Project Report EEC264

Toshinori Kitamura

March 15, 2019

I. Introduction

This report explains about a paper "Optimal Bayesian Kalman Filtering with Prior Update"[1]. The optimal bayesian Kalman Filter(OBKF) is an advanced version of the intrinsically bayesian robust Kalman filter(IBRKF)[2], which is a Kalman filter exploiting the prior knowledge for the model.

The Kalman filter(classic KF)[3] is a widely used technique to estimate the state vector of a linear dynamics system from its previous estimation and the measurement. Although it provides the best estimation in some condition, it has a problem. That is, the classic KF is highly sensitive to the noise covariance of the target linear dynamic system[4]. To manage this problem, mainly two robust approaches, bayesian approach and non-bayesian approach, have been studied.

The adaptive Kalman filter[5][6] is a non-bayesian approach to achieve the robustness towards the uncertain system. It achieves the robustness by estimating the covariance matrixes during the state estimation. Although it doesn't require any prior knowledge, it needs a lot of observation data to tune the parameter. For a certain problem like gene regulatory network, the cost to obtain the data is expensive, so algorithms which doesn't require a lot of data are preferred.

On the other hand, because of its prior knowledge, the bayesian approach doesn't require so many data comared with non-bayesian approaches. The bayesian approach Kalman filter, IBRKF, is robust in the sense of that it minimizes the average MSE relative to the prior distribution.

The IBRKF achieves the robustness in bayesian sense, but it doesn't utilizes the any information obtained from the observation. The OBKF exploits the both prior knowledge and the observation data, and it achieves the optimal estimation relative to the posterior distribution.

The rest of this paper is organized as follows. Section II shows the overview of the classic Kalman filter and its problem. In section III, a bayesian approach, IBR Kalman filter is explained. Section IV describes the OBKF, which is the main algorithm in this paper.

II. Kalman filter

The classic KF is the optimal linear estimator for linear dynamic systems. A linear dynamic system is written as

$$\mathbf{x}_{k+1} = \boldsymbol{\Phi}_k \mathbf{x}_k + \boldsymbol{\Gamma}_k \mathbf{u}_k \tag{1}$$
$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \tag{2}$$

Where,

- $\mathbf{x}_k$: *state vector*
- $\mathbf{y}_k$: *observation vector*
- $\mathbf{u}_k$, $\mathbf{v}_k$: *zero-mean gaussian noise vector*
- $\boldsymbol{\Phi}_k$, $\boldsymbol{\Gamma}_k$, $\boldsymbol{H}_k$: *transition matrix*

And the noise statistics are,

$$\mathbb{E}[\mathbf{u}_k \mathbf{u}_l^T] = \mathbf{Q}\delta_{kl},\ \mathbb{E}[\mathbf{v}_k \mathbf{v}_l^T] = \mathbf{R}\delta_{kl},\ \forall k, l = 0, 1, \ldots \tag{3}$$
$$\mathbb{E}[\mathbf{v}_k \mathbf{x}_l^T] = \mathbf{0}, \quad \mathbb{E}[\mathbf{u}_k \mathbf{v}_l^T] = \mathbf{0}\ \forall k, l = 0, 1, \ldots \tag{4}$$
$$\mathbb{E}[\mathbf{u}_k \mathbf{y}_l^T] = \mathbf{0},\ \ 0 \le l \le k, \tag{5}$$

For such linear dynamic systems, the classic KF provides the linear estimation $\hat{\boldsymbol{x}}_k$ minimizing the mean squared error based on the observations $\boldsymbol{y}_l$, $l \le k-1$. The estimated state vector is given as,

$$\hat{\boldsymbol{x}}_k = \arg\min_{\hat{\boldsymbol{x}}_k} \mathbb{E}[|\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k|^2] \tag{6}$$

Let the mean of the estimated $\hat{\boldsymbol{x}}_k$ be itself $\hat{\boldsymbol{x}}_k$ and the covariance of it be as $\mathbf{P^x}_k$. Then, the classic KF algorithm is obtained by the Algorithm 1.

