JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 42, No. 8, August 2019

Gradient-, Ensemble-, and Adjoint-Free Data-Driven Parameter Estimation

Ankit Goel* and Dennis S. Bernstein†
University of Michigan, Ann Arbor, Michigan 48109

DOI: 10.2514/1.G004158

Nonlinear estimation methods, such as the extended Kalman filter, unscented Kalman filter, and ensemble Kalman filter, can be used for parameter estimation by viewing the unknown parameters as constant states. This paper presents an alternative approach to this problem based on retrospective cost parameter estimation (RCPE), which uses the difference between the output of the physical system and the output of the model to update the parameter estimate. The parameter update is based on a retrospective cost function, whose minimizer updates the coefficients of the estimator. The present paper extends RCPE to the case where the model depends nonlinearly on multiple unknown parameters. The main contribution is to demonstrate the need for choosing a permutation matrix that correctly associates each parameter estimate with the corresponding unknown parameter. RCPE is illustrated through several numerical examples, including the Burgers equation.

Nomenclature

G_{f}	=	filter
N_i	=	filter coefficient
$\mathcal{O}_{\mathrm{p}}$	=	permutation matrix
R	=	adaptive integrator gain
u	=	measured input
y	=	measured output
z	=	output error
λ	=	forgetting factor
μ	=	unknown parameter
$\hat{\mu}$	=	parameter estimate
ν	=	parameter pre-estimate
ϕ	=	integrator states
θ	=	parameter estimator coefficient

I. Introduction

IN MANY applications, models of physical systems have known structure but unknown parameters. By viewing the unknown parameters as constant states, nonlinear estimation methods can be used to estimate the states of the augmented system, thereby providing estimates of the parameters along with the dynamic states [1]. The extended Kalman filter (EKF), unscented Kalman filter (UKF), and ensemble Kalman filter (EnKF) can be applied to these problems [2–7]. An alternative approach to parameter estimation is variational methods [8–10], which require an adjoint formulation of the dynamics. These methods tend to be computationally expensive due to the need for multiple iterations of the forward model and backward adjoint.

As a special case of this problem, a linear system may have uncertain entries in its state space representation. Because the parameter states multiply the dynamic states, the resulting estimation dynamics are nonlinear despite the fact that the "original" dynamics are linear. For this problem, a two-step procedure is used in [11], where a black-box model is first constructed based on the input-output data, and a similarity transformation is used to recover the unknown parameters. In [12], a sequential convex relaxation method is used to estimate unknown entries in the matrices of a state space realization.

The present paper focuses on retrospective cost parameter estimation (RCPE), which is a variation of retrospective cost model refinement (RCMR) developed in [13–15] and based on retrospective cost adaptive control [16]. RCPE is applicable to parameter estimation in linear and nonlinear models, where the parameterization may be either affine or nonaffine. To update the parameter estimate, RCPE uses an error signal given by the difference between the output of the physical system and the output of the estimation model. The parameter update is obtained by minimizing a retrospective cost function whose minimizer provides an update of the gains of an integrator. The output of the adaptive integrator consists of the parameter pre-estimates, whose absolute values are the parameter estimates. However, the parameter estimates may be permuted in an unknown way, and thus a permutation is needed to correctly associate each parameter estimate with the corresponding unknown parameter.

Like UKF but unlike EKF, RCPE does not require a Jacobian of the dynamics in order to update the parameter estimates. However, unlike UKF, RCPE does not require an ensemble of models. In particular, for parameter estimation, the unscented Kalman filter (UKF) is based on an ensemble of $2N+1$ models, where $N = n + p$, in which n is the number of dynamic states and p is the number of unknown parameters. The total number of states that must be propagated at each iteration is thus $2N^2 + N = \mathcal{O}(N^2)$. Consequently, for a system with n dynamic states and p unknown parameters, it follows that $N = n + p$, and thus the total number of states that must be propagated at each iteration is $(2N+1)N = [2(n+p)+1](n+p)$. In contrast to UKF, RCPE requires the propagation of only a single copy of the "original" system dynamics so that the number of states that must be propagated at each iteration is simply n. For both UKF and RCPE, this model needs only to be given as an executable simulation; explicit knowledge of the equations and source code underlying the simulation is not required. However, the price paid for not requiring an explicit model or an ensemble of models is the need within RCPE to select a permutation matrix that correctly associates each parameter estimate with the corresponding unknown parameter. Finally, unlike variational methods, RCPE does not require an adjoint model.

The contribution of the current paper is to present, analyze, and demonstrate the RCPE algorithm for estimating multiple unknown parameters in linear and nonlinear systems with affine or nonaffine parameterizations. RCPE is shown to be applicable without explicit knowledge of the system equations, and thus is implementable using only an executable simulation. The paper analyzes the effect of the filter coefficients in determining the search directions leading to the parameter estimates. Most importantly, this paper demonstrates the need for the permutation matrix in problems with multiple unknown parameters. Finally, a numerical example with 101 dynamic states and 2 parameter states shows that the computation required by RCPE

Received 20 October 2018; revision received 23 January 2019; accepted for publication 10 March 2019; published online 25 April 2019.

*Ph.D. Candidate, Aerospace Engineering Department; ankgoel@umich.edu.

†Professor, Aerospace Engineering Department; dsbaero@umich.edu.

Table 1 Summary of the numerical examples

Example	System	Parameterization	l_μ	l_y	Objective
V.1	Linear	Affine	1	1	Effect of $x(0)$ and u on the choice of N_1
V.2	Linear	Nonaffine	1	1	Effect of noise and comparison with UKF
VI.1	Linear	Nonaffine	2	2	Effect of sparse G_{f}
VII.1	Linear	Affine	2	1	Choice of N_1, N_2, and $\mathcal{O}_{\mathrm{p}}$
VII.2	Nonlinear	Nonaffine	2	1	Choice of $\mathcal{O}_{\mathrm{p}}$ with fixed N_1, N_2
VIII.1	Linear	Affine	3	1	Choice of N_1, N_2, N_3 with fixed $\mathcal{O}_{\mathrm{p}}$
VIII.2	Nonlinear	Affine	3	1	Choice of $\mathcal{O}_{\mathrm{p}}$ with fixed N_1, N_2, N_3
IX	Nonlinear	Affine	2	1	High-dimensional application

(202 propagated states) is substantially less than the computation required by UKF (21,321 propagated states).

The present paper extends and refines prior conference publications. In particular, early versions of RCPE were demonstrated in [17–19]. However, the role of the permutation matrix in RCPE is unique to the present paper. Finally, RCPE is distinct from the smoothing technique developed in [20].

The paper is structured as follows. Section II describes the parameter-estimation problem. Section III describes the RCPE algorithm. Next, Sec. IV analyzes the effect of the user-defined filter in RCPE on the performance of the parameter estimator. Sections V–IX present several numerical examples (summarized in Table 1) demonstrating the application of RCPE and its features. Finally, the paper concludes with a discussion of the results and future work.

II. Parameter-Estimation Problem

Consider the discrete-time system

$$x(k+1) = f(x(k), u(k), \mu) + w_1(k) \tag{1}$$

$$y(k) = h(x(k), u(k), \mu) + w_2(k) \tag{2}$$

where $x(k) \in \mathbb{R}^{l_x}$ is the state, $u(k) \in \mathbb{R}^{l_u}$ is the measured input, $y(k) \in \mathbb{R}^{l_y}$ is the measured output, $w_1(k) \in \mathbb{R}^{l_x}$ is the process noise, $w_2(k) \in \mathbb{R}^{l_y}$ is the measurement noise, and $\mu = [\mu_1 \ \cdots \ \mu_{l_\mu}]^T \in \mathcal{M} \subseteq \mathbb{R}^{l_\mu}$ is the *true parameter*, which is unknown. The set $\mathcal{M}$ is assumed to be known and satisfy $\mathcal{M} \subseteq [0, \infty)^{l_\mu}$; that is, $\mathcal{M}$ is contained in the nonnegative orthant. If $\mathcal{M}$ does not satisfy this condition, then it may be possible to replace $\mathcal{M}$ by $\mathcal{M}' \triangleq \bar{\mu} + \mathcal{M}$ and μ by $\mu - \bar{\mu}$ in (1), (2), where $\bar{\mu} \in \mathbb{R}^{l_\mu}$ shifts $\mathcal{M}$ such that $\mathcal{M}'$ is contained in the nonnegative orthant. With this transformation, which can always be done if $\mathcal{M}$ is bounded, it can be assumed that μ is an element of the nonnegative orthant. The system (1), (2) is viewed as the *truth model* of a physical system.

Based on (1), (2), the *estimation model* is constructed as

$$\hat{x}(k+1) = f(\hat{x}(k), u(k), \hat{\mu}(k)) \tag{3}$$

$$\hat{y}(k) = h(\hat{x}(k), u(k), \hat{\mu}(k)) \tag{4}$$

where $\hat{x}(k)$ is the computed state, $\hat{y}(k)$ is the computed output of (3), (4), and $\hat{\mu}(k)$ is the *parameter estimate*. It is assumed that f and h are known, and thus they can be used to construct (3), (4). Since $w_1(k)$ and $w_2(k)$ are unknown, they do not appear in (3), (4). Since μ is unknown, it is replaced by $\hat{\mu}(k)$ in (3), (4). The objective is to construct $\hat{\mu}(k)$ based on the *output error* $z(k) \in \mathbb{R}^{l_y}$ defined by

$$z(k) \triangleq \hat{y}(k) - y(k) \tag{5}$$

The ability to estimate μ is based on the assumption that (1), (2) is structurally identifiable [21–23] and the data are sufficiently persistent [24,25].

Since measurements of only y are available, the state x is unknown, and thus $x(0)$ is unknown. For all examples in this paper, the initial state of the estimation model (3), (4) is chosen to be zero to reflect the absence of additional modeling information. However, the initial state of (1), (2) is unknown and nonzero.

