A Single-Pass Noise Covariance Estimation Algorithm in Nonswitching Multiple-Model Adaptive Kalman Filters for Nonstationary Systems

HEE-SEUNG KIM
ADAM BIENKOWSKI
LINGYI ZHANG
KRISHNA R. PATTIPATI

This paper presents a single-pass stochastic gradient descent (SGD) algorithm for estimating unknown noise covariances. The proposed algorithm is designed for nonswitching multiple-model adaptive Kalman filters, where the noise covariances can occasionally jump up or down by an unknown magnitude. Compared to our previous batch estimation or multipass decision-directed estimation methods, the proposed algorithm has the advantage of reading measurement data exactly once, leading to a significant improvement in computational efficiency and practicality. Moreover, the algorithm achieves an acceptable level of root mean square error (RMSE) in state estimates, making it suitable for real-time industrial applications. The proposed algorithm utilizes recursive fading memory estimates of the sample cross-correlations of the innovations and employs the root mean square propagation (RMSprop) accelerated SGD algorithm. The combination of these techniques enables the algorithm to achieve high accuracy in estimating the unknown noise covariances while maintaining superior computational efficiency over iterative batch methods. The paper presents a comparative evaluation of the proposed method on several test cases, which demonstrate its effectiveness in terms of computational efficiency and estimation accuracy. Overall, the proposed algorithm offers a promising approach for real-time noise covariance estimation in multiple-model adaptive Kalman filters.

Manuscript received February 5, 2024; revised May 21, 2024; released for publication October 14, 2024

Refereeing of this contribution was handled by Erik Leitinger.

This work was supported in part by the U.S. Office of Naval Research (ONR), in part by the U.S. Naval Research Laboratory (NRL) under Grant N00014-18-1-1238, N00014-21-1-2187 and Grant N00173-16-1-G905. This paper was presented in part at the Twenty-Fourth International Conference on Information Fusion, Sun City, South Africa, November, 2021.

The authors are with Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269, USA, (e-mail: hee-seung.kim@uconn.edu; adam.bienkowski@uconn.edu; lingyi.zhang@uconn.edu; krishna.pattipati@uconn.edu). (Corresponding author: Hee-Seung Kim.)

1557-6418/2024/$17.00

I. INTRODUCTION

The Kalman filter (KF) [13] is a widely used algorithm that provides an optimal minimum mean square error (MMSE) estimate for discrete-time linear dynamic systems under the Gaussian assumption, provided that the mean and covariance are known. When the noises entering a system are non-Gaussian but still have known first and second moments, i.e., mean and covariance, the KF remains the best linear state estimator due to its ability to efficiently incorporate prior knowledge of the system dynamics and noise statistics. As a consequence of numerous research studies, the KF has gained significant interest and attention in many industrial applications, including fault diagnosis, robotics, signal processing, navigation, and target tracking, to name a few [2], [3]. However, in many real-world circumstances, the statistics of noise processes are either completely unknown or partially known.

In order to estimate unknown noise covariance parameters, Zhang *et al*. [32] derived the necessary and sufficient conditions for their identifiability, and then proposed an iterative *batch* optimization algorithm that ensures uncorrelated innovations. The rank of a matrix formed from the cross-correlations of the weighted sum of innovations, where the weights are the coefficients of the minimal polynomial of any closed-loop filter matrix, optimal or suboptimal, is required for the noise covariance identifiability. The innovation sequence of an optimal KF under the Gaussian assumption is orthogonal, indicating that the innovations are strictly white and independent of each other [3]. Zhang *et al*. [32] formulated an objective function using normalized temporal cross-correlations of the innovations based on this attribute to determine the optimal gain and subsequently the innovation (preresidual) and postresidual covariances, the measurement noise covariance and the process noise covariance.

We presented an improved method for estimating noise covariances, employing a sequential mini-batch stochastic gradient descent (SGD) algorithm that requires multiple passes through the data. Moreover, we showed a technique to detect changes in noise covariances when applying this estimation method to nonstationary systems [17]. To avoid multiple passes through the data required by the batch and multipass algorithms, we propose herein a *single-pass* real-time adaptive Kalman filtering approach designed for nonstationary systems. The proposed method is suitable for

scenarios in which process and measurement noise covariances occasionally fluctuate by an unknown magnitude and the system behavior is associated with one of a finite number of known models.

A. Prior Work

The relationship between the covariance of the state estimation error and the innovations in any suboptimal filter is the key to process and measurement noise covariance estimation. This relationship serves as a fundamental building block for correlation-based approaches. Pioneering contributions using this approach were made by [5], [21], [24], [25].

In linear state space models, Sarkka and Nummenmaa [26] presented a variational Bayesian approach for the joint recursive estimate of the dynamic state and measurement noise parameters. The method is implemented by forming separable variance approximations to the joint posterior distribution of state and noise parameters at each time step. This approach, however, does not take variations in process noise into consideration. Because of their mode-seeking behavior, the variational algorithms often converge to local minima and typically need tuning parameters to converge to the correct parameters.

Our sequential mini-batch estimation method [17] enhanced the computational efficiency and accuracy of the batch estimation algorithm in [32] by applying dynamic convergence thresholds and adaptive step size rules. To update the filter gain, we used sequential fading memory mini-batch estimates of the innovation correlations. For nonstationary systems, a change-point detection algorithm described in [15] was used for determining the time points of abrupt changes in unknown noise covariances based on the innovation sequence.

In the multiple-model adaptive estimation method proposed herein, the system is assumed to obey one of a finite number of models, and each model has its own nonswitching dynamics [3]. The overall estimate of the system state is obtained by taking a convex combination of the estimates from multiple parallel filters. The weights used in the convex combination correspond to the posterior model probabilities.

The following are the limits of previous research [17] on noise covariance estimation in nonstationary systems. First, the previous methods are computationally expensive because they require multiple passes through the observation data and are not suitable for online streaming data applications. Second, since the sequential estimation method is used for samples between two consecutive change points, the accuracy of the decision-directed noise covariance estimation method is reliant on the accuracy of the change-point detection algorithm. Third, the previous methods assumed that the structure of the dynamic model was known. In this paper, we relax this assumption and propose a streaming algorithm that extends the method to multiple-model settings.

B. Contribution and Organization of the Paper

We present a single-pass sequential mini-batch noise covariance estimation algorithm suitable for streaming data as an extension of the work in [17], [18] for nonstationary and nonswitching multiple-model systems. Our proposed method enables the estimation of the measurement and process noise covariances without the use of a change-point detection algorithm. We enhance the computational efficiency of the method via a single-pass through the observation data. The only caveat is that jumps are assumed to occur occasionally, and after the filter has reached a steady state, that is, the jumps are infrequent. However, small variations in the noise covariances are allowed in between jumps. More significantly, the structure of the dynamic model is unknown, but is assumed to belong to one of a finite number of known models.

We validate the proposed method on several nonstationary and multiple-model system test cases. In addition, we derive noise covariance identifiability conditions in terms of prefit residual (innovation) correlations as in [32], as well as postfit residual correlations and output correlations; the latter was used in the covariance estimation algorithm in [25]. We also prove the convergence of the iterative algorithm for process noise covariance, which was not established in our prior work.

The paper is organized as follows: In Section 2, we provide an overview of the multiple-model KF. Then, in Section 3, we discuss the identifiability conditions in terms of prefit and postfit residuals and outputs for estimating the unknown noise covariances in each individual model. Section 4 provides approaches for obtaining the unknown covariance parameters in a multiple-model system using the sequential mini-batch SGD method, including the fading memory filter-based correlation estimation, and the SGD update of the Kalman gain. In Section 5, numerical results[1] show the evidence that our method can track unknown noise covariances in nonstationary systems, as well as systems exhibiting dynamics from a finite number of known models, and that the single-pass algorithm is computationally efficient. Lastly, we conclude the paper and discuss potential avenues for future work in Section 6.

II. PLANT AND MEASUREMENT MODEL FOR THE MULTIPLE-MODEL KF

The multiple-model approach assumes that the system obeys one of a finite number of fixed models. Formally, the approach assumes that the linear discrete-time stochastic dynamic system can assume one of J models, $j = 1, 2, ..., J$, given by

$$x(k+1) = F^j x(k) + \Gamma^j v^j(k), \tag{1}$$

[1]Numerical results suggest that the occasional jump assumption may be relaxed in practice.

$$z(k) = H^j x(k) + w^j(k), \tag{2}$$

where $x(k)$ is the n_x-dimensional state vector, $z(k)$ is the n_z-dimensional measurement vector, and j is the candidate model. Here F^j and H^j are the $n_x \times n_x$ state transition matrix and the $n_z \times n_x$ measurement matrix of the system, respectively, and Γ^j is the noise gain matrix. We assume that the process noise $v^j(k)$ and the measurement noise $w^j(k)$ processes are the sequences of zero-mean white Gaussian noises with *unknown* process noise covariance $Q^j(k)$ and *unknown* measurement noise covariance $R^j(k)$, respectively. Note that the initial state error and the two noise processes are assumed to be mutually independent. We assume that $Q^j(k)$ and $R^j(k)$ are piecewise constants such that the filter reaches a steady state between any two jumps and that the jump is of an unknown magnitude.

Given $Q^j(k)$ and $R^j(k)$, the multiple-model adaptive KF involves the consecutive processes of prediction and update given by [3], [12], [13], [20], [27]

$$\hat{x}^j(k+1|k) = F^j \hat{x}^j(k|k), \tag{3}$$

$$v^j(k+1) = z(k+1) - H^j \hat{x}^j(k+1|k), \tag{4}$$

$$\hat{x}^j(k+1|k+1) = \hat{x}^j(k+1|k) + W^j(k+1) v^j(k+1), \tag{5}$$

$$P^j(k+1|k) = F^j P^j(k|k) F^{j\prime} + \Gamma^j Q^j(k) \Gamma^{j\prime}, \tag{6}$$

$$S^j(k+1) = H^j P^j(k+1|k) H^{j\prime} + R^j(k), \tag{7}$$

$$W^j(k+1) = P^j(k+1|k) H^{j\prime} S^j(k+1)^{-1}, \tag{8}$$

$$\begin{aligned} P^j(k+1|k+1) &= (I_{n_x} - W^j(k+1)H^j) P^j(k+1|k) \\ &\quad (I_{n_x} - W^j(k+1)H^j)' + W^j(k+1) R^j(k) W^j(k+1)', \end{aligned} \tag{9}$$

$$\Lambda^j(k) = \frac{1}{\sqrt{|2\pi S^j(k)|}} \exp\left(-\frac{1}{2} v^j(k)' S^j(k)^{-1} v^j(k)\right), \tag{10}$$

$$p^j(k) = \frac{\Lambda^j(k) p^j(k-1)}{\sum_{l=1}^{J} \Lambda^l(k) p^l(k-1)}. \tag{11}$$

The KF predicts the next state estimate at time index $(k+1)$, given the observations up to time index k in (3) and the concomitant predicted state estimation error covariance in (6), using model-specific system dynamics, the updated state error covariance $P^j(k|k)$ at time index k and the process noise covariance, $Q^j(k)$. The updated state estimate at time $(k+1)$ in (5) incorporates the measurement at time $(k+1)$ via the Kalman gain matrix in (8), which depends on the innovation covariance $S^j(k+1)$ (which in turn depends on the measurement noise covariance $R^j(k)$ and the predicted state error covariance $P^j(k+1|k)$). The updated state error covariance $P^j(k+1|k+1)$ is computed via (9). This corresponds to Joseph form in [3], [8], which guarantees that the updated state covariance matrix will remain positive definite.

