Distributed Estimation with Relative Measurements

Fundamental Limitations, Algorithms, Application to Power Systems

Paolo Frasca

UNIVERSITY OF TWENTE.

based on joint works with
F. Fagnani, H. Ishii, C. Ravazzi, W.S. Rossi, and R. Tempo

partly supported by
Joint International Lab COOPS and by JST CREST

SICE International Symposium on Control Systems 2015
March 4-7, 2015, Tokyo, Japan

Outline

1. Estimation from relative measurements
   - Problem statement & graph representation
   - Least-squares formulation
   - Fundamental limitations: error of optimal estimator
2. Gradient algorithm
   - Finite-time optimality of the expected error
3. Ergodic randomized algorithm
   - Asynchronous *gossip* randomization
   - Oscillations and ergodicity: sample-averages from time-averages
   - An asynchronous distributed algorithm exploiting ergodicity
4. Application & extension: power systems

Problem statement: relative estimation

- $\mathcal{V}$ is a set of sensors of cardinality N
- $\xi \in \mathbb{R}^{\mathcal{V}}$ 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 $v \in \mathcal{V}$, estimate the scalar value ξ_v*

Applications:

- clock synchronization

  A. Giridhar and P. R. Kumar. Distributed clock synchronization over wireless networks: Algorithms and analysis. In *IEEE Conference on Decision and Control*, pages 4915–4920, San Diego, CA, USA, December 2006
- self-localization of mobile robots

  P. Barooah and J. P. Hespanha. Estimation from relative measurements: Algorithms and scaling laws. *IEEE Control Systems Magazine*, 27(4):57–74, 2007
- statistical ranking in databases

  B. Osting, C. Brune, and S. J. Osher. Optimal data collection for improved rankings expose well-connected graphs. *Journal of Machine Learning Research*, 15:2981–3012, 2014

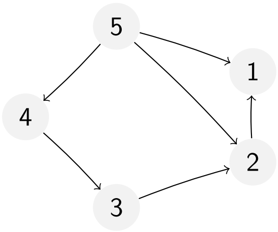
Relative estimation as a graph problem

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

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

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

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



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

Questions:

Q1 How good is the estimate $x^\star$?

Q2 How can the sensor network compute $x^\star$?

Estimator error and effective resistance

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

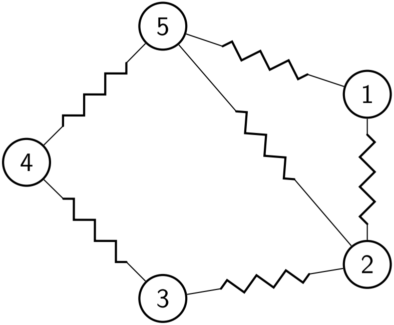
where $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_N$ are the eigenvalues of L
σ^2 is variance of noise

Observation from graph theory:

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

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

The error is determined by the topology of the measurement graph: e.g., the scaling in N depends on the graph dimension



Gradient algorithm

Gradient descent algorithm

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

We define, choosing a parameter $\tau > 0$,

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

Proposition (Convergence)

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

Gradient descent algorithm

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

We define, choosing a parameter $\tau > 0$,

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

Proposition (Convergence)

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

The gradient algorithm is

- *distributed*: each node only needs to know the states of its neighbors

$$x(k+1) = \underbrace{(I - \tau L)}_{\text{consensus algorithm}} x(k) + \underbrace{\tau A^\top b}_{\text{constant input}}$$

- *synchronous*: all nodes update their states at the same time

Finite-time optimality of the expected error

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

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

Finite-time optimality of the expected error

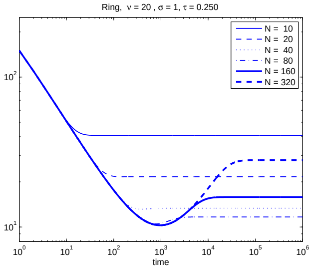
Assume ξ_vs are i.i.d. with zero mean and variance ν^2

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

Results:

- If $k \geq \dfrac{\nu^2}{\tau\sigma^2}$, then $J(k+1) \geq J(k)$ (eventual increase)
  - the error $J(k)$ has a minimum at a finite time $k_{\min}$
  - $k_{\min}$ has an upper bound which does not depend on N or on $\mathcal{G}$

Surprising conclusion:
the algorithm should not be run until convergence, but stopped earlier, irrespective of the measurement graph!



