A Size Independent Formulation for Automatic Performance Evaluation of Decentralized Optimization Algorithms

Sebastien Colla and Julien M. Hendrickx
ICTEAM Institute, UCLouvain, 1348 Louvain-la-Neuve, Belgium.
Email: { sebastien.colla, julien.hendrickx } @uclouvain.be

We develop a methodology that automatically provides numerically tight performance bounds for first-order decentralized methods, by solving a semidefinite program whose size is independent of the number of agents in the network [1].

In decentralized optimization, we consider a set of agents $\{1,\dots,N\}$, working together to minimize the average of their local functions $f(x) = \frac{1}{N}\sum_{i=1}^{N} f_i(x)$. This problem can be written using a separable function $F_s : \mathbb{R}^{Nd} \to \mathbb{R}$:

$$\min_{x_1\dots,x_N \in \mathbb{R}^d} \quad F_s(x_1,\dots,x_N) = \frac{1}{N}\sum_{i=1}^{N} f_i(x_i), \qquad (1)$$
$$\text{such that} \quad x_1 = \dots = x_N.$$

where $f_i : \mathbb{R}^d \to \mathbb{R}$ is the private function locally held by agent i and $x_i \in \mathbb{R}^d$ is its local decision variable. Each agent i performs local computations and exchanges local information with its neighbors to come to an agreement on the minimizer x^* of the global function f. Exchanges of information often take the form of an average consensus on some quantity, e.g., on the x_i. These consensuses can be represented using a multiplication by a matrix $W \in \mathbb{R}^{N\times N}$, typically assumed symmetric and doubly stochastic and characterized by its second largest eigenvalue λ.

In general, the quality of an optimization algorithm is evaluated *via* a worst-case guarantee. Accurate worst-case guarantees can be difficult to compute theoretically. Recently, an alternative computational approach, called Performance Estimation Problem (PEP), has been developed to automatically compute tight worst-case guarantees by solving an optimization problem [2]. A PEP is looking for the function and the initial point maximizing a given performance criterion when running K iterations of the algorithm. In our previous work [3], we have extended this PEP framework to decentralized optimization, resulting in problems whose size grows with the number of agents N. In our new work [1], we develop a new PEP formulation for decentralized optimization whose problem size is now independent of N, though N could still appear as a scaling coefficient in some cases. This allows to obtain worst-case guarantees valid for any number of agents N by solving compact problems.

This new PEP formulation takes a global view on the decentralized problem (1), by neglecting the separability of the function F_s, which a priori corresponds to a relaxation. The problem becomes then one of optimizing a function $F : \mathbb{R}^{Nd} \to \mathbb{R}$ over $\mathbf{x} \in \mathbb{R}^{Nd}$, under the constraint that $\mathbf{x}$ belongs to the consensus subspace (all $x_i \in \mathbb{R}^d$ are equal). This will admit a simple representation that only requires decomposing $\mathbf{x}$ into two blocks of components, respectively along the consensus subspace and along its orthogonal complement, as opposed to the N blocks of components along each x_i in [3]. See [1] for details about this new formulation.

We have demonstrated this new agent-independent PEP formulation on several examples (algorithms DGD, DIGing and EXTRA), and observe empirically its tightness for symmetric and generalized[1] doubly stochastic averaging matrices [1]. This tool allows to answer rapidly to a large diversity of questions about the performance of decentralized algorithms. For example, it can help in algorithm comparison: Figure 1 shows that EXTRA seems to perform better than DIGing for time-constant averaging matrices but is not robust to time-varying matrices.

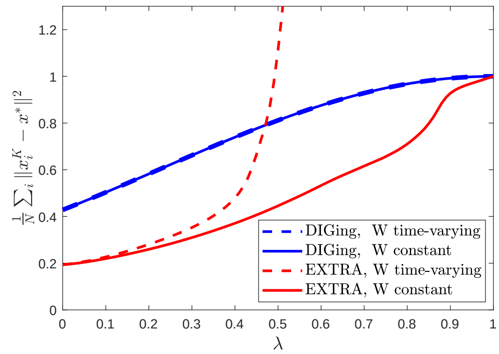

Figure 1: Comparison between algorithms DIGing and EXTRA for 10 iterations with constant and time-varying averaging matrices W, for different values of λ. Optimized step-sizes and f_i are 0.1-strongly convex and 1-smooth, for all i.

Acknowledgements

S. Colla is supported by the French Community of Belgium through a FRIA fellowship (F.R.S.-FNRS).

References

[1] S. Colla, J. M. Hendrickx, "Automated Performance Estimation for Decentralized Optimization via Network Size Independent Problems", *61th IEEE CDC*, 2022.

[2] A. B. Taylor, J. M. Hendrickx and F. Glineur, "Exact Worst Case Performance of First-Order Methods for Composite Convex Optimization," 2015.

[3] S. Colla, J. M. Hendrickx, "Automatic Performance Estimation for Decentralized Optimization", *Preprint*, 2022.

[1] A generalized doubly stochastic matrix has rows and columns that sum to one, without the need to be non-negative, see [3, Definition 2]