Distributed algorithms over graphs

Consensus, gossiping, estimation

Paolo Frasca

CNRS
Grenoble, France

Peyresq Summer School 2023

Outline

1. Algorithms for distributed computation/estimation/control
2. The (average) consensus problem
   - Linear consensus dynamics
   - Variations & applications
3. Gossiping and randomized algorithms
4. Estimation from relative measurements
   - Fundamental limitations: error of optimal estimator
   - Gradient algorithm
   - Gossiping randomized algorithm

1. Algorithms for distributed computation/estimation/control

2. The (average) consensus problem
   - Linear consensus dynamics
   - Variations & applications

3. Gossiping and randomized algorithms

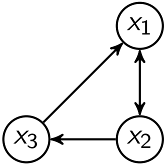
4. Estimation from relative measurements
   - Fundamental limitations: error of optimal estimator
   - Gradient algorithm
   - Gossiping randomized algorithm

What are distributed algorithms?

We are given a directed graph $G = (V, E)$ with N nodes (a.k.a. agents)

Each node $v \in V$ has the ability to

- store information ('state' x_v)
- perform computations to update their state
- communicate with neighbors in $\mathcal{N}_v$



The nodes have a task that requires them to cooperate. For instance,

- if the nodes are *sensors*, computing a function $f(y_1, \ldots, y_N)$ that depends on measurements y_u taken by all nodes
- if the nodes are *mobile robots*, reach a rendez-vous position or deploy effectively in an environment.

The goal of the algorithms is to achieve the task despite the communication constraints.
Performance is measured as speed of convergence, accuracy in reaching the task, computation and communication cost.

The consensus problem

Prototypical coordination problem is consensus stabilization, i.e. making agents' states $x_v(t)$ converge to a common value:

$$\lim_{t\to\infty} x_v(t) = \alpha \text{ for some } \alpha \in \mathbb{R}$$

by defining a dynamics that respects the communication constraints.

We require translation invariance: if $x_v(0)$ becomes $x_v(0) + b$, then the consensus point should become $\alpha + b$

The average consensus problem

An important special case of consensus is average consensus, i.e. making agents' states $x_v(t)$ converge to the mean of their original positions

$$\lim_{t \to \infty} x_v(t) = \frac{1}{N} \sum_{u \in V} x_u(0)$$

Motivation

1. Going to the average is likely to be the "best" choice for mobile agents that seek to minimize effort
2. The average is immediately relevant in networks of sensors that take noisy measurements $\theta + n_v$ of a true state θ.

A linear algorithm

We seek a linear algorithm

$$x_v(t+1) = \sum_{w \in V} P_{vw} x_w(t)$$

or equivalently

$$x(t+1) = Px(t)$$

so that

$$x(t) = P^t x(0)$$

Necessary conditions on P

1. We want consensus to be an equilibrium: then we need $P\mathbf{1} = \mathbf{1}$ (each row of P sums to 1).
2. We want the dynamics to respect the graph: P needs to have the same structure as the adjacency matrix of G

A linear algorithm

We seek a linear algorithm

$$x_v(t+1) = \sum_{w \in V} P_{vw} x_w(t)$$

or equivalently

$$x(t+1) = Px(t)$$

so that

$$x(t) = P^t x(0)$$

Solving the consensus problem:

If G has a globally reachable vertex $v^\star$, then the consensus problem can be solved by any matrix $P \in \mathbb{R}^{V \times V}$ such that:

(Pa) $P_{vw} \geq 0$ for every $v, w \in V$; *(non-negativity)*

(Pb) $P\mathbf{1} = \mathbf{1}$; *(translation invariance)*

(Pc) *For every* $v \neq w$, $P_{vw} > 0 \Leftrightarrow (v, w) \in E$; *(P adapted to the graph)*

(Pd) $P_{v^\star v^\star} > 0$.

Stochastic matrices and Laplacians

Positive matrices such that $P\mathbf{1} = \mathbf{1}$ are called stochastic matrices and appear in many different contexts, e.g. Markov chains.

Laplacian interpretation

Being P stochastic, its Laplacian is $L(P) = I - P$.

