Automated Worst-Case Performance Analysis of Decentralized Optimization Methods

Sebastien Colla and Julien M. Hendrickx
ICTEAM, 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 primal first-order decentralized optimization methods on convex functions [3].

Decentralized optimization has received an increasing attention due to its useful applications in large-scale machine learning and sensor networks, see survey [1]. In decentralized methods for separable objective functions, we consider a set of agents $\{1,\dots,N\}$, working together to minimize $f(x) = \frac{1}{N}\sum_{i=1}^{N} f_i(x)$, where f_i is the private local function 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 seek to reach agreement on the minimizer x^* of the global function f. In the system, the exchanges of information often take the form of a multiplication by a given matrix $W \in \mathbb{R}^{N\times N}$, typically assumed symmetric and doubly-stochastic.

A classical way for evaluating the quality of an optimization method is to obtain worst-case guarantees regarding a chosen quality metric. Obtaining theoretical worst-case performance bounds for decentralized algorithms can often be a challenging task, requiring to combine the effects of the optimization and of the interconnection network. This can result in performance bounds that are not very tight.

In this work, we follow an alternative computational approach that finds the worst-case performance of an algorithm by solving an optimization problem, whose variables are the iterates, functions values and (sub)gradients. This is known as the performance estimation problem - PEP - and has been studied for centralized fixed step first-order methods in [2]. Current PEP theory does not allow matrix multiplications in the methods it analyzes, which are needed for representing communications in decentralized algorithms.

2 Main results

When the communication matrix W is given *a priori*, the performance estimation problem (PEP) can be directly formulated as an SDP, in the same way as presented in [2]. This formulation provides the exact worst-case performance for the given decentralized method and the specific communication matrix given *a priori*. This can be useful for trying different communication matrices and observing their impact on the performance of the algorithm. However, we would like to obtain more general results, valid over entire classes of communication matrices.

We now consider that the matrix W is not *a priori* given, but is one of the decision variables of PEP. As constraints, we impose W to belong to the commonly used set of symmetric, doubly-stochastic matrices with a given range of eigenvalues $[-\lambda,\lambda]$. We also impose this matrix W to represent the different interactions that occur in the given decentralized algorithm. We consider only algorithms in which an interaction k takes the form of a matrix-vector multiplication $y^k = Wx^k$. The vector x^k regroups the N local variables that are subject to a communication, typically the iterates x_i^k of each agent.We suppose the same matrix W is used for each interaction k, then we have $Y = WX$, where $Y = [x^k]$ and $X = [x^k]$, for $k = 1,\dots,K$.

We do not have a direct way for representing these constraints into PEP but we construct a relaxation, that is often close to tight. For that, from these constraints, we derive the following new necessary conditions involving only variables X and Y and allowing to remove W from the problem:

$$\overline{X} = \overline{Y} \qquad \text{and} \qquad Y_\perp^T Y_\perp \preceq \lambda^2 X_\perp^T X_\perp,$$
$$\text{with} \qquad Y_\perp = Y - \mathbf{1}\overline{Y}^T \qquad X_\perp = X - \mathbf{1}\overline{X}^T,$$

where $\overline{X}$ and $\overline{Y}$ denote the vectors with the agents average of X and Y. In short, these new constraints says the agents average is preserved during a communication, while the variance is reduced by a factor at least λ^2. See [3] for details.

We apply our new PEP formulation to the decentralized gradient descent (DGD) and obtain worst-case performance bounds that largely improve the one from [1]. For DGD, by comparing the second relaxed PEP formulation with the first exact one, it appears to be tight if we allow the communication matrix to have negative elements. See [3] for details.

Acknowledgements

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

References

[1] A. Nedić, A. Olshevsky and M. G. Rabbat, "Network Topology and Communication-Computation Tradeoffs in Decentralized Optimization,"textitProceedings of the IEEE, vol. 106, no. 5, pp. 953-976, 2018.

[2] A. B. Taylor, J. M. Hendrickx and F. Glineur, "Exact Worst-Case Performance of First-Order Methods for Composite Convex Optimization," *SIAM Journal on Optimization*, vol.27, 2015.

[3] S. Colla, J. M. Hendrickx, "Automated Worst-Case Performance Analysis of Decentralized Gradient Descent", preprint, 2021.