Scalable Bayesian Algorithms for Distributed Estimation and Inference

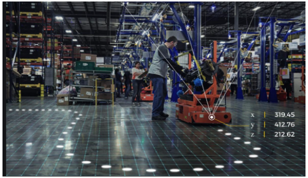
Presented by: Parth Paritosh
Chair / Co-Chair: Sonia Martínez, Nikolay Atanasov
Committee members: Jorge Cortés, Patricia Hidalgo-Gonzalez, Melvin Leok

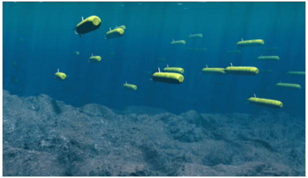
Contextual Robotics Institute
University of California San Diego (UCSD)
Supported by NSF, ARL, DARPA, and ONR

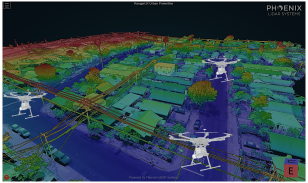
Dec 5 2023

Motivation: Distributed Estimation for Autonomy

Estimation tasks with naturally distributed structure:





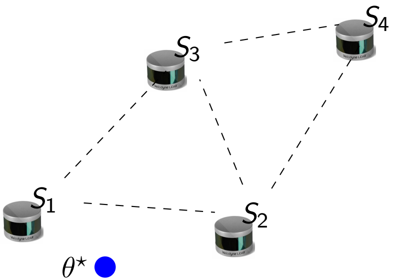


Key capabilities for inference algorithms:

- Online localized signals data
- Fast computation and low storage at nodes
- Interconnected heterogeneous systems
- Large networks with temporal variations

Example Problem: Estimation in Sensor Network

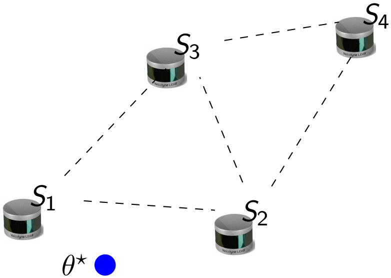
- Sensing agents $\mathcal{N} = \{1, \cdots, n\}$ with neighbor set $\mathcal{N}_i$
- Local communication network, (Weighted adjacency matrix: A)
- Unknown variable $\theta \in \mathbb{R}^m$
- Agent measurements models $\ell_i(z_i|\theta)$
- Measurements $z_i \in \mathbb{R}^d$ from $\ell_i(z_i|\theta^\star)$



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

Example Problem: Estimation in Sensor Network

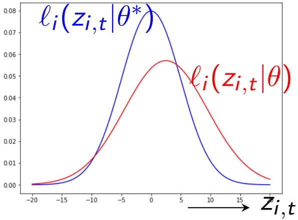
- Sensing agents $\mathcal{N} = \{1, \cdots, n\}$ with neighbor set $\mathcal{N}_i$
- Local communication network, (Weighted adjacency matrix: A)
- Unknown variable $\theta \in \mathbb{R}^m$
- Agent measurements models $\ell_i(z_i|\theta)$
- Measurements $z_i \in \mathbb{R}^d$ from $\ell_i(z_i|\theta^\star)$



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

Estimation problem

- Samples: $z_{i,t}$
- Likelihood: $\ell_i(z_{i,t}|\theta)$
- $\ell^\star(z_{1:n,t}) = \prod_{i\in\mathcal{N}} \ell_i(z_{i,t}|\theta^*)$: Data generating density
- $\ell(z_{1:n,t}|\theta) = \prod_{i\in\mathcal{N}} \ell(z_{i,t}|\theta)$: Known likelihood functions

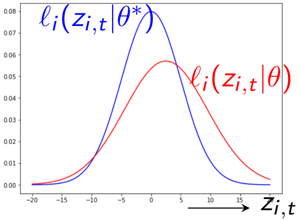


Find an estimator

$\theta_t = f(z_{1:n,1}, \cdots, z_{1:n,t})$ such that $\theta_t \to \theta^\star$

Estimation problem (continued)

- $\ell^\star(z_{1:n,t}) = \prod_{i\in\mathcal{N}} \ell_i(z_{i,t}|\theta^*)$: Data generating density
- $\ell(z_{1:n,t}|\theta) = \prod_{i\in\mathcal{N}} \ell(z_{i,t}|\theta)$: Known likelihood functions



Agent estimation error

$$H_i(\theta^\star,\theta) = \mathrm{D_{KL}}(\ell_i(\cdot|\theta^\star)||\,\ell_i(\cdot|\theta)) \equiv \int \ell_i(\cdot|\theta^\star)\log\frac{\ell_i(\cdot|\theta^\star)}{\ell_i(\cdot|\theta)}$$

Optimal parameters

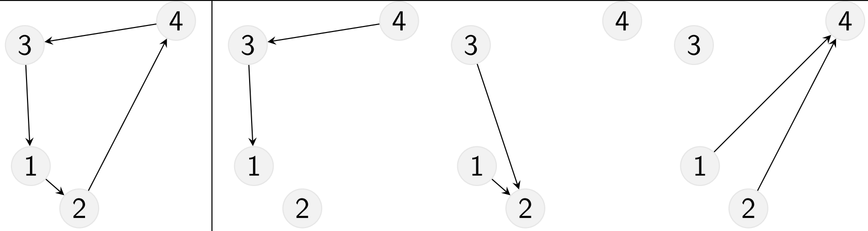
Agent-specific optimal set: $\mathcal{X}_i^\star = \arg\min_\theta H_i(\theta^\star,\theta)$

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

Distributed Estimation: Static and Time-Varying Networks

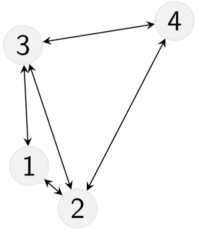
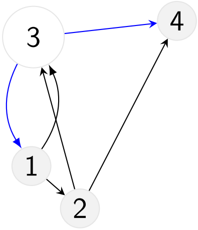
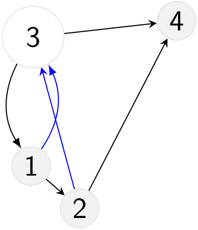
Network $\mathcal{G}$ with nodes, edges $\{\mathcal{N}, \mathcal{E}_t\}$:

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.

Literature survey: Distributed Estimation

How to combine private observations and neighbor opinions?

Bayesian

- Bayesian
  Anderson and Moore, 2005
  Bandopadhyay and Chung, 2018
- Non-Bayesian
  Jadbabaie et al., (2012, 2018)
  Nedic et al., (2015, 2017)
  Mitra et al., 2020

Network architecture

- Static Olfati-Saber et al., 2006; Moreau, 2008
- Time-varying Kia et al., 2016, Nedic et al., 2015

Distributed optimization

- In networks
  Shahrampour and Jadbabaie, 2016
  Pu et al., 2020
  Saadatniaki et al., 2020
  Uribe at al., 2022

Mixing

- General Minka, 2005; Cortes, 2008
- Arithmetic Jadbabaie et al., 2012, Parasnis et al., 2021
- Geometric Magnesius et al., 2016

Context: Existing Work and Contributions

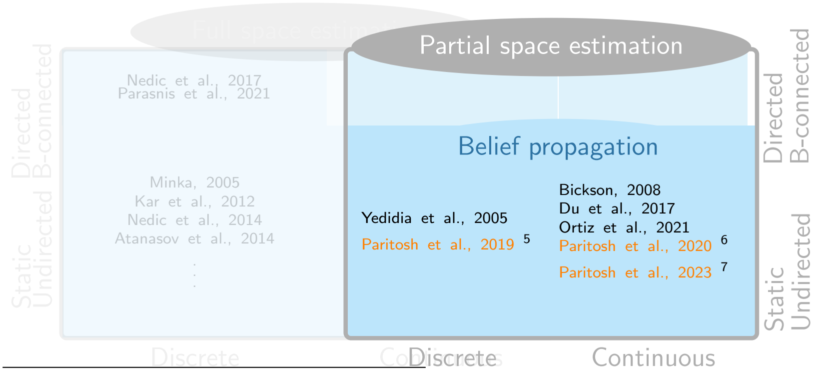
Full space estimation

	Discrete	Continuous
Directed B-connected	Nedic et al., 2017 Parasnis et al., 2021	Paritosh et al., 2022 [3]
Static Undirected	Minka, 2005 Kar et al., 2012 Nedic et al., 2014 Atanasov et al., 2014 ⋮	Lalitha et al., 2018 Uribe et al., 2022 Paritosh et al., 2023 [4]

[3] **P. Paritosh**, Atanasov N, Martinez S. Distributed Bayesian Estimation of Continuous Variables Over Time-Varying Directed Networks. IEEE Control Systems Letters. 2022 Apr 14;6:2545-50. Joint submission with IEEE CDC 2023.

[4] **P. Paritosh**, N. Atanasov, and S. Martínez, "Distributed Bayesian Estimation in Sensor Networks: Consensus on Marginal Densities", In peer review at IEEE Transactions on Network Science and Engineering.

Context: Existing work and Contributions

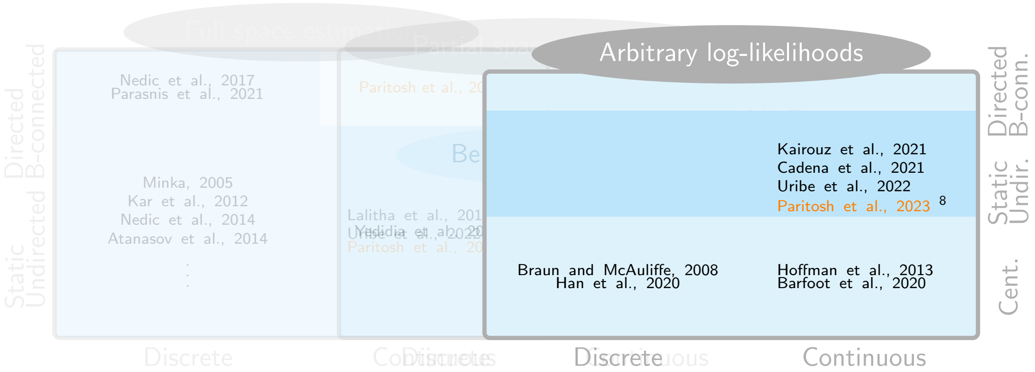


[5] **P. Paritosh**, N. Atanasov, and S. Martínez, , “Hypothesis assignment and partial likelihood averaging for cooperative estimation”, In IEEE Conference on Decision and Control, 2019, pp. 7850-7856.

[6] **P. Paritosh**, N. Atanasov, and S. Martínez, “Marginal density averaging for distributed node localization from local edge measurements”, In IEEE Conference on Decision and Control, 2020, pp. 2404-2410.

[7] **P. Paritosh**, N. Atanasov, and S. Martínez, “Distributed Bayesian Estimation in Sensor Networks: Consensus on Marginal Densities”, In peer review at IEEE Transactions on Network Science and Engineering.

