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

HEE-SEUNG KIM*, LINGYI ZHANG[†], ADAM BIENKOWSKI[‡], KRISHNA R. PATTIPATI[§]

Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269, USA

Email: *hee-seung.kim@uconn.edu, [†]lingyi.zhang@uconn.edu,

[‡]adam.bienkowski@uconn.edu, [§]krishna.pattipati@uconn.edu

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.

Index Terms—Adaptive Kalman filtering, noise covariance estimation, RMS Prop, stochastic gradient descent, sequential, fading memory, non-stationary

I. INTRODUCTION

The Kalman filter (KF) [1] is the optimal minimum mean square error (MMSE) state estimator for linear systems with Gaussian white process and measurement noises, and is the best linear state estimator when the noise variables are Gaussian with known covariances. This algorithm has received a huge interest from numerous industrial applications, such as robotics, system diagnosis, navigation and weather forecasting, to name a few [15], [16]. However, the noise covariances are mostly unknown or only partially known in many practical applications.

For estimating the unknown covariances, in [7], we first derived the necessary and sufficient conditions for identifiability of noise covariances, and then presented a *batch* optimization algorithm for estimating them to minimize the sum of the normalized temporal cross-correlations of the innovation sequence. The motivation for this optimization metric stems from the fact that the innovations of an optimal Kalman filter are white, meaning that they are uncorrelated over time [16]. In [19], 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 and a method to detect changes in noise covariances when applied to non-stationary systems. In this paper, we propose

a single-pass adaptive Kalman filtering algorithm for non-stationary systems wherein process and measurement noise covariances can occasionally jump up or down by an unknown magnitude.

A. Prior work

The key to process noise and measurement noise covariance estimation is an expression for the covariance of the state estimation error and of the innovations of any stable suboptimal filter as a function of noise covariances. This expression serves as a fundamental building block for the correlation-based methods for covariance estimation. Pioneering contributions using this approach were made by [3]–[6].

In [19], we proposed a sequential mini-batch estimation method with adaptive step size rules to improve the computational efficiency of the batch algorithm in [7]. We applied the sequential fading memory mini-batch estimates of the innovation correlations when updating the filter gain. We also enhanced the computational efficiency of the sequential algorithm via iteration-dependent dynamic thresholds for convergence. For non-stationary systems, the method in [19] required a change-point detection algorithm. The change-point detection algorithm described in [17] extracts the time points of abrupt changes in unknown noise covariances, given the innovation sequence.

Särkkä and Nummenmaa [2] proposed a variational Bayesian method for the joint recursive estimation of the dynamic state and measurement noise parameters in linear state space models. This method is based on forming a separable variational approximation to the joint posterior distribution of states and noise parameters at each time step separately. This method does not consider changes in process noise. In general, the variational methods require tuning parameters to converge to the correct parameters and often converge to a local minimum.

The limitation of previous work [19] is that the performance of noise covariance estimation algorithm depends on the accuracy of the change-point detection algorithm because the sequential estimation algorithm is invoked for samples between two consecutive change points. This approach also requires multiple passes through the data for the noise covariance estimation to converge, and it has a relatively high computation cost. The present paper seeks to overcome these two limitations.

B. Contribution and organization of the paper

In this paper, we present a single-pass, sequential mini-batch estimation method as an extension of the work in [19] for non-stationary systems. Our new approach enables the estimates of the measurement and process noise covariances without the use of a change-point detection algorithm. We enhance the computational efficiency by using a single-pass through the 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. We validate the proposed method on several non-stationary system test cases.

The paper is organized as follows. In Section II, we provide an overview of the sequential mini-batch gradient descent method for estimating the unknown noise covariances, including the fading memory filter-based innovation correlation estimation, and stochastic gradient descent (SGD) update of the Kalman gain. Section III provides evidence that the proposed method can track noise covariances in non-stationary systems and that our single-pass algorithm is computationally-efficient. Lastly, we conclude the paper and discuss our future work in Section IV.

