Bayesian filtering of state-space models with unknown covariance matrices

Tomáš Vlk

FIT ČVUT v Praze

9.12.2021

Overview

- Introduction into filtering
- Proposed approach on single node
  - Results
- Introduction into distributed settings
- Proposed approach for distributed setting
  - Results
- Further resources

Introduction

- Bayesian approach is quite simillar to the way humans learn
  - Having some form prior knowleadge
  - Discovering some new information
  - Intercorporating the new knowleadge into what we have already known

Current knowleadge & new information → update knowleadge

State-space models

- General state-space model

$$x_t = A_t x_{t-1} + B_t u_t + \omega_\mathrm{t} \qquad x_t \sim \mathcal{N}(A_t x_{t-1} + B_t u_t, Q_t)$$

$$y_{i,t} = H_t x_t + \epsilon_{i,t} \qquad y_{i,t} \sim \mathcal{N}(H_t x_t, R_t)$$

- Example can be a constant velocity model (CVM)

$$x_k = \begin{bmatrix} 1 & 0 & \Delta k & 0 \\ 0 & 1 & 0 & \Delta k \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} [pos_{x,k-1} \quad pos_{y,k-1} \quad vel_{x,k-1} \quad vel_{y,k-1}] + w_k$$

$$y_k^{(i)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} [pos_{x,k} \quad pos_{y,k} \quad vel_{x,k} \quad vel_{y,k}] + v_k^{(i)}$$

Filtration: estimation of state by Kalman filter

- Starting with prior distribution

$\mathcal{N}(\hat{x}^+_{i,t-1}, P^+_{i,t-1})$

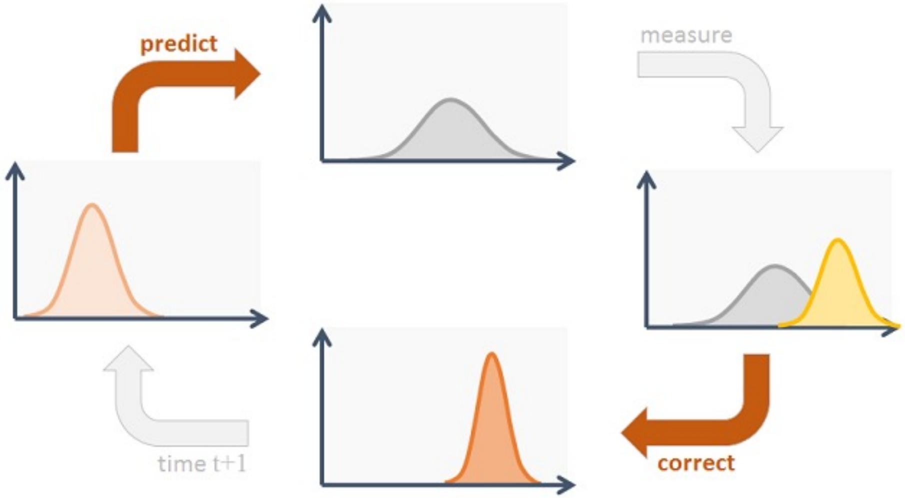
- Transform it during prediction

$\mathcal{N}\left(\hat{x}^-_{i,t}, P^-_{i,t}\right) = \mathcal{N}(A_t \hat{x}^-_{i,t}, A_t P^-_{i,t} A^T_t + Q_t)$

- Update by the Bayes' theorem

$\mathcal{N}(\hat{x}^+_{i,t}, P^+_{i,t})$

- Problem is the need-to-know covariances R_t a Q_t in both steps



Variational inference

- Unknown $\theta_t = [x_t, P_t, R_t]$
- Bayes' theorem does not lead to analytically tractable posterior distribution
- Approximation using $\rho_i(\theta_t) \equiv \rho_i(x_t)\rho_i(P_t)\rho_i(R_t)$
- Using variational inference to minimize divergence

$$\mathcal{D}\left[\rho_i(\theta_t)||\pi_i\left(\theta_t\middle|\Delta_{i,t}\right)\right] = \mathbb{E}_{\rho_i(\theta_t)}\left[\log\frac{\rho_i(\theta_t)}{\pi_i(\theta_t|\Delta_{i,t})}\right]$$

$$= -\mathcal{L}[\rho_i(\theta_t)] + logf(y_{i,t}|\Delta_{i,t-1}, u_t)$$

Variational inference

- Optimization of divegence is equivalent to maximalization of negative evidence lower bound (ELBO)
- Can be done by coordinate-ascent variational inference (CAVI)
  - Iterative optimization algorithm
- Need to have matching conjugate priors

Variable		**Prior**
x_t	$\sim$	$\mathcal{N}(\hat{x}^-_{i,t}, \hat{P}^\star_{i,t})$
P_t	$\sim$	$i\mathcal{W}(\Psi^-_{i,t}, \psi^-_{i,t})$
R_t	$\sim$	$i\mathcal{W}(\Phi^-_{i,t}, \phi^-_{i,t})$