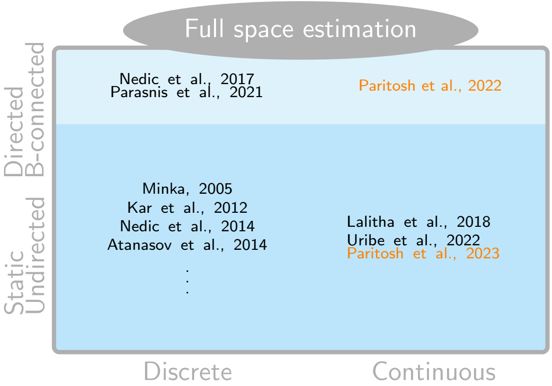
Context: Existing work and Contributions



[8] **P. Paritosh**, N. Atanasov, and S. Martínez, , "Distributed Variational Inference for Online Supervised Learning", In peer review at IEEE Transactions on Signal Processing

Outline

1 Introduction
2 Estimation in Continuous Spaces
3 Distributed Density Estimation
4 Distributed Variational Inference
5 Distributed Marginal Estimation
6 Future Directions



Research Question and Motivation

Research Question

How to estimate the unknown density $\bar{p}(\theta)$ with $\int \bar{p} = 1$ over $\theta \in \mathbb{R}^m$ based on data $z_{1:n,\leq t}$ collected by the network?

Posing Estimation Problem as Optimization

Divergence based objective

- KL-divergence objective function

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

$$\equiv \arg\min_{\bar{p}} \left\{ \mathop{\mathbb{E}}_{z_{1:n,t} \sim \ell^\star(z_{1:n,t})} \mathop{\mathbb{E}}_{\theta \sim \bar{p}} [-\log(\ell(z_{1:n,t}|\theta))] \right\}$$

Objective function is revealed online

Centralized objective

- Time-averaged objective function

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

$$F[\bar{p}; z_{1:n,t}] = \mathop{\mathbb{E}}_{\theta\sim\bar{p}}[-\log(\ell(z_{1:n,t}|\theta))]$$

Summable property of the objective function

Using independence between agent observations $z_{i,t}$,

$$F[\bar{p}; z_{1:n,t}] = \sum_{i=1}^{n} F_i[\bar{p}_i, z_{i,t}] = \sum_{i=1}^{n} \mathop{\mathbb{E}}_{\theta\sim\bar{p}_i}[-\log(\ell_i(z_{i,t}|\theta))]$$

Mirror Descent Yields Bayesian Updates

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

Stochastic mirror descent

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

$$\implies p_{t+1}(\theta) \propto \ell(z_{1:n,t}|\theta)^{\alpha_t} p_t(\theta)$$

Theorem: Convergence Guarantees

Assumptions

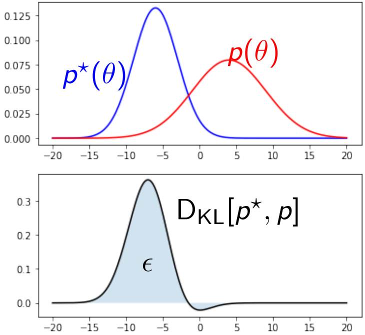
- Bounded likelihoods (lower)
- Independent Observations
- Positive priors

Functional convergence

For square summable α_t, the PDF sequence $\{p_t\}$ converges almost surely to,

$$\mathcal{B}(\mathcal{F}^\star, \epsilon) = \{p \in \mathcal{F}_d | \min_{p^\star \in \mathcal{F}^\star} \mathrm{D}_{\mathrm{KL}}[p^\star, p] \le \epsilon\},$$

an ϵ-divergence neighborhood of the minimizers $\mathcal{F}^\star$ for any $\epsilon > 0$.



Theorem: Convergence Guarantees

Assumptions

- Bounded likelihoods (lower)
- Independent Observations
- Positive priors

Functional convergence

For square summable α_t, the PDF sequence $\{p_t\}$ converges almost surely to,

$$\mathcal{B}(\mathcal{F}^\star, \epsilon) = \{p \in \mathcal{F}_d | \min_{p^\star \in \mathcal{F}^\star} \mathrm{D}_{\mathrm{KL}}[p^\star, p] \leq \epsilon\},$$

an ϵ-divergence neighborhood of the minimizers $\mathcal{F}^\star$ for any $\epsilon > 0$.

Convergence rate

For the step sizes $\alpha_t < (f[p_t] - f[p^\star])/2L^2$, the time average $\bar{p}_t = \frac{1}{t}\sum_{k=1}^{t} p_k$ satisfies,

$$f[\bar{p}_t] - f[p^\star] \leq \mathcal{O}(1/\sqrt{t}),$$

where minimizer $p^\star \in \mathcal{F}^\star$.

Outline

1 Introduction

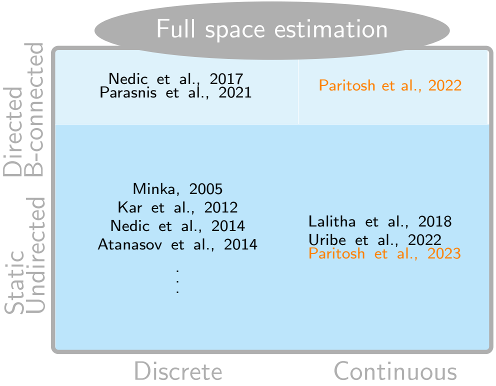
2 Estimation in Continuous Spaces
- Distributed Estimation as Optimization
- Convergence guarantees

3 Distributed Density Estimation
- Proposed algorithm and convergence guarantees
- Cooperative localization and parameter estimation

4 Distributed Variational Inference
- Distributed ELBO
- Distributed Gaussian Variational Inference
- Distributed Mapping: Simulation and Implementation

5 Distributed Marginal Estimation
- Research Question
- Decentralized Communication and Storage
- Gaussian Marginal Consensus Algorithm
- Simulation

6 Future Directions



Research Question and Motivation

Research Question

How to estimate the unknown probability density $p_i(\theta)$ over $\theta \in \mathbb{R}^m$ at each agent $i \in \mathcal{N}$ based on private data $z_{i,\leq t}$ and neighbor inferences?

Distributed Optimization Requires Consensus

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

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

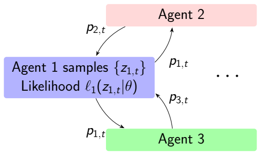
$$p_i(\theta) = p_j(\theta), \quad \forall j \in \mathcal{N}_i \qquad \text{(Consensus constraint)}$$

- Static connectivity: The graph adjacency matrix A satisfies $A\mathbf{1} = A^\top \mathbf{1} = \mathbf{1}$, and diagonal entries $A_{t,ii} > 0\,, \forall i \in \{1, \ldots, n\}$.

Proposed Distributed Estimation Algorithm

Modified Stochastic mirror descent

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



Proposed algorithm

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

$$p_{i,t+1}(\theta) = \ell_i(z_{i,t}|\theta)^{\alpha_t} v_{i,t}(\theta) \Big/ \left(\int \ell_i(z_{i,t}|\theta)^{\alpha_t} v_{i,t}(\theta) d\theta \right) \qquad \text{(Likelihood update)}$$

Theorem: Weak Convergence Guarantees

Assumptions

- Static connectivity
- Positive priors
- Independent observations
- Bounded likelihoods

Functional convergence

Then the estimated PDF sequence $\{p_{i,t}\}$ converges almost surely to ϵ-divergence neighborhood $\mathcal{B}(\mathcal{F}^\star, \epsilon)$ around optimal PDF set $\mathcal{F}^\star$ for any $\epsilon > 0$.

- Divergence neighborhood: $\mathcal{B}(\mathcal{F}^\star, \epsilon) = \{p \in \mathcal{F}_d \,|\, \min_{p^\star \in \mathcal{F}^\star} \mathrm{D}_{\mathrm{KL}}[p^\star, p] \leq \epsilon\}$.

Theorem: Weak Convergence Guarantees

Assumptions

- Static connectivity
- Positive priors
- Independent observations
- Bounded likelihoods

Functional convergence

Then the estimated PDF sequence $\{p_{i,t}\}$ converges almost surely to ϵ-divergence neighborhood $\mathcal{B}(\mathcal{F}^\star, \epsilon)$ around optimal PDF set $\mathcal{F}^\star$ for any $\epsilon > 0$.

- Divergence neighborhood: $\mathcal{B}(\mathcal{F}^\star, \epsilon) = \{p \in \mathcal{F}_d \,|\, \min_{p^\star \in \mathcal{F}^\star} \mathrm{D}_{\mathrm{KL}}[p^\star, p] \leq \epsilon\}$.

Relaxing Connectivity and Likelihood Bounds

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 moment generating functions (MGF)

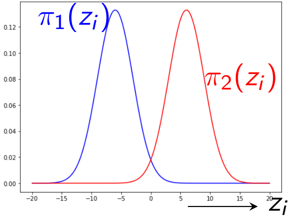
The MGF of log-likelihood ratios $g_{i,t}(x)$ is finite.

Step sizes

Fixed step sizes $\alpha_t = \alpha > 0$.

Log-Likelihoods can be Unbounded

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



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: $\mathbb{E}\left[\exp\left(bg_{12}(z_i)\right)\right] < \infty$

Proposed Distributed Estimation Algorithm

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

Proposed algorithm

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

$$p_{i,t+1}(\theta) = \ell_i(z_{i,t}|\theta)^{\alpha} v_{i,t}(\theta) \Big/ \left(\int \ell_i(z_{i,t}|\theta) v_{i,t}(\theta) d\theta \right) \qquad \text{(Likelihood update)}$$

Theorem: Pointwise Convergence Rate is Exponential

Assumptions

- Uniform connectivity
- Positive priors
- Independent observations
- Finite moment generating functions on likelihood

Claims

Then, for each $\theta \notin \mathcal{X}_\star$, $\theta_\star \in \mathcal{X}_\star$, there exists a time $t_0 \in \mathbb{N}$ such that $\forall\, t \geq t_0$, the estimated PDF $p_{i,t}$ satisfies,

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

- Exponential convergence rate $\bar{a}(\theta,\theta_\star) = -c\delta\|H(\theta,\theta_\star)\|_1 < 0$
- $\|H(\theta,\theta_\star)\|_1 = \sum_{j\in\mathcal{N}} \mathrm{D}_{\mathrm{KL}}(\ell_j(\cdot|\theta_\star)\|\,\ell_j(\cdot|\theta))$
- Any $c \in (0,1)$ ensures $J_{t_0}(\bar{a}(\theta,\theta_\star)) > 0$.

Theorem: Pointwise Convergence Rate is Exponential

Assumptions

- Uniform connectivity
- Positive priors
- Independent observations
- Finite moment generating functions on likelihood

Claims

Then, for each $\theta \notin \mathcal{X}_\star$, $\theta_\star \in \mathcal{X}_\star$, there exists a time $t_0 \in \mathbb{N}$ such that $\forall\, t \geq t_0$, the estimated PDF $p_{i,t}$ satisfies,

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

- Exponential convergence rate $\bar{a}(\theta,\theta_\star) = -c\delta\|H(\theta,\theta_\star)\|_1 < 0$
- $\|H(\theta,\theta_\star)\|_1 = \sum_{j\in\mathcal{N}} \mathrm{D}_{\mathrm{KL}}(\ell_i(\cdot|\theta_\star)\|\,\ell_i(\cdot|\theta))$
- Any $c \in (0,1)$ ensures $J_{t_0}(\bar{a}(\theta,\theta_\star)) > 0$.

