Distributed estimation from relative and absolute measurements

W.S. Rossi  P. Frasca  F. Fagnani

UNIVERSITY OF TWENTE.

Osaka University, Japan
December 8, 2015
Outline of this talk

1 The estimation problem
   - Relative measurements
   - Absolute measurements
   - Applications

2 Mean square errors of the optimal estimators
   - Effective resistance, size, and dimension

3 Gradient algorithm
   - Scaling on large graphs
The (scalar) localization problem

Each node \( i \in V = \{1, \ldots, N\} \) has position \( \bar{x}_i \in \mathbb{R} \) (unknown)

**Goal:** Each node seeks an estimate \( \hat{x}_i \) of its own position

**Available measurements:**
The (scalar) localization problem

Each node $i \in V = \{1, \ldots, N\}$ has position $\bar{x}_i \in \mathbb{R}$ (unknown)

**Goal:** Each node seeks an estimate $\hat{x}_i$ of its own position

**Available measurements:**

- Some pairs of nodes $\{i, j\}$ measure their relative positions

  $$b_{\{i,j\}} = \bar{x}_i - \bar{x}_j + n_{\{i,j\}}$$

  $n_{\{i,j\}}$ i.i.d. $\mathcal{N}(0, \sigma^2_r)$
The (scalar) localization problem

Each node $i \in V = \{1, \ldots, N\}$ has position $\bar{x}_i \in \mathbb{R}$ (unknown)

**Goal:** Each node seeks an estimate $\hat{x}_i$ of its own position

**Available measurements:**

- Some pairs of nodes $\{i, j\}$ measure their relative positions

  \[ b_{\{i,j\}} = \bar{x}_i - \bar{x}_j + n_{\{i,j\}} \]

  \[ n_{\{i,j\}} \text{ i.i.d. } N(0, \sigma_r^2) \]
The (scalar) localization problem

Each node $i \in V = \{1, \ldots, N\}$ has position $\bar{x}_i \in \mathbb{R}$ (unknown)

**Goal:** Each node seeks an estimate $\hat{x}_i$ of its own position

**Available measurements:**

- Some pairs of nodes $\{i, j\}$ measure their relative positions
  
  \[ b_{\{i,j\}} = \bar{x}_i - \bar{x}_j + n_{\{i,j\}} \]

  $n_{\{i,j\}}$ i.i.d. $N(0, \sigma_r^2)$

- Every node $i$ measures $\bar{x}_i$ (absolute meas.)
  
  \[ x_{0i} = \bar{x}_i + n_i \]

  $n_i$ i.i.d. $N(0, \sigma_a^2)$
Graph representation of the relative measurements

- $G = (V, E)$ connected undirected graph
- Edges in $E$ are pairs $e = \{i, j\}$ that share a measurement
Graph representation of the relative measurements

- **$G = (V, E)$** connected undirected graph
- Edges in $E$ are pairs $e = \{i, j\}$ that share a measurement
- **Incidence** matrix $A \in \mathbb{R}^{E \times V}$ constructed by choosing an arbitrary orientation, without loss of generality:
  
  $A_{e, i} = -1$ and $A_{e, j} = +1$ if $i < j$

\[
A = \begin{bmatrix}
-1 & +1 & 0 & 0 & 0 \\
0 & -1 & +1 & 0 & 0 \\
0 & 0 & -1 & +1 & 0 \\
0 & 0 & 0 & -1 & +1 \\
-1 & 0 & 0 & 0 & +1 \\
-1 & 0 & 0 & +1 & 0 \\
-1 & 0 & +1 & 0 & 0 \\
\end{bmatrix}
\]

\[
L = A^\top A = \begin{bmatrix}
4 & -1 & -1 & -1 & -1 \\
-1 & 2 & -1 & 0 & 0 \\
-1 & -1 & 3 & -1 & 0 \\
-1 & 0 & -1 & 3 & -1 \\
-1 & 0 & 0 & -1 & 2 \\
\end{bmatrix}
\]
Graph representation of the relative measurements

