# Provably Accelerated Randomized Gossip Algorithms

Nicolas Loizou* Michael Rabbat♯ Peter Richtárik♮†
*University of Edinburgh ♯ Facebook AI Research ♮ KAUST †MIPT

## 1. Average Consensus Problem (ACP)

**SETUP:** $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a connected network with $|\mathcal{V}| = n$ nodes (e.g., sensors) and $|\mathcal{E}| = m$ edges (e.g., communication links). Node $i \in \mathcal{V}$ stores a private value $c_i \in \mathbb{R}$ (e.g., temperature).

**GOAL:** Compute the average of the private values (i.e., the quantity $\bar{c} := \frac{1}{n}\sum_i c_i$) in a **distributed** fashion. That is, exchange of information can only occur along the edges.

## 2. Optimization Formulation of ACP

The optimal solution of the optimization problem

$$\min_{x \in \mathbb{R}^n} \quad \tfrac{1}{2}\|x - c\|^2 = \tfrac{1}{2}\textstyle\sum_i (x_i - c_i)^2 \quad \text{subject to} \quad x_i = x_j \quad \text{for all} \quad (i,j) \in \mathcal{E} \tag{1}$$

is $x_i^* = \bar{c}$ for all $i$. The constraints can be written compactly as $\mathbf{A}x = 0$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$, and the rows of the system enforce the constraints $x_i = x_j$ for $(i,j) \in \mathcal{E}$.
**QUESTIONS:** Can we interpret Randomized Gossip (RG) algorithms for ACP as instances of specific randomized optimization methods for (1)? Can new accelerated **RG** methods be developed this way?

## 3. Accelerated Randomized Kaczmarz Method

**Best Approximation Problem:** $\min_{x \in \mathbb{R}^n} \quad \frac{1}{2}\|x - x^0\|^2 \quad \text{subject to} \quad \mathbf{A}x = b$

**Algorithm 1** Accelerated Randomized Kaczmarz Method (AccRK)

Data: Matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$; vector $b \in \mathbb{R}^m$
Choose $x^0 \in \mathbb{R}^n$ and set $v^0 = x^0$
Parameters: Evaluate the sequences of the scalars $\alpha_k, \beta_k, \gamma_k$ following one of two possible options.
**for** $k = 0, 1, 2, \dots,$ **do**
  $y^k = \alpha_k v^k + (1 - \alpha_k) x^k$
  Draw a fresh sample $i_k \in [m]$ with probability $1/m$
  $x^{k+1} = y^k - \frac{\mathbf{A}_{i_k:} y^k - b_{i_k}}{\|\mathbf{A}_{i_k:}\|^2} \mathbf{A}_{i_k:}^\top$
  $v^{k+1} = \beta_k v^k + (1 - \beta_k) y^k - \gamma_k \frac{\mathbf{A}_{i_k:} y^k - b_{i_k}}{\|\mathbf{A}_{i_k:}\|^2} \mathbf{A}_{i_k:}^\top$
**end for**

**Option 1, [4]:** $\lambda \in [0, \lambda_{\min}^+(\mathbf{A}^\top \mathbf{A})]$ and $\gamma_{-1} = 0$. Generate $\{\gamma_k\}$: $\gamma_k$ largest root of

$$\gamma_k^2 - \frac{\gamma_k}{m} = (1 - \frac{\gamma_k}{\lambda} m)\gamma_{k-1}^2.$$

Generate $\{\alpha_k\}$ and $\{\beta_k\}$ by setting,

$$\alpha_k = (m - \gamma_k \lambda)/(\gamma_k(m^2 - \lambda)), \qquad \beta_k = 1 - (\gamma_k \lambda)/m.$$

**Convergence Rate:** For $k \to \infty$,

$$\mathbb{E}[\|x^k - x^*\|^2] \leq \left(1 - \sqrt{\lambda}/m\right)^k 4\lambda \|x^0 - x^*\|^2_{(\mathbf{A}^\top \mathbf{A})^\dagger}$$

