Scalable Average Consensus with Compressed Communications

Mohammad Taha Toghani, César A. Uribe

Department of Electrical and Computer Engineering, Rice University, Houston, TX, USA

RICE ENGINEERING
Electrical and Computer Engineering

Problem Setup

A group of n agents interacting over a fixed, undirected, and connected network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, seek to solve problem (1), defined as follows: Each agent $i \in [n]$

(a) can only communicate to its neighbors $\mathcal{N}_i$, over graph $\mathcal{G}$,

(b) maintains a set of initial parameters $\mathbf{x}_i \in \mathbb{R}^d$,

$$\mathbf{x}^\star := \arg\min_{\mathbf{x}\in\mathbb{R}^d} \frac{1}{2n}\sum_{i=1}^{n}\|\mathbf{x}-\mathbf{x}_i\|^2. \tag{1}$$

Challenges:

- Classically, consensus algorithms have convergence rates proportional to their corresponding networks' spectral gaps.
- Structures such as Ring and Path have spectral gaps of $\mathcal{O}(n^2)$.
- Scalability is a generic term for methods that improve the convergence rate dependence on the number of agents (n).

Main Objective

Design a communication-efficient and scalable algorithm for the average consensus problem in (1).

Compression Model

We consider ω-contracted randomized compression operators $Q : \mathbb{R}^d \to \mathbb{R}^d$ that satisfy

$$\mathbb{E}\big\|Q(\mathbf{x})-\mathbf{x}\big\|^2 \le \omega^2\big\|\mathbf{x}\big\|^2, \qquad \forall \mathbf{x}\in\mathbb{R}^d, \tag{2}$$

where $\omega \in [0, 1)$, and $\omega = 0$ means no compression.

Example Operators:

- Sparsification: rand_k (or top_k) operator that selects k random (or top) out of d entries.
- Quantization: qsgd_k operator that rounds each entry to one of the $2^{k-1}+1$ quantized levels.

Aggregation Model

For an undirected graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, we consider its associated *Metropolis-Hasting* mixing matrix $\mathbf{W} = \mathcal{MH}(\mathcal{G})$ as follows:

$$\mathbf{W}_{ij} = \begin{cases} \frac{1}{\max\{|\mathcal{N}_i|,|\mathcal{N}_j|\}+1}, & \text{if } (i,j)\in\mathcal{E} \\ 1-\sum_{j\neq i}\mathbf{W}_{ij}, & \text{if } i=j \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

which is doubly-stochastic and symmetric.

Algorithm & Comparisons

Algorithm 1: Scalable Compressed Gossip (SCG)

Input: initial parameters $\mathbf{x}_i(0) \in \mathbb{R}^d$, for all $i \in [n]$, network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mixing matrix $\mathbf{W}$, stepsize $\gamma \in (0, 1]$, operator Q, momentum $\sigma \in [0, 1)$.

1: $\hat{\mathbf{x}}_i(0) := \mathbf{0},\ \mathbf{y}_i(0) := \mathbf{x}_i(0), \quad \forall i \in [n]$

2: **for** t **in** $0, \ldots, T-1$, in parallel $\forall i \in [n]$ **do**

3: $\mathbf{q}_i(t) := Q(\mathbf{x}_i(t) - \hat{\mathbf{x}}_i(t))$

4: Send $\mathbf{q}_i(t)$ and receive $\mathbf{q}_j(t)$, for all $j \in \mathcal{N}_i$

5: $\hat{\mathbf{x}}_j(t+1) := \hat{\mathbf{x}}_j(t) + \mathbf{q}_j(t)$, for all $j \in \mathcal{N}_i \cup \{i\}$

6: $\mathbf{y}_i(t+1) := \mathbf{x}_i(t) + \gamma \sum_{j\in\mathcal{N}_i\cup\{i\}} \mathbf{W}_{ij}\left(\hat{\mathbf{x}}_j(t+1) - \hat{\mathbf{x}}_i(t+1)\right)$

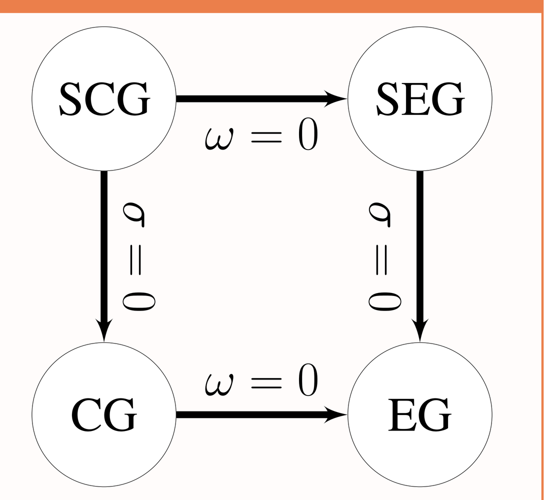
7: $\mathbf{x}_i(t+1) := (1+\sigma)\,\mathbf{y}_i(t+1) - \sigma\,\mathbf{y}_i(t)$

