# Optimal Bayesian Kalman Filtering with Prior Update

Toshinori Kitamura

University of California Davis

*tkitamura@ucdavis.edu*

March 10, 2019

# Overview

# What's this Paper?

- Introduce a new Kalman Filter: Optimal Bayesian Kalman Filter(OBKF)[Dehghannasiri et al., 2018]
- OBKF is an advanced Kalman Filter of IBR Kalman Filter[Dehghannasiri et al., 2017]

# Kalman Filter: Basic idea

- Outputs the optimal estimation of the state of a dynamic system
- Exploits both the prediction based on the model and the measurement
- Robot Localization

Works for linear dynamic systems:

## Linear Dynamic System Example

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

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

The red parameters are necessary for the Kalman Filter

---

**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:** $\hat{\mathbf{x}}_{k+1}$, $\mathbf{P^x}_{k+1}$

---

- $\hat{\mathbf{x}}_k$: *estimated mean*
- $\mathbf{P^x}_k$: *estimated covariance*
- $\mathbf{R}$: *Noise covariance of* $\mathbf{u}_k$
- $\mathbf{Q}$: *Noise covariance of* $\mathbf{v}_k$

# Problems of Kalman Filter

- The performance is sensitive to the accuracy of the noise covariance matrixes: $\mathbf{Q}$ and $\mathbf{R}$ [Sangsuk-Iam and Bullock, 1990]

## Uncertainty of the noise covariances

Assume the covariances are parameterized by $\theta = [\theta_1, \theta_2]$

- The covariance matrix of **u** is $\mathbf{Q}^{\theta_1}$ (e.g. $\mathbf{Q}^{\theta_1} = \theta_1 \mathbf{I}$)
- The covariance matrix of **v** is $\mathbf{R}^{\theta_2}$ (e.g. $\mathbf{R}^{\theta_2} = \theta_2 \mathbf{I}$)
- Unknown prameter: $\theta = [\theta_1, \theta_2]$

If the $\theta$ used in the algorithm is very different from the true $\theta$, Kalman Filter provides a poor estimation

# Solutions of the Problem

- Non-Bayesian Approach: Adaptive Kalman Filter
  - Doesn't require any prior knowledge ($\theta$ is not a R.V.)
  - Requires a lot measured data

# Solutions of the Problem

- Non-Bayesian Approach: Adaptive Kalman Filter
  - Doesn't require any prior knowledge ($\theta$ is not a R.V.)
  - Requires a lot measured data
- Bayesian Approach: IBR Kalman Filter
  - Require prior knowledge ($\theta$ is a R.V.)
  - Robust: Guarantees the best average performance relative to the prior distribution

# Solutions of the Problem

- Non-Bayesian Approach: Adaptive Kalman Filter
  - Doesn't require any prior knowledge ($\theta$ is not a R.V.)
  - Requires a lot measured data
- Bayesian Approach: IBR Kalman Filter
  - Require prior knowledge ($\theta$ is a R.V.)
  - Robust: Guarantees the best average performance relative to the prior distribution
- Bayesian Approach: Optimal Bayesian Kalman Filter
  - Requires prior knowledge ($\theta$ is a R.V.)
  - Utilizes measured data to estimate unknown parameters
  - Optimal over the posterior distribution obtained from measured data

# Basic Idea

- Assume the prior distribution $\pi(\theta)$ for the uncertainty of the covariance matrices
- Robust estimation relative to $\pi(\theta)$: $\underset{\hat{\mathbf{x}}_\theta(k)}{\arg\min}\, E_\theta[E[|\mathbf{x}_\theta(k) - \hat{\mathbf{x}}_\theta(k)|^2]]$

- IBR Kalman Filter is obtained just replacing $\mathbf{Q}$ and $\mathbf{R}$ in Classic Kalman Filter with $E_{\theta_1}[\mathbf{R}^{\theta_1}]$ and $E_{\theta_2}[\mathbf{Q}^{\theta_2}]$ respectively

---

**Algorithm 2** IBR Kalman Filter

---

**Input:** $\hat{\mathbf{x}}_k^\theta$, $E_\theta[\mathbf{P}^{\mathbf{x},\theta}{}_k]$, $\mathbf{y}_k^\theta$

1: $\tilde{\mathbf{z}}_k^\theta = \mathbf{y}_k^\theta - \mathbf{H}_k \hat{\mathbf{x}}_k^\theta$
2: $\mathbf{K}_k^\Theta = E_\theta[\mathbf{P}^{\mathbf{x},\theta}{}_k] \mathbf{H}_k^T E_\theta^{-1}[\mathbf{H}_k \mathbf{P}^{\mathbf{x},\theta}{}_k \mathbf{H}_k^T + \mathbf{R}^{\theta_2}]$
3: $\hat{\mathbf{x}}_{k+1}^\theta = \mathbf{\Phi}_k \hat{\mathbf{x}}_k^\theta + \mathbf{\Phi}_k \mathbf{K}_k^\Theta \tilde{\mathbf{z}}_k^\theta$
4: $E_\theta[\mathbf{P}^{\mathbf{x},\theta}{}_{k+1}] = \mathbf{\Phi}_k(\mathbf{I} - \mathbf{K}_k^\Theta \mathbf{H}_k)E_\theta[\mathbf{P}^{\mathbf{x},\theta}{}_k]\mathbf{\Phi}_k^T + \mathbf{\Gamma}_k E_{\theta_1}[\mathbf{Q}^{\theta_1}{}_k]\mathbf{\Gamma}_k^T$

