(t-t_j)\beta + (t-t_j)^2K^2) \in (0,1)$ and $B := (1 - 2\tau_{max}K) \in (0,1)$.

Proof: Given $t \in (t_j, t_{j+1})$, the solution ϕ has flowed some distance given by $(t-t_j)\dot{\phi}$ where $\dot{\phi}$ is constant due to the sample-and-hold methodology. Applying our definition of f in (4) gives (7) and (8).
Proof of (9): Using (7) and the fact that $\nabla L(x^*) = 0$, we can rewrite $\|\phi_{z_1}(t,j) - x^*\|^2$ as

$$\begin{aligned}\|\phi_{z_1}(t,j) - x^*\|^2 &= \|\phi_{z_2}(t_j,j) - (t-t_j)\nabla L(\phi_{z_2}(t_j,j)) - x^* + (t-t_j)\nabla L(x^*)\|^2 \\ &= \|\phi_{z_2}(t_j,j) - x^*\|^2 - 2(t-t_j)(\nabla L(\phi_{z_2}(t_j,j)) - \nabla L(x^*))^T(\phi_{z_2}(t_j,j) - x^*) \\ &\quad + (t-t_j)^2\|\nabla L(\phi_{z_2}(t_j,j)) - \nabla L(x^*)\|^2,\end{aligned}$$

where the second equality follows from expanding the norm squared. Using the β-strong convexity of L and the Lipschitz property of ∇L, we upper bound this with

$$\begin{aligned}\|\phi_{z_1}(t,j) - x^*\|^2 &\leq \|\phi_{z_2}(t_j,j) - x^*\|^2 - 2(t-t_j)\beta\|\phi_{z_2}(t_j,j) - x^*\|^2 + (t-t_j)^2K^2\|\phi_{z_2}(t_j,j) - x^*\|^2 \\ &= (1 - 2(t-t_j)\beta + (t-t_j)^2K^2)\|\phi_{z_2}(t_j,j) - x^*\|^2.\end{aligned}$$

For contraction, we must show $(1-2(t-t_j)\beta+(t-t_j)^2K^2) < 1$. To derive a sufficient condition for this, note that $(1-2(t-t_j)\beta+(t-t_j)^2K^2) < 1$ may be rewritten as $(t-t_j)K^2 < 2\beta$ by subtracting 1 from both sides, dividing by $t-t_j$, and then adding 2β to both sides. Using $\beta \leq K$ and $\tau_{max} < \frac{\beta^2}{3K^3}$, we have $(t-t_j) \leq \tau_{max} < \frac{\beta^2}{3K^3} < \frac{2\beta}{K^2}$, and therefore $(1-2(t-t_j)\beta+(t-t_j)^2K^2) < 1$. To show that this term is also positive, it is sufficient to show that $(1 - 2(t-t_j)\beta) \geq 1 - 2\tau_{max}K > 0$. This is satisfied for $\tau_{max} < \frac{\beta^2}{3K^3} \leq \frac{1}{3K} < \frac{1}{K}$. Thus, $q(t,t_j) = (1 - 2(t-t_j)\beta + (t-t_j)^2K^2) \in (0,1)$.

Proof of (10): From (7)–(8), we have $\|\phi_{z_1}(t,j)-\phi_{z_2}(t,j)\| = \|\phi_{z_2}(t_j,j)-(t-t_j)\nabla L(\phi_{z_2}(t_j,j))-\phi_{z_2}(t_j,j)\|$. Simplifying and bounding $t-t_j$ above by τ_{max} gives the final result.

Proof of (11): Using (7) and the fact that $\nabla L(x^*)=0$, we can rewrite $\|\phi_{z_1}(t,j)-x^*\|^2$ as $\|\phi_{z_2}(t_j,j)-(t-t_j)\nabla L(\phi_{z_2}(t_j,j))-x^*+(t-t_j)\nabla L(x^*)\|^2$. Expanding the norm squared gives

$$\begin{aligned}\|\phi_{z_1}(t,j)-x^*\|^2 = \|\phi_{z_2}(t_j,j)-x^*\|^2 &+ (t-t_j)^2\|\nabla L(\phi_{z_2}(t_j,j))-\nabla L(x^*)\|^2\\ &-2(t-t_j)(\phi_{z_2}(t_j,j)-x^*)^T(\nabla L(\phi_{z_2}(t_j,j))-\nabla L(x^*)).\end{aligned}$$

Dropping the middle term (which is positive) and using $(\phi_{z_2}(t_j,j)-x^*)^T(\nabla L(\phi_{z_2}(t_j,j))-\nabla L(x^*)) \le K\|\phi_{z_2}(t_j,j)-x^*\|^2$, which follows from ∇L being K-Lipschitz, we can derive a lower bound:

$$\begin{aligned}\|\phi_{z_1}(t,j)-x^*\|^2 &\ge (1-2(t-t_j)K)\|\phi_{z_2}(t_j,j)-x^*\|^2\\ &\ge (1-2\tau_{max}K)\|\phi_{z_2}(t_j,j)-x^*\|^2.\end{aligned}$$

Let $B := 1-2\tau_{max}K$. Then using $\beta \le K$, we have $\tau_{max} < \frac{\beta^2}{6K^3} < \frac{1}{2K}$. Therefore, $B \in (0,1)$. ■

In preparation for establishing the convergence properties of $\mathcal{H}$, we also show that the angle between the gradient of the current state and the gradient of the previously communicated state is never greater than 90 degrees as a result of the bound on τ_{max}. This is formally stated in Lemma 4.

