# Distributed Bayesian Estimation of Continuous Variables Over Time-Varying Directed Networks

**Presented by: Parth Paritosh**

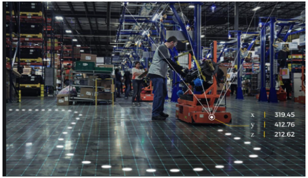
**In collaboration with: Nikolay Atanasov, Sonia Martínez**

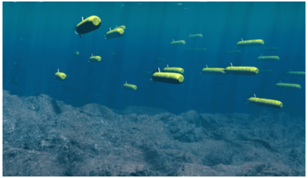
University of California San Diego (UCSD)
Mechanical and Aerospace Engineering

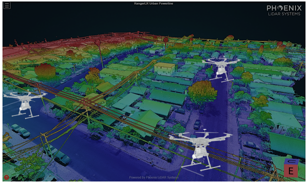
Dec 9 2022

# Motivation: distributed estimation for autonomy

Estimation tasks with naturally distributed structure:



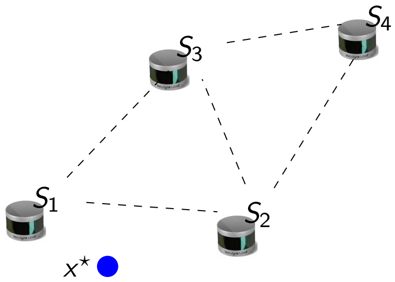



Key capabilities for distributed inference:

- Rely on localized signals
- Fast computation and low storage at nodes
- Communication efficiency
- Large scale networks with temporal variations

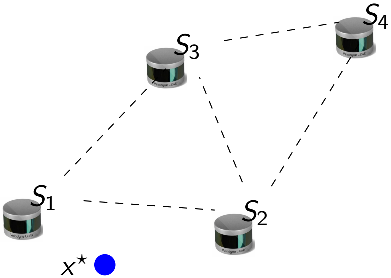
# Example problem: Estimation in sensor network

- Sensing agents $\mathcal{N} = \{1, \cdots, n\}$ with neighbor set $\mathcal{N}_i$
- Local communication network, (Weighted adjacency matrix: $A$)
- Unknown variable $\boldsymbol{x} \in \mathbb{R}^m$
- Agent measurements models $\mathrm{q}_i(\boldsymbol{z}_i|\boldsymbol{x})$
- $\boldsymbol{z}_i$: Measurements sampled from $\mathrm{q}_i(\boldsymbol{z}_i|\boldsymbol{x}^\star)$



# Example problem: Estimation in sensor network

- Sensing agents $\mathcal{N} = \{1, \cdots, n\}$ with neighbor set $\mathcal{N}_i$
- Local communication network, (Weighted adjacency matrix: $A$)
- Unknown variable $\boldsymbol{x} \in \mathbb{R}^m$
- Agent measurements models $\mathrm{q}_i(\boldsymbol{z}_i|\boldsymbol{x})$
- $\boldsymbol{z}_i$: Measurements sampled from $\mathrm{q}_i(\boldsymbol{z}_i|\boldsymbol{x}^\star)$




- How to find the true value $\boldsymbol{x}^\star$ of the unknown variable using measurements received over the given communication network?

# Estimation problem

- Samples: $\{\boldsymbol{z}_i\}_{i\in\mathcal{N}}$
- Likelihood: $\mathrm{q}_i(\boldsymbol{z}_i|\boldsymbol{x})$
- $\mathrm{q}^\star(\boldsymbol{z}_{1:n,t}) = \prod_{i\in\mathcal{N}} \mathrm{q}_i(\boldsymbol{z}_{i,t}|\boldsymbol{x}^*)$: Sampled data generating density
- $\mathrm{q}(\boldsymbol{z}_{1:n,t}|\boldsymbol{x}) = \prod_{i\in\mathcal{N}} \mathrm{q}(\boldsymbol{z}_{i,t}|\boldsymbol{x})$: Known observation likelihood for agent $i$

