Collaborative Kalman Filtration
Bayesian Perspective

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

Keywords: Bayesian analysis, Estimation Theory, Distributed estimation, Kalman filter.

Abstract: The contribution studies the problem of collaborative Kalman filtering over distributed networks with or without a fusion center from the theoretically consistent Bayesian perspective. After presenting the Bayesian derivation of the basic Kalman filter, we develop a versatile method allowing exchange of observations among the network nodes and their local incorporation. A probabilistic nodes selection technique based on prior knowledge of nodes performance is proposed to reduce the communication requirements.

1 INTRODUCTION

The theory of distributed parameter estimation has attained tremendous attention in the last decade, particularly due to still cheaper and increasingly powerful (wireless) sensor networks. According to the network topology, three main types of networks and hence algorithms can be distinguished. First, the networks with a fusion center, responsible for information processing. In these networks, the nodes do not necessarily evaluate any modelling/estimation. Second, the networks with a Hamiltonian cycle, similar to the token ring computer networks. There exists only one path in these networks; the information circulates in the network and the nodes incorporate own information (observations) into it. Third, the diffusion networks, avoiding both the fusion center and the Hamiltonian cycle. These networks with a higher degree of connectedness employ cooperation among nodes within subsets called neighborhoods. The other two topology types can be viewed as highly degraded diffusion networks. Unlike them, the (non-degraded) diffusion networks have the highest robustness due to the avoidance of single points of failure (SPOFs). Therefore, we focus on filtering in diffusion networks, while keeping in mind that the centralized and Hamiltonian types can be solved with the proposed results as well.

Distributed Kalman filtering we focus on is closely related to the distributed recursive least squares, first proposed for the diffusion networks in (Cattivelli et al., 2008) in the classical paradigm and in (Dedecius and Sečkárová, 2013) from the Bayesian point of view. For a totally connected (hence decentralized) network, the Kalman filter was proposed by (Speyer, 1978; Ribeiro et al., 2006). However, the requirement of total connectedness is relatively prohibitive. Three types of the consensus Kalman filters, avoiding this requirement, were proposed, e.g., in the seminal paper (Olfati-Saber, 2007). These solutions rely on the so-called microfilter architecture. The consensus algorithms typically impose the need of intermediate averaging iterations among nodes, demanding additional in-network communication. The diffusion Kalman filter, (Cattivelli and Sayed, 2010), avoids them.

The main problem associated with distributed estimation is the communication burden. Several strategies for its alleviation were proposed, however, mostly for centralized networks. For example, (Gupta et al., 2006) considers the case where only one node can take measurements at a time and proposes a stochastic scheme for its selection. Another, also centralized scheme, was proposed in (Mo et al., 2006), with the goal of minimizing an objective function related to the Kalman filter error covariance matrix. The most recent *distributed* solution (Yang et al., 2014) considers minimization of the mean square estimation error.

The purpose of this paper is twofold: first, we review the formal derivation of the Kalman filter assimilating measurements obtained from the network (or its part), given in a detail in (Dedecius, 2014). This derivation follows the basic Bayesian approach

to the Kalman filter, e.g. (Meinhold and Singpurwalla, 1983) or (Peterka, 1981). Second, and more importantly, a probabilistic method for selection of a subset of network nodes is developed, that allows to significantly reduce the communication resources. This contribution focuses on the measurements incorporation only. The state-of-art methods often involve a step when the resulting estimates are merged as well. We leave the probabilistic selection of network nodes for merging as the future work.

2 KALMAN FILTER: BAYESIAN FORMULATION

Consider a system with an observable real multivariate output y_t linearly dependent on multivariate latent system state x_t and, if present, a known multivariate input (control) u_t. Suppose, that the system is described by the state-space model

$$\begin{aligned} x_t &= Ax_{t-1} + Bu_t + w_t \\ y_t &= Cx_t + e_t, \end{aligned}$$

where A, B and C are known matrices of appropriate dimensions and the mutually independent noise variables

$$\begin{aligned} w_t &\sim \mathcal{N}(0, R_w) \\ e_t &\sim \mathcal{N}(0, R_e). \end{aligned}$$

From the Bayesian viewpoint, the state variable x_t is a model parameter with the probability density function (pdf) $p(x_t|x_{t-1}, u_t)$ given by

$$\mathcal{N}(A\hat{x}_{t-1} + Bu_t, R_w), \tag{1}$$

while the output y_t obeys the model

$$y_t|x_t \sim \mathcal{N}(C\hat{x}_t, R_e) \tag{2}$$

.