Then, we can write $x(t+1) = x(t) - L(P)x(t)$, which becomes

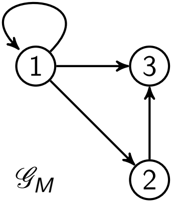
$$x_v(t+1) = x_v(t) + \sum_w P_{vw}(x_w(t) - x_v(t))$$

Relative information is enough!

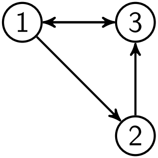
Some more graph theory notions

Given a positive matrix M, we can associate a graph $\mathscr{G}_M$ with edges corresponding to non-zero entries

$$M = \begin{bmatrix} 0.3 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$



Aperiodicity: A vertex is said to be aperiodic if the lengths of the cycles going through it are co-prime.



General convergence result

Let P be a stochastic matrix such that $\mathscr{G}_P$ has a globally reachable aperiodic node. Then, the following two equivalent facts hold true.

1. For any initial condition $x(0) = x^0 \in \mathbb{R}^V$, there exist $\alpha \in \mathbb{R}$ such that

$$x(t) = P^t x(0) \to \alpha \mathbf{1} \qquad t \to +\infty.$$

   All components $x_v(t)$ converge to the same value α.

2. There exists a vector $\pi \in \mathbb{R}^V$ such that $\pi \geq 0$, $\sum_v \pi_v = 1$, and

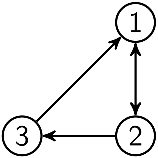
$$\lim_{t \to +\infty} P^t = \mathbf{1}\pi^\top.$$

   P^t converges to a matrix with all rows equal to row vector $\pi^\top$.

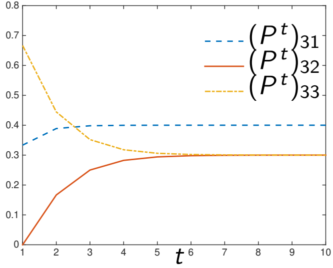
Furthermore, π is the only normalized vector such that $\pi^\top P = \pi^\top$ and $\alpha = \pi^\top x(0)$.

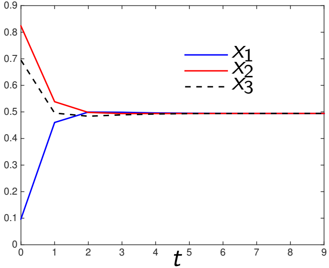
Example

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 2/3 \end{bmatrix}$$



The invariant probability π (such that $\pi^\top P = \pi^\top$) is $\pi = \left(\frac{2}{5}, \frac{3}{10}, \frac{3}{10}\right)^\top$.





Left: entries of the third row of P^t that converges to $\pi^\top$.
Right: dynamics from random initial conditions to consensus.

Convergence speed

Spectrum of P

Provided $\mathscr{G}_P$ has a globally reachable node,

(i) 1 is an algebraically simple eigenvalue whose eigenspace is generated by $\mathbf{1}$;

(ii) Any other eigenvalue μ of P is such that $|\mu| < 1$.

Spectrum and convergence rate

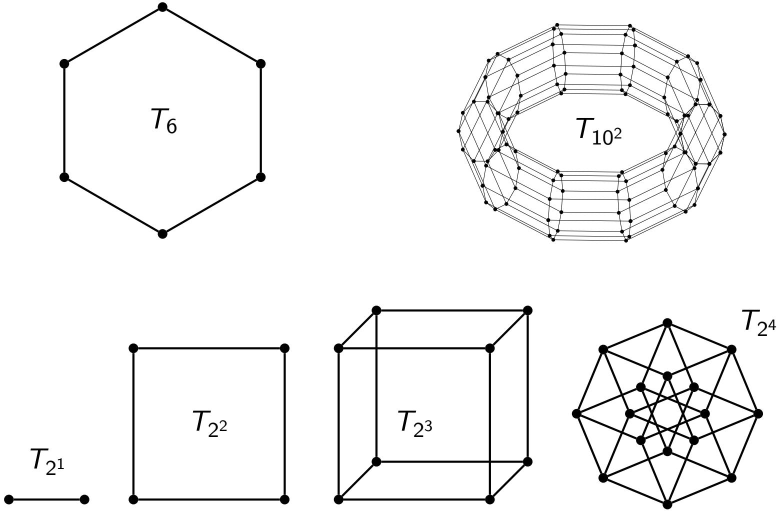
If $P \in \mathbb{R}^{V \times V}$ is a symmetric stochastic matrix such that $\mathscr{G}_P$ is strongly connected and aperiodic, then

$$||(P^t - N^{-1}\mathbf{1}\mathbf{1}^\top)x_0||_2 \leq {\rho_2}^t ||x_0||_2$$

where ρ_2 is the second-largest eigenvalue of P.

