Asynchronous Decentralized Optimization with Constraints: Achievable Speeds of Convergence for Directed Graphs

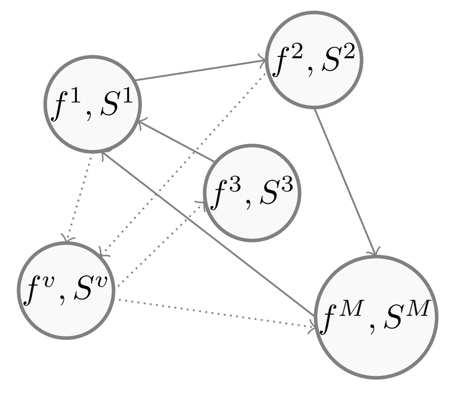
Firooz Shahriari-mehr and Ashkan Panahi
Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg

Abstract

We address decentralized convex optimization problems, where every agent has its local objective function and constraint set. Agents compute at different speeds, and their communication may be delayed and directed. We propose a novel algorithm handling difficult scenarios such as message failures, by employing local buffers. We guarantee the convergence speed of our algorithm using linear quadratic performance estimation problems. This approach simplifies the analysis of smooth convex optimization problems, going beyond Lyapunov function analysis and avoiding restrictive assumptions such as strong-convexity.

1. Asynchronous Decentralized Constrained Optimization

Decentralized setup:



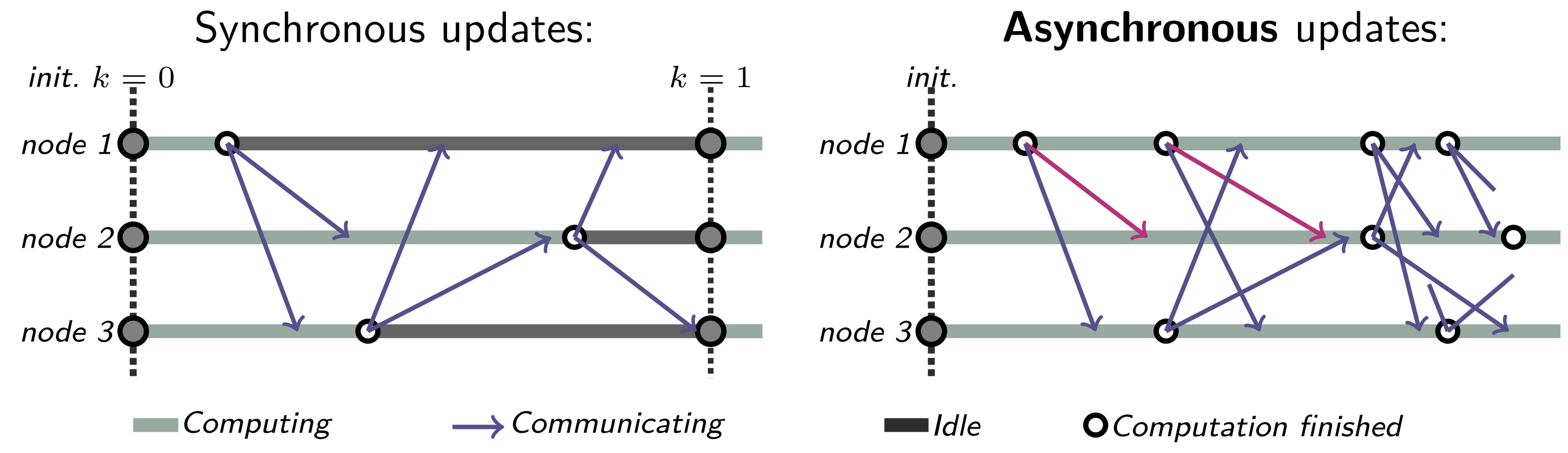
Constrained optimization:

$$\min_{\mathbf{x}\in\mathbb{R}^m} \frac{1}{M}\sum_{v=1}^{M} f^v(\mathbf{x})$$

$$\text{s.t.} \quad \mathbf{x} \in \bigcap_{v=1}^{M} S^v$$

Synchronous updates:

Asynchronous updates:



Goal: $\mathbf{0} \in \sum_{v=1}^{M}\left(\partial I_{S^v}(\mathbf{x}^*) + \nabla f^v(\mathbf{x}^*)\right)$

2. Asynchronous Double Averaging and Gradient Projection

DAGP[1] is modified for asynchronous setup since it is the first algorithm addressing local constraints on directed graphs, with a fixed-step size.