Definition II.1: The system (1), (2) is *affinely parameterized* if there exist functions $f_0, f_1, \ldots, f_{l_\mu}$ and $h_0, h_1, \ldots, h_{l_\mu}$ such that

$$f(x, u, \mu) = f_0(x, u) + \sum_{i=1}^{l_\mu} \mu_i f_i(x, u) \tag{6}$$

$$h(x, u, \mu) = h_0(x, u) + \sum_{i=1}^{l_\mu} \mu_i h_i(x, u) \tag{7}$$

Otherwise, (1), (2) is nonaffinely parameterized.

A specialization of (1), (2) is given by the linear discrete-time system

$$x(k+1) = A(\mu)x(k) + B(\mu)u(k) + w_1(k) \tag{8}$$

$$y(k) = C(\mu)x(k) + D(\mu)u(k) + w_2(k) \tag{9}$$

In this case, the estimation model (3), (4) becomes

$$\hat{x}(k+1) = A(\hat{\mu}(k))\hat{x}(k) + B(\hat{\mu}(k))u(k) \tag{10}$$

$$\hat{y}(k) = C(\hat{\mu}(k))\hat{x}(k) + D(\hat{\mu}(k))u(k) \tag{11}$$

Definition II.2: The linear system (8), (9) is *affinely* parameterized if there exist constant matrices $A_0, A_1, \ldots, A_{l_\mu} \in \mathbb{R}^{l_x \times l_x}$, B_0, $B_1, \ldots, B_{l_\mu} \in \mathbb{R}^{l_x \times l_u}$, $C_0, C_1, \ldots, C_{l_\mu} \in \mathbb{R}^{l_y \times l_x}$, and $D_0, D_1, \ldots, D_{l_\mu} \in \mathbb{R}^{l_y \times l_u}$ such that

$$A(\mu) = A_0 + \sum_{i=1}^{l_\mu} \mu_i A_i, \qquad B(\mu) = B_0 + \sum_{i=1}^{l_\mu} \mu_i B_i \tag{12}$$

$$C(\mu) = C_0 + \sum_{i=1}^{l_\mu} \mu_i C_i, \qquad D(\mu) = D_0 + \sum_{i=1}^{l_\mu} \mu_i D_i \tag{13}$$

Otherwise, (8), (9) is nonaffinely parameterized.

III. Retrospective Cost Parameter Estimation

This section presents RCPE. RCPE uses the estimation model (3), (4) along with a parameter estimator to construct $\hat{\mu}(k)$. The parameter estimator constructs $\hat{\mu}(k)$ by minimizing a cost function based on the output error z.

A. Parameter Estimator

The *parameter estimator* consists of an adaptive integrator and an output nonlinearity. In particular, the *parameter pre-estimate* $\boldsymbol{\nu}$ is given by

$$\nu(k) = R(k)\phi(k) \tag{14}$$

where the *integrator state* $\phi(k) \in \mathbb{R}^{l_y}$ is updated by

$$\phi(k) = \phi(k-1) + z(k-1) \tag{15}$$

The *adaptive integrator gain* $R(k) \in \mathbb{R}^{l_\mu \times l_y}$ is updated by RCPE as described later in this section. Since $\nu(k)$ is not necessarily an element

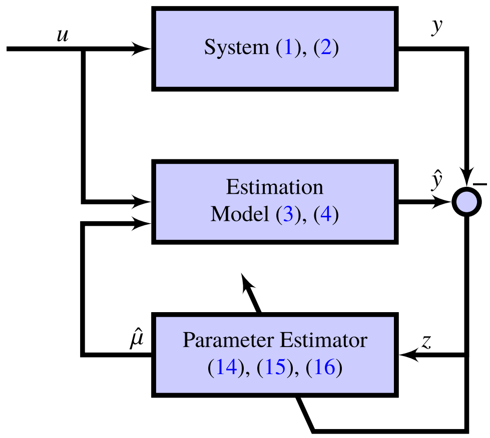


Fig. 1 Retrospective cost parameter estimation.

of the nonnegative orthant, an output nonlinearity is used to transform $\nu(k)$. In particular, the parameter estimate $\hat{\mu}(k)$ is given by

$$\hat{\mu}(k) = \mathcal{O}_{\mathrm{p}}|\nu(k)| \tag{16}$$

where the absolute value is applied componentwise. The matrix $\mathcal{O}_{\mathrm{p}}$ is explained below. The parameter estimator, which consists of Eqs. (14–16), is represented in Fig. 1. Since $z(k) \to 0$ is a necessary condition for ϕ to converge, the integrator (15) allows z to converge to zero while ϕ converges to a finite value. Consequently, the parameter pre-estimate ν given by Eq. (14) can converge to a nonzero value, which, in turn, allows the parameter estimate $\hat{\mu}$, given by Eq. (16), to converge to μ.

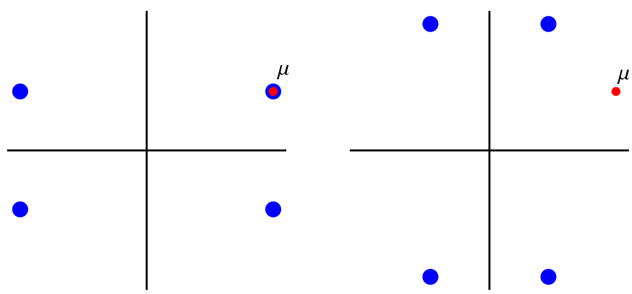
Let the l_μ-tuple $\mathrm{p} = (i_1, \ldots, i_{l_\mu})$ denote a permutation of $(1, \ldots, l_\mu)$. Then the matrix $\mathcal{O}_{\mathrm{p}} \in \mathbb{R}^{l_\mu \times l_\mu}$ maps $(1, \ldots, l_\mu)$ to $(i_1, \ldots, i_{l_\mu})$. Since $\mathcal{O}_{\mathrm{p}}$ is a permutation matrix, each of its rows and columns contains exactly one "1" and the remaining entries are all zero. Specifically, row j of $\mathcal{O}_{\mathrm{p}}$ is row i_j of the identity matrix I_{l_μ}. Now, define the set

$$\mathcal{S}_{\mathcal{O}_{\mathrm{p}}} \triangleq \{s \in \mathbb{R}^{l_\mu}\colon \mathcal{O}_{\mathrm{p}}|s| = \mu\} \tag{17}$$

whose elements are the vectors that are mapped to μ by the componentwise absolute value and the permutation O_{p}. For illustration, Fig. 2a shows the elements of $\mathcal{S}_{\mathcal{O}_{12}}$, and Fig. 2b shows the elements of $\mathcal{S}_{\mathcal{O}_{21}}$.

To facilitate the subsequent development, note that the parameter pre-estimate (14) can be rewritten as

$$\nu(k) = \Phi(k)\theta(k) \tag{18}$$



a) p = 12 **b) p = 21**

Fig. 2 The set $\mathcal{S}_{\mathcal{O}_{\mathrm{p}}}$ for $l_\mu = 2$ consists of the blue dots; μ is shown in red.

where the regressor matrix $\Phi(k)$ is defined by

$$\Phi(k) \triangleq I_{l_\mu} \otimes \phi^T(k) \in \mathbb{R}^{l_\mu \times l_\theta} \tag{19}$$

and the coefficient vector $\theta(k)$ is defined by

$$\theta(k) \triangleq \mathrm{vec} R(k) \in \mathbb{R}^{l_\theta} \tag{20}$$

where $l_\theta \triangleq l_\mu l_y$, "$\otimes$" is the Kronecker product, and "vec" is the column-stacking operator. Note that $\theta(k)$ is an alternative representation of the adaptive integrator gain $R(k)$.

B. Retrospective Cost Optimization

The retrospective error variable is defined by

$$\hat{z}(k, \hat{\theta}) \triangleq z(k) + G_{\mathrm{f}}(\boldsymbol{q})[\Phi(k)\hat{\theta} - \nu(k)] \tag{21}$$

where $\boldsymbol{q}$ is the forward-shift operator and $\hat{\theta} \in \mathbb{R}^{l_\theta}$ is determined by optimization to obtain the updated coefficient vector $\theta(k+1)$ The filter G_{f} has the form

$$G_{\mathrm{f}}(\boldsymbol{q}) = \sum_{i=1}^{n_f} \frac{1}{\boldsymbol{q}^i} N_i \tag{22}$$

where $N_1, \ldots, N_{n_f} \in \mathbb{R}^{l_y \times l_\mu}$ are the *filter coefficients*. Note that G_{f} is an $l_y \times l_\mu$ finite impulse response filter. The retrospective error variable (21) can thus be rewritten as

$$\hat{z}(k, \hat{\theta}) = z(k) + N\bar{\Phi}(k)\hat{\theta} - N\bar{V}(k) \tag{23}$$

where

$$N \triangleq [N_1 \quad \cdots \quad N_{n_f}] \in \mathbb{R}^{l_y \times n_f l_\mu} \tag{24}$$

$$\bar{\Phi}(k) \triangleq \begin{bmatrix} \Phi(k-1) \\ \vdots \\ \Phi(k-n_f) \end{bmatrix} \in \mathbb{R}^{l_\mu n_f \times l_\theta}, \qquad \bar{V}(k) \triangleq \begin{bmatrix} \nu(k-1) \\ \vdots \\ \nu(k-n_f) \end{bmatrix} \in \mathbb{R}^{l_\mu n_f} \tag{25}$$

The retrospective cost function is defined by

$$J(k, \hat{\theta}) \triangleq \sum_{i=1}^{k} \lambda^{k-i} \hat{z}(i, \hat{\theta})^T \hat{z}(i, \hat{\theta}) + \lambda^k \hat{\theta}^T R_\theta \hat{\theta} \tag{26}$$

where $R_\theta \in \mathbb{R}^{l_\theta \times l_\theta}$ is positive definite and $\lambda \in (0, 1]$ is the forgetting factor. The following result uses recursive least squares (RLS) to minimize (26).

