Diffusion Estimation of State-Space Models

Bayesian Formulation

Kamil Dedecius dedecius@utia.cas.cz

Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic

Studied problem

We study the problem of *distributed sequential estimation (filtration) of state-space models* in diffusion networks from the Bayesian viewpoint. The network is a (un/directed) graph of nodes $i = 1, \ldots, N$, cooperating within neighborhoods $\mathcal{N}_i$, i.e., one edge distance. The diffusions avoid intermediate averaging iterations between two subsequent data updates, typical for the consensus algorithms, thus saving communication and computation resources [1, 2]. The present contribution formulates the diffusion approach for general state-space models with Markov-type transitions and proposes an adaptive method for determination of combiners weights. A special focus is given to the Kalman filter.

Bayesian filtration

Bayesian formulation of models

Assume filtration of the latent state $x_t, t = 1, 2, \ldots$ of the state-space models with observable output y_t given by

$$
\begin{aligned}
x_t &= f(x_{t-1}, u_t, w_t)\\
y_t &= g(x_t, v_t)
\end{aligned}
$$

where f and g are known functions, u_t denotes a known input, v_t and w_t are zero-centered mutually independent noise variables.
In the Bayesian realm, this model takes the form of conditional probability distributions

$$
x_t | x_{t-1}, u_t \sim \pi(x_t | x_{t-1}, u_t) \tag{1}
$$

$$
y_t | x_t \sim p(y_t | x_t). \tag{2}
$$

Steps of the Bayesian filtration

Prediction using the prior pdf $\pi(x_{t-1}|Y_{t-1}, U_{t-1})$, the conditional distribution (1) and the Chapman-Kolmogorov equation we obtain

$$
\pi(x_t | U_t, Y_{t-1}) = \int \pi(x_t | x_{t-1}, u_t) \pi(x_{t-1} | Y_{t-1}, U_{t-1}) dx_{t-1}.
$$

Update by the obtained observation y_t using the Bayes' theorem

$$
\pi(x_t | U_t, Y_t) = \frac{\pi(x_t | U_t, Y_{t-1}) p(y_t | x_t)}{\int \pi(x_t | U_t, Y_{t-1}) p(y_t | x_t) dx_t}.
$$

(Analytically tractable under certain conditions – see below)

Bayes in the exponential family

Definition 1 (Exponential family)
An exponential family of distributions of a variable y_t conditioned by a state x_t is a family of distributions with the probability density function (pdf) of the form

$$
f(y_t | x_t) = h(y_t) g(x_t) \exp\left[\eta(x_t) T(y_t)\right],
$$

where $h(\cdot)$ is a known function, $g(\cdot)$ is a known normalization function, $\eta(x_t)$ is a natural parameter and $T(\cdot)$ is a sufficient statistic.

Definition 2 (Conjugate prior pdf)
The conjugate prior pdf for a state x_t with the hyperparameters ξ_{t-1} of the same dimension as $T(\cdot)$ and $\nu_{t-1} \in \mathbb{R}^+$ has the form

$$
\pi(x_t | \xi_{t-1}, \nu_{t-1}) = q(\xi_{t-1}, \nu_{t-1}) g(x_t)^{\nu_{t-1}} \exp\left[\eta(x_t) \xi_{t-1}\right],
$$

where $q(\cdot)$ is a normalization function and $g(\cdot)$ has the same form as in the exponential family.

The **Bayes' theorem**

$$
\pi(x_t | \xi_t, \nu_t) \propto \pi(x_t | \xi_{t-1}, \nu_{t-1}) f(y_t | x_t)
$$

updates the hyperparameters

$$
\begin{aligned}
\xi_t &= \xi_{t-1} + T(y_t) \\
\nu_t &= \nu_{t-1} + 1
\end{aligned} \tag{3}
$$

Adaptation step

Node i incorporates the observations $y_{j,t}$, $j \in \mathcal{N}_i$ weighted by $c_{ij} \in [0, 1]$ (subjective probability that jth node's information is valid). Using the Bayes' theorem,

$$
\pi_i(x_t | \tilde{U}_{i,t}, \tilde{Y}_{i,t}) \propto \pi_i(x_t | \tilde{U}_{i,t}, \tilde{Y}_{i,t-1}) \prod_{j \in \mathcal{N}_i} p(y_{j,t} | x_t)^{c_{ij}},
$$

where tilde denotes the variables affected by the shared information. Under conjugacy, the diffusion update counterpart of (3) reads

$$
\begin{aligned}
\xi_{i,t} &= \xi_{i,t-1} + \sum_{j \in \mathcal{N}_i} c_{ij} T(y_{j,t}) \\
\nu_{i,t} &= \nu_{i,t-1} + 1.
\end{aligned}
$$

