2016 American Control Conference (ACC)
Boston Marriott Copley Place
July 6-8, 2016. Boston, MA, USA

Kalman-Filter-Based Time-Varying Parameter Estimation via Retrospective Optimization of the Process Noise Covariance

Frantisek M. Sobolic and Dennis S. Bernstein

Abstract— **Retrospective estimation of the process noise covariance is performed by minimizing the cumulative state-estimation error based on the innovations. This technique is applied to parameter estimation problems, where the parameters to be estimated are time-varying and thus do not fit in the classical Kalman filter noise structure. This technique is compared to the standard Kalman filter with a fixed process noise covariance as well as an innovations-based adaptive Kalman filter.**

I. Introduction

The Kalman filter is a stochastically optimal observer that utilizes statistical information about the measurement and process noise. In particular, the measurement and process noise are assumed to be white Gaussian processes with zero mean and covariances R and Q, respectively. Although R is relatively easy to characterize in practice, it is often challenging to estimate Q. Consequently, there is an extensive amount of literature on adaptive Kalman filtering, where the innovations signal is used to construct an estimate of Q [1, 5, 7, 8].

In the present paper, we develop an alternative approach to estimating Q and apply this technique to parameter estimation. First, we show that the process noise covariance Q that minimizes the cumulative innovations does not (in general) correspond to the value of Q that, when used in the Kalman filter, minimizes the cumulative state-estimate error. However, for parameter estimation, we show that under certain assumptions the value of Q that minimizes the cumulative state-estimate error does in fact minimize the cumulative innovations. This insight provides the basis for a novel approach to online estimation of the process noise covariance for parameter estimation problems.

The estimation technique uses a past window of data to rerun the Kalman filter along with an optimization routine that retrospectively determines the value of Q that minimizes the cumulative innovations at the present time. Based on the previous analysis, this estimate is assumed to minimize the state-estimate error. Since the optimization does not lend itself to analytical gradients, we use gradient-free optimization techniques.

This technique is applied to the system identification problem, specifically, estimation of the coefficients of an input-output model. The dynamical system state components for the Kalman filter are thus the coefficients of the input-output model, and the dynamics matrix is the identity. The ultimate objective is to use this technique to identify linear systems whose coefficients are time-varying.

Adaptive Kalman filter techniques have been developed for this sort of problem in [1, 4, 5, 8, 9], where the covariances R and Q are determined in real-time. These methods are also innovations-based and are not derived in the sense of a minimization optimization problem of the innovation itself. Rather, they are based on an averaging inside a moving estimation window. Another method commonly used for covariance identification is the maximum likelihood estimation (MLE). This method assumes that the noise is characterized by a Gaussian model and that the best fit is determined by minimizing the error for the given noise structure [9]. The type of parameter uncertainty investigated in the present paper are only constrained in the structure of Q.

II. The State Estimation Problem

Consider state estimation for the linear system

$$x_{k+1} = A_k x_k + w_k, \tag{1}$$
$$y_k = C_k x_k + v_k, \tag{2}$$

where $x_k \in \mathbb{R}^n$ is the state to be estimated, $w_k \in \mathbb{R}^n$ is the process noise, $y_k \in \mathbb{R}^m$ is the measurement, and $v_k \in \mathbb{R}^m$ is the measurement noise. We assume that the measurement and process noise are stationary and have zero mean with covariances $R = \mathbb{E}[v_k v_k^{\rm T}]$ and $Q = \mathbb{E}[w_k w_k^{\rm T}]$, respectively. Assuming (A_k, C_k) is observable, the Kalman filter is given by

$$\hat{x}_{k+1|k} = A_k \hat{x}_{k|k}, \tag{3}$$
$$= A_k(\hat{x}_{k|k-1} + K_k z_k), \tag{4}$$
$$P_{k+1|k} = A_k P_{k|k} A_k^{\rm T} + Q, \tag{5}$$
$$= A_k[(I - K_k C_k)P_{k|k-1}]A_k^{\rm T} + Q, \tag{6}$$

where the error covariance is defined as

$$P_{k|k-1} \triangleq \mathbb{E}[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})^{\rm T}], \tag{7}$$

and where

$$K_k \triangleq P_{k|k-1} C_k^{\rm T}(C_k P_{k|k-1} C_k^{\rm T} + R)^{-1}, \tag{8}$$
$$z_k \triangleq y_k - C_k \hat{x}_{k|k-1}, \tag{9}$$

are the optimal gain and innovations, respectively [3].

F. M. Sobolic is a graduate student with the Aerospace Engineering Department at the University of Michigan, 1320 Beal Ave., Ann Arbor, MI 48109, USA fsobolic@umich.edu

D. S. Bernstein is a Professor with the Aerospace Engineering Department, University of Michigan, dsbaero@umich.edu

In the present paper, we use the innovations to estimate the process noise covariance. We define the cumulative state-estimate error as

$$e_{x,k+k_0} \triangleq \sum_{i=k_0}^{k+k_0} (x_i - \hat{x}_{i|i-1})^{\mathrm{T}}(x_i - \hat{x}_{i|i-1}), \tag{10}$$

with expected value

$$\begin{aligned}\mathbb{E}[e_{x,k+k_0}] &= \sum_{i=k_0}^{k_0+k} \mathrm{tr}(P_{i|i-1}), \\ &= \sum_{i=k_0}^{k+k_0} \mathrm{tr}(A_i P_{i-1|i-1} A_i^{\mathrm{T}}) + k\ \mathrm{tr}(Q). \end{aligned} \tag{11}$$

Next, we define the cumulative innovations as

$$e_{z,k+k_0} \triangleq \sum_{i=k_0}^{k+k_0} z_i^{\mathrm{T}} z_i, \tag{12}$$

with expected value

$$\begin{aligned}\mathbb{E}[e_{z,k+k_0}] &= \sum_{i=k_0}^{k+k_0} \mathbb{E}[(y_i - C_i\hat{x}_{i|i-1})^{\mathrm{T}}(y_i - C_i\hat{x}_{i|i-1})], \\ &= \sum_{i=k_0}^{k+k_0} \mathrm{tr}(C_i(A_i P_{i-1|i-1} A_i^{\mathrm{T}} + Q)C_i^{\mathrm{T}}) + k\mathrm{tr}(R). \end{aligned} \tag{13}$$

Note that, since the state error at step i is independent of the measurement noise at step i, it follows that $\mathbb{E}[(x_i - \hat{x}_{i|i-1})^{\mathrm{T}} C_i^{\mathrm{T}} v_i] = \mathbb{E}[(x_i - \hat{x}_{i|i-1})^{\mathrm{T}}] C_i^{\mathrm{T}} \mathbb{E}[v_i] = 0$. Also, note that (11) and (13) both depend on Q due to (5). Furthermore, (11) and (13) are closely related since both involve $P_{i|i-1}$. However, unless $C_i = I_n$, there is no guarantee that minimizing (10) implies that (12) is minimized and vice versa.

III. The Parameter Estimation Problem

In this section, we specialize the state estimation problem to a problem of parameter estimation, which is subsequently applied to system identification. We show that, under some assumptions, the value of Q that minimizes the cumulative state-estimate error also minimizes the cumulative innovations for the parameter estimation problem. In particular, consider the parameter estimation problem

$$\theta_{k+1} = \theta_k + w_k, \tag{14}$$
$$y_k = \phi_k^{\mathrm{T}} \theta_k + v_k, \tag{15}$$

where $\theta_k \in \mathbb{R}^n$ are the parameters to be estimated, $y_k \in \mathbb{R}^m$ are the measurements, $\phi_k \in \mathbb{R}^{n\times m}$ is the regression matrix, $v_k \in \mathbb{R}^m$ is the measurement noise, and $w_k \in \mathbb{R}^n$ is the process noise. Note that (14), (15) is a special case of (1), (2) with $A_k = I_n$ and $C_k = \phi_k^{\mathrm{T}}$. Furthermore, if the parameters to be estimated are constant, that is, $w_k = 0$, and the measurement noise $v_k \sim \mathcal{N}(0, 1)$, then the parameter estimation problem becomes the standard least-squares problem [4]. For the case where the parameters are time-varying, that is $w_k \neq 0$, we write (3) – (9) recursively as

$$\hat{\theta}_{k+1|k} = \hat{\theta}_{k|k-1} + K_k z_k, \tag{16}$$
$$P_{k+1|k} = (I - K_k \phi_k^{\mathrm{T}}) P_{k|k-1} + Q, \tag{17}$$

where

$$K_k = P_{k|k-1}\phi_k(\phi_k^{\mathrm{T}} P_{k|k-1}\phi_k + R)^{-1}, \tag{18}$$
$$z_k = y_k - \phi_k^{\mathrm{T}} \hat{\theta}_{k|k-1}. \tag{19}$$

Since $A_k = I_n$, the expected value of the cumulative state-estimate error (11) and the expected value of the cumulative innovations (13) becomes

$$\mathbb{E}[e_{\theta,k+k_0}] = \sum_{i=k_0}^{k+k_0} \mathrm{tr}(P_{i-1|i-1}) + k\ \mathrm{tr}(Q), \tag{20}$$

and

$$\mathbb{E}[e_{z,k+k_0}] = \sum_{i=k_0}^{k+k_0} \mathrm{tr}(\phi_i^{\mathrm{T}}(P_{i-1|i-1} + Q)\phi_i) + k\ \mathrm{tr}(R), \tag{21}$$

respectively. Within the context of parameter estimation, the notion of observability depends on persistent excitation [4]. To show that the same value of Q minimizes (20) and (21), we introduce the following assumptions.

Assumption 1. The measurement noise covariance R is known.

Assumption 2. There exists a positive integer N such that the regressor matrix ϕ_k satisfies

$$\beta_l I \leq \sum_{i=n}^{n+N} \phi_i \phi_i^{\mathrm{T}} \leq \beta_u I, \tag{22}$$

where β_l and β_u are positive constants [4].

Assumption 3. The process noise w_k is uncorrelated at each time step k.

In practice, Assumption 1 can be met by performing a sensor characterization. Assumption 2 requires that the system whose parameters are to be identified is persistently excited and hence (I_n, ϕ_k) is observable. Assumption 3 states that the process noise is uncorrelated at each time step, thus making Q diagonal.

Lemma 1. Consider the parameter estimation problem (14), (15) that satisfies Assumptions 1-3. Then, the value of Q that minimizes the expected value of the cumulative state-estimate error (20) also minimizes the expected value of the cumulative innovations (21).

Proof. For all $k \geq N$, the expected value of the cumulative innovations (21) is written as

$$\begin{aligned}\mathbb{E}[e_{z,k_0+k}] &= \sum_{i=k_0}^{k+k_0} \mathrm{tr}((P_{i-1|i-1} + Q)\phi_i\phi_i^{\mathrm{T}}) + k\ \mathrm{tr}(R),\\ &= \sum_{i=k_0}^{k+k_0} \mathrm{tr}(P_{i-1|i-1}\phi_i\phi_i^{\mathrm{T}}) + \mathrm{tr}(Q\phi_i\phi_i^{\mathrm{T}}) + k\ \mathrm{tr}(R),\\ &\geq \sum_{i=k_0}^{k+k_0} \mathrm{tr}(P_{i-1|i-1}\phi_i\phi_i^{\mathrm{T}}) + k\beta_l \mathrm{tr}(Q) + k\ \mathrm{tr}(R),\end{aligned} \tag{23}$$

which is a linear combination of the expected value of the cumulative state estimate (20) and the innovations matrix $\phi_i\phi_i^{\mathrm{T}}$. Since the sum of N innovation matrices are full rank, it follows that the value of Q that minimizes (20) also minimizes (23).

In the next example, we examine the residuals of the state and innovations for a parameter estimation problem.

Example 3.1: Consider the ARMAX model described in (14), (15) as

$$\theta_{k+1} = \theta_k + w_k, \tag{24}$$

$$y_k = \begin{bmatrix} u_{k-1} & u_{k-2} & y_{k-1} & y_{k-2} \end{bmatrix} \theta_k + v_k, \tag{25}$$

where $\theta_k \in \mathbb{R}^4$ with initial condition $\theta_0 = [0\ 0\ 0\ 0]^{\mathrm{T}}$. The input u_k is a zero-mean Gaussian with standard deviation of 1. The true process noise w_k has zero mean with covariance $Q = 2.5 \times 10^{-5} I_4$. We assume that $v_k \sim \mathcal{N}(0, R)$, where $R = 0.01$ is known. Fig. 1 shows the cumulative state-estimate error and the cumulative innovations as functions of α, where $Q = \alpha I_4$. The simulation is run for 6000 time steps for varying values of α. The minimizing values of α for the cumulative state-estimate error and cumulative innovations are 2.70×10^{-5} and 2.37×10^{-5}, respectively. Note that since the parameter estimation problem (25) is a statistical process, the estimate of Q becomes more accurate as the amount of data increases and the value of α will converge to the true value 2.5×10^{-5}. ∎

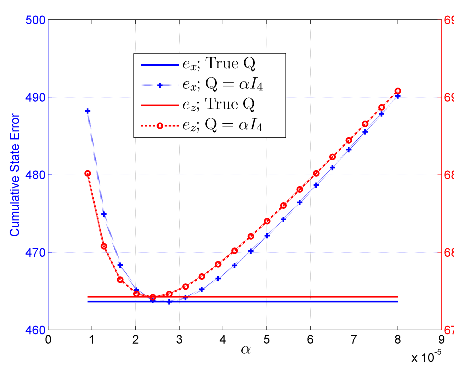


Fig. 1: Results for Example 3.1. The cumulative state-estimate error and cumulative innovations as a function of the entry α of Q for Example 3.1. Each simulation is run for 6000 time steps for varying values of α. The minimizing value of α for the cumulative state-estimation error and the cumulative innovations correspond to the true value $Q = 2.5 \times 10^{-5} I$. The cumulative state-estimation error achieves its minimum at $\alpha = 2.70 \times 10^{-5}$, while the cumulative innovations achieves its minimum at $\alpha = 2.37 \times 10^{-5}$.

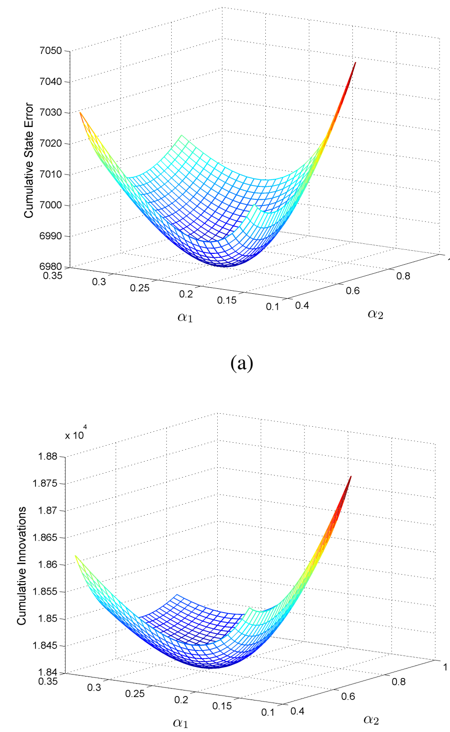


Fig. 2: Results for Example 3.2. The cumulative state-estimate error and cumulative innovations as a function of the entries α_1 and α_2 of Q for Example 3.2. Each simulation is run for 5000 time steps for varying values and combinations of α. (a) shows the cumulative state-estimate error as a function of α, with the minimizer $\alpha = [0.225\ 0.578]$. (b) shows the cumulative innovations, with the minimizer $\alpha = [0.25\ 0.62]$.

Example 3.2: Consider the finite impulse response (FIR) filter described in (14), (15) as

$$\theta_{k+1} = \theta_k + w_k, \tag{26}$$

$$y_k = \begin{bmatrix} u_{k-1} & u_{k-2} \end{bmatrix} \theta_k + v_k, \tag{27}$$

where $\theta_k \in \mathbb{R}^2$ with parameter initial condition $\theta_0 = [0\ 0]^{\mathrm{T}}$. The system input u_k is a white noise sequence with zero mean and standard deviation of 1. The true process noise w_k has zero mean with the covariance matrix $Q = \mathrm{diag}([0.25\ 0.64])$. We assume that the statistical properties of the measurement are known and that $v_k \sim \mathcal{N}(0, R)$, with $R = 1$. Fig. 2 shows the cumulative state-estimate error and cumulative innovations as a function of $\alpha = [\alpha_1\ \alpha_2]$, where $Q = \mathrm{diag}(\alpha)$. The simulation is run for 5000 time steps for varying values and combinations of α. The minimizing values of α for the cumulative state-estimate error and the cumulative innovations is $\alpha = [0.225\ 0.578]$ and $\alpha = [0.25\ 0.62]$, respectively. Note that since the parameter estimation problem (27) is a statistical process, the estimate of Q becomes more accurate as the amount of data increases and will converge to the true value of $[0.25\ 0.64]$. ∎

IV. INNOVATIONS-BASED SLIDING WINDOW-Q OPTIMIZATION

The innovations-based sliding window-Q optimization (ISW-QO) is a methid that minimizes the cumulative innovations by using a retrospective optimization to update the process noise covariance at each time step. Note that a version of this algorithm has been mentioned in [2] for state estimation but no formal proof of convergence is given nor is it used for online estimation. Thus, the problem is to minimize

$$J(Q) = \sum_{j=j_0}^{k} z_j^{\mathrm{T}} z_j, \tag{28}$$

$$\text{s.t. } Q \succeq 0, \tag{29}$$

where $j_0 = k - N + 1$, and N is the window size determined by the user that guarantees Assumption 2. Fig 3 illustrates the method via a flow diagram. The Kalman filter is first initialized and is run for $N-1$ time steps. At time step $k = k+1$, the initial estimates $\hat{\theta}$, error covariance P, input u, and measurement y for that interval is sent to the optimizer. The current value of Q_k is used as an initial guess in the optimizer. In Fig 3, the data $y_{k-N+1:k} = [y_{k-N+1} \ \dots \ y_k]$ and $u_{k-N+1:k} = [u_{k-N+1} \ \dots \ u_k]$. Once the optimal Q is found, the optimization routine passes the optimized parameter estimates, error covariance, and Q to the Kalman filter to process the next step. The index k is incremented and the process loops again.

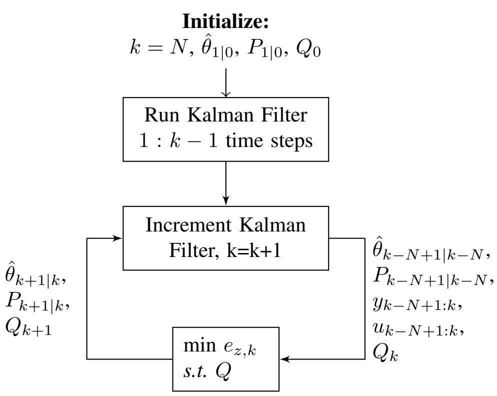


Fig. 3: Block diagram illustrating the innovation-based sliding window Q optimization (ISW-QO) iteration process.

For the following examples, we use the gradient-free optimization method *fminsearch* in MATLAB to optimize Q based on the cumulative innovations. This method is a multidimensional unconstrained nonlinear minimization routine [6]. Since this method is unconstrained, we must impose the positive semidefiniteness of Q. First, Q is enforced to have a diagonal structure, e.g. $Q = \mathrm{diag}(\alpha)$, where $\alpha = [\alpha_1 \ \dots \ \alpha_n]$. Second, at every iteration of the algorithm, the diagonal entries of Q are squared to ensure positive semidefiniteness.

Example 4.1: We again consider the FIR filter presented in Example 3.2. To test the ISW-QO method, we allow the optimizer to calculate Q based on the cumulative innovations

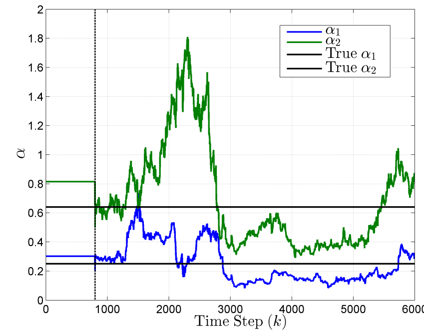
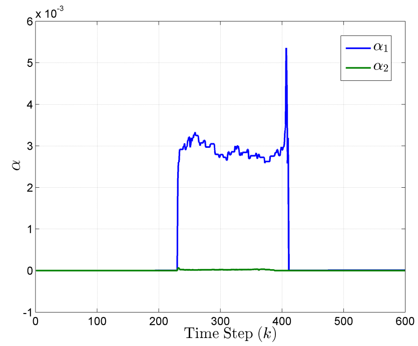


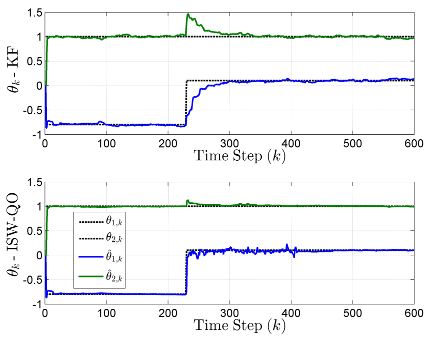
Fig. 4: Results for Example 4.1. The optimized values of α and the parameters θ_k and $\hat{\theta}_k$ for Example 4.1. The Q optimizer is run with a window size of 800 time steps after the initial 800 time steps. In this example, the true value of Q is held constant throughout the entire simulation. The true α and the optimized value of α are shown.

with a window size of 800 time steps, running the simulation for a total of 6000 time steps. Fig. 4 shows the optimal α for each time step and a comparison between the true and the estimated parameters. Note that in this example, α is initialized at $[0.3 \ 0.81]$ and is used in the optimizer until $k = 800$. After $k = 800$, the optimizer updates α at each time step. The true value of α is shown as reference. Note that the window used for optimization is only 800 time steps, whereas in Example 3.1, the optimization uses the entire 6000 time steps. Although the optimized values of α do not converge to the true values, the cumulative state-estimation error is 1.1×10^4 whereas, if the true value of Q were used, the cumulative state-estimation error is 7.0×10^3 after the 6000 time steps. ■

Example 4.2: In this example, we compare a static Q Kalman filter to the ISW-QO method to the same FIR filter as in Example 3.2, except with $R = 0.01$. The parameters are initialized at $\theta_0 = [-0.8 \ 1.0]^{\mathrm{T}}$ and there is no process noise added to the system, that is $w_k = 0$. The parameter θ_1 induces a sudden step change at $k = 250$ from -0.8 to 0.1, while θ_2 remains constant at 1. Note that unless the exact time of the parameter step change is known, their is no way to create an accurate noise model of the parameter variation. Thus, implementing $Q = 0_{2\times 2}$ is the best option and leads to recursive least squares (RLS), which has slow convergence properties. In this example, the Kalman filter is initialized with the same initial conditions as the ISW-QO method except that the process noise covariance is constant, that is $Q = 1 \times 10^{-4} I_2$, with an anticipation that a parameter might change. The ISW-QO is initialized at $Q = 0_{2\times 2}$ with a sliding window of 180 time steps. Fig. 5 shows how the parameters α evolve over the length of the simulation as well as how the parameter estimates compare to the actual values. Notice the correlation between the the parameter that changes θ_1 and the covariance value that changes α_1. ISW-QO correctly identifies the parameter in Q, namely α_1, that aided in reducing the innovations and since the sliding window is of size $N = 180$ time steps, the value of α_1 remains active until the step outside the window. Also notice that with a small disturbance in $\hat{\theta}_2$, $\hat{\theta}_1$ converges faster to the

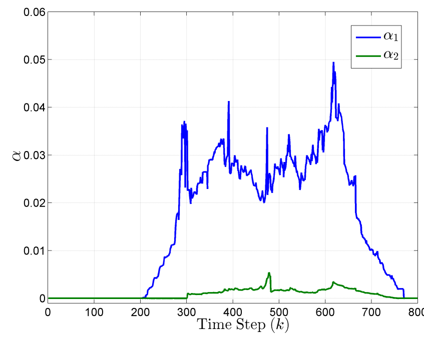


(a)

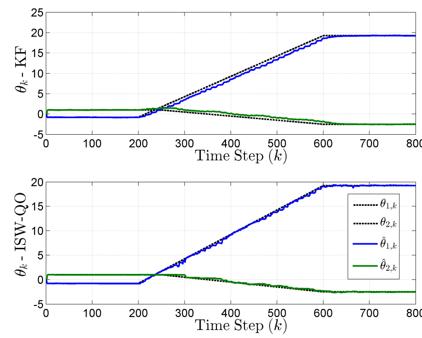


(b)

Fig. 5: Results for Example 4.2. The optimized values of α and the parameters θ_k and $\hat{\theta}_k$ for Example 4.2. (a) shows the optimized Q parameters with α_1 increased due to the step change in the parameter θ_1. (b) compares the Kalman filter to ISW-QO estimates. Notice the convergence of $\hat{\theta}_1$ and the minimal disruption to $\hat{\theta}_2$ in ISW-QO compared to the Kalman Filter with constant $Q = 1 \times 10^{-4} I_2$.



(a)



(b)

Fig. 6: Results for Example 4.3. The optimized values of α and the estimated and true parameters $\hat{\theta}_k$ and θ_k, respectively for Example 4.3. (a) shows the evolution of α_1 and α_2 due to a ramp increase in both the θ parameters, with the magnitude of α_1 being greater due to the more aggressive ramp. (b) compares the Kalman filter to ISW-QO estimates. The Kalman filter looks like a delayed version of the true parameters, unlike the estimated parameters of ISW-QO, which tracks the ramp.

true parameter than the Kalman filter with an added small process noise covariance. ■

Example 4.3: In this example, we revisit Example 4.2 but allow the parameters θ_1 and θ_2 to ramp. At $k = 200$, the parameters ramp with a slope of 0.05 and -0.01, respectively and remain constant after $k = 600$ at $\theta_1 = 19.25$ and $\theta_2 = -3.01$. Fig. 6 shows the how the parameters in Q vary during and after the course of the ramp as well as how the estimated parameters track the true values. Notice that the Kalman filter estimation looks like a delayed version of the true parameters, unlike the estimated parameters of the ISW-QO method, which tracks the ramp. ■

V. Innovations-based Adaptive Kalman filter

The innovations-based adaptive Kalman filter (IAKF) adapts the covariance matrices as the measurements evolve with time [1, 5, 7, 8]. In this method, the estimate of Q is

$$\hat{Q}_k = \frac{1}{N} \sum_{j=j_0}^{k} \Delta x_j \Delta x_j^{\mathrm{T}} + P_{k|k} - A_{k-1} P_{k-1|k-1} A_{k-1}^{\mathrm{T}}, \tag{30}$$

where Δx_k is a first order, steady-state term [8] defined as

$$\Delta x_k \triangleq K_k z_k, \tag{31}$$

and $j_0 = k - N + 1$ is the first sample within the estimation of moving window size N. Note that in the parameter estimation problem, $A_k = I_n$, and (30) reduces to

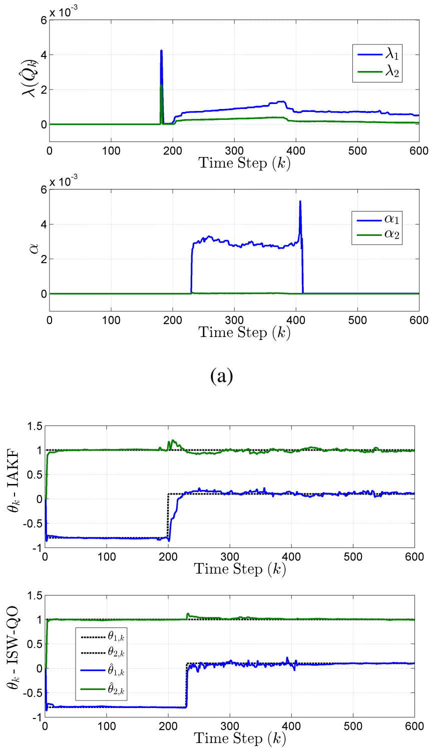
$$\hat{Q}_k = \frac{1}{N} \sum_{j=j_0}^{k} \Delta x_j \Delta x_j^{\mathrm{T}} + P_{k|k} - P_{k-1|k-1}. \tag{32}$$

In this method, it is also assumed that the measurement covariance R is known but the structure of Q is not forced to be diagonal.

VI. ISW-QO Compared to IAKF

We compare ISW-QO and IAKF by revisiting Examples 4.2 and 4.3. In these examples, both ISW-QO and IAKF have the same moving window size of 180 time steps. Note that unlike ISW-QO, where the user has the ability to define the structure of Q to be optimized, IAFK does not have this feature. In order to compare how the estimate $\hat{Q}$ evolves over the length of the simulation, we chose to compare the eigenvalues $\lambda(\hat{Q}_k)$ of $\hat{Q}_k$, to the values of α_k in ISW-QO where $Q_k = \mathrm{diag}(\alpha_{1,k}, \alpha_{2,k})$. This comparison is used to identify which parameters IAKF views as having the largest uncertainty.

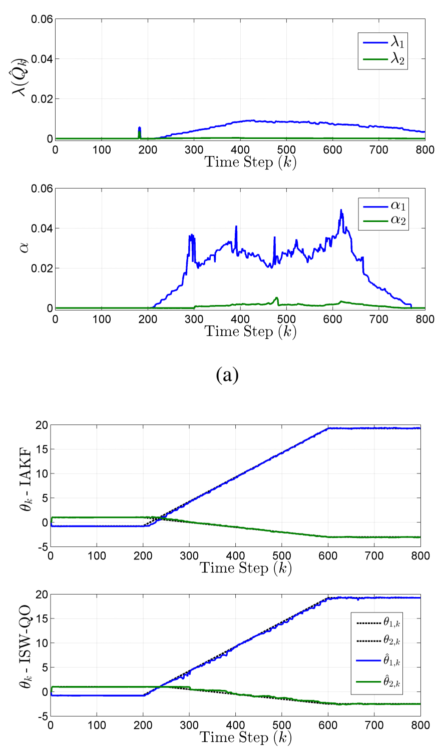
Example 6.1: In this example, we compare ISW-QO to IAKF for a sudden step change in θ_1. Fig. 7 compares the evolution of Q along with the parameter estimates. Notice that the two methods have different eigenvalue profiles but both identify which value needs to be greater in order to



(a)

(b)

Fig. 7: Results for Example 6.1. A comparison of ISW-QO to IAKF for a sudden step change in θ_1. (a) shows how the eigenvalue profiles are different but share in common which value is greater. (b) shows the parameter estimates of the two methods compared to the true values. ISW-QO has a faster response but IAKF performs better compared to the Kalman filter in Fig. 5. The cumulative state-estimation errors for the full 600 time steps for IAKF and ISW–QO is 33.2 and 19.2, respectively.



(a)

(b)

Fig. 8: Results for Example 6.2. A comparison ISW-QO and IAKF for a ramp change in both of the parameters θ_1 and θ_2. (a) shows how the eigenvalue profiles are different but both identify which value is needed to be greater to minimize the innovations. (b) shows the parameter estimates of the two methods compared to the true values. The cumulative state-estimation errors for the full 800 time steps for IAKF and ISW-QO are 104.8 and 156.7, respectively.

minimize the innovations. The parameter estimates for IAKF take longer to converge to the true parameters than ISW-QO but has a faster response then the Kalman filter in Fig. 5. The cumulative state-estimation errors for the full 600 time steps for IAKF and ISW-QO are 33.2 and 19.2, respectively. ■

Example 6.2: In this example, we compare ISW-QO to IAKF for a ramp change in both θ_1 and θ_2. Fig. 8 shows the evolution of Q as well as the parameter estimates. Although the methods produce different eigenvalue profiles, they both identify which value needs to be greater in order to minimize the innovations. The parameter estimates for IAKF are smoother than those of ISW-QO and the cumulative state-estimation errors for the full 800 time steps for IAKF and ISW-QO are 104.8 and 156.7, respectively. ■

VII. CONCLUSIONS

In this present paper, we proposed a technique that minimizes the innovations based on retrospective optimization of the process noise covariance. We showed that minimizing the cumulative innovations is equivalent to minimizing the cumulative state-estimation error for the parameter estimation problem under certain assumptions. This technique is applied to system identification problems where the parameters to be estimated can be time-varying and thus have an unknown Q value. We compared our technique to the standard Kalman filter and IAKF. Results show that when applied to time-varying parameter estimation problem, this technique performs as well if not better than IAKF.

VIII. ACKNOWLEDGMENTS

The authors thank Dr. R. Fuentes, M. Unger, and anonymous reviewers.

REFERENCES

[1] A. Almagbile, J. Wang, and W. Ding. Evaluating the performances of adaptive kalman filter methods in gps/ins integration. *J. of G.P.S.*, 9(1):33–40, 2010.

[2] P. Axelsson, U. Orguner, F. Gustafsson, and M. Norrlöf. Ml estimation of process noise variance in dynamic systems. 2010.

[3] R. G. Brown and P. Y.C. Hwang. *Introduction to random signals and applied Kalman filtering: with MATLAB exercises and solutions*. John Wiley and Sons, New York, 3rd edition, 1997.

[4] L. Guo. Estimating time-varying parameters by the kalman filter based algorithm: stability and convergence. *IEEE Trans. on Autom. Contr.*, 35(2):141–147, 1990.

[5] C. Hide, T. Moore, and M. Smith. Adaptive kalman filtering for low-cost ins/gps. *The J. of Nav.*, 56(01):143–152, 2003.

[6] MATLAB. *version 7.9.0.529 (R2009b)*. The MathWorks Inc., Natick, Massachusetts, 2009.

[7] R. K. Mehra. Approaches to adaptive filtering. *IEEE Trans. on Autom. Contr.*, 17(5):693–698, 1972.

[8] A.H. Mohamed and K.P. Schwarz. Adaptive kalman filtering for ins/gps. *J. of Geo.*, 73(4):193–203, 1999.

[9] Megan A Zagrobelny and James B Rawlings. Identification of disturbance covariances using maximum likelihood estimation. Technical report, No. 2014-02, 2014.