W. S. Rossi, P. Frasca, and F. Fagnani. Limited benefit of cooperation in distributed relative localization. In *IEEE Conference on Decision and Control*, pages 5427–5431, Florence, Italy, December 2013

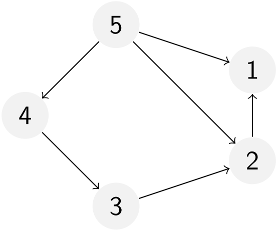
Ergodic randomized algorithm

Asynchronous randomized algorithm

We take a pairwise "gossip" approach

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

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

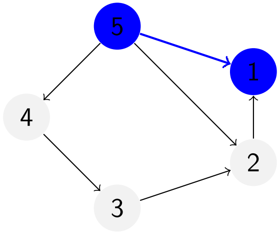


Asynchronous randomized algorithm

We take a pairwise "gossip" approach

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

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



Asynchronous randomized algorithm

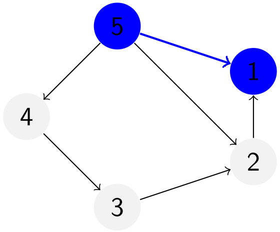
We take a pairwise "gossip" approach

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

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

and the states are updated:

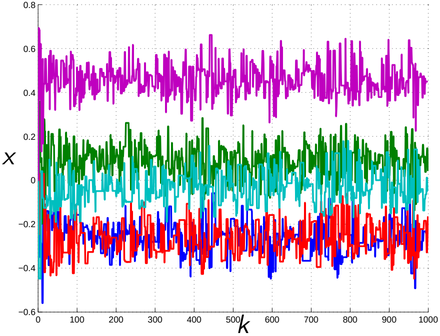
$$x_u(k+1) = (1-\gamma)x_u(k) + \gamma x_v(k) + \gamma b_{(u,v)}$$
$$x_v(k+1) = (1-\gamma)x_v(k) + \gamma x_u(k) - \gamma b_{(u,v)}$$
$$x_w(k+1) = x_w(k) \qquad \text{if } w \notin \{u,v\}$$

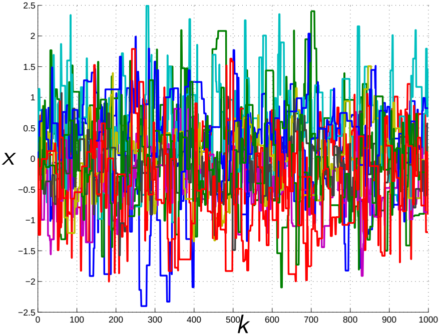


This is not a standard coordinate gradient
This reminds a gossip consensus algorithm with a constant input

Simulations: no convergence

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



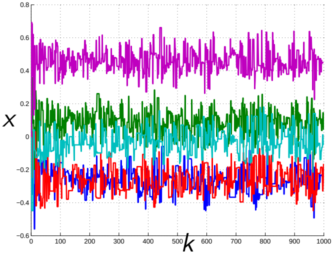


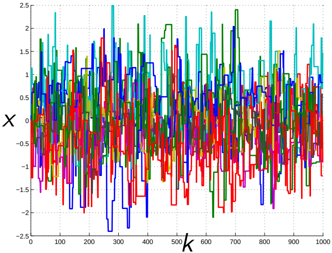
Can we still use this algorithm?

