Randomized consensus algorithms over large scale networks

Fabio Fagnani

Dipartimento di Matematica, Politecnico di Torino, ITALY

joint work with

Sandro Zampieri

Department of Information Engineering, Università di Padova, ITALY

EE-YALE 10/05/07

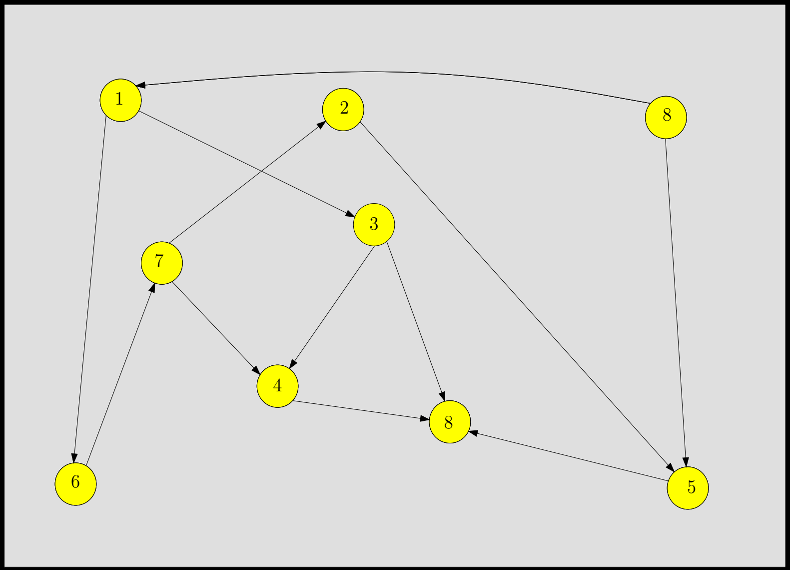
Outline

- Problem formulation and motivation
- Deterministic consensus algorithms
- Random consensus algorithms
- Performance issues
- Mean square analysis, examples
- Concentration results
- Conclusions

Problem formulation

$\mathcal{G} = (V, E)$ directed graph.
$V = \{1, \ldots, N\}$

$\forall i \in V$, a measure $x_i \in \mathbb{R}$



GOAL:
compute $X_A := N^{-1} \sum x_i$,
iteratively, exchanging information along available edges.
in a decentralized fashion

APPLICATIONS:

- Load balancing in computer networks
- Data fusion in sensor networks
- Coordination of multi-agent systems

Consensus algorithms I

$P(t)$ $N \times N$ stochastic matrix. ($P(t)_{ij} \geq 0$, $P(t)\mathbb{1} = \mathbb{1}$)

$x(t)_i$ estimation of x_A by agent i at time t. $x(t) \in \mathbb{R}^N$

$$x(t+1) = P(t)x(t)\,, \qquad x(0)_i = x_i$$

$x(t) = Q(t)x(0)$, $Q(t) = \prod_{s=0}^{t-1} P(s)$,

- CONSENSUS: $Q(t) \to \mathbb{1}\rho^T \Rightarrow x(t) \to \mathbb{1}\rho^T x(0)$
- AVERAGE CONSENSUS: $\rho = N^{-1}\mathbb{1}$
- $P(t)$ ADAPTED to $\mathcal{G}$: $P(t)_{ij} > 0 \Rightarrow (j,i) \in E$.

Consensus algorithms II

$\mathcal{G} = (V, E)$ strongly connected

P stochastic $P_{ij} > 0 \ \Leftrightarrow \ (j, i) \in E$, $P_{ii} > 0 \ \forall i$

- $P(t) = P$ achieves consensus
- If $\mathbb{1}^T P = \mathbb{1}^T$ then achieves av. consensus

A possible construction:

$A_\mathcal{G} \in \{0,1\}^{N \times N}$: $(A_\mathcal{G})_{ij} = 1 \Leftrightarrow (j, i) \in E$

$D_\mathcal{G} = \mathrm{diag}(\nu_1, \ldots \nu_N), \quad \nu_i = |\{j | (j, i) \in E\}| = \sum_j (A_\mathcal{G})_{ij}$

- $P = kI + (1-k) D_\mathcal{G}^{-1} A_\mathcal{G}$ achieves consensus
- $\mathcal{G}$ regular, out-deg.=in-deg. $\Rightarrow$ av. consensus

Tsitsiklis, Cybenko, Morse, Olfati-Saber, Murray, Francis,...

