UC San Diego

UC San Diego Previously Published Works

Title

Smoothed estimation of unknown inputs and states in dynamic systems with application to oceanic flow field reconstruction

Permalink

<https://escholarship.org/uc/item/1xq35553>

Journal

International Journal of Adaptive Control and Signal Processing, 29(10)

ISSN

0890-6327

Authors

Fang, H
de Callafon, RA
Franks, PJS

Publication Date

2015-10-01

DOI

10.1002/acs.2529

Peer reviewed

Smoothed estimation of unknown inputs and states in dynamic systems with application to oceanic flow field reconstruction

Huazhen Fang^{1,‡}, Raymond A. de Callafon^{1,*} and Peter J. S. Franks²

¹*Department of Mechanical & Aerospace Engineering, University of California, San Diego, CA 92093, USA*

²*Scripps Institution of Oceanography, University of California, San Diego, CA 92093, USA*

SUMMARY

Forward-backward smoothing based unknown input and state estimation for dynamic systems is studied in this paper, motivated by reconstruction of an oceanographic flow field using a swarm of buoyancy-controlled drifters. The development is conducted in a Bayesian framework. A Bayesian paradigm is constructed first to offer a probabilistic view of the unknown quantities given the measurements. Then a maximum *a posteriori* is established to build a means for simultaneous input and state smoothing, which can be solved by the classical Gauss–Newton method in the nonlinear case. Application to reconstruction of a complex three-dimensional flow field is presented and investigated *via* simulation studies. Copyright © 2014 John Wiley & Sons, Ltd.

Received 22 November 2013; Accepted 8 November 2014

KEY WORDS: input estimation; state estimation; forward-backward smoothing; nonlinear systems; ocean observing

1. INTRODUCTION

Oceanic flow field reconstruction is extremely important in oceanography, because of the fundamental role of flows in the marine physical, chemical and biological environment and the climate change.

High-accuracy flow-field reconstruction is aimed to be achieved using a group of submersible drifters. A schematic illustration of the scenario is shown in Figure 1(a). The drifters are released at different locations in a region of the ocean, and each then travels in cycling movement patterns of submerging/surfacing (Figure 1(b)), with lateral motion driven by ocean currents and vertical motion regulated by buoyancy. The position and acceleration of a drifter during the travel will be measured and recorded. The former is obtained by the satellite-based Global Positioning System (GPS), and the latter by on-board accelerometers. While underwater, each drifter stores a time record of its current depth, acceleration and other oceanographic quantities such as temperature and salinity. When it surfaces, it determines its location by GPS and then transmits all the data *via* a communication satellite to a central server, where analysis and computation are performed.

To reconstruct the flow field and monitor the traveling status of a drifter, it is necessary to estimate simultaneously the flow velocities, which act as external inputs to the drifter, and the drifter's trajectory and velocities, which are the states, from its position and acceleration measurements. Thus, the problem of estimation of unknown inputs and states is to be investigated. It is noteworthy that the same issue also arises in other fields such as fault detection, automotive engineering and weather

*Correspondence to: Raymond A. de Callafon, Department of Mechanical & Aerospace Engineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0411, USA.

†E-mail: callafon@ucsd.edu

‡Present address: Department of Mechanical Engineering, University of Kansas, Lawrence, KS 66045, USA.

Figure 1. (a) The scenario for flow-field reconstruction based on submersible drifters; (b) the traveling profile (submerging/surfacing) of a drifter.

forecasting [1–3], where both input and state variables of a dynamical system are unknown. This, together with the specific application of flow field reconstruction, motivates the development of a general-purpose solution in this paper.

Literature review: Research on *simultaneous input and state estimation (SISE)* dates back to 1970s, with the development of estimators and smoothers of white noise inputs to a linear state-space system in [4]. Since then, it has been gaining momentum. A large body of more recent work has considered completely unknown inputs, and a majority of them are based on existing state estimation methods. A two-stage Kalman filter is proposed in [5] for SISE through decoupled design of input and state filters. Sliding mode state observers are extended in [6, 7] such that unknown input variables can also be estimated. The notion of moving horizon estimation is used in [8] for hybrid systems modeled in the piecewise affine form. In [9, 10], minimum-variance unbiased estimation (MVUE) is applied to derive SISE filters for systems with and without direct input-output feedthrough, respectively, with asymptotic stability properties analyzed in [11]. The same technique is also used in [12] to construct suboptimal SISE filters with proven stability properties. Compared to the linear case, investigation on SISE for nonlinear systems has been paid less attention to. SISE for a special class of nonlinear systems composed of a nominally linear part and a nonlinear part is studied in [13, 14]. Nonlinear SISE can be tackled using Bayesian statistics and Maximum *a Posteriori* (MAP) estimation [15], the algorithms proposed therein being applicable to nonlinear systems in general form. Another problem of relevance to SISE is state estimation with unknown inputs. In this case, the states are estimated from the output measurements, despite no knowledge of inputs. Among an abundance of works on this topic, we underscore [16–18] on MVUE and [19] on sliding mode-based approaches. SISE has a wide range of applications such as fault detection and diagnosis [20, 21], maneuvering target tracking [22], geophysics and environmentology [23] and so on.

Statement of contributions: This paper is partially facilitated by our previous research on nonlinear *simultaneous input and state filtering (SISF)* [15, 24]. The inputs are assumed to be completely unknown, thus admitting no model-based state augmentation for inputs. Compared to the existing literature, our work has two major contributions.

First, a lead is taken in this paper with the study on *simultaneous input and state smoothing (SISS)*, while almost all the work highlighted earlier is focused on SISF. Although both filtering and smoothing are about estimation, the former uses a batch of measurements up until the current time instant with real-time implementation, and the latter uses measurements inclusive of those lagging behind that time instant *via* off-line computing. Hence, SISS can yield more accurate estimates. Forward-backward SISS algorithms are proposed for systems with and without direct input-output feedthrough, respectively, within a Bayesian framework, comprising the first known study on SISS to the authors' best knowledge.

Second, the proposed SISS algorithms are applied to reconstruct an ocean flow field using observation data collected by a group of buoyancy-controlled drifters, as a promising means to address

the pressing need for high-accuracy ocean monitoring. It is noteworthy that SISS, compared to SISF, is a more desirable choice, because a typical oceanic flow or circulation allows off-line data analysis as a result of changes on time scales ranging from a few days to a season. A simulation-based study is conducted to examine the performance of the SISS algorithm with direct feedthrough in this application, and the results validate its effectiveness.

Organization: The remainder of the paper is organized as follows. Sections 2 and 3 consider systems with direct feedthrough. Section 2 proposes the Bayesian paradigm for forward-backward smoothing of the input and state variables. In Section 3, smoothing algorithms are developed for nonlinear and linear systems, respectively, as a continuation of the Bayesian smoother to computationally feasible implementation. Similar lines are followed in Section 4 to develop SISS algorithms for systems without direct feedthrough. Section 5 presents the application to oceanographic flow field reconstruction to illustrate the effectiveness of the proposed algorithms. Finally, Section 6 gathers our conclusions.

Notation: The notation used throughout the paper is fairly standard. The superscripts ‘T’ and ‘−1’ stand for matrix transposition and inverse, respectively; p denotes the probability density function (pdf); $\mathbb{R}^n$ denotes the n -dimensional Euclidean space; $\|\mathbf{x}\|_{\Sigma}^2 = \mathbf{x}^T \Sigma^{-1} \mathbf{x}$ is the square of weighted 2-norm of $\mathbf{x}$; $\mathcal{G}(\mathbf{x}; \bar{\boldsymbol{\mu}}, \Sigma)$ is a Gaussian function of $\mathbf{x}$ given by $\mathcal{G}(\mathbf{x}; \bar{\boldsymbol{\mu}}, \Sigma) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2} \|\mathbf{x} - \bar{\boldsymbol{\mu}}\|_{\Sigma}^2\right)$; $\nabla_{\mathbf{x}} \mathbf{y}$ represents the derivative of $\mathbf{y}$ with respect to $\mathbf{x}$.

2. BAYESIAN INPUT AND STATE SMOOTHING

Let us consider the following nonlinear discrete-time system with direct feedthrough:

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) + \mathbf{w}_k, \\ \mathbf{y}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{u}_k) + \mathbf{v}_k, \end{cases} \quad (1)$$

where $\mathbf{x}_k \in \mathbb{R}^{n_x}$ is the unknown state vector, $\mathbf{u}_k \in \mathbb{R}^{n_u}$ is the unknown input vector and $\mathbf{y}_k \in \mathbb{R}^{n_y}$ is the output vector. The process noise $\{\mathbf{w}_k\}$ and the measurement noise $\{\mathbf{v}_k\}$ are assumed to be mutually independent, zero-mean white Gaussian sequences with covariances $\mathbf{Q}_k \geq 0$ and $\mathbf{R}_k > 0$, respectively. More specifically, we have $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$, that is, $p(\mathbf{w}_k) = \mathcal{G}(\mathbf{w}_k; \mathbf{0}, \mathbf{Q}_k)$, and $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$, that is, $p(\mathbf{v}_k) = \mathcal{G}(\mathbf{v}_k; \mathbf{0}, \mathbf{R}_k)$. The nonlinear mappings $\mathbf{f}: \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_x}$ and $\mathbf{h}: \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_y}$ characterize the process dynamics and measurement model, respectively. Both of them are assumed to be C^1 .

From a statistical point of view, $\mathbf{u}_k$, $\mathbf{x}_k$ and $\mathbf{y}_k$ are random variables, the change of which over time leads to corresponding sequences equivalent to different stochastic processes. As output measurements, the sequence $\{\mathbf{y}_k\}$ contains information about $\mathbf{u}_k$ and $\mathbf{x}_k$. Define $\mathbb{Y}_{1:k} := \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k\}$, which is the set of all measurements up until time k . SISF is aimed to compute the joint pdf of $\mathbf{u}_k$ and $\mathbf{x}_k$ conditioned on $\mathbb{Y}_{1:k}$, $p(\mathbf{u}_k, \mathbf{x}_k | \mathbb{Y}_{1:k})$, or more specifically, the passing to $p(\mathbf{u}_k, \mathbf{x}_k | \mathbb{Y}_{1:k})$ from $p(\mathbf{u}_{k-1}, \mathbf{x}_{k-1} | \mathbb{Y}_{1:k-1})$. Fixed-interval SISS, however, considers $p(\mathbf{u}_k, \mathbf{x}_k | \mathbb{Y}_{1:N})$, where $N \geq k$ is fixed. Here, additional measurement data are used in this case, in the hope of obtaining better estimates of $\mathbf{u}_k$ and $\mathbf{x}_k$.

For notational convenience, we define an augmented vector $\boldsymbol{\xi}_k = [\mathbf{u}_k^T \mathbf{x}_k^T]^T$ that is to be estimated. For fully unknown inputs, we establish the following assumption to proceed further:

A1 $\{\mathbf{u}_k\}$ is a *white* process, independent of $\mathbf{x}_0$, $\{\mathbf{w}_k\}$ and $\{\mathbf{v}_k\}$.