Proposition III.1: Let $P(0) = R_\theta^{-1}$ and $\theta(0) = 0$. Then, for all $k \geq 1$, the retrospective cost function (26) has a unique global minimizer $\theta(k+1)$, which is given by

$$P(k+1) = \lambda^{-1}[P(k) - P(k)\bar{\Phi}(k)^T N^T \Gamma(k)^{-1} N\bar{\Phi}(k)P(k)] \tag{27}$$

$$\theta(k+1) = \theta(k) - P(k+1)\bar{\Phi}(k)^T N^T [N\bar{\Phi}(k)\theta(k) + z(k) - N\bar{V}(k)] \tag{28}$$

where

$$\Gamma(k) \triangleq \lambda I_{l_y} + N\bar{\Phi}(k)P(k)\bar{\Phi}(k)^T N^T \tag{29}$$

Furthermore, the parameter estimate at step $k+1$ is given by

$$\hat{\mu}(k+1) = \mathcal{O}_{\rm p}|\nu(k+1)| = \mathcal{O}_{\rm p}|\Phi(k+1)\theta(k+1)| \tag{30}$$

Since $\theta(0)=0$, it follows that $\nu(0)=0$ and thus $\hat{\mu}(0)=0$.

IV. Analysis of RCPE

This section analyzes the role of the filter $G_{\rm f}$ in the update of the parameter pre-estimate ν. In particular, it is shown that the filter coefficients determine the subspace of $\mathbb{R}^{l_\mu}$ that contains ν.

To analyze the role of $G_{\rm f}$, the cost function (26) is rewritten as

$$J(k,\hat{\theta}) = \hat{\theta}^T A_\theta(k)\hat{\theta} + 2b_\theta(k)^T\hat{\theta} + c_\theta(k) \tag{31}$$

where

$$A_\theta(k) \triangleq \sum_{i=1}^{k} \lambda^{k-i}\bar{\Phi}(i)^T N^T N\bar{\Phi}(i) + \lambda^k R_\theta \tag{32}$$

$$b_\theta(k) \triangleq \sum_{i=1}^{k} \lambda^{k-i}\bar{\Phi}(i)^T N^T (z(i) - N\bar{V}(i)) \tag{33}$$

$$c_\theta(k) \triangleq \sum_{i=1}^{k} \lambda^{k-i}(z(i) - N\bar{V}(i))^T(z(i) - N\bar{V}(i)) \tag{34}$$

At step k, the batch least squares minimizer $\theta(k+1)$ of Eq. (26) is given by

$$\theta(k+1) = -A_\theta(k)^{-1}b_\theta(k) \tag{35}$$

which is equal to $\theta(k+1)$, given by Eq. (28).

The following result shows that the parameter pre-estimate $\nu(k)$, and thus the estimate $\hat{\mu}(k)$, is constrained to lie in a subspace determined by the coefficients of $G_{\rm f}$.

Lemma IV.1: Let $\beta > 0$, $R_\theta = \beta I_{l_\theta}$, $\nu(k)$ be given by Eq. (18), $\Phi(k)$ be given by Eq. (19), $N, \overline{\Phi}(k), \overline{V}(k)$ be given by Eq. (24) and (25), and $\theta(k+1)$ be given by Eq. (35). Then, for all $k \ge 1$,

$$\begin{aligned}\nu(k+1) = &-\frac{1}{\beta}[N_1^T \ \cdots \ N_{n_f}^T]\sum_{i=1}^{k}\lambda^{-i}\Psi(k,i)\\ &\times[z(i) + N\bar{\Phi}(i)\theta(k+1) - N\bar{V}(i)] \in \mathcal{R}([N_1^T \ \cdots \ N_{n_f}^T])\end{aligned} \tag{36}$$

where

$$\Psi(k,i) \triangleq \begin{bmatrix} \phi(k+1)^T\phi(i-1)\otimes I_{l_y} \\ \vdots \\ \phi(k+1)^T\phi(i-n_f)\otimes I_{l_y}\end{bmatrix} \tag{37}$$

Proof. Note that

$$\begin{aligned}\Phi(k+1)&A_\theta(k)\theta(k+1)\\ &= \sum_{i=1}^{k}\lambda^{k-i}\Phi(k+1)(\bar{\Phi}(i)^T N^T N\bar{\Phi}(i))\theta(k+1)\\ &+ \lambda^k\beta\Phi(k+1)\theta(k+1)\\ &= \sum_{i=1}^{k}\left(\lambda^{k-i}\sum_{j=1}^{n_f}(I_{l_u}\otimes\phi(k+1)^T)(I_{l_u}\otimes\phi(i-j))N_j^T\right)\\ &\times N\bar{\Phi}(i)\theta(k+1) + \lambda^k\beta\Phi(k+1)\theta(k+1)\\ &= \sum_{i=1}^{k}\left(\lambda^{k-i}\sum_{j=1}^{n_f}N_j^T\phi(k+1)^T\phi(i-j)\right)N\bar{\Phi}(i)\theta(k+1)\\ &+ \lambda^k\beta\Phi(k+1)\theta(k+1)\\ &= [N_1^T \ \cdots \ N_{n_f}^T]\sum_{i=1}^{k}\lambda^{k-i}\Psi(k,i)N\bar{\Phi}(i)\theta(k+1)\\ &+ \lambda^k\beta\Phi(k+1)\theta(k+1)\end{aligned} \tag{38}$$

and

$$\begin{aligned}\Phi(k+1)b_\theta(k) &= \Phi(k+1)\sum_{i=1}^{k}\lambda^{k-i}\bar{\Phi}(i)^T N^T(z(i) - N\bar{V}(i))\\ &= [N_1^T \ \cdots \ N_{n_f}^T]\sum_{i=1}^{k}\lambda^{k-i}\Psi(k,i)(z(i) - N\bar{V}(i))\end{aligned} \tag{39}$$

Writing Eq. (35) as $A_\theta(k)\theta(k+1) = -b_\theta(k)$, multiplying by $\Phi(k+1)$, and using Eqs. (38) and (39) yield (36).

It follows from Lemma IV.1 that the parameter pre-estimate ν is constrained to lie in the subspace of $\mathbb{R}^{l_\mu}$ spanned by the coefficients of the filter used by RCPE. In addition to the subspace constraint, the numerical examples in Secs. V–VIII show that the feasible region is determined by the choice of the filter coefficients. The *feasible region* is the set of parameter pre-estimates in $\mathbb{R}^{l_\mu}$ that are asymptotically reachable by the estimator. Consequently, the permutation matrix $\mathcal{O}_{\rm p}$ must be chosen such that at least one element of $\mathcal{S}_{\mathcal{O}_{\rm p}}$, defined in Eq. (17), lies in the feasible region.

In view of Lemma IV.1, for all examples in this paper where $l_z = 1$, n_f is set to be equal to l_μ, and each filter coefficient is chosen to be an element of $\{e_1, e_2, \ldots, e_{l_\mu}\}$, where e_i is the ith row of the identity matrix I_{l_μ}. For $l_z > 1$, the filter coefficients must be selected such that $\mu \in \mathcal{R}([N_1^T \ \cdots \ N_{n_f}^T])$.

V. Examples with $l_\mu = l_y = 1$

In this section, RCPE is used to estimate one parameter in an affinely and nonaffinely parameterized linear systems.

Example V.1: Affinely parameterized linear dynamics with one unknown parameter in the dynamics matrix. This example shows the effect of u, $x(0)$, and N_1 on the feasible region. Consider the linear system (8), (9), where

$$A(\mu) = \begin{bmatrix}\mu & 0.2\\ 0.1 & 0.6\end{bmatrix}, \quad B = \begin{bmatrix}0.9\\ 0.3\end{bmatrix}, \quad C = [1.1 \quad 0.5] \tag{40}$$

and $\mu = 0.3$. The initial state is $x(0) = [10 \quad 10]^T$, and the input is

$$u(k) = 2 + \sum_{j=1}^{15}\sin\left(\frac{2\pi j}{100}k\right) \tag{41}$$

$N_1 = 1$, $\lambda = 1$, and $R_\theta = 10^6$. Furthermore, $\mathcal{O}_1 = 1$, and thus $\mathcal{S}_1 = \{\pm\mu\}$. Figure 3 shows the output error, true parameter, parameter pre-estimate, parameter estimate, state-estimate error, and

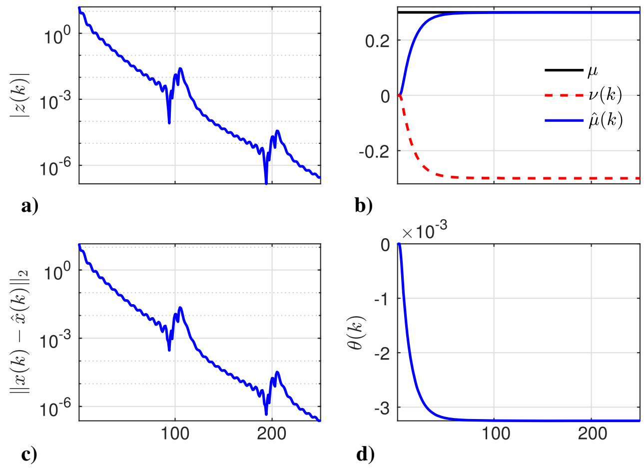


Fig. 3 Example V.1: a) output error; b) true parameter, parameter pre-estimate, and parameter estimate; c) state-estimate error; d) parameter estimator coefficient.

estimator coefficient. (b) shows that $\nu(k)$ converges to $-\mu$, and thus, by Eq. (16), $\hat{\mu}(k) = |\nu(k)|$ converges to μ. Unless stated otherwise, the abscissa of all plots denotes the iteration step.

To investigate the effect of u and N_1 on the feasible region, μ is estimated with the input αu, where u is given by Eq. (41), $\alpha = \pm 1$, and $N_1 = \pm 1$. For all four cases, Fig. 4 shows the true parameter, parameter pre-estimate, and parameter estimate. Note that, for a given input u and filter coefficient N_1, $\nu(k)$ converges to either μ or $-\mu$, and thus $\hat{\mu}(k)$ converges to μ.