**Option 2, [3]:** Let $\mathbf{W} = \frac{\mathbf{A}^\top \mathbf{A}}{2m}$ and

$$\nu = \max_{u \in \text{Range}(\mathbf{A}^\top)} \frac{u^\top \left[\sum_{i=1}^m \mathbf{A}_{i:}^\top \mathbf{A}_{i:} (\mathbf{A}^\top \mathbf{A})^\dagger \mathbf{A}_{i:}^\top \mathbf{A}_{i:}\right] u}{u^\top \frac{\mathbf{A}^\top \mathbf{A}}{m} u},$$

where $1 \leq \nu \leq m$. $\{\beta_k\} \equiv 1 - \sqrt{\lambda_{\min}^+(\mathbf{W})/\nu}$, $\{\gamma_k\} \equiv \sqrt{\frac{1}{\lambda_{\min}^+(\mathbf{W})\nu}}$, $\{\alpha_k\} \equiv \frac{1}{1+\gamma\nu} \in (0,1)$.
Let $\Psi^k = \mathbb{E}[\|v^k - x^*\|^2_{\mathbf{W}^\dagger} + \frac{1}{\mu}\|x^k - x^*\|^2]$.
**Convergence Rate:**

$$\Psi^k \leq \left(1 - \sqrt{\lambda_{\min}^+(\mathbf{W})/\nu}\right)^k \Psi^0$$

## 4. Iteration Complexity

Let $\mathbf{L}$ be the Laplacian matrix of the network $\mathcal{G}$. Then, for solving $\mathbf{A}x = 0$, where $\mathbf{A}$ the incidence matrix,
**RK** [5, 2]: $O\left(\left(\frac{2m}{\lambda_{\min}^+(\mathbf{L})}\right)\log(1/\epsilon)\right)$,
**AccRK (Option 1)**: $O\left(m\sqrt{2/\lambda_{\min}^+(\mathbf{L})}\log(1/\epsilon)\right)$, **AccRK (Option 2)**: $O\left(\sqrt{2m\nu/\lambda_{\min}^+(\mathbf{L})}\log(1/\epsilon)\right)$

## 5. Accelerated Randomized Gossip Algorithms

**Algorithm 2** Accelerated Randomized Gossip Algorithm (AccGossip)

**Data:** Matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ (normalized incidence matrix); vector $b = 0 \in \mathbb{R}^m$
Choose $x^0 \in \mathbb{R}^n$ and set $v^0 = x^0$
**Parameters:** Evaluate the sequences $\alpha_k, \beta_k, \gamma_k$ following option 1 or 2.
**for** $k = 0, 1, 2, \dots,$ **do**
  Each node $\ell \in \mathcal{V}$ evaluates

$$y_\ell^k = \alpha_k v_\ell^k + (1 - \alpha_k) x_\ell^k.$$

  Pick an edge $e = (i,j) \in \mathcal{E}$ uniformly at random.
- Nodes $i$ and $j$ update their values via:

$$x_i^{k+1} = x_j^{k+1} = (y_i^k + y_j^k)/2$$
$$v_i^{k+1} = \beta_k v_i^k + (1-\beta_k) y_i^k - \gamma_k (y_i^k - y_j^k)/2$$
$$v_j^{k+1} = \beta_k v_j^k + (1-\beta_k) y_j^k - \gamma_k (y_j^k - y_i^k)/2$$

- Any other node $\ell \in \mathcal{V}/\{i,j\}$:

$$x_\ell^{k+1} = y_\ell^k \quad , \quad v_\ell^{k+1} = \beta_k v_\ell^k + (1-\beta_k) y_\ell^k$$