**Output:** $\hat{\mathbf{x}}_{k+1}^\theta$, $E_\theta[\mathbf{P}^{\mathbf{x},\theta}{}_{k+1}]$

---

# Basic Idea: OBKF

- Based on the IBR Kalman Filter
- Utilize measured data $\mathcal{Y}_k = \{\mathbf{y}_0, ..., \mathbf{y}_k\}$ to obtain the posterior distribution $\pi(\theta|\mathcal{Y}_k)$
- Best estimation relative to $\pi(\theta|\mathcal{Y}_{k-1})$:
$$\arg\min_{\hat{\mathbf{x}}_\theta(k)} E_\theta[E[|\mathbf{x}_\theta(k) - \hat{\mathbf{x}}_\theta(k)|^2]|\mathcal{Y}_{k-1}]$$

- OBKF is obtained just replacing $\mathbf{Q}$ and $\mathbf{R}$ in Classic Kalman Filter with $E_{\theta_1}[\mathbf{Q}^{\theta_1}|\mathcal{Y}_k]$ and $E_{\theta_2}[\mathbf{R}^{\theta_2}|\mathcal{Y}_k]$ respectively

---

**Algorithm 3** OBKF

---

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

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

2: $\mathbf{K}_k^{\mathbf{\Theta}^*} = E_\theta[\mathbf{P}^{\mathbf{x},\theta}{}_k|\mathcal{Y}_{k-1}]\mathbf{H}_k^T E_\theta^{-1}[\mathbf{H}_k\mathbf{P}^{\mathbf{x},\theta}{}_k\mathbf{H}_k^T + \mathbf{R}^{\theta_2}|\mathcal{Y}_{k-1}]$

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

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

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

---

- OBKF is obtained just replacing $\mathbf{Q}$ and $\mathbf{R}$ in Classic Kalman Filter with $E_{\theta_1}[\mathbf{Q}^{\theta_1}|\mathcal{Y}_k]$ and $E_{\theta_2}[\mathbf{R}^{\theta_2}|\mathcal{Y}_k]$ respectively

---

**Algorithm 4** OBKF

---

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

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

2: $\mathbf{K}_k^{\Theta^*} = E_\theta[\mathbf{P}^{\mathbf{x},\theta}{}_k|\mathcal{Y}_{k-1}]\mathbf{H}_k^T E_\theta^{-1}[\mathbf{H}_k\mathbf{P}^{\mathbf{x},\theta}{}_k\mathbf{H}_k^T + \mathbf{R}^{\theta_2}|\mathcal{Y}_{k-1}]$

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

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

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

---

How are the posterior expectations found: $E_{\theta_1}[\mathbf{Q}^{\theta_1}|\mathcal{Y}_k]$ and $E_{\theta_2}[\mathbf{R}^{\theta_2}|\mathcal{Y}_k]$ ?

- Approximate $E_{\theta_1}[\mathbf{Q}^{\theta_1}|\mathcal{Y}_k]$ and $E_{\theta_2}[\mathbf{R}^{\theta_2}|\mathcal{Y}_k]$ using Metropolis Hastings MCMC
- MCMC requires the likelihood function $f(\mathcal{Y}_k|\theta)$
- $f(\mathcal{Y}_k|\theta)$ is calculated by sum-product algorithm[Kschischang et al., 2001]

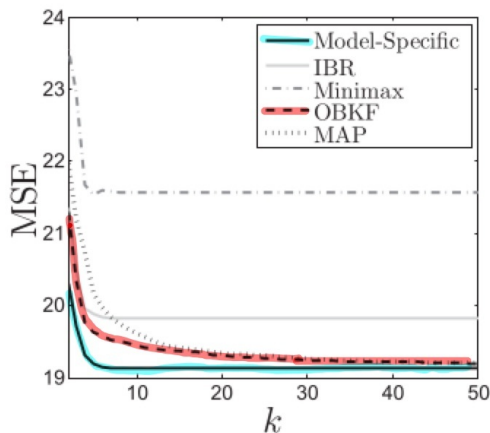**Algorithm 5** Factor-Graph-Based Likelihood Function Calculation

