A New Perspective on Randomized Gossip Algorithms

Nicolas Loizou & Peter Richtárik
School of Mathematics, University of Edinburgh

1. Average Consensus Problem (ACP)

SETUP: Let $G = (V, E)$ be a connected network with $|V| = n$ nodes (e.g. sensors) and $|E| = m$ edges (e.g. communication). All nodes $i \in V$ store 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.

Algorithms for solving ACP: **Randomized Gossip Algorithms (RGA)**

2. Optimization Formulation of ACP

The optimal solution of the optimization problem

$$\text{minimize} \quad \tfrac{1}{2}\|x - c\|^2 \quad \text{subject to} \quad x_i = x_j \quad \text{for all} \quad e = (i,j) \in E \qquad (1)$$

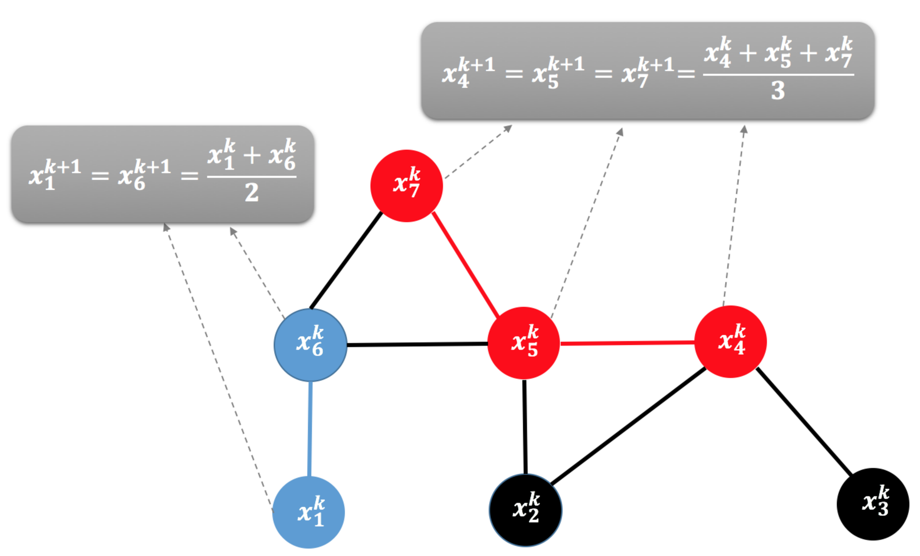
is $x_i^* = \bar{c}$ for all i. So, **RGA** solves the above optimization problem. The constraints can be written compactly as $\mathbf{A}x = 0$, with each row of the system enforcing $x_i = x_j$ for one edge $(i,j) \in E$.

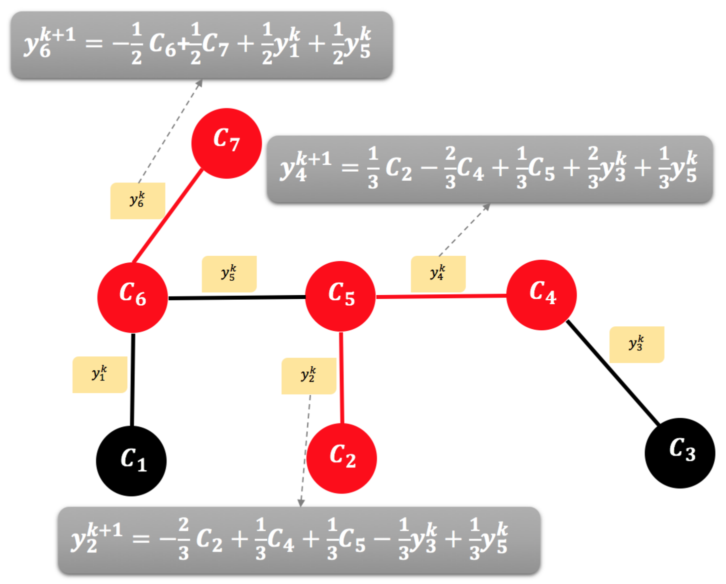
QUESTION: By formulate the constraints of problem (1) as linear system can we get new variants of Randomized Gossip Algorithms?

5. Randomized Block Kaczmarz (RBK)/Randomized Newton (RN)

NEW GOSSIP METHODS: We can now formulate many new variants of **RGA**, by applying SDA to (1) with various choices of random matrices $\mathbf{S}$. We also naturally obtain dual interpretation of such new gossip methods.

