# Towards Computation-Aware Distributed Optimization over Networks

**Daniel E. Ochoa** [1] **Jorge I. Poveda** [1] **César A. Uribe** [2]

## Abstract

In this workshop, we extend on our work presented in (Ochoa et al., 2021). Specifically, we study tools for the analysis of robustness certificates for computationally-aware dual-based distributed optimization algorithms over networks. Contrary to existing literature, we follow a hybrid dynamical systems approach to analyze the stability properties of the distributed Nesterov's ODE, by explicitly taking into account the computational resources and time required by a dual first-order oracle to generates an approximate gradient. We show that, in such scenario, the distributed Nesterov's ODE can be rendered unstable under arbitrarily small disturbances, i.e., there exist an arbitrarily bounded perturbation function for which the inexact Oracle drives the system unstable. Moreover, we study modified dynamics that are provable stable and robust, and which also minimize smooth and strongly convex functions with suitable acceleration properties.

## 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. The application of CPS spans several domains, from autonomous driving and intelligent transportation systems to deep-space exploration, quantum computing, and machine learning. 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.

[1]Department of Electrical, Computer and Energy Engineering, University of Colorado Boulder, Colorado, USA [2]Department of Electrical and Computer Engineering, Rice University, Texas, USA. Correspondence to: Daniel E. Ochoa <daniel.ochoa@colorado.edu>.

*Proceedings of the LatinX in AI (LXAI) Research workshop at ICML 2021*. Copyright 2021 by the author(s).

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 algorithms that can be safely deployed over different types of infrastructures with practical computational limitations.

Motivated by the previous background, in this workshop we investigate the stability and performance properties of a certain 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) (Ochoa et al., 2020) and local computational blocks assigned to the nodes of the network to generate 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 over networks 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 smooth and strongly convex function $f_i:\mathbb{R}^p:\to\mathbb{R}$, 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

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

where $\mathbf{z} = [z_1^\top, z_2^\top, \cdots, z_n^\top]^\top \in \mathbb{R}^{np}$, $z_i \in \mathbb{R}$ is a local estimate of the network state $z$ for every node $i\in\mathcal{V}$, $\mathbf{L}:=\mathcal{L}\otimes I_p\in\mathbb{R}^{np}\times\mathbb{R}^{np}$ 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 methods (Mokhtari et al., 2017), primal-dual dynamics (Cortes & Niederlander, 2019), and projected dynamics (Nedic et al., 2010). 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 (Ochoa et al., 2020), 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 (Su et al., 2016), which induces suitable acceleration properties for smooth convex cost functions, but which, as shown in (Poveda & Li, 2019), may suffer from instability under small disturbances. By using the formalism presented in (Ochoa et al., 2020; 2021), the optimization Problem (1) can be studied in the dual domain by leveraging the so-called *dual-friendly* functions (Uribe et al., 2020), 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 (Ochoa et al., 2020).

This work builds on the preliminary results presented in (Ochoa et al., 2020), and aims to incorporate the computational time needed by the nodes of the network to obtain an $\varepsilon$-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 study 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 $\boldsymbol{\eta}{=}[\eta_1,\ldots,\eta_i,\ldots,\eta_n]$ to the Gradient Computational Block, where $\eta_i$ is computed by the $i^{th}$ node of the network and is assumed to converge, in finite-time, to an $\varepsilon$-approximation of the true gradient of the dual cost function, and it , i.e., on compact sets, and for $\varepsilon{>}0$ there exists $J_i{\in}\mathbb{N}$ such that $\eta_i(J_i)$ satisfies $|\eta_i(J_i)-h_i(\mathbf{Lx})|{<}\varepsilon$, with

$$h(\mathbf{Lx}) := \operatorname*{arg\,max}_{\mathbf{z}\in\mathbb{R}^{np}} \left\{\langle \mathbf{Lx},\,\mathbf{z}\rangle - F(\mathbf{z})\right\}. \qquad (2)$$

The state $\eta$ serves as input to the accelerated optimization dynamics of the nodes, which are given (in vector form) by

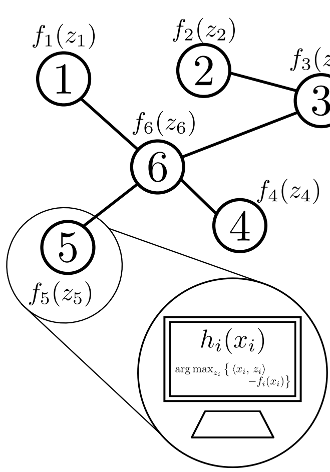
$$\dot{\mathbf{p}} = F_A(\mathbf{p}) := \alpha\cdot\begin{pmatrix} 2\mathbf{D}(\boldsymbol{\tau})^{-1}(\mathbf{y}-\mathbf{x}) \\ -2\gamma\Psi(\boldsymbol{\tau},\mathbf{x}) \\ \frac{1}{2}\mathbf{1}_n \end{pmatrix}, \qquad (3)$$

where $\alpha{\in}\{0,1\}$, and $\Psi(\boldsymbol{\tau},\mathbf{x}){:=}\mathbf{LD}(\boldsymbol{\tau}\otimes\mathbf{1}_p)\boldsymbol{\eta}$. The logic state $\alpha$ 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-data 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 (Goebel et al., 2012). 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 (Ochoa et al., 2020) to characterize robustness margins for the nominal HARDD and formulate the modified HARDD as a perturbed version of the nominal dynamics to establish two main results:

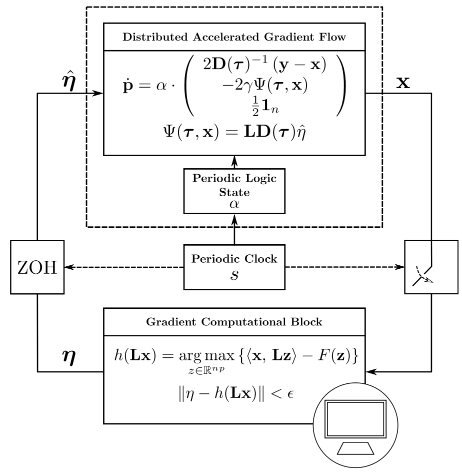
- **Main Result 1:** On compact sets, for every $\varepsilon>0$ and each sampling period $T>0$, there exists a bounded perturbation function satisfying $\|d_t(T)\|<\varepsilon$ such that the distributed Nesterov's ODE is unstable for an inexact dual first-order oracle $h(\mathbf{Lx}(t))+d_t(T)$, c.f. (2).
- **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 in (3) is stable, robust, and the closed-loop system solves Problem (1), modulo a small residual error.

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 (Sanfelice & Teel, 2010, Thm. 2).

**Sketch of the Proof of Main Result 1:** Consider a distributed version of Nesterov's ODE (Su et al., 2016) implemented in a sampled-data structure with an inexact Oracle having sampling time $T$. Let $t_k=t_0+kT$, be the sampling



(a) The distributed optimization problem. Each node of the network has a local smooth and strongly convex function $f_i$. The objective is to minimize the sum of the individual functions. Each agent has access to approximate dual oracles.



(b) The hybrid dynamical systems view of the distributed optimization problem over networks. The HARDD dynamics is a continuous dynamics systems with a feedback loop driven by an approximate sampled dual first-order oracle.

*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

times, where $k \in \mathbb{Z}_{\geq 0}$. The evolution in time of Nesterov's ODE with the inexact oracles, between times $(t_k, t_{k+1})$, is then given by

$$\dot{\mathbf{x}}_1(t) = \mathbf{x}_2(t) \tag{4a}$$

$$\dot{\mathbf{x}}_2(t) = -\frac{c}{\tau(t)}\mathbf{x}_2(t) - \gamma \mathbf{L}\eta(t_k) \tag{4b}$$

$$\dot{\tau} = 1, \tag{4c}$$

where $\eta(t_k) = h(\mathbf{L}\mathbf{x}(t_k)) + d_k(T)$. Suppose that he primal cost function is given by $F(z) = \sum_{i=1}^{N} |z_i|^2$. It follows that $h(\mathbf{L}\mathbf{x}_1) = \mathbf{L}\Gamma\mathbf{x}_1$ for some positive definite matrix $\Gamma$. Under these assumptions, system (4) is reduced to

$$\dot{\mathbf{x}}_1(t) = \mathbf{x}_2(t) \tag{5a}$$

$$\dot{\mathbf{x}}_2(t) = -\frac{c}{\tau(t)}\mathbf{x}_2(t) - \gamma \mathbf{L}^2\Gamma\mathbf{x}_1(t_k) + d_k(T). \tag{5b}$$

Suppose the solutions of (5) are defined for all $t > 0$, and that $\mathbf{x}(0,0) \in K$, with $K \subset \mathbb{R}^n$ a compact set. Additionally, assume that $\mathbf{x}_2$ remains uniformly bounded, otherwise there is nothing to prove. We claim that, the origin of System (5) is not uniformly attractive no matter how small we select $T > 0$. Indeed, let $\varepsilon > 0$ be given. Then, there exists $T^*$ such that for all $t \geq T^*$ the damping term in (5) satisfies

$$\left| \frac{c}{\tau(t)}\mathbf{x}_2(t) \right| \leq \varepsilon \tag{6}$$

for all $t \geq T^*$. It follows that System (5) behaves as a perturbed system of the form

$$\dot{\mathbf{x}}_1(t) = \mathbf{x}_2(t) \tag{7a}$$

$$\dot{\mathbf{x}}_2(t) \in -\gamma \mathbf{L}^2\Gamma\mathbf{x}_1(t) + d_K(T) + e(t), \tag{7b}$$

where $||e(t)|| \leq 2\varepsilon$. When $d_k(T) + 2\varepsilon = 0$, system (7) behaves as a **marginally** stable system. Since the system is marginally stable, no matter how small we select $e(t)$ the dynamics can be rendered unstable.

