Comparison of Various Approaches for Joint Wiener/Kalman Filtering and Parameter Estimation with Application to BASS

Siouar Bensaid and Dirk Slock
Mobile Communications Department
EURECOM, Sophia Antipolis, France
Email: {bensaid, slock}@eurecom.fr

***Abstract*—In recent years, the Kalman filter (KF) has encountered renewed interest, due to an increasing range of applications. Even though in many cases the state-space model may be linear, it is often only known up to the values of some parameters, usually related to the vector autoregressive process of the state evolution equation. In this paper, after finding motivation in some applications, we review a number of approaches for adaptive Kalman filtering (AKF), in which state and parameters get estimated jointly. We propose an improved version of the Extended KF (EKF) in which the estimation error covariance matrix is computed exactly assuming overall joint Gaussianity. We also compare the performance and Cramer Rao bounds (CRBs) of joint Maximum A Posteriori Maximum Likelihood (MAP-ML) estimation of Bayesian state and deterministic parameters, and marginalized ML estimation of the parameters, and relate this to the Expectation-Maximization KF (EM-KF). The perspectives involve also the Variational Bayesian KF (VB-KF).**

I. Introduction

Since Rudolf E. Kalman published his famous paper in 1960, the Kalman filter has become the work horse of many estimation processes in different application areas. The Kalman filter (KF) considers the estimation of a first-order vector autoregressive (AR(1)) (Markov) process from linear measurements in white noise. The KF performs this estimation recursively by alternating between filtering (measurement update) and (one step ahead) prediction (time update). An alternative viewpoint is that the Kalman filter recursively generates the innovations of the measurement signal (by a structured Gram-Schmidt approach that decorrelates the consecutive measurements). The KF corresponds to optimal (Minimum Mean Squared Error (MMSE) or Maximum A Posteriori (MAP)) Bayesian estimation of the state sequence if all random sources involved (measurement noise, state noise and state initial conditions) are Gaussian. In the non-Gaussian case, the KF performs Linear MMSE (LMMSE) estimation. The linear state-space model can be written as

$$\begin{cases} \text{state update equation:} \\ \qquad \boldsymbol{x}_{k+1} \;=\; \boldsymbol{F}_k(\theta)\,\boldsymbol{x}_k + \boldsymbol{G}_k(\theta)\,\boldsymbol{w}_k \\ \text{measurement equation:} \\ \qquad \boldsymbol{y}_k \;=\; \boldsymbol{H}_k(\theta)\,\boldsymbol{x}_k + \boldsymbol{v}_k \end{cases} \tag{1}$$

for discrete time $k = 1, 2, \ldots$, where the initial state $\boldsymbol{x}_0 \sim \mathcal{N}(\widehat{\boldsymbol{x}}_0, \boldsymbol{P}_0(\theta))$, the measurement noise $\boldsymbol{v}_k \sim \mathcal{N}(0, \boldsymbol{R}_k(\theta))$, and the state noise $\boldsymbol{w}_k \sim \mathcal{N}(0, \boldsymbol{Q}_k(\theta))$ and all these random quantities are mutually uncorrelated. As indicated, the state-space model is often specified up to the value of some parameters θ. Often the $\boldsymbol{F}_k(\theta)$, $\boldsymbol{G}_k(\theta)$, $\boldsymbol{H}_k(\theta)$ are linear in θ, which would correspond to the bilinear case. Although this signal model seems very simple, the applications are numerous. We will cite here three examples:

- Bayesian adaptive filtering [18] (or wireless channel estimation [9], [12]):
  in this case, $\boldsymbol{x}_k$ = FIR filter response, and θ contains e.g. the Power Delay Profile (diagonal of a diagonal filter coefficient covariance matrix $\boldsymbol{P}_0 = \boldsymbol{P}_k$, and the AR(1) dynamics in e.g. diagonal $\boldsymbol{F}$ and $\boldsymbol{Q}$.
- Position tracking (GPS) (see [5] and references therein):
  $$\mathrm{x}_{t+1} = \begin{bmatrix} 1 & \Delta t & \frac{1}{2}\Delta t^2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} \cdot \mathrm{x}_t = \begin{bmatrix} x_t + \Delta t \cdot v_t + \frac{1}{2}\Delta t^2 a \\ v_t + \Delta t \cdot a \\ a \end{bmatrix}$$
  the state contains position, velocity and possible acceleration and θ contains acceleration model parameters (e.g. white noise, AR(1))
- Blind Audio Source Separation (BASS) [4]: x_k = source signals, θ: (short+long term) AR parameters, reverb filters

In the literature, variations on the KF theme have been derived to handle the joint filtering and parameter estimation problem, such as e.g. the widely used EM-KF algorithm ([6], [8], [9]) which uses the famous Expectation Maximization technique (EM), and alternating optimization technique for ML estimation. Another well-known variation is the EKF algorithm, which can handle general nonlinear state space models. In this case, the state is extended with the unknown parameters, rendering the new state update equation nonlinear. A third derivation is the truncated Second-Order EKF (SOEKF) introduced by [3], [11] in which nonlinearities are expanded up to second order, third and higher order statistics being neglected. A corrected derivation of this filter is presented in [10]. In ([2], [11]), the Gaussian SOEKF is derived in which fourth-order terms in the Taylor series expansions are retained and approximated by assuming that the underlying joint probability distribution is Gaussian. In [21], Villares at al. introduced the Quadratic Extended Kalman Filter (QEKF) where they extend the EKF to a new algorithm using quadratic processing and incorporating fourth order statistics of the input signal. The problem of uncertainty about the process noise

EURECOMs research is partially supported by its industrial members: ORANGE, BMW Group, Swisscom, Cisco, SFR, ST Ericsson, Thales, Symantec, SAP, Monaco Telecom, and also by the EU FP7 project WHERE2.

and measurement noise covariance matrices was also tackled in [16] where a test of Kalman filter optimality is used in order to estimate the unknown noise covariance matrices. The performance of some of these Adaptive KF (AKF) approaches was studied in the literature. In [7], the EM approach is proved to converge to the ML performance. The asymptotic behavior of the EKF for AKF has been treated in [13] where it is proved that no global convergence is garanteed. The performance analysis of linear and nonlinear KF has also been treated in terms of Cramer Rao Bound (CRB) computations. In [19], the Posterior CRB (PCRB) is developed for the discrete nonlinear KF. Recursive Bayesian CRBs were also developed for continuous and discrete nonlinear filtering for many problems. We can refer to [20] for an overview. This paper is organized as follows: a review of some AKF approaches is proposed in section II. An improved version of the EKF algorithm is developed in section III and some performance orderings are provided in section IV.