II. SEQUENTIAL MINI-BATCH GRADIENT DESCENT METHOD FOR ESTIMATING Q AND R

The linear discrete-time stochastic dynamic system is given by

$$x(k+1) = Fx(k) + \Gamma v(k) \quad (1)$$

$$z(k) = Hx(k) + w(k) \quad (2)$$

where $x(k)$ is the state vector, $z(k)$ is the measurement vector, F and H are the state transition matrix and the measurement matrix of the system, respectively, and Γ is the noise gain matrix. Here, the process noise $v(k)$ and the measurement noise $w(k)$ are zero-mean white Gaussian noises with *unknown* process noise covariance $Q(k)$ and *unknown* measurement noise covariance $R(k)$, respectively. Note that the initial state error and the two noise processes are assumed to be mutually independent. We assume that $Q(k)$ and $R(k)$ are piece-wise constant such that the filter reaches a steady state between any two jumps of unknown magnitude.

When $Q(k)$ and $R(k)$ are known, the Kalman filter involves consecutive processes of prediction and update given by

$$\hat{x}(k+1|k) = F\hat{x}(k|k) \quad (3)$$

$$\nu(k+1) = z(k+1) - H\hat{x}(k+1|k) \quad (4)$$

$$\hat{x}(k+1|k+1) = \hat{x}(k+1|k) + W(k+1)\nu(k+1) \quad (5)$$

$$P(k+1|k) = FP(k|k)F' + \Gamma Q(k)\Gamma' \quad (6)$$

$$S(k+1) = HP(k+1|k)H' + R(k) \quad (7)$$

$$W(k+1) = P(k+1|k)H'S(k+1)^{-1} \quad (8)$$

$$P(k+1|k+1) = (I_{n_x} - W(k+1)H)\bar{P}(I_{n_x} - W(k+1)H)' + W(k+1)R(k)W(k+1)' \quad (9)$$

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 system dynamics, the updated state error covariance $P(k|k)$ at time index k and the process noise covariance, $Q(k)$. 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 $S(k+1)$ (which in turn depends on the measurement noise covariance, $R(k)$ and the predicted state error covariance $P(k+1|k)$). The updated state error covariance $P(k+1|k+1)$ is computed via (9); this corresponds to Joseph form, which is less sensitive to round-off error because it guarantees that the updated state covariance matrix will remain positive definite.

A. Recursive fading memory-based Innovation Correlation Estimation

We compute the sample correlation matrix $\hat{C}_{seq}^k(i)$ at sample k for time lag i as a weighted combination of the correlation matrix $\hat{C}_{seq}^{k-1}(i)$ at the previous sample $(k-1)$ for time lag i and the samples of innovations $\nu(k-i)$ and $\nu(k)$. 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 proposed algorithm makes it amenable in estimating Q and R in non-stationary systems.

The current M sample correlation matrices are recursively used as the initial values for the next pairs of samples.

$$\hat{C}_{seq}^k(i) = (1-\lambda)\nu(k-i)\nu(k)' + \lambda\hat{C}_{seq}^{k-1}(i), \quad (10)$$

$$\hat{C}_{seq}^0(i) = 0, i = 0, 1, 2, \dots, M-1; k = M, \dots, N_k \quad (11)$$

B. Objective function and the Gradient

The ensemble cross-correlation of a steady-state suboptimal Kalman filter is related to the closed-loop filter matrix $\bar{F}$, the matrix F , the measurement matrix H , the predicted covariance matrix $\bar{P}$, filter gain W and the innovation covariance, $C(0)$ via [5], [6]

$$C(i) = E[\nu(k)\nu(k-i)'] = H\bar{F}^{i-1}F[\bar{P}H' - WC(0)] \quad (12)$$

The objective function J , formulated in [7], involves minimization of the sum of normalized $C(i)$ with respect to the corresponding diagonal elements of $C(0)$ for $i > 0$. Formally, we can define the objective function J to be minimized with respect to W as