Denote $\hat{x}_{t-1}$ and P_{t-1} the mean and covariance of the state given past observations $X_{t-1} = \{x_0, x_1, \ldots, x_{t-1}\}$ and $U_t = \{u_0, \ldots, u_t\}$

$$x_{t-1}|U_{t-1}, Y_{t-1} \sim \mathcal{N}(\hat{x}_{t-1}, P_{t-1}).$$

The state prediction follows from (1) and (2) using the chain rule and marginalization

$$\begin{aligned} &p(x_t|U_t, Y_{t-1}) \\ &\quad = \int p(x_t|x_{t-1}, u_t) p(x_{t-1}|U_{t-1}, Y_{t-1}) dx_{t-1}, \end{aligned} \tag{3}$$

yielding a conditional distribution for x_t

$$x_t|U_t, Y_{t-1} \sim \mathcal{N}(A\hat{x}_{t-1} + Bu_t, AP_{t-1}A^\intercal + R_w). \tag{4}$$

We adopt the convention to denote the predicted covariance by P_t^-,

$$P_t^- = AP_{t-1}A^\intercal + R_w.$$

Similarly, we denote

$$\hat{x}_t^- = A\hat{x}_{t-1} + Bu_t.$$

The Bayesian estimation of x_t given observations Y_t and U_t incorporates the latest observation y_t into the prior pdf $p(x_t|U_t, Y_{t-1})$ via the Bayes' theorem

$$p(x_t|U_t, Y_t) \propto p(y_t|x_t) p(x_t|U_t, Y_{t-1}), \tag{5}$$

where $\propto$ denotes proportionality, i.e., equality up to a normalizing constant.

The posterior pdf in (5) is proportional to a product of two normal pdfs with known variances, hence again a normal pdf. The exponent from the product reads

$$\begin{aligned} -\frac{1}{2}\big[&(y_t - C\hat{x}_t^-)^\intercal R_e^{-1}(y_t - C\hat{x}_t^-) \\ &+ (x_t - \hat{x}_t)^\intercal P_t^-(x_t - \hat{x}_t)\big] \end{aligned}$$

After completion of squares we conclude that the posterior inverse covariance P_t has the form

$$P_t^{-1} = \left(P_t^-\right)^{-1} + C^\intercal R_e^{-1} C. \tag{6}$$

The non-inverse form is obtained using the Sherman–Morrison–Woodburry lemma (Lemma 1 in Appendix),

$$P_t = (I - K_t C) P_t^- \tag{7}$$

where

$$K_t = P_t^- C^\intercal (R_e + CP_t^- C^\intercal)^{-1}$$

is the Kalman gain.

The estimator $\hat{x}_t$ follows simultaneously from the relations

$$\begin{aligned} \hat{x}_t &= \left[C^\intercal R_e^{-1} C + P_t^-\right]^{-1} \left[C^\intercal R_e^{-1} y_t + (P_t^-)^{-1} \hat{x}_t^-\right] \\ &= P_t \left[C^\intercal R_e^{-1} y_t + \left(P_t^-\right)^{-1} \hat{x}_t^-\right] \\ &= \hat{x}_t^- + P_t C^\intercal R_e^{-1} (y_t - C\hat{x}_t^-), \end{aligned} \tag{8}$$

where we use (6) to substitute for P_t^- on the second line. The obtained Kalman filter summarizes Algorithm 1. Its two phases, (i) the prediction, when the estimates of x_t and P_t are found, and (ii) the correction, incorporating latest measurements, are clearly distinguished. In the presented form, these phases coincide with equation (3) yielding (4), and (5) resulting in (7) and (8), respectively.