Theorem: Pointwise Convergence Rate is Exponential

Assumptions

- Uniform connectivity
- Positive priors
- Independent observations
- Finite moment generating functions on likelihood

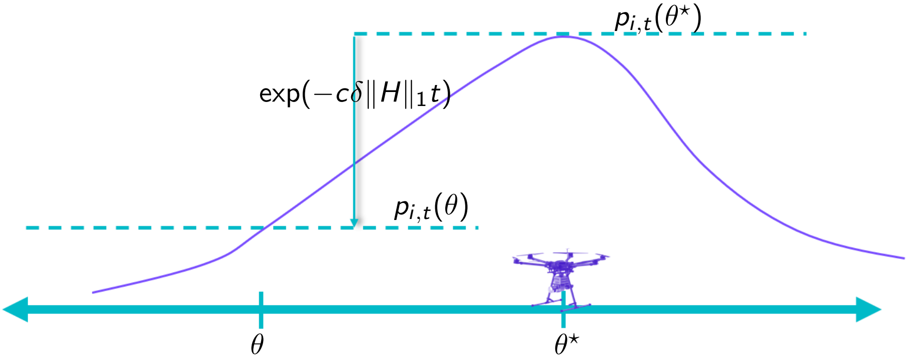
Claims

Then, for each $\theta \notin \mathcal{X}_\star$, $\theta_\star \in \mathcal{X}_\star$, there exists a time $t_0 \in \mathbb{N}$ such that $\forall\, t \geq t_0$, the estimated PDF $p_{i,t}$ satisfies,

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

- Exponential convergence rate $\bar{a}(\theta,\theta_\star) = -c\delta\|H(\theta,\theta_\star)\|_1 < 0$
- $\|H(\theta,\theta_\star)\|_1 = \sum_{j\in\mathcal{N}} \mathrm{D}_{\mathrm{KL}}(\ell_i(\cdot|\theta_\star)\|\,\ell_i(\cdot|\theta))$
- Any $c \in (0,1)$ ensures $J_{t_0}(\bar{a}(\theta,\theta_\star)) > 0$.

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 $p_{i,t}(\theta)$ estimated by agent i almost surely lies in the set of optimal parameters $\theta_\star$.

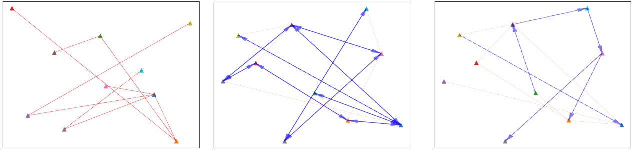
Corollary: Discrete probabilities

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 $\theta_1 \in \theta \backslash \theta_\star$ satisfy, $p_{i,t}(\theta_1) \to 0$ a.s.

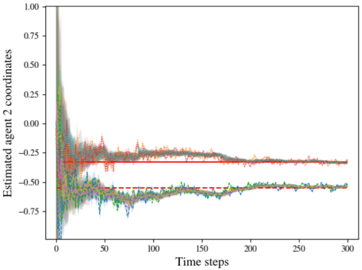
Example 1: Cooperative Localization

- A 10-node network with unknown locations $\theta = [\theta_i]_{i \in \mathcal{N}}, \theta_i \in \mathbb{R}^2$.
- Observations: $z_{ij} = (\theta_j - \theta_i) + \epsilon, \epsilon \sim \mathbf{N}(0, V_i), V_i = \mathbb{I}_2$
- Agent observation model $\ell_i(z_i|\theta) = \phi(z_i - H_i\theta, V_i)$, $H_i \in \{-1, 0, 1\}^{d_z|\theta_i| \times 2|\theta_i|}$
- $p_{i,t}(\theta) = \phi(\theta|\mu_{i,t}, \Omega_{i,t}^{-1})$: Estimated normal density representing variables in θ with mean $\mu_{i,t}$ and covariance $\Omega_{i,t}^{-1}$

Example 1: Cooperative Localization



Observation network and time-varying communication network at times $t \in \{1, 2\}$.



Estimated positions of agent 2.

Example 2: Target Tracking

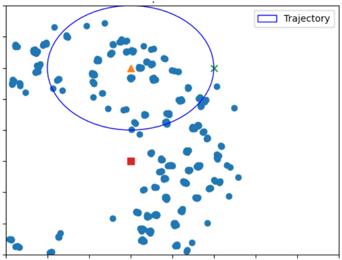
- Target position $\boldsymbol{y}_t^d = \theta_\star + r[\cos(\beta_t), \sin(\beta_t)]^\top$, $\beta_t = \beta_{t-1} + \omega \Delta t$
- Sensor i at $\boldsymbol{y}_i^s$ measures $z_{i,t}(\boldsymbol{y}_i^s, \boldsymbol{y}_t^d) = |\boldsymbol{y}_i^s - \boldsymbol{y}_t^d|_2 + \eta, \eta \sim \mathbf{N}(0,1)$

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

$$p_{i,t+1|t}(\theta_\star) \propto \ell_i(z_{i,t}|\theta_\star) \sum_{m=1}^{M} \alpha_{i,t}^m \delta(\theta_\star | \theta_{i,t}^m)$$

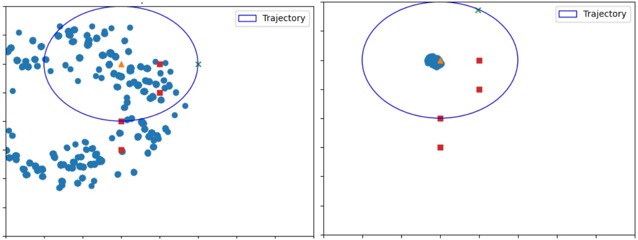
$$\alpha_{i,t+1}^m = \left(\ell_i(z_{i,t}|\theta_{i,t}^m)\alpha_{i,t}^m \Big/ \sum_{m=1}^{M} \ell_i(z_{i,t}|\theta_{i,t}^m)\alpha_{i,t}^m \right)$$

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

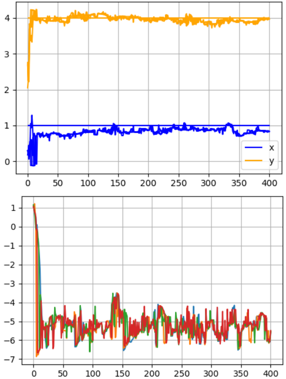


Trajectory and sensor particles.

Example 2: Target Tracking



Cooperatively estimated particle-filter distribution of the target's center after 1 and 200 iterations. Estimating the trajectory center (orange triangle) using a uniformly connected network of four sensors (red squares).



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
- Weak and pointwise convergence results for distributed estimation of continuous probability densities
- Presented the Gaussian and a modified particle version of the algorithm

Publications:

- P. Paritosh, N. Atanasov, and S. Martínez, "Distributed Bayesian Estimation of Continuous Variables Over Time-Varying Directed Networks", in IEEE Control Systems Letters, vol. 6, pp. 2545-2550, 2022. (Joint submission with IEEE CDC)
- P. Paritosh, N. Atanasov, and S. Martínez, "Distributed Bayesian Estimation in Sensor Networks: Consensus on Marginal Densities", Under review at IEEE Transactions on Network Science and Engineering.

Outline

1 Introduction

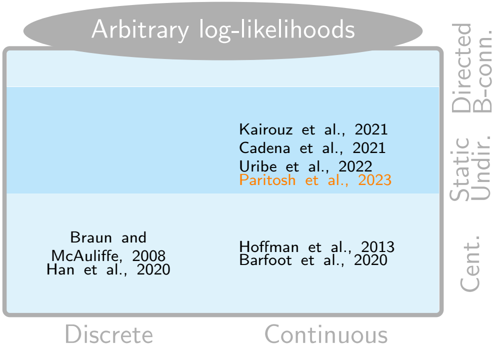
2 Estimation in Continuous Spaces
- Distributed Estimation as Optimization
- Convergence guarantees

3 Distributed Density Estimation
- Proposed algorithm and convergence guarantees
- Cooperative localization and parameter estimation

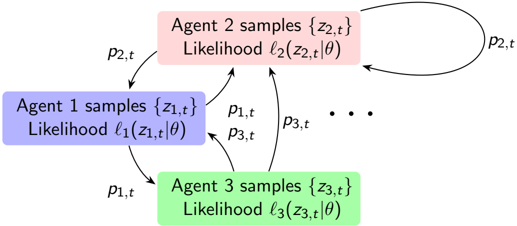
4 Distributed Variational Inference
- Distributed ELBO
- Distributed Gaussian Variational Inference
- Distributed Mapping: Simulation and Implementation

5 Distributed Marginal Estimation
- Research Question
- Decentralized Communication and Storage
- Gaussian Marginal Consensus Algorithm
- Simulation

6 Future Directions



Problem Setup: Real-Time Bayesian Inference



- Non-linear heterogeneous likelihoods
- Distributed communication
- Online probabilistic inference

Goal: Design a distributed real-time approximate inference algorithm for learning probability density function $p(\theta)$ over unknown θ.

Variational Inference

- Bayes' rule: Posterior on θ satisfies $p(\theta|z_{\leq t}) = \dfrac{\overbrace{\ell(z_t|\theta)}^{\text{Likelihood}}\,\overbrace{p(\theta|z_{<t})}^{\text{Prior}}}{\underbrace{p(z_t|z_{<t})}_{\text{Normalization factor}}}$
- Computing normalization factor is intractable (unless conditionally conjugate)
- Approximate posterior via a variational family of distributions $q(\theta) \in \mathcal{F}$
- Maximize Evidence Lower Bound (ELBO) on the normalization factor,

$$p(z_t|z_{<t}) \geq \mathop{\mathbb{E}}_{q(\theta)}[\log \ell(z_t|\theta) - \log(q(\theta)) + \log p(\theta|z_{<t})].$$

- In recursive settings, replace prior $p(\theta|z_{<t})$ with $q_{t-1}(\theta)$

Variational Inference

- Bayes' rule: Posterior on θ satisfies $p(\theta|z_{\leq t}) = \dfrac{\overbrace{\ell(z_t|\theta)}^{\text{Likelihood}}\overbrace{p(\theta|z_{<t})}^{\text{Prior}}}{\underbrace{p(z_t|z_{<t})}_{\text{Normalization factor}}}$
- Computing normalization factor is intractable (unless conditionally conjugate)
- Approximate posterior via a variational family of distributions $q(\theta) \in \mathcal{F}$
- Maximize Evidence Lower Bound (ELBO) on the normalization factor,

$$p(z_t|z_{<t}) \geq \mathop{\mathbb{E}}_{q(\theta)}[\log \ell(z_t|\theta) - \log(q(\theta)) + \log p(\theta|z_{<t})].$$

- In recursive settings, replace prior $p(\theta|z_{<t})$ with $q_{t-1}(\theta)$

Theorem: Distributed Evidence Lower Bound (DELBO)

Assuming:

- Independent Observations,
- Connected network,
- Agent PDFs $q_{i,t}(\theta) = q_t(\theta)$ for some PDF $q_t(\theta)$,

the separable distributed evidence lower bound (DELBO) on the normalization factor is,

$$p(z_t|z_{<t}) \geq \sum_{i\in\mathcal{N}} \mathbb{E}_{q_{i,t}(\theta)}\Big[\ell_i(z_{i,t}|\theta) - \frac{1}{n}\log(q_{i,t}(\theta)) + \sum_{j\in\mathcal{N}} \frac{A_{ij}}{n} \log p_j(\theta|z_{<t})\Big],$$

where A is the adjacency matrix representing connected networks.

Theorem: Distributed Evidence Lower Bound (DELBO)

Assuming:

- Independent Observations,
- Connected network,
- Agent PDFs $q_{i,t}(\theta) = q_t(\theta)$ for some PDF $q_t(\theta)$,

the separable distributed evidence lower bound (DELBO) on the normalization factor is,

$$p(z_t|z_{<t}) \geq \sum_{i \in \mathcal{N}} \mathop{\mathbb{E}}_{q_{i,t}(\theta)} [\ell_i(z_{i,t}|\theta) - \frac{1}{n} \log(q_{i,t}(\theta)) + \sum_{j \in \mathcal{N}} \frac{A_{ij}}{n} \log p_j(\theta|z_{<t})],$$

where A is the adjacency matrix representing connected networks.

