DIFFUSION ESTIMATION OF STATE-SPACE MODELS: BAYESIAN FORMULATION

Kamil Dedecius

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 1143/4, 182 08 Prague, Czech Republic
email: dedecius@utia.cas.cz

ABSTRACT

The paper studies the problem of decentralized distributed estimation of the state-space models from the Bayesian viewpoint. The adopted diffusion strategy, consisting of collective adaptation to new data and combination of posterior estimates, is derived in general model-independent form. Its particular application to the celebrated Kalman filter demonstrates the ease of use, especially when the measurement model is rewritten into the exponential family form and a conjugate prior describes the estimated state.

***Index Terms*—** Distributed estimation, state-space models, Bayesian estimation, diffusion networks.

1. INTRODUCTION

In the last decade, the rapid development of wireless sensor networks has induced a considerable research effort in the field of the fully distributed (decentralized) estimation. In the state-space domain, dominated by the celebrated Kalman filter, the initial works date back to 1978 when Speyer proposed the distributed Kalman filter for a totally connected network [1]. In order to alleviate the restrictive topological requirements, Olfati-Saber proposed three types of consensus Kalman filters with the so-called microfilter architecture [2]. A similar approach is adopted in the Alricksson's weighted-averaging of the distributed KF [3]. Both Olfati-Saber's and Alriksson's algorithms combine only the state estimators, leaving the covariances intact, which may principally lead to flawed estimators. This was noticed, e.g., by Carli *at al.* [4], Schizas' *et al.* [5] and Ribeiro *et al.* [6], who involve the fusion of covariance matrices in their consensus algorithms.

The drawback of the consensus algorithms consists in the necessity of intermediate averaging iterations between two subsequent data updates. The diffusion Kalman filter by Cattivelli and Sayed [7] avoids them, saving the communication resources. While [7] explains how local state estimates can be fused by using local covariance matrices, it nevertheless suggests a simplified covariance-independent combination step to reduce the communication burden between nodes. Hu, Xie and Zhang [10] examine the covariance-based fusion method at an increased communication cost.

The present contribution adopts the diffusion approach, develops it for the general state-space models with Markov-type transitions and specializes to the case of the Kalman filter. Generally, the diffusion algorithms consist of two steps: the adaptation and combination, when the network nodes exchange measurements and estimates, respectively. Unlike most of the state-of-art approaches, the method proposed in this paper formulates both steps using the Bayesian paradigm, independent of a particular model type. If the observation model adheres to the exponential family of distributions, and the knowledge of the state is characterized by a conjugate prior distribution, then the diffusion estimator follows analytically. This is demonstrated on the particular case of the Kalman filter. If the model and the prior are not a conjugate pair and approximate methods (e.g. particle filters) are required, the method is applicable as well. Due to the intricacy of these methods, we leave the nonconjugate cases beyond the scope of the paper.

2. BAYESIAN ESTIMATION OF DISCRETE-TIME STATE-SPACE MODELS

Let us focus on the problem of dynamic estimation of the latent state x_t of the discrete-time state space models with observable output y_t given by

$$x_t = f(x_{t-1}, u_t, w_t) \tag{1}$$

$$y_t = g(x_t, v_t) \tag{2}$$

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.

Example (Linear state-space model)**.** *An example of the lin-*

This work is supported by the Czech Science Foundation, grant no. 14–06678P. The example source code is available on http://diffest.utia.cas.cz.

978-1-4799-3694-6/14/$31.00

ear state-space model is

$$
\begin{aligned}
x_t &= A_t x_{t-1} + B_t u_t + w_t \\
y_t &= H_t x_t + v_t
\end{aligned} \tag{3}
$$

where A_t, B_t and H_t are matrices of appropriate dimensions and both v_t and w_t are independent and identically distributed. Many important models may be written in this form, e.g. the stochastic volatility models, the autoregressive moving average (ARMA) models and others.