The mode likelihood function $\Lambda^j(k)$ is computed via (10), which depends on the innovation sequence $v^j(k)$ and innovation covariance $S^j(k)$. In (10), $|\cdot|$ is the determinant, and the determinant of any scalar value times a matrix is equal to the determinant of the matrix times the scalar raised to the dimension of the matrix. This means that $|2\pi S^j(k)| = (2\pi)^{n_z} |S^j(k)|$ for a multidimensional random variable. Note that n_z here is the dimension of the measurement (or innovation) vector. The mode probability $p^j(k)$ corresponding to each candidate model at time index k is computed via (11). Without loss of generality, we assume the initial mode probability $p^j(0) = 1/J$.

III. NECESSARY AND SUFFICIENT CONDITIONS FOR THE IDENTIFIABILITY OF UNKNOWN COVARIANCES

We derive the necessary and sufficient conditions to estimate the unknown covariance matrices in terms of prefit residual (innovation) correlations, as well as postfit residual correlations and output correlations. Note that the identifiability conditions of the multiple-model approach depend on each candidate model since the corresponding KFs are noninteracting.

A. Innovation-Based Identifiability Conditions

Consider model j. Assume that Q^j and R^j are unknown but are piecewise constants such that the filter reaches the steady state before any jump to a new value (in practice, they can vary in between jumps as demonstrated in illustrative examples). Let us define the coefficients of the m^{th} order minimal polynomial of the closed-loop filter matrix $\bar{F}^j$, $\sum_{i=0}^{m} a_i^j (\bar{F}^j)^{m-i} = 0$, $a_0^j = 1$. Now, consider the innovations corresponding to a stable, suboptimal closed-loop filter matrix $\bar{F}^j = F^j(I_{n_x} - W^j H^j)$ given by [29], [32]

$$\begin{aligned} v^j(k) &= H^j (\bar{F}^j)^m \tilde{x}^j(k-m|k-m-1) + H^j \sum_{\ell=0}^{m-1} \Big\{ (\bar{F}^j)^{m-1-\ell} \\ &\quad \times \left[\Gamma^j v^j(k-m+\ell) - F^j W^j w^j(k-m+\ell) \right] \Big\} + w^j(k), \end{aligned} \tag{12}$$

where $\tilde{x}^j(k-m|k-m-1) = x^j(k-m) - \hat{x}^j(k-m|k-m-1)$ is the predicted error at time $(k-m)$. Given the innovation sequence (12), a weighted sum of innovations,

$\xi^j(k)$, is obtained as follows:

$$\xi^j(k) = \sum_{i=0}^{m} a_i^j v^j(k-i)$$
$$= \sum_{l=1}^{m} B_l^j v^j(k-l) + \sum_{l=0}^{m} G_l^j w^j(k-l). \quad (13)$$

It is easy to see that $\xi^j(k)$ is the sum of two moving average processes driven by the process noise and measurement noise, respectively [29], [32]. Here, B_l^j and G_l^j are given by

$$B_l^j = H^j\Big(\sum_{i=0}^{l-1} a_i^j (\bar{F}^j)^{l-i-1}\Big)\Gamma^j, \quad (14)$$

$$G_l^j = \Big[a_l^j I_{n_z} - H^j\Big(\sum_{i=0}^{l-1} a_i^j (\bar{F}^j)^{l-i-1}\Big)F^j W^j\Big],\ G_0^j = I_{n_z}. \quad (15)$$

Then, if we define the cross-covariance between $\xi^j(k)$ and $\xi^j(k-\ell)$ as L_ℓ^j, we obtain

$$L_\ell^j = E\big[\xi^j(k)\xi^j(k-\ell)'\big]$$
$$= \sum_{i=\ell+1}^{m} B_i^j Q^j (B_{i-\ell}^j)' + \sum_{i=\ell}^{m} G_i^j R (G_{i-\ell}^j)'. \quad (16)$$

The noise covariance matrices $Q^j = [q_{i\ell}]$ of dimension $n_v \times n_v$ and $R^j = [r_{i\ell}]$ of dimension $n_z \times n_z$ are positive definite and symmetric. By converting noise covariance matrices and the L_ℓ^j matrices as vectors as in Zhang *et al*. [32], they are related to the noise covariance identifiability matrix I^j as in (17).

$$I^j \begin{bmatrix} vec(Q^j) \\ vec(R^j) \end{bmatrix} = \begin{bmatrix} L_0^j \\ L_1^j \\ \vdots \\ L_m^j \end{bmatrix}. \quad (17)$$

As shown in [32], if matrix I^j has full column rank, then the unknown noise covariance matrices, Q^j and R^j, are uniquely identifiable. When W^j is optimal, L_ℓ^j are multiples of the innovation covariance S^j, where the scaling factor involves the minimal polynomial coefficients. For an optimal filter, it is easy to show that

$$L_\ell^j = \Big(\sum_{i=0}^{m-l} a_i^j a_{i+l}^j\Big)S^j;\ l = 0, 1, 2, ..., m. \quad (18)$$

B. Postfit Residual-Based Identifiability Conditions

Let us define $\mu^j(k)$ as the postfit residual sequence of the KF. This sequence is related to the innovation sequence $\nu^j(k)$, $k = 1, 2, ..., N$ via

$$\mu^j(k) = z(k) - H^j\hat{x}^j(k|k) = [I_{n_z} - H^j W^j(k)]\nu^j(k). \quad (19)$$

We can rewrite (19) as

$$\mu^j(k) = H^j(\tilde{F}^j)^m e^j(k-m|k-m) + \Big\{H^j\sum_{p=0}^{m-1}(\tilde{F}^j)^p$$
$$\times\big[(I_{n_x} - W^j H^j)\Gamma^j v^j(k-p-1) - W^j w^j(k-p)\big]\Big\}$$
$$+ w^j(k);\ k \geq m, \quad (20)$$

where $e^j(k+1|k+1) = \tilde{F}e^j(k|k) + (I_{n_x} - WH)\Gamma v(k) - Ww(k+1)$ is the postfit error at time $(k+1)$. Note that $\tilde{F}^j = (I_{n_x} - W^j H^j)F^j$ and $\bar{F}^j = F^j(I_{n_x} - W^j H^j)$ are similar because $\tilde{F}^j = (F^j)^{-1}\bar{F}^j F^j$.

Given the postfit residual sequence (20), let $\zeta^j(k)$ be a weighted sum of postfit residuals (see Appendix A) as,

$$\zeta^j(k) = \sum_{i=0}^{m} a_i^j \mu^j(k-i)$$
$$= \sum_{l=1}^{m} \tilde{B}_l v^j(k-l) + \sum_{l=0}^{m} \tilde{G}_l w^j(k-l), \quad (21)$$

where $\tilde{B}_l{}^j$ and $\tilde{G}_l{}^j$ are given by

$$\tilde{B}_l^j = (I_{n_z} - H^j W^j)B_l^j, \quad (22)$$

$$\tilde{G}_l^j = (I_{n_z} - H^j W^j)G_l^j. \quad (23)$$

Note that $\zeta^j(k) = (I_{n_z} - H^j W^j)\xi^j(k)$. Identifiability conditions in terms of postfit residual correlations similar to (17) ensue because $(I_{n_z} - H^j W^j)$ is invertible.

C. Output Correlations-Based Identifiability Conditions

The identifiability conditions using output correlations can be derived by using outputs only for stable open-loop systems or by using postfit residuals when the state estimation error is stabilizable when the open-loop system is unstable or marginally stable (e.g., a constant velocity target model). We will use the latter approach here.

Given (3) and (4), we can rewrite (5), the updated state estimate at time k, as

$$\hat{x}^j(k|k) = F^j\hat{x}^j(k-1|k-1) + W^j[z(k) - H^j F^j\hat{x}^j(k-1|k-1)]$$
$$= \tilde{F}^j\hat{x}^j(k-1|k-1) + W^j z(k). \quad (24)$$

We can write (19), the postfit residual sequence, as

$$\mu^j(k) = -H^j(\tilde{F}^j)^m\hat{x}^j(k-m|k-m)$$
$$- H^j\sum_{\ell=0}^{m-1}(\tilde{F}^j)^{m-1-\ell}W^j w^j(k-m+\ell+1) + z(k). \quad (25)$$

Given the postfit residual sequence (25), a weighted sum of postfit residual based on the output correlations, $\vartheta^j(k)$, can be obtained as

$$\begin{aligned}\vartheta^j(k)&=\sum_{i=0}^{m} a_i^j \mu^j(k-i)\\ &=\sum_{i=0}^{m} a_i^j\Big\{-H^j\tilde{F}^{m-i}\hat{x}^j(k-m|k-m)\\ &\quad -H^j\sum_{p=0}^{m-1-i}(\tilde{F}^j)^p W^j z(k-p)+z(k-i)\Big\}\\ &=\sum_{l=1}^{m}\hat{B}_l v^j(k-l)+\sum_{l=0}^{m}\hat{G}_l w^j(k-l)\\ &=\sum_{l=0}^{m}\hat{G}_l z(k-l). \end{aligned} \tag{26}$$

Here, the cross-covariance of $\vartheta^j(k)$ is the same as the cross-covariance of $\sum_{l=0}^{m}\hat{G}_l z(k-l)$. Identifiability conditions in terms of output correlations similar to (17) are obtained.

IV. ESTIMATING UNKNOWN FILTER PARAMETERS

A. Recursive Fading Memory-Based Innovation Correlation Estimation

We compute the sample correlation matrix $\hat{C}_{seq}^{j,k}(i)$ at sample k for model j and time lag i as a weighted combination of the correlation matrix $\hat{C}_{seq}^{j,k-1}(i)$ at the previous sample $(k-1)$ for model j and time lag i, and the samples of innovations $v^j(k-i)$ and $v^j(k)$. The tuning parameter λ, a positive constant between 0 and 1, is the weight associated with the previous sample correlation matrix. The recursive nature of the proposed algorithm makes it amenable to estimate slowly varying Q^j and R^j in nonstationary systems.