Optimazation of $\hat{Q}_{i,t}$

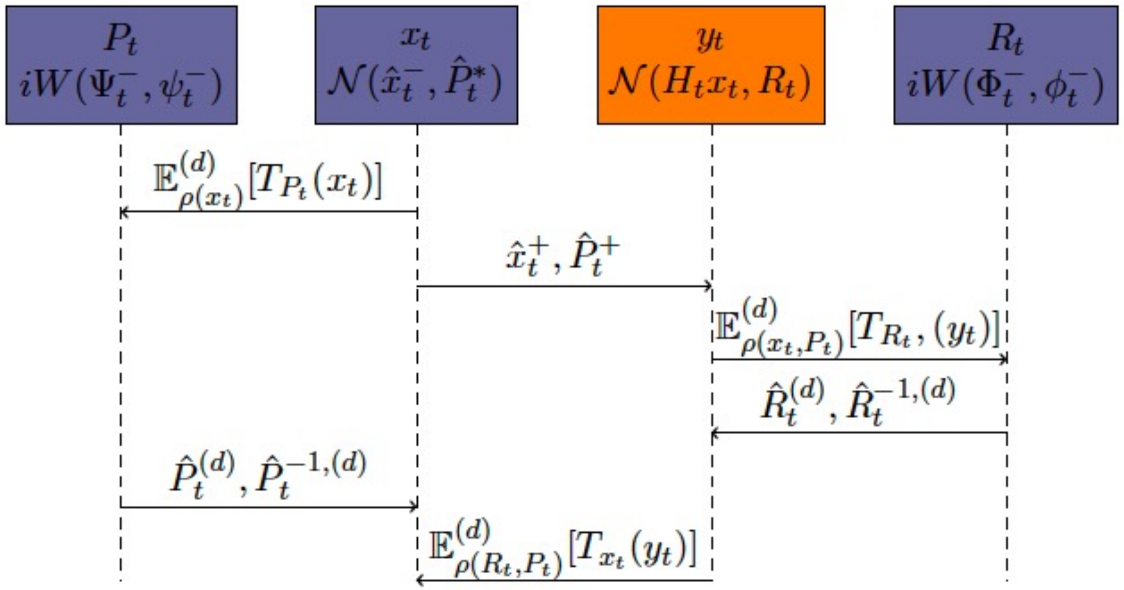
- User provided set of $\hat{Q}_{i,t}$ matrices
  $\mathcal{Q}_{i,t} = \{\hat{Q}_{i,t}^{(1)}, \dots, \hat{Q}_{i,t}^{(C)}\}$
- Selection of the matrix that maximizes

$$\hat{Q}_{i,t} = \underset{\tilde{Q}_t \in \mathcal{Q}_{i,t}}{\arg\max} \log \mathcal{N}\left(y_{i,t} | H_t \hat{x}_{i,t}^{-}, R(\tilde{Q}_t)\right)$$