Countermeasure: time-averages

Time-averages smooth out the oscillations

$$\bar{\bar{x}}(k) := \frac{1}{k+1}\sum_{\ell=0}^{k} x(\ell) \quad \Longrightarrow \bar{\bar{x}}(k) \to x^\star \text{ as } k \to +\infty$$

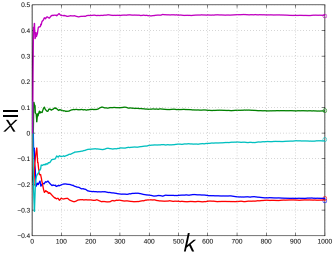


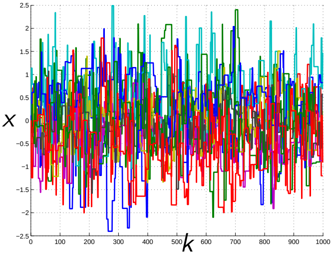


Countermeasure: time-averages

Time-averages smooth out the oscillations

$$\overline{\overline{x}}(k) := \frac{1}{k+1}\sum_{\ell=0}^{k} x(\ell) \quad \Longrightarrow \overline{\overline{x}}(k) \to x^\star \text{ as } k \to +\infty$$

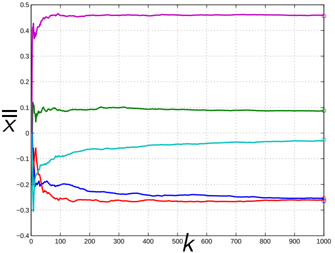


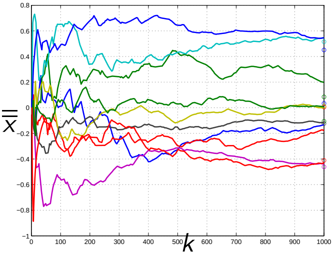


Countermeasure: time-averages

Time-averages smooth out the oscillations

$$\overline{\overline{x}}(k) := \frac{1}{k+1}\sum_{\ell=0}^{k} x(\ell) \quad \Longrightarrow \overline{\overline{x}}(k) \to x^\star \text{ as } k \to +\infty$$

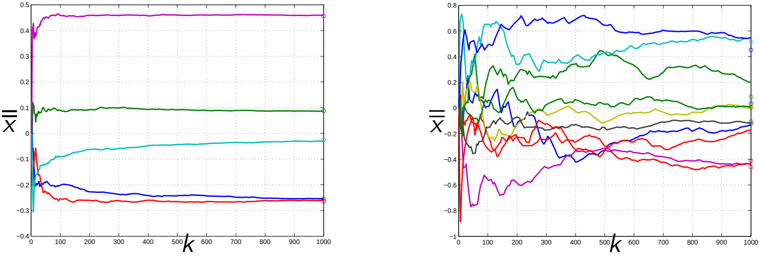




Countermeasure: time-averages

Time-averages smooth out the oscillations

$$\overline{\overline{x}}(k) := \frac{1}{k+1} \sum_{\ell=0}^{k} x(\ell) \quad \Longrightarrow \overline{\overline{x}}(k) \to x^{\star} \text{ as } k \to +\infty$$



thanks to:

- *ergodicity* of $x(\cdot)$: sample averages $\Longleftrightarrow$ time averages
- simple average dynamics (like the gradient algorithm)

$$\mathbb{E}[x(k+1)] = \left(I - \frac{\gamma}{|\mathcal{E}|} L \right) \mathbb{E}[x(k)] + \frac{\gamma}{|\mathcal{E}|} A^{\top} b$$

- $\mathbb{E}[x(k)] \to x^{\star}$ as $k \to +\infty$

Local vs global clocks

To compute time-averages $\overline{\overline{x}}$ each sensor needs to know the absolute time k

We can overcome this drawback by defining two auxiliary dynamics:

- local times: $\kappa_w(0) = 1$ for all $w \in \mathcal{V}$

$$\begin{aligned} \kappa_u(k+1) &= \kappa_u(k) + 1 \\ \kappa_v(k+1) &= \kappa_v(k) + 1 \\ \kappa_w(k+1) &= \kappa_w(k) \qquad \text{if } w \notin \{u, v\} \end{aligned}$$

- "local" time-averages: $\overline{x}_w(0) = 0$ for all $w \in \mathcal{V}$

$$\begin{aligned} \overline{x}_u(k+1) &= \frac{1}{\kappa_u(k+1)}\big(\kappa_u(k)\overline{x}_u(k) + x_u(k+1)\big) \\ \overline{x}_v(k+1) &= \frac{1}{\kappa_v(k+1)}\big(\kappa_v(k)\overline{x}_v(k) + x_v(k+1)\big) \\ \overline{x}_w(k+1) &= \overline{x}_w(k) \qquad \text{if } w \notin \{u, v\} \end{aligned}$$

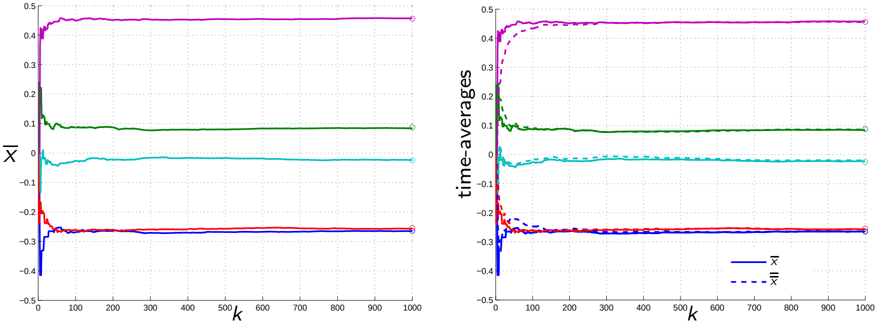
Convergence of "local" time-averages

We obtain a correct algorithm:

Theorem (Ergodicity & Convergence)

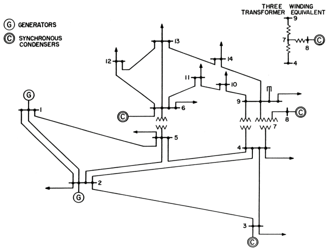
$$\lim_{k\to+\infty} \overline{x}(k) = x^\star \textit{ almost surely}$$

$$\lim_{k\to+\infty} \mathbb{E}[||\overline{x}(k) - x^\star||_2^2] = 0 \textit{ and } \mathbb{E}\left[||\overline{x}(k) - x^\star||_2^2\right] \leq \frac{C(\gamma, \mathcal{G})}{k}$$



Estimation in power systems

Power systems estimation



IEEE 14-bus test system

- x is the state of the network (frequencies at the nodes)
- linearized model of measurement

$$z = Hx + \nu$$

  - $\mathbb{E}[\nu\nu^\top] = R$, H full rank (observable)
  - measurements can be power flows, power injections, PMU data, ...

Least-squares problem: $x^\star = \arg\min_x (z - Hx)^\top R^{-1}(z - Hx)$

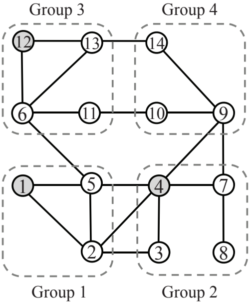
Closed-form solution: $x^\star = L^{-1}u$, where $L = H^\top R^{-1} H$ and $u = H^\top R^{-1} z$