There are several forms of the basic Kalman filter derived above, some of which are given in (Simon, 2006).

Algorithm 1: Basic KF.

Initialization:
Set initial $\hat{x}_0$ and P_0^-.

Online mode:
(While obtaining measurements y_t)
Get measurements y_t.

Prediction

$$\hat{x}_t^- = A\hat{x}_{t-1} + Bu_{t-1}$$
$$P_t^- = AP_{t-1}A^\intercal + R_w$$

Correction

$$K_t = P_t^- C^\intercal (R_e + CP_t^- C^\intercal)^{-1}$$
$$P_t = (I - K_t C)P_t^-$$
$$\hat{x}_t = \hat{x}_t^- + P_t C^\intercal R_e^{-1}(y_t - C\hat{x}_t^-)$$

3 COLLABORATIVE KALMAN FILTERING

Consider a network represented by a connected undirected or weakly connected directed graph of N spatially distributed nodes. That is, there always exists a path from from any node to all others. In the directed case, this connectivity is weaker in the sense of the path existence under the theoretical assumption of all vertices being undirected.

Fix some node i. Its neighborhood, denoted by $\mathcal{Z}_i$, consists of all the nodes $j \in \mathcal{Z}_i$ with which i can directly communicate and exchange information; $i \in \mathcal{Z}_i$ too. We call the elements of the set $\mathcal{Z}_i$ neighbors. A special case of such setting is the diffusion network, where the nodes communicate with neighbors within 1-hop distance, Figure 1. Assume, for simplicity, that all nodes in the network employ identical state-space model for both y_t and x_t, have the same matrices A, B, C and u; the difference consists in the observation noise covariances and, potentially, in the initial setting of $\hat{x}_0$ and P_0. This can become useful in ad-hoc networks, where the nodes attach or detach during the runtime. These assumptions can be easily relaxed and serve only for notation simplification.

The collaboration consists in exchange and local incorporation of the information from the neighbors. This information includes the actual measurements $y_{j,t}$ and the observation covariance matrices $R_{e,j}$, $j \in \mathcal{Z}_i$. From the Bayesian viewpoint this means, that the whole pdfs $p_j(y_{j,t}|x_t)$ – Equation (2) – are known to ith node and a convenient alternative of the Bayes' theorem (5) remains to be defined. The following section deals with this issue.

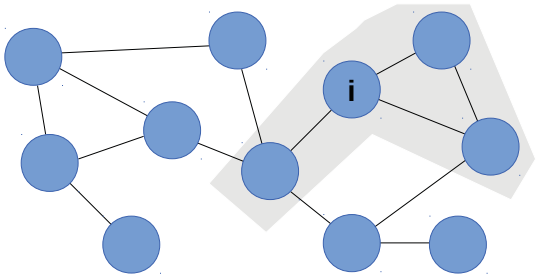

Figure 1: Example of a diffusion network: Fixed node i and its neighborhood (grey).

3.1 Bayes Theorem & Collaboration

Let us proceed with i fixed (the same rules derived below apply to all other nodes). It is quite natural to assume, that the nodes $j \in \mathcal{Z}_i$ may have different credibility, be it due to their observation noise, nodes' and connection reliabilities, occurrence of outliers etc. From the probabilistic aspect, this credibility can be expressed as a probability of the j's information being correct (true) from the ith node's viewpoint, with respect to the rest of $\mathcal{Z}_i$. That means, that i assigns the nodes $j \in \mathcal{Z}_i$ nonnegative weights $\omega_{ij} \in [0,1]$. In Example (Section 4), we will consider a total ignorance of i regarding the neighbors. Some more convenient choices of weights, for instance based on neighbors degrees, can be found, e.g., in (Cattivelli and Sayed, 2011). Quite natural is also setting the weights according to the observation noise properties (e.g., variances), similarly to the weighted least squares (Simon, 2006).

