A Single-pass Noise Covariance Estimation Algorithm in Multiple-model Adaptive Kalman Filtering for Non-stationary Systems

Hee-Seung Kim 1,1, Lingyi Zhang 2, Adam Bienkowski 2, and Krishna Pattipati 2

1University of Connecticut
2Affiliation not available

November 8, 2023

Abstract

Estimation of unknown noise covariances in a Kalman filter is a problem of significant practical interest in a wide array of applications. This paper presents a single-pass stochastic gradient descent (SGD) algorithm for noise covariance estimation for use in adaptive Kalman filters applied to non-stationary systems where the noise covariances can occasionally jump up or down by an unknown magnitude. Unlike our previous batch method or our multi-pass decision-directed algorithm, the proposed streaming algorithm reads measurement data exactly once and has similar root mean square error (RMSE). The computational efficiency of the new algorithm stems from its one-pass nature, recursive fading memory estimation of the sample cross-correlations of the innovations, and the RMSprop accelerated SGD algorithm. The comparative evaluation of the proposed method on a number of test cases demonstrates its computational efficiency and accuracy.
A Single-pass Noise Covariance Estimation Algorithm in Multiple-model Adaptive Kalman Filtering for Non-stationary Systems

HEE-SEUNG KIM, Member, IEEE
ADAM BIENKOWSKI, Member, IEEE
KRISHNA R. PATTIPATI, Life Fellow, IEEE
University of Connecticut, Storrs, CT 06269, USA

Abstract—Estimation of unknown noise covariances in a Kalman filter is a problem of significant practical interest in a wide array of industrial applications. This paper presents a single-pass stochastic gradient descent (SGD) algorithm for noise covariance estimation for use in non-switching multiple-model adaptive Kalman filters, where the noise covariances can occasionally jump up or down by an unknown magnitude. Unlike our previous batch estimation or multi-pass decision-directed estimation methods, the proposed streaming algorithm reads measurement data exactly once and has acceptable root mean square error (RMSE) of state estimates. The computational efficiency of our proposed algorithm stems from its single-pass nature, recursive fading memory estimation of the sample cross-correlations of the innovations, and the root mean square propagation (RMSprop) accelerated SGD algorithm. The comparative evaluation of the proposed method on a number of test cases demonstrates its computational efficiency and estimation accuracy.

Index Terms—Adaptive Kalman filtering, Covariance matrices, Sequential estimation

I. INTRODUCTION

Kalman filter (KF) [10] is an optimal minimum mean square error (MMSE) state estimator under the Gaussian assumption with known mean and covariance for discrete-time linear dynamic systems. It is also the best linear state estimator if the noises entering into the system are non-Gaussian with known first and second moments, i.e., mean and covariance. As a result of numerous research studies, Kalman filter has received substantial interest in many industrial applications, such as diagnosis, robotics, signal processing, navigation, and target tracking, to name a few [2], [3]. However, the statistics of noise processes are mostly unknown or partially known in many practical situations.

In [23], we first derived the necessary and sufficient conditions and then presented a batch optimization algorithm for estimating unknown noise covariances. The noise covariance identifiability conditions involve the minimization of the sum of the normalized temporal cross-correlations of the innovations. With the Gaussian assumption, the innovation sequence of an optimal Kalman filter is orthogonal, implying that the innovations are independent of each other, and are strictly white [3].

In [12], we proposed an enhanced covariance estimation method based on a sequential mini-batch stochastic gradient descent (SGD) algorithm that required multiple passes through the data as well as a method to detect changes in noise covariances when the covariance estimation method was applied to non-stationary systems. In this paper, we propose a single pass adaptive Kalman filtering algorithm for non-stationary systems, wherein the process and measurement noise covariances can occasionally jump up or down by an unknown magnitude, and the system behavior can stem from a member of a known subset of models.

A. Prior work

The key to process and measurement noise covariance estimation is a relationship linking the covariance of the state estimation error and of the innovations in any suboptimal filter. This relationship serves as a fundamental building block for the correlation-based approaches. Pioneering contributions using this approach were made by [5], [16]–[18].

Sarkka and Nummenmaa [19] proposed a variational Bayesian method for the joint recursive estimation of the dynamic state and measurement noise parameters in linear state space models. The method is implemented by forming separable variance approximations to the joint posterior distribution of state and noise parameters at each time step. However, this method does not account for changes in process noise. The variational approaches generally require tuning parameters to converge to the correct parameters and they often converge to local minima.

In [12], we improved the computational efficiency and accuracy of the batch estimation algorithm in [23] by...
a sequential mini-batch estimation method with adaptive step size rules. When updating the filter gain, we applied a sequential fading memory mini-batch estimates of the innovation correlations. We also applied dynamic convergence thresholds to enhance the computational efficiency of the sequential estimation algorithm. Application of this method to non-stationary systems requires a change-point detection algorithm described in [11] to extract the time points of abrupt changes in unknown noise covariances based on the innovation sequence.

The multiple-model estimation algorithm assumes that the system obeys one of a finite number of models, and each model has its own non-switching dynamics [3]. The overall estimate of multiple-model algorithm is a convex combination of the estimates from multiple parallel filters, each based on the individual models, where the weights of convex combination are the concomitant posterior model probabilities.

For the noise covariance estimation in non-stationary systems, the limitations of previous research [12] are the following: First, the previous methods require multiple passes through the observation data and thus are computationally expensive. Second, the accuracy of decision-directed noise covariance estimation method depends on the accuracy of the change-point detection algorithm because the sequential estimation algorithm is invoked for samples in between two consecutive change points. Third, the previous algorithms assumed that the structure of the dynamic model is known. In this paper, we seek to overcome these three limitations by developing a streaming algorithm that reads measurements exactly once and extends the approach to multiple models.

B. Contribution and organization of the paper

We present a single-pass, sequential mini-batch noise covariance estimation algorithm as an extension of the work in [12], [13] for non-stationary systems. Our proposed method enables the estimates 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. More significantly, the structure of the dynamic model is unknown, but is known to belong to a known subset of models. We validate the proposed method on several non-stationary and multiple-model system test cases.

The paper is organized as follows. In Section II, we provide an overview of the sequential mini-batch stochastic gradient descent (SGD) method in multiple-model system, including identifiability conditions for estimating the unknown noise covariances in each individual model, the fading memory filter-based innovation correlation estimation, and the SGD update of the Kalman gain. Section III shows evidence that our method can track unknown noise covariances in non-stationary systems, as well as systems exhibiting dynamics from a known subset, and that the single-pass algorithm is computationally efficient. Lastly, we conclude the paper and discuss potential avenues for future work in Section IV.

