Average-case optimization analysis for distributed consensus algorithms on regular graphs

Nhat Trung Nguyen **Alexander Rogozin** **Alexander Gasnikov**

Moscow Institute of Physics and Technology

Consensus Problem

Consider a network of agents represented by an undirected finite graph $G = (V, E)$, where $V = \{1, \ldots, n\}$ represents the set of vertices (agents) and E represents the set of edges (communication links). Each agent i holds an initial vector $x_0^{(i)} \in \mathbb{R}^d$. We denote by $x_0 = \left(\left(x_0^{(1)}\right)^\top, \ldots, \left(x_0^{(n)}\right)^\top\right)^\top$. The goal is to design efficient algorithms that allow each agent to quickly compute the average value $\overline{x}_0 = \frac{1}{n}\sum_{i=1}^n x_0^{(i)}$, with the constraint that at each iteration of the algorithm, agents can only exchange their vectors with their neighbors.

To achieve consensus on the graph G, we solve the following problem starting with the initial vector x_0:

$$\min_{x \in \mathbb{R}^{nd}} f(x) = \frac{1}{2} x^T \mathbf{L} x, \tag{1}$$

where $\mathbf{L} = L \otimes I_d$, the symbol $\otimes$ denotes the Kronecker product, and L is a *gossip matrix*, which is defined as follows

Definition 1. *A gossip matrix $L \in \mathbb{R}^{n\times n}$ on the graph $G = (V, E)$ is a matrix satisfying following properties:*

1. *L is an $n \times n$ symmetric matrix,*
2. *L is positive semi-definite,*
3. *$\ker(L) = \operatorname{span}(\mathbf{1})$, where $\mathbf{1} = (1, \ldots, 1)^\top$,*
4. *L is defined on the edges of the network: $L_{ij} \neq 0$ only if $i = j$ or $(i, j) \in E$.*

Polynomial-Based Iterative Methods

We consider *first-order methods* or *gradient-based methods* to solve the problem (1). These are methods in which the sequence of iterates x_t is in the span of previous gradients, i.e.,

$$x_{t+1} \in x_0 + \operatorname{span}\{\nabla f(x_0), \ldots, \nabla f(x_t)\}. \tag{2}$$

Lemma 1. *Let x_t be generated by a first-order method of kind (2). Then there exists a polynomial P_t of degree t such that $P_t(0) = 1$ and it verifies*

$$x_t - x_* = P_t(\mathbf{L})(x_0 - x_*) \tag{3}$$

The polynomial P_t is called the residual polynomial.

Average-case analysis

Definition 2. *Let L be a random matrix with eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$. The empirical spectral distribution of L is the probability measure*

$$\mu_L(\lambda) = \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i}(\lambda), \tag{4}$$

where δ_{λ_i} is the Dirac delta. Since L is random, the empirical spectral distribution μ_L is a random measure. Its expectation over L,

$$\mu = \mathbb{E}_L[\mu_L] \tag{5}$$

is called the expected spectral distribution

Theorem 1. *Let x_t be generated by a first-order method, associated to the polynomial P_t. Then we can decompose the expected error at iteration t as*

$$\mathbb{E}\|x_t - x_*\|^2 = R^2 \int P_t^2 d\mu. \tag{6}$$

Spectrum of regular graph

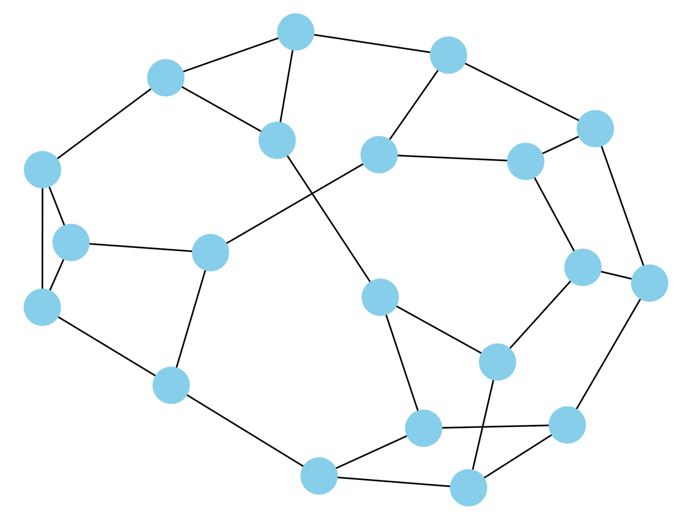

Figure 1: Regular graphs with $n = 20$, $k = 3$.