The aforementioned assumption is justifiable because of the unavailability of the *a priori* knowledge about $\mathbf{u}_k$ — $\mathbf{u}_k$ may take arbitrary values and contains no information about $\mathbf{u}_l$ for $k \neq l$. By A1, $\mathbf{u}_k$ is independent of $\mathbb{X}_k := \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k\}$ and $\mathbb{Y}_{k-1}$ [25, Theorem 10.4, p. 71].

The Bayesian paradigm for SISS can be established using the Bayes’ rule to build the backward recursion of $p(\boldsymbol{\xi}_k | \mathbb{Y}_{1:N})$ from $p(\mathbf{x}_{k+1} | \mathbb{Y}_{1:N})$, as shown in the following theorem.

Theorem 1

Suppose that A1 holds. For the system in (1), the Bayesian smoothing paradigm for input and state estimation is given by

$$p(\xi_k | \mathbb{Y}_{1:N}) = p(\xi_k | \mathbb{Y}_{1:k}) \int \frac{p(\mathbf{x}_{k+1} | \xi_k) p(\mathbf{x}_{k+1} | \mathbb{Y}_{1:N})}{p(\mathbf{x}_{k+1} | \mathbb{Y}_{1:k})} d\mathbf{x}_{k+1}. \quad (2)$$

Proof

It is seen that

$$\begin{aligned} p(\xi_k | \mathbb{Y}_{1:N}) &= \int p(\xi_k, \xi_{k+1} | \mathbb{Y}_{1:N}) d\xi_{k+1} \\ &= \int p(\xi_k | \xi_{k+1}, \mathbb{Y}_{1:N}) p(\xi_{k+1} | \mathbb{Y}_{1:N}) d\xi_{k+1}. \end{aligned}$$

Note that ξ_k is independent of $\mathbb{Y}_{k+1:N}$ and $\mathbf{u}_{k+1}$ given $\mathbf{x}_{k+1}$, because of the Markovian propagation of the state and the assumption A1. Hence,

$$p(\xi_k | \xi_{k+1}, \mathbb{Y}_{1:N}) = p(\xi_k | \mathbf{x}_{k+1}, \mathbb{Y}_{1:k}).$$

Then we have

$$\begin{aligned} p(\xi_k | \mathbb{Y}_{1:N}) &= \int p(\xi_k | \mathbf{x}_{k+1}, \mathbb{Y}_{1:k}) p(\xi_{k+1} | \mathbb{Y}_{1:N}) d\xi_{k+1} \\ &= \int p(\xi_k | \mathbf{x}_{k+1}, \mathbb{Y}_{1:k}) p(\mathbf{x}_{k+1} | \mathbb{Y}_{1:N}) d\mathbf{x}_{k+1}. \end{aligned} \quad (3)$$

Meanwhile, note that

$$\begin{aligned} p(\xi_k | \mathbf{x}_{k+1}, \mathbb{Y}_{1:k}) &= \frac{p(\xi_k, \mathbf{x}_{k+1}, \mathbb{Y}_{1:k})}{p(\mathbf{x}_{k+1}, \mathbb{Y}_{1:k})} \\ &= \frac{p(\mathbf{x}_{k+1} | \xi_k, \mathbb{Y}_{1:k}) p(\xi_k | \mathbb{Y}_{1:k})}{p(\mathbf{x}_{k+1} | \mathbb{Y}_{1:k})} \\ &= \frac{p(\mathbf{x}_{k+1} | \xi_k) p(\xi_k | \mathbb{Y}_{1:k})}{p(\mathbf{x}_{k+1} | \mathbb{Y}_{1:k})}. \end{aligned} \quad (4)$$

Inserting (4) into (3), we obtain (2). $\square$

Illustrating how to calculate the conditional pdf of ξ_k given $\mathbb{Y}_{1:N}$, the Bayesian paradigm in (2) is an input and state smoother in a statistical sense. However, direct, analytical evaluation of the pdf's is known to be quite difficult, if not impossible, for nonlinear systems, because of the intractable nonlinear transformation of random variables. Hence, we will seek a numerically feasible solution by introducing certain approximations.

3. SMOOTHING ALGORITHMS

In this section, we present the development of the SISS algorithm for the nonlinear system in (1). Then we specialize the algorithm to the linear case.

3.1. Smoothing for nonlinear systems

Gaussian approximations are imposed on the following conditional probability distributions:

- A2** $\mathbf{x}_{k+1} | \xi_k \sim \mathcal{N}(\mathbf{f}(\xi_k), \mathbf{Q}_k)$;
- A3** $\xi_k | \mathbb{Y}_{1:k} \sim \mathcal{N}(\hat{\xi}_{k|k}, \mathbf{P}_{k|k}^{\xi})$;
- A4** $\mathbf{x}_{k+1} | \mathbb{Y}_{1:N} \sim \mathcal{N}(\hat{\mathbf{x}}_{k+1|N}, \mathbf{P}_{k+1|N}^{\mathbf{x}})$;
- A5** $\mathbf{x}_{k+1} | \mathbb{Y}_{1:k} \sim \mathcal{N}(\hat{\mathbf{x}}_{k+1|k}, \mathbf{P}_{k+1|k}^{\mathbf{x}})$.

Here, $\hat{\xi}_{k|k}$ is the filtered estimate of ξ_k given $\mathbb{Y}_{1:k}$, $\mathbf{P}_{k|k}^{\xi}$ is the filtering error covariance, $\hat{\mathbf{x}}_{k+1|N}$ is the smoothed estimate of $\mathbf{x}_{k+1}$ given $\mathbb{Y}_{1:N}$ and $\mathbf{P}_{k+1|N}^{\mathbf{x}}$ is the smoothing error covariance.

The assumption that $\mathbf{w}_k$ is zero-mean and Gaussian implies A2. Random variables with unknown distributions are often approximated to be Gaussian, based on the central limit theorem. The Gaussian distribution functions are also convenient to deal with. Therefore, the distribution of $\xi_k|\mathbb{Y}_{1:k}$ is approximated by a Gaussian distribution with the mean $\hat{\xi}_{k|k}$ and covariance $\mathbf{P}_{k|k}^{\xi}$, and similar approximations are made in A4 and A5. Such a treatment is prevalent in the development of a substantial number of nonlinear estimation algorithms [26, 27]. The assumptions A2–A5, as will be seen, bridge the gap from the Bayesian paradigm in (2) to the numerical algorithm.

It is appealing to consider a joint smoother of input $\mathbf{u}_k$ and state $\mathbf{x}_k$ that maximizes $p(\xi_k|\mathbb{Y}_{1:N})$. It can be expressed as follows:

$$\hat{\xi}_{k|N} = \arg \max_{\xi_k} p(\xi_k|\mathbb{Y}_{1:N}). \quad (5)$$

The aforementioned maximization of $p(\xi_k|\mathbb{Y}_{1:N})$ can be transformed into the familiar form of minimization of a cost function with assistance of the assumptions A1–A5, as shown in the next theorem.

Theorem 2

Suppose that A1–A5 hold. For the system in (1), an equivalence of the smoother in (5) is given by

$$\hat{\xi}_{k|N} = \arg \min_{\xi_k} \ell(\xi_k), \quad (6)$$

where

$$\ell(\xi_k) := \left\| \xi_k - \hat{\xi}_{k|k} \right\|_{\mathbf{P}_{k|k}^{\xi}}^2 + \left\| \mathbf{f}(\xi_k) - \delta_k \right\|_{\Delta_k}^2, \quad (7)$$

$$\Delta_k = \left[\left(\mathbf{P}_{k+1|N}^{\mathbf{x}} \right)^{-1} - \left(\mathbf{P}_{k+1|k}^{\mathbf{x}} \right)^{-1} \right]^{-1} + \mathbf{Q}_k, \quad (8)$$

$$\begin{aligned} \delta_k &= \left[\left(\mathbf{P}_{k+1|N}^{\mathbf{x}} \right)^{-1} - \left(\mathbf{P}_{k+1|k}^{\mathbf{x}} \right)^{-1} \right]^{-1} \\ &\quad \cdot \left[\left(\mathbf{P}_{k+1|N}^{\mathbf{x}} \right)^{-1} \hat{\mathbf{x}}_{k+1|N} - \left(\mathbf{P}_{k+1|k}^{\mathbf{x}} \right)^{-1} \hat{\mathbf{x}}_{k+1|k} \right]. \end{aligned} \quad (9)$$

Proof

Consider the Bayesian paradigm in (2), where the probabilities can be substituted using the assumptions A2–A5. By Lemma A.2 in Appendix, we have

$$\frac{\mathcal{G}(\mathbf{x}_{k+1}; \hat{\mathbf{x}}_{k+1|N}, \mathbf{P}_{k+1|N}^{\mathbf{x}})}{\mathcal{G}(\mathbf{x}_{k+1}; \hat{\mathbf{x}}_{k+1|k}, \mathbf{P}_{k+1|k}^{\mathbf{x}})} \propto \mathcal{G}(\mathbf{x}_{k+1}; \bar{\mathbf{x}}_{k+1}, \bar{\mathbf{P}}_{k+1}^{\mathbf{x}}),$$

where

$$\begin{aligned} \bar{\mathbf{P}}_{k+1}^{\mathbf{x}} &= \left[\left(\mathbf{P}_{k+1|N}^{\mathbf{x}} \right)^{-1} - \left(\mathbf{P}_{k+1|k}^{\mathbf{x}} \right)^{-1} \right]^{-1}, \\ \bar{\mathbf{x}}_{k+1} &= \bar{\mathbf{P}}_{k+1}^{\mathbf{x}} \left[\left(\mathbf{P}_{k+1|N}^{\mathbf{x}} \right)^{-1} \hat{\mathbf{x}}_{k+1|N} - \left(\mathbf{P}_{k+1|k}^{\mathbf{x}} \right)^{-1} \hat{\mathbf{x}}_{k+1|k} \right]. \end{aligned}$$

It then follows from Lemma A.3 that

$$\int \mathcal{G}(\mathbf{x}_{k+1}; \mathbf{f}(\xi_k), \mathbf{Q}_k) \cdot \mathcal{G}(\mathbf{x}_{k+1}; \bar{\mathbf{x}}_{k+1}, \bar{\mathbf{P}}_{k+1}^{\mathbf{x}}) d\mathbf{x}_{k+1} \propto \mathcal{G}(\mathbf{f}(\xi_k); \delta_k, \Delta_k),$$

where

$$\begin{aligned}\Delta_k &= \bar{\mathbf{P}}_{k+1}^x + \mathbf{Q}_k, \\ \delta_k &= \bar{\mathbf{x}}_{k+1}.\end{aligned}$$

By (2), the likelihood function $L(\xi_k) = p(\xi_k | \mathbb{Y}_{1:N})$ is then

$$L(\xi_k) \propto \mathcal{G}(\xi_k; \hat{\xi}_{k|k}, \mathbf{P}_{k|k}^\xi) \mathcal{G}(\mathbf{f}(\xi_k); \delta_k, \Delta_k).$$

Hence, the log-likelihood function $\ell(\xi_k)$ to minimize is exactly the one in (7). $\square$

Remark 1

It is noted that the estimator designed in (5) or (6) is exactly based on the MAP estimation, because $p(\xi_k | \mathbb{Y}_{1:N})$ is the *a posteriori* distribution of ξ_k given $\mathbb{Y}_{1:N}$. Therefore, $\ell(\xi_k)$ in (7) is a log-MAP cost function. MAP estimation formulated within Bayesian framework has often been used to develop optimal filtering methods, for example, [28]. Here, we extend this technique to address the smoothing problem and especially for the purpose of joint input and state estimation.