II. SEQUENTIAL MINI-BATCH GRADIENT DESCENT METHOD FOR ESTIMATING PROCESS AND MEASUREMENT NOISE COVARIANCE PARAMETERS IN MULTIPLE-MODEL SYSTEMS

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, $j = 1, 2, \cdots, J$, given by

\begin{align*}
    x(k + 1) &= F^j x(k) + \Gamma^j \nu^j(k) \\
    z(k) &= H^j x(k) + \omega^j(k)
\end{align*}

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. $F^j$ 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 $\Gamma^j$ is the noise gain matrix. Here, the process noise $\nu^j(k)$ and the measurement noise $\omega^j(k)$ 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 piece-wise constant such that the filter reaches a steady state between any two jumps and that the jump is of unknown magnitude.

Given $Q^j(k)$ and $R^j(k)$, the multiple-model adaptive Kalman filter involves the consecutive processes of prediction and update given by [3], [9], [10], [15]

\begin{align*}
    \hat{x}^j(k+1|k) &= F^j \hat{x}^j(k|k) \\
    \nu^j(k+1) &= z(k + 1) - H^j \hat{x}^j(k + 1|k) \\
    \hat{x}^j(k+1|k+1) &= \hat{x}^j(k + 1|k) + W^j(k+1)\nu^j(k+1) \\
    \Delta^j(k) &= \frac{1}{\sqrt{2\pi S^j(k)}} \exp(-\frac{1}{2} \nu^j(k)^T S^j(k)^{-1} \nu^j(k)) \\
    \lambda^j(k) &= \frac{\Delta^j(k) p^j(k - 1)}{\sum_{j=1}^J \lambda^j(k) p^j(k - 1)}
\end{align*}

The Kalman filter 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...
Kalman filters are non-interacting.

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 \( P_j(k+1) \)). The updated state error covariance \( P_j(k+1|k) \) is computed via (9). This corresponds to Joseph form, 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 \( \nu^j(k) \) and innovation covariance \( S_j(k) \). 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 \( p^j(0) = 1/J \). The identifiability conditions of the multiple-model approach depend on each candidate model since the corresponding Kalman filters are non-interacting.

### A. Identifiability conditions for \( Q \) and \( R \)

Consider model \( j \). Assume that \( Q^j \) and \( R^j \) are unknown, but are piece-wise constant such that the filter reaches the steady state. Now, consider the innovations corresponding to a stable, suboptimal closed-loop filter matrix \( F^j = F^j(I_{n_x} - W^jH^j) \) given by [20], [23]

\[
\nu^j(k) = H^j(F^j)^m \bar{\nu}^j(k-m|k-m-1) + \left\{ H^j \sum_{\ell=0}^{m-1} (F^j)^{m-\ell} \times \left[ \Gamma^j \nu^j(k+m-\ell) - F^jW^j \nu^j(k+m+\ell) \right] \right\} + \nu^j(k)
\]

where \( \bar{\nu}^j(k-m|k-m-1) = x^j(k-m) - \bar{x}^j(k-m|k-m-1) \) is the predicted error at time \( (k-m) \). Given the innovation sequence (12), let us define a weighted sum of innovations, \( \xi^j(k) = \sum_{i=0}^{m} a_i^j \nu^j(k-i) \), where the weights are the coefficients of the minimal polynomial of the closed-loop filter matrix \( F^j \). \( \sum_{i=0}^{m} a_i^j (F^j)^{m-i} = 0, a_0^j = 1 \). 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, given by [20], [23]

\[
\xi^j(k) = \sum_{i=1}^{m} B_i^j \nu^j(k-i) + \sum_{i=0}^{m} G_i^j \nu^j(k-i) \quad (13)
\]

Here, \( B_i^j \) and \( G_i^j \) are given by

\[
B_i^j = H^j \left( \sum_{i=0}^{l-1} a_i^j (F^j)^{l-i-1} \right) \Gamma^j
\]

\[
G_i^j = \left[ a_i^j I_{n_x} - H^j \left( \sum_{i=0}^{l-1} a_i^j (F^j)^{l-i-1} \right) F^j W^j \right], G_0^j = I_{n_x}
\]

Then, if we define the cross-covariance between \( \xi^j(k) \) and \( \xi^j(k-l) \) as \( L^j_{k,l} \), we obtain

\[
L^j_{k,l} = E[\xi^j(k)\xi^j(k-l)] = \sum_{i=\ell+1}^{m} B_i^j Q^j(B_{i-\ell}^j)^t + \sum_{i=\ell}^{m} G_i^j R(G_{i-\ell}^j)^t \quad (16)
\]

The noise covariance matrices \( Q^j = [q_{ij}] \) of dimension \( n_x \times n_x \) and \( R^j = [r_{ij}] \) of dimension \( n_z \times n_z \) are positive definite and symmetric. By converting noise covariance matrices and the \( L^j_{k,l} \) matrices as vectors, Zhang et al. [23] show that they are related by the noise covariance identifiability matrix \( I^j \) as in (17).

\[
I^j \begin{bmatrix} \text{vec}(Q^j) \\ \text{vec}(R^j) \end{bmatrix} = \begin{bmatrix} L_1^j \\ \vdots \\ L_m^j \end{bmatrix}
\]

As shown in [23], if matrix \( I^j \) has full column rank, then the unknown noise covariance matrices, \( Q^j \) and \( R^j \), are uniquely identifiable.

### B. Recursive fading memory-based Innovation Correlation Estimation

We compute the sample correlation matrix \( \hat{C}^{j,k}_{\text{seq}}(i) \) at sample \( k \) for model \( j \) and time lag \( i \) as a weighted combination of the correlation matrix \( \hat{C}^{j,k-1}_{\text{seq}}(i) \) at the previous sample \( (k-1) \) for model \( j \) and time lag \( i \), and the samples of innovations \( \nu^j(k-i) \) and \( \nu^j(k) \). The tuning parameter \( \lambda \), a positive constant between 0 and 1, is the weight associated with the previous sample correlation matrix. The recursive nature of proposed algorithm makes it amenable to estimating \( Q^j \) and \( R^j \) for non-stationary 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 \( N \).

\[
\hat{C}^{j,k}_{\text{seq}}(i) = (1-\lambda) \hat{C}^{j,k}_{\text{seq}}(i) + \lambda \hat{C}^{j,k-1}_{\text{seq}}(i) \quad (18)
\]

\[
\hat{C}^{j}_{\text{seq}}(0) = 0, i = 0, 1, 2, \cdots M-1; k = M, \cdots , N \quad (19)
\]