$$J = \frac{1}{2} \text{tr} \left\{ \sum_{i=1}^{M-1} \left[\text{diag}(C(0)) \right]^{-\frac{1}{2}} C(i)' \times \left[\text{diag}(C(0)) \right]^{-1} C(i) \left[\text{diag}(C(0)) \right]^{-\frac{1}{2}} \right\} \quad (13)$$

where $\text{diag}(C)$ denotes the diagonal matrix of C or equivalently the Hadamard product of an identity matrix with C . We can rewrite the objective function by substituting (13) into (12) as

$$J = \frac{1}{2} \text{tr} \left\{ \sum_{i=1}^{M-1} \Theta(i) X \varphi X' \right\} \quad (14)$$

where

$$\Theta(i) = [H \bar{F}^{i-1} F]' \varphi [H \bar{F}^{i-1} F] \quad (15)$$

$$X = \bar{P} H' - W C(0) \quad (16)$$

$$\varphi = [\text{diag}(C(0))]^{-1} \quad (17)$$

The gradient of objective function $\nabla_W J$ can be computed as [7]

$$\begin{aligned} \nabla_W J = & - \sum_{i=1}^{M-1} [H \bar{F}^{i-1} F]' \varphi C(i) \varphi C(0) - F' Z F X \\ & - \sum_{l=0}^{i-2} [C(l+1) \varphi C(i)' \varphi H \bar{F}^{i-l-2}]' \end{aligned} \quad (18)$$

The Z term in (18) is computed by the Lyapunov equation.

$$\begin{aligned} Z = & \bar{F}' Z \bar{F} + \frac{1}{2} \sum_{i=1}^{M-1} (H \bar{F}^{i-1} F)' \varphi C(i) \varphi H \\ & + ((H \bar{F}^{i-1} F)' \varphi C(i) \varphi H)' \end{aligned} \quad (19)$$

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

C. Updating gain W sequentially

Let B be the mini-batch size and let $K = N_k/B$ be the number of mini-batches (we assume that N_k is divisible by B for simplicity). While the mini-batch gradient descent sequentially updates the M sample covariance matrices at every input 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. The generic form of gain update is

$$W^{(r+1)} = W^{(r)} - \alpha^{(r)} \nabla_{W^{(r)}} J \quad (20)$$

The incremental gradient algorithm in (20) can be sped up by adaptively selecting the step size $\alpha^{(r)}$. Our results in [19] showed that Adam [9] and RMSProp [10] have the best accuracy and convergence among all the accelerated SGD algorithms (e.g., bold driver [11], constant, subgradient [18], Adadelta [8]) studied. Here, we show the performance results of our algorithm using the RMSProp update with a mini-batch size of 64.

RMSProp keeps track of the moving average of the squared incremental gradients for each gain element for adapting the step size element-wise.

$$\tau_{r,ij} = \gamma \tau_{r-1,ij} + (1 - \gamma) [(\nabla_{W^{(r)}} J)_{ij}]^2; \tau_0 = 0 \quad (21)$$

$$\alpha_{ij}^{(r)} = \frac{\alpha^{(0)}}{\sqrt{\tau_{r,ij} + \epsilon}} \quad (22)$$

Here, $\gamma = 0.9$ is the default value and $\epsilon = 10^{-8}$ to prevent division by zero.

D. Estimation of Q and R

1) *Estimation of R* : Let us define $\mu(k), k = 1, 2, \dots, N_k$, as the post-fit residual sequence of the Kalman filter, which is related to the innovations $\nu(k), k = 1, 2, \dots, N_k$ via

$$\mu(k) = z(k) - H \hat{x}(k|k) = (I_{n_z} - HW(k)) \nu(k) \quad (23)$$

From the joint covariance of the innovation sequence $\nu(k)$ and the post-fit residual sequence $\mu(k)$, and the Schur determinant identity [12], [13], one can show that in the steady-state (assuming constant gain, W and constant Q and R over large time intervals) [7]

$$G = E[\mu(k) \mu(k)'] = RS^{-1}R \quad (24)$$

where S is the steady-state innovation covariance. Because (24) can be interpreted as a simultaneous diagonalization problem in linear algebra [13] or as a continuous-time algebraic Riccati equation, the measurement covariance R 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 [7], [14].