The current M sample correlation matrices at time k are used as the initial values for the next pairs of samples for recursive computation. Let us define the number of samples as N. The recursive expressions for the correlation matrices $\hat{C}_{seq}^{j,k}(i)$ are

$$\hat{C}_{seq}^{j,k}(i)=(1-\lambda)v^j(k-i)v^j(k)'+\lambda\hat{C}_{seq}^{j,k-1}(i), \tag{27}$$

$$\hat{C}_{seq}^{j,0}(i)=0;\ i=0,1,...,M-1;\ k=M,...,N. \tag{28}$$

B. Objective Function and the Gradient

The ensemble cross-correlations of a steady-state suboptimal KF are related to the closed-loop filter matrix $\bar{F}^j=F^j(I_{n_x}-W^jH^j)$, the matrix F^j, the measurement matrix H^j, the steady-state predicted covariance matrix $\bar{P}^j$, filter gain W^j, and the innovation covariance, $C^j(0)$ via [5], [21]

$$\begin{aligned}C^j(i)&=E[v^j(k)v^j(k-i)']\\ &=H^j(\bar{F}^j)^{i-1}F^j[\bar{P}^j(H^j)'-W^jC^j(0)]. \end{aligned} \tag{29}$$

The objective function Ψ^j, formulated in [32], involves minimization of the sum of normalized $C^j(i)$ with respect to the corresponding diagonal elements of $C^j(0)$ for $i>0$. The objective function is dimensionless and is zero when the filter gain is optimal and the innovation sequence is decorrelated. Formally, we can define the decorrelating objective function Ψ^j to be minimized with respect to W^j as

$$\begin{aligned}\Psi^j=&\frac{1}{2}\mathrm{tr}\Big\{\sum_{i=1}^{M-1}\big[\mathrm{diag}(C^j(0))\big]^{-\frac{1}{2}}C^j(i)'\\ &\times\big[\mathrm{diag}(C^j(0))\big]^{-1}C^j(i)\big[\mathrm{diag}(C^j(0))\big]^{-\frac{1}{2}}\Big\}, \end{aligned} \tag{30}$$

where $\mathrm{diag}(C^j)$ denotes the Hadamard product of an identity matrix with C^j. We can rewrite the objective function by substituting (29) into (30) as

$$\Psi^j=\frac{1}{2}\mathrm{tr}\Big\{\sum_{i=1}^{M-1}\phi^j(i)X^j\varphi^j(X^j)'\Big\}, \tag{31}$$

where

$$\phi^j(i)=[H^j(\bar{F}^j)^{i-1}F^j]'\varphi^j[H^j(\bar{F}^j)^{i-1}F^j], \tag{32}$$

$$X^j=\bar{P}^j(H^j)'-W^jC^j(0), \tag{33}$$

$$\varphi^j=[\mathrm{diag}(C^j(0))]^{-1}. \tag{34}$$

The gradient of objective function $\nabla_W\Psi^j$ can be computed as [32]

$$\begin{aligned}\nabla_W\Psi^j=&-\sum_{i=1}^{M-1}[H^j(\bar{F}^j)^{i-1}F^j]'\varphi^jC^j(i)\varphi^jC^j(0)-(F^j)'Z^jF^jX^j\\ &-\sum_{l=0}^{i-2}[C^j(l+1)\varphi^jC^j(i)'\varphi^jH^j(\bar{F}^j)^{i-l-2}]'. \end{aligned} \tag{35}$$

The Z^j term in (35) is computed by the Lyapunov equation.

$$\begin{aligned}Z^j=&(\bar{F}^j)'Z^j\bar{F}^j+\frac{1}{2}\sum_{i=1}^{M-1}(H^j(\bar{F}^j)^{i-1}F^j)'\varphi^jC^j(i)\varphi^jH^j\\ &+[(H^j(\bar{F}^j)^{i-1}F^j)'\varphi^jC^j(i)\varphi^jH^j]'. \end{aligned} \tag{36}$$

In computing the objective function and the gradient, we replace $C^j(i)$ by their sample estimates, $\hat{C}^{j,k}_{seq}(i)$ in (27). Evidently, the covariance estimation is a stochastic optimization problem because the cost function and the gradient depend on the realized sample paths.

C. Updating Filter Gain Sequentially

Let B be the mini-batch size and let $K = N/B$ be the number of mini-batches (we assume that N is divisible by B for simplicity). While the mini-batch gradient descent sequentially updates the M sample covariance matrices at every sample, we update the KF gain W^j when the sample index k is divisible by the size of the mini-batch B using the gradient of the objective function at sample k. Sequential mini-batch gradient descent allows more opportunities to converge to a better local minimum by frequent updates of the gain than the batch algorithm and is much less noisy than a single sample stochastic gradient algorithm [17]. Let r denote the updating index, starting with $r = 0$. The generic form of gain update is

$$(W^j)^{r+1} = (W^j)^r - (\alpha^j)^r (\nabla_W \Psi^j)^r. \tag{37}$$

The incremental gradient algorithm in (37) can be sped up by adaptively selecting the step size $(\alpha^j)^r$. Our results in [17] showed that Adam [19] and RMSProp [30] have the best accuracy and rapid convergence among all the accelerated SGD algorithms (e.g., bold driver [4], constant, subgradient [7], and Adadelta [31]) studied. Here, we show the performance results of our algorithm using the RMSProp update. RMSProp keeps track of the moving average of the squared incremental gradients for each gain element by adapting the step size elementwise.

$$\tau^j_{r,i\ell} = \gamma^j \tau^j_{r-1,i\ell} + (1-\gamma^j)[(\nabla_W \Psi^j)^r_{i\ell}]^2, \tag{38}$$

$$(\alpha^j_{i\ell})^r = \frac{(\alpha^j)^0}{\sqrt{\tau^j_{r,i\ell} + \epsilon}}; \; \tau^j_0 = 0; \; (\alpha^j)^0 = \frac{c}{K}, \tag{39}$$

where $c > 0$ is a constant and K is the number of mini-batches. Here, $\gamma = 0.9$ is the default value and $\epsilon = 10^{-8}$ to prevent division by zero. When N is unknown, as in streaming data, K is absorbed into the constant c. This is not a restriction, as mini-batch size B is all we need to implement the SGD algorithm.

D. Estimation of Process and Measurement Noise Covariances

Assuming that the necessary and sufficient conditions for the identifiability of covariances are satisfied for each model [32], here we explore the noise covariance estimation using a single-pass SGD algorithm and validate it with three illustrative examples. Unlike the algorithm in [32], this algorithm is applicable to nonstationary and multiple-model systems.

From the joint covariance of the innovation sequence $\nu^j(k)$ and the postfit residual sequence $\mu^j(k)$ in (19), and the Schur determinant identity [6], [11], one can show that at the steady state (assuming constant gain, W^j and constant Q^j and R^j over large enough time intervals such that the filter achieves steady state)[32]

$$G^j = E[\mu^j(k)\mu^j(k)'] = R^j (S^j)^{-1} R^j, \tag{40}$$

where S^j is the steady-state innovation covariance. Because (40) can be interpreted as a simultaneous diagonalization problem in linear algebra [11] or as a continuous-time algebraic Riccati equation, the measurement covariance R^j can be estimated by solving the simultaneous diagonalization problem via Cholesky decomposition and eigen decomposition, or by solving a continuous-time Riccati equation as in [1], [32].

Given the estimated R^j, we can compute the process noise covariance Q^j and the steady-state updated state covariance P^j. This requires an iterative process because Q^j and P^j are coupled in the general case [32]. Let t and l denote the iteration indices starting with $t = 0$ and $l = 0$, and using an initial $(Q^j)^0 = W^j S^j W^{j\prime}$ (exact solution in the Wiener process case [32]), we initialize the steady-state updated covariance matrix P^j as the solution of the Lyapunov equation in (41)

$$\begin{aligned}(P^j)^0 =& \tilde{F}^j (P^j)^0 (\tilde{F}^j)' + W^j R^j (W^j)' \\ &+ (I_{n_x} - W^j H^j)\Gamma^j (Q^j)^t (\Gamma^j)' (I_{n_x} - W^j H^j)', \end{aligned} \tag{41}$$

where $\tilde{F}^j = (I_{n_x} - W^j H^j) F^j$. We iteratively update P^j as in (42) until convergence

$$(P^j)^{l+1} = \left[\left(F^j (P^j)^l (F^j)' + \Gamma^j (Q^j)^t (\Gamma^j)' \right)^{-1} + (H^j)' (R^j)^{-1} H^j \right]^{-1}. \tag{42}$$

Given the converged P^j, Q^j will be updated in the t-loop until the estimate of Q^j converges. Proof of convergence is included in Appendix B.

$$(Q^j)^{t+1} = (\Gamma^j)^\dagger \left[(P^j + W^j S^j (W^j)' - F^j P^j (F^j)')^{t+1} \right] ((\Gamma^j)')^\dagger. \tag{43}$$

V. NUMERICAL EXAMPLES

Our prior study explored the effects of varying the batch sizes and the number of observation samples for accurately estimating the unknown variance parameters in nonstationary systems [18]. Here, jumps in the estimated noise covariance for the one that is not changing may be due to the changes in the other one (for example, an estimate of R may jump when R is static but Q is changing or vice versa). This is because the Kalman gain $W = PH'(HPH' + R)^{-1}$ is impacted by both Q (through P) and R.

In this section, we explore the problem of tracking the position and velocity of an aircraft in an air traffic control (ATC) system (see Section 5.1). We also consider a three-state system for estimating the unknown

Q and R (see Section 5.2). For comparison with the algorithms in the literature, we also compare the estimation performance with noise covariance estimation algorithms in stationary systems (see Section 5.3). Finally, we explore a multiple-model scenario with a set of single-pass adaptive KFs for each mode. The multiple-model method estimates the noise covariance parameters in parallel, and then selects the probable model by the concomitant mode probability (see Section 5.4). Additional application examples may be found in [16].

In the data generation process, the system is assumed to have nonstationary noise covariance matrices. We define subgroups where each subgroup has a subset of observation samples during which the noise covariances remain constant. The noise covariances abruptly change by an unknown magnitude when one subgroup of samples ends and another starts.

Note that we present the performance of the proposed method using a single model ($J = 1$) in Sections 5.1–5.3, and then consider the multiple-model case ($J = 2$) in Section 5.4. In the estimation procedure, we set the number of burn-in samples $N_b = 50$, and the number of lags $M = 5$. All computational simulations were run on a computer with an Intel Core i7-8665U processor and 16 GB of RAM.

A. A Nonlinear ATC Scenario

We consider an ATC scenario used in [3]. The ground truth is a target moving with a constant speed of 250 m/s with an initial state specified in Cartesian coordinates. The sampling interval is $T = 1$ second. A total of 500 measurement samples were collected (500 seconds of data). The target starts a left turn of 2°/s for 30 seconds at $k = 100$, then continues straight for 70 seconds (until $k = 200$), at which time it turns right with 1°/s for 45 seconds (until $k = 245$), then left with 1°/s for 90 seconds (until $k = 335$), then right with 1°/s for 45 seconds (until $k = 380$), then continues straight for 120 seconds (until $k = 500$).

The target position measurements are generated starting from $k = 0$, and they are in polar coordinates (range r and azimuth θ) by a radar located at $[\xi_0, \eta_0] = [-10^4, 0]$, with

$$r = \sqrt{(\xi - \xi_0)^2 + (\eta - \eta_0)^2}, \tag{44}$$

$$\theta = \tan^{-1}\left(\frac{\eta - \eta_0}{\xi - \xi_0}\right), \tag{45}$$

with additive white Gaussian noise with covariance $R = \text{diag}([2500\text{m}^2, (1^\circ)^2])$. Note that the noise is added to the Cartesian converted measurements, and the true values of Q and R are used for the methods which do not estimate Q and R. For this example, we used a KF based on a second order linear kinematic model (WNA) with process noise of standard deviation 1 m/s^2 described in (46).

$$F = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix}, H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \Gamma = \begin{bmatrix} T^2/2 & 0 \\ T & 0 \\ 0 & T^2/2 \\ 0 & T \end{bmatrix}. \tag{46}$$

An interacting multiple-model (IMM) estimator with one WNA (a constant velocity model with process noise standard deviation 1 m/s^2) for the uniform motion (UM) and a nearly coordinated turn (CT) model described in (47) and (48) are used. The process noise standard deviations used in the CT model were 3 m/s^2 and 0.1°/s^2 for the UM and turn rate of the state, respectively.

