Adaptive Kalman filter for detectable linear time invariant systems

Rahul Moghe *, Renato Zanetti†, and Maruthi R. Akella‡
The University of Texas at Austin, Austin, Texas 78712

A novel covariance matching technique is proposed for estimating the states and unknown entries of the process and measurement noise covariance matrices for additive white Gaussian noise elements in a linear Kalman filter. Under this assumption of detectability (i.e., unobservable modes remain stable), the stability and convergence properties of the covariance matching Kalman filter are established. It is shown that the measurement covariance matrix cannot be unambiguously estimated if the measurement model contains linearly dependent measurements. Monte Carlo simulations evaluate the numerical properties of the proposed algorithm.

I. Introduction

The Kalman filter is an optimal estimation algorithm in the mean squared sense for a linear system with additive white Gaussian noises in the process and measurement channels [1, 2]. However, optimality is lost when one or more of the underlying assumptions are violated as the case with: (a) nonlinearities in state or measurement model; (b) non-Gaussian or correlated noise sequences; and (c) white Gaussian noise with inaccurate covariance matrices. Accordingly, nonlinear filtering and estimation algorithms have been extensively studied in the literature [3–6]. The issue of non-Gaussian noises has traditionally been handled by artificially inflating the covariance matrices. Instead of these ad hoc inflation techniques, more sophisticated algorithms have also been developed [7–9]. The case when the noise covariance matrices are unknown has been shown to cause filter divergence [10–15]. These challenges associated with filter divergence provided motivation for the development of adaptive filtering algorithms that simultaneously estimate the system states along with the covariance matrices. Although several of these adaptive formulations have found their application in many individual settings, rigorous theoretical foundations are usually absent and if present, are accompanied by restrictive assumptions such as steady state conditions, invertibility of the state transition and observation matrix or availability of the true state estimate. Addressing this gap in the existing literature, this paper presents an adaptive covariance matching technique for detectable linear time-invariant (LTI) systems based on correlation methods to recursively estimate the state vector and certain number of unknown elements within both the noise covariance

*Ph.D. Student, Aerospace Engineering and Engineering Mechanics Department; rahul.moghe@utexas.edu, Student Member AIAA.
†Assistant Professor, Aerospace Engineering and Engineering Mechanics Department; renato@utexas.edu, Senior Member AIAA.
‡E. P. Schoch Professor, Aerospace Engineering and Engineering Mechanics Department; makella@mail.utexas.edu, Associate Fellow AIAA.

matrices.

An early study of adaptive algorithms to estimate noise covariance matrices is summarized in [16]. Broadly speaking, there exist four approaches to adaptive filtering, namely, Bayesian methods, maximum likelihood estimation, correlation methods, and covariance matching techniques. The algorithm presented in this paper adopts concepts based on correlation methods, specifically in terms of forming a linear time series using the measurement sequence. We also embrace certain ideas from covariance matching techniques by way of formally enforcing covariance consistency. It should be noted that for LTI systems, steady state convergence of an adaptive filter using the innovation sequence for estimation of the noise covariance matrix was originally reported in [17]. However, the relaxation of certain assumptions central to [17] such as steady-state conditions, observability, and invertibility of the state transition matrix makes our result much stronger. Moreover, technical insufficiencies concerning the whiteness condition in [17] as clearly pointed out in [10, 18] play no role whatsoever in our formulation and its underlying convergence analysis. Specific differences between [17] and our work are described in further detail in the sequel. An approach to estimate only the process noise covariance using a backward shift operator is presented in [19]. This method uses a left co-prime factorization which is either not guaranteed to exist or can be difficult to analytically calculate. H_∞ filtering is used to estimate the statistics of the noises in [20]. This approach can be classified as offline since it uses the H_∞ filter estimate at future times to calculate the estimate of covariance matrices. The approach given in [21] uses linear regression to derive a robust Kalman filter. Our proposed algorithm significantly extends the work of [22] by way of providing rigorous convergence guarantees for both the state and covariance estimates subject to satisfaction of the detectability condition. Moreover, our result provides the explicit characterization of necessary conditions for estimability of the process and measurement noise covariance matrices.

A recent survey of covariance estimation results for LTI systems using correlation methods can be found in [23]. Standing out from among the various methods surveyed in [23] is the measurement averaging correlation method (MACM) which forms a stacked measurement model wherein the observability matrix is referred to as the observation matrix. This paper shares some of the algebraic features of the MACM approach wherein we too construct a measurement stack as part of our overall formulation. Another approach to estimate the process noise covariance matrix of LTI systems assuming a left-invertible observation matrix is presented in [24]. In this case, a linear stationary time series is formed by inverting the observation matrix and the covariance is estimated by squaring the sequence. A similar method which assumes full column rank of the observation matrix is presented in [25].

More recently, a covariance estimation result is presented in [26] for linear time-varying (LTV) systems using the linear time series construction. Specifically, in [26], a stacked measurement model is formed to estimate the N-step predicted measurement and the measurement prediction error is shown to form a linear time series. The sample covariance of the linear time series is used to calculate estimates of the covariance matrices. Our paper also formulates a time series with the important distinction that it uses minimal number of measurements required. Moreover, compared to [26], we take a significant next step – that of simultaneously estimating the state of the system. A major contribution of this work is a proof of convergence for the state estimator. For purposes of calculating the covariance estimate using the covariance of the time series, full column rank of a certain coefficient matrix in assumed in [26]. In this paper, we mathematically characterize how this full column rank condition corresponds to certain intrinsic physical properties of the system pertaining to estimability.

A preliminary version of the filter derived here was reported earlier by the authors as a conference paper in [27]. In that prior work, full observability for the system was assumed and only one of either the process or measurement noise covariance matrix was to be estimated given the other covariance matrix. In this paper, we estimate a set of unknown elements which can be from either the process or measurement noise covariance matrices while also relaxing the observability requirement to uniform detectability. Weakening the observability assumption to detectability allows for applicability of our results to a larger class of systems since some states of a detectable system are not observable while remaining stable.

The paper is organized as follows. The baseline Kalman filter, the problem statement, and the various technical assumptions are listed in section II. A linear stationary time series is then derived using the modified stacked measurement model. The filter is subsequently formulated in section III. In section IV, we present stability results and the stochastic convergence analysis. The estimability of noise covariance matrices is analyzed in section V. Having presented our main result, further discussion on the significance of our result and its contrast with related past literature is presented in section VI. A simulated example is described in section VII to study the effectiveness of the proposed algorithm. Finally, section VIII presents some concluding remarks to motivate fruitful directions for future research.