These weights may be of considerable interest if the nodes $j \in \mathcal{Z}_i$ exhibit very heterogeneous statistical properties of their measurements. This mostly means that the measurements $y_{j,t}$ are corrupted by noise terms with different variance (we leave the case of a systematic error, i.e. non-centered noise aside for this moment). There are several possible strategies to reflect this during the collaborative data assimilation process. First, the Kalman filter naturally weights the measurements by the known noise variance matrix. Second, one may employ the Bayes' theorem with the weighted likelihood

$$p_i^*(\tilde{y}_{i,t}|x_t) = \prod_{j \in \mathcal{Z}_i} p_j(y_{j,t}|x_t)^{\omega_{ij}},$$

where $\omega_{ij} \in [0,1]$ are weights. Then, the posterior of

(5) under collaboration has the form

$$p_i(x_t|\tilde{U}_t,\tilde{Y}_t) \propto p_i(x_t|\tilde{U}_t,\tilde{Y}_{t-1}) \prod_{j\in\mathcal{Z}_i} p_j(y_{j,t}|x_t)^{\omega_{ij}}, \quad (9)$$

where tilde denotes the past values of the respective variables $u_{\cdot,t}$ and $y_{\cdot,t}$.

Exponentiation of the normal pdf by $\omega_{ij} \in [0,1]$ is equivalent to its flattening, i.e. increasing the variance $R_{j,e}$ to $\omega_{ij}^{-1}R_{j,e}$. The weighted update (9) is thus equivalent to a sequence of Bayesian updates with flattened likelihoods. We can derive the one-shot update by all data relevant to time t following the same steps as in Section 2. The completion of squares and a little algebra yield

$$K_{i,t} = P_{i,t}^- C^\intercal \left[\left(\sum_{j\in\mathcal{Z}_i} \omega_{ij} R_{e,j}^{-1} \right)^{-1} + C P_{i,t}^- C^\intercal \right]^{-1}$$

$$P_{i,t} = (I - K_{i,t}C) P_{i,t}^-$$

$$\hat{x}_{i,t} = \hat{x}_{i,t}^- + P_{i,t} C^\intercal \left[\sum_{j\in\mathcal{Z}_i} \omega_{ij} R_{e,j}^{-1} \left(y_{j,t} - C\hat{x}_{i,t}^- \right) \right]$$

The prediction phase of the ordinary Kalman filter remains unaltered. The correction exploits the three equations above. Algorithm 2 summarizes the resulting collaborative Kalman filter.

3.2 Stochastic Neighbors Selection

The above-considered correction step inevitably imposes high communication requirements. Each node i needs to obtain data from all of its neighbors $j \in \mathcal{Z}_i$, regardless how well do they fit the true underlying model. Their reliability is only afterwards reflected by the weights ω_{ij}. However, with a reasonable record of past data reliability one can adopt a method for a significant reduction in communication requirements.

The basic idea is to randomly select a fixed number of neighbors from the neighborhood $\mathcal{Z}_i$. This selection is hence a random choice (sampling) without replacement and it should respect all available information about the nodes reliability (or, more precisely, the reliability of the incoming information). Let us denote the cardinality of the neighborhood $|\mathcal{Z}_i| = M$ and its elements (the nodes) by $n_{(1)},\ldots,n_{(M)}$. A particular node $n_{(j)} \in \mathcal{Z}_i$ is to be chosen with a probability

$$\pi_{(j)} = \Pr(n_{(j)}|\tilde{U}_t,\tilde{Y}_{t-1}).$$

Naturally, considering all nodes at once, $\pi_{(1)},\ldots,\pi_{(M)}$ take values from a unit M-simplex. Then, the Dirichlet distribution with the pdf

$$p(\pi_{(1)},\ldots,\pi_{(M)}|\tilde{U}_t,\tilde{Y}_{t-1}) \propto \prod_{m=1}^{M} \pi_{(m)}^{\kappa_{(m)}-1}$$

Algorithm 2: Collaborative KF.