Lemma 4. *Let Assumption 1 hold. Consider the hybrid system given by $\mathcal{H}$ with data (C,f,D,G) defined in* (4)–(5)*. Let τ_{min} and τ_{max} be such that $0<\tau_{min}\le\tau_{max}<\frac{\beta^2}{3K^3}$, where β is the strong convexity constant of L and K is the Lipschitz constant of ∇L. Denote the unique fixed point of ∇L by x^*. Pick a solution ϕ such that $\phi_{z_1}(0,0)=\phi_{z_2}(0,0)$. For each $I^j := \{t:(t,j)\in \text{dom }\phi\}$ with nonempty interior and with $t_{j+1}>t_j$ such that $[t_j,t_{j+1}]=I^j$, we have*

$$\nabla L(\phi_{z_1}(t,j))^T\nabla L(\phi_{z_2}(t,j)) \ge A\|\phi_{z_2}(t_j,j)-x^*\|^2,$$

for all $t\in(t_j,t_{j+1})$, where $A := \beta^2(1-2\tau_{max}K)-\tau_{max}K^3>0$.

Proof: We begin by expanding $\|\nabla L(\phi_{z_1}(t,j))\|^2$:

$$\begin{aligned}\|\nabla L(\phi_{z_1}(t,j))\|^2 &= \nabla L(\phi_{z_1}(t,j))^T\Big(\nabla L(\phi_{z_2}(t,j))+\nabla L(\phi_{z_1}(t,j))-\nabla L(\phi_{z_2}(t,j))\Big)\\ &\le \nabla L(\phi_{z_1}(t,j))^T\nabla L(\phi_{z_2}(t,j)) + \|\nabla L(\phi_{z_1}(t,j))\|\|\nabla L(\phi_{z_1}(t,j))-\nabla L(\phi_{z_2}(t,j))\|\\ &\le \nabla L(\phi_{z_1}(t,j))^T\nabla L(\phi_{z_2}(t,j)) + K\|\nabla L(\phi_{z_1}(t,j))\|\|\phi_{z_1}(t,j)-\phi_{z_2}(t,j)\|\\ &\le \nabla L(\phi_{z_1}(t,j))^T\nabla L(\phi_{z_2}(t,j)) + \tau_{max}K\|\nabla L(\phi_{z_1}(t,j))\|\|\nabla L(\phi_{z_2}(t_j,j))\|, \qquad (12)\end{aligned}$$

where the last two inequalities are from the K-Lipschitz property of ∇L which states $\|\nabla L(x)-\nabla L(y)\| \le K\|x-y\|$, for all $x,y\in\mathbb{R}^n$ and (10) in Lemma 3. Note that the K-Lipschitz property also gives the inequal-

ities $\|\nabla L(\phi_{z_1}(t,j))\| \leq K\|\phi_{z_1}(t,j) - x^*\|$ and $\|\nabla L(\phi_{z_2}(t_j,j))\| \leq K\|\phi_{z_2}(t_j,j) - x^*\|$ because $\nabla L(x^*) = 0$. Thus, (12) becomes

$$\begin{aligned}\|\nabla L(\phi_{z_1}(t,j))\|^2 &\leq \nabla L(\phi_{z_1}(t,j))^T \nabla L(\phi_{z_2}(t_j,j)) + \tau_{max} K^3 \|\phi_{z_1}(t,j) - x^*\| \|\phi_{z_2}(t_j,j) - x^*\| \\ &\leq \nabla L(\phi_{z_1}(t,j))^T \nabla L(\phi_{z_2}(t_j,j)) + \tau_{max} K^3 \|\phi_{z_2}(t_j,j) - x^*\|^2, \end{aligned} \tag{13}$$

where the last inequality follows from (9) in Lemma 3. Because L is β-strongly convex, we have $\|\nabla L(\phi_{z_1}(t,j)) - \nabla L(x^*)\|^2 \geq \beta^2 \|\phi_{z_1}(t,j) - z^*\|^2$. To lower bound (13) in terms of $\|\phi_{z_2}(t_j,j) - x^*\|^2$, we combine this with $\nabla L(x^*) = 0$ and (11) in Lemma 3. This gives the inequality

$$\begin{aligned}\|\nabla L(\phi_{z_1}(t,j))\|^2 &\geq \beta^2 \|\phi_{z_1}(t,j) - x^*\|^2 \\ &\geq \beta^2 B \|\phi_{z_2}(t_j,j) - x^*\|^2, \end{aligned} \tag{14}$$

where $B := (1 - 2\tau_{max} K)$. Then using (14) in conjunction with (13) allows us to lower bound $\nabla L(\phi_{z_1}(t,j))^T \nabla L(\phi_{z_2}(t,j))$:

$$\begin{aligned}\nabla L(\phi_{z_1}(t,j))^T \nabla L(\phi_{z_2}(t,j)) &\geq \beta^2 B \|\phi_{z_2}(t_j,j) - x^*\|^2 - \tau_{max} K^3 \|\phi_{z_2}(t_j,j) - x^*\|^2 \\ &= (\beta^2(1 - 2\tau_{max} K) - \tau_{max} K^3) \|\phi_{z_2}(t_j,j) - x^*\|^2. \end{aligned}$$

This lower bound is positive for $\tau_{max} < \frac{\beta^2}{2\beta^2 K + K^3}$. Because $\beta \leq K$, the denominator $2\beta^2 K + K^3 \leq 3K^3$. This lower bound is positive since $\tau_{max} < \frac{\beta^2}{3K^3} \leq \frac{\beta^2}{2\beta^2 K + K^3}$. ■

C. Bound on the Lyapunov Function

Central to proving our main result is a Lyapunov function that is bounded above and below by $\mathcal{K}_\infty$ comparison functions α_1, α_2 given in Lemma 5.