II. Problem Statement

An LTI system is considered of the form

$$\left.\begin{aligned} x_{k+1} &= \tilde{F} x_k + \tilde{w}_k \\ y_k &= \tilde{H} x_k + v_k \end{aligned}\right\} \tag{1}$$

wherein $\tilde{w}_k$ and v_k are the process and measurement noises. We assume that the pair $(\tilde{F}, \tilde{H})$ is uniformly detectable. Hence, there exists an invertible state transformation matrix W such that the system can be transformed into

$$\left.\begin{aligned} z_{k+1} &= \begin{bmatrix} F_{11} & \mathbf{0}_{l\times s} \\ F_{21} & F_{22} \end{bmatrix} \begin{bmatrix} z_k^O \\ z_k^{UO} \end{bmatrix} + W\tilde{w}_k \\ y_k &= \begin{bmatrix} H_1 & \mathbf{0}_{p\times s} \end{bmatrix} \begin{bmatrix} z_k^O \\ z_k^{UO} \end{bmatrix} + v_k \end{aligned}\right\} \tag{2}$$

wherein, the state and the transformed state at time t_k are respectively denoted by x_k and $z_k \in \mathbb{R}^n$. The integers l and s are the dimensions of the observable $(^O)$ and unobservable $(^{UO})$ subspace of the state respectively such that $s + l = n$. Accordingly, the observable state $z_k^O \in \mathbb{R}^l$ and the unobservable state $z_k^{UO} \in \mathbb{R}^s$. The measurement at time t_k is given by $y_k \in \mathbb{R}^p$. The pair (F_{11}, H_1) is uniformly observable and F_{22} is stable (i.e. all eigenvalues within the unit sphere). The matrix W is such that $z_k = Wx_k$ and we define $w_k = W\tilde{w}_k = \begin{bmatrix} w_k^{O^T} & w_k^{UO^T} \end{bmatrix}^T$ to be the transformed process noise. The transformed state transition matrix is given by

$$F \triangleq W\tilde{F}W^{-1} = \begin{bmatrix} F_{11} & \mathbf{0}_{l\times s} \\ F_{21} & F_{22} \end{bmatrix} \tag{3}$$

and the observation matrix is defined as

$$H \triangleq \tilde{H}W^{-1} = \begin{bmatrix} H_1 & \mathbf{0}_{p\times s} \end{bmatrix}. \tag{4}$$

The process noise $w_k \in \mathbb{R}^n$ and the measurement noise $v_k \in \mathbb{R}^p$ are white Gaussian and uncorrelated with each other. The transformed state at time t_0 is denoted by z_0. Hence, $w_k \sim \mathcal{N}(\mathbf{0}_{n\times 1}, Q)$ and $v_k \sim \mathcal{N}(\mathbf{0}_{p\times 1}, R)$, wherein Q can be partitioned into observable and unobservable subspaces as $Q = \begin{bmatrix} Q^O & Q^{O,UO} \\ Q^{UO,O} & Q^{UO} \end{bmatrix}$.

The baseline Kalman filter equations for known Q and R matrices are given by [28]:

$$\left.\begin{aligned} \hat{z}^{O}_{k|k-1} &= F_{11}\hat{z}^{O}_{k-1|k-1} \\ \hat{z}^{O}_{k|k} &= \hat{z}^{O}_{k|k-1} + K_k(y_k - H_1\hat{z}^{O}_{k|k-1}) \\ P_{k|k-1} &= F_{11}P_{k-1|k-1}F_{11}^T + Q^O \\ K_k &= P_{k|k-1}H_1^T(H_1P_{k|k-1}H_1^T + R)^{-1} \\ P_{k|k} &= P_{k|k-1} - K_kH_1P_{k|k-1} \end{aligned}\right\} \tag{5}$$

wherein K_k is the Kalman gain, $P_{k|k-1}$ and $P_{k|k}$ are the state error covariance matrices after the prediction and update step respectively. The baseline Kalman filter equations contain the observable part of the Q matrix only. The unobservable state z_k^{UO} cannot be estimated as the system is not fully observable. The noise covariance matrices Q and R are constant for all time t_k. If the process and measurement noises are white Gaussian, the baseline Kalman filter in (5) is optimal in the mean squared error sense and the state error covariance $P_{k|k}$ converges to a steady-state value [29].

A. Problem Description

Given full knowledge of the system matrices (F_{11}, F_{21}, F_{22}, H_1, Q and R), the Kalman filter is the *best* estimator of the system given by (1). However, most practical applications approximate the values of Q and R. Thus, the problem addressed in this paper can be stated as follows: given F_{11}, F_{21}, F_{22}, H_1, and measurements y_k, we formulate an adaptive algorithm to estimate both the state x_k and unknown elements of R and Q matrices subject to certain restrictions on the number of elements that can be estimated as stated through the following assumptions.

B. Assumptions

The following assumptions are made.

Assumption 1 *The matrices forming the transformed state transition matrix, i.e., F_{11}, F_{21}, and F_{22}, and the observation matrix H_1 are assumed to be completely known. The measurement sequence y_k is also assumed to be accessible.*

Assumption 2 *It is assumed that the matrices F and H, the process noise covariance matrix $Q_k = Q \succ 0$, and the measurement covariance matrix $R_k = R \succ 0$ are constant with time.*

Assumption 3 *The pair $(\tilde{F}, \tilde{H})$ is assumed completely uniformly detectable, i.e., the pair (F_{11}, H_1) is uniformly observable and the matrix F_{22} has all its eigenvalues inside the unit circle in the complex plane. The pair $(F, Q^{\frac{1}{2}})$ is assumed to be stabilizable.*

Remark 1 *The preceding assumption ensures that the baseline Kalman filter given in (5) converges to a steady state [30].*

Assumption 4 *The pair* $(F, Q^{\frac{1}{2}})$ *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 for the algebraic Riccati equation [10, 31].*

There is an additional assumption that is required for our proposed result to hold which is stated in the sequel.

III. Filter Derivation

The filter is derived in 3 subsections, namely, formulating modified stacked measurement model, forming a linear strictly stationary time series and finally, estimating the covariance to calculate the unknown elements.

A. Modified measurement model

For the n-dimensional discrete-time stochastic linear system given by (1), the observability matrix is defined by

$$O = \begin{bmatrix} HF^{n-1} \\ HF^{n-2} \\ \vdots \\ HF \\ H \end{bmatrix} \tag{6}$$

Since the (F, H) pair is detectable, it follows from assumption 3 that O is column-rank deficient. However, since every system given by (1) can be transformed into one given by (2), the observability matrix for the transformed system is given by

$$O = \begin{bmatrix} H_1 F_{11}^{n-1} & \mathbf{0}_{p\times s} \\ \vdots & \vdots \\ H_1 F_{11} & \mathbf{0}_{p\times s} \\ H_1 & \mathbf{0}_{p\times s} \end{bmatrix} = \begin{bmatrix} O_1 & \mathbf{0}_{np\times s} \end{bmatrix} \tag{7}$$

where the O_1 corresponds to the observability matrix of the pair (F_{11}, H_1). In this paper, $\mathbf{0}_{a\times b}$ is matrix of zeros with size $a \times b$ for some positive integers a and b. Without loss of generality, we can state that

$$\exists\, m \in \mathbb{N},\ 1 \leq m \leq n : M_o = \begin{bmatrix} H_1 F_{11}^{m-1} \\ \vdots \\ H_1 F_{11} \\ H_1 \end{bmatrix} : \ rank(M_o) = l \tag{8}$$

wherein $M_o \in \mathbb{R}^{mp\times l}$. This follows from the definition of detectability for LTI systems. The value of m corresponds to the buffer size, i.e., the number of past measurements that are stored in memory at each time step. In order to minimize the memory storage, a smallest such m which satisfies the rank condition. However, one may choose a larger value of m so that the matrix M_0 is well conditioned. Using this result, a stacked measurement model is formulated by stacking precisely m measurements,

$$z_{k+1} = F z_k + w_k \tag{9}$$

$$z_{k+2} = F^2 z_k + F w_k + w_{k+1} \tag{10}$$

$$z_{k+i} = F^i z_k + \sum_{j=0}^{j=i-1} F^j w_{k+i-1-j} \tag{11}$$

Now, the measurement equations for the corresponding times are given by

$$y_k = H z_k + v_k \tag{12}$$

$$y_{k+1} = HF z_k + H w_k + v_{k+1} \tag{13}$$

$$y_{k+2} = HF^2 z_k + HF w_k + H w_{k+1} + v_{k+2} \tag{14}$$

$$y_{k+i} = HF^i z_k + \sum_{j=0}^{j=i-1} HF^{k+i-1-j} w_{k+j} + v_{k+i} \tag{15}$$

wherein $i = 0, 1, \ldots, m-1$. Coagulating all the equations and the measurements for m time steps, we get

$$\underbrace{\begin{bmatrix} y_{k+m-1} \\ y_{k+m-2} \\ \vdots \\ y_k \end{bmatrix}}_{\triangleq \mathcal{Y}_k} = \underbrace{\begin{bmatrix} HF^{m-1} \\ HF^{m-2} \\ \vdots \\ H \end{bmatrix}}_{O} z_k + \begin{bmatrix} H & HF & HF^2 & \cdots & HF^{m-2} \\ \mathbf{0}_{p\times n} & H & HF & \cdots & HF^{m-3} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & H \\ \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} \end{bmatrix} \begin{bmatrix} w_{k+m-2} \\ w_{k+m-3} \\ \vdots \\ w_k \end{bmatrix} + \underbrace{\begin{bmatrix} v_{k+m-1} \\ v_{k+m-2} \\ \vdots \\ v_k \end{bmatrix}}_{\triangleq V_k} \tag{16}$$