Issues: possibly very large state; computing capacities not everywhere

Grouping

We divide nodes into groups endowed with computing power

- partition $\mathcal{V}$ into N disjoint groups $\mathcal{V}_i \subset \mathcal{V}$ for $i \in \widetilde{\mathcal{V}} = \{1, 2, \ldots, N\}$
- decompose vectors z, x, u and matrix H into $x = [x_1^\top, \ldots, x_N^\top]^\top$ where $x_i \in \mathbb{R}^{\mathcal{V}_i}$, etc.
- define neighborhoods, for each group $i \in \widetilde{\mathcal{V}}$
  - out-neighbors: $\widetilde{\mathcal{N}}_i = \{j \in \widetilde{\mathcal{V}} : \ H_{ji} \neq 0\}$
  - in-neighbors: $\widetilde{\mathcal{M}}_i = \{j \in \widetilde{\mathcal{V}} : \ H_{ij} \neq 0\}$



Issue: the resulting graph can be directed ($H_{ij} \neq H_{ji}$)

we need to design a new ergodic randomized algorithm...

A more involved randomized algorithm

Initialization:

$(x_i, \kappa_i, \overline{x}_i) = (0, 0, 0)$

$u_i = (H^\top R^{-1} z)_i = \sum_{\ell \in \widetilde{\mathcal{N}}_i} H_{\ell i}^\top R_\ell^{-1} z_\ell$ measurements z_ℓ from out-neighbors

At time $k \in \mathbb{Z}_{\geq 0}$:

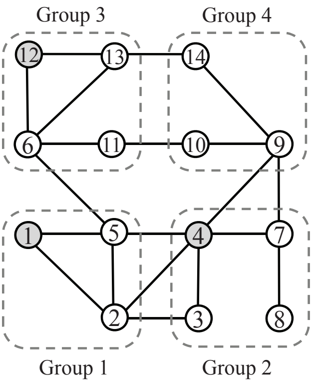
- one group $j \in \widetilde{\mathcal{V}}$ is randomly chosen to initiate updates
- group j sends its current estimate $x_j(k)$ to its out-neighbors $\ell \in \widetilde{\mathcal{N}}_j$
- groups ℓ compute $y_\ell^{(i,j)} := H_{\ell i}^\top R_\ell^{-1} H_{\ell j} x_j(k)$
- in-neighbors $i \in \widetilde{\mathcal{M}}_\ell$ update by

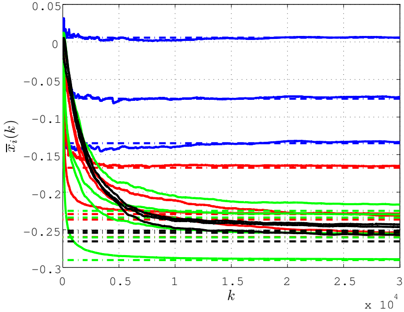
$$x_i(k+1) = x_i(k) - \tau \sum_{m \in \widetilde{\mathcal{N}}_i \cap \widetilde{\mathcal{N}}_j} y_m^{(i,j)} + \tau u_i, \quad i \in \widetilde{\mathcal{M}}_\ell,\ \ell \in \widetilde{\mathcal{N}}_j$$

- the states of the other nodes remain unchanged

$$x_i(k+1) = x_i(k), \quad i \notin \widetilde{\mathcal{M}}_\ell,\ \ell \in \widetilde{\mathcal{N}}_j$$

Simulations





P. Frasca, H. Ishii, C. Ravazzi, and R. Tempo. Distributed randomized algorithms for opinion formation, centrality computation and power systems estimation. *European Journal of Control*, 2015. submitted

Take-home summary & final comments

Estimation from relative measurements

- is ubiquitous: sensor networks, clock networks, statistical ranking
- graph theory is useful to formulate the problem
- estimation error depends on the measurement graph
- stopping times of (gradient) algorithms do not depend on size
- can be solved by distributed, asynchronous, randomized algorithms, which also apply to similar least-squares problems, like in power systems

Note: Dynamics with randomization + averaging also relevant in PageRank computation and in social networks

C. Ravazzi, P. Frasca, R. Tempo, and H. Ishii. Ergodic randomized algorithms and dynamics over networks, 2015. To appear in *IEEE Transactions on Control of Network Systems*. URL: http://arxiv.org/abs/1309.1349