**Find an online estimator**

$\boldsymbol{x}_t = f(\boldsymbol{z}_{1:n,1}, \cdots, \boldsymbol{z}_{1:n,t})$ such that $\boldsymbol{x}_t \to \boldsymbol{x}^\star$

# Estimation problem

- $\mathrm{q}^{\star}(\boldsymbol{z}_{1:n,t}) = \prod_{i\in\mathcal{N}} \mathrm{q}_i(\boldsymbol{z}_{i,t}|\boldsymbol{x}^{*})$: Sampled data generating density
- $\mathrm{q}(\boldsymbol{z}_{1:n,t}|\boldsymbol{x}) = \prod_{i\in\mathcal{N}} \mathrm{q}(\boldsymbol{z}_{i,t}|\boldsymbol{x})$: Known observation likelihood for agent $i$
- Estimation error: $H_i(\boldsymbol{x}^{\star}, \boldsymbol{x}) = \mathrm{D}_{\mathrm{KL}}(q_i(\cdot|\boldsymbol{x}^{\star})||q_i(\cdot|\boldsymbol{x})) \equiv \int q_i(\cdot|\boldsymbol{x}^{\star}) \log \frac{q_i(\cdot|\boldsymbol{x}^{\star})}{q_i(\cdot|\boldsymbol{x})}$

## Optimal parameters

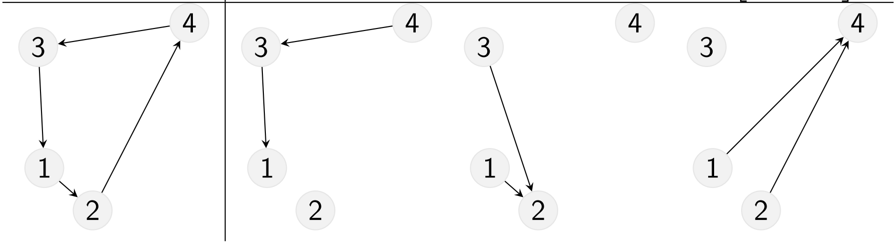
Agent-specific optimal set: $\mathcal{X}_i^{\star} = \arg\min_{\boldsymbol{x}} H_i(\boldsymbol{x}^{\star}, \boldsymbol{x})$

Network optimal set: $\mathcal{X}^{\star} \equiv \cap_{i\in\mathcal{N}} \mathcal{X}_i^{\star}$

# Distributed estimation: Static and time-varying networks

Network $\mathcal{G}$ with nodes, edges $\{\mathcal{N}, \mathcal{E}_t\}$:

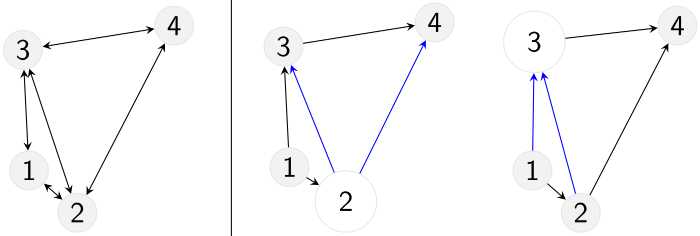
| Static | Time-varying | | |
|---|---|---|---|
| $\mathcal{E}$ | $\mathcal{E}_t$ | $\cup_{k=t}^{t+B}\mathcal{E}_k$ | $\mathbb{E}\left[\cup_{k=t}^{t+B}\mathcal{E}_k\right]$ |



# Distributed estimation: Directed networks

Network $\mathcal{G}$ with matrix model $\{A_t\}$ :