SETUP: Choose $\mathbf{S} = \mathbf{I}_{\mathcal{S}_k}$, where $\mathbf{I}_{\mathcal{S}_k}$ is a column submatrix of the $m \times m$ identity matrix corresponding to columns indexed by a random subset of edges $\mathcal{S}_k \subseteq E$.





Primal Iterates of SDA = Randomized Block Kaczmarz Algorithm

$$x^{k+1} = x^k - \mathbf{A}^\top \mathbf{I}_{\mathcal{S}_k}(\mathbf{I}_{\mathcal{S}_k}^\top \mathbf{A}\mathbf{A}^\top \mathbf{I}_{\mathcal{S}_k})^\dagger \mathbf{I}_{\mathcal{S}_k}^\top \mathbf{A}x^k$$

1. Form a subgraph G_k of G by selecting a random set of edges $\mathcal{S}_k \subseteq E$
2. For each connected component of G_k, replace node values with their average

Dual Iterates of SDA = Randomized Newton Algorithm

$$y^{k+1} = y^k - \mathbf{I}_{\mathcal{S}_k}(\mathbf{I}_{\mathcal{S}_k}^\top \mathbf{A}\mathbf{A}^\top \mathbf{I}_{\mathcal{S}_k})^\dagger \mathbf{A}(c + \mathbf{A}^\top y^k)$$

1. Form a subgraph G_k of G by selecting a random set of edges $\mathcal{S}_k \subseteq E$
2. Modify the dual variables y_e for $e \in \mathcal{S}_k$ (see the image)

3. Duality for Linear Systems

Problem (1) is special case of the more general problem:

PRIMAL PROBLEM:

$$\min_{x \in \mathbb{R}^n} \tfrac{1}{2}\|x - c\|^2 \quad \text{s.t.} \quad \mathbf{A}x = b$$

where $\mathbf{A}$ can be any matrix such that $\mathbf{A}x = b$ has a solution.

DUAL PROBLEM:

$$\max_{y \in \mathbb{R}^m} \quad D(y) := (b - \mathbf{A}c)^\top y - \tfrac{1}{2}\|\mathbf{A}^\top y\|^2$$

4. Stochastic Dual Ascent [1]

DUAL METHOD (SDA):

$$y^{k+1} \leftarrow y^k + \mathbf{S}_k \lambda^k$$

where $\mathbf{S}_k$ is a random matrix with m rows, and λ^k is chosen so that $D(y^k + \mathbf{S}_k\lambda^k)$ is maximized.

PRIMAL METHOD: With the dual iterates $\{y^k\}$ we can associate primal iterates $\{x^k\}$:

$$x^k \leftarrow c + A^\top y^k$$

6. Theoretical Results and Numerical Experiments

Convergence Rate:

Theorem [1]. RN and RBK converge as:

$$\mathbb{E}[D(y^*) - D(y^k)] \leq \rho^k (D(y^*) - D(y^0)),$$

$$\mathbb{E}[\tfrac{1}{2}\|x^k - x^*\|^2] \leq \rho^k \tfrac{1}{2}\|x^0 - x^*\|^2,$$

where the rate is given by

$$\rho := 1 - \lambda_{\min}^+ \left(\mathbf{A}^\top \mathbb{E}[I_{\mathbf{S}_k}(I_{\mathbf{S}_k}^\top \mathbf{A}\mathbf{A}^\top I_{\mathbf{S}_k})^\dagger I_{\mathbf{S}_k}^\top]\mathbf{A}\right)$$

Theorem: (ε-**Averaging Time**)
$T_{ave}(\varepsilon) \leq 3\log(1/\varepsilon)/\log(1/\rho) \leq \frac{3}{1-\rho}\log(1/\epsilon)$