**end for**

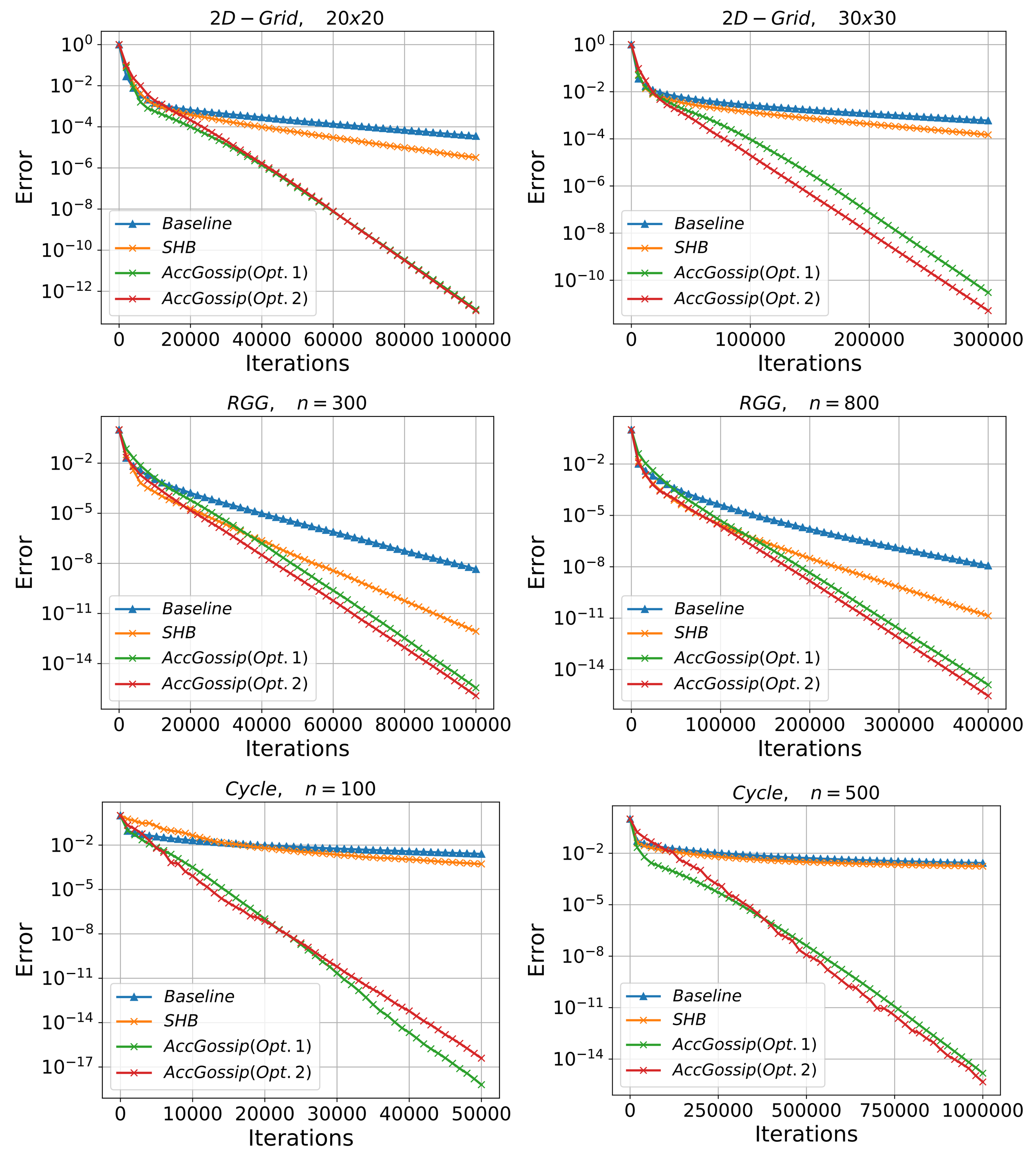


**Figure 1:** Performance of AccGossip in a 2-dim grid, random geometric graph (RGG) and a cycle graph. The baseline method corresponds to the randomized pairwise gossip algorithm proposed in [1] and SHB to the fast gossip algorithm proposed in [6]. The $n$ in the title of each plot indicates the # of nodes of the network. For the grid graph, this is $n \times n$.

## 6. References

[1] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Trans. Inf. Theory*, 2006.

[2] R. M. Gower and P. Richtárik. Stochastic dual ascent for solving linear systems. *arXiv preprint arXiv:1512.06890*, 2015.

[3] R.M. Gower, F. Hanzely, P. Richtárik, and S. Stich. Accelerated stochastic matrix inversion: general theory and speeding up BFGS rules for faster second-order optimization. *NeurIPS*, 2018.

[4] J. Liu and S. Wright. An accelerated randomized Kaczmarz algorithm. *Math. Comput.*, 85(297):153–178, 2016.

[5] N. Loizou and P. Richtárik. A new perspective on randomized gossip algorithms. In *GlobalSIP*, 2016.

[6] N. Loizou and P. Richtárik. Accelerated gossip via stochastic heavy ball method. In *Allerton*, 2018.

[7] N. Loizou and P. Richtárik. Revisiting randomized gossip algorithms: General framework, convergence rates and novel block and accelerated protocols. In *Manuscript*, 2019.