$$d\mu(\lambda) = \frac{k}{2\pi}\frac{\sqrt{\frac{4(k-1)}{k^2} - (1-\lambda)^2}}{1-(1-\lambda)^2}\, d\lambda \tag{7}$$

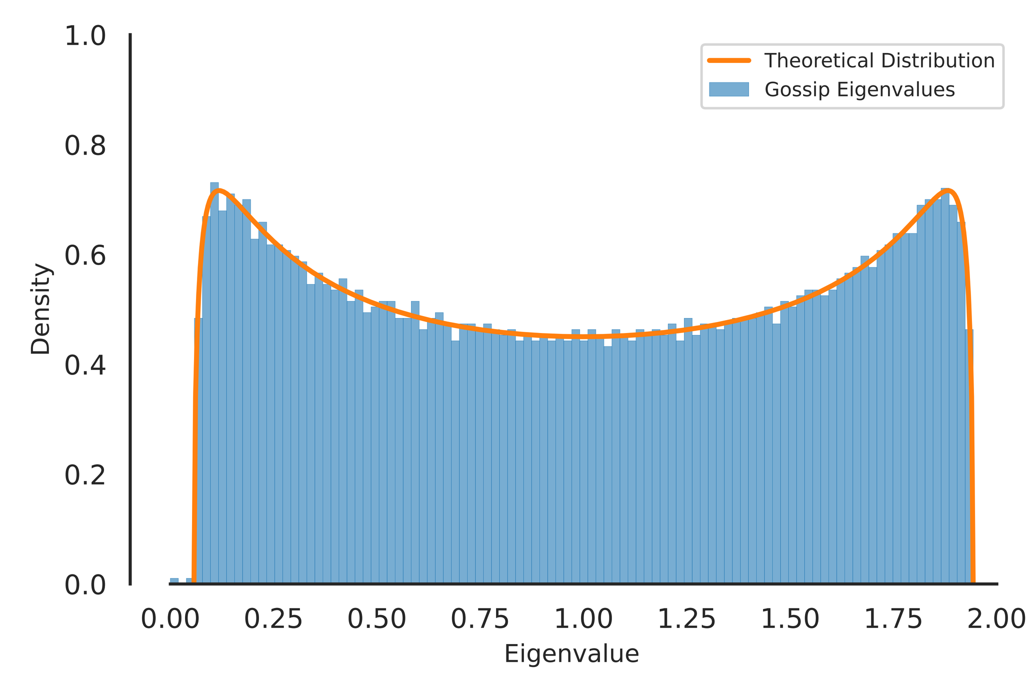

Figure 2: Spectrum of regular graph with $n = 5000$, $k = 3$.

Optimal method

Algorithm 1 Optimal average-case method for regular graphs

Input: starting guess x_0, regular parameter k, $\delta_0 = \frac{k}{k+1}$.
Initialize: $x_1 = x_0 - \delta_0 \cdot Lx_0$
for $t = 1, 2, \ldots$ **do**
$\quad \delta_t = \left(1 - \frac{k-1}{k^2}\cdot \delta_{t-1}\right)^{-1}$
$\quad x_{t+1} = x_t + (\delta_t - 1)(x_t - x_{t-1}) - \delta_t \cdot Lx_t$
end for

Theorem 2. *If we apply Algorithm 1 to problem (1), where L is the gossip matrix of random k-regular graphs, then*

$$\mathbb{E}\|x_t - x_*\|^2 = \Theta\left(\left(\frac{1}{k-1}\right)^t \cdot \left(\frac{1}{1 + \frac{2}{k-2}\left(1 - \frac{1}{(k-1)^t}\right)}\right)^2\right). \tag{8}$$

