Covariance matching Kalman filter for observable LTI systems

Rahul Moghe, Renato Zanetti and Maruthi Akella[1]

***Abstract*— It is well known that the Kalman filter is globally optimal for linear time-invariant systems excited with white Gaussian noise along the process and measurement channels. However, its application to practical situations is often sub-optimal because the noise covariance matrices typically are not accurately known. This paper presents a new covariance matching Kalman filter (CMKF) algorithm which estimates the measurement (R) or the process (Q) noise covariance matrix in addition to the state. The convergence of the filter for estimating Q or R matrices respectively is proved for discrete stochastic linear time-invariant systems under relatively mild conditions of observability.**

I. Introduction

The Kalman filter formulation ensures convergence of the state error covariance for the case when the system dynamics are linear and the additive process and measurement noise are white Gaussian [1], [2]. However, a complete knowledge about the system dynamics and the statistics of the process and measurement errors is assumed. In practical situations, however, these quantities are uncertain. This paper focuses on the problem of identifying the distribution of the process and measurement errors.

An erroneous noise model may result in divergence of the filter under certain conditions [3], [4], [5], [6], [7], [8]. The conditions for stability and semi-stability of the Kalman filter under incorrect noise covariance have also been studied [9], [10], [11]. Prior methods pre-calculate or estimate the noise statistics offline and use it in the filter. However, for real-time applications, studies suggest that an online adaptation of noise covariances can improve performance [12], [13]. Hence, online estimation of the covariance matrices is necessary towards ensuring filter convergence. We introduce the CMKF which simultaneously estimates the states and the covariance of either the white Gaussian process or measurement noises given the covariance matrix of the other.

Reference [14] summarizes early approaches to covariance estimation via Innovation-Based Estimation (IAE). In the covariance matching approach the covariance of the innovation sequence is estimated to calculate the noise covariance matrices online. A rigorous proof of convergence for this method is given in [15]. A similar approach was used to estimate the process covariance matrix of systems with a left-invertible observation matrix [16]. Our method only assumes that the system is observable. Various other adaptive Kalman filters were introduced which used the backward shift operator [17], H∞ filter estimate [18], M-estimation [19] and an optimal adaptive filter [20]. Our method only uses the measurement sequence to calculate the estimates of the states and the covariance matrix and proves convergence under conditions of observability and number of unknown elements in the system.

The major contribution of this paper is an adaptive covariance matching technique derived for observable LTI systems. The paper is organized as follows. The problem statement and the assumptions are listed in section II. The algorithm for recursively estimating both the state and the noise covariance is given in section III. Sections IV and V present stability results and stochastic convergence analysis. A simulated example is presented in section VII to show the effectiveness of the proposed algorithm. Finally, section VIII summarizes the contributions and motivates future directions of this research.

[1]The authors are with Department of Aerospace Engineering and Engineering Mechanics, University of Texas at Austin, Austin, TX 78712, USA `rahul.moghe@utexas.edu`, `renato@utexas.edu`, `makella@mail.utexas.edu`

II. Problem Statement

A. Kalman Filter

The state $x_k \in \mathbb{R}^n$, to be estimated using the measurements $y_k \in \mathbb{R}^p$ at time t_k evolve according to the following linear discrete-time difference equations:

$$\begin{aligned} x_{k+1} &= Fx_k + w_k \\ y_k &= Hx_k + v_k \end{aligned} \tag{1}$$

Here the process noise $w_k \in \mathbb{R}^n$, and measurement noise $v_k \in \mathbb{R}^p$ are both white Gaussian and uncorrelated with each other and the initial state x_0. The constant state transition matrix $F \in \mathbb{R}^{n\times n}$ and the observation matrix $H \in \mathbb{R}^{p\times n}$ respectively are assumed to be perfectly known. The Kalman filter equations provide the estimate of the state $\hat{x}_{k|k}$ and the error covariance matrix $P_{k|k}$ at each time t_k as follows [21]:

$$\begin{aligned} \hat{x}_{k|k-1} &= F\hat{x}_{k-1|k-1} \\ \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - H\hat{x}_{k|k-1}) \\ P_{k|k-1} &= FP_{k-1|k-1}F^T + Q \\ K_k &= P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1} \\ P_{k|k} &= P_{k|k-1} - K_kHP_{k|k-1} \end{aligned} \tag{2}$$

The constant process and measurement noise covariance matrices are given by Q and R respectively. If the process and measurement noises are white Gaussian, the state estimation in Eqs. (2) are optimal in the mean squared error sense [22].

B. Problem Description

Given a full knowledge of the system (F, H, Q and R), the Kalman filter is the best estimator of the system given in Eq. (1). However, most practical applications approximate the values of Q and R. Given F, H, and the measurements y_k, we develop an algorithm to estimate the state x_k and R given Q and also to estimate x_k and Q given R.

C. Assumptions

For this work, the following assumptions are made.

Assumption 1: Matrices $F, H, Q_k = Q \succ 0$, and $R_k = R \succ 0$ are constant with time.

Assumption 2: The pair (F,H) is assumed completely observable and the pair $(F, Q^{\frac{1}{2}})$ is assumed to be stabilizable.

Remark 1: Detectability and stabilizability assumptions above ensure that the Kalman filter from Eq. (2) converges [23].