### C. Objective function and the Gradient

The ensemble cross-correlations of a steady-state suboptimal Kalman filter is related to the closed-loop filter matrix \( \tilde{F}^j = F^j(I_{n_x} - W^jH^j) \), the matrix \( F^j \), the measurement matrix \( H^j \), the steady-state predicted covariance matrix \( \tilde{P}^j \), filter gain \( W^j \) and the innovation covariance, \( C^j(0) \) via [5], [16]

\[
C^j(i) = E[\nu^j(k)\nu^j(k-i)^t] = H^j (F^j)^{i-1} F^j (\tilde{P}^j(H^j)^t - W^j C^j(0)) \quad (20)
\]

The objective function \( \Psi^j \) formulated in [23], involves minimization of the sum of normalized \( C^j(i) \) with respect to the corresponding diagonal elements of \( C^j(0) \) for \( i > 0 \).
0. Formally, we can define the objective function \( \Psi^j \) to be minimized with respect to \( W^j \) as

\[
\Psi^j = \frac{1}{2} \text{tr} \left\{ \sum_{i=1}^{M-1} \left[ \text{diag}(C^j(0)) \right]^{-\frac{1}{2}} C^j(i) \right\} \times \left[ \text{diag}(C^j(0)) \right]^{-\frac{1}{2}} \}
\]

(21)

where \( \text{diag}(C^j) \) denotes the diagonal matrix of \( C^j \) or equivalently the Hadamard product of an identity matrix with \( C^j \). We can rewrite the objective function by substituting (20) into (21) as

\[
\Psi^j = \frac{1}{2} \text{tr} \left\{ \sum_{i=1}^{M-1} \phi^j(i) X^j \varphi^j(X^j)' \right\}
\]

(22)

where

\[
\phi^j(i) = [H^j(\tilde{F}^j)^{-1} F^j] \varphi^j \left[ H^j(\tilde{F}^j)^{-1} F^j \right]' \]

(23)

\[
X^j = \tilde{P}^j [H^j]' - W^j C^j(0)
\]

(24)

\[
\varphi^j = [\text{diag}(C^j(0))]^{-1}
\]

(25)

The gradient of objective function \( \nabla_W \Psi^j \) can be computed as [23]

\[
\nabla_W \Psi^j = - \sum_{i=1}^{M-1} \left[ H^j(\tilde{F}^j)^{-1} F^j \right]' \varphi^j C^j(i) \varphi^j C^j(0) - (F^j)Z^j F^j X^j
\]

\[- \sum_{l=0}^{l-2} C^j(l+1) \phi^j C^j(i)' \phi^j H^j(\tilde{F}^j)^{-1-2}(i)'
\]

(26)

The \( Z^j \) term in (26) is computed by the Lyapunov equation.

\[
Z^j = (\tilde{F}^j)' Z^j \tilde{F}^j + \frac{1}{2} \sum_{i=1}^{M-1} \left[ H^j(\tilde{F}^j)^{-1} F^j \right]' \varphi^j C^j(i) \varphi^j H^j
\]

\[+ \left[ (H^j(\tilde{F}^j)^{-1} F^j)' \varphi^j C^j(i) \varphi^j H^j \right]'
\]

(27)

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

D. Updating gain \( W \) 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 Kalman filter gain \( W \) when the sample index \( k \) is divisible by the size of 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 update of the gain than the batch algorithm and is much less noisy than a single sample stochastic gradient algorithm [12]. Letting \( r \) denotes the updating index starting with \( r = 0 \). The generic form of gain update is

\[
(W^j)^{r+1} = (W^j)^r - (\alpha^j)^r (\nabla_W \Psi^j)^{r}
\]

(28)

The incremental gradient algorithm in (28) can be sped up by adaptively selecting the step size \( (\alpha^j)^r \). Our results in [12] showed that Adam [14] and RMSProp [21] have the best accuracy and rapid convergence among all the accelerated SGD algorithms (e.g., bold driver [4], constant, subgradient [7], Adadelta [22]) 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 element-wise.

\[
\tau^j_{r,i} = \gamma^j \tau^j_{r-1,i} + (1 - \gamma^j) [(\nabla_W \Psi^j)^r]_i^2
\]

(29)

\[
(\alpha^j)^r = \frac{(\alpha^j)^0}{\sqrt{\tau^j_{r,i} + \epsilon}}; \tau^j_{0} = 0; (\alpha^j)^0 = \frac{c}{K}
\]

(30)

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 \( B \) size is all we need to implement the stochastic gradient descent algorithm.

E. Estimation of \( Q \) and \( R \)

Assuming that the necessary and sufficient conditions for the identifiability of covariances are satisfied [23], here we explore the noise covariance estimation using a single-pass SGD algorithm and validate it with standard examples described in the literature. Unlike the algorithm in [23], this algorithm is applicable to non-stationary and multi-model systems.

1. Estimation of \( R \)

Let us define \( \mu^j(k) \), as the post-fit residual sequence of the Kalman filter, which is related to the innovations \( \nu^j(k), k = 1, 2, \cdots, N \) via

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

(31)

From the joint covariance of the innovation sequence \( \nu^j(k) \) and the post-fit residual sequence \( \mu^j(k) \), and the Schur determinant identity [6], [8], one can show that in the steady-state (assuming constant gain, \( W \) and constant \( Q^j \) and \( R^j \) over large enough time intervals such that the filter achieves steady state) [23]

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

(32)

where \( S^j \) is the steady-state innovation covariance. Because (32) can be interpreted as a simultaneous diagonalization problem in linear algebra [8] 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], [23].

2. Estimation of Q

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 [23]. Letting \( t \) and \( l \) denote the iteration indices starting with \( t = 0 \) and \( l = 0 \), and using an initial \( (Q^j)^0 = W^j S^j W^j \) (exact solution in the Wiener process case [23]), we initialize the steady-state updated covariance matrix \( P^j \) as the solution of the Lyapunov equation in (33)

\[
(P^j)^0 = \tilde{F}^j (P^j)^0 (\tilde{F}^j)^T + W^j R^j (W^j)^T + (I_n - W^j H^j)^T \Gamma^j (Q^j)^0 (\Gamma^j)^T (I_n - W^j H^j)^T \tag{33}
\]

where \( \tilde{F}^j = (I_n - W^j H^j) F^j \). We iteratively update \( P^j \) as in (34) until convergence

\[
(P^j)^{t+1} = \left( (F^j (P^j)^t (F^j)^T + \Gamma^j (Q^j)^t (\Gamma^j)^T) \right)^{-1} (\Gamma^j)^T (I_n - W^j H^j)^T \tag{34}
\]

Given the converged \( P^j, Q^j \) will be updated in the \( t \)-loop until the estimate of \( Q^j \) converges.

\[
(Q^j)^{t+1} = (\Gamma^j)^T \left[ (P^j + W^j S^j (W^j)^T - F^j P^j (F^j)^T)^{t+1} \right] (\Gamma^j)^T \tag{35}
\]

The pseudocode for the sequential mini-batch SGD estimation algorithm for noise covariance parameters in multiple-model non-stationary systems is included as Algorithm 1.