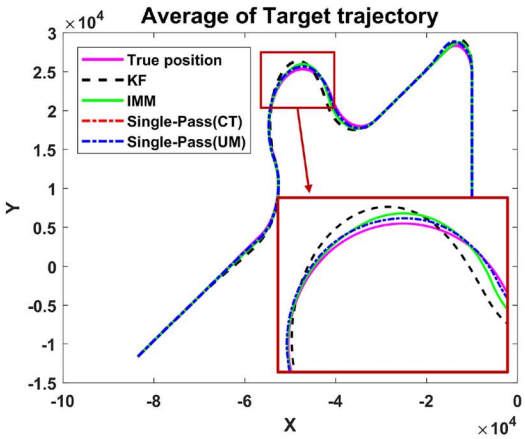
$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \Gamma = \begin{bmatrix} T^2/2 & 0 & 0 \\ T & 0 & 0 \\ 0 & T^2/2 & 0 \\ 0 & T & 0 \\ 0 & 0 & T \end{bmatrix}. \tag{48}$$

The mode transition probability matrix π in (49) is used for IMM estimator.

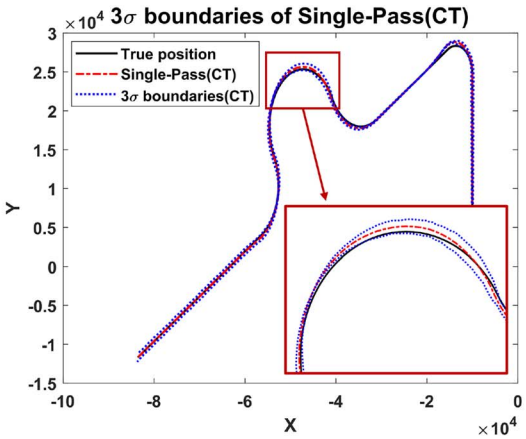
$$\pi = \begin{bmatrix} 0.95 & 0.05 \\ 0.10 & 0.90 \end{bmatrix}, \tag{49}$$

Fig. 1a shows the averaged tracking results of target motion over 100 Monte Carlo (MC) runs by KF, IMM, and the proposed method. For the single-pass SGD estimation algorithm, we considered two models using either the UM or the CT model. The proposed approach can track the target close to its true trajectory when compared to both KF and IMM. For the statistical analysis, we consider the trajectory of the single-pass SGD method with three-σ boundaries as shown in Fig. 1b. We calculate the upper and lower limits by three standard deviations (three-σ) from the mean computed over 100 MC runs. The estimates based on the CT model (even the UM model as well) are within these boundaries, which indicates that the estimates are close to the mean values. This may suggest that adapting the covariance of noise processes may overcome a lack of knowledge of the dynamics of the target to a certain extent.

$$F = \begin{bmatrix} 1 & \frac{\sin\hat{\Omega}(k)T}{\hat{\Omega}(k)} & 0 & -\frac{1-\cos\hat{\Omega}(k)T}{\hat{\Omega}(k)} & f_{\Omega,1}(k) \\ 0 & \cos\hat{\Omega}(k)T & 0 & -\sin\hat{\Omega}(k)T & f_{\Omega,2}(k) \\ 0 & \frac{1-\cos\hat{\Omega}(k)T}{\hat{\Omega}(k)} & 1 & \frac{\sin\hat{\Omega}(k)T}{\hat{\Omega}(k)} & f_{\Omega,3}(k) \\ 0 & \sin\hat{\Omega}(k)T & 0 & \cos\hat{\Omega}(k)T & f_{\Omega,4}(k) \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} f_{\Omega,1}(k) \\ f_{\Omega,2}(k) \\ f_{\Omega,3}(k) \\ f_{\Omega,4}(k) \end{bmatrix} = \begin{bmatrix} \frac{\cos\hat{\Omega}(k)T^2\hat{\dot{\xi}}(k)}{\hat{\Omega}(k)} - \frac{\sin\hat{\Omega}(k)T\hat{\dot{\xi}}(k)}{(\hat{\Omega}(k))^2} - \frac{\sin\hat{\Omega}(k)T^2\hat{\eta}(k)}{\hat{\Omega}(k)} - \frac{(-1+\cos\hat{\Omega}(k)T)\hat{\dot{\eta}}(k)}{(\hat{\Omega}(k))^2} \\ -(\sin\hat{\Omega}(k)T)T\hat{\dot{\xi}}(k) - (\cos\hat{\Omega}(k)T)T\hat{\dot{\eta}}(k) \\ \frac{\sin\hat{\Omega}(k)T^2\hat{\dot{\xi}}(k)}{\hat{\Omega}(k)} - \frac{(1-\cos\hat{\Omega}(k)T)\hat{\dot{\xi}}(k)}{(\hat{\Omega}(k))^2} + \frac{\cos\hat{\Omega}(k)T^2\hat{\eta}(k)}{\hat{\Omega}(k)} - \frac{\sin\hat{\Omega}(k)T\hat{\dot{\eta}}(k)}{(\hat{\Omega}(k))^2} \\ (\cos\hat{\Omega}(k)T)T\hat{\dot{\xi}}(k) - (\sin\hat{\Omega}(k)T)T\hat{\dot{\eta}}(k) \end{bmatrix}. \tag{47}$$



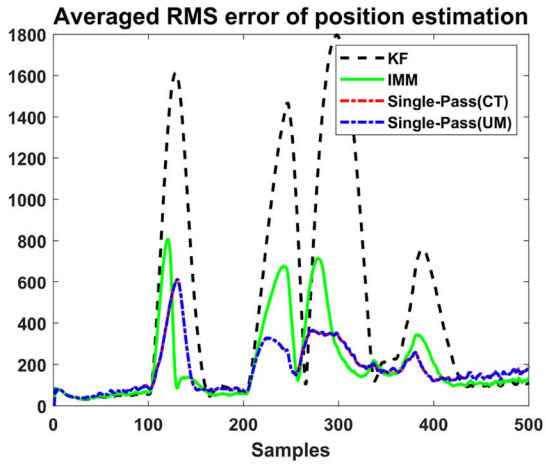
(a) Comparison of target trajectories (100 MC runs)



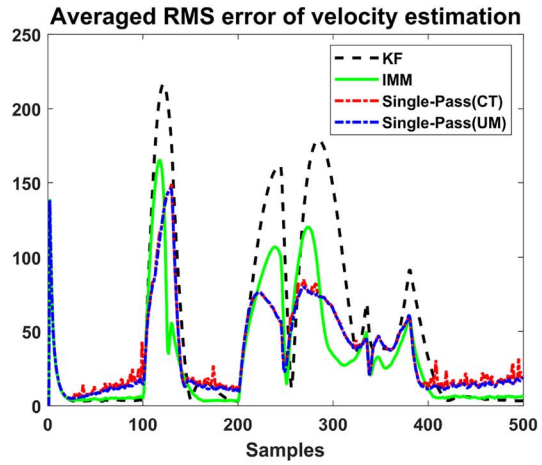
(b) Target trajectory of the proposed method with 3σ boundaries (100 MC runs)

Figure 1. Averaged tracking results of ATC motion scenario.

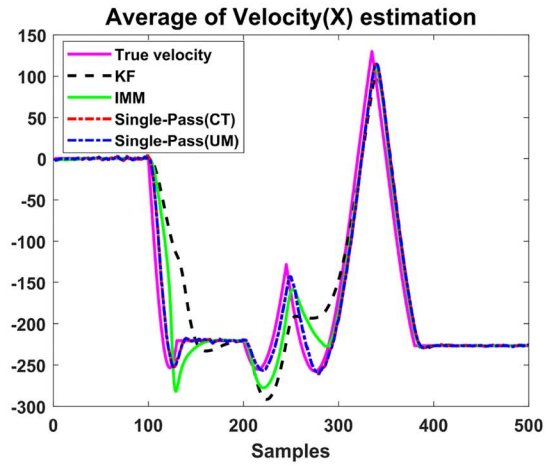
As shown in Fig. 2a, the proposed approach based on the CT model (even the UM model as well) has the peak root mean square (RMS) position error of about 200 m in the scenario considered. The proposed method reduces the RMS position error by a factor of nine when compared to a KF and by a factor of four when compared to an IMM estimator when the aircraft is maneuvering. The proposed approach shows an acceptable RMS error of velocity estimation, as shown in Fig. 2b. The proposed approach can also track the target velocity close to its true value as shown in Fig. 2c and 2d.



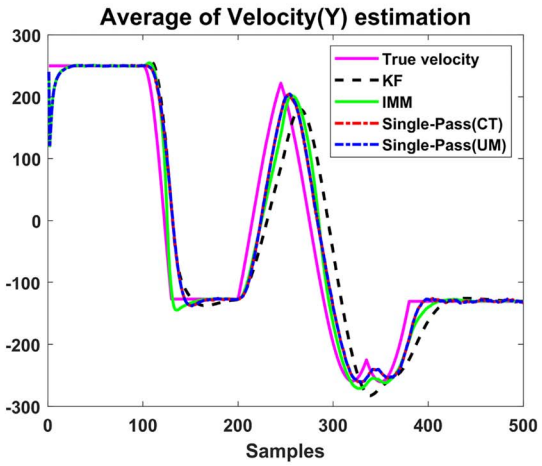
(a) RMS position estimation error



(b) RMS velocity estimation error



(c) X-velocity estimation



(d) Y-velocity estimation

Figure 2. Comparison of optimization algorithms for ATC motion estimation (100 MC runs).