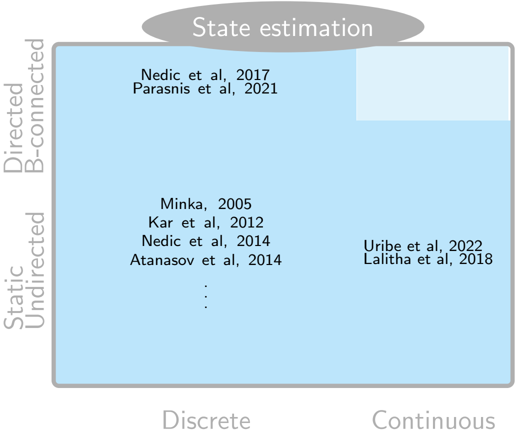
| Undirected | Directed | |
|---|---|---|
| Doubly Stochastic[1] | Column stochastic | Row stochastic[2] |
| $A\mathbf{1} = A^\top\mathbf{1} = \mathbf{1}$ | $A^\top\mathbf{1} = \mathbf{1}$ | $A\mathbf{1} = \mathbf{1}$ |




[1] B. Gharesifard and J. Cortes. When does a digraph admit a doubly stochastic adjacency matrix? In Proceedings of American Control Conference, pages 2440–2445, 2010.

[2] J. M. Hendrickx and J. N. Tsitsiklis. Fundamental limitations for anonymous distributed systems with broadcast communications. In Annu. Allert. Conf. Commun. Control Comput., pages 9–16, 2015.

# Context: Existing work



# Posing estimation problem as optimization

- $\bar{p}(\boldsymbol{x})$: Unknown probability density function with $\int \bar{p} = 1$ over $\boldsymbol{x} \in \mathbb{R}^m$

## Divergence based objective

- KL-divergence objective function

$$\arg\min_{\bar{p}} \left\{ \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\bar{p}} [\mathrm{D_{KL}}(\mathrm{q}^\star(\boldsymbol{z}_{1:n,t}) || \, \mathrm{q}(\boldsymbol{z}_{1:n,t} | \boldsymbol{x}))] \right\}$$

$$\equiv \arg\min_{\bar{p}} \left\{ \mathop{\mathbb{E}}_{\boldsymbol{z}_{1:n,t}\sim \mathrm{q}(\boldsymbol{z}_{1:n,t})} \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\bar{p}} [-\log(\mathrm{q}(\boldsymbol{z}_{1:n,t} | \boldsymbol{x}))] \right\}$$

# Objective function is revealed online

## Centralized objective

- Time-averaged objective function

$$\arg\min_{\bar{p}\in\mathcal{F}} \left\{ \frac{1}{T} \sum_{t=1}^{T} F[\bar{p}; \boldsymbol{z}_{1:n,t}] \right\}$$

$$F[\bar{p}; \boldsymbol{z}_{1:n,t}] = \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\bar{p}}[-\log(\mathrm{q}(\boldsymbol{z}_{1:n,t}|\boldsymbol{x}))]$$

## Summable property of the objective function

Using independence between agent observations $\boldsymbol{z}_{i,t}$,

$$F[\bar{p}; \boldsymbol{z}_{1:n,t}] = \sum_{i=1}^{n} F_i[\bar{p}_i, \boldsymbol{z}_{i,t}] = \sum_{i=1}^{n} \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\bar{p}_i}[-\log(\mathrm{q}_i(\boldsymbol{z}_{i,t}|\boldsymbol{x}))]$$

# Mirror descent yields Bayesian updates

- $p_t$: Estimated probability density function at time $t$
- $\frac{\delta F}{\delta p}[p_t, \boldsymbol{z}_{1:n,t}]$: Gradient of centralized objective function
- The sequence $\{\alpha_t\}$ is square-summable but non-summable.