III. NUMERICAL EXAMPLES

We apply our proposed method to the system used in [17], but modified to have non-stationary noise covariance matrices. The system matrices are assumed to be as follows.

\[
F = \begin{bmatrix} 0.8 & 1 \\ -0.4 & 0 \end{bmatrix}, \quad H = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} \tag{36}
\]

The non-stationarity is assumed to arise from occasional abrupt changes in process and measurement noise covariances of unknown magnitude. Here “occasionally” implies that the jumps are infrequent enough that the Kalman filter is in the steady-state prior to a jump in the noise covariance. In this paper, we call the subset of samples where the noise covariances are not changing as a subgroup. Let \( N_{sg} \) be the number of subgroups and let \( N \) be the total number of samples. Here, each subgroup has \( N/N_{sg} \) samples during which the noise covariances remain constant.

Algorithm 1 Pseudocode of single-pass estimation algorithm in multiple-model adaptive Kalman filtering for a non-stationary system

1: input: \( (W^j)^0, (Q^j)^0, (R^j)^0, \alpha^j, B, p^j(0) \)
2: \( (W^j)^0 \): initial gain, \( (Q^j)^0 \): initial \( Q^j \), \( (R^j)^0 \): initial \( R^j \), \( \alpha^j \): initial step size, \( B \): batch size, \( p^j(0) \): initial mode probability
3: \( r = 0 \) \( \{ \triangleright \) Initialize the updating index \( r \) \( \}\)
4: for \( k = 1 \) to \( N \) \( \{ \triangleright \) Number of samples \( \}\)
5: compute the innovation correlations \( \nu^j(k) \)
6: compute the mode likelihood function \( \Lambda^j(k) \)
7: if \( k > N_b + M \) then \( \{ \triangleright \) \( N_b \): Number of burn-in samples \( \}\)
8: compute \( \hat{\Theta}_{\alpha_j}^j(i) \)
9: if \( \text{Mod}(k, B) = 0 \) then
10: compute the objective function \( \Psi^j \)
11: compute the gradient \( \nabla_W \Psi^j \)
12: update the step size \( \alpha^j \)
13: update the gain
14: update \( (R^j)^{r+1} + (Q^j)^{r+1} \)
15: \( r = r + 1 \)
16: end if
17: end if
18: end for
19: compute the mode probability \( p^j(k) \)
20: mix the state estimates
21: end for

First, we consider three non-stationary scenarios for this system with \( N \) up to 50,000 measurement samples and \( N_{sg} = 5 \) subgroups: 1) both \( Q \) and \( R \) change every \( N/N_{sg} \) samples (see Section A), 2) Only \( Q \) changes every \( N/N_{sg} \) samples, and 3) Only \( R \) changes every \( N/N_{sg} \) samples (see Section B). Next, we consider a scenario using 4,000 measurement samples in which the measurement noise covariance changes continuously and compare our algorithm with variational Bayesian method and the interacting multiple model approach (see Section C). We also explore the problem of tracking the position and velocity of an aircraft in an air traffic control (ATC) system (see Section D). Finally, we explore a multiple model scenario with a set of single-pass adaptive Kalman filters for each mode. The multiple-model method estimates the noise covariance parameters in parallel, and then finds the probable model by the concomitant mode probability (see Section E).

Note that we present the performance of proposed method using a single model \( \{ j = 1 \} \) in Sections A to D, and then consider the multiple-model approach \( \{ j = 1, 2 \} \) in Section E. 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. Scenarios when both $Q$ and $R$ vary

We show that the proposed algorithm works with varying mini-batch sizes. We also simulate 100 MC runs with varying number of observation samples to establish the minimum number of samples required for accurately estimating the unknown variance parameters.

1. Varying the mini-batch sizes

Table I shows Monte Carlo simulation results when $Q$ and $R$ are varied. The noise covariance estimation algorithm was run for varying mini-batch sizes. The minimum mini-batch size for acceptable RMSE for this problem is 64, since the accuracy of estimating $P_{11}$ is relatively low when mini-batch size $\leq 32$ (shown highlighted). For all subsequent results, we use RMSProp update with a mini-batch size of 64 in our estimation algorithm.

2. Varying the number of samples observed

In order to find the minimum number of observation samples for tracking time-varying $Q$ and $R$, we compared the estimated $Q$ and $R$ values with their concomitant true values, while varying the number of observation samples $N=5,000, 10,000, 20,000, 30,000$ and $50,000$, and keeping the number of subgroups constant at $N_{sp}=5$. Fig. 1 shows the results of 100 Monte Carlo simulations in estimating $Q$ and $R$ when the number of observed samples is varied. The five dotted lines indicate true values corresponding to each subgroup. Note that the true $Q$ values by subgroups are $[0.04, 0.64, 0.25, 1.00, 0.09]$, and the true $R$ values by subgroups are $[0.42, 0.81, 0.49, 0.16, 0.64]$ for all observation samples given. The solid lines show the estimated noise covariance parameters based on the single-pass SGD algorithm. Estimated $Q$ and $R$ are closer to the corresponding true values when $N \geq 30,000$. Fig. 1 indicates that the accuracy of estimates increases with an increase in the number of samples, as one should expect.

3. Comparison of different algorithms for noise covariance estimation

Table II shows the performance comparison of noise covariance estimation algorithms in non-stationary systems. As shown in Table IIa, the batch estimation algorithms presented in [17], [23] are not suitable for non-stationary systems because they assume the covariances to be constant and require the entire observation sequence to compute the objective function and the gradient. In Table IIb, multi-pass sequential mini-batch algorithm presented in [12] can accurately estimate the noise parameters, but the algorithm requires an anomaly detection algorithm to detect the change points, and is computationally expensive.