Combine step

The Kullback-Leibler optimal approximation

$$
\sum_{j \in \mathcal{N}_i} a_{ij} \, \mathrm{D}\left(\pi_i^* || \pi_j\right)
$$

(nonnegative weights a_{ij} sum to unity for fixed i) is given by the weighted geometric mean [3]

$$
\pi_i^*(x_t | \tilde{U}_{i,t}, \tilde{Y}_{i,t}) \propto \prod_{j \in \mathcal{N}_i} \pi_j(x_t | \tilde{U}_{j,t}, \tilde{Y}_{j,t})^{a_{ij}}.
$$

Again, under conjugate priors,

$$
\xi_{i,t}^* = \sum_{j \in \mathcal{N}_i} a_{ij} \xi_{j,t} \quad \text{and} \quad \nu_{i,t}^* = \sum_{j \in \mathcal{N}_i} a_{ij} \nu_{j,t}.
$$

Determination of weights a_{ij} and c_{ij}

c_{ij} may be interpreted as the probabilities that the measurements from $j \in \mathcal{N}_i$ obey the model $p(y_{j,t}|x_t)$ given $\pi_i(x_t|\tilde{U}_{i,t}, \tilde{Y}_{i,t-1})$ thus $c_{ij} \sim \mathrm{Beta}(r_{ij,t}, s_{ij,t})$. This distribution may be updated based on the presence of $y_{j,t}$ in a high-credibility region $\mathcal{Y}_{ij,t}$ of the predictive pdf $p(y_{j,t}|\tilde{U}_{i,t}, \tilde{Y}_{i,t-1})$. The Bayesian update (again in the exp. family) then reads

$$
r_{ij,t} = r_{ij,t-1} + \mathbb{1}[y_{j,t} \in \mathcal{Y}_{ij,t}] \quad \text{and} \quad s_{ij,t} = s_{ij,t-1} + (1 - \mathbb{1}[y_{j,t} \in \mathcal{Y}_{ij,t}]), \quad \text{yielding} \quad \hat{c}_{ij,t} = \frac{r_{ij,t}}{r_{ij,t} + s_{ij,t}}
$$

Adaptation with highly reliable data (due to optimized $\hat{c}_{ij,t}$) suppresses the sensitivity to the choice of a_{ij}. It is possible to proceed, e.g., with strategies proposed in [2] or [4]. Note that while a_{ij} sum to unity, c_{ij} do not need to.

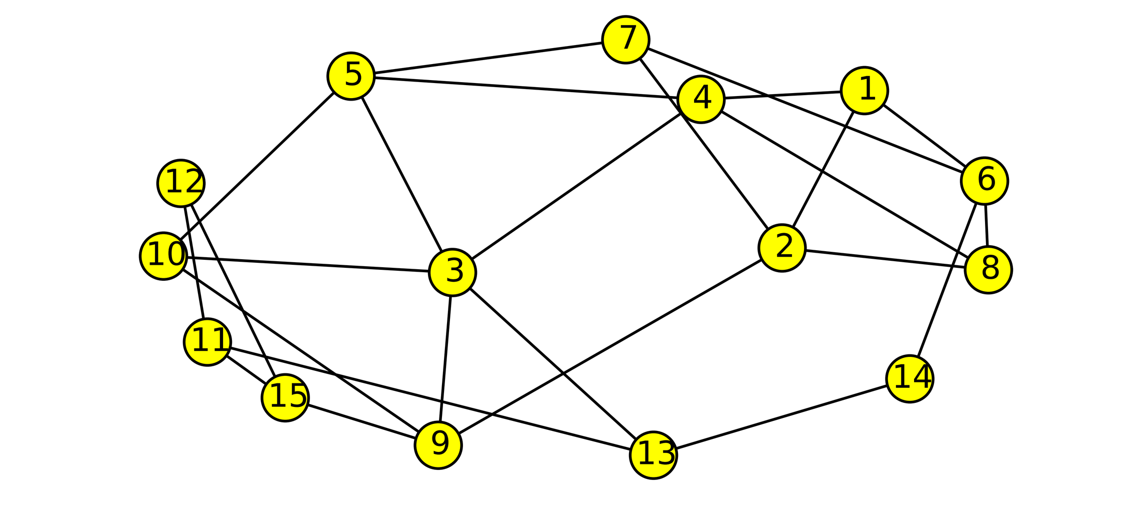
Numerical example (Kalman filter)