Rewriting (16) using the detectability definition and eliminating the unobservable states, we get

$$\mathcal{Y}_k = \underbrace{\begin{bmatrix} H_1F_{11}^{m-1} & \mathbf{0}_{p\times s} \\ H_1F_{11}^{m-2} & \mathbf{0}_{p\times s} \\ \vdots & \mathbf{0}_{p\times s} \\ H_1 & \mathbf{0}_{p\times s} \end{bmatrix}}_{O} \begin{bmatrix} z_k^O \\ z_k^{UO} \end{bmatrix} + \begin{bmatrix} \begin{bmatrix} H_1 & \mathbf{0}_{p\times s} \end{bmatrix} & \begin{bmatrix} H_1F_{11} & \mathbf{0}_{p\times s} \end{bmatrix} & \cdots & \begin{bmatrix} H_1F_{11}^{m-2} & \mathbf{0}_{p\times s} \end{bmatrix} \\ \mathbf{0}_{p\times n} & \begin{bmatrix} H_1 & \mathbf{0}_{p\times s} \end{bmatrix} & \cdots & \begin{bmatrix} H_1F_{11}^{m-3} & \mathbf{0}_{p\times s} \end{bmatrix} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \begin{bmatrix} H_1 & \mathbf{0}_{p\times s} \end{bmatrix} \\ \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} \end{bmatrix} \begin{bmatrix} w_{k+m-2}^O \\ w_{k+m-2}^{UO} \\ w_{k+m-3}^O \\ \vdots \\ w_k^{UO} \end{bmatrix} + V_k \tag{17}$$

$$\mathcal{Y}_k = M_o z_k^O + \underbrace{\begin{bmatrix} H_1 & H_1F_{11} & \cdots & H_1F_{11}^{m-2} \\ \mathbf{0}_{p\times l} & H_1 & \cdots & H_1F_{11}^{m-3} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{p\times l} & \mathbf{0}_{p\times l} & \mathbf{0}_{p\times l} & H_1 \\ \mathbf{0}_{p\times l} & \mathbf{0}_{p\times l} & \mathbf{0}_{p\times l} & \mathbf{0}_{p\times l} \end{bmatrix}}_{\triangleq M_w} \underbrace{\begin{bmatrix} w_{k+m-2}^O \\ w_{k+m-3}^O \\ \vdots \\ w_k^O \end{bmatrix}}_{\triangleq W_k^O} + V_k. \tag{18}$$

Writing the modified measurement equation at the next time step, we get

$$\mathcal{Y}_k = M_o z_k^O + M_w W_k^O + V_k \tag{19}$$

$$\mathcal{Y}_{k+1} = M_o z_{k+1}^O + M_w W_{k+1}^O + V_{k+1} \tag{20}$$

B. Formulating linear stationary time series

Since, the system is detectable, the matrix M_o is full column rank and hence its pseudo-inverse is unique and is defined by $M_o^\dagger = (M_o^T M_o)^{-1}(M)_o^T$. Projecting (19) and (20) onto the state space, we get

$$M_o^\dagger \mathcal{Y}_k = z_k^O + M_o^\dagger M_w W_k^O + M_o^\dagger V_k \tag{21}$$

$$M_o^\dagger \mathcal{Y}_{k+1} = z_{k+1}^O + M_o^\dagger M_w W_{k+1}^O + M_o^\dagger V_{k+1} \tag{22}$$

Eliminating the observable state z_k^O by subtracting (21) from (22), we form a time series given by

$$\underbrace{M_o^\dagger \mathcal{Y}_{k+1} - F_{11} M_o^\dagger \mathcal{Y}_k}_{\triangleq \mathcal{Z}_k} = \underbrace{w_k^O + M_o^\dagger M_w W_{k+1}^O - F_{11} M_o^\dagger M_w W_k^O}_{\triangleq \mathcal{W}_k} + \underbrace{M_o^\dagger V_{k+1} - F_{11} M_o^\dagger V_k}_{\triangleq \mathcal{V}_k} \tag{23}$$

wherein $\mathcal{Z}_k$, $\mathcal{W}_k$ and $\mathcal{V}_k$ are concatenated of the measurement sequence, the process noise and the measurement noise respectively at different times. Although there is an abuse of notation in using the subscript k here, the actual time histories of the noise and measurement sequences is clearly specified in (16).

We will now write the $\mathcal{W}_k$ and $\mathcal{V}_k$ as a co-efficient matrix multiplied by a concatenated vector of noise terms

$$\mathcal{W}_k = w_k^O + M_o^\dagger M_w W_{k+1}^O - F_{11} M_o^\dagger M_w W_k^O \tag{24}$$

$$\mathcal{W}_k = \underbrace{\left[\begin{bmatrix} M_o^\dagger M_w & \mathbb{I}_{l\times l} \end{bmatrix} - \begin{bmatrix} \mathbf{0}_{l\times l} & F_{11} M_o^\dagger M_w \end{bmatrix} \right]}_{\triangleq \mathcal{A}} \begin{bmatrix} w_{k+m-1}^O \\ w_{k+m-2}^O \\ \vdots \\ w_k^O \end{bmatrix} \tag{25}$$

$$\text{Let } \mathcal{A} = \begin{bmatrix} A_m & A_{m-1} & \cdots & A_1 \end{bmatrix}, \tag{26}$$

wherein, the matrix $\mathbb{I}_{l\times l}$ is the identity matrix of dimension l. The sequence $\mathcal{W}_k$ and its covariance can be written as follows.

$$\mathcal{W}_k = A_1 w_k + A_2 w_{k+1} + \cdots + A_m w_{k+m-1} \tag{27}$$

The sequence $\mathcal{V}_k$ and its covariance can be expressed in a similar way,

$$\mathcal{V}_k = M_o^{\dagger} V_{k+1} - F_{11} M_o^{\dagger} V_k \tag{28}$$

$$\mathcal{V}_k = \underbrace{\left[\begin{bmatrix} M_o^{\dagger} & \mathbf{0}_{l\times p} \end{bmatrix} - \begin{bmatrix} \mathbf{0}_{l\times p} & F_{11} M_o^{\dagger} \end{bmatrix} \right]}_{\triangleq \mathcal{B} \ \in \mathbb{R}^{l\times(m+1)p}} \begin{bmatrix} v_{k+m} \\ v_{k+m-1} \\ \vdots \\ v_k \end{bmatrix} \tag{29}$$

$$\text{Let } \mathcal{B} = \begin{bmatrix} B_m & B_{m-1} & \cdots & B_0 \end{bmatrix} \tag{30}$$

$$\mathcal{V}_k = B_0 v_k + B_1 v_{k+1} + \cdots + B_m v_{k+m} \tag{31}$$

C. Estimating the covariance matrices

Since the white Gaussian noise sequences w_k^O and v_k are both independent and identically distributed (i.i.d.), the sequence $\mathcal{Z}_k$, which is a function of the measurements, is a zero mean strictly stationary time series with these noise terms as inputs. Therefore, the covariance of $\mathcal{Z}_k$ is given by

$$Cov(\mathcal{Z}_k) = Cov(\mathcal{W}_k) + Cov(\mathcal{V}_k). \tag{32}$$

Writing down the expressions for the covariances in (32) using (27) and (31), we get

$$Cov(\mathcal{W}_k) = A_1 Q^O A_1^T + \cdots A_m Q^O A_m^T \tag{33}$$

$$Cov(\mathcal{V}_k) = B_0 R B_0^T + \cdots B_m R B_m^T \tag{34}$$

Note that the covariance matrices are constant for all time. Since $\mathcal{Z}_k$ is a function of the measurements, its covariance can be estimated using the following unbiased estimator Λ_k.