2) *Estimation of Q* : Given the estimated R , we can compute the process noise covariance Q and the steady-state state updated covariance P . This requires an iterative process because Q and P are coupled in the general case, Wiener process being an exception where an explicit non-iterative solution $Q = WSW'$ is possible in the steady-state [7]. Letting t and l denote the iteration indices starting with $t = 0$ and $l = 0$, and using an initial $Q^{(0)} = WSW'$, we initialize the steady-state updated covariance matrix P as the solution of the Lyapunov equation in (25)

$$\begin{aligned} P^{(0)} = & \tilde{F} P^{(0)} \tilde{F}' + W R W' \\ & + (I_{n_x} - W H) \Gamma Q^{(t)} \Gamma' (I_{n_x} - W H) \end{aligned} \quad (25)$$

where $\tilde{F} = (I_{n_x} - W H) F$. We iteratively update P as in (26) until convergence

$$P^{(l+1)} = \left[(F P^{(l)} F' + \Gamma Q^{(t)} \Gamma')^{-1} + H' R^{-1} H \right]^{-1} \quad (26)$$

Given the converged P , Q will be updated in the t -loop until the estimate of Q converges. Note that λ_Q is a regularization parameter used for ill-conditioned estimation problems.

$$Q^{(t+1)} = \Gamma^\dagger \left[(P + W S W' - F P F')^{(t+1)} + \lambda_Q I_{n_x} \right] (\Gamma')^\dagger \quad (27)$$

The pseudo code for the sequential mini-batch SGD estimation algorithm for a non-stationary system is included as Algorithm 1.

Algorithm 1 Pseudo code of sequential mini-batch SGD algorithm for a non-stationary system

---

```

1: input:  $W^0, Q^0, R^0, \alpha^0, B \triangleright W^0$ : initial gain,  $Q^0$ : initial
    $Q, R^0$ : initial  $R, \alpha^0$ : initial step size,  $B$ : batch size
2:  $r = 0$   $\triangleright$  Initialize the updating index  $r$ 
3: for  $k = 1$  to  $N_k$  do  $\triangleright N_k$ : Num. samples
4:   compute innovation correlations  $\nu(k)$ 
5:   if  $k > N_b + M$  then  $\triangleright N_b$ : Num. burn-in samples
6:     compute  $\hat{C}$ 
7:     if  $\text{Mod}(k, B) = 0$  then
8:       compute objective function  $J$ 
9:       compute gradient  $\nabla_W J$ 
10:      update the step size  $\alpha^{(r)}$ 
11:       $W_{ij}^{(r+1)} = W_{ij}^{(r)} - \alpha_{ij}^{(r)} [\nabla_{W^{(r)}} J]_{ij}$ 
12:      update  $R^{(r+1)}$  and  $Q^{(r+1)}$ 
13:       $r = r + 1$ 
14:    end if
15:  end if
16: end for

```

III. NUMERICAL EXAMPLES

We apply our proposed method to the system used in [3], but modified to have non-stationary noise covariance matrices. The system matrices are assumed to be as follows.

$$F = \begin{bmatrix} 0.8 & 1 \\ -0.4 & 0 \end{bmatrix}, H = \begin{bmatrix} 1 & 0 \end{bmatrix}, \Gamma = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} \quad (28)$$

In this paper, the non-stationarity is assumed to arise from abrupt changes in noise covariances occasionally, but 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. We considered three non-stationary scenarios for this system with 50,000 measurement samples: 1) both Q and R change every 10,000 samples (see Section III-A), 2) Only Q changes every 10,000 samples, and 3) Only R changes every 10,000 samples (see Section III-B).

In the estimation procedure, we set the number of burn-in samples $N_b = 50$, and the number of lags $M = 5$. Here, we show the performance results of our algorithm using the RMSProp update. 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 will show that the proposed algorithm works with varying mini-batch sizes. We also simulate 100 MC runs with varying number of observation samples to find the minimum number of samples required for estimating the unknown variance parameters accurately.