Lemma 5. *Let Assumption 1 hold. Let τ_{min} and τ_{max} be such that $0 < \tau_{min} \leq \tau_{max} < \frac{\beta^2}{3K^3}$, where β is the strong convexity constant of L and K is the Lipschitz constant of ∇L. Let $V : \mathcal{X} \to \mathbb{R}_{\geq 0}$ be a Lyapunov function candidate for the hybrid system $\mathcal{H} = (C, f, D, G)$ defined in* (4)–(5)*, given by*

$$V(\xi) = (L(z_1) - L(x^*))^2 + (L(z_2) - L(x^*))^2,$$

for all $\xi = (z_1, z_2, \tau) \in \mathcal{X}$, where L is the objective function and x^* is the unique fixed point of ∇L. Then there exist $\alpha_1, \alpha_2 \in \mathcal{K}_\infty$ such that

$$\alpha_1(|\xi|_{\mathcal{A}}) \leq V(\xi) \leq \alpha_2(|\xi|_{\mathcal{A}})$$

for all $\xi \in C \cup D \cup G(D)$. In particular, α_1 and α_2 may be given by, for each $s \geq 0$,

$$\alpha_1(s) = \frac{\beta^2}{16} s^4 \quad \text{and} \quad \alpha_2(s) = \frac{K^2}{2} s^4.$$

Proof: The minimizer of L is $\xi^* = (z_1^*, z_2^*, \tau) = (x^*, x^*, \tau)$ for any τ, and, by construction, $V(\xi)$ is zero only for $\xi = \xi^*$ and is positive otherwise.

Because ∇L is K-Lipschitz, $L(x) - L(x^*) \leq \nabla L(x^*)^T (x - x^*) + \frac{K}{2}\|x - x^*\|^2 = \frac{K}{2}\|x - x^*\|^2$, for all $x \in \mathbb{R}^n$. Thus, $V(\xi)$ may be bounded as

$$V(\xi) \leq \frac{K^2}{4}\|z_1 - x^*\|^4 + \frac{K^2}{4}\|z_2 - x^*\|^4. \tag{15}$$

First consider the case where $\|z_1 - x^*\| \leq \|z_2 - x^*\|$. Then (15) becomes

$$V(\xi) \leq \frac{K^2}{2}\|z_2 - x^*\|^4 \leq \frac{K^2}{2}|\xi|_{\mathcal{A}}^4, \tag{16}$$

where the last inequality follows from $\|z_2 - x^*\|^2 \leq |\xi|_{\mathcal{A}}^2$ by definition of $|\xi|_{\mathcal{A}}$. If $\|z_1 - x^*\| > \|z_2 - x^*\|$ instead, then (15) is bounded above by

$$V(\xi) < \frac{K^2}{2}\|z_1 - x^*\|^4 \leq \frac{K^2}{2}|\xi|_{\mathcal{A}}^4,$$

where the definition of $|\xi|_{\mathcal{A}}$ is again used in the last inequality. Thus, $V(\xi) \leq \frac{K^2}{2}|\xi|_{\mathcal{A}}^4$ for all $\xi \in \mathcal{X}$ and we set $\alpha_2(s) = \frac{K^2}{2}s^4 \in \mathcal{K}_\infty$ for all $s \geq 0$.

The β-strong convexity of L and $\nabla L(x^*) = 0$ allow us to write $L(z_1) - L(x^*) \geq \frac{\beta}{2}\|z_1 - x^*\|^2$ and $L(z_2) - L(x^*) \geq \frac{\beta}{2}\|z_2 - x^*\|^2$. Applying both inequalities to the definition of V gives

$$V(\xi) \geq \frac{\beta^2}{4}\|z_1 - x^*\|^4 + \frac{\beta^2}{4}\|z_2 - x^*\|^4. \tag{17}$$

First, consider the case where $\|z_1 - x^*\| \geq \|z_2 - x^*\|$. Then dropping the second term in (17) and using $|\xi|_{\mathcal{A}}^2 = \|z_1 - x^*\|^2 + \|z_2 - x^*\|^2 \leq 2\|z_1 - x^*\|^2$,

$$V(\xi) \geq \frac{\beta^2}{4}\|z_1 - x^*\|^4 \geq \frac{\beta^2}{16}|\xi|_{\mathcal{A}}^4. \tag{18}$$

Now consider the case that $\|z_1 - x^*\| < \|z_2 - x^*\|$. Then the same steps apply and $V(\xi) > \frac{\beta^2}{16}|\xi|_{\mathcal{A}}^4$. Thus, $V(\xi) \geq \frac{\beta^2}{16}|\xi|_{\mathcal{A}}^4$ for all $\xi \in \mathcal{X}$. Accordingly, $\alpha_1(s) = \frac{\beta^2}{16}s^4$ for all $s \geq 0$. ■

D. Exponential Convergence

Using Lemmas 3, 4, and 5, we are able to bound the distance to the minimizer of L over time for a class of initial conditions in Proposition 1. This result will then be expanded to include all possible solutions and initial conditions in Theorem 1, thus showing exponential convergence to the minimizer of L.