II. Adaptive Kalman Filtering Approaches

A. Basic Kalman Filter (KF)

In the following, we introduce the notation $\boldsymbol{y}_{1:k} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k\}$. The KF performs Gram-Schmidt orthogonalization (decorrelation) of the measurement variables $\boldsymbol{y}_k$. This is done by computing the LMMSE predictor $\widehat{\boldsymbol{y}}_{k|k-1}$ of $\boldsymbol{y}_k$ on the basis of $\boldsymbol{y}_{1:k-1}$, leading to the orthogonalized prediction error (or innovation) $\widetilde{\boldsymbol{y}}_k = \widetilde{\boldsymbol{y}}_{k|k-1} = \boldsymbol{y}_k - \widehat{\boldsymbol{y}}_{k|k-1}$. We introduce the correlation matrix notation $R_{\boldsymbol{x}\boldsymbol{y}} = \mathrm{E}\,\boldsymbol{x}\,\boldsymbol{y}^T$ (correlation matrices will usually also be covariance matrices here since the processes $\boldsymbol{y}_k$ and $\boldsymbol{x}_k$ have zero mean and also various estimation errors will have (conditional) zero mean). We denote the covariance matrix $R_{\widetilde{\boldsymbol{y}}_k \widetilde{\boldsymbol{y}}_k} = \boldsymbol{S}_k$. The idea of the innovations approach is that (linear) estimation in terms of $\boldsymbol{y}_{1:k}$ is equivalent to estimation in terms of $\widetilde{\boldsymbol{y}}_{1:k}$ since one set is obtained from the other by an invertible linear transformation. Now, since the $\widetilde{\boldsymbol{y}}_k$ are decorrelated, estimation in terms of $\widetilde{\boldsymbol{y}}_{1:k}$ simplifies:

$$\widehat{\boldsymbol{x}}_{|k} = \sum_{i=1}^{k} R_{\boldsymbol{x}\widetilde{\boldsymbol{y}}_i} R^{-1}_{\widetilde{\boldsymbol{y}}_i \widetilde{\boldsymbol{y}}_i} \widetilde{\boldsymbol{y}}_i = \boldsymbol{x}_{|k-1} + R_{\boldsymbol{x}\widetilde{\boldsymbol{y}}_k} \boldsymbol{S}_k^{-1} \widetilde{\boldsymbol{y}}_k \ .$$

This will be used to obtain *predicted* estimates $\widehat{\boldsymbol{x}}_{k|k-1}$ with estimation error $\widetilde{\boldsymbol{x}}_{k|k-1} = \boldsymbol{x}_k - \widehat{\boldsymbol{x}}_{k|k-1}$ with covariance matrix $\boldsymbol{P}_{k|k-1} = R_{\widetilde{\boldsymbol{x}}_{k|k-1} \widetilde{\boldsymbol{x}}_{k|k-1}}$ and also *filtered* estimates $\widehat{\boldsymbol{x}}_{k|k}$ with estimation error $\widetilde{\boldsymbol{x}}_{k|k} = \boldsymbol{x}_k - \widehat{\boldsymbol{x}}_{k|k}$ with covariance matrix $\boldsymbol{P}_{k|k} = R_{\widetilde{\boldsymbol{x}}_{k|k} \widetilde{\boldsymbol{x}}_{k|k}}$.

Now exploiting the correlation structure in the signal model, this leads to the following two-step recursive procedure to go from $|k-1$ to $|k$:

Measurement Update

$$\begin{array}{rcl} \widehat{\boldsymbol{y}}_{k|k-1} &=& \boldsymbol{H}_k\, \widehat{\boldsymbol{x}}_{k|k-1} \\ \widetilde{\boldsymbol{y}}_k &=& \boldsymbol{y}_k - \widehat{\boldsymbol{y}}_{k|k-1} \\ \boldsymbol{S}_k &=& \boldsymbol{H}_k\, \boldsymbol{P}_{k|k-1}\, \boldsymbol{H}_k^T + \boldsymbol{R}_k \\ \boldsymbol{K}_k &=& \boldsymbol{P}_{k|k-1}\, \boldsymbol{H}_k^T\, \boldsymbol{S}_k^{-1} \\ \widehat{\boldsymbol{x}}_{k|k} &=& \widehat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_k\, \widetilde{\boldsymbol{y}}_k \\ \boldsymbol{P}_{k|k} &=& \boldsymbol{P}_{k|k-1} - \boldsymbol{K}_k\, \boldsymbol{H}_k\, \boldsymbol{P}_{k|k-1} \end{array} \tag{2}$$

Time Update (prediction)

$$\begin{array}{rcl} \widehat{\boldsymbol{x}}_{k+1|k} &=& \boldsymbol{F}_k\, \widehat{\boldsymbol{x}}_{k|k} \\ \boldsymbol{P}_{k+1|k} &=& \boldsymbol{F}_k\, \boldsymbol{P}_{k|k}\, \boldsymbol{F}_k^T + \boldsymbol{G}_k\, \boldsymbol{Q}_k\, \boldsymbol{G}_k^T \end{array} \tag{3}$$

In the usual case of total absence of prior information on the initial state, one can choose $\widehat{\boldsymbol{x}}_0 = 0$, $\boldsymbol{P}_0 = p_0\, I$ with p_0 a (very) large number.

B. Extended Kalman Filter (EKF)

For the case of a nonlinear state-space model, the idea of the EKF is to apply the KF to a linearized version of the state-space model, via a first-order Taylor series expansion. So we get

state update equation:

$$\boldsymbol{x}_{k+1} \;=\; f(\boldsymbol{x}_k, \boldsymbol{w}_k) \;\approx\; \boldsymbol{F}_k\, \boldsymbol{x}_k + \boldsymbol{G}_k\, \boldsymbol{w}_k$$

measurement equation:

$$\boldsymbol{y}_k \;=\; h(\boldsymbol{x}_k) + \boldsymbol{v}_k \;\approx\; \boldsymbol{H}_k\, \boldsymbol{x}_k + \boldsymbol{v}_k \tag{4}$$

where

$$\boldsymbol{F}_k = \left.\frac{\partial\, f(\boldsymbol{x}, \boldsymbol{w})}{\partial\, \boldsymbol{x}^T}\right|_{(\boldsymbol{x},\boldsymbol{w})=(\boldsymbol{x}_k,\boldsymbol{w}_k)} \quad \boldsymbol{G}_k = \left.\frac{\partial\, f(\boldsymbol{x}, \boldsymbol{w})}{\partial\, \boldsymbol{w}^T}\right|_{(\boldsymbol{x},\boldsymbol{w})=(\boldsymbol{x}_k,\boldsymbol{w}_k)}$$
$$\boldsymbol{H}_k = \left.\frac{\partial\, h(\boldsymbol{x})}{\partial\, \boldsymbol{x}^T}\right|_{\boldsymbol{x}=\boldsymbol{x}_k} . \tag{5}$$