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 $\bar{P}$ is relatively low when mini-batch size ≤ 32 . 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 with their concomitant true values, while varying the number of observation samples $N_k = 5000, 10000, 20000, 30000$ and 50000. Fig. 1 shows the results of 100 Monte Carlo simulations in estimating Q and R when the number of observed samples is varied. Estimated Q and R are closer to the corresponding true values when $N_k \geq 30000$. Fig. 1 indicates the accuracy of estimating unknown noise covariances increases with an increase in the number of samples as expected.

Fig. 1: Monte Carlo Simulation in varying observation samples in varying Q and R

The proposed algorithm can track Q and R accurately as shown in Fig. 2. The trajectory of Q and R estimates can be smoothed by a simple first order fading memory filter with a smoothing weight of 0.9. Fig. 2c 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 transient peaks when the noise covariances change abruptly.

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

(a) Mini-batch size of 16

	R			Q			P_{11}			P_{22}			W_{11}			W_{21}		
	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE
1	0.42	0.42	0.18	0.04	0.04	0.15	0.11	0.25	0.32	0.02	0.05	0.06	0.21	0.23	0.16	-0.01	-0.01	0.08
2	0.81	0.79		0.64	0.65		1.27	1.05		0.24	0.20		0.61	0.61		0.07	0.07	
3	0.49	0.48		0.25	0.25		0.54	0.71		0.10	0.14		0.53	0.54		0.03	0.04	
4	0.16	0.17		1.00	1.02		1.31	1.32		0.27	0.27		0.89	0.88		0.30	0.29	
5	0.64	0.62		0.09	0.09		0.24	0.25		0.05	0.05		0.28	0.29		-0.01	-0.01	

(b) Mini-batch size of 32

	R			Q			P_{11}			P_{22}			W_{11}			W_{21}		
	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE
1	0.42	0.42	0.17	0.04	0.04	0.15	0.11	0.16	0.29	0.02	0.03	0.05	0.21	0.22	0.16	-0.01	-0.01	0.08
2	0.81	0.80		0.64	0.65		1.27	1.11		0.24	0.21		0.61	0.61		0.07	0.07	
3	0.49	0.49		0.25	0.25		0.54	0.60		0.10	0.12		0.53	0.53		0.03	0.04	
4	0.16	0.18		1.00	1.02		1.31	1.33		0.27	0.28		0.89	0.88		0.30	0.29	
5	0.64	0.64		0.09	0.09		0.24	0.25		0.05	0.05		0.28	0.28		-0.01	-0.01	

(c) Mini-batch size of 64

	R			Q			P_{11}			P_{22}			W_{11}			W_{21}		
	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE
1	0.42	0.42	0.16	0.04	0.04	0.15	0.11	0.11	0.30	0.02	0.02	0.05	0.21	0.21	0.16	-0.01	-0.01	0.09
2	0.81	0.80		0.64	0.65		1.27	1.28		0.24	0.24		0.61	0.61		0.07	0.07	
3	0.49	0.49		0.25	0.25		0.54	0.55		0.10	0.10		0.53	0.53		0.03	0.03	
4	0.16	0.17		1.00	1.03		1.31	1.34		0.27	0.28		0.89	0.89		0.30	0.30	
5	0.64	0.64		0.09	0.09		0.24	0.25		0.05	0.05		0.28	0.28		-0.01	-0.01	

(d) Mini-batch size of 128

	R			Q			P_{11}			P_{22}			W_{11}			W_{21}		
	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE
1	0.42	0.42	0.17	0.04	0.04	0.15	0.11	0.11	0.31	0.02	0.02	0.06	0.21	0.21	0.16	-0.01	-0.01	0.09
2	0.81	0.81		0.64	0.65		1.27	1.28		0.24	0.24		0.61	0.61		0.07	0.07	
3	0.49	0.49		0.25	0.26		0.54	0.56		0.10	0.11		0.53	0.54		0.03	0.03	
4	0.16	0.18		1.00	1.02		1.31	1.36		0.27	0.28		0.89	0.88		0.30	0.30	
5	0.64	0.64		0.09	0.09		0.24	0.25		0.05	0.05		0.28	0.28		-0.01	-0.01	

(a) Estimated Q

(b) Estimated R

(c) Averaged NIS of SGD method in the non-stationary system

Fig. 2: Trajectory of Q and R in varying Q and R system

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.