Assumption 3: Either R or Q is perfectly known (no uncertainty), while the other is taken to be unknown.

Assumption 4: $(F, Q^{\frac{1}{2}})$ has no unreachable nodes on the unit circle in the complex plane.

Remark 2: This assumption along with the others stated above ensure the existence of a stabilizing solution of the algebraic Riccati equation [3], [24].

Assumption 5: The matrix S defined in Eq. (27) and matrix T defined in Eq. (31) has full column rank.

Remark 3: This assumption has to do with the estimability of Q and R matrices. A similar assumption was made about the number of unknown elements in Q matrix by [15]. The implications of this assumptions are discussed in section VI.

III. The CMKF Algorithm

The CMKF algorithm is derived in three steps, namely, formulating stacked dynamics, eliminating the state, and then estimating R or Q matrix.

A. Stacked Dynamics

Stacking n measurements vertically allows us to have the observability matrix M_o as the state transition matrix. The dynamics of the stacked measurements are derived below.

$$x_{k+1} = Fx_k + w_k \tag{3}$$
$$x_{k+2} = F^2 x_k + Fw_k + w_{k+1} \tag{4}$$
$$x_{k+n-1} = F^{n-1}x_k + F^{n-2}w_k + \cdots + w_{k+n-1} \tag{5}$$
$$y_k = Hx_k + v_k \tag{6}$$
$$y_{k+1} = HFx_k + Hw_k + v_{k+1} \tag{7}$$
$$y_{k+2} = HF^2 x_k + HFw_k + Hw_{k+1} + v_{k+2} \tag{8}$$
$$\begin{aligned} y_{k+n-1} = {} & HF^{n-1}x_k + HF^{n-2}w_k + \dots \\ & \dots + Hw_{k+n-2} + v_{k+n-1} \end{aligned} \tag{9}$$

$$\underbrace{\begin{bmatrix} y_{k+n-1} \\ y_{k+n-2} \\ \vdots \\ y_k \end{bmatrix}}_{\triangleq \mathscr{Y}_k} = \underbrace{\begin{bmatrix} HF^{n-1} \\ HF^{n-2} \\ \vdots \\ H \end{bmatrix}}_{\triangleq M_o} x_k + \underbrace{\begin{bmatrix} H & HF & HF^2 & \cdots & HF^{n-2} \\ \mathbf{0} & H & HF & \cdots & HF^{n-3} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & H \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}}_{\triangleq M_w} \underbrace{\begin{bmatrix} w_{k+n-2} \\ w_{k+n-3} \\ \vdots \\ w_k \end{bmatrix}}_{\triangleq W_k} + \underbrace{\begin{bmatrix} v_{k+n-1} \\ v_{k+n-2} \\ \vdots \\ v_k \end{bmatrix}}_{\triangleq V_k} \tag{10}$$

Eq. (10) can be compactly expressed as follows:

$$\mathscr{Y}_k = M_o x_k + M_w W_k + V_k \tag{11}$$
$$\mathscr{Y}_{k+1} = M_o x_{k+1} + M_w W_{k+1} + V_{k+1} \tag{12}$$

B. Eliminating the state

Assumption 2 ensures that the observability matrix M_o has full column rank. Hence, the Moore-Penrose pseudo inverse defined by $M_o^\dagger$ is such that $M_o^\dagger M_o$ is the identity matrix. Pre-multiplying by $M_o^\dagger$ in Eqs. (11) and (12),

$$M_o^\dagger \mathscr{Y}_k = x_k + M_o^\dagger M_w W_k + M_o^\dagger V_k \tag{13}$$
$$FM_o^\dagger \mathscr{Y}_k = Fx_k + FM_o^\dagger M_w W_k + FM_o^\dagger V_k \tag{14}$$
$$M_o^\dagger \mathscr{Y}_{k+1} = x_{k+1} + M_o^\dagger M_w W_{k+1} + M_o^\dagger V_{k+1} \tag{15}$$
$$= Fx_k + w_k + M_o^\dagger M_w W_{k+1} + M_o^\dagger V_{k+1} \tag{16}$$

Subtracting Eq. (16) and (14) and eliminating the state,

$$\begin{aligned} \underbrace{M_o^\dagger \mathscr{Y}_{k+1} - FM_o^\dagger \mathscr{Y}_k}_{\triangleq \mathscr{Z}_k} = {} & \underbrace{w_k + M_o^\dagger M_w W_{k+1} - FM_o^\dagger M_w W_k}_{\triangleq \mathscr{W}_k} \\ & + \underbrace{M_o^\dagger V_{k+1} - FM_o^\dagger V_k}_{\triangleq \mathscr{V}_k} \end{aligned} \tag{17}$$

wherein $\mathscr{Z}_k$, $\mathscr{W}_k$ and $\mathscr{V}_k$ are functions of the measurement sequence, the process noise and the measurement noise respectively. Note that here the time subscript k is used for simplicity even though the constituents have different time subscripts.