2D tracking problem using a network of 15 nodes, one of them being faulty. $T = 100$ simulated measurements were generated with an input-free linear state space model

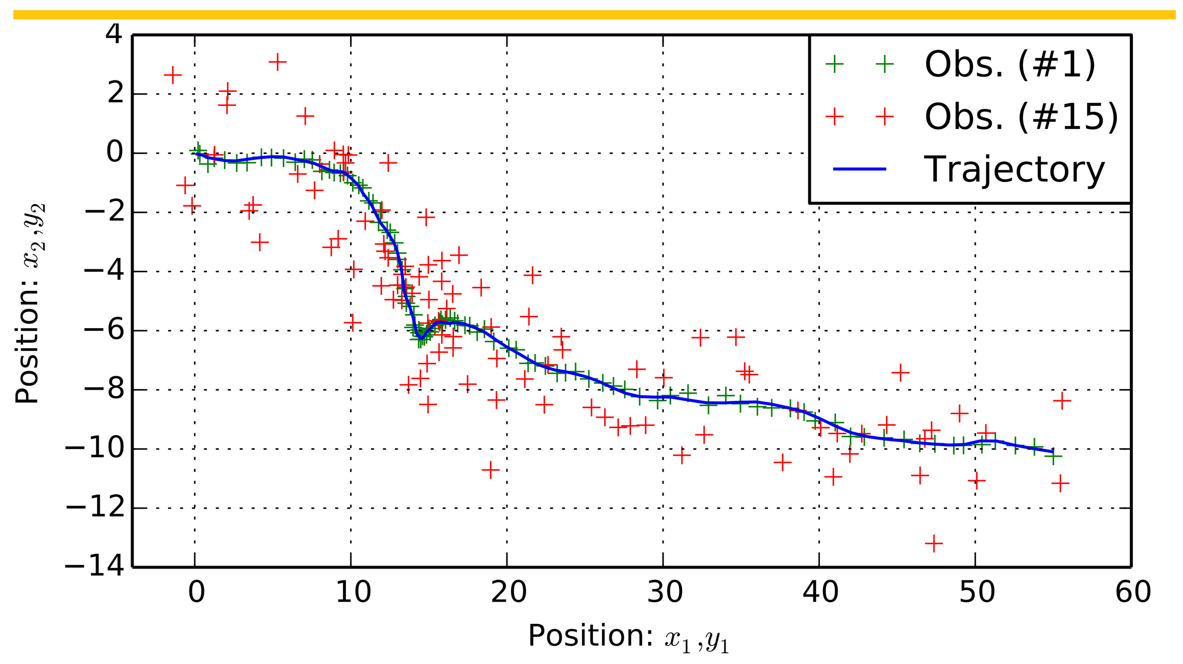
$$
x_t | x_{t-1}, u_t \sim \mathcal{N}(A_t x_{t-1}, Q_t) \quad \text{and} \quad y_t | x_t \sim \mathcal{N}(H_t x_t, R_t) \quad \text{where}
$$

$$
A = \begin{bmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad
Q = q \cdot \begin{bmatrix} \frac{dt^3}{3} & 0 & \frac{dt^2}{2} & 0 \\ 0 & \frac{dt^3}{3} & 0 & \frac{dt^2}{2} \\ \frac{dt^2}{2} & 0 & dt & 0 \\ 0 & \frac{dt^2}{2} & 0 & dt \end{bmatrix} \quad
H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad
R = r^2 \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
$$