Proposition 1. *Let Assumption 1 hold and consider the hybrid system $\mathcal{H}$ defined in* (4)-(5)*. Let $\mathcal{A}$ be as defined in* (6) *and let τ_{min} and τ_{max} be such that $0 < \tau_{min} \leq \tau_{max} < \frac{\beta^2}{3K^3}$, where β is the strong convexity constant of L and K is the Lipschitz constant of ∇L. For each solution ϕ to $\mathcal{H}$ such that $\phi_{z_1}(0,0) = \phi_{z_2}(0,0)$, for all $(t,j) \in \text{dom}\ \phi$, the following is satisfied:*

$$|\phi(t,j)|_{\mathcal{A}} \leq \sqrt{\frac{K}{\beta}}\sqrt[4]{8}\exp\Big(-\frac{\beta AB}{8K^2}t\Big)|\phi(0,0)|_{\mathcal{A}},$$

where $A = \beta^2(1 - 2\tau_{max}K) - \tau_{max}K^3 > 0$ and $B = (1 - 2\tau_{max}K) \in (0,1)$.

Proof: We first consider $\xi \in C$ and the Lyapunov function V defined in Lemma 5. The partial derivatives of V with respect to z_i are given by $\nabla_{z_i}V(\xi) = 2\nabla L(z_i)(L(z_i) - L(x^*)) \in \mathbb{R}^n$, for $i = 1, 2$, where x^* is the unique fixed point of ∇L. This leads to

$$\langle \nabla V(\xi), f(\xi)\rangle = -2(L(z_1) - L(x^*))\nabla L(z_1)^T \nabla L(z_2). \tag{19}$$

We now pick a solution ϕ such that $\phi_{z_1}(0,0) = \phi_{z_2}(0,0)$. For each $I^j := \{t : (t,j) \in \text{dom } \phi\}$ with nonempty interior and with $t_{j+1} > t_j$ such that $[t_j, t_{j+1}] = I^j$, we have from Lemma 4,

$$\nabla L(\phi_{z_1}(t,j))^T \nabla L(\phi_{z_2}(t,j)) \geq A\|\phi_{z_2}(t_j,j) - x^*\|^2,$$

where $A = \beta^2(1 - 2\tau_{max}K) - \tau_{max}K^3 > 0$. Combining this with $(L(\phi_{z_1}(t,j)) - L(x^*)) \geq \frac{\beta}{2}\|\phi_{z_1}(t,j) - x^*\|^2$ from the β-strong convexity of L, and using (19) we can write

$$\begin{aligned}\langle \nabla V(\phi(t,j)), f(\phi(t,j))\rangle &\leq -\beta\|\phi_{z_1}(t,j) - x^*\|^2 \nabla L(\phi_{z_1}(t,j))^T \nabla L(\phi_{z_2}(t,j)) \\ &\leq -\beta A\|\phi_{z_1}(t,j) - x^*\|^2 \|\phi_{z_2}(t_j,j) - x^*\|^2 \\ &\leq -\beta AB\|\phi_{z_2}(t_j,j) - x^*\|^4, \end{aligned} \tag{20}$$

where the last inequality is from applying (11) in Lemma 3. Additionally, for each (t,j) in the interval of flow $[t_j, t_{j+1}] \times \{j\}$ in dom ϕ, we have $\|\phi_{z_1}(t,j) - x^*\| \leq \|\phi_{z_2}(t_j,j) - x^*\|$ from (9) in Lemma 3. Applying this to the definition of $|\phi(t,j)|^2_{\mathcal{A}}$ gives the relationship $|\phi(t,j)|^2_{\mathcal{A}} \leq 2\|\phi_{z_2}(t_j,j) - x^*\|^2$. Squaring both sides allows us to rewrite (20) as

$$\langle \nabla V(\phi(t,j)), f(\phi(t,j))\rangle \leq -\frac{\beta AB}{4}|\phi(t,j)|^4_{\mathcal{A}}. \tag{21}$$

This bound may be related back to $V(\phi(t,j))$ using the comparison function α_2 from Lemma 5, which leads to

$$\langle \nabla V(\phi(t,j)), f(\phi(t,j))\rangle \leq -\frac{\beta AB}{2K^2}V(\phi(t,j)). \tag{22}$$

We now consider the change of V at jumps. For each $\xi \in D$, and $g = (g_{z_1}, g_{z_2}, g_\tau) \in G(\xi)$, $V(g) = (L(g_{z_1}) - L(x^*))^2 + (L(g_{z_2}) - L(x^*))^2$. Thus, we can write the change of V at jumps as $V(g) - V(\xi) = (L(g_{z_1}) - L(x^*))^2 - (L(z_2) - L(x^*))^2$ because $g_{z_1} = z_1$ and $g_{z_2} = z_1$ from (5).