A relevant family of graphs

T_{M^d} is the toroidal grid with dimension d and $N = M^d$ nodes



Graph spectra and convergence speed

Since toroidal graphs T_{M^d} are regular (with degree $k = 2d$), we can define

$$P = \frac{1}{2d+1}(I + A) = I - \frac{1}{2d+1}L.$$

$$\text{Then, } \rho_2 = 1 - \frac{\lambda_2}{2d+1}$$

Dimension and convergence time:

$d = 1$: $\rho_2 = \frac{1}{3} + \frac{2}{3}\cos\frac{2\pi}{N} = 1 - \frac{2\pi^2}{N^2} + o(\frac{1}{N^2})$

Convergence time $\Theta(N^2)$

$d = 2$: $\rho_2 = \frac{1}{5}(3 + 2\cos\frac{2\pi}{\sqrt{N}})$ Convergence time $\Theta(N)$

any d: Convergence time $\Theta(N^{2/k})$

hypercube: $\rho_2 = 1 - \dfrac{2}{\log_2 N + 1}$ Convergence time $\Theta(\log N)$

Constructing stochastic matrices

Starting from A, we need to construct the stochastic matrix P.

In order to solve the *average* consensus problem, we also need $\mathbf{1}^\top P = \mathbf{1}^\top$ (this is possible if and only if the graph is strongly connected).

Methods:

- Simple random walk: $P = D^{-1}A$
- Lazy random walk: $P = (1-\tau)I + \tau D^{-1}A$ for $\tau \in (0,1)$
- Metropolis random walk (only for undirected graphs): For any $v \neq w$,

$$P_{vw} := A_{vw} \min\left\{\frac{1}{\deg_v}, \frac{1}{\deg_w}\right\} \qquad P_{vv} = 1 - \sum_{w \neq v} P_{vw}.$$

Time-varying graphs

Time-dependent graphs

A sequence of graphs $G(t) = (V, E(t))$ ensures consensus if "enough connectivity across time" is guaranteed.

However, none of the individual graphs needs to be connected!

Time-varying graphs

Time-dependent graphs

A sequence of graphs $G(t) = (V, E(t))$ ensures consensus if "enough connectivity across time" is guaranteed.
However, none of the individual graphs needs to be connected!

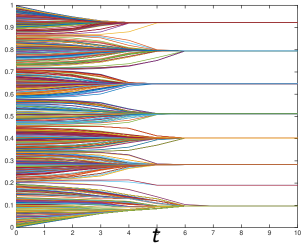
State-dependent graphs

Examples of state-dependent graphs $G(x(t)) = \big(V, E(x(t))\big)$:

- $\mathcal{N}_u(t) = \{v \in V : |x_u(t) - x_v(t)| < d\}$ (disk graph)
- $\mathcal{N}_u(t) = \{k\text{-nearest neighbors of } u \text{ at time } t\}$

Applications:

- Opinion dynamics in social networks with "bounded confidence" between the individuals
- communication in ad hoc mobile networks
- "naive" clustering



Consensus with limited data rate

Communication may be quantized because of limited data rate in the communication between the nodes.

1. Using uniform quantizer $q : \mathbb{R} \to \mathbb{Z}$, $q(x) = \lfloor x + \frac{1}{2} \rfloor$, we can use

$$x_v(t+1) = \sum_{w \in V} P_{vw} q(x_w(t))$$

or

$$x_v(t+1) = P_{vv} x_v(t) + \sum_{w \neq v} P_{vw} q(x_w(t))$$

or

$$x_v(t+1) = x_v(t) - (1 - P_{vv}) q(x_v(t)) + \sum_{w \neq v} P_{vw} q(x_w(t))$$

Which one is better for average consensus?