- Based on the available measurements of $y_{i,t}$

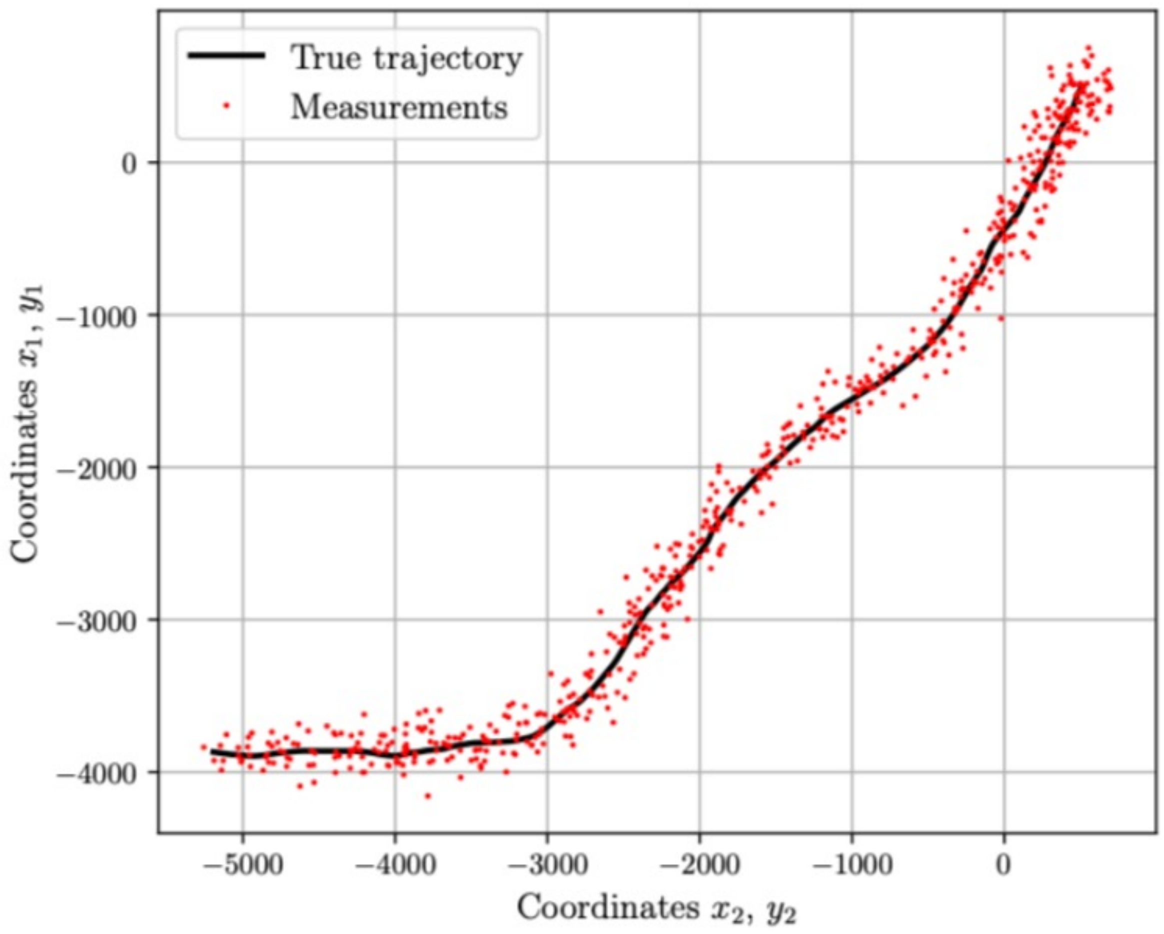
Summary of the proposed solution

- Approximation of x_t, P_t and R_t using variational inference by means of message passing.
- Approximation of Q_t is done using a cheap hypotheses-testing procedure and an intrinsic optimization of a relevant prior distribution

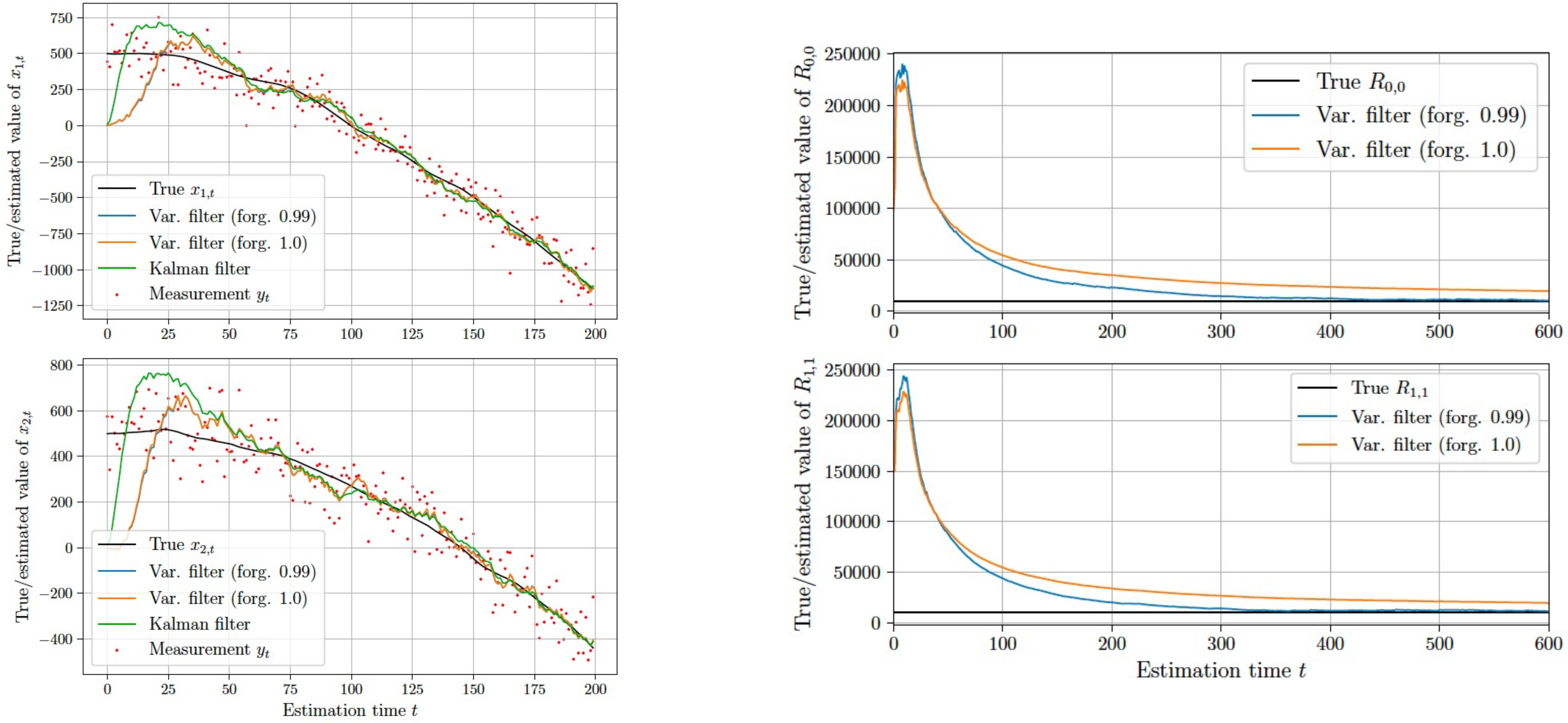


Testing

- 2 dimensional target tracking
- Simulated using constant velocity model
- CAVI algorithm is always set to 4 itterations