Table I
Single-Pass SGD Estimation When Both Q and R Change Continuously (100 Runs; RMSProp Update)

Subgroup	R			Q			$\bar{P}_{11}$			$\bar{P}_{22}$			$\bar{P}_{33}$			W_{11}			W_{21}			W_{31}		
index	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE
1st	0.06	0.07	0.01	0.36	0.35	0.03	0.38	0.37	0.06	1.46	1.39	0.22	3.36	3.20	0.50	1.21	1.18	0.11	2.37	2.28	0.14	3.60	3.45	0.20
2nd	0.04	0.04		0.20	0.21		0.22	0.23		0.82	0.86		1.90	1.97		1.15	1.22		2.25	2.32		3.41	3.51	
3rd	0.06	0.05		0.25	0.27		0.27	0.30		1.02	1.13		2.35	2.60		1.03	1.16		2.02	2.24		3.07	3.40	
4th	0.10	0.10		0.46	0.46		0.49	0.50		1.86	1.89		4.28	4.36		1.07	1.10		2.09	2.16		3.18	3.32	
5th	0.12	0.12		0.56	0.57		0.61	0.61		2.29	2.32		5.28	5.35		1.10	1.09		2.15	2.15		3.27	3.32	

B. Scenario Where Process and Measurement Noise Covariances Change Continuously

In this scenario, we consider a three-state system using 50 000 samples where both Q and R change continuously as in the example used in [25]. The system, which has a well-conditioned observable matrix, is assumed to be as follows:

$$F = \begin{bmatrix} 0.1 & 0 & 0.1 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.3 \end{bmatrix}, \; H = \begin{bmatrix} 0.1 & 0.2 & 0 \end{bmatrix}, \; \Gamma = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}. \tag{50}$$

The true values of Q and R are generated by first starting with piecewise constant variance values for the five subgroups of samples as $Q = [0.36, 0.20, 0.25, 0.46, 0.56]$, and $R = [0.06, 0.04, 0.06, 0.10, 0.12]$ with the values changing every 10 000 samples. The Gaussian-weighted moving average algorithm with a window size of 10,000 samples is applied to the piecewise constant noise covariances to generate smoothed continuous values.

As shown in Table I, the proposed algorithm can track the noise covariance parameters accurately when both Q and R change continuously. Because the noise covariance is changing continuously, the table values are provided only in the middle of the subgroups.

Fig. 3 shows the trajectory of noise parameters when both Q and R change continuously. Our sequential algorithm can track Q and R correctly with a smoothing weight of 0.7, and the KF is consistent when evaluated with respect to the normalized innovation squared (NIS) metric, as shown in Fig. 3c.

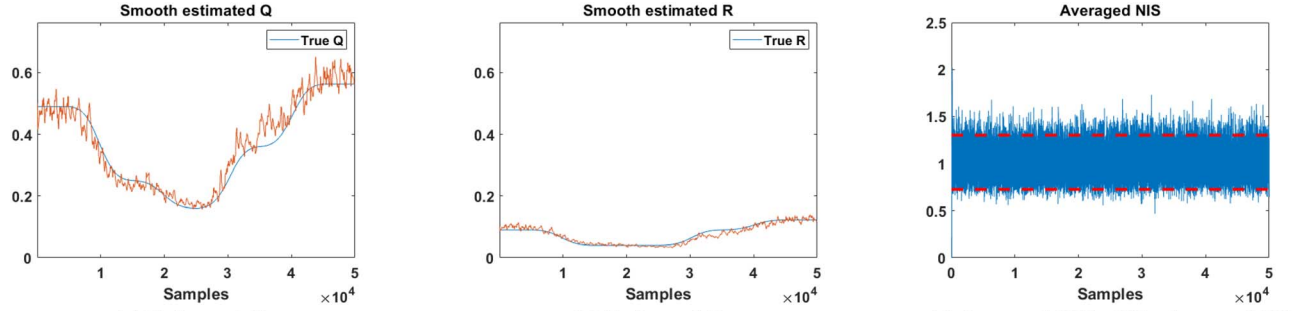
C. Comparison of Noise Covariance Estimation Algorithms on Stationary and Nonstationary Systems

Since most noise covariance estimation algorithms assume constant Q and R, in this scenario, we consider a stationary system as in the example used in [10] with 10 000 samples. We compare our single pass, multiple pass, and batch estimation algorithms with the noise covariance estimation algorithms based on the Bayesian method, the covariance matching methods (CMMs), correlation methods, and the maximum likelihood methods (MLMs). The system is assumed to be as follows:

$$F = \begin{bmatrix} 0.9 & 0 \\ -0.3 & 0.8 \end{bmatrix}, \; H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \; \Gamma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{51}$$

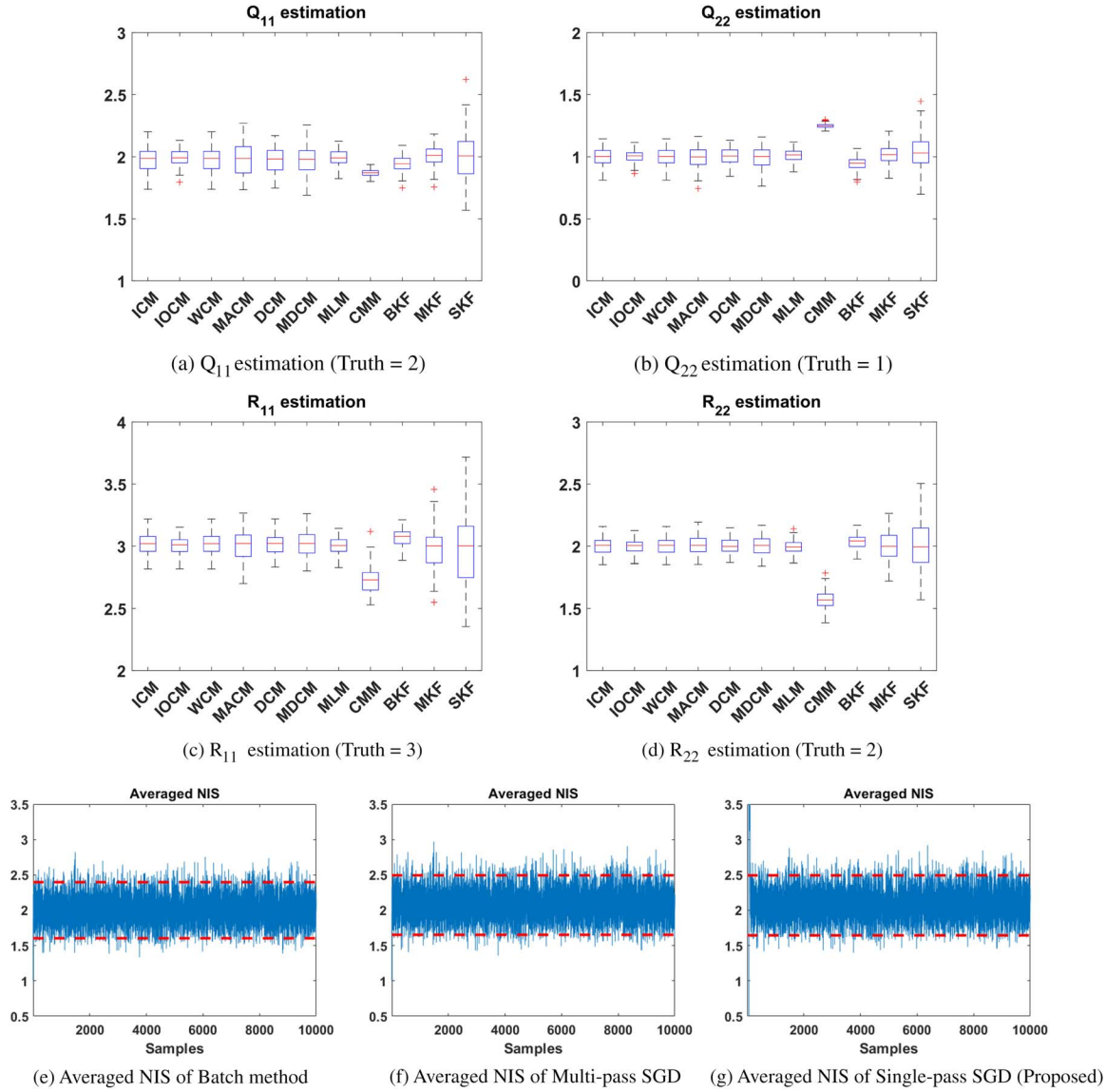
The indirect correlation method (ICM) [21], [22] relies on examining the autocovariance function (ACF) of the innovations of a linear estimator. The weighted correlation method (WCM) [5] is based on an analysis of the innovation sequence in the linear estimator, and the direct correlation method (DCM) [25] estimates the noise covariances of the innovation sequence of a stable linear estimator. In this scenario, we set an initial gain $W^0 = 0.8I_2$ for ICM, WCM, and DCM algorithms. The input–output correlation method (IOCM) [14] is designed for the linear Gaussian models by a minimization of the measurement prediction error related to an input–output model. The measurement matrix H is the identity matrix as required by IOCM algorithm, and the initial condition for estimating the coefficient matrices $\mathcal{B}^0 = O_2$.

The measurement average correlation method (MACM) [33] is based on an analysis of the covariance



(a) Estimated Q (b) Estimated R (c) Averaged NIS of Single-pass SGD

Figure 3. Trajectory of noise parameters when Q and R change continuously.



(a) Q_{11} estimation (Truth = 2)

(b) Q_{22} estimation (Truth = 1)

(c) R_{11} estimation (Truth = 3)

(d) R_{22} estimation (Truth = 2)

(e) Averaged NIS of Batch method

(f) Averaged NIS of Multi-pass SGD

(g) Averaged NIS of Single-pass SGD (Proposed)

Figure 4. Comparison of covariance estimation algorithms in stationary systems (100 MC Runs; 10 000 samples).

sequences of measurement estimate error. Two initial weight matrices $M^1_{wgt} = I_4$ and $M^2_{wgt} = I_6$ are given in this scenario. The measurement difference correlation method (MDCM) [9] directly derives the measurement estimate from other measurements without requiring state estimation. However, instead of weighting multiple measurements as MACM does, this method predicts the measurement through forward-in-time propagation of the measurement. Here, the number of measurement predictions for MDCM is set to 1. The MLM [28] relies on maximizing the likelihood function directly associated with the state space models via numerical optimization. The CMM [23] is designed for a *linear time-varying* (LTV) models with time-varying noise covariances, employing the filtering and predictive steps of a linear estimator. We set the initial noise covariances $Q^0 = I_2$ and $R^0 = I_2$ for the CMM and the single-pass SGD Kalman filter (SKF) algorithms. The batch and multipass versions of our approach (BKF and MKF) are also included for comparison purposes.