Because of the nonlinearities, it would be extremely difficult to obtain an analytical solution to the optimization problem in (6), motivating the development of an approximate numerical solution. By observation, the sum of the weighted 2-norms in (7) can be rewritten as

$$\ell(\xi_k) = \|\mathbf{s}_k(\xi_k)\|^2, \quad (10)$$

where

$$\mathbf{s}_k(\xi_k) = \begin{bmatrix} \left(\mathbf{P}_{k|k}^\xi\right)^{-\frac{1}{2}} (\xi_k - \hat{\xi}_{k|k}) \\ \Delta_k^{-\frac{1}{2}} (\mathbf{f}(\xi_k) - \delta_k) \end{bmatrix}.$$

The classical Gauss–Newton method, which has been developed for nonlinear least squares problems, can then be applied here. It is an iterative searching process that linearizes around the current arrival point, determines the best search direction and then moves forward to the next point. The following theorem thus presents the SISS algorithm.

Theorem 3

For the input and state smoother designed in (6), the Gauss–Newton-based solution $\hat{\xi}_{k|N} = \hat{\xi}_{k|N}^{(i_{\max})}$ is iteratively computed by

$$\hat{\xi}_{k|N}^{(i+1)} = \hat{\xi}_{k|N}^{(i)} - \left[\nabla_{\xi}^T \mathbf{s}(\hat{\xi}_{k|N}^{(i)}) \nabla_{\xi} \mathbf{s}(\hat{\xi}_{k|N}^{(i)}) \right]^{-1} \nabla_{\xi}^T \mathbf{s}(\hat{\xi}_{k|N}^{(i)}) \mathbf{s}(\hat{\xi}_{k|N}^{(i)}), \quad (11)$$

where (i) denotes the iteration number, $i_{\max}$ is the prespecified max number of iterations and

$$\nabla_{\xi} \mathbf{s}(\xi_k) = \begin{bmatrix} \left(\mathbf{P}_{k|k}^\xi\right)^{-\frac{1}{2}} \\ \Delta_k^{-\frac{1}{2}} \nabla_{\xi} \mathbf{f}(\xi_k) \end{bmatrix}.$$

The first-order approximation of the associated error covariance is

$$\mathbf{P}_{k|N}^\xi = \left[\left(\mathbf{P}_{k|k}^\xi\right)^{-1} + \nabla_{\xi}^T \mathbf{f}(\hat{\xi}_{k|N}) \Delta_k^{-1} \nabla_{\xi} \mathbf{f}(\hat{\xi}_{k|N}) \right]^{-1}. \quad (12)$$

Proof

Applying the classical Gauss–Newton method to (25) leads to (11) directly [29]. It is noted that the assumptions A1–A3 ensure $p(\xi_k | \mathbb{Y}_{1:N})$ to be Gaussian as well. Thus, according to the MAP estimation theory, the error covariance can be approximately estimated by the inverse of the Fisher information matrix $\mathcal{F}$ evaluated at the estimate:

$$\mathbf{P}_{k|N}^\xi = \mathcal{F}^{-1}(\hat{\xi}_{k|N}),$$

where $\mathcal{F}$ is given by

$$\mathcal{F}(\xi_k) = \left(\mathbf{P}_{k|k}^\xi \right)^{-1} + \nabla_{\xi}^T \mathbf{f}(\xi_k) \mathbf{\Delta}_k^{-1} \nabla_{\xi} \mathbf{f}(\xi_k).$$

This completes the proof. $\square$

We say that (11) and (12) are the backward smoothing equations for input and state estimation. The corresponding forward filtering equations are presented in [15, 24]. The derivation follows similar lines: a Bayesian framework is constructed first, and then a MAP estimation problem is formulated within the framework and solved.

Remark 2

The Gauss–Newton method is used for an iterative search over a designed cost function to find the best estimates. This principle also applies to some nonlinear state estimation algorithms, for example, the iterated extended Kalman filter [30].

Remark 3

The Gauss–Newton method can be modified in several ways for improvement of the computational performance. Among the improvements, one is the damped Gauss–Newton method, which has better convergence performance by adding a damping coefficient. Another one is the trust-region method, which overcomes the singularity problems that may arise in the computing process. The reader is referred to [29] for more details.

3.2. Specialization for linear systems

Developed for nonlinear systems, the SISS algorithm proposed in Theorem 3 is examined in the linear case in this section. Consider a linear system of the following form

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{F}_k \mathbf{x}_k + \mathbf{G}_k \mathbf{u}_k + \mathbf{w}_k, \\ \mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{J}_k \mathbf{u}_k + \mathbf{v}_k, \end{cases} \quad (13)$$

or equivalently,

$$\begin{cases} \mathbf{x}_{k+1} = \bar{\mathbf{F}}_k \xi_k + \mathbf{w}_k, \\ \mathbf{y}_k = \bar{\mathbf{H}}_k \xi_k + \mathbf{v}_k, \end{cases} \quad (14)$$

where

$$\bar{\mathbf{F}}_k = [\mathbf{F}_k \ \mathbf{G}_k], \quad \bar{\mathbf{H}}_k = [\mathbf{H}_k \ \mathbf{J}_k].$$

Because the system is linear, the smoother in (11) and (12) can find the best input and state estimates in a single iteration. We then obtain a linear smoother expressed succinctly in Corollary 1.

Corollary 1

The input and state smoother for the linear system in (14) is given by

$$\hat{\xi}_{k|N} = \hat{\xi}_{k|k} + \mathbf{K}_k (\hat{\mathbf{x}}_{k+1|N} - \bar{\mathbf{F}}_k \hat{\xi}_{k|k}), \quad (15)$$

$$\mathbf{K}_k = \mathbf{P}_{k|k}^\xi \bar{\mathbf{F}}_k^T \left(\mathbf{P}_{k+1|k}^\mathbf{x} \right)^{-1}, \quad (16)$$

with the associated smoothing error covariance given by

$$\mathbf{P}_{k|N}^\xi = \mathbf{P}_{k|k}^\xi + \mathbf{K}_k (\mathbf{P}_{k+1|N}^\mathbf{x} - \mathbf{P}_{k+1|k}^\mathbf{x}) \mathbf{K}_k^T. \quad (17)$$

Proof

Given the linear system in (13), we have

$$\mathbf{s}(\xi_k) = \begin{bmatrix} \left(\mathbf{P}_{k|k}^\xi \right)^{-\frac{1}{2}} \left(\xi_k - \hat{\xi}_{k|k} \right) \\ \mathbf{\Delta}_k^{-\frac{1}{2}} \left(\bar{\mathbf{F}}_k \xi_k - \delta_k \right) \end{bmatrix}, \quad \nabla_{\xi} \mathbf{s}(\xi_k) = \begin{bmatrix} \left(\mathbf{P}_{k|k}^\xi \right)^{-\frac{1}{2}} \\ \mathbf{\Delta}_k^{-\frac{1}{2}} \bar{\mathbf{F}}_k \end{bmatrix}.$$

Let $\hat{\xi}_{k|N}^{(0)}$ be the initial guess of $\hat{\xi}_{k|N}$. As the minimum point can be reached in a single iteration, the smoother in (11) becomes

$$\begin{aligned}\hat{\xi}_{k|N} &= \hat{\xi}_{k|N}^{(0)} - \left[\left(\mathbf{P}_{k|k}^{\xi} \right)^{-1} + \bar{\mathbf{F}}_k^T \Delta_k^{-1} \bar{\mathbf{F}}_k \right]^{-1} \left[\left(\mathbf{P}_{k|k}^{\xi} \right)^{-1} \left(\hat{\xi}_{k|N}^{(0)} - \hat{\xi}_{k|k} \right) + \bar{\mathbf{F}}_k^T \Delta_k^{-1} \left(\bar{\mathbf{F}}_k \hat{\xi}_{k|N}^{(0)} - \delta_k \right) \right] \\ &= \left[\left(\mathbf{P}_{k|k}^{\xi} \right)^{-1} + \bar{\mathbf{F}}_k^T \Delta_k^{-1} \bar{\mathbf{F}}_k \right]^{-1} \left[\left(\mathbf{P}_{k|k}^{\xi} \right)^{-1} \hat{\xi}_{k|k} + \bar{\mathbf{F}}_k^T \Delta_k^{-1} \delta_k \right] \\ &= \hat{\xi}_{k|k} + \left[\left(\mathbf{P}_{k|k}^{\xi} \right)^{-1} + \bar{\mathbf{F}}_k^T \Delta_k^{-1} \bar{\mathbf{F}}_k \right]^{-1} \bar{\mathbf{F}}_k^T \Delta_k^{-1} (\delta_k - \bar{\mathbf{F}}_k \hat{\xi}_{k|k}) \\ &= \hat{\xi}_{k|k} + \mathbf{P}_{k|k}^{\xi} \bar{\mathbf{F}}_k^T \left(\Delta_k + \bar{\mathbf{F}}_k \mathbf{P}_{k|k}^{\xi} \bar{\mathbf{F}}_k^T \right)^{-1} (\delta_k - \bar{\mathbf{F}}_k \hat{\xi}_{k|k}).\end{aligned}$$