Initialization:
forall the $i = 1,\ldots,N$ **do**
  Set initial $\hat{x}_{i,0}$ and $P_{i,0}^-$.
  Assign weights ω_{ij} to all $j \in \mathcal{Z}_i$.
  Pull covariance matrices $R_{e,j}$ from $j \in \mathcal{Z}_i$.
end

Online mode:
(While obtaining measurements $y_{j,t}$)
forall the *nodes* $i = 1,\ldots,N$ **do**

Prediction

$$\hat{x}_{i,t}^- = A\hat{x}_{i,t-1} + Bu_{t-1}$$
$$P_{i,t}^- = AP_{i,t-1}A^\intercal + R_{w,i}$$

Stochastic diffusion correction

if *Stochastic selection* **then**
  Randomly sample neighbors $n_{(j)} \in \mathcal{Z}_i$ according to their probabilities, Eq. (10).
  Get measurements $y_{(j),t}$ from these neighbors.
  Perform collaborative update:
end
else if *Update neighbors probabilities* **then**
  Update of Dirichlet hyperparameters, Eq. (13)
end

$$K_{i,t} = P_{i,t}^- C^\intercal \left[\left(\sum_{j\in\mathcal{Z}_i} \omega_{ij} R_{e,j}^{-1} \right)^{-1} + C P_{i,t}^- C^\intercal \right]^{-1}$$

$$P_{i,t} = (I - K_{i,t}C) P_{i,t}^-$$

$$\hat{x}_{i,t} = \hat{x}_{i,t}^- + P_{i,t} C^\intercal \left[\sum_{j\in\mathcal{Z}_i} \omega_{ij} R_{e,j}^{-1} \left(y_{j,t} - C\hat{x}_{i,t}^- \right) \right]$$

end

with the mean values

$$\mathbb{E}[\pi_{(j)}|\cdot] = \frac{\kappa_{(j)}}{\sum_{m=1}^{M} \kappa_{(m)}} \quad (10)$$

is a legitimate choice for their modelling. The sampling of neighbors (without replacement) from $\mathcal{Z}_i$ can then be implemented as proportional to their probabilities.

Consider the vector $[n_{(1)},\dots,n_{(M)}]$ as a multinomial random variable with parameters $\pi_{(1)},\dots,\pi_{(M)}$. This provides means for obtaining the values of the hyperparameters $\kappa_{(j)}$, as the Dirichlet distribution is conjugate to the multinomial one. Having obtained measurements $y_{(j),t}$ from all neighbors $n_{(j)}$, one may determine their current credibility based on (i) the prior knowledge about their reliability and (ii) their current predictive likelihood,

$$\Pr(y_{j,(t)}|\tilde{U}_t,\tilde{Y}_{t-1}) \propto p(\pi_{(j)}|\tilde{U}_t,\tilde{Y}_{t-1}) \times p_i(y_{(j),t}|\tilde{U}_t,\tilde{Y}_{t-1}) \quad (11)$$

normalized over all $j=1,\dots,M$, where

$$p_i(y_{j,t}|\tilde{U}_t,\tilde{Y}_{t-1}) = \int p_i(y_{j,t}|x_t)p_i(x_t|\tilde{U}_t,\tilde{Y}_{t-1})dx_t \quad (12)$$

is the predictive likelihood of $y_{(j),t}$ with respect to the ith node state. These probabilities serve as the weights in the Bayesian update similarly to (9), yielding the Dirichlet posterior hyperparameters $\kappa_{(j)}$

$$\kappa_{(j),t} = \kappa_{(j),t-1} + \Pr(y_{j,(t)}|\tilde{U}_t,\tilde{Y}_{t-1}). \quad (13)$$

The strategy for updating the Dirichlet hyperparameters depends on the user. It is reasonable to first learn from the complete neighborhood data and then update from time to time, e.g. after each $k\cdot t$ measurements where k is a positive integer.

We stress that this approach avoids the need of setting ω_{ij} (they can be considered equal to one) and directly provides means for avoidance of nodes with systematic (non-zero-centered) error term.

3.3 Properties

Let us now briefly focus on the properties of the derived collaborative Kalman filter with stochastic neighbors selection. The used notions of "worst" and "best" estimator are understood in the user-imposed terms (biasedness, consistency etc.). Thorough analysis would require definition of Bayesian estimators properties and is beyond the scope of this paper. In all cases, the properties are driven by the weights ω_{ij} and/or by the probabilities of neighbors $\pi_{(j)}$.

