2013 American Control Conference (ACC)
Washington, DC, USA, June 17-19, 2013

Simultaneous Input and State Smoothing and Its Application to Oceanographic Flow Field Reconstruction

Huazhen Fang and Raymond A. de Callafon

Abstract— **Forward-backward smoothing of unknown inputs and states of a nonlinear system is studied in this paper, motivated by oceanographic flow field reconstruction using a swarm of buoyancy-controlled drogues. A Bayesian paradigm is developed first to provide a statistics based solution framework. A nonlinear maximum a posteriori (MAP) optimization problem is established within the framework as a means to achieve simultaneous input and state smoothing, which is solved by the iteration based Gauss-Newton method. Application of the proposed method to reconstruction of a complex three-dimensional flow field is investigated via simulation studies.**

I. Introduction

Joint estimation of unknown inputs and states is of considerable importance in many applications that arise in fault detection, automotive engineering, weather forecasting and oceanography [1–4]. Consequently, research on simultaneous input and state estimation (SISE) has been gaining increasing momentum in the past few years.

Literature review: A lead has been taken in [5] with the development of estimators and smoothers of white noises regarded as external inputs to a linear system. A large body of more recent work has considered completely unknown inputs, and a majority of them are built upon existing state estimation methods. A two-stage KF is proposed in [6] for SISE through decoupled design of input and state filters. Sliding-mode state observers are extended in [7; 8] such that unknown input variables can also be estimated. The notion of moving horizon estimation (MHE) is used in [9]. In [10; 11], minimum-variance unbiased estimation (MVUE) is applied to derive SISE filters for systems with and without direct input-output feedthrough, respectively. The same technique is also used in [12] to construct suboptimal SISE filters with proven stability properties. While the works highlighted above consider only linear systems, SISE for nonlinear systems has also been discussed, though far from extensively, in the literature. SISE for a special class of nonlinear systems composed of a nominally linear part and a nonlinear part is studied in [13; 14]. It is pointed out in [4] that SISE can be tackled using Bayesian statistics and Maximum a Posteriori (MAP) estimation, the proposed algorithms being applicable to nonlinear systems in general form. Another problem of relevance to SISE is state estimation with unknown inputs. In this case, despite no knowledge of inputs, only the states are estimated from the output measurements. Among an abundance of works on this topic, we underscore [3; 15; 16] on MVUE and [17] on sliding mode based approaches.

H. Fang and R.A. de Callafon are with Department of Mechanical and Aerospace Engineering, University of California, San Diego, CA 92093, USA. {hzfang,callafon}@ucsd.edu

Statement of contributions: Simultaneous input and state smoothing (SISS) for nonlinear systems is investigated in this paper. The inputs are assumed completely unknown. We pose this problem in a Bayesian setting, which has been an important framework for developing various estimation techniques. On the basis of the obtained Bayesian forward-backward smoother, we develop a practically implementable algorithm through MAP optimization to carry out SISS.

II. Bayesian Input and State Smoothing

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

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

where $\mathbf{x}_k \in \mathbb{R}^{n_\mathbf{x}}$ is the unknown state vector, $\mathbf{u}_k \in \mathbb{R}^{n_\mathbf{u}}$ is the unknown input vector and $\mathbf{y}_k \in \mathbb{R}^{n_\mathbf{y}}$ is the output vector. The process noise $\mathbf{w}_k$ and the measurement noise $\mathbf{v}_k$ are mutually independent, zero-mean white Gaussian sequences with covariances $\mathbf{Q}_k \geq 0$ and $\mathbf{R}_k > 0$, respectively. For a Gaussian random vector $\mathbf{x} \in \mathbb{R}^n$, we use the notation

$$\mathcal{N}(\mathbf{x}; \boldsymbol{\sigma}, \boldsymbol{\Sigma}) := (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2} \|\mathbf{x} - \boldsymbol{\sigma}\|_{\boldsymbol{\Sigma}}^2\right),$$

where $\mathbf{x}, \boldsymbol{\sigma} \in \mathbb{R}^n$, $|\boldsymbol{\Sigma}|$ denotes the determinant of $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$, and

$$\|\mathbf{x} - \boldsymbol{\sigma}\|_{\boldsymbol{\Sigma}}^2 = (\mathbf{x} - \boldsymbol{\sigma})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\sigma}).$$