Compared to the batch and multi-pass algorithms, the proposed single-pass method can estimate the noise

---

**TABLE I**: Monte Carlo Simulation when $Q$ and $R$ are varied (100 Runs; 50,000 samples; RMSProp update)

(a) Mini-batch size of 16

| Subgroup index | $R$ | $Q$ | $P_{11}$ | $P_{22}$ | $W_{11}$ | $W_{21}$ |
|----------------|-----|-----|---------|---------|---------|---------|
| $1^{st}$       | 0.42| 0.42| 0.04    | 0.11    | 0.02    | 0.04    |
| $2^{nd}$       | 0.81| 0.79| 0.64    | 1.27    | 0.24    | 0.61    |
| $3^{rd}$       | 0.74| 0.48| 0.25    | 0.54    | 0.10    | 0.53    |
| $4^{th}$       | 0.16| 0.17| 1.10    | 0.24    | 0.27    | 0.89    |
| $5^{th}$       | 0.64| 0.62| 0.09    | 0.25    | 0.05    | 0.28    |

(b) Mini-batch size of 32

| Subgroup index | $R$ | $Q$ | $P_{11}$ | $P_{22}$ | $W_{11}$ | $W_{21}$ |
|----------------|-----|-----|---------|---------|---------|---------|
| $1^{st}$       | 0.42| 0.42| 0.04    | 0.11    | 0.02    | 0.04    |
| $2^{nd}$       | 0.81| 0.80| 0.64    | 1.27    | 0.24    | 0.61    |
| $3^{rd}$       | 0.49| 0.49| 0.25    | 0.54    | 0.10    | 0.53    |
| $4^{th}$       | 0.16| 0.18| 1.10    | 0.24    | 0.27    | 0.89    |
| $5^{th}$       | 0.64| 0.64| 0.09    | 0.25    | 0.05    | 0.28    |

(c) Mini-batch size of 64

| Subgroup index | $R$ | $Q$ | $P_{11}$ | $P_{22}$ | $W_{11}$ | $W_{21}$ |
|----------------|-----|-----|---------|---------|---------|---------|
| $1^{st}$       | 0.42| 0.42| 0.04    | 0.11    | 0.02    | 0.04    |
| $2^{nd}$       | 0.81| 0.80| 0.64    | 1.27    | 0.24    | 0.61    |
| $3^{rd}$       | 0.49| 0.49| 0.25    | 0.54    | 0.10    | 0.53    |
| $4^{th}$       | 0.16| 0.17| 1.10    | 0.24    | 0.27    | 0.89    |
| $5^{th}$       | 0.64| 0.64| 0.09    | 0.25    | 0.05    | 0.28    |

(d) Mini-batch size of 128

| Subgroup index | $R$ | $Q$ | $P_{11}$ | $P_{22}$ | $W_{11}$ | $W_{21}$ |
|----------------|-----|-----|---------|---------|---------|---------|
| $1^{st}$       | 0.42| 0.42| 0.04    | 0.11    | 0.02    | 0.04    |
| $2^{nd}$       | 0.81| 0.81| 0.64    | 1.27    | 0.24    | 0.61    |
| $3^{rd}$       | 0.49| 0.49| 0.25    | 0.54    | 0.10    | 0.53    |
| $4^{th}$       | 0.16| 0.18| 1.10    | 0.24    | 0.27    | 0.89    |
| $5^{th}$       | 0.64| 0.64| 0.09    | 0.25    | 0.05    | 0.28    |
Fig. 1: Estimation of noise parameters with varying the number of observation samples for varying Q and R

Fig. 2: Comparison of optimization algorithms in varying system

TABLE II: Performance comparison of optimization algorithms in non-stationary systems

(a) Batch estimation (100 Monte Carlo Runs; 50,000 samples; M = 100)

| Computation time (sec) | Subgroup index | R | Q | P_11 | P_22 | W_1 | W_2 | Truth | Mean | RMSE | Truth | Mean | RMSE | Truth | Mean | RMSE | Truth | Mean | RMSE | Truth | Mean | RMSE | Truth | Mean | RMSE |
|------------------------|----------------|---|---|------|------|-----|-----|-------|------|------|-------|------|------|-------|------|------|-------|------|------|-------|------|------|-------|------|------|-------|
| 18.005                 | 1st            | 0.42| 0.51| 0.04| 0.40| 0.11| 0.80| 0.02| 0.15| 0.21| 0.61| -0.01| 0.07|
|                        | 2nd            | 0.81| 0.51| 0.64| 0.40| 1.27| 0.80| 0.24| 0.15| 0.64| 0.61| 0.03| 0.07|
|                        | 3rd            | 0.49| 0.51| 0.25| 0.40| 0.54| 0.80| 0.10| 0.15| 0.53| 0.61| 0.03| 0.07|
|                        | 5th            | 0.64| 0.51| 1.00| 0.40| 1.31| 0.80| 0.27| 0.15| 0.89| 0.61| 0.30| 0.05|
| 15.011                 | 1st            | 0.42| 0.42| 0.04| 0.04| 0.11| 0.12| 0.02| 0.02| 0.21| 0.22| -0.01| 0.02|
|                        | 2nd            | 0.81| 0.82| 0.64| 0.63| 1.27| 1.27| 0.24| 0.24| 0.81| 0.61| 0.03| 0.07|
|                        | 3rd            | 0.49| 0.48| 0.25| 0.25| 0.54| 0.54| 0.10| 0.10| 0.53| 0.53| 0.03| 0.03|
|                        | 5th            | 0.64| 0.63| 0.09| 0.09| 0.24| 0.25| 0.05| 0.05| 0.28| 0.29| -0.04| 0.02|
| 416                    | 1st            | 0.42| 0.42| 0.04| 0.04| 0.11| 0.11| 0.02| 0.02| 0.21| 0.21| -0.01| 0.01|
|                        | 2nd            | 0.81| 0.80| 0.64| 0.65| 1.27| 1.28| 0.24| 0.24| 0.61| 0.61| 0.07| 0.07|
|                        | 3rd            | 0.49| 0.49| 0.25| 0.25| 0.54| 0.55| 0.10| 0.10| 0.53| 0.53| 0.03| 0.03|
|                        | 5th            | 0.64| 0.64| 0.09| 0.09| 0.24| 0.25| 0.05| 0.05| 0.28| 0.28| -0.04| 0.01|

(b) Multi-pass SGD estimation (100 Monte Carlo Runs; 50,000 samples; M = 5; RMSProp; Batch size = 64)

(c) Single-pass SGD estimation (100 Monte Carlo Runs; 50,000 samples; M = 5; RMSProp; Batch size = 64)

Unlike the batch algorithm that is only applicable to stationary systems, the proposed algorithm and the multi-pass algorithm can track Q and R accurately as shown in Fig. 2e and Fig. 2b. The trajectories of Q and R estimates can be smoothed by a simple first order fading memory filter with a smoothing weight of 0.9. Fig. 2e shows the averaged NIS of the SGD algorithm (RMSProp; a mini-batch size of 64) when Q and R are varied. The RMSProp SGD algorithm-based Kalman filter is consistent, although there are transients when the noise covariances change abruptly.
B. Scenarios where either $Q$ or $R$ is constant

1. Varying $Q$

In this scenario, the process noise covariance $Q$ changes every 10,000 samples, while the measurement noise covariance $R$ is a constant over all 50,000 samples, as shown in Fig. 3. The sequential mini-batch algorithm can track $Q$ and $R$ quite well and provides a consistent state estimator.