## Stochastic mirror descent

$$p_{t+1} = \arg\min_{p \in \mathrm{L}^1} \left\{ \left\langle \frac{\delta F}{\delta p}[p_t, \boldsymbol{z}_{1:n,t}], p \right\rangle + \frac{1}{\alpha_t} \mathrm{D}_{\mathrm{KL}}(p || p_t) \right\}$$

$$\implies p_{t+1}(\boldsymbol{x}) \propto \mathrm{q}(\boldsymbol{z}_{1:n,t} | \boldsymbol{x})^{\alpha_t} p_t(\boldsymbol{x})$$

# Distributed optimization requires consensus

- Consensus: Ensuring that agents achieve the same estimate
- Likelihood update: Including likelihood information at each time step

$$p_i(\boldsymbol{x}) = \arg\min_{\bar{p}_i} \left[ \mathop{\mathbb{E}}_{\boldsymbol{z}_{i,t} \sim \mathrm{q}_i^\star(\boldsymbol{z}_{i,t})} \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \bar{p}_i} [-\log(\mathrm{q}_i(\boldsymbol{z}_{i,t} | \boldsymbol{x}))] \right] \qquad \text{(Agent objective)}$$

$$\text{subject to} \quad p_i(\boldsymbol{x}) = p_j(\boldsymbol{x}), \quad \forall j \in \mathcal{N}_i \qquad \text{(Consensus constraint)}$$

- (Assumption) The communication network is connected over $B$-time steps and the graph adjacency matrix $A_t$ satisfies $A_t\mathbf{1} = \mathbf{1}$, and diagonal entries $A_{t,ii} > 0\,, \forall i \in \{1, \ldots, n\}$, where $\mathbf{1}$ is a vector of ones.

# Proposed distributed communication algorithm

## Modified Stochastic mirror descent

$$p_{i,t+1} = \arg\min_{p \in \mathcal{F}} \left\{ \left\langle \frac{\delta F}{\delta p}[p_{i,t}, \boldsymbol{z}_{i,t}], p \right\rangle + \frac{1}{\alpha} \sum_{j=1}^{n} A_{t,ij} \, \mathrm{D}_{\mathrm{KL}}(p \| p_{j,t}) \right\}$$

## Proposed algorithm

$$v_{i,t}(\boldsymbol{x}) = \frac{1}{Z_{i,t}^{v}} \prod_{j=1}^{n} p_{j,t}(\boldsymbol{x})^{A_{t,ij}}, \quad Z_{i,t}^{v} = \int_{\boldsymbol{x} \in \mathbb{R}^m} \left( \prod_{j=1}^{n} p_{j,t}(\boldsymbol{x})^{A_{t,ij}} \right) \qquad \text{(Mixing step)}$$

$$p_{i,t+1}(\boldsymbol{x}) = \mathrm{q}_i(\boldsymbol{z}_{i,t} | \boldsymbol{x})^{\alpha} v_{i,t}(\boldsymbol{x}) \Big/ \left( \int \mathrm{q}_i(\boldsymbol{z}_{i,t} | \boldsymbol{x}) v_{i,t}(\boldsymbol{x}) d\boldsymbol{x} \right) \qquad \text{(Likelihood update)}$$

# Proof elements

- Define log-probability and log-likelihood terms,
  $r_{i,t}(\boldsymbol{x}) = \log\left[\frac{p_{i,t}(\boldsymbol{x})}{p_{i,t}(\boldsymbol{x}_\star)}\right]$, $g_{i,t}(\boldsymbol{x}) = \log\left[\frac{\mathrm{q}_i(\boldsymbol{z}_{i,t}|\boldsymbol{x})}{\mathrm{q}_i(\boldsymbol{z}_{i,t}|\boldsymbol{x}_\star)}\right]$

$$\boldsymbol{r}_{t+1}(\boldsymbol{x}) = A_t \dots A_0 \boldsymbol{r}_0(\boldsymbol{x}) + \alpha \sum_{k=1}^{t} A_t \dots A_k \boldsymbol{g}_k(\boldsymbol{x}).$$

**Network assumption**

Row stochastic weights: $A_t\mathbf{1} = \mathbf{1}$, $[A_t]_{ii} > 0$,
B-connectivity: $(\mathcal{N}, \cup_{k=t}^{t+B}\mathcal{E}_k)$ is connected $\forall t > 0$.