The Bayesian approach to estimation of the state-space models reformulates (1) and (2) to the forms of conditional probability distributions with probability density functions (pdfs)

$$x_t|x_{t-1}, u_t \sim \pi(x_t|x_{t-1}, u_t) \tag{4}$$

$$y_t|x_t \sim p(y_t|x_t), \tag{5}$$

respectively. Denote the past values $Y_t = \{y_1, \ldots, y_t\}$, $U_t = \{u_1, \ldots, u_t\}$ and the prior pdf of x_{t-1} as $\pi(x_{t-1}|Y_{t-1}, U_{t-1})$. The Bayesian sequential estimation runs in two steps:

(i) Prediction – using the prior pdf $\pi(x_{t-1}|Y_{t-1}, U_{t-1})$, the conditional distribution (4) and the Chapman-Kolmogorov equation we obtain

$$
\begin{aligned}
&\pi(x_t|U_t, Y_{t-1}) \\
&\quad = \int \pi(x_t|x_{t-1}, u_t)\pi(x_{t-1}|Y_{t-1}, U_{t-1}) dx_{t-1}.
\end{aligned} \tag{6}
$$

(ii) Update – also called correction, as it corrects the predicted (here again prior) estimator using the obtained observation of y_t by virtue of the Bayes' theorem

$$\pi(x_t|U_t, Y_t) = \frac{\pi(x_t|U_t, Y_{t-1})p(y_t|x_t)}{\int \pi(x_t|U_t, Y_{t-1})p(y_t|x_t)dx_t}. \tag{7}$$

The update step inherits the intrinsic computational issues associated with the Bayes' theorem: unless the involved pdfs have rather special forms, the posterior pdf is not analytically tractable and needs to be approximated, e.g. using Markov chain Monte Carlo or sequential Monte Carlo methods. The tractability is preserved if the prior distribution $\pi(x_t|U_t, Y_{t-1})$ is conjugate to the exponential family model $p(y_t|x_t)$, see Appendix, Definitions 1 and 2. Assuming $T(y_t)$ to be the sufficient statistics of $p(y_t|x_t)$ and ξ_{t-1} and ν_{t-1} to be the hyperparameters of the prior $\pi(x_t|U_t, Y_{t-1})$, the Bayesian update (7) reduces to two summations,

$$
\begin{aligned}
\xi_t &= \xi_{t-1} + T(y_t) \\
\nu_t &= \nu_{t-1} + 1.
\end{aligned} \tag{8}
$$

Later in Section 4, we focus on the celebrated Kalman filter, enjoying this analytical type of update.

3. DIFFUSION ESTIMATION

Assume a network represented by an undirected graph of N nodes (vertices). Fixing some node i, its adjacent neighbors $j \in \{1, \ldots, N\}$ form a *neighborhood* $\mathbf{N}_i$ of cardinality $|\mathbf{N}_i|$. The node i is a member of $\mathbf{N}_i$, too. In the diffusion network, the node i exchanges measurements and estimates with $j \in \mathbf{N}_i$ and incorporates them into its own statistical knowledge. The measurements exchange and incorporation is called the *adaptation* step; the exchange and fusion of estimates is called the *combination* step [7]. It is possible to employ either one or both of them. In the Bayesian framework, both steps are defined on probability distributions, yielding methods consistent from the probability theory viewpoint [9].

3.1. Adaptation

During the adaptation step the node i acquires the observations $y_{j,t}$ from its neighbors $j \in \mathbf{N}_i$. Each of them is independently assigned a weight $c_{ij} \in [0, 1]$, expressing the degree of the ith node's belief in jth node's information. In other words, it is a (subjective) probability that jth node's information is true. The observations $y_{j,t}$ are incorporated into the distribution $\pi_i(x_t|\tilde{U}_{i,t}, \tilde{Y}_{i,t-1})$ using the Bayes' theorem (7) in the form

$$\pi_i(x_t|\tilde{U}_{i,t}, \tilde{Y}_{i,t}) \propto \pi_i(x_t|\tilde{U}_{i,t}, \tilde{Y}_{i,t-1}) \prod_{j \in \mathbf{N}_i} p(y_{j,t}|x_t)^{c_{ij}}, \tag{9}$$

where tilde denotes the variables affected by the shared information. Interestingly, there are two equivalent interpretations of (9). First, the product of likelihoods can be viewed as as a *fusion of models* (possibly different due to noise heterogeneity across the network) prescribed by Proposition 1. Second, it is a sequence of $|\mathbf{N}_i|$ *weighted Bayesian updates*.

Under conjugacy, the diffusion update counterpart of (8) reads

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

3.2. Combination

The combination step proceeds with the posterior pdfs resulting from the adaptation step (9). The ith node now merges $\pi_j(x_t|\tilde{U}_{j,t}, \tilde{Y}_{j,t})$ of all $j \in \mathbf{N}_i$ exploiting the Kullback-Leibler optimal fusion prescribed by the following proposition (proved in [9]).