For a solution ϕ such that $\phi_{z_1}(0,0) = \phi_{z_2}(0,0)$, this is equivalent to

$$V(G(\phi(t_{j+1},j))) - V(\phi(t_{j+1},j)) = (L(\phi_{z_1}(t_{j+1},j+1)) - L(x^*))^2 - (L(\phi_{z_2}(t_{j+1},j)) - L(x^*))^2,$$

for all $(t_{j+1},j), (t_{j+1},j+1) \in$ dom ϕ. For this quantity to be nonpositive, it is sufficient to show that $L(\phi_{z_1}(t_{j+1},j+1)) \leq L(\phi_{z_2}(t_{j+1},j))$.

Towards doing this, we leverage $L(x) \leq L(y) + \nabla L(y)^T(x-y) + \frac{K}{2}\|x-y\|^2$, for all $x, y \in \mathbb{R}^n$ from the K-Lipschitz property of ∇L. Applying this for $x = \phi_{z_1}(t_{j+1}, j+1), y = \phi_{z_2}(t_{j+1},j)$, we have

$$\begin{aligned} L(\phi_{z_1}(t_{j+1},j+1)) &\leq L(\phi_{z_2}(t_{j+1},j)) + \nabla L(\phi_{z_2}(t_{j+1},j))^T(\phi_{z_1}(t_{j+1},j+1) - \phi_{z_2}(t_{j+1},j)) \\ &\quad + \frac{K}{2}\|\phi_{z_1}(t_{j+1},j+1) - \phi_{z_2}(t_{j+1},j)\|^2 \\ &= L(\phi_{z_2}(t_{j+1},j)) + \nabla L(\phi_{z_2}(t_{j+1},j))^T(\phi_{z_2}(t_j,j) - (t_{j+1}-t_j)\nabla L(\phi_{z_2}(t_j,j)) - \phi_{z_2}(t_{j+1},j)) \\ &\quad + \frac{K}{2}\|\phi_{z_2}(t_j,j) - (t_{j+1}-t_j)\nabla L(\phi_{z_2}(t_j,j)) - \phi_{z_2}(t_j,j)\|^2 \\ &= L(\phi_{z_2}(t_{j+1},j)) - \Big(1 - \frac{K}{2}(t_{j+1}-t_j)\Big)(t_{j+1}-t_j)\|\nabla L(\phi_{z_2}(t_j,j))\|^2, \end{aligned}$$

where the first equality uses $\phi_{z_1}(t_{j+1}, j+1) = \phi_{z_1}(t_{j+1}, j) = \phi_{z_2}(t_j, j) - (t_{j+1} - t_j)\nabla L(\phi_{z_2}(t_j, j))$ from (5) and (7), and the last equality uses $\phi_{z_2}(t_{j+1}, j) = \phi_{z_2}(t_j, j)$. Since $t_{j+1} - t_j \leq \tau_{max} \leq \frac{\beta^2}{3K^3} \leq \frac{1}{3K} \leq \frac{2}{K}$, we have the desired property: $L(\phi_{z_1}(t_{j+1}, j+1)) \leq L(\phi_{z_2}(t_{j+1}, j))$. Thus,

$$V(G(\phi(t_{j+1}, j))) - V(\phi(t_{j+1}, j)) \leq 0. \tag{23}$$

Following the work done in [15] and [14], we are able to perform direct integration in order to upper bound $V(\phi(t, j))$ in terms of $V(\phi(0, 0))$ using (22) and (23) as bounds. Thus,

$$V(\phi(t, j)) \leq \exp\Big(-\frac{\beta AB}{2K^2}t\Big)V(\phi(0, 0)).$$

Using the comparison functions given in Lemma 5, we get a bound for $|\phi(t, j)|^4$:

$$\begin{aligned} |\phi(t, j)|^4_{\mathcal{A}} &\leq \frac{16}{\beta^2}\exp\Big(-\frac{\beta AB}{2K^2}t\Big)V(\phi(0, 0)) \\ &\leq \frac{8K^2}{\beta^2}\exp\Big(-\frac{\beta AB}{2K^2}t\Big)|\phi(0, 0)|^4_{\mathcal{A}}. \end{aligned}$$

Taking the fourth root gives the final answer. ■

In practice, this preliminary result is useful when agreeing on initial values is easy to implement. However, it does not show our desired result, namely, exponential convergence to the minimizer of L, regardless of initialization. By examining all possible scenarios at the first jump, we show in Theorem 1 below that exponential convergence to the minimizer of L still applies after the first jump.

Theorem 1 (Exponential Convergence)**.** *Let Assumption 1 hold and consider the hybrid system $\mathcal{H}$ defined in* (4)-(5)*. Let $\mathcal{A}$ be as defined in* (6) *and choose τ_{min} and τ_{max} such that $0 < \tau_{min} \leq \tau_{max} < \frac{\beta^2}{3K^3}$, where β is the strong convexity constant of L and K is the Lipschitz constant of ∇L. For each solution ϕ and for all $(t, j) \in \text{dom}\ \phi$ such that $j \geq 1$, the following is satisfied:*

$$|\phi(t, j)|_{\mathcal{A}} \leq \frac{8}{3}\sqrt[4]{2}\sqrt{\frac{K}{\beta}}\exp\Big(-\frac{\beta AB}{8K^2}t\Big)|\phi(0, 0)|_{\mathcal{A}},$$

where $A = \beta^2(1 - 2\tau_{max}K) - \tau_{max}K^3 > 0$ and $B = (1 - 2\tau_{max}K) \in (0, 1)$.

