Scalable Average Consensus with Compressed Communications
Mohammad Taha Toghani, César A. Uribe
Department of Electrical and Computer Engineering, Rice University, Houston, TX, USA

Problem Setup
A group of $n$ agents interacting over a fixed, undirected, and connected network $G = (\mathcal{V}, \mathcal{E})$, 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 $G$,
(b) maintains a set of initial parameters $x_i \in \mathbb{R}^d$.

\[
x^* := \arg\min_{x \in \mathbb{R}^d} \frac{1}{2n} \sum_{i=1}^{n} \|x - x_i\|^2.
\]  

$\gamma \in (0,1)$, operator $Q$, momentum $\sigma \in [0,1]$.

Challenges:
• Classically, consensus algorithms have convergence rates proportional to their corresponding networks’ spectral gaps.
• Structures such as Ring and Path have spectral gaps of $O(n^2)$.
• 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 $\omega$-contracted randomized compression operators $Q : \mathbb{R}^d \to \mathbb{R}^d$ that satisfy
\[
E\|Q(x) - x\|^2 \leq \omega \|x\|^2, \quad \forall x \in \mathbb{R}^d,
\]  

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

Example Operators:
• Sparsification: rand$k_d$(or top$k$) operator that selects $k$ random (or top) out of $d$ entries.
• Quantization: spq$k_d$ operator that rounds each entry to one of the $2^k + 1$ quantized levels.

Aggregation Model
For an undirected graph $G = (\mathcal{V}, \mathcal{E})$, we consider its associated Metropolis-Hasting mixing matrix $W = M\mathcal{H}(G)$ as follows:
\[
W_{ij} = \begin{cases} 
\frac{1}{\max\{\mathcal{N}_i \cup \mathcal{N}_j\} + 1} & \text{if } (i,j) \in \mathcal{E} \\
1 - \sum_{j \neq i} W_{ij}, & \text{if } i = j \\
0, & \text{otherwise}
\end{cases}.
\]

which is doubly-stochastic and symmetric.

Algorithm & Comparisons
Algorithm 1: Scalable Compressed Gossip (SCG)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Linear Rate</th>
<th>Stepsize ($\gamma$)</th>
<th>$\omega$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EG [1]</td>
<td>$O(1 - \gamma n^{-2})$</td>
<td>$(0, 1]$</td>
<td>no compression</td>
</tr>
<tr>
<td>SEG [3]</td>
<td>$O(1 - \gamma^2 n^{-1})$</td>
<td>$(0, 1]$</td>
<td>no compression</td>
</tr>
<tr>
<td>CG [2]</td>
<td>$O(1 - n^{-2})$</td>
<td>$O(n^{-1})$</td>
<td>$(0, 1]$</td>
</tr>
</tbody>
</table>

This Work

Comparison

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

Convergence Analysis

Theorem 1: Scalable Compressed Gossip
Let operator $Q$ satisfies (2), $y(0) = x_i(0)$, $x_i(0) = 0$, for all $i \in [n]$, $\gamma \in (0,1)$, $\sigma \in [0,1]$, and $W = M\mathcal{H}(G)$. The iterates of Algorithm 1 satisfy
\[
\|\Psi_x(t)\| \leq C_0 \beta_{\omega} \Psi_x(0),
\]
where $\beta_{\omega} = \sqrt{2\sigma^2 + 2\sigma^2 + 1}$, $C_0 = \|W - I\|$, $\lambda = 1 - \frac{\gamma}{\sigma_{\min}}$, $\beta_{\omega} = \sum_{i=1}^{n} \|x_i(t) - x^*\|^2$, and $C_0, C > 0$, for $\omega \leq (2(\kappa + \gamma^2 \kappa^2)(\lambda^2 + \gamma \kappa C^{-2}(1 - \lambda^{-2})^2))^{-1}$.

Numerical Result

Number of iterations required for $\varepsilon$-accuracy

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

References