Optimizing DELBO

- Replace prior $p_i(\theta|z_{<t})$ with its approximation $q_{i,t-1}(\theta)$
- Separable objective $J_t[q_{1,t},\dots,q_{n,t}] = \sum_{i\in\mathcal{N}} J_{i,t}[q_{i,t}]$,

$$J_{i,t}[q_{i,t}] = \mathop{\mathbb{E}}_{q_{i,t}(\theta)}[\log[\ell_i(z_{i,t}|\theta)\prod_{j\in\mathcal{N}} q_{j,t-1}(\theta)^{\frac{A_{ij}}{n}}] - \log q_{i,t}(\theta)^{\frac{1}{n}}].$$

- Optimal PDF for agent i is $q_{i,t}(\theta) \propto \ell_i(z_{i,t}|\theta)q_i^g(\theta) \in \arg\max_p J_{i,t}[p]$
  - Mixed PDF $q_i^g(\theta) \propto \prod_{j\in\mathcal{N}_i} q_{j,t-1}(\theta)^{\frac{A_{ij}}{n}}$ with likelihood exponent $\alpha = n$.

Optimizing DELBO

- Replace prior $p_i(\theta|z_{<t})$ with its approximation $q_{i,t-1}(\theta)$
- Separable objective $J_t[q_{1,t}, \ldots, q_{n,t}] = \sum_{i\in\mathcal{N}} J_{i,t}[q_{i,t}]$,

$$J_{i,t}[q_{i,t}] = \mathop{\mathbb{E}}_{q_{i,t}(\theta)}[\log[\ell_i(z_{i,t}|\theta)\prod_{j\in\mathcal{N}} q_{j,t-1}(\theta)^{\frac{A_{ij}}{n}}] - \log q_{i,t}(\theta)^{\frac{1}{n}}].$$

- Optimal PDF for agent i is $q_{i,t}(\theta) \propto \ell_i(z_{i,t}|\theta)q_i^g(\theta) \in \arg\max_p J_{i,t}[p]$
  - Mixed PDF $q_i^g(\theta) \propto \prod_{j\in\mathcal{N}_i} q_{j,t-1}(\theta)^{\frac{A_{ij}}{n}}$ with likelihood exponent $\alpha = n$.

Computing Variational Densities

At agent $i \in \mathcal{N}$,

$$q_{i,t}(\theta) = \ell_i(z_{i,t}|\theta)q_i^g(\theta) \Big/ \int \ell_i(z_{i,t}|\theta)q_i^g(\theta)d\theta,$$

$$q_i^g(\theta) \propto \prod_{j \in \mathcal{N}_i} q_{j,t-1}(\theta)^{\frac{A_{ij}}{n}}$$

How to handle non-conjugate likelihoods?
Approximate Gaussian variational densities with arbitrary differentiable likelihoods

Lemma: Distributed Gaussian variational inference (DGVI)

At agent i and time t, given:

- observation $z_{i,t}$ with likelihood $\ell(z_{i,t}|\theta)$,
- neighbor estimates $q_{j,t-1}(\theta) = \mathbf{N}(\theta|\mu_{j,t-1}, \Omega_{j,t-1}^{-1})$,
- Neighbor weights in communication matrix A,

the mean $\mu_{i,t}$ and information matrix $\Omega_{i,t}$ of the PDF $q_{i,t}$ minimizing DELBO is,

$$\Omega_{i,t}^{g} = \sum_{j \in \mathcal{N}} A_{ij}\Omega_{j,t-1}, \Omega_{i,t}^{g}\mu_{i,t}^{g} = \sum_{j \in \mathcal{N}} A_{ij}\Omega_{j,t-1}\mu_{j,t-1}$$

$$\Omega_{i,t} = \Omega_{i,t}^{g} - \mathbb{E}_{q_{i,t}^{g}}[\nabla_{\theta}^{2} \log \ell(z_{i,t}|\theta)],$$

$$\mu_{i,t} = \mu_{i,t}^{g} + (\Omega_{i,t}^{g})^{-1}\mathbb{E}_{q_{i,t}^{g}}[\nabla_{\theta} \log \ell(z_{i,t}|\theta)].$$

Adapting DGVI to Supervised Learning

Problem: Approximate $\mathbb{E}_{q^g_{i,t}}[\nabla_\theta \log \ell(z_{i,t}|\theta)]$ for real-time computation:

- Computating expectation by sampling $q^g_{i,t}$ is computationally prohibitive
- Define kernel based classification/regression model as agent likelihoods
- Compute expectation w.r.t. the mixed Gaussian PDF $q^g_{i,t} = \phi(\theta|\mu^g_{i,t}, (\Omega^g_{i,t})^{-1})$

Classification Model

- Observed data $z = (x, y)$ with input $x \in \mathbb{R}^d$ and label $y \in \{0, 1\}$
- Model features $\Phi_x \in \mathbb{R}^{l+1}$ with kernel elements:
  $\Phi_x = [1, k_1(x), \ldots, k_l(x)], k_s(x) = \exp(-\gamma \|x - x^{(s)}\|^2)$
- Agent likelihood model with parameters θ and sigmoid function σ:

$$\ell(z|\theta) = \sigma(\Phi_x^\top \theta)^y (1 - \sigma(\Phi_x^\top \theta))^{1-y}$$

DGVI for Kernel Classification

For agent i's observation $z = (x, y)$ with classification likelihood, and neighbor estimates $\phi(\theta|\mu_{j,t}, \Omega_{j,t}^{-1})$,

the mean $\mu_{i,t}$ and information matrix $\Omega_{i,t}$ of the PDF $q_{i,t}$ maximizing DELBO is,

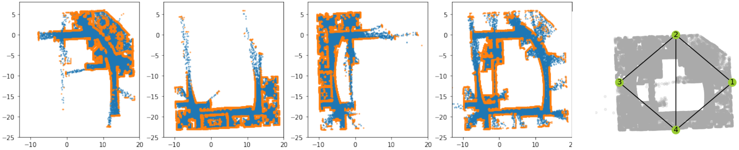
$$\Omega_{i,t}^{g} = \sum_{j\in\mathcal{N}} A_{ij}\Omega_{j,t-1},\ \Omega_{i,t}^{g}\mu_{i,t}^{g} = \sum_{j\in\mathcal{N}} A_{ij}\Omega_{j,t-1}\mu_{j,t-1},\ \Sigma_{i,t}^{g} = (\Omega_{i,t}^{g})^{-1}$$

$$\Omega_{i,t} = \Omega_{i,t}^{g} + \gamma\Phi_x\Phi_x^{\top}, \Omega_{i,t}^{-1} = \Sigma_{i,t}^{g} - \frac{\gamma}{\gamma_1}\Sigma_{i,t}^{g}\Phi_x\Phi_x^{\top}\Sigma_{i,t}^{g}$$

$$\mu_{i,t} = \mu_{i,t}^{g} + \left(y - \Gamma\left(\xi\Phi_x^{\top}\mu_{i,t}^{g}/\sqrt{\beta}\right)\right)\Omega_{i,t}^{-1}\Phi_x$$

with unit normal cdf Γ, $\beta = 1 + \xi^2\Phi_x^{\top}(\Omega_{i,t}^{g})^{-1}\Phi_x$, $\gamma_1 = 1 + \gamma\Phi_x^{\top}(\Omega_{i,t}^{g})^{-1}\Phi_x$ and $\gamma = \sqrt{\frac{\xi^2}{2\pi\beta}}\exp\left(-0.5[\frac{\xi^2}{\beta}(\mu_{i,t}^{g})^{\top}\Phi_x\Phi_x^{\top}\mu_{i,t}^{g}]\right)$.

Distributed Mapping with Intel LIDAR Dataset[9]



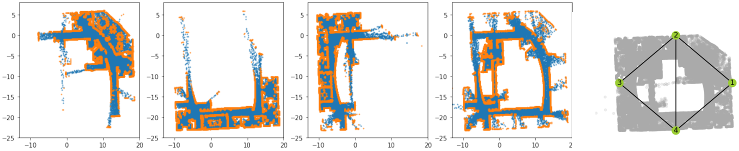
Training data distributed among 4 agents sharing their inferences, Communication network.

- Observed data $z = (x, y)$ with position $x \in \mathbb{R}^2$ and occupancy label $y \in \{0, 1\}$
- Model features $\Phi_x \in \mathbb{R}^{l+1}$ with kernels: $\Phi_x = [1, k_1(x), \ldots, k_l(x)]$
- Kernel $k_s(x) = \exp(-\gamma \|x - x^{(s)}\|^2)$ centered at $x^{(s)}$ with lengthscale γ
- Agent likelihood model with parameters θ and sigmoid function σ:

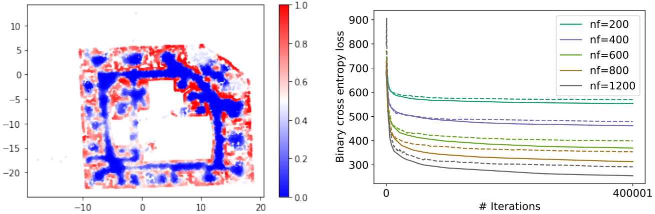
$$\ell(z|\theta) = \sigma(\Phi_x^\top \theta)^y (1 - \sigma(\Phi_x^\top \theta))^{1-y}$$

[8] A. Howard and N. Roy. The robotics data set repository (radish), 2003.

Distributed Mapping with Intel LIDAR Dataset

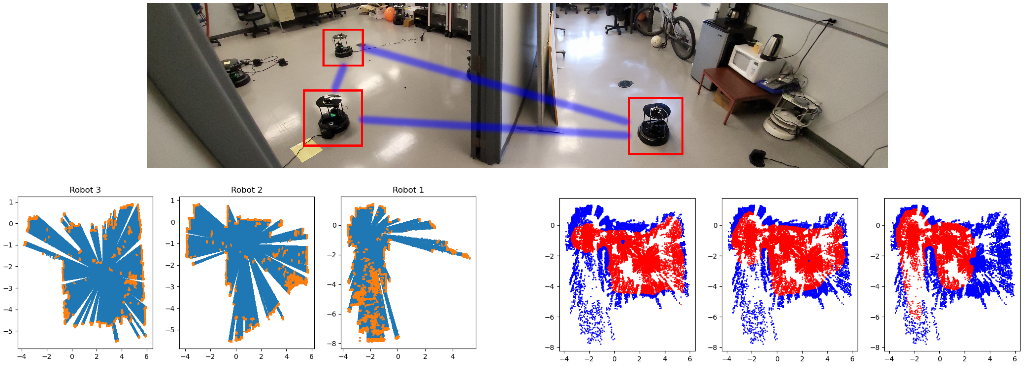


Training data distributed among 4 agents sharing their inferences, Communication network.



Free and occupied spaces in blue and orange color respectively with a 1500 features model. Comparing verification loss with diagonalized covariance model.

Implementation: Distributed Mapping with MURO Lab Turtlebots



Indoor lab space with directed communication (top), Training data collected and maps predicted by the 3 Turtlebots (bottom).

Contributions

- Compute a separable version of evidence lower bound for inference
- Distributed Gaussian updates with tractable expectation terms in supervised learning setting
- Simulation and implementation for distributed robot mapping

Publication:

- P. Paritosh, N. Atanasov and S. Martinez. Distributed Variational Inference for Online Supervised Learning. Under review at IEEE Transactions on Signal Processing.
- P. Paritosh, S. Lau, N. Atanasov and S. Martinez. Distributed Variational Inference for Online Estimation: A Distributed Mapping Implementation on Turtlebot4s. Poster at Southern California Robotics Symposium 2023.