![Estimated Q and R](image)

Fig. 3: Trajectory of noise parameters in varying $Q$ system

2. Varying $R$

In this scenario, measurement noise covariance $R$ changes every 10,000 observation samples, but the process noise covariance $Q$ is kept constant over the 50,000 samples. Fig. 4 shows that the sequential mini-batch gradient descent algorithm can track $Q$ and $R$ correctly.

![Estimated Q and R](image)

Fig. 4: Trajectory of noise parameters in varying $R$ system

Table III shows the 100 MC simulation results when either $Q$ or $R$ is constant. The trajectories of estimated process noise covariance and the estimated measurement noise covariance are very close to their true trajectories by the sequential algorithm.

![Monte Carlo Simulation](image)

**Table III: Monte Carlo Simulation when either $Q$ or $R$ is constant (100 Runs; 50,000 samples; RMSProp update)**

(a) The case of varying $Q$

| Subgroup index | Truth | Mean | RMSE | Truth | Mean | RMSE | Truth | Mean | RMSE | $P_{11}$ | Truth | Mean | RMSE | Truth | Mean | RMSE | Truth | Mean | RMSE | $W_{11}$ | Truth | Mean | RMSE | Truth | Mean | RMSE |
|----------------|-------|------|------|-------|------|------|-------|------|------|---------|-------|------|------|-------|------|------|-------|------|------|---------|-------|------|------|-------|------|------|
| 1$^{st}$       | 0.36  | 0.35 | 0.04 | 0.04  | 0.11 | 0.12 | 0.02  | 0.03 | 0.30 | 0.05    | 0.04  | 0.04 | 0.04 | 0.04  | 0.38 | 0.38 | 0.01  | 0.001|
| 2$^{nd}$       | 0.36  | 0.37 | 0.64 | 0.64  | 1.07 | 1.07 | 0.20  | 0.20 | 0.13 | 0.13    | 0.75 | 0.74 | 0.15 | 0.15  | 0.20 | 0.20 | 0.08  | 0.001|
| 3$^{rd}$       | 0.36  | 0.36 | 0.25 | 0.25  | 0.51 | 0.51 | 0.10  | 0.10 | 0.13 | 0.13    | 0.59 | 0.59 | 0.06 | 0.06  | 0.20 | 0.20 | 0.08  | 0.001|
| 4$^{th}$       | 0.36  | 0.38 | 1.00 | 0.99  | 1.52 | 1.52 | 0.30  | 0.30 | 0.13 | 0.13    | 0.81 | 0.79 | 0.20 | 0.20  | 0.08  | 0.001|
| 5$^{th}$       | 0.36  | 0.36 | 0.09 | 0.09  | 0.22 | 0.23 | 0.04  | 0.04 | 0.13 | 0.13    | 0.38 | 0.38 | 0.001 | 0.001|

(b) The case of varying $R$

| Subgroup index | Truth | Mean | RMSE | Truth | Mean | RMSE | Truth | Mean | RMSE | $P_{11}$ | Truth | Mean | RMSE | Truth | Mean | RMSE | Truth | Mean | RMSE | $W_{11}$ | Truth | Mean | RMSE | Truth | Mean | RMSE |
|----------------|-------|------|------|-------|------|------|-------|------|------|---------|-------|------|------|-------|------|------|-------|------|------|---------|-------|------|------|-------|------|------|
| 1$^{st}$       | 0.04  | 0.05 | 0.36 | 0.36  | 0.45 | 0.46 | 0.10  | 0.10 | 0.14 | 0.14    | 0.92 | 0.91 | 0.34 | 0.33  | 0.72 | 0.71 | 0.13  | 0.11 |
| 2$^{nd}$       | 0.64  | 0.63 | 0.36 | 0.38  | 0.73 | 0.77 | 0.15  | 0.15 | 0.16 | 0.16    | 0.55 | 0.55 | 0.04 | 0.04  | 0.46 | 0.46 | 0.02  | 0.01 |
| 3$^{rd}$       | 0.52  | 0.52 | 0.36 | 0.36  | 0.63 | 0.62 | 0.12  | 0.12 | 0.16 | 0.16    | 0.72 | 0.71 | 0.13  | 0.11 | 0.46 | 0.46 | 0.02  | 0.01 |
| 4$^{th}$       | 1.00  | 0.96 | 0.36 | 0.38  | 0.84 | 0.86 | 0.16  | 0.17 | 0.22 | 0.22    | 0.85 | 0.84 | 0.24  | 0.24 | 0.85 | 0.84 | 0.24  | 0.24 |
| 5$^{th}$       | 0.99  | 0.99 | 0.36 | 0.37  | 0.51 | 0.51 | 0.10  | 0.10 | 0.34  | 0.34    | 0.85 | 0.84 | 0.24  | 0.24 | 0.85 | 0.84 | 0.24  | 0.24 |

C. Scenario where $R$ changes continuously

We explore the scenario where $R$ changes continuously (but $Q$ is constant), as in the example used in [19].

The system matrices are

$$
F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.1 \\ 0 & -0.002 & 1 \end{bmatrix},
H = \begin{bmatrix} 0.1 \\ 1 \\ 0 \end{bmatrix}, \Gamma = \begin{bmatrix} 0.1 \\ 1 \\ 0.1 \end{bmatrix}
$$

(37)

The sampling interval, $h$, is 0.1 seconds. A total of 4,000 measurement samples were collected (400 seconds of data). The measurement noise covariance $R$ is 0.2, except for the area where it starts to increase rapidly to 1 around the sample index $k = 1,000$ (100 seconds) and then decreases again rapidly around $k = 2,000$ (200 seconds).

In our data generation process, we model the measurement noise variance, $R(k)$ as a continuous function that approximates this change:

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

(38)

Variational Bayesian Adaptive Kalman Filter (VB-AKF) algorithm described in [19] models this change by simply “spreading” its previous approximate posteriors via a heuristic factor $\rho$. Our approach does not require such heuristic approximations.