So, at this point, the basic KF can be applied to the thus obtained approximate linear state-space model. The EKF approach can be used to adapt some parameters in an otherwise linear state-space model $\boldsymbol{x}_{k+1}^{'} = \boldsymbol{F}^{'}\, \boldsymbol{x}_k^{'} + \boldsymbol{G}^{'}\, \boldsymbol{w}_k$. For instance, consider the case in which one wants to adapt parameters appearing (e.g.) linearly in the matrix $\boldsymbol{F}^{'} = \boldsymbol{F}^{'}(\boldsymbol{\theta})$. One can jointly estimate the unknown constant parameter vector $\boldsymbol{\theta}$ by considering the following state update for them: $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k$. Then one can introduce the augmented state and system matrices

$$\boldsymbol{x}_k = \begin{bmatrix} \boldsymbol{x}_k^{'} \\ \boldsymbol{\theta}_k \end{bmatrix}\ ,\ \boldsymbol{F}_k = \begin{bmatrix} \boldsymbol{F}^{'}(\boldsymbol{\theta}_k) & \boldsymbol{C}(\boldsymbol{x}_k^{'}) \\ 0 & I \end{bmatrix}\ ,\ \boldsymbol{G}_k = \begin{bmatrix} \boldsymbol{G}^{'} \\ 0 \end{bmatrix} \tag{6}$$

where $\boldsymbol{C}(\boldsymbol{x}_k^{'}) = \dfrac{\partial\, \boldsymbol{F}^{'}(\boldsymbol{\theta})\boldsymbol{x}_k^{'}}{\partial\, \boldsymbol{\theta}^T}$. When running the EKF, the state-dependent system matrices have to be filled with the latest state estimates, so in this case

$$\boldsymbol{F}_k = \begin{bmatrix} \boldsymbol{F}^{'}(\widehat{\boldsymbol{\theta}}_{k|k}) & \boldsymbol{C}(\widehat{\boldsymbol{x}}_{k|k}^{'}) \\ 0 & I \end{bmatrix} . \tag{7}$$

The parameters $\boldsymbol{\theta}$ are often not really constant and hence need to be tracked adaptively. This can be done either by introducing some process noise in $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k$ (ramdom walk time evolution) or by introducing exponential weighting (at least for the $\boldsymbol{\theta}$ portion) into the KF updates [1]. The EKF approach allows fairly straightforwardly to estimate parameters in $\boldsymbol{F}_k$, $\boldsymbol{H}_k$, or $\boldsymbol{G}_k$, but much less so in $\boldsymbol{Q}_k$, $\boldsymbol{R}_k$.

For adapting (parameters in) $\boldsymbol{Q}$ and $\boldsymbol{R}$, one needs to consider the innovations representation $\widehat{\boldsymbol{x}}_{k+1|k} = \boldsymbol{F}_k\, \widehat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{F}_k \boldsymbol{K}_k\, \widetilde{\boldsymbol{y}}_k$ and consider gradients of the Kalman gain $\boldsymbol{K}_k$ w.r.t. these matrices.

C. Recursive Prediction Error Method (RPEM-KF)

The RPEM [15], [14] is an adaptive implementation of Maximum Likelihood (ML) parameter estimation. The negative loglikelihood becomes a least-squares criterion in the prediction errors (innovations) and RPEM performs one iteration per new sample. Applied to KF, the RPEM can be seen as a more rigorous version of EKF and computes gradients more precisely [23]. Indeed, for the case of a state transition matrix $\boldsymbol{F}_k = \boldsymbol{F}_k(\boldsymbol{\theta})$, the EKF would consider the gradient

$$\frac{\partial \boldsymbol{x}_{k+1}}{\partial \boldsymbol{\theta}^T} = \frac{\partial \boldsymbol{F}_k(\boldsymbol{\theta})\boldsymbol{x}_k}{\partial \boldsymbol{\theta}^T} \tag{8}$$

where only the explicit dependence of $\boldsymbol{F}$ on $\boldsymbol{\theta}$ would be considered, whereas the RPEM would consider more correctly

$$\frac{\partial \boldsymbol{x}_{k+1}}{\partial \boldsymbol{\theta}^T} = \frac{\partial \boldsymbol{F}_k(\boldsymbol{\theta})\boldsymbol{x}_k}{\partial \boldsymbol{\theta}^T} + \boldsymbol{F}_k(\boldsymbol{\theta})\,\frac{\partial \boldsymbol{x}_k}{\partial \boldsymbol{\theta}^T}\ . \tag{9}$$

RPEM for KF can be found in the references above, but will not be pursued here further. One characteristic of the RPEM is a higher complexity.

D. Expectation-Maximization (EM-KF)

In EM [7], the parameters are estimated by minimizing expected values of negative loglikelihoods, see e.g. [22] for an application involving KF. For the state update, since $\boldsymbol{G}_k$ is typically a tall matrix, $\boldsymbol{G}_k\,\boldsymbol{w}_k$ has a singular covariance matrix. The state update equation can be rewritten as

$$\boldsymbol{G}_k^+\,\boldsymbol{x}_{k+1} = \boldsymbol{G}_k^+\,\boldsymbol{F}_k\,\boldsymbol{x}_k + \boldsymbol{w}_k \tag{10}$$

where $\boldsymbol{G}_k^+ = (\boldsymbol{G}_k^T\boldsymbol{G}_k)^{-1}\boldsymbol{G}_k^T$ is the pseudo-inverse of $\boldsymbol{G}_k$. For the parameters involved in the state update equation, hence the following negative loglikelihood is applicable:

$$\sum_k \{\ln\det(\boldsymbol{Q}_k) + (\boldsymbol{x}_{k+1} - \boldsymbol{F}_k\,\boldsymbol{x}_k)^T\boldsymbol{G}_k^{+T}\boldsymbol{Q}_k^{-1}\boldsymbol{G}_k^+(\boldsymbol{x}_{k+1} - \boldsymbol{F}_k\,\boldsymbol{x}_k)\} \tag{11}$$

For the parameters involved in the measurement equation, the appropriate loglikelihood is

$$\sum_k \{\ln\det(\boldsymbol{R}_k) + (\boldsymbol{y}_k - \boldsymbol{H}_k\,\boldsymbol{x}_k)^T\boldsymbol{R}_k^{-1}(\boldsymbol{y}_k - \boldsymbol{H}_k\,\boldsymbol{x}_k)\}\ . \tag{12}$$