Table IIa shows the MC simulation results for the system considered here with varying Q . The estimated process noise covariance trajectory and the estimated measurement noise covariance are very close to their true values.

(a) Estimated Q

(b) Estimated R

Fig. 3: Trajectory of Q and R 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. As shown in Table IIb, the noise parameters can be estimated quite accurately by the sequential algorithm.

TABLE II: Monte Carlo Simulation (100 Runs; 50,000 samples; $M = 5$; RMSProp; Batch size = 64)(a) The case of varying Q

	R			Q			$\bar{P}_{11}$			$\bar{P}_{22}$			W_{11}			W_{21}		
	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE
1	0.36	0.35	0.15	0.04	0.04	0.13	0.11	0.12	0.30	0.02	0.02	0.05	0.24	0.25	0.16	-0.01	-0.02	0.08
2	0.36	0.37		0.64	0.64		1.07	1.07		0.20	0.20		0.75	0.74		0.15	0.15	
3	0.36	0.36		0.25	0.25		0.51	0.51		0.10	0.10		0.59	0.59		0.06	0.05	
4	0.36	0.38		1.00	0.99		1.52	1.52		0.30	0.30		0.81	0.79		0.20	0.20	
5	0.36	0.36		0.09	0.09		0.22	0.23		0.04	0.04		0.38	0.38		0.001	-0.001	

(b) The case of varying R

	R			Q			$\bar{P}_{11}$			$\bar{P}_{22}$			W_{11}			W_{21}		
	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE
1	0.04	0.05	0.14	0.36	0.36	0.11	0.45	0.46	0.42	0.10	0.10	0.08	0.92	0.91	0.14	0.34	0.33	0.11
2	0.64	0.63		0.36	0.38		0.77	0.77		0.15	0.15		0.55	0.55		0.04	0.03	
3	0.25	0.25		0.36	0.36		0.63	0.62		0.12	0.12		0.72	0.71		0.13	0.11	
4	1.00	0.96		0.36	0.38		0.84	0.86		0.16	0.17		0.46	0.47		0.02	0.01	
5	0.09	0.09		0.36	0.37		0.51	0.51		0.10	0.10		0.85	0.84		0.24	0.24	

(a) Estimated Q (b) Estimated R Fig. 4: Trajectory of Q and R in varying R system

C. Scenario where R changes continuously

We explore the scenario where R changes continuously (but Q is constant), as in the example used in [2]. The system matrices are

$$F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.1 \\ 0 & -0.002 & 1 \end{bmatrix}, H = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}, \Gamma = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} \quad (29)$$

The sampling interval, h is 0.1 second. 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:

$$R(k) = \begin{cases} 0.2 + 0.4(1 + \tanh(0.1h(k - 1000))), & k \leq 1500 \\ 0.2 + 0.4(1 + \tanh(0.1h(2000 - k))), & \text{otherwise} \end{cases} \quad (30)$$

Variational Bayesian Adaptive Kalman Filter (VB-AKF) algorithm described in [2] models this change by simply “spreading” its previous approximate posteriors via a heuristic

factor ρ . Our approach does not require such heuristic approximations.

Table III shows the RMSE of estimated R by VB-AKF, IMM and the proposed method for various values of ρ (for VB-AKF) and the number of noise models (for IMM). The performance of VB-AKF is very sensitive to the selection of ρ as seen in Table III; the RMSE can change by a factor of 3 if the wrong value is selected for ρ . 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 [2] 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 expensive.