Then we have $p(\mathbf{w}_k) = \mathcal{N}(\mathbf{w}_k; \mathbf{0}, \mathbf{Q}_k)$ and $p(\mathbf{v}_k) = \mathcal{N}(\mathbf{v}_k; \mathbf{0}, \mathbf{R}_k)$, where p represents the probability density function (pdf). The nonlinear mappings $\mathbf{f} : \mathbb{R}^{n_\mathbf{x}} \times \mathbb{R}^{n_\mathbf{u}} \to \mathbb{R}^{n_\mathbf{x}}$ and $\mathbf{h} : \mathbb{R}^{n_\mathbf{x}} \times \mathbb{R}^{n_\mathbf{u}} \to \mathbb{R}^{n_\mathbf{y}}$ characterize the process dynamics and the measurement model, respectively.

From the statistical point of view, the states $\{\mathbf{x}_k\}$ and the output measurements $\{\mathbf{y}_k\}$ of the system in (1) form two different stochastic processes, propagating according to the state space equations. Information about the states and inputs is hidden in the measurements and needs to be mined. Define the set $\mathcal{Y}_{1:k} := \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_k\}$, which contains all the measurements up until time k. Thus essentially, real-time filtering, i.e., SISE, is aimed to compute the joint pdf of $\mathbf{u}_k$ and $\mathbf{x}_k$ conditioned on $\mathcal{Y}_{1:k}$, i.e., $p(\mathbf{u}_k, \mathbf{x}_k | \mathcal{Y}_{1:k})$. When real-time data processing is not required, SISS is more desirable, which uses additional measurement data or the data after time k to obtain the joint pdf of $\mathbf{u}_k$ and $\mathbf{x}_k$ with higher accuracy. That is, $p(\mathbf{u}_k, \mathbf{x}_k | \mathcal{Y}_{1:N})$ for $N \geq k$ is of particular interest to us in this situation.

978-1-4799-0176-0/$31.00 ©2013 AACC

In the sequel, we shall denote for notational convenience $\boldsymbol{\xi}_k := \left[\mathbf{u}_k^\top\ \mathbf{x}_k^\top\right]^\top$ that is the augmented vector to be estimated. The Bayesian smoothing paradigm for SISE can be obtained using the Bayes' rule to build the backward recursion of $p(\boldsymbol{\xi}_k|\mathcal{Y}_{1:N})$ from $p\left(\boldsymbol{\xi}_{k+1}|\mathcal{Y}_{1:N}\right)$. 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 assumption A1 is made 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. We have 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

$$p(\boldsymbol{\xi}_k|\mathcal{Y}_{1:N}) = p(\boldsymbol{\xi}_k|\mathcal{Y}_{1:k}) \int \frac{p(\mathbf{x}_{k+1}|\boldsymbol{\xi}_k) p\left(\mathbf{x}_{k+1}|\mathcal{Y}_{1:N}\right)}{p(\mathbf{x}_{k+1}|\mathcal{Y}_{1:k})} \mathrm{d}\mathbf{x}_{k+1}. \tag{2}$$

Proof: It is seen that

$$\begin{aligned} p(\boldsymbol{\xi}_k|\mathcal{Y}_{1:N}) &= \int p(\boldsymbol{\xi}_k, \boldsymbol{\xi}_{k+1}|\mathcal{Y}_{1:N}) \mathrm{d}\mathbf{x}_{k+1} \\ &= \int p\left(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k+1}, \mathcal{Y}_{1:N}\right) p\left(\boldsymbol{\xi}_{k+1}|\mathcal{Y}_{1:N}\right) \mathrm{d}\boldsymbol{\xi}_{k+1}. \end{aligned} \tag{3}$$