- $G = (V, E)$ connected undirected graph
- Edges in $E$ are pairs $e = \{i, j\}$ that share a measurement
- **Incidence** matrix $A \in \mathbb{R}^{E \times V}$
  constructed by choosing an arbitrary orientation, without loss of generality:
  
  $$A_{e,i} = -1 \text{ and } A_{e,j} = +1 \text{ if } i < j$$

  

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

  

  $$L = A^\top A = \begin{bmatrix}
  4 & -1 & -1 & -1 & -1 \\
  -1 & 2 & -1 & 0 & 0 \\
  -1 & -1 & 3 & -1 & 0 \\
  -1 & 0 & -1 & 3 & -1 \\
  -1 & 0 & 0 & -1 & 2 \\
  \end{bmatrix}$$

- Vector of measurements $b = A\bar{x} + n$
We define two least-squares problems:

\[ \Phi^r(x) = \frac{1}{\sigma_r^2} \| Ax - b \|_2^2 \]

\[ \hat{x}^r := \arg\min_{x: \sum_i x_i = \sum_i \bar{x}_i} \Phi^r(x) \]

if absolute measurements are not available

\[ \Phi^{ra}(x) = \frac{1}{\sigma_r^2} \| Ax - b \|_2^2 + \frac{1}{\sigma_a^2} \| x - x_0 \|_2^2 \]

\[ \hat{x}^{ra} := \arg\min \Phi^{ra}(x) \]

if absolute measurements are available
Relative estimation as a least-squares problem

We define two least-squares problems:

\[ \Phi^r(x) = \frac{1}{\sigma_r^2} \|Ax - b\|_2^2 \]
\[ \hat{x}^r := \arg\min_{x: \sum_i x_i = \sum_i \bar{x}_i} \Phi^r(x) \]

if absolute measurements are not available

\[ \Phi^{ra}(x) = \frac{1}{\sigma_r^2} \|Ax - b\|_2^2 + \frac{1}{\sigma_a^2} \|x - x_0\|_2^2 \]
\[ \hat{x}^{ra} := \arg\min \Phi^{ra}(x) \]

if absolute measurements are available

Questions:

Q1 How good are the estimates \( \hat{x}^\bullet \)? That is, \( \frac{1}{N} \mathbb{E}[\|\hat{x}^\bullet - \bar{x}\|_2^2] =? \)

Q2 How can the network compute \( \hat{x}^\bullet \)?
Applications

- **Self-localization of mobile robots**
  

- **Clock synchronization**
  

- **Statistical ranking in databases: “Netflix problem”, sport tournaments**
  

- **Power systems estimation**
  
Errors of the optimal estimators
Optimal estimators

The least-squares solutions can be explicitly computed

\[ \hat{x}^r = (A^\top A)^\dagger A^\top b \]

\[ \hat{x}^{ra} = \left( \frac{1}{\sigma_r^2} A^\top A + \frac{1}{\sigma_a^2} I \right)^{-1} \left( \frac{1}{\sigma_r^2} A^\top b + \frac{1}{\sigma_a^2} x_0 \right) \]

The estimation errors are

\[ H^r := \frac{1}{N} \mathbb{E}[\|\hat{x}^r - \bar{x}\|^2_2] = \frac{\sigma_r^2}{N} \text{tr}(A^\top A)^\dagger \]

\[ = \frac{\sigma_r^2}{N} \sum_{h=2}^{N} \frac{1}{\lambda_h} \]

\[ H^{ra} := \frac{1}{N} \mathbb{E}[\|\hat{x}^{ra} - \bar{x}\|^2_2] = \frac{\sigma_r^2}{N} \text{tr}(A^\top A + \gamma I)^{-1} \]

\[ = \frac{\sigma_r^2}{N} \sum_{h=1}^{N} \frac{1}{\gamma + \lambda_h} \]

where \( \gamma = \frac{\sigma_r^2}{\sigma_a^2} \) and \( 0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_N \) are the eigenvalues of \( L = A^\top A \)