Results (non distributed setting)

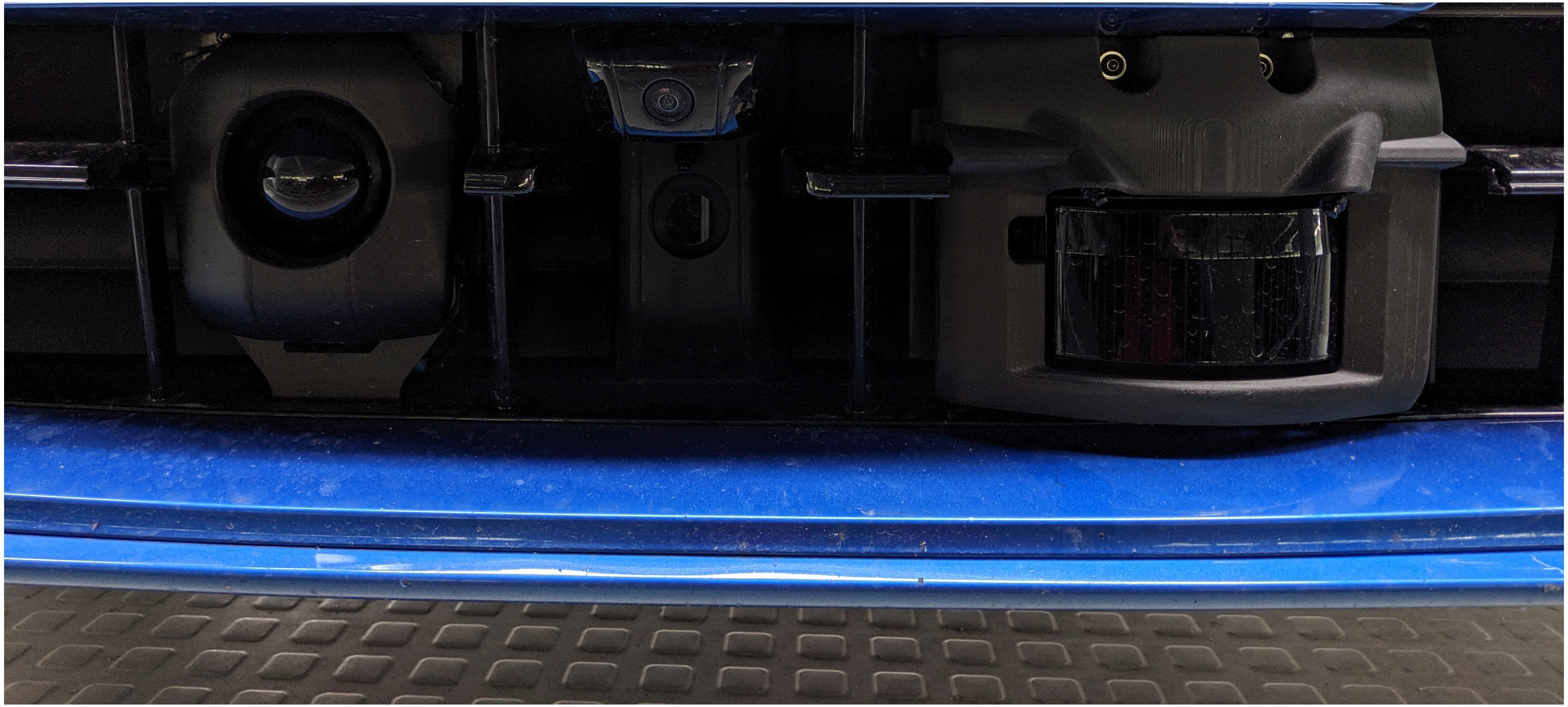


What about distributed setting?