By A1 and the Markovian propagation of the state, we have

$$p\left(\boldsymbol{\xi}_k|\mathbf{x}_{k+1}, \mathcal{Y}_{1:N}\right) = p\left(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{k+1}, \mathcal{Y}_{1:k}\right).$$

It follows that

$$p(\boldsymbol{\xi}_k|\mathcal{Y}_{1:N}) = \int p\left(\boldsymbol{\xi}_k|\mathbf{x}_{k+1}, \mathcal{Y}_{1:k}\right) p\left(\mathbf{x}_{k+1}|\mathcal{Y}_{1:N}\right) \mathrm{d}\mathbf{x}_{k+1}. \tag{4}$$

Meanwhile, it is seen by the Bayes' theorem that

$$p\left(\boldsymbol{\xi}_k|\mathbf{x}_{k+1}, \mathcal{Y}_{1:k}\right) = \frac{p(\mathbf{x}_{k+1}|\boldsymbol{\xi}_k) p(\boldsymbol{\xi}_k|\mathcal{Y}_{1:k})}{p(\mathbf{x}_{k+1}|\mathcal{Y}_{1:k})}. \tag{5}$$

Inserting (5) into (4), we obtain (2). ■

Illustrating how to calculate the conditional pdf of $\boldsymbol{\xi}_k$ given $\mathcal{Y}_{1:N}$, the Bayesian paradigm in (2) is an input and state smoother in a statistical sense. However, direct computation of pdf's is known to be both analytically and computationally intractable, if not impossible, for nonlinear systems. Hence, we will seek a numerically feasible solution by introducing certain approximations.

III. Smoothing Algorithms

In this section, we present the development of the SISS algorithm for the nonlinear system in (1).

A. Smoothing for Nonlinear Systems

To proceed further, we make the following Gaussianity approximations:

A2 $p(\boldsymbol{\xi}_k|\mathcal{Y}_{1:k}) = \mathcal{N}\left(\boldsymbol{\xi}_k; \hat{\boldsymbol{\xi}}_{k|k}, \mathbf{P}_{k|k}^{\boldsymbol{\xi}}\right)$;

A3 $p(\mathbf{x}_{k+1}|\boldsymbol{\xi}_k) = \mathcal{N}(\mathbf{x}_{k+1}; \mathbf{f}(\boldsymbol{\xi}_k), \mathbf{Q}_k)$;

A4 $p(\mathbf{x}_{k+1}|\mathcal{Y}_{1:N}) = \mathcal{N}\left(\mathbf{x}_{k+1}; \hat{\mathbf{x}}_{k+1|N}, \mathbf{P}_{k+1|N}^{\mathbf{x}}\right)$.

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

Due to justification the central limit theorem, Gaussianity assumptions analogous to the above are prevalent in statistical estimation, on which a substantial number of nonlinear estimation algorithms have been built [18; 19]. In addition, Gaussian functions are easy to manipulate mathematically, facilitating the ensuing derivation. The assumptions A2-A4, as will be seen, bridge the gap from the Bayesian paradigm in (2) to numerical algorithms.

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

$$\hat{\boldsymbol{\xi}}_{k|N} = \arg\max_{\boldsymbol{\xi}_k} p(\boldsymbol{\xi}_k|\mathcal{Y}_{1:N}). \tag{6}$$

The above maximization of $p(\boldsymbol{\xi}_k|\mathcal{Y}_{1:N})$ can be transformed into the familiar form of minimization of a cost function, as shown in the next theorem.

Theorem 2: For the system in (1), given A1-A4, an equivalence of the smoother in (6) is given by

$$\hat{\boldsymbol{\xi}}_{k|N} = \arg\min_{\boldsymbol{\xi}_k} \ell(\boldsymbol{\xi}_k), \tag{7}$$

where

$$\ell(\boldsymbol{\xi}_k) := \left\|\boldsymbol{\xi}_k - \hat{\boldsymbol{\xi}}_{k|k}\right\|^2_{\mathbf{P}_{k|k}^{\boldsymbol{\xi}}} + \left\|\mathbf{f}(\boldsymbol{\xi}_k) - \boldsymbol{\Delta}_k \boldsymbol{\delta}_k\right\|^2_{\boldsymbol{\Omega}_k}, \tag{8}$$

