

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,200

Open access books available

168,000

International authors and editors

185M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



Chapter

# Sequential Mini-Batch Noise Covariance Estimator

*Hee-Seung Kim, Lingyi Zhang, Adam Bienkowski,  
Krishna R. Pattipati, David Sidoti, Yaakov Bar-Shalom  
and David L. Kleinman*

## Abstract

Noise covariance estimation in an adaptive Kalman filter is a problem of significant practical interest in a wide array of industrial applications. Reliable algorithms for their estimation are scarce, and the necessary and sufficient conditions for identifiability of the covariances were in dispute until very recently. This chapter presents the necessary and sufficient conditions for the identifiability of noise covariances, and then develops sequential mini-batch stochastic optimization algorithms for estimating them. The optimization criterion involves the minimization of the sum of the normalized temporal cross-correlations of the innovations; this is based on the property that the innovations of an optimal Kalman filter are uncorrelated over time. Our approach enforces the structural constraints on noise covariances and ensures the symmetry and positive definiteness of the estimated covariance matrices. Our approach is applicable to non-stationary and multiple model systems, where the noise covariances can occasionally jump up or down by an unknown level. The validation of the proposed method on several test cases demonstrates its computational efficiency and accuracy.

**Keywords:** Adaptive Kalman filtering, Minimal polynomial, Noise covariance estimation, Stochastic gradient descent (SGD), Mini-batch SGD

## 1. Introduction

This chapter addresses the following learning problem: Given a vector time series and a library of models for the time evolution of the data (e.g., a Wiener process, a white noise acceleration model, also called nearly constant velocity model, or a white noise jerk model, also called nearly constant acceleration model), find suitable process and measurement noise covariances and select the best dynamic model for the time series. This problem is of considerable interest in a number of applications, such as fault diagnosis, robotics, signal processing, navigation, and target tracking, to name a few [1, 2].

The Kalman filter (KF) [3] is the optimal minimum mean square error (MMSE) state estimator for linear systems with mutually uncorrelated Gaussian white process

and measurement noises, and is the best linear state estimator when the noises are non-Gaussian with known covariances. However, the noise covariances are unknown or only partially known in many practical applications.

We derived the necessary and sufficient conditions for the identifiability of unknown noise covariances, and presented a batch optimization algorithm for their estimation using the sum of the normalized temporal cross-correlations of the innovation sequence as the optimization criterion [4]. 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 [2]. In [5], we proposed a sequential mini-batch stochastic gradient descent (SGD) algorithm that required multiple passes through the measurements for estimating noise covariances. We also presented its applicability to non-stationary systems by detecting changes in noise covariances. In this chapter, we present a practical single-pass stochastic gradient descent algorithm for noise covariance estimation in non-stationary systems. Extensions to multiple models where the system behavior can stem from a member of a known subset of models are discussed in [6].

## **1.1 Prior work**

The key to noise covariance estimation is an expression for the covariance of the state estimation error and of the innovations of any stable, but not necessarily optimal, filter as a function of noise covariances. This expression serves as a foundational building block for the correlation-based methods for noise covariance estimation. Pioneering contributions using this approach were made by [7–9]. Sarkka and Nummenmaa [10] proposed a recursive noise-adaptive Kalman filter for linear state space models using variational Bayesian approximations. However, the variational methods generally require tuning hyper-parameters to converge to the correct covariance parameters and these algorithms often converge to local minima.

In [5], we presented a computationally efficient and accurate sequential estimation algorithm that is a major improvement over the batch estimation algorithm in [4]. The novelties of this algorithm stem from its sequential nature and the use of mini-batches, adaptive step size rules and dynamic thresholds for convergence in the stochastic gradient descent (SGD) algorithm. The innovation cross-correlations are obtained by a sequential fading memory filter. We applied a change-point detection algorithm described in [11] to extract the change points in noise covariances for non-stationary systems.

This chapter seeks to develop a streaming algorithm that reads measurements exactly once, thus making it real-time and practical. The only caveat is that the changes in noise covariances are assumed to occasionally jump up or down by an unknown magnitude. Extensions of this algorithm to a multiple model setting may be found in [6].

## **1.2 Organization of the chapter**

The organisation of the chapter is as follows. Section 2 presents the mathematical formulation of the sequential mini-batch gradient descent algorithm for estimating the unknown noise covariances. In this section, we also present an overview of our approach based on a fading memory filter-based innovation correlation estimation, and an accelerated SGD update of the Kalman gain. In Section 3, we show that our

single-pass method can track unknown noise covariances in non-stationary systems. Lastly, we conclude the chapter with a brief summary of the contributions in Section 4.

## 2. Sequential mini-batch SGD method for estimating process and measurement noise covariances

Consider a discrete-time linear dynamic system

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

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

where  $x(k)$  is the  $n_x$ -dimensional state vector,  $v(k)$  is the sequence of zero-mean white Gaussian process noise with unknown process noise covariance  $Q(k)$  in the plant equation. The measurement equation,  $z(k)$  with  $n_z$ -dimensional vector, is given in (2). Here,  $w(k)$  is the sequence of zero-mean white Gaussian measurement noise with unknown measurement noise covariance  $R(k)$ . In the system,  $F$  and  $\Gamma$  are the  $n_x \times n_x$  state transition matrix and the noise gain matrix, respectively, and  $H$  is the  $n_z \times n_x$  measurement matrix. Here, the two noise sequences and the initial state error are assumed to be mutually uncorrelated. We assume that noise covariances  $Q(k)$  and  $R(k)$  are piecewise constant such that the filter reaches a steady-state between any two jumps of unknown magnitude.

Given  $Q(k)$  and  $R(k)$ , the Kalman filter involves the consecutive processes of prediction and update given by [2, 3].

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

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

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

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

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

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

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

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 is the Joseph form, which is less sensitive to round-off error because it guarantees that the updated state covariance matrix will remain positive definite.

## 2.1 Identifiability conditions for estimating $Q$ and $R$

The necessary and sufficient conditions for identifiability of the covariances in adaptive Kalman filters were in dispute until very recently [4, 7–9, 12]. When  $Q$  and  $R$  are unknown, consider the innovations corresponding to a stable, suboptimal closed-loop filter matrix  $\bar{F} = F(I_{n_x} - WH)$  given by [4, 13].

$$\nu(k) = H\bar{F}^m [x(k-m) - \hat{x}(k-m|k-m-1)] + \left\{ H \sum_{j=0}^{m-1} \bar{F}^{m-1-j} [\Gamma v(k-m+j) - FWw(k-m+j)] \right\} + w(k) \quad (10)$$

Given the innovation sequence (10), a weighted sum of innovations,  $\xi(k)$ , can be computed as

$$\xi(k) = \sum_{i=0}^m a_i \nu(k-i) \quad (11)$$

where the weights are the coefficients of the minimal polynomial of the closed-loop filter matrix  $\bar{F}$ ,  $\sum_{i=0}^m a_i \bar{F}^{m-i} = 0, a_0 = 1$ . It is easy to see that  $\xi(k)$  is the sum of two moving average processes driven by the process noise and measurement noise, respectively, given by [4].

$$\xi(k) = \sum_{l=1}^m \mathcal{B}_l v(k-l) + \sum_{l=0}^m \mathcal{G}_l w(k-l) \quad (12)$$

Here,  $\mathcal{B}_l$  and  $\mathcal{G}_l$  are given by

$$\mathcal{B}_l = H \left( \sum_{i=0}^{l-1} a_i \bar{F}^{l-i-1} \right) \Gamma \quad (13)$$

$$\mathcal{G}_l = \left[ a_l I_{n_z} - H \left( \sum_{i=0}^{l-1} a_i \bar{F}^{l-i-1} \right) FW \right]; \mathcal{G}_0 = I_{n_z} \quad (14)$$