Outline

1 Introduction

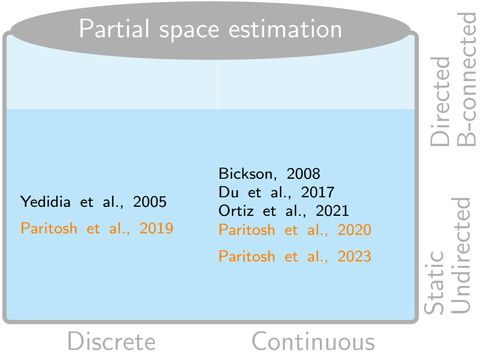
2 Estimation in Continuous Spaces
- Distributed Estimation as Optimization
- Convergence guarantees

3 Distributed Density Estimation
- Proposed algorithm and convergence guarantees
- Cooperative localization and parameter estimation

4 Distributed Variational Inference
- Distributed ELBO
- Distributed Gaussian Variational Inference
- Distributed Mapping: Simulation and Implementation

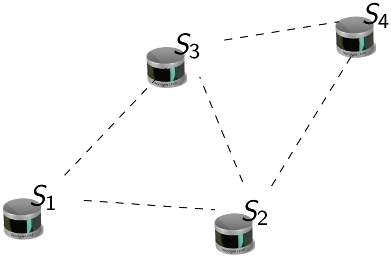
5 Distributed Marginal Estimation
- Research Question
- Decentralized Communication and Storage
- Gaussian Marginal Consensus Algorithm
- Simulation

6 Future Directions



Problem Setup

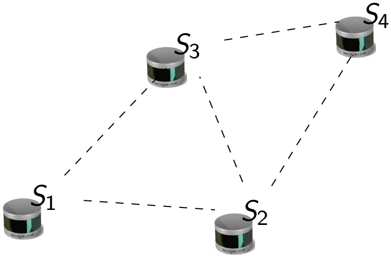
- Sensing agents $\mathcal{N} = \{1, \cdots, n\}$ with neighbor set $\mathcal{N}_i$
- Agent state: $\theta_i \in \mathbb{R}^m$
- Neighbor based measurement models $\mathrm{p}_i(z_i | \{\theta_j\}_{j \in \mathcal{N}_i}) = \prod_{j \in \mathcal{N}_i} \mathrm{p}_i(z_{ij} | \theta_i, \theta_j)$
- Local communication network, (Weighted adjacency matrix: A)



- How to find the true value of agent states $\theta_1, \cdots, \theta_n$ with relative measurements received over the given communication network?

Problem Setup

- Sensing agents $\mathcal{N} = \{1, \cdots, n\}$ with neighbor set $\mathcal{N}_i$
- Agent state: $\theta_i \in \mathbb{R}^m$
- Neighbor based measurement models $\mathrm{p}_i(z_i | \{\theta_j\}_{j\in\mathcal{N}_i}) = \prod_{j\in\mathcal{N}_i} \mathrm{p}_i(z_{ij}|\theta_i, \theta_j)$
- Local communication network, (Weighted adjacency matrix: A)



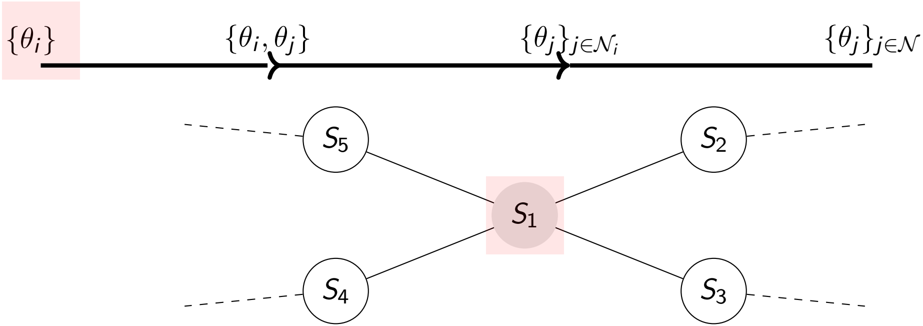
- How to find the true value of agent states $\theta_1, \cdots, \theta_n$ with relative measurements received over the given communication network?

Research Question

How to design an inference algorithm to learn true value of variables $\Theta_{\mathcal{N}_i}$ at agent i using noisy measurements and neighbor estimates at any time?

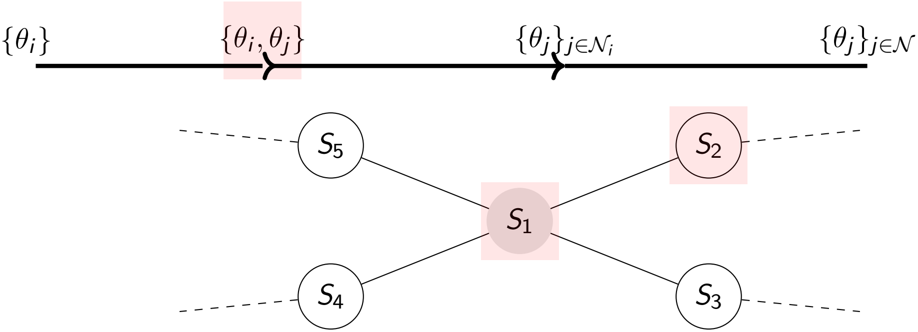
Agent Domains for Distributed Estimation

How do we select the domain $\mathcal{X}_i$ of agent i's estimate?



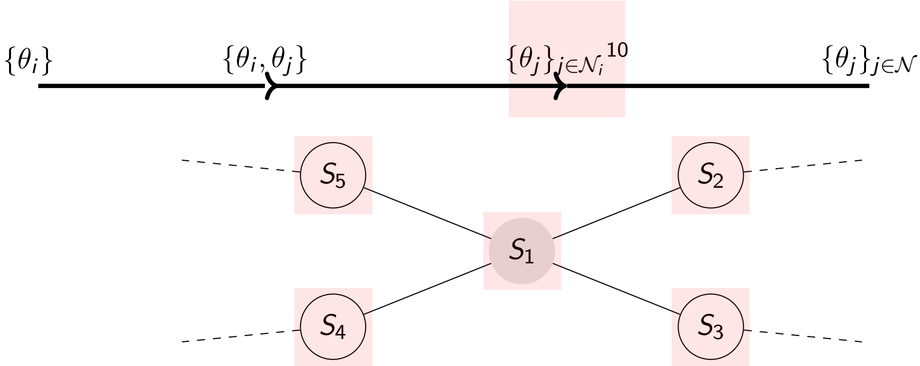
Agent Domains for Distributed Estimation

How do we select the domain $\mathcal{X}_i$ of agent i's estimate?



Agent Domains for Distributed Estimation

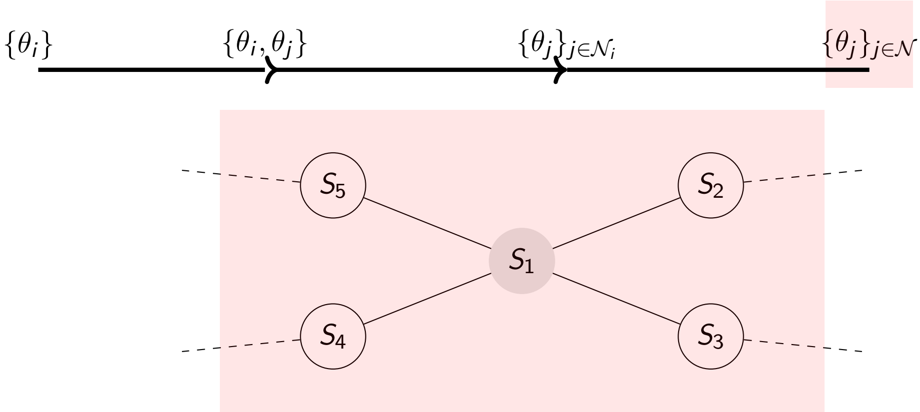
How do we select the domain $\mathcal{X}_i$ of agent i's estimate?



[10] **P. Paritosh**, N. Atanasov, and S. Martínez, , "Hypothesis assignment and partial likelihood averaging for cooperative estimation", In IEEE Conference on Decision and Control, 2019, pp. 7850-7856.

Agent Domains for Distributed Estimation

How do we select the domain $\mathcal{X}_i$ of agent i's estimate?



Existing Solutions

How do we select the domain $\mathcal{X}_i$ of agent i's estimate?

$\{\theta_i\}$ $\{\theta_i, \theta_j\}$ $\{\theta_j\}_{j\in\mathcal{N}_i}$ $\{\theta_j\}_{j\in\mathcal{N}}$

Belief propagation - - - - - - - - - - - - - - - - - ? - Geometric updates

- Yedidia et al.[10](2003): Learning marginal density at each agent state via Belief propagation in forest type graphs
- Nedich et al.[11](2017): Convergence rates of geometric averaging of inferences for the decentralized communication problem

[11] Jonathan S. Yedidia, William T. Freeman, and Yair Weiss. "Understanding belief propagation and its generalizations." Exploring artificial intelligence in the new millennium 8 (2003): 236-239.

[11] Angelia Nedić, Alex Olshevsky, and César A. Uribe. "Fast convergence rates for distributed non-bayesian learning." IEEE Transactions on Automatic Control 62.11 (2017): 5538-5553.

Decentralized Marginal Objective

Decentralized communication with marginal state estimates ($\Theta_{\mathcal{N}_i} = \{\Theta_k\}_{k\in\mathcal{N}_i}$)

- Consensus: Ensuring that agents achieve the same estimate on common domain $\Theta_{\mathcal{N}_{ij}} = \{\theta_k\}_{k\in\mathcal{N}_i} \cap \{\theta_k\}_{k\in\mathcal{N}_j}$
- Likelihood update: Including likelihood information at each time step

$$p_i = \arg\min_{\bar{p}_i} \left[\mathop{\mathbb{E}}_{z_{i,t}\sim\ell_i^\star(z_{i,t})} \mathop{\mathbb{E}}_{\Theta_{\mathcal{N}_i}\sim\bar{p}_i} [-\log(\ell_i(z_{i,t}|\Theta_{\mathcal{N}_i}))] \right] \quad \text{(Agent objective)}$$

$$s.t. \quad p_i(\Theta_{\mathcal{N}_{ij}}) = p_j(\Theta_{\mathcal{N}_{ij}}), \quad \forall j \in \mathcal{N}_i \quad \text{(Marginal consensus constraint)}$$

Marginal Consensus Step

- $p_{i,t}(\Theta_{\mathcal{N}_i})$: Estimated density function by agent i on variables contained in $\mathcal{X}_i$
- $p_{i,t}(\Theta_{\mathcal{N}_{ij}})$: Marginal density of $p_{i,t}(\Theta_{\mathcal{N}_i})$ computed over the common set of variables at agent j

Geometric marginal mixing with stochastic weights

$$p_{i,t}(\Theta_{\mathcal{N}_{ij}}) = \int_{\Theta_{\mathcal{N}_i} \setminus \Theta_{\mathcal{N}_{ij}}} p_{i,t}(\Theta_{\mathcal{N}_i}) \qquad \text{(Common marginal)}$$

$$p_{i,t}(\Theta_{\mathcal{N}_i} | \Theta_{\mathcal{N}_{ij}}) = \frac{p_{i,t}(\Theta_{\mathcal{N}_i})}{p_{i,t}(\Theta_{\mathcal{N}_{ij}})} \qquad \text{(Conditional density)}$$

$$\tilde{p}_{ji,t}(\Theta_{\mathcal{N}_i}) = p_{i,t}(\Theta_{\mathcal{N}_i} | \Theta_{\mathcal{N}_{ij}}) p_{j,t}(\Theta_{\mathcal{N}_{ij}}) \qquad \text{(Include marginal information)}$$

$$v_{i,t}(\Theta_{\mathcal{N}_i}) \propto \prod_{j \in \mathcal{N}_i} \tilde{p}_{ji,t}^{A_{ij}}(\Theta_{\mathcal{N}_i}) \qquad \text{(Mixing step)}$$

Analyzing Marginal Consensus Step

- **Marginal consensus manifold**
  - set of marginals consistent with some joint $\bar{p}$
  - $\mathcal{M} = \{\{p_{i,t}\}_{i=1}^{n} | \sum_{i=1}^{n} \mathrm{D_{KL}}[\bar{p}_i, p_{i,t}] = 0, p_{i,t} \in \mathcal{F}_{\mathfrak{d}_i}, \bar{p} \in \mathcal{F}\}$
- For any PDF $p \in \mathcal{F}$, the mixed and original PDFs $\{v_{i,t}\}, \{p_{i,t}\}$,
  - $\sum_{i=1}^{n} \mathrm{D_{KL}}[p_i, v_{i,t}] \leq \sum_{i=1}^{n} \mathrm{D_{KL}}[p_i, p_{i,t}]$
  - equality iff the original PDFs $\{p_{i,t}\} \in \mathcal{M}$
- With marginal consensus steps to PDFs $\{p_{i,t}\}$ in connected networks,
  - the resulting PDF $\lim_{k\to\infty} p_{i,t}^{(k)}$ lies in the consensus manifold $\mathcal{M}$.