Proof: Two initialization scenarios must be considered: $\phi_{z_1}(0, 0) = \phi_{z_2}(0, 0)$ and $\phi_{z_1}(0, 0) \neq \phi_{z_2}(0, 0)$.

For the first case, $\phi_{z_1}(0, 0) = \phi_{z_2}(0, 0)$, Proposition 1 applies in its original form. This is the best-case scenario that results in the smallest upper bound.

Now consider the second case, when $\phi_{z_1}(0, 0) \neq \phi_{z_2}(0, 0)$. We note that after the first jump, all assumptions of Proposition 1 hold. Thus, for any solution ϕ where $(t, j) \in \text{dom}\ \phi$ such that $j \geq 1$, we have the following:

$$|\phi(t, j)|_{\mathcal{A}} \leq \sqrt{\frac{K}{\beta}}\sqrt[4]{8}\exp\Big(-\frac{\beta AB}{8K^2}t\Big)|\phi(t_1, 1)|_{\mathcal{A}}, \tag{24}$$

where t_1 denotes the time of the first jump.

We now seek to bound $|\phi(t_1, 1)|_{\mathcal{A}}$ in terms of $|\phi(0, 0)|_{\mathcal{A}}$. We begin by expanding and simplifying $|\phi(t_1, 1)|^2_{\mathcal{A}}$ using that $\phi_{z_2}(t_1, 1) = \phi_{z_1}(t_1, 1)$ after the first jump:

$$|\phi(t_1, 1)|^2_{\mathcal{A}} = \|\phi_{z_1}(t_1, 1) - x^*\|^2 + \|\phi_{z_2}(t_1, 1) - x^*\|^2 = 2\|\phi_{z_1}(t_1, 1) - x^*\|^2, \tag{25}$$

where x^* is the unique fixed point of ∇L. Applying the flow dynamics and using $\phi_{z_1}(t_1,1) = \phi_{z_1}(t_1,0)$ allows us to rewrite the distance of ϕ_{z_1} from the minimizer of L after the first jump as

$$\begin{aligned}\|\phi_{z_1}(t_1,1) - x^*\|^2 &= \|\phi_{z_1}(0,0) - t_1\nabla L(\phi_{z_2}(0,0)) - x^*\|^2 \\ &\leq \|\phi_{z_1}(0,0) - x^*\|^2 + 2t_1\|\phi_{z_1}(0,0) - x^*\|\|\nabla L(\phi_{z_2}(0,0))\| + t_1^2\|\nabla L(\phi_{z_2}(0,0))\|^2 \\ &\leq \|\phi_{z_1}(0,0) - x^*\|^2 + 2t_1K\|\phi_{z_1}(0,0) - x^*\|\|\phi_{z_2}(0,0) - x^*\| + t_1^2K^2\|\phi_{z_2}(0,0) - x^*\|^2, \quad (26)\end{aligned}$$

where the last inequality comes from the K-Lipschitz property of ∇L and the fact that $\nabla L(x^*) = 0$. If $\|\phi_{z_1}(0,0) - x^*\| \leq \|\phi_{z_2}(0,0) - x^*\|$, then (26) becomes

$$\begin{aligned}\|\phi_{z_1}(t_1,1) - x^*\|^2 &\leq (1 + 2t_1K + t_1^2K^2)\|\phi_{z_2}(0,0) - x^*\|^2 \\ &\leq (1 + t_1K)^2|\phi(0,0)|_{\mathcal{A}}^2.\end{aligned}$$

Similarly, if $\|\phi_{z_1}(0,0) - x^*\| > \|\phi_{z_2}(0,0) - x^*\|$, then (26) becomes $\|\phi_{z_1}(t_1,1) - x^*\|^2 < (1 + t_1K)^2|\phi(0,0)|_{\mathcal{A}}^2$. Thus, in both cases,

$$\|\phi_{z_1}(t_1,1) - x^*\|^2 \leq (1 + t_1K)^2|\phi(0,0)|_{\mathcal{A}}^2$$

Applying this to (25) and using $t_1 \leq \tau_{max} < \frac{\beta^2}{3K^3} < \frac{1}{3K}$, we get the bound

$$|\phi(t_1,1)|_{\mathcal{A}}^2 \leq 2\left(\frac{4}{3}\right)^2 |\phi(0,0)|_{\mathcal{A}}^2. \quad (27)$$

Taking the square root and applying to (24) gives the final result for any $(t,j) \in \text{dom } \phi$ such that $j \geq 1$. ■

VI. Numerical Validation

We consider $N = n$ agents for various values of n. Each agent updates a scalar and they minimize

$$L(x) = \frac{1}{2}x^TQx + b^Tx,$$

where $x \in \mathbb{R}^n$, Q is a $n \times n$ symmetric, positive definite matrix, and $b \in \mathbb{R}^n$. To form Q, we decompose a random $n \times n$ matrix into an unitary orthogonal matrix U and a matrix D that contains only our desired eigenvalues on the diagonal. We use these two resulting matrices to set $Q = U^TDU$. Our choice of eigenvalues varies by trial (discussed below) with the minimum eigenvalue corresponding to β and the maximum eigenvalue corresponding to K. The entries of b are set to random values between 1 and 5. Simulations used the HyEq Toolbox (Version 2.04) [16][1].