If the state-space models are identical in all terms, then regardless the weights the update (9) reduces to the ordinary Bayesian update (5), i.e., it is an admissible estimator. Hence the update is Bayes-optimal under such situation. Assuming the models have identical means, then the estimator is unbiased. The variance of the estimator is driven by weights ω_{ij}; the higher weight is assigned to the factor with high variance, the higher is the variance of the posterior pdf (and vice versa). In the worst case, the posterior is proportional to the highest-variance model times the prior, which is, under usual situations, still acceptable. The stochastic neighbors selection scheme provides means to updating by data from the better nodes.

4 SIMULATION EXAMPLE

Assume a network consisting of $N=10$ nodes denoted $i=0,\dots,9$. Its topology is depicted in Fig. 2. The nodes communicate with their neighbors within 1-hop distance. Three cases are considered: (i) no cooperation, (ii) cooperation with uniform weights, that is, a node i with a neighborhood of cardinality K assigns $\omega_{ij}=\frac{1}{K}$ to all $j\in\mathcal{Z}_i$, and (iii) cooperation with a stochastic selection of two neighboring nodes and weights equal to one. The Dirichlet distribution was learned from the first 15 data and then updated every fifth time step.

The state-space model is an approximate free fall model with the state equation

$$x_t = \underbrace{\begin{bmatrix}1 & \Delta t\\ 0 & 1-\frac{k}{m}\Delta t\end{bmatrix}}_{A}\underbrace{\begin{bmatrix}h_{t-1}\\ v_{t-1}\end{bmatrix}}_{x_{t-1}} + \underbrace{\begin{bmatrix}1 & 0\\ 0 & -g\Delta t\end{bmatrix}}_{B}\underbrace{\begin{bmatrix}0\\ 1\end{bmatrix}}_{u} + w_t$$

and the observation equation

$$y_t = \underbrace{\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}}_{C} x_t + e_t$$

where $g\doteq 9.8m\cdot s^{-2}$ is the gravitational constant, $m=1kg$ is the body mass of the observed object, $k=10N\cdot m^{-1}s$ is the frictional coefficient, v in $[m\cdot s^{-1}]$ is the speed and h in $[m]$ is the position. The sampling period $\Delta t=0.01s$; 100 samples are generated. The measurement noise covariances $R_{e,i}$ are diagonal matrices with elements $0.04(i+1)^2$ where $i=0,\dots,9$ indicates the number of the node. The zero-mean process noise has standard deviation 0.02 for both state variables.

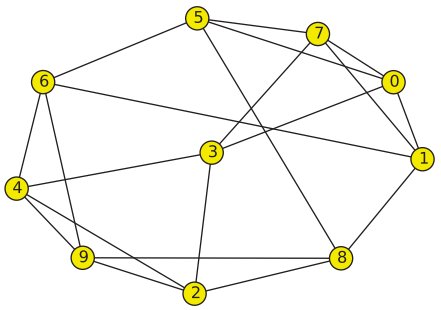

Figure 2: The example network topology.

The cumulative mean squared error (CMSE) of the estimates is computed for the whole network as the sum of individual nodes' MSEs. Without cooperation, the CMSE=[0.647, 0.252], while with cooperation over the network, the CMSE=[0.138, 0.112]; cooperation with a stochastic selection of nodes yields CMSE=[0.194, 0.128]. That indicates a significant improvement of filtration with cooperation and indicates, that the stochastic selection is able to reach results close to the case when the measurements from all neighbors are incorporated. In other words, the stochastic selection provides good performance at smaller communication and computational burden. This is evident from Figure 2: the neighborhoods are of cardinality 5, hence the stochastic selection saves 3/5 of resources under no Dirichlet update.

Figures 3, 4 and 5 depict the difference between the no-cooperation, cooperation and cooperation with stochastic selection for the nodes $i = 0$ with the least observation noise and $i = 9$ with the highest. The figures also show the nodes' MSEs of both estimates. Apparently, under cooperation, node 0 has very slightly worse MSE connected with the first state variable, and very slight or no improvement in the second variable. With respect to the MSEs' order 10^{-3}, the change in this node, best in terms of observation noise variance, is negligible. On the other side, the filtration in the "worst" node 9 is significantly better.