Distributed setting intuition

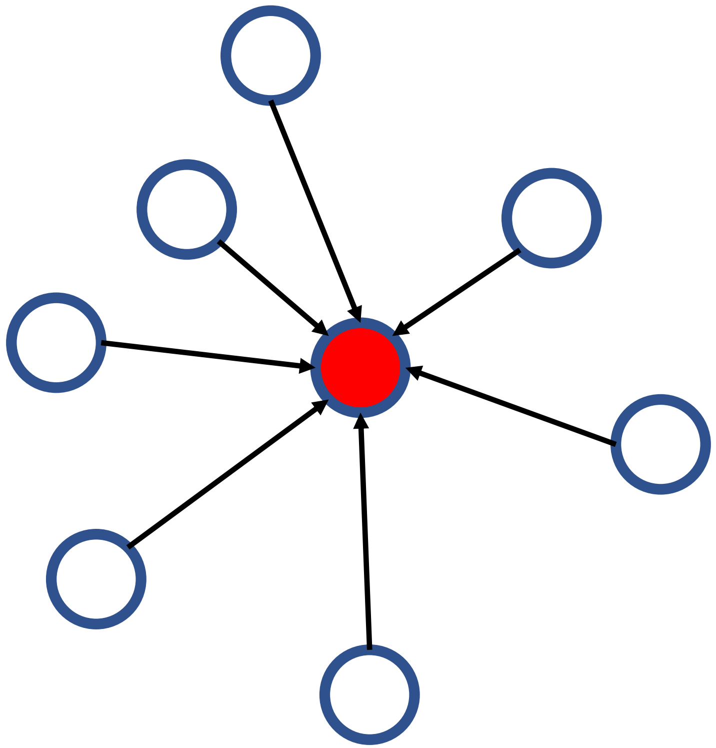
- Combining information from more sources
  - Using information from multiple sources can prove usefull
  - Problem can be with source with high trust but incorrect information
  - A lot of different ways to communicate
- Many open topics
  - How to detect that both sources observe the same object?
  - How to handle various sources in different conditions?
  - What is the optimal way to weight informations provided by different sources?
  - What is the best way to intercorporate the indidual weeknesses of various sources?

Does such a situation even occur?



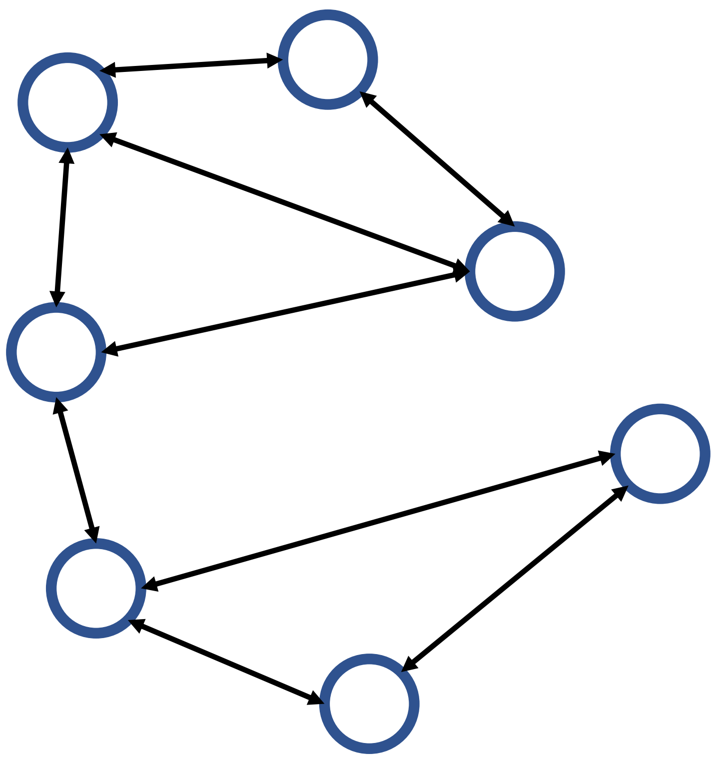
Communication approaches: Fusion Center

- All nodes send information to single processing node
- All computations done in one node only
- Risky, since processing node is single point of failure



Comunication approaches: Diffusion

- No specialized node for processing
- Nodes share information with their neighbours (usually in 1-step distance)
- Any node can fail and the network will still run
- Done in two steps:
  - Adaptation
  - Combination



Diffusion strategy

- Adaptation step
  - Interoperating observations from the neighbours into the nodes knowledge
- Combination step
  - Posterior information of the node is shared back to its neighbour nodes

- Two variants possible:
  - ATC – Adapt Then Combine (used in this case)
  - CTA – Combine Then Adapt

Adaptation step

- Using the available measurements of $y_{j,t}$ from neighbours node to accelerate convergence of the estimates of θ_t
- The measurement update step gets extended by the measurements from neighbours
- CAVI updates are replaced by expected sufficient statistics

$$\mathbb{E}^{(d)}_{\rho_i(R_t,P_t)}\left[\sum_{j\in\mathcal{I}_i} T_{x_t}(y_{j,t})\right] = \sum_{j\in\mathcal{I}_i} \begin{bmatrix} y_{j,t}^{\intercal} \\ H_t^{\intercal} \end{bmatrix} \left(\hat{R}^{+,(d)}_{i,t}\right)^{-1} \begin{bmatrix} y_{j,t}^{\intercal} \\ H_t^{\intercal} \end{bmatrix}^{\intercal}$$

$$\mathbb{E}^{(d)}_{\rho_i(x_t,P_t)}\left[\sum_{j\in\mathcal{I}_i} T_{R_t}(y_{j,t})\right]$$

$$= \sum_{j\in\mathcal{I}_i} \begin{bmatrix} \left(y_{j,t} - H_t \hat{x}^{+,(d-1)}_{i,t}\right)\left(\bullet\right)^{\intercal} + H_t \hat{P}^{+,(d-1)}_{i,t} H_t^{\intercal} \\ 1 \end{bmatrix}$$