- $B$-connectivity $\implies |[A_t \dots A_k]_{ij} - \phi_{k,j}| \leq \lambda^k$, where $\lambda \in (0,1)$ and $\phi_{k,j} > \delta > 0$

# Log-likelihoods can be unbounded

- Agent observation models: $\pi_i(\boldsymbol{z}_i|\mu_i,1) = \exp(-0.5(\boldsymbol{z}_i - \mu_i)^2)$
- Log- likelihood ratio $g_{12}(\boldsymbol{z}_i) = \log\left(\pi_1(\boldsymbol{z}_i)/\pi_2(\boldsymbol{z}_i)\right) = 2\boldsymbol{z}_i(\mu_1 - \mu_2) + (\mu_2^2 - \mu_1^2)$

**Definition: Moment generating functions (mgf)**

For a random variable $X$ with density $p_X$, mgf $\psi(b) = \mathbb{E}[\exp(bX)]$ for any $b \in \mathbb{R}$.

- Log likelihoods have a bounded mgf: $\mathbb{E}\left[\exp\left(bg_{12}(\boldsymbol{z}_i)\right)\right] < \infty$

# Assumptions

### Networks

Row stochastic weights: $A_t\mathbf{1} = \mathbf{1}$, $[A_t]_{ii} > 0$,
B-connectivity: $(\mathcal{N}, \cup_{k=t}^{t+B}\mathcal{E}_k)$ is connected $\forall t > 0$.

### Finite mgf

The mgf of log-likelihood ratios $g_{i,t}(x)$ is finite.

### Other assumptions

Positive priors Agents' prior pdfs $p_{i,0}(x^\star) > 0$ at optimal values $x^\star \in \boldsymbol{x}^\star$.

Independent observations Independence across time and agents: $z_{i,t} \sim q_i(\cdot|x^\star)$.

# Pointwise convergence rate is exponential

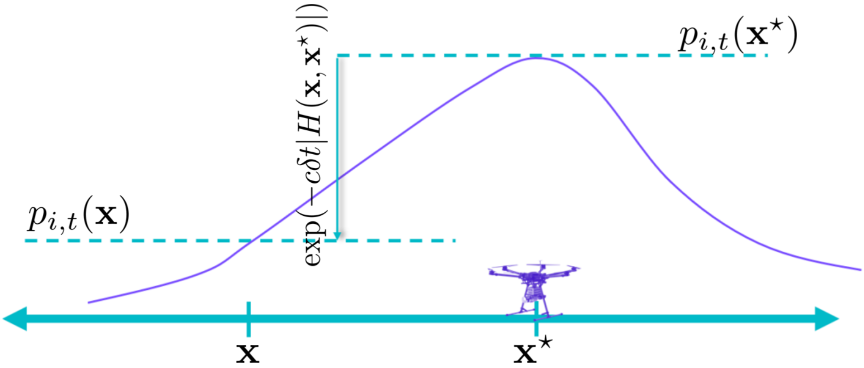
**Theorem**

Let uniform connectedness, independent observations, positive priors, and finite mgf assumptions hold. For each $\boldsymbol{x} \notin \mathcal{X}_\star$, $\boldsymbol{x}_\star \in \mathcal{X}_\star$, there is a $t_0 \in \mathbb{N}$ s.t. $\forall\, t \geq t_0$, estimated pdf $p_{i,t}$ satisfies,

$$\mathbb{P}\left( \frac{p_{i,t}(\boldsymbol{x})}{p_{i,t}(\boldsymbol{x}_\star)} < \exp(\bar{a}(\boldsymbol{x}, \boldsymbol{x}_\star)t) \right) \geq 1 - \exp(-tJ_{t_0}(\bar{a}(\boldsymbol{x}, \boldsymbol{x}_\star))).$$