Then, the cross-covariance between  $\xi(k)$  and  $\xi(k-j)$ ,  $\mathcal{L}_j$ , can be obtained as

$$\mathcal{L}_j = E[\xi(k)\xi(k-j)'] = \sum_{i=j+1}^m \mathcal{B}_i Q \mathcal{B}_{i-j}' + \sum_{i=j}^m \mathcal{G}_i R \mathcal{G}_{i-j}' \quad (15)$$

The noise covariance matrices  $Q = [q_{ij}]$  of dimension  $n_v \times n_v$  and  $R = [r_{ij}]$  of dimension  $n_z \times n_z$  are positive definite and symmetric. By converting the noise covariance matrices and the  $\mathcal{L}_j$  matrices to vectors, Zhang et al. [4] show that they are related by the noise covariance identifiability matrix  $\mathcal{I}$  given by

$$\mathcal{I} \begin{bmatrix} \text{vec}(Q) \\ \text{vec}(R) \end{bmatrix} = \begin{bmatrix} \mathcal{L}_0 \\ \mathcal{L}_1 \\ \vdots \\ \mathcal{L}_m \end{bmatrix} \quad (16)$$

As shown in [4], if the matrix  $\mathcal{I}$  has full rank, then the unknown noise covariance matrices,  $Q$  and  $R$ , are identifiable. Direct solution of linear equations in (16) for  $Q$  and  $R$  is highly ill-conditioned and is prone to numerical errors.

## 2.2 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 ( $k - 1$ ) and time lag  $i$ , and the samples of innovations  $\nu(k - i)$  and  $\nu(k)$ . The tuning parameter  $\lambda$ , a positive constant between 0 and 1, is the weight associated with the previous sample correlation matrix. The current  $M$  sample correlation matrices at time  $k$  are used as the initial values for the next pairs of samples for the recursive computation. Let us define the number of measurement samples as  $N$ . Then,

$$\hat{C}_{seq}^k(i) = (1 - \lambda)\nu(k - i)\nu(k)' + \lambda\hat{C}_{seq}^{k-1}(i), \quad (17)$$

$$\hat{C}_{seq}^0(i) = 0, i = 0, 1, 2, \dots, M - 1; k = M, \dots, N \quad (18)$$

## 2.3 Objective function and the gradient

The ensemble cross-correlation of a steady-state suboptimal Kalman filter is related to the closed-loop filter matrix  $\bar{F} = F(I_{n_x} - WH)$ , the matrix  $F$ , the measurement matrix  $H$ , the steady-state predicted covariance matrix  $\bar{P}$ , steady-state filter gain  $W$  and the steady-state innovation covariance,  $C(0)$  via [8, 9].

$$C(i) = E[\nu(k)\nu(k - i)'] = H\bar{F}^{i-1}F[\bar{P}H' - WC(0)] \quad (19)$$

To avoid the scaling effects of measurements, the objective function  $\Psi$  formulated in [4] involves a 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  $\Psi$  to be minimized with respect to  $W$  as

$$\Psi = \frac{1}{2}tr \left\{ \sum_{i=1}^{M-1} [diag(C(0))]^{-\frac{1}{2}} C(i)' [diag(C(0))]^{-1} C(i) [diag(C(0))]^{-\frac{1}{2}} \right\} \quad (20)$$

where  $diag(C)$  denotes the diagonal matrix of  $C$ . We can rewrite the objective function by substituting (20) into (19) as

$$\Psi = \frac{1}{2}tr \left\{ \sum_{i=1}^{M-1} \phi(i) X \phi X' \right\} \quad (21)$$

where

$$\phi(i) = [H\bar{F}^{i-1}F]' \varphi [H\bar{F}^{i-1}F] \quad (22)$$

$$X = \bar{P}H' - WC(0) \quad (23)$$



$$\varphi = [\text{diag}(C(0))]^{-1} \quad (24)$$

The gradient of objective function,  $\nabla_W \Psi$ , can be computed as [4].

$$\nabla_W \Psi = - \sum_{i=1}^{M-1} [H\bar{F}^{i-1}F]' \varphi C(i) \varphi C(0) - F'ZFX - \sum_{l=0}^{i-2} [C(l+1)\varphi C(i)' \varphi H\bar{F}^{i-l-2}]' \quad (25)$$

where

$$Z = \bar{F}'Z\bar{F} + \frac{1}{2} \sum_{i=1}^{M-1} (H\bar{F}^{i-1}F)' \varphi C(i) \varphi H + \left( (H\bar{F}^{i-1}F)' \varphi C(i) \varphi H \right)' \quad (26)$$

The  $Z$  term in (26) is computed by a Lyapunov equation; it is often small and can be neglected in (25) for computational efficiency.

In computing the objective function and the gradient, we replace  $C(i)$  by their sample estimates,  $\hat{C}_{seq}(i)$ . With this replacement, the noise covariance estimation becomes a data-dependent stochastic optimization/learning problem.

## 2.4 Estimation of $Q$ and $R$

### 2.4.1 Estimation of $R$

We define  $\mu(k)$  as the post-fit residual sequence of the Kalman filter, which is related to the innovations  $\nu(k)$  via

$$\mu(k) = z(k) - H\hat{x}(k|k) = (I_{n_z} - HW)\nu(k); k = 1, 2, \dots, N \quad (27)$$

From the joint covariance of the innovation sequence  $\nu(k)$  and the post-fit residual sequence  $\mu(k)$ , and the Schur determinant identity [14, 15], one can show that [4].

$$G = E[\mu(k)\mu(k)'] = RS^{-1}R \quad (28)$$

where  $S$  is the innovation covariance. Knowing the sampled estimates of  $G$  and  $S = \hat{C}_{seq}(0)$ , the measurement noise covariance  $R$  is estimated. Because (28) can be interpreted as a continuous-time algebraic Riccati equation or as a simultaneous diagonalization problem in linear algebra [15], the measurement noise covariance  $R$  can be estimated by solving a continuous-time Riccati equation as in [4, 16] or by solving the simultaneous diagonalization problem via Cholesky decomposition and eigen decomposition.

### 2.4.2 Estimation of $Q$

Since the process noise covariance  $Q$  and the steady-state updated covariance  $P$  are generally coupled,  $Q$  and  $P$  can be obtained via a Gauss–Seidel type iterative computation given the estimated  $R$ . Wiener process is an exception where an explicit non-iterative solution  $Q = WSW'$  is possible [4]. Let  $t$  and  $l$  denote the indices of iteration

starting with  $t = 0$  and  $l = 0$ . The initial steady-state updated covariance,  $P^0$ , can be computed as the solution of the Lyapunov equation given by

$$P^{(0)} = \tilde{F}P^{(0)}\tilde{F}' + WRW' + (I_{n_x} - WH)\Gamma Q^{(t)}\Gamma'(I_{n_x} - WH)'; Q^{(0)} = WSW' \quad (29)$$

where  $\tilde{F} = (I_{n_x} - WH)F$ . We iteratively update  $P$  as in (30) until convergence

$$P^{(l+1)} = \left[ \left( FP^{(l)}F' + \Gamma Q^{(t)}\Gamma' \right)^{-1} + H'R^{-1}H \right]^{-1} \quad (30)$$

Given the converged  $P$ ,  $Q$  will be updated in the  $t$ -loop until the estimate of  $Q$  converges.

$$Q^{(t+1)} = \Gamma^\dagger \left[ (P + WSW' - FPF')^{(t+1)} + \lambda_Q I_{n_x} \right] (\Gamma')^\dagger \quad (31)$$

where  $\lambda_Q$  is a regularization parameter used for ill-conditioned estimation problems.