The error is determined by the topology of the measurement graph!
We start from the relative measurement graph.
Estimator error and effective resistance

- We start from the relative measurement graph
- We add a virtual node 0 to include the absolute measurements
Estimator error and effective resistance

- We start from the relative measurement graph
- We add a virtual node 0 to include the absolute measurements
- We assign the variances as resistances to the edges and construct a circuit
Estimator error and effective resistance

- We start from the relative measurement graph
- We add a virtual node 0 to include the absolute measurements
- We assign the variances as resistances to the edges and construct a circuit
- We compute effective resistances $R_{\ell m}^{\text{eff}}$ for every pair of nodes $\{\ell, m\}$
Estimator error and effective resistance

- We start from the relative measurement graph
- We add a virtual node 0 to include the absolute measurements
- We assign the variances as resistances to the edges and construct a circuit
- We compute effective resistances $R_{\ell m}^{\text{eff}}$ for every pair of nodes $\{\ell, m\}$

\[
H^r = \frac{1}{N^2} \sum_{i \neq 0, j > i} R_{ij}^{\text{eff}}
\]

\[
H^{\text{ra}} = \frac{1}{N} \sum_{i=1}^{N} R_{0i}^{\text{eff}}
\]
$H^r$ and graph topology

Based on the formulas: $H^r = \frac{\sigma_r^2}{N} \sum_{h=2}^{N} \frac{1}{\lambda_h} = \frac{1}{N^2} \sum_{i \neq 0, j > i} R_{ij}^{\text{eff}}$

Results:

- as $N \to \infty$, the error $H^r$ can diverge, stay bounded, or go to zero
- scaling in $N$ depends on the graph dimension $d$
  - in dimension $d = 1$: $H^r = \Theta(N)$
  - in dimension $d = 2$: $H^r = \Theta(\log N)$
  - in dimension $d \geq 3$: $H^r = \Theta(1/d)$
- on $d$-dimensional hypercube: $H^r \sim \frac{\sigma_r^2}{\log_2 N}$
- on complete graphs: $H^r \sim \frac{\sigma_r^2}{N}$
$H^{ra}$ and graph topology

Based on the formulas:

$$H^{ra} = \frac{\sigma_a^2}{N} \sum_{h=1}^{N} \frac{\gamma}{\gamma + \lambda_h} = \frac{1}{N} \sum_{i=1}^{N} R_{0i}^{\text{eff}}$$

Results:

- $\frac{\sigma_r^2}{2d_{\text{max}} + \gamma} \leq H^{ra} \leq \sigma_a^2$ where $d_{\text{max}}$ is the largest degree of the graph $G$

- On cycle graphs: $\lim_{N \to \infty} H^{ra} = \sigma_a^2 \sqrt{\frac{\gamma}{\gamma + 4}}$ (decreasing in $N$)

- On hypercube graphs: $H^{ra} \sim \sigma_a^2 \frac{\gamma}{\gamma + \log_2 N}$ as $N \to \infty$

- On complete graphs: $H^{ra} = \sigma_a^2 \frac{1 + \gamma}{N + \gamma}$
Gradient algorithm (with absolute measurements)
Gradient descent algorithm

We define, choosing a parameter $\tau > 0$, the gradient descent algorithm

$$x[0] = x_0$$

$$x[t + 1] = x[t] - \frac{\tau}{2} \nabla \Phi(x[t])$$

$$= \left( I - \frac{\tau}{\sigma_r^2} A^\top A - \frac{\tau}{\sigma_a^2} I \right) x[t] + \frac{\tau}{\sigma_r^2} A^\top b + \frac{\tau}{\sigma_a^2} x_0$$
Gradient descent algorithm