The exponential rate of convergence $\bar{a}(\boldsymbol{x}, \boldsymbol{x}_\star) = -c\delta |H(\boldsymbol{x}, \boldsymbol{x}_\star)|_1 < 0$ is defined via the bound $\delta \in (0, 1)$ and KL-divergence sum $|H(\boldsymbol{x}, \boldsymbol{x}_\star)|_1 = \sum_{j \in \mathcal{N}} \mathrm{D_{KL}}(\mathrm{q}_i(\cdot|\boldsymbol{x}_\star) \| \, \mathrm{q}_i(\cdot|\boldsymbol{x}))$. Any choice of $c \in (0, 1)$ ensures $J_{t_0}(\bar{a}(\boldsymbol{x}, \boldsymbol{x}_\star))$ is positive.

# Pointwise convergence rate is exponential



# Mode of the estimated pdf is optimal

**Theorem: Mode of probability densities**

As $t \to \infty$, a mode of the pdf $p_{i,t}(\boldsymbol{x})$ estimated by agent $i$ almost surely lies in the set of optimal parameters $\boldsymbol{x}_\star$.

**Corollary: Discrete probabilities**

If the estimated probability density $p_{i,t}$ is bounded above by some $\gamma > 0$ as is the case for probability mass functions, then the probability estimated at any $\boldsymbol{x}_1 \in \boldsymbol{x} \backslash \boldsymbol{x}_\star$ satisfy, $p_{i,t}(\boldsymbol{x}_1) \to 0$ a.s.

# Cooperative localization

- A 10-node network with unknown locations $\boldsymbol{x} = [\boldsymbol{x}_i]_{i\in\mathcal{N}}, \boldsymbol{x}_i \in \mathbb{R}^2$.
- Observations: $\boldsymbol{z}_{ij} = (\boldsymbol{x}_j - \boldsymbol{x}_i) + \epsilon, \epsilon \sim \mathbf{N}(0, V_i), V_i = \mathbb{I}_2$
- Agent observation model $\mathrm{q}_i(\boldsymbol{z}_i|\boldsymbol{x}) = \phi(\boldsymbol{z}_i - H_i\boldsymbol{x}, V_i),\ H_i \in \{-1, 0, 1\}^{d_z|\boldsymbol{x}_i|\times 2|\boldsymbol{x}_i|}$
- $p_{i,t}(\boldsymbol{x}) = \phi(\boldsymbol{x}|\mu_{i,t}, \Omega_{i,t}^{-1})$: Estimated normal density representing variables in $\boldsymbol{x}$ with mean $\mu_{i,t}$ and covariance $\Omega_{i,t}^{-1}$
- Mean and covariance updates:

$$\Omega_{i,t+1} = \sum_{j\in\mathcal{N}} A_{t,ij}\Omega_{j,t-1} + \alpha H_i^\top \Omega_i^z H_i,$$

$$\mu_{i,t+1} = \Omega_{i,t+1}^{-1}\Big(\sum_{j\in\mathcal{N}} A_{t,ij}\Omega_{j,t-1}\mu_{j,t-1} + \alpha H_i^\top \Omega_i^z \boldsymbol{z}_{i,t}\Big).$$

# Cooperative localization

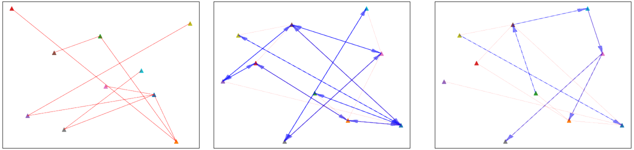


Figure: (left) Observation network and (center, right) time-varying communication network at times $t \in \{1, 2\}$.

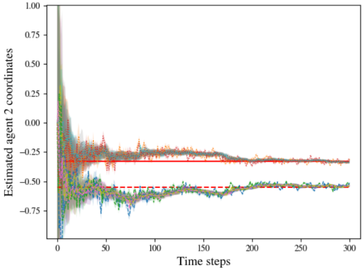



Figure: A ten-agent time varying network estimates relative positions of agent 2. The horizontal dashed and dotted lines represent true $(x_2, y_2)$ positions.