Next, to investigate the effect of the initial conditions of Eqs. (8) and (9) on the performance of RCPE, μ is estimated with $x_1(0)$ and $x_2(0)$ varied from -100 to 100. The input $u(k)$ is given by Eq. (41) in all cases. Each point in Figs. 5a and 5b corresponds to an initial condition of Eqs. (8) and (9), where green indicates that $\nu(k)$ converges to μ, blue indicates that $\nu(k)$ converges to $-\mu$, and red indicates that $\nu(k)$ diverges. Note that all cases are obtained by running RCPE under the same values of λ and R_θ; however, the set of initial conditions $x(0)$ for which $\nu(k)$ converges can be expanded by varying these parameters. Figures 4 and 5 suggest that, for the given input u and filter coefficient N_1, the feasible region is either $(-\infty, 0]$

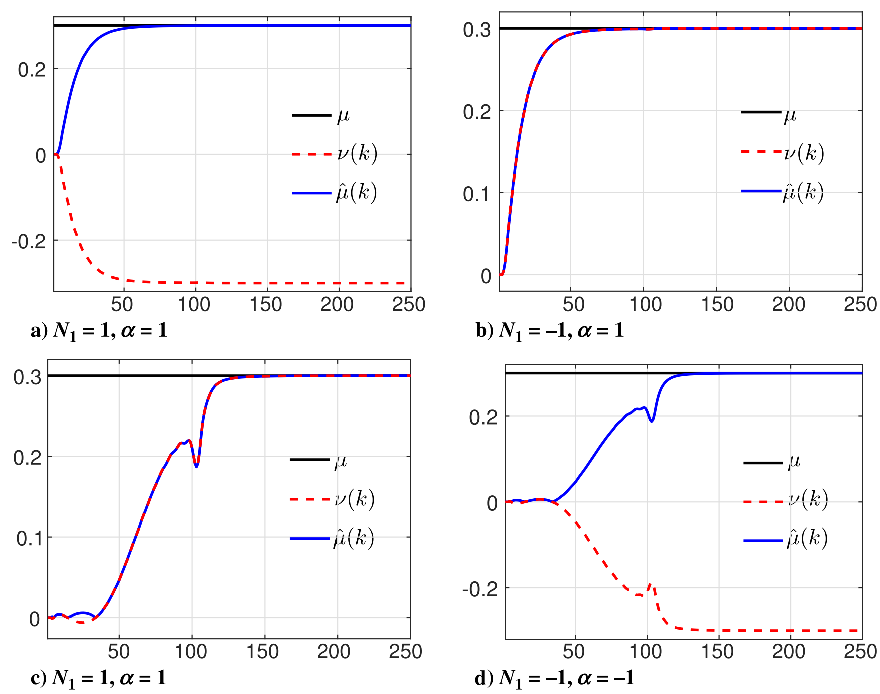


Fig. 4 Example V.1. Parameter pre-estimate $\nu(k)$ and the estimate $\hat{\mu}(k)$ for various choices of the filter coefficient N_1 and the input u determined by the parameter α.

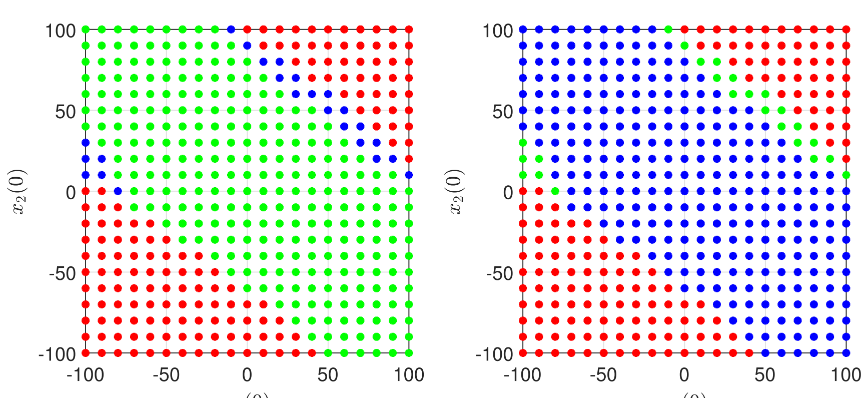


Fig. 5 Example V.1. Convergence of the parameter estimate for a grid of initial conditions of (8), (9).

or $[0, \infty)$, and thus cannot be determined a priori. Consequently, Eq. (16) ensures that there exists $s \in \mathcal{S}_{\mathcal{O}_p}$ in the feasible region such that $\nu(k)$ converges to s, and thus $\hat{\mu}(k)$ converges to μ. □

Example V.2: Nonaffinely parameterized linear dynamics with one unknown parameter in the dynamics matrix. This example investigates the effect of noise, and compares the performance of RCPE with UKF. Consider the linear system (8), (9), where

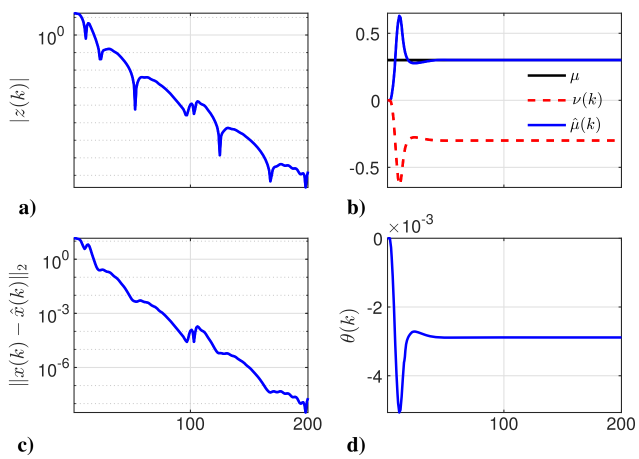


Fig. 6 Example V.2: a) output error; b) true parameter, parameter pre-estimate, and parameter estimate; c) state-estimate error; d) parameter estimator coefficient.

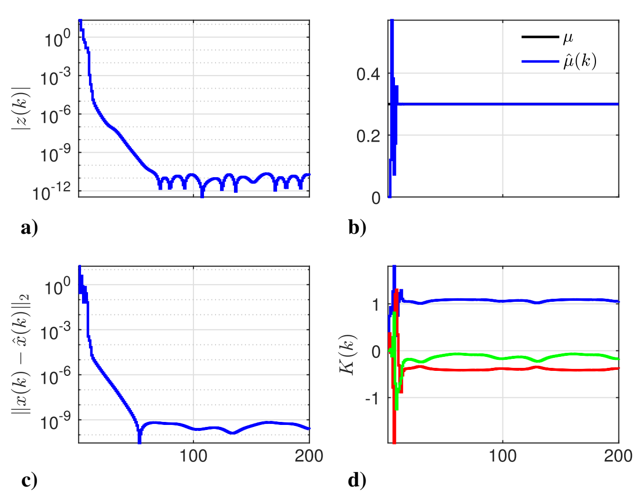


Fig. 7 Example V.2. UKF-based parameter estimation: a) output error; b) true parameter and parameter estimate; c) state-estimate error; d) components of the UKF-based estimator gain.

$$A(\mu) = \begin{bmatrix} \sin\mu & \cos\frac{\mu}{3} \\ \frac{e^{-\mu}}{3} & \frac{0.5}{1.1+\mu^2} \end{bmatrix} \tag{42}$$

$\mu = 0.3$, and B and C are given by Eq. (40). The initial state is $x(0) = [10 \quad 10]^T$, $u(k)$ is given by Eq. (41), $N_1 = 1$, $\lambda = 1$, and $R_\theta = 10^6$. Furthermore, $\mathcal{O}_1 = 1$, and thus $\mathcal{S}_1 = \{\pm\mu\}$. Figure 6 shows the output error, true parameter, parameter pre-estimate, parameter estimate, state-estimate error, and estimator coefficient. (b) shows that $\nu(k)$ converges to $-\mu$.

Next, μ is estimated using UKF with $\alpha = 1.01$, $P(0)$diag$(10^3, 10^3, 1)$, $Q = 10^{-10} I_2$, and $R = 10^{-10}$, where α affects the distribution of sigma points, and P, Q, R are the augmented state, process, and measurement covariance matrices. Figure 7 shows the output error, true parameter, parameter pre-estimate, parameter estimate, state-estimate error, and estimator coefficient. Note that the UKF gain $K(k)$ does not converge.

Next, to compare the accuracy of RCPE and UKF in the presence of noise, μ is estimated with process noise $w_1 \sim \mathcal{N}(0, \sigma_1^2 I_2)$ and measurement noise $w_2 \sim \mathcal{N}(0, \sigma_2^2)$. For RCPE, $N_1 = 1$, $\lambda = 1$, and $R_\theta = 10^6$; for UKF, $\alpha = 1.01$, $P(0) = \mathrm{diag}(10^3, 10^3, 1)$, $Q = \sigma_1 I_2$, and $R = \sigma_2$. For a range of values of σ_1 and σ_2, Fig. 8 shows

$$\varepsilon_\mu \triangleq \frac{1}{100}\sqrt{\sum_{i=901}^{1000} (\hat{\mu}(i) - \mu)^2} \tag{43}$$

Note that, unlike UKF, RCPE uses no knowledge of the noise statistics Q and R to compute $\hat{\mu}$. □

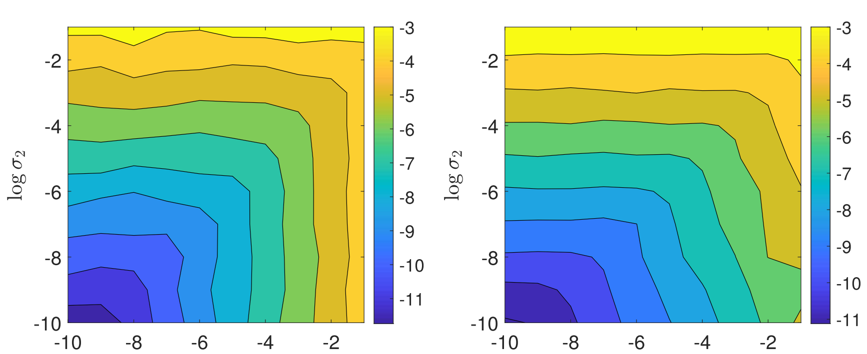
VI. Example with $l_\mu = l_y = 2$

In this section, RCPE is used to estimate two unknown parameters in linear systems that are affinely and nonaffinely parameterized with two measurements.