Variables $(\mathbf{x}^u, \mathbf{p}^u)$ are transmitted across the network. Due to asynchrony, a node v may receive none, one, or multiple messages from u. Hence, each node holds a distinct local buffer $\mathcal{B}^{vu}$ for every neighbor, storing received messages. These messages are utilized at the start of a computation round to compute the estimates $(\mathbf{a}^{vu}, \mathbf{b}^{vu})$, then the buffer is cleared. Estimates are updated as:

$$\mathbf{a}^{vu} = \frac{1}{|\mathcal{B}^{vu}|}\sum_{\mathbf{x}^u\in\mathcal{B}^{vu}} \mathbf{x}^u \qquad \mathbf{b}^{vu} = \frac{1}{|\mathcal{B}^{vu}|}\sum_{\mathbf{p}^u\in\mathcal{B}^{vu}} \mathbf{p}^u$$

Then, ASY-DAGP updates its variables close to DAGP as:

$$\mathbf{z}^v = \boxed{\mathbf{x}^v - \sum_{u\in\mathcal{N}_{\text{in}}^v} w_{vu}\mathbf{a}^{vu}} - \boxed{\mu\left(\nabla f^v(\mathbf{x}^v) - \mathbf{g}^v\right)} \qquad \mathbf{x}^v = \boxed{P_{S^v}\left(\mathbf{z}^v\right)}$$

$$\mathbf{g}^v = \boxed{\mathbf{g}^v + \rho\left[\nabla f^v(\mathbf{x}^v) - \mathbf{g}^v + \frac{1}{\mu}(\mathbf{z}^v - \mathbf{x}^v)\right]} + \boxed{\alpha\left(\mathbf{h}^v - \mathbf{g}^v\right)}$$

$$\mathbf{h}^v = \boxed{\boxed{\gamma}\mathbf{h}^v - \sum_{u\in\mathcal{N}_{\text{in}}^v} \boxed{q_{vu}}\mathbf{b}^{vu}} \qquad \mathbf{p}^v = \boxed{\mathbf{p}^v - \eta\sum_{u\in\mathcal{N}_{\text{in}}^v} q_{vu}\mathbf{b}^{vu} + \eta(\gamma-1)\mathbf{g}^v}$$

ASY-DAGP concepts: Local buffers, Weighted averaging, Projection, Constrained gradient tracking, and Modified Distributed null projection.

3. LQ-PEP: A new convergence analysis tool

Performance Estimation Problem (PEP) evaluates the worst-case performance of optimization algorithms after K iterations [5]:

$$\max_{f,\mathbf{x}_*,\{\mathbf{x}_k\}_{k=0}^{K-1}} \Phi(\mathbf{x}_{K-1}, \mathbf{x}_*) \quad \text{s.t.} \begin{cases} f \text{ is a function from a family } \mathcal{F}, \\ \mathbf{x}_* \text{ is the minimizer of } f, \\ \text{Algorithm generates } \mathbf{x}_1, \ldots, \mathbf{x}_{K-1}, \\ \|\mathbf{x}_0 - \mathbf{x}_*\| \le R. \end{cases}$$

Linear Quadratic (LQ-)PEP is a relaxation of PEP with a quadratic objective function and linear constraints. It is achieved by

- Considering the class of linear optimization algorithms.
- Finding an upper bound for the performance measure function Φ.

$$\min_{\{\mathbf{\Psi}_k\}_{k=0}^{K-1}} \sum_{k=0}^{K-1}\langle \mathbf{\Psi}_k, \mathbf{S}\mathbf{\Psi}_k\rangle \quad \text{s.t.} \quad \mathbf{\Psi}_{k+1} - \bar{\mathbf{R}}\mathbf{\Psi}_k - \mathbf{P}\mathbf{X}_{k+2} = \sum_{l=0}^{K-2}\tilde{\mathbf{R}}_{k,k+1-l}\mathbf{\Psi}_l.$$

LQ-PEP bridges the gap between PEP and Lyapunov-based analyses.