Proposed Marginal Consensus Estimation Algorithm

$$v_{i,t}(\Theta_{\mathcal{N}_i}) \propto \prod_{j \in \mathcal{N}_i} \left(\frac{p_{i,t}(\Theta_{\mathcal{N}_i})}{p_{i,t}(\Theta_{\mathcal{N}_{ij}})} p_{j,t}(\Theta_{\mathcal{N}_{ij}}) \right)^{A_{ij}} \qquad \text{(Mixing step)}$$

$$\begin{aligned} p_{i,t+1}(\Theta_{\mathcal{N}_i}) &= \arg\min_{p \in \mathcal{F}_m} \left\{ \alpha_t \left\langle \frac{\delta F}{\delta p}(p_{i,t}, z_{i,t}), p \right\rangle + \mathrm{D_{KL}}(p || v_{i,t}) \right\} \\ &= \ell_i(z_{i,t}|\Theta_{\mathcal{N}_i})^{\alpha_t} v_{i,t}(\Theta_{\mathcal{N}_i}) \Big/ \left(\int \ell_i(z_{i,t}|\Theta_{\mathcal{N}_i})^{\alpha_t} v_{i,t}(\Theta_{\mathcal{N}_i}) d\Theta_{\mathcal{N}_i} \right) \end{aligned}$$

Analyzing marginal consensus estimation algorithm

- Assuming variable independence $p_{i,t}(\Theta_{\mathcal{N}_i}) = \prod_{\theta \in \Theta_{\mathcal{N}_i}} p_{i,t}(\theta)$
  - convergent PDF on **Marginal consensus manifold**,
  - $\bar{p}_t(\Theta) = \prod_{\theta \in \Theta} \bar{p}_t(\theta)$, $\bar{p}_t(\theta) \propto \prod_{j \in \mathcal{N}(\theta)} p_{j,t}(\theta)^{\frac{1}{|\mathcal{N}(\theta)|}}$.
  - where set of agents observing θ is $\mathcal{N}(\theta)$

Analyzing marginal consensus estimation algorithm

- Assuming variable independence $p_{i,t}(\Theta_{\mathcal{N}_i}) = \prod_{\theta \in \Theta_{\mathcal{N}_i}} p_{i,t}(\theta)$
  - convergent PDF on **Marginal consensus manifold**,
  - $\bar{p}_t(\Theta) = \prod_{\theta \in \Theta} \bar{p}_t(\theta)$, $\bar{p}_t(\theta) \propto \prod_{j \in \mathcal{N}(\theta)} p_{j,t}(\theta)^{\frac{1}{|\mathcal{N}(\theta)|}}$.
  - where set of agents observing θ is $\mathcal{N}(\theta)$
- With independent observations, connected network, bounded gradients, square summable step-sizes and variable independence,
  - the marginal PDF estimates $\{p_{i,t}\}_{i \in \mathcal{N}} \xrightarrow{a.s.} \mathcal{B}(\mathcal{F}_i^\star, \epsilon)$,
  - where the ϵ-partial neighborhood of optimal PDF $p^\star \in \mathcal{F}^\star$ is,

$$\mathbb{B}_i(\mathcal{F}^\star, \epsilon) = \left\{ p_i \in \mathcal{F}_i \middle| \min_{p^\star \in \mathcal{F}^\star} \mathrm{D_{KL}}[p_i^\star, p_i] \leq \epsilon, p_i^\star = \int_{\mathcal{X} \setminus \mathcal{X}_i} p^\star \right\}.$$

Gaussian Marginal Consensus Algorithm

Gaussian estimates with log-linear likelihoods lead to Gaussian posterior.

Consider a Gaussian PDF $\phi\left(\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix}\middle|\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix},\begin{bmatrix}\Omega_{11} & \Omega_{12}\\ \Omega_{21} & \Omega_{22}\end{bmatrix}^{-1}\right)$

Gaussian Marginal Consensus Algorithm

Common marginal

$$p_{i,t}(\Theta_{\mathcal{N}_{ij}}) = \int_{\Theta_{\mathcal{N}_i} \setminus \Theta_{\mathcal{N}_{ij}}} p_{i,t}(\Theta_{\mathcal{N}_i})$$

Conditional density

$$p_{i,t}(\Theta_{\mathcal{N}_i} | \Theta_{\mathcal{N}_{ij}}) = \frac{p_{i,t}(\Theta_{\mathcal{N}_i})}{p_{i,t}(\Theta_{\mathcal{N}_{ij}})}$$

Conditional marginal product:

$$\tilde{p}_{ji,t}(\Theta_{\mathcal{N}_i}) = p_{i,t}(\Theta_{\mathcal{N}_i} | \Theta_{\mathcal{N}_{ij}}) p_{j,t}(\Theta_{\mathcal{N}_{ij}})$$

Marginal density w.r.t. θ_1:

$$\phi\left(\theta_1 \middle| \boldsymbol{\mu}_1, \left(\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}\right)^{-1}\right)$$

Conditional distribution

$$(X_1 | X_2 = \theta_2) \sim \mathbf{N}\left(\mu_1 - \Omega_{11}^{-1}\Omega_{12}(\theta_2 - \mu_2), \Omega_{11}^{-1}\right)$$

Marginal distribution of X_2: $\mathcal{N}(\bar{\mu}_2, \bar{\Omega}_{22}^{-1})$
Joint distribution of (X_1, X_2)

$$\left(\begin{bmatrix} \mu_1 + \Omega_{11}^{-1}\Omega_{12}(\mu_2 - \bar{\mu}_2) \\ \bar{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}^{\top} & \bar{\Omega}_{22} + \Omega_{12}^{\top}\Omega_{11}^{-1}\Omega_{12} \end{bmatrix}^{-1}\right)$$

Gaussian Marginal Consensus Algorithm

Common marginal

$$p_{i,t}(\Theta_{\mathcal{N}_{ij}}) = \int_{\Theta_{\mathcal{N}_i} \setminus \Theta_{\mathcal{N}_{ij}}} p_{i,t}(\Theta_{\mathcal{N}_i})$$

Conditional density

$$p_{i,t}(\Theta_{\mathcal{N}_i}|\Theta_{\mathcal{N}_{ij}}) = \frac{p_{i,t}(\Theta_{\mathcal{N}_i})}{p_{i,t}(\Theta_{\mathcal{N}_{ij}})}$$

Conditional marginal product:

$$\tilde{p}_{ji,t}(\Theta_{\mathcal{N}_i}) = p_{i,t}(\Theta_{\mathcal{N}_i}|\Theta_{\mathcal{N}_{ij}})p_{j,t}(\Theta_{\mathcal{N}_{ij}})$$

Marginal density w.r.t. θ_1:

$$\phi\left(\theta_1 \middle| \boldsymbol{\mu}_1, \left(\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}\right)^{-1}\right)$$

Conditional distribution

$$(X_1|X_2 = \theta_2) \sim \mathbf{N}\left(\mu_1 - \Omega_{11}^{-1}\Omega_{12}(\theta_2 - \mu_2), \Omega_{11}^{-1}\right)$$

Marginal distribution of X_2: $\mathcal{N}(\bar{\mu}_2, \bar{\Omega}_{22}^{-1})$
Joint distribution of (X_1, X_2)

$$\left(\begin{bmatrix} \mu_1 + \Omega_{11}^{-1}\Omega_{12}(\mu_2 - \bar{\mu}_2) \\ \bar{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}^{\top} & \bar{\Omega}_{22} + \Omega_{12}^{\top}\Omega_{11}^{-1}\Omega_{12} \end{bmatrix}^{-1}\right)$$

Gaussian Marginal Consensus Algorithm [12]

- Mixing step

$$v_{i,t}(\Theta_{\mathcal{N}_i}) \propto \prod_{j \in \mathcal{N}_i} \tilde{p}_{ji,t}^{A_{ij}}(\Theta_{\mathcal{N}_i})$$

- Likelihood update

$$p_{i,t+1}(\Theta_{\mathcal{N}_i}) \propto \ell_i(z_{i,t}|\Theta_{\mathcal{N}_i})^{\alpha_t} v_{i,t}(\Theta_{\mathcal{N}_i})$$

- Gaussians $\phi(\theta|\mu_i, \Omega_i^{-1})$, $\Omega_w = \sum_{i=1}^n A_i \Omega_i$,

$$\prod_{i=1}^n \phi(\theta|\mu_i, \Omega_i^{-1})^{A_i} = \phi\left(\theta \middle| \Omega_w^{-1} \sum_{i=1}^n A_i \Omega_i \mu_i, \Omega_w^{-1}\right)$$

- Likelihood $\ell_i(z_{i,t}|\Theta_{\mathcal{N}_i}) = \phi\left(z_{i,t}|H_i \Theta_{\mathcal{N}_i}, V_i\right)$,

$$\phi(z_{i,t}|H_i \Theta_{\mathcal{N}_i}, V_i^{-1}) \phi\left(\Theta_{\mathcal{N}_i}; \mu, \Omega_i^{-1}\right) =$$
$$\mathbf{N}\left(\left(H_i^\top V_i H_i + \Omega_i\right)^{-1} (H_i^\top V_i z_{i,t} + \Omega_i \mu_i)\right.$$
$$\left., \left(H_i^\top V_i H_i + \Omega_i\right)^{-1}\right)$$

[11] **P. Paritosh**, N. Atanasov, and S. Martínez, "Marginal density averaging for distributed node localization from local edge measurements", In IEEE Conference on Decision and Control, 2020, pp. 2404-2410.