Now the expectation is taken, in principle with the conditional distribution given all data. Hence $\mathrm{E}_{|n}$ involving all data $\boldsymbol{y}_k$ up to the last sample n. This leads to an iterative algorithm with in each iteration a whole fixed-interval smoothing operation. An adaptive version [22], [9] can be obtained by replacing fixed-interval smoothing by fixed-lag smoothing and performing one iteration per time sample. Since the state update equation corresponds to a vector AR(1) model, one may expect (as in [9]) that a lag of 1 should be enough (to guarantee convergence). In [22], complexity is reduced further by suggesting that filtering might be enough. In that case, the (presumably) slowly varying $\widehat{\boldsymbol{Q}}_{k+1}$, $\widehat{\boldsymbol{F}}_{k+1}$ (for use in the KF at time $k{+}1$) get determined by minimizing $\sum_{i=1}^k \lambda^{k-i}\ \mathrm{E}_{|i}\{\text{Terms in (11)}\}$ w.r.t. $\boldsymbol{Q}$, $\boldsymbol{F}$ ($\boldsymbol{G}$ is known) where we introduced an exponential forgetting factor $\lambda \lessapprox 1$. This is equivalent to

$$\gamma_k^{-1}\ln\det(\widehat{\boldsymbol{Q}}) + \sum_{i=1}^k \lambda^{k-i}\ \mathrm{tr}\{\boldsymbol{G}_i^{+T}\widehat{\boldsymbol{Q}}^{-1}\boldsymbol{G}_i^+\ \mathrm{E}_{|i}(\boldsymbol{x}_{i+1} - \widehat{\boldsymbol{F}}\,\boldsymbol{x}_i)(\boldsymbol{x}_{i+1} - \widehat{\boldsymbol{F}}\,\boldsymbol{x}_i)^T\} \tag{13}$$

where we introduced $\gamma_k^{-1} = \sum_{i=1}^k \lambda^{k-i} = \lambda\,\gamma_{k-1}^{-1} + 1$. γ_k^{-1} behaves initially as $1/k$ but saturates eventually at $\gamma_\infty^{-1} = 1-\lambda$. We shall need

$$\begin{aligned}
\mathrm{E}_{|i}\,\boldsymbol{x}_i\boldsymbol{x}_i^T &= \widehat{\boldsymbol{x}}_{i|i}\widehat{\boldsymbol{x}}_{i|i}^T + \boldsymbol{P}_{i|i}\\
\mathrm{E}_{|i}\,\boldsymbol{x}_{i+1}\boldsymbol{x}_i^T &= \boldsymbol{F}_i\,\widehat{\boldsymbol{x}}_{i|i}\widehat{\boldsymbol{x}}_{i|i}^T + \boldsymbol{F}_i\,\boldsymbol{P}_{i|i}\\
\mathrm{E}_{|i}\,\boldsymbol{x}_i\boldsymbol{x}_{i+1}^T &= \widehat{\boldsymbol{x}}_{i|i}\widehat{\boldsymbol{x}}_{i|i}^T\boldsymbol{F}_i^T + \boldsymbol{P}_{i|i}\,\boldsymbol{F}_i^T\\
\mathrm{E}_{|i}\,\boldsymbol{x}_{i+1}\boldsymbol{x}_{i+1}^T &= \boldsymbol{F}_i\,\widehat{\boldsymbol{x}}_{i|i}\widehat{\boldsymbol{x}}_{i|i}^T\boldsymbol{F}_i^T + \boldsymbol{P}_{i+1|i}\\
&= \boldsymbol{F}_i\,(\widehat{\boldsymbol{x}}_{i|i}\widehat{\boldsymbol{x}}_{i|i}^T + \boldsymbol{P}_{i|i})\,\boldsymbol{F}_i^T + \boldsymbol{G}_i\boldsymbol{Q}_i\boldsymbol{G}_i^T
\end{aligned} \tag{14}$$

In case of time-invariant $\boldsymbol{G}_k \equiv \boldsymbol{G}$, we can rewrite (13) as

$$\begin{aligned}
&\ln\det(\widehat{\boldsymbol{Q}}) + \mathrm{tr}\{\boldsymbol{G}^{+T}\widehat{\boldsymbol{Q}}^{-1}\boldsymbol{G}^+(\boldsymbol{M}_k^{11} - \widehat{\boldsymbol{F}}\,\boldsymbol{M}_k^{01} - \boldsymbol{M}_k^{10}\,\widehat{\boldsymbol{F}}^T\\
&+ \widehat{\boldsymbol{F}}\,\boldsymbol{M}_k^{00}\,\widehat{\boldsymbol{F}}^T)\}\quad\text{where}\\
&\boldsymbol{M}_k^{00} = (1-\gamma_k)\,\boldsymbol{M}_{k-1}^{00} + \gamma_k\,(\widehat{\boldsymbol{x}}_{k|k}\widehat{\boldsymbol{x}}_{k|k}^T + \boldsymbol{P}_{k|k})\\
&\boldsymbol{M}_k^{10} = (1-\gamma_k)\,\boldsymbol{M}_{k-1}^{10} + \gamma_k\,\boldsymbol{F}_k\,(\widehat{\boldsymbol{x}}_{k|k}\widehat{\boldsymbol{x}}_{k|k}^T + \boldsymbol{P}_{k|k})\\
&\boldsymbol{M}_k^{01} = (1-\gamma_k)\,\boldsymbol{M}_{k-1}^{01} + \gamma_k\,(\widehat{\boldsymbol{x}}_{k|k}\widehat{\boldsymbol{x}}_{k|k}^T + \boldsymbol{P}_{k|k})\,\boldsymbol{F}_k^T\\
&\boldsymbol{M}_k^{11} = (1-\gamma_k)\,\boldsymbol{M}_{k-1}^{11} + \gamma_k(\widehat{\boldsymbol{x}}_{k+1|k}\widehat{\boldsymbol{x}}_{k+1|k}^T + \boldsymbol{P}_{k+1|k})
\end{aligned} \tag{15}$$

In case of furthermore time-invariant $\boldsymbol{F}_k \equiv \boldsymbol{F}$, $\boldsymbol{Q}_k \equiv \boldsymbol{Q}$, then

$$\begin{aligned}
\boldsymbol{M}_k^{10} &= \boldsymbol{F}\,\boldsymbol{M}_k^{00}\\
\boldsymbol{M}_k^{01} &= \boldsymbol{M}_k^{00}\,\boldsymbol{F}^T\\
\boldsymbol{M}_k^{11} &= \boldsymbol{F}\,\boldsymbol{M}_k^{00}\,\boldsymbol{F}^T + \boldsymbol{G}\,\boldsymbol{G}\,\boldsymbol{G}^T
\end{aligned} \tag{16}$$

