THE UNIVERSITY *of* EDINBURGH
School of Engineering

HERIOT WATT UNIVERSITY

Distributed Estimation of Latent Parameters in State Space Models Using Separable Likelihoods

Murat Üney†, Bernard Mulgrew†, Daniel Clark‡

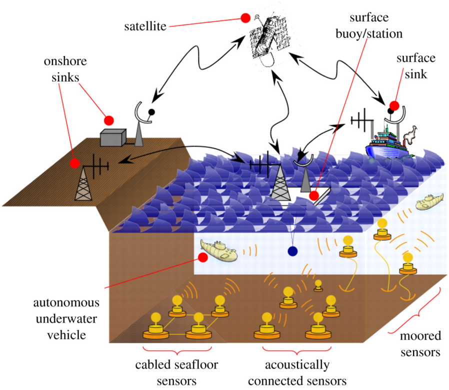
† The University of Edinburgh
‡ Heriot-Watt University

24/03/2016

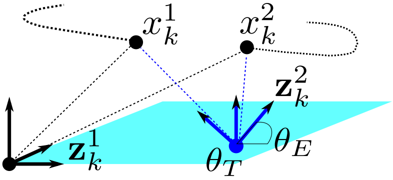
UDRC

ERPem
Edinburgh Research Partnership in Engineering and Mathematics

Motivation



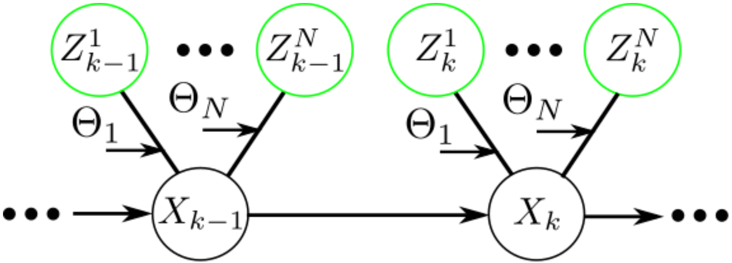
An underwater surveillance network (Heidemann et. al, 2011, adapted from Akyildiz, et.al, 2005)



Opportunistic sensor registration: Estimation of respective sensor position θ_T and orientation θ_E using measurements $\mathbf{z}$ from objects x over $k = 1, ..., t$.

Example: Sensor localisation in GPS denying environments such as underwater.

Multi-sensor state space model



- The object state X_k evolves as a Markov chain with transition density $\pi(x_k|x_{k-1})$ and initial density $\pi(x_1)$
- Sensor i measures z^i_k with a likelihood $l_i(z^i_k|x_k;\theta_i)$
- If latent parameters $\theta = [\theta_1, \ldots, \theta_N]$ are known, only unknown is X_k which is estimated using sensor histories $z^1_{1:k}, \ldots, z^N_{1:k}$
- Solved by finding $p(x_k|z^1_{1:k}, \ldots, z^N_{1:k}, \theta)$ for $k = 1, \ldots, t$ using Bayesian prediction and update recursions, i.e., "filtering".

Centralised solution for estimating unknown θ

The likelihood for “parameter estimation in state space models”

$$l\left(z^1_{1:t}, ..., z^N_{1:t} | \theta = [\theta_1, \ldots, \theta_N]\right) = \prod_{k=1}^{t} p(z^1_k, \ldots, z^N_k | z^1_{1:k-1}, \ldots, z^N_{1:k-1}, \theta)$$

$$p(z^1_k, \ldots, z^N_k | z^1_{1:k-1}, \ldots, z^N_{1:k-1}, \theta) = \int \left(\prod_{j=1}^{N} l(z^j_k | x_k, \theta) \right) \times \underbrace{p(x_k | z^1_{1:k-1}, \ldots, z^N_{1:k-1}, \theta)}_{\text{Prediction distribution of a (centralised) filter.}} \mathrm{d}(x_k)$$

- The likelihood is a product of update terms over $k = 1, \ldots, t$
- Computational cost is dominated by joint multi-sensor filtering.
- The filtering complexity is combinatorial with the number of sensors N (dimensionality of θ), when the state-space model is equipped with models tackling uncertainties in the number of objects, measurement-object associations etc.

Separable likelihoods

- Provide approximate models based on single sensor filtering, i.e., use local prediction $p(x_k|z^j_{1:k-1})$ and update $p(x_k|z^j_{1:k})$ for $j \in \mathcal{V}$.
  - provide scalability with the number of sensors
  - align well with distributed fusion architectures
  - exploit recent advances in single sensor filtering algorithms
- First, we consider a pair of sensors i and j.