Example VI.1: Nonaffinely parameterized linear dynamics with two measurements and two unknown parameters. This example shows how RCPE can be implemented with a sparse $R(k)$. Consider the linear system (8), (9), where

$$A(\mu) = \begin{bmatrix} \sin\mu_1 & \cos\frac{\mu_1}{3} \\ \frac{e^{-\mu_1}}{3} & \frac{0.5}{1.1+\mu_1^2} \end{bmatrix}, \qquad B(\mu) = \begin{bmatrix} \log(1+\mu_2^2) \\ 1 + \sin\mu_2 \end{bmatrix},$$
$$C(\mu) = \begin{bmatrix} \mu_2 & 4\mu_1^2 \\ \sin\mu_1 & 2\mu_2 \end{bmatrix} \tag{44}$$

and $\mu = [\mu_1 \quad \mu_2]^T = [0.4 \quad 0.2]^T$. The initial state is $x(0) = [10 \quad 10]^T$, $u(k)$ is given by Eq. (41), $N_1 = I_2$, $\lambda = 1$, and $R_\theta = 10^8 I_4$. Furthermore, p = 12, and thus $\mathcal{S}_{\mathcal{O}_p} = \{[\pm\mu_1 \pm \mu_2]^T\}$. Note that $R(k) \in \mathbb{R}^{2\times 2}$, and thus the estimates of μ_1 and μ_2 are determined by both z_1 and z_2. Figure 9 shows the output error,



a) RCPE b) UKF

Fig. 8 Example V.2. Effect of noise on the estimation accuracy of a) RCPE and b) UKF; the color scale denotes values of $\log\varepsilon_\mu$ as a function of σ_1 and σ_2.

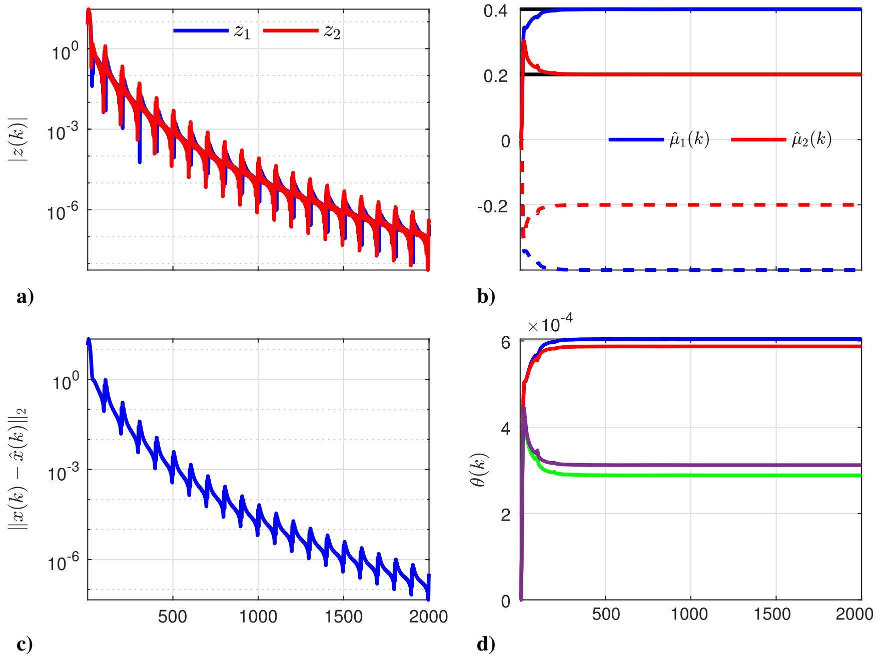


Fig. 9 Example VI.1: a) output error; b) parameter estimates; c) state-estimation error; d) parameter estimator coefficients.

true parameter, parameter pre-estimate, parameter estimate, state-estimate error, and estimator coefficient. (b) shows that $\nu(k)$ converges to $-\mu$.

Next, the adaptive integrator (14) is constrained such that

$$\nu(k) = \begin{bmatrix} R_{11}(k) & 0 \\ 0 & R_{22}(k) \end{bmatrix} \phi(k) \tag{45}$$

Since $R(k)$ is sparse, it follows that μ_1 is determined by z_1 only, and μ_2 is determined by z_2 only. Furthermore, $p = 12$, $N_1 = I_2$, $\lambda = 1$, and $R_\theta = 10^8$. Figure 10 shows the output error, true parameter, parameter pre-estimate, parameter estimate, state-estimate error, and estimator coefficient. □

VII. Examples with $l_\mu = 2$ and $l_y = 1$

In this section, RCPE is used to estimate two unknown parameters in affinely and nonaffinely parameterized systems with one measurement. These examples show that the feasible region is determined by the choice and ordering of the filter coefficients.

Example VII.1: Affinely parameterized linear dynamics with two unknown parameters in the dynamics matrix. This example investigates the effect of N_1, N_2, and $\mathcal{O}_\mathrm{p}$ on the feasible region. Consider the linear system (8), (9), where

$$A(\mu) = \begin{bmatrix} \mu_1 & \mu_2 \\ 0.1 & 0.6 \end{bmatrix} \tag{46}$$

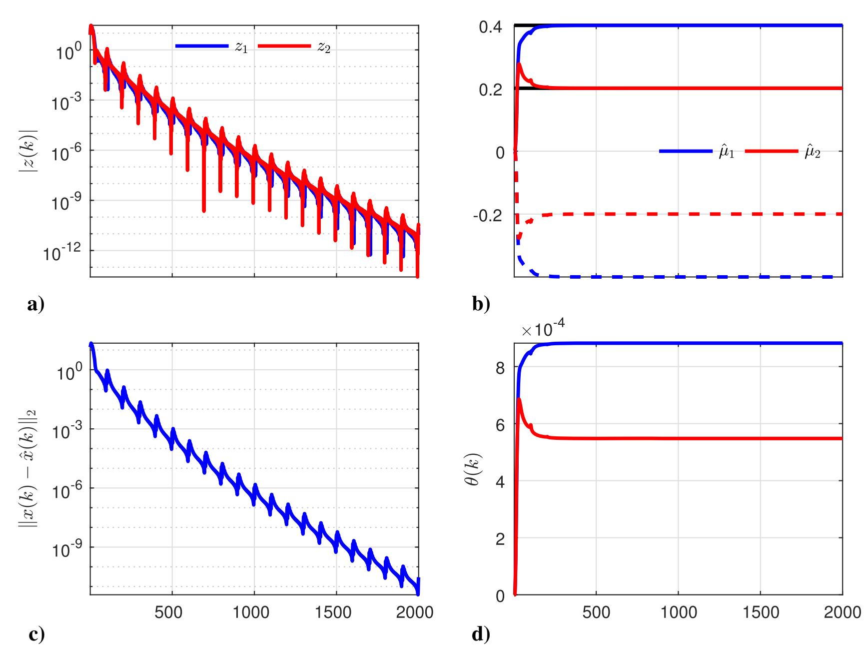


Fig. 10 Example VI.1. Sparse integrator: a) output error; b) parameter estimates; c) state-estimation error; d) parameter estimator coefficients.

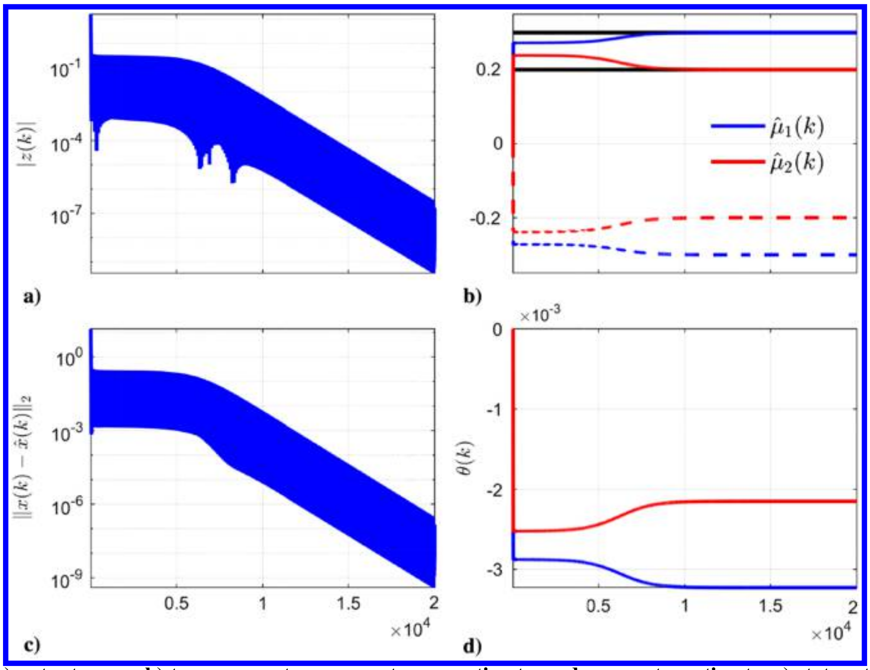


Fig. 11 Example VII.1: a) output error; b) true parameter, parameter pre-estimate, and parameter estimate; c) state-estimate error; d) parameter estimator coefficients.

the input and output matrices are given by Eq. (40), and $\mu = [\mu_1 \quad \mu_2]^T = [0.3 \quad 0.2]^T$. The initial state is $x(0) = [10 \quad 10]^T$, $u(k)$ is given by Eq. (41), $N_1 = e_1$, $N_2 = e_2$, $\lambda = 0.999$, and $R_\theta = 10^6 I_2$. Furthermore, $p = 12$, and thus $\mathcal{S}_{\mathcal{O}_{12}} = \{[\pm\mu_1 \pm \mu_2]^T\}$. Figure 11 shows the output error, true parameter, parameter pre-estimate, parameter estimate, state-estimate error, and estimator coefficient. (b) shows that $\nu(k)$ converges to $-\mu$.