$$\Lambda_k = \frac{1}{k} \sum_{i=1}^{k} \mathcal{Z}_i \mathcal{Z}_i^T \tag{35}$$

In order to calculate the covariance recursively, the following recursive estimator is used with $\Lambda_0 = 0$,

$$\Lambda_k = \frac{k-1}{k} \Lambda_{k-1} + \frac{1}{k} \mathcal{Z}_k \mathcal{Z}_k^T \tag{36}$$

Note that first m measurements are used to get the first estimate of the covariance matrices. In order to estimate the unknown elements of the covariance matrices, $Cov(\mathcal{Z}_k)$ is replaced with its estimator Λ_k. The entire equation is then vectorized and the right hand side is split into a known part Θ_{known} and a product of matrix S and a vector of concatenated unknown elements to be estimated $\hat{\theta}_k$.

$$vec(\Lambda_k) = \Theta_{known} + S\hat{\theta}_k \tag{37}$$

The S matrix here is constructed using the matrices A_i and B_i from (27) and (31). We are now ready to state our final assumption.

Assumption 5 *The matrix S used in (37) has full column rank.*

This assumption ensures the estimability of the unknown elements in the noise covariance matrices Q and R. This is well in line with a similar assumption was made about the number of unknown elements in the Q matrix by Mehra [17]. The implications of this assumptions are discussed in further detail in section V. Hence, the unknown elements are calculated at each time using

$$\hat{\theta}_k = S^{\dagger}(vec(\Lambda_k) - \Theta_{known}) \tag{38}$$

wherein $S^{\dagger} = (S^T S)^{-1} S^T$ is the unique pseudo-inverse. This is a generalization of the case handled in our previous work [27]. It is clear that the estimability of the unknown elements depends on whether or not S is full column rank. However, since S is a constant matrix depending only on the state transition matrix F and the observation matrix H, it can be precalculated and the bounds on the number of unknown elements or the conditions for estimability of the covariance matrices can be checked. This analysis is performed in section V.

It is noteworthy that this algorithm can incorporate linear constraints between the elements of the unknown matrix. For example, if two of the elements of the unknown covariance matrix are known to be equal even though their value is unknown, the two unknowns can be denoted by the same variable while splitting the equation into known and unknown parts as in (37).

The special cases when only one of R or Q^O is unknown is given by

$$\Lambda_k - Cov(\mathcal{W}_k) = B_0 \hat{R}_k B_0^T + \cdots B_m \hat{R}_k B_m^T. \tag{39}$$

Vectorizing the above equation (denoted $vec(\cdot)$), let $vec(\Lambda_k - Cov(\mathcal{W}_k)) \triangleq U_k^r$, we get

$$U_k^r = vec(B_0 \hat{R}_k B_0^T + \cdots B_m \hat{R}_k B_m^T) \tag{40}$$

$$U_k^r = \underbrace{(B_0 \otimes B_0 + \cdots B_m \otimes B_m)}_{\triangleq T_1} vec(\hat{R}_k) \tag{41}$$

$$vec(\hat{R}_k) = T_1^{\dagger} U_k^r \tag{42}$$

wherein, '$\otimes$' is the Kronecker product. We assume that the matrix T_1 has full column rank.

For unknown Q^O case, we formulate a similar equation,

$$\Lambda_k - Cov(\mathcal{V}_k) = A_1 \hat{Q}_k^O A_1^T + \cdots A_m \hat{Q}_k^O A_m^T. \tag{43}$$

Let $vec(\Lambda_k - Cov(\mathcal{V}_k)) \triangleq U_k^q$. Hence, vectorizing the above equation, we get

$$U_k^q = vec(A_1 \hat{Q}_k^O A_1^T + \cdots A_m \hat{Q}_k^O A_m^T) \tag{44}$$

$$U_k^q = \underbrace{(A_1 \otimes A_1 + \cdots A_m \otimes A_m)}_{\triangleq T_2} vec(\hat{Q}_k^O) \tag{45}$$

Assuming that T_2 has full column rank, the estimate $\hat{Q}_k^O$ is given by

$$vec(\hat{Q}_k^O) = T^{\dagger} vec(U_k^q). \tag{46}$$

D. Algorithm outline

The pseudo-code of the algorithm is summarized in Algorithm 1. Note that since the covariance matrices estimated using the measurements use the $vec(\cdot)$ operation, they may not positive definite due to the randomness in the measurements. A simple condition check is embedded into the algorithm which ensures positive definiteness of the estimated covariance matrices.

IV. Convergence Analysis for noise covariance matrices

This section analyzes the stability of the algorithm under the stated assumptions. The convergence of the covariance estimates to their true values are investigated first. Then, the stability of the state error covariance matrix sequences is evaluated and compared to the state error covariance of the baseline Kalman filter.

Algorithm 1 The adaptive Kalman filter

Initialization: F_{11}, F_{21}, F_{22}, H_1, $\hat{z}_0^O$, P_0, $Cov(\mathcal{Z})_0 = 0$
Input: measurement sequence $\{y_k\}$
Output: state estimate $\{\hat{z}_k^O\}$, $\{\hat{Q}_k^O\}$, $\{\hat{R}_k\}$
for $k = 1$ **to** n **do**
  Using $\{y_k\}$ calculate $\mathcal{Y}_k$ $\{(16)\}$
  Using $\mathcal{Y}_k$ calculate $\mathcal{Z}_k\{(23)\}$
  Using $\mathcal{Z}_k$ calculate $\Lambda_k\{(36)\}$
  Using Λ_k calculate $\hat{R}_k$ and $\hat{Q}_k^O\{(38)\}$
  if $\hat{R}_k \leq 0$ **then**
    $\hat{R}_k = \hat{R}_{k-1}$
  end if
  if $\hat{Q}_k^O \leq 0$ **then**
    $\hat{Q}_k^O = \hat{Q}_{k-1}^O$
  end if
  $\hat{z}_{k|k-1}^O = F_{11}\hat{z}_{k-1|k-1}^O$
  $\hat{z}_{k|k}^O = \hat{z}_{k|k-1}^O + K_k(y_k - H_1\hat{z}_{k|k-1}^O)$
  $P_{k|k-1} = F_{11}P_{k-1|k-1}F_{11}^T + \hat{Q}_k^O$
  $K_k = P_{k|k-1}H_1^T(H_1P_{k|k-1}H_1^T + \hat{R}_k)^{-1}$
  $P_{k|k} = P_{k|k-1} - K_kH_1P_{k|k-1}$
end for