Randomized algorithms

$P(t)$ chosen randomly at every time step

- PROBABILISTIC CONSENSUS:
  $Q(t) \longrightarrow \mathbb{1}\rho^T$ a.s. ρ random stochastic vector

Mean evolution: $\overline{P} = \overline{P(t)}, \ \overline{x(t)} = \overline{P}^t x(0).$

THEOREM: (Cogburn 1987) Assume that

- $\mathcal{G} = (V, E)$ strongly connected
- $\overline{P}_{ij} > 0 \ \Leftrightarrow \ (j, i) \in E,$
- $P(t)_{ii} > 0 \ \forall i$ almost surely.

Then, $P(t)$ achieves probabilistic consensus.

Example I

SYMMETRIC GOSSIP (Boyd et al. AC-2006)
$\mathcal{G}$ undirected, strongly connected

At every time t:

- Choose $i \in V$ randomly, and a neighbor j of i randomly
- i and j exchange their current estimations
- $x(t+1)_i = kx(t)_i + (1-k)x(t)_j$
- $x(t+1)_j = kx(t)_j + (1-k)x(t)_i$

$$\overline{P}_{ij} = \frac{1}{N}\left[\frac{1}{\nu_i} + \frac{1}{\nu_j}\right] k, \quad (i,j) \in E$$

$$\Downarrow$$

Average probabilistic consensus

Example II

ASYMMETRIC GOSSIP $\mathcal{G}$ strongly connected

At every time t:

- Choose $i \in V$ randomly, and an in-neighbor j of i randomly
- i receives the estimation of j
- $x(t+1)_i = kx(t)_i + (1-k)x(t)_j$

$$\overline{P}_{ij} = \frac{1}{2N}\left[\frac{1}{\nu_i} + \frac{1}{\nu_j}\right] k, \quad (i,j) \in E$$

$$\Downarrow$$

Probabilistic consensus

Average is not preserved in general!

Example III

SYNCRONOUS GOSSIP $\mathcal{G}$ strongly connected

At every time t:

- Every $i \in V$ chooses an in-neighbor j randomly
- i receives the estimation of j
- $x(t+1)_i = kx(t)_i + (1-k)x(t)_j$

$$\overline{P} = (1-k)I + kD_{\mathcal{G}}^{-1}A_{\mathcal{G}}$$

$$\Downarrow$$

Probabilistic consensus

Average is not preserved in general!

Examples IV

BROADCASTING $\mathcal{G}$ strongly connected

At every time t:

- A node $i \in V$ is chosen randomly
- i sends its estimation to all its out-neighbors
- $x(t+1)_j = kx(t)_j + (1-k)x(t)_i$, if $(i,j) \in E$

$$\overline{P} = I + \frac{k}{N}[A_{\mathcal{G}} - D_{\mathcal{G}}]$$

$$\Downarrow$$

Probabilistic consensus

Average is not preserved in general!

Complexity and performance

WHY RANDOMNESS?

- Computation and transmission consume energy. You want to keep them low.
- In many applications, an agent can only receive data from just one neighbor at a time.
- Under these limitations, random algorithms perform better

Complexity and performance

WHY RANDOMNESS?

- Computation and transmission consume energy. You want to keep them low.
- In many applications, an agent can only receive data from just one neighbor at a time.
- Under these limitations, random algorithms perform better

WHY ASYMMETRIC SCHEMES?

- Transmission links can be quite asymmetric (sensor networks)
- In many applications it is sufficient to get close to the average

Performance issues

$$d(t) = N^{-1}||x(t) - \mathbb{1}x_A(t)||^2\,, \qquad \beta(t) = |x_A(t) - x_A(0)|^2\,.$$

Notice that: $\frac{1}{N}||x(t) - \mathbb{1}x_A(0)||^2 = d(t) + \beta(t)$.

Rate of convergence can always be defined:

$$\lim_{t \to +\infty} d(t)^{1/t} \quad \text{is constant a.s.}\,.$$

It equals the second Lyapunov exponent of $P(t)$.

Is the right performance index?