We define, choosing a parameter $\tau > 0$, the gradient descent algorithm

$$x[0] = x_0$$

$$x[t + 1] = x[t] - \frac{\tau}{2} \nabla \Phi(x[t])$$

$$= \left( I - \frac{\tau}{\sigma_r^2} A^\top A - \frac{\tau}{\sigma_a^2} I \right) x[t] + \frac{\tau}{\sigma_r^2} A^\top b + \frac{\tau}{\sigma_a^2} x_0$$

Properties:

1. convergence: if $\tau < \frac{\sigma_r^2}{d_{\text{max}} + \gamma}$, then
   - the matrix $Q = I - \frac{\tau}{\sigma_r^2} A^\top A - \frac{\tau}{\sigma_a^2} I$ is sub-stochastic (thus, Schur stable)
   - $\lim_{t \to +\infty} x[t] = \hat{x}^{\text{ra}}$

2. the algorithm is distributed: $Q_{ij} > 0 \iff \{i, j\} \in E$
   - each node only needs to communicate with its neighbors in $G$

3. the mean square error $H_t := \frac{1}{N} \mathbb{E}[\|x[t] - \hat{x}^{\text{ra}}\|^2_2]$ is strictly decreasing in $t$
(1 + \(\epsilon\))-approximation time

Cycle graph, \(N = 160\), \(\gamma = .003\), \(\tau = .25\)
(1 + \epsilon)-approximation time

Cycle graph, \( N = 160, \gamma = .003, \tau = .25, \epsilon = .01 \)
(1 + \(\epsilon\))-approximation time

\[
H_t \leq \frac{1}{N} \|x[t] - \bar{x}\|_2^2
\]

\[
t^*_\epsilon := \inf \{ t : H_t < (1 + \epsilon) H_\infty \}
\]

**Theorem**

\[
t^*_\epsilon \leq \frac{\sigma_a^2}{2 \tau} \log \left( \frac{2 \sigma_a^2}{\tau \epsilon} \right)
\]

Bound independent of topology or number of nodes

Cycle graph, \(N = 160\), \(\gamma = .003\), \(\tau = .25\), \(\epsilon = .01\)
Let $\theta \in \mathbb{R}$ be a common parameter, measured by the nodes as $y_i = \theta + n_i$

Optimal estimator is $\hat{\theta} = \frac{1}{N} \sum_{i=1}^{N} y_i$

Nodes can run average consensus dynamics such that $x[t] \to \hat{\theta} \mathbf{1}$
Let $\theta \in \mathbb{R}$ be a common parameter, measured by the nodes as $y_i = \theta + n_i$

Optimal estimator is $\hat{\theta} = \frac{1}{N} \sum_{i=1}^{N} y_i$

Nodes can run average consensus dynamics such that $x[t] \to \hat{\theta} 1$

On some graphs, the error is $\frac{1}{N} \mathbb{E}[\|x[t] - \theta 1\|_2^2] = \max\{\frac{1}{N}, \frac{1}{\sqrt{t}}\}$

Then, stopping time must grow with $N$!

Conclusion
Take-home summary & references

Estimation from relative and absolute measurements

- is ubiquitous: vehicle networks, clock networks, ranking
- graph theory is useful to formulate the problem
- the estimation error depends on the measurement graph
- estimate can be computed by distributed algorithms
- stopping time of (gradient) algorithm does not depend on size or graph

Preliminary results:


Full story:


(including heterogeneous $\sigma_{ij}^r$ and $\sigma_{i}^{ra}$ and details on effective resistance analysis)
Open problems

1. Design algorithms for time-varying graphs

2. Design asynchronous *gossip* algorithms
   as recently done without absolute measurements:

3. Extend analysis and algorithms to directed graphs

4. Estimation + classification: include measurements with unknown variance (possibly, outliers)