Table IV shows the RMSE of estimated $R$ by VB-AKF, interacting multiple-model (IMM) filters and the proposed method for various values of $\rho$ (for VB-AKF) and the number of noise models (for IMM). The performance of VB-AKF is very sensitive to the selection of $\rho$ as shown in Table IV; the RMSE can change by a factor.
of 3 if the wrong value is selected for $\rho$. For the VB-AKF algorithm, we obtain the least RMSE value when $\rho = 1 - \exp(-4)$.

For estimating $R$ based on the IMM filter, the range of noise levels is selected to be between 0.1 and 1.2 to be consistent with [19] and the quantized levels for the various numbers of noise models between 2 and 111 are selected from this range uniformly. The RMSE of estimated $R$ is the least when IMM uses 111 noise models. Our experiments show that even 64 noise models would have provided better RMSE than the VB-AKF method. However, IMM with a large number of models is computationally very expensive.

**TABLE IV:** RMSE of estimated $R$ (100 MC Runs; 4,000 samples)

| Method      | Computation time (sec) | RMSE  | Description                   |
|-------------|-------------------------|-------|-------------------------------|
| VB-AKF      | 42                      | 0.3631 | $\rho = 1 - \exp(-2)$        |
|             | 42                      | 0.1625 | $\rho = 1 - \exp(-3)$        |
|             | 42                      | 0.1178 | $\rho = 1 - \exp(-4)$        |
|             | 42                      | 0.1569 | $\rho = 1 - \exp(-5)$        |
|             | 43                      | 0.2435 | $\rho = 1 - \exp(-6)$        |
| IMM         | 24                      | 0.3746 | 2 noise models                |
|             | 49                      | 0.3352 | 4 noise models                |
|             | 113                     | 0.3081 | 8 noise models                |
|             | 315                     | 0.2335 | 16 noise models               |
|             | 1,039                   | 0.1508 | 32 noise models               |
|             | 3,887                   | 0.1009 | 64 noise models               |
|             | 11,435                  | 0.0836 | 111 noise models              |
| Single-pass SGD | 36                    | 0.0941 | RMSProp update; Batch size = 64 |

The computational efficiency of the proposed method (with RMSProp update; a mini-batch size of 64) is almost the same as that of VB-AKF, but the proposed method performs slightly better than VB-AKF in terms of RMSE values. More significantly, our approach does not require the knowledge of hyper-parameter $\rho$, which is rarely known in practice. If the computational cost of tuning the hyper-parameter $\rho$ is taken into account, our algorithm is much faster and more accurate than the VB-AKF. Also, unlike VB-AKF, our approach allows for the estimation of both $Q$ and $R$.

Fig. 5 shows the trajectory of estimated $R$ by VB-AKF (for $\rho = 1 - \exp(-4)$), IMM (for 111 noise models) and the proposed method. The proposed method and IMM provide measurement noise variance estimates that can track the true variance more closely than the VB-AKF. The correct choice of $\rho$ is critical to the success of VB-AKF.

**D. An Air Traffic Control (ATC) Scenario**

We consider an air traffic control (ATC) scenario used in [3]. The ground truth is a target moving with a constant speed of 250 $m/s$ with 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 $2^\circ/s$ for 30 seconds at $k = 100$, then continues straight for 70 seconds (until $k = 200$), at which time it turns right with $1^\circ/s$ for 45 seconds (until $k = 245$), then left with $1^\circ/s$ for 90 seconds (until $k = 335$), then right with $1^\circ/s$ 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 $\theta$) by a radar located at $[\xi_0, \eta_0] = [-10^4, 0]$, with

$$r = \sqrt{(\xi - \xi_0)^2 + (\eta - \eta_0)^2} \quad (39)$$

$$\theta = \tan^{-1}\left(\frac{\eta - \eta_0}{\xi - \xi_0}\right) \quad (40)$$

with additive white Gaussian noise with covariance $R = \text{diag}[2500 m^2, (1^\circ)^2]$. In this example, we used a Kalman filter based on a second order linear kinematic model (WNA) with process noise of standard deviation 1 $m/s^2$ described in (41).

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \ H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \ \Gamma = \begin{bmatrix} T^2/2 & 0 & 0 \\ T & 0 & 0 \\ 0 & T^2/2 & 0 \\ 0 & 0 & T \end{bmatrix} \quad (41)$$

An 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 (42) and (43) are used. The process noise standard deviations used in the CT model were 3 $m/s^2$ and 0.1 $^\circ/s^2$ for the UM and turn rate of the state, respectively.

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \ \Gamma = \begin{bmatrix} T^2/2 & 0 & 0 \\ T & 0 & 0 \\ 0 & T^2/2 & 0 \\ 0 & 0 & T \end{bmatrix} \quad (43)$$

The mode transition probability matrix $\pi$ in (44) is used for IMM estimator.

$$\pi = \begin{bmatrix} 0.95 & 0.05 \\ 0.10 & 0.90 \end{bmatrix} \quad (44)$$
Fig. 6 shows the averaged tracking results of target motion over 100 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.

![Target trajectory](image)

**Fig. 6: Target trajectory (100 MC runs)**

As shown in Fig. 7a, the proposed approach based on CT model (even 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 Kalman filter 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. 7b. The proposed approach can also track the target velocity close to its true value as shown in Fig. 7c and Fig. 7d.

![Comparison of optimization algorithms for ATC motion estimation](image)

**Fig. 7: Comparison of optimization algorithms for ATC motion estimation (100 MC runs)**

1. **Scenario when both Q and R vary**
   In this example, we assume that the observation samples are generated by Model 2 as in (46). Model 1 is assumed to be

   \[
   F^1 = \begin{bmatrix} 1 & 1 \\ -0.1 & 0.1 \end{bmatrix}, \quad H^1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \Gamma^1 = \begin{bmatrix} 1 \\ 0.4 \end{bmatrix}
   \]  

   Model 2 is

   \[
   F^2 = \begin{bmatrix} 0.8 & 1 \\ -0.4 & 0 \end{bmatrix}, \quad H^2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \Gamma^2 = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}
   \]  

   Fig. 8 shows the trajectory of estimated parameters. The mode probability of 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 where R changes continuously**
   We consider the same scenario as in (37) with two dynamic models.
Model 1;
\[ F^1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.1 \\ 0 & -0.002 & 1 \end{bmatrix}, \quad H^1 = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}, \quad \Gamma^1 = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} \quad (47) \]

Model 2;
\[ F^2 = \begin{bmatrix} 0.99 & 0 & 0 \\ 0.001 & 1 & 0.1 \\ 0 & 0 & 1 \end{bmatrix}, \quad H^2 = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}, \quad \Gamma^2 = \begin{bmatrix} 0.01 \\ 0.05 \\ 0.1 \end{bmatrix} \quad (48) \]

Table V shows the RMSE of estimated \( R \) over 100 MC runs by the multiple-model approach. The multiple-model method estimates the parameters in parallel, and then selects a suitable model based on mode probabilities. Note that the measurements are generated using Model 1 as in (47). The VB-AKF provides the best estimate when the hyper-parameter \( \rho = 1 - \exp(-3) \), but the RMSE value is quite sensitive to the selection of \( \rho \). When the computation time needed for tuning the hyper-parameter \( \rho \) 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 a better computational efficiency by a factor of 66 over the IMM.