$$\boldsymbol{\Omega}_k := \boldsymbol{\Delta}_k + \mathbf{Q}_k \tag{9}$$

$$\boldsymbol{\Delta}_k := \left[\left(\mathbf{P}_{k+1|N}^{\mathbf{x}}\right)^{-1} - \left(\mathbf{P}_{k+1|k}^{\mathbf{x}}\right)^{-1}\right]^{-1}, \tag{10}$$

$$\boldsymbol{\delta}_k := \left(\mathbf{P}_{k+1|N}^{\mathbf{x}}\right)^{-1} \hat{\mathbf{x}}_{k+1|N} - \left(\mathbf{P}_{k+1|k}^{\mathbf{x}}\right)^{-1} \hat{\mathbf{x}}_{k+1|k}. \tag{11}$$

Proof: For notational convenience, we use vectors $\mathbf{q}$, $\mathbf{z}$, $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, symmetric positive definite matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, and a mapping $\mathbf{g}$. All have compatible dimensions as considered in the following equation, which is formulated by virtue of (2):

$$L(\mathbf{q}) := \mathcal{N}(\mathbf{q}; \mathbf{a}, \mathbf{A}) \int \frac{\mathcal{N}\left(\mathbf{z}; \mathbf{b}, \mathbf{B}\right) \cdot \mathcal{N}\left(\mathbf{z}; \mathbf{c}, \mathbf{C}\right)}{\mathcal{N}\left(\mathbf{z}; \mathbf{d}, \mathbf{D}\right)} \mathrm{d}\mathbf{z}, \tag{12}$$

where $\mathbf{b} = \mathbf{g}(\mathbf{q})$. Compared with (2), we indeed let $\mathbf{A} \to \mathbf{P}_{k|k}^{\boldsymbol{\xi}}$, $\mathbf{B} \to \mathbf{Q}_k$, $\mathbf{C} \to \mathbf{P}_{k+1|N}^{\mathbf{x}}$, $\mathbf{D} \to \mathbf{P}_{k+1|k}^{\mathbf{x}}$, $\mathbf{q} \to \boldsymbol{\xi}_k$, $\mathbf{a} \to \hat{\boldsymbol{\xi}}_{k|k}$, $\mathbf{b} \to \mathbf{f}(\boldsymbol{\xi}_k)$, $\mathbf{c} \to \hat{\mathbf{x}}_{k+1|N}$, and $\mathbf{d} \to \hat{\mathbf{x}}_{k+1|k}$. Here, $\to$ denotes the corresponding relationship.

As indicated by (A.1) in Appendix, we have

$$\frac{\mathcal{N}\left(\mathbf{z}; \mathbf{b}, \mathbf{B}\right) \cdot \mathcal{N}\left(\mathbf{z}; \mathbf{c}, \mathbf{C}\right)}{\mathcal{N}\left(\mathbf{z}; \mathbf{d}, \mathbf{D}\right)} = \lambda \cdot \mathcal{N}(\mathbf{z}; \mathbf{e}, \mathbf{E}),$$

where

$$\begin{aligned} \mathbf{E} &= \left(\mathbf{B}^{-1} + \mathbf{C}^{-1} - \mathbf{D}^{-1}\right)^{-1}, \\ \mathbf{e} &= \mathbf{E}\left(\mathbf{B}^{-1}\mathbf{b} + \mathbf{C}^{-1}\mathbf{c} - \mathbf{D}^{-1}\mathbf{d}\right), \\ \lambda(\mathbf{q}) &= |\mathbf{B}|^{-\frac{1}{2}} |\mathbf{C}|^{-\frac{1}{2}} |\mathbf{D}|^{\frac{1}{2}} |\mathbf{E}|^{\frac{1}{2}} \end{aligned}$$

$$\cdot \exp\left[-\frac{1}{2}\left(\|\mathbf{b}\|_{\mathbf{B}}^2+\|\mathbf{c}\|_{\mathbf{C}}^2-\|\mathbf{d}\|_{\mathbf{D}}^2-\|\mathbf{e}\|_{\mathbf{E}}^2\right)\right].$$