Algorithm 1 Classic Kalman Filter

Input: $\hat{\mathbf{x}}_k$, $\mathbf{P^x}_k$, $\mathbf{y}_k$
1: $\tilde{\mathbf{z}}_k = \mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_k$
2: $\mathbf{K}_k = \mathbf{P^x}_k \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P^x}_k \mathbf{H}_k^T + \mathbf{R})^{-1}$
3: $\hat{\mathbf{x}}_{k+1} = \boldsymbol{\Phi}_k \hat{\mathbf{x}}_k + \boldsymbol{\Phi}_k \mathbf{K}_k \tilde{\mathbf{z}}_k$
4: $\mathbf{P^x}_{k+1} = \boldsymbol{\Phi}_k (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P^x}_k \boldsymbol{\Phi}_k^T + \boldsymbol{\Gamma}_k \mathbf{Q} \boldsymbol{\Gamma}_k^T$
Output: $\hat{\mathbf{x}}_{k+1}$, $\mathbf{P^x}_{k+1}$

This algorithm provides the best estimation which minimizes the cost function (6) and provides the optimal estimation. However, the algorithm estimates poorly if the covariance matrix $\mathbf{Q}$ or $\mathbf{R}$ are very different from the exact value.

III. IBR Kalman Filter

The IBRKF is a robust Kalman filter exploiting the prior knowledge. To begin with, suppose the noise covariance are expressed by unknown parameter vector $\boldsymbol{\theta} = [\theta_1, \theta_2]$. It is governed by the prior distribution $\pi(\boldsymbol{\theta})$, and the noise statistics is written as

$$\mathbb{E}[\mathbf{u}_k^{\theta_1}(\mathbf{u}_l^{\theta_1})^T] = \mathbf{Q}^{\theta_1}\delta_{kl} \tag{7}$$

$$\mathbb{E}[\mathbf{v}_k^{\theta_2}(\mathbf{v}_l^{\theta_2})^T] = \mathbf{R}^{\theta_2}\delta_{kl} \tag{8}$$

The IBRKF is robust in the sense of that it minimizes the cost function (6) relative to the prior distribution. That means, the IBRKF produces the $\hat{x}_k$ as

$$\hat{x}_k = \arg\min_{\hat{x}_k} \mathbb{E}_{\boldsymbol{\theta}}[\mathbb{E}[|x_k - \hat{x}_k|^2]] \tag{9}$$

The algorithm which provides (9) is pretty similar to the one of classic KF. It's obtained by the Algorithm 2.

Algorithm 2 IBR Kalman Filter

Input: $\hat{\mathbf{x}}_k^{\boldsymbol{\theta}}$, $\mathbb{E}_{\boldsymbol{\theta}}[\mathbf{P}_k^{x,\boldsymbol{\theta}}]$, $\mathbf{y}_k^{\boldsymbol{\theta}}$

1: $\tilde{\mathbf{z}}_k^{\boldsymbol{\theta}} = \mathbf{y}_k^{\boldsymbol{\theta}} - \mathbf{H}_k\hat{\mathbf{x}}_k^{\boldsymbol{\theta}}$

2: $\mathbf{K}_k^{\Theta} = \mathbb{E}_{\boldsymbol{\theta}}[\mathbf{P}_k^{x,\boldsymbol{\theta}}]\mathbf{H}_k^T\mathbb{E}_{\boldsymbol{\theta}}^{-1}[\mathbf{H}_k\mathbf{P}_k^{x,\boldsymbol{\theta}}\mathbf{H}_k^T + \mathbf{R}^{\boldsymbol{\theta}_2}]$

3: $\hat{\mathbf{x}}_{k+1}^{\boldsymbol{\theta}} = \Phi_k\hat{\mathbf{x}}_k^{\boldsymbol{\theta}} + \Phi_k\mathbf{K}_k^{\boldsymbol{\Theta}}\tilde{\mathbf{z}}_k^{\boldsymbol{\theta}}$

4: $\mathbb{E}_{\boldsymbol{\theta}}[\mathbf{P}_{k+1}^{x,\boldsymbol{\theta}}] = \Phi_k(\mathbf{I} - \mathbf{K}_k^{\Theta}\mathbf{H}_k)\mathbb{E}_{\boldsymbol{\theta}}[\mathbf{P}_k^{x,\boldsymbol{\theta}}]\Phi_k^T + \Gamma_k\mathbb{E}_{\theta_1}[\mathbf{Q}_k^{\theta_1}]\Gamma_k^T$