2. Using a dynamic quantizer, 1 bit of exchange per time step is enough!

Randomized updates

We consider time-varying consensus algorithms in which the update matrix is selected at each time step by a random process:

$$x(t+1) = P(t)x(t) \qquad t \in \mathbb{Z}_{\geq 0},$$

where

- $P(t)$ is a stochastic matrix for each $t \geq 0$
- $\{P(t)\}_{t\geq 0}$ is a sequence of independent and identically distributed random variables.

Example: In lossy networks, messages exchanged between nodes can be randomly lost.

Gossiping algorithm

Let a symmetric/undirected graph $G=(V,E)$ be given, and for each time step $t \geq 0$, let an edge (v,w) be chosen according to a uniform distribution over E. Define

$$x_v(t+1) = \frac{1}{2}x_v(t) + \frac{1}{2}x_w(t),$$
$$x_w(t+1) = \frac{1}{2}x_w(t) + \frac{1}{2}x_v(t),$$
$$x_u(t+1) = x_u(t), \quad \text{for } u \neq v, w.$$

This is a simple, asynchronous algorithm.

This algorithm can be written in the general form by defining

$$\Pr[P(t) = P^{(v,w)}] = \frac{1}{|E|} \qquad P^{(v,w)} = I - \frac{1}{2}(e_v e_v^\top - e_v e_w^\top - e_w e_v^\top + e_w e_w^\top),$$

where e_u is the u-th vector of the canonical basis of $\mathbb{R}^V$

Dynamics of the expectation

What is the dynamics of $\mathbb{E}[x(t)]$?

By independence among $P(t)$, $\mathbb{E}[x(t+1)|x(t)] = \mathbb{E}[P(t)]x(t)$ for all t. Then,

$$\mathbb{E}[x(t+1)] = \bar{P}\,\mathbb{E}[x(t)] \qquad \text{with } \bar{P} := \mathbb{E}[P(t)]$$

If the graph associated with $\bar{P}$ has a globally reachable aperiodic node, then $\mathbb{E}[x(t)]$ converges to a consensus point $c\mathbf{1}$.
The convergence rate is given by $\rho_2(\bar{P})$ and $c = v^\top x(0)$, where v is the normalized dominant left eigenvector of $\bar{P}$.

Convergence of the expected dynamics does not, by itself, guarantee convergence of the random dynamics...

Convergence result

If the matrices $P(t)$ are i.i.d. and have positive diagonal entries, then the following three facts are equivalent:

1. for every initial condition, there exists a scalar random variable x_∞ such that $x(t)$ converges almost surely to $x_\infty \mathbf{1}$;
2. $\rho_2(\bar{P}) < 1$;
3. the "expected graph" $\mathscr{G}_{\bar{P}}$ has a globally reachable node.

Convergence speed

Let us denote the current empirical variance as

$$x_{\text{var}}(t) := \frac{1}{N}||x(t) - x_{\text{ave}}(t)\mathbf{1}||^2 = \frac{1}{N}||\Omega x(t)||^2,$$

where $\Omega = I - \frac{1}{N}\mathbf{1}\mathbf{1}^\top$, and define the mean square rate of convergence as

$$R := \sup_{x(0)} \limsup_{t\to+\infty} \mathbb{E}[x_{\text{var}}(t)]^{1/t}.$$

Mean square convergence rate:

$$\rho_2(\bar{P})^2 \le R \le \text{sr}\left(\mathbb{E}[P(t)^\top \Omega P(t)]\right),$$

where $\text{sr}(\cdot)$ denotes the spectral radius.

Gossiping algorithm: generalization and convergence rate

Let a weighted graph $G = (I, E, W)$ and $q \in (0,1)$ be given, such that W is *symmetric* and $\mathbf{1}^\top W \mathbf{1} = 1$.
For every $t \geq 0$, one edge $(v,w) \in E$ is sampled from a distribution such that the probability of selecting (v,w) is W_{vw}. Then,

$$\begin{aligned} x_v(t+1) &= (1-q)\,x_v(t) + q\,x_w(t) \\ x_w(t+1) &= (1-q)\,x_w(t) + q\,x_v(t) \\ x_u(t+1) &= x_u(t) \quad \text{for } u \neq v, w. \end{aligned}$$