## 2.5 Updating the gain $W$ sequentially

The estimation algorithm sequentially computes the  $M$  sample covariance matrices at every measurement sample  $k$  as in (17). Let  $B$  be the mini-batch size for updating the Kalman filter gain  $W$  in the SGD. Our proposed method updates the gain  $W$  when the sample index  $k$  is divisible by the mini-batch size  $B$ . When compared to the batch estimation algorithm, the sequential mini-batch SGD algorithm allows more opportunities to converge to a better local minimum of (20) by frequently updating the filter the gain [5]. The generic form of the gain update is given by

$$W^{(r+1)} = W^{(r)} - \alpha^{(r)} \nabla_{W^{(r)}} \Psi \quad (32)$$

where  $r$  is the updating index starting with  $r = 0$ . In our previous research [5], we explored the performance of accelerated SGD methods (e.g., bold driver [17], constant, subgradient [18], RMSProp [19], Adam [20], Adadelata [21]) for updating adaptive step size  $\alpha^{(r)}$  in (32). The root mean square propagation (RMSProp) method is applied for the estimation procedure in this chapter. The RMSProp keeps track of the moving average of the squared incremental gradients for each gain element by adapting the step size element-wise as in the following.

$$\tau_{r,ij} = \gamma \tau_{r-1,ij} + (1 - \gamma) \left[ \left( \nabla_{W^{(r)}} \Psi \right)_{ij} \right]^2; \tau_0 = 0 \quad (33)$$

$$\alpha_{ij}^{(r)} = \frac{\alpha^{(0)}}{\sqrt{\tau_{r,ij} + \varepsilon}} \quad (34)$$

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

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



**Algorithm 1** Pseudocode of sequential mini-batch SGD algorithm.

- 
- 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  $N$  **do** { $\triangleright N$ : Number of samples}.
  - 4:   compute innovation correlations  $\nu(k)$ .
  - 5:   **if**  $k > N_b + M$  **then** { $\triangleright N_b$ : Number of burn-in samples}.
  - 6:     compute  $\hat{C}_{seq}^k(i), i = 0, 1, 2, \dots, M-1$ .
  - 7:     **if**  $Mod(k, B) = 0$  **then**.
  - 8:       compute the objective function  $\Psi$ .
  - 9:       compute the gradient  $\nabla_W \Psi$ .
  - 10:       update the step size  $\alpha^{(r)}$ .
  - 11:       update the gain  $W_{ij}^{(r+1)} = W_{ij}^{(r)} - \alpha_{ij}^{(r)} [\nabla_{W^{(r)}} \Psi]_{ij}$ .
  - 12:       update  $R^{(r+1)}$  and  $Q^{(r+1)}$ .
  - 13:        $r = r + 1$ .
  - 14:     **end if**.
  - 15:   **end if**.
  - 16: **end for**.
- 

### 3. Numerical examples

In [5], we provided the evidence that the multi-pass sequential mini-batch stochastic gradient descent (SGD) algorithms improve the computational efficiency of the batch estimation algorithm via a number of test cases used in [2, 7–9, 12], and also showed their applicability to non-stationary systems when coupled with a change-point detection algorithm [11]. In [22], we proposed a single-pass sequential mini-batch SGD estimation algorithm by accessing measurements exactly once for non-stationary systems by modifying the example used in [12] to periodically change the process and measurement noise covariances.

In this section, we illustrate the utility of our proposed single-pass sequential mini-batch SGD estimation algorithm by applying it to general diverse examples involving detectable (but not completely observable) systems, non-stationary systems and a bearings-only tracking problem.

For the non-stationary systems, we assumed the process and measurement noise covariances occasionally change by an unknown level. Here “occasionally” implies the jumps are infrequent enough that the Kalman filter is in the steady-state prior to a jump in the noise covariance. We define the number of subgroups in which the noise covariances are not changing as  $N_{sg}$ . Given the number of observation samples,  $N$ , each subgroup has constant noise covariances with  $N/N_{sg}$  samples. In this section, we consider two non-stationary scenarios for tracking time-varying  $Q$  and  $R$  with  $N_{sg} = 5$  subgroups. We also consider the bearings-only tracking problem where  $Q$  changes continuously.

Note that the number of “burn-in” samples and the number of lags are  $N_b = 50$  and  $M = 5$ , respectively in the estimation procedure. The root mean square propagation (RMSProp) method is applied to update the filter gain. All Monte Carlo simulations were run using a computer with an Intel Core i7-8665U processor and 16 GB of RAM.

We used the averaged normalized innovation squared (NIS) metric [2] to measure the consistency of the proposed algorithm.

$$\bar{\varepsilon}_v(k) = \frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} v(k)' S^{-1} v(k) \quad (35)$$

where  $N_{mc}$  is the number of Monte Carlo runs. The root mean square error (RMSE) in resultant position and velocity is computed using

$$RMSE(k) = \sqrt{\frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} (x(k)_{true} - \hat{x}(k))^2} \quad (36)$$

### 3.1 Case 1: A detectable (but not completely observable) system that satisfies the identifiability conditions

Mehra [8] stated, without proof, that a necessary and sufficient condition for noise covariance estimation is that the system satisfies the observability property. This example, due to Odelson et al. [7], demonstrates that this condition is not necessary. The example does satisfy the full column rank condition for the identifiability matrix in (16).

Odelson et al. [7] proposed a noise covariance estimation method based on the autocovariance least-squares formulation by using the Kronecker operator  $\delta$ . This method computes the covariances from the residuals of the state estimation. Note that the incompletely observable (but detectable<sup>1</sup>) system used in [7] is described by

$$F = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.2 \end{bmatrix}, H = [1 \ 0], \Gamma = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad (37)$$

where  $F$  is the non-singular transition matrix,  $H$  is the constant output matrix, and  $\Gamma$  is the constant input matrix in (1) and (2). Note that this system is a hypothetical numerical example. The process noise  $v(k)$  and the measurement noise  $w(k)$  are supposed to be uncorrelated Gaussian white noise sequences with zero-mean and covariances as in the following

$$E[v(k)v(j)'] = Q\delta_{kj} \quad (38)$$

$$E[w(k)w(j)'] = R\delta_{kj} \quad (39)$$

In this scenario, the true  $R$  values for the five subgroups are [0.30, 0.81, 0.49, 0.72, 0.42], and the true  $Q$  values for the five subgroups are [0.16, 0.49, 0.25, 0.36, 0.20]. The values are changed every 10,000 samples. **Table 1** shows the results of 100 Monte Carlo simulations based on the single-pass SGD algorithm in estimating  $Q$  and  $R$ . As can be seen, the estimated  $Q$  and  $R$  are close to their corresponding true values. In this

<sup>1</sup> The pair (F, H) in the system should be detectable in order for the continuous-time algebraic Riccati equation to have at least one positive semidefinite solution and in this case at least one solution results in a marginally stable steady-state KF [23, 24].