Output: $\hat{\mathbf{x}}_{k+1}^{\boldsymbol{\theta}}$, $\mathbb{E}_{\boldsymbol{\theta}}[\mathbf{P}_{k+1}^{x,\boldsymbol{\theta}}]$

It's worth nothing that the IBR Kalman Filter is obtained just replacing $\mathbf{Q}$ and $\mathbf{R}$ in Classic Kalman Filter with $\mathbb{E}_{\theta_1}[\mathbf{R}^{\theta_1}]$ and $\mathbb{E}_{\theta_2}[\mathbf{Q}^{\theta_2}]$ respectively. Since the covariance matrixes are same during the algorithm, the computational cost is exactly same as the classic KF.

IV. Optimal Bayesian Kalman Filter

Although IBRKF achieves the robustness relative to its prior knowledge, it doesn't exploit the whole information provided by the observation. The OBKF utilizes both the prior and posterior knowledge to achieve a better estimation. Compared with the IBRKF, it minimizes the cost relative to the posterior distribution, i.e.,

$$\hat{x}_k = \arg\min_{\hat{x}_k} \mathbb{E}_{\boldsymbol{\theta}}[\mathbb{E}[|x_k - \hat{x}_k|^2]|\mathcal{Y}_k] \tag{10}$$

where $\mathcal{Y}_k$ is $\mathcal{Y}_k = \{y_0, y_1, ..., y_k\}$

Same as the IBRKF, the algorithm providing the $\hat{x}_k$ satisfying (10) is obtained just by replacing $\mathbf{Q}$ and $\mathbf{R}$ in classic KF with $\mathbb{E}_{\theta_1}[\mathbf{R}^{\theta_1}|\mathcal{Y}_k]$ and $\mathbb{E}_{\theta_2}[\mathbf{Q}^{\theta_2}|\mathcal{Y}_k]$ respectively, shown as Algorithm 3.

Algorithm 3 OBKF

Input: $\hat{\mathbf{x}}_k^{\theta}$, $\mathbb{E}_{\theta}[\boldsymbol{P}_k^{x,\boldsymbol{\theta}}|\mathcal{Y}_{k-1}]$, $\mathcal{Y}_k$

1: $\tilde{\mathbf{z}}_k^{\boldsymbol{\theta}} = \mathbf{y}_k^{\boldsymbol{\theta}} - \mathbf{H}_k\hat{\mathbf{x}}_k^{\boldsymbol{\theta}}$

2: $\mathbf{K}_k^{\boldsymbol{\Theta}^*} = \mathbb{E}_{\theta}[\boldsymbol{P}_k^{x,\boldsymbol{\theta}}|\mathcal{Y}_{k-1}]\mathbf{H}_k^T\mathbb{E}_{\theta}^{-1}[\mathbf{H}_k\boldsymbol{P}_k^{x,\boldsymbol{\theta}}\mathbf{H}_k^T + \grave{\mathbf{R}}^{2}|\mathcal{Y}_{k-1}]$

3: $\hat{\mathbf{x}}_{k+1}^{\theta} = \boldsymbol{\Phi}_k\hat{\mathbf{x}}_k^{\theta} + \boldsymbol{\Phi}_k\mathbf{K}_k^{\boldsymbol{\Theta}}\tilde{\mathbf{z}}_k^{\theta}$

4: $\mathbb{E}_{\theta}[\boldsymbol{P}_{k+1}^{x,\boldsymbol{\theta}}|\mathcal{Y}_k] = \boldsymbol{\Phi}_k(\mathbf{I} - \mathbf{K}_k^{\boldsymbol{\Theta}^*}\mathbf{H}_k)\mathbb{E}_{\theta}[\boldsymbol{P}_k^{x,\boldsymbol{\theta}}|\mathcal{Y}_k]\boldsymbol{\Phi}_k^T + \boldsymbol{\Gamma}_k\mathbb{E}_{\theta_1}[\boldsymbol{Q}^{\boldsymbol{\theta}_1}|\mathcal{Y}_k]\boldsymbol{\Gamma}_k^T$

Output: $\hat{\mathbf{x}}_{k+1}^{\theta}$, $\mathbb{E}_{\theta}[\boldsymbol{P}_{k+1}^{x,\boldsymbol{\theta}}|\mathcal{Y}_k]$

Now the only remaining problem is how to find the posterior expectations: $\mathbb{E}_{\theta_1}[\boldsymbol{Q}^{\boldsymbol{\theta}_1}|\mathcal{Y}_k]$ and $\mathbb{E}_{\theta_2}[\boldsymbol{R}^{\boldsymbol{\theta}_2}|\mathcal{Y}_k]$.