C. Estimating R Matrix

Given independent and identically distributed (i.i.d.) white Gaussian process and measurement noise sequences, the sequence $\mathscr{Z}_k$ is a zero mean stationary process. Therefore, the covariance of $\mathscr{Z}_k$ which is constant in time can be written as follows.

$$Cov(\mathscr{Z}) = Cov(\mathscr{W}) + Cov(\mathscr{V}) \tag{18}$$

Simplifying Eq. (17), we get

$$\mathscr{W}_k = A_1 w_k + A_2 w_{k+1} + \cdots + A_{n-1} w_{k+n-1} \tag{19}$$
$$\mathscr{V}_k = B_0 v_k + B_1 v_{k+1} + \cdots + B_n v_{k+n} \tag{20}$$

Here the constant matrices A_i and B_i can be pre-calculated. Note that this is a linear strictly stationary moving average time series with output $\mathscr{Z}_k$ and inputs w_k and v_k.

$$Cov(\mathscr{W}) = A_1 Q A_1^T + \cdots A_{n-1} Q A_{n-1}^T \tag{21}$$
$$Cov(\mathscr{V}) = B_0 R B_0^T + \cdots B_n R B_n^T \tag{22}$$

The time subscripts are dropped because of the stationarity assumptions in place. In order to estimate R given the Q matrix, $Cov(\mathscr{Z})$ has to be estimated. Given that the sequence $\mathscr{Z}_k$ is zero mean, the following estimator is used.

$$Cov(\mathscr{Z})_k = \frac{1}{k}\sum_{i=1}^{k} \mathscr{Z}_i \mathscr{Z}_i^T \tag{23}$$

Assuming $Cov(\mathscr{Z})_0 = 0$, Eq. (23) can be recursively calculated as follows.

$$Cov(\mathscr{Z})_k = \frac{k-1}{k} Cov(\mathscr{Z})_{k-1} + \frac{1}{k}\mathscr{Z}_k \mathscr{Z}_k^T \tag{24}$$

From Eqs. (18) and (22), $\hat{R}_k$ satisfies,

$$Cov(\mathscr{Z})_k - Cov(\mathscr{W}) = B_0 \hat{R}_k B_0^T + \cdots B_n \hat{R}_k B_n^T \tag{25}$$

Let $C_k = Cov(\mathscr{Z})_k - Cov(\mathscr{W})$ and let $vec(\cdot)$ be the operation of vectorizing a matrix. For example, $vec(A)_l = A_{ij}$ where $l = n(i-1) + j$ for $A \in \mathbb{R}^{n \times n}$.

$$vec(C_k) = vec(B_0 \hat{R}_k B_0^T + \cdots B_n \hat{R}_k B_n^T) \tag{26}$$

$$vec(C_k) = \underbrace{(B_0 \otimes B_0 + \cdots B_n \otimes B_n)}_{\triangleq S} vec(\hat{R}_k) \tag{27}$$

wherein, '$\otimes$' is the Kronecker product. Using assumption 5, we calculate the estimate $\hat{R}_k$ as follows.

$$vec(\hat{R}_k) = S^{\dagger} vec(C_k) \tag{28}$$

D. Estimating Q matrix

$$\underbrace{Cov(\mathscr{Z})_k - Cov(\mathscr{V})}_{D_k} = A_1 \hat{Q}_k A_1^T + \cdots A_n \hat{Q}_k A_n^T \tag{29}$$

$$vec(D_k) = vec(A_1 \hat{Q}_k A_1^T + \cdots A_n \hat{Q}_k A_n^T) \tag{30}$$

$$vec(D_k) = \underbrace{(A_1 \otimes A_1 + \cdots A_n \otimes A_n)}_{\triangleq T} vec(\hat{Q}_k) \tag{31}$$

From assumption 5, we calculate the estimate $\hat{Q}_k$ as follows.

$$vec(\hat{Q}_k) = T^{\dagger} vec(D_k) \tag{32}$$

E. CMKF Algorithm

The pseudo-code of the CMKF algorithm is as Algorithm 1. Note that $\hat{R}_k$ may turn out to be occasionally negative semidefinite because of a outstanding measurement. In the CMKF algorithm, the **if** statement retains the most recent positive definite $\hat{R}_k$ in the filter equations.

Algorithm 1 CMKF algorithm

Initialization: F, H, Q, $\hat{x}_0$, P_0, $Cov(\mathscr{Z})_0 = 0$
Input: measurement sequence $\{y_k\}$
Output: state estimate $\{\hat{x}_k\}$
for $k = 1$ to n **do**
  Using $\{y_k\}$ calculate $\mathscr{Z}_k$ ▷ Eq. (17)
  Using $\mathscr{Z}_k$ calculate $Cov(\mathscr{Z})_k$ ▷ Eq. (24)
  Using $Cov(\mathscr{Z})_k$ calculate $\hat{R}_k$ ▷ Eq. (28)
  if $\hat{R}_k \preceq 0$ **then**
    $\hat{R}_k = \hat{R}_{k-1}$
  end if
  $\hat{x}_{k|k-1} = F\hat{x}_{k-1|k-1}$
  $\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(y_k - H\hat{x}_{k|k-1})$
  $P_{k|k-1} = FP_{k-1|k-1}F^T + Q$
  $K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + \hat{R}_k)^{-1}$
  $P_{k|k} = P_{k|k-1} - K_k H P_{k|k-1}$
end for

IV. Convergence Analysis for $\hat{R}_k$

This section is divided into two subsections. The convergence in probability of the estimate $\hat{R}_k$ to the actual covariance R is shown. Then, the convergence of the new filter using the estimate is shown.