Performance issues: an example

Symmetric gossip. $\mathcal{G}$ complete, $k = 1/2$, $N = 2^r$.

$$i \leftrightarrow j \implies P(t) = R^{ij} = I - 2^{-1}(e_i - e_j)(e_i - e_j)^*$$

There exist edges $(i_1, j_1), \cdots, (i_s, j_s)$ such that

$$R^{i_1, j_1} \dots R^{i_s, j_s} = N^{-1}\mathbb{1}\mathbb{1}^*$$

Case $r = 2$: $R^{2,4}R^{1,3}R^{3,4}R^{1,2} = 4^{-1}\mathbb{1}\mathbb{1}^*$

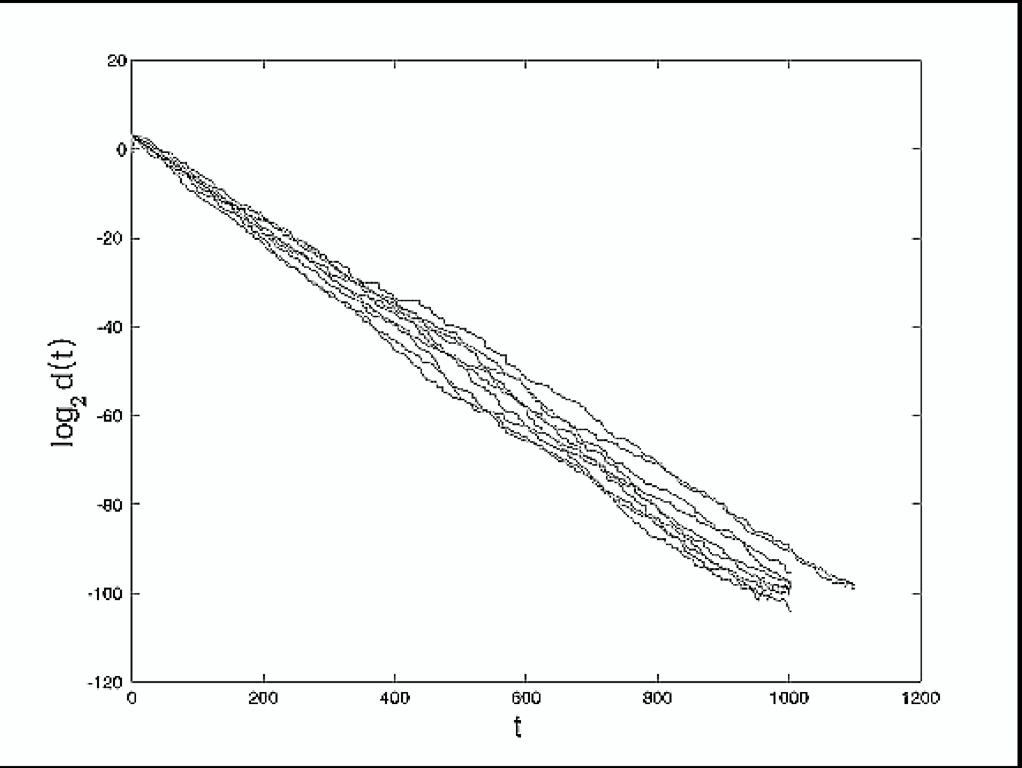
Consensus is achieved in finite time, almost surely!

Hence, $\lim_{t\to+\infty} d(t)^{1/t} = 0$ a.s.

Performance issues: an example

Plot of $\log d(t)$:

Case $N = 8$



Time to consensus $T \sim 1000$ is very big.

For time $t < T$ we have an exponential convergence with rate $\simeq 0.93$.

It can be caught by analyzing $\mathbb{E}[d(t)]$.