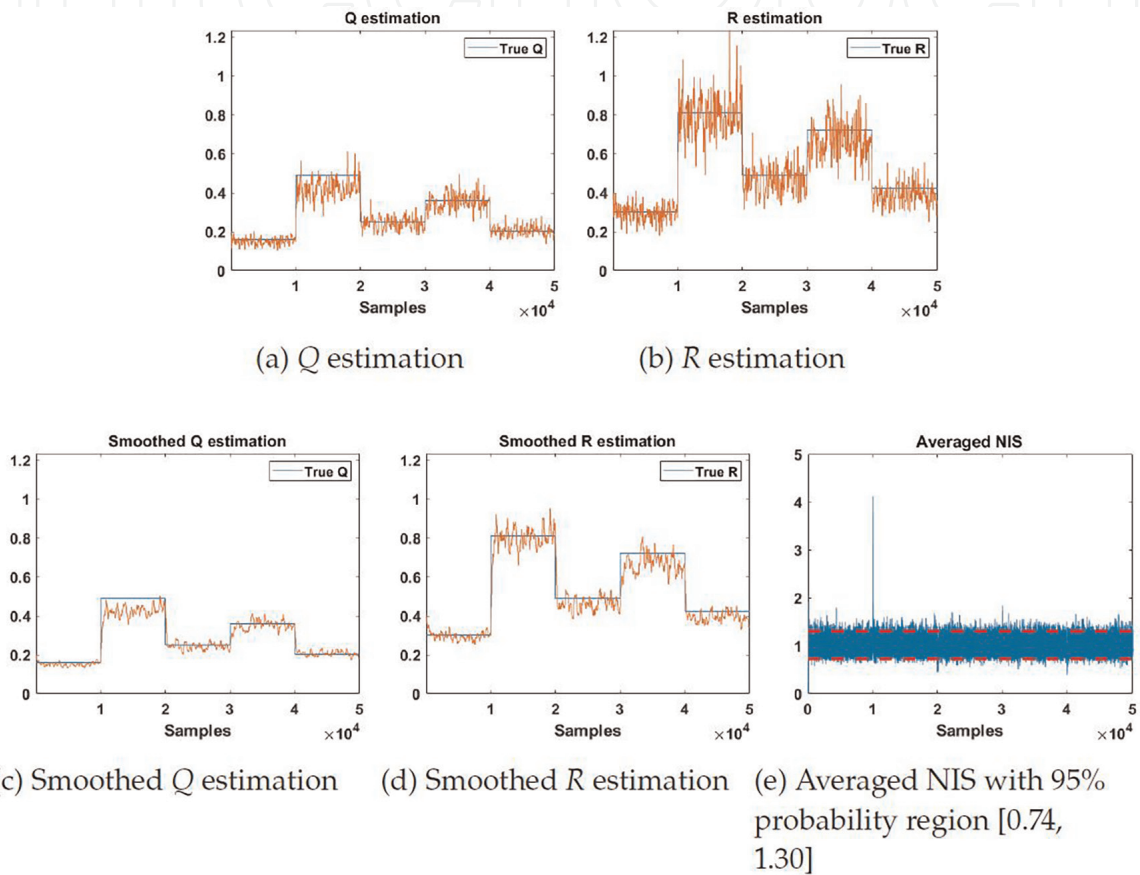
Subgroup index	R			Q			$\bar{P}_{11}$			$\bar{P}_{22}$			$W_{11}$			$W_{21}$		
	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE	Truth	Mean	RMSE
1 <sup>st</sup>	0.30	0.31	0.06	0.16	0.15	0.04	0.16	0.15	0.04	0.66	0.62	0.18	0.35	0.33	0.03	0.70	0.67	0.03
2 <sup>nd</sup>	0.81	0.86		0.49	0.42		0.49	0.43		2.01	1.74		0.38	0.34		0.76	0.67	
3 <sup>rd</sup>	0.49	0.49		0.25	0.24		0.25	0.24		1.03	0.97		0.34	0.34		0.68	0.67	
4 <sup>th</sup>	0.72	0.71		0.36	0.35		0.36	0.35		1.48	1.44		0.33	0.34		0.67	0.67	
5 <sup>th</sup>	0.42	0.41		0.20	0.20		0.20	0.21		0.83	0.84		0.33	0.34		0.66	0.67	

**Table 1.**

Single-pass SGD estimation for Case 1 (100 MC Runs; 50,000 samples;  $M = 5$ ; RMSProp; Batch size = 64).

scenario, the single-pass SGD estimation method has a speedup factor of 31 over the batch and multiple-pass SGD estimation methods (not shown).

**Figure 1** demonstrates that the sequential mini-batch gradient descent algorithm can track  $Q$  and  $R$  correctly. Here, the trajectories of  $Q$  and  $R$  estimates are smoothed by a simple first order fading memory filter with a smoothing weight of 0.7. **Figure 1e** shows the averaged NIS of SGD (RMSProp; batch size of 64) method with the 95% probability region  $[0.74, 1.30]$ , and shows that the SGD-based Kalman filter is consistent. The only place at which the NIS values are large are immediately after the jump in the noise variances. This is because adaptation requires a few samples.



**Figure 1.** Trajectories of  $Q$  and  $R$  estimates without signal smoothing and with a smoothing weight of 0.7 for Case 1.

### 3.2 Case 2: a five-state inertial navigation system with diagonal $Q$ and $R$

For estimating the unknown noise covariance parameters and the optimal Kalman filter gain for part of an inertial navigation system (INS), Mehra [8] proposed an iterative innovation correlations-based method starting from an arbitrary initial stabilizing gain. Inertial navigation [25] involves tracking the position and orientation of an object relative to a known starting orientation and velocity and it uses measurements provided by accelerometers and gyroscopes. These systems have found universal use in military and commercial applications [26].

Since the earth is not flat, the inertial navigation systems need to keep tilting the platform (with respect to inertial space) to keep the axes of the accelerometers horizontal. Here, small error sources that drive the Schuler-loop cause the navigation errors, and these errors are "damped" by making use of external velocity measurements, such as are furnished by a Doppler radar [27, 28].

In this problem, Mehra [8] used a system based on the damped Schuler-loop error propagation forced by exponentially correlated as well as white noise input. The system matrices for this navigation system are given by

$$F = \begin{bmatrix} 0.75 & -1.74 & -0.3 & 0 & -0.15 \\ 0.09 & 0.91 & -0.0015 & 0 & -0.008 \\ 0 & 0 & 0.95 & 0 & 0 \\ 0 & 0 & 0 & 0.55 & 0 \\ 0 & 0 & 0 & 0 & 0.905 \end{bmatrix}; H = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix};$$

$$\Gamma = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 24.64 & 0 & 0 \\ 0 & 0.835 & 0 \\ 0 & 0 & 1.83 \end{bmatrix} \quad (40)$$

where the system is discretized using a time step of 0.1 seconds. In this five-state system, the first two states represent the a velocity damping term and velocity error, respectively, and the other three states model the correlated noise processes. The noise corresponding to states 3 and 5 impacts both the velocity error and the velocity damping term; the fourth state impacts the sensor error in state 2 only.

In this problem, the true values corresponding to each subgroup with  $N_{sg} = 5$  subgroups and  $N = 100,000$  samples are as in (41). Each parameter of  $Q$  and  $R$  changes every 20,000 samples.

$$\begin{bmatrix} R_{11} \\ R_{22} \end{bmatrix} = \begin{bmatrix} [0.25, 0.56, 0.64, 0.42, 0.36] \\ [0.25, 0.25, 0.49, 0.16, 0.04] \end{bmatrix}; \begin{bmatrix} Q_{11} \\ Q_{22} \\ Q_{33} \end{bmatrix} = \begin{bmatrix} [0.25, 0.64, 0.49, 0.25, 0.49] \\ [0.25, 0.36, 0.56, 0.16, 0.04] \\ [0.36, 0.49, 0.64, 0.25, 0.09] \end{bmatrix} \quad (41)$$

**Table 2** shows the results of 100 Monte Carlo simulations for estimating the noise parameters using RMSProp update. The estimated parameters are close to the corresponding true values. Given 100,000 samples, the proposed method with a batch-size of 64 requires 1,891 seconds for 100 Monte Carlo simulations, i.e., 18.91 seconds per run. The batch and multi-pass SGD estimation methods need more than 3,000 seconds for a single MC run (not shown); the single-pass SGD algorithm has a speedup factor of 158.6.