Combination step

- Agent acquires posterior estimates for its neighbours
- In this case it is represented by variational factors
  - $\rho_j(x_t)$
  - $\rho_j(R_t)$
- Through out time by means of fusion the information is diffused between all interconnected nodes
- Different combination rules are possible with various properties

$$\widetilde{\rho}_i(x_t) \propto \exp\left\{\eta_{x_t}^{\intercal} \cdot \widetilde{\Xi}_{x_t,i}^{+}\right\} = \exp\left\{\eta_{x_t}^{\intercal} \cdot \frac{1}{|\mathcal{I}_i|} \sum \Xi_{x_t,j}^{+}\right\}$$

$$\widetilde{\rho}_i(R_t) \propto \exp\left\{\eta_{R_t}^{\intercal} \cdot \widetilde{\Xi}_{R_t,i}^{+}\right\} = \exp\left\{\eta_{R_t}^{\intercal} \cdot \frac{1}{|\mathcal{I}_i|} \sum \Xi_{R_t,j}^{+}\right\}$$

Distributed optimization of $\hat{Q}_{i,t}$

- Advantage of having increased amount of measurents provided by its neighbourghs
- They are independent identically distributed (iid)
- Joint predictive density is just a product of individual densities

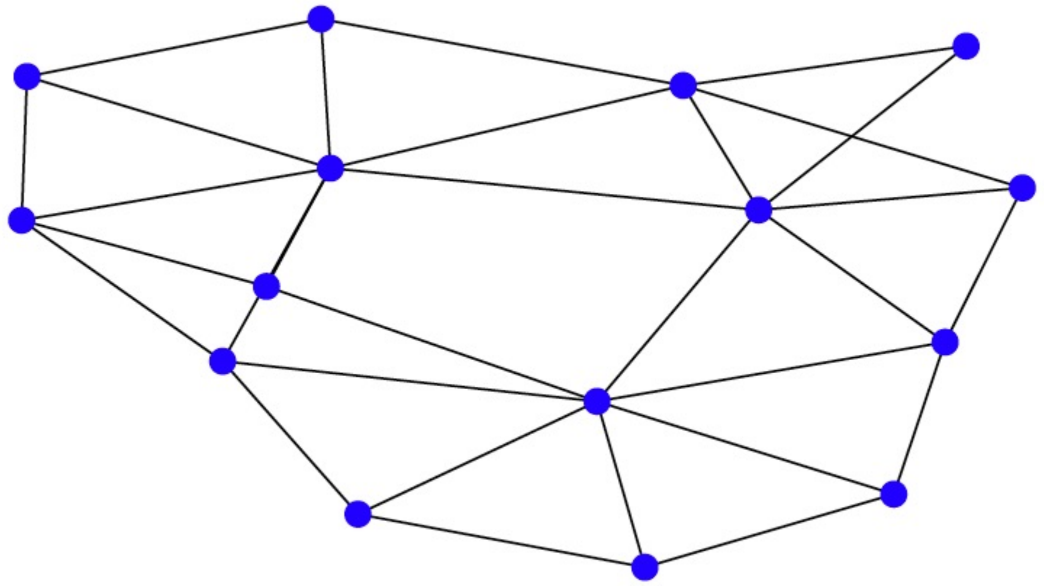
$$f\left(\{y_{j,t}\}_{j\in\mathcal{I}_i}|\Delta_{i,t-1},u_t\right)=\prod_{j\in\mathcal{I}_i} f(y_{j,t}|\Delta_{i,t-1},u_t)$$