A. Convergence of $\hat{R}$

Using Eqs. (19) and (20), we get,

$$\mathscr{Z}_k = \sum_{i=1}^{n-1} A_i w_{k+i-1} + \sum_{i=0}^{n} B_i v_{k+i} \tag{33}$$

Since the noises are white Gaussian and uncorrelated with each other, Eq. (33) can be interpreted as a linear strictly stationary Gaussian time series. The central limit theorem for linear stationary time series ensures the following element-wise convergence result proved in [25].

$$\sqrt{k}\{Cov(\mathscr{Z})_k - Cov(\mathscr{Z})\}_{ij} \xrightarrow{D} \mathscr{N}(\mathbf{0}, \Omega_{ij}) \tag{34}$$

Here, the subscript ij denotes the element corresponding to the i^{th} row and j^{th} column of the matrix and the ^{D} denotes convergence in distribution. From Eqs. (18) and (25), we get,

$$Cov(\mathscr{Z})_k - Cov(\mathscr{Z}) = \sum_{l=0}^{n} B_l(\hat{R}_k - R)B_l^T \tag{35}$$

$$\sqrt{k}\Big\{\sum_{l=0}^{n} B_l(\hat{R}_k - R)B_l^T\Big\}_{ij} \xrightarrow{D} \mathscr{N}(\mathbf{0}, \Omega_{ij}) \tag{36}$$

$$\sqrt{k}\{S\ vec(\hat{R}_k - R)\}_t \xrightarrow{D} \mathscr{N}(\mathbf{0}, \Omega_{ij}) \tag{37}$$

wherein, $t = p(i-1) + j$. Using assumption 5 and pre-multiplying by $S^{\dagger}$, the limiting distribution is a sum of white Gaussian distributions which is again white Gaussian with a finite covariance $\bar{\Omega}_{ij}$ (independent of k).

$$\sqrt{k}\{\hat{R}_k - R\}_{ij} \xrightarrow{D} \mathscr{N}(\mathbf{0}, \bar{\Omega}_{ij}) \tag{38}$$

$$\lim_{k \longrightarrow \infty} Pr\big\{\{\hat{R}_k - R\}_{ij} > \varepsilon\big\} = 0, \quad \forall\ \varepsilon > 0 \tag{39}$$

B. Convergence of the state error covariance

Three different error covariance matrix sequences are considered here, namely, $\hat{P}_k$, the one-step predictor error covariance, P_k, the true error covariance of the filter which uses the true R to propagate the uncertainty but uses $\hat{R}_k$ to calculate the filter gain, and finally $P_{k,opt}$, the optimal error covariance matrix of the filter given full knowledge of the noise statistics. More specifically,

$$\hat{P}_{k+1} = \hat{\bar{F}}_k \hat{P}_k \hat{\bar{F}}_k^T + \hat{K}_k \hat{R}_k \hat{K}_k^T + Q \tag{40}$$

$$P_{k+1} = \hat{\bar{F}}_k P_k \hat{\bar{F}}_k^T + \hat{K}_k R \hat{K}_k^T + Q \tag{41}$$

$$P_{k+1,opt} = \bar{F}_k P_{k,opt} \bar{F}_k^T + K_k R K_k^T + Q \tag{42}$$

wherein,

$$\hat{K}_k = F\hat{P}_k H^T(H\hat{P}_k H^T + \hat{R}_k)^{-1}, \quad (43)$$

$$K_k = FP_{k,opt}H^T(HP_{k,opt}H^T + R)^{-1}, \quad (44)$$

$$\hat{\bar{F}}_k = F - \hat{K}_k H, \ and \ \bar{F}_k = F - K_k H \quad (45)$$

Let us first look at the asymptotics of the matrix sequence $\hat{P}_k - P_k$. Subtracting equations (40) and (41), we get

$$\hat{P}_{k+1} - P_{k+1} = \hat{\bar{F}}_k(\hat{P}_k - P_k)\hat{\bar{F}}_k^T + \hat{K}_k(\hat{R}_k - R)\hat{K}_k^T \quad (46)$$

The initial knowledge about the state is the same in each of the three matrix sequences. Hence, using $\hat{P}_0 = P_0 = P_{0,opt}$,

$$\hat{P}_{k+1} - P_{k+1} = \sum_{i=0}^{k} \hat{\phi}_i K_i^r(\hat{R}_i - R)\hat{K}_i^T \hat{\phi}_i^T \quad (47)$$

$$\lim_{k\longrightarrow\infty}(\hat{P}_{k+1} - P_{k+1}) = \sum_{i=0}^{\infty} \hat{\phi}_i K_i^r(\hat{R}_i - R)\hat{K}_i^T \hat{\phi}_i^T \quad (48)$$

where $\hat{\phi}_i = \hat{\bar{F}}_i\hat{\bar{F}}_{i-1}\cdots\hat{\bar{F}}_0$ is the state transition matrix corresponding to $\hat{\bar{F}}_k$ from initial time to the i^{th} time. Consider the partial sum $\Delta P_{m,n}$ from m to n as follows.

$$\Delta P_{m,n} = \sum_{i=m}^{n} \hat{\phi}_i \hat{K}_i(\hat{R}_i - R)\hat{K}_i^T \hat{\phi}_i^T \quad (49)$$