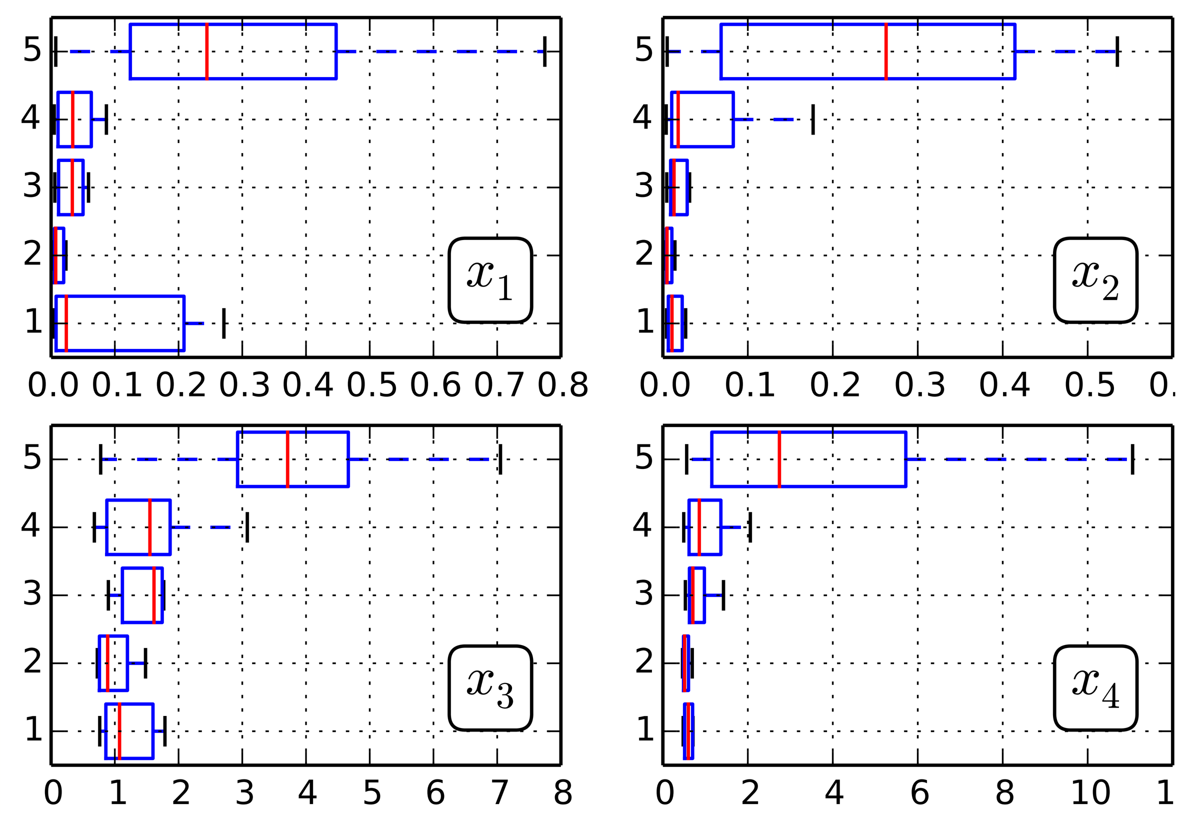
where $dt = 0.1$, $q = 5.0$, $r = 0.1n$ with $n = 1, \ldots, 15$ being the node's number. Additionally, the node 15 suffers drop-outs: at each t, it measures with probability 0.4 zeros. Initialization: $P_{i,0}^+$ with values 1000 and zero vectors $x_{i,0}^+, i = 1, \ldots, 15$; prior $c_{ij} \sim \mathrm{Beta}(10, 1)$, the 0.99999-confidence regions were computed per individual x_t elements and their marginal pdfs. $a_{ij} = |\mathcal{N}_i|^{-1}$ for simplicity.



Network topology.



True trajectory and noisy observations of nodes 1 and 15 with the least and highest observation noise, respectively.



MSD: (1) diffKF [1], (2) ATC, (3) combination-only, (4) adaptation-only, (5) no cooperation. Outliers filtered out.

Strategy	x_1	x_2	x_3	x_4
Bayesian ATC	0.0326	0.0107	1.1438	0.5836
Bayesian A	0.2336	0.0525	1.7870	1.2067
Bayesian C	0.2223	0.0296	2.1371	0.8290
No coop.	7.6446	0.6323	7.1186	5.0680
diffKF [1]	0.5016	0.0390	2.2800	0.6970

Final MSD of estimates of $x_1, \ldots, x_4$ averaged over all nodes.

Future work

Application of the method to more elaborate methods of unscented, particle and Rao-Blackwellized particle filtration would be very interesting. Also, the problem of weights deserves more attention.

This work is supported by the Czech Science Foundation, grant no. 14-06678P.

References

[1] F. S. Cattivelli and A. H. Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Trans. Autom. Control*, vol. 55, no. 9, pp. 2069–2084, 2010.

[2] J. Hu, L. Xie, and C. Zhang, "Diffusion Kalman filtering based on covariance intersection," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 891–902, Feb. 2012.

[3] K. Dedecius and V. Sečkárová, "Dynamic diffusion estimation in exponential family models," *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1114–1117, 2013.

[4] F.S. Cattivelli and A.H. Sayed, "Diffusion LMS Strategies for Distributed Estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.