# Estimating motion model

- Target position $\boldsymbol{y}_t^d = \boldsymbol{x}_\star + r[\cos(\theta_t), \sin(\theta_t)]^\top$, $\theta_t = \theta_{t-1} + \omega \Delta t$
- Sensor $i$ at $\boldsymbol{y}_i^s$ measures $z_{i,t}(\boldsymbol{y}_i^s, \boldsymbol{y}_t^d) = |\boldsymbol{y}_i^s - \boldsymbol{y}_t^d|_2 + \eta, \eta \sim \mathbf{N}(0,1)$

- Prior $p_{i,0}(\boldsymbol{x}_\star) = \sum_{m=1}^M \alpha_{i,0}^m \delta(\boldsymbol{x}_\star | \boldsymbol{x}_{i,0}^m)$

$$p_{i,t+1|t}(\boldsymbol{x}_\star) \propto \mathrm{q}_i(\boldsymbol{z}_{i,t} | \boldsymbol{x}_\star) \sum_{m=1}^{M} \alpha_{i,t}^m \delta(\boldsymbol{x}_\star | \boldsymbol{x}_{i,t}^m)$$

$$\alpha_{i,t+1}^m = \left( \mathrm{q}_i(\boldsymbol{z}_{i,t} | \boldsymbol{x}_{i,t}^m) \alpha_{i,t}^m \middle/ \sum_{m=1}^{M} \mathrm{q}_i(\boldsymbol{z}_{i,t} | \boldsymbol{x}_{i,t}^m) \alpha_{i,t}^m \right)$$

- Distributed resampling weights: $A_{ij} \alpha_{j,t}^m$

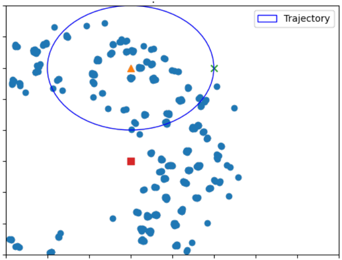



Figure: Trajectory and sensor particles.

# Estimating motion model

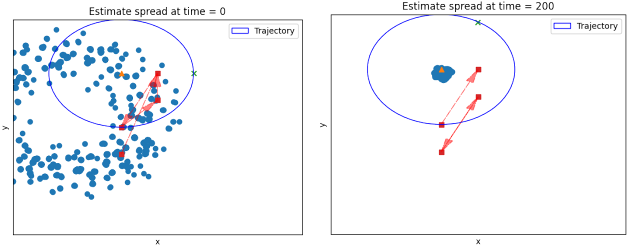



Figure: Estimating the center of a circular trajectory (orange triangle at $[1, 4]$) using a time-varying uniformly connected network of four sensors (red squares at $[1, 1], [1, 2], [2, 4], [2, 3]$). The subfigures show the cooperatively estimated particle-filter distribution of the circle center after 1 (left) and 200 (right) iterations.

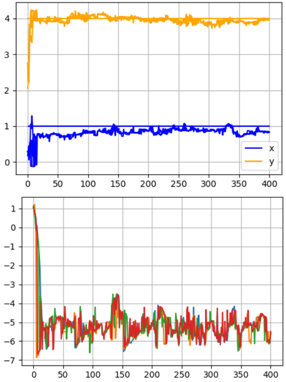



Figure: Evolution of the mean and log-maximum eigenvalue of the covariance of the particle-filter estimates.