Next, the effect of the choice of $\mathcal{O}_\mathrm{p}$ and N_1 and N_2 is investigated. For $l_\mu = 2$, there are two choices of $\mathcal{O}_\mathrm{p}$ and two ways to order the filter coefficients e_1 and e_2. Further, for each ordering, there are four ways to allocate signs. Table 2 shows all such filter choices. Figure 12a shows $\nu(k)$ for p = 12, and Fig. 12b shows $\nu(k)$ for $p = 21$, where the corresponding filter coefficients are given in Table 2. Note that, for a fixed ordering of the filter coefficients, there is exactly one permutation matrix $\mathcal{O}_\mathrm{p}$ such that the parameter pre-estimate ν converges to an element of $\mathcal{S}_{\mathcal{O}_\mathrm{p}}$. Conversely, for a fixed permutation matrix $\mathcal{O}_\mathrm{p}$, there is exactly one ordering of the filter coefficients such that the parameter pre-estimate ν converges to an element of $\mathcal{S}_{\mathcal{O}_\mathrm{p}}$. □

Example VII.2: Nonaffinely parameterized nonlinear dynamics with two unknown parameters. This example investigates the effect of N_1, N_2, and $\mathcal{O}_\mathrm{p}$ on the feasible region. Consider the (3, 3) type nonlinear system ([26] p. 183)

$$x(k+1) = \begin{bmatrix} x_2(k) \\ \frac{1+0.8x_2(k)+x_1(k)}{1+\mu_1 x_2(k)+\mu_2 x_1(k)} \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(k) \tag{47}$$

$$y(k) = x_1(k) \tag{48}$$

where $\mu = [\mu_1 \quad \mu_2]^T = [0.6 \quad 1.1]^T$. The initial state is $x(0) = [10 \quad 10]^T$, $u(k)$ is given by

$$u(k) = 2 + \sum_{j=1}^{5} \frac{1}{j} \sin\left(\frac{2\pi j}{100} k + j^2\right) \tag{49}$$

$N_1 = e_2$, $N_2 = e_1$, $\lambda = 0.999$, and $R_\theta = 10^{12} I_2$. Furthermore, $p = 21$, and thus $\mathcal{S}_{\mathcal{O}_\mathrm{p}} = \{[\pm\mu_2 \pm \mu_1]^T\}$. Figure 13 shows the output

Table 2 Filter coefficients for Example VII.1

	N_1	N_2		N_1	N_2
G_{f1}	e_1	e_2	G_{f5}	e_2	e_1
G_{f2}	$-e_1$	e_2	G_{f6}	$-e_2$	e_1
G_{f3}	e_1	$-e_2$	G_{f7}	e_2	$-e_1$
G_{f4}	$-e_1$	$-e_2$	G_{f8}	$-e_2$	$-e_1$

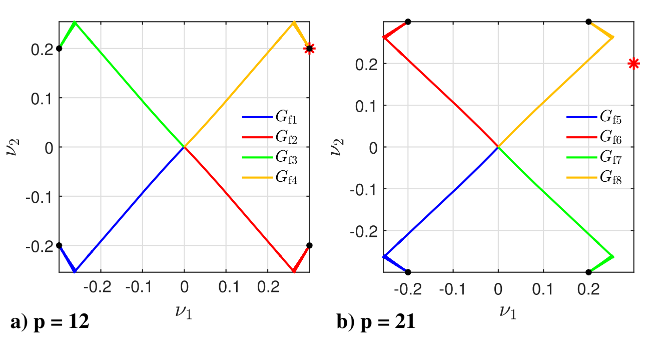


Fig. 12 Example VII.1. Parameter pre-estimate $\nu(k)$ for various choices of G_f. The true parameter μ is shown in red.

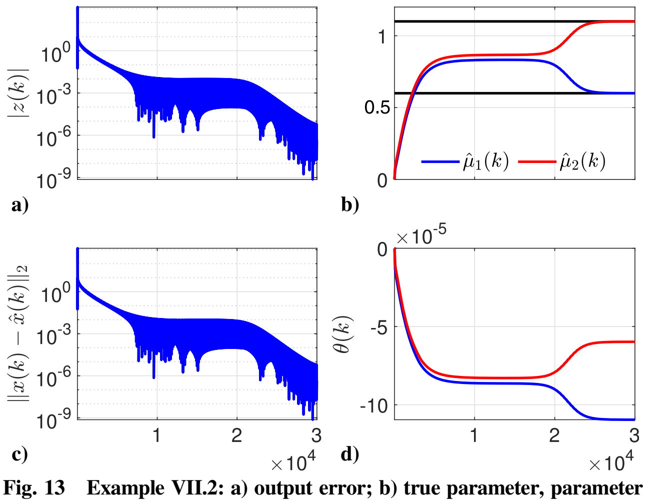


Fig. 13 Example VII.2: a) output error; b) true parameter, parameter pre-estimate, and parameter estimate; c) state-estimate error; d) parameter estimator coefficients.

error, true parameter, parameter pre-estimate, parameter estimate, state-estimate error, and estimator coefficient. (b) shows that $\nu(k)$ converges to $\mathcal{O}_{\rm p}^{-1}\mu$. Analogous results shown in Fig. 12 are obtained for other choices of the filter coefficients and permutation matrix $\mathcal{O}_{\rm p}$. □

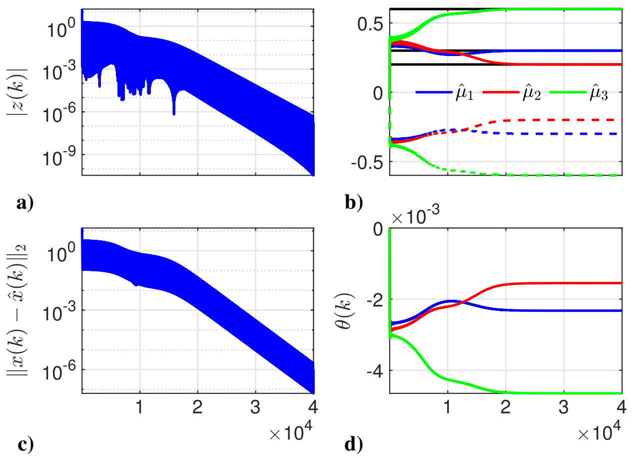


Fig. 14 Example VIII.1: a) output error; b) true parameter, parameter pre-estimate, and parameter estimate; c) state-estimate error; d) parameter estimator coefficients.

VIII. Examples with $l_\mu = 3$ and $l_y = 1$

In this section, RCPE is used to estimate three unknown parameters in affinely and nonaffinely parameterized systems with one measurement.

Example VIII.1: Affinely parameterized linear dynamics with three unknown parameters in the dynamics matrix. This example investigates the effect of N_1, N_2, N_3, and $\mathcal{O}_{\rm p}$ on the feasible region. Consider the linear system (8), (9), where

$$A(\mu) = \begin{bmatrix} \mu_1 & \mu_2 \\ 0.1 & \mu_3 \end{bmatrix} \tag{50}$$

the input and output matrices are given by Eq. (40), and $\mu = [\mu_1 \quad \mu_2 \quad \mu_3]^T = [0.3 \quad 0.2 \quad 0.6]^T$. The initial state is $x(0) = [10 \quad 10 \quad 10]^T$, $u(k)$ is given by Eq. (41), $N_1 = e_3$, $N_2 = e_2$, $N_3 = e_1$, $\lambda = 0.999$, and $R_\theta = 10^6 I_2$. Furthermore, $p = 123$, and thus $\mathcal{S}_{\mathcal{O}_{\rm p}} = \{[\pm\mu_1 \pm \mu_2 \pm \mu_3]^T\}$. Figure 14 shows the output error, true parameter, parameter pre-estimate, parameter estimate, state-estimate error, and estimator coefficient. (b) shows that $\nu(k)$ converges to $-\mu$.

Next, the effect of the choice of $\mathcal{O}_{\rm p}$, N_1, N_2, and N_3 is investigated. For $l_\mu = 3$, there are six ways of ordering e_1, e_2, and e_3 and six choices of $\mathcal{O}_{\rm p}$. Figure 15 shows $\nu(k)$ for each ordering of the filter coefficients, where the corresponding choice of $\mathcal{O}_{\rm p}$ is given in Table 3. Note that, in each case, $|\nu|$ converges to $\mathcal{O}_{\rm p}^{-1}\mu$. □

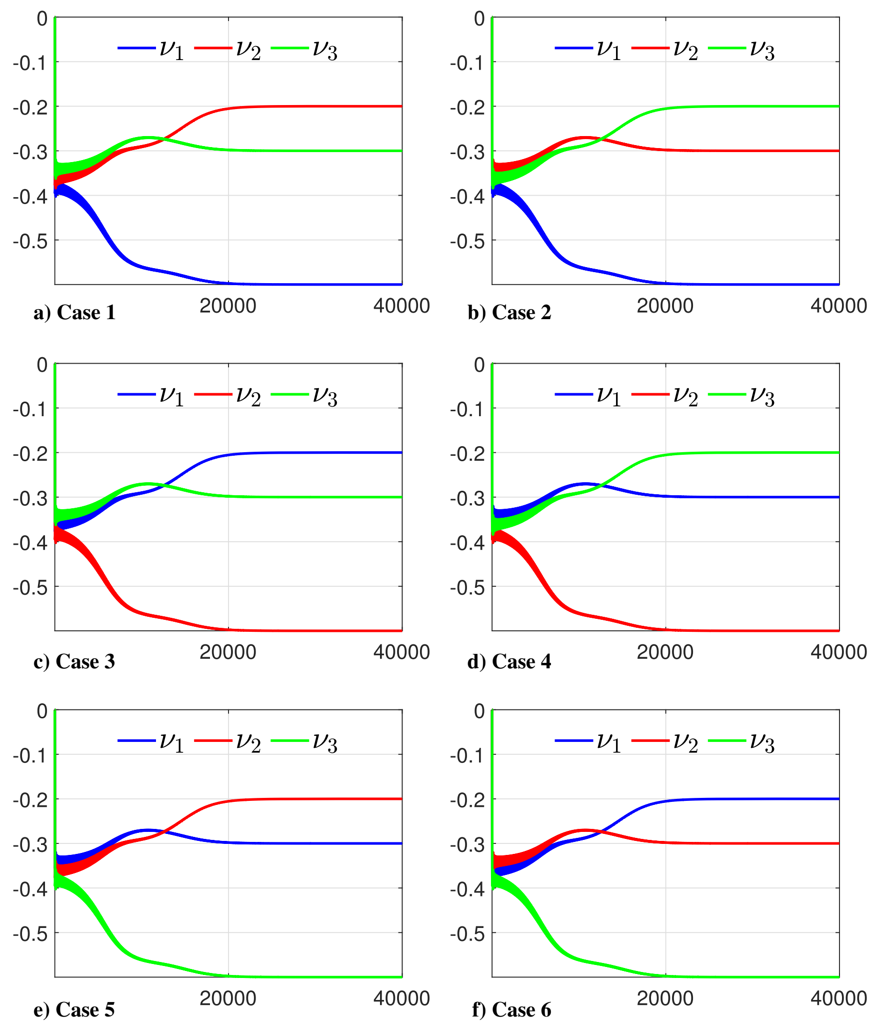


Fig. 15 Example VIII.1. Parameter pre-estimate $\nu(k)$ for various choices of $G_{\rm f}$ and $\mathcal{O}_{\rm p}$ given in Table 3.