Proposition 1. *Given pdfs π_j with weights $a_{ij}, j \in \mathbf{N}_i$, their approximating pdf π_i^* optimal in the Kullback-Leibler sense minimizing the loss*

$$\sum_{j \in \mathbf{N}_i} a_{ij} \, \mathrm{D}\left(\pi_i^* \middle|\middle| \pi_j\right) \tag{10}$$

is given by the weighted geometric mean

$$\pi_i^*(x_t|\tilde{U}_{i,t}, \tilde{Y}_{i,t}) \propto \prod_{j \in \mathbf{N}_i} \pi_j(x_t|\tilde{U}_{j,t}, \tilde{Y}_{j,t})^{a_{ij}}, \tag{11}$$

where $a_{ij} \in [0, 1]$ *summing to unity are the weights (probabilities) of neighbors' estimates from* i*'s perspective.*

In this respect, the result is a shrinkage estimator. If the posterior pdf is conjugate and hence possesses hyperparameters $\xi_{i,t}$ and $\nu_{i,t}$, then the combination step yields

$$\begin{aligned} \xi_{i,t}^* &= \sum_{j \in \mathbf{N}_i} a_{ij} \xi_{j,t} \\ \nu_{i,t}^* &= \sum_{j \in \mathbf{N}_i} a_{ij} \nu_{j,t}. \end{aligned} \tag{12}$$

This is in accordance with the diffusion estimation of the exponential family models by Dedecius and Sečkárová [9]. Moreover, the Bayesian-update interpretation applies here as well: the i's posterior pdf $p_i(x_t|\tilde{U}_{i,t}, \tilde{Y}_{i,t})$ can be viewed as the prior knowledge, enriched by the knowledge of $\mathbf{N}_i \setminus \{i\}$.

3.3. Determination of c_{ij} and a_{ij}

The *adaptation weights* c_{ij} can be interpreted as the probabilities that the measurements from $j \in \mathbf{N}_i$ obey the model $p(y_{j,t}|x_t)$ given $\pi_i(x_t|\tilde{U}_{i,t}, \tilde{Y}_{i,t-1})$. From the Bayesian viewpoint this can be modelled by as a beta-distributed variable $c_{ij} \sim \mathrm{Beta}(r_{ij,t}, s_{ij,t})$ updated by the information whether $y_{j,t}$ is contained in a high-credibility (e.g. 0.99999) region $\mathcal{Y}_{ij,t}$ of the predictive pdf

$$p(y_{j,t}|\tilde{U}_{i,t}, \tilde{Y}_{i,t-1}) = \int p(y_{j,t}|x_t) \pi_i(x_t|\tilde{U}_{i,t}, \tilde{Y}_{i,t-1}) dx_t.$$

The Bayesian update of the distribution of c_{ij} then reads

$$\begin{aligned} r_{ij,t} &= r_{ij,t-1} + \mathbb{1}[y_{j,t} \in \mathcal{Y}_{ij,t}] \\ s_{ij,t} &= s_{ij,t-1} + (1 - \mathbb{1}[y_{j,t} \in \mathcal{Y}_{ij,t}]) \end{aligned}$$

where $\mathbb{1}$ is the indicator function and the estimate

$$\hat{c}_{ij,t} = \frac{r_{ij,t}}{r_{ij,t} + s_{ij,t}}.$$

This approach coincides with the 0-1 Bayesian test that $y_{j,t}$ follows the model.

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 [10] or [11]. Note that while a_{ij} sum to unity, c_{ij} do not need to.

4. APPLICATION TO THE KALMAN FILTER

In this section, we first review the basic Kalman filter from the Bayesian perspective, exploiting the exponential family forms of involved distributions. The diffusion algorithm then easily follows using the theory given above.

4.1. Kalman Filter

Consider the linear state-space model (3) given in Section 2 with mutually independent normal noises v_t and w_t, written in terms of normal distributions.

$$x_t|x_{t-1}, u_t \sim \mathcal{N}(A_t x_{t-1} + B_t u_t, Q_t) \tag{13}$$

$$y_t|x_t \sim \mathcal{N}(H_t x_t, R_t), \tag{14}$$

where we assume compatible dimensions. Furthermore assume starting the filtration from the initial $x_0 \sim \mathcal{N}(x_0^+, P_0^+)$. Below, we denote by the superscript '$-$' the variables after the prediction step and by '$+$' after the update step.