We know that,

$$\lim_{k\longrightarrow\infty} Pr\big\{\{\hat{R}_k - R\}_{ij} > \varepsilon\big\} = 0 \quad (50)$$

Therefore, $\forall\ \varepsilon > 0,\ \forall\ \delta > 0,\ \exists N;\ \forall\ n \geq N$,

$$Pr\{\|\hat{R}_k - R\| < \varepsilon\} > 1 - \delta \quad (51)$$

Now using Eq. (49),

$$\|\Delta P_{m,n}\| = \|\sum_{i=m}^{n} \hat{\phi}_i \hat{K}_i(\hat{R}_i - R)\hat{K}_i^T \hat{\phi}_i^T\| \quad (52)$$

$$\leq \sum_{i=m}^{n} \|\hat{\phi}_i \hat{K}_i(\hat{R}_i - R)\hat{K}_i^T \hat{\phi}_i^T\| \quad (53)$$

$$\leq \sum_{i=m}^{n} \|\hat{\phi}_i\|^2 \|\hat{K}_i\|^2 \|\hat{R}_i - R\| \quad (54)$$

From Eq. (51), one can choose a fixed δ and an ε_k corresponding to the $\hat{R}_k$ such that,

$$\varepsilon_k = \frac{\varepsilon}{(n-m)\|\hat{\phi}_k\|_\infty^2 \|\hat{K}_k\|_\infty^2}$$

We know that $\exists\ N_k$ satisfying Eq. (51) for $\varepsilon = \varepsilon_k$. Using $N = \max_{k\in[n,m]}\{N_k\}$, we get the following result.

For all $\varepsilon > 0,\ \forall\ \delta > 0,\ \exists N; \forall\ n \geq N$,

$$Pr\{\|\Delta P_{m,n}\| < \varepsilon\} > 1 - \delta \quad (55)$$

Therefore, $\lim_{m,n\longrightarrow\infty} Pr\{\|\Delta P_{m,n}\| < \varepsilon\} = 1 \quad (56)$

This proves that the sequence $\Delta P_{0,n}$ is a Cauchy sequence. This implies that this sequence has a limit point or there exists a convergent subsequence $\Delta P_{0,n_k}$. However, the limit point of this subsequence is in fact, $\lim_{k\longrightarrow\infty} Pr\{\|\Delta P_{0,n_k}\| < \varepsilon\}$. Hence,

$$\lim_{k\longrightarrow\infty} Pr\{\|\hat{P}_k - P_k\| < \varepsilon\} = 1 \quad (57)$$

$$\hat{P}_k \xrightarrow{P} P_k \text{ as } k \to \infty \quad (58)$$

We just proved the following,

$$\hat{K}_k \xrightarrow{P} FP_k H^T(HP_k H^T + R)^{-1} \text{ as } k \to \infty \quad (59)$$

Hence, we can say that Eq. (41) converges in probability to the Algebraic Riccati equation (Eq. (42)).

$$P_{k+1} = FP_kF^T + FP_kH^T(HP_kH^T + R)^{-1}HP_kF^T + Q \quad (60)$$

This equation is exactly the same as Eq. (42), and hence, both share a solution at steady state. However, we still have to discuss the contribution of the terms during transition at lower k values. Using the assumption that (F,H) is observable, we know that the state transition matrix corresponding to $\hat{\bar{F}}_k$ and $\bar{F}_k$ are stable ([3] Lemma 3.1). Therefore, the contribution of the transient terms, weighted by a stable state transition matrix diminishes in the limit as $k \to \infty$. This is also evident when eq. (41) and (42) are expanded as follows:

$$P_{k+1} = \hat{\phi}_k P_0 \hat{\phi}_k^T + \sum_{i=0}^{k} \hat{\phi}_i \big[\hat{K}_{k-i} R \hat{K}_{k-i}^T + Q\big] \hat{\phi}_i^T \quad (61)$$

$$P_{k+1,opt} = \phi_k P_0 \phi_k^T + \sum_{i=0}^{k} \phi_i \big[K_{k-i} R K_{k-i}^T + Q\big] \phi_i^T \quad (62)$$

$$\lim_{k\to\infty} P_{k+1} = \sum_{i=0}^{\infty} \hat{\phi}_i \big[\hat{K}_{k-i} R \hat{K}_{k-i}^T + Q\big] \hat{\phi}_i^T \quad (63)$$

$$\lim_{k\to\infty} P_{k+1,opt} = \sum_{i=0}^{\infty} \phi_i \big[K_{k-i} R K_{k-i}^T + Q\big] \phi_i^T \quad (64)$$

Here, $\phi_k = \bar{F}_k\bar{F}_{k-1}\cdots\bar{F}_0$. The initial terms of both the infinite series have low contribution to the series sum. Hence,

$$P_k \xrightarrow{P} P_{k,opt} \text{ as } k \to \infty \quad (65)$$

Using Eqs. (58) and (65), we get

$$\hat{P}_k \xrightarrow{P} P_{k,opt} \text{ as } k \to \infty \quad (66)$$

V. Convergence Analysis for $\hat{Q}_k$

A. Convergence of $\hat{Q}$

Using similar arguments from section IV-A, we get a similar set of equations for $\hat{Q}_k$.