Quad-term separable likelihood

Approximate $l\left(z^i_{1:t}, z^j_{1:t}|\theta\right)$ with $\prod_{k=1}^{t} q(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta)$

$$q(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta) \triangleq \frac{1}{\kappa_k(\theta)} \left(p(z^i_k|z^j_{1:k}, \theta)p(z^j_k|z^j_{1:k-1}, \theta)\right)^{1/2} \times \left(p(z^j_k|z^i_{1:k}, \theta)p(z^i_k|z^i_{1:k-1}, \theta)\right)^{1/2}$$

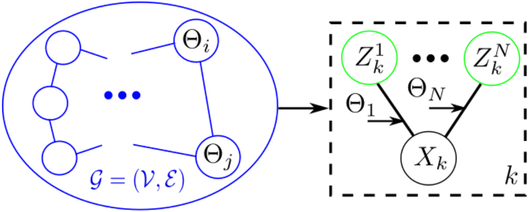
where $\kappa_k(\theta)$ is the normalisation constant.

Theorem (Kullback-leibler divergence of the approximation)

$$
\begin{aligned}
D(p||q) \leqslant \frac{1}{2}\Bigg(& \Big(H(X_k|Z^j_{1:k-1},\Theta) - H(X_k|Z^j_{1:k-1}, Z^i_{1:k-1},\Theta)\Big) \\
& + \Big(H(X_k|Z^i_{k-1},\Theta) - H(X_k|Z^j_{1:k-1}, Z^i_{1:k-1},\Theta)\Big)\Bigg) \\
& + \frac{1}{2}\Bigg(\Big(H(X_k|Z^j_{1:k},\Theta) - H(X_k|Z^j_{1:k}, Z^i_{1:k-1},\Theta)\Big) \\
& + \Big(H(X_k|Z^i_{1:k},\Theta) - H(X_k|Z^i_{1:k}, Z^j_{1:k-1},\Theta)\Big)\Bigg), \qquad (1)
\end{aligned}
$$

- Better approximation when local prediction and estimation at sensors i and j are accurate (small difference in Shannon Entropies H with respect to joint filtering).

Pairwise MRF Model with Separable Likelihood Edge Potentials



- Second, we assume that the local latent parameters $\mathbf{\Theta}_i$ are Markov with respect to $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with edges associated with, e.g., available communication links, neighbourhood relations etc...

$$p(\theta | Z^1_{1:t}, ..., Z^N_{1:t}) \propto \prod_{i \in \mathcal{V}} \psi_i(\theta_i) \prod_{(i,j) \in \mathcal{E}} \psi^t_{ij}(\theta_i, \theta_j),$$

$$\psi_i(\theta_i) = p_{0,i}(\theta_i), \quad \psi^t_{ij}(\theta_i, \theta_j) = \prod_{k=1}^{t} q(z^i_k, z^j_k | z^i_{1:k-1}, z^j_{1:k-1}, \theta_i, \theta_j)$$

Message passing algorithm for sensor localisation

for all $j \in \mathcal{V}$ **do** $\quad\triangleright$ Local filtering

  Find $p(x_k|z^j_{1:k})$ for $k = 1, \ldots t$

end for

for all $j \in \mathcal{V}$ **do** $\quad\triangleright$ Sample from priors

  Sample $\theta_i^{(l)} \sim p_{0,i}(\theta_i)$ for $l = 1, \ldots, L$

end for

for $s = 1, ..., S$ **do** $\quad\triangleright$ S-steps of loopy belief propagation (LBP)

  for all $(i,j) \in \mathcal{E}$ **do** $\quad\triangleright$ Evaluate separable likelihood edge potentials

    Find $\psi^t_{i,j}(\theta_i^{(l)}, \theta_j^{(l)}) = \prod_{k=1}^{t} q(z^i_k, z^j_k | z^i_{1:k-1}, z^j_{1:k-1}, \theta_i^{(l)}, \theta_j^{(l)})$ for $l = 1, \ldots, L$

  end for

  for all $(i,j) \in \mathcal{E}$ **do** $\quad\triangleright$ Find LBP messages m_{ji}s

    Sample $\tilde{\theta}_i^{(l)}$ from $m_{ji}(\theta_i) = \int \psi^t_{ij}(\theta_i, \theta_j)\psi_j(\theta_j) \prod_{i' \in ne(j)\setminus i} m_{i'j}(\theta_j)\,\mathrm{d}\theta_j$ for $l = 1, \ldots, L$

  end for

  for all $i \in \mathcal{V}$ **do** $\quad\triangleright$ Update local marginals $p_i(\theta_i)$s

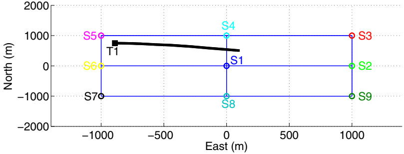
    Sample $\theta_i^{(l)}$ from $p_i(\theta_i) \propto \psi_i(\theta_i) \prod_{j \in ne(i)} m_{ji}(\theta_i)$ for $l = 1, \ldots, L$

    $\hat{\theta}_i \leftarrow \frac{1}{L}\sum_{l=1}^{L} \theta_i^{(l)}$

  end for

end for

Example

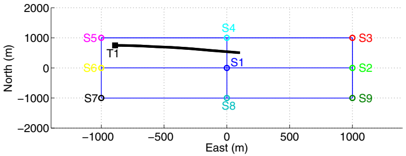


$$\pi(x_k|x_{k-1}) = \mathcal{N}(x_k; \mathbf{F}x_{k-1}, \mathbf{Q})$$

$$l_i(z_k^i|x_k; \theta_i) = \mathcal{N}(z_k^i; \mathbf{H}_i(x_k - \theta_i), \mathbf{R}_i)$$