As a result, (15) can be rewritten as $\ln\det(\widehat{\boldsymbol{Q}}) + \mathrm{tr}\{\widehat{\boldsymbol{Q}}^{-1}\boldsymbol{Q}\} + \mathrm{tr}\{\boldsymbol{G}^{+T}\widehat{\boldsymbol{Q}}^{-1}\boldsymbol{G}_i^+(\boldsymbol{F} - \widehat{\boldsymbol{F}})\,\boldsymbol{M}_k^{00}\,(\boldsymbol{F} - \widehat{\boldsymbol{F}})^T)\}$, the optimization of which now clearly leads to $\widehat{\boldsymbol{F}} = \boldsymbol{F}$, $\widehat{\boldsymbol{Q}} = \boldsymbol{Q}$. So we just get back the quantities that we use in the KF, without any additional information. Hence, just Kalman filtering in the EM-KF is not enough to adapt the state update parameters.

E. Fixed-Lag Smoothing

Using the innovations approach, we have

$$\widehat{\boldsymbol{x}}_{k-1|k} = \widehat{\boldsymbol{x}}_{k-1|k-1} + R_{\boldsymbol{x}_{k-1}\widetilde{\boldsymbol{y}}_k}\,\boldsymbol{S}_k^{-1}\,\widetilde{\boldsymbol{y}}_k\ . \tag{17}$$

After a few steps, we get the following lag-1 smoothing equations that need to be added to the basic Kalman Filter equations (to be inserted between the Measurement Update and the Time Update)

$$\begin{aligned}
\boldsymbol{K}_{k;1} &= \boldsymbol{P}_{k-1|k-1}\,\boldsymbol{F}_{k-1}^T\,\boldsymbol{H}_k^T\\
\widehat{\boldsymbol{x}}_{k-1|k} &= \widehat{\boldsymbol{x}}_{k-1|k-1} + \boldsymbol{K}_{k;1}\,\boldsymbol{S}_k^{-1}\,\widetilde{\boldsymbol{y}}_k\\
\boldsymbol{P}_{k-1|k} &= \boldsymbol{P}_{k-1|k-1} - \boldsymbol{K}_{k;1}\,\boldsymbol{S}_k^{-1}\,\boldsymbol{K}_{k;1}^T\ .
\end{aligned} \tag{18}$$

F. Adaptive EM-KF with Fixed-Lag Smoothing

Consider now the case in which the state-space model is essentially time-invariant (or slowly time-varying). In that case the time index of the system matrices $\boldsymbol{F}_k$ etc. just reflects at

which time the (unknown) system matrices have been adapted. The resulting KF equations with lag-1 smoothing become

$$\begin{array}{rcl}
\widehat{\boldsymbol{y}}_{k|k-1} &=& \boldsymbol{H}_{k-1}\,\widehat{\boldsymbol{x}}_{k|k-1} \\
\widetilde{\boldsymbol{y}}_{k} &=& \boldsymbol{y}_{k} - \widehat{\boldsymbol{y}}_{k|k-1} \\
\boldsymbol{S}_{k} &=& \boldsymbol{H}_{k-1}\,\boldsymbol{P}_{k|k-1}\,\boldsymbol{H}_{k-1}^{T} + \boldsymbol{R}_{k-1} \\
\boldsymbol{K}_{k;1} &=& \boldsymbol{P}_{k-1|k-1}\,\boldsymbol{F}_{k-1}^{T}\,\boldsymbol{H}_{k-1}^{T} \\
\widehat{\boldsymbol{x}}_{k-1|k} &=& \widehat{\boldsymbol{x}}_{k-1|k-1} + \boldsymbol{K}_{k;1}\,\boldsymbol{S}_{k}^{-1}\,\widetilde{\boldsymbol{y}}_{k} \\
\boldsymbol{P}_{k-1|k} &=& \boldsymbol{P}_{k-1|k-1} - \boldsymbol{K}_{k;1}\,\boldsymbol{S}_{k}^{-1}\,\boldsymbol{K}_{k;1}^{T} \\
\boldsymbol{K}_{k} &=& \boldsymbol{P}_{k|k-1}\,\boldsymbol{H}_{k-1}^{T}\,\boldsymbol{S}_{k}^{-1} \\
\widehat{\boldsymbol{x}}_{k|k} &=& \widehat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_{k}\,\widetilde{\boldsymbol{y}}_{k} \\
\boldsymbol{P}_{k|k} &=& \boldsymbol{P}_{k|k-1} - \boldsymbol{K}_{k}\,\boldsymbol{H}_{k-1}\,\boldsymbol{P}_{k|k-1} \\
& & \text{parameter update} \\
\widehat{\boldsymbol{x}}_{k+1|k} &=& \boldsymbol{F}_{k}\,\widehat{\boldsymbol{x}}_{k|k} \\
\boldsymbol{P}_{k+1|k} &=& \boldsymbol{F}_{k}\,\boldsymbol{P}_{k|k}\,\boldsymbol{F}_{k}^{T} + \boldsymbol{G}_{k}\,\boldsymbol{Q}_{k}\,\boldsymbol{G}_{k}^{T}
\end{array} \tag{19}$$

So, the system matrices ($\boldsymbol{F}$, $\boldsymbol{G}$, $\boldsymbol{Q}$) should be adapted after the smoothing step and before the filtering and prediction steps. We now adapt the system matrices $\boldsymbol{F}$, $\boldsymbol{G}$, $\boldsymbol{Q}$ from the equivalent of (13) with $\mathrm{E}_{|i}$ replaced by $\mathrm{E}_{|i+1}$. This leads to the matrix updates

$$\begin{aligned}
\boldsymbol{M}_{k}^{00} &= (1-\gamma_k)\,\boldsymbol{M}_{k-1}^{00} + \gamma_k\,(\widehat{\boldsymbol{x}}_{k-1|k}\widehat{\boldsymbol{x}}_{k-1|k}^{T} + \boldsymbol{P}_{k-1|k}) \\
\boldsymbol{M}_{k}^{10} &= (\boldsymbol{M}_{k}^{01})^{T} = (1-\gamma_k)\,\boldsymbol{M}_{k-1}^{10} + \gamma_k\,(\widehat{\boldsymbol{x}}_{k|k}\widehat{\boldsymbol{x}}_{k-1|k}^{T} \\
&\quad + \boldsymbol{F}_{k-1}\boldsymbol{P}_{k-1|k} - \boldsymbol{G}_{k-1}\boldsymbol{Q}_{k-1}\boldsymbol{G}_{k-1}^{T}\boldsymbol{H}_{k-1}^{T}\boldsymbol{S}_{k}^{-1}\boldsymbol{K}_{k,1}^{T}) \\
\boldsymbol{M}_{k}^{11} &= (1-\gamma_k)\,\boldsymbol{M}_{k-1}^{11} + \gamma_k(\widehat{\boldsymbol{x}}_{k|k}\widehat{\boldsymbol{x}}_{k|k}^{T} + \boldsymbol{P}_{k|k})\;.
\end{aligned}$$