**IV. CONCLUSION AND FUTURE WORK**

In this paper, we presented a single-pass stochastic gradient descent (SGD) algorithm for the noise covariance estimation in adaptive Kalman filters that are an order of magnitude faster than the batch method or the multi-pass sequential algorithm and has acceptable state estimation root mean square error (RMSE). This algorithm is valid for streaming data from non-stationary systems, where the noise covariances can occasionally exhibit abrupt, but finite, changes, as well as for multiple models. The computational efficiency of the new algorithm stems from recursive fading memory estimation of the sample cross-correlations of the innovations, accelerated SGD algorithms and single-pass computations. Evaluation of the
The proposed method on a number of test cases demonstrated its computational efficiency, accuracy and filter consistency over 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; and 3) explore the utility of the covariance estimation algorithm as an alternative to IMMs.

REFERENCES

[1] W. Arnold and A. Laub, “Generalized eigenproblem algorithms and software for algebraic riccati equations,” Proceedings of the IEEE, vol. 72, no. 12, pp. 1746–1754, 1984.

[2] F. Auger, M. Hilaire, J. M. Guerrero, E. Monmasson, T. Orlowska-Kowalska, and S. Katsura, “Industrial applications of the kalman filter: A review,” IEEE Transactions on Industrial Electronics, 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. John Wiley & Sons, 2004.

[4] R. Battiti, “Accelerated backpropagation learning: Two optimization methods.” Complex systems, 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, ser. Athena scientific optimization and computation series. Athena Scientific, 2016.

[7] S. Boyd, L. Xiao, and A. Mutapcic, “Subgradient methods,” lecture notes of EE392o, Stanford University, Autumn Quarter, vol. 2004, pp. 2004–2005, 2003.

[8] G. H. Golub and C. F. Van Loan, Matrix computations. JHU press, 2013, vol. 3.

[9] P. Hanlon and P. Maybeck, “Multiple-model adaptive estimation using a residual correlation kalman filter bank.” IEEE Transactions on Aerospace and Electronic Systems, vol. 36, no. 2, pp. 393–406, 2000.

[10] R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems,” Journal of Basic Engineering, vol. 82, no. 1, pp. 35–45, 03 1960.

[11] R. Killick, P. Fearnhead, and I. A. Eckley, “Optimal detection of changepoints with a linear computational cost,” Journal of the American Statistical Association, vol. 107, no. 500, pp. 1590–1598/12, 2012.

[12] 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. 99 220–99 234, 2021.

[13] ——, “A single-pass noise covariance estimation algorithm in adaptive kalman filtering for non-stationary systems,” 2021 IEEE 24th International Conference on Information Fusion (FUSION), pp. 556–563, 2021.

[14] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.

[15] X.-R. Li and Y. Bar-Shalom, “Multiple-model estimation with variable structure,” IEEE Transactions on Automatic control, vol. 41, no. 4, pp. 478–493, 1996.

[16] R. Mehra, “On the identification of variances and adaptive kalman filtering,” IEEE Transactions on automatic control, vol. 15, no. 2, pp. 175–184, 1970.

[17] C. Neethling and P. Young, “Comments on identification of optimum filter steady-state gain for systems with unknown noise covariances,” IEEE Transactions on Automatic control, vol. 19, no. 5, pp. 623–625, 1974.

[18] B. J. Odellson, M. R. Rajamani, and J. B. Rawlings, “A new autocovariance least-squares method for estimating noise covariances,” Automatica, vol. 42, no. 2, pp. 303–308, 2006.

[19] S. Sarkka and A. Nummenmaa, “Recursive noise adaptive kalman filtering by variational bayesian approximations,” IEEE Transactions on Automatic control, vol. 54, no. 3, pp. 596–600, 2009.

[20] K. Tajima, “Estimation of steady-state kalman filter gain,” IEEE Transactions on Automatic Control, vol. 23, no. 5, pp. 944–945, 1978.

[21] T. Tieleman and G. Hinton, “Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude,” COURSERA: Neural networks for machine learning, vol. 4, no. 2, pp. 26–31, 2012.

[22] M. D. Zeiler, “Adadelta: an adaptive learning rate method,” arXiv preprint arXiv:1212.5701, 2012.

[23] 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. 59 362–59 388, 2020.

HEE-SEUNG KIM (Member, IEEE) received the B.S. and M.S. degrees in the department of computer and communication engineering from the Chungbuk National University, Cheongju, South Korea, in 2011 and 2013, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT, USA.

He was a researcher with 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 (Member, IEEE) received the B.S. and M.Eng. degrees in electrical and computer engineering from the University of Connecticut, Storrs, CT in 2013 and 2017, respectively, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering Department, under the advisement of Dr. K. R. Pattipati.

He was an Electrical Engineer at General Dynamics Electric Boat, Groton, CT 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.

KRISHNA R. PATTIPATI (Life Fellow, IEEE) received the B. Tech. degree in Electrical engineering with highest honors from the Indian Institute of Technology, Kharagpur, in 1975, and the M.S. and Ph.D. degrees in Systems Engineering from UCONN, Storrs, in 1977 and 1980, respectively. He was with ALPATECH, Inc., Burlington, MA from 1980 to 1986. He has been with the department of Electrical and Computer Engineering at UCONN since 1986, where he is currently an Emeritus 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 over 500 scholarly journal and conference papers in these areas. He is a co-founder 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 has served as the Editor-
in-Chief of 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 Diag-
nosis," 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, and the 2003 AAUP Research Excellence Award at UCONN.
He is an elected Fellow of IEEE for his contributions to discrete-
optimization algorithms for large-scale systems and team decision-
making and of the Connecticut Academy of Science and Engineering.

H.-S. Kim et al.: A Single-pass Noise Covariance Estimation Algorithm in Multiple-model Adaptive Kalman Filtering 13