**Figure 2** shows the trajectories of the estimated  $Q$  and  $R$  with a signal smoothing with a smoothing weight of 0.7. For this example, it is known that accurate estimation of  $R_{11}$  is hard as shown in **Figure 2d**. The reason is that  $R_{11}$  is dominated by the state uncertainty, i.e., the measurement noise is “buried” in a much larger innovation [4]. In spite of the difficulty in estimating  $R_{11}$ , the filter is consistent as measured by NIS as shown in **Figure 2f**.

(a)  $R$ ,  $Q$  and  $\bar{P}$  Estimates for Case 2

Subgroup index	$R_{11}$		RMSE	$R_{22}$		RMSE	$Q_{11}$		RMSE	$Q_{22}$		RMSE	$Q_{33}$		RMSE
	Truth	Mean		Truth	Mean		Truth	Mean		Truth	Mean		Truth	Mean	
1 <sup>st</sup>	0.25	0.23	0.34	0.25	0.24	0.04	0.25	0.24	0.05	0.25	0.23	0.05	0.36	0.35	0.06
2 <sup>nd</sup>	0.56	0.52		0.25	0.25		0.64	0.62		0.36	0.31		0.49	0.44	
3 <sup>rd</sup>	0.64	0.85		0.49	0.50		0.49	0.50		0.56	0.52		0.64	0.71	
4 <sup>th</sup>	0.42	0.44		0.16	0.16		0.25	0.26		0.16	0.16		0.25	0.25	
5 <sup>th</sup>	0.36	0.30		0.04	0.04		0.49	0.48		0.04	0.04		0.09	0.08	

Subgroup index	$\bar{P}_{11}$		RMSE	$\bar{P}_{22}$		RMSE	$\bar{P}_{33}$		RMSE	$\bar{P}_{44}$		RMSE	$\bar{P}_{55}$		RMSE
	Truth	Mean		Truth	Mean		Truth	Mean		Truth	Mean		Truth	Mean	
1 <sup>st</sup>	19.46	18.84	4.12	0.33	0.31	0.05	306.26	297.69	73.27	0.23	0.22	0.05	4.02	3.86	0.71
2 <sup>nd</sup>	43.76	42.11		0.43	0.40		766.59	744.98		0.34	0.30		5.45	4.99	
3 <sup>rd</sup>	37.81	40.15		0.66	0.67		600.91	621.04		0.52	0.49		7.33	8.00	
4 <sup>th</sup>	18.49	18.96		0.22	0.22		303.75	312.24		0.15	0.15		2.78	2.78	
5 <sup>th</sup>	29.31	28.55		0.07	0.07		574.37	562.90		0.04	0.04		0.97	0.93	

(b)  $W$  Estimates for Case 2

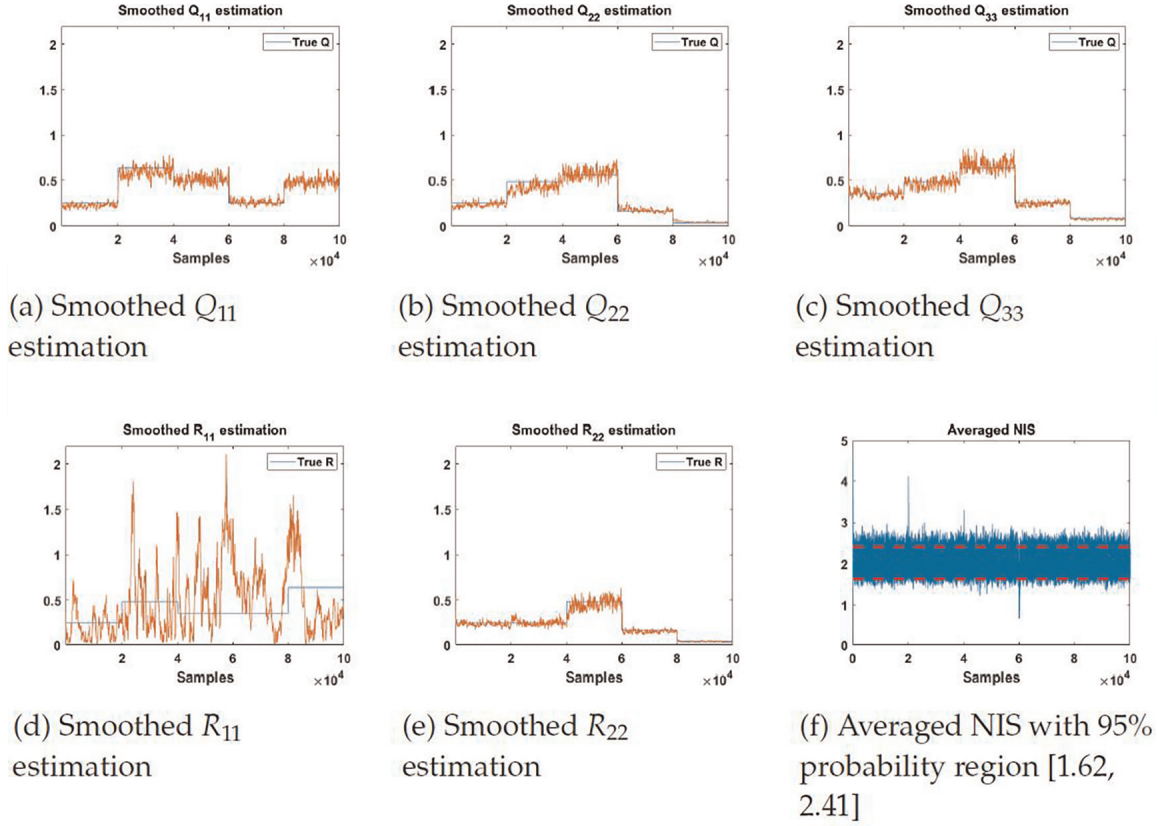
Subgroup index	$W_{11}$		RMSE	$W_{21}$		RMSE	$W_{31}$		RMSE	$W_{41}$		RMSE	$W_{51}$		RMSE
	Truth	Mean		Truth	Mean		Truth	Mean		Truth	Mean		Truth	Mean	
1 <sup>st</sup>	0.94	0.96	0.02	$2.78 \cdot 10^{-3}$	$-3.56 \cdot 10^{-3}$	0.01	-2.80	-2.75	0.04	$-2.22 \cdot 10^{-5}$	$6.50 \cdot 10^{-3}$	0.01	0.04	0.04	0.01
2 <sup>nd</sup>	0.96	0.98		$3.22 \cdot 10^{-3}$	$-4.32 \cdot 10^{-3}$		-2.91	-2.88		$-8.76 \cdot 10^{-4}$	$8.37 \cdot 10^{-3}$		0.03	0.01	
3 <sup>rd</sup>	0.94	0.95		$2.84 \cdot 10^{-3}$	$-3.84 \cdot 10^{-3}$		-2.79	-2.83		$-1.19 \cdot 10^{-4}$	$8.79 \cdot 10^{-3}$		0.04	0.03	
4 <sup>th</sup>	0.94	0.95		$3.84 \cdot 10^{-3}$	$2.08 \cdot 10^{-4}$		-2.80	-2.84		$-6.80 \cdot 10^{-4}$	$3.89 \cdot 10^{-4}$		0.03	0.02	
5 <sup>th</sup>	0.98	1.00		$3.50 \cdot 10^{-3}$	$-1.33 \cdot 10^{-4}$		-3.03	-3.00		$-1.09 \cdot 10^{-3}$	$2.10 \cdot 10^{-4}$		0.01	0.01	