Let $\mu_k = \left(\Delta_k + \bar{\mathbf{F}}_k \mathbf{P}_{k|k}^{\xi} \bar{\mathbf{F}}_k^T \right)^{-1} (\delta_k - \bar{\mathbf{F}}_k \hat{\xi}_{k|k})$. Referring to [15], we have $\hat{\mathbf{x}}_{k+1|k} = \bar{\mathbf{F}}_k \hat{\xi}_{k|N}$. Then

$$\begin{aligned}\mu_k &= \left(\left[\left(\mathbf{P}_{k+1|N}^{\mathbf{x}} \right)^{-1} - \left(\mathbf{P}_{k+1|k}^{\mathbf{x}} \right)^{-1} \right]^{-1} + \mathbf{P}_{k+1|k}^{\mathbf{x}} \right)^{-1} \\ &\quad \cdot \left(\left[\left(\mathbf{P}_{k+1|N}^{\mathbf{x}} \right)^{-1} - \left(\mathbf{P}_{k+1|k}^{\mathbf{x}} \right)^{-1} \right]^{-1} \left[\left(\mathbf{P}_{k+1|N}^{\mathbf{x}} \right)^{-1} \hat{\mathbf{x}}_{k+1|N} - \left(\mathbf{P}_{k+1|k}^{\mathbf{x}} \right)^{-1} \bar{\mathbf{F}}_k \hat{\xi}_{k|k} \right] - \bar{\mathbf{F}}_k \hat{\xi}_{k|k} \right) \\ &= \left(\left[\left(\mathbf{P}_{k+1|N}^{\mathbf{x}} \right)^{-1} - \left(\mathbf{P}_{k+1|k}^{\mathbf{x}} \right)^{-1} \right]^{-1} + \mathbf{P}_{k+1|k}^{\mathbf{x}} \right)^{-1} \\ &\quad \cdot \left[\left(\mathbf{P}_{k+1|N}^{\mathbf{x}} \right)^{-1} - \left(\mathbf{P}_{k+1|k}^{\mathbf{x}} \right)^{-1} \right]^{-1} \left(\mathbf{P}_{k+1|N}^{\mathbf{x}} \right)^{-1} (\hat{\mathbf{x}}_{k+1|N} - \bar{\mathbf{F}}_k \hat{\xi}_{k|k}) \\ &= \left(\mathbf{P}_{k+1|k}^{\mathbf{x}} \right)^{-1} (\hat{\mathbf{x}}_{k+1|N} - \bar{\mathbf{F}}_k \hat{\xi}_{k|k}).\end{aligned}$$

Hence,

$$\begin{aligned}\hat{\xi}_{k|N} &= \hat{\xi}_{k|k} + \mathbf{P}_{k|k}^{\xi} \bar{\mathbf{F}}_k^T \mu_k \\ &= \hat{\xi}_{k|k} + \mathbf{P}_{k|k}^{\xi} \bar{\mathbf{F}}_k^T \left(\mathbf{P}_{k+1|k}^{\mathbf{x}} \right)^{-1} (\hat{\mathbf{x}}_{k+1|N} - \bar{\mathbf{F}}_k \hat{\xi}_{k|k}) \\ &= \hat{\xi}_{k|k} + \mathbf{K}_k (\hat{\mathbf{x}}_{k+1|N} - \bar{\mathbf{F}}_k \hat{\xi}_{k|k}).\end{aligned}$$

By (12), the error covariance is

$$\mathbf{P}_{k|N}^{\xi} = \left[\left(\mathbf{P}_{k|k}^{\xi} \right)^{-1} + \bar{\mathbf{F}}_k^T \Delta_k^{-1} \bar{\mathbf{F}}_k \right]^{-1}$$

It is known from [15] that $\mathbf{P}_{k+1|k}^{\mathbf{x}} = \bar{\mathbf{F}}_k \mathbf{P}_{k|k}^{\xi} \bar{\mathbf{F}}_k^T + \mathbf{Q}_k$. Then

$$\begin{aligned}\mathbf{P}_{k|N}^{\xi} &= \mathbf{P}_{k|k}^{\xi} - \mathbf{P}_{k|k}^{\xi} \bar{\mathbf{F}}_k^T \left(\left[\left(\mathbf{P}_{k+1|N}^{\mathbf{x}} \right)^{-1} - \left(\mathbf{P}_{k+1|k}^{\mathbf{x}} \right)^{-1} \right]^{-1} + \mathbf{P}_{k+1|k}^{\mathbf{x}} \right)^{-1} \bar{\mathbf{F}}_k \mathbf{P}_{k|k}^{\xi} \\ &= \mathbf{P}_{k|k}^{\xi} + \mathbf{P}_{k|k}^{\xi} \bar{\mathbf{F}}_k^T \left(\mathbf{P}_{k+1|k}^{\mathbf{x}} \right)^{-1} \left(\mathbf{P}_{k+1|N}^{\mathbf{x}} - \mathbf{P}_{k+1|k}^{\mathbf{x}} \right) \left(\mathbf{P}_{k+1|k}^{\mathbf{x}} \right)^{-1} \bar{\mathbf{F}}_k \mathbf{P}_{k|k}^{\xi} \\ &= \mathbf{P}_{k|k}^{\xi} + \mathbf{K}_k \left(\mathbf{P}_{k+1|N}^{\mathbf{x}} - \mathbf{P}_{k+1|k}^{\mathbf{x}} \right) \mathbf{K}_k^T.\end{aligned}$$

The proof is thus completed. $\square$

The forward filtering formulae corresponding to the smoothing ones in Corollary 1 are available in [10]. The methodology used in [10] is based on filter design and MVUE-based filter gain selection, but the obtained filter can also be interpreted as the result of Bayesian MAP estimation [24].

4. EXTENSION TO SYSTEMS WITHOUT DIRECT FEEDTHROUGH

In this section, we examine extension of the Bayesian smoothing of inputs and states to systems without direct feedthrough. Consider

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) + \mathbf{w}_k, \\ \mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k. \end{cases} \quad (18)$$

The above system is different from (1) in that the first measurement containing the information of $\mathbf{u}_k$ is $\mathbf{y}_{k+1}$ instead of $\mathbf{y}_k$. This, however, will render nontrivial the extension of results in Sections 2 and 3. The following discussion will begin with the Bayesian SISS paradigm for (18), followed by an examination of SISS for linear systems and possible solutions to the nonlinear case.

We continue to use the notation $\xi_k = [\mathbf{u}_{k-1}^T \mathbf{x}_k^T]^T$. The same assumption as A1 is adopted, because an unknown signal such as $\{\mathbf{u}_k\}$ can be regarded white as a result of its unpredictable variation in magnitude over time scale. Accordingly, the independence of $\mathbf{u}_k$ from $\mathbb{X}_k$ and $\mathbb{Y}_k$ is implied.

The next theorem establishes the Bayesian paradigm for computation of $p(\xi_k | \mathbb{Y}_{1:N})$.

Theorem 4

Suppose that A1 holds. Given the system in (18), the Bayesian smoothing paradigm for input and state estimation is given by

$$p(\xi_k | \mathbb{Y}_{1:N}) = p(\xi_k | \mathbb{Y}_{1:k}) \iint \frac{p(\mathbf{x}_{k+1} | \mathbf{x}_k, \mathbf{u}_k) p(\xi_{k+1} | \mathbb{Y}_{1:N})}{\int p(\mathbf{x}_{k+1} | \mathbf{x}_k, \mathbf{u}_k) p(\mathbf{x}_k | \mathbb{Y}_{1:k}) d\mathbf{x}_k} d\mathbf{u}_k d\mathbf{x}_{k+1}. \quad (19)$$

Proof

It is seen that

$$\begin{aligned} p(\xi_k | \mathbb{Y}_{1:N}) &= \iiint p(\mathbf{u}_{k-1}, \mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{x}_k, \mathbf{x}_{k+1} | \mathbb{Y}_{1:N}) d\mathbf{x}_{k-1} d\mathbf{u}_k d\mathbf{x}_{k+1} \\ &= \iiint p(\mathbf{u}_{k-1}, \mathbf{x}_{k-1} | \mathbf{u}_k, \mathbf{x}_k, \mathbf{x}_{k+1}, \mathbb{Y}_{1:N}) \\ &\quad \cdot p(\mathbf{x}_k | \mathbf{u}_k, \mathbf{x}_{k+1}, \mathbb{Y}_{1:N}) p(\mathbf{u}_k, \mathbf{x}_{k+1} | \mathbb{Y}_{1:N}) d\mathbf{x}_{k-1} d\mathbf{u}_k d\mathbf{x}_{k+1}. \end{aligned} \quad (20)$$

The following equations hold:

$$\begin{aligned} p(\mathbf{u}_{k-1}, \mathbf{x}_{k-1} | \mathbf{u}_k, \mathbf{x}_k, \mathbf{x}_{k+1}, \mathbb{Y}_{1:N}) &= p(\mathbf{u}_{k-1}, \mathbf{x}_{k-1} | \mathbf{x}_k, \mathbb{Y}_{1:k}) \\ &= \frac{p(\mathbf{x}_{k-1} | \mathbf{u}_{k-1}, \mathbf{x}_k, \mathbb{Y}_{1:k}) p(\mathbf{u}_{k-1}, \mathbf{x}_k | \mathbb{Y}_{1:k})}{p(\mathbf{x}_k | \mathbb{Y}_{1:k})}, \end{aligned} \quad (21)$$

$$\begin{aligned} p(\mathbf{x}_k | \mathbf{u}_k, \mathbf{x}_{k+1}, \mathbb{Y}_{1:N}) &= p(\mathbf{x}_k | \mathbf{u}_k, \mathbf{x}_{k+1}, \mathbb{Y}_{1:k}) \\ &= \frac{p(\mathbf{x}_{k+1} | \mathbf{x}_k, \mathbf{u}_k) p(\mathbf{x}_k, \mathbf{u}_k | \mathbb{Y}_{1:k})}{\int p(\mathbf{x}_{k+1} | \mathbf{x}_k, \mathbf{u}_k) p(\mathbf{x}_k, \mathbf{u}_k | \mathbb{Y}_{1:k}) d\mathbf{x}_k} \\ &= \frac{p(\mathbf{x}_{k+1} | \mathbf{x}_k, \mathbf{u}_k) p(\mathbf{x}_k | \mathbb{Y}_{1:k})}{\int p(\mathbf{x}_{k+1} | \mathbf{x}_k, \mathbf{u}_k) p(\mathbf{x}_k | \mathbb{Y}_{1:k}) d\mathbf{x}_k}, \end{aligned} \quad (22)$$

due to the Markovian state propagation and the assumption A6. Substituting (21) and (22) into (20) and doing integration with respect to $\mathbf{x}_{k-1}$, we obtain (19). $\square$

Theorem 4 establishes the Bayesian solution framework to address SISS for systems without direct feedthrough. However, a careful observation reveals that it is an arguably formidable task to derive a viable SISS method based on Theorem 4 for a nonlinear system of general form as shown

in (18). The fundamental reason is $\mathbf{u}_k$ and $\mathbf{x}_{k+1}$ being coupled nonlinearly. Hence, we consider a linear system as follows:

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{F}_k \mathbf{x}_k + \mathbf{G}_k \mathbf{u}_k + \mathbf{w}_k, \\ \mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k, \end{cases} \quad (23)$$

where $\{\mathbf{w}_k\}$ and $\mathbf{v}_k$ are zero-mean white Gaussian noise variables with covariances $\mathbf{Q}_k$ and $\mathbf{R}_k$, respectively, and $\mathbf{u}_k$ is assumed white Gaussian yet without knowledge of its statistical moments.