To calculate them, the posterior distribution $\pi(\boldsymbol{\theta}|\mathcal{Y}_k)$ is necessary. Since there is no closed-form solution for $\pi(\boldsymbol{\theta}|\mathcal{Y}_k)$, the author of this paper employs MCMC to approximate $\mathbb{E}_{\theta_1}[\boldsymbol{Q}^{\boldsymbol{\theta}_1}|\mathcal{Y}_k]$ and $\mathbb{E}_{\theta_2}[\boldsymbol{R}^{\boldsymbol{\theta}_2}|\mathcal{Y}_k]$.

The posterior distribution is proportinal to the product of its likelihood and prior distribution, i.e., $\pi(\boldsymbol{\theta}|\mathcal{Y}_k) \propto f(\mathcal{Y}_k|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$. Then, once the likelihood function $f(\mathcal{Y}_k|\boldsymbol{\theta})$ is obtained, we can run the MCMC.

From the Markov assumption of the system, it can be assumed that $\boldsymbol{y}_k$ depends only on $\boldsymbol{x}_k$ and $\boldsymbol{\theta}$. Then, the likelihood function is calculated as,

$$\begin{aligned} & f(\mathcal{Y}_k|\boldsymbol{\theta}) \\ & = \int\cdots\int f(\mathcal{Y}_k, \mathcal{X}_k|\boldsymbol{\theta})dx_0...dx_k \\ & = \int\cdots\int \prod_{i=0}^{k} f(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{\theta}) \\ & \qquad \prod_{i=1}^{k} f(\boldsymbol{x}_i|\boldsymbol{x}_{i-1}, \boldsymbol{\theta})f(\boldsymbol{x}_0)dx_0...dx_k \end{aligned} \tag{11}$$

Since this equation (11) is a factorization, it can be expressed as a factor-graph, and we can use sum-product algorithm[7] to compute $f(\mathcal{Y}_k|\boldsymbol{\theta})$. Finally, the likelihood function is obtained as the Algorithm 4.

Algorithm 4 Factor-Graph-Based Likelihood Function Calculation

Input: $\boldsymbol{\theta}$, $\mathcal{Y}_k$

$\boldsymbol{M}_0 \leftarrow \mathbb{E}[\boldsymbol{x}_0]$

$S_0 \leftarrow 1$

$\boldsymbol{\Sigma}_0 \leftarrow \text{cov}[\boldsymbol{x}_0]$

$i \leftarrow 0$

while $i \leq k-1$ **do**

$\quad \boldsymbol{W}_i \leftarrow \boldsymbol{H}_i^T(\boldsymbol{R}^{\theta_2})^{-1}\boldsymbol{y}_i + \boldsymbol{\Sigma}_i^{-1}\boldsymbol{M}_i$

$\quad \boldsymbol{\Lambda}_i^{-1} \leftarrow \boldsymbol{\Phi}_i^T(\tilde{\boldsymbol{Q}}_i^{\theta_1})^{-1}\boldsymbol{\Phi}_i + \boldsymbol{H}_i^T(\boldsymbol{R}^{\theta_2})^{-1}\boldsymbol{H}_i + \boldsymbol{\Sigma}_i^{-1}$

$\quad \boldsymbol{\Sigma}_{i+1}^{-1} \leftarrow (\tilde{\boldsymbol{Q}}_i^{\theta_1})^{-1} - (\tilde{\boldsymbol{Q}}_i^{\theta_1})^{-1}\boldsymbol{\Phi}_i\boldsymbol{\Lambda}_i\boldsymbol{\Phi}_i^T(\tilde{\boldsymbol{Q}}_i^{\theta_1})^{-1}$