If for example $\boldsymbol{G}$ would be fixed and invertible, minimization of the expected loglikelihood w.r.t. $\widehat{\boldsymbol{F}}$, $\widehat{\boldsymbol{Q}}$ would lead to the following minimizers

$$\begin{array}{rcl}
\boldsymbol{F}_{k} &=& \boldsymbol{M}_{k}^{10}\,(\boldsymbol{M}_{k}^{00})^{-1} \\
\boldsymbol{Q}_{k} &=& \boldsymbol{G}^{+}(\boldsymbol{M}_{k}^{11} - \boldsymbol{M}_{k}^{10}\,(\boldsymbol{M}_{k}^{00})^{-1}\boldsymbol{M}_{k}^{01})\boldsymbol{G}^{+\,T}\;.
\end{array} \tag{20}$$

For adapting the parameters in the measurement equation on the other hand, Kalman filtering should be sufficient. A similar derivation from the expected measurement logliklihood leads to

$$\begin{aligned}
&\widehat{\boldsymbol{H}}_{k} = \widehat{R}_{\boldsymbol{yx},k}\,\widehat{R}_{\boldsymbol{xx},k}^{-1} \\
&\widehat{\boldsymbol{R}}_{k} = \widehat{R}_{\boldsymbol{yy},k} - \widehat{R}_{\boldsymbol{yx},k}\,\widehat{R}_{\boldsymbol{xx},k}^{-1}\,\widehat{R}_{\boldsymbol{xy},k} \quad \text{where}
\end{aligned} \tag{21}$$

$$\begin{array}{rcl}
\widehat{R}_{\boldsymbol{yy},k} &=& (1-\gamma_k)\widehat{R}_{\boldsymbol{yy},k-1} + \gamma_k \boldsymbol{y}_k\boldsymbol{y}_k^{T} \\
\widehat{R}_{\boldsymbol{xy},k} &=& (1-\gamma_k)\widehat{R}_{\boldsymbol{xy},k-1} + \gamma_k \widehat{\boldsymbol{x}}_{k|k}\boldsymbol{y}_k^{T} \\
\widehat{R}_{\boldsymbol{yx},k} &=& \widehat{R}_{\boldsymbol{xy},k}^{T} \\
\widehat{R}_{\boldsymbol{xx},k} &=& (1-\gamma_k)\widehat{R}_{\boldsymbol{xx},k-1} + \gamma_k(\widehat{\boldsymbol{x}}_{k|k}\widehat{\boldsymbol{x}}_{k|k}^{T} + \boldsymbol{P}_{k|k})
\end{array} \tag{22}$$

For the initialization, in absence of any side information, one can take $\boldsymbol{M}_{0}^{00} = 1/p_0\boldsymbol{I}$, $\boldsymbol{M}_{0}^{10} = \boldsymbol{0}$, $\boldsymbol{M}_{0}^{11} = \boldsymbol{0}$, $\widehat{R}_{\boldsymbol{xx},0} = 1/p_0\boldsymbol{I}$, $\widehat{R}_{\boldsymbol{xy},0} = \boldsymbol{0}$, $\widehat{R}_{\boldsymbol{yy},0} = \boldsymbol{0}$ where again p_0 is a very large number.

Adaptive processing means one iteration per new sample. Now, esp. in BASS, the parameters to be adapted represent filters. For short+long term AR processes with narrow bandwidths (well separable signals), these filters have a long memory, hence convergence transients are long and estimation efficiency is low.

G. Alternating MAP-ML KF (AMAPMLKF)

Joint MAP estimate for the state sequence $\boldsymbol{x}_k$ and ML estimate for the parameters θ, which is typically solved by alternating optimization between the two parts. The ML estimate of θ is then obtained by performing least-squares (LS) estimation, given the state sequence, which is replaced by its estimate. The resulting algorithm is similar to the EM-KF with only the $\widehat{\boldsymbol{x}}$ terms kept in the matrices $\boldsymbol{M}_{k}^{ij}$.

H. Variational Bayes KF (VB-KF)

This is again an application of alternating optimization, but this time applied to the Kullback-Leibler distance between the true joint posterior pdf of state and parameters, and an approximate product form using Gaussian pdfs for both state and parameters. In this case, not only does the estimation of the parameters account for the estimation errors in the state, but symmetrically the state estimation now also accounts for the estimation errors in the parameters, so the state estimation now uses a modified form of the KF. The fixed-interval vs. fixed-lag smoothing issue also applies to VB-KF (and AMAPMLKF).

I. Gaussian SOEKF for bilinear system

For a discussion of the Gaussian SOEKF, consider the bilinear state update equation $\boldsymbol{x}_{k+1} = \boldsymbol{f}(\boldsymbol{x}_k) + \boldsymbol{G}_k\boldsymbol{w}_k = \boldsymbol{F}_k\boldsymbol{x}_k + \boldsymbol{G}_k\boldsymbol{w}_k$. where $\boldsymbol{F}_k$ depends on the (augmented) $M\times 1$ state vector $\boldsymbol{x}_k$ linearly. We get for the SOEKF time update equations:

$$\begin{aligned}
\widehat{\boldsymbol{x}}_{k+1|k} &= \boldsymbol{f}(\widehat{\boldsymbol{x}}_{k|k}) + \tfrac{1}{2}\sum_{i=1}^{M}\phi_i\,tr\left\{\mathbf{H}_{i,k}\boldsymbol{P}_{k|k}\right\} \\
\boldsymbol{P}_{k+1|k} &= \mathbf{D}_k\boldsymbol{P}_{k|k}\mathbf{D}_k^{T} + \boldsymbol{G}_k\boldsymbol{Q}_k\boldsymbol{G}_k^{T} \\
&\quad + \tfrac{1}{2}\sum_{i,j=1}^{M}\phi_i\phi_j^{T}\,tr\left\{\mathbf{H}_{i,k}\boldsymbol{P}_{k|k}\mathbf{H}_{j,k}\boldsymbol{P}_{k|k}\right\} \\
\mathbf{H}_{i,k} &= \tfrac{\partial^2 \boldsymbol{f}_i}{\partial\boldsymbol{x}_k\partial\boldsymbol{x}_k^{T}}|_{\boldsymbol{x}_k=\widehat{\boldsymbol{x}}_{k|k}},\; \mathbf{D}_k = \tfrac{\partial\,\boldsymbol{f}_i}{\partial\boldsymbol{x}_k^{T}}|_{\boldsymbol{x}_k=\widehat{\boldsymbol{x}}_{k|k}}
\end{aligned} \tag{23}$$

where $\boldsymbol{f}_i$ is the i$^{\text{th}}$ component of the vector $\boldsymbol{f}$ and ϕ_i is the $M\times 1$ vector with all zeros except for 1 in the i$^{\text{th}}$ element. The terms involving the ϕ's represent correction terms w.r.t. the EKF.