This generalised gossiping converges almost surely with rate

$$1 - 4q\lambda \leq R \leq 1 - 4q(1-q)\lambda,$$

with λ the smallest non-zero eigenvalue of $L(W)$

Problem statement: relative estimation

- V is a set of sensors of cardinality N
- $\xi \in \mathbb{R}^V$ is an unknown vector
- each sensor u obtains noisy relative measurements with some other nodes v,

$$b_{uv} = \xi_u - \xi_v + \eta_{uv} \qquad \eta_{uv} \text{ are i.i.d. noise}$$

Goal: *for each sensor $v \in V$, estimate the scalar value ξ_v*

Applications:

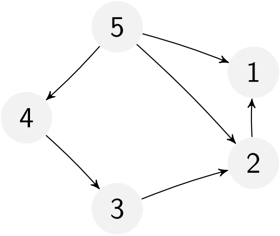
- clock synchronization
- self-localization of moving robots
- statistical ranking (e.g. sport tournaments)

Relative estimation as a graph problem

Measurements $\longrightarrow$ edges E of an oriented connected graph $G = (V, E)$

Incidence matrix $A \in \{0, \pm 1\}^{E \times V}$

$$A_{ew} = \begin{cases} +1 & \text{if } e = (v, w) \\ -1 & \text{if } e = (w, v) \\ 0 & \text{otherwise} \end{cases}$$



$$A = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

Laplacian matrix

$$L = A^\top A = \begin{bmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 3 & -1 & 0 & -1 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ -1 & -1 & 0 & -1 & 3 \end{bmatrix}$$

Relative estimation as a least-squares problem

We define the least-squares problem

$$\min_z ||Az - b||^2$$

Matrix A has rank $N-1 \implies$ affine space of solutions (up to a constant)

The minimum-norm solution $x^\star = L^\dagger A^\top b$ best explains the measurements

How good is the estimate $x^\star$?

How can the sensor network compute $x^\star$?

Estimator error and effective resistance

Estimator error:

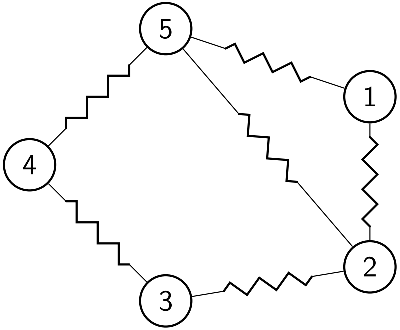
$$\frac{1}{N}\mathbb{E}\|x^\star - \xi\|^2 = \sigma^2 \frac{1}{N}\sum_{i\geq 2}\frac{1}{\lambda_i}$$

where $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_N$ are the eigenvalues of L

Remark from graph theory:

$$\frac{1}{N}\sum_{i\geq 2}\frac{1}{\lambda_i} = R_{\text{ave}}(G)$$

$R_{\text{ave}}(G)$ is the average of all resistances between all pairs of nodes if the graph was an electrical network of unit resistors



Effective resistance of d-dimensional graphs

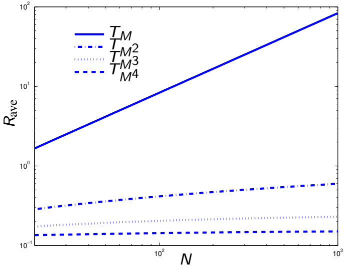
$$R_{\text{ave}}(T_M) \sim \frac{1}{12} M \quad \text{for } M \to +\infty$$

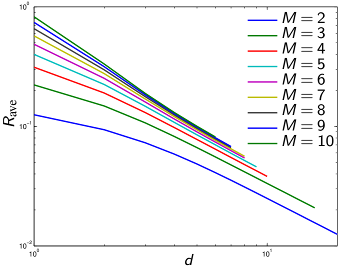
$$R_{\text{ave}}(T_{M^2}) \sim \frac{1}{4\pi} \ln M \quad \text{for } M \to +\infty$$

$$R_{\text{ave}}(T_{M^d}) = O(1) \quad \text{for } M \to +\infty \quad \text{if } d \geq 3$$

$$R_{\text{ave}}(T_{M^d}) = \Theta\left(\frac{1}{d}\right) \quad \text{for } d \to +\infty,\ M \ \text{fixed}$$

$$R_{\text{ave}}(T_{2^d}) \sim \frac{1}{d} \quad \text{for } d \to +\infty$$





More connected measurement graphs ensure better performance

1. Algorithms for distributed computation/estimation/control

2. The (average) consensus problem
   - Linear consensus dynamics
   - Variations & applications

3. Gossiping and randomized algorithms

4. Estimation from relative measurements
   - Fundamental limitations: error of optimal estimator
   - Gradient algorithm
   - Gossiping randomized algorithm