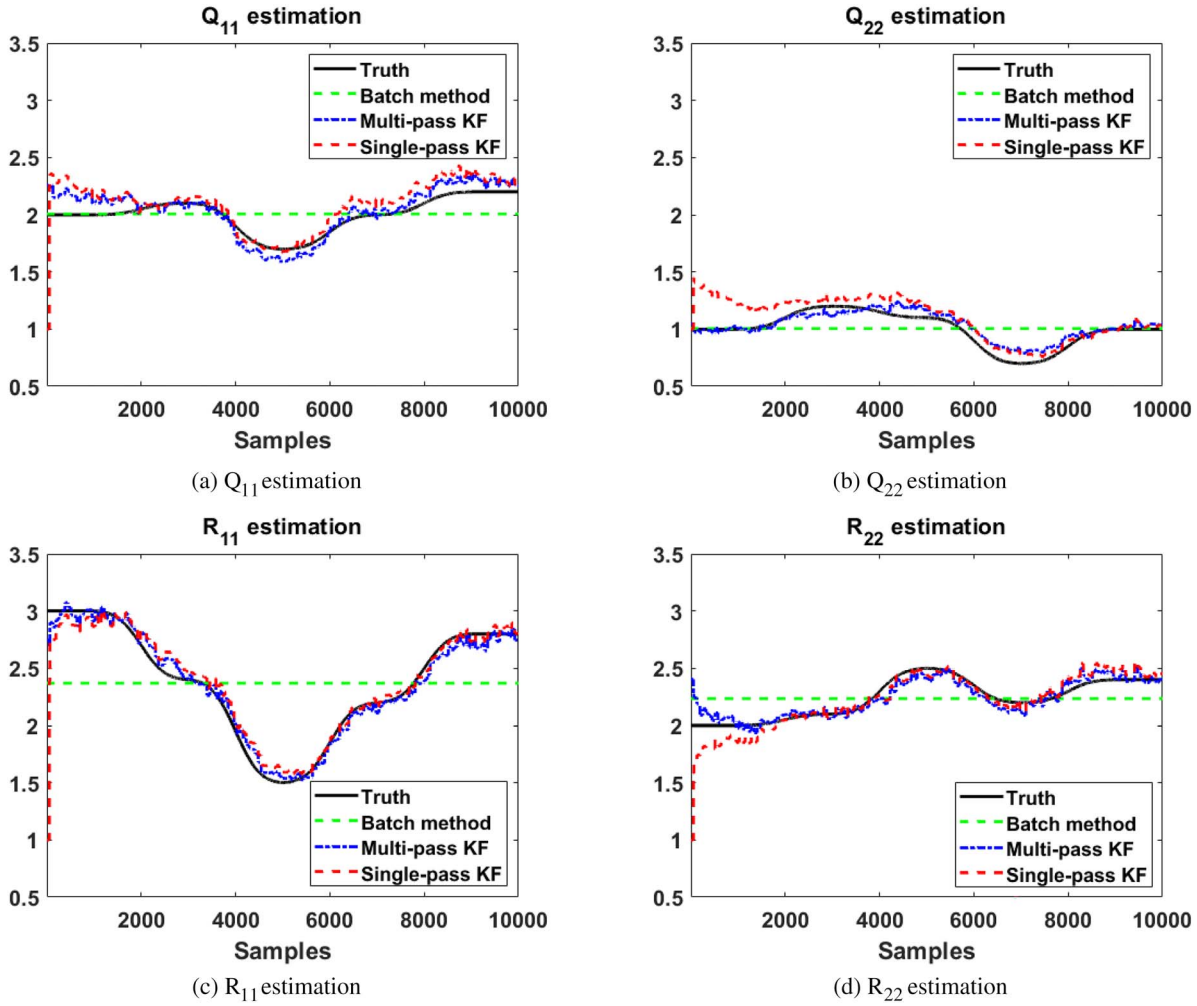
Table II shows the performance comparison of our proposed method (shown highlighted) with other algorithms for estimating noise covariances Q and R for this system averaged over 100 MC simulation runs.

Fig. 4 shows the box plots of the estimates for the various covariance estimation algorithms. Each method shows the estimates of noise covariances, with the red central mark being the median, the edges of the box being the (blue) 25th and (black) 75th percentiles, and the red crosses corresponding to the not considered outliers. Note that the batch and multipass methods estimate the parameters as well as any other method, while the single-pass SGD method estimates the noise covariances reasonably well; indeed, all filters are consistent as measured by the averaged NIS for this stationary system.

Table II
Estimates of Noise Covariances Q and R in Stationary Systems (100 MC Runs; 10 000 Samples)

Method	Q_{11} (Truth = 2)		Q_{22} (Truth = 1)		R_{11} (Truth = 3)		R_{22} (Truth = 2)	
	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
ICM	1.98	1.04E-02	1.00	4.72E-03	3.02	7.84E-03	2.00	4.69E-03
IOCM	1.99	4.20E-03	1.01	2.71E-03	3.01	4.68E-03	2.00	3.59E-03
WCM	1.98	1.04E-02	1.00	4.72E-03	3.02	7.84E-03	2.00	4.69E-03
MACM	1.98	1.86E-02	1.00	7.22E-03	3.01	1.49E-02	2.01	5.68E-03
DCM	1.98	9.22E-03	1.00	4.14E-03	3.02	7.26E-03	2.00	4.34E-03
MDCM	1.97	1.57E-02	1.00	6.92E-03	3.02	9.81E-03	2.00	5.51E-03
MLM	1.99	4.26E-03	1.01	2.37E-03	3.01	4.16E-03	1.99	3.23E-03
CMM	1.87	7.56E-04	1.25	3.05E-04	2.73	1.27E-02	1.57	4.80E-03
BKF	1.94	4.19E-03	0.95	2.95E-03	3.07	4.54E-03	2.04	3.55E-03
MKF	2.00	6.48E-03	1.01	5.00E-03	2.97	2.56E-02	2.00	1.39E-02
SKF	2.00	3.72E-02	1.04	1.89E-02	2.95	7.09E-02	1.99	4.01E-02

Fig. 5 shows the trajectories of noise parameters in nonstationary systems. As discussed in [18], the batch estimation method is not well-suited for nonstationary systems due to its assumption of constant noise covariances and the need for the availability of the entire observation sequence to compute both the objective function and the gradient. For this example, a change-point detection algorithm [15] is not applied to the multipass SGD method because noise covariances change continuously. The single-pass SGD method is consistent as measured by averaged NIS (not shown) and can estimate the noise covariances correctly for online streaming data.



(a) Q_{11} estimation

(b) Q_{22} estimation

(c) R_{11} estimation

(d) R_{22} estimation

Figure 5. Trajectories of noise parameters in nonstationary systems (100 MC Runs; 10 000 samples).

D. Application to Multiple-Model Cases

We investigate a situation involving an unknown motion model, utilizing a multiple-model approach. Within this algorithm, two KFs are utilized to estimate the noise covariance parameters, each tuned to a distinct model. The mode probabilities are integral to determining the active model. It is important to highlight that we regard a model as valid if its mode probability exceeds 0.66, employing the corresponding state estimate in such instances. When mode probabilities fall between 0.33 and 0.66, we combine the state estimates proportionally based on their respective posterior probabilities.

1. Scenario When Process and Measurement Noise Covariances Vary: For the multiple-model approach, we apply our proposed algorithm to the system used in [24]. Here, we assume that the observation samples are generated by Model 2, as in (53). Model 1 is assumed to be

$$F^1 = \begin{bmatrix} 1 & 1 \\ -0.1 & 0.1 \end{bmatrix},\ H^1 = \begin{bmatrix} 1 & 0 \end{bmatrix},\ \Gamma^1 = \begin{bmatrix} 1 \\ 0.4 \end{bmatrix}. \quad (52)$$

Model 2 is

$$F^2 = \begin{bmatrix} 0.8 & 1 \\ -0.4 & 0 \end{bmatrix},\ H^2 = \begin{bmatrix} 1 & 0 \end{bmatrix},\ \Gamma^2 = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}. \quad (53)$$

Fig. 6 shows the trajectory of estimated parameters with 50 000 measurements, and the noise parameters change every 10,000 samples. Here, the piecewise constant functions with five subgroups for true Q and R are generated based on Q=[0.04, 0.64, 0.25, 1.00, 0.09], and $R = [0.42, 0.81, 0.49, 0.16, 0.64]$. The mode probability of the second model is higher than the first model as expected. We find that the single-pass multiple-model approach can track Q and R accurately.

2. Scenario When Measurement Noise Covariance Changes Continuously: We consider a scenario used in [26] with two dynamic models using 4000 measurement samples in which the measurement noise covariance changes continuously and compare our algorithm with the variational Bayesian method and the IMM approach.

Model 1:

$$F^1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.1 \\ 0 & -0.002 & 1 \end{bmatrix},\ H^1 = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix},\ \Gamma^1 = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix}. \quad (54)$$

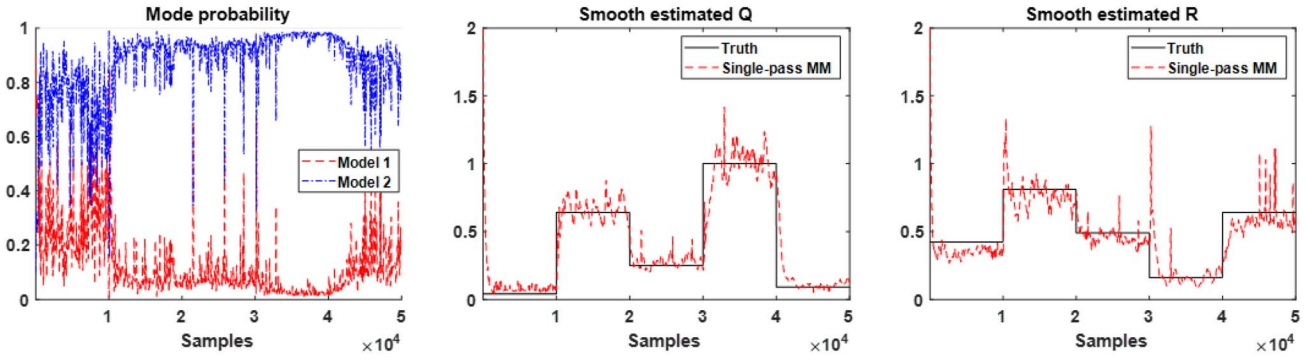
Model 2:

$$F^2 = \begin{bmatrix} 0.99 & 0 & 0 \\ 0.001 & 1 & 0.1 \\ 0 & 0 & 1 \end{bmatrix},\ H^2 = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix},\ \Gamma^2 = \begin{bmatrix} 0.01 \\ 0.05 \\ 0.1 \end{bmatrix}. \quad (55)$$

In our data generation process, we model the measurement noise variance, $R(k)$, as a continuous function in the range between 0.2 and 1 when the sampling interval, h, is 0.1 seconds, as follows:

$$R(k) = \begin{cases} 0.2 + 0.4(1 + \tanh(0.1h(k - 1000))), k \le 1500 \\ 0.2 + 0.4(1 + \tanh(0.1h(2000 - k))), \text{otherwise.} \end{cases} \quad (56)$$

Table III shows the RMSE of estimated R over 100 MC runs by the multiple-model approach. The multiple-model method estimates the noise covariance parameters in parallel, and then finds the probable model using the concomitant mode probability. Note that the measurements are generated using Model 1, as in (54). The variational bayesian adaptive kalman filter (VB-AKF) algorithm provides the best estimate when the hyper parameter $\rho = 1 - \exp(-4)$, but the RMSE value is quite sensitive to the selection of ρ. When the computation time needed for tuning the hypre parameter ρ is considered, our algorithm is superior to VB-AKF in RMSE and computational efficiency. For estimating R by the IMM filter with a multiple-model approach, each model used 111 noise models that changed uniformly between 0.1 and 1.2. The IMM filter with a large number of models (111 noise models) shows a 10% lower RMSE than our proposed method, but the IMM filter is very expensive computationally by as much as a factor of 197. Even IMM filter with 64 noise models shows slightly worse RMSE than our proposed method, but our method has better computational efficiency by a factor of 66 over the IMM.



(a) Mode probability (b) Estimated Q (c) Estimated R

Figure 6. Estimated trajectories of Q and R based on the multiple-model approach.