A. Convergence of noise covariance

Substituting the values of $\mathcal{W}_k$ and $\mathcal{V}_k$ using (27) and (31), we get,

$$\mathcal{Z}_k = \sum_{i=1}^{m} A_i w_{k+i-1}^O + \sum_{i=0}^{m} B_i v_{k+i} \tag{47}$$

The above equation is a linear strictly stationary time series because of the zero mean, white Gaussian, and uncorrelated noise assumptions in place. Consider the autocovariance function of zero mean time series $\mathcal{Z}_k$ given by

$$C(k, k+\tau) \coloneqq E[\mathcal{Z}_k\mathcal{Z}_{k+\tau}^T]. \tag{48}$$

If $|\tau| > m$, $C(k, k+\tau) = 0$. Hence, the autocovariance function $C(k_1, k_2)$ decays to 0 as k_1 and k_2 grow farther away from each other. Hence, the central limit theorem for linear stationary time series applies here which uses the weak law of large numbers [32]. This theorem ensures an element-wise convergence given by

$$\sqrt{k}\{\hat{C}(k,k) - C(k,k)\}_{ij} \xrightarrow{D} \mathcal{N}(\mathbf{0}, \Omega_{ij}) \tag{49}$$

wherein, $\hat{C}(k,k)$ is equal to the Λ_k calculated recursively in (36), the subscript ij denotes the element corresponding to the i^{th} row and j^{th} column of the matrix and the ^{D} signifies convergence in distribution. This result motivates the following convergence,

$$\forall \epsilon > 0,\ Pr(|\{\hat{C}(k,k) - C(k,k)\}_{ij}| > \epsilon) \stackrel{k\to\infty}{\longrightarrow} 0 \tag{50}$$

The rate of convergence is directly proportional to $k^{-\frac{1}{2}}$ for all i and j. Using (38) and the existence of the pseudo-inverse for S matrix, we conclude that

$$\forall \epsilon > 0,\ Pr(|\{\hat{\theta}_k - \theta\}_i| > \epsilon) \stackrel{k\to\infty}{\longrightarrow} 0 \tag{51}$$

wherein θ is the vector of true values of the unknown elements. Hence, the convergence of the covariance matrices is given by

$$\forall \epsilon > 0,\ Pr(|\{\hat{Q}^O_k - Q^O\}_{ij}| > \epsilon) \stackrel{k\to\infty}{\longrightarrow} 0, \tag{52}$$

$$\forall \epsilon > 0,\ Pr(|\{\hat{R}_k - R\}_{ij}| > \epsilon) \stackrel{k\to\infty}{\longrightarrow} 0 \tag{53}$$

where i and j correspond to the unknown elements of the covariance matrix. A similar convergence holds for autocovariance function with a lag $\tau \neq 0$,

$$\forall \epsilon > 0,\ Pr(|\{\hat{C}(k,k+\tau) - C(k,k+\tau)\}_{ij} > \epsilon) \stackrel{k\to\infty}{\longrightarrow} 0 \tag{54}$$

wherein, $\hat{C}(k,k+\tau) = \frac{1}{k}\sum_{i=1}^{k} \mathcal{Z}_i\mathcal{Z}^T_{i+\tau}$. Using additional autocovariance functions with non-zero τ augment and improve the ability to estimate covariance matrices. However, more measurements will have to be stored in the memory which might not be desirable. A trade off between the accuracy and memory may give the best performance. It is important to note that the positive definiteness checks in the algorithm do not affect the convergence of the covariance estimates. Retaining the previous estimate when positive definiteness is violated affects state estimation and the state error covariance matrix. However, since the autocovariance estimate $\hat{C}(k,k)$ is independent of the state estimate and is only dependent on the measurements, the convergence of $\hat{C}(k,k)$ to its true value is guaranteed. This causes the covariance estimate to converge to its true value. Hence, our result holds regardless of the checks.

B. Convergence of the state error covariance

Three different error covariance matrix sequences are compared in this subsection. The matrix $\hat{P}_k$ is the one-step predictor error covariance of the filter which is calculated by propagating the initial covariance matrix P_0 using the estimated Kalman gain $\hat{K}_k$ and the estimated $\hat{Q}^O_k$ and $\hat{R}_k$ covariance matrices. The P_k matrix is the true error covariance of the filter which propagates the covariance using the estimated Kalman gain $\hat{K}_k$ and the true Q^O and R. Finally, $P_{k,opt}$ is the optimal error covariance matrix of the baseline Kalman filter given full knowledge of the noise statistics. Writing these matrix sequences down, we get the following

$$\hat{P}_{k+1} = \hat{\bar{F}}_k \hat{P}_k \hat{\bar{F}}_k^T + \hat{K}_k \hat{R}_k \hat{K}_k^T + \hat{Q}_k^O \tag{55}$$

$$P_{k+1} = \hat{\bar{F}}_k P_k \hat{\bar{F}}_k^T + \hat{K}_k R \hat{K}_k^T + Q^O \tag{56}$$

$$P_{k+1,opt} = \bar{F}_k P_{k,opt} \bar{F}_k^T + K_k R K_k^T + Q^O \tag{57}$$

$$\text{wherein, } \hat{K}_k = F_{11} \hat{P}_k H_1^T (H_1 \hat{P}_k H_1^T + \hat{R}_k)^{-1}, \tag{58}$$

$$K_k = F_{11} P_{k,opt} H_1^T (H_1 P_{k,opt} H_1^T + R)^{-1}, \tag{59}$$

$$\hat{\bar{F}}_k = F_{11} - \hat{K}_k H_1, \ and \ \bar{F}_k = F_{11} - K_k H_1 \tag{60}$$

and the initial error covariances are equal, $\hat{P}_0 = P_0 = P_{0,opt}$. Subtracting (56) from (55), the asymptotics of the matrix sequence $\hat{P}_k - P_k$ can be analyzed by

$$\hat{P}_{k+1} - P_{k+1} = \hat{\bar{F}}_k (\hat{P}_k - P_k) \hat{\bar{F}}_k^T + \hat{K}_k (\hat{R}_k - R) \hat{K}_k^T + (\hat{Q}_k^O - Q^O) \tag{61}$$

Since the initial error covariance is the same, the matrix sequence can be expanded as

$$\hat{P}_{k+1} - P_{k+1} = \sum_{i=0}^{k} \{ \hat{\phi}_i \hat{K}_i (\hat{R}_i - R) \hat{K}_i^T \hat{\phi}_i^T + \hat{\phi}_i (\hat{Q}_i^O - Q^O) \hat{\phi}_i^T \} \tag{62}$$

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 defined as

$$\Delta P_{m,n} = (\hat{P}_{n+1} - P_{n+1}) - (\hat{P}_m - P_m) \tag{63}$$

$$\Delta P_{m,n} = \sum_{i=m}^{n} \{ \hat{\phi}_i \hat{K}_i (\hat{R}_i - R) \hat{K}_i^T \hat{\phi}_i^T + \hat{\phi}_i (\hat{Q}_i^O - Q^O) \hat{\phi}_i^T \} \tag{64}$$

Each element of $\Delta P_{m,n}$ is a function of the elements of $\hat{R}_i - R$ and $\hat{Q}_i^O - Q^O$. Since, $\hat{\phi}_i$ and $\hat{K}_i$ are bounded from above, there exist ϵ_{ij}^r, ϵ_{ij}^q, δ_{ij}^r, δ_{ij}^q and corresponding N_{ij}^r and N_{ij}^q such that

$$\forall k > N_{ij}^q, Pr(|\{\hat{Q}_k^O - Q^O\}_{ij}| < \epsilon_{ij}^q) > 1 - \delta_{ij}^q \tag{65}$$

$$\forall k > N_{ij}^r, Pr(|\{\hat{R}_k - R\}_{ij}| < \epsilon_{ij}^r) > 1 - \delta_{ij}^r \tag{66}$$

$$\forall m, n > \max_{i,j}(N_{ij}^q, N_{ij}^r), \quad Pr(|\{\Delta P_{m,n}\}_{ij}| < \epsilon_{ij}^p) > 1 - \delta_{ij}^p \tag{67}$$