III. Improved EKF (IEKF)

The EKF performs AKF by using the parameter estimate $\widehat{\boldsymbol{\theta}}$ without accounting for its estimation error $\widetilde{\boldsymbol{\theta}}$. We can correct that. With the extended state update equation being of the form $\boldsymbol{x}_{k+1} = \boldsymbol{F}_k\,\boldsymbol{x}_k + \boldsymbol{G}\boldsymbol{w}_k$, the predicted state $\widehat{\boldsymbol{x}}_{k+1|k}$ is computed as $\widehat{\boldsymbol{x}}_{k+1|k} = \boldsymbol{F}(\widehat{\boldsymbol{x}}_{k|k})\,\widehat{\boldsymbol{x}}_{k|k}$. Hence, we get for the state prediction error $\widetilde{\boldsymbol{x}}_{k+1|k} = \boldsymbol{x}_{k+1} - \widehat{\boldsymbol{x}}_{k+1|k}$ the following evolution:

$$\begin{aligned}
\widetilde{\boldsymbol{x}}_{k+1|k} &= (\boldsymbol{F}(\widehat{\boldsymbol{x}}_{k|k}) + \boldsymbol{F}(\widetilde{\boldsymbol{x}}_{k|k}))\,\boldsymbol{x}_k - \boldsymbol{F}(\widehat{\boldsymbol{x}}_{k|k})\,\widehat{\boldsymbol{x}}_{k|k} + \boldsymbol{G}\boldsymbol{w}_k \\
&= \boldsymbol{F}(\widehat{\boldsymbol{x}}_{k|k})\,\widetilde{\boldsymbol{x}}_{k|k} + \boldsymbol{G}\boldsymbol{w}_k + \boldsymbol{F}(\widetilde{\boldsymbol{x}}_{k|k})\,\boldsymbol{x}_k
\end{aligned} \tag{24}$$

The last term in the second line is not considered in the classic EKF. Considering the independence of the predicted state and the state error estimation (assuming Gaussian signals), the prediction error covariance gets updated as follows

$$\boldsymbol{P}_{k+1|k} = \boldsymbol{F}(\widehat{\boldsymbol{x}}_{k|k})\boldsymbol{P}_{k|k}\boldsymbol{F}(\widehat{\boldsymbol{x}}_{k|k})^{T} + \boldsymbol{G}_k\boldsymbol{Q}_k\boldsymbol{G}_k^{T} + \boldsymbol{O}_k \tag{25}$$

where $\boldsymbol{F}(\widehat{\boldsymbol{x}}_{k|k})$ equals $\mathbf{D}_k$ in (23) and $\boldsymbol{O}_k$ is the covariance matrix of $\boldsymbol{F}(\widetilde{\boldsymbol{x}}_{k|k})\,\boldsymbol{x}_k$ and is a correction term w.r.t. the basic EKF. Since $\boldsymbol{F}(\widetilde{\boldsymbol{x}}_{k|k})\,\boldsymbol{x}_k$ is linear in $\widetilde{\boldsymbol{x}}_{k|k}$, we can find a matrix function $\boldsymbol{D}$ such that

$$\boldsymbol{F}(\widetilde{\boldsymbol{x}}_{k|k})\,\boldsymbol{x}_k = \boldsymbol{D}(\boldsymbol{x}_k)\,\widetilde{\boldsymbol{x}}_{k|k} = \boldsymbol{D}(\widehat{\boldsymbol{x}}_{k|k})\,\widetilde{\boldsymbol{x}}_{k|k} + \boldsymbol{D}(\widetilde{\boldsymbol{x}}_{k|k})\,\widetilde{\boldsymbol{x}}_{k|k}.$$

Consequently, $\boldsymbol{O}_k$ can be computed as

$$\boldsymbol{O}_k = \boldsymbol{D}(\widehat{\boldsymbol{x}}_{k|k})\boldsymbol{P}_{k|k}\boldsymbol{D}(\widehat{\boldsymbol{x}}_{k|k})^T + \boldsymbol{N}_k$$

where $\boldsymbol{N}_k$ is the covariance of $\boldsymbol{D}(\widetilde{\boldsymbol{x}}_{k|k})\,\widetilde{\boldsymbol{x}}_{k|k}$ which is a fourth order moment that can be computed assuming a Gaussian distribution for $\widetilde{\boldsymbol{x}}_{k|k}$. $\boldsymbol{O}_k$ represents increased (awareness of) estimation error covariance which tends to dampen the filters (rendering them less frequency-selective), hence potentially allowing faster convergence transients.

IV. Performance Analysis: Back to Basics

The $\boldsymbol{y}$ are the measurements, the state sequence $\boldsymbol{x}$ are random (Gaussian) parameters (in some applications they are nuisance parameters, in others parameters of interest), and θ are the deterministic parameters. In a *joint* estimation approach, we maximize the likelihood

$$f(\boldsymbol{y}, \boldsymbol{x}|\theta) \tag{26}$$

which means MAP for $\boldsymbol{x}$ and ML for θ, with (joint) ML error covariance matrix $C^J_{\widetilde{\theta}\widetilde{\theta}}$ and CRB^J_θ. In a *marginalized* estimation approach, we maximize the likelihood

$$f(\boldsymbol{y}|\theta) \tag{27}$$

which means ML for θ, with (marginalized) ML error covariance matrix $C^M_{\widetilde{\theta}\widetilde{\theta}}$ and associated CRB^M_θ. Indeed, since the state $\boldsymbol{x}$ is random, it can be eliminated from the likelihood. Asymptotically (in the amount of data $\boldsymbol{y}$), we get

$$C^J_{\widetilde{\theta}\widetilde{\theta}} \overset{(i)}{\geq} C^M_{\widetilde{\theta}\widetilde{\theta}} \overset{(ii)}{=} \mathrm{CRB}^M_\theta \overset{(iii)}{\geq} \mathrm{CRB}^J_\theta \tag{28}$$

where (ii) is due to $\widehat{\theta}^M_{ML}$ being consistent, (i) is due to the inconsistency of $\widehat{\boldsymbol{x}}_{MAP}$ which prevents $\widehat{\theta}^J_{ML}$ from reaching its CRB. (iii) on the other hand follows from

$$\mathrm{CRB}^{-M}_\theta = \mathrm{CRB}^{-J}_\theta - \mathrm{E}_{\boldsymbol{y}|\theta} Cov_{\boldsymbol{x}|\boldsymbol{y},\theta}\left(\frac{\partial}{\partial\theta}\ln f(\boldsymbol{y}|\boldsymbol{x},\theta)\right).$$