Table III
RMSE of Estimated R Using the Multiple-model Approach (100 MC Runs; 4000 Samples)

Method by MM	Computation time (sec)	RMSE	Description
VB-AKF	85	0.1624	$\rho = 1 - \exp(-3)$
	86	0.1178	$\rho = 1 - \exp(-4)$
	85	0.1506	$\rho = 1 - \exp(-5)$
IMM	9038	0.1009	64 noise models
	26 941	0.0836	111 noise models
Single-pass SGD	137	0.0922	RMSProp update, Batch size = 16, Smoothing weight = 0.7

Fig. 7 shows the averaged estimated trajectory of R over 100 MC runs by VB-AKF (for $\rho = 1 - \exp(-4)$), IMM (for 111 noise models) and our single-pass SGD (for RMSProp update with a mini-batch size of 16) in the multiple-model scenario. For estimating the noise covariance parameters, all optimization methods can track R correctly, but VB-AKF requires knowledge of the heuristic factor ρ, and the computation cost of IMM is substantially high. Here, the estimated trajectory of R by the single-pass algorithm was smoothed by a smoothing weight of 0.7.

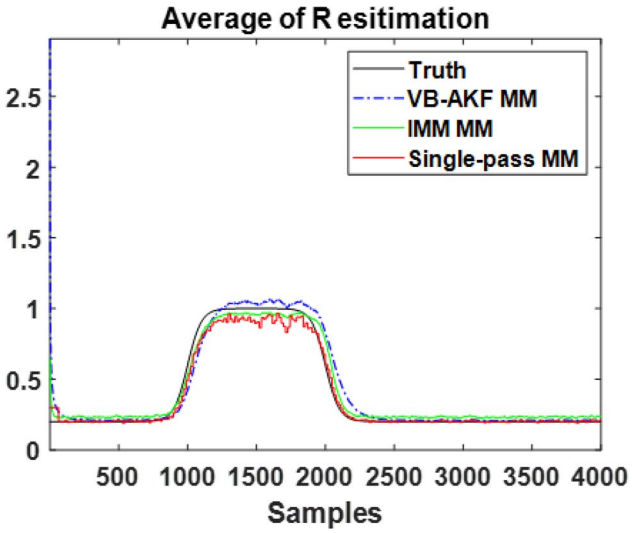


Figure 7. Trajectory of estimated R by multiple-model approach (100 MC runs).

VI. CONCLUSION AND FUTURE WORK

In this paper, we presented a single-pass SGD algorithm that estimates the noise covariance in adaptive KFs with streaming data. Compared to the batch method or multipass sequential algorithm, our proposed streaming method is an order of magnitude faster, while still achieving acceptable root mean square error (RMSE) of the state estimates. This algorithm is suitable for nonstationary systems where noise covariances vary slowly and can occasionally exhibit abrupt changes, as well as for multiple models. The efficiency of the algorithm comes from the recursive fading memory estimation of sample cross-correlations of the innovations, along with an accelerated SGD algorithms and single-pass computations. Our proposed method has been evaluated on several test cases to demonstrate its computational efficiency, accuracy, and filter consistency when compared to extant approaches.

In the future, a number of research avenues can be pursued, including (1) estimating Q and R using one-step lag smoothed residuals; (2) automatic model selection from a library of dynamic models for model adaptation; (3) use of (16) directly in a stochastic gradient algorithm, while ensuring the positive definiteness of R and positive semi-definiteness of Q (although preliminary results are not promising); (4) use of maximum likelihood criterion instead of the normalized time correlations of innovations; and (5) explore the utility of the covariance estimation algorithm as an alternative to IMMs.

APPENDIX A. PROOF OF POSTFIT RESIDUAL-BASED IDENTIFIABILITY CONDITIONS

Let us assume the postfit error $e^j(k|k) = x(k) - \hat{x}^j(k|k)$. Note that $\tilde{F}^j = (I_{n_x} - W^j H^j)F^j$ and $\bar{F}^j = F^j(I_{n_x} - W^j H^j)$ are similar because $\tilde{F}^j = (F^j)^{-1}\bar{F}^j F^j$ given by [29], [32].

$$\begin{aligned}\mu^j(k) &= H^j(\tilde{F}^j)^m e^j(k-m|k-m) + \Big\{H^j \sum_{\ell=0}^{m-1}(\tilde{F}^j)^{m-1-\ell} \\ &\quad \times \left[(I_{n_x} - W^j H^j)\Gamma^j v^j(k-m+\ell) - W^j \right. \\ &\quad \left. \times w^j(k-m+\ell+1)\right]\Big\} + w^j(k) \\ &= H^j(\tilde{F}^j)^m e^j(k-m|k-m) + \Big\{H^j \sum_{p=0}^{m-1}(\tilde{F}^j)^p \\ &\quad \times \left[(I_{n_x} - W^j H^j)\Gamma^j v^j(k-p-1) - W^j w^j(k-p)\right]\Big\} \\ &\quad + w^j(k);\ k \geq m. \end{aligned} \tag{57}$$

The process $\zeta^j(k)$, the weighted sum of postfit residual, can be obtained as

$$\begin{aligned}\zeta^j(k) &= \sum_{i=0}^{m} a_i^j \mu^j(k-i)\\ &= \sum_{i=0}^{m} a_i^j \Big\{ H^j (\tilde{F}^j)^{m-i} e^j(k-m|k-m)\\ &\quad + H^j \sum_{p=0}^{m-1-i} (\tilde{F}^j)^p \big[(I_{n_x} - W^j H^j)\Gamma^j v^j(k-i-p-1)\\ &\quad - W^j w^j(k-i-p)\big] + w^j(k-i) \Big\}. \end{aligned} \tag{58}$$

Let $l = i + p + 1$ and use the minimal polynomial

$$\begin{aligned}\zeta^j(k) &= \sum_{i=0}^{m} a_i^j \Big\{ H^j \sum_{l=i+1}^{m} (\tilde{F}^j)^{l-i-1} \big[(I_{n_x} - W^j H^j)\\ &\quad \times \Gamma^j v^j(k-l) - W^j w^j(k-l+1)\big] + w^j(k-i) \Big\}\\ &= H^j \sum_{l=1}^{m} \sum_{i=0}^{l-1} a_i^j (\tilde{F}^j)^{l-i-1} (I_{n_z} - W^j H^j)\Gamma^j v^j(k-l)\\ &\quad - H^j \sum_{l=0}^{m-1} \sum_{i=0}^{l} a_i^j (\tilde{F}^j)^{l-i} W^j w^j(k-l) + \sum_{l=0}^{m} a_l^j w^j(k-l)\\ &= \sum_{l=1}^{m} \tilde{B}_l^j v^j(k-l) + \sum_{l=0}^{m} \tilde{G}_l^j w^j(k-l), \end{aligned} \tag{59}$$

where $\tilde{B}_l{}^j$ and $\tilde{G}_l{}^j$ are given by

$$\begin{aligned}\tilde{B}_l{}^j &= H^j \left(\sum_{i=0}^{l-1} a_i^j (\tilde{F}^j)^{l-i-1} \right) (I_{n_z} - W^j H^j)\Gamma^j\\ &= (I_{n_z} - H^j W^j) \left(\sum_{i=0}^{l-1} a_i^j (\bar{F}^j)^{l-i-1} \right) \Gamma^j\\ &= (I_{n_z} - H^j W^j) B_l^j. \end{aligned} \tag{60}$$

$$\begin{aligned}\tilde{G}_l{}^j &= -H^j \left(\sum_{i=0}^{l} a_i^j \tilde{F}^{l-i} \right) W^j + a_l^j I_{n_z}\\ &= -(I_{n_z} - H^j W^j) H^j \Big(\sum_{i=0}^{l} a_i^j (\bar{F}^j)^{l-i-1} F^j W^j + a_l^j I_{n_z}\\ &= -(I_{n_z} - H^j W^j) H^j \left(\sum_{i=0}^{l-1} a_i^j (\bar{F}^j)^{l-i-1} \right) F^j W^j\\ &\quad - a_l^j (I_{n_z} - H^j W^j) H^j (\bar{F}^j)^{-1} F^j W^j + a_l^j I_{n_z}\\ &= -(I_{n_z} - H^j W^j) H^j \left(\sum_{i=0}^{l-1} a_i^j (\bar{F}^j)^{l-i-1} \right) F^j W^j\\ &\quad - a_l^j H^j W^j + a_l^j I_{n_z}\\ &= -(I_{n_z} - H^j W^j) H^j \left(\sum_{i=0}^{l-1} a_i^j (\bar{F}^j)^{l-i-1} \right) F^j W^j\\ &\quad + a_l^j (I_{n_z} - H^j W^j)\\ &= (I_{n_z} - H^j W^j) G_l^j. \end{aligned} \tag{61}$$

APPENDIX B. PROOF OF CONVERGENCE OF THE ITERATIVE ALGORITHM FOR Q

Since the predicted error covariance is related to the updated error covariance via $\bar{P}^j = P^j + W^j S^j (W^j)'$, we have

$$P^j = F^j P^j (F^j)' + \Gamma^j Q^j (\Gamma^j)' - W^j S^j (W^j)'. \tag{62}$$

Since the Kalman gain is related to the updated error covariance via $W^j = P^j (H^j)' (R^j)^{-1}$, the first iteration of the updated state error covariance, P_1^j, can be obtained by solving the Riccati equation as

$$\begin{aligned}P_1^j &= F^j P_1^j (F^j)' + \Gamma^j Q_1^j (\Gamma^j)' - W^j S^j (W^j)'\\ &= F^j P_1^j (F^j)' + \Gamma^j Q_1^j (\Gamma^j)'\\ &\quad - P_1^j (H^j)' (R^j)^{-1} S (R^j)^{-1} H^j P_1^j\\ &= F^j P_1^j (F^j)' + \Gamma^j Q_1^j (\Gamma^j)' - P_1^j (H^j)' (G^j)^{-1} H^j P_1^j. \end{aligned} \tag{63}$$

With this solution, $W^j S^j (W^j)' \geq P_1^j (H^j)' (G^j)^{-1} H^j P_1^j$ because otherwise $P_1^j \leq 0$ at the initial iteration (Recall $\Gamma^j Q_1^j (\Gamma^j)' = W^j S^j (W^j)'$ at the initial iteration). Given $\Gamma^j Q_2^j (\Gamma^j)' = P_1^j - F^j P_1^j (F^j)' - W^j S^j (W^j)' \geq \Gamma^j Q_1^j (\Gamma^j)'$, we have the second iteration of the updated state error covariance, P_2^j, as

$$\begin{aligned}P_2^j &= F^j P_2^j (F^j)' + \Gamma^j Q_2^j (\Gamma^j)' - P_2^j (H^j)' (G^j)^{-1} H^j P_2^j\\ &= F^j P_2^j (F^j)' + P_1^j - F^j P_1^j (F^j)' - W^j S^j (W^j)'\\ &\quad - P_2^j (H^j)' (G^j)^{-1} H^j P_2^j. \end{aligned} \tag{64}$$

Evidently,

$$\begin{aligned}\delta P^j &= P_1^j - P_2^j\\ &= F^j \delta P^j (F^j)' + \Gamma^j Q_1^j (\Gamma^j)' - P_1^j + F^j P_1^j (F^j)'\\ &\quad - W^j S^j (W^j)' - P_1^j (H^j)' (G^j)^{-1} H^j P_1^j\\ &\quad + P_2^j (H^j)' (G^j)^{-1} H^j P_2^j\\ &= F^j \delta P^j (F^j)' + P_2^j (H^j)' (G^j)^{-1} H^j P_2^j - W^j S^j (W^j)'\\ &\leq 0. \end{aligned} \tag{65}$$

As the iterations proceed, $P_n^j (H^j)' (G^j)^{-1} H^j P_n^j - W^j S^j (W^j)' \to 0$, and P^j monotonically approaches P^{j*} from below.