Theorem 5

For the system in (23), the smoother $\xi_{k|N} = \arg \max_{\xi_k} p(\xi_k | \mathbb{Y}_{1:N})$ can be rewritten as

$$\hat{\xi}_{k|N} = \arg \min_{\xi_k} \ell(\xi_k), \quad (24)$$

where

$$\ell(\xi_k) := \left\| \xi_k - \hat{\xi}_{k|k} \right\|_{\mathbf{P}_{k|k}^{\xi}}^2 + \left\| \mathbf{F}_k \mathbf{N} \xi_k - \delta_k \right\|_{\Delta_k}^2, \quad (25)$$

$$\mathbf{M}_k = [-\mathbf{G}_k \ \mathbf{I}], \quad \mathbf{N} = [\mathbf{0} \ \mathbf{I}], \quad (26)$$

$$\Delta_k = \left[\left(\mathbf{M}_k \mathbf{P}_{k+1|N}^{\xi} \mathbf{M}_k^T \right)^{-1} - \left(\mathbf{P}_{k+1|k}^{\mathbf{x}*} \right)^{-1} \right]^{-1} + \mathbf{Q}_k \quad (27)$$

$$\delta_k = \mathbf{F}_k \hat{\mathbf{x}}_{k|k} - \mathbf{P}_{k+1|k}^{\mathbf{x}*} \left(\mathbf{M}_k \mathbf{P}_{k+1|N}^{\xi} \mathbf{M}_k^T - \mathbf{P}_{k+1|k}^{\mathbf{x}*} \right)^{-1} \left(\mathbf{M}_k \hat{\xi}_{k+1|N} - \mathbf{F}_k \hat{\mathbf{x}}_{k|k} \right), \quad (28)$$

$$\mathbf{P}_{k+1|k}^{\mathbf{x}*} = \mathbf{F}_k \mathbf{P}_{k|k}^{\mathbf{x}} \mathbf{F}_k^T + \mathbf{Q}_k. \quad (29)$$

Furthermore, the solution to (24) is

$$\mathbf{K}_k = \mathbf{P}_{k|k}^{\xi} \mathbf{N}^T \mathbf{F}_k^T \left(\mathbf{P}_{k+1|k}^{\mathbf{x}*} \right)^{-1}, \quad (30)$$

$$\hat{\xi}_{k|N} = \hat{\xi}_{k|k} + \mathbf{K}_k \left(\mathbf{M}_k \hat{\xi}_{k+1|N} - \mathbf{F}_k \hat{\mathbf{x}}_{k|k} \right), \quad (31)$$

$$\mathbf{P}_{k|N}^{\xi} = \mathbf{P}_{k|k}^{\xi} + \mathbf{K}_k \left[\mathbf{M}_k \mathbf{P}_{k+1|N}^{\xi} \mathbf{M}_k^T - \left(\mathbf{P}_{k+1|k}^{\mathbf{x}*} \right) \right]^{-1} \mathbf{K}_k^T. \quad (32)$$

Proof

The proof is grounded on the Bayesian paradigm in (19). With the assumptions A6–A9, all the involved probabilities in (19) can be expressed as Gaussian functions. First consider

$$\int \mathcal{G}(\mathbf{x}_{k+1}; \mathbf{F}_k \mathbf{x}_k + \mathbf{G}_k \mathbf{u}_k, \mathbf{Q}_k) \cdot \mathcal{G}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}^{\mathbf{x}}) d\mathbf{x}_k. \quad (33)$$

By Lemma A.5, it is proportional to

$$\mathcal{G}(\mathbf{M}_k \xi_{k+1}; \hat{\mathbf{x}}_{k+1|k}^*, \mathbf{P}_{k+1|k}^{\mathbf{x}*}),$$

where

$$\begin{aligned} \hat{\mathbf{x}}_{k+1|k}^* &= \mathbf{F}_k \hat{\mathbf{x}}_{k|k}, \\ \mathbf{P}_{k+1|k}^{\mathbf{x}*} &= \mathbf{F}_k \mathbf{P}_{k|k}^{\mathbf{x}} \mathbf{F}_k^T + \mathbf{Q}_k. \end{aligned}$$

Then it is seen from Lemma A.4 that

$$\frac{\mathcal{G}(\xi_{k+1}; \hat{\xi}_{k+1|N}, \mathbf{P}_{k+1|N}^{\xi})}{\mathcal{G}(\mathbf{M}_k \xi_{k+1}; \hat{\mathbf{x}}_{k+1|K}^*, \mathbf{P}_{k+1|k}^{\mathbf{x}*})} = \mathcal{G}(\xi_{k+1}; \bar{\xi}_{k+1}, \bar{\mathbf{P}}_{k+1}^{\xi}),$$

where

$$\begin{aligned}\bar{\mathbf{P}}_{k+1}^{\xi} &= \left[\left(\mathbf{P}_{k+1|N}^{\xi} \right)^{-1} - \mathbf{M}_k^T \left(\mathbf{P}_{k+1|k}^{\mathbf{x}^*} \right)^{-1} \mathbf{M}_k \right]^{-1}, \\ \bar{\xi}_{k+1} &= \bar{\mathbf{P}}_{k+1}^{\xi} \left[\left(\mathbf{P}_{k+1|N}^{\xi} \right)^{-1} \hat{\xi}_{k+1|N} - \mathbf{M}_k^T \left(\mathbf{P}_{k+1|k}^{\mathbf{x}^*} \right)^{-1} \hat{\mathbf{x}}_{k+1|k}^* \right].\end{aligned}$$

Following Lemma A.4, we have

$$\begin{aligned}& \iint \mathcal{G}(\mathbf{x}_{k+1}; \mathbf{F}_k \mathbf{x}_k + \mathbf{G}_k \mathbf{u}_k, \mathbf{Q}_k) \cdot \mathcal{G}(\xi_{k+1}; \bar{\xi}_{k+1}, \bar{\mathbf{P}}_{k+1}^{\xi}) d\mathbf{u}_k d\mathbf{x}_{k+1} \\ & \propto \int \mathcal{G}(\mathbf{M}_k \xi_{k+1}; \mathbf{F}_k \mathbf{x}_k, \mathbf{Q}_k) \cdot \mathcal{G}(\xi_{k+1}; \bar{\xi}_{k+1}, \bar{\mathbf{P}}_{k+1}^{\xi}) d\xi_{k+1} \\ & \propto \mathcal{G}(\mathbf{F}_k \mathbf{x}_k; \delta_k, \Delta_k),\end{aligned}$$

where

$$\begin{aligned}\Delta_k &= \mathbf{M}_k \bar{\mathbf{P}}_{k+1}^{\xi} \mathbf{M}_k^T + \mathbf{Q}_k \\ &= \mathbf{M}_k \left[\left(\mathbf{P}_{k+1|N}^{\xi} \right)^{-1} - \mathbf{M}_k^T \left(\mathbf{P}_{k+1|k}^{\mathbf{x}^*} \right)^{-1} \mathbf{M}_k \right]^{-1} \mathbf{M}_k^T + \mathbf{Q}_k \\ &= \left[\left(\mathbf{M}_k \mathbf{P}_{k+1|N}^{\xi} \mathbf{M}_k^T \right)^{-1} - \left(\mathbf{P}_{k+1|k}^{\mathbf{x}^*} \right)^{-1} \right]^{-1} + \mathbf{Q}_k, \\ \delta_k &= \mathbf{M}_k \bar{\xi}_{k+1} - \eta_k \\ &= \mathbf{M}_k \left[\left(\mathbf{P}_{k+1|N}^{\xi} \right)^{-1} - \mathbf{M}_k^T \left(\mathbf{P}_{k+1|k}^{\mathbf{x}^*} \right)^{-1} \mathbf{M}_k \right]^{-1} \\ & \quad \cdot \left[\left(\mathbf{P}_{k+1|N}^{\xi} \right)^{-1} \hat{\xi}_{k+1|N} - \mathbf{M}_k^T \left(\mathbf{P}_{k+1|k}^{\mathbf{x}^*} \right)^{-1} \hat{\mathbf{x}}_{k+1|k}^* \right] \\ &= \mathbf{x}_{k+1|k}^* - \mathbf{P}_{k+1|k}^{\mathbf{x}^*} \left(\mathbf{M}_k \mathbf{P}_{k+1|N}^{\xi} \mathbf{M}_k^T - \mathbf{P}_{k+1|k}^{\mathbf{x}^*} \right)^{-1} \left(\mathbf{M}_k \hat{\xi}_{k+1|N} - \hat{\mathbf{x}}_{k+1|k}^* \right) \\ &= \mathbf{F}_k \hat{\mathbf{x}}_{k|k} - \mathbf{P}_{k+1|k}^{\mathbf{x}^*} \left(\mathbf{M}_k \mathbf{P}_{k+1|N}^{\xi} \mathbf{M}_k^T - \mathbf{P}_{k+1|k}^{\mathbf{x}^*} \right)^{-1} \left(\mathbf{M}_k \hat{\xi}_{k+1|N} - \hat{\mathbf{x}}_{k+1|k}^* \right).\end{aligned}$$

Then the likelihood function $L(\xi_k)$ is

$$L(\xi_k) = p(\xi_k | \mathbb{Y}_{1:N}) = \mathcal{G}(\xi_k; \hat{\xi}_{k|k}, \mathbf{P}_{k|k}^{\xi}) \cdot \mathcal{G}(\mathbf{F}_k \mathbf{N} \xi_k; \delta_k, \Delta_k),$$

the logarithmic version of which, $\ell(\xi_k)$, is given by (25).