Gradient descent algorithm

The gradient of $\Psi(z) = ||Az - b||^2$ is $\nabla\Psi(z) = 2Lz - 2A^\top b$

Then, choosing a parameter $\tau > 0$,

$$\begin{cases} x(0) = 0 \\ x(t+1) = (I - \tau L)x(t) + \tau A^\top b \end{cases}$$

Convergence:

$\lim_{t\to+\infty} x(t) = x^\star$ if $\tau < 1/d_{\max}$, where $d_{\max}$ is the largest degree in G

The gradient algorithm is

- *distributed*, *i.e.*, each node only needs for its update to know the states of its neighbors (measurement graph = communication graph)
- *synchronous*, *i.e.*, all nodes update their states at the same time

Finite-time optimality (i.e. oversmoothing)

Assume the intrinsic values ξ_vs are i.i.d. with zero mean and variance ν^2

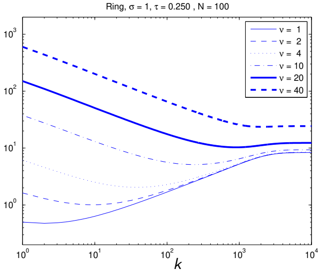
$$J(t) := \frac{1}{N}\mathbb{E}_\xi \mathbb{E}_\eta \|x(t) - \xi\|^2 \qquad \text{expected cost}$$

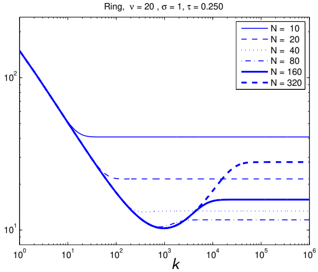
Eventual Monotonicity:

Assume $\tau < 1/d_{\max}$. If $t \geq \dfrac{\nu^2}{\tau\sigma^2}$, then $J(t+1) \geq J(t)$

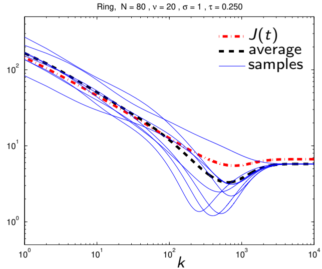
- the cost $J(t)$ has a minimum at a finite time $t_{\min}$
- $t_{\min}$ has an upper bound which does not depend on N or on G

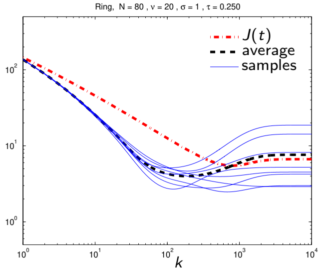
Role of parameters





Sample averages with fixed initial condition





An asynchronous randomized algorithm

We take a pairwise gossiping approach

Fix a real number $\gamma \in (0,1)$
At every time instant $t \in \mathbb{Z}_+$, an edge $(u,v) \in E$ is sampled according to

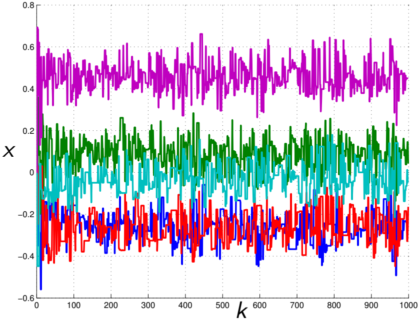
$$\mathbb{P}[(u,v) \text{ is selected at time } t] = \frac{1}{|E|}$$