4. Convergence Results by LQ-PEP

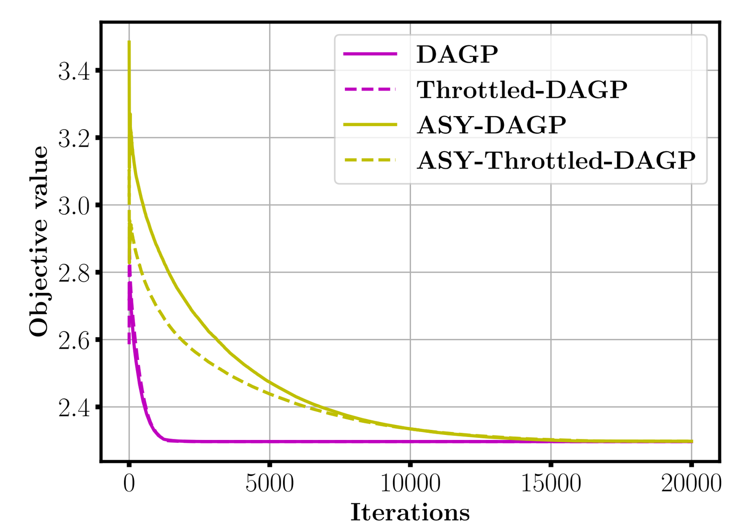
1. We establish that ASY-DAGP achieves an ϵ gap in consensus, optimality and feasibility in $O(1/\epsilon^2)$ iterations, under fairly general asynchrony and delays. We require no restrictive assumption such as strong convexity.

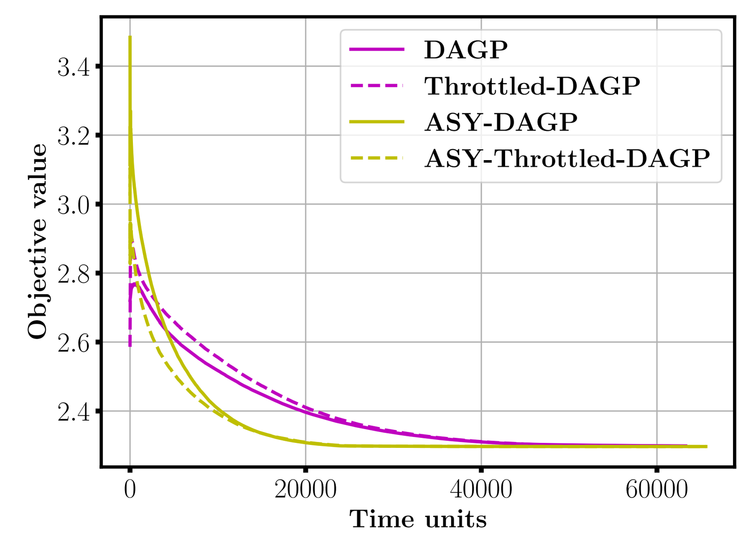
2. We study an arbitrary delay pattern. We calculate a quantity named delay factor for any possible delay pattern. We can guarantee convergence only if the delay factor is finite (and the gossip matrices are scaled accordingly).

$$\sum_v \text{dist}^2(\bar{\mathbf{x}}_K, S_v) \le \sum_\nu \|\bar{\mathbf{x}}_K^v - \bar{\mathbf{x}}_K\|_2^2 = O\left(\frac{\kappa C_0}{\zeta K}\right)$$

$$\sum_v f^v(\bar{\mathbf{x}}_K^v) - \sum_v f^v(\mathbf{x}^*) = O\left(\frac{C_0}{\mu K} + \sqrt{\frac{\kappa C_0 C_1}{\zeta K}}\right)$$