5 CONCLUSION

The contribution presented the Bayesian approach to cooperative Kalman filtering in distributed networks, where the node collaborate in terms of sharing their measurements. Besides the derivation of the basic cooperative Kalman filter, a scheme for stochastic selection of adjacent nodes was discussed. It provides a reasonable way towards decreasing the communication and computational burden, while retaining the ability to adapt to evolving statistical properties of neighbors measurements.

Although the theory was developed under quite restrictive assumptions (common model, invariant transition matrices), their releasing is very simple: it can be achieved by assuming time-variability and heterogeneity in the derivation of the basic Kalman filter in Section 2.

ACKNOWLEDGEMENT

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

APPENDIX

Lemma 1 (Sherman-Morrison-Woodburry)**.** *Let A, B, C and D be matrices of appropriate dimensions. Then the following equality holds:*

$$(A+BCD)^{-1} = A^{-1}B(C^{-1}+DA^{-1}B)^{-1}DA^{-1}.$$

The (well-known) proof can be found, e.g., in (Meyer, 2000).

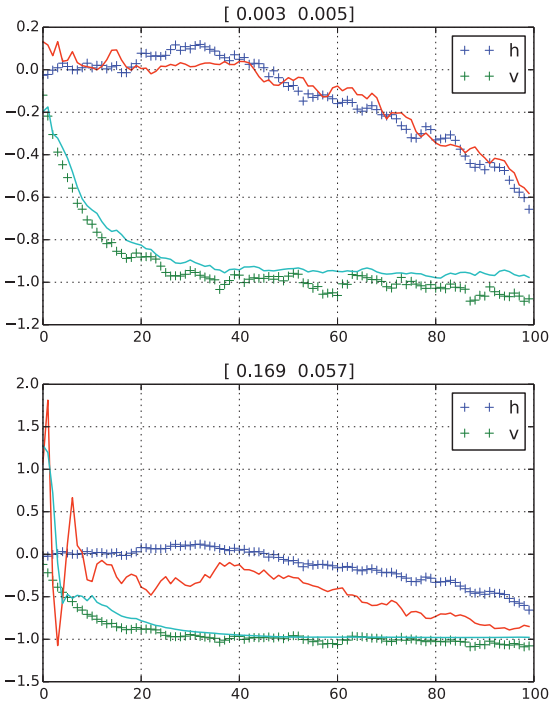

Figure 3: Filtering without cooperation: nodes $i = 0$ and $i = 9$ (with least and maximum measurement noises). The numbers indicate MSEs. Solid line depicts estimates, '+' true noisy values.

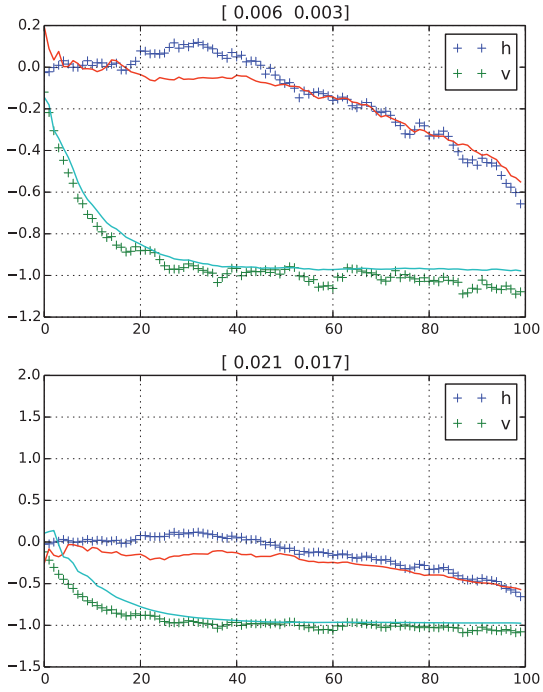

Figure 4: Filtering with cooperation: The same two nodes.