$$\sqrt{k}\{Cov(\mathscr{Z})_k - Cov(\mathscr{Z})\}_{ij} \to \mathscr{N}(\mathbf{0}, \Lambda_{ij}) \quad (67)$$

$$Cov(\mathscr{Z})_k - Cov(\mathscr{Z}) = \sum_{l=1}^{n} A_l(\hat{Q}_k - Q)A_l^T \quad (68)$$

$$\sqrt{k}\{\sum_{l=1}^{n} A_l(\hat{Q}_k - Q)A_l^T\}_{ij} \to \mathscr{N}(\mathbf{0}, \Lambda_{ij}) \quad (69)$$

$$\sqrt{k}\{T\ \ vec(\hat{Q}_k - Q)\}_t \to \mathscr{N}(\mathbf{0}, \Lambda_{ij}) \quad (70)$$

wherein $t = n(i-1) + j$. Again, using assumption 5,

$$\sqrt{k}\{\hat{Q}_k - Q\}_{ij} \xrightarrow{D} \mathscr{N}(\mathbf{0}, \bar{\Lambda}_{ij}) \quad (71)$$

$$\lim_{k\longrightarrow\infty} Pr\big\{\{\hat{Q}_k - Q\}_{ij} > \varepsilon\big\} = 0, \quad \forall\ \varepsilon > 0 \quad (72)$$

B. Convergence of error covariance

We consider three error covariances as before with the only difference being that the R matrix is perfectly known and the $\hat{Q}_k$ is being updated recursively.

$$\hat{P}_{k+1} = \hat{\bar{F}}_k \hat{P}_k \hat{\bar{F}}_k^T + \hat{K}_k R \hat{K}_k^T + \hat{Q}_k \tag{73}$$

$$P_{k+1} = \hat{\bar{F}}_k P_k \hat{\bar{F}}_k^T + \hat{K}_k R \hat{K}_k^T + Q \tag{74}$$

$$P_{k+1,opt} = \bar{F}_k P_{k,opt} \bar{F}_k^T + K_k R K_k^T + Q \tag{75}$$

wherein,

$$\hat{K}_k = F \hat{P}_k H^T (H \hat{P}_k H^T + R)^{-1}, \tag{76}$$

$$K_k = F P_{k,opt} H^T (H P_{k,opt} H^T + R)^{-1}, \tag{77}$$

$$\hat{\bar{F}}_k = F - \hat{K}_k H, \text{ and } \bar{F}_k = F - K_k H \tag{78}$$

Subtracting Eq. (73) and (74), we get

$$\hat{P}_{k+1} - P_{k+1} = \hat{\bar{F}}_k (\hat{P}_k - P_k) \bar{F}_k^T + (\hat{Q}_k - Q) \tag{79}$$

$$\hat{P}_{k+1} - P_{k+1} = \sum_{i=0}^{k} \hat{\phi}_i (\hat{Q}_{k-i} - Q) \phi_i^T \tag{80}$$

Using arguments similar from those in section IV-B, we can show that the above sequence is a Cauchy sequence and further prove the following convergence.

$$Pr\{\|\hat{P}_k - P_k\| > \varepsilon\} \xrightarrow{k\to\infty} 0 \tag{81}$$

$$\hat{P}_k \xrightarrow{P} P_k \text{ as } k \to \infty \tag{82}$$

Now comparing the actual error covariance (Eq. (74)) with the optimal Kalman filter error covariance (Eq. (75)), we use arguments from section IV-B to prove the following.

$$P_k \xrightarrow{P} P_{k,opt} \text{ as } k \to \infty$$

$$\text{Therefore, } \hat{P}_k \xrightarrow{P} P_{k,opt} \text{ as } k \to \infty$$

VI. Estimability of R and Q

If assumption 5 fails, S and T matrices can be modified in some cases by using the structure of the R and Q matrices to be estimated, e.g., symmetry, known elements and constraints on the elements. In these cases, the corresponding columns of S and T matrices can be subtracted or averaged out from the left hand side to remove the rank deficiency. Therefore, we modify Eqs. (28) and (32) to get,

$$vec(\hat{R}_k)_u = \bar{S}^\dagger vec(C_k) \tag{83}$$

$$vec(\hat{Q}_k)_u = \bar{T}^\dagger vec(D_k) \tag{84}$$

wherein $(\cdot)_u$ denotes the unknown and unique entries and $\bar{S}$ and $\bar{T}$ are the modified matrices.

VII. Simulations

We adopt an example from [16] and choose a non-invertible H matrix.

$$\begin{aligned} x_k &= \begin{bmatrix} 1 & T & 0.5T^2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix} x_{k-1} + w_{k-1} \\ y_k &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} x_k + v_k \end{aligned} \tag{85}$$

Assume that $T = 0.05$, and $w_k \sim \mathcal{N}(0,Q)$, and $v_k \sim \mathcal{N}(0,R)$ are both i.i.d white Gaussian noises. We assume

$$R = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix} \text{ and } Q = \begin{bmatrix} 0.25 & 0.04 & 0.04 \\ 0.04 & 0.25 & 0.04 \\ 0.04 & 0.04 & 0.25 \end{bmatrix}$$

The initial conditions in their respective simulations were $\hat{R}_0 = 5I_{2\times 2}$ and $\hat{Q}_0 = 5I_{3\times 3}$. From simulation results, the 2 norm of the error between $\hat{R}_k$ and R converges to within 10^{-1} while for estimating Q matrix it converges to 10^0 as shown in Fig. (1). For both R and Q estimation, the three state error covariance converge to the optimal Kalman filter values Figs. (4) and (6) which represents the 3σ values for the state estimation error shown in Figs. (3) and (5).