We first compare convergence results for different initial values of ϕ_{z_1} and ϕ_{z_2} for five agents. For the first trial, we consider the case where $\phi_{z_1}(0,0) = \phi_{z_2}(0,0) = (2,2,2,2,2)^T$. In Trial 2, we consider the "worst-case" initialization scenario: when $\phi_{z_1}(0,0) = (2,2,2,2,2)^T$ is some distance from the optimum but $\phi_{z_2}(0,0) = x^*$, resulting in an increase in the distance from the minimizer of L before the first jump. We consider $\beta = K = 5$ and set $\tau_{max} = \frac{\beta^2}{3K^3+1}$ and $\tau_{min} = \frac{1}{2}\tau_{max}$. Figure 1 shows the distance from optimum through the first twenty jumps for both trials. There is a consistent decrease in the distance to the minimizer, even at jumps, for the first trial. In

[1]Simulation code for this section may be found at www.github.com/kathendrickson/DistrHybridGD.

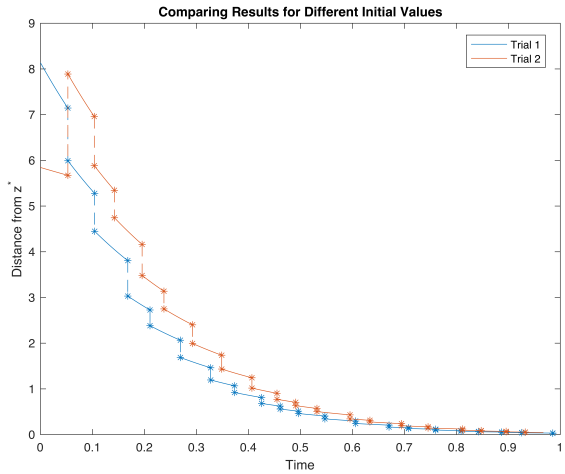

Fig. 1. Effect of initial values on convergence for two trials, where flows are denoted with solid lines and jumps with stars and dashed lines. Trial 1 sets $\phi_{z_1}(0,0) = \phi_{z_2}(0,0)$, while Trial 2 sets $\phi_{z_2}(0,0) = x^*$ instead. The first jump does not increase the distance from the minimizer in Trial 1 but does increase this distance in Trial 2. However, after the first jump, progress continues toward the optimum and differences between trials diminish.

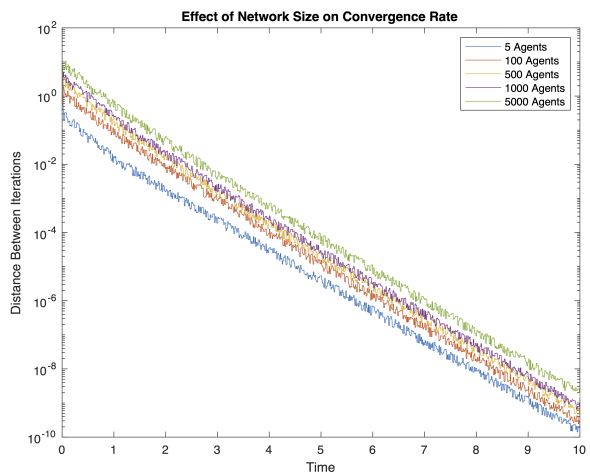

Fig. 2. Effect of network size on convergence. Convergence results still hold even for very large network sizes, demonstrating the scalability of our algorithm.

contrast, when initial values for ϕ_{z_1} and ϕ_{z_2} are not equal, there is an increase in distance to the minimizer after the first jump in the second trial. However, as expected, distance to the optimum decreases exponentially thereafter, with the difference between the two trials decreasing over time.

We then examined the effects of varying the network size from 5 agents to 100, 500, 1000, and 5000 agents. We set $\beta = 2$ and $K = 4$ and chose to initialize ϕ_{z_1} and ϕ_{z_2} with vectors of twos in $\mathbb{R}^n$. For each network size, the matrix Q and vector b were randomly generated. As shown in Figure 2, drastically expanding the network size does not have a significant impact on convergence. This demonstrates our algorithm's scalability and convergence

results that hold regardless of network size.

VII. Conclusion

This paper presented a hybrid systems framework for analyzing continuous-time multi-agent optimization with discrete-time communications. Using this framework, we established that every maximal solution is complete, as well as the exponential convergence of a block coordinate descent law to the minimizer of a strongly convex and smooth objective function. Future work in this area includes the use of heterogeneous timers and exploration of other update laws, as well as constrained problems.

