Robust Optimization over Networks Using Distributed Restarting of Accelerated Dynamics

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

Abstract

Motivated by the several engineering applications it has in areas such as machine learning, power, transportation, and water distribution systems, and distributed network control, see [1] and references therein, 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. Specifically, we consider the setting where each node i has a local function $f_i : \mathbb{R}^p :\to \mathbb{R}$, 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

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

which is also known in the literature as the *consensus-optimization* problem [2].

Discrete-time and continuous-time approaches to solve problem (1) have been extensively studied using gradient descent and Newton-based dynamics [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, 10, 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 matter 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 mentioned in [16], it remains an

*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.

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

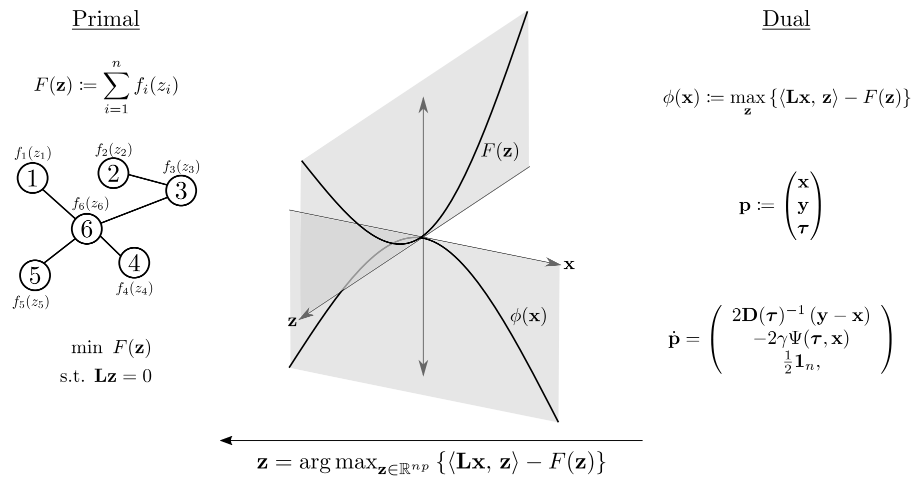

Figure 1: We design an algorithm to solve the consensus-optimization problem from a complete dual perspective, by constructing dynamics for the *dual* variable $\mathbf{x} \in \mathbb{R}^{np}$. The proposed accelerated dynamics make use of a momentum variable $\mathbf{y} \in \mathbb{R}^{np}$, and timers 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 $\mathcal{G}$.

open question whether or not similar techniques could be pursued for distributed optimization problems of the form (1). As we show in [17], the answer to this question turns out to be positive.

Hence, the main contribution of our 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 and discrete-time dynamics, they are modeled as set-valued hybrid dynamical systems [18], 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], see Figure 1. 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.

We highlight that besides the theoretical contribution to the field of distributed optimization, applications on different fronts of the applied and theoretical machine learning, resource allocation and robotics fields may benefit by the usage of novel distributed and accelerated dynamics, see [19, 20, 21, 22, 23] and references therein. For example, they can suitably be implemented for the solution of learning problems involving large amounts of data, which have been proven to benefit from the flexibilities that the distributed setup brings to the table [24, 25, 26, 27].

References

[1] 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.

[2] 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.

[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," *American Control Conference, to appear*, 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] D. E. Ochoa, J. I. Poveda, C. A. Uribe, and N. Quijano, "Robust optimization over networks using distributed restarting of accelerated dynamics," *IEEE Control Systems Letters*, vol. 5, no. 1, pp. 301–306, 2020.

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

[19] J. Konečnỳ, B. McMahan, and D. Ramage, "Federated optimization: Distributed optimization beyond the datacenter," *arXiv preprint arXiv:1511.03575*, 2015.

[20] T. Kraska, A. Talwalkar, J. C. Duchi, R. Griffith, M. J. Franklin, and M. I. Jordan, "Mlbase: A distributed machine-learning system.," in *Cidr*, vol. 1, pp. 2–1, 2013.

[21] L. Xiao and S. Boyd, "Optimal scaling of a gradient method for distributed resource allocation," *Journal of optimization theory and applications*, vol. 129, no. 3, pp. 469–488, 2006.

[22] M. G. Rabbat and R. D. Nowak, "Decentralized source localization and tracking [wireless sensor networks]," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. iii–921, IEEE, 2004.

[23] A. Nedić, A. Olshevsky, and C. A. Uribe, "Distributed learning for cooperative inference," *arXiv preprint arXiv:1704.02718*, 2017.

[24] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

[25] A. Nedich, A. Olshevsky, and W. Shi, "A geometrically convergent method for distributed optimization over time-varying graphs," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 1023–1029, IEEE, 2016.

[26] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.

[27] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*, pp. 177–186, Springer, 2010.