Estimation with Log-Linear Marginal Likelihoods

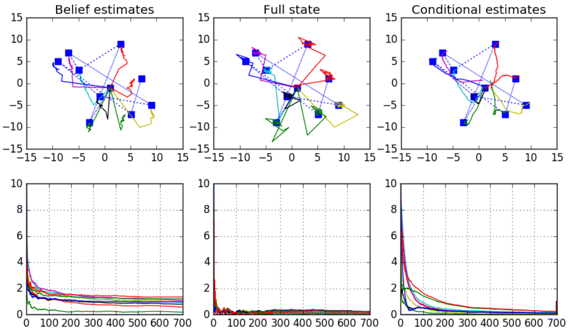
- A 10-node network with unknown locations $\theta = [\theta_i]_{i \in \mathcal{N}}, \theta_i \in \mathbb{R}^2$.
- Observations: $z_{ij} = (\theta_j - \theta_i) + \epsilon, \epsilon \sim \mathbf{N}(0, V_i), V_i = \mathbb{I}_2$
- Agent observation model $\ell_i(z_i|\theta) = \phi(z_i - H_i\theta, V_i)$,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

Existing algorithms

- Belief propagation
- Full state updates
- Proposed algorithm

Self-State Estimates

Belief propagation(BP), full state(FS) and marginal state estimates(MS) for a 10-agent ring network



(Row 1) Convergence to true positions
(Row 2) Estimation error across time

Effect of Edge Density

Estimation in multiple 10-node graphs with number of edges in $\{9, \cdots, 45\}$.

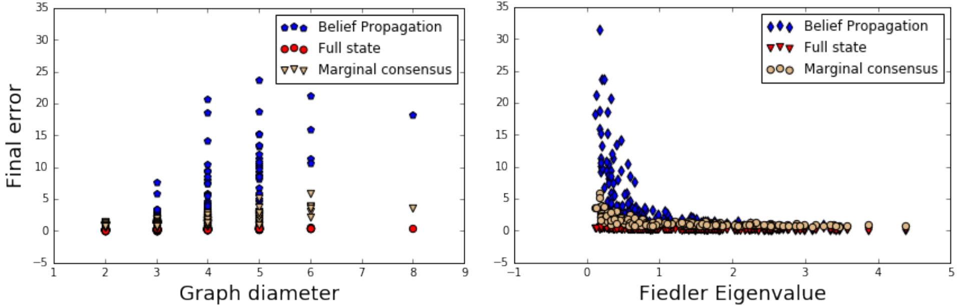


Figure: Error in self position estimates via BP, full and partial state estimation algorithms. (a) With increasing graph diameter after 500 steps. (b) With increasing connectivity captured by Fiedler Eigenvalue.

Marginal Distributed Mapping

Classification model on parameters Θ:

- Observed data $z = (x, y)$ with input $x \in \mathbb{R}^{\ell-1}$ and label $y \in \{0, 1\}$
- Model features $\Phi_i(x) \in \mathbb{R}^{l+1}$ with kernel elements:
  $\Phi_i(x) = [1, k_1(x), \ldots, k_l(x)], k_s(x) = \exp(-\gamma\|x - x^{(s)}\|^2)$
- Agent likelihood model with parameters $\Theta_{\mathcal{N}_i}$ and sigmoid function σ:

$$\ell(z|\Theta_{\mathcal{N}_i}) = \sigma(\Phi_i(x)^\top \Theta_{\mathcal{N}_i})^y (1 - \sigma(\Phi_i(x)^\top \Theta_{\mathcal{N}_i}))^{1-y}.$$

Agent estimates pdf $p_{i,t+1} = \phi(\cdot|\mu_{i,t+1}, \Omega_{i,t+1})$ from mixed pdf $p^v_{i,t} = \phi(\cdot|\mu^v_{i,t}, \Omega^v_{i,t})$ using variational inference,

$$\Omega_{i,t+1} = \Omega^v_{i,t} - \mathbb{E}_{p^v_{i,t}}[\nabla^2_{\Theta_i} \log \ell_i(z_{i,t+1}|\Theta_{\mathcal{N}_i})],$$

$$\mu_{i,t+1} = \mu^v_{i,t} + (\Omega^v_{i,t})^{-1}\mathbb{E}_{p^v_{i,t}}[\nabla_{\Theta_i} \log \ell_i(z_{i,t+1}|\Theta_{\mathcal{N}_i})].$$

Marginal Distributed Mapping

Classification model on parameters Θ:

- Observed data $z = (x, y)$ with input $x \in \mathbb{R}^{\ell-1}$ and label $y \in \{0, 1\}$
- Model features $\Phi_i(x) \in \mathbb{R}^{l+1}$ with kernel elements: $\Phi_i(x) = [1, k_1(x), \dots, k_l(x)]$, $k_s(x) = \exp(-\gamma\|x - x^{(s)}\|^2)$
- Agent likelihood model with parameters $\Theta_{\mathcal{N}_i}$ and sigmoid function σ:

$$\ell(z|\Theta_{\mathcal{N}_i}) = \sigma(\Phi_i(x)^\top \Theta_{\mathcal{N}_i})^y (1 - \sigma(\Phi_i(x)^\top \Theta_{\mathcal{N}_i}))^{1-y}.$$

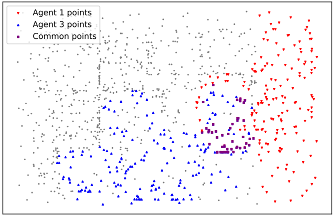
Agent estimates pdf $p_{i,t+1} = \phi(\cdot|\mu_{i,t+1}, \Omega_{i,t+1})$ from mixed pdf $p^v_{i,t} = \phi(\cdot|\mu^v_{i,t}, \Omega^v_{i,t})$ using variational inference,

$$\Omega_{i,t+1} = \Omega^v_{i,t} - \mathbb{E}_{p^v_{i,t}}[\nabla^2_{\Theta_i} \log \ell_i(z_{i,t+1}|\Theta_{\mathcal{N}_i})],$$
$$\mu_{i,t+1} = \mu^v_{i,t} + (\Omega^v_{i,t})^{-1}\mathbb{E}_{p^v_{i,t}}[\nabla_{\Theta_i} \log \ell_i(z_{i,t+1}|\Theta_{\mathcal{N}_i})].$$

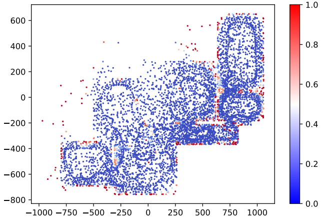
Distributed Mapping with Intel LIDAR Dataset

- Classification model with 1000 feature points
- Trajectory adjacent models at agents with $(208, 195, 247, 188, 180, 224, 216)$ points

(a) Training data



(b) Distinct/ shared feature points



(c) Predicted occupancy probability

Figure: Seven agent LiDAR dataset for distributed mapping.

Contributions

- Introduced distributed inference algorithm on **partial set of variables**
- Marginal consensus characterization and almost sure convergence guarantees
- Developed Gaussian version of the marginal consensus algorithm
- Simulations studying trade-offs with Belief propagation and Full-state algorithms

Publications:

- P. Paritosh, N. Atanasov, and S. Martínez, "Hypothesis assignment and partial likelihood averaging for cooperative estimation", In IEEE CDC, 2019, pp. 7850-7856.
- P. Paritosh, N. Atanasov and S. Martinez. Marginal Density Averaging for Distributed Node Localization from Local Edge Measurements. In IEEE CDC, 2020, pp. 2404-2410.
- P. Paritosh, N. Atanasov and S. Martinez. Distributed Bayesian Estimation in Sensor Networks: Consensus on Marginal Densities. Submitted to IEEE Transactions on Network Science and Engineering.

Summary

- Distributed estimation with mild requirements on connectivity and likelihood
- Multi-agent Gaussian and particle estimation algorithms
- Distributed variational inference for differentiable log-likelihoods
- Multi-robot mapping demonstration in simulation and on Turtlebots
- Distributed marginal estimation algorithm analysis
- Simulations studying trade-offs with Belief propagation and Full-state algorithms

1 Introduction

2 Estimation in Continuous Spaces
- Distributed Estimation as Optimization
- Convergence guarantees

3 Distributed Density Estimation
- Proposed algorithm and convergence guarantees
- Cooperative localization and parameter estimation

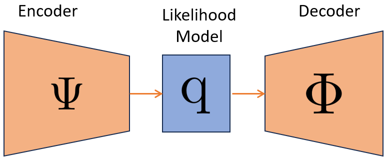
4 Distributed Variational Inference
- Distributed ELBO
- Distributed Gaussian Variational Inference
- Distributed Mapping: Simulation and Implementation

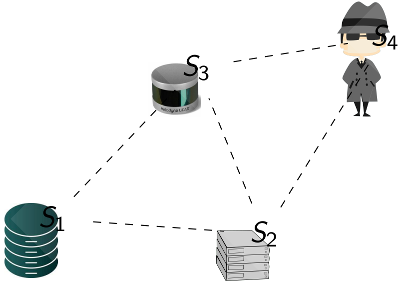
5 Distributed Marginal Estimation
- Research Question
- Decentralized Communication and Storage
- Gaussian Marginal Consensus Algorithm
- Simulation

6 Future Directions

Future Directions: Scalable Algorithms for Heterogeneous Sensing Networks

- Physics-informed modeling of sensing likelihoods
- Information fusion in distributed heterogeneous networks
- Reasoning with trust in sensing networks
- Providing convergence guarantees for real-time operation





Publications

1. P. Paritosh, N. Atanasov, and S. Martínez, "Distributed Bayesian Estimation in Sensor Networks: Consensus on Marginal Densities", Under review at IEEE Transactions on Network Science and Engineering.
2. P. Paritosh, N. Atanasov, and S. Martínez, "Distributed Variational Inference for Online Supervised Learning", Under review at Transactions on Signal Processing.
3. P. Paritosh, N. Atanasov, and S. Martínez, "Distributed Bayesian Estimation of Continuous Variables Over Time-Varying Directed Networks", in IEEE Control Systems Letters, vol. 6, pp. 2545-2550, 2022. (Joint submission with IEEE CDC)
4. P. Paritosh, N. Atanasov, and S. Martínez, "Marginal density averaging for distributed node localization from local edge measurements", In IEEE Conference on Decision and Control, 2020, pp. 2404-2410.
5. P. Paritosh, N. Atanasov, and S. Martínez, "Hypothesis assignment and partial likelihood averaging for cooperative estimation", In IEEE Conference on Decision and Control, 2019, pp. 7850-7856.

Thank you

Theorem: Convergence Guarantees

Assumptions

- Bounded likelihoods (lower)
- Independent Observations
- Positive priors

Functional convergence

For square summable step sizes α_t, the PDF sequence $\{p_t\}$ converges almost surely to an ϵ-divergence neighborhood,

$$\mathcal{B}(\mathcal{F}^\star, \epsilon) = \{p \in \mathcal{F}_d | \min_{p^\star \in \mathcal{F}^\star} \mathrm{D_{KL}}[p^\star, p] \leq \epsilon\}$$

around the set of minimizers in $\mathcal{F}^\star$ for any $\epsilon > 0$.

Convergence rate

There exis $\alpha_t < (f[p_t] - f[p^\star])/2L^2$, the expected objective function satisfies,

$$f[\bar{p}_t] - f[p^\star] \leq \sqrt{\frac{8L^2 \,\mathrm{D_{KL}}[p^\star, p_0]}{t}},$$

where $\bar{p}_t = \frac{1}{t}\sum_{k=1}^{t} p_k$ and minimizer $p^\star \in \mathcal{F}^\star$.