Subgroup index	$W_{12}$		RMSE	$W_{22}$		RMSE	$W_{32}$		RMSE	$W_{42}$		RMSE	$W_{52}$		RMSE
	Truth	Mean		Truth	Mean		Truth	Mean		Truth	Mean		Truth	Mean	
1 <sup>st</sup>	0.94	0.94	0.08	0.38	0.39	0.02	-1.63	-1.65	0.25	0.23	0.23	0.02	-0.94	-0.93	0.05
2 <sup>nd</sup>	1.02	1.00		0.40	0.42		-1.65	-1.70		0.27	0.25		-1.02	-0.98	
3 <sup>rd</sup>	0.86	0.96		0.36	0.37		-1.56	-1.68		0.26	0.25		-0.86	-0.93	
4 <sup>th</sup>	0.99	0.98		0.39	0.39		-1.57	-1.69		0.23	0.23		-0.98	-0.99	
5 <sup>th</sup>	1.19	1.06		0.44	0.46		-1.26	-1.77		0.20	0.22		-1.17	-1.11	

**Table 2.**

Single-pass SGD estimation for Case 2 (100 MC Runs, 100,000 samples;  $M = 5$ ; RMSProp; Batch size = 64).



**Figure 2.** Trajectories of  $Q$  and  $R$  estimates with a signal smoothing at smoothing weight = 0.7 for Case 2.

### 3.3 Case 3: Bearings-only tracking problem

In many practical situations, it is generally hard to get a closed-form solution for state estimation because the noise covariances are often unknown and the dynamics are nonlinear. Arasaratnam et al. [29] proposed a nonlinear filter using bearings-only measurements for estimating the position and velocity of a target in a high-dimensional state. This method is based on the measurements from a passive sensor that measures only the direction of arrival of a signal emitted by the target [2]. This so-called bearings-only tracking problem arises in a variety of practical applications, such as air traffic control, underwater sonar tracking and aircraft surveillance [2, 30, 31].

In this example, we consider a two-dimensional bearings-only tracking problem of a nearly-constant velocity target from a single moving observer used in [32]. The dynamics of the target (relative to the observer) are described by

$$x(k+1) = Fx(k) + \Gamma v(k) - U(k) \quad (42)$$

$$z(k) = h(x(k)) + w(k) \quad (43)$$

Formally, if the state vector of the target is  $x^t(k) = [\zeta^t, \eta^t, \dot{\zeta}^t, \dot{\eta}^t]'$ , and the state vector of the observer is  $x^o(k) = [\zeta^o, \eta^o, \dot{\zeta}^o, \dot{\eta}^o]'$  for position and velocity along the  $\zeta$  and  $\eta$  axes,  $x(k) = x^t(k) - x^o(k)$  represents the relative state vector of the target with respect to the observer and the input vector  $U(k) = x^o(k) - Fx^o(k-1)$ ;  $w(k)$  is a zero-mean white Gaussian noise with variance  $\sigma_w^2$ . The nonlinear measurement involves the bearing of the target from the observer's platform, given by  $h(x(k)) = \tan^{-1}(\zeta/\eta)$ . Here,  $\Gamma$  is the identity matrix with ones on the diagonal and zeros elsewhere.

The system matrices for this problem are given by

$$F = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \Gamma = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, Q = \begin{bmatrix} T^3/3 & 0 & T^2/2 & 0 \\ 0 & T^3/3 & 0 & T^2/2 \\ T^2/2 & 0 & T & 0 \\ 0 & T^2/2 & 0 & T \end{bmatrix} \tilde{q}(k) \quad (44)$$

where the sampling interval,  $T$  is 1 second. The zero-mean white process noise intensity  $\tilde{q}(k)$  is  $\tilde{q}_0 = 9 \cdot 10^{-12} \text{ km}^2/\text{s}^3$ , except for the interval where it starts to increase rapidly to  $1.5 \cdot \tilde{q}_0$  around the sample index  $k = 480$  and then decreases again rapidly around  $k = 960$  as below:

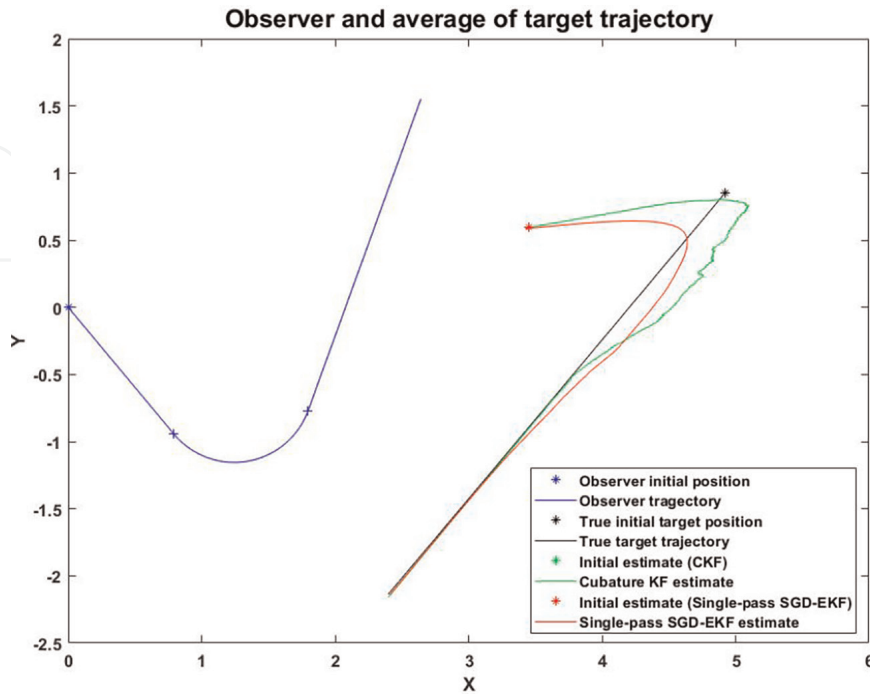
$$\tilde{q}(k) = \begin{cases} \tilde{q}_0 + 0.25\tilde{q}_0(1 + \tanh(0.015(k - 480))), & k \leq 720 \\ \tilde{q}_0 + 0.25\tilde{q}_0(1 + \tanh(0.015(960 - k))), & \text{otherwise} \end{cases} \quad (45)$$

The linearized measurement matrix,  $H(k)$ , is the Jacobian of the measurement function given by

$$H_k = \frac{\partial h(x(k))}{\partial x(k)} = \begin{bmatrix} \frac{\eta(k)}{\zeta^2(k) + \eta^2(k)} & \frac{-\zeta(k)}{\zeta^2(k) + \eta^2(k)} & 0 & 0 \end{bmatrix} \quad (46)$$

A total of 1920 measurement samples were generated for this scenario. The observer moves straight with a speed of 5 *knots*, except for 480 seconds (between  $k = 480$  and  $k = 960$ ), where it turns with  $2.4^\circ/\text{s}$  as shown in **Figure 3** (these times are marked by cross sign).