Noting that the integral in (12) will be equal to λ, we consider only λ now. It is obvious that λ can be decomposed into two parts, one of which is a proportional coefficient, and the other one of which is a function $\mathbf{q}$, i.e.,

$$\lambda(\mathbf{q}) = \lambda_o \bar{\lambda}(\mathbf{q}).$$

Here,

$$\bar{\lambda}(\mathbf{q}) = \exp\Big[-\frac{1}{2}\big\|\mathbf{b}-(\mathbf{B}^{-1}-\mathbf{B}^{-1}\mathbf{E}\mathbf{B}^{-1})^{-1} \cdot\mathbf{B}^{-1}\mathbf{E}(\mathbf{C}^{-1}\mathbf{c}-\mathbf{D}^{-1}\mathbf{d})\big\|^2_{(\mathbf{B}^{-1}-\mathbf{B}^{-1}\mathbf{E}\mathbf{B}^{-1})^{-1}}\Big],$$

which, through using the Matrix Inversion Lemma, is found equal to

$$\bar{\lambda}(\mathbf{q}) = \exp\Big[-\frac{1}{2}\,\mathbf{b}-(\mathbf{C}^{-1}-\mathbf{D}^{-1})^{-1} (\mathbf{C}^{-1}\mathbf{c}-\mathbf{D}^{-1}\mathbf{d})\big\|^2_{(\mathbf{C}^{-1}-\mathbf{D}^{-1})^{-1}+\mathbf{B}}\Big],$$

Incorporating the results above and considering the logarithm of $L(\mathbf{q})$, we obtain the following cost function to minimize:

$$\ell(\mathbf{q}) = \|\mathbf{q}-\mathbf{a}\|^2_{\mathbf{A}} + \|\mathbf{g}(\mathbf{q}) -(\mathbf{C}^{-1}-\mathbf{D}^{-1})^{-1}(\mathbf{C}^{-1}\mathbf{c}-\mathbf{D}^{-1}\mathbf{d})\big\|^2_{(\mathbf{C}^{-1}-\mathbf{D}^{-1})^{-1}+\mathbf{B}}.$$

Referring to the above derivation, we obtain (7)-(11) from (2), (6) and the assumptions A1-A3. ∎

Remark 1: It is noted that (2) computes $p(\boldsymbol{\xi}_k|\mathcal{Y}_{1:N})$ that is the *a posteriori* distribution of $\boldsymbol{\xi}_k$ given $\mathcal{Y}_{1:N}$. Thus the estimator designed in (6) or (7) is exactly based on the MAP estimation. Therefore, $\ell(\boldsymbol{\xi}_k)$ in (8) is a log-MAP cost function. MAP estimation formulated within Bayesian framework has often been used to develop optimal filtering methods, e.g., [20]. Here, we extend this technique to address the smoothing problem and especially for the purpose of joint input and state estimation.

It is often extremely difficult to obtain an analytical solution to the optimization problem in (7) due to the nonlinearities, motivating the development of an approximate numerical solution.

First note that the sum of the weighted 2-norms in (8) can be rewritten as

$$\ell(\boldsymbol{\xi}_k) = \|\mathbf{s}_k(\boldsymbol{\xi}_k)\|^2, \tag{13}$$

where

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

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.

Theorem 3: For the input and state smoother designed in (7), the Gauss-Newton based solution $\hat{\boldsymbol{\xi}}_{k|N} = \hat{\boldsymbol{\xi}}^{(i_{\max})}_{k|N}$ is iteratively computed by

$$\hat{\boldsymbol{\xi}}^{(i+1)}_{k|N} = \hat{\boldsymbol{\xi}}^{(i)}_{k|N} - \left[\nabla^{\mathrm{T}}_{\boldsymbol{\xi}}\mathbf{s}\left(\hat{\boldsymbol{\xi}}^{(i)}_{k|N}\right)\nabla_{\boldsymbol{\xi}}\mathbf{s}\left(\hat{\boldsymbol{\xi}}^{(i)}_{k|N}\right)\right]^{-1} \cdot\nabla^{\mathrm{T}}_{\boldsymbol{\xi}}\mathbf{s}\left(\hat{\boldsymbol{\xi}}^{(i)}_{k|N}\right)\mathbf{s}\left(\hat{\boldsymbol{\xi}}^{(i)}_{k|N}\right), \tag{14}$$