By Lemma A.6, the solution to (24) is

$$\begin{aligned}\hat{\xi}_{k|N} &= \hat{\xi}_{k|k} + \mathbf{P}_{k|k}^{\xi} \mathbf{N}^T \mathbf{F}_k^T \left(\Delta_k + \mathbf{F}_k \mathbf{N} \mathbf{P}_{k|k}^{\xi} \mathbf{N}^T \mathbf{F}_k^T \right)^{-1} \left(\delta_k - \mathbf{F}_k \mathbf{N} \hat{\xi}_{k|k} \right) \\ &= \hat{\xi}_{k|k} - \mathbf{P}_{k|k}^{\xi} \mathbf{N}^T \mathbf{F}_k^T \left[\left(\left(\mathbf{M}_k \mathbf{P}_{k+1|N}^{\xi} \mathbf{M}_k^T \right)^{-1} - \left(\mathbf{P}_{k+1|k}^{\mathbf{x}^*} \right)^{-1} \right)^{-1} + \mathbf{P}_{k+1|k}^{\mathbf{x}^*} \right]^{-1} \\ & \quad \cdot \mathbf{P}_{k+1|k}^{\mathbf{x}^*} \left(\mathbf{M}_k \mathbf{P}_{k+1|N}^{\xi} \mathbf{M}_k^T - \mathbf{P}_{k+1|k}^{\mathbf{x}^*} \right)^{-1} \left(\mathbf{M}_k \hat{\xi}_{k+1|N} - \hat{\mathbf{x}}_{k+1|k}^* \right) \\ &= \hat{\xi}_{k|k} + \mathbf{P}_{k|k}^{\xi} \mathbf{N}^T \mathbf{F}_k^T \left(\mathbf{P}_{k+1|k}^{\mathbf{x}^*} \right)^{-1} \left(\mathbf{M}_k \hat{\xi}_{k+1|N} - \hat{\mathbf{x}}_{k+1|k}^* \right) \\ &= \hat{\xi}_{k|k} + \mathbf{K}_k \left(\mathbf{M}_k \hat{\xi}_{k+1|N} - \hat{\mathbf{x}}_{k+1|k}^* \right),\end{aligned}$$

where $\mathbf{K}_k = \mathbf{P}_{k|k}^\xi \mathbf{N}^T \mathbf{F}_k^T (\mathbf{P}_{k+1|k}^{x*})^{-1}$. The associated smoothing covariance is then given by

$$\begin{aligned} \mathbf{P}_{k|N}^\xi &= \mathbf{P}_{k|k}^\xi - \mathbf{P}_{k|k}^\xi \mathbf{N}^T \mathbf{F}_k^T (\Delta_k + \mathbf{F}_k \mathbf{N} \mathbf{P}_{k|k}^\xi \mathbf{N}^T \mathbf{F}_k^T)^{-1} \mathbf{F}_k \mathbf{N} \mathbf{P}_{k|k}^\xi \\ &= \mathbf{P}_{k|k}^\xi - \mathbf{P}_{k|k}^\xi \mathbf{N}^T \mathbf{F}_k^T \left[\left((\mathbf{M}_k \mathbf{P}_{k+1|N}^\xi \mathbf{M}_k^T)^{-1} - (\mathbf{P}_{k+1|k}^{x*})^{-1} \right)^{-1} + \mathbf{P}_{k+1|k}^{x*} \right]^{-1} \mathbf{F}_k \mathbf{N} \mathbf{P}_{k|k}^\xi \\ &= \mathbf{P}_{k|k}^\xi + \mathbf{K}_k \left[\mathbf{M}_k \mathbf{P}_{k+1|N}^\xi \mathbf{M}_k^T - \mathbf{P}_{k+1|k}^{x*} \right]^{-1} \mathbf{K}_k^T. \end{aligned}$$

This concludes the proof. $\square$

Theorem 5 formulates and solves a MAP-optimization-based SISS problem for the linear system (23). The backward smoothing is governed by (30)–(32). The reader is referred to [9] for the associated forward filtering equations. Like in [10], the SISE filter [9] is designed on the basis of MVUE, but it can also be derived using MAP estimation in a Bayesian framework, as shown in [24].

However, it should be pointed out that SISS for the nonlinear system in (18) is rather difficult. Because $\mathbf{u}_{k-1}$ and $\mathbf{x}_k$ differ by one time unit and are nonlinearly coupled, it remains a challenge to build an optimization problem based on Theorem 4 without introducing approximation that compromises the smoothing accuracy. While leaving the problem beyond the scope of this paper, we suggest two possible alternative solutions. First, if the process noise $\mathbf{w}_k$ is approximately ignorable, we can apply the results of Theorem 3 after rewriting (18) as

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), \\ \mathbf{z}_k = \mathbf{h}(\mathbf{f}(\mathbf{x}_k, \mathbf{u}_k)) + \bar{\mathbf{v}}_k, \end{cases} \quad (34)$$

where $\mathbf{z}_k = \mathbf{y}_{k+1}$ and $\bar{\mathbf{v}}_k = \mathbf{v}_{k+1}$. The other one is to perform one-step fixed-lag smoothing, which can be combined with filtering indeed. It can be proven that

$$p(\mathbf{u}_{k-1}, \mathbf{x}_{k-1}, \mathbf{x}_k | \mathbb{Y}_k) = \frac{p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{u}_{k-1}) p(\mathbf{x}_{k-1} | \mathbb{Y}_{k-1}) p(\mathbf{u}_{k-1})}{p(\mathbf{y}_k | \mathbb{Y}_{k-1})},$$

upon which an optimization problem can be built and solved analogously to estimate $\mathbf{u}_{k-1}$, $\mathbf{x}_{k-1}$ and $\mathbf{x}_k$ from $\mathbb{Y}_k$ simultaneously.

5. APPLICATION EXAMPLE

Reconstruction of flow fields has been a research subject of intense interest for its vital role in helping people, especially the oceanographers, understand the ocean. Flows are known to be crucial for fishing, shipping, navigation, weather forecasting, environmental monitoring and climate change. To study the flows, a novel ocean observing system has been conceived and is being built at the University of California, San Diego. It is based on a swarm of inexpensive buoyancy-controlled drifters [31, 32], which are deployed to traverse a domain of an ocean. The drifters are capable of arbitrary vertical migration while traveling along the flows. During the travel, each drifter measures and stores a time record of its depth, acceleration, position and some oceanographic quantities such as temperature and salinity. Its position information is measurable only intermittently when the drifter is over water surface, because GPS signals attenuate seriously in underwater circumstances. The record of data will be sent to a central server for processing when the drifter is at the surface. In this section, a three-dimensional flow domain is considered (Figure 2), the space of which is mainly occupied by two adjacent eddies of opposite direction of rotation.

5.1. Drifter dynamics

For simplicity, we only examine a drifter's motion along x-direction. The same results can apply to y-direction because of the independence of perpendicular components of motion. For a drifter, the

Figure 2. (a) The three-dimensional flow field; (b) top view of the flow field.

flow velocity $v(d_x, z)$ at its x -displacement d_x is time-stationary and dependent only on its depth z . The dynamics of a drifter within the flow field is described in [33]:

$$m\ddot{d}_x = c \cdot \text{sign} \left(v(d_x, z) - \dot{d}_x \right) \cdot \left(v(d_x, z) - \dot{d}_x \right)^2, \quad (35)$$

where m is the constant rigid mass and c is the drag parameter. The right-hand side of the earlier equation represents the drag force that quantifies the resistance exercised on the drifter in the flow field.

From (35), we define two state variables $x_1 := d_x$ and $x_2 := \dot{d}_x$. Further, $v(d_x, z)$ can be viewed as the unknown external input into the drifter, naturally implying the definition of $u := v(d_x, z)$. Then (35) can be rewritten as,

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= \frac{c}{m} \cdot \text{sign} (u - x_2) \cdot (u - x_2)^2. \end{aligned} \quad (36)$$

Its discrete-time representation, obtained by hypothetically holding the input constant over half open intervals $[kT, (k+1)T)$, and doing forward finite difference and measuring output measurements at kT , is given by

$$\begin{aligned} x_{1,k+1} &= x_{1,k} + T \cdot x_{2,k}, \\ x_{2,k+1} &= x_{2,k} + T \cdot \frac{c}{m} \cdot \text{sign} (u_k - x_{2,k}) \cdot (u_k - x_{2,k})^2, \end{aligned} \quad (37)$$

where $u_k := u(kT)$ and $x_{i,k} := x_i(kT)$ for $i = 1, 2$. The preceding equation can be expressed as

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, u_k), \quad (38)$$

where $\mathbf{f}$ can be determined from the context.

The motion of the drifter is characterized by an irregularly cycling submerging/surfacing pattern — it submerges and moves underwater for a certain duration, then resurfaces and repeats the process over time. No matter whether it is underwater or on the surface, the depth $z_k := z(kT)$ and acceleration $\ddot{d}_{x,k} := \ddot{d}_x(kT)$ are measurable; however, the position $d_{x,k} := d_x(kT)$ can only be measured when it is at surface. Thus, irregularly sampled measurements arise as a result, with the fast one $\tau_k := \ddot{d}_{x,k}$ and slow one $\eta_k := d_{x,k}$ given by, respectively,

$$\begin{aligned} \tau_k &= \frac{c}{m} \cdot \text{sign} (u_k - x_{2,k}) \cdot (u_k - x_{2,k})^2, \\ \eta_k &= x_{1,k}. \end{aligned} \quad (39)$$

For simplicity of notation, we rewrite (39) as

$$\begin{aligned}\tau_k &= \varphi(u_k, x_k), \\ \eta_k &= \phi(x_k).\end{aligned}\quad (40)$$

Combining (37) and (39), we obtain the state-space model to describe the dynamics of the drifter:

$$\Sigma : \begin{cases} \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, u_k) + \mathbf{w}_k, \\ \mathbf{y}_k = \mathbf{h}(\mathbf{x}_k, u_k) + \mathbf{v}_k, \end{cases} \quad (41)$$

Here, when the drifter is underwater, $\mathbf{y}_k = \tau_k$ and $\mathbf{h} = \varphi$, when at surface, $\mathbf{y}_k = [\tau_k^T \eta_k^T]^T$ and $\mathbf{h} = [\varphi^T \phi^T]^T$.

In addition, $\mathbf{w}$ and $\mathbf{v}$ are added to account for noise in real world. They are assumed to be white Gaussian and independent of each other. The algorithm proposed in Section 3 is applicable to the system Σ in (41) to acquire the information estimates of not only the velocities of the flow field (unknown input variables) but also the trajectory and velocity profile of the drifter (state variables).

5.2. Numerical simulation

The flow field considered has dimensions of $(-1000, 1000) \text{ m} \times (-1000, 1000) \text{ m} \times (0, 30) \text{ m}$, and the eddies are centered at $(500, 500) \text{ m}$ and $(-500, -500) \text{ m}$, respectively, as shown in Figure 2. Compared to a typical one in real life, its scale is intentionally narrowed to reduce computational burden, but this does not restrict application of the proposed SISS algorithm to larger flow fields. Let

Figure 3. Smoothing results for the drifters released at $(-400, -600) \text{ m}$: (a) x_1 — displacement along x -direction; (b) x_2 — velocity along x -direction; (c) u — flow velocity; (d) trajectory of the drifter (the circle denotes the location where the drifter is released).

20 drifters be deployed evenly along the line segment from $(-800, -1000)$ m to $(1000, 800)$ m. The mass of a drifter is 1.5 kg, the drag coefficient c is $2 \text{ N s}^2/\text{m}^2$ and the sampling period T is 0.05 s. The total traversing duration is 52 min.

The SISS algorithm presented in Theorem 3 is applied to smoothing the estimates of the inputs and states in the state-space model of the drifter. Let us examine the drifter released at $(-400, -600)$ m and consider its motion in the x-direction. Figure 3(a) and (b) demonstrate the estimation of the x-displacement, x-velocity and the flow velocity, respectively. The filtered estimates based on the algorithm in [24] are offered for comparison purpose. We observe that the smoothed estimates are overall closer to the truth than the filtered ones. This is particularly evident for input estimation in Figure 3(a) — the filtered input estimates are quite noisy, but the smoothing reduces

Table I. A quantified comparison between filtering and smoothing errors.