Here each of ϵ^p_{ij} and δ^p_{ij} are functions of all the ϵ^q_{ij} and ϵ^r_{ij}, and δ^q_{ij} and δ^r_{ij} respectively for all i and j. Using (67) we can conclude that

$$\forall i, j\ \forall \epsilon^p_{ij} > 0\ \lim_{m,n\to\infty} Pr(|\{\Delta P_{m,n}\}_{ij}| > \epsilon^p_{ij}) \longrightarrow 0 \tag{68}$$

Hence, using the Cauchy criterion for random sequences [33], we conclude that

$$\forall i, j\ \forall \epsilon^p_{ij} > 0\ \lim_{k\to\infty} Pr(|\{\hat{P}_k - P_k\}_{ij}| > \epsilon^p_{ij}) \longrightarrow 0 \tag{69}$$

Hence, we further infer that

$$\hat{P}_k - P_k \xrightarrow{P} \mathbf{0}_{l\times l} \tag{70}$$

Now consider the matrices defined in (58) and (60). The gain matrix $\hat{K}_k$ which is a continuous function of $\hat{P}_k$, $\hat{Q}^O_k$, and $\hat{R}_k$. Hence using the continuous mapping theorem [33], we get

$$\hat{K}_k - K_k \xrightarrow{P} \mathbf{0}_{l\times p} \tag{71}$$

$$\hat{K}_k - K_k \xrightarrow{P} \mathbf{0}_{l\times p} \implies \hat{\bar{F}}_k - \bar{F}_k \xrightarrow{P} \mathbf{0}_{l\times l} \tag{72}$$

Hence using the above results, we conclude that

$$P_k - P_{k,opt} \xrightarrow{P} \mathbf{0}_{l\times l} \tag{73}$$

$$\hat{P}_k - P_{k,opt} \xrightarrow{P} \mathbf{0}_{l\times l} \tag{74}$$

V. Estimability of noise covariance matrices

In (38), the existence of a pseudo-inverse of S matrix was assumed so that the unknown elements of Q^O and R matrix could be estimated. The S matrix can be precalculated and the estimability of the unknown elements can be checked before the filter is deployed. However, a physical insight behind this assumption is uncovered via mathematical analysis in this section.

A. Estimability of R matrix

Say that the entire R is unknown and needs to be estimated while Q^O matrix is known. Consider the case when the system given by (1) has linearly dependent measurements. This is mathematically expressed as

$$\exists\, \xi \neq \mathbf{0}_{p\times 1} \;:\; \xi^T H_1 = \mathbf{0}_{1\times l} \tag{75}$$

$$\text{Thus,} \quad \begin{bmatrix} \xi^T & \dots & \xi^T \end{bmatrix}_{1\times mp} M_o = \mathbf{0}_{1\times l} \implies M_o^\dagger \begin{bmatrix} \xi^T & \dots & \xi^T \end{bmatrix}^T_{mp\times 1} = \mathbf{0}_{l\times 1} \tag{76}$$

Hence, using the matrices defined in (30), the above condition translates to

$$B_i \xi = \mathbf{0}_{l\times 1}, \forall\, i = 0, 1, \dots, m \implies \sum_{i=0}^{m} (B_i \xi \xi^T B_i^T) = \mathbf{0}_{l\times l} \tag{77}$$

$$\underbrace{(B_0 \otimes B_0 + \cdots B_m \otimes B_m)}_{T_1} vec(\xi\xi^T) = \mathbf{0}_{l^2\times 1} \tag{78}$$

Hence, we found a common null space for all the coefficient matrices B_i for $i = 1, 2, \dots, m$. Therefore, the matrix $vec(\xi\xi^T)$ belongs to the null space of the matrix T_1 defined in (41). In case of repeated or linearly dependent measurement, estimation of the R matrix is ambiguous. It is important to note that this unintentionally establishes a hard limit on the number of measurements available to the system. If the number of measurements is greater than the number of states, there always exists a ξ which satisfies the above condition and the R matrix cannot be unambiguously estimated. Intuitively, for example, say that the system has two identical sensors, This algorithm is unable to estimate the cross-covariance between the two noises of the identical sensors (with possibly different noise covariance matrices). The measurement equation can be modified by averaging out the linearly dependent measurements and estimating their aggregate covariance matrix. However, note that this is not a sufficient condition for estimability of R. There can be other cases which make $rank(T_1) = 0$.

B. Estimability of Q^O matrix

For estimability of the Q^O matrix, it has been established that the number of unknown elements that need be estimated cannot be more than $l \times p$ where p is the number of independent measurements [17].

Again let us assume that the entire Q^O matrix is to be estimated while R matrix is completely known. Let the matrix $X = (M_o^T M_o)^{-1}$ and $M_o^\dagger$ is stated as

$$M_o^\dagger = \begin{bmatrix} X(F_{11}^T)^{m-1} H_1^T & X(F_{11}^T)^{m-2} H_1^T & \dots & X H_1^T \end{bmatrix} \tag{79}$$

Note that X is an invertible matrix. Similar to the case for unknown R matrix, the idea is to find a null space common to all the matrix $A_i,\ i = 1, 2, \ldots, m$. First, the expression for $M_o^\dagger$ can be expressed as

$$M_o^\dagger = \begin{bmatrix} C_1 & C_2 & \ldots & C_m \end{bmatrix} \tag{80}$$

$$C_i = X(F_{11}^T)^{m-i} H_1^T \quad i = 1, 2, \ldots, m \tag{81}$$

Using the above expression, the matrix $M_o^\dagger M_w$ is evaluated as

$$M_o^\dagger M_w = \begin{bmatrix} D_1 & D_2 & \ldots & D_m \end{bmatrix} \tag{82}$$

$$D_1 = C_1 H_1 \tag{83}$$

$$D_2 = C_1 H_1 F_{11} + C_2 H_1 \tag{84}$$

$$\vdots$$

$$D_m = C_1 H_1 F_{11}^{m-2} + C_2 H_1 F_{11}^{m-3} + \ldots + C_{m-1} H_1 \tag{85}$$

Evaluating $\mathcal{A}$ matrix using the expression above, we get

$$\mathcal{A} = \begin{bmatrix} \begin{bmatrix} M_o^\dagger M_w & \mathbb{I} \end{bmatrix} - \begin{bmatrix} \mathbf{0}_{l\times l} & F_{11} M_o^\dagger M_w \end{bmatrix} \end{bmatrix} \tag{86}$$

$$A_m = D_1 = C_1 H_1 \tag{87}$$

$$\vdots$$

$$A_2 = D_m - F_{11} D_{m-1} \tag{88}$$

$$A_1 = \mathbb{I} - F_{11} D_m \tag{89}$$

Note that all matrices $A_i \in \mathbb{R}^{l\times l}$ for $i = 1, 2, \ldots, m$. However, due to the matrix multiplication $H_1^T H_1$, their rank cannot be more than p with an exception of A_1. The estimability of Q^O thus directly depends on $rank(A_1 \otimes A_1 + \ldots A_m \otimes A_m)$ and one of the cases in which this rank is 0 is if the null spaces of the matrices $A_i,\ i = 1, 2, \ldots, m$ intersect. Let us say this is the case and there exists a non-zero vector $\kappa \in \mathbb{R}^{l\times 1}$ such that $A_i \kappa = \mathbf{0}_{l\times 1}$ for $i = 1, 2, \ldots, m$. Trying to find an expression for κ, we get

$$\forall i = 1, 2, \ldots, m \quad A_i \kappa = \mathbf{0}_{l\times 1} \implies D_i \kappa = \mathbf{0}_{l\times 1} \tag{90}$$

However, note that in that case, $A_1 \kappa \neq \mathbf{0}_{l\times 1}$. Hence, we have a contradiction and the matrices A_i do not share a part of their null space. However, since this is only a necessary condition, this analysis does not provide a condition for estimability of Q^O matrix.