4.1.1. Prediction

The prediction equation (6) can be divided into two stages: first, the integrand is formed, which is a product of the normal conditional pdf of the distribution (13) and the normal pdf from the previous time step (or the initial pdf of x_0 if $t = 1$). This results in a joint multivariate normal pdf of both x_t and x_{t-1}, the latter being subsequently integrated out. By virtue of Lemma 2 (Appendix), the marginalization (integration) yields just the terms relevant to x_t. Hence the two stages can be identically viewed as a simple normality preserving linear transformation of the variable x_{t-1} giving by Lemma 1 (Appendix) a normal pdf with the parameters

$$\begin{aligned} x_t^- &= A_t x_{t-1}^+ + B_t u_t \\ P_t^- &= A_t P_{t-1}^+ A_t + Q_t. \end{aligned}$$

Obviously, this is what the "traditional" Kalman filter prediction does.

4.1.2. Update

Most Bayesian derivations of the update step involve tedious algebraic manipulations or tricks avoiding them. Below, an alternative algebraically easier approach (later straightforwardly generalized to diffusion) is given. It consists in reformulation of the problem to the update of conjugate prior's hyperparameters by sufficient statistics.[1]

Recall that $y_t|x_t \sim \mathcal{N}(H_t x_t, R_t)$. That is, the pdf $p(y_t|x_t)$ has the form

$$\begin{aligned} p(y_t|x_t) &\propto \exp\left\{ -\frac{1}{2}(y_t - H_t x_t)^\intercal R_t^{-1} (y_t - H_t x_t) \right\} \\ &= \exp\left\{ -\frac{1}{2} \mathrm{Tr}\left(\begin{bmatrix} -1 \\ x_t \end{bmatrix} \begin{bmatrix} -1 \\ x_t \end{bmatrix}^\intercal \underbrace{\begin{bmatrix} y_t^\intercal \\ H_t^\intercal \end{bmatrix} R_t^{-1} \begin{bmatrix} y_t^\intercal \\ H_t^\intercal \end{bmatrix}^\intercal}_{T(y_t)} \right) \right\}, \end{aligned} \tag{15}$$

[1] The author is not aware of any publication adopting this form (but assumes it exists).

the latter variant being the exponential family representation. The distribution $\pi(x_t|U_t, Y_{t-1})$ obtained after the prediction step is also the normal distribution. Its conjugate form according to Definition 2 (see Appendix) reads

$$p(x_t|U_t, Y_{t-1}) \propto \exp\left\{-\frac{1}{2}(x_t - x_t^-)^\intercal (P_t^-)^{-1}(x_t - x_t^-)\right\}$$

$$= \exp\left\{-\frac{1}{2}\,\mathrm{Tr}\left(\begin{bmatrix}-1\\ x_t\end{bmatrix}\begin{bmatrix}-1\\ x_t\end{bmatrix}^\intercal \underbrace{\begin{bmatrix}(x_t^-)^\intercal\\ I\end{bmatrix}(P_t^-)^{-1}\begin{bmatrix}(x_t^-)^\intercal\\ I\end{bmatrix}^\intercal}_{\xi_t}\right)\right\},$$

where I is a unit matrix of the appropriate shape.

The Bayesian update (7) then reduces to the update of the hyperparameters according to Equation (8),

$$\xi_t = \xi_{t-1} + T(y_t)$$
$$= \begin{bmatrix}(x_t^-)^\intercal (P_t^-)^{-1} x_t^- + y_t^\intercal R_t^{-1} y_t, & (x_t^-)^\intercal (P_t^-)^{-1} + y_t^\intercal R_t^{-1} H_t\\ (P_t^-)^{-1}(x_t^-)^\intercal + H_t^\intercal R_t^{-1} y_t, & (P_t^-)^{-1} + H_t^\intercal R_t^{-1} H_t\end{bmatrix} \quad (16)$$