	Filtering	Smoothing
Err_u	340.02	90.61
Err_{x_1}	33931.97	19123.04
Err_{x_2}	83.08	26.89

Figure 4. (a) True surface flow velocity along x-direction; (b) smoothed surface flow velocity along x-direction; (c) true surface flow velocity along y-direction; (b) smoothed surface flow velocity along y-direction.

the errors significantly, thus improving the reconstruction accuracy. A quantitative comparison is further presented in Table I. The metric is accumulative estimation error. It is defined as

$$Err_s = \frac{1}{N} \sum_{k=1}^N (\hat{s}_k - s_k)^2,$$

where s_k for $k = 0, 1, \dots, N$ is a discrete-time signal and $\hat{s}_k$ is its estimate. It is seen from Table I that the smoothing errors are considerably smaller than the filtering errors, evidencing the enhancement of accuracy achieved by smoothing. Figure 3(d) shows the smoothed trajectory in comparison to the true one. It is seen that the smoothing algorithm exhibits high accuracy, with the estimated values agreeing well with the truth.

Further, the estimated inputs of all drifters, which are the smoothed flow velocity data at different locations, are put together and used to reconstruct the flow field *via* the tessellation-based linear interpolation. The first and second rows in Figure 4 illustrate the true and reconstructed flow velocity fields along x-direction and y-direction, respectively. It is seen that the estimated velocities at the two eddies and the transition area are quite close to the actual case. Despite some minor differences observed, the overall reconstruction accuracy is quite satisfactory. This indicates that the proposed SISS algorithm is able to provide reliable input and state estimates.

6. CONCLUSIONS

Simultaneous input and state estimation are challenging issues encountered in a broad range of areas and applications. Existing research has been concerned with the *filtering for SISE*. This paper, however, studies the *smoothing for SISE*, as demanded by the practical application of high accuracy but nonreal-time flow field reconstruction.

Forward-backward smoothing for SISE is considered for systems with and without direct feedthrough. It is treated from a statistical perspective, with a Bayesian framework constructed in the first place. MAP-based smoothing algorithms are then developed within the framework to obtain smoothed input and state estimates. The proposed smoothing algorithm for nonlinear systems with direct feedthrough is applied *via* a simulation-based study to flow field reconstruction where flow velocity profiles are estimated from the motion information of a group of buoyancy-controlled drifters, with the results illustrating their soundness. Further potential for this work is foreseen to address other multidisciplinary applications, which necessitate simultaneous estimation of the inputs and states. Additionally, the same methodology can be applied to develop fixed-internal or fixed-point input and state smoothing algorithms.

APPENDIX

Lemma A.1

(Matrix inversion lemma) [34] Let $\mathbf{A}$, $\mathbf{C}$ and $\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B}$ be nonsingular matrices. Then

$$(\mathbf{A} + \mathbf{B}\mathbf{C}\mathbf{D})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}\mathbf{A}^{-1}.$$

Lemma A.2

Let $\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2 > 0$. Given two Gaussian functions of $\mathbf{x} \in \mathbb{R}^n$, their multiplication leads to another Gaussian function, that is,

$$\mathcal{G}(\mathbf{x}; \bar{\mathbf{x}}_1, \bar{\mathbf{X}}_1) \cdot \mathcal{G}(\mathbf{x}; \bar{\mathbf{x}}_2, \bar{\mathbf{X}}_2) = \lambda \cdot \mathcal{G}(\mathbf{x}; \tilde{\mathbf{x}}, \tilde{\mathbf{X}}),$$

where

$$\begin{aligned} \tilde{\mathbf{X}} &= (\bar{\mathbf{X}}_1^{-1} + \bar{\mathbf{X}}_2^{-1})^{-1}, \\ \tilde{\mathbf{x}} &= \tilde{\mathbf{X}}(\bar{\mathbf{X}}_1^{-1}\bar{\mathbf{x}}_1 + \bar{\mathbf{X}}_2^{-1}\bar{\mathbf{x}}_2), \\ \lambda &= (2\pi)^{-\frac{n}{2}} |\bar{\mathbf{X}}_1|^{-\frac{1}{2}} |\bar{\mathbf{X}}_2|^{-\frac{1}{2}} |\tilde{\mathbf{X}}|^{\frac{1}{2}} \exp \left[-\frac{1}{2} \left(\|\bar{\mathbf{x}}_1\|_{\bar{\mathbf{X}}_1}^2 + \|\bar{\mathbf{x}}_2\|_{\bar{\mathbf{X}}_2}^2 - \|\tilde{\mathbf{x}}\|_{\tilde{\mathbf{X}}}^2 \right) \right]. \end{aligned}$$

Proof

The proof is straightforward and thus omitted here. $\square$

Lemma A.3

Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\bar{\mathbf{X}}, \mathbf{Y} > 0$, the following equation holds:

$$\mathcal{G}(\mathbf{x}; \bar{\mathbf{x}}, \bar{\mathbf{X}}) \cdot \mathcal{G}(\mathbf{x}; \mathbf{y}, \mathbf{Y}) \propto \mathcal{G}(\mathbf{x}; \tilde{\mathbf{x}}, \tilde{\mathbf{X}}) \cdot \mathcal{G}(\mathbf{y}; \tilde{\mathbf{y}}, \tilde{\mathbf{Y}}),$$

where

$$\begin{aligned}\tilde{\mathbf{X}} &= (\bar{\mathbf{X}}^{-1} + \mathbf{Y}^{-1})^{-1}, \\ \tilde{\mathbf{x}} &= \tilde{\mathbf{X}}(\bar{\mathbf{X}}^{-1}\bar{\mathbf{x}} + \mathbf{Y}^{-1}\mathbf{y}), \\ \tilde{\mathbf{Y}} &= \bar{\mathbf{X}} + \mathbf{Y}, \\ \tilde{\mathbf{y}} &= \bar{\mathbf{x}}.\end{aligned}$$

Proof

It follows from Lemma A.2 that

$$\mathcal{G}(\mathbf{x}; \bar{\mathbf{x}}, \bar{\mathbf{X}}) \cdot \mathcal{G}(\mathbf{x}; \mathbf{y}, \mathbf{Y}) = \lambda \cdot \mathcal{G}(\mathbf{x}; \tilde{\mathbf{x}}, \tilde{\mathbf{X}}).$$

In the preceding equation,

$$\begin{aligned}\lambda &= (2\pi)^{-\frac{n}{2}} |\bar{\mathbf{X}}|^{-\frac{1}{2}} |\mathbf{Y}|^{-\frac{1}{2}} |\tilde{\mathbf{X}}|^{\frac{1}{2}} \exp \left[-\frac{1}{2} \left(\|\bar{\mathbf{x}}\|_{\bar{\mathbf{X}}}^2 + \|\mathbf{y}\|_{\mathbf{Y}}^2 - \|\tilde{\mathbf{x}}\|_{\tilde{\mathbf{X}}}^2 \right) \right] \\ &= (2\pi)^{-\frac{n}{2}} |\bar{\mathbf{X}}|^{-\frac{1}{2}} |\mathbf{Y}|^{-\frac{1}{2}} |\tilde{\mathbf{X}}|^{\frac{1}{2}} \exp \left[-\frac{1}{2} \left(\|\mathbf{y} - \mathbf{U}\mathbf{V}\bar{\mathbf{x}}\|_{\mathbf{U}}^2 - \|\bar{\mathbf{x}}\|_{(\mathbf{V}\mathbf{U}\mathbf{V})^{-1}}^2 + \|\tilde{\mathbf{x}}\|_{\tilde{\mathbf{W}}}^2 \right) \right],\end{aligned}$$

where

$$\begin{aligned}\mathbf{U} &= \left(\mathbf{Y}^{-1} - \mathbf{Y}^{-1} (\bar{\mathbf{X}}^{-1} + \mathbf{Y}^{-1})^{-1} \mathbf{Y}^{-1} \right)^{-1} = \bar{\mathbf{X}} + \mathbf{Y}, \\ \mathbf{V} &= \mathbf{Y}^{-1} (\bar{\mathbf{X}}^{-1} + \mathbf{Y}^{-1})^{-1} \bar{\mathbf{X}}^{-1} = (\bar{\mathbf{X}} + \mathbf{Y})^{-1}, \\ \mathbf{W} &= \left(\bar{\mathbf{X}}^{-1} - \bar{\mathbf{X}}^{-1} (\bar{\mathbf{X}}^{-1} + \mathbf{Y}^{-1})^{-1} \bar{\mathbf{X}}^{-1} \right)^{-1} = \bar{\mathbf{X}} + \mathbf{Y}.\end{aligned}$$

It is noted that $\mathbf{U}\mathbf{V} = \mathbf{I}$, $(\mathbf{V}\mathbf{U}\mathbf{V})^{-1} = \mathbf{V}^{-1} = \mathbf{W}$. Hence,

$$\begin{aligned}\lambda &= |\bar{\mathbf{X}}|^{-\frac{1}{2}} |\mathbf{Y}|^{-\frac{1}{2}} |\tilde{\mathbf{X}}|^{\frac{1}{2}} |\bar{\mathbf{X}} + \mathbf{Y}|^{-\frac{1}{2}} \cdot (2\pi)^{-\frac{n}{2}} |\bar{\mathbf{X}} + \mathbf{Y}|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \|\mathbf{y} - \bar{\mathbf{x}}\|_{\bar{\mathbf{X}} + \mathbf{Y}}^2 \right] \\ &\propto \mathcal{G}(\mathbf{y}; \bar{\mathbf{x}}, \bar{\mathbf{X}} + \mathbf{Y}),\end{aligned}$$

which concludes the proof. $\square$

Lemma A.4

For $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\bar{\mathbf{X}}, \mathbf{Y} > 0$, the following equation holds:

$$\mathcal{G}(\mathbf{x}; \bar{\mathbf{x}}, \bar{\mathbf{X}}) \cdot \mathcal{G}(\mathbf{A}\mathbf{x}; \mathbf{y}, \mathbf{Y}) \propto \mathcal{G}(\mathbf{x}; \tilde{\mathbf{x}}, \tilde{\mathbf{X}}) \cdot \mathcal{G}(\mathbf{y}; \tilde{\mathbf{y}}, \tilde{\mathbf{Y}}),$$

where

$$\begin{aligned}\tilde{\mathbf{X}} &= (\bar{\mathbf{X}}^{-1} + \mathbf{A}^T \mathbf{Y}^{-1} \mathbf{A})^{-1}, \\ \tilde{\mathbf{x}} &= \tilde{\mathbf{X}}(\bar{\mathbf{X}}^{-1}\bar{\mathbf{x}} + \mathbf{A}^T \mathbf{Y}^{-1}\mathbf{y}), \\ \tilde{\mathbf{Y}} &= \mathbf{A}\bar{\mathbf{X}}\mathbf{A}^T + \mathbf{Y}, \\ \tilde{\mathbf{y}} &= \mathbf{A}\bar{\mathbf{x}}.\end{aligned}$$