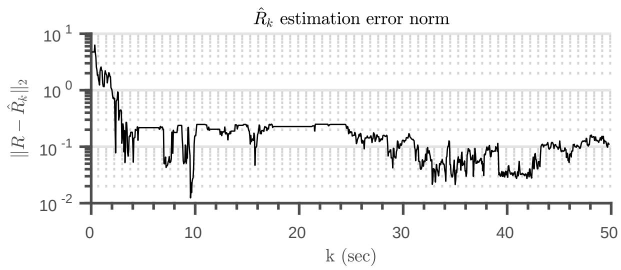

Fig. 1. Estimation error norm of the process covariance matrix R.

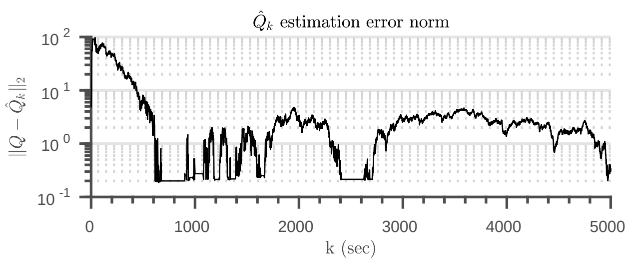

Fig. 2. Estimation error norm of the process covariance matrix Q. The flat portions show the cases where the calculated $\hat{Q}_k$ is negative definite and hence, not updated.

VIII. Conclusion

A novel adaptive Kalman filter algorithm with proved convergence was developed to estimate the process or measurement noise covariance matrix of a LTI stochastic system. The convergence of the states and the covariance matrix was shown using a numerical simulation. Future work would explore generalizations to towards LTV systems and weakening the observability assumption to detectability.

References

[1] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Journal of basic engineering*, vol. 83, no. 1, pp. 95–108, 1961.

[2] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[3] S. Sangsuk-Iam and T. E. Bullock, "Analysis of discrete-time kalman filtering under incorrect noise covariances," *IEEE Transactions on Automatic Control*, vol. 35, no. 12, pp. 1304–1309, 1990.

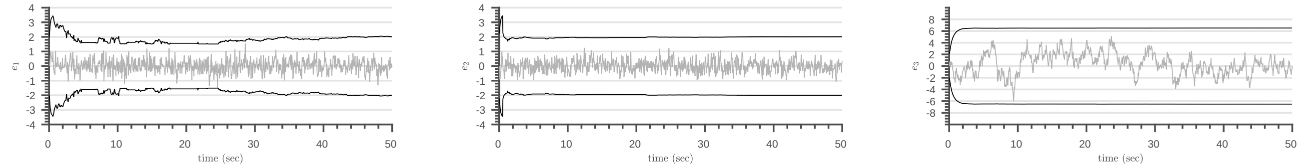

Fig. 3. $\mathbf{\hat{R}_k}$ **case**: Estimation errors for the three states with the 3σ values corresponding to their predictor error covariance $\hat{P}_k$

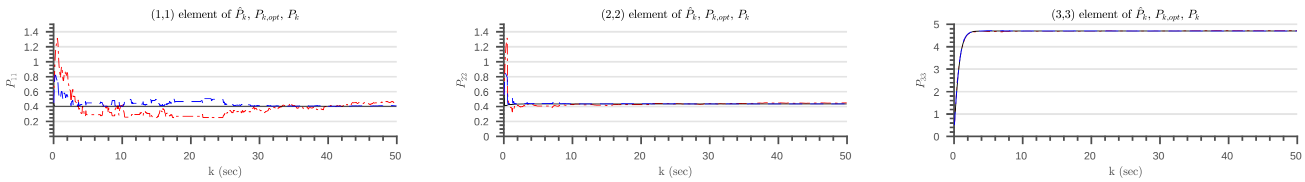

Fig. 4. $\mathbf{\hat{R}_k}$ **case**: Comparison of the three different error covariances from Eqs. (40) (red), (41) (blue) and (42) (black).

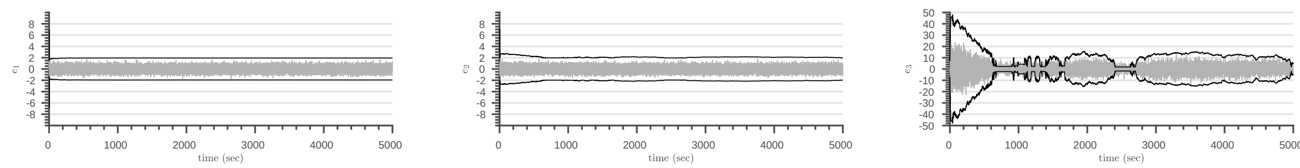

Fig. 5. $\mathbf{\hat{Q}_k}$ **case**: Estimation errors for the three states with the 3σ values corresponding to their predictor error covariance $\hat{P}_k$