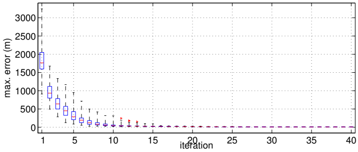
- Linear model & additive Gaussian uncertainties conditioned on θ
- θ_is are unknown sensor locations
- sensor 1 is the origin of the network coordinate frame

Example

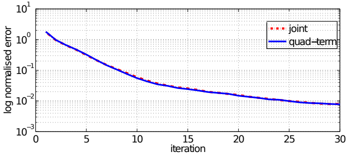


$$\pi(x_k|x_{k-1}) = \mathcal{N}(x_k; \mathbf{F}x_{k-1}, \mathbf{Q})$$
$$l_i(z_k^i|x_k;\theta_i) = \mathcal{N}(z_k^i; \mathbf{H}_i(x_k - \theta_i), \mathbf{R}_i)$$

- Linear model & additive Gaussian uncertainties conditioned on θ
- θ_is are unknown sensor locations
- sensor 1 is the origin of the network coordinate frame



- $< 10m$ average error
- Average of 12.2s per edge per iteration (compared to 28.3s with joint filtering edge pot.)



- Similar performance with the joint filtering edge pot., in this example.

Conclusion

- We consider multi-sensor state space models underpinning fusion networks and surveillance applications and address scalability of parameter estimation with the number of sensors.
- We propose a quad-term separable likelihood which together with pairwise MRFs facilitate scalability and distributed estimation.
- We relate the approximation quality to the accuracy of object state estimation and prediction using local filtering.
- The proposed likelihood can be used with hypothesis based multi-object filters in more complicated scenarios [1], and can be contrasted with a dual-term approximation introduced recently [2].

[1] Uney, Mulgrew, Clark "Latent parameter estimation in fusion networks using separable likelihoods," IEEE TSIPN, submitted to the special issue on inference and learning over networks.

[2] Uney, Mulgrew, Clark "A cooperative approach to sensor localisation in distributed fusion networks," IEEE TSP, March 2016.

Thank you very much for your attention, Questions & Comments?

Quad-term separable likelihood

Approximate $l\left(z^i_{1:t}, z^j_{1:t}|\theta\right)$ with $\prod_{k=1}^{t} q(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta)$

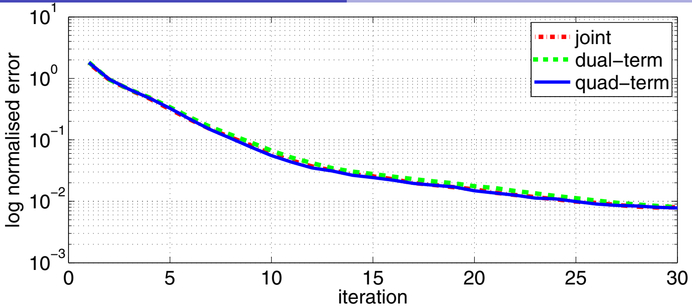
$$q(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta) \triangleq \frac{1}{\kappa_k(\theta)}\left(p(z^i_k|z^j_{1:k}, \theta)p(z^j_k|z^j_{1:k-1}, \theta)\right)^{1/2}$$
$$\times\left(p(z^j_k|z^i_{1:k}, \theta)p(z^i_k|z^i_{1:k-1}, \theta)\right)^{1/2}$$

where $\kappa_k(\theta)$ is the normalisation constant.

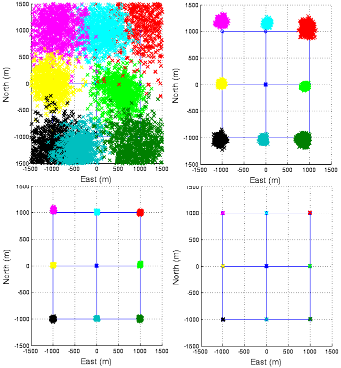
Dual-term separable likelihood

Approximate $l\left(z^i_{1:t}, z^j_{1:t}|\theta\right)$ with $\prod_{k=1}^{t} s(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta)$

$$s(z^i_k, z^j_k|z^i_{1:k-1}, z^j_{1:k-1}, \theta) \triangleq \underbrace{p(z^i_k|z^j_{1:k-1}, \theta)}_{\int p(z^i_k|x_k;\theta))\, \underbrace{p(x_k|z^j_{1:k-1})}_{\text{info from sensor } j \text{ to } i}\, \mu(\mathrm{d}x_k)} \times p(z^j_k|z^i_{1:k-1}, \theta)$$



- The performances of the joint filtering edge-potential, dual-term separable likelihood and the quad-term separable likelihood are similar, in this example scenario.



- $< 10m$ average error
- Average of 12.2s per edge per iteration (compared to 28.3s with joint filtering edge pot.)