The scalar ν_t counting the update steps is not of interest for the purpose of this paper. Now, the "classical" Kalman filter update equations are recovered with the least-squares estimator, exploiting the blocks of the matrix ξ_t:

$$P_t^+ = (\xi_{t;[2,2]})^{-1} \quad (17)$$
$$= \left[(P_t^-)^{-1} + H_t^\intercal R_t^{-1} H_t\right]^{-1} \quad (18)$$
$$= (I - K_t H_t) P_t^- \quad (19)$$
$$x_t^+ = (\xi_{t;[2,2]})^{-1}\xi_{t;[2,1]} \quad (20)$$
$$= P_t^+\left[(P_t^-)^{-1}(x_t^-)^\intercal + H_t^\intercal R_t^{-1} y_t\right]$$
$$= x_t^- + P_t^+ H_t^\intercal R_t^{-1}(y_t - H_t x_t^-) \quad (21)$$

where (19) follows from the Sherman–Morrison-Woodburry lemma (Appendix, Lemma 3) and

$$K_t = P_t^- H_t^\intercal (R_t + H_t P_t^- H_t^\intercal)$$

is the Kalman gain.

The Bayesian derivation of the extended Kalman filter with additive noise variables is obtained by essentially the same reasoning using a linearization of the state and measurement functions.

4.2. Diffusion Kalman Filter

The application of the diffusion estimation theory developed in Section 3 to the Kalman filter written in the above-given form is very straightforward. The adapt step – Equation (9) – incorporates the sufficient statistics $T(y_{j,t})$ – Equation (15) – of nodes $j \in \mathbf{N}_i$ into the ith node's prior pdf with hyperparameter $\xi_{i,t-1}$ – Equation (16) via the update (12). This simply means to evaluate

$$\xi_{i,t} = \xi_{i,t-1} + \sum_{j\in\mathbf{N}_i} c_{ij} T(y_{j,t})$$
$$= \xi_{i,t-1} + \sum_{j\in\mathbf{N}_i} c_{ij}\begin{bmatrix}y_{j,t}\\ H_{j,t}\end{bmatrix} R_{j,t}^{-1}\begin{bmatrix}y_{j,t}\\ H_{j,t}\end{bmatrix}^\intercal$$

resulting in the diffusion alternatives of (18) and (21),

$$P_{i,t}^+ = \left[(P_{i,t}^-)^{-1} + \left(\sum_{j\in\mathbf{N}_i} c_{ij} H_{j,t}^\intercal R_{j,t}^{-1} H_{j,t}\right)\right]^{-1}$$

$$x_{i,t}^+ = x_{i,t}^- + P_{i,t}^+\left[\sum_{j\in\mathbf{N}_i} c_{ij} H_{j,t}^\intercal R_{j,t}^{-1}\left(y_{j,t} - H_{j,t} x_{i,t}^-\right)\right].$$

If the combine step is used, rather than evaluating the above equations for $P_{i,t}^+$ and $x_{i,t}^+$ it is better to stick with the matrix $\xi_{i,t}$ (or elements involved in point estimators (17) and (20)) and merge according to (12),

$$\xi_{i,t}^* = \sum_{j\in\mathbf{N}_i} a_{ij}\xi_{j,t},$$

with evaluation of these estimators using (17) and (20) afterwards. This is somewhat similar to the covariance intersection approach adopted by Hu, Xie and Zhang [10].

5. NUMERICAL EXAMPLE

The numerical example demonstrates the effects of the diffusion Kalman filtering applied to a 2D tracking problem using a network of 15 nodes, one of them being faulty. The $T = 100$ simulated measurements were generated with an input-free linear state space model with constant matrices

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

$$H = \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\end{bmatrix} \qquad R = r^2\cdot\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

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 value $y_t = [0,0]^\intercal$. Figure 1 shows the obtained trajectory, along with the observations of nodes 1 and 15 with the least and highest observation noise variance, respectively. The Kalman filters are initialized with diagonal covariance matrices $P_{i,0}^+$ with values 1000 and zero vectors $x_{i,0}^+, i = 1,\ldots,15$. The prior for c_{ij} is $\mathrm{Beta}(10,1)$, the 0.99999-confidence regions were computed per individual x_t elements and their marginal

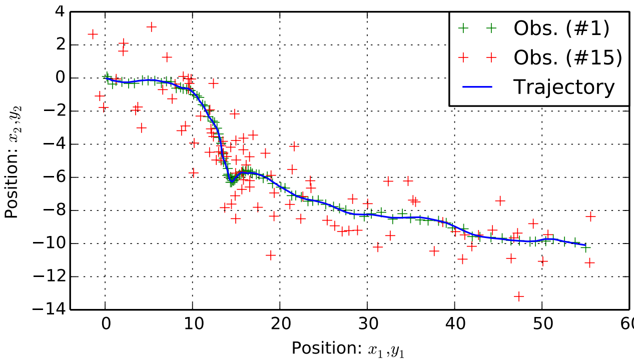

Fig. 1. True trajectory and noisy observations of nodes 1 and 15 with the least and highest observation noise, respectively.