Proof

The proof can be obtained following similar lines to that of Lemma A.3. $\square$

Lemma A.5

For $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{z} \in \mathbb{R}^l$, $\mathbf{A} \in \mathbb{R}^{l \times n}$, $\mathbf{B} \in \mathbb{R}^{l \times m}$ and $\bar{\mathbf{X}}, \mathbf{Z} > 0$, the following equation holds:

$$\mathcal{G}(\mathbf{x}; \bar{\mathbf{x}}, \bar{\mathbf{X}}) \cdot \mathcal{G}(\mathbf{z}; \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} + \mathbf{c}, \mathbf{Z}) \propto \mathcal{G}(\mathbf{x}; \tilde{\mathbf{x}}, \tilde{\mathbf{X}}) \cdot \mathcal{G}(\mathbf{D}\boldsymbol{\sigma}; \tilde{\boldsymbol{\sigma}}, \tilde{\boldsymbol{\Sigma}}),$$

where $\mathbf{D} = [-\mathbf{B} \ \mathbf{I}]$, $\boldsymbol{\sigma} = [\mathbf{y}^T \ \mathbf{z}^T]^T$, and

$$\begin{aligned}\tilde{\mathbf{X}} &= (\bar{\mathbf{X}}^{-1} + \mathbf{A}^T \mathbf{Z}^{-1} \mathbf{A})^{-1}, \\ \tilde{\mathbf{x}} &= \tilde{\mathbf{X}}(\bar{\mathbf{X}}^{-1} \bar{\mathbf{x}} + \mathbf{A}^T \mathbf{Z}^{-1} (\mathbf{D}\boldsymbol{\sigma} - \mathbf{c})), \\ \tilde{\boldsymbol{\Sigma}} &= \mathbf{A} \bar{\mathbf{X}} \mathbf{A}^T + \mathbf{Z}, \\ \tilde{\boldsymbol{\sigma}} &= \mathbf{A} \bar{\mathbf{x}} + \mathbf{c}.\end{aligned}$$

Proof

It is noted that

$$\mathcal{G}(\mathbf{z}; \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} + \mathbf{c}, \mathbf{Z}) = \mathcal{G}(\mathbf{A}\mathbf{x}; \mathbf{D}\boldsymbol{\sigma} - \mathbf{c}, \mathbf{Z}).$$

By Lemma A.4, we have

$$\begin{aligned}\mathcal{G}(\mathbf{x}; \bar{\mathbf{x}}, \bar{\mathbf{X}}) \cdot \mathcal{G}(\mathbf{A}\mathbf{x}; \mathbf{D}\boldsymbol{\sigma} - \mathbf{c}, \mathbf{Z}) &= \mathcal{G}(\mathbf{x}; \tilde{\mathbf{x}}, \tilde{\mathbf{X}}) \cdot \mathcal{G}(\mathbf{D}\boldsymbol{\sigma} - \mathbf{c}; \mathbf{A}\bar{\mathbf{x}}, \mathbf{A}\bar{\mathbf{X}}\mathbf{A}^T + \mathbf{Z}) \\ &= \mathcal{G}(\mathbf{x}; \tilde{\mathbf{x}}, \tilde{\mathbf{X}}) \cdot \mathcal{G}(\mathbf{D}\boldsymbol{\sigma}; \mathbf{A}\bar{\mathbf{x}} + \mathbf{c}, \mathbf{A}\bar{\mathbf{X}}\mathbf{A}^T + \mathbf{Z}) \\ &= \mathcal{G}(\mathbf{x}; \tilde{\mathbf{x}}, \tilde{\mathbf{X}}) \cdot \mathcal{G}(\mathbf{D}\boldsymbol{\sigma}; \tilde{\boldsymbol{\sigma}}, \tilde{\boldsymbol{\Sigma}}),\end{aligned}$$

which concludes the proof. □

Lemma A.6

For the minimization problem

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x} - \bar{\mathbf{x}}\|_{\bar{\mathbf{X}}}^2 + \|\mathbf{T}\mathbf{x} - \mathbf{y}\|_{\mathbf{Y}}^2,$$

the solution is given by

$$\hat{\mathbf{x}} = \bar{\mathbf{x}} + \bar{\mathbf{X}}\mathbf{T}^T (\mathbf{Y} + \mathbf{T}\bar{\mathbf{X}}\mathbf{T}^T)^{-1} (\mathbf{y} - \mathbf{T}\bar{\mathbf{x}}),$$

with the associated covariance

$$\begin{aligned}\mathbf{P} &= (\bar{\mathbf{X}}^{-1} + \mathbf{T}^T \mathbf{Y}^{-1} \mathbf{T})^{-1} \\ &= \bar{\mathbf{X}} - \bar{\mathbf{X}}\mathbf{T}^T (\mathbf{Y} + \mathbf{T}\bar{\mathbf{X}}\mathbf{T}^T)^{-1} \mathbf{T}\bar{\mathbf{X}}.\end{aligned}$$

Proof

The aforementioned minimization problem is equivalent to the standard maximum likelihood estimation. The derivation of its solution is available in [28, Section 5.3]. □

ACKNOWLEDGEMENTS

This work was partially supported by NSF award OCE-0941692.

REFERENCES

1. Schubert U, Kruger U, Wozny G, Arellano-Garcia Harvey. Input reconstruction for statistical-based fault detection and isolation. *AIChE Journal* 2012; **58**(5):1513–1523.
2. Imine H, Delanne Y, M'Sirdi NK. Road profile input estimation in vehicle dynamics simulation. *Vehicle System Dynamics* 2006; **44**(4):285–303.
3. Kitanidist PK. Unbiased minimum-variance linear state estimation. *Automatica* 1987; **23**(6):775–778.
4. Mendel J. White-noise estimators for seismic data processing in oil exploration. *Automatic Control, IEEE Transactions on* 1977; **22**(5):694–706.

5. Hsieh Chien-Shu. Robust two-stage Kalman filters for systems with unknown inputs. *IEEE Transactions on Automatic Control* 2000; **45**(12):2374–2378.
6. Floquet T, Edwards C, Spurgeon SK. On sliding mode observers for systems with unknown inputs. *International Journal of Adaptive Control and Signal Processing* 2007; **21**(8–9):638–656.
7. Kalsi K, Lian J, Hui S, Žak SH. Sliding-mode observers for systems with unknown inputs: a high-gain approach. *Automatica* 2010; **46**(2):347–353.
8. Lee H, Tahk M-J. Generalized input-estimation technique for tracking maneuvering targets. *IEEE Transactions on Aerospace and Electronic Systems* 1999; **35**(4):1388–1402.
9. Gillijns S, Moor BD. Unbiased minimum-variance input and state estimation for linear discrete-time systems. *Automatica* 2007; **43**(1):111–116.
10. Gillijns S, Moor BD. Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough. *Automatica* 2007; **43**(5):934–937.
11. Fang H, de Callafon RA. On the asymptotic stability of minimum-variance unbiased input and state estimation. *Automatica* 2012; **48**(12):3183–3186.
12. Fang H, Shi Y, Yi J. On stable simultaneous input and state estimation for discrete-time linear systems. *International Journal of Adaptive Control and Signal Processing* 2011; **25**(8):671–686.
13. Corless M, Tu J. State and input estimation for a class of uncertain systems. *Automatica* 1998; **34**(6):757–764.
14. Ha QP, Trinh H. State and input simultaneous estimation for a class of nonlinear systems. *Automatica* 2004; **40**(10):1779–1785.
15. Fang H, de Callafon RA. Nonlinear simultaneous input and state estimation with application to flow field estimation. *IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, Orlando, Florida, 2011; 6013–6018.
16. Kitanidis PK. Unbiased minimum-variance linear state estimation. *Automatica* 1987; **23**(6):775–778.
17. Darouach M, Zasadzinski M. Unbiased minimum variance estimation for systems with unknown exogenous inputs. *Automatica* 1997; **33**(4):717–719.
18. Darouach M, Zasadzinski M, Boutayeb M. Extension of minimum variance estimation for systems with unknown inputs. *Automatica* 2003; **39**(5):867–876.
19. Bejarano FJ, Fridman L, Poznyak A. Exact state estimation for linear systems with unknown inputs based on hierarchical super-twisting algorithm. *International Journal of Robust and Nonlinear Control* 2007; **17**(18):1734–1753.
20. Gao Z., Ho DWC. Proportional multiple-integral observer design for descriptor systems with measurement output disturbances. *IET Control Theory and Applications* 2004; **151**(3):279–288.
21. Gao Z, Wang H. Descriptor observer approaches for multivariable systems with measurement noises and application in fault detection and diagnosis. *Systems & Control Letters* 2006; **55**(4):304–313.
22. Lee H, Tahk M-J. Generalized input-estimation technique for tracking maneuvering targets. *IEEE Transactions on Aerospace and Electronic Systems* 1999; **35**(4):1388–1402.
23. Kitanidis PK. Unbiased minimum-variance linear state estimation. *Automatica* 1987; **23**(6):775–778.
24. Fang H, de Callafon RA, Cortés J. Simultaneous input and state estimation for nonlinear systems with applications to flow field estimation. *Automatica* 2013; **49**(9):2805–2812.
25. Gut A. *Probability: A Graduate Course*. Springer: New York, 2005.
26. Anderson BDO, Moore JB. *Optimal Filtering*. Prentice-Hall: Englewood Cliffs, 1979.
27. Ito K, Xiong K. Gaussian filters for nonlinear filtering problems. *IEEE Transactions on Automatic Control* 2000; **45**(5):910–927.
28. Candy JV. *Bayesian Signal Processing: Classical, Modern and Particle Filtering Methods*. Wiley-Interscience: New York, NY, USA, 2009.
29. Björck Å. *Numerical Methods for Least Squares Problems*. SIAM: Philadelphia, 1996.
30. Bell BM, Cathey FW. The iterated Kalman filter update as a Gauss–Newton method. *IEEE Transactions on Automatic Control* 1993; **38**(2):294–297.
31. Colgan C. Underwater laser shows. *Explorations, Scripps Institution of Oceanography* 2006; **12**(4):20–27.
32. Han Y, de Callafon RA, Cortés J, Jaffe J. Dynamic modeling and pneumatic switching control of a submersible drogue. *International Conference on Informatics in Control, Automation and Robotics*, Vol. 2, Funchal, Madeira, Portugal, June 2010; 89–97.
33. Booth DA. On the use of drogues for measuring subsurface ocean currents. *Ocean Dynamics* 1981; **34**:284–294.
34. Henderson HV, Searle SR. On deriving the inverse of a sum of matrices. *SIAM Review* 1981; **23**:53–60.