REFERENCES

[1] W. Arnold and A. Laub
"Generalized eigenproblem algorithms and software for algebraic Riccati equations,"
Proc. IEEE, vol. 72, no. 12, pp. 1746–1754, 1984.

[2] F. Auger, M. Hilairet, J. M. Guerrero, E. Monmasson, T. Orlowska-Kowalska, and S. Katsura
"Industrial applications of the Kalman filter: A. review,"
IEEE Trans. Ind. Electron., vol. 60, no. 12, pp. 5458–5471, 2013.

[3] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan
Estimation with applications to tracking and navigation: Theory algorithms and software. Hoboken, NJ, USA: Wiley, 2004.

[4] R. Battiti
"Accelerated backpropagation learning: Two optimization methods,"
Complex Syst., vol. 3, no. 4, pp. 331–342, 1989.

[5] P. R. Belanger
"Estimation of noise covariance matrices for a linear time-varying stochastic process,"
Automatica, vol. 10, no. 3, pp. 267–275, 1974.

[6] D. Bertsekas
Nonlinear Programming (Athena Scientific Optimization and Computation Series). Belmont, MA, USA: Athena Scientific. 2016.

[7] S. Boyd, L. Xiao, and A. Mutapcic
Subgradient methods. *lecture notes of EE392o, Stanford University, Autumn Quarter*, 2004(01), 2003.

[8] R. S. Bucy and P. D. Joseph
Filtering for Stochastic Processes with Applications to Guidance, vol. 326. Providence, RI, USA: American Mathematical Society. 2005.

[9] J. Duník, O. Straka, and O. Kost
Measurement difference autocovariance method for noise covariance matrices estimation.
In *Proc. IEEE 55th Conf. Decis. Control*, 2016, pp. 365–370.

[10] J. Duník, O. Straka, O. Kost, and J. Havlík
"Noise covariance matrices in state-space models: A. survey and comparison of estimation methods—Part I,"
Int. J. Adaptive Control Signal Process., vol. 31, no. 11, pp. 1505–1543, 2017.

[11] G. H. Golub and C. F. Van Loan
Matrix Computations, vol. 3. Baltimore, MD, USA: The Johns Hopkins Univ. Press. 2013.

[12] P. Hanlon and P. Maybeck
"Multiple-model adaptive estimation using a residual correlation Kalman filter bank,"
IEEE Trans. Aerosp. Electron. Syst., vol. 36, no. 2, pp. 393–406, 2000.

[13] R. E. Kalman
"A new approach to linear filtering and prediction problems,"
J. Basic Eng., vol. 82, no. 1, pp. 35–45, Mar. 1960.

[14] R. Kashyap
"Maximum likelihood identification of stochastic linear systems,"
IEEE Trans. Autom. Control, vol. 15, no. 1, pp. 25–34, 1970.

[15] R. Killick, P. Fearnhead, and I. A. Eckley
"Optimal detection of changepoints with a linear computational cost,"
J. Amer. Stat. Assoc., vol. 107, no. 500, pp. 1590–1598, 2012.

[16] H.-S Kim, A. Bienkowski, and K. R. Pattipati
"A. single-pass noise covariance estimation algorithm in multiple-model adaptive Kalman filtering. In,
Proc., 2023 IEEE Aerospace Conference, 2023.

[17] H.-S Kim, L. Zhang, A. Bienkowski, and K. R. Pattipati
"Multi-pass sequential mini-batch stochastic gradient descent algorithms for noise covariance estimation in adaptive Kalman filtering,"
IEEE Access, vol. 9, pp. 99220–99234, 2021.

[18] H.-S Kim, L. Zhang, A. Bienkowski, and K. R. Pattipati
A. single-pass noise covariance estimation algorithm in adaptive Kalman filtering for nonstationary systems.
In Proc. *IEEE 24th Int. Conf. Inf. Fusion* , 2021, pp. 556–563.

[19] D. P. Kingma and J. Ba
Adam: A. method for stochastic optimization. arXiv, 2014.

[20] X.-R Li and Y. Bar-Shalom
Multiple-model estimation with variable structure.
IEEE Trans. Autom. Control, vol. 41, no. 4, pp. 478–493, Apr. 1996.

[21] R. Mehra
"On the identification of variances and adaptive Kalman filtering,"
IEEE Trans. Autom. Control, vol. 15, no. 2, pp. 175–184, Apr. 1970.

[22] R. Mehra
"Approaches to adaptive filtering,"
IEEE Trans. Autom. Control, vol. 17, no. 5, pp. 693–698, Oct. 1972.

[23] K. Myers and B. Tapley
"Adaptive sequential estimation with unknown noise statistics,"
IEEE Trans. Autom. Control, vol. 21, no. 4, pp. 520–523, Aug. 1976.

[24] C. Neethling and P. Young
" Comments on identification of optimum filter steady-state gain for systems with unknown noise covarianc,"
IEEE Trans. Autom. Control, vol. 19, no. 5, pp. 623–625, Oct. 1974.

[25] "A. new autocovariance least-squares method for estimating noise covariances,,"
Automatica, vol. 42, no. 2, pp. 303–308, 2006.

[26] S. Sarkka and A. Nummenmaa
"Recursive noise adaptive Kalman filtering by variational bayesian approximations,"
IEEE Trans. Autom. Control, vol. 54, no. 3, pp. 596–600, Mar. 2009.

[27] S. Servadio and R. Zanetti
"Differential algebra-based multiple Gaussian particle filter for orbit determination,"
J. Optim. Theory Appl., vol. 191, no. 2, pp. 459–485, 2021.

[28] R. H. Shumway, D. S. Stoffer, and D. S. Stoffer
Time Series Analysis and Its Applications, vol 3. Berlin, Germany: Springer 2000.

[29] K. Tajima
"Estimation of steady-state kalman filter gain,"
IEEE Trans. Autom. Control, vol. 23, no. 5, pp. 944–945, 1978.

[30] T. Tieleman and G. Hinton
"Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude, *COURSERA: Neural Netw. Mach. Learn.*, vol. 4, no. 2, pp. 26–31, 2012.

[31] M. D. Zeiler
Adadelta: An adaptive learning rate method. arXiv, 2012.

[32] L. Zhang, D. Sidoti, A. Bienkowski, K. R. Pattipati, Y. Bar-Shalom, and D. L. Kleinman
"On the identification of noise covariances and adaptive Kalman filtering: A. new look at a 50 year-old problem,"
IEEE Access, vol. 8, pp. 59362–59388, 2020.

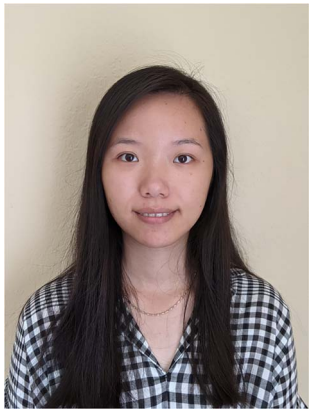
[33] J. Zhou and R. Luecke
"Estimation of the co variances of the process noise and measurement noise for a linear discrete dynamic system,"
Comput. Chem. Eng., vol. 19, no. 2, pp. 187–195, 1995.

Hee-Seung Kim received the B.S. and M.S. degrees in computer and communication engineering from the Chungbuk National University, Cheongju, South Korea, in 2011 and 2013, respectively. He is currently pursuing a Ph.D. degree in the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT, USA.

He was a researcher at the Electronics and Telecommunications Research Institute (ETRI), Gwangju, South Korea, from 2013 to 2016. His research interests include machine learning, image processing, and adaptive Kalman filtering.

Adam Bienkowski received the B.S. and M.Eng. degrees in electrical and computer engineering from the University of Connecticut, Storrs, CT, USA, in 2013 and 2017, respectively, where he is currently pursuing a Ph.D. degree in Department of Electrical and Computer Engineering University of Connecticut, Storrs, CT, USA, under the advisement of Dr. K. R. Pattipati.

He was an Electrical Engineer at General Dynamics Electric Boat, Groton, CT, USA, from 2013 to 2017. His current research interests include modeling dynamic and uncertain environments for asset allocation and path planning, context aware decision support systems, and optimization and machine learning-based techniques for mission planning and coordination.



Lingyi Zhang received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from the University of Connecticut, Storrs, CT, USA, in 2014, 2019, and 2020, respectively. She was the co-recipient of the Tammy Blair award for best student paper at FUSION 2016. Her current research interests include modeling dynamic and uncertain environments for asset allocation and path planning, optimization-based techniques for mission planning and coordination, and adaptive Kalman filtering.

Krishna R. Pattipati received the B. Tech. degree in electrical engineering with highest honors from the Indian Institute of Technology, Kharagpur, West Bengal, India, in 1975, and the M.S. and Ph.D. degrees in systems engineering from the University of Connecticut, Storrs, CT, USA, in 1977 and 1980, respectively. He was with ALPHATECH, Inc., Burlington, USA, MA from 1980 to 1986. Since 1986, he has been with the Department of Electrical and Computer Engineering, University of Connecticut where he is currently a Distinguished Professor Emeritus and the Collins Chair Professor of Systems engineering. His research interests are in the application of systems theory, optimization and inference techniques to agile planning, anomaly detection, diagnostics, and prognostics. He has published more than 500 scholarly journal and conference papers in these areas. He is a cofounder of Qualtech Systems, Inc., a firm specializing in advanced integrated diagnostics software tools (TEAMS, TEAMS-RT, TEAMS-RDS, TEAMATE, PackNGo), and serves on the board of Aptima, Inc. Dr. Pattipati was selected by the IEEE Systems, Man, and Cybernetics (SMC) Society as the Outstanding Young Engineer of 1984 and received the Centennial Key to the Future award. He served as the *IEEE Transactions on Systems, Man, and Cybernetics Part B* from 1998 to 2001. He was co-recipient of the Andrew P. Sage Award for the Best SMC Transactions Paper for 1999, the Barry Carlton Award for the Best AES Transactions Paper for 2000, the 2002 and 2008 NASA Space Act Awards for "A Comprehensive Toolset for Model-based Health Monitoring and Diagnosis," and "Real-time Update of Fault-Test Dependencies of Dynamic Systems: A Comprehensive Toolset for Model-Based Health Monitoring and Diagnostics," the 2005 School of Engineering Outstanding Teaching Award, the 2003 AAUP Research Excellence Award at UCONN, and the 2023 Distinguished Alumnus Award from the Indian Institute of Technology, Kharagpur. He is an elected Fellow of IEEE for his contributions to discrete-optimization algorithms for large-scale systems and team decision-making, as well as a Fellow of the Connecticut Academy of Science and Engineering.