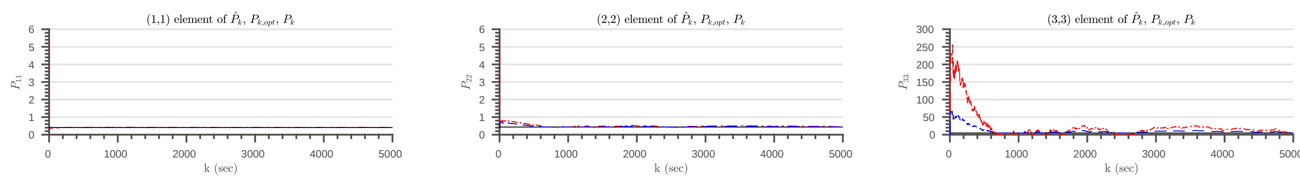

Fig. 6. $\mathbf{\hat{Q}_k}$ **case**: Comparison of the three different error covariances from Eq. (73) (red), (74) (blue) and (75) (black).

[4] H. Heffes, "The effect of erroneous models on the kalman filter response," *IEEE Transactions on Automatic Control*, vol. 11, no. 3, pp. 541–543, 1966.

[5] N. Madjarov and L. Mihaylova, "Kalman filter sensitivity with respect to parametric noises uncertainty," *Kybernetika*, vol. 32, no. 3, pp. 307–322, 1996.

[6] S. S. Saab, "Discrete-time kalman filter under incorrect noise covariances," in *American Control Conference, Proceedings of the 1995*, vol. 2. IEEE, 1995, pp. 1152–1156.

[7] S. Sangsuk-Iam and T. E. Bullock, "Analysis of continuous-time kalman filtering under incorrect noise covariances," *Automatica*, vol. 24, no. 5, pp. 659–669, 1988.

[8] J. L. Willems and F. M. Callier, "Divergence of the stationary kalman filter for correct and for incorrect noise variances," *IMA Journal of Mathematical Control and Information*, vol. 9, no. 1, pp. 47–54, 1992.

[9] E. F. Costa and A. Astolfi, "A necessary and sufficient condition for semi-stability of the recursive kalman filter," in *American Control Conference, 2008*. IEEE, 2008, pp. 1280–1285.

[10] ——, "On the stability of the recursive kalman filter for linear time-invariant systems," in *American Control Conference, 2008*. IEEE, 2008, pp. 1286–1291.

[11] ——, "Bounds related to the riccati difference equation for linear time varying systems," in *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*. IEEE, 2009, pp. 2321–2326.

[12] C. Hide, T. Moore, and M. Smith, "Adaptive kalman filtering for low-cost ins/gps," *The Journal of Navigation*, vol. 56, no. 1, pp. 143–152, 2003.

[13] K. Myers and B. Tapley, "Adaptive sequential estimation with unknown noise statistics," *IEEE Transactions on Automatic Control*, vol. 21, no. 4, pp. 520–523, 1976.

[14] R. Mehra, "Approaches to adaptive filtering," *IEEE Transactions on automatic control*, vol. 17, no. 5, pp. 693–698, 1972.

[15] ——, "On the identification of variances and adaptive kalman filtering," *IEEE Transactions on automatic control*, vol. 15, no. 2, pp. 175–184, 1970.

[16] B. Feng, M. Fu, H. Ma, Y. Xia, and B. Wang, "Kalman filter with recursive covariance estimation-sequentially estimating process noise covariance," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 11, pp. 6253–6263, 2014.

[17] H. Wang, Z. Deng, B. Feng, H. Ma, and Y. Xia, "An adaptive kalman filter estimating process noise covariance," *Neurocomputing*, vol. 223, pp. 12–17, 2017.

[18] X. Wang, M.-q. Liu, Z. Fan, and S.-l. Zhang, "A novel approach of noise statistics estimate using h∞ filter in target tracking," *Frontiers of Information Technology & Electronic Engineering*, vol. 17, no. 5, pp. 449–457, 2016.

[19] Z. M. Durovic and B. D. Kovacevic, "Robust estimation with unknown noise statistics," *IEEE Transactions on Automatic Control*, vol. 44, no. 6, pp. 1292–1296, 1999.

[20] Y. Yang and W. Gao, "An optimal adaptive kalman filter," *Journal of Geodesy*, vol. 80, no. 4, pp. 177–183, 2006.

[21] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.

[22] B. D. Anderson and J. B. Moore, "Optimal filtering," *Englewood Cliffs*, vol. 21, pp. 22–95, 1979.

[23] B. Anderson and J. B. Moore, "Detectability and stabilizability of time-varying discrete-time linear systems," *SIAM Journal on Control and Optimization*, vol. 19, no. 1, pp. 20–32, 1981.

[24] S. Chan, G. Goodwin, and K. Sin, "Convergence properties of the riccati difference equation in optimal filtering of nonstabilizable systems," *IEEE Transactions on Automatic Control*, vol. 29, no. 2, pp. 110–118, 1984.

[25] Y. Hosoya and M. Taniguchi, "A central limit theorem for stationary processes and the parameter estimation of linear processes," *The Annals of Statistics*, pp. 132–153, 1982.