Table 3 Filter coefficients and $\mathcal{O}_p$ for Example VIII.1

Case	N_1	N_2	N_3	p
1	e_1	e_2	e_3	321
2	e_1	e_3	e_2	231
3	e_2	e_1	e_3	312
4	e_2	e_3	e_1	132
5	e_3	e_2	e_1	123
6	e_3	e_1	e_2	213

Example VIII.2: Affinely parameterized nonlinear dynamics with three unknown parameters. This example investigates the effect of N_1, N_2, N_3, and $\mathcal{O}_p$ on the feasible region. Consider the (3, 3) type nonlinear system ([26] p. 183)

$$x(k+1) = \begin{bmatrix} x_2(k) \\ \frac{\mu_1+\mu_2 x_2(k)+\mu_3 x_1(k)}{1+0.6x_2(k)+1.1x_1(k)} \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(k) \tag{51}$$

$$y(k) = x_1(k) \tag{52}$$

where $\mu = [\mu_1 \quad \mu_2 \quad \mu_3]^T = [0.5 \quad 0.8 \quad 1.0]^T$. The initial state is $x(0) = [10 \quad 10 \quad 10]^T$, $u(k)$ is given by Eq. (41), $N_1 = e_1$, $N_2 = e_2$, $N_3 = e_3$, $\lambda = 0.9999$, and $R_\theta = 10^6 I_2$. Furthermore, $p = 213$, and thus $\mathcal{S}_{\mathcal{O}_p} = \{[\pm\mu_2 \pm \mu_1 \pm \mu_3]^T\}$. Figure 16 shows the output error, true parameter, parameter pre-estimate, parameter estimate, state-estimate error, and estimator coefficient. Analogous results shown in Fig. 15 are obtained for other choices of the filter coefficients and the permutation matrix $\mathcal{O}_p$. □

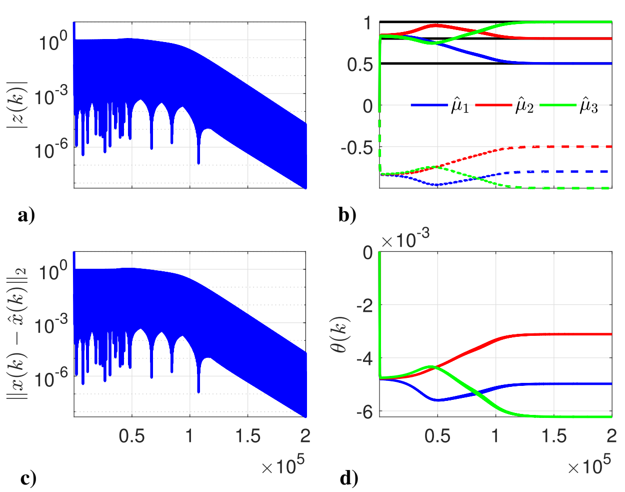


Fig. 16 Example VIII.2: a) output error; b) true parameter, parameter pre-estimate, and parameter estimate; c) state-estimate error; d) parameter estimator coefficients.

IX. Parameter Estimation for the Generalized Burgers Equation

In this section, we consider the generalized one-dimensional viscous Burgers equation [27]

$$\frac{\partial u}{\partial t} + \mu_1 \frac{\partial}{\partial x}\frac{u^2}{2} = \frac{\partial}{\partial x}\left(\mu_2 \frac{\partial u}{\partial x}\right) \tag{53}$$

where $u(x,t)$ is a function of space and time with domain $[0,1]\times[0,\infty)$, $\mu_1 > 0$ is the convective constant, and $\mu_2 > 0$ is the viscosity. Note that there is no external input to this system and u is used to denote the solution of this partial differential equation. The initial condition is $u(x,0) = 0$ for all $x \in [0,1]$, and the boundary conditions are $u(0,t) = 0$ and $u(1,t) = \sin(5t) + 0.25\sin(10t)$ for all $t \geq 0$. The objective is to estimate the unknown parameter $\mu \triangleq [\mu_1 \quad \mu_1]^T$ using measurements of u at a single location.

The Burgers Eq. (53) is discretized using a forward Euler approximation for the time derivative, a second-order-accurate upwind method for the convective term, and a second-order-accurate central difference scheme for the viscous term. The spatial domain [0, 1] is discretized using N equally spaced grid points; thus $\Delta x \triangleq 1/(N-1)$. The time step Δt is chosen to satisfy the CFL condition, that is,

$$\Delta t < \frac{C_{\max}\Delta x}{|\max(u)|} \tag{54}$$

where the Courant number $C_{\max}$ depends on the discretization scheme [28]. Finally, the discrete variable $u_j(k) \triangleq u((j-1)\Delta x, k\Delta t)$ is defined on the grid points $j \in \{1, \ldots, N\}$ for all time steps $k \geq 0$. Hence, at each grid point, $j \in \{3, \ldots, N-1\}$,

$$u_j(k+1) = u_j(k) - \mu_1 \frac{\Delta t}{2\Delta x}(1.5u_j(k)^2 - 2u_{j-1}(k)^2 + 0.5u_{j-2}(k)^2) + \mu_2 \frac{\Delta t}{\Delta x^2}(u_{j+1}(k) - 2u_j(k) + u_{j-1}(k)) \tag{55}$$

For all $k \geq 0$, the discretized boundary conditions are

$$u_1(k) = u_2(k) = 0, \quad u_N(k) = \sin(5\Delta tk) + 0.25\sin(10\Delta tk) \tag{56}$$

and, for all $j \in \{3, \ldots, N-1\}$, the initial condition is

$$u_j(0) = 0 \tag{57}$$

In this example, $\mu_1 = 1.4$, $\mu_2 = 0.3$, $C_{\max} = 0.25$, $N = 100$, and $\Delta t = 10^{-4}$ s. Figure 17a shows the numerical solution of Eq. (55) with the boundary conditions (56) and initial conditions (57), where the solid black line shows the measurement location. Figure 17b shows the measurement $y(k) \triangleq u_{87}(k) = u(0.87, k\Delta t)$.

To start the estimation model, nonzero values of $\hat{\mu}_1(0)$ and $\hat{\mu}_2(0)$ are needed. A simple way to ensure this is to replace μ by $\hat{\mu}(k) = \bar{\mu} + \mathcal{O}_p \nu(k)$, where $\bar{\mu} = [\bar{\mu}_1 \quad \bar{\mu}_1]^T = [1 \quad 0.01]^T$, so that $\hat{\mu}(0) \neq 0$. Furthermore, $N_1 = e_1$, $N_2 = e_2$, $\lambda = 0.9999$, and $R_\theta = 10^6 I_2$. Let $p = 21$ so that $\mathcal{S}_{\mathcal{O}_p} = \{[\pm(\mu_2 - \bar{\mu}_2) \pm (\mu_1 - \bar{\mu}_2)]^T\}$. Figure 18 shows

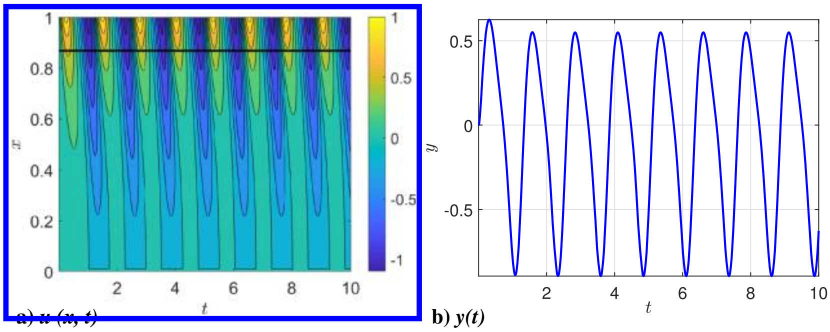


Fig. 17 Simulation of the generalized Burgers equation with the discretization (55).

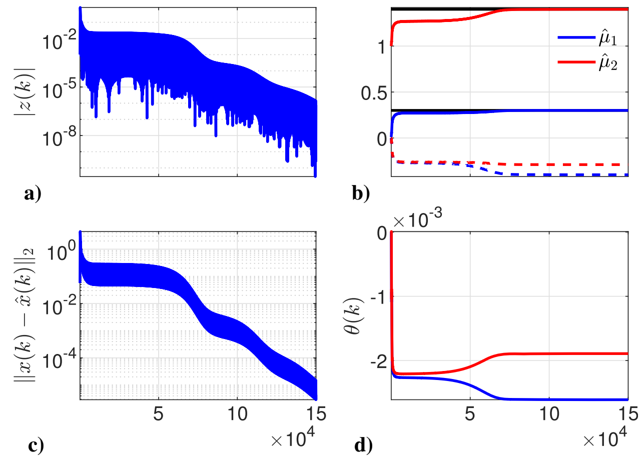


Fig. 18 Generalized Burgers equation: a) output error; b) true parameter, parameter pre-estimate, and parameter estimate; c) state-estimate error; d) parameter estimator coefficients.

the output error, true parameter, parameter pre-estimate, parameter estimate, state-estimate error, and estimator coefficient.

