Computation-Aware Distributed Optimization over Networks: A Hybrid Dynamical Systems Approach

Daniel E. Ochoa
University of Colorado, Boulder
Colorado, USA
daniel.ochoa@colorado.edu

Jorge I. Poveda
University of Colorado, Boulder
Colorado, USA
jorge.poveda@colorado.edu

César A. Uribe
Rice University
Texas, USA
cauribe@rice.edu

KEYWORDS

Distributed Optimization, Accelerated Algorithms, Hybrid Dynamical Systems, Computation-Aware Algorithms

ACM Reference Format:

Daniel E. Ochoa, Jorge I. Poveda, and César A. Uribe. 2021. Computation-Aware Distributed Optimization over Networks: A Hybrid Dynamical Systems Approach. In *2021 CPS-IoT Week: Workshop on Computation-Aware Algorithmic Design for Cyber-Physical Systems, May 18-21, 2021*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3457335.3461710>

1 INTRODUCTION

The area of Cyber-Physical Systems (CPS) has emerged as a general discipline that studies complex dynamical systems that incorporate computation, control, and communication technologies [10]. The application of CPS spans several domains, from autonomous driving and intelligent transportation systems to deep-space exploration and quantum computing. However, traditionally, the algorithmic design and the implementation of CPS algorithms have been studied separately. While this approach facilitates the theoretical analysis of the system, it naturally poses critical issues when deploying the designed systems in their intended domain, i.e., the physical world.

For instance, the study of distributed optimization problems over networks usually neglects the computational time needed by the nodes to perform local auxiliary operations. While small-scale optimization problems can usually be studied under enough time scale separation between communication and computation dynamics, for large-scale problems this approach is usually unfeasible. Thus, there is a critical need to design high-performance optimization algorithm that can be safely deployed over different types of infrastructures with practical computational limitations.

Motivated by the previous background, this work investigates the stability and performance properties associated to a class of distributed accelerated gradient algorithms implemented over networks with limited computational resources. In particular, we study the convergence properties of the interconnection between a class of recently proposed Accelerated Restarting Distributed Dynamics (HARDD) [5] and local computational blocks assigned to the nodes of the network to generate local estimates of the gradients.

2 PROBLEM STATEMENT

We study distributed accelerated gradient algorithms for the solution of multi-agent optimization problems. The algorithms are deployed on networks characterized by connected and undirected graphs $\mathcal{G} := (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \dots, n\}$ 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 smooth and strongly convex function $f_i : \mathbb{R}^p \rightarrow \mathbb{R}$, and the nodes cooperate with each other 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

$$\min_{z \in \mathbb{R}^{np}} F(z) := \sum_{i=1}^n f_i(z_i), \text{ s.t. } Lz = 0, \quad (1)$$

where $z = [z_1^\top, z_2^\top, \dots, z_n^\top]^\top \in \mathbb{R}^{np}$, $z_i \in \mathbb{R}^p$ is a local copy of the network state z for every node $i \in \mathcal{V}$, $L := \mathcal{L} \otimes I_p \in \mathbb{R}^{np} \times \mathbb{R}^{np}$ is called the interaction matrix, and $\mathcal{L}$ is the Laplacian matrix of the graph $\mathcal{G}$ that encodes the sparsity of the network.

Discrete-time and continuous-time approaches to solve problem (1) have been extensively studied using gradient descent and Newton-based [3], primal-dual [1], and projected dynamics [4]. However, a persistent challenge in the solution of (1) is to achieve fast rates of convergence without sacrificing essential robustness properties of the algorithms. To address this challenge, in [5], the authors introduced a class of distributed restarting-based accelerated dynamics using the formalism of hybrid dynamical systems. The algorithm is inspired by Nesterov's ODE [8], which induces suitable acceleration properties for smooth convex cost functions, but which, as shown in [6], may suffer from instability under small disturbances. By using the formalism presented in [5], the optimization problem (1) can be studied in the dual domain by leveraging the so-called *dual-friendly* functions [9], i.e., functions for which explicit minimization of their Legendre-Fenchel conjugate can be computed. Moreover, the computation of the gradient of the Lagrangian dual function can be carried out locally in each node of the network. However, in order to establish suitable acceleration properties while maintaining stability, the time taken by such computations was ignored in the analysis of the HARDD algorithm of [5]. This work builds on the preliminary results presented in [5], and aims to incorporate the computational time needed by the nodes of the network to obtain an ε -approximation of the local gradients. In particular, we model the interconnection between the accelerated gradient dynamics and the local computational mechanisms as sampled-data systems with sampling times related to the iterations needed by the nodes to obtain suitable approximations of the gradients. By leveraging this formalism, as well as Lyapunov-based tools for hybrid dynamical systems, we present the

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike International 4.0 License.
2021 CPS-IoT Week, May 18-21, 2021,
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8399-8/21/05.
<https://doi.org/10.1145/3457335.3461710>

Figure 1: The interconnection with the gradient computers is studied as a sampled-dynamical system where the result of the computation is sampled periodically and fed back into the distributed gradient system

first characterization of the minimum sampling time needed by the distributed accelerated optimization dynamics to preserve stability and to guarantee convergence to the solution of problem (1).