$\quad \boldsymbol{M}_{i+1} \leftarrow \boldsymbol{\Sigma}_{i+1}(\tilde{\boldsymbol{Q}}_i^{\theta_1})^{-1}\boldsymbol{\Phi}_i\boldsymbol{\Lambda}_i(\boldsymbol{H}_i^T(\boldsymbol{R}^{\theta_2})^{-1}\boldsymbol{y}_i + \boldsymbol{\Sigma}_i^{-1}\boldsymbol{M}_i)$

$\quad S_{i+1} \leftarrow S_i\sqrt{\frac{|\boldsymbol{\Lambda}_i||\boldsymbol{\Sigma}_{i+1}|}{|\tilde{\boldsymbol{Q}}_i^{\theta_1}||\boldsymbol{\Sigma}_i|)}}\mathcal{N}(\boldsymbol{y}_i;\boldsymbol{0}_{m\times 1},\boldsymbol{R}^{\theta_2}) \times \exp(\frac{\boldsymbol{M}_{i+1}^T\boldsymbol{\Sigma}_{i+1}^{-1}\boldsymbol{M}_{i+1}+\boldsymbol{W}_i^T\boldsymbol{\Lambda}_i\boldsymbol{W}_i-\boldsymbol{M}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{M}_i}{2})$

$\quad i \leftarrow i+1$

end while

$\boldsymbol{\Delta}_k^{-1} \leftarrow \boldsymbol{H}_k^T(\boldsymbol{R}^{\theta_2})^{-1}\boldsymbol{H}_k + \boldsymbol{\Sigma}_k^{-1}$

$\boldsymbol{G}_k \leftarrow \boldsymbol{\Delta}_k(\boldsymbol{H}_k^T(\boldsymbol{R}^{\theta_2})^{-1}\boldsymbol{y}_k + \boldsymbol{\Sigma}_k^{-1}\boldsymbol{M}_k)$

$f(\mathcal{Y}_k|\theta) \leftarrow S_k\sqrt{\frac{|\boldsymbol{\Delta}_k|}{|\boldsymbol{\Sigma}_k|}}\mathcal{N}(\boldsymbol{y}_k;\boldsymbol{0}_{m\times 1},\boldsymbol{R}^{\theta_2}) \times \exp(\frac{\boldsymbol{G}_k^T\boldsymbol{\Delta}_k^{-1}\boldsymbol{G}_k-\boldsymbol{M}_k^T\boldsymbol{\Sigma}_k^{-1}\boldsymbol{M}_k}{2})$

Output: $f(\mathcal{Y}_k|\theta)$

V. Simulation and Performance

The author compares the OBKF with the classic Kalman filter with known parameter, IBRKF, minimax Kalman filter, and MAP Kalman filter. To evaluate the estimation, the covariance of the estimation error is used. In the equation (12), $\mathbf{P^x}_{k+1}(\boldsymbol{\theta};\boldsymbol{\theta}')$ is the covariance of the estimation error $\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k$.

$$\mathbf{P^x}_{k+1}(\boldsymbol{\theta};\boldsymbol{\theta}') = \boldsymbol{\Phi}_k(\mathbf{I}-\mathbf{K}_k^{\boldsymbol{\theta}'}\mathbf{H}_k)\mathbf{P^x}_k(\boldsymbol{\theta};\boldsymbol{\theta}')(\mathbf{I}-\mathbf{K}_k^{\boldsymbol{\theta}'}\mathbf{H}_k)^T\boldsymbol{\Phi}_k^T + \boldsymbol{\Gamma}_k\mathbf{Q}^{\theta_1}\boldsymbol{\Gamma}_k^T + \boldsymbol{\Phi}_k\boldsymbol{K}_k^{\boldsymbol{\theta}'}\boldsymbol{R}^{\theta_2}(\boldsymbol{K}_k^{\boldsymbol{\theta}'})^T\boldsymbol{\Phi}_k^T \tag{12}$$

The trace of $\mathbf{P^x}_{k+1}(\boldsymbol{\theta};\boldsymbol{\theta}')$ is computed as the MSE of the estimation, and it is used as the metric.

i. Simulation of the algorithm

To evaluate the performance of the OBKF, consider a tracking problem with an unknown parameter. To the equations (1) and (2), substitute the following parameters.

$$\boldsymbol{\Phi}_k = \begin{bmatrix} 1 & \tau & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \tau \\ 0 & 0 & 0 & 1 \end{bmatrix}, \ \boldsymbol{H}_k = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \ \boldsymbol{\Gamma}_k = \boldsymbol{I}$$

$$\boldsymbol{Q} = q \times \begin{bmatrix} \tau^3/3 & \tau^2/2 & 0 & 0 \\ \tau^2/2 & \tau & 0 & 0 \\ 0 & \tau & 0 & 0 \\ 0 & 0 & \tau^2/2 & \tau \end{bmatrix}, \ \boldsymbol{R} = r \times \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Here, τ is the measurement interval. q and r are the covariance noise intensity.