Mean square analysis

Another performance index:

$$R = \limsup_{t\to+\infty} (\mathbb{E}[d(t)])^{1/t} \geq \lim_{t\to+\infty} d(t)^{1/t}$$

We can write

$$\mathbb{E}[d(t)] = N^{-1} x^*(0)\Delta(t)x(0)$$

where

$$\Delta(t+1) = \mathcal{L}(\Delta(t)) := \mathbb{E}[P(0)^*\Delta(t)P(0)],$$
$$\Delta(0) = I - N^{-1}\mathbb{1}\mathbb{1}^*$$

$R =$ spectral radius of $\mathcal{L}$ on $\{S \text{ sym}\,, S\mathbb{1} = 0\}$

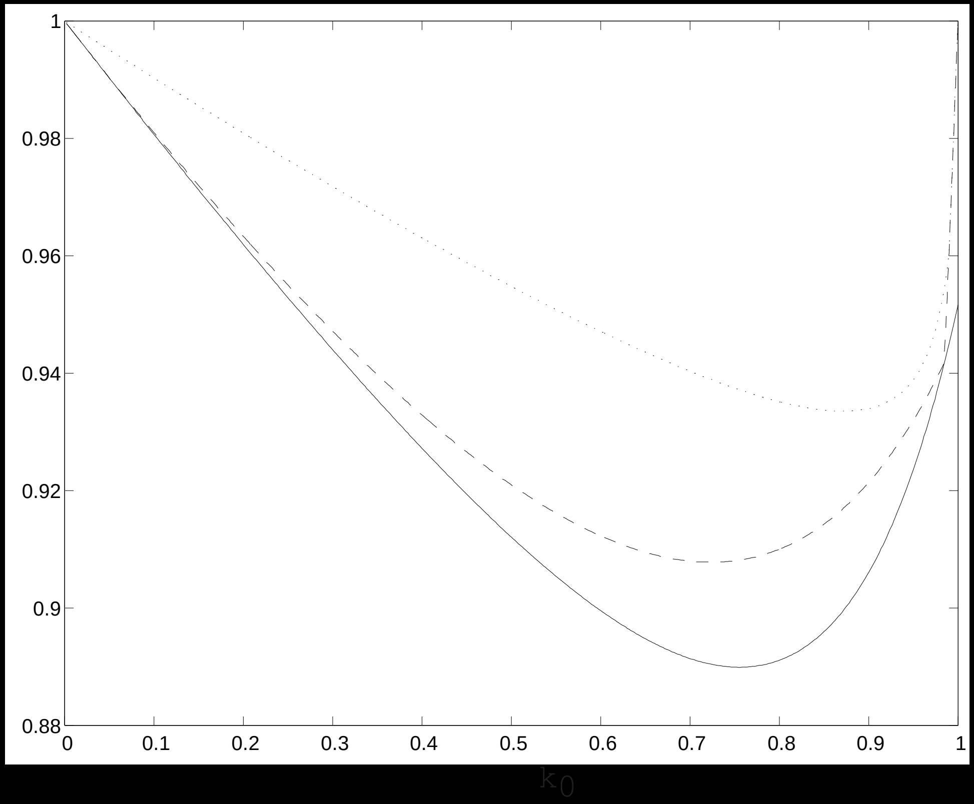
The case when $\mathcal{G}$ is complete

- SYMMETRIC GOSSIP $k = 1/2$
  $R_{sym} = 1 - \frac{1}{(N-1)}$
- SYNCR. ASYM. GOSSIP $k = 1/2$
  $R_{in} = \frac{1}{2} + O(N^{-1})$
- BROADCASTING
  $R_{br} = (1-k)^2$.

REMARKS:

- For $N = 16$, $R_{sym} \simeq 0.93$
- $R_{sym}^N \to e^{-1} < R_{in}$
- Some analytical results for Cayley graphs.

The cycle graph



R as a function of k: symmetric gossip (dashed line), syncronous gossip (dotted line), broadcasting (solid line) for $N = 20$.
For symmetric and broadcasting, rates are powered to N.

Concentration results I

Assume $\overline{P}$ achieves average consensus.

$d(t) = N^{-1}||x(t) - \mathbb{1}x_A(t)||^2, \;\; \beta(t) = |x_A(t) - x_A(0)|^2.$

$$\mathbb{P}[|d(t) - \mathbb{E}[d(t)]| \geq \delta] \leq \exp\left(-\frac{\delta^2 N^\alpha}{K||x(0)||_\infty^4 t}\right)$$

$$\mathbb{P}[\beta(t) \geq \delta] \leq \exp\left(-\frac{\delta N^\alpha}{K||x(0)||_\infty^2 t}\right).$$

Azuma's inequality + edge exposure technique.

- SYM., ASYM. GOSSIP: $\alpha = 2$
- SYNCR. GOSSIP, : $\alpha = 1$
- BROADCASTING, : $1 \leq \alpha \leq 2$.