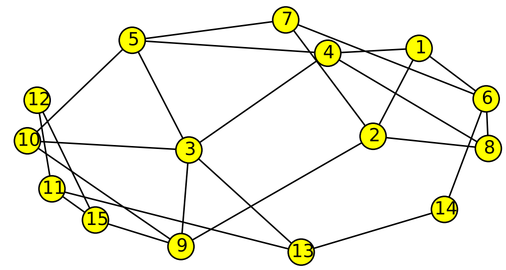

Fig. 2. Network topology.

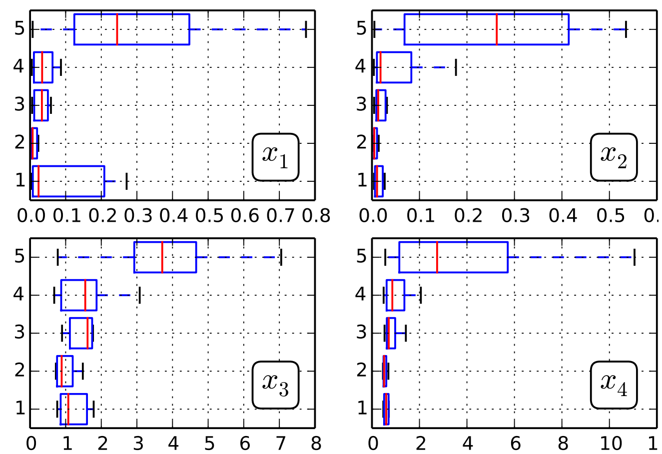

Fig. 3. Boxplots of MSDs of estimates of $x_1, \ldots, x_4$ of all considered scenarios: (1) diffKF [7], (2) ATC, (3) combination-only, (4) adaptation-only, (5) no cooperation. The red lines indicate the medians, the edges of each box are the lower and the upper quartiles. The outliers were filtered out.

pdfs. $a_{ij} = |\mathbf{N}_i|^{-1}$ for simplicity. The topology of the network depicts Figure 2.

Four scenarios are inspected: (A) – adaptation only, (C) – combination only, ATC – adapt-then-combine and finally no-cooperation mode. The original diffusion Kalman filter (diffKF) of Cattivelli and Sayed [7] is used for comparison; the appealing filter of Hu, Xie and Zhang's [10] is avoided for its higher flexibility, deserving much deeper analyses – it is a part of the future work.

The differences among these scenarios clearly depicts Figure 3, showing the boxplots of the mean squared deviations (MSD) of the estimates of $x_1, \ldots, x_4$ of all network nodes, defined (per node) as

$$MSD(x_l) = \frac{1}{T}\sum_{t=1}^{T}(\hat{x}_{l,t} - x_{l,t})^2, \qquad l \in \{1, 2, 3, 4\},$$

where $\hat{x}_{l,t}$ denotes the final estimate of x_l at time t. According to the strategy, it results from the update step (adaptation-only and no-cooperation) or from the combination step (ATC and diffKF).

All cooperation strategies lead to a significant improvement of the tracking ability, the ATC strategy dominates even the diffKF. This is obvious from the mean MSDs showing the performance of the estimation of all four parameters. The reason lies in the ability of the proposed method to reflect the drop-outs (node 15) and the more effective combination strategy. The results could be even more improved by employing variance-based weights c_{ij}, discriminating the nodes with a higher observation noise variance.

6. DISCUSSION

It is naturally interesting to compare the obtained Bayesian version of the diffusion Kalman filter with the one proposed by Cattivelli and Sayed [7]. If $c_{ij} = 1$ for all neighboring nodes, then the adaptation step of the developed Bayesian diffusion Kalman filter and the diffusion Kalman filter [7] coincide. However, the ability to reflect the degree of belief in network nodes makes the proposed solution more robust to unreliable nodes. The proposed combination step consistently combines both state estimates and their covariances. The communication requirements of the adapt step are the same as in Cattivelli and Sayed's version [7], the increased communication requirements are equivalent to Hu, Xie and Zhang's filter [10].