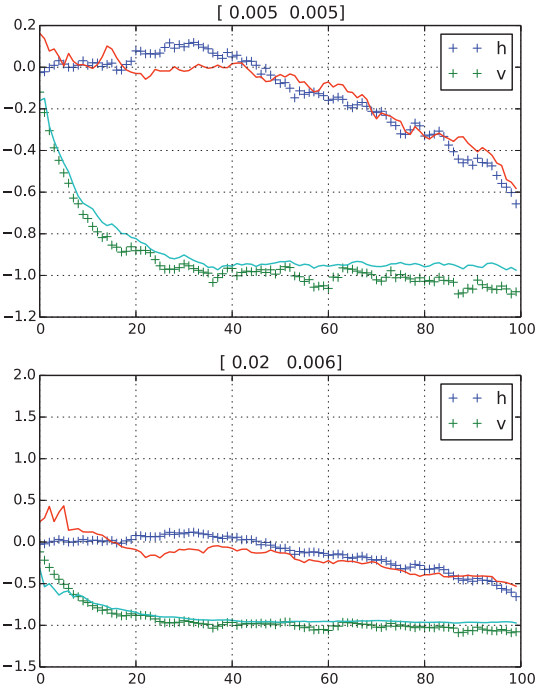

Figure 5: Filtering with cooperation and stochastic neighbors selection: The same two nodes.

REFERENCES

Cattivelli, F. S., Lopes, C. G., and Sayed, A. H. (2008). Diffusion Recursive Least-Squares for Distributed Estimation over Adaptive Networks. *IEEE Trans. Signal Processing*, 56(5):1865–1877.

Cattivelli, F. S. and Sayed, A. H. (2010). Diffusion Strategies for Distributed Kalman Filtering and Smoothing. *IEEE Trans. Automat. Contr.*, 55(9):2069–2084.

Cattivelli, F. S. and Sayed, A. H. (2011). Analysis of Spatial and Incremental LMS Processing for Distributed Estimation. *IEEE Trans. Signal Processing*, 59(4):1465–1480.

Dedecius, K. and Sečkárová, V. (2013). Dynamic Diffusion Estimation in Exponential Family Models. *IEEE Signal Process. Lett.*, 20(11):1114–1117.

Gupta, V., Chung, T. H., Hassibi, B. and Murray, R. M. (2006). On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage. *Automatica*, 42(2), 251-260.

Mo, Y., Ambrosino, R. and Sinopoli, B. (2011). Sensor selection strategies for state estimation in energy constrained wireless sensor networks. *Automatica*, 47(7), 1330-1338.

Yang, W., Chen, G., Wang, X. and Shi, L. (2014). Stochastic sensor activation for distributed state estimation over a sensor network. *Automatica [in Press]*. doi:10.1016/j.automatica.2014.05.025

Dedecius, K. (2014). Diffusion estimation of state-space models: Bayesian formulation. *Submitted to 2014 IEEE Machine Learning for Signal Processing Workshop*.

Meinhold, R. J. and Singpurwalla, N. D. (1983). Understanding the Kalman Filter. *The American Statistician*, 37(2):123–127.

Meyer, C. (2000). *Matrix analysis and applied linear algebra*. SIAM.

Olfati-Saber, R. (2007). Distributed Kalman filtering for sensor networks. In *2007 46th IEEE Conference on Decision and Control*, pages 5492–5498. IEEE.

Peterka, V. (1981). Bayesian approach to system identification. In Eykhoff, P., editor, *Trends and Progress in System Identification*, pages 239–304. Pergamon Press, Oxford, U.K.

Ribeiro, A., Giannakis, G., and Roumeliotis, S. (2006). SOI-KF: Distributed Kalman Filtering With Low-Cost Communications Using the Sign of Innovations. *IEEE Trans. Signal Processing*, 54(12):4782–4795.

Simon, D. (2006). *Optimal state estimation: Kalman, H_∞, and nonlinear approaches*. Wiley.

Speyer, J. L. (1978). Computation and Transmission Requirements for a Decentralized Linear-Quadratic-Gaussian Control Problem. *IEEE Trans. Autom. Contr.*, 24(2):266–269.