# On the Identification of Noise Covariances and Adaptive Kalman Filtering: A New Look at a 50 Year-old Problem

Lingyi Zhang [1,1,1,1], David Sidoti [2], Adam Bienkowski [2], Krishna Pattipati [2], Yaakov Bar-Shalom [2], and David Kleinman [2]

[1]University of Connecticut
[2]Affiliation not available

November 8, 2023



**Abstract**

The Kalman filter requires knowledge of the noise statistics; however, the noise covariances are generally *unknown*. Although this problem has a long history, reliable algorithms for their estimation are scant, and necessary and sufficient conditions for identifiability of the covariances are in dispute. We address both of these issues in this paper. We first present the necessary and sufficient condition for unknown noise covariance estimation; these conditions are related to the rank of a matrix involving the auto and cross-covariances of a weighted sum of innovations, where the weights are the coefficients of the the minimal polynomial of the closed-loop system transition matrix of a stable, but not necessarily optimal, Kalman filter. We present an optimization criterion and a novel six-step approach based on a successive approximation, coupled with a gradient algorithm with adaptive step sizes, to estimate the steady-state Kalman filter gain, the unknown noise covariance matrices, as well as the state prediction (and updated) error covariance matrix. Our approach enforces the structural assumptions on unknown noise covariances and ensures symmetry and positive definiteness of the estimated covariance matrices. We provide several approaches to estimate the unknown measurement noise covariance $R$ via *post-fit residuals*, an approach not yet exploited in the literature. The validation of the proposed method on five different test cases from the literature demonstrates that the proposed method significantly outperforms previous state-of-the-art methods. It also offers a number of novel machine learning motivated approaches, such as sequential (one sample at a time) and mini-batch-based methods, to speed up the computations.

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

*Digital Object Identifier 10.1109/ACCESS.2017.DOI*

# On the Identification of Noise Covariances and Adaptive Kalman Filtering: A New Look at a 50 Year-old Problem

**LINGYI ZHANG[1], DAVID SIDOTI[2], ADAM BIENKOWSKI[1], KRISHNA R. PATTIPATI[1], (Fellow, IEEE), YAAKOV BAR-SHALOM[1], (Fellow, IEEE), DAVID L. KLEINMAN[1], (Fellow, IEEE)**

[1]Dept. of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269 USA (e-mail:krishna.pattipati@uconn.edu)
[2]Marine Meteorology Division, US Naval Research Laboratory, Monterey, CA 93940 USA (e-mail:david.sidoti@nrlmry.navy.mil)

Corresponding author: Lingyi Zhang (e-mail: lingyi.zhang@uconn.edu).

This work was supported in part by the U.S. Office of Naval Research and US Naval Research Laboratory under Grants #N00014-18-1-1238, #N00173-16-1-G905 and #HPCM034125HQU. Research of Pattipati is also partially supported by a Space Technology Research Institutes grant (number 80NSSC19K1076) from NASA's Space Technology Research Grants Program.

**ABSTRACT** The Kalman filter requires knowledge of the noise statistics; however, the noise covariances are generally *unknown*. Although this problem has a long history, reliable algorithms for their estimation are scant, and necessary and sufficient conditions for identifiability of the covariances are in dispute. We address both of these issues in this paper. We first present the necessary and sufficient condition for unknown noise covariance estimation; these conditions are related to the rank of a matrix involving the auto and cross-covariances of a weighted sum of innovations, where the weights are the coefficients of the the minimal polynomial of the closed-loop system transition matrix of a stable, but not necessarily optimal, Kalman filter. We present an optimization criterion and a novel six-step approach based on a successive approximation, coupled with a gradient algorithm with adaptive step sizes, to estimate the steady-state Kalman filter gain, the unknown noise covariance matrices, as well as the state prediction (and updated) error covariance matrix. Our approach enforces the structural assumptions on unknown noise covariances and ensures symmetry and positive definiteness of the estimated covariance matrices. We provide several approaches to estimate the unknown measurement noise covariance $R$ via *post-fit residuals*, an approach not yet exploited in the literature. The validation of the proposed method on five different test cases from the literature demonstrates that the proposed method significantly outperforms previous state-of-the-art methods. It also offers a number of novel machine learning motivated approaches, such as sequential (one sample at a time) and mini-batch-based methods, to speed up the computations.

**INDEX TERMS** adaptive filtering, Kalman filter, minimal polynomial, noise covariance estimation, adaptive gradient descent.

## I INTRODUCTION

The Kalman filter (KF) [23] is the optimal state estimator for linear dynamic systems driven by Gaussian white noise with measurements corrupted by Gaussian white noise [1]. In the classical design of a Kalman filter, the noise covariance matrices are assumed known and they, along with the system dynamics, determine the achievable filter's accuracy. However, in many practical situations, including noisy feature data in machine learning, the statistics of the noise covariances are often unknown or only *partially* known. Thus, noise identification

[1]The KF is also the best linear estimation algorithm when the noises are non-Gaussian with known covariances [2].

is an essential part of adaptive filtering. Adaptive filtering has numerous applications in engineering [36], machine learning [9], econometrics [11], weather forecasting [10], [20], [35], [44], to name a few.

We were motivated by the following learning problem: Given a vector time series and a library of models of system dynamics for 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 a suitable process noise and measurement noise model and the best system dynamics for the time series. The problem we consider in this paper is limited to finding a suitable process noise and measurement noise covaraiance for a given dynamic model.

### I-A PREVIOUS WORK

The approaches for estimating the noise covariance matrices for a Kalman filter can be broadly classified into four general categories: Bayesian inference, maximum likelihood estimation, covariance-matching, and correlation methods. The first two categories pose the noise covariance estimation problem as a parameter estimation problem.

In the Bayesian inference approach [19], the covariance estimation problem is solved by obtaining the posterior probability density function (pdf) of the unknown parameters (in this case, the noise covariance matrix elements) from their prior pdf and the observed measurements using the Bayes' formula recursively. Recently, Matisko and Havlena [32] proposed a new Bayesian method to estimate the unknown covariance matrices. They first use a Monte Carlo method to generate a grid of possible unknown covariance matrix pairs ($Q$, $R$) with more density near the highest prior probability. Then, they compute the likelihood and posterior probability after performing state estimation for each pair using a Kalman filter. In general, the Bayesian approach suffers from the curse of dimensionality and is computationally intractable due to the fact that it involves numerical integration or Monte Carlo simulations over a very large parameter space.

In the maximum likelihood estimation [25], [46], the noise statistics are obtained by maximizing the probability density function of the measurement residuals generated by the filter, which is the likelihood of the filter parameters [2]. These filter-based maximum likelihood methods require nonlinear programming based optimization and are computationally intractable. Shumway and Stoffer [45] utilize the expectation maximization (EM) algorithm [12], which requires the smoothed estimates of the system state. This approach starts with the smoothed estimation of the system state given an estimate of the initial state and noise covariance matrices. Then, the unknown parameters are estimated via maximum likelihood estimation using the smoothed state estimates obtained from the expectation step. Later, Ghahramani and Hinton [17] present an extension of [45] that can account for an unknown observation matrix in linear dynamic systems. They then go on to use forward and backward recursions to estimate the noise covariance matrices. This process is repeated until the estimated parameters converge. In addition to computational complexity, this method suffers from convergence to a local optimum.

The basic idea of the covariance-matching techniques [38] is that the sample covariance of the innovations should be consistent with its theoretical value. In [38], the unknown noise covariances are estimated from the sample covariance computed from the innovation sequences accumulated over the entire historical data (or in a moving time window). In this method, if the estimated innovation covariance value is much larger than the theoretical value, then the process noise covariance is increased. The convergence has never been proved for this method.

With regard to correlation methods, Heffes [18] derived an expression for the covariance of the state error and of the innovations of any suboptimal filter as a function of noise covariances. This expression serves as a fundamental building block in the correlation methods. The first innovation-based technique to estimate the optimal Kalman filter gain and the unknown noise covariance matrices via the correlations of innovations from an arbitrary initial stabilizing filter gain was introduced by Mehra [33]. Another procedure to carry out the identification of unknowgn optimal Kalman filter gain and the noise covariance matrices is by Carew and Bélanger [7]. Their strategy calculates the Kalman filter gain based on the estimation error that is defined as the discrepancy between the optimal state estimates obtained from the optimal Kalman filter gain and the state estimates obtained from an arbitrary suboptimal Kalman filter gain. There is a question as to whether the correlation method is sensitive to the initial Kalman filter

gain selection. Mehra suggested to repeat the noise covariance estimation steps with the obtained gain from the first attempt to improve the estimation. However, Carew and Bélanger [7] claim that if the optimal Kalman filter gain is used as the initial condition, then the approximations in Mehra's approach are such that the correctness of the optimal gain will not be confirmed.

Later, Neethling and Young [39] suggested to combine the noise covariance matrices in a vector and solve a single least squares or weighted least squares problem to improve the performance of Mehra and Carew–Bélanger's approaches. Recently, Odelson et al. [41], [42] developed the autocovariance least squares method to estimate the noise covariance matrices by applying the suggestions of [39] on Mehra's approach and using the Kronecker operator. The algorithm defines a multistep autocovariance function between the measurements, which is used to develop a linear least squares formulation to estimate the noise covariance matrices. Dunik et al. [14] compared the method presented by Odelson, Rajamani, and Rawlings [42] to a combined state and parameter estimation approach.

An interesting variant of the correlation methods is to utilize the output correlations. In 1972, Mehra [34] proposed an output correlation technique to directly estimate the optimal Kalman filter gain. This method has the advantage of being non-recursive compared to the innovation correlation techniques. However, the poor estimates of sample output correlation functions can lead to an ill-conditioned Riccati equation.

The contributions of the present paper are as follows:

1) A necessary and sufficient condition for the identifiability of unknown noise covariances is provided. This involves the rank of the auto and cross-covariances of the weighted sum of innovations of a suboptimal filter, where the weights are the coefficients of the minimal polynomial of the state transition matrix.
2) A novel six-step solution approach via a successive approximation and adaptive gradient descent scheme with a new objective function to obtain the unknown noise covariance matrices $Q$ and $R$, as well as the steady-state Kalman filter gain $W$, and the steady-state state prediction covariance matrix $\bar{P}$ or the updated state covariance matrix $P$, is proposed. They ensure positive definite $Q$ and positive definite $R$, as well as $\bar{P}$ and $P$.
3) Several novel approaches to estimate the unknown noise covariance matrix $R$ are derived via utilization of the *post-fit residual*, which has not yet been discussed in the literature.
4) Convergence proofs in [7] assumed that time averages are the same as ensemble averages. This is only approximate with finite data. Consequently, these methods either diverge or result in largely inaccurate estimates of unknown covariances.
5) Our approach can enforce the structural assumptions on $Q$ and $R$ (e.g., diagonality of $Q$ and $R$, symmetry and positive definiteness).

The paper is organized as follows. In Section II, we provide an overview of the Kalman filter and derive a new Riccati equation for the updated state covariance. Then, in Section III, we discuss the necessary and sufficient condition for the unknown noise covariances' estimation. We briefly discuss different approaches to obtaining the unknown covariance parameters in Section IV. Then, in Section V, we discuss a convergent version of Mehra's method to estimate the optimal Kalman filter gain. In Section VI, we derive five different ways to obtain $R$. Section VII provides a method to estimate the process noise covariance matrix $Q$ and the steady-state updated state covariance $P$, iteratively. All these methods are combined in Section VIII, where we present a systematic process to find the optimal filter gain $W$, the innovation covariance $S$, the measurement noise covariance $R$, the steady-state state prediction (or updated state) covariance $\bar{P}$ ($P$) and the process noise covariance $Q$. In Section IX, we specialize the approach to estimate $W$, $R$, $Q$ and $P$ for a process, where all the states are observed and for a Wiener process. Lastly, we apply our approach to five numerical examples from the literature in Section X, and conclude the paper in Section XI. In this paper, all the subscripts denote matrix indexes. The iteration variable is superscript with $(\cdot)$ to differentiate the notation from exponents.

## II PLANT AND MEASUREMENTS MODEL FOR THE KALMAN FILTER

Consider the discrete-time linear dynamic system

$$x(k+1) = Fx(k) + \Gamma v(k) \tag{1}$$

$$z(k) = Hx(k) + w(k) \tag{2}$$

where $x(k)$ is an $n_x$ -dimensional state vector, $F$ is the state transition matrix of the system, $H$ is the $n_z \times n_x$ measurement matrix, and $\Gamma$ is the $n_x \times n_v$ dimensional noise gain matrix. The sequences $v(k), k = 0, 1, \ldots,$ and $w(k), k = 0, 1, \ldots,$ are zero-mean white Gaussian noises with covariance matrices $Q$ and $R$, respectively. The two noise sequences and the initial state are assumed to be mutually independent. The matrices $Q$ and $R$ are assumed to be positive definite. Note that even if $Q$ is positive definite, $\Gamma Q \Gamma'$ need not be; it can be positive semi-definite. We assume that the system is observable and $(F, \Gamma Q^{1/2})$ is controllable [2].

Given the estimate $\hat{x}(k|k)$, the Kalman filter [2], [23] estimates the state at the next time instant $k+1$ as

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

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

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

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

$$S(k+1) = HP(k+1|k)H' + R \tag{7}$$

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

$$P(k+1|k+1) = P(k+1|k) - W(k+1)S(k+1)W(k+1)' \tag{9}$$

where the estimate $\hat{x}(k+1|k)$ is the one-step extrapolated estimate of the state vector $x(k)$ based on the measurements up to $k$, $W(k), k = 1, \ldots, N$ is the sequence of Kalman filter gains, $\nu(k), k = 1, \ldots, N$ is the innovation sequence, $P(k+1|k)$ is the state prediction covariance, $S(k+1)$ is the measurement prediction (or innovation) covariance, and $P(k+1|k+1)$ is the updated state error covariance.

The six-step approach in this paper is designed specifically for a steady-state Kalman filter. The steady-state state prediction covariance matrix $\bar{P}$ satisfies the Riccati equation.

$$\bar{P} = F[\bar{P} - \bar{P}H'(H\bar{P}H' + R)^{-1}H\bar{P}]F' + \Gamma Q \Gamma' \tag{10}$$

[2]Detectability and stabilizability are all that are needed for a stable Kalman filter.

The steady-state updated state covariance, denoted as $P$, can also be computed via a Riccati equation (see Appendix A)

$$P = FPF' - PH'(R - HPH')^{-1}HP' + \Gamma Q \Gamma' \tag{11}$$

Evidently,

$$P = \bar{P} - WSW' \tag{12}$$

$$= (I_{n_x} - WH)\bar{P}(I_{n_x} - WH)' + WRW' \tag{13}$$

where (13) is known as the Joseph form; $W$ and $S$ are the steady-state optimal gain, and the steady-state innovation covariance, respectively, and are given by

$$\begin{aligned} W &= \bar{P}H'S^{-1} \\ &= \bar{P}H'(H\bar{P}H' + R)^{-1} \\ &= PH'R^{-1} \end{aligned} \tag{14}$$

$$S = E[\nu(k)\nu(k)'] = H\bar{P}H' + R \tag{15}$$

Note that $(I_{n_x} - WH)$ is invertible, but need not be stable (i.e., eigenvalues need not be inside the unit circle).

## III IDENTIFIABILITY OF $Q$ AND $R$

One major issue in the previous literature involves the necessary conditions to estimate the unknown covariance matrices. Mehra [33] claimed that the system must be observable and controllable; however, Odelson [42] provided a counter-example wherein the system was observable and controllable, but the full $Q$ matrix was not estimable. Following the ideas in [47], we prove that the necessary and sufficient condition to estimate the unknown covariance matrices in a system is *directly related* to its minimal polynomial of $\bar{F}$, and a transformation of the innovations based on the coefficients of the minimal polynomial. Let us define $\tilde{x}(k+1|k)$ to be the predicted error between the state $x(k+1)$ and its predicted state $\hat{x}(k+1|k)$, that is,

$$\tilde{x}(k+1|k) = x(k+1) - \hat{x}(k+1|k) \tag{16}$$

We can rewrite $\hat{x}(k+1|k)$ in terms of $\tilde{x}$, that is,

$$\begin{aligned} \hat{x}(k+1|k) &= F\hat{x}(k|k-1) \\ &+ FWH\tilde{x}(k|k-1) + FWw(k) \end{aligned} \tag{17}$$

Then, substituting (17) into (16) and using (1), we have

$$\tilde{x}(k+1|k) = \bar{F}\tilde{x}(k|k-1) + \Gamma v(k) - FWw(k) \tag{18}$$

where $\bar{F}$ is the stable closed-loop filter matrix

$$\bar{F} = F(I_{n_x} - WH) \tag{19}$$

We can also write $\nu(k)$ in terms of $\tilde{x}$, that is

$$\nu(k) = H\tilde{x}(k|k-1) + w(k) \tag{20}$$

Let us define the $m^{th}$ order minimal polynomial of $\bar{F}$ as

$$\sum_{i=0}^{m} a_i \bar{F}^{m-i} = 0; \qquad a_0 = 1 \tag{21}$$

Then, $\nu(k)$ can be written as

$$\nu(k) = H\bar{F}^m\tilde{x}(k-m|k-m-1) + \left\{ H\sum_{j=0}^{m-1} \bar{F}^{m-1-j}\left[\Gamma v(k-m+j) - FWw(k-m+j)\right] \right\} + w(k) \tag{22}$$

and let us define $\xi(k)$ as

$$\xi(k) = \sum_{i=0}^{m} a_i \nu(k-i) \tag{23}$$

$$= \sum_{i=0}^{m} a_i \left[ H\left\{ \sum_{j=0}^{m-i-1} \bar{F}^{m-i-1-j}\left[\Gamma v(k-m+j) - FWw(k-m+j)\right] \right\} + w(k-i) \right] \tag{24}$$

$$= \sum_{i=0}^{m} a_i \left[ H\left\{ \sum_{l=i+1}^{m} \bar{F}^{l-i-1}\left[\Gamma v(k-l) - FWw(k-l)\right] \right\} + w(k-i) \right] \tag{25}$$

$$= \sum_{l=1}^{m} H\left( \sum_{i=0}^{l-1} a_i \bar{F}^{l-i-1} \right)\left[\Gamma v(k-l) - FWw(k-l)\right] + \sum_{l=0}^{m} a_l w(k-l) \tag{26}$$

$$= \sum_{l=1}^{m} \mathscr{B}_l v(k-l) + \sum_{l=0}^{m} \mathscr{G}_l w(k-l) \tag{27}$$

where $\mathscr{B}_l$ and $\mathscr{G}_l$ are the sum of two moving average processes driven by the process noise and the measurement noise, that is,

$$\mathscr{B}_l = H\left( \sum_{i=0}^{l-1} a_i \bar{F}^{l-i-1} \right)\Gamma \tag{28}$$

$$\mathscr{G}_l = \left[ a_l I_{n_z} - H\left( \sum_{i=0}^{l-1} a_i \bar{F}^{l-i-1} \right) FW \right] \tag{29}$$

$$\mathscr{G}_0 = I_{n_z} \tag{30}$$

Denoting $L_j = E\left[\xi(k)\xi(k-j)'\right]$, for $j = 0, 1, 2, \ldots, m$, we have

$$L_j = \sum_{i=j+1}^{m} \mathscr{B}_i Q \mathscr{B}'_{i-j} + \sum_{i=j}^{m} \mathscr{G}_i R \mathscr{G}'_{i-j} \tag{31}$$

We know that $Q = [q_{ij}]$ is an $n_v \times n_v$ positive semi-definite and symmetric matrix, and $R = [r_{ij}]$ is an $n_z \times n_z$ positive definite and symmetric matrix. Utilizing the symmetry of $Q$ and $R$, and letting $b_{i,l}$ and $g_{i,l}$ denote the $l$-th column of $\mathscr{B}_i$ and $\mathscr{G}_i$, respectively, we can rewrite (31) as

$$L_j = \sum_{l=1}^{n_v}\sum_{p=1}^{n_v} q_{lp}\left[ \sum_{i=j+1}^{m} b_{i,l} b'_{i-j,p} \right] + \sum_{l=1}^{n_z}\sum_{p=1}^{n_z} r_{lp}\left[ \sum_{i=j}^{m} g_{i,l} g'_{i-j,p} \right] \tag{32}$$

$$= \sum_{l=1}^{n_v} \left\{ \sum_{p=1}^{l} q_{lp} \left[ \sum_{i=j+1}^{m} b_{i,l} b'_{i-j,p} \right] + \sum_{p=l+1}^{m} q_{lp} \left[ \sum_{i=j+1}^{m} b_{i,l} b'_{i-j,p} \right] \right\} + \sum_{l=1}^{n_z} \left\{ \sum_{p=1}^{l} r_{lp} \left[ \sum_{i=j}^{m} g_{i,l} g'_{i-j,p} \right] + \sum_{p=l+1}^{m} r_{lp} \left[ \sum_{i=j}^{m} g_{i,l} g'_{i-j,p} \right] \right\} \tag{33}$$

$$= \sum_{l=1}^{n_v} \left\{ q_{ll} \left[ \sum_{i=j+1}^{m} b_{i,l} b'_{i-j,l} \right] + \sum_{p=1}^{l-1} q_{lp} \left[ \sum_{i=j+1}^{m} b_{i,l} b'_{i-j,p} + b_{i,p} b'_{i-j,l} \right] \right\} + \sum_{l=1}^{n_z} \left\{ r_{ll} \left[ \sum_{i=j}^{m} g_{i,l} g'_{i-j,p} \right] + \sum_{p=1}^{l-1} r_{lp} \left[ \sum_{i=j}^{m} g_{i,l} g'_{i-j,p} + g_{i,p} g'_{i-j,l} \right] \right\} \tag{34}$$

From (34), we can form a matrix $\Xi$ of dimension $(m+1)n_z^2 \times \frac{1}{2}\left[n_v(n_v+1)+n_z(n_z+1)\right]$ to estimate the unknowns in $Q$ and $R$, that is,

$$\Xi \begin{bmatrix} \mathrm{vec}(Q) \\ \mathrm{vec}(R) \end{bmatrix} = \begin{bmatrix} L_0 \\ L_1 \\ \vdots \\ L_m \end{bmatrix} \tag{35}$$

The $\mathrm{vec}(A)$ function converts a matrix $A$ into a column vector. For a $p \times n$ matrix $A$,

$$\mathrm{vec}(A) \triangleq [a_{11}, \ldots, a_{p1}, a_{12}, \ldots, a_{p2}, \ldots, a_{1n}, \ldots, a_{pn}]' \tag{36}$$

Since $R$ is always estimable because $\mathcal{G}_m$ (recall that $m$ is the order of minimal polynomial) is invertible, the maximum number of unknowns in $Q$ that can be estimated must be less than or equal to the minimum number of independent measurements minus the number of unknowns in $R$. That is

$$\mathrm{rank}(\Xi) - n_R > n_Q \tag{37}$$

where $n_R$ is the number of unknowns in $R$, and $n_Q$ is the number of unknowns in $Q$

To illustrate the necessity and sufficiency of this condition, consider an example system from [42],

$$x(k) = \begin{bmatrix} 0.9 & 0 & 0 \\ 1 & 0.9 & 0 \\ 0 & 0 & 0.9 \end{bmatrix} x(k-1) + v(k-1) \tag{38}$$

$$z(k) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} x(k) + w(k) \tag{39}$$

with $Q$ being a full 3-by-3 positive definite symmetric matrix and $R$ being a full 2-by-2 positive definite symmetric matrix. Since the rank of $\Xi$ is not affected by $W$, one can examine the rank of $\Xi$ for $W = 0$ for convenience. In this case, the minimal polynomial coefficients are

$$\begin{bmatrix} a_0 & a_1 & a_2 \end{bmatrix}' = \begin{bmatrix} 1 & -1.8 & 0.81 \end{bmatrix}' \tag{40}$$

The $\mathcal{B}$ and $\mathcal{G}$ matrices are

$$\mathcal{B}_1 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathcal{B}_2 = \begin{bmatrix} 1 & -0.9 & 0 \\ 0 & 0 & -0.9 \end{bmatrix} \tag{41}$$

$$\mathcal{G}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathcal{G}_1 = \begin{bmatrix} -1.8 & 0 \\ 0 & -1.8 \end{bmatrix} \tag{42}$$

$$\mathcal{G}_2 = \begin{bmatrix} 0.81 & 0 \\ 0 & 0.81 \end{bmatrix} \tag{43}$$

Here, $\Xi$ is a 12-by-9 matrix with a rank of 8. Since there are 9 unknown variables (6 $q$s and 3 $r$s), the covariance matrix elements are not identifiable. However, if $E[v(k)v(k)']$ is diagonal, as is typically the case, then the covariance matrix elements are identifiable because there are only 6 unknown variables (full $R$ matrix and three diagonal elements of $Q$).

Note that the minimal polynomial can be used to estimate the unknown covariances $R$ and $Q$ via quadratic programming techniques. Furthermore, it can be used to estimate the optimal gain $W$, as in [47] and Appendix B; however, reliable and accurate estimation of the parameters of vector moving average processes is still an unresolved problem [16], [24], [31], [43].

## IV APPROACHES TO OBTAIN FILTER PARAMETERS

There are two competing approaches for the estimation of the filter parameters $W$, $R$, $Q$, and $\bar{P}$. The first approach is to estimate the noise covariance

matrices first and subsequently the Kalman filter gain $W$ and the predicted state covariance $\bar{P}$ are computed given the estimated noise covariance matrices [32], [46]. This method has an underlying problem in that it involves the sum of two moving average processes. Additionally, the autoregressive moving average (ARMA) approach, pioneered in the econometric literature, does not extend naturally to sums of moving average processes and we have found the resulting algorithms [16], [24], [31], [43] to have erratic computational behavior.

The second approach is to estimate the Kalman filter gain $W$ from the measured data first [7], [33]. Given the optimal $W$, we can compute $R$, $Q$ and $\bar{P}$ (this approach is applied in this paper). The proposed $R$, $Q$ and $\bar{P}$ estimates in this paper are valid as long as an optimal gain $W$ is provided. There are many ways to obtain the optimal Kalman filter gain $W$. The techniques listed in this paper to obtain the optimal $W$, that is, Section V and Appendix B are by no means all-inclusive, and several such methods may be suitable for a given problem. For example, the optimal gain $W$ can be obtained from the suboptimal Kalman filter residual [8], solving the minimal polynomial problem [47], and utilizing the least squares method on the observable form [6], to name a few.

## V ESTIMATION OF $W$

This section includes the discussion of two different approaches to estimate the optimal Kalman filter gain $W$, namely, the minimal polynomial approach and the successive approximation, coupled with an adaptive gradient descent scheme, on a criterion based on innovation correlations. The derivation of the minimal polynomial approach is detailed in Appendix B. This approach assumes the system to be purely driven by the optimal innovation. In doing so, the estimation of the optimal Kalman gain can be achieved via a vector auto-regressive model approximation of a vector moving average process. However, from limited testing on examples chosen in this paper, this approach was found to be numerically unstable, only performing well on systems with no eigenvalues close to unity. In fact, the vector auto-regressive model has various numerical problems and an accurate and reliable algorithm to obtain the solution still remains to be developed [24]. Therefore, we omit this approach from the paper and focus on minimization of the innovation correlations using a successive approximation and adaptive gradient descent method.

In the sequel, we describe in detail the approach of our paper using the correlation-based criterion. If the Kalman filter gain $W$ is not optimal, the innovation sequence $\{\nu(k)\}_{k=1}^{N}$ is correlated. We can use the innovation sequence of any stable suboptimal Kalman filter and compute $M$ sample covariance matrices, as in [33]:

$$\hat{C}(i) = \frac{1}{N-M}\sum_{j=1}^{N-M}\nu(j)\nu(j+i)' \qquad i = 0, 1, 2, \ldots, M-1 \tag{44}$$

We know that the optimal Kalman filter gain $W$ makes the autocorrelation function $\hat{C}(i), i = 0, 1, 2, \ldots, M-1$ vanish for all $i \neq 0$. Given the correlation matrix for $i \geq 1$ as in [33], that is

$$\begin{aligned} C(i) &= E[\nu(k)\nu(k-i)'] \\ &= H\bar{F}^{i-1}F\left[\bar{P}H' - WC(0)\right] \end{aligned} \tag{45}$$

where $\bar{F}$ is as in (19). We define the objective function $J$ as

$$J = \frac{1}{2}\sum_{i=1}^{M-1}\left[\operatorname{diag}\left(\hat{C}(0)\right)\right]^{-\frac{1}{2}} \hat{C}(i)'\left[\operatorname{diag}\left(\hat{C}(0)\right)\right]^{-1}\hat{C}(i)\left[\operatorname{diag}\left(\hat{C}(0)\right)\right]^{-\frac{1}{2}} \tag{46}$$

where $\operatorname{diag}(C)$ is the Hadamard product of an identity matrix, of same dimension as $C$, with $C$

$$\operatorname{diag}(C) = I \odot C \tag{47}$$

The optimal $J$ becomes 0 as the sample size $N$ tends to $\infty$ because the time averages are the same as ensemble averages given infinite data. Substituting (45) into (46) and utilizing the cyclic property of trace, we have

$$J = \frac{1}{2}\operatorname{tr}\left\{\sum_{i=1}^{M-1}\Theta(i)XE^2X'\right\} \tag{48}$$

where

$$\Theta(i) = \Phi(i)'E^2\Phi(i) \tag{49}$$

$$\Phi(i) = H\bar{F}^{i-1}F \tag{50}$$

$$X = \Psi - WC(0) \tag{51}$$

$$\Psi = \bar{P}H' \tag{52}$$

$$E = \left[\operatorname{diag}\left(C(0)\right)\right]^{-\frac{1}{2}} \tag{53}$$

For ill-conditioned systems, a regularization term $\lambda_W \mathrm{tr}(WW')$ can be added to convexify the objective function. Taking the gradient[3] of (48) with respect to $W$, we get

$$\nabla_W J = -\sum_{i=1}^{M-1} \Phi(i)' E^2 C(i) E^2 C(0) - F' Z F X - \sum_{\ell=0}^{i-2} \left[ C(\ell+1) E^2 C(i)' E^2 H \bar{F}^{i-\ell-2} \right]' \tag{54}$$

and $Z$ is given by the Lyapunov equation

$$Z = \bar{F}' Z \bar{F} + \frac{1}{2} \sum_{i}^{M-1} \Phi(i)' E^2 \hat{C}(i) E^2 H + \left( \Phi(i)' E^2 \hat{C}(i) E^2 H \right)' \tag{55}$$

and $X$ is obtained by rewriting (45) as

$$\begin{bmatrix} HF \\ H\bar{F}F \\ \vdots \\ H\bar{F}^{M-1}F \end{bmatrix} X = \begin{bmatrix} \hat{C}(1) \\ \hat{C}(2) \\ \vdots \\ \hat{C}(M-1) \end{bmatrix} \tag{56}$$

Then, we can obtain $X$ as

$$X = \begin{bmatrix} HF \\ H\bar{F}F \\ \vdots \\ H\bar{F}^{M-1}F \end{bmatrix}^{\dagger} \begin{bmatrix} \hat{C}(1) \\ \hat{C}(2) \\ \vdots \\ \hat{C}(M-1) \end{bmatrix} \tag{57}$$

where $A^{\dagger}$ is the pseudoinverse of $A$, defined by

$$A^{\dagger} = (A'A)^{-1} A' \tag{58}$$

which exists, since we assume the system to be completely observable and $M \geq n_x$. The gradient direction can be used to obtain the optimal Kalman filter gain $W$ iteratively through the bold driver method in [3], [29], [50]. Details of this application can be found in Section VIII-C2

## VI ESTIMATION OF $R$

### VI-A GENERAL $R$

Given the steady-state optimal gain $W$ and the innovation covariance $S$, whose estimation is explained later in Section VIII, let $u(k), k = 1, \ldots, N$ be the sequence of post-fit residuals of the Kalman filter, that is,

$$u(k) = z(k) - H\hat{x}(k|k) \tag{59}$$
$$= (I_{n_z} - HW)\nu(k) \tag{60}$$

Note that $(I_{n_z} - HW)$ is invertible (rank $n_z$) because $(I_{n_z} - HW) = RS^{-1}$ (proven below) and due to the assumption that $R$ is positive definite.

**PROPOSITION 1**: Given the optimal steady-state Kalman filter gain $W$, the post-fit residual sequence $u(k)$, and the innovation sequence $\nu(k)$, the joint covariance of these two sequences is

$$\mathrm{Cov}\left( \begin{bmatrix} v(k) \\ u(k) \end{bmatrix} \right) = \begin{bmatrix} S & R \\ R & R - HPH' \end{bmatrix} \tag{61}$$

**PROOF:** On the right hand side of (61), the (1,1) block is simply the definition of the innovation covariance matrix in (15). Using (60), the (1,2) block in (61) is, given by

$$E[u(k)\nu(k)'] = (I_{n_z} - HW)E[\nu(k)\nu(k)'] = (I_{n_z} - HW)S \tag{62}$$

Using (7) and (8),

$$E[u(k)\nu(k)'] = (I - H\bar{P}H'S^{-1})S \tag{63a}$$
$$= S - H\bar{P}H' = R \tag{63b}$$

Equations (63a)-(63b) were derived by Heffes [18] and Kailath [22].

The (2,2) block of (61) is obtained as follows.

$$G = E[u(k)u(k)'] \tag{64}$$
$$= E\left\{ [(I_{n_z} - HW)\,\nu(k)] \; [(I_{n_z} - HW)\,\nu(k)]' \right\} \tag{65}$$
$$= (I_{n_z} - HW)S(I_{n_z} - HW)' \tag{66}$$
$$= R(I_{n_z} - HW)' = R - RW'H' \tag{67}$$

which, given (14), simplifies to

$$G = R - HPH' \tag{68}$$

Note that the determinant of (61) is

$$\begin{vmatrix} S & R \\ R & R - HPH' \end{vmatrix} = |G - RS^{-1}R| = 0 \tag{69}$$

where the relationship $G = RS^{-1}R$ is proved in Proposition 2 below.

**PROPOSITION 2**: Given the optimal steady-state Kalman filter gain $W$ and the corresponding

[3]Detailed steps on the gradient computation are provided in Appendix E.

post-fit residual $u(k)$ and innovation $\nu(k)$ sequences, the covariance matrix $R$ can be computed in the following five ways:

$$R1 : R = (I_{n_z} - HW)S \tag{70}$$

$$R2 : R = \frac{1}{2}\left\{E[u(k)\nu(k)' + E[\nu(k)u(k)']\right\} \tag{71}$$

$R3$ : Exploiting the fact that

$$G = RS^{-1}R \tag{72}$$

$$R4 : R = \frac{1}{2}\left[G + S - HWSW'H'\right] \tag{73}$$

$$R5 : R = \frac{1}{2}\left\{G(I_{n_z} - W'H')^{-1} + (I_{n_z} - HW)^{-1}G\right\} \tag{74}$$

**PROOF: R1** is proven in (62). Method **R2** to estimate $R$ is by symmetrizing (62). For method **R3** to estimate $R$, we can substitute (8) in (68) and rewrite $G$ as

$$G = R - H\bar{P}H' + HWSW'H' \tag{75}$$

Then, by substituting (14) into (75)

$$G = R - H\bar{P}H' + H\bar{P}H'S^{-1}H\bar{P}H' \tag{76}$$

We also know from (15)

$$(S - R) = H\bar{P}H' \tag{77}$$

By substituting (77) into (76), we can write $G$, defined in (65), as

$$G = R - (S - R) + (S - R)S^{-1}(S - R) \tag{78}$$

$$= RS^{-1}R \tag{79}$$

which is a continuous-time algebraic Riccati equation. Therefore, we can estimate $R$ by solving the continuous-time Riccati equation, as in [1], or Kleinman's method [27]. Some additional methods to solve the continuous-time algebraic Riccati equation can be found in [30]. We can also interpret (72) in terms of a Linear Quadratic Regulator (LQR) optimal control problem, where we can obtain $R$ as the solution of the continuous-time algebraic Riccati equation associated with the optimal gain in the LQR problem. The computation of $R$ is also related to the simultaneous diagonalization problem[4] in linear algebra [49]. Note that, in the scalar case, $R$ is the geometric mean of the covariance of the post-fit residual and the innovation, as in the (1,2) block of (61).

For **R4**, we substitute (77) into (75) and rewrite $G$ as

$$G = R - (S - R) + HWSW'H' \tag{80}$$

$$= 2R - S + HWSW'H' \tag{81}$$

Solving for $R$, we obtain

$$R = \frac{1}{2}\left\{G + S - HWSW'H'\right\} \tag{82}$$

thus, proving **R4**.

For **R5**, recall (62). We can rewrite (66) as

$$G = (I_{n_z} - HW)S(I_{n_z} - HW)' \tag{83}$$

$$= R(I_{n_z} - HW)' = (I_{n_z} - HW)R \tag{84}$$

Thus, we can compute $R$ as

$$\hat{R} = (I_{n_z} - HW)^{-1}G \tag{85}$$

$$= G\,(I_{n_z} - W'H')^{-1} \tag{86}$$

We can symmetrize the estimate of $R$ by

$$\hat{\hat{R}} = \frac{1}{2}\left\{G\,(I_{n_z} - W'H')^{-1} + (I_{n_z} - HW)^{-1}\,G\right\} \tag{87}$$

proving **R5**.

Note that **R1**–**R5** are theoretically the same; however, they are numerically different. We recommend **R3**, since it ensures the positive definiteness of $R$.

### VI-B DIAGONAL $R$

When $R$ is diagonal, we want to solve the least squares problem of

$$\min_{X \geq 0} ||X - R||_{\mathrm{F}}^2 \tag{88}$$

where F, indicates the Frobenius norm. The positive definite $R$ can be estimated from **R3**, given in Proposition 2. The solution is simply the diagonal elements of the estimated $R$ from **R3**.

### VI-C USE OF SMOOTHED STATE ESTIMATE WITH ONE-STEP-LAG POST-FIT RESIDUALS

Note that $R$ can also be estimated using one-step-lag smoothing on the post-fit residuals. Let us define the one-step-lag smoothed residual $s(k)$ as in [37], that is,

$$s(k) = z(k) - H\hat{x}(k|k+1) \tag{89}$$

$$= z(k) - H\hat{x}(k|k) - HW_1\nu(k+1) \tag{90}$$

$$W_1 = \bar{P}\tilde{F}'\bar{P}^{-1}W \tag{91}$$

[4]The solution via Cholesky decomposition and eigen decomposition or simultaneous diagonalization can be found in Appendix C and Appendix D, respectively.

where $\tilde{F}$ is defined as

$$\tilde{F} = (I_{n_x} - WH)F = F^{-1}\bar{F}F \tag{92}$$

From (59), we can also write $s(k)$ as a one-step moving average process

$$\begin{aligned} s(k) &= u(k) - HW_1\nu(k+1) && (93)\\ &= (I_{n_z} - HW)\nu(k) - HW_1\nu(k+1) && (94)\end{aligned}$$

Therefore,

$$E[s(k)\nu(k)'] = (I_{n_z} - HW)C(0) - HW_1C(1)' \tag{95}$$

and for the optimal Kalman filter gain $W$, we can write (95) as

$$E[s(k)\nu(k)'] = (I_{n_z} - HW)S = R \tag{96}$$

A similar expression can be derived for $E[s(k)u(k)']$, that is,

$$\begin{aligned} E[s(k)u(k)'] = (I_{n_z} - HW)C(0)(I_{n_z} - HW)' \\ - HW_1C(1)'(I_{n_z} - HW)' \end{aligned} \tag{97}$$

and for the optimal Kalman filter gain $W$, we have

$$\begin{aligned} E[s(k)u(k)'] &= (I_{n_z} - HW)S(I_{n_z} - HW)' \\ &= RS^{-1}R = G \end{aligned} \tag{98}$$

Lastly, the expression for $E[s(k)s(k)']$ is

$$\begin{aligned} E[s(k)s(k)'] = (I_{n_z} - HW)C(0)(I_{n_z} - HW)' \\ + HW_1C(0)W_1'H' \\ - (I_{n_z} - HW)C(1)(W_1)'H' \\ - HW_1C(1)'(I_{n_z} - HW)' \end{aligned} \tag{99}$$

and with the optimal Kalman filter gain $W$, combined with (14), we get

$$\begin{aligned} E[s(k)s(k)'] &= RS^{-1}R + H\bar{P}\tilde{F}'\bar{P}^{-1}WSW'\bar{P}^{-1}WSW'\bar{P}^{-1}\tilde{F}\bar{P}H' && (100)\\ &= RS^{-1}R + RW'F'(I_{n_x} - H'W')H'R^{-1}SR^{-1}H(I_{n_x} - WH)FWR && (101)\\ &= RS^{-1}R + RW'F'H'S^{-1}SS^{-1}HFWR && (102)\\ &= R(S^{-1} + W'F'H'S^{-1}HFW)R && (103)\end{aligned}$$

Note that $E[s(k)s(k)']$ can be used in a manner similar to the algorithm in Section V to obtain the optimal Kalman filter gain $W$. More investigation is needed into this approach.

## VII ESTIMATION OF $Q$, $P$ AND $\bar{P}$

In this section, we discuss a method to estimate the process noise covariance $Q$ and the state prediction (updated) covariance $\bar{P}$ ($P$). Unlike the case of a Wiener process and for a process with $H = I$, where both $Q$ and $\bar{P}$ can be estimated separately and without iteration, as shown in Section IX-A3, $Q$ and $\bar{P}$ ($P$) are coupled in the general case, requiring multiple iterations for the estimation to converge. The relationship between the steady-state state prediction covariance matrix $\bar{P}$ and the steady-state updated state covariance matrix $P$ with the process noise covariance matrix $Q$ is

$$\begin{aligned} \bar{P} &= FPF' + \Gamma Q\Gamma' && (104)\\ &= F\left(\bar{P}^{-1} + H'R^{-1}H\right)^{-1}F' + \Gamma Q\Gamma' && (105)\\ &= \bar{F}\bar{P}\bar{F}' + FWRW'F' + \Gamma Q\Gamma' && (106)\end{aligned}$$

Similarly, the steady-state updated state covariance matrix can be written as

$$\begin{aligned} P &= \tilde{F}P\tilde{F}' + WRW' \\ &\quad + (I_{n_x} - WH)\Gamma Q\Gamma'(I_{n_x} - WH)' && (107)\\ &= \left(\bar{P}^{-1} + H'R^{-1}H\right)^{-1} && (108)\\ &= \left[(FPF' + \Gamma Q\Gamma')^{-1} + H'R^{-1}H\right]^{-1} && (109)\end{aligned}$$

where $\tilde{F}$ is defined as in (92) and (108) is derived utilizing (14) and the fact (from [2]) that

$$P = (I_{n_x} - WH)\bar{P} \tag{110}$$

Note that we can also define $\tilde{P}$ as

$$\tilde{P} = FPF' = \bar{F}\tilde{P}\bar{F}' + FWRW'F' + \bar{F}\Gamma Q\Gamma'\bar{F}' \tag{111}$$

Given $\tilde{P}$ and $S$, or $P$ and $S$, or $\bar{P}$ and $S$, we can compute $\Gamma Q\Gamma'$ in the following ways:

$$Q1 : \Gamma Q \Gamma' = F^{-1}\tilde{P}(F^{-1})' + WSW' - \tilde{P} \tag{112}$$

$$Q2 : \Gamma Q \Gamma' = P + WSW' - FPF' \tag{113}$$

$$Q3 : \Gamma Q \Gamma' = \bar{P} - F\bar{P}F' + FWSW'F' \tag{114}$$

where **Q1** – **Q3** are derived from (6).

In this paper, we utilize the updated state covariance matrix to estimate $Q$ and $P$, iteratively. Let $t = 0, 1, \ldots$ and $\ell = 0, 1, \ldots$ denote the iteration indices, and let us assume the initial estimate $\Gamma Q^{(0)} \Gamma' = WSW'$ (this is the Wiener process solution for the estimation of $Q$, as shown in Section IX). Let us initialize $P$ by solving the Lyapunov equation (starting with $t = 0$ and $\ell = 0$)

$$P^{(0)} = \tilde{F} P^{(0)} \tilde{F}' + WRW' + (I_{nx} - WH)\,\Gamma Q^{(t)} \Gamma' (I_{nx} - WH)' \tag{115}$$

for $P^{(0)}$. We compute $P^{(\ell+1)}$ utilizing (109) until the value converges, that is,

$$P^{(\ell+1)} = \left[ \left( F P^{(\ell)} F' + \Gamma Q^{(t)} \Gamma' \right)^{-1} + H' R^{-1} H \right]^{-1} \tag{116}$$

Given the converged $P$, let us denote $D^{(t+1)}$ as

$$D^{(t+1)} = P + WSW' - FPF' \tag{117}$$

Then, we can update $Q^{(t+1)}$ from (113)

$$Q^{(t+1)} = \Gamma^{\dagger} D^{(t+1)} (\Gamma')^{\dagger} \tag{118}$$

A mask matrix $A$ can shape $Q$ into a corresponding structure. The mask matrix comprises binary matrix elements with a 1 in the desired positions and 0, elsewhere, for example, as in an identity matrix. Then $Q$ is structured by

$$Q^{(t+1)} = A \odot Q^{(t+1)} \tag{119}$$

where $\odot$ is the Hadamard product. We subsequently set $\ell = 0$ and recompute $P$ using $Q^{(t+1)}$ in (116), and this process repeats until the estimate of $Q$ converges. For ill-conditioned systems, a tuning (regularization) parameter $\lambda_Q$ can be used in (118), that is

$$Q^{(t+1)} = \Gamma^{\dagger} \left[ D^{(t+1)} + \lambda_Q I_{n_x} \right] (\Gamma')^{\dagger} \tag{120}$$

After the estimate of $Q$ converges, we can estimate $\bar{P}$ using either (104), (105) or (106).

## VIII ITERATIVE ALGORITHM TO ESTIMATE STEADY-STATE $W$, $S$, $P$ ($\bar{P}$), $Q$ AND $R$

Given the methods to obtain estimates of $R$ and $Q$ in Sections VI and VII, we summarize our method into a six-step solution approach to obtain the optimal steady-state $W$, $S$, $P$ ($\bar{P}$), $Q$, and $R$.

### VIII-A STEP 1

Start with iteration $r = 0$ and initialize a $W_0$ to stabilize the system as in [28]. We execute the Kalman filter for samples $k = 1, 2, \ldots, N$ as

$$\hat{x}^{(r)}(k+1|k) = F\hat{x}^{(r)}(k|k) \tag{121}$$

$$\nu^{(r)}(k+1) = z(k+1) - H\hat{x}^{(r)}(k+1|k) \tag{122}$$

$$\hat{x}^{(r)}(k+1|k+1) = \hat{x}^{(r)}(k+1|k) + W^{(r)}\nu^{(r)}(k+1) \tag{123}$$

$$u^{(r)}(k+1) = z(k+1) - H\hat{x}^{(r)}(k+1|k+1) \tag{124}$$

### VIII-B STEP 2

Compute $M$ sample covariance matrices, as in (44).

### VIII-C STEP 3

In this step, we check whether any of the termination conditions given below are met. If none of the termination conditions are met, we update the Kalman filter gain via the proposed method, detailed later in Section VIII-C2.

#### VIII-C1 Termination Conditions

There are four conditions that result in algorithm termination, subsequently yielding a Kalman filter gain $W$ for $R$, $Q$ and $\bar{P}$ estimates in later steps:

Condition 1: The converged Kalman filter gain is within a specified threshold $\zeta_W$.

Condition 2: The gradient of Kalman filter gain (54) is within a specified threshold $\zeta_\Delta$.

Condition 3: The objective function value in (46) is within a specified threshold $\zeta_J$ from zero.

Condition 4: The objective function value stops improving, given a specified "patience" (number of epochs, detailed in Section VIII-C2) for the adaptive gradient method.

Condition 5: The maximum number of iterations is reached.

#### VIII-C1a Condition 1:

Let $\Delta W$ be the change in the Kalman filter gain from iteration $r$ to $r+1$, that is

$$\Delta W = W^{(r+1)} - W^{(r)} \tag{125}$$

then

$$\delta_W = ||\Delta W./(W^{(r)} + \epsilon_W)|| \tag{126}$$

where $./$ indicates element-wise division and $||\cdot||$ is a matrix norm (In this paper, the authors use the Euclidean norm) and $\epsilon_W$ is a very small value to protect against zeros in the denominator. When $\delta_W$ is less than a specified threshold $\zeta_W$, the Kalman filter gain is assumed to have converged and we terminate the algorithm; otherwise, we update the Kalman filter gain $W$ for the next iteration.

#### VIII-C1b Condition 2:

We also examine the gradient of the Kalman filter gain $\nabla_W J$ for convergence. We assume the Kalman filter gain to be converged when the Euclidean norm of $\nabla_W J$ is less than a sufficiently small threshold $\zeta_\Delta$, that is,

$$\|\nabla_W J\|_2 < \zeta_\Delta \tag{127}$$

#### VIII-C1c Condition 3:

Similar to $W$, we can compute the change in the objective function $J$ from iteration $r$ to $r+1$. The Kalman filter gain is assumed to have converged when $J^{(r)}$ is less than a specified threshold $\zeta_J$; otherwise, we update the Kalman filter gain for the next iteration.

#### VIII-C1d Condition 4:

The fourth termination condition is related to the step size for the proposed approximation method. We adapt the bold driver method in [3], [29], [50] and the method considers a "patience" parameter to indicate that the objective function value $J^{(r)}$ has stopped improving (detailed in Section VIII-C2). The algorithm is terminated with the Kalman filter gain corresponding to minimum $J^{(r)}$.

#### VIII-C1e Condition 5:

This condition is implemented to ensure that the algorithm terminates within a reasonable number of iterations, denoted by $n_L$. Typically, the number of iterations required to reach the optimal Kalman filter gain $W$ increases proportionally with $n_x$.

### VIII-C2 Kalman Filter Gain Update

When any of the above conditions are met, we terminate the algorithm. Otherwise, we update the Kalman filter gain $W$ for the next iteration $r+1$ via the gradient direction in (54). Given the gradient direction, the Kalman filter gain at iteration $r+1$ is updated by

$$W^{(r+1)} = W^{(r)} - \alpha^{(r)} \nabla_W J \tag{128}$$

where $\alpha^{(r)}$ is the step size for the proposed method. The step size is initialized as

$$\alpha^{(0)} = \min\left(c\left(\frac{N}{N_s}\right)^{\beta}, c\right) \tag{129}$$

where $c$ is a positive constant and is used to update the Kalman filter gain in the first iteration, $N_s$ is a hyperparameter on the number of observations, and $\beta$ is a positive constant to adapt the initial step size to the number of observations.

Subsequently, $\alpha^{(r)}$ is computed using the bold driver method in [3], [29], [50]. That is, after each iteration, we compare the $J^{(r)}$ to its previous value, $J^{(r-1)}$, and set

$$\alpha^{(r)} = \begin{cases} 0.5\alpha^{(r-1)}, & \text{if } J^{(r)} > J^{(r-1)} \\ \max(1.1\alpha^{(r-1)}, \bar{c}), & \text{otherwise} \end{cases} \tag{130}$$

where $\bar{c}$ is the maximum step size defined as,

$$\bar{c} = \min\left(\left(\frac{N}{N_s}\right)^{\beta}, c_{\max}\right) \tag{131}$$

and $c_{\max}$ is a positive constant between 0 and 1.

Once we update the Kalman filter gain $W$, we go back to Step 1 by setting $r = r+1$ and repeat the same process until any of the five termination conditions are met.

Note that each time $J^{(r)} \leq J^{(r-1)}$, we save the corresponding Kalman filter gain $W^{(r)}$ and $J^{(r)}$, and we halve the step size each time $J^{(r)} > J^{(r-1)}$ in the hope of observing a decrease in $J^{(r)}$. If the value of $J^{(r)}$ has consecutively increased for a specified number of iterations (i.e., given a "patience" factor), we select the best Kalman filter gain $W$ by

$$W = \arg\min_r J^{(r)} \tag{132}$$

We then terminate the iteration and move onto Step 4 after repeating Steps 1 and 2 with the corresponding $W$. Note that adaptive stochastic gradient descent methods can be applied to compute the optimal Kalman filter gain $W$ as in [21], [26], [40], [48], [51].

### VIII-D STEP 4

Once we obtain the optimal steady-state Kalman filter gain $W$ and the corresponding innovation covariance $S$, we can compute the unknown $R$, as in Section VI.

### VIII-E STEP 5

Given the covariance matrix $R$, computed in Step 4, we can compute the covariance matrix $Q$ and steady-state state prediction covariance matrix $\bar{P}$, as detailed in Section VII.

### VIII-F STEP 6

We implement a successive approximation as follows: an outer-loop is used to reinitialize with the $R$ and $Q$ obtained from Step 5 and then reinvoke Steps 1-5. We keep track of the best $J^{(r)}$ among the outer-loop iterations. The Kalman filter gain associated with the minimum $J^{(r)}$ is selected to be the optimal Kalman filter gain. The algorithm terminates when the difference between the best $J^{(r)}$ from each outer-loop is less than $\zeta_J$ or the maximum number of outer-loop iterations is reached.

## IX SPECIAL CASES: WIENER PROCESS AND $H = I_{n_x}$ CASES

In this section, we consider two special cases below. The first case is when the state transition matrix $F$ and the measurement matrix $H$ are both identity matrices, $I_{n_x}$ and $I_{n_z}$, respectively. This considerably simplifies our method to estimate $R$ and $Q$. The second special case is when only the measurement matrix $H$ is the identity matrix, while the state transition matrix $F$ remains general.

### IX-A CASE 1: WIENER PROCESS

For a Wiener process, we have $F = I_{n_x}$ and $H = I_{n_z}$.

#### IX-A1 Kalman Filter Gain Update for a Wiener Process

To get the optimal Kalman filter gain, for $k = 1, 2, \ldots, N$,

$$\hat{x}(k|k-1) = \hat{x}(k-1|k-2) + W\nu(k-1) \tag{133}$$

$$\hat{z}(k|k-1) = \hat{x}(k|k-1) \tag{134}$$

$$\begin{aligned} z(k) &= \hat{z}(k|k-1) + \nu(k) \\ &= \hat{x}(k|k-1) + \nu(k) \end{aligned} \tag{135}$$

Define

$$\begin{aligned} \xi(k) &= z(k) - z(k-1) && (136)\\ &= \hat{x}(k|k-1) + \nu(k) - \hat{x}(k-1|k-2) \\ &\quad - \nu(k-1) && (137)\\ &= \nu(k) + (W - I_{n_x})\nu(k-1) && (138) \end{aligned}$$

Then, let us define $L_0$ and $L_1$ as

$$\begin{aligned} L_0 &= E\left[\xi(k)\xi(k)'\right] \\ &= S + (W - I_{n_x})S(W - I_{n_x})' && (139)\\ L_1 &= E\left[\xi(k)\xi(k-1)'\right] = (W - I_{n_x})S && (140) \end{aligned}$$

Note that both $L_0$ and $L_1$ can be computed from samples. Additionally, we can obtain the optimal $W$ from $L_1$ as

$$W = I_{n_x} + L_1 S^{-1} \tag{141}$$

Substituting $W$ in (141) into (139), we can write the relationship between $L_0$ and $L_1$ as

$$\begin{aligned} L_0 &= S + L_1 S^{-1} S S^{-1} L_1' && (142)\\ &= S + L_1 S^{-1} L_1' && (143) \end{aligned}$$

Note that (143) is in a form related to the discrete algebraic Riccati equation and has a positive definite solution [15].

#### IX-A2 Estimation of $R$ for a Wiener Process

**PROPOSITION 3**: For a Wiener process where both the state transition matrix $F$ and the measurement matrix $H$ are both the identity matrices, $I_{n_x}$ and $I_{n_z}$, respectively, and given the optimal steady-state Kalman filter gain $W$, and the concomitant post-fit residual sequence $u(k)$ and innovation sequence $\nu(k)$, the covariance matrix $R$ can be computed in the following ways:

$$SR1: R = (I_{n_z} - W)S \tag{144}$$

$$SR2: R = \frac{1}{2}\left\{E[u(k)\nu(k)' + E[\nu(k)u(k)']\right\} \tag{145}$$

$$SR3: G = RS^{-1}R \tag{146}$$

$$SR4: R = \frac{1}{2}\left[G + S - WSW'\right] \tag{147}$$

$$SR5: R = G - WSW' + \frac{1}{2}(WS + S'W') \tag{148}$$

**PROOF: SR1-SR4** are directly proven by substituting $H = I_{n_z}$ into **R1**–**R4**. For **SR5**, we know from (8) that

$$WS = \bar{P} \tag{149}$$

and we also know from (15) that,

$$S = \bar{P} + R \tag{150}$$

Then,

$$\begin{aligned} G &= (I_{n_x} - W)S(I_{n_x} - W)' && (151)\\ &= (I_{n_x} - W)S - (I_{n_x} - W)SW' && (152)\\ &= S - WS - SW' + WSW && (153)\\ &= R - SW' + WSW' && (154)\end{aligned}$$

Then, we can compute $R$ as

$$R = G + WS - WSW' \tag{155}$$

Symmetrizing (155),

$$R = G - WSW' + \frac{1}{2}(WS + S'W') \tag{156}$$

hence, **SR5** is proven.

### IX-A3 Estimation of $\bar{P}$ and $Q$ for a Wiener Process

Unlike the general case, where multiple iterations are needed to estimate both $Q$ and $\bar{P}$, in the case of a Wiener process, we can estimate them with no iteration.

**PROPOSITION 4**: For a Wiener process, where the state transition matrix $F$ and the measurement matrix $H$ are both identity matrices, $I_{n_x}$ and $I_{n_z}$, respectively, and given the optimal steady-state Kalman filter gain $W$, and the corresponding innovation sequence $\nu(k)$, the steady-state state prediction covariance and the process noise covariance $Q$ can be computed as:

$$\bar{P} = WS \tag{157}$$

$$Q = WSW' \tag{158}$$

**PROOF:** Given the relationship in (8) and knowing that, for a Wiener process $H = I_{n_z}$, using (8), we have (157).

For a Wiener process, we can rewrite the Riccati equation (10) as

$$\bar{P} = \bar{P} - \bar{P}(\bar{P} + R)^{-1}\bar{P} + Q \tag{159}$$

Using the relationship of (15) and (157) in (159) yields

$$\bar{P} = \bar{P} - WSW' + Q \tag{160}$$

Thus, for a Wiener process, $Q$ can be estimated as

$$Q = WSW' \tag{161}$$

Hence, (158) is proven. Note that (158) is used as $Q^{(0)}$ in the general case for iteratively computing $Q$.

## IX-B CASE 2: $H = I_{n_x}$

In the second case, only $H$ is the identity matrix, but $F$ is not necessarily so.

### IX-B1 Kalman filter Gain Update for the $H = I_{n_x}$ Case

To get the optimal Kalman filter gain, for $k = 1, 2, \ldots, N$,

$$\hat{x}(k+1|k) = F\hat{x}(k|k-1) + FW\nu(k) \tag{162}$$

$$\begin{aligned} z(k) &= x(k) + w(k)\\ &= \hat{x}(k|k-1) + \nu(k) && (163)\end{aligned}$$

Let $\xi(k)$ be

$$\xi(k) = z(k) - Fz(k-1) \tag{164}$$

Define

$$\begin{aligned} \xi(k) &= \hat{x}(k|k-1) + \nu(k)\\ &\quad - F\hat{x}(k-1|k-2) - F\nu(k-1) && (165)\\ &= \nu(k) - \bar{F}\nu(k-1) && (166)\end{aligned}$$

where

$$\bar{F} = F(I_{n_x} - W) \tag{167}$$

We can write $L_0 = E\{\xi(k)\xi(k)'\}$ as

$$L_0 = S + \bar{F}S\bar{F}' \tag{168}$$

Similarly, $L_1 = E\{\xi(k)\xi(k-1)'\}$ can be computed based on (166) as,

$$L_1 = -\bar{F}S \tag{169}$$

From the right hand side of (168), we can find $S$ by solving

$$S + L_1 S^{-1} L_1' = L_0 \tag{170}$$

Upon calculating $S$, we can find the optimal Kalman filter gain $W$ as

$$W = I_{n_z} + F^{-1}L_1 S^{-1} \tag{171}$$

and we can calculate $R$ from $R3$, in (79). $G$ can be obtained by running the filter given the optimal Kalman filter gain. Note that, we can also write $\xi(k)$ as

$$\begin{aligned} \xi(k) &= x(k) - Fx(k-1)\\ &\quad + w(k) - Fw(k-1) && (172)\\ &= \Gamma v(k-1) + w(k) - Fw(k-1) && (173)\end{aligned}$$

Then, $L_0$ is

$$L_0 = \Gamma Q \Gamma' + R + FRF' \tag{174}$$

Equating (168) and (174), we can compute $\Gamma Q \Gamma'$ as

$$\Gamma Q \Gamma' = S + \bar{F} S \bar{F}' - (R + FRF') \quad (175)$$
$$= S + FGF' - (R + FRF') \quad (176)$$

Equation (176) follows from

$$G = (I_{n_z} - W) S (I_{n_z} - W)' \quad (177)$$

## X NUMERICAL EXAMPLES

In this section, we illustrate our method on the following five cases:

1) A second-order kinematic system (a white noise acceleration or nearly constant velocity model) by varying the lags $M$ in the correlation.
2) A system described in Neethling [39].
3) A five-state system from [33] and [4] with diagonal $Q$ and $R$.
4) A detectable, but not completely observable, system from [42].
5) A three-state system from [42].

Each case is simulated with 100 Monte Carlo (MC) runs with an assumed "patience" of 5, $\zeta_J = 10^{-6}$, $\zeta_W = 10^{-6}$, $\zeta_\Delta = 10^{-6}$, $c = 0.01$, $c_{\max} = 0.2$, $\beta = 2$ and the maximum outer-loop iteration limit is set to 20. Case 5 is simulated with 200 MC runs to be compatible with the results in [42].

For each test case, we examine the condition number of the system's observability and controllability matrices, as well as matrix $\Xi$. The condition number of matrix $A$ is computed as

$$\kappa(A) = \|A\| \|A^\dagger\| \quad (178)$$

where $A^\dagger$ is the pseudoinverse of $A$ and $||\cdot||$ is a Euclidean norm. The rank of matrix $\Xi$ is also examined for each test case. For each test case result, we compute the 95% probability interval (PI) and denote by $\underline{r}$ and $\overline{r}$ the corresponding lower and upper limits, respectively. We also provide the mean and the root mean squared error (RMSE) of each distribution. The averaged normalized innovation squared (NIS) is also provided to measure the consistency of the Kalman filter,

$$\bar{\epsilon}(k) = \frac{1}{n_{MC}} \sum_{i=1}^{n_{MC}} \nu(k)' S^{-1} \nu(k) \quad (179)$$

where $n_{MC}$ is the number of MC runs. The elements of each matrix $A$ are denoted as $a_{ij}$, representing the element in the $i^{th}$ row and the $j^{th}$ column of $A$.

### X-A CASE 1

We simulated a second-order kinematic system described by

$$x(k) = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} x(k-1) + \begin{bmatrix} \frac{1}{2}T^2 \\ T \end{bmatrix} v(k-1) \quad (180)$$

$$z(k) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(k) + w(k) \quad (181)$$

$$(182)$$

with sampling period $T = 0.1$, where

$$E[v(k)v(j)'] = 0.0025 \quad (183)$$
$$E[w(k)w(j)'] = 0.01 \quad (184)$$

The mean of the process and the measurement noises are assumed to be zero and the corresponding variances are given in (183) and (184), respectively. Note that the system has the condition number of 20.1 for its observability matrix and 20.2 for its controllability matrix. The matrix $\Xi$, given the initial Kalman filter gain in (186), is

$$\Xi = \begin{bmatrix} 5 \cdot 10^{-5} & 6 \\ 2.5 \cdot 10^{-5} & -4 \\ 0 & 1 \end{bmatrix} \quad (185)$$

which has a rank of 2, and we have 2 unknown variables to estimate, implying that $Q$ and $R$ are identifiable. The condition number for $\Xi$ is $1.5 \cdot 10^5$. The least squares problem using the minimal polynomial approach is ill-conditioned.

#### X-A1 Varying the Number of Lags in the Correlations

We performed 100 MC runs, where each run contained $N = 1000$ sample observations. We set $n_L = 100$, $N_s = 1000$, and vary the lags, $M = 10, 20, 30, 40, 50, 100$, with an initial Kalman Filter gain

$$W^{(0)} = \begin{bmatrix} 0.1319 \\ 0.0932 \end{bmatrix}, \quad (186)$$

obtained by solving the Riccati equation with $Q^{(0)} = 0.1$ and $R^{(0)} = 0.1$. Figs.1 and 2 show the box plots of the estimated $R$ using **R3**[5] and $Q$ of 100 MC runs, respectively, with varying $M$.

The bottom and top of each "box" are the first (denoted $\mathbb{Q}_1$) and third (denoted $\mathbb{Q}_3$) quartiles of the estimate, respectively. The line in the middle of each box is the median estimate. The distances between the tops and bottoms are the interquartile ranges (IQR= $\mathbb{Q}_3 - \mathbb{Q}_1$). The whiskers are lines

[5] All (**R1**–**R5**) obtain the same values.

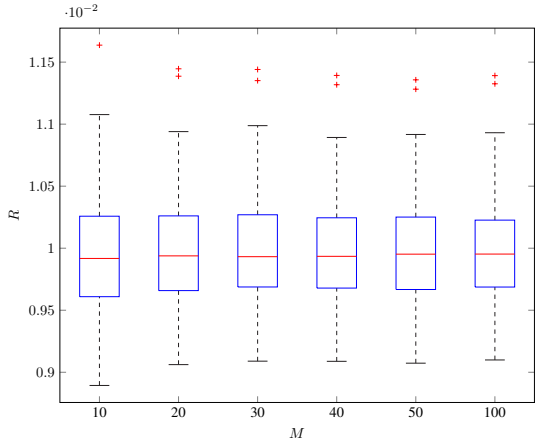



FIGURE 1: 100 Monte Carlo runs for the Kalman filter $R$ estimation using method **R3** with various $M$.

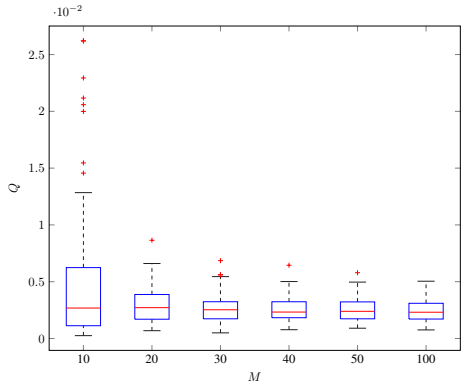



FIGURE 2: 100 Monte Carlo runs for the Kalman filter $Q$ estimation with various $M$.

extending above and below each box and are drawn from each end of the interquartile ranges to the the upper ($\mathbb{Q}_3 + 1.5IQR$) and lower ($\mathbb{Q}_1 - 1.5IQR$) adjacent values. Estimates beyond the whisker length are marked as outliers (indicated by the "+" symbols). The accuracies of the estimates of both $R$ and $Q$ increase with an increase in $M$. Table 1 shows the mean value of the estimates of both $R$ and $Q$. The smallest error of the median of the estimates of $R$ and the variability of the estimates of $Q$ are obtained with $M \geq 100$.

### X-A2 Estimation of $W$ and $\bar{P}$

Given $M = 100$, for 100 MC runs with the initial Kalman Filter gain as in (186), we found that **R1**–**R5** estimate the same $R$ values. The true values of $R$ all lie within the 95% PI associated with the

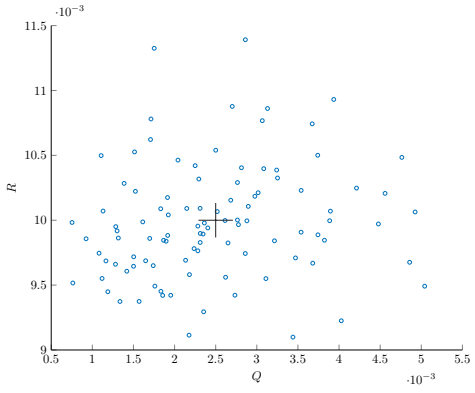



FIGURE 3: $Q$ and $R$ estimation for Case 1.

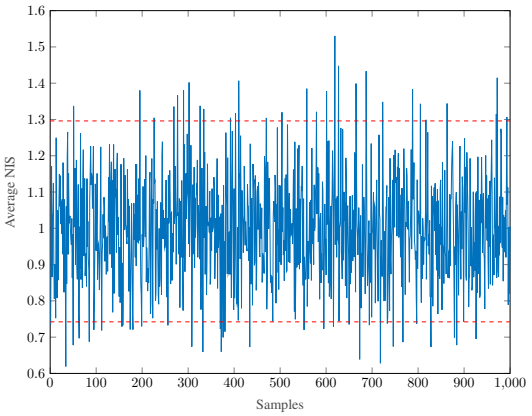



FIGURE 4: Averaged NIS for Case 1.

distribution of estimates. Fig.3 shows the $Q$ versus $R$ plot of each estimate. The true values are marked by "+" symbols. The reason the estimated $Q$ varies so much is that its value is very small compared to the measurement noise. Fig.4 shows the averaged NIS and its 95% probability region, which proves that the filter is consistent.

### X-B CASE 2

We simulated the system described in Neethling [39],

$$x(k) = \begin{bmatrix} 0.8 & 1 \\ -0.4 & 0 \end{bmatrix} x(k-1) + \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} v(k-1) \tag{187}$$

$$z(k) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(k) + w(k) \tag{188}$$

where

$$E[v(k)v(j)'] = 1 \tag{189}$$

$$E[w(k)w(j)'] = 1 \tag{190}$$

TABLE 1: Monte Carlo Simulation for Case 1 Varying the Number of Lags $M$ (Method **R3)**

| | $M$ | | | | | |
|---|---|---|---|---|---|---|
| | 10 | 20 | 30 | 40 | 50 | 100 |
| $R$ | 0.0100 | 0.0100 | 0.0100 | 0.0100 | 0.0100 | 0.0100 |
| $Q$ | 0.0048 | 0.0030 | 0.0027 | 0.0026 | 0.0025 | 0.0025 |

TABLE 2: Monte Carlo Simulation for Case 1 with $M = 100$ and PI= $2\sigma$ (100 Runs)

| | $R$ | | | | | $Q$ |
|---|---|---|---|---|---|---|
| | **R1** | **R2** | **R3** | **R4** | **R5** | |
| Truth | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.0025 |
| $\underline{r}$ | 0.0092 | 0.0092 | 0.0092 | 0.0092 | 0.0092 | $6.49 \cdot 10^{-4}$ |
| Mean | 0.0100 | 0.0100 | 0.0100 | 0.0100 | 0.0100 | 0.0025 |
| $\overline{r}$ | 0.0109 | 0.0109 | 0.0109 | 0.0109 | 0.0109 | 0.0046 |
| RMSE | $4.41 \cdot 10^{-4}$ | $4.41 \cdot 10^{-4}$ | $4.41 \cdot 10^{-4}$ | $4.41 \cdot 10^{-4}$ | $4.41 \cdot 10^{-4}$ | 0.0010 |
| | $W_{11}$ | | | $W_{21}$ | | |
| Truth | 0.0952 | | | 0.0476 | | |
| $\underline{r}$ | 0.0697 | | | 0.0255 | | |
| Mean | 0.0925 | | | 0.0465 | | |
| $\overline{r}$ | 0.1250 | | | 0.0634 | | |
| RMSE | 0.0147 | | | 0.0100 | | |
| | $\bar{P}_{11}$ | | | $\bar{P}_{22}$ | | |
| Truth | 0.0011 | | | $5.13 \cdot 10^{-4}$ | | |
| $\underline{r}$ | $8.47 \cdot 10^{-4}$ | | | $2.82 \cdot 10^{-4}$ | | |
| Mean | 0.0011 | | | $5.13 \cdot 10^{-4}$ | | |
| $\overline{r}$ | 0.0013 | | | $8.72 \cdot 10^{-4}$ | | |
| RMSE | $1.26 \cdot 10^{-4}$ | | | $1.60 \cdot 10^{-4}$ | | |

The system's condition numbers for its observability and controllability matrices are 2.18 and 2.56, respectively. Here, $\Xi$, given the initial Kalman filter gain, is

$$\Xi = \begin{bmatrix} 1.25 & 1.8 \\ 0.5 & -1.12 \\ 0 & 0.4 \end{bmatrix} \tag{191}$$

and the rank is 2. The number of unknown variables is 2, therefore, the system noise variances are estimable. The condition number of $\Xi$ is 2.3 and indeed the minimal polynomial approach works well for this problem. We simulated 100 Monte Carlo runs with $N = 1000$, $n_L = 100$, $N_s = 1000$, and an initial suboptimal Kalman filter gain

$$W^{(0)} = \begin{bmatrix} 0.9 \\ 0.5 \end{bmatrix} \tag{192}$$

(193)

Table 3 shows the estimated noise variances. Similar to the Case 1 result, the mean values of each of the estimated parameters are very close to their corresponding true values. As seen in Table 3, the true values lie within the 95% PI associated with the distribution of estimates for each variable $Q$, $R$, $W$ and $P_{ii}$. Fig.5 shows the $Q$ and $R$ estimates for each MC run. As shown in Fig.6, the Kalman filter is considered consistent.

### X-C CASE 3

In Case 3, we test on the example in [33]. The system matrices are assumed to be as follows.

TABLE 3: Monte Carlo Simulation for Case 2 with $M = 100$ and PI$= 2\sigma$ (100 Runs)

| | $R$ | | | | | $Q$ |
|---|---|---|---|---|---|---|
| | **R1** | **R2** | **R3** | **R4** | **R5** | |
| Truth | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $\underline{r}$ | 0.56 | 0.56 | 0.56 | 0.56 | 0.56 | 0.78 |
| Mean | 1.05 | 1.05 | 1.05 | 1.05 | 1.05 | 0.97 |
| $\overline{r}$ | 1.51 | 1.51 | 1.51 | 1.51 | 1.51 | 1.18 |
| RMSE | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.11 |
| | $W_1$ | | | $W_2$ | | |
| Truth | 0.65 | | | 0.09 | | |
| $\underline{r}$ | 0.49 | | | $-0.03$ | | |
| Mean | 0.63 | | | 0.10 | | |
| $\overline{r}$ | 0.80 | | | 0.25 | | |
| RMSE | 0.08 | | | 0.07 | | |
| | $\bar{P}_{11}$ | | | $\bar{P}_{22}$ | | |
| Truth | 1.89 | | | 0.35 | | |
| $\underline{r}$ | 1.59 | | | 0.30 | | |
| Mean | 1.87 | | | 0.35 | | |
| $\overline{r}$ | 2.07 | | | 0.39 | | |
| RMSE | 0.12 | | | 0.02 | | |

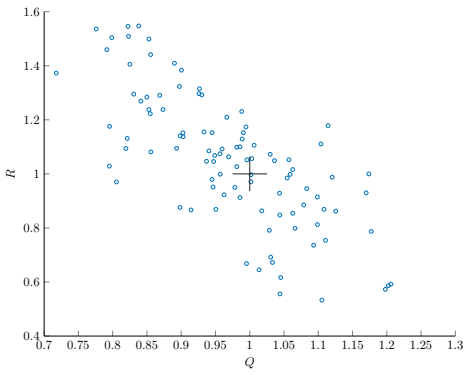



FIGURE 5: $Q$ and $R$ estimation for Case 2.

$$F = \begin{bmatrix} 0.75 & -1.74 & -0.3 & 0 & -0.15 \\ 0.09 & 0.91 & -0.0015 & 0 & -0.008 \\ 0 & 0 & 0.95 & 0 & 0 \\ 0 & 0 & 0 & 0.55 & 0 \\ 0 & 0 & 0 & 0 & 0.905 \end{bmatrix} \tag{194}$$

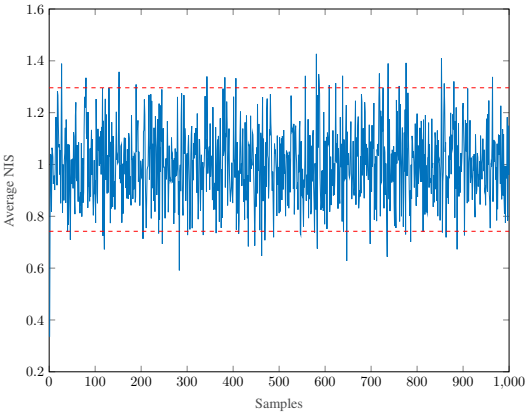



FIGURE 6: Averaged NIS for Case 2.

$$\Gamma = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 24.64 & 0 & 0 \\ 0 & 0.835 & 0 \\ 0 & 0 & 1.83 \end{bmatrix} \tag{195}$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix} \tag{196}$$

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{197}$$

The condition number for the observability matrix is 42.6, and the condition number for the controllability matrix is 54.6. The system has a $\mathrm{rank}(\Xi)$ equal to 5 (utilizing the constraint that both $R$ and $Q$ are diagonal), with a total of 5 unknowns. Hence, the $Q$ and $R$ parameters are identifiable. The condition number of $\Xi$ is 808. The initial Kalman filter gain is obtained by solving the Riccati equation with

$$Q^{(0)} = \begin{bmatrix} 0.25 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.75 \end{bmatrix} \tag{198}$$

$$R^{(0)} = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.6 \end{bmatrix} \tag{199}$$

### X-C1 Minimum Number of Observation Samples Needed for Mehra's and Bélanger's Methods to Converge

Both Mehra's [33] and Bélanger's [4] methods to update the Kalman filter gain $W$ can be unstable unless a large number of data samples are observed. This is due to the fact that the time average converges slowly to the ensemble average. We conducted 100 MC simulations with

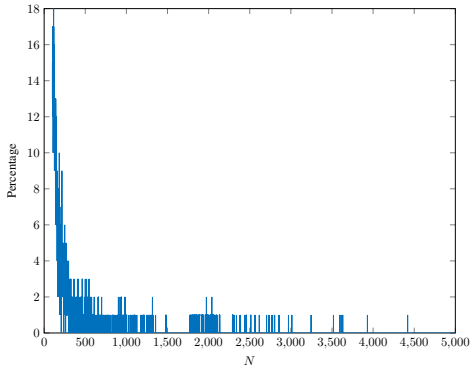



FIGURE 7: Percentage of unstable Kalman filter gains obtained from [33] for varying the total number of observed samples (M = 40).

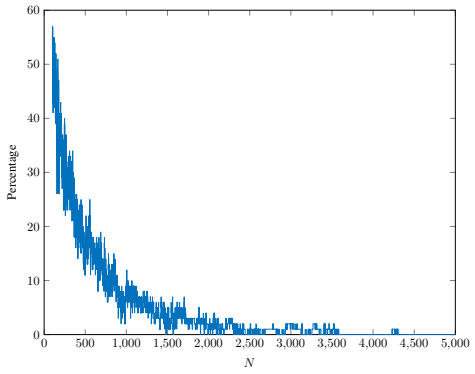



FIGURE 8: Percentage of unstable Kalman filter gains obtained from [4] for varying the total number of observed samples (M = 5).

10,000 data samples in each run given the five-state system described in (194)–(197). We then varied the number of observed samples from 100 to 10,000 and updated the Kalman filter gain using both the Mehra [33] and Bélanger [4] methods. We measure the percentage of unstable Kalman filter gains by checking if any of the eigenvalues of $\bar{F}$ are outside of the unit circle for each run over the 100 MC runs. The results are shown in Figs.7 and 8. We only display up to 5,000 samples for both methods because each approach terminated with a stable gain when the total observation samples exceeded 5,000. The minimum number of samples required to obtain a stable gain from these methods were about 4,500. Our proposed method always results in a stable Kalman filter gain; hence, it is not included in the comparison of methods.

### X-C2 Comparison of Proposed, Mehra's, and Belanger's Gain Update Methods

Given the 100 MC simulations with 10,000 observation samples generated in X-C1 and setting $n_L = 500$, $N_s = 10000$, Table 4 shows the estimation of the Kalman filter gain $W$ over 100 Monte Carlo runs, given three different gain update methods: the proposed method with $M = 40$, Mehra's method with $M = 40$ [33] and Bélanger's method with $M = 5$ [4].

In Table 4, we see that all methods have the true values staying within its 95% PI; however, our proposed method is able to obtain the Kalman filter gain closest to the optimal Kalman filter gain and the RMSE are, on average, 8 and 4 times smaller compared to Mehra's and Bélanger's, respectively. The very small gains $W_{21}$ and $W_{41}$ are (similarly to the small $Q$ from Case 1) are very hard to estimate — they are essentially buried in noise.

We test and compare the proposed method with that of Mehra's and Bélanger's for the estimation of $R, Q$ and $P$ using the methodology described in Sections VI-B and VII, combined with the converged Kalman filter gain from Table 4. The results are shown in Table 5 and Mehra's method results in the true value of $P_{33}$ staying outside of the 95% PI. In comparison to Bélanger's method, the proposed method is vastly more accurate with lower RMSE (2 to 9 times smaller) for all $R$, $Q$, and $\bar{P}$, while Mehra's method obtained a result that is less accurate than Bélanger's method, as expected from the Kalman filter gain results. The reason $r_1$ is so difficult to estimate is that $S_1$ is dominated by the state uncertainty ($S_1 = 65$, $r_1 = 1$), i.e. the measurement noise is "buried" in a much larger innovation. In the case of $r_2 = 1$, one has $S_2 = 2.45$, i.e., $r_2$ is "visible" in the innovations.

### X-C3 Varying The Number of Samples Observed

In this section, we vary the number of samples observed, $N = 1000, 2500, 5000, 10000$ using our six-step approach. The results are detailed in Tables 6 and 7. As expected, the accuracy increases with an increase in $N$. The estimation is greatly degraded for $N < 5000$. Fig.9 illustrates that the Kalman filter is consistent.

TABLE 4: $W$ Estimation Monte Carlo Simulation for Case 3 (100 Runs; 10,000 Samples)

| | $W_{11}$ | $W_{21}$ | $W_{31}$ | $W_{41}$ | $W_{51}$ | $W_{12}$ | $W_{22}$ | $W_{32}$ | $W_{42}$ | $W_{52}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Truth | 0.95 | $2.80{\cdot}10^{-3}$ | $-2.86$ | $-1.76 \cdot 10^{-4}$ | 0.03 | 0.77 | 0.34 | $-1.49$ | 0.25 | -0.77 |
| Proposed Method | $W_{11}$ | $W_{21}$ | $W_{31}$ | $W_{41}$ | $W_{51}$ | $W_{12}$ | $W_{22}$ | $W_{32}$ | $W_{42}$ | $W_{52}$ |
| $\underline{r}$ | 0.93 | $-8.65 \cdot 10^{-3}$ | $-2.94$ | $-0.02$ | 0.02 | 0.71 | 0.31 | $-1.57$ | 0.18 | $-0.84$ |
| Mean | 0.95 | $2.52 \cdot 10^{-3}$ | $-2.86$ | $5.29 \cdot 10^{-4}$ | 0.03 | 0.77 | 0.34 | $-1.50$ | 0.25 | $-0.76$ |
| $\overline{r}$ | 0.97 | 0.01 | $-2.80$ | 0.02 | 0.05 | 0.84 | 0.38 | $-1.41$ | 0.30 | $-0.68$ |
| RMSE | 0.01 | $5.33 \cdot 10^{-3}$ | 0.04 | $9.31 \cdot 10^{-3}$ | $9.60 \cdot 10^{-3}$ | 0.03 | 0.02 | 0.05 | 0.03 | 0.04 |
| Mehra's Method | $W_{11}$ | $W_{21}$ | $W_{31}$ | $W_{41}$ | $W_{51}$ | $W_{12}$ | $W_{22}$ | $W_{32}$ | $W_{42}$ | $W_{52}$ |
| $\underline{r}$ | 0.92 | $-0.04$ | $-3.38$ | $-0.07$ | $-0.11$ | 0.18 | 0.04 | $-2.61$ | $-0.12$ | $-1.17$ |
| Mean | 1.01 | $3.79 \cdot 10^{-4}$ | $-3.15$ | $3.59 \cdot 10^{-3}$ | $-0.02$ | 0.62 | 0.31 | $-1.34$ | 0.28 | $-0.62$ |
| $\overline{r}$ | 1.08 | 0.06 | $-2.83$ | 0.08 | 0.06 | 1.11 | 0.60 | 0.19 | 0.80 | $-0.11$ |
| RMSE | 0.07 | 0.03 | 0.33 | 0.04 | 0.07 | 0.30 | 0.14 | 0.79 | 0.22 | 0.34 |
| Bélanger's Method | $W_{11}$ | $W_{21}$ | $W_{31}$ | $W_{41}$ | $W_{51}$ | $W_{12}$ | $W_{22}$ | $W_{32}$ | $W_{42}$ | $W_{52}$ |
| $\underline{r}$ | 0.89 | $-0.02$ | $-3.02$ | $-0.05$ | $-0.04$ | 0.45 | 0.13 | $-2.27$ | $-0.02$ | $-1.15$ |
| Mean | 0.96 | $1.46 \cdot 10^{-4}$ | $-2.85$ | $3.85 \cdot 10^{-3}$ | 0.03 | 0.77 | 0.33 | $-1.44$ | 0.26 | $-0.77$ |
| $\overline{r}$ | 1.01 | 0.04 | $-2.70$ | 0.04 | 0.10 | 1.13 | 0.52 | $-0.32$ | 0.56 | $-0.32$ |
| RMSE | 0.03 | 0.02 | 0.09 | 0.02 | 0.03 | 0.17 | 0.09 | 0.48 | 0.14 | 0.20 |

TABLE 5: $R$, $Q$ and $\bar{P}$ Estimation Monte Carlo Simulation for Case 3 (100 Runs; 10,000 Samples)

| Method | $R$ | | $Q$ | | | $\bar{P}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $r_1$ | $r_2$ | $q_1$ | $q_2$ | $q_3$ | $\bar{P}_{11}$ | $\bar{P}_{22}$ | $\bar{P}_{33}$ | $\bar{P}_{44}$ | $\bar{P}_{55}$ |
| Truth | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 72.31 | 1.143 | 1213 | 0.932 | 11.74 |
| Proposed Method | $r_1$ | $r_2$ | $q_1$ | $q_2$ | $q_3$ | $\bar{P}_{11}$ | $\bar{P}_{22}$ | $\bar{P}_{33}$ | $\bar{P}_{44}$ | $\bar{P}_{55}$ |
| $\underline{r}$ | 0.043 | 0.918 | 0.941 | 0.662 | 0.802 | 66.06 | 0.930 | 1141 | 0.633 | 9.639 |
| Mean | 1.067 | 1.008 | 0.998 | 1.000 | 0.994 | 72.36 | 1.146 | 1212 | 0.933 | 11.72 |
| $\overline{r}$ | 1.976 | 1.107 | 1.058 | 1.261 | 1.166 | 77.97 | 1.334 | 1290 | 1.172 | 13.76 |
| RMSE | 0.554 | 0.052 | 0.031 | 0.170 | 0.097 | 2.906 | 0.106 | 37.87 | 0.153 | 1.083 |
| Mehra's Method | $r_1$ | $r_2$ | $q_1$ | $q_2$ | $q_3$ | $\bar{P}_{11}$ | $\bar{P}_{22}$ | $\bar{P}_{33}$ | $\bar{P}_{44}$ | $\bar{P}_{55}$ |
| $\underline{r}$ | 0.102 | 0.676 | 0.995 | 0.060 | 0.153 | 69.01 | 0.540 | 1270 | 0.060 | 0.791 |
| Mean | 1.597 | 1.024 | 1.224 | 1.788 | 0.989 | 89.88 | 1.744 | 1505 | 1.715 | 14.22 |
| $\overline{r}$ | 3.681 | 1.199 | 1.420 | 4.432 | 2.240 | 120.7 | 4.067 | 1855 | 4.138 | 34.21 |
| RMSE | 1.484 | 0.212 | 0.251 | 1.464 | 0.652 | 23.38 | 1.280 | 330.0 | 1.419 | 10.61 |
| Bélanger's Method | $r_1$ | $r_2$ | $q_1$ | $q_2$ | $q_3$ | $\bar{P}_{11}$ | $\bar{P}_{22}$ | $\bar{P}_{33}$ | $\bar{P}_{44}$ | $\bar{P}_{55}$ |
| $\underline{r}$ | 0.0270 | 0.755 | 0.885 | 0.043 | 0.292 | 60.37 | 0.564 | 1069 | 0.043 | 2.507 |
| Mean | 1.171 | 1.008 | 0.992 | 1.319 | 1.117 | 74.58 | 1.416 | 1216 | 1.238 | 13.86 |
| $\overline{r}$ | 2.631 | 1.254 | 1.126 | 3.160 | 2.198 | 92.54 | 2.765 | 1370 | 2.902 | 27.078 |
| RMSE | 0.789 | 0.122 | 0.064 | 0.829 | 0.516 | 9.461 | 0.667 | 81.21 | 0.764 | 6.829 |

TABLE 6: $R$, $Q$ and $\bar{P}$ Estimation when Varying the Number of Samples Observed $N$, Monte Carlo Simulation for Case 3 (100 Runs; 1,000–10,000 Samples)

| | $R$ | | $Q$ | | | $\bar{P}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $r_1$ | $r_2$ | $q_1$ | $q_2$ | $q_3$ | $\bar{P}_{11}$ | $\bar{P}_{22}$ | $\bar{P}_{33}$ | $\bar{P}_{44}$ | $\bar{P}_{55}$ |
| Truth | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 72.31 | 1.143 | 1213 | 0.932 | 11.74 |
| $N$ =1,000 | $r_1$ | $r_2$ | $q_1$ | $q_2$ | $q_3$ | $\bar{P}_{11}$ | $\bar{P}_{22}$ | $\bar{P}_{33}$ | $\bar{P}_{44}$ | $\bar{P}_{55}$ |
| $\underline{r}$ | 0.075 | 0.708 | 0.715 | 0.593 | 0.431 | 62.97 | 0.965 | 855.5 | 0.579 | 7.505 |
| Mean | 2.246 | 1.014 | 0.919 | 1.343 | 1.120 | 73.25 | 1.396 | 1139 | 1.263 | 13.45 |
| $\overline{r}$ | 5.221 | 1.358 | 1.137 | 2.847 | 1.542 | 89.88 | 2.204 | 1336 | 2.677 | 20.37 |
| RMSE | 2.060 | 0.182 | 0.133 | 0.663 | 0.308 | 6.998 | 0.398 | 136.6 | 0.626 | 3.511 |
| $N$ =2,500 | $r_1$ | $r_2$ | $q_1$ | $q_2$ | $q_3$ | $\bar{P}_{11}$ | $\bar{P}_{22}$ | $\bar{P}_{33}$ | $\bar{P}_{44}$ | $\bar{P}_{55}$ |
| $\underline{r}$ | 0.041 | 0.767 | 0.864 | 0.635 | 0.719 | 64.94 | 0.941 | 1071 | 0.660 | 8.554 |
| Mean | 1.453 | 1.010 | 0.977 | 1.127 | 1.049 | 73.24 | 1.245 | 1195 | 1.056 | 12.49 |
| $\overline{r}$ | 3.354 | 1.173 | 1.103 | 1.685 | 1.431 | 80.49 | 1.653 | 1342 | 1.638 | 16.44 |
| RMSE | 1.076 | 0.094 | 0.070 | 0.302 | 0.181 | 4.472 | 0.203 | 77.80 | 0.280 | 2.088 |
| $N$ =5,000 | $r_1$ | $r_2$ | $q_1$ | $q_2$ | $q_3$ | $\bar{P}_{11}$ | $\bar{P}_{22}$ | $\bar{P}_{33}$ | $\bar{P}_{44}$ | $\bar{P}_{55}$ |
| $\underline{r}$ | 0.043 | 0.885 | 0.919 | 0.578 | 0.726 | 65.55 | 0.856 | 1126 | 0.550 | 8.864 |
| Mean | 1.100 | 1.011 | 0.997 | 1.008 | 0.981 | 72.27 | 1.148 | 1211 | 0.942 | 11.62 |
| $\overline{r}$ | 2.580 | 1.161 | 1.084 | 1.397 | 1.265 | 80.13 | 1.390 | 1329 | 1.293 | 14.85 |
| RMSE | 0.757 | 0.077 | 0.045 | 0.218 | 0.140 | 3.692 | 0.136 | 52.74 | 0.197 | 1.530 |
| $N$ =10,000 | $r_1$ | $r_2$ | $q_1$ | $q_2$ | $q_3$ | $\bar{P}_{11}$ | $\bar{P}_{22}$ | $\bar{P}_{33}$ | $\bar{P}_{44}$ | $\bar{P}_{55}$ |
| $\underline{r}$ | 0.043 | 0.918 | 0.941 | 0.662 | 0.802 | 66.06 | 0.930 | 1141 | 0.633 | 9.639 |
| Mean | 1.067 | 1.008 | 0.998 | 1.000 | 0.994 | 72.36 | 1.146 | 1212 | 0.933 | 11.72 |
| $\overline{r}$ | 1.976 | 1.107 | 1.058 | 1.261 | 1.166 | 77.97 | 1.334 | 1290 | 1.172 | 13.76 |
| RMSE | 0.554 | 0.052 | 0.031 | 0.170 | 0.097 | 2.906 | 0.106 | 37.87 | 0.153 | 1.083 |

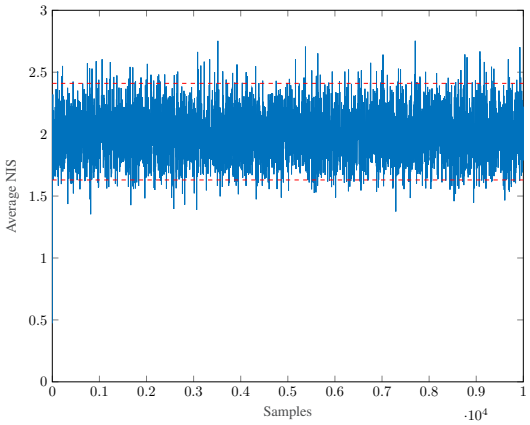



FIGURE 9: Averaged NIS for Case 3.

### X-D CASE 4

For case 4, we simulate the unobservable (but detectable) system in [42],

$$x(k) = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.2 \end{bmatrix} x(k-1) + \begin{bmatrix} 1 \\ 2 \end{bmatrix} v(k-1) \tag{200}$$

$$z(k) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(k) + w(k) \tag{201}$$

with

$$E[v(k)v(j)'] = 1 \tag{202}$$

$$E[w(k)w(j)'] = 1 \tag{203}$$

With the initial Kalman filter gain, the system has

$$\Xi = \begin{bmatrix} 1.04 & 1.09 \\ -0.20 & -0.31 \\ 0 & 0.02 \end{bmatrix} \tag{204}$$

The rank of $\Xi$ is 2 and we have a total of 2 unknown variables. The condition number for $\Xi$ is 23.4. We simulated 100 MC runs with observed samples $N = 1000$ in each run. We set $n_L = 100$, $N_s = 1000$, and $\lambda_Q = 0.1$. Table 8 shows the estimated parameters with the initial Kalman filter gain obtained by solving the Riccati equation with $R^{(0)} = 0.2$, and $Q^{(0)} = 0.4$. Note that the system is *not* fully observable, i.e., the condition number for the observability matrix is infinity, while that for the controllability matrix is 25.8. In Table 8, the true values lie within the 95% PI associated with

TABLE 7: $W$ Estimation when Varying the Number of Samples Observed $N$, Monte Carlo Simulation for Case 3 (100 Runs; 1,000–10,000 Samples)

| | $W_{11}$ | $W_{21}$ | $W_{31}$ | $W_{41}$ | $W_{51}$ | $W_{12}$ | $W_{22}$ | $W_{32}$ | $W_{42}$ | $W_{52}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Truth | 0.95 | $2.80{\cdot}10^{-3}$ | −2.86 | $-1.76 \cdot 10^{-4}$ | 0.03 | 0.77 | 0.34 | −1.49 | 0.25 | −0.77 |
| $N$ =1,000 | $W_{11}$ | $W_{21}$ | $W_{31}$ | $W_{41}$ | $W_{51}$ | $W_{12}$ | $W_{22}$ | $W_{32}$ | $W_{42}$ | $W_{52}$ |
| $\underline{r}$ | 0.90 | −0.04 | −2.97 | −0.05 | −0.03 | 0.56 | 0.27 | −1.64 | 0.14 | −0.95 |
| Mean | 0.96 | $-1.57 \cdot 10^{-3}$ | −2.74 | $7.53 \cdot 10^{-3}$ | 0.04 | 0.79 | 0.33 | −1.49 | 0.26 | −0.79 |
| $\overline{r}$ | 1.01 | 0.03 | −2.56 | 0.08 | 0.09 | 1.00 | 0.41 | −1.27 | 0.44 | −0.54 |
| RMSE | 0.03 | 0.02 | 0.17 | 0.03 | 0.03 | 0.11 | 0.03 | 0.11 | 0.08 | 0.12 |
| $N$ =2,500 | $W_{11}$ | $W_{21}$ | $W_{31}$ | $W_{41}$ | $W_{51}$ | $W_{12}$ | $W_{22}$ | $W_{32}$ | $W_{42}$ | $W_{52}$ |
| $\underline{r}$ | 0.92 | −0.02 | −2.97 | −0.03 | $-6.51 \cdot 10^{-3}$ | 0.64 | 0.28 | −1.64 | 0.14 | −0.90 |
| Mean | 0.96 | $8.95 \cdot 10^{-4}$ | −2.82 | $2.17 \cdot 10^{-3}$ | 0.03 | 0.77 | 0.33 | −1.50 | 0.25 | −0.78 |
| $\overline{r}$ | 1.01 | 0.02 | −2.62 | 0.05 | 0.07 | 0.95 | 0.40 | −1.31 | 0.36 | −0.59 |
| RMSE | 0.02 | 0.01 | 0.10 | 0.02 | 0.02 | 0.08 | 0.03 | 0.09 | 0.05 | 0.08 |
| $N$ =5,000 | $W_{11}$ | $W_{21}$ | $W_{31}$ | $W_{41}$ | $W_{51}$ | $W_{12}$ | $W_{22}$ | $W_{32}$ | $W_{42}$ | $W_{52}$ |
| $\underline{r}$ | 0.92 | −0.02 | −2.98 | -0.03 | $8.68 \cdot 10^{-3}$ | 0.67 | 0.30 | −1.63 | 0.15 | −0.89 |
| Mean | 0.95 | $2.02 \cdot 10^{-3}$ | −2.85 | $1.15 \cdot 10^{-3}$ | 0.03 | 0.77 | 0.34 | −1.49 | 0.25 | −0.76 |
| $\overline{r}$ | 0.98 | 0.01 | −2.77 | 0.03 | 0.06 | 0.90 | 0.39 | −1.32 | 0.32 | −0.64 |
| RMSE | 0.02 | $7.08 \cdot 10^{-3}$ | 0.06 | 0.01 | 0.01 | 0.06 | 0.02 | 0.08 | 0.05 | 0.06 |
| $N$ =10,000 | $W_{11}$ | $W_{21}$ | $W_{31}$ | $W_{41}$ | $W_{51}$ | $W_{12}$ | $W_{22}$ | $W_{32}$ | $W_{42}$ | $W_{52}$ |
| $\underline{r}$ | 0.93 | $-8.65 \cdot 10^{-3}$ | −2.94 | −0.02 | 0.02 | 0.71 | 0.31 | −1.57 | 0.18 | −0.84 |
| Mean | 0.95 | $2.52 \cdot 10^{-3}$ | −2.86 | $5.29 \cdot 10^{-4}$ | 0.03 | 0.77 | 0.34 | −1.50 | 0.25 | −0.76 |
| $\overline{r}$ | 0.97 | 0.01 | −2.80 | 0.02 | 0.05 | 0.84 | 0.38 | −1.41 | 0.30 | −0.68 |
| RMSE | 0.01 | $5.33 \cdot 10^{-3}$ | 0.04 | $9.31 \cdot 10^{-3}$ | $9.60 \cdot 10^{-3}$ | 0.03 | 0.02 | 0.05 | 0.03 | 0.04 |

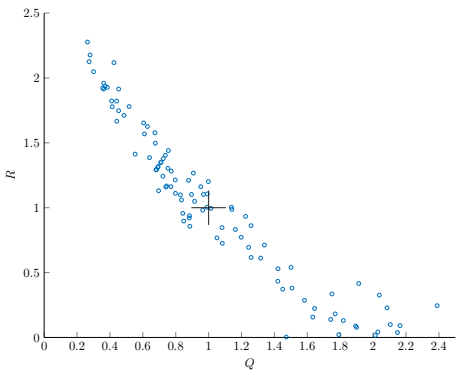



FIGURE 10: $Q$ and $R$ estimation for Case 4.

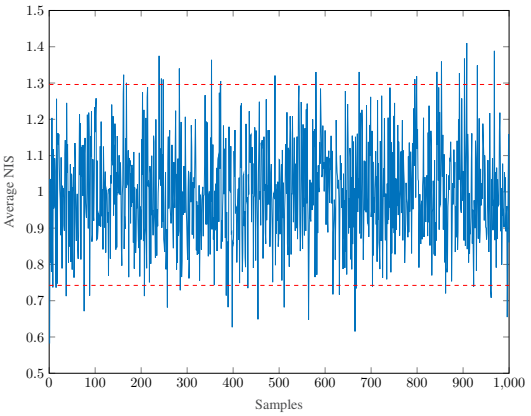



FIGURE 11: Averaged NIS for Case 4.

each distribution. Fig.10 shows a wide variation of $Q$ and $R$ estimates; however, the NIS in Fig.11 shows that the Kalman filter is consistent.

### X-E CASE 5

In Case 5, we simulate the system from [42],

$$x(k) = \begin{bmatrix} 0.1 & 0 & 0.1 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.3 \end{bmatrix} x(k-1) + \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} v(k-1) \tag{205}$$

$$z(k) = \begin{bmatrix} 0.1 & 0.2 & 0 \end{bmatrix} x(k) + w(k) \tag{206}$$

TABLE 8: Monte Carlo Simulation for Case 4 with $M = 100$ and PI$= 2\sigma$ (100 Runs; 1000 Samples)

| | $R$ | $Q$ | $W_1$ | $W_2$ | $\bar{P}_{11}$ | $\bar{P}_{22}$ |
|---|---|---|---|---|---|---|
| Truth | 1.00 | 1.00 | 0.50 | 1.01 | 1.01 | 4.08 |
| $\underline{r}$ | $3.27 \cdot 10^{-3}$ | 0.26 | 0.05 | 0.32 | 0.27 | 1.09 |
| Mean | 1.04 | 1.02 | 0.50 | 1.00 | 1.03 | 4.16 |
| $\bar{r}$ | 2.05 | 2.09 | 1.22 | 2.00 | 2.09 | 8.37 |
| RMSE | 0.60 | 0.53 | 0.32 | 0.52 | 0.53 | 2.11 |

with

$$E[v(k)v(j)'] = 0.5 \tag{207}$$
$$E[w(k)w(j)'] = 0.1 \tag{208}$$

The condition number for observability and controllability matrices are 362 and 561, respectively; hence it is an ill-conditioned case. With the initial Kalman filter gain, $\Xi$ is

$$\Xi = \begin{bmatrix} 0.28 & 1.37 \\ -0.09 & -0.67 \\ 0.006 & 0.11 \\ 0 & -0.006 \end{bmatrix} \tag{209}$$

The rank of $\Xi$ is 2 and we have a total of 2 unknown variables indicating that both $Q$ and $R$ are identifiable (albeit due to the high condition number, not very well relative to the other systems tested). The condition number for $\Xi$ is 36.4. We simulated 200 MC runs with $N = 1000$ observed samples for each run. We set $M = 15$ to be consistent with the setup in [42]. We also set the maximum number of iterations $n_L = 100$, $N_s = 1000$, and the regularization term from (120) is $\lambda_Q = 0.3$. Table 9 shows the estimated parameters with the initial Kalman filter gain obtained by solving the Riccati equation with $R^{(0)} = 0.1$, and $Q^{(0)} = 0.5$. The results are detailed in Table 9, where the true value stays within the 95% PI. Fig.12 shows the scatter plot for the estimates of $R$ and $Q$ of each MC run. The plot is similar to the estimates in [42]. However, the upper bound on $Q$ is less than that of [42] (about 0.2), which does not provide the detailed results presented in Table 9. Fig.13 shows that the Kalman filter is consistent.

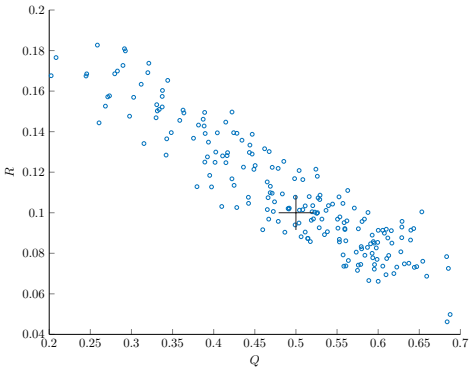



FIGURE 12: $Q$ and $R$ estimation for Case 5.

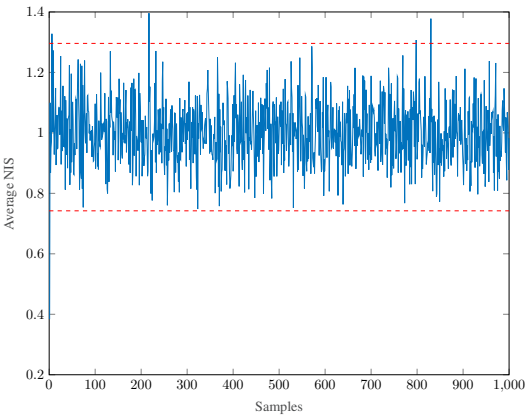



FIGURE 13: Averaged NIS for Case 5.

## XI CONCLUSION AND FUTURE WORK

In this paper we derived necessary and sufficient conditions for the identification of the process and measurement noise covariances. We also provide a novel six-step successive approximation method, coupled with an adaptive gradient method, to estimate the steady-state Kalman filter gain $W$, unknown noise covariance matrices $R$, and $Q$, as well as the state prediction (or updated) covariance matrix $\bar{P}$ (or $P$) when $Q$ and $R$ are identifiable. Moreover, we developed a novel iterative approach to obtain positive definite $Q$, $R$ and $\bar{P}$, while ensuring that the structural assumptions on $Q$ and $R$ are enforced (e.g., diagonality of $Q$ and $R$, if appropriate, symmetry and positive definiteness).

TABLE 9: Monte Carlo Simulation for Case 5 with $M = 100$ and PI$= 2\sigma$ (100 Runs; 1000 Samples)

| | $R$ | $Q$ | $W_1$ | $W_2$ | $W_3$ | $\bar{P}_{11}$ | $\bar{P}_{22}$ | $\bar{P}_{33}$ |
|---|---|---|---|---|---|---|---|---|
| Truth | 0.10 | 0.50 | 1.14 | 2.24 | 3.39 | 0.54 | 2.04 | 4.69 |
| $\underline{r}$ | 0.07 | 0.26 | 0.54 | 1.05 | 1.60 | 0.29 | 1.07 | 2.50 |
| Mean | 0.11 | 0.49 | 1.05 | 2.06 | 3.12 | 0.52 | 1.99 | 4.58 |
| $\bar{r}$ | 0.17 | 0.65 | 1.40 | 2.76 | 4.13 | 0.69 | 2.65 | 6.09 |
| RMSE | 0.03 | 0.11 | 0.27 | 0.54 | 0.80 | 0.11 | 0.45 | 1.00 |

We provided several approaches to estimate the unknown noise covariance $R$ via *post-fit residuals*. We examined previous methods from the literature and heretofore undiscussed assumptions of these methods that result in largely inaccurate or unstable estimates of the unknown parameters. The proposed method significantly outperformed the previous ones, given the same system assumptions.

We validated the proposed method on five different test cases and were able to obtain parameter estimates where the truth stays within the 95% probability interval based on the estimates.

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) exploring vector moving average estimation algorithms using the minimal polynomial approach and/or truncating the effects of state; 3) replacing the batch innovation covariance estimates by their individual or mini-batch estimates, as is done in machine learning; 4) investigating accelerated gradient methods (e.g., Adam [26], AdaGrad [13], RMSProp [48], conjugate gradient, memoryless quasi-Newton, and trust region methods [5]); 5) automatic model selection; and 6) extension to nonlinear dynamic models.

.

## Appendix A STEADY-STATE UPDATED STATE COVARIANCE RICCATI EQUATION

From (6) and (9), we can write the steady-state updated state covariance matrix as

$$\bar{P} = P + WSW' = FPF' + \Gamma Q \Gamma' \tag{210}$$

Thus,

$$P = FPF' - WSW' + \Gamma Q \Gamma' \tag{211}$$

Given (14) and (68), we can rewrite (211) as

$$P = FPF' - PH'R^{-1}SR^{-1}HP' + \Gamma Q \Gamma' \tag{212}$$

$$= FPF' - PH'G^{-1}HP' + \Gamma Q \Gamma' \tag{213}$$

$$= FPF' - PH'(R - HPH')^{-1}HP' + \Gamma Q \Gamma' \tag{214}$$

## Appendix B OBTAIN $W$ USING MINIMAL POLYNOMIAL

Let $W_s$ be the suboptimal Kalman filter gain and $\tilde{e}$ be the difference of the state predictions between the optimal and suboptimal filter, that is,

$$\tilde{e}(k+1|k) = \bar{F}_s \tilde{e}(k|k-1) + F(W - W_s)\nu(k) \tag{215}$$

where $\bar{F}_s$ is defined as

$$\bar{F}_s = F(I_{n_x} - W_s H) \tag{216}$$

We can write the suboptimal innovation $\nu_s(k)$ in terms of $\tilde{e}(k|k-1)$

$$\nu_s(k) = H\tilde{e}(k|k-1) + \nu(k) \tag{217}$$

Then, using the minimal polynomial of $\bar{F}_s$ from (21), $\nu_s(k-i)$ can be written as

$$\begin{aligned}\nu_s(k-i) &= H\Bigg[\bar{F}_s^{m-i}\tilde{e}(k-m|k-m-1) \\ &\quad + \sum_{l=i+1}^{m} \bar{F}_s^{l-i-1}F(W - W_s)\nu(k-l)\Bigg] \\ &\quad + \nu(k-i)\end{aligned} \tag{218}$$

Let us define $\xi(k)$ as

$$\xi(k) = \sum_{i=0}^{m} a_i \nu_s(k-i) \tag{219}$$

$$= \sum_{i=0}^{m} a_i \left\{ \left[ H \sum_{l=i+1}^{m} \bar{F}_s^{l-i-1} F(W - W_s)\nu(k-l) \right] + \nu(k-i) \right\} \tag{220}$$

$$= \sum_{l=0}^{m} \left[ a_l I_{n_z} + H \sum_{i=0}^{l-1} a_i \bar{F}_s^{l-i-1} F(W - W_s) \right] \nu(k-l) \tag{221}$$

$$= \sum_{l=0}^{m} V_l \nu(k-l) \tag{222}$$

where

$$V_l = a_l I_{n_z} + H \sum_{i=0}^{l-1} a_i \bar{F}_s^{l-i-1} F(W - W_s) \tag{223}$$

From (15), we can write (222) in terms of z-transform, that is

$$\xi(\mathrm{z}) = \sum_{l=0}^{m} V_l \mathrm{z}^{-l} \nu(\mathrm{z}) \tag{224}$$

Note that we can write $\xi(k)$ as a vector auto-regressive process of infinite order (which can be truncated to $\mathscr{M}^{th}$ order), that is,

$$\xi(k) = \sum_{j=1}^{\infty} Y_j \xi(k-j) + \nu(k) \tag{225}$$

The z-transform of (225) is,

$$\xi(\mathrm{z}) = \left[ I_{n_z} - \sum_{j=1}^{\infty} Y_j \mathrm{z}^{-j} \right]^{-1} \nu(k) \tag{226}$$

Also, note the relationship between (226) and (224),

$$\left( I_{n_z} - \sum_{j=1}^{\infty} Y_j \mathrm{z}^{-j} \right) \sum_{l=0}^{m} V_l \mathrm{z}^{-l} = I_{n_z} \tag{227}$$

By equating coefficients, we have

$$Y_j = V_j - \sum_{l=1}^{j-1} Y_{j-l} V_l \qquad j = 0, 1, 2, \ldots, m \tag{228}$$

$$= -\sum_{l=1}^{m} Y_{j-l} V_l \qquad j = m+1 \tag{229}$$

We can truncate the infinite vector auto-regressive model at $\mathscr{M} \gg m$, for $i = 1, 2, \ldots, \mathscr{M}$,

$$E[\xi(k)\xi(k-i)'] = E\left\{ \sum_{j=1}^{\mathscr{M}} Y_j \xi(k-j)\xi(k-i)' + \nu(k)\xi(k-i)' \right\} \tag{230}$$

Then, we obtain the estimates of $\{Y_i\}_{i=1}^{\mathscr{M}}$ by solving

$$\sum_{j=1}^{i} Y_j L_{i-j} + \sum_{j=i+1}^{m+i} Y_j L'_{j-i} = L_i \tag{231}$$
$$i = 1, 2, \ldots, m$$

$$\sum_{j=i-m}^{i} Y_j L_{m-j+1} + \sum_{j=i+1}^{m+i} Y_j L'_{j-i} = 0 \tag{232}$$
$$i = m+1, m+2, \ldots, \mathscr{M}$$

Let $\hat{\nu}(k)$ be

$$\hat{\nu}(k) = \xi(k) - \sum_{j=1}^{\mathscr{M}} Y_j \xi(k-j) \tag{233}$$

and recall (223) and note that

$$\hat{C}_l = H \sum_{i=0}^{l-1} a_i \bar{F}_s^{l-i-1} F \tag{234}$$

where $\{\hat{C}_l\}$ are the sample covariance matrices. Then,

$$\mathrm{vec}\left[ \xi(k) - \sum_{l=0}^{m} a_l \hat{\nu}(k-l) + \sum_{l=1}^{m} \hat{C}_l F W_s \hat{\nu}(k-l) \right] \tag{235}$$

$$= \left[ \sum_{l=0}^{m} \hat{\nu}(k-l)' \otimes \hat{C}_l \right] \mathrm{vec}(W) \tag{236}$$

Alternately, $W$ can be computed via $\{V_l\}$. To compute $V_l$, we know that

$$V_0 = I_{n_z} \tag{237}$$

$$V_l = \sum_{i=0}^{l-1} V_i Y_{l-i} \qquad l = 1, 2, \ldots, m \tag{238}$$

Recalling (223), we have the following relationship,

$$\tilde{V}_l = V_l - a_l I_{n_z} + \hat{C}_l W_s = \hat{C}_l W \quad l = 1, 2, \ldots, m \tag{239}$$

Then,

$$\text{vec}\left(\begin{bmatrix}\tilde{V}_1\\ \tilde{V}_2\\ \vdots\\ \tilde{V}_m\end{bmatrix}\right) = \begin{bmatrix}I_{n_z}\otimes\hat{C}_1\\ I_{n_z}\otimes\hat{C}_2\\ \vdots\\ I_{n_z}\otimes\hat{C}_m\end{bmatrix} \text{vec}\,(W) \tag{240}$$

$$= \hat{C}_a \text{vec}\,(W) \tag{241}$$

where the $\text{vec}(\cdot)$ function converts $W$ into a column vector as in (36) and $\otimes$ is the Kronecker product. We can obtain the optimal Kalman filter gain $W$ by solving the least squares problem, where a unique solution exists if $\hat{C}_a$ has full column rank.

## Appendix C CHOLESKY DECOMPOSITION AND EIGEN DECOMPOSITION

To solve for $R$ using **R3**, we first perform Cholesky decomposition of $S^{-1}$. That is,

$$S^{-1} = LL' \tag{242}$$

Then,

$$L'RS^{-1}RL = (L'RL)^2 = L'GL \tag{243}$$

Let us perform eigen decomposition on (243), that is

$$L'GL = U\Lambda U' \tag{244}$$

Then, we have

$$L'RL = U\Lambda^{1/2}U' \tag{245}$$

and $R$ can be computed as

$$R = (L')^{-1}U\Lambda^{1/2}U'L^{-1} \tag{246}$$

## Appendix D SIMULTANEOUS DIAGONALIZATION

To solve for $R$ using **R3**, we first perform eigen decomposition on $S^{-1}$. That is,

$$S^{-1} = U_1\Lambda_1 U_1' \tag{247}$$

$$= (U_1\Lambda_1^{1/2}U_1')^2 \tag{248}$$

Noting that

$$S^{-1/2}GS^{-1/2} = \left(S^{-1/2}RS^{-1/2}\right)^2 \tag{249}$$

we perform another eigen decomposition on $U_1\Lambda_1^{1/2}U_1'GU_1\Lambda_1^{1/2}U_1'$ to get

$$U_1\Lambda_1^{1/2}U_1'GU_1\Lambda_1^{1/2}U_1' = (U_2\Lambda_2^{1/2}U_2')^2 \tag{250}$$

$$= (U_1\Lambda_1^{1/2}U_1'RU_1\Lambda_1^{1/2}U_1')^2 \tag{251}$$

and $R$ can be computed as

$$R = U_1\Lambda_1^{-1/2}U_1'U_2\Lambda_2^{1/2}U_2'U_1\Lambda_1^{-1/2}U_1' \tag{252}$$

## Appendix E OBJECTIVE FUNCTION GRADIENT COMPUTATION

Note that $\Theta(i)$, $\Psi$, and $W\hat{C}(0)$ are all functions of $W$ in (48). Thus,

$$\delta J = \frac{1}{2}\text{trace}\left\{\sum_{i=1}^{M-1}\left[\delta\Theta(i)\Omega + \Theta(i)\delta\Omega\right]\right\} \tag{253}$$

where

$$\Omega = \left[\Psi - WC(0)\right]E^2\left[\Psi' - C(0)W'\right] \tag{254}$$

$$\delta\Omega = \sum_{i=1}^{M-1}\left\{\left[\delta\Psi - \delta WC(0)\right]E^2\left[\Psi' - C(0)W'\right] + \left[\Psi - WC(0)\right]E^2\left[\delta\Psi' - C(0)\delta W'\right]\right\} \tag{255}$$

and

$$\sum_{i=1}^{M-1}\delta\Theta(i)\Omega = \sum_{i=1}^{M-1}\Big\{\left[F'\left(I_{n_x} - (W+\delta W)H\right)'F'\right]^{i-1}H'E^2H\left[F\left(I_{n_x} - (W+\delta W)H\right)\right]^{i-1}F - F'(\bar{F}')^{i-1}H'E^2H\bar{F}^{i-1}F\Big\}\Omega \tag{256}$$

To first order, (256) can be approximated by

$$\begin{aligned}\sum_{i=1}^{M-1} &-F'\left[H'\delta W'(\bar{F}')^{i-2}+\bar{F}'H'\delta W'(\bar{F}')^{i-3}+\cdots+(\bar{F}')^{i-2}H'\delta W'\right]H'E^2H\bar{F}^{i-1}F\Omega \\ &-F'(\bar{F}')^{i-1}H'E^2H\left[\delta WH\bar{F}^{i-2}+\bar{F}\delta WH\bar{F}^{i-3}+\cdots+\bar{F}^{i-2}\delta WH\right]F\Omega\end{aligned} \tag{257}$$

Then,

$$\begin{aligned}\sum_{i=1}^{M-1}\delta\Theta(i)\Omega \approx &-\sum_{i=1}^{M-1}\sum_{\ell=0}^{i-2}\left[F'(\bar{F}')^{\ell}H'\delta W'(\bar{F}')^{i-2-\ell}\right]H'E^2H\bar{F}^{i-1}F\Omega \\ &+F'(\bar{F}')^{i-1}H'E^2H\left[\bar{F}^{r}\delta WH\bar{F}^{i-2-\ell}\right]F\Omega\end{aligned} \tag{258}$$

So,

$$\frac{1}{2}\text{trace}\left(\sum_{i=1}^{M-1}\delta\Theta(i)\Omega\right)=-\text{trace}\left[\delta W'\sum_{i=1}^{M-1}\sum_{\ell=0}^{i-2}(\bar{F}')^{i-2-\ell}H'E^2H\bar{F}^{i-1}F\Omega F'(\bar{F}')^{\ell}H'\right] \tag{259}$$

$$=-\text{trace}\left[\delta W'\sum_{i=1}^{M-1}\sum_{\ell=0}^{i-2}(\bar{F}')^{i-2-\ell}H'E^2C(i)E^2C(\ell+1)'\right] \tag{260}$$

For $\sum_{i=1}^{M-1}\Theta(i)\delta\Omega$, we have

$$\begin{aligned}\sum_{i=1}^{M-1}\Theta(i)\delta\Omega=\sum_{i=1}^{M-1}\Theta(i)\big\{&\left[\delta PH'-\delta WC(0)\right]E^2(\Psi'-C(0)W') \\ &-\left[\Psi-WC(0)\right]E^2\left[H\delta P-C(0)\delta W'\right]\big\}\end{aligned} \tag{261}$$

where,

$$\delta P=\bar{F}P\bar{F}'-F\delta W(\Psi'-C(0)W')F'+F(\Psi-WC(0))\delta W'F' \tag{262}$$

$$=-\sum_{b=0}^{\infty}\bar{F}^{b}\left[F\delta W(\Psi'-C(0)W')F'+F(\Psi-WC(0))\delta W'F'\right](\bar{F}')^{b} \tag{263}$$

Then,

$$\begin{aligned}\frac{1}{2}\text{trace}\left(\sum_{i=1}^{M-1}\Theta(i)\delta\Omega\right)=\text{trace}\Bigg\{&-\delta W'\sum_{i=1}^{M-1}\Theta(i)\left[\Psi-WC(0)\right]E^2\hat{C}(0) \\ &+\frac{1}{2}\left[\Theta(i)(\Psi-WC(0))E^2H+H'E^2(\Psi'-C(0)W')\Theta(i)\right]\delta P\Bigg\}\end{aligned} \tag{264}$$

Substituting (263) into (267), we get

$$\begin{aligned}\frac{1}{2}\text{trace}\left(\sum_{i=1}^{M-1}\Theta(i)\delta\Omega\right)=&-\text{trace}\left\{\delta W'\sum_{i=1}^{M-1}\Theta(i)(\Psi-WC(0))E^2C(0)\right\} \\ &-\text{trace}\left\{\left[F\delta W(\Psi'-C(0)W')F'+F(\Psi-WC(0))\delta W'F'\right]Z\right\}\end{aligned} \tag{265}$$

$$Z=\sum_{b=0}^{\infty}(\bar{F}')^{b}\left[\frac{1}{2}\sum_{i=1}^{M-1}\left[\Theta(i)(\Psi-WC(0))E^2H+H'E^2(\Psi'-C(0)W')\Theta(i)\right]\right]\bar{F}^{b} \tag{266}$$

We can solve for $Z$ via a Lyapunov equation as in (55). Then, by substituting $Z$ into (265), we have

$$\frac{1}{2}\text{trace}\left(\sum_{i=1}^{M-1}\Theta(i)\delta\Omega\right)=-\text{trace}\left\{-\delta W'\left[\sum_{i=1}^{M-1}\Theta(i)XE^2C(0)+F'ZFX\right]\right\} \tag{267}$$

where $X$ can be estimated using (57). Then, by subsituting (260) and (267) into (253), we get (54).

## REFERENCES

[1] W. F. Arnold and A. J. Laub, "Generalized Eigenproblem Algorithms and Software for Algebraic Riccati Equations," Proceedings of the IEEE, vol. 72, no. 12, pp. 1746–1754, 1984.

[2] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, Estimation with applications to Tracking and Navigation: Theory Algorithms and Software. John Wiley & Sons, 2001.

[3] R. Battiti, "Accelerated Backpropagation Learning: Two Optimization Methods," Complex Systems, vol. 3, no. 4, pp. 331–342, 1989.

[4] P. R. Bélanger, "Estimation of Noise Covariance Matrices for a Linear Time-Varying Stochastic Process," Automatica, vol. 10, no. 3, pp. 267–275, 1974.

[5] D. P. Bertsekas, "Nonlinear Programming. Athena Scientific Belmont," Massachusets, USA, 1999.

[6] J. I. Canelon, R. S. Provence, N. Mehta, and L. S. Shieh, "An Alternative Kalman Innovation Filter Approach for Receiver Position Estimation Based on GPS Measurements," International Journal of Systems Science, vol. 38, no. 12, pp. 977–990, 2007.

[7] B. Carew and P. Belanger, "Identification of optimum filter steady-state gain for systems with unknown noise covariances," IEEE Transactions on Automatic Control, vol. 18, no. 6, pp. 582–587, 1973.

[8] C.-W. Chen and J.-K. Huang, "Estimation of Steady-State Optimal Filter Gain from Nonoptimal Kalman Filter Residuals," Journal of Dynamic Systems, Measurement, and Control, vol. 116, no. 3, pp. 550–553, 1994.

[9] S. Chen, "Kalman Filter for Robot Vision: a Survey," IEEE Transactions on Industrial Electronics, vol. 59, no. 11, pp. 4409–4420, 2011.

[10] R. Daley, "The Lagged Innovation Covariance: A Performance Diagnostic for Atmospheric Data Assimilation," Monthly Weather Review, vol. 120, no. 1, pp. 178–196, 1992.

[11] A. Das and T. K. Ghoshal, "Market Risk Beta Estimation Using Adaptive Kalman Filter," International Journal of Engineering Science and Technology, vol. 2, no. 6, pp. 1923–1934, 2010.

[12] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," Journal of the Royal Statistical Society: Series B (Methodological), vol. 39, no. 1, pp. 1–22, 1977.

[13] J. Duchi, E. Hazan, and Y. Singer, "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization," Journal of Machine Learning Research, vol. 12, no. Jul, pp. 2121–2159, 2011.

[14] J. Duník, M. Ŝimandl, and O. Straka, "Methods for Estimating State and Measurement Noise Covariance Matrices: Aspects and Comparison," IFAC Proceedings Volumes, vol. 42, no. 10, pp. 372–377, 2009.

[15] J. C. Engwerda, A. M. Ran, and A. L. Rijkeboer, "Necessary and Sufficient Conditions for the Existence of a Positive Definite Solution of the Matrix Equation $X+A^*X^{-1}A=Q$," Linear Algebra and its Applications, vol. 186, pp. 255–275, 1993.

[16] J. W. Galbraith, A. Ullah, and V. Zinde-Walsh, "Estimation of the Vector Moving Average Model by Vector Autoregression," Econometric Reviews, vol. 21, no. 2, pp. 205–219, 2002.

[17] Z. Ghahramani and G. E. Hinton, "Parameter Estimation for Linear Dynamical Systems," Technical Report CRG-TR-96-2, University of Totronto, Dept. of Computer Science, Tech. Rep., 1996.

[18] H. Heffes, "The Effect of Erroneous Models on the Kalman Filter Response," IEEE Transactions on Automatic Control, vol. 11, no. 3, pp. 541–543, 1966.

[19] C. G. Hilborn and D. G. Lainiotis, "Optimal Estimation in the Presence of Unknown Parameters," IEEE Transactions on Systems Science and Cybernetics, vol. 5, no. 1, pp. 38–43, 1969.

[20] P. L. Houtekamer and H. L. Mitchell, "Data Assimilation Using an Ensemble Kalman Filter Technique," Monthly Weather Review, vol. 126, no. 3, pp. 796–811, 1998.

[21] C. Igel and M. Hüsken, "Improving the Rprop Learning Algorithm," in Proceedings of the Second International ICSC Symposium on Neural Computation (NC 2000), vol. 2000. Citeseer, 2000, pp. 115–121.

[22] T. Kailath, "An Innovations Approach to Least-Squares Estimation–Part I: Linear Filtering in Additive White Noise," IEEE Transactions on Automatic Control, vol. 13, no. 6, pp. 646–655, 1968.

[23] R. E. Kalman, "A New Approach To Linear Filtering and Prediction Problems," Journal of Basic Engineering, vol. 82, no. 1, pp. 35–45, March 1960.

[24] C. Kascha, "A Comparison of Estimation Methods for Vector Autoregressive Moving-Average Models," Econometric Reviews, vol. 31, no. 3, pp. 297–324, 2012.

[25] R. Kashyap, "Maximum Likelihood Identification of Stochastic Linear Systems," IEEE Transactions on Automatic Control, vol. 15, no. 1, pp. 25–34, 1970.

[26] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," arXiv preprint arXiv:1412.6980, 2014.

[27] D. Kleinman, "On an Iterative Technique for Riccati Equation Computations," IEEE Transactions on Automatic Control, vol. 13, no. 1, pp. 114–115, 1968.

[28] ——, "Stabilizing a Discrete, Constant, Linear System with Application to Iterative Methods for Solving the Riccati Equation," IEEE Transactions on Automatic Control, vol. 19, no. 3, pp. 252–254, 1974.

[29] A. Lapedes and R. Farber, "A self-optimizing, nonsymmetrical neural net for content addressable memory and pattern recognition," Physica D: Nonlinear Phenomena, vol. 22, no. 1-3, pp. 247–259, 1986.

[30] Y. Lim, "Factorizations and geometric means of positive definite matrices," Linear Algebra and its Applications, vol. 437, no. 9, pp. 2159–2172, 2012.

[31] H. Lütkepohl and D. S. Poskitt, "Estimating Orthogonal Impulse Responses via Vector Autoregressive Models," Econometric Theory, vol. 7, no. 4, pp. 487–496, 1991.

[32] P. Matisko and V. Havlena, "Noise Covariance Estimation for Kalman Filter Tuning using Bayesian Approach and Monte Carlo," International Journal of Adaptive Control and Signal Processing, vol. 27, no. 11, pp. 957–973, 2013.

[33] R. Mehra, "On The Identification Of Variances and Adaptive Kalman Filtering," IEEE Transactions on Automatic Control, vol. 15, no. 2, pp. 175–184, 1970.

[34] ——, "Approaches to Adaptive Filtering," IEEE Transactions on Automatic Control, vol. 17, no. 5, pp. 693–698, 1972.

[35] R. Ménard and R. Daley, "The Application of Kalman Smoother Theory to the Estimation of 4DVAR Error Statistics," Tellus A, vol. 48, no. 2, pp. 221–237, 1996.

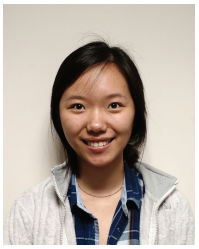
[36] A. Mohamed and K. Schwarz, "Adaptive Kalman Filtering for INS/GPS," Journal of geodesy, vol. 73, no. 4, pp. 193–203, 1999.

[37] J. B. Moore, "Discrete-Time Fixed-Lag Smoothing Algorithms," Automatica, vol. 9, no. 2, pp. 163–173, 1973.

[38] K. Myers and B. Tapley, "Adaptive Sequential Estimation with Unknown Noise Statistics," IEEE Transactions on Automatic Control, vol. 21, no. 4, pp. 520–523, 1976.

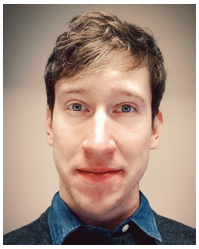
[39] C. Neethling and P. Young, "Comments on: Identification of Optimum Filter Steady-State Gain for Systems with Unknown Noise Covariances," IEEE Transactions on Automatic Control, vol. 19, no. 5, pp. 623–625, 1974.

[40] Y. Nesterov, "A Method for Unconstrained Convex Minimization Problem with the Rate of Convergence $O(1/k^2)$," in Doklady AN USSR, vol. 269, 1983, pp. 543–547.
[41] B. J. Odelson, A. Lutz, and J. B. Rawlings, "The Autocovariance Least-Squares Method for Estimating Covariances: Application to Model-Based Control of Chemical Reactors," IEEE Transactions on Control Systems Technology, vol. 14, no. 3, pp. 532–540, 2006.
[42] 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, 2006.
[43] F. Poloni and G. Sbrana, "Closed-Form Results for Vector Moving Average Models with a Univariate Estimation Approach," Econometrics and Statistics, vol. 10, pp. 27–52, 2019.
[44] R. H. Reichle, W. T. Crow, and C. L. Keppenne, "An Adaptive Ensemble Kalman Filter for Soil Moisture Data Assimilation," Water resources research, vol. 44, no. 3, 2008.
[45] R. H. Shumway and D. S. Stoffer, "An Approach to Time Series Smoothing and Forecasting using the EM Algorithm," Journal of Time Series Analysis, vol. 3, no. 4, pp. 253–264, 1982.
[46] P. L. Smith, "Estimation of the Covariance Parameters of Non-Stationary Time-Discrete Linear Systems," Aerospace Corp., El Segundo CA, Engineering Science Operations, Tech. Rep., 1971.
[47] K. Tajima, "Estimation of Steady-State Kalman Filter Gain," IEEE Transactions on Automatic Control, vol. 23, no. 5, pp. 944–945, 1978.
[48] 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.
[49] C. F. Van Loan and G. H. Golub, Matrix Computations, 4th ed. Johns Hopkins University Press, 2013.
[50] T. P. Vogl, J. Mangis, A. Rigler, W. Zink, and D. Alkon, "Accelerating the Convergence of the Back-Propagation Method," Biological Cybernetics, vol. 59, no. 4-5, pp. 257–263, 1988.
[51] M. D. Zeiler, "ADADELTA: An Adaptive Learning Rate Method," arXiv preprint arXiv:1212.5701, 2012.



LINGYI ZHANG received the B.S. and M.S. degrees in Electrical and Computer Engineering from the University of Connecticut, Storrs in 2014 and 2019, respectively. She is currently pursuing the Ph.D. degree from the Department of Electrical and Computer Engineering at the same university.

She was the co-recipient of the Tammy Blair award for best student paper at FUSION 2016. Her current research interests include modeling dynamic and uncertain environments for asset allocation and path planning, optimization-based techniques for mission planning and coordination, and adaptive Kalman filtering.



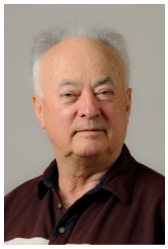
DAVID SIDOTI received the B.S., M.S., and Ph.D. degrees in Electrical and Computer Engineering from the University of Connecticut, Storrs, CT, USA, in 2011, 2016, and 2018, respectively. He is a Computer Scientist under the Meteorological Applications Development Branch at the U.S. Naval Research Laboratory, Marine Meteorology Division (MMD) in Monterey, CA. He is NRLMMD's primary Machine Learning subject matter expert and contributes to a broad portfolio of Machine Learning projects relating to atmospheric and oceanographic forecasting.

He was the co-recipient of the Tammy Blair award for best student paper at FUSION 2016. In 2018, he was recognized and awarded a Distinguished Scholar Jerome and Isabella Karle's Fellowship. His current interests include multi-objective algorithms for dynamic scheduling and resource management in weather-impacted environments, and deep learning applications to numerical weather prediction.

ADAM BIENKOWSKI 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 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 ALPHATECH, Inc., Burlington, MA from 1980 to 1986. He has been with the department of Electrical and Computer Engineering at UConn, where he is currently the Board of Trustees Distinguished Professor and the UTC Chair Professor in Systems Engineering. Dr. Pattipati's research activities are in the areas of proactive decision support, uncertainty quantification, smart manufacturing, autonomy, knowledge representation, and optimization-based learning and inference. A common theme among these applications is that they are characterized by a great deal of uncertainty, complexity, and computational intractability. He is a cofounder of Qualtech Systems, Inc., a firm specializing in advanced integrated diagnostics software tools (TEAMS, TEAMS-RT, TEAMS-RDS, TEAMATE), 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, Vice-President for Technical Activities of the IEEE SMC Society from 1998 to 1999, and as Vice-President for Conferences and Meetings of the IEEE SMC Society from 2000 to 2001. He was co-recipient of the Andrew P. Sage Award for the Best SMC Transactions Paper for 1999, the Barry Carlton Award for the Best AES Transactions Paper for 2000, the 2002 and 2008 NASA Space Act Awards for "A Comprehensive Toolset for Model-based Health Monitoring and Diagnosis," and "Real-time Update of Fault-Test Dependencies of Dynamic Systems: A Comprehensive Toolset for Model-Based Health Monitoring and Diagnostics," and the 2003 AAUP Research Excellence Award at UCONN. He also won the best technical paper awards at the 1985, 1990, 1994, 2002, 2004, 2005 and 2011 IEEE AUTOTEST Conferences, and at the 1997, 2004 Command and Control Conference. He is an elected Fellow of IEEE and of the Connecticut Academy of Science and Engineering.



YAAKOV BAR-SHALOM was born on May 11, 1941. He received the B.S. and M.S. degrees from the Technion, Israel Institute of Technology, in 1963 and 1967 and the Ph.D. degree from Princeton University in 1970, all in electrical engineering. From 1970 to 1976 he was with Systems Control, Inc., Palo Alto, California. Currently he is Board of Trustees Distinguished Professor in the Dept. of Electrical and Computer Engineering and Marianne E. Klewin Professor in Engineering at the University of Connecticut. He is also Director of the ESP (Estimation and Signal Processing) Lab.

His current research interests are in estimation theory, target tracking and data fusion. He has published over 600 papers and book chapters in these areas and in stochastic adaptive control. He coauthored the monograph Tracking and Data Association (Academic Press, 1988), the graduate texts Estimation and Tracking: Principles, Techniques and Software (Artech House, 1993; translated into Russian, MGTU Bauman, Moscow, Russia, 2011), Estimation with Applications to Tracking and Navigation: Algorithms and Software for Information Extraction (Wiley, 2001), the advanced graduate texts MultitargetMultisensor Tracking: Principles and Techniques (YBS Publishing, 1995), Tracking and Data Fusion (YBS Publishing, 2011), and edited the books MultitargetMultisensor Tracking: Applications and Advances (Artech House, Vol. I, 1990; Vol. II, 1992; Vol. III, 2000).

He has been elected Fellow of IEEE for "contributions to the theory of stochastic systems and of multi target tracking". He has been consulting to numerous companies and government agencies, and originated the series of MultitargetMultisensor Tracking short courses offered via UCLA Extension, at Government Laboratories, private companies and overseas.

During 1976 and 1977 he served as Associate Editor of the IEEE Transactions on Automatic Control and from 1978 to 1981 as Associate Editor of Automatica. He was Program Chairman of the 1982 American Control Conference, General Chairman of the 1985 ACC, and CoChairman of the 1989 IEEE International Conference on Control and Applications. During 198387 he served as Chairman of the Conference Activities Board of the IEEE Control Systems Society and during 198789 was a member of the Board of Governors of the IEEE CSS. He was a member of the Board of Directors of the International Society of Information Fusion (1999–2004) and served as General Chairman of FUSION 2000, President of ISIF in 2000 and 2002 and Vice President for Publications in 2004-13.

In 1987 he received the IEEE CSS Distinguished Member Award. Since 1995 he is a Distinguished Lecturer of the IEEE AESS and has given numerous keynote addresses at major national and international conferences. He is corecipient of the M. Barry Carlton Award for the best paper in the IEEE Transactions on Aerospace and Electronic Systems in 1995 and 2000 and recipient of the 1998 University of Connecticut AAUP Excellence Award for Research. In 2002 he received the J. Mignona Data Fusion Award from the DoD JDL Data Fusion Group. He is a member of the Connecticut Academy of Science and Engineering.

In 2008 he was awarded the IEEE Dennis J. Picard Medal for Radar Technologies and Applications, and in 2012 the Connecticut Medal of Technology. He has been listed by academic.research.microsoft (top authors in engineering) as #1 among the researchers in Aerospace Engineering based on the citations of his work. He is the recipient of the 2015 ISIF

Award for a Lifetime of Excellence in Information Fusion. In 2016 The ISIF Award for Lifetime of Excellence in Information Fusion has been renamed as "ISIF Yaakov Bar-Shalom Award for Lifetime of Excellence in Information Fusion" for recipients starting that year.

DAVID L. KLEINMAN received the B.E.E. degree from Cooper Union, New York, NY, USA, in 1962, and the M.S. and Ph.D. degrees in electrical engineering (control systems) from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1963 and 1967, respectively.

He is a part-time Research Professor with the Department of Information Sciences, Naval Postgraduate School, Monterey, CA, USA, and a Professor Emeritus with the Electrical and Computer Engineering Department, University of Connecticut, Storrs, CT, USA, where he directed the CYBERLAB, from 1973 to 1994. From 1967 to 1971, he was with Bolt Beranek and Newman, Cambridge, where he pioneered in the application of modern control and estimation theory to develop and validate an analytical model for describing human control and information processing performance in manned vehicle systems. From 1971 to 1973, he directed the East Coast Office of Systems Control Inc., Cambridge, MA, USA, where he led applied research projects in both manual and automatic control. He was the Co-founder of ALPHATECH Inc., Burlington, MA, USA, Qualtech Systems Inc., Rocky Hill, CT, USA, and Aptima Inc., Woburn, MA, USA. He continues to serve on the Board of Directors of the latter two companies. He has authored or co-authored over 200 journal and conference papers, many of which are in the area of military command and control. His current research interests include team decision making/coordination in dynamic task environments and structural adaptation processes in human organizations.

Dr. Kleinman has been the Vice President of Finance and Long-range Planning for the IEEE-SMC Society. He was the Program Chairman of the 1989 IEEE International Conference on Systems, Man, and Cybernetics and the 1990 JDL Symposium on Command and Control Research.