8: **end for**

Initialization Compression Communication
Error-Feedback Aggregation Extrapolation

Comparison

- **Exact Gossip (EG) [1]**
- **Compressed Gossip (CG) [2]**
- **Scalable Exact Gossip (SEG) [3]**
- **Scalable Compressed Gossip (SCG)**
  This Work



Algorithm	Linear Rate	Stepsize (γ)	ω
EG [1]	$\mathcal{O}(1-\gamma n^{-2})$	$(0,1]$	no compression
SEG [3]	$\mathcal{O}(1-\gamma^{\frac{1}{2}} n^{-1})$	$(0,\frac{1}{2}]$	no compression
CG [2]	$\mathcal{O}(1-n^{-4})$	$\mathcal{O}(n^{-4})$	$[0,1)$
CG flexible γ	$\mathcal{O}(1-\gamma n^{-2})$	$(0,1]$	$\left[0, \Theta\left(\frac{1}{(1+\gamma)n^2}\right)\right]$
SCG This Work	$\mathcal{O}(1-\gamma^{\frac{1}{2}} n^{-1})$	$(0,\frac{1}{2}]$	$\left[0, \Theta\left(\frac{1}{(1+\gamma)n^2}\right)\right]$

n is the number of agents

Convergence Analysis

Theorem 1: Scalable Compressed Gossip

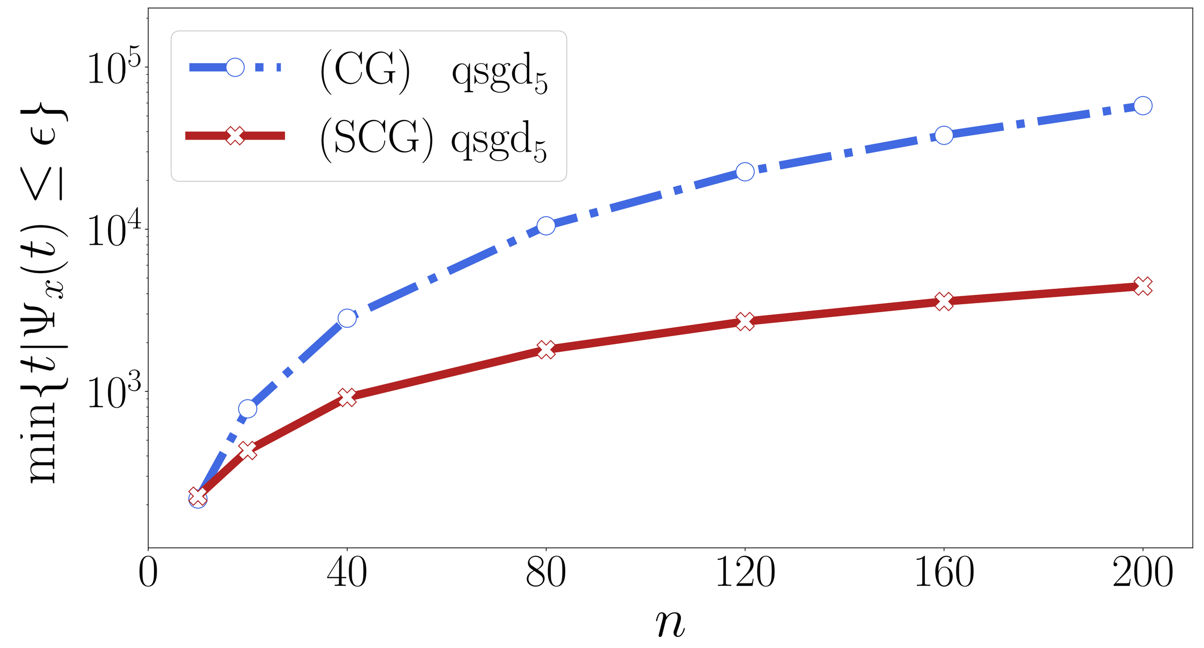
Let operator Q satisfies (2), $\mathbf{y}_i(0) = \mathbf{x}_i(0)$, $\hat{\mathbf{x}}_i(0) = \mathbf{0}$, for all $i\in[n]$, $\gamma \in (0, 1]$, $\sigma=\frac{5n-\sqrt{\gamma}}{5n+\sqrt{\gamma}}$, and $\mathbf{W} = \mathcal{MH}(\mathcal{G})$. The iterates of Algorithm 1 satisfy

