DECENTRALIZED OPTIMIZATION ON TIME-VARYING DIRECTED GRAPHS UNDER COMMUNICATION CONSTRAINTS

#4252

The University of Texas at Austin
Electrical and Computer Engineering
Cockrell School of Engineering

Yiyue Chen, Abolfazl Hashemi and Haris Vikalo

Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX, USA

TEXAS
The University of Texas at Austin

BACKGROUND

- Decentralized optimization problems: all clients in the network to collaboratively learn the model via communication

- Potential issues on privacy, unreliable communication and resource constraint

PROBLEM FORMULATION

- Decentralized problems over directed and time-varying networks:

$$\min_{\mathbf{x}\in\mathbb{R}^d}\left[f(\mathbf{x}) := \frac{1}{n}\sum_{i=1}^{n} f_i(\mathbf{x})\right]$$

- The agents collaborate to solve the problem by exchanging information over a network
- The network is modeled by a time-varying directed graph, $\mathcal{G}(t) = (|n|, \mathcal{E}(t))$
- The exchanged information is compressed before communication

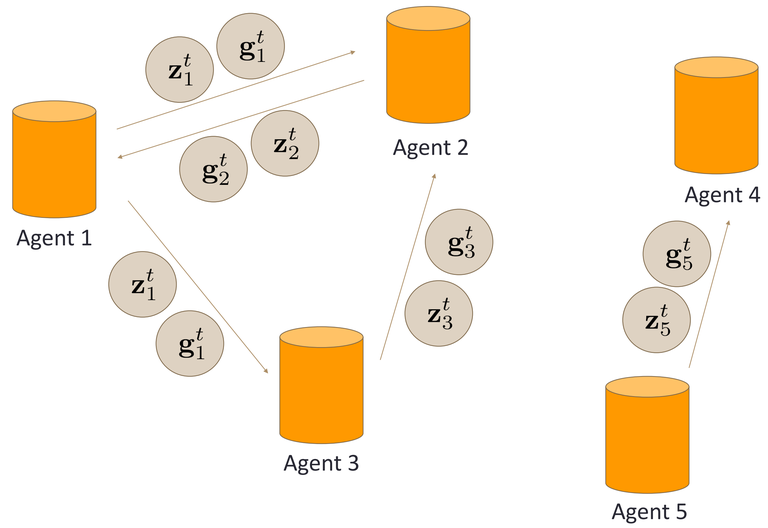
CHALLENGES

- Our algorithm is a communication-compression algorithm over time-varying directed network:
  - Communication imbalance in directed time-varying networks
  - Bias induced by the compression operator

- Compression operator: a sparsification operator that uniformly select k out of d entries from a d-dimensional message

$$Q : \mathbb{R}^d \to \mathbb{R}^d$$

ALGORITHM



$\mathbf{z}_i^t$ → Model-related vector

$\mathbf{g}_i^t$ → Gradient vector

$$\mathbf{z}_i^t = \begin{cases} \mathbf{x}_i^t, & i \in \{1, ..., n\} \\ \mathbf{y}_{i-n}^t, & i \in \{n+1, ..., 2n\} \end{cases}$$

$\mathbf{x}_i^t$ Vector of parameters

$\mathbf{y}_i^t$ Auxiliary vector to record the network imbalance bias

- Elementwise update:

$$z_{im}^{t+1} = \sum_{j=1}^{2n} [\bar{M}_m^t]_{ij} [Q(\mathbf{z}_j^t)]_m + \mathbf{1}_{\{t \bmod \mathcal{B}=\mathcal{B}-1\}} \epsilon [F]_{ij} z_{jm}^{\mathcal{B}\lfloor t/\mathcal{B}\rfloor} - \mathbf{1}_{\{t \bmod \mathcal{B}=\mathcal{B}-1\}} \alpha_{\lfloor t/\mathcal{B}\rfloor} g_{im}^{\mathcal{B}\lfloor t/\mathcal{B}\rfloor}$$

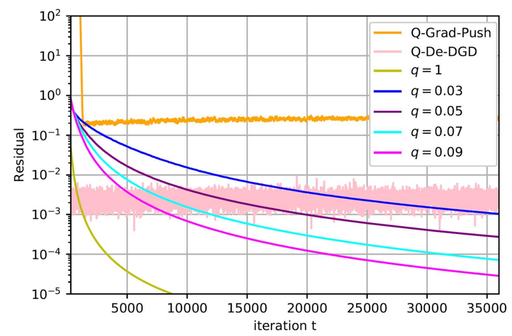
- Reweighting mixing matrix to cancel out compression bias
- Stored message facilitating consensus convergence in jointly-connected networks
- Display local gradient descent in jointly-connected network

CONVERGENCE GUARANTEE

For the stepsize $\alpha_t = \mathcal{O}(1/\sqrt{t})$, the algorithm attains the convergence rate $\mathcal{O}(\frac{\ln T}{\sqrt{T}})$.

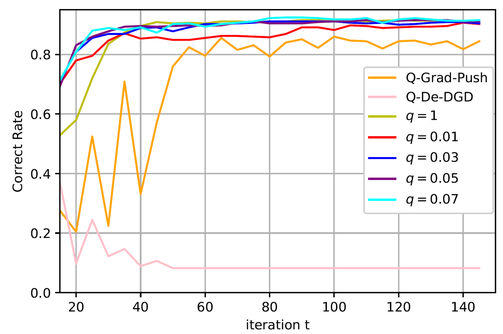
SIMULATION RESULTS

- Decentralized linear regression simulation:



Algorithm with different sparsification levels

- Decentralized logistic regression simulation:



Algorithm with different sparsification levels

CONCLUSION AND FUTURE WORK

- Proposed a communication-sparsifying algorithm for decentralized convex optimization over directed time-varying graphs
- Proved the convergence rate of the proposed algorithm
- Justified the performance of the algorithm in simulations

Future work

- Study other gradient methods to accelerate convergence rate
- Extend the algorithm to non-convex optimization problems