7. FURTHER WORK

The generality of the presented diffusion framework allows its application to more elaborate methods of unscented, particle and Rao-Blackwellized particle filtration. These topics, together with the further possibilities of dynamic determination of a_{ij} and c_{ij} allowing communication savings remain for further research. So do thorough analyses of the method. Some initial results on communication reductions are given in [12].

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 [7]	0.5016	0.0390	2.2800	0.6970

Table 1. Final MSD of estimates of $x_1, \ldots, x_4$ averaged over all nodes.

8. APPENDIX

Definition 1 (Exponential family distributions)**.** *The probability distribution of a random variable X and parameter θ is said to be an exponential family distribution with a natural parameter $\eta = \eta(\theta)$ and a sufficient statistic $T = T(X)$ if its pdf can be written in the form*

$$p(X|\theta) = \exp\left(\eta^\intercal T - A(\eta) - b(X)\right)$$

where $A(\eta)$ is the cumulant function, assuring unity of the integral of $p(X)$ and $b(X)$ is the link function independent of θ.

Definition 2 (Conjugate prior)**.** *Assume the exponential family distribution with the pdf $p(X|\theta)$ of the form given by Def. 1. The distribution of θ is said to be conjugate to it, if it can be written in the form*

$$\pi(\theta) = \exp\left(\eta^\intercal \xi - \nu A(\eta) - c(\eta)\right)$$

where again η and A are the same as in the exponential family and ξ and ν are hyperparameters (statistics) of appropriate dimensions (the latter is nonnegative real scalar).

Lemma 1 (Linear transformation of random variables)**.** *Given a random variable $X \sim \mathcal{N}(\mu, \Sigma)$, and a and b of appropriate shapes. Then, the random variable $Y = aX + b$ has the distribution $\mathcal{N}(a\mu + b, a\Sigma a^\intercal)$.*

Lemma 2 (Multivariate normal marginals)**.** *Assume random variables X, Y with a multivariate normal distribution*

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{XY}^\intercal & \Sigma_Y \end{bmatrix} \right).$$

with a pdf $p_{XY}(X, Y)$. The marginal distribution of X obtained as

$$p_X(X) = \int p_{XY}(X, Y)\, \mathrm{d}Y$$

is a (multivariate) normal distribution

$$X \sim \mathcal{N}\left(\mu_X, \Sigma_X\right).$$

Lemma 3 (Sherman–Morrison-Woodburry)**.** *Assume matrices A, B and C of appropriate dimensions. Then*

$$(A + B^\intercal C B)^{-1} = A^{-1} B^\intercal (C^{-1} + B A^{-1} B^\intercal)^{-1} B A^{-1}$$

9. REFERENCES

[1] J. Speyer, "Computation and transmission requirements for a decentralized linear-quadratic-Gaussian control problem," *IEEE Transactions on Automatic Control*, vol. 24, no. 2, pp. 266–269, Apr. 1979.

[2] R. Olfati-Saber, "Distributed Kalman filtering for sensor networks," in *2007 46th IEEE Conference on Decision and Control*. 2007, pp. 5492–5498, IEEE.

[3] P. Alriksson and A. Rantzer, "Distributed Kalman filtering using weighted averaging," in *Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems*, 2006, pp. 1–60.

[4] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed Kalman filtering based on consensus strategies," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 4, pp. 622–633, May 2008.

[5] I. D. Schizas, G. B. Giannakis, S. I. Roumeliotis, and A. Ribeiro, "Consensus in ad-hoc WSNs with noisy links. Part II: Distributed estimation and smoothing of random signals," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1650–1666, 2008.

[6] A. Ribeiro, I. Schizas, S. Roumeliotis, and G. Giannakis, "Kalman filtering in wireless sensor networks," *IEEE Control Systems Magazine*, vol. 30, no. 2, pp. 66–86, Apr. 2010.

[7] 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.

[8] Y. Zheng, R. Niu, and P. K. Varshney, "Sequential Bayesian estimation with censored data for multi-sensor systems," *IEEE Trans. Signal Process.*, vol. 62, no. 10, pp. 2626–2641, May 2014.

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

[10] 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.

[11] 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.

[12] K. Dedecius, "Collaborative Kalman filtration: Bayesian perspective," in *Proceedings of the 11th International Conference on Informatics in Control, Automation and Robotics*, 2014.