- Optimal solution is therefore

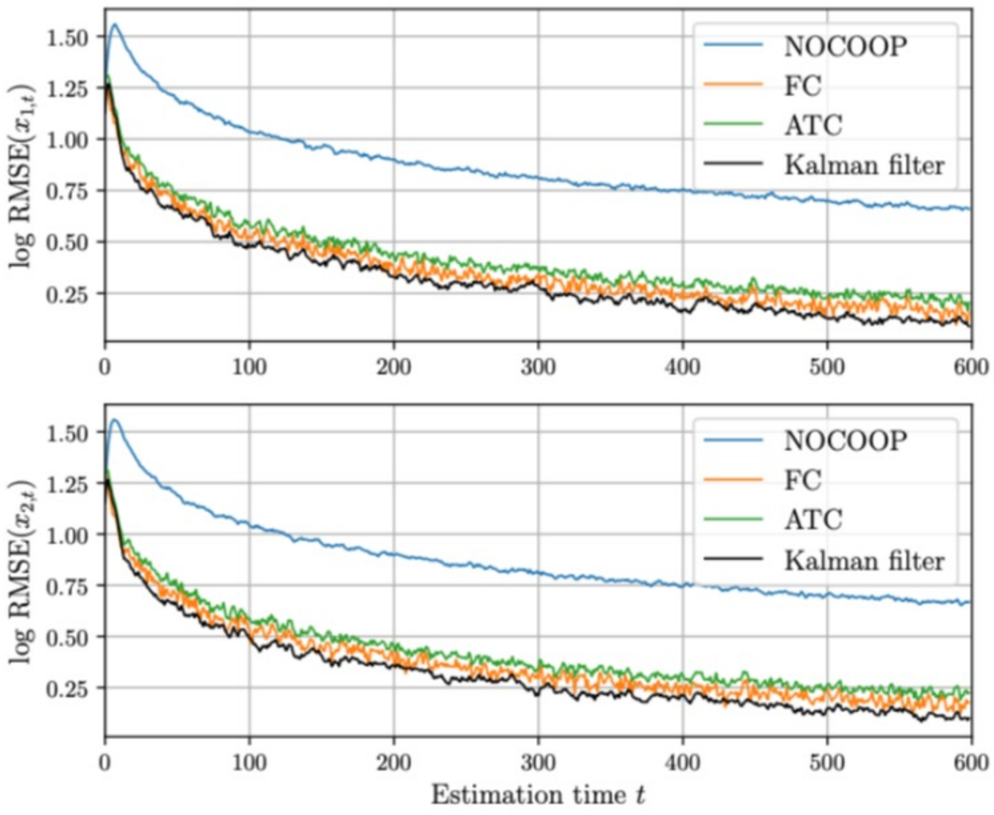
$$\begin{aligned}\hat{Q}_{i,t} &= \arg\max_{\tilde{Q}_t\in\mathcal{Q}_{i,t}} \log\prod_{j\in\mathcal{I}_i}\mathcal{N}\left(y_{j,t}|H_t\hat{x}^-_{i,t},R(\tilde{Q}_t)\right)\\ &= \arg\max_{\tilde{Q}_t\in\mathcal{Q}_{i,t}} \sum_{j\in\mathcal{I}_i}\log\mathcal{N}(y_{j,t}|H_t\hat{x}^-_{i,t},R(\tilde{Q}_t))\end{aligned}$$

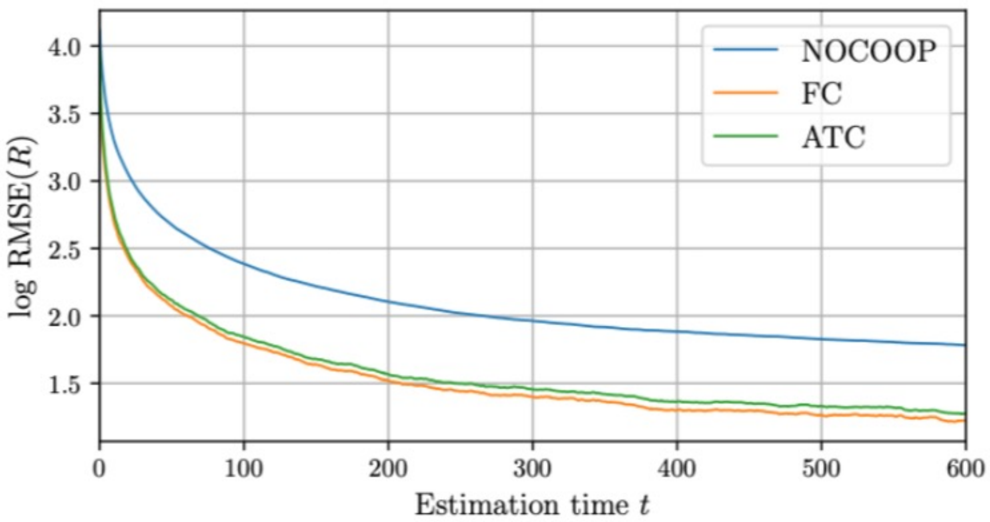
Results (distributed setting)

- Simulated data 2 dimensional target tracking
- Generated using constant velocity model
- 15 agents with independent observations of the target
- Results of 300 independent runs
- Two experiments
  - Static R matrix
  - Time-varying R matrix