For a fair comparison of the estimation algorithms, we initialized all filters with the same mean and covariance using the prior knowledge of the initial target range



**Figure 3.** Observer and target trajectory (100 MC runs).

and the initial bearing measurement [33, 34]. Here, the initial target range and the initial bearing measurement are generated as  $\bar{r} \sim N(r, \sigma_r^2)$  and  $\bar{\theta}_0 \sim N(\theta, \sigma_\theta^2)$ , respectively, where  $r$  is the true initial target range and  $\theta$  is the true initial bearing measurement. The initial target speed is initialized as  $\bar{s} \sim N(s, \sigma_s^2)$ , where  $s$  is the true initial target speed. Let  $\sigma_{\langle \cdot \rangle}$  denotes the standard deviation of the parameter. Assuming that the target is moving towards the observer, the initial course estimate can be obtained as  $\bar{c} = \bar{\theta}_0 + \pi$ . The initial state vector and the initial covariance are given by

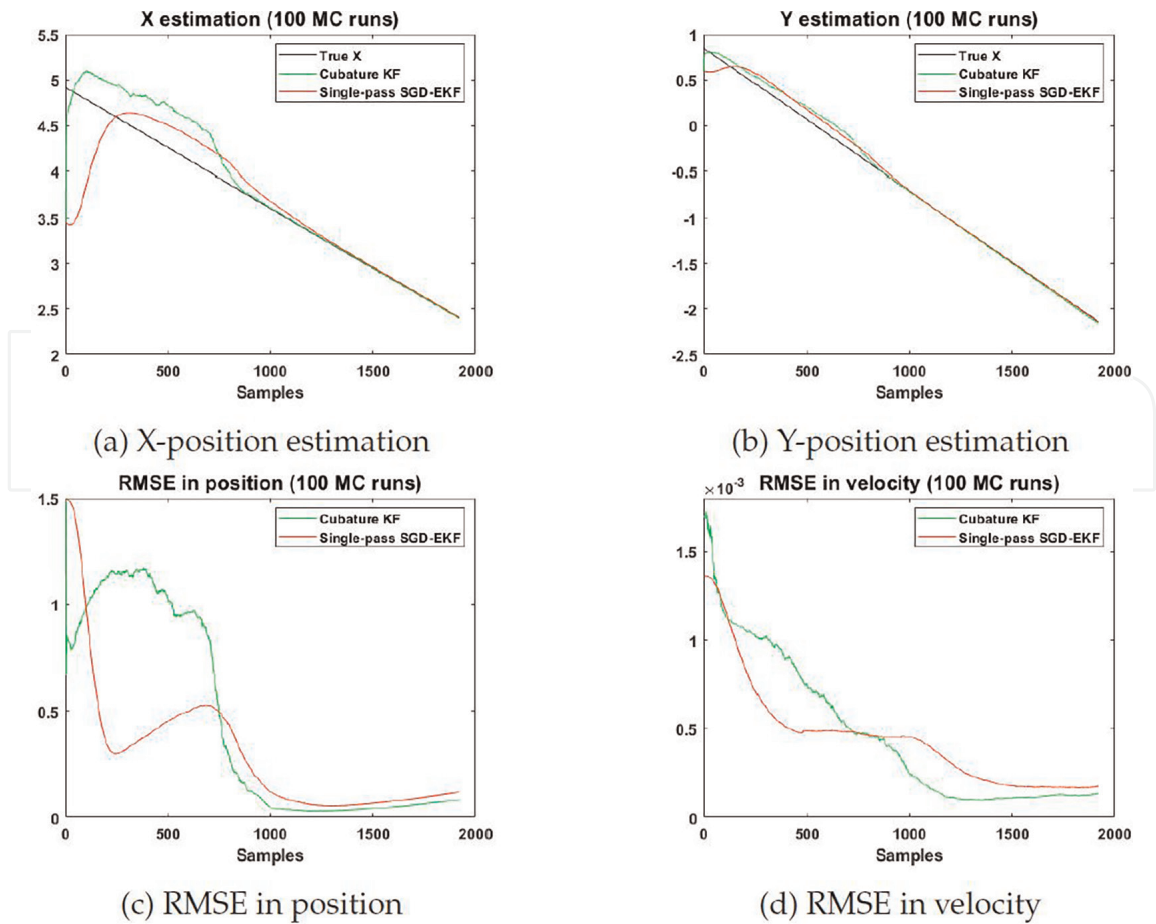
$$\hat{x}_0 = \begin{bmatrix} \hat{\zeta} \\ \hat{\eta} \\ \hat{\zeta} \\ \hat{\eta} \end{bmatrix} = \begin{bmatrix} \bar{r} \cos(\bar{\theta}_0) \\ \bar{r} \sin(\bar{\theta}_0) \\ \bar{s} \sin(\bar{c}) - \dot{\zeta}_0^o \\ \bar{s} \cos(\bar{c}) - \dot{\eta}_0^o \end{bmatrix}; P_0 = \begin{bmatrix} P_{\zeta\zeta} & P_{\zeta\eta} & 0 & 0 \\ P_{\eta\zeta} & P_{\eta\eta} & 0 & 0 \\ 0 & 0 & P_{\dot{\zeta}\dot{\zeta}} & P_{\dot{\zeta}\dot{\eta}} \\ 0 & 0 & P_{\dot{\eta}\dot{\zeta}} & P_{\dot{\eta}\dot{\eta}} \end{bmatrix} \quad (47)$$

where

$$P_{\zeta\zeta} = \bar{r}^2 \sigma_\theta^2 \cos^2(\bar{\theta}_0) + \sigma_r^2 \sin^2(\bar{\theta}_0); P_{\eta\eta} = \bar{r}^2 \sigma_\theta^2 \sin^2(\bar{\theta}_0) + \sigma_r^2 \cos^2(\bar{\theta}_0) \quad (48)$$

$$P_{\zeta\eta} = P_{\eta\zeta} = (\sigma_r^2 - \bar{r}^2 \sigma_\theta^2) \sin(\bar{\theta}_0) \cos(\bar{\theta}_0); P_{\dot{\zeta}\dot{\zeta}} = \bar{s}^2 \sigma_c^2 \cos^2(\bar{c}) + \sigma_s^2 \sin^2(\bar{c}) \quad (49)$$

$$P_{\dot{\eta}\dot{\eta}} = \bar{s}^2 \sigma_c^2 \sin^2(\bar{c}) + \sigma_s^2 \cos^2(\bar{c}); P_{\dot{\zeta}\dot{\eta}} = P_{\dot{\eta}\dot{\zeta}} = (\sigma_s^2 - \bar{s}^2 \sigma_c^2) \sin(\bar{c}) \cos(\bar{c}) \quad (50)$$



**Figure 4.** Comparison of estimation algorithms for the bearings-only tracking problem (100 MC runs).

where  $r$  and  $s$  are 5 km and 4 *knots*, respectively, and the target course is  $-140^\circ$ . Here,  $\sigma_r$  is 2 km,  $\sigma_\theta$  is  $1.5^\circ$ ,  $\sigma_s$  is 2 *knots* and  $\sigma_c = \pi/\sqrt{12}$  for this problem.

**Figure 4** shows a comparison of algorithms for the bearings-only tracking problem. The cubature Kalman filter (CKF) uses a third-degree spherical-radial cubature rule that provides the set of cubature points scaling linearly with the state-vector dimension [29]. The cubature Kalman filter and our single-pass SGD extended KF (EKF) method can track the target well, but our proposed method shows better computational efficiency compared to CKF by a factor of 2.5 (not shown). Root mean square error (RMSE) in position and velocity over 100 Monte Carlo runs are shown in **Figure 4c** and **Figure 4d**. During the whole maneuver, the RMSE of the proposed single-pass SGD-EKF algorithm was slightly lower than that with the CKF method.

## 4. Conclusions

In this chapter, we derived a single-pass sequential mini-batch SGD algorithm for estimating the noise covariances in an adaptive Kalman filter. We demonstrated the utility of the method using diverse examples involving a detectable (but not completely observable) system, a non-stationary system, and a nonlinear bearings-only tracking problem. The evaluation showed that the proposed method has acceptable state estimation root mean square error (RMSE) and exhibits filter consistency as measured by the normalized innovation squared (NIS) criterion.

## Acknowledgements

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.

## Abbreviations

CKF	Cubature Kalman filter
EKF	Extended Kalman filter
INS	Inertial navigation system
KF	Kalman filter
MMSE	Minimum mean square error
NIS	Normalized innovation squared
RMSE	Root mean square error
RMSProp	Root mean square propagation
SGD	Stochastic gradient descent

# IntechOpen

## Author details

Hee-Seung Kim<sup>1\*</sup>, Lingyi Zhang<sup>2</sup>, Adam Bienkowski<sup>1</sup>, Krishna R. Pattipati<sup>1</sup>,  
David Sidoti<sup>3</sup>, Yaakov Bar-Shalom<sup>1</sup> and David L. Kleinman<sup>1</sup>

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


2 HRL Laboratories, Intelligent System Laboratory, Malibu, CA, USA

3 U.S. Naval Research Laboratory, Marine Meteorology Division, Monterey, CA, USA

\*Address all correspondence to: [hee-seung.kim@uconn.edu](mailto:hee-seung.kim@uconn.edu)

## IntechOpen

---

© 2022 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 



## References

- [1] Auger F, Hilairet M, Guerrero JM, Monmasson E, Orłowska-Kowalska T, Katsura S. Industrial applications of the Kalman filter: A review. *IEEE Transactions on Industrial Electronics*. 2013;**60**(12):5458-5471
- [2] Bar-Shalom Y, Li XR, Kirubarajan T. Estimation with applications to tracking and navigation: theory algorithms and software. New York, NY: John Wiley & Sons; 2001
- [3] Kalman RE. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*. 1960;**82**(1): 35-45
- [4] Zhang L, Sidoti D, Bienkowski A, Pattipati KR, Bar-Shalom Y, Kleinman DL. On the identification of noise covariances and adaptive Kalman Filtering: A new look at a 50 year-old problem. *IEEE Access*. 2020;**8**: 59362-59388
- [5] Kim HS, Zhang L, Bienkowski A, Pattipati KR. Multi-pass sequential mini-batch stochastic gradient descent algorithms for noise covariance estimation in adaptive Kalman filtering. *IEEE Access*. 2021;**9**:99220-99234
- [6] Kim HS, Bienkowski A, Pattipati KR. A single-pass noise covariance estimation algorithm in multiple-model adaptive Kalman filtering for non-stationary systems. *TechRxiv*. 2022. DOI: 10.36227/techrxiv.14761005.v2
- [7] Odelson BJ, Rajamani MR, Rawlings JB. A new autocovariance least-squares method for estimating noise covariances. *Automatica*. 2006;**42**(2): 303-308
- [8] Mehra R. On the identification of variances and adaptive Kalman filtering. *IEEE Transactions on automatic control*. 1970;**15**(2):175-184
- [9] Belanger PR. Estimation of noise covariance matrices for a linear time-varying stochastic process. *Automatica*. 1974;**10**(3):267-275
- [10] Sarkka S, Nummenmaa A. Recursive noise adaptive Kalman filtering by variational Bayesian approximations. *IEEE Transactions on Automatic Control*. 2009;**54**(3):596-600
- [11] Killick R, Fearnhead P, Eckley IA. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*. 2012;**107**(500):1590-1598
- [12] Neethling C, Young P. Comments on "Identification of optimum filter steady-state gain for systems with unknown noise covariances". *IEEE Transactions on Automatic Control*. 1974;**19**(5):623-625
- [13] Tajima K. Estimation of steady-state Kalman filter gain. *IEEE Transactions on Automatic Control*. 1978;**23**(5):944-945
- [14] Bertsekas DP. Athena scientific optimization and computation series. In: *Nonlinear Programming*. Belmont, MA: Athena Scientific; 2016. Available from: <http://www.athenasc.com/nonlinbook.html>
- [15] Golub GH, Van Loan CF. *Matrix computations*. Vol. 3. Baltimore, MD: JHU press; 2013
- [16] Arnold WF, Laub AJ. Generalized eigenproblem algorithms and software for algebraic Riccati equations. *Proceedings of the IEEE*. 1984;**72**(12):1746-1754
- [17] Battiti R. Accelerated backpropagation learning: Two

optimization methods. *Complex Systems*. 1989;3(4):331-342

[18] Boyd S, Xiao L, Mutapcic A. Subgradient methods. lecture notes of EE392o, Stanford University, Autumn Quarter. 2003;2004:2004-2005. Available from: [www.stanford.edu/class/ee392o/subgrad\\_method.pdf](http://www.stanford.edu/class/ee392o/subgrad_method.pdf)

[19] Tieleman T, Hinton G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural networks for machine learning. 2012; 4(2):26-31

[20] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv. 2014. DOI: 10.48550/arXiv.1412.6980

[21] Zeiler MD. Adadelta: an adaptive learning rate method. arXiv. 2012. DOI: 10.48550/arXiv.1212.5701

[22] Kim HS, Zhang L, Bienkowski A, Pattipati KR. A single-pass noise covariance estimation algorithm in adaptive Kalman filtering for non-stationary systems. In: 2021 IEEE 24th International Conference on Information Fusion (FUSION). IEEE; 2021. pp. 556-563. DOI: 10.23919/FUSION49465.2021.9626861

[23] Naets F, Cuadrado J, Desmet W. Stable force identification in structural dynamics using Kalman filtering and dummy-measurements. *Mechanical Systems and Signal Processing*. 2015;50: 235-248

[24] Simon D. Optimal state estimation: Kalman, H infinity, and nonlinear approaches. Hoboken, NJ: John Wiley & Sons; 2006

[25] Woodman OJ. An introduction to inertial navigation. Technical Report, Computer Laboratory. Cambridge, UK:

University of Cambridge; 2007 UCAM-CL-TR-696. DOI: 10.48456/tr-696

[26] Kuritsky MM, Goldstein MS. Inertial navigation. *Proceedings of the IEEE*. 1983;71:1156-1176

[27] Heller WG. Free-Inertial and Damped-Inertial Navigation Mechanization and Error Equations. Reading, MA: Analytic Sciences Corp; 1975

[28] King A. Inertial navigation-forty years of evolution. *GEC review*. 1998; 13(3):140-149

[29] Arasaratnam I, Haykin S. Cubature Kalman Filters. *IEEE Transactions on Automatic Control*. 2009;54(6): 1254-1269

[30] Farina A. Target tracking with bearings – Only measurements. *Signal Processing*. 1999;78(1):61-78

[31] Aidala VJ. Kalman Filter Behavior in Bearings-Only Tracking Applications. *IEEE Transactions on Aerospace and Electronic Systems*. 1979;AES-15(1):29-39

[32] Leong PH, Arulampalam S, Lamahewa TA, Abhayapala TD. A Gaussian-Sum Based Cubature Kalman Filter for Bearings-Only Tracking. *IEEE Transactions on Aerospace and Electronic Systems*. 2013;49(2): 1161-1176

[33] Ristic B, Arulampalam S, Gordon N. Beyond the Kalman filter: Particle filters for tracking applications. Norwood, MA: Artech house; 2003

[34] Kumar K, Bhaumik S, Arulampalam S. Tracking an Underwater Object with Unknown Sensor Noise Covariance Using Orthogonal Polynomial Filters. *Sensors*. 2022;22(13):4870