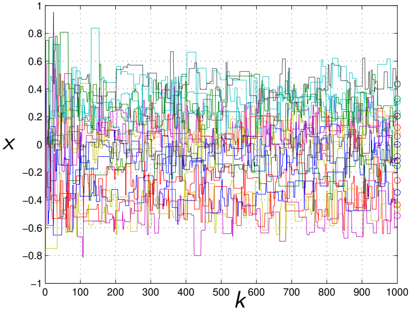
and the states are updated according to

$$\begin{aligned}
x_u(t+1) &= (1-\gamma)x_u(t) + \gamma x_v(t) + \gamma b_{(u,v)} \\
x_v(t+1) &= (1-\gamma)x_v(t) + \gamma x_u(t) - \gamma b_{(u,v)} \\
x_w(t+1) &= x_w(t) \qquad \text{if } w \notin \{u,v\}
\end{aligned}$$

Simulations: no convergence

The states $x(t)$ persistently oscillate!!!





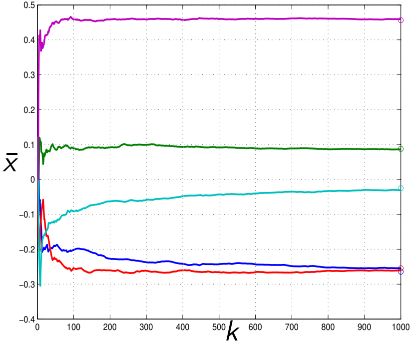
Can we still use this algorithm?

Countermeasure: time-averages

Time-averages $\bar{x}(t) := \frac{1}{t+1}\sum_{s=0}^{t} x(s)$

smooth out the oscillations

$\Longrightarrow \bar{x}(t) \to x^\star$ as $t \to +\infty$



thanks to the *ergodicity* of $x(t)$:

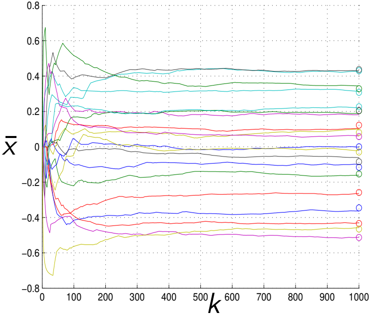
- sample averages $\Longleftrightarrow$ time averages
- the average dynamics is $\mathbb{E}[x(t+1)] = \left(I - \frac{\gamma}{|E|}L\right)\mathbb{E}[x(t)] + \frac{\gamma}{|E|}A^\top b$
- $\mathbb{E}[x(t)] \to x^\star$ as $t \to +\infty$

Countermeasure: time-averages

Time-averages $\bar{x}(t) := \frac{1}{t+1} \sum_{s=0}^{t} x(s)$

smooth out the oscillations

$\Longrightarrow \bar{x}(t) \to x^\star$ as $t \to +\infty$



thanks to the *ergodicity* of $x(t)$:

- sample averages $\Longleftrightarrow$ time averages
- the average dynamics is $\mathbb{E}[x(t+1)] = \left(I - \frac{\gamma}{|E|}L\right)\mathbb{E}[x(t)] + \frac{\gamma}{|E|}A^\top b$
- $\mathbb{E}[x(t)] \to x^\star$ as $t \to +\infty$

Some references

Fagnani, F. and Frasca, P. (2018). *Introduction to Averaging Dynamics over Networks*. Springer.

Frasca, P., Carli, R., Fagnani, F., and Zampieri, S. (2009). Average consensus on networks with quantized communication. *International Journal of Robust and Nonlinear Control*, 19(16):1787–1816.

Levin, D. A., Peres, Y., and Wilmer, E. L. (2009). *Markov Chains and Mixing Times*. American Mathematical Society.

Morarescu, I.-C. and Girard, A. (2010). Opinion dynamics with decaying confidence: Application to community detection in graphs. *IEEE Transactions on Automatic Control*, 56(8):1862–1873.

Ravazzi, C., Frasca, P., Tempo, R., and Ishii, H. (2014). Ergodic randomized algorithms and dynamics over networks. *IEEE Transactions on Control of Network Systems*, 2(1):78–87.

Rossi, W. S., Frasca, P., and Fagnani, F. (2012). Transient and limit performance of distributed relative localization. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pages 2744–2748. IEEE.