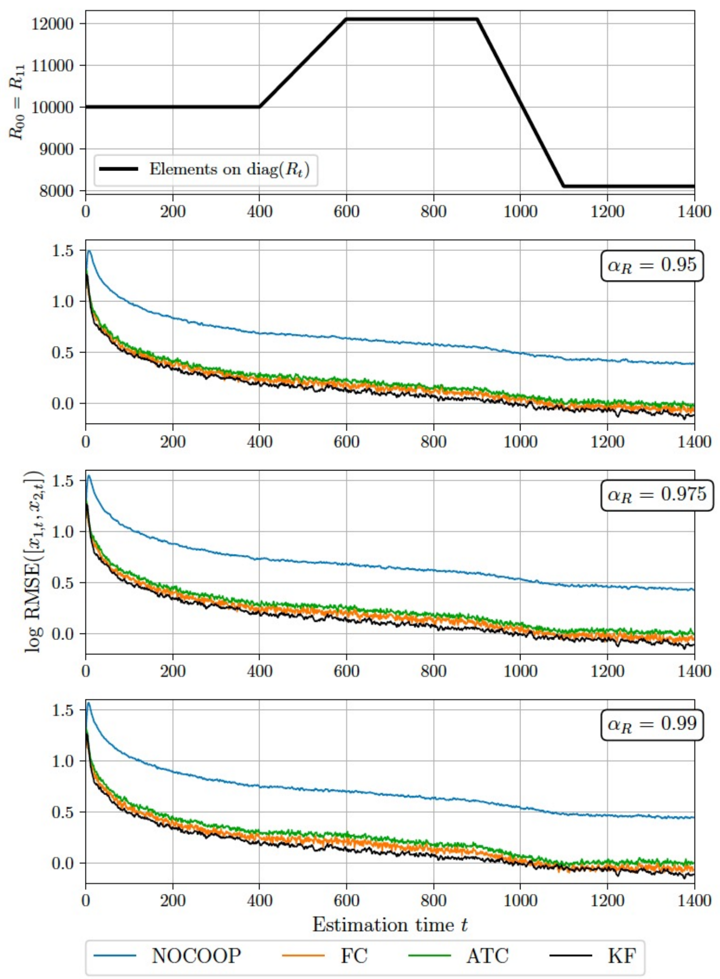


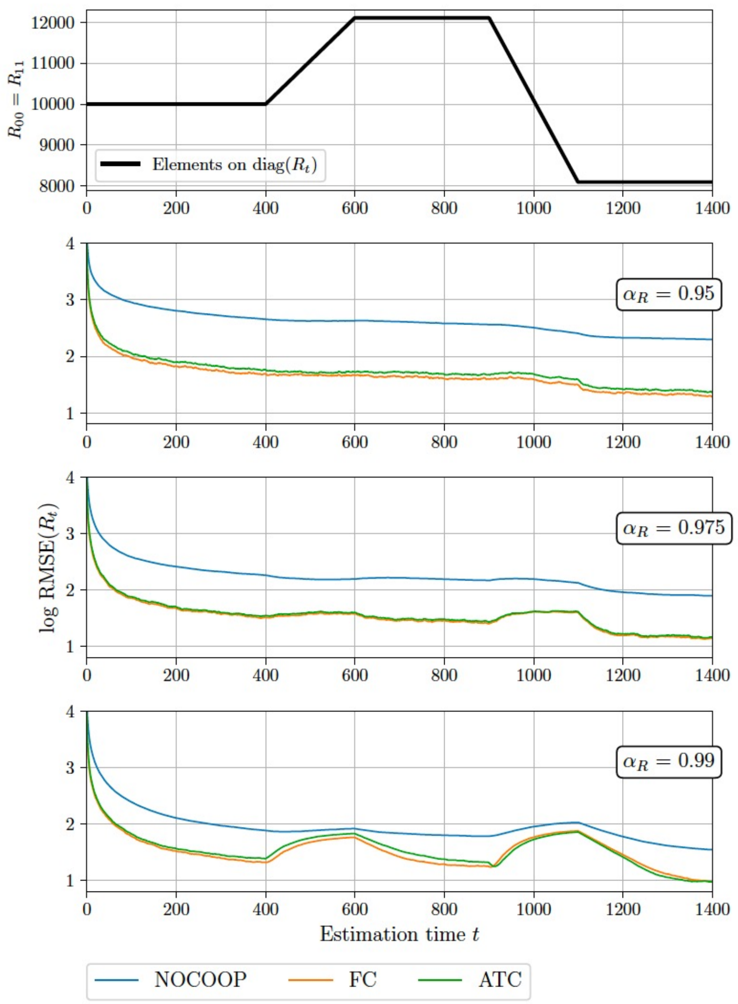
Results (static R matrix)





Results (time varying R matrix)





Further reading

1. Y. Huang, Y. Zhang, Z. Wu, N. Li, and J. Chambers, “A Novel Adaptive Kalman Filter With Inaccurate Process and Measurement Noise Covariance Matrices,” IEEE Trans. Automat. Contr., vol. 63, no. 2, pp. 594-601, Feb. 2018.
2. K. Dedecius and O. Tich ́y, “Collaborative sequential state estimation under unknown heterogeneous noise covariance matrices,” IEEE Trans.
   Signal Process., vol. 68, pp. 5365–5378, 2020.

Thanks for your attention