Performance Estimation Problem for Decentralized Optimization Methods

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

1 Introduction

We develop a methodology that automatically provides nearly tight performance bounds for first-order decentralized methods on convex functions and we demonstrate its usefulness on different existing methods [1]. Decentralized optimization has received an increasing attention due to its useful applications in large-scale machine learning and sensor networks, see e.g. [2] for a survey. In decentralized methods for separable objective functions, we consider a set of agents $\{1,\dots,N\}$, working together to solve the following optimization problem:

$$\underset{x\in\mathbb{R}^d}{\text{minimize}} \quad f(x) = \frac{1}{N}\sum_{i=1}^{N} f_i(x),$$

where $f_i : \mathbb{R}^d \to \mathbb{R}$ is the private function locally held by agent i. Each agent i holds its own version x_i of the decision variable x, performs local computations and exchanges local information with its neighbors to come 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.

One of the simplest decentralized optimization method is the distributed (sub)gradient descent (DGD) [2] where agents successively perform an average consensus step (1) and a local gradient step (2). We have, for all $i \in \{1,\dots,N\}$,

$$y_i^k = \sum_{j=1}^{N} w_{ij} x_j^k; \tag{1}$$

$$x_i^{k+1} = y_i^k - \alpha \nabla f_i(x_i^k), \tag{2}$$

where $\alpha > 0$ is a constant step-size.

2 Contributions and results

In general, the quality of an optimization method is evaluated *via* a worst-case guarantee. Obtaining theoretical worst-case performance bounds for decentralized algorithms can often be a challenging task, requiring combining the impact of the optimization component and the interconnection network. This can result in performance bounds that are complex and not very tight. For example, we have empirically shown that the existing performance bounds for DGD are significantly worse than the actual worst-cases. However, accurate performance bounds are important to correctly understand the impact of the network topology and algorithm parameters on the performance of the algorithm.

In this work, we follow an alternative computational approach that finds a worst-case performance guarantee of an algorithm by solving an optimization problem, known as the performance estimation problem (PEP). The PEP approach has led to many results in centralized optimization, see e.g. [3], but it has never been applied to decentralized optimization methods. The current PEP framework lacks for ways of representing the communications between the agents. We therefore propose two formulations of the average consensus steps that can be embedded in a solvable PEP [1].

The first formulation uses a given averaging matrix W to directly incorporate the updates of the chosen method as constraints over the iterates. This leads to performance bounds that are tight, but specific to the given matrix.
The second formulation is a relaxation that considers entire spectral classes of possible symmetric matrices. This allows PEP to provide spectral upper bounds on the performance that are valid over an entire class of networks and can thus be compared with the bounds of the literature. This also allows looking for the worst communication network from the given class. This formulation is a relaxation because it replaces the set of consensus steps (e.g. (1)) with a set of constraints that are only proven to be necessary for having an averaging matrix in the given class. Although it is a relaxation, the performance guarantees it provides for the decentralized algorithms we have experimented, such as DGD and DIGing, significantly improve on the theoretical existing ones and are numerically tight. They are also independent of the number of agents in the problem and they help for better tuning of the parameters of the algorithms.

Using our two new formulations, the PEP approach can be applied directly to many decentralized algorithms.

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, “Automatic Performance Estimation for Decentralized Optimization”, *Preprint*, 2022.

[2] A. Nedić, A. Olshevsky and M. G. Rabbat, “Network Topology and Communication-Computation Tradeoffs in Decentralized Optimization,” 2018.

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