**Sketch of the Proof of Main Result 2:** We take inspiration from the work presented in (Sanfelice & Teel, 2006). Particularly, in order to account for the interconnection between the dual Oracles and the HARDD algorithm described in Figure 1b, we define the $\zeta := (x, \hat{\eta}, s)$, where $x$ is the state of the HARDD optimization dynamics, $\hat{\eta}$ represents the truncated computation of the local gradients, and $s \in [0, T]$ is the timer variable that conditions when the information is fed-back to the optimization dynamics. Moreover, we describe the hybrid dynamics of $\eta$ by using the system proposed in Example 2.14 of (Sanfelice, 2021), which corresponds to a sample-and-hold system connecting a hybrid plant, in this case the nominal HARDD, with a state-feedback controller, which in our setup corresponds to the state $\hat{\eta}$. Since uniform global asymptotic stability was proven in (Ochoa et al., 2020) for the set of minimizers of the cost function $F(\mathbf{z})$, and the HARDD dynamics are well-posed by design, existence of a smooth Lyapunov function $V : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is guaranteed for the nominal hybrid dynamics. Using this nominal Lyapunov function, we define

$$W(\zeta) := e^{\lambda_s s} V(x),$$

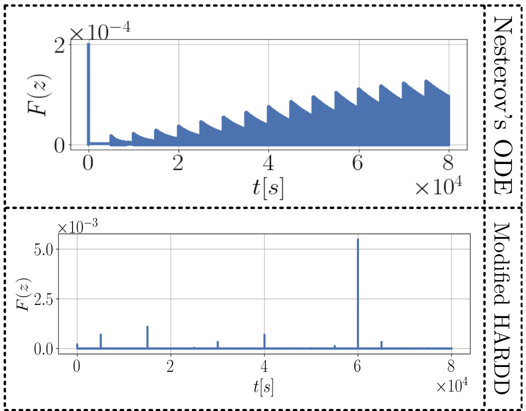



*Figure 2.* Instability of (4) under interconnection with inexact dual first-order oracle (top). The instability can be addressed by implementing a hybrid-regularized version of the system, and imposing a suitable sampling period $T^*$ (bottom).

as a Lyapunov function candidate for the interconnected sample-and-hold system, and leverage the properties of $V$ during the flows and jumps of the nominal HARDD dynamics. We prove that, for every compact set of initial conditions, there exists a sufficiently small sampling period $T$, for which the stability and convergence properties of the nominal dynamics are preserved in a practical sense as $T \to 0$.

## 4. Conclusions

In this workshop, we continue to study the robustness properties 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. Ongoing research includes the exploration of an adaptive sampling time that takes into account local computational limitations, as well as asynchronous sampling of the local gradients.

## References

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

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

Mokhtari, A., Ling, Q., and Ribeiro, A. Network Newton distributed optimization methods. *IEEE Transactions on Signal Processing*, 65:146–161, 2017.

Nedic, A., Ozdaglar, A., and Parrilo, P. Constrained consensus and optimization in multi-agent networks. *IEEE Trans. Autom. Contr.*, 55:922–938, 2010.

Ochoa, D. E., Poveda, J. I., Uribe, C. A., and Quijano, N. Robust optimization over networks using distributed restarting of accelerated dynamics. *IEEE Control Systems Letters*, 5(1):301–306, 2020.

Ochoa, D. E., Poveda, J. I., and Uribe, C. A. Computation-aware distributed optimization over networks: A hybrid dynamical systems approach. Workshop on Computation-Aware Algorithmic Design for Cyber-Physical Systems, May 2021.

Poveda, J. I. and Li, N. Inducing uniform asymptotic stability in non-autonomous accelerated optimization dynamics via hybrid regularization. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 3000–3005. IEEE, 2019.

Sanfelice, R. G. *Hybrid Feedback Control*. Princeton University Press, 2021.

Sanfelice, R. G. and Teel, A. R. Lyapunov analysis of sample-and-hold hybrid feedbacks. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pp. 4879–4884. IEEE, 2006.

Sanfelice, R. G. and Teel, A. R. Dynamical properties of hybrid systems simulators. *Automatica*, 46(2):239–248, 2010.

Su, W., Boyd, S., and Candes, E. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. *Journal of Machine Learning Research*, 17(153):1–43, 2016.

Uribe, C. A., Lee, S., Gasnikov, A., and Nedić, A. A dual approach for optimal algorithms in distributed optimization over networks. *Optimization Methods and Software*, pp. 1–40, 2020.