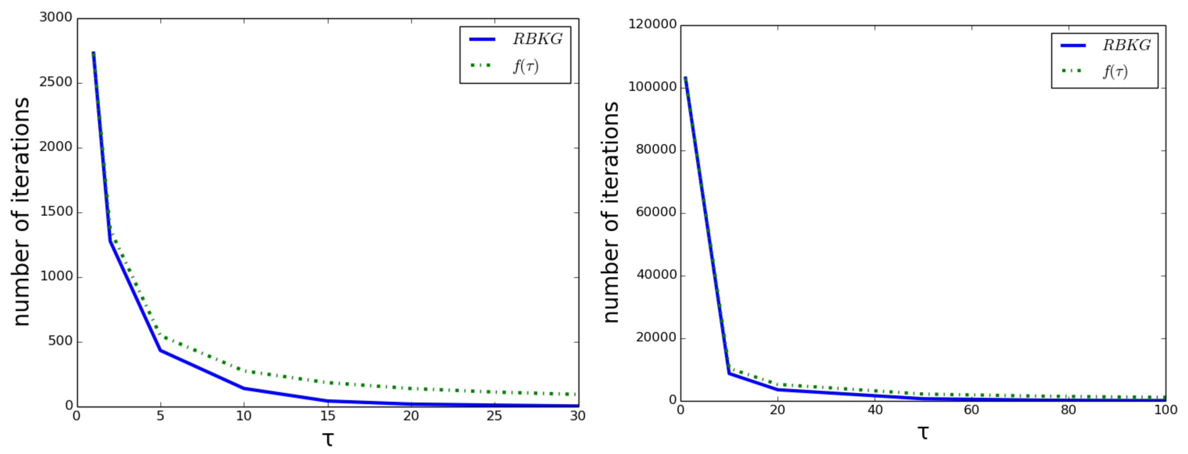
Importance of Duality:

Theorem: RBK enjoys superlinear speedup in $\tau = |\mathcal{S}|$. That is, as τ increases by some factor, the iteration complexity drops by a factor that is at least as large.

On Numerical Experiments:

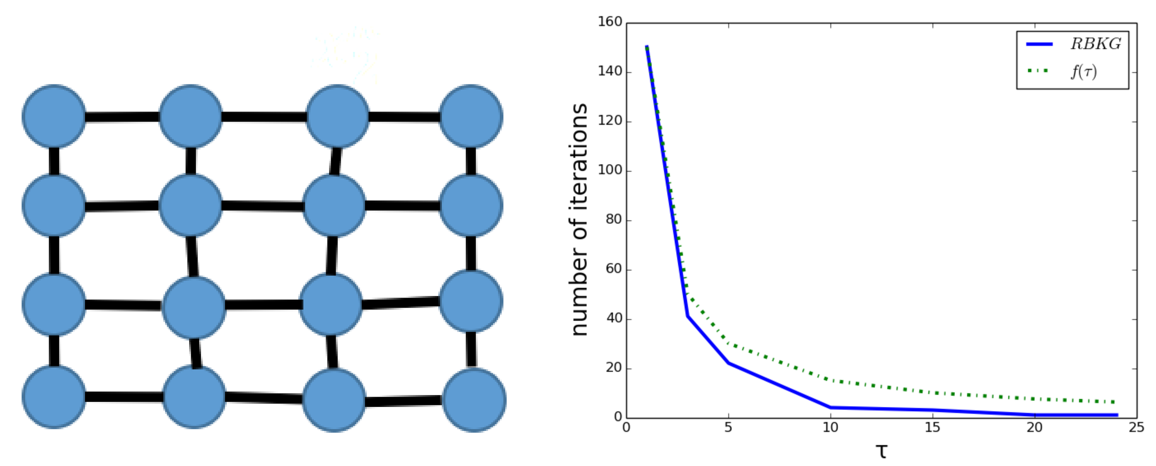
Blue solid line: The actual number of iterations (after running the code)

Green dotted line: Represents the function $f(\tau) := \frac{\ell}{\tau}$, where ℓ is the number of iterations of RBK with $\tau = 1$. (Linear Speedup)



(a) Ring graph with $n = 30$ **(b)** Ring graph with $n = 100$

Figure 1: Superlinear speedup of RBK on the ring graph.



(a) 4×4 grid graph **(b)** Speedup in τ

Figure 2: Superlinear speedup of RBK on the 4×4 grid graph

7. References

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

[2] N. Loizou and P. Richtárik. Randomized gossip algorithms: Complexity, duality and new variants. *In Progress*, 2016.

[3] D. Needell and J.A. Tropp. Paved with good intentions: analysis of a randomized block Kaczmarz method. *Linear Algebra Appl.*, 441:199–221, 2014.

[4] Z. Qu, P. Richtárik, M. Takáč, and O. Fercoq. SDNA: stochastic dual Newton ascent for empirical risk minimization. *ICML*, 2016.