Analyzing marginal consensus estimation algorithm

- Assuming variable independence $p_{i,t}(\Theta_{\mathcal{N}_i}) = \prod_{\theta \in \Theta_{\mathcal{N}_i}} p_{i,t}(\theta)$
  - convergent PDF on **Marginal consensus manifold**,
  - $\bar{p}_t(\Theta) = \prod_{\theta \in \Theta} \bar{p}_t(\theta)$, $\bar{p}_t(\theta) \propto \prod_{j \in \mathcal{N}(\theta)} p_{j,t}(\theta)^{\frac{1}{|\mathcal{N}(\theta)|}}$.
  - where set of agents observing θ is $\mathcal{N}(\theta)$
- With independent observations, connected network, bounded gradients, square summable step-sizes and variable independence,
  - the marginal PDFs $\{p_{i,t}\}_{i \in \mathcal{N}}$ converge almost surely to partial neighborhood $\mathcal{B}(\mathcal{F}_i^\star, \epsilon)$ around optimal PDF set $\mathcal{F}^\star$ for any $\epsilon > 0$,
  - where the ϵ-partial neighborhood of PDF $p^\star \in \mathcal{F}^\star$ is,

$$\mathbb{B}_i(\mathcal{F}^\star, \epsilon) = \left\{ p_i \in \mathcal{F}_i \middle| \min_{p^\star \in \mathcal{F}^\star} \mathrm{D}_{\mathrm{KL}}[p_i^\star, p_i] \leq \epsilon, p_i^\star = \int_{\mathcal{X} \setminus \mathcal{X}_i} p^\star \right\}.$$

Proof Elements

- Define log-probability and log-likelihood terms, $r_{i,t}(\theta) = \log\left[\frac{p_{i,t}(\theta)}{p_{i,t}(\theta_\star)}\right]$, $g_{i,t}(\theta) = \log\left[\frac{\ell_i(z_{i,t}|\theta)}{\ell_i(z_{i,t}|\theta_\star)}\right]$

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

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

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: $\mathbb{E}\left[\exp\left(bg_{12}(z_i)\right)\right] < \infty$

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 $x^\star \in \theta^\star$.

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: $e_0 = [A_t \dots A_0 \boldsymbol{r}_0]_i, e_k = \alpha[A_t \dots A_k \boldsymbol{g}_k]_i, \psi_k(b) = \mathbb{E}[\exp(be_k)]$

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

Gaussian full state updates

$\mathbf{N}(\mu_{i,t}, \Omega_{i,t}^{-1})$: Normal density representing agent i's estimate over the space $\mathcal{X}$

$$\Omega_{i,t} = \sum_{j \in \mathcal{N}_i} \Omega_{j,t-1};\ \mu_{i,t} = \Omega_{i,t}^{-1}\left(\sum_{j \in \mathcal{N}_i} \Omega_{j,t-1}\mu_{j,t-1}\right).$$

Belief Propagation

- $m_{t,ij}(\theta_i)$: Message from agent i to agent j
- $p_{i,t}(\theta_i)$: Agent i's estimate over the variable θ_i

$$m_{t,ij}(\theta_i) = \sum_{x_i} \ell_i(z_{ij}|\theta_i, \theta_j) p_{i,t}(\theta_i) \prod_{k \in \mathcal{N}_j \setminus i} m_{t-1,kj}(\theta_i),$$

$$p_{i,t}(\theta_i) = \frac{p_{i,t-1}(\theta_i) \prod_{k \in \mathcal{N}_i} m_{ki}(\theta_i)}{\sum_{j=1}^{n} p_{j,t-1}(\theta_j) \prod_{k \in \mathcal{N}_j} m_{kj}(\theta_j)}.$$

Gaussian Belief Propagation

A Gaussian BP algorithm for agents with observation model $z_i = H\begin{bmatrix}\theta_i & \theta_j\end{bmatrix}^\top + \epsilon, \epsilon \sim \mathbf{N}(0_{d\times 1}, \Omega_i^z)$, with $H = \begin{bmatrix}-1, 1\end{bmatrix} \otimes \mathbb{I}_d$, where $\otimes$ is a kronecker product. The update rule for each agent is given as

$$\Omega_{jj,t} = \sum_{i\in\mathcal{N}_j} \Omega_{ij,t-1};\ \mu_{jj,t} = \Omega_{jj,t}^{-1}\left(\sum_{i\in\mathcal{N}_j} \Omega_{ij,t-1}\mu_{ij,t-1}\right),$$

which depends on the messages sent to j from $i \in \mathcal{N}_j$:

$$\Omega_{ij,t} = \begin{bmatrix}\Omega_{ii,t} - \Omega_{ji,t-1} & 0 \\ 0 & 0\end{bmatrix} + H_i^\top \Omega_i^z H_i,$$

$$\mu_{ij,t} = \Omega_{ij,t}^{-1}\left(\begin{bmatrix}\sum_{k\in\{\mathcal{N}_i\setminus j\}} \Omega_{ki,t-1}\mu_{ki,t-1} \\ 0\end{bmatrix} + H_i^\top \Omega_i^z z_{ij,t}\right).$$

Marginal Averaging Algorithm

We present the Gaussian estimate equivalent to the four algorithm steps in the following lemmas. Here, we denote a Gaussian random variable $\mathbf{N}(\mu, \Omega^{-1})$ with mean μ and information matrix as Ω, and its associated density function as $\phi(\cdot|\mu, \Omega^{-1})$.

Neighbor messages

The marginal density of the Gaussian PDF $\phi\left(\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix} \middle| \begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}, \begin{bmatrix}\Omega_{11} & \Omega_{12}\\ \Omega_{21} & \Omega_{22}\end{bmatrix}^{-1}\right)$ with respect to θ_1 is given as,

$$\phi\left(\theta_1 \middle| \boldsymbol{\mu}_1, \left(\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}\right)^{-1}\right).$$

Gaussian Marginal Algorithm

Pre-edge merging

Let (X_1, X_2) be random vectors represented by a joint Gaussian distribution with mean $\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}$ and information matrix $\Omega = \begin{bmatrix}\Omega_{11} & \Omega_{12}\\ \Omega_{21} & \Omega_{22}\end{bmatrix}$. The PDF associated with conditional distribution is,$(X_1|X_2 = \theta_2) \sim \mathbf{N}\left(\mu_1 - \Omega_{11}^{-1}\Omega_{12}(\theta_2 - \mu_2), \Omega_{11}^{-1}\right)$

Edge merging

Let X_1, X_2 be random vectors with a joint Gaussian distribution. Assume that X_1 conditioned on $X_2 = x_2$ is distributed as $\mathcal{N}(\mu_1 - \Omega_{11}^{-1}\Omega_{12}(x_2 - \mu_2), \Omega_{11}^{-1})$ and that the marginal distribution of X_2 is $\mathcal{N}(\bar{\mu}_2, \bar{\Omega}_{22}^{-1})$. Then, X_1 and X_2 joint distribution is

$$\mathcal{N}\left(\begin{bmatrix}\mu_1 + \Omega_{11}^{-1}\Omega_{12}(\mu_2 - \bar{\mu}_2)\\ \bar{\mu}_2\end{bmatrix}, \begin{bmatrix}\Omega_{11} & \Omega_{12}\\ \Omega_{12}^{\top} & \bar{\Omega}_{22} + \Omega_{12}^{\top}\Omega_{11}^{-1}\Omega_{12}\end{bmatrix}^{-1}\right).$$

Gaussian Marginal Algorithm

Lemma (Geometric averaging)

Let $\Omega_w = \sum_{i=1}^{n} A_i \Omega_i$. The weighted geometric product of Gaussian density functions $\phi(\theta|\mu_i, \Omega_i^{-1}), \forall i \in \{1, \ldots, n\}$ with corresponding weights A_i is given as,

$$\prod_{i=1}^{n} \phi(\theta|\mu_i, \Omega_i^{-1})^{A_i} = \phi\left(\theta \middle| \Omega_w^{-1} \sum_{i=1}^{n} A_i \Omega_i \mu_i, \Omega_w^{-1}\right).$$

Gaussian Marginal Algorithm

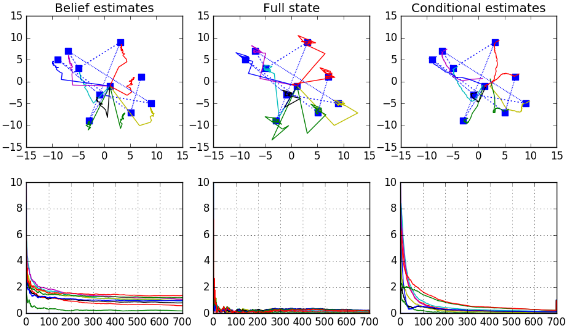
Lemma (Likelihood update)

Let the likelihood density be described as $\ell_i(z_{i,t}|\mathcal{X}_i) = \phi\left(z_{i,t}|H_i\mathcal{X}_i, V_i\right)$. Then the posterior Gaussian density obtained as likelihood prior product $\phi\left(z_{i,t}|H_i\mathcal{X}_i, V_i^{-1}\right)\phi\left(\mathcal{X}_i; \mu, \Omega_i^{-1}\right)$ is

$$\mathbf{N}\left(\left(H_i^\top V_i H_i + \Omega_i\right)^{-1}\left(H_i^\top V_i z_{i,t} + \Omega_i \mu_i\right),\ \left(H_i^\top V_i H_i + \Omega_i\right)^{-1}\right)$$

Self-State Estimates

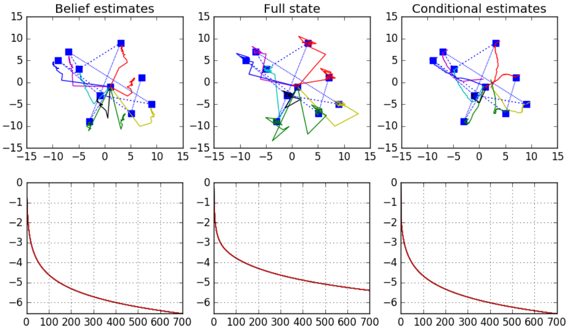
Belief propagation, full state and partial state estimates for a 10-agent ring network



(Row 1) Convergence to true positions
(Row 2) Estimation error across time

Self-State Estimates

Belief propagation, full state and partial state estimates for a 10-agent ring network



(Row 1) Convergence to true positions
(Row 2) Logarithm of maximum eigenvalue of the self-covariance estimates

Comparing the Communicated Information

Transmitting an m-dimensional Gaussian density requires transmitting $m + m^2$ floating point numbers.

Table: Comparing the iterations and communicated units for convergence to the error $\epsilon = 0.1$ in a 25-node graph

	Iterations			Information units		
	BP	FS	CS	BP	FS	CS
Line	NA	18	1356	NA	2203k	1301k
100 edges	9	2	28	29.8k	846.6k	291k
287 edges	7	2	15	51k	1856k	2709k

Decentralized Update

Consensus: Geometric mixing with stochastic weights

$$v_{i,t} = \frac{1}{Z^v_{i,t}} \prod_{j=1}^{n} p_{j,t}^{A_{ij}}, \quad Z^v_{i,t} = \int_{\theta \in \mathcal{X}} \left(\prod_{j=1}^{n} p_{j,t}^{A_{ij}} \right) \qquad \text{(Mixing step)}$$

Likelihood update: SMD algorithm

$$\begin{aligned} p_{i,t+1} &= \arg\min_{p \in \mathcal{F}_m} \left\{ \alpha_t \left\langle \frac{\delta F}{\delta p}(p_{i,t}, z_{i,t}), p \right\rangle + \mathrm{D_{KL}}(p || v_{i,t}) \right\} \\ &= \exp\left(\alpha_t \frac{\delta F}{\delta p} \right) v_{i,t} \Big/ \left(\int \exp(\alpha_t \frac{\delta F}{\delta p}) v_{i,t}(\theta) d\theta \right) \end{aligned}$$