Appendix

Proof of Lemma 2: Using Proposition 6.10 in [14] with $U = C$, let $\xi = (z_1, z_2, \tau) \in C \backslash D$. Then $f(\xi) \subset T_C(\xi)$. Because $G(D) \subset C$, case (c) in Proposition 6.10 does not apply. We avoid case (b) of Proposition 6.10 by showing that every solution lies entirely in a compact subset $W \subset C$. To do this, we verify is that there is no finite escape time for any solution. Consider a solution ϕ. Then $\phi_{z_1}(0,0)$ and $\phi_{z_2}(0,0)$ denote the initial values of ϕ_{z_1} and ϕ_{z_2}, respectively. Let t_1 denote the continuous-time at which the first jump occurs; then the value of ϕ after the first jump may be written as $\phi(t_1, 1)$. We first show that there is no finite escape time from initialization through the first jump. Towards doing this, we use (27), $|\phi(t_1,1)|^2_{\mathcal{A}} \leq 2(\frac{4}{3})^2 |\phi(0,0)|^2_{\mathcal{A}}$, and apply the comparison functions in Lemma 5. This gives the set of inequalities

$$\begin{aligned} V(\phi(t_1,1)) &\leq \frac{K^2}{2} |\phi(t_1,1)|^4_{\mathcal{A}} \\ &\leq \frac{K^2}{2} 4 \left(\frac{4}{3}\right)^4 |\phi(0,0)|^4_{\mathcal{A}} \\ &\leq \frac{K^2}{2} 4 \left(\frac{4}{3}\right)^4 \frac{16}{\beta^2} V(\phi(0,0)), \end{aligned}$$

which follow from (16), (27), and (18), respectively. Thus, through the first jump, $V(\phi(t_1,1)) \leq \frac{8192K^2}{81\beta^2} V(\phi(0,0))$.

After this first jump, $\phi_{z_1}(t_1,1) = \phi_{z_2}(t_1,1)$ then holds. By construction and Assumption 1, ∇L is Lipschitz and thus the map f is Lipschitz as well. Applying (21) after hybrid time $(t_1, 1)$, we have

$$\dot{V}(\phi(t,j)) = \langle \nabla V(\phi(t,j)), f(\phi(t,j)) \rangle \leq 0, \tag{28}$$

for all $(t,j) \in \text{dom } \phi$ such that $j \geq 1$. Thus, for any solution ϕ, we see that

$$V(\phi(t,j)) \leq \frac{8192K^2}{81\beta^2} V(\phi(0,0)) \tag{29}$$

for all $(t,j) \in \text{dom } \phi$.

Now consider a solution to $\dot{\xi} = f(\xi)$ that starts from some c-sublevel set $W_1 = \{\xi \in \mathcal{X} : V(\xi) \leq c\}$. Then, from (28) and (29), we see that all such solutions remain in the sub-level set $W_2 = \{\xi \in \mathcal{X} : V(\xi) \leq \max\{c, \frac{8192K^2}{81\beta^2} V\big(\phi(0,0)\big)\}\}$. Because V is continuous and radially unbounded by Lemma 5, W_2 is compact. From (28), W_2 is forward invariant for $\mathcal{H}$. Thus any trajectory that starts in the subset W_1 remains in W_2. Thus by Theorem 3.3 in [17], there is no finite escape time from C. ■

REFERENCES

[1] S. Sra, S. Nowozin, and S. J. Wright, *Optimization for Machine Learning*. MIT Press, 2012.

[2] Z. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1426–1438, 2006.

[3] D. Verscheure, B. Demeulenaere, J. Swevers, J. De Schutter, and M. Diehl, "Time-optimal path tracking for robots: A convex optimization approach," *IEEE Transactions on Automatic Control*, vol. 54, no. 10, pp. 2318–2327, 2009.

[4] W. Su, S. Boyd, and E. J. Candès, "A differential equation for modeling nesterov's accelerated gradient method: Theory and insights," *Journal of Machine Learning Research*, vol. 17, no. 153, pp. 1–43, 2016. [Online]. Available: http://jmlr.org/papers/v17/15-084.html

[5] S. Rahili and W. Ren, "Distributed continuous-time convex optimization with time-varying cost functions," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1590–1605, 2017.

[6] K. Garg and D. Panagou, "Fixed-time stable gradient flows: Applications to continuous-time optimization," *IEEE Transactions on Automatic Control*, 2020.

[7] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.

[8] J. Cortes, S. Martinez, T. Karatas, and F. Bullo, "Coverage control for mobile sensing networks," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 2, pp. 243–255, 2004.

[9] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. USA: Prentice-Hall, Inc., 1989.

[10] S. Phillips and R. G. Sanfelice, "Robust distributed synchronization of networked linear systems with intermittent information," *Automatica*, vol. 105, pp. 323–333, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0005109819301414

[11] B. Gharesifard and J. Cortés, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 781–786, 2014.

[12] J. Lu and C. Y. Tang, "Zero-gradient-sum algorithms for distributed convex optimization: The continuous-time case," *IEEE Transactions on Automatic Control*, vol. 57, no. 9, pp. 2348–2354, 2012.

[13] S. S. Kia, J. Cortés, and S. Martínez, "Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication," *Automatica*, vol. 55, pp. 254–264, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0005109815001053

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

[15] J. Chai, P. Casau, and R. G. Sanfelice, "Analysis and design of event-triggered control algorithms using hybrid systems tools," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 2017, pp. 6057–6062.

[16] R. G. Sanfelice, D. Copp, and P. Nanez, "A toolbox for simulation of hybrid systems in Matlab/Simulink: Hybrid equations (HyEQ) toolbox," in *Proceedings of the 16th International Conference on Hybrid Systems: Computation and Control*, 2013. [Online]. Available: https://doi.org/10.1145/2461328.2461346

[17] H. K. Khalil, *Nonlinear systems; 3rd ed.* Upper Saddle River, NJ: Prentice-Hall, 2002.