# Contributions

- Proposed distributed estimation algorithm for uniformly connected directed graphs
- Proved probability on non-optimal domain decays exponentially, even for continuous likelihood models
- Demonstrated that the mode of estimated pdf converges to true value using Borel-Cantelli arguments
- Presented the Gaussian and a modified particle version of the algorithm

# Thank you

# Proof elements

- Define log-probability and log-likelihood terms, $r_{i,t}(\boldsymbol{x}) = \log\left[\frac{p_{i,t}(\boldsymbol{x})}{p_{i,t}(\boldsymbol{x}_\star)}\right]$, $g_{i,t}(\boldsymbol{x}) = \log\left[\frac{\mathrm{q}_i(\boldsymbol{z}_{i,t}|\boldsymbol{x})}{\mathrm{q}_i(\boldsymbol{z}_{i,t}|\boldsymbol{x}_\star)}\right]$

$$\boldsymbol{r}_{t+1}(\boldsymbol{x}) = A_t \dots A_0 \boldsymbol{r}_0(\boldsymbol{x}) + \alpha \sum_{k=1}^{t} A_t \dots A_k \boldsymbol{g}_k(\boldsymbol{x}).$$

**Network assumption**

Row stochastic weights: $A_t\mathbf{1} = \mathbf{1}$, $[A_t]_{ii} > 0$,
B-connectivity: $(\mathcal{N}, \cup_{k=t}^{t+B}\mathcal{E}_k)$ is connected $\forall t > 0$.

- Matrix product: $|[A_t \dots A_k]_{ij} - \phi_{k,j}| \leq \lambda^k$, where $\lambda \in (0,1)$ and $\phi_{k,j} > \delta > 0$

# Log-likelihoods can be unbounded

- Agent observation models: $\pi_i(\boldsymbol{z}_i|\mu_i, 1) = \exp(-0.5(\boldsymbol{z}_i - \mu_i)^2)$
- Log- likelihood ratio $g_{12}(\boldsymbol{z}_i) = \log\left(\pi_1(\boldsymbol{z}_i)/\pi_2(\boldsymbol{z}_i)\right) = 2\boldsymbol{z}_i(\mu_1 - \mu_2) + (\mu_2^2 - \mu_1^2)$

**Definition: Moment generating functions (mgf)**

For a random variable $X$ with density $p_X$, mgf $\psi(b) = \mathbb{E}[\exp(bX)]$ for any $b \in \mathbb{R}$.

- Log likelihoods have a bounded mgf: $\mathbb{E}\left[\exp\left(b g_{12}(\boldsymbol{z}_i)\right)\right] < \infty$

# Assumptions

### Networks

Row stochastic weights: $A_t\mathbf{1} = \mathbf{1}$, $[A_t]_{ii} > 0$,
B-connectivity: $(\mathcal{N}, \cup_{k=t}^{t+B}\mathcal{E}_k)$ is connected $\forall t > 0$.

### Finite mgf

The mgf of log-likelihood ratios $g_{i,t}(x)$ is finite.

### Other assumptions

Positive priors Agents' prior pdfs $p_{i,0}(x^\star) > 0$ at optimal values $x^\star \in \boldsymbol{x}^\star$.

Independent observations Independence across time and agents: $z_{i,t} \sim q_i(\cdot|x^\star)$.

# Large deviations from the mean is improbable

## Cramer's theorem

Assume that the mgf $\psi(b)$ of a random variable $X_t$ is finite for some $b > 0$ and let $\mu = \mathbb{E}[X_t]$. Then, for any $a > \mu$ and a running sum $S_t = \sum_{k=1}^{t} X_t$,

$$\mathbb{P}(S_t > at) \leq \exp(-tI(a)),$$

where $I(a) = \sup_{b>0}\{ab - \log(\psi(b))\} > 0$.

- Relating to convergence rates in Cramer's theorem:
  $e_0 = [A_t \dots A_0 \boldsymbol{r}_0]_i, e_k = \alpha[A_t \dots A_k \boldsymbol{g}_k]_i, \psi_k(b) = \mathbb{E}[\exp(be_k)]$

$$J_t(a) = \sup_{b>0} \left( D_t(a, b) \equiv ab - \frac{1}{t} \sum_{k=0}^{t} \log(\psi_k(b)) \right)$$