VI. Further comparison with prior literature

The problem of estimating the state and certain unknown elements of the process and measurement noise covariance matrices has received significant attention in prior literature, most notably in [17]. Several important facts that highlight crucial differences between [17] and our work are stated below.

- In [17], the pair $(F, Q^{\frac{1}{2}})$ is assumed to be controllable and pair (F, H) is taken to be observable. Our work respectively assumes stabilizability and detectability of the same matrix pairs as given in assumptions 3 which are clearly weaker technical hypotheses. This allows our results to be applicable to a wider class of systems.
- Both the optimal (case of known covariance matrices) and suboptimal (case of unknown covariance matrices) are assumed to have reached steady-state conditions in [17]. This assumption is central to the developments therein since it is used to calculate the covariance matrix estimate. We make no such assumption as we prove the steady-state properties for our estimator and guarantee its convergence.
- The state transition matrix F is assumed to be nonsingular in [17] which is arguably a mild restriction. This assumption is subsequently used to calculate the estimate of the noise covariance matrices and hence is crucial to the formulation in [17]. On the other hand, our assumption of detectability is sufficient for convergence and we do not place any additional non-singularity restrictions upon the state transition matrix.
- We note that [10] analyzed discrete-time Kalman filtering with incorrect noise covariances. Corollary 3.3 in [10] states that for incorrect noise covariances obtained simply by multiplying with a positive scalar, the sequences are asymptotically white. This result, as the authors state, shows the insufficiency of the whiteness test used in [17] for estimating the noise covariance matrices that are required for constructing the steady-state filter. For related discussion, we also refer to [18]. This inadequacy of the whiteness condition should be noted to be a major limitation of the approach in [17] for estimating covariance matrices using the autocovariance function in case of sub-optimality of the filter. Our method is impervious to these arguments involving the whiteness tests as we do not use the residual autocovariance properties for our proposed estimation schemes.
- In [17], convergence is proved for the asymptotic case, i.e., when the batch size (N) becomes infinitely large. However, such a filter can be potentially impractical (large memory buffer usage), and moreover, one has to always initiate the filter with a fixed batch size, the chosen batch size may not be large enough to guarantee convergence. In [17], it is assumed that the batch size N is much larger than n. All the results that follow prove convergence of the estimates for a large value of N. In our work, as given in (8), at most n measurements are stacked and the number of stacked measurements is pre-calculated using the system matrices (F and H).

VII. Simulations

We construct a fictitious detectable system which satisfies the assumptions for this algorithm to converge.

$$x_k = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.3 & 0.5 & 0 \\ 0.1 & 0.9 & 0.7 \end{bmatrix} x_{k-1} + w_{k-1} \tag{91}$$

$$y_k = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} x_k + v_k \tag{92}$$

It is assumed that $w_k \sim \mathcal{N}(0, Q)$, and $v_k \sim \mathcal{N}(0, R)$ are both i.i.d white Gaussian noises. We assume

$$R = \begin{bmatrix} 5 & 0 \\ 0 & 4 \end{bmatrix} \text{ and } Q = \begin{bmatrix} 3 & 0.2 & 0 \\ 0.2 & 2 & 0 \\ 0 & 0 & 7.5 \end{bmatrix}$$

The elements R_{11}, Q_{11}, and Q_{22} were to be estimated and all other elements were assumed to be known. The initial estimates for all these elements was chosen to be 10. From Monte Carlo simulation results, the 2 norm of the error between $\hat{R}_k$ and R as well as $\hat{Q}_k^O$ and Q^O is shown to converge Figs. (1) and (2) respectively. The predictor and actual state error covariance converges to the optimal Kalman filter values Fig. (4) as shown in one of the Monte Carlo simulations. The state estimation error and their respective 3σ bounds of predictor state error covariances for the three states are shown in Fig. (3).

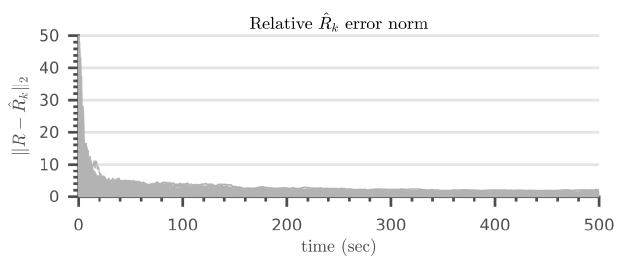

Fig. 1 Estimation error norm of R matrix for 100 Monte Carlo simulations.

VIII. Conclusions

A novel algorithm to adaptively estimate the state and certain unknown elements of the process and measurement noise covariance matrices of a discrete linear time invariant stochastic system is formulated. The algorithm presented here is derived using a judicious combination of established adaptive filtering approaches such as correlation techniques

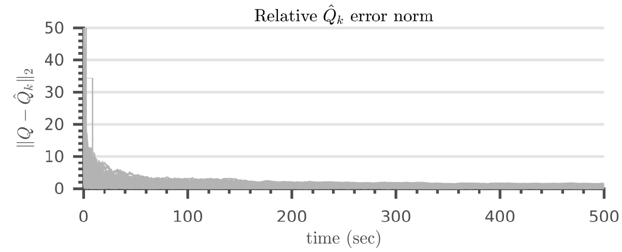

Fig. 2 Estimation error norm of Q matrix for 100 Monte Carlo simulations.

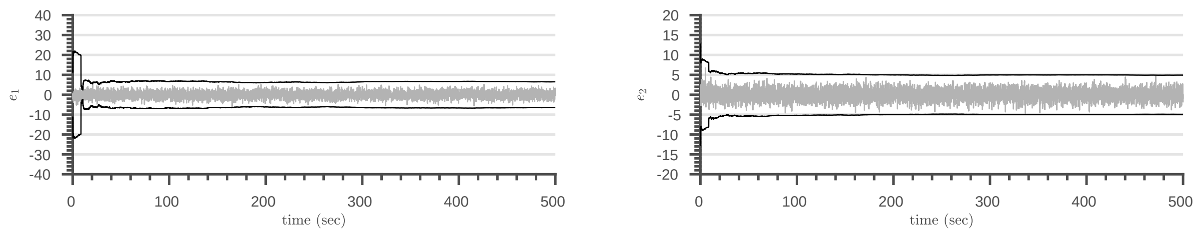

Fig. 3 Estimation errors for the two observable states with the 3σ bounds corresponding to their predictor error covariance $\hat{P}_k$

and covariance matching techniques. The detectability property of the system is utilized for observing the state and formulating a time series containing measurement and process noise sequences independent of the state. A proof for the probabilistic convergence of the new algorithm is presented under additional assumptions of existence of a pseudo inverse used for uniqueness of the estimates of the covariance matrices. Our work bears significant contrasts to approaches for adaptive covariance estimation algorithms reported in existing literature. Firstly, we note that our technical assumptions of detectability rather than observability, no non-singularity requirements on the state transition and observation matrix like invertibility, and no requirement of reaching steady-state are less restrictive. Secondly, our proposed algorithm is independent of suboptimality tests for whiteness toward constructing the state-state filter. Lastly, our results do not require arbitrarily large batch sizes for ensuring convergence but rather, the memory usage and buffer size demands imposed by our algorithm can be a priori benchmarked in terms of the dimensionality of the state-space. Monte Carlo simulations demonstrate the effectiveness of our algorithm. Further generalizations of the proposed algorithm to the continuous-time domain and applications for linear time-varying systems represent fruitful directions for further research in this field.