A digression

When does $\overline{P}$ achieve average consensus?

$\overline{P}$ needs to be doubly stochastic.

GOSSIP MODELS, BROADCASTING:
$\mathrm{out}-\mathrm{degree}_i = \mathrm{in}-\mathrm{degree}_i$ for all $i \in V$.

True for symmetric (undirected) graphs and for Cayley graphs.

For the gossip models, a variation of the proposed algorithms yields $\overline{P}$ doubly stochastic whenever $\mathcal{G}$ is strongly connected.

Concentration results II

Theorem: Let $T_N = O\left(N^{\alpha-\epsilon}\right)$. Then, a. s.,

$$\limsup_{N\to+\infty} \sup_{t\le T_N} |d(t) - \mathbb{E}[d(t)]| = 0$$

$$\limsup_{N\to+\infty} \sup_{t\le T_N} \beta(t) = 0\,.$$

- For $t \le T_N$ and large N, $d(t)$ is well approximated by its mean $\mathbb{E}[d(t)]$. This gives motivation for considering the rate of convergence of $\mathbb{E}[d(t)]$ for analysis and optimization.
- For $t \le T_N$ and large N, the displacement $\beta(t)$ is uniformly small.

Is $t \le T_N$ the interesting time range?

Concentration results III

ϵ-averaging time:

$$T_N(\epsilon) = \inf\{t \in \mathbb{N} \mid \mathbb{P}(||x(t) - x_A(0)\mathbb{1}|| \geq \epsilon ||x(0)||) \leq \epsilon\}$$

In many cases: $T_N(\epsilon) \asymp N^\theta \ln \epsilon^{-1}$.

- SYM. GOSSIP:(Boyd et al.)
  - complete graph: $\theta = 1$
  - n-torus: $\theta = n^{-1}(2+n)$.
- ASYM. GOSSIP:Same results (for complete).
- SYNCR. GOSSIP, BROAD.:
  - complete graph: $\theta = 0$
  - (Conjectured) n-torus: $\theta = 2/n$ and $\theta = n^{-1}(2+n)$.

Concentration results IV

Concentration results valid up to $T_N(\epsilon)$ if:

- ALL GOSSIP: complete, n-torus for $n \geq 3$.
- BROADCASTING: n-torus for $n \geq 3$.

$$\limsup_{N \to +\infty} \sup_{t \leq T_N(\epsilon)} |d(t) - \mathbb{E}[d(t)]| = 0$$

$$\limsup_{N \to +\infty} \sup_{t \leq T_N(\epsilon)} \beta(t) = 0\,.$$

- For the broadcasting model with complete graph: $\beta(t) \not\to 0$ for $N \to +\infty$.
- What about the n-torus for $n = 1, 2$??

Large time behavior

$$\beta(t) \to \beta(\infty) = |(\rho^* - N^{-1}\mathbb{1}^*)x(0)|^2$$

$$\mathbb{E}[\beta(\infty)] = x(0)^* B x(0), \quad B = \mathbb{E}[\rho\rho^*] - N^{-2}\mathbb{1}\mathbb{1}^*$$

$$\mathcal{L}(\mathbb{E}[\rho\rho^*]) = \mathbb{E}[\rho\rho^*], \quad \mathcal{L} \Rightarrow B$$

SYNCR. GOSSIP complete graph, $k = 1/2$:

$$B = O(N^{-2})(I - N^{-1}\mathbb{1}\mathbb{1}^*)$$

$$\mathbb{E}[\beta(\infty)] = N^{-1}||x(0)||^2 O(N^{-1})$$

$$\limsup_{N\to+\infty} \sup_t \beta(t) = 0$$

Summary

- Random consensus algorithms which do not need symmetric communication but yield consensus near the average.
- In some cases precise analytical results
- Concentration results which validate the mean square analysis for the purpose of analysis and optimization
- Theory also applies to deterministic consensus models in the context of random packet drops or agent failures.

Future research

- Complete the analysis for Cayley graphs
- Extend analysis to other examples (random graphs)
- Investigate those cases where concentration seems to fail.
- Threshold phenomena.
- Analysis for other consensus algorithms (with memory, Van Roy’s model)
- Models with noisy or quantized information.