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 ρ , which is rarely known in practice. 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 can track close to the true variance when compared to VB-AKF. The correct choice of ρ is critical to the success of VB-AKF.

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 of faster than the batch method or the multi-pass sequential algorithm for similar root mean square error (RMSE). This

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

Type	Run time (sec)	RMSE	Description
vbakf	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	0.3354	4 noise models
	113	0.3051	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
SGD	36	0.0941	RMSProp update, Mini-batch size = 64

Fig. 5: Trajectory of estimated R

algorithm is valid for streaming data from non-stationary systems, where the noise covariances can occasionally exhibit abrupt, but finite, changes. 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 proposed method on a number of test cases demonstrated its computational efficiency, accuracy and filter consistency.

In the future, we plan to pursue a number of research avenues, 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 in adaptive interacting multiple model (IMM) filters or as an alternative to IMMs.

ACKNOWLEDGMENT

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, and in part by the NASA's Space Technology Research Grants Program under Grant 80NSSC19K1076.

REFERENCES

- [1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, Mar. 1960.
- [2] S. Särkkä and A. Nummenmaa, "Recursive Noise Adaptive Kalman Filtering by Variational Bayesian Approximations," in *IEEE Transactions on Automatic Control*, vol. 54, no. 3, pp. 596–600, March 2009.
- [3] C. Neethling and P. Young, "Comments on 'Identification of optimum filter steady-state gain for systems with unknown noise covariances,'" *IEEE Trans. Autom. Control*, vol. AC-19, no. 5, pp. 623–625, Oct. 1974.
- [4] B. J. Odelson, 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, Feb. 2006.
- [5] R. Mehra, "On the identification of variances and adaptive Kalman filtering," *IEEE Trans. Autom. Control*, vol. AC-15, no. 2, pp. 175–184, Apr. 1970.
- [6] P. R. Bélanger, "Estimation of noise covariance matrices for a linear time-varying stochastic process," *Automatica*, vol. 10, no. 3, pp. 267–275, May 1974.
- [7] 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, Mar. 2020.
- [8] M. D. Zeiler, "ADADELTA: An adaptive learning rate method," 2012, arXiv:1212.5701. [Online]. Available: <http://arxiv.org/abs/1212.5701>
- [9] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, arXiv:1412.6980. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [10] T. Tieleman and G. Hinton, "Lecture 6.5-RMSPROP: Divide the gradient by a running average of its recent magnitude," COURSE, Neural Netw. Mach. Learn., vol. 4, no. 2, pp. 26–31, 2012.
- [11] R. Battiti, "Accelerated backpropagation learning: Two optimization methods," *Complex Syst.*, vol. 3, no. 4, pp. 331–342, 1989.
- [12] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.
- [13] C. F. Van Loan and G. H. Golub, *Matrix Computations*, 4th ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 2013.
- [14] W. F. Arnold and A. J. Laub, "Generalized eigenproblem algorithms and software for algebraic Riccati equations," *Proc. IEEE*, vol. 72, no. 12, pp. 1746–1754, Dec. 1984.
- [15] F. Auger, M. Hilaret, J. M. Guerrero, E. Monmasson, T. Orlowska-Kowalska and S. Katsura, "Industrial Applications of the Kalman Filter: A Review," in *IEEE Transactions on Industrial Electronics*, vol. 60, no. 12, pp. 5458–5471, Dec. 2013.
- [16] Y. Bar-Shalom, X.R. Li and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York: John Wiley Sons, 2001.
- [17] R. Killick, P. Fearnhead I. A. Eckley, "Optimal Detection of Change-points With a Linear Computational Cost," *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1590–1598, 2012.
- [18] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," Lecture notes of EE392a, Stanford University, Autumn Quarter, vol. 2004, pp. 2004–2005, 2003.
- [19] H.-S. Kim, L. Zhang, A. Bienkowski, K. R. Pattipati (2021): Noise Covariance Estimation in Adaptive Kalman Filtering via sequential Mini-batch Stochastic Gradient Descent Algorithms. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.14671887.v1>