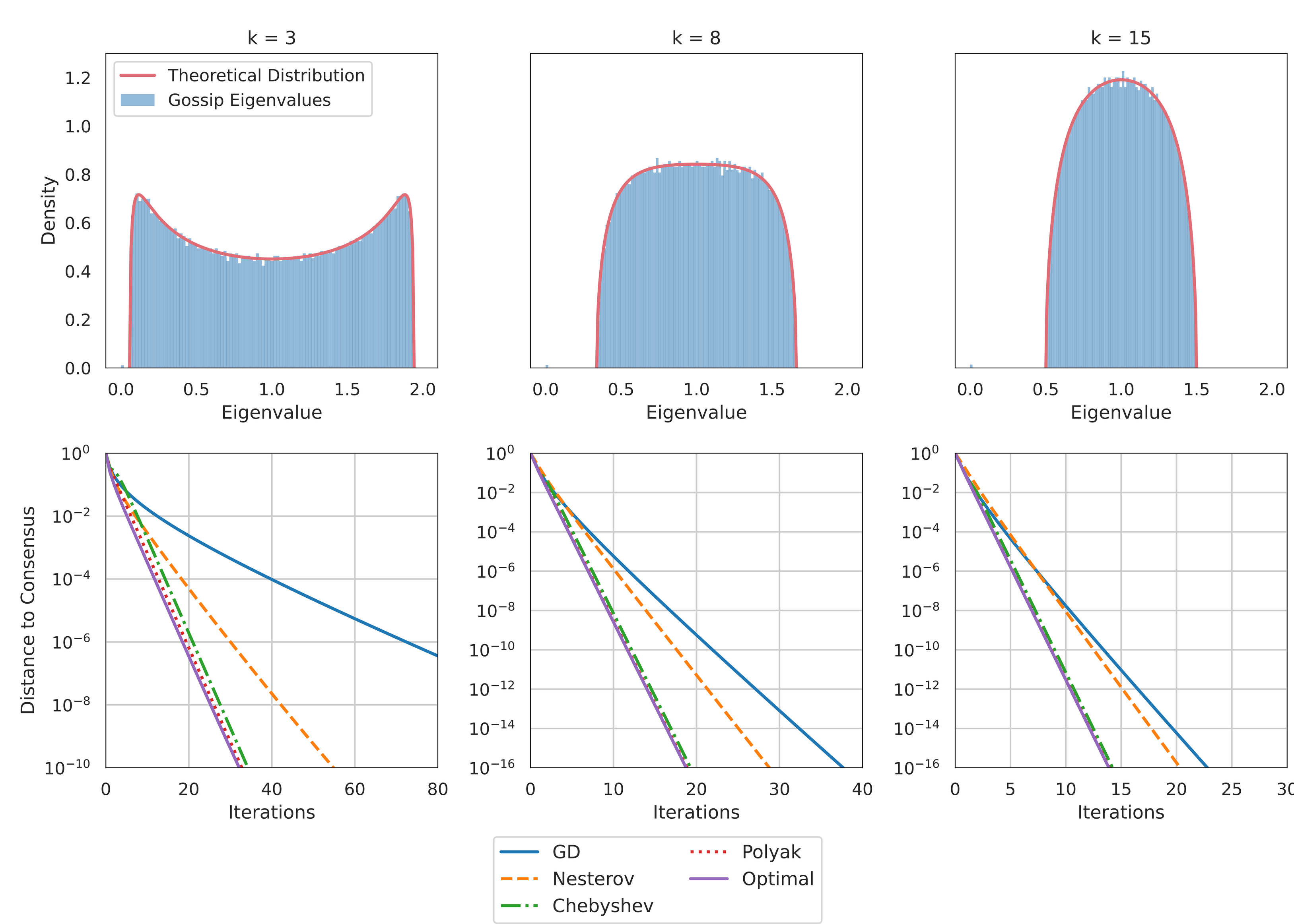

Figure 3: Comparison of convergence speeds of algorithms on regular graphs.