**Input:** $\theta$, $\mathcal{Y}_k$
1: $\mathbf{M}_0 \leftarrow E[\mathbf{x}_0]$, $S_0 \leftarrow 1$, $\boldsymbol{\Sigma}_0 \leftarrow cov[\mathbf{x}_0]$, $i \leftarrow 0$
2: **while** $i \leq k - 1$ **do**
3: $\quad \mathbf{W}_i \leftarrow \mathbf{H}_i^T (\mathbf{R}^{\theta_2})^{-1} \mathbf{y}_i + \boldsymbol{\Sigma}_i^{-1} \mathbf{M}_i$
4: $\quad \boldsymbol{\Lambda}_i^{-1} \leftarrow \boldsymbol{\Phi}_i^T (\tilde{\mathbf{Q}}_i^{\theta_1})^{-1} \boldsymbol{\Phi}_i + \mathbf{H}_i^T (\mathbf{R}^{\theta_2})^{-1} \mathbf{H}_i + \boldsymbol{\Sigma}_i^{-1}$
5: $\quad \boldsymbol{\Sigma}_{i+1}^{-1} \leftarrow (\tilde{\mathbf{Q}}_i^{\theta_1})^{-1} - (\tilde{\mathbf{Q}}_i^{\theta_1})^{-1} \boldsymbol{\Phi}_i \boldsymbol{\Lambda}_i \boldsymbol{\Phi}_i^T (\tilde{\mathbf{Q}}_i^{\theta_1})^{-1}$
6: $\quad \mathbf{M}_{i+1} \leftarrow \boldsymbol{\Sigma}_{i+1} (\tilde{\mathbf{Q}}_i^{\theta_1})^{-1} \boldsymbol{\Phi}_i \boldsymbol{\Lambda}_i (\mathbf{H}_i^T (\mathbf{R}^{\theta_2})^{-1} \mathbf{y}_i + \boldsymbol{\Sigma}_i^{-1} \mathbf{M}_i)$
7: $\quad S_{i+1} \leftarrow$ using Eq.(29) in the paper
8: $\quad i \leftarrow i + 1$
9: **end while**
10: $\boldsymbol{\Delta}_k^{-1} \leftarrow \mathbf{H}_k^T (\mathbf{R}^{\theta_2})^{-1} \mathbf{H}_k + \boldsymbol{\Sigma}_k^{-1}$
11: $\mathbf{G}_k \leftarrow \boldsymbol{\Delta}_k (\mathbf{H}_k^T (\mathbf{R}^{\theta_2})^{-1} \mathbf{y}_k + \boldsymbol{\Sigma}_k^{-1} \mathbf{M}_k)$
12: $f(\mathcal{Y}_k | \theta) \leftarrow$ using Eq.(34) in the paper
**Output:** $f(\mathcal{Y}_k | \theta)$

# Performance: Accuracy



Figure: Performance analysis for specific $\theta_1 = 1$ and known $\theta_2$
OBKF achieves the lowest MSE (cited from [Dehghannasiri et al., 2018])
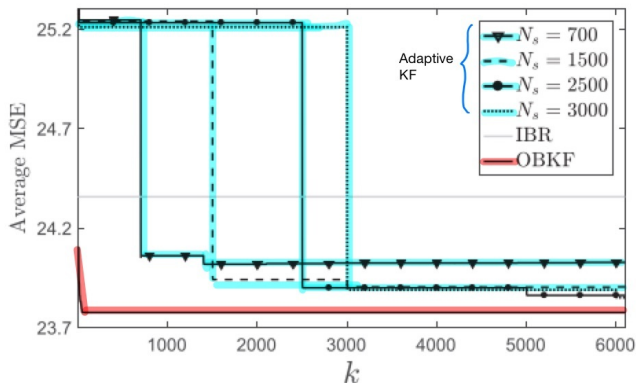
Figure: Unknown $\theta_1$ and comparison with an adaptive Kalman Filter (cited from [Dehghannasiri et al., 2018])

# Problems and Future Works

- If the prior distribution doesn't include the true value, the estimation won't converge
- Factor-graph and MCMC are computationally expensive. Finding an efficient approach is a future work.

# Summary

- OBKF converges to the optimal estimation as long as the prior distribution includes the true value
- OBKF requires much less data compared to Non-Bayesian methods
- Computational cost and deciding the prior distribution are problems

📄 Analysis of Discrete-Time Kalman Filterinng Under Incorrect Noise Covariances.

📄 Dalton, L. A. and Dougherty, E. R. (2014).
Intrinsically optimal bayesian robust filtering.
*IEEE Transactions on Signal Processing*, 62(3):657–670.

📄 Dehghannasiri, R., Esfahani, M. S., and Dougherty, E. R. (2017).
Intrinsically Bayesian robust kalman filter: An innovation process approach.
*IEEE Transactions on Signal Processing*, 65(10):2531–2546.

📄 Dehghannasiri, R., Esfahani, M. S., Qian, X., and Dougherty, E. R. (2018).
Optimal Bayesian Kalman Filtering with Prior Update.
*IEEE Transactions on Signal Processing*, 66(8):1982–1996.

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

📄 Kschischang, F. R., Frey, B. J., and Loeliger, H.-A. (2001).
Factor Graphs and the Sum-Product Algorithm.
Technical Report 2.

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