In the simulation, the author uses $\tau = 1$ second. The initial conditions are set as $\mathbb{E}[\boldsymbol{x}_0] = [100\,10\,30\;-10]^T$ and $\text{cov}[\boldsymbol{x}_0] = \text{diag}([25\,2\,25\,2])$ where $\text{diag}(v)$ is a diagonal matrix whose diagonal elements are v. The parameter q is set to 2 and r is defined as a random variable. As for the unknown parameter q and r, q is assumed as known, and r is uniformly distributed over [0.25,4].

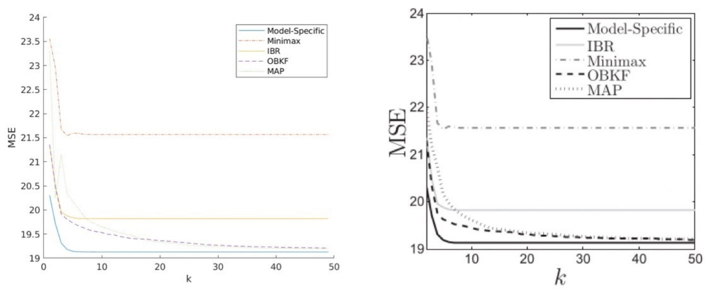

Figure 1: *caption*

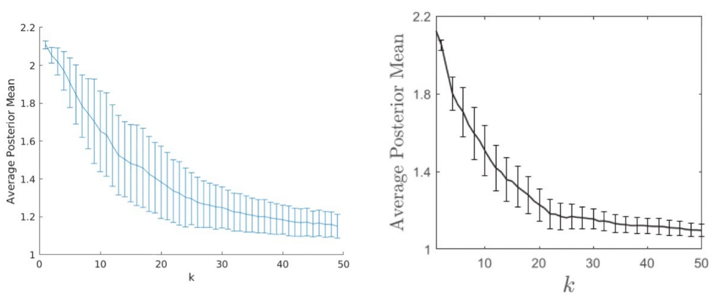

Figure 2: *caption*

Fig. compares the simulated MSE with the figure from the paper. You can see that in the both images, the OBKF outperforms the other Kalman filters.

Fig. compares the estimated r in the simulation and the figure from the paper. In the both image, the estimated r converges to the true value $r = 1$ as k increases.

VI. Problems and Future Works

VII. Conclusion

References

[1] Roozbeh Dehghannasiri, Mohammad Shahrokh Esfahani, Xiaoning Qian, and Edward R. Dougherty. Optimal Bayesian Kalman Filtering with Prior Update. *IEEE Transactions on Signal Processing*, 66(8):1982–1996, 2018.

[2] Roozbeh Dehghannasiri, Mohammad Shahrokh Esfahani, and Edward R. Dougherty. Intrinsically Bayesian robust kalman filter: An innovation process approach. *IEEE Transactions on Signal Processing*, 65(10):2531–2546, 2017.

[3] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35, 1960.

[4] S. Sangsuk-Iam and T.E. Bullock. Analysis of discrete-time Kalman filtering under incorrect noise covariances. *IEEE Transactions on Automatic Control*, 35(12):1304–1309, 1990.

[5] Kenneth A. Myers and Byron D. Tapley. Adaptive Sequential Estimation with Unknown Noise Statistics. *IEEE Transactions on Automatic Control*, 21(4):520–523, 1976.

[6] Raman K. Mehra. Approaches to Adaptive Filtering. *IEEE Transactions on Automatic Control*, 17(5):693–698, 1972.

[7] Frank R Kschischang, Brendan J Frey, and Hans-Andrea Loeliger. Factor Graphs and the Sum-Product Algorithm. Technical Report 2, 2001.