(vectorized [17]) where $\mathrm{CRB}^{-M}_\theta = -\mathrm{E}_{\boldsymbol{y}|\theta}\frac{\partial^2}{\partial\theta\partial\theta^T}\ln f(\boldsymbol{y}|\theta)$, $\mathrm{CRB}^{-J}_\theta = -\mathrm{E}_{\boldsymbol{y}|\theta}\,\mathrm{E}_{\boldsymbol{x}|\boldsymbol{y},\theta}\frac{\partial^2}{\partial\theta\partial\theta^T}\ln f(\boldsymbol{y}|\boldsymbol{x},\theta)$. In other words, even though the CRBs would indicate otherwise, in terms of actual performance, joint estimation of the state and the parameters leads to worse parameter estimates than when the parameters are estimated in a marginalized fashion. So we get

- AMAPMLKF: $\widehat{\theta}$ from $\widehat{\boldsymbol{x}}$ only, $\widehat{\boldsymbol{x}}$ from $\widehat{\theta}$.
  Converges to joint MAP-ML (ML-KF).
- EM-KF: $\widehat{\theta}$ from $\widehat{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{x}}$, $\widehat{\boldsymbol{x}}$ from $\widehat{\theta}$.
  Now, it is known that the EM approach converges to the marginalized ML approach, so the EM-KF would be one approach to get this optimal performance.
- VB-KF: $\widehat{\theta}$ from $\widehat{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{x}}$, $\widehat{\boldsymbol{x}}$ from $\widehat{\theta}$ and $\widetilde{\theta}$.

Open issues: Can VB-KF do better than EM-KF? In VB-KF, the state estimate is improved, but the question is whether this allows an improved estimation of the parameters. Relative performance of EKF, RPEM-KF etc? And finally, relation of the IEKF to SOEKF, which also handles the bias due to the bilinearity.

References

[1] B. Anderson and J. Moore, *Optimal Filtering*. Prentice Hall (or more recently: Dover, 2005), 1979.

[2] M. Athans, R. Wishner, and A. Bertolini, "Suboptimal state estimation for continuous-time nonlinear systems from discrete noisy measurements," *Automatic Control, IEEE Transactions on*, vol. 13, no. 5, pp. 504 – 514, oct 1968.

[3] R. D. Bass, V. D. Norum, and L. Swartz, "Optimal multichannel nonlinear filtering," *J. Mufh. Anal. Appl.*, vol. 16, pp. 152 – 164, 1966.

[4] S. Bensaid, A. Schutz, and D. Slock, "Single Microphone Blind Audio Source Separation Using EM-Kalman Filter and Short+Long Term AR Modeling," in *Proc. Int'l Conf. on Latent Variable Analysis and Signal Separation (LVA-ICA)*, Saint-Malo, France, Sept. 2010.

[5] Consortium Partners, "Deliverable D2.3 "Hybrid Localization Techniques"," EC FP7 project WHERE, Tech. Rep., May 2010, URL: http://www.kn-s.dlr.de/where/public_documents_deliverables.php.

[6] C. Couvreur and Y. Bresler, "Decomposition of a mixture of gaussian ar processes," *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, vol. 3, pp. 1605–1608, 1995.

[7] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, vol. 39, no. 1, pp. 1–38, 1977.

[8] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the em algorithm," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 36, no. 4, pp. 477–489, Apr 1988.

[9] W. Gao, T. S., and J. Lehnert, "Diversity combining for ds/ss systems with time-varying, correlated fading branches," *Communications, IEEE Transactions on*, vol. 51, no. 2, pp. 284–295, Feb 2003.

[10] R. Henriksen, "The truncated second-order nonlinear filter revisited," *Automatic Control, IEEE Transactions on*, vol. 27, no. 1, pp. 247 – 251, feb 1982.

[11] A. H. Jazwinski, *Stochastic processes and filtering theory*, 1970.

[12] M. Lenardi and D. Slock, "Estimation of Time-Varying Wireless Channels and Application to the UMTS W-CDMA FDD Downlink," in *Proc. European Wireless (EW)*, Florence, Italy, Feb. 2002.

[13] L. Ljung, "Asymptotic behavior of the extended kalman filter as a parameter estimator for linear systems," *IEEE Transactions on Automatic Control*, vol. 24, no. 1, pp. 36 – 50, feb 1979.

[14] ——, *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall, 2002, 2nd edition.

[15] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press, 1983.

[16] R. Mehra, "On the identification of variances and adaptive kalman filtering," *IEEE Transactions on Automatic Control*, vol. 15, no. 2, pp. 175 – 184, apr 1970.

[17] M. Moeneclaey, "On the True and the Modified Cramer-Rao Bounds for the Estimation of a Scalar Parameter in the Presence of Nuisance Parameters," *IEEE Trans. Communications*, vol. 46, Nov. 1998.

[18] T. Sadiki and D. Slock, ""Bayesian Adaptive Filtering: Principles and Practical Approaches"," in *Proc. 12th European Signal Processing Conf. (EUSIPCO)*, Vienna, Austria, Sept. 2004.

[19] P. Tichavsky, C. Muravchik, and A. Nehorai, "Posterior cramer-rao bounds for discrete-time nonlinear filtering," *Signal Processing, IEEE Transactions on*, vol. 46, no. 5, pp. 1386 –1396, may 1998.

[20] H. L. V. Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. Wiley-IEEE Press, 2007.

[21] J. Villares and G. Vazquez, "The quadratic extended kalman filter," in *Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2004*, july 2004, pp. 480 – 484.

[22] E. Weinstein, A. Oppenheim, M. Feder, and J. Buck, "Iterative and sequential algorithms for multisensor signal enhancement," *Signal Processing, IEEE Transactions on*, vol. 42, no. 4, pp. 846–859, Apr 1994.

[23] J. Wiklander, "Performance comparison of the Extended Kalman Filter and the Recursive Prediction Error Method," Master's thesis, Linköping Univ., 2003, reg nr: LiTH-ISY-EX-3351, http://www.essays.se/essay/5eb48d45e1/.