X. Conclusions

This paper presented retrospective cost parameter estimation (RCPE), which is an iterative, data-driven technique for estimating unknown parameters in linear and nonlinear dynamical systems. Unlike the extended Kalman filter, RCPE is gradient-free; unlike the unscented Kalman filter, RCPE is ensemble-free; and unlike variational methods, RCPE is adjoint-free. The main contribution relative to prior work is the extension to multiple unknown parameters. This extension requires a permutation matrix to correctly associate each parameter estimate with the corresponding unknown parameter. The need to select the permutation matrix is the price paid for not requiring gradient information, an ensemble of models, or an adjoint model. The potential usefulness of RCPE was demonstrated by application to the Burgers equation, for which unscented Kalman filter requires an ensemble of 207 models each with 101 states, requiring a total of 21,321 state updates. In contrast, RCPE required one model and testing of two permutations, which requires a total of 202 state updates (101 for each permutation).

Future research will focus on several key issues. First, the number of permutation matrices that must be tested is potentially $l_\mu!$, which is prohibitive when l_μ is large. Fortunately, divergence of the parameter estimates can be used to rule out incorrect permutation matrices. Next, RCPE is based on recursive least squares (RLS) with a forgetting factor. Although this is a standard technique, the performance of RLS tends to be sensitive to the choice of λ. It may be possible to overcome this sensitivity by using the extension of RLS with an a priori covariance bound given in [29]. Finally, all of the examples considered in this paper involve asymptotically stable dynamics. Preliminary results suggest that RCPE is effective for systems that are Lyapunov stable; however, in this case, the performance of RCPE is sensitive to the weighting parameters. Better understanding of this case is needed to mitigate the sensitivity.

Acknowledgments

This research was supported by AFOSR under DDDAS grant FA9550-16-1-0071 (Dynamic Data-Driven Applications Systems; http://www.1dddas.org/). The authors thank Karthik Duraisamy for assistance with Burgers equation. They also thank the reviewers for helpful comments.

References

[1] Van der Merwe, R., and Wan, E. A., "The Square-Root Unscented Kalman Filter for State and Parameter-Estimation," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 6, IEEE Publ., Piscataway, NJ, 2001, pp. 3461–3464. doi:10.1109/ICASSP.2001.940586

[2] Ljung, L., "Asymptotic Behavior of the Extended Kalman Filter as a Parameter Estimator for Linear Systems," *IEEE Transactions on Automatic Control*, Vol. 24, No. 1, 1979, pp. 36–50. doi:10.1109/TAC.1979.1101943

[3] Plett, G. L., "Extended Kalman Filtering for Battery Management Systems of LiPB-Based HEV Battery Packs—Part 3. State and Parameter Estimation," *Journal of Power Sources*, Vol. 134, No. 2, 2004, pp. 277–292. doi:10.1016/j.jpowsour.2004.02.033

[4] Wan, E. A., and Van Der Merwe, R., "The Unscented Kalman Filter for Nonlinear Estimation," *Adaptive Systems for Signal Processing, Communications, and Control Symposium*, IEEE Publ., Piscataway, NJ, 2000, pp. 153–158. doi:10.1109/ASSPCC.2000.882463

[5] Evensen, G., "The Ensemble Kalman Filter for Combined State and Parameter Estimation," *IEEE Control Systems Magazine*, Vol. 29, No. 3, 2009, pp. 83–104. doi:10.1109/MCS.2009.932223

[6] Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R., "Dual State and Parameter Estimation of Hydrological Models Using Ensemble Kalman Filter," *Advances in Water Resources*, Vol. 28, No. 2, 2005, pp. 135–147. doi:10.1016/j.advwatres.2004.09.002

[7] Madankan, R., Singla, P., Singh, T., and Scott, P. D., "Polynomial-Chaos-Based Bayesian Approach for State and Parameter Estimations," *Journal of Guidance, Control, and Dynamics*, Vol. 36, No. 4, 2013, pp. 1058–1074. doi:10.2514/1.58377

[8] Jaakkola, T. S., and Jordan, M. I., "Bayesian Parameter Estimation via Variational Methods," *Statistics and Computing*, Vol. 10, No. 1, 2000, pp. 25–37. doi:10.1023/A:1008932416310

[9] Raffard, R. L., Amonlirdviman, K., Axelrod, J. D., and Tomlin, C. J., "An Adjoint-Based Parameter Identification Algorithm Applied to Planar Cell Polarity Signaling," *IEEE Transactions on Automatic Control*, Vol. 53, Jan. 2008, pp. 109–121. doi:10.1109/TAC.2007.911362

[10] Eknes, M., and Evensen, G., "Parameter Estimation Solving a Weak Constraint Variational Formulation for an Ekman Model," *Journal of Geophysical Research: Oceans*, Vol. 102, No. C6, June 1997, pp. 12479–12491. doi:10.1029/96JC03454

[11] Mercere, G., Prot, O., and Ramos, J. A., "Identification of Parameterized Gray-Box State-Space Systems: from a Black-Box Linear Time-Invariant Representation to a Structured One," *IEEE Transactions on Automatic Control*, Vol. 59, No. 11, 2014, pp. 2873–2885. doi:10.1109/TAC.2014.2351853

[12] Yu, C., Ljung, L., and Verhaegen, M., "Identification of Structured State-Space Models," *Automatica*, Vol. 90, April 2018, pp. 54–61. doi:10.1016/j.automatica.2017.12.023

[13] D'Amato, A. M., Ridley, A. J., and Bernstein, D. S., "Retrospective-Cost-Based Adaptive Model Refinement for the Ionosphere and Thermosphere," *Statistical Analysis and Data Mining*, Vol. 4, No. 4, 2011, pp. 446–458. doi:10.1002/sam.v4.4

[14] Zhou, X., Bernstein, D. S., Stein, J. L., and Ersal, T., "Battery State of Health Monitoring by Estimation of Side Reaction Current Density via Retrospective-Cost Subsystem Identification," *Journal of Dynamic Systems, Measurement, and Control*, Vol. 139, No. 9, 2017, Paper 091007. doi:10.1115/1.4036030

[15] Goel, A., Duraisamy, K., and Bernstein, D. S., "Parameter Estimation in the Burgers Equation Using Retrospective-Cost Model Refinement," *Proceedings of American Control Conference*, IEEE Publ., Piscataway, NJ, 2016, pp. 6983–6988. doi:10.1109/ACC.2016.7526773

[16] Rahman, Y., Xie, A., and Bernstein, D. S., "Retrospective Cost Adaptive Control: Pole Placement, Frequency Response, and Connections with LQG Control," *IEEE Control System Magazine*, Vol. 37, No. 5, Oct. 2017, pp. 28–69. doi:10.1109/MCS.2017.2718825

[17] Goel, A., and Bernstein, D. S., "Parameter Estimation for Nonlinearly Parameterized Gray-Box Models," *Proceedings of American Control Conference*, IEEE Publ., Piscataway, NJ, 2018, pp. 5280–5285. doi:10.23919/ACC.2018.8431502

[18] Goel, A., Ridley, A., and Bernstein, D. S., "Estimation of the Eddy Diffusion Coefficient Using Total Electron Content Data," *Proceedings of American Control Conference*, IEEE Publ., Piscataway, NJ, 2018, pp. 3298–3303.
doi:10.23919/ACC.2018.8431184

[19] Goel, A., and Bernstein, D. S., "Data-Driven Parameter Estimation for Models with Nonlinear Parameter Dependence," *Proceedings of Conference on Decision and Control*, IEEE Publ., Piscataway, NJ, Dec. 2018, pp. 1470–1475.
doi:10.1109/CDC.2018.8619566

[20] Yu, M.-J., and Bernstein, D. S., "Retrospective Cost Subsystem Estimation and Smoothing for Linear Systems with Structured Uncertainty," *Journal of Aerospace Information Systems*, Vol. 15, No. 10, 2018, pp. 566–584.
doi:10.2514/1.I010581

[21] Bellman, R., and Åström, K. J., "On Structural Identifiability," *Mathematical Biosciences*, Vol. 7, Nos. 3–4, 1970, pp. 329–339.
doi:10.1016/0025-5564(70)90132-X

[22] Grewal, M., and Glover, K., "Identifiability of Linear and Nonlinear Dynamical Systems," *IEEE Transactions on Automatic Control*, Vol. 21, No. 6, 1976, pp. 833–837.
doi:10.1109/TAC.1976.1101375

[23] Stanhope, S., Rubin, J. E., and Swigon, D., "Identifiability of Linear and Linear-in-Parameters Dynamical Systems from a Single Trajectory," *SIAM Journal on Applied Dynamical Systems*, Vol. 13, No. 4, 2014, pp. 1792–1815.
doi:10.1137/130937913

[24] Mareels, I. M. Y., Bitmead, R. R., Gevers, M., Johnson, C. R., Kosut, R. L., and Poubelle, M. A., "How Exciting Can a Signal Really Be?" *Systems & Control Letters*, Vol. 8, No. 3, 1987, pp. 197–204.
doi:10.1016/0167-6911(87)90027-2

[25] Willems, J. C., Rapisarda, P., Markovsky, I., and De Moor, B. L., "A Note on Persistency of Excitation," *Systems & Control Letters*, Vol. 54, No. 4, 2005, pp. 325–329.
doi:10.1016/j.sysconle.2004.09.003

[26] Kulenović, M. R. S., and Ladas, G. E., *Dynamics of Second Order Rational Difference Equations: With Open Problems and Conjectures*, Chapman & Hall/CRC, Boca Raton, FL, 2002, p. 183.

[27] Blackstock, D. T., "Generalized Burgers Equation for Plane Waves," *The Journal of the Acoustical Society of America*, Vol. 77, No. 6, 1985, pp. 2050–2053.
doi:10.1121/1.391778

[28] Courant, R., Friedrichs, K., and Lewy, H., "On the Partial Difference Equations of Mathematical Physics," *IBM Journal of Research and Development*, Vol. 11, No. 2, 1967, pp. 215–234.
doi:10.1147/rd.112.0215

[29] Salgado, M., Goodwin, G., and Middleton, R., "Modified Least Squares Algorithm Incorporating Exponential Resetting and Forgetting," *International Journal of Control*, Vol. 47, No. 2, 1988, pp. 477–491.
doi:10.1080/00207178808906026