$$\mathbb{E}\Psi_x(t) \le C_0\tilde{\lambda}^t\Psi_x(0),$$

where $\kappa_2=\sqrt{2\sigma^2+2\sigma+1}$, $\kappa_3 = \sqrt{2\sigma^2+2}$, $\beta = \|\mathbf{W}-\mathbf{I}\|$, $\lambda=1-\frac{\sqrt{\gamma}}{5n}$, $\tilde{\lambda}=1-\frac{\sqrt{\gamma}}{10n}$, $\Psi_x(t)^2=\sum_{i=1}^{n}\left\|\mathbf{x}_i(t)-\mathbf{x}^\star\right\|^2$, and $C_0, C>0$, for $\omega \le \left(2\left(\kappa_3+\gamma\beta\kappa_2\right)\left(\lambda^{-\frac{1}{2}}+\gamma\beta\kappa_2 C\lambda^{-1}(1-\lambda^{\frac{1}{2}})^{-2}\right)\right)^{-1}$.

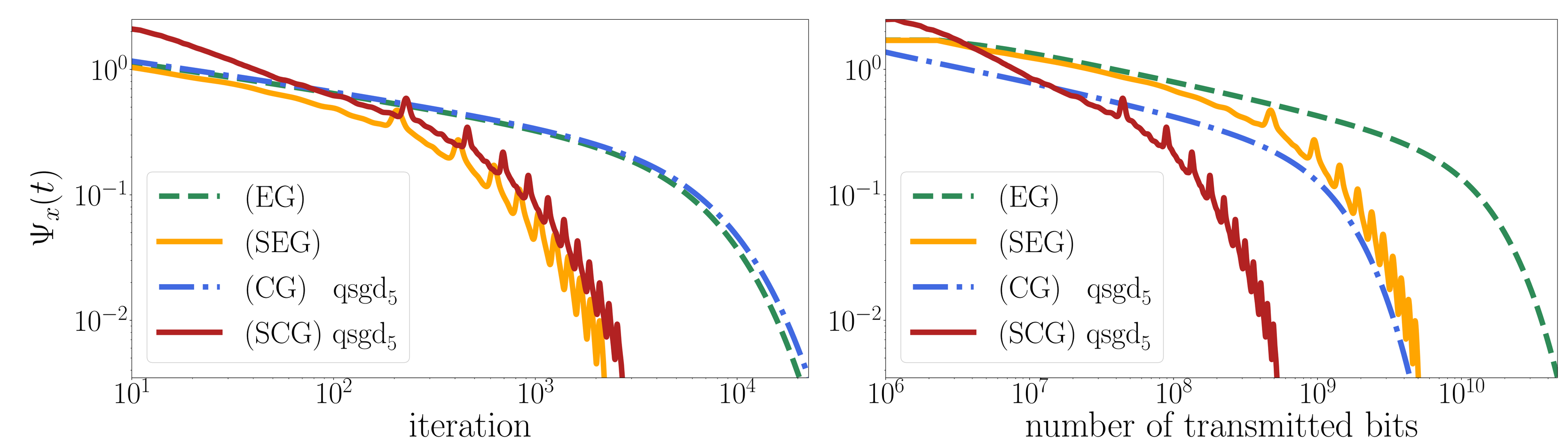
Numerical Result

Number of iterations required for ϵ-accuracy



- Path graph
- $n = 10$ to 200
- $d = 150$
- $\epsilon = 10^{-4}$
- 5-bit quantization

Consensus error



- Ring graph, $n = 120$, $d = 150$, 5-bit quantization

References

[1] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65–78, 2004.

[2] A. Koloskova, S. Stich, and M. Jaggi, "Decentralized Stochastic Optimization and Gossip Algorithms with Compressed Communication," in *International Conference on Machine Learning*, 2019, pp. 3478–3487.

[3] A. Olshevsky, "Linear Time Average Consensus and Distributed Optimization on Fixed Graphs," *SIAM J. Control. Optim.*, vol. 55, pp. 3990–4014, 2017.

[4] M.T. Toghani and C. Uribe, "Scalable average consensus with compressed communications," *arXiv preprint arXiv:2109.06996*, 2021.