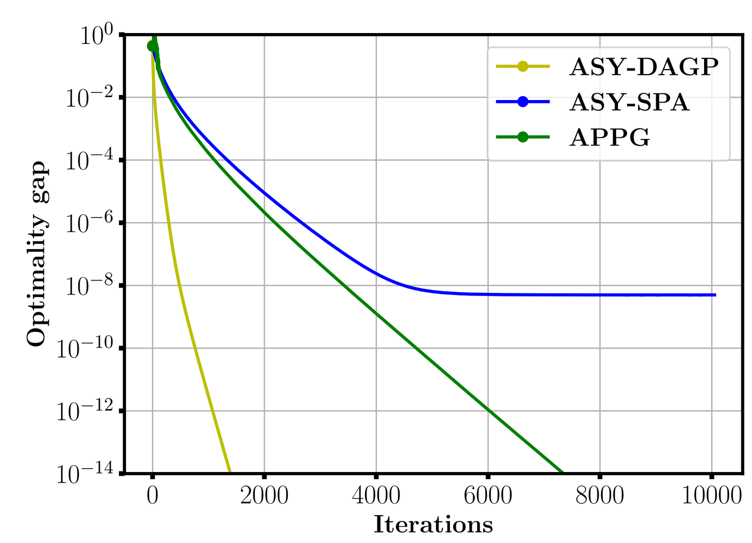
5. Experimental Results



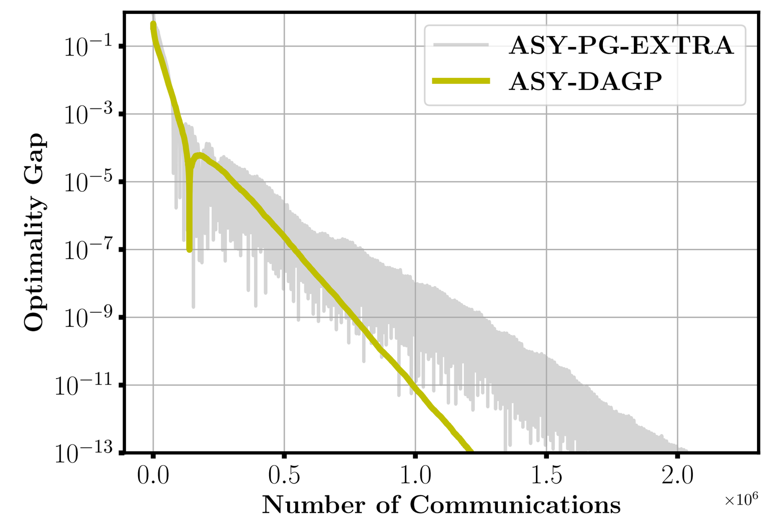


Our constrained problem: compare to DAGP and its throttled version.

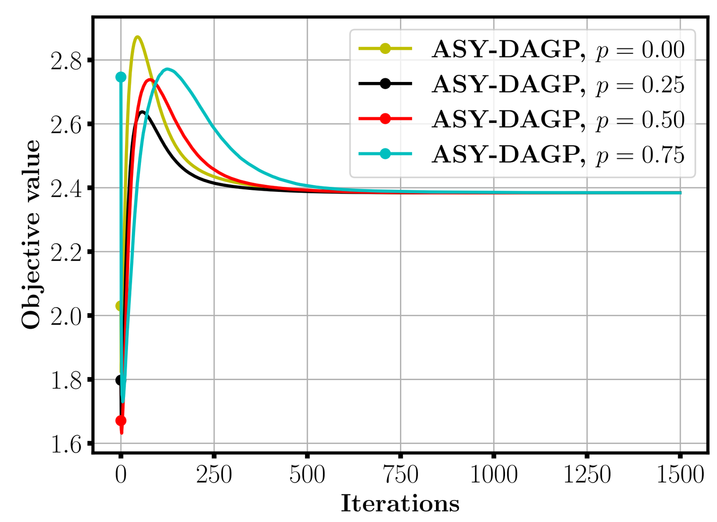
$$f^v(\mathbf{x}) = \log\left(\cosh(\mathbf{a}_v^T\mathbf{x} - b_v)\right), \quad S^v = \{\mathbf{x}\in\mathbb{R}^{10} \mid \mathbf{c}_v^T\mathbf{x} \le d_v\}. \quad v = 1, \ldots, 20$$



Unconstrained Logistic Regression: compare to APPG[2] and ASY-SPA[3].



Our constrained problem, undirected graphs: compare to ASY-PG-EXTRA[4].



Robustness to message losses with the communication failure probability p.

References

[1] Decentralized Constrained Optimization: Double Averaging and Gradient Projection, F. Shahriari-Mehr, D. Bosch, A. Panahi, Conference on Decision and Control, Dec. 2021.

[2] Fully asynchronous distributed optimization with linear convergence in directed networks, J. Zhang, K. You, arXiv preprint:1901.08215, 2019

[3] AsySPA: An Exact Asynchronous Algorithm for Convex Optimization Over Digraphs, J. Zhang, K. You, IEEE Transactions on Automatic Control, vol. 65, June 2020.

[4] Decentralized consensus optimization with asynchrony and delays, T. Wu, K. Yuan, L. Qing, A.H. Sayed, IEEE Transactions on Signal and Information Processing over Networks, April 2017.

[5] Convex interpolation and performance estimation of first-order methods, Adrien Taylor's Dissertation, Catholic University of Louvain, 2017.

Download full paper here:

SCAN ME

WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM