

# On the Accuracy of State Estimators for Constant and Time-Varying Parameter Estimation

Yousaf Rahman\*, Jiapeng Zhong†, Alexey Morozov‡, and Dennis S. Bernstein§

*University of Michigan, 1320 Beal Ave., Ann Arbor, MI 48109*

Nonlinear estimation techniques are often used to estimate constant and time-varying parameters. The purpose of this paper is to use illustrative examples to compare the accuracy of several estimation techniques (the extended Kalman filter, the unscented Kalman filter, and the ensemble adjustment Kalman filter) along with retrospective cost model refinement. Both constant and time-varying examples are considered. Each algorithm is tuned to illustrate its capabilities for the given examples.

## I. Introduction

In many modeling and control applications, the structure of the model is known, but the parameters may be uncertain. Within the context of system identification, models of this type are called *white box models*. In contrast, models whose structure is either partially or fully unknown are called *grey-box* and *black-box models*, respectively.

Parameter-estimation is related to, but distinct from, state estimation, where states evolve due to external inputs and their interaction with other states. In contrast, an unknown parameter may either be constant or time-varying in a pre-specified manner that is independent of initial conditions and outputs. Although a constant or time-varying parameter is not technically a state, it can be modeled as a state by assigning it fictitious dynamics and stochastic forcing. In continuous time, these dynamics are  $\dot{x} = w$ , whereas, in discrete time, these dynamics are  $x_{k+1} = x_k + w_k$ , where  $w$  is the external forcing. For a system with linear dynamics, the resulting state estimation problem is nonlinear due to the multiplication between “real” and “fictitious” states.

State-estimation techniques are widely used for parameter estimation [1–3]. Among the earliest works is the classic paper [4], which analyzes the accuracy of the extended Kalman filter within the context of linear dynamics. Convergence analysis of the extended Kalman filter is provided in [5].

Beyond the extended Kalman filter, nonlinear estimation techniques have been developed based on a wide variety of techniques, including stochastic ensembles [6–8], deterministic ensembles [9,10], Gaussian mixtures [11], density estimators [12], Fokker-Planck solutions [13], moving horizon techniques [14], and adaptive estimators [15,16]. Each of these techniques can potentially be applied to parameter estimation.

\*Graduate Student, Aerospace Engineering Department, University of Michigan, Ann Arbor

†Graduate Student, Control Science and Engineering Department, Harbin Institute of Technology, Harbin, China

‡Graduate Student, Aerospace Engineering Department, University of Michigan, Ann Arbor

§Professor, Aerospace Engineering Department, University of Michigan, Ann Arbor

With this plethora of techniques available, it is of interest to evaluate the relative accuracy of these methods for parameter estimation. The goal of this paper is thus to apply several nonlinear estimation techniques to two examples, namely, a mass-spring structure and linearized aircraft dynamics. These low-order examples are chosen to provide a transparent setting for numerical studies. The methods we consider are the extended Kalman filter [3–5], the unscented Kalman filter [2], the ensemble adjustment Kalman filter [6], and retrospective cost model refinement. We consider both constant and time-varying parameters as well as the effects of noise.

## II. Problem Statement

Consider the multi-input, multi-output discrete-time system

$$x(k+1) = A(\kappa(k))x(k) + Bw(k), \quad (1)$$

$$y(k) = Cx(k) + v(k), \quad (2)$$

where  $x(k) \in \mathbb{R}^n$  is the unknown state,  $w(k) \in \mathbb{R}^m$  is an unknown input,  $y(k) \in \mathbb{R}^p$  is the output,  $v(k)$  is measurement noise,  $\kappa(k) \in \mathbb{R}$  is an uncertain possibly time-varying parameter, and  $A(\kappa(k)) \in \mathbb{R}^{n \times n}$ ,  $B \in \mathbb{R}^{n \times m}$ , and  $C \in \mathbb{R}^{p \times n}$ . The goal of this study is to estimate  $\kappa(k)$ . For simplicity in comparing estimation algorithms, we consider only the case where a single entry of  $A(\kappa(k))$  is uncertain.

## III. Parameter Identification Algorithms

To estimate the unknown parameter in (1), we use four algorithms, namely, the extended Kalman filter (EKF), the unscented Kalman filter (UKF), the ensemble adjustment Kalman filter (EAKF), and retrospective cost model refinement (RCMR). For the Kalman filter approaches, we treat  $\kappa(k)$  as an unknown state, and we augment the state vector to include  $\kappa(k)$ . In RCMR we model the unknown parameter as an unknown subsystem. In this section, we briefly describe UKF, EAKF, and RCMR.

### III.A. Unscented Kalman Filter

The UKF approach to state estimation of nonlinear systems is developed in [9]. UKF does not use the Jacobian of the dynamics or a factorization of the dynamics to propagate a pseudo error covariance. The starting point for UKF is a set of sample points, that is, a collection of state estimates that capture the initial probability distribution of the state. Let  $x \in \mathbb{R}^n$ , and let  $P \in \mathbb{R}^{n \times n}$  be positive semidefinite. The unscented transformation provides  $2n + 1$  ensembles  $X_i \in \mathbb{R}^n$  and corresponding weights  $\gamma_{x,i}$  and  $\gamma_{y,i}$ , for  $0 = 1, \dots, 2n$ , such that the weighted mean and weighted variance of the ensembles are  $x$  and  $P$ , respectively. Specifically, let  $S \in \mathbb{R}^{n \times n}$  satisfy

$$SS^T = P, \quad (3)$$

and, for all  $i = 1, \dots, n$ , let  $S_i$  denote the  $i$ th column of  $S$ . For  $\alpha > 0$ , the unscented transformation  $X = \Psi(x, S, \alpha) \in \mathbb{R}^{n \times (2n+1)}$  of  $x$  with covariance  $P = SS^T$  is defined by

$$X \triangleq \begin{cases} x, & i = 0, \\ x + \sqrt{\alpha}S_i, & i = 1, \dots, n, \\ x - \sqrt{\alpha}S_{i-n} & i = n + 1, \dots, 2n. \end{cases} \quad (4)$$

The parameter  $\alpha$  determines the spread of the ensembles around  $x$ . Next, define the weights  $\gamma_i \in \mathbb{R}$  by

$$\gamma_0 \triangleq \frac{\alpha - n}{n}, \quad \gamma_i \triangleq \frac{1}{2\alpha}, \quad i = 1, \dots, 2n, \quad (5)$$

Then,

$$\sum_{i=0}^{2n} \gamma_i X_i = x, \quad \sum_{i=0}^{2n} \gamma_i (X_i - x)(X_i - x)^T = P. \quad (6)$$

The unscented transformation (3)-(6) is the scaled unscented transformation given in [17]. This technique ensures that the distance between  $x$  and the sample point  $X_i$  does not increase as  $n$  increases.

UKF uses the unscented transformation to approximate the error covariance and estimate the state  $x_k$ . Letting  $x_0^-$  be an initial estimate of  $x_0$  with error covariance  $P_0^-$ , the data assimilation step of UKF is given by

$$x_k^+ = x_k^- + K_F(Y_k - y_k^-), \quad (7)$$

$$y_k^- = C_k x_k^-, \quad (8)$$

$$X_k^+ = \Psi(x_k^+, S_k^+, \alpha), \quad (9)$$

$$P_k^+ = P_k^- - K_k P_{yy,k} K_k^T, \quad (10)$$

where

$$K_k = P_{xy,k} P_{yy,k}^{-1}, \quad (11)$$

$$P_{xy,k} = \sum_{i=0}^{2n} \gamma_i (X_{i,k}^- - x_k^-)(Y_{i,k}^- - y_k^-)^T, \quad (12)$$

$$P_{yy,k} = \sum_{i=0}^{2n} \gamma_i (Y_{i,k}^- - y_k^-)(Y_{i,k}^- - y_k^-)^T + R_k, \quad (13)$$

$$Y_{i,k}^- = h(X_{i,k}^-, k), i = 0, \dots, 2n, \quad (14)$$

where  $h(X_{i,k}^-, k)$  maps the input to the output and  $S_k^+ \in \mathbb{R}^{n \times n}$  satisfies

$$S_k^+ (S_k^+)^T = P_k^+. \quad (15)$$

The forecast step of UKF is given by

$$X_{i,k+1}^- = f(X_{i,k}^+, u_k, k), i = 0, \dots, 2n, \quad (16)$$

$$x_{k+1}^- = \sum_{i=0}^{2n} \gamma_i X_{i,k+1}^-, \quad (17)$$

$$P_{k+1}^- = \sum_{i=0}^{2n} \gamma_i (X_{i,k+1}^- - x_{k+1}^-)(X_{i,k+1}^- - x_{k+1}^-)^T = Q_k, \quad (18)$$

where  $f(X_{i,k}^+, u_k, k)$  represents the dynamics of the nonlinear system. If the dynamics are linear, then UKF is equivalent to the Kalman filter. Furthermore, in the linear case,  $P_k^+$  and  $P_k^-$  are the covariances of the errors  $x_k - x_k^+$  and  $x_k - x_k^-$ , respectively. However, in the nonlinear case,  $P_k^+$  and  $P_k^-$  are pseudo-error covariances.

At each time step  $k$ , the ensemble  $X_k^+$  is constructed using the unscented transformation based on a square root  $S_k^+$  of  $P_k^+$  satisfying (15). However, the factor  $S_k^+$  satisfying (15) is not unique. For example, the singular value decomposition or the Cholesky factorization can be used to obtain a square root of the pseudo-error covariance  $P_k^+$ . Moreover, if  $S_k^+ = \hat{S}_k^+$  satisfies (15), then, for any orthogonal matrix  $U \in \mathbb{R}^{n \times n}$ , the matrix  $S_k^+ = \hat{S}_k^+ U$  also satisfies (15). For linear dynamics, UKF is equivalent to the Kalman filter, and the performance of UKF does not depend on the choice of  $S_k^+$ . However, for nonlinear dynamics, the performance of UKF depends on the choice of  $S_k^+$ , although simulation results indicate that the performance of UKF is similar for different choices of  $S_k^+$ .

### III.B. Ensemble Adjustment Kalman Filter

In this section we give a brief overview of the Ensemble Adjustment Kalman Filter (EAKF), as described in [6]. Parameter estimation is performed in EAKF by augmenting the state estimate  $\hat{x}_k$  with the parameter estimate  $\hat{\kappa}_k$ . The augmented state vector  $\hat{X}_k = \begin{bmatrix} \hat{x}_k \\ \hat{\kappa}_k \end{bmatrix}$  is initialized with  $N = 20$  realizations of a zero-mean normal random variable with a variance of 1. The time evolution of the state for each initial condition constitutes an ensemble member. In addition to defining an augmented state, we define the joint measurement-state vector  $\hat{z}_k$  by

$$\hat{z}_k = \begin{bmatrix} \hat{X}_k \\ \hat{y}_k \end{bmatrix} = \begin{bmatrix} \hat{x}_k \\ \hat{\kappa}_k \\ \hat{y}_k \end{bmatrix}, \quad (19)$$

where  $\hat{y}_k$  is the measurement estimate given by  $\hat{y}_k = C\hat{x}_k$ . Accordingly, the dimension of  $\hat{z}_k$  is  $n + 1 + l_y$ . By using the parameter-dependent matrices  $A(\hat{\kappa}_k)$  and  $B(\hat{\kappa}_k)$ , the joint vector for  $i^{th}$  ensemble member is propagated in time by means of

$$\hat{z}_{k,i}^- = \begin{bmatrix} A(\hat{\kappa}_k^+) \hat{x}_{k-1,i}^+ + B(\hat{\kappa}_k^+) w_{k-1,i} \\ \hat{\kappa}_k^+ \\ C \hat{x}_k^+ \end{bmatrix}, \quad (20)$$

where the superscript "+" denotes a quantity after the assimilation step (posterior estimates) and the superscript "-" denotes a quantity prior to assimilation (prior estimates).

Next, the prior error covariance  $P_k^-$  can be approximated by calculating the sample covariance between the ensemble joint vectors at step  $k$  as given by

$$P_k^- = \lambda \sum_{i=1}^N (\hat{z}_{k,i}^- - \mu[\hat{z}_k^-]) (\hat{z}_{k,i}^- - \mu[\hat{z}_k^-])^T / (N - 1), \quad (21)$$

where  $\lambda$  is the inflation coefficient,  $N$  is the number of ensemble members, and  $\mu[\cdot]$  denotes the sample mean of its argument. At the next step, the matrix

$$\mathcal{A}_k = (\mathcal{F}_k^T)^{-1} \mathcal{G}_k^T (U_k^T)^{-1} B_k^T (\mathcal{G}_k^T)^{-1} \mathcal{F}_k^T \quad (22)$$

is used to update the posterior error covariance matrix  $P_k^+$ , where  $\mathcal{F}_k$  is obtained from the singular value decomposition (SVD) of  $P_k^- = \mathcal{F}_k D_k \mathcal{F}_k^T$ ,  $\mathcal{G}_k$  is a square root of  $D_k$ ,  $U_k$  is obtained from the SVD of  $\mathcal{G}_k^T \mathcal{F}_k^T \mathcal{H}^T R^{-1} \mathcal{H} \mathcal{F}_k \mathcal{G}_k = U_k J_k U_k^T$ , and  $B_k$  is a square root of  $(I + J_k)^{-1}$ . The matrix  $\mathcal{H} = [0_{l_y \times n} \quad I_{l_y}]$  selects the measurement estimate from the vector  $\hat{z}_k$ . The matrix  $\mathcal{A}_k$  can be thought of as an operator performing a series of rotations and scalings in convenient reference frames. The posterior error covariance is then computed using

$$P_k^+ = \mathcal{A}_k P_k^- \mathcal{A}_k^T. \quad (23)$$

The updated mean of the ensemble is given by

$$\mu[\hat{z}_k^+] = P_k^+ [(P_k^-)^{-1} \mu[\hat{z}_k^-] + \mathcal{H}^T R_k^{-1} y_k]. \quad (24)$$

The final step in EAKF procedure consists of scaling and shifting the ensemble to have the updated covariance (23) and mean (24), that is, for  $i = 1 \dots N$ , defined

$$\hat{z}_{k,i}^+ = \mathcal{A}_k^T (\hat{z}_{k,i}^- - \mu[\hat{z}_k^-]) + \mu[\hat{z}_k^+]. \quad (25)$$

We note that (20) implies that the prior parameter estimate is not updated by the model, that is,  $\hat{\kappa}_k^- = \hat{\kappa}_k^+$ . This update is used since the time evolution of the parameter is unknown, which necessitates

the use of inflation. More precisely, the fact that the model does not update the parameter estimate causes the ensemble spread to approach zero, which can be remedied by manually spreading the ensemble. In the following experiments we keep the ensemble variance of the stiffness estimate  $\hat{k}_k^-$  constant by setting the corresponding entry in  $P_k^-$  to a positive constant  $\sigma_i^2$ . We also check the shape of the estimate sample distribution and correct it if outliers are present. More precisely, if the maximum (minimum) ensemble member reaches values above (below) three standard deviations from the ensemble mean, this ensemble member is reset to a value of the mean plus (minus) one standard deviation. This solution is implemented to prevent the estimate distribution from becoming highly non-Gaussian.

### III.C. Retrospective Cost Model Refinement

Consider the MIMO discrete-time main system

$$x(k+1) = Ax(k) + Bw(k) + D_1u(k), \quad (26)$$

$$y(k) = Cx(k) + v(k), \quad (27)$$

$$y_0(k) = E_1x(k), \quad (28)$$

where  $x(k) \in \mathbb{R}^n$ ,  $y(k) \in \mathbb{R}^{l_y}$ ,  $y_0(k) \in \mathbb{R}^{l_{y_0}}$ ,  $u(k) \in \mathbb{R}^{l_u}$ ,  $w(k) \in \mathbb{R}^{l_w}$ , and  $k \geq 0$ . The main system (26)–(28) is interconnected with the unknown subsystem modeled by

$$u(k) = G_s(\mathbf{q})y_0(k), \quad (29)$$

where  $\mathbf{q}$  is the forward shift operator. The system (26)–(29) represents the true system. The dynamics of the true system can also be expressed as

$$x(k+1) = (A + D_1\kappa E_1)x(k) + Bw(k). \quad (30)$$

where  $\kappa$  is the unknown parameter. We assume that the excitation signal  $w(k)$  is known.  $v(k)$  denotes measurement noise.

Next, we assume a model of the main system of the form

$$\hat{x}(k+1) = \hat{A}\hat{x}(k) + \hat{B}w(k) + \hat{D}_1\hat{u}(k), \quad (31)$$

$$\hat{y}(k) = \hat{C}\hat{x}(k), \quad (32)$$

$$\hat{y}_0(k) = \hat{E}_1\hat{x}(k), \quad (33)$$

where  $\hat{x}(k) \in \mathbb{R}^n$ ,  $\hat{y}(k) \in \mathbb{R}^{l_{\hat{y}}}$ ,  $\hat{y}_0(k) \in \mathbb{R}^{l_{\hat{y}_0}}$ ,  $\hat{u}(k) \in \mathbb{R}^{l_{\hat{u}}}$ . The model of the main system is interconnected with the subsystem model

$$\hat{u}(k) = \hat{G}_s(\mathbf{q})\hat{y}_0(k). \quad (34)$$

We choose  $\hat{G}_s(\mathbf{q})$  to be an FIR first order approximation of  $G_s(\mathbf{q})$ . The goal is to estimate the subsystem model  $\hat{G}_s(\mathbf{q})$  by minimizing a cost function based on the performance variable

$$z(k) \triangleq \hat{y}(k) - y(k) \in \mathbb{R}^{l_z}. \quad (35)$$

We estimate  $\hat{G}_s(\mathbf{q})$  by retrospectively reconstructing the signal  $\hat{u}(k)$  that minimizes the performance at the current time step. The reconstruction of  $\hat{u}(k)$  uses minimal modeling information about the true system (26)–(28), namely, a limited number of Markov parameters. We then use  $\hat{u}(k)$  and  $\hat{y}_0(k)$  to construct  $\hat{G}_s(\mathbf{q})$ . Figure 1 illustrates the model-refinement architecture.

We begin by defining Markov parameters of the main system model  $\hat{G}(\mathbf{q})$ . For  $i \geq 1$ , let

$$H_i \triangleq \hat{C}\hat{A}^{i-1}\hat{D}_1. \quad (36)$$

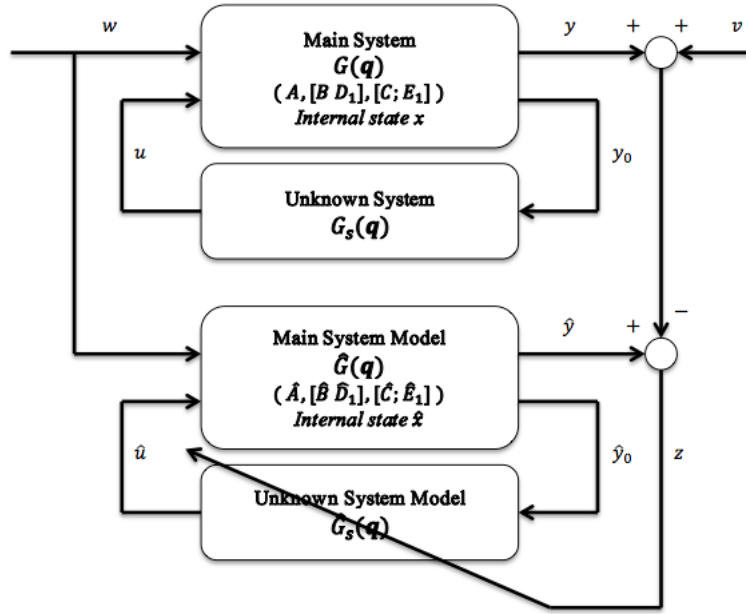


Figure 1. Model-refinement architecture.

Therefore,  $H_1 = \hat{C}\hat{D}_1$  and  $H_2 = \hat{C}\hat{A}\hat{D}_1$ . Let  $r$  be a positive integer. Then, for all  $k \geq r$ ,

$$\hat{x}(k) = \hat{A}^r \hat{x}(k-r) + \sum_{i=1}^r \hat{A}^{i-1} \hat{B}w(k-i) + \sum_{i=1}^r \hat{A}^{i-1} \hat{D}_1 \hat{u}(k-i), \quad (37)$$

and thus

$$z(k) = \hat{C}\hat{A}^r \hat{x}(k-r) + \sum_{i=1}^r \hat{C}\hat{A}^{i-1} \hat{B}w(k-i) - y(k) + \bar{\mathcal{H}}\bar{U}(k-1), \quad (38)$$

where  $\bar{\mathcal{H}} \triangleq \begin{bmatrix} H_1 & \dots & H_r \end{bmatrix} \in \mathbb{R}^{l_z \times r l_a}$ , and  $\bar{U}(k-1) \triangleq \begin{bmatrix} \hat{u}^T(k-1) & \dots & \hat{u}^T(k-r) \end{bmatrix}^T$ .

Next, we rearrange the columns of  $\bar{\mathcal{H}}$  and the components of  $\bar{U}(k-1)$  and partition the resulting matrix and vector so that

$$\bar{\mathcal{H}}\bar{U}(k-1) = \mathcal{H}'U'(k-1) + \mathcal{H}U(k-1), \quad (39)$$

where  $\mathcal{H}' \in \mathbb{R}^{l_z \times (r l_a - l_U)}$ ,  $\mathcal{H} \in \mathbb{R}^{l_z \times l_U}$ ,  $U'(k-1) \in \mathbb{R}^{r l_a - l_U}$ , and  $U(k-1) \in \mathbb{R}^{l_U}$ . Then, we can rewrite (38) as

$$z(k) = \mathcal{S}(k) + \mathcal{H}U(k-1), \quad (40)$$

where

$$\mathcal{S}(k) \triangleq \hat{C}\hat{A}^r \hat{x}(k-r) + \sum_{i=1}^r \hat{C}\hat{A}^{i-1} \hat{B}w(k-i) - y(k) + \mathcal{H}'U'(k-1). \quad (41)$$

For example,  $\bar{\mathcal{H}} = \begin{bmatrix} H_1 & H_2 & H_3 \end{bmatrix}$ ,

$$\mathcal{H}' = \begin{bmatrix} H_1 & H_2 \end{bmatrix}, \quad U'(k-1) = \begin{bmatrix} \hat{u}(k-1) \\ \hat{u}(k-2) \end{bmatrix},$$

and  $\mathcal{H} = H_3$ ,  $U(k-1) = \hat{u}(k-3)$ . Next, we rewrite (40) with a delay of  $k_j$  time steps, where  $0 \leq k_1 \leq k_2 \leq \dots \leq k_s$ , in the form

$$z(k-k_j) = \mathcal{S}_j(k-k_j) + \mathcal{H}_j U_j(k-k_j-1), \quad (42)$$

where (41) becomes

$$\mathcal{S}_j(k-k_j) \triangleq \hat{C} \hat{A}^r \hat{x}(k-k_j-r) + \sum_{i=1}^r \hat{C} \hat{A}^{i-1} \hat{B} w(k-k_j-i) - y_0(k-k_j) + \mathcal{H}'_j U'_j(k-k_j-1) \quad (43)$$

and (39) becomes

$$\tilde{\mathcal{H}} \tilde{U}(k-k_j-1) = \mathcal{H}'_j U'_j(k-k_j-1) + \mathcal{H}_j U_j(k-k_j-1), \quad (44)$$

where  $\mathcal{H}'_j \in \mathbb{R}^{l_z \times (r l_{\hat{u}} - l_{U_j})}$ ,  $\mathcal{H}_j \in \mathbb{R}^{l_z \times l_{U_j}}$ ,  $U'_j(k-k_j-1) \in \mathbb{R}^{r l_{\hat{u}} - l_{U_j}}$ , and  $U_j(k-k_j-1) \in \mathbb{R}^{l_{U_j}}$ . Now, by stacking  $z(k-k_1), \dots, z(k-k_s)$ , we define the *extended performance*

$$Z(k) \triangleq \begin{bmatrix} z^T(k-k_1) & \dots & z^T(k-k_s) \end{bmatrix}^T \in \mathbb{R}^{s l_z}. \quad (45)$$

Therefore,

$$Z(k) \triangleq \tilde{\mathcal{S}}(k) + \tilde{\mathcal{H}} \tilde{U}(k-1), \quad (46)$$

where  $\tilde{\mathcal{S}}(k) \triangleq \begin{bmatrix} \mathcal{S}^T(k-k_1) & \dots & \mathcal{S}^T(k-k_s) \end{bmatrix}^T \in \mathbb{R}^{s l_z}$ ,  $\tilde{\mathcal{H}} \in \mathbb{R}^{s l_z \times l_{\tilde{U}}}$ , and  $\tilde{U}(k-1) \in \mathbb{R}^{l_{\tilde{U}}}$ . The vector  $\tilde{U}(k-1)$  is formed by stacking  $U_1(k-k_1-1), \dots, U_s(k-k_s-1)$  and removing repetitions of components. For example, with  $k_1 = 0$  and  $k_2 = 1$ , stacking  $U_1(k-1) = \begin{bmatrix} \hat{u}(k-1) \\ \hat{u}(k-2) \end{bmatrix}$  and  $U_2(k-2) = \hat{u}(k-2)$  results in  $\tilde{U}(k-1) = \begin{bmatrix} \hat{u}(k-1) \\ \hat{u}(k-2) \end{bmatrix}$ . The coefficient matrix  $\tilde{\mathcal{H}}$  consists of the entries of  $\mathcal{H}_1, \dots, \mathcal{H}_s$  arranged according to the structure of  $\tilde{U}(k-1)$ . Furthermore, we assume that the last entry of  $\tilde{U}(k-1)$  is a component of  $\hat{u}(k-r)$ .

Next, we define the *retrospective performance*

$$\hat{z}(k-k_j) \triangleq \mathcal{S}_j(k-k_j) + \mathcal{H}_j U_j^*(k-k_j-1), \quad (47)$$

where the actual past subsystem outputs  $U_j(k-k_j-1)$  in (42) are replaced by the retrospective subsystem outputs  $U_j^*(k-k_j-1)$ . The *extended retrospective performance* for (47), which is defined as

$$\hat{Z}(k) \triangleq \begin{bmatrix} \hat{z}^T(k-k_1) & \dots & \hat{z}^T(k-k_s) \end{bmatrix}^T \in \mathbb{R}^{s l_z}, \quad (48)$$

is given by

$$\hat{Z}(k) = \tilde{\mathcal{S}}(k) + \tilde{\mathcal{H}} \tilde{U}^*(k-1), \quad (49)$$

where the components of  $\tilde{U}^*(k-1) \in \mathbb{R}^{l_{\tilde{U}}}$  are components of  $U_1^*(k-k_1-1), \dots, U_s^*(k-k_s-1)$  ordered in the same way as the components of  $\tilde{U}(k-1)$ . Subtracting (46) from (49) yields

$$\hat{Z}(k) = Z(k) - \tilde{\mathcal{H}} \tilde{U}(k-1) + \tilde{\mathcal{H}} \tilde{U}^*(k-1). \quad (50)$$

Finally, we define the *retrospective cost function*

$$J(\tilde{U}^*(k-1), k) \triangleq \hat{Z}^T(k) R(k) \hat{Z}(k), \quad (51)$$

where  $R(k) \in \mathbb{R}^{s l_z \times s l_z}$  is a positive-definite performance weighting. The goal is to determine refined subsystem outputs  $\tilde{U}^*(k-1)$  that would have provided better performance than the subsystem outputs  $U(k)$  that were applied to the system. The refined subsystem outputs values  $\tilde{U}^*(k-1)$  are subsequently used to update the subsystem estimate.

### III.C.1. Cost Function Optimization with Adaptive Regularization

To ensure that (51) has a global minimizer, we consider the regularized cost

$$\bar{J}(\tilde{U}^*(k-1), k) \triangleq \hat{Z}^T(k)R(k)\hat{Z}(k) + \eta(k)\tilde{U}^{*\text{T}}(k-1)\tilde{U}^*(k-1), \quad (52)$$

where  $\eta(k) = \bar{\eta}z^T(k)z(k)$  and  $\bar{\eta} \geq 0$ . Substituting (50) into (52) yields

$$\bar{J}(\tilde{U}^*(k-1), k) = \tilde{U}^*(k-1)^T \mathcal{A}(k) \tilde{U}^*(k-1) + \mathcal{B}(k) \tilde{U}^*(k-1) + \mathcal{C}(k), \quad (53)$$

where

$$\mathcal{A}(k) \triangleq \tilde{\mathcal{H}}^T R(k) \tilde{\mathcal{H}} + \eta(k) I_{l_{\tilde{U}}}, \quad (54)$$

$$\mathcal{B}(k) \triangleq 2\tilde{\mathcal{H}}^T R(k) [Z(k) - \tilde{\mathcal{H}}\tilde{U}(k-1)], \quad (55)$$

$$\mathcal{C}(k) \triangleq Z^T(k)R(k)Z(k) - 2Z^T(k)R(k)\tilde{\mathcal{H}}\tilde{U}(k-1) + \tilde{U}^T(k-1)\tilde{\mathcal{H}}^T R(k)\tilde{\mathcal{H}}\tilde{U}(k-1). \quad (56)$$

If either  $\tilde{\mathcal{H}}$  has full column rank or  $\eta(k) > 0$ , then  $\mathcal{A}(k)$  is positive definite. In this case,  $\bar{J}(\tilde{U}^*(k-1), k)$  has the unique global minimizer

$$\tilde{U}^*(k-1) = -\frac{1}{2}\mathcal{A}^{-1}(k)\mathcal{B}(k). \quad (57)$$

### III.C.2. Subsystem Modeling

The subsystem output  $\hat{u}(k)$  is given by the strictly proper FIR time-series model of order  $n_c$  given by

$$\hat{u}(k) = \sum_{i=1}^{n_c} N_i(k)\hat{y}(k-i), \quad (58)$$

where, for all  $i = 1, \dots, n_c$ ,  $N_i(k) \in \mathbb{R}^{l_u \times l_{\hat{y}}}$ . The subsystem output (58) can be expressed as  $\hat{u}(k) = \theta(k)\phi(k-1)$ , where  $\theta(k) \in \mathbb{R}^{l_u \times n_c l_{\hat{y}}}$  and  $\phi(k) \in \mathbb{R}^{n_c l_{\hat{y}}}$ ,

$$\theta(k) \triangleq [N_1(k) \ \cdots \ N_{n_c}(k)], \quad (59)$$

$$\phi(k-1) \triangleq [\hat{y}^T(k-1) \ \cdots \ \hat{y}^T(k-n_c)]^T. \quad (60)$$

If  $n_c = 1$ , then

$$\hat{u}(k) = \theta(k)\hat{y}(k-1) = N_1(k)\hat{y}(k-1). \quad (61)$$

Thus,  $\theta(k)$  is the estimate of the uncertain parameter  $\kappa$ .

### III.C.3. Recursive Least Squares Update

Let  $d > 0$  such that  $\tilde{U}^*(k-1)$  contains  $u^*(k-d)$ , and define the cost

$$J_R(\theta(k)) \triangleq \sum_{i=1}^k \lambda^{k-i} \|u^{*\text{T}}(k-d) - \phi^T(k-d-1)\theta^T(k)\|^2 + \lambda^k (\theta(k) - \theta(0))P^{-1}(0)(\theta(k) - \theta(0))^T, \quad (62)$$



where  $\phi(k-d)$  is given by (60),  $\|\cdot\|$  is the Euclidean norm, and  $\lambda(k) \in (0, 1]$  is the forgetting factor. Minimizing the cumulative cost function recursively yields

$$\theta^T(k) = \theta^T(k-1) + P(k)\phi(k-d-1) \cdot (u^*(k-d) - \phi^T(k-d-1)\theta^T(k-1)), \quad (63)$$

$$P(k) = \lambda^{-1}(k)P(k-1) - \lambda^{-1}(k)P(k-1)\phi(k-d-1) \cdot [\phi^T(k-d-1)P(k-1)\phi(k-d) + \lambda(k)]^{-1} \cdot \phi^T(k-d-1)P(k-1). \quad (64)$$

We initialize  $P(0) = \beta I$ , where  $\beta > 0$ .

## IV. Mass-Spring-Damper Example with Uncertain Constant Stiffness

Consider the mass-spring-damper system

$$\begin{aligned} m\ddot{q} + c\dot{q} + \kappa q &= w, \\ y &= \dot{q} + \nu, \end{aligned} \quad (65)$$

where  $q$  and  $\dot{q}$  are the position and velocity of the mass, respectively,  $m = 1$ ,  $c = 5$ , and  $\kappa = 10$  are the mass, damping, and spring constants, respectively,  $w$  is the force input, and  $\nu$  is white Gaussian noise with mean  $\mu_\nu$  and variance  $\sigma_\nu^2$ . The stiffness is assumed to be constant and uncertain. The state space representation of (65) is given by

$$\begin{bmatrix} \dot{q} \\ \ddot{q} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{\kappa}{m} & -\frac{c}{m} \end{bmatrix} \begin{bmatrix} q \\ \dot{q} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix} w, \quad (66)$$

$$y = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} q \\ \dot{q} \end{bmatrix} + v. \quad (67)$$

Euler discretization of (66) and (67) with sampling time  $T_s = 0.1$  yields

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} 1 & T_s \\ -\frac{\kappa T_s}{m} & 1 - \frac{c T_s}{m} \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{T_s}{m} \end{bmatrix} w(k), \quad (68)$$

$$y(k) = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + v(k), \quad (69)$$

where  $x_1(k) \triangleq q(kT_s)$  and  $x_2(k) \triangleq \dot{q}(kT_s)$ . For each algorithm, we use  $\hat{\kappa}(0) = 0$ ,  $\hat{x}_1(0) = 0.1$  and  $\hat{x}_2(0) = 0.01$ . For EKF, we use  $R = \sigma_\nu^2$  and  $Q = \text{diag}([0.1, 0.1, 5000])$ , for UKF, we use  $R = \sigma_\nu^2$  and  $Q = \text{diag}([0.1, 0.1, 200])$ , for EAKF, we use  $\hat{R} = 10^{-5}$ ,  $\lambda = 1.10$  and  $\sigma_i^2 = 0.25$ , and, for RCMR, we use  $\beta = 0.42$ ,  $\lambda = 1$ ,  $\eta = 0$ , and  $\tilde{\mathcal{H}} = [H_1 \ H_2]$ , unless otherwise specified.

**Example IV.1.** ( $\mu_\nu = 0$  and  $\sigma_\nu^2 = 0$ ) We first consider the case where there is no measurement noise. Figure 2 shows that all algorithms estimate the unknown parameter with roughly the same accuracy.

**Example IV.2.** ( $\mu_\nu = 0$  and  $\sigma_\nu^2 = 10^{-5}$ ) We now consider measurement noise where  $\sigma_\nu^2 = 10^{-5}$ . Figure 3 shows that EKF, UKF, and RCMR yield estimates of  $\kappa$  with smaller error than EAKF.

**Example IV.3.** ( $\mu_\nu = 10^{-3}$  and  $\sigma_\nu^2 = 0$ ) We now consider the case where  $v$  represents sensor bias. Specifically,  $\mu_\nu = 10^{-3}$  and  $\sigma_\nu^2 = 0$ . Figure 4 shows that EKF, UKF, and RCMR yield estimates of  $\kappa$  with smaller error than EAKF.

**Example IV.4.** ( $\mu_\nu = 10^{-3}$  and  $\sigma_\nu^2 = 10^{-5}$ ) We now consider the case where  $\mu_\nu = 10^{-3}$  and  $\sigma_\nu^2 = 10^{-5}$ . Figure 5 shows that EKF, UKF, and RCMR yield estimates of  $\kappa$  with smaller error than EAKF.

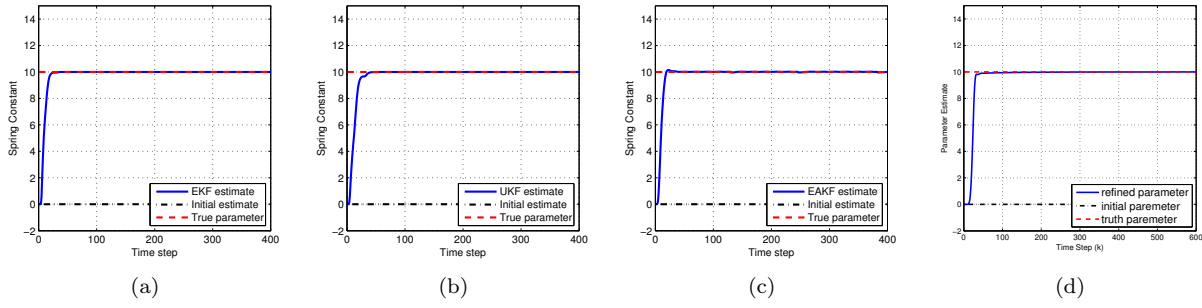


Figure 2. Example IV.1 Estimates of constant stiffness  $\kappa$  with  $\mu_v = 0$  and  $\sigma_v^2 = 0$ . (a) EKF, (b) UKF, (c) EAKF, (d) RCMR.

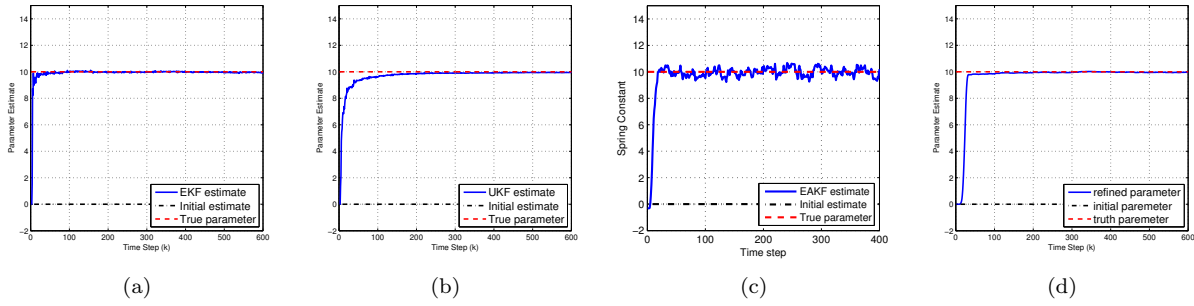


Figure 3. Example IV.2. Estimates of constant stiffness  $\kappa$  with  $\mu_v = 0$  and  $\sigma_v^2 = 10^{-5}$ . (a) EKF, (b) UKF, (c) EAKF, (d) RCMR.

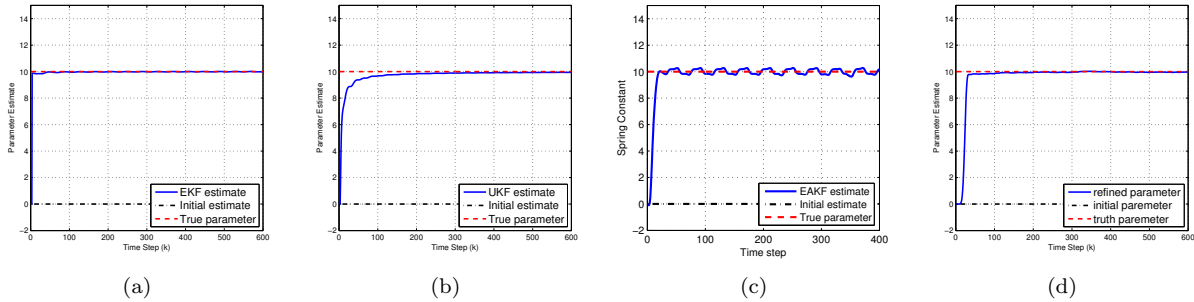


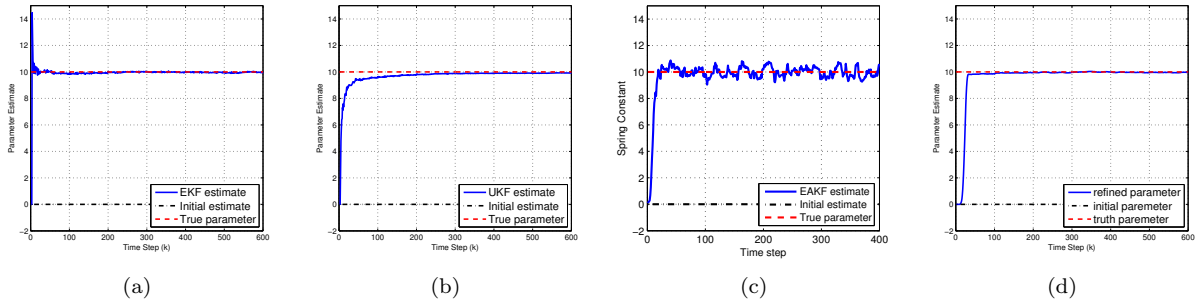
Figure 4. Example IV.3. Estimates of constant stiffness  $\kappa$  with  $\mu_v = 10^{-3}$  and  $\sigma_v^2 = 0$ . (a) EKF, (b) UKF, (c) EAKF, (d) RCMR.

## V. Mass-Spring-Damper with Uncertain Time-Varying Stiffness

We consider the same model as in Section IV, but with time-varying  $\kappa$ . Specifically,

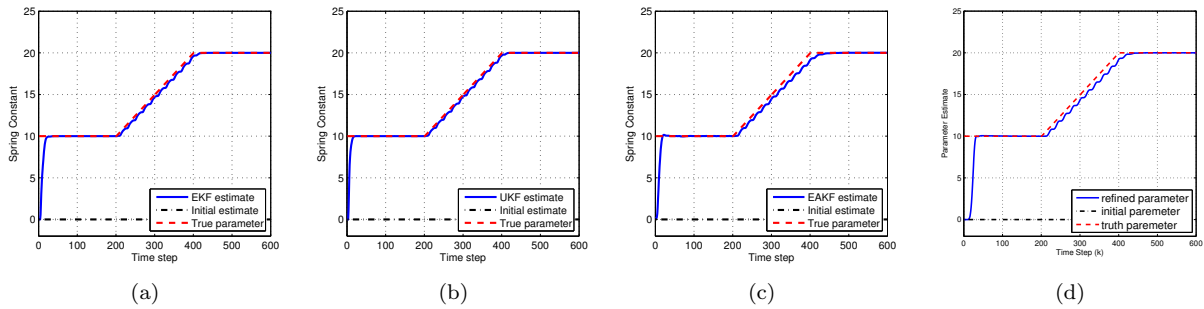
$$\kappa = \begin{cases} 10, & k < 200, \\ 10 + (k - 200)/20, & 200 \leq k \leq 400, \\ 20, & k > 400. \end{cases}$$

For each algorithm, we use  $\hat{\kappa}(0) = 0$ ,  $\hat{x}_1(0) = 0.1$  and  $\hat{x}_2(0) = 0.01$ . For EKF, we use  $R = \sigma_v^2$  and  $Q = \text{diag}([0.1, 0.1, 5000])$ , for UKF, we use  $R = \sigma_v^2$  and  $Q = \text{diag}([0.1, 0.1, 200])$ , for EAKF, we use  $\hat{R} = 10^{-5}$ ,  $\lambda = 1.10$  and  $\sigma_i^2 = 0.25$ , and, for RCMR, we use  $\beta = 0.3$ ,  $\lambda = 0.985$ ,  $\eta = 0$ , and  $\mathcal{H} = [H_1 \ H_2]$ , unless otherwise specified.



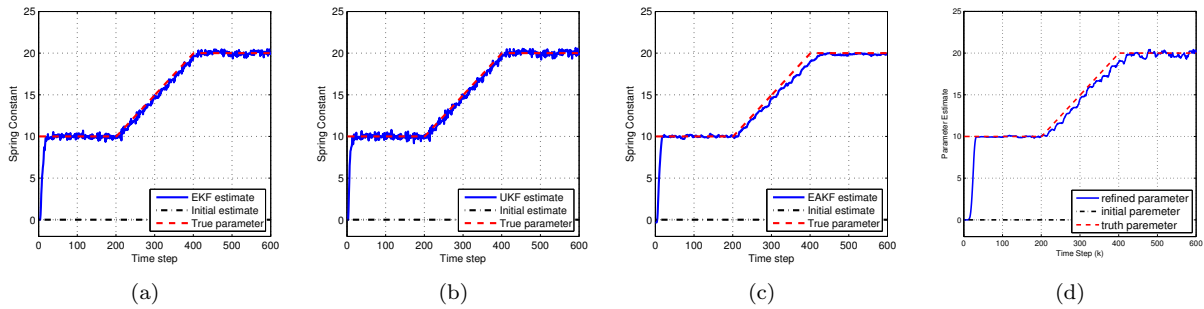
**Figure 5. Example IV.4. Estimates of constant stiffness  $\kappa$  with  $\mu_v = 10^{-3}$  and  $\sigma_v^2 = 10^{-5}$ . (a) EKF, (b) UKF, (c) EAKF, (d) RCMR.**

**Example V.1.** ( $\mu_v = 0$  and  $\sigma_v^2 = 0$ ) We first consider the case where there is no measurement noise. Figure 6 shows that all algorithms identify the unknown parameter with roughly the same accuracy.



**Figure 6. Example V.1. Estimates of time-varying stiffness  $\kappa$  with  $\mu_v = 0$  and  $\sigma_v^2 = 0$ . (a) EKF, (b) UKF, (c) EAKF, (d) RCMR.**

**Example V.2.** ( $\mu_v = 0$  and  $\sigma_v^2 = 10^{-5}$ ) We now consider the case where  $\sigma_v^2 = 10^{-5}$ . Figure 7 shows that RCMR yields an estimate of  $\kappa$  with smaller error than EKF, UKF, and EAKF.



**Figure 7. Example V.2. Estimates of time-varying stiffness  $\kappa$  with  $\mu_v = 0$  and  $\sigma_v^2 = 10^{-5}$ . (a) EKF, (b) UKF, (c) EAKF, (d) RCMR.**

**Example V.3.** ( $\mu_v = 10^{-3}$  and  $\sigma_v^2 = 0$ ) We now consider the case where  $v$  represents sensor bias. Specifically,  $\mu_v = 10^{-3}$ . Figure 8 shows that all algorithms identify the unknown parameter with roughly the same accuracy.

**Example V.4.** ( $\mu_v = 10^{-3}$  and  $\sigma_v^2 = 10^{-5}$ ) We now consider the case where  $\mu_v = 10^{-3}$  and  $\sigma_v^2 = 10^{-5}$ . Figure 9 shows that RCMR yields an estimate of  $\kappa$  with smaller error than EKF, UKF, and EAKF.

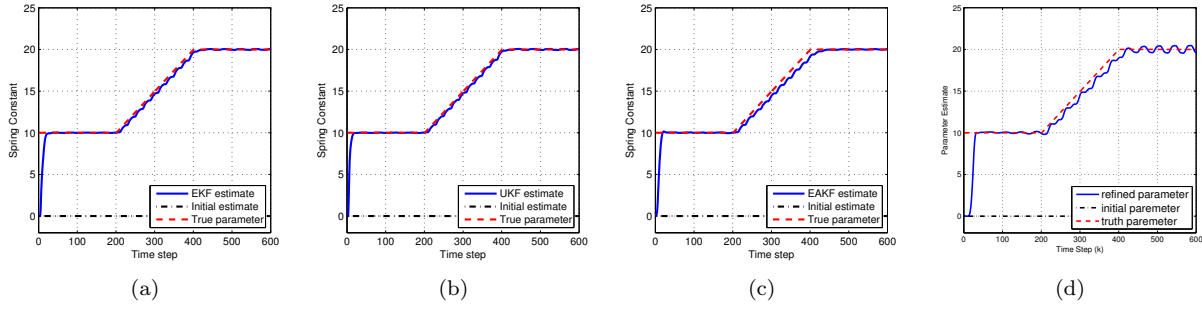


Figure 8. Example V.3. Estimates of time-varying stiffness  $\kappa$  with  $\mu_v = 10^{-3}$  and  $\sigma_v^2 = 0$ . (a) EKF, (b) UKF, (c) EAKF, (d) RCMR.

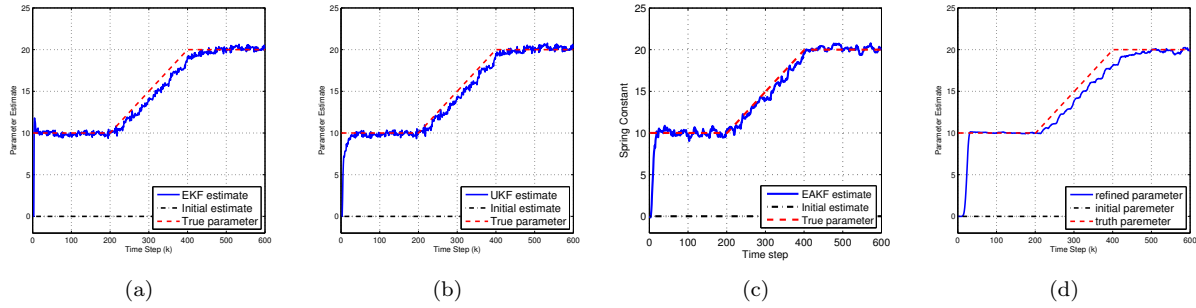


Figure 9. Example V.4. Estimates of time-varying stiffness  $\kappa$  with  $\mu_v = 10^{-3}$  and  $\sigma_v^2 = 10^{-5}$ . (a) EKF, (b) UKF, (c) EAKF, (d) RCMR.

## VI. Estimation of Linearized Aircraft Dynamics in the Presence of Modeling Errors

In this section we consider an example where entries of  $A$  and  $B$  that are not being estimated have modeling errors. We consider a discretized model of a hypersonic aircraft, where  $x = \begin{bmatrix} \beta & P & R & \phi \end{bmatrix}^T$ . The true system model is obtained by Euler discretization of a continuous-time plant with  $T_s = 0.001$ . The discrete-time matrices are given by

$$A_{\text{true}} = \begin{bmatrix} 0.9999 & 0.000269 & -0.000963 & -0.000039 \\ A_{2,1} & 1.00021 & 0.000099 & 0 \\ 0.00041 & 0.000037 & 0.9973 & 0 \\ 0 & 0.00100 & -0.00042 & 1.0000 \end{bmatrix}, B_{\text{true}} = \begin{bmatrix} 0 \\ 0.002519 \\ -0.0000665 \\ 0 \end{bmatrix}, \quad (70)$$

$$C_{\text{true}} = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}, \quad (71)$$

where

$$A_{2,1} = \begin{cases} -0.02560 & k < 25000, \\ -0.02560 - 1.6640 \times 10^{-6}k & 25000 \leq k \leq 75000, \\ -0.1088 & k > 75000. \end{cases} \quad (72)$$

However, aside from  $A_{2,1}$ , we assume that the (3, 1) entry of  $A$  and the (3, 1) component of  $B$  are erroneous and are given by

$$A_{\text{model}} = \begin{bmatrix} 0.9999 & 0.000269 & -0.000963 & -0.000039 \\ \hat{A}_{2,1} & 1.00021 & 0.000099 & 0 \\ 0.000616 & 0.000037 & 0.9973 & 0 \\ 0 & 0.00100 & -0.00042 & 1.0000 \end{bmatrix}, B_{\text{model}} = \begin{bmatrix} 0 \\ 0.002519 \\ -0.0000222 \\ 0 \end{bmatrix}, \quad (73)$$

$$C_{\text{model}} = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}. \quad (74)$$

The goal is to thus estimate  $A_{2,1}$ , despite  $A_{\text{model}}$  and  $B_{\text{model}}$  having erroneous entries.

In all subsequent examples, the input signal  $w$  is measurement noise with  $\mu_w = 0$  and  $\sigma_w^2 = 0.001$ , the initial estimate is  $\hat{A}_{2,1}(0) = -0.2560$ , and the initial conditions are  $x(0) = 0$ . The examples below consider parameter estimation for each algorithm under various scenarios of noise and bias in the measurement.

In all cases, for RCMR,  $\beta = 1$  and  $\eta = 0$ . We do not compare UKF or EAKF since these were found to be difficult to tune. In fact, considerable retuning was required for each case when using EKF.

**Example VI.1.** We first consider the case where no measurement noise is present. We choose  $\lambda = 0.97$  for this case. Figure 10 shows the accuracy of EKF and RCMR.

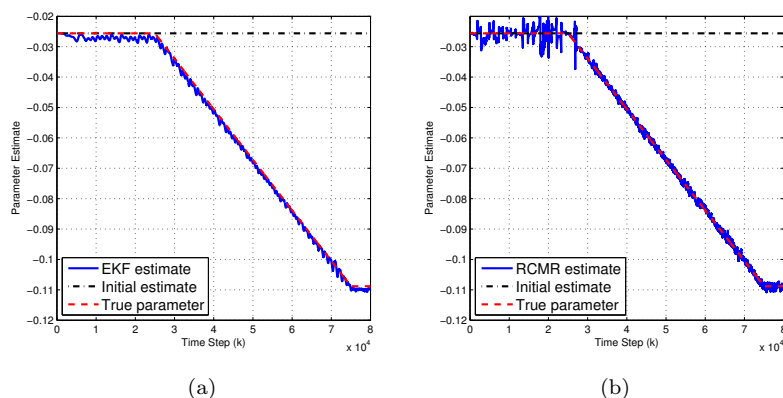


Figure 10. Estimates of  $A_{2,1}$  with  $\mu_v = 0$  and  $\sigma_v^2 = 0$ . (a) EKF, (b) RCMR.

**Example VI.2.** We next consider the case where zero-mean measurement noise is present. We assume  $\mu_v = 0$ ,  $\sigma_v^2 = 1 \times 10^{-8}$  and choose  $\lambda = 0.999$ . Figure 11 shows the accuracy of EKF and RCMR for the hypersonic aircraft in this case.

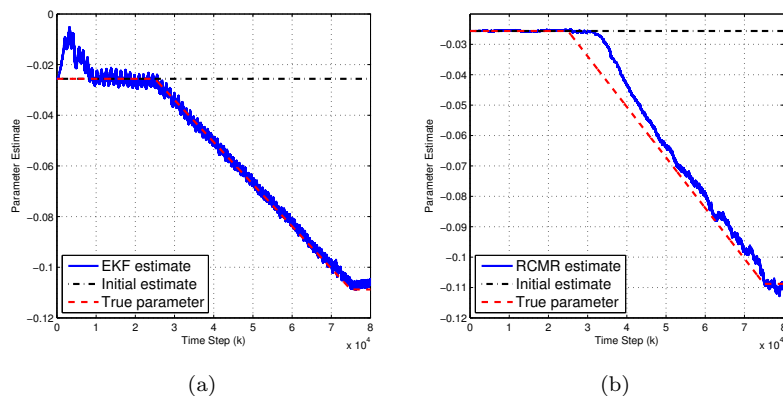


Figure 11. Estimates of  $A_{2,1}$  with  $\mu_v = 0$  and  $\sigma_v^2 = 1 \times 10^{-8}$ . (a) EKF, (b) RCMR.

**Example VI.3.** We next consider the case where biased measurement noise is present. We assume  $\mu_v = 1 \times 10^{-2}$ ,  $\sigma_v^2 = 0$  and choose  $\lambda = 0.97$ . Figure 12 shows the accuracy of RCMR. Note that EKF is not able to estimate the parameter correctly despite considerable tuning effort. Note that we do not attempt to measure the measurement bias when using the EKF.

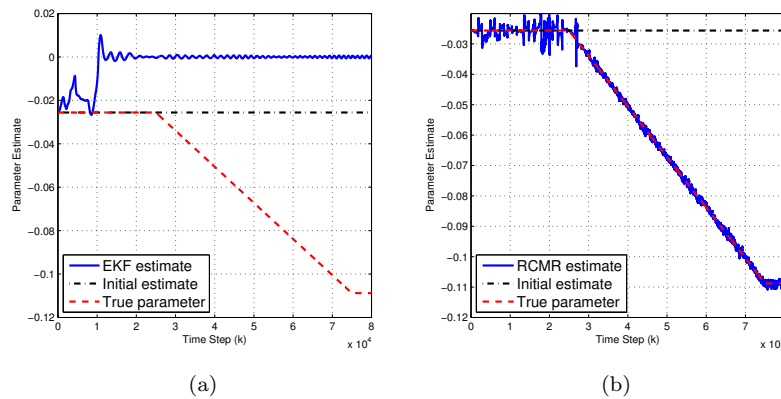


Figure 12. Estimates of  $A_{2,1}$  with  $\mu_v = 1 \times 10^{-2}$  and  $\sigma_v^2 = 0$ . (a) EKF, (b) RCMR.

**Example VI.4.** We next consider the case of nonzero-mean measurement noise. We assume  $\mu_v = 1 \times 10^{-2}$ ,  $\sigma_v^2 = 1 \times 10^{-8}$  and choose  $\lambda = 0.995$  for this case. Figure 13 shows the estimate accuracy of RCMR for the hypersonic aircraft in this case.

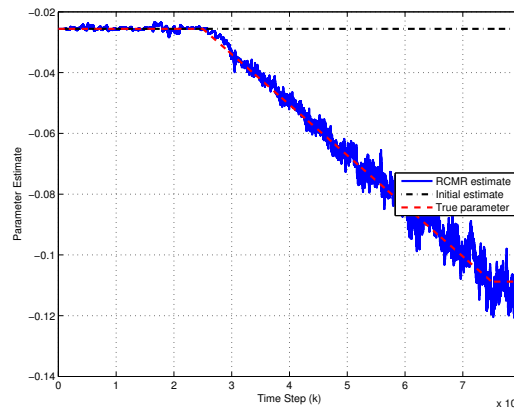


Figure 13. Estimates of  $A_{2,1}$  with  $\mu_v = 1 \times 10^{-2}$  and  $\sigma_v^2 = 1 \times 10^{-8}$ .

## VII. Conclusions

The goal of this paper is to compare the accuracy of techniques for parameter estimation. A plethora of techniques exists for nonlinear estimation, and all of these algorithms are candidates for parameter estimation. In this paper we focused on three nonlinear estimation algorithms. We also considered retrospective cost model refinement (RCMR), which is not an estimation technique but rather is intended for subsystem identification, of which parameter estimation is a special case. These results suggest that RCMR is less sensitive to measurement noise, particularly, biased measurements. No attempt was made to estimate the measurement bias. Since RCMR does not involve an ensemble of models, it is computationally efficient compared to EAKF and UKF. However, unlike EAKF and UKF, RCMR does not provide an error probability distribution for the parameter estimates. The results of this paper can also provide a starting point for comparison with additional nonlinear estimation techniques in future work.

## VIII. Acknowledgment

The authors wish to thank John Burken for providing the example considered in Section VI.

## References

- <sup>1</sup>Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R., "Dual state-parameter estimation of hydrological models using ensemble Kalman filter," *Advances in Water Resources*, Vol. 28, 2005, pp. 135–147.
- <sup>2</sup>Merwe, R. V. D., and Wan, E. A., "The square-root unscented Kalman filter for state and parameter-estimation," *Acoustics, Speech, and Signal Processing*, Beaverton, OR, May 2001, pp. 3461–3464.
- <sup>3</sup>Lefebvre, T., Bruyninckx, H., and Schutter, J. D., "Kalman Filters for non-linear systems: a comparison of performance," *International Journal of Control*, Vol. 77, No. 7, 2004.
- <sup>4</sup>Ljung, L., "Asymptotic Behavior of the Extended Kalman Filter as a Parameter Estimator for Linear Systems," *IEEE Trans. Autom. Control*, Vol. 24, 1979, pp. 36–50.
- <sup>5</sup>Grizzle, J. W. and Song, Y., "The Extended Kalman Filter as a Local Asymptotic Observer for Nonlinear Discrete-Time Systems," *Journal of Mathematical Systems, Estimation and Control*, Vol. 5, 1995, pp. 59–78.
- <sup>6</sup>Anderson, J. L., "An Ensemble Adjustment Kalman Filter for Data Assimilation," *Monthly Weather Review*, Vol. 129, 2001, pp. 2884–2903.
- <sup>7</sup>Evensen, G., "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *Journal of Geophysical Research*, Vol. 99, 1994, pp. 10143–10162.
- <sup>8</sup>Crisan, D. and Doucet, A., "A Survey of Convergence Results on Particle Filtering Methods for Practitioners," *IEEE Transactions on Signal Processing*, Vol. 50, 2002, pp. 736–746.
- <sup>9</sup>Julier, S. J. and Uhlmann, J. K., "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, Vol. 92, 2004, pp. 401–422.
- <sup>10</sup>Thrun, S., Burgard, W., and Fox, D., *Probabilistic Robotics*, MIT Press, 2005.
- <sup>11</sup>Sorenson, H. W. and Alspach, D. L., "Recursive Bayesian Estimation Using Gaussian Sums," *Automatica*, Vol. 7, 1971, pp. 465–479.
- <sup>12</sup>Hanebeck, U. D., Briechle, K., and Rauh, A., "Progressive Bayes: A New Framework for Nonlinear State Estimation," *Proceedings of SPIE*, Vol. 5099, 2003, pp. 256–267.
- <sup>13</sup>Kumar, M., Singla, P., Chakravorty, S., and Junkins, J., "The Partition of Unity Finite Element Approach to the Stationary Fokker-Planck Equation," *AIAA/AAS Astrodynamics Specialist Conference and Exhibit*, Keystone, CO, August 2006, AIAA-2006-6285.
- <sup>14</sup>Rao, C. V., Rawlings, J., and Mayne, D. Q., "Constrained State Estimation for Nonlinear Discrete-Time Systems: Stability and Moving Horizon Approximations," *IEEE Transactions on Automatic Control*, Vol. 48, 2003, pp. 246–258.
- <sup>15</sup>D'Amato, A. M., Springmann, J., Ali, A. A., Cutler, J. W., Ridley, A. J., and Bernstein, D. S., "Adaptive State Estimation for Nonminimum-Phase Systems with Uncertain Harmonic Inputs," *Proc. AIAA Guid. Nav. Contr. Conf.*, Portland, OR, August 2011, AIAA-2011-6315.
- <sup>16</sup>Agarwal, K., Ali, A. A., D'Amato, A. M., Ridley, A. J., and Bernstein, D. S., "Retrospective-Cost-Based Adaptive State Estimation and Input Reconstruction for the Global Ionosphere-Thermosphere Model," *Proc. AIAA Guid. Nav. Contr. Conf.*, Minneapolis, MN, August 2012, AIAA-2012-4601-996.
- <sup>17</sup>Julier, S. J., "The Scaled Unscented Transformation," *Proc. Amer. Contr. Conf.*, Anchorage, May 2002, pp. 4555–4559.