3 MAIN RESULTS

The main ideas behind the analysis are illustrated in Figure 1. In particular, as shown in the figure, we assign an output state $\eta = [\eta_1, \dots, \eta_i, \dots, \eta_n]$ to the Gradient Computational Block, where η_i converges in finite-time to an ε -approximation of the true gradient of the dual cost function, and it is computed by the i^{th} node of the network, i.e., for $\varepsilon > 0$ there exists $J_i \in \mathbb{N}$ such that $\eta_i(J_i)$ satisfies $|\eta_i(J_i) - h_i(Lx)| < \varepsilon$, with $h(Lx) := \arg \max_{z \in \mathbb{R}^{np}} \{ \langle Lx, z \rangle - F(z) \}$. The state η serves as input to the accelerated optimization dynamics of the nodes, which are given (in vector form) by

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

where $\alpha \in \{0, 1\}$, and $\Psi(\tau, x) := LD(\tau \otimes \mathbf{1}_p)\eta$. The logic state α is used to model “active” and “in-active” modes, which are assigned depending on the current state of the Gradient Computational Block. We model the interconnected system as a sampled-dynamical system where the modified HARDD algorithm now plays the role of the plant, and the local Gradient Computers plays the role of the controller. The resulting closed-loop system combines continuous-time dynamics and discrete-time dynamics, and therefore it is naturally modeled as a set-valued hybrid dynamical system [2]. For this closed-loop system, suitable stability, convergence, and robustness properties can be established by using Lyapunov functions for accelerated gradient flows, and the hybrid invariance principle. To do so, we leverage the properties of the Lyapunov function used in [5] to characterize robustness margins for the nominal HARDD and formulate the modified HARDD as a perturbed version of the nominal dynamics. Using these margins, we derive graph-dependent conditions on the periodicity of α that guarantees the stability of the interconnected system, and establish two main results:

Main Result 1: No matter how small is the sampling period, the distributed version of Nesterov’s ODE is unstable whenever it is implemented in closed-loop with a computational block that generates evaluations of the gradient every T seconds; see Fig. 1.

Main Result 2: There exists a $T^* > 0$ such that for each sampling time $T \in (0, T^*)$, the closed-loop sampled-data system composed by the modified HARDD and the Gradient Computational blocks is stable, robust, and the closed-loop system solves Problem (1).

Finally, we note that the implementation of continuous-time gradient flows to model the distributed gradient-based optimization dynamics is motivated by the availability of general theoretical tools for the stability and robustness analysis of continuous-time and sampled-data systems. However, discrete-time results can be obtained by using discretization mechanisms with sufficiently small step-sizes, which have been shown to preserve the convergence properties of well-posed hybrid dynamical systems (in a semi-global practical sense); see [7, Thm. 2].

4 CONCLUSIONS

In this work, we study for the first time the robustness analysis of distributed accelerated gradient flows with respect to computational limitations in Oracles that generate evaluations of the gradients. We first show that standard accelerated ODEs with vanishing damping have zero-margins of robustness with respect to arbitrarily small gradient errors. This observation motivates the development of alternative computation-aware algorithms for which suitable robustness properties can be established under a general class of Gradient Computational Models. The alternative algorithms are synthesized and analyzed using tools from hybrid dynamical systems theory and graph theory. Future research directions include the exploration of an adaptive sampling time that takes into account local computational limitations, as well as asynchronous sampling of the local gradients.

REFERENCES

- [1] J. Cortes and S. K. Niederlander. 2019. Distributed Coordination for Nonsmooth Convex Optimization via Saddle-Point Dynamics. *Journal of Nonlinear Science* 29 (2019), 1247–1272.
- [2] R. Goebel, R. G. Sanfelice, and A. R. Teel. 2012. *Hybrid Dynamical Systems: Modeling, Stability, and Robustness*. Princeton University Press.
- [3] A. Mokhtari, Q. Ling, and A. Ribeiro. 2017. Network Newton Distributed Optimization Methods. *IEEE Transactions on Signal Processing* 65 (2017), 146–161.
- [4] A. Nedic, A. Ozdaglar, and P. Parrilo. 2010. Constrained Consensus and Optimization in Multi-Agent Networks. *IEEE Trans. Autom. Contr.* 55 (2010), 922–938. Issue 4.
- [5] Daniel E Ochoa, Jorge I Poveda, César A Uribe, and Nicanor Quijano. 2020. Robust optimization over networks using distributed restarting of accelerated dynamics. *IEEE Control Systems Letters* 5, 1 (2020), 301–306.
- [6] Jorge I Poveda and Na Li. 2019. Inducing uniform asymptotic stability in non-autonomous accelerated optimization dynamics via hybrid regularization. In *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 3000–3005.
- [7] Ricardo G Sanfelice and Andrew R Teel. 2010. Dynamical properties of hybrid systems simulators. *Automatica* 46, 2 (2010), 239–248.
- [8] W. Su, S. Boyd, and E. Candes. 2016. A differential equation for modeling Nesterov’s accelerated gradient method: Theory and Insights. *Journal of Machine Learning Research* 17, 153 (2016), 1–43.
- [9] César A Uribe, Soomin Lee, Alexander Gasnikov, and Angelia Nedić. 2020. A dual approach for optimal algorithms in distributed optimization over networks. *Optimization Methods and Software* (2020), 1–40.
- [10] Jiafu Wan, Hehua Yan, Hui Suo, and Fang Li. 2011. Advances in cyber-physical systems research. *KSII Transactions on Internet and Information Systems (TIIS)* 5, 11 (2011), 1891–1908.