IX. Acknowledgment

This work was sponsored in part by the NASA Johnson Space Center under contract number NNX17AI35A.

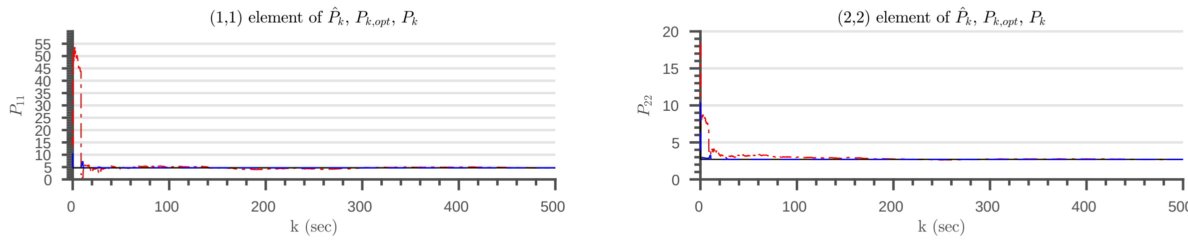

Fig. 4 Comparison of the three different error covariances from (55) (red), (56) (blue) and (57) (black).

References

[1] Kalman, R. E., “A new approach to linear filtering and prediction problems,” *Transactions of the ASME, Journal of Basic Engineering*, Vol. 82, No. 1, 1960, pp. 35–45.

[2] Kalman, R. E., and Bucy, R. S., “New results in linear filtering and prediction theory,” *Transactions of the ASME, Journal of Basic Engineering*, Vol. 83, No. 1, 1961, pp. 95–108.

[3] Jazwinski, A. H., *Stochastic processes and filtering theory*, Courier Corporation, 2007.

[4] Julier, S., Uhlmann, J., and Durrant-Whyte, H. F., “A new method for the nonlinear transformation of means and covariances in filters and estimators,” *IEEE Transactions on Automatic Control*, Vol. 45, No. 3, 2000, pp. 477–482.

[5] Arasaratnam, I., Haykin, S., and Elliott, R. J., “Discrete-time nonlinear filtering algorithms using Gauss–Hermite quadrature,” *Proceedings of the IEEE*, Vol. 95, No. 5, 2007, pp. 953–977.

[6] Arasaratnam, I., and Haykin, S., “Cubature kalman filters,” *IEEE Transactions on Automatic Control*, Vol. 54, No. 6, 2009, pp. 1254–1269.

[7] Gordon, N. J., Salmond, D. J., and Smith, A. F., “Novel approach to nonlinear/non-Gaussian Bayesian state estimation,” *IEE Proceedings F (Radar and Signal Processing)*, Vol. 140, IET, 1993, pp. 107–113.

[8] Doucet, A., Godsill, S., and Andrieu, C., “On sequential Monte Carlo sampling methods for Bayesian filtering,” *Statistics and computing*, Vol. 10, No. 3, 2000, pp. 197–208.

[9] Alspach, D., and Sorenson, H., “Nonlinear Bayesian estimation using Gaussian sum approximations,” *IEEE Transactions on Automatic Control*, Vol. 17, No. 4, 1972, pp. 439–448.

[10] Sangsuk-Iam, S., and Bullock, T. E., “Analysis of discrete-time Kalman filtering under incorrect noise covariances,” *IEEE Transactions on Automatic Control*, Vol. 35, No. 12, 1990, pp. 1304–1309.

[11] Heffes, H., “The effect of erroneous models on the Kalman filter response,” *IEEE Transactions on Automatic Control*, Vol. 11, No. 3, 1966, pp. 541–543.

[12] Madjarov, N., and Mihaylova, L., “Kalman filter sensitivity with respect to parametric noises uncertainty,” *Kybernetika*, Vol. 32, No. 3, 1996, pp. 307–322.

[13] Saab, S. S., “Discrete-time Kalman filter under incorrect noise covariances,” *American Control Conference, Proceedings of the 1995*, Vol. 2, IEEE, 1995, pp. 1152–1156.

[14] Sangsuk-Iam, S., and Bullock, T. E., “Analysis of continuous-time Kalman filtering under incorrect noise covariances,” *Automatica*, Vol. 24, No. 5, 1988, pp. 659–669.

[15] Willems, J. L., and Callier, F. M., “Divergence of the stationary Kalman filter for correct and for incorrect noise variances,” *IMA Journal of Mathematical Control and Information*, Vol. 9, No. 1, 1992, pp. 47–54.

[16] Mehra, R., “Approaches to adaptive filtering,” *IEEE Transactions on Automatic Control*, Vol. 17, No. 5, 1972, pp. 693–698.

[17] Mehra, R., “On the identification of variances and adaptive Kalman filtering,” *IEEE Transactions on Automatic Control*, Vol. 15, No. 2, 1970, pp. 175–184.

[18] Alspach, D., “Comments on" On the identification of variances and adaptive Kalman filtering",” *IEEE Transactions on Automatic Control*, Vol. 17, No. 6, 1972, pp. 843–845.

[19] Wang, H., Deng, Z., Feng, B., Ma, H., and Xia, Y., “An adaptive Kalman filter estimating process noise covariance,” *Neurocomputing*, Vol. 223, 2017, pp. 12–17.

[20] Wang, X., Liu, M.-q., Fan, Z., and Zhang, S.-l., “A novel approach of noise statistics estimate using H∞ filter in target tracking,” *Frontiers of Information Technology & Electronic Engineering*, Vol. 17, No. 5, 2016, pp. 449–457.

[21] Durovic, Z. M., and Kovacevic, B. D., “Robust estimation with unknown noise statistics,” *IEEE Transactions on Automatic Control*, Vol. 44, No. 6, 1999, pp. 1292–1296.

[22] Yang, Y., and Gao, W., “An optimal adaptive Kalman filter,” *Journal of Geodesy*, Vol. 80, No. 4, 2006, pp. 177–183.

[23] Duník, J., Straka, O., Kost, O., and Havlík, J., “Noise covariance matrices in state-space models: A survey and comparison of estimation methods-Part I,” *International Journal of Adaptive Control and Signal Processing*, Vol. 31, No. 11, 2017, pp. 1505–1543.

[24] Feng, B., Fu, M., Ma, H., Xia, Y., and Wang, B., “Kalman filter with recursive covariance estimation-Sequentially estimating process noise covariance,” *IEEE Transactions on Industrial Electronics*, Vol. 61, No. 11, 2014, pp. 6253–6263.

[25] Duník, J., Straka, O., and Kost, O., “Measurement difference autocovariance method for noise covariance matrices estimation,” *Decision and Control (CDC), 2016 IEEE 55th Conference on*, IEEE, 2016, pp. 365–370.

[26] Duník, J., Kost, O., and Straka, O., “Design of measurement difference autocovariance method for estimation of process and measurement noise covariances,” *Automatica*, Vol. 90, 2018, pp. 16–24.

[27] Moghe, R., Zanetti, R., and Akella, M., “Covariance matching Kalman filter for observable LTI systems,” , 2018. Accepted at the 2018 IEEE 57th Annual Conference on Decision and Control (CDC).

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

[29] Anderson, B., and Moore, J., *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ, 1979.

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

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

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

[33] Resnick, S. I., *A probability path*, Springer Science & Business Media, 2013.