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

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

The first-order approximation of the associated smoothed error covariance is

$$\mathbf{P}^{\boldsymbol{\xi}}_{k|N} = \left[\left(\mathbf{P}^{\boldsymbol{\xi}}_{k|k}\right)^{-1} + \nabla^{\top}_{\boldsymbol{\xi}}\mathbf{f}(\hat{\boldsymbol{\xi}}_{k|N})\boldsymbol{\Omega}_k^{-1}\nabla_{\boldsymbol{\xi}}\mathbf{f}(\hat{\boldsymbol{\xi}}_{k|N})\right]^{-1}. \tag{15}$$

Proof: Applying the classical Gauss-Newton method to (13), (14) follows directly [21]. It is noted that the assumptions A1-A3 ensure $p(\boldsymbol{\xi}_k|\mathcal{Y}_{1:N})$ to be Gaussian. Thus according to the MAP estimation theory, the error covariance is covariance can be estimated as the inverse of the Fisher information matrix $\boldsymbol{\mathcal{F}}$ evaluated at the estimate:

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

where $\boldsymbol{\mathcal{F}}$ is the given by

$$\boldsymbol{\mathcal{F}}(\boldsymbol{\xi}_k) = \left(\mathbf{P}^{\boldsymbol{\xi}}_{k|k}\right)^{-1} + \nabla^{\top}_{\boldsymbol{\xi}}\mathbf{f}(\boldsymbol{\xi}_k)\boldsymbol{\Omega}_k^{-1}\nabla_{\boldsymbol{\xi}}\mathbf{f}(\boldsymbol{\xi}_k).$$

This completes the proof. ∎

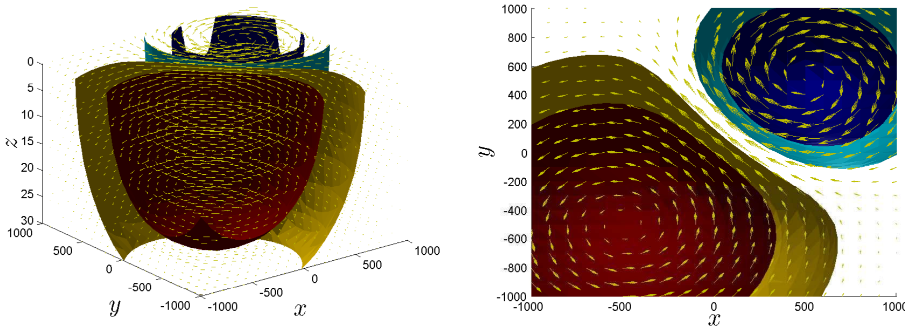
Here, we say that (14)-(15) are the *backward smoothing equations* for input and state estimation. The corresponding forward filtering equations are presented in [4]. Derivations of the proposed smoother and the filter in [4] follow similar lines: a Bayesian framework is constructed first, and then a MAP estimation problem is formulated within the framework and solved.

Remark 2: The Gauss-Newton method is used for iterative search over a designed cost function to find the best estimates. This principle also applies to some nonlinear state estimation algorithms, e.g., the iterated extended Kalman filter (IEKF) [22].

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 worth mentioning is the trust-region method, which overcomes the singularity problems that may arise in the computing process. The reader is referred to [21] for more details.

IV. Application Example

A fundamental problem in oceanography is flow field reconstruction. Flows are known to be crucial to fishing, shipping and weather forecasting. To study the flows, a swarm of buoyancy-controlled drogues [23; 24] are deployed



(a) (b)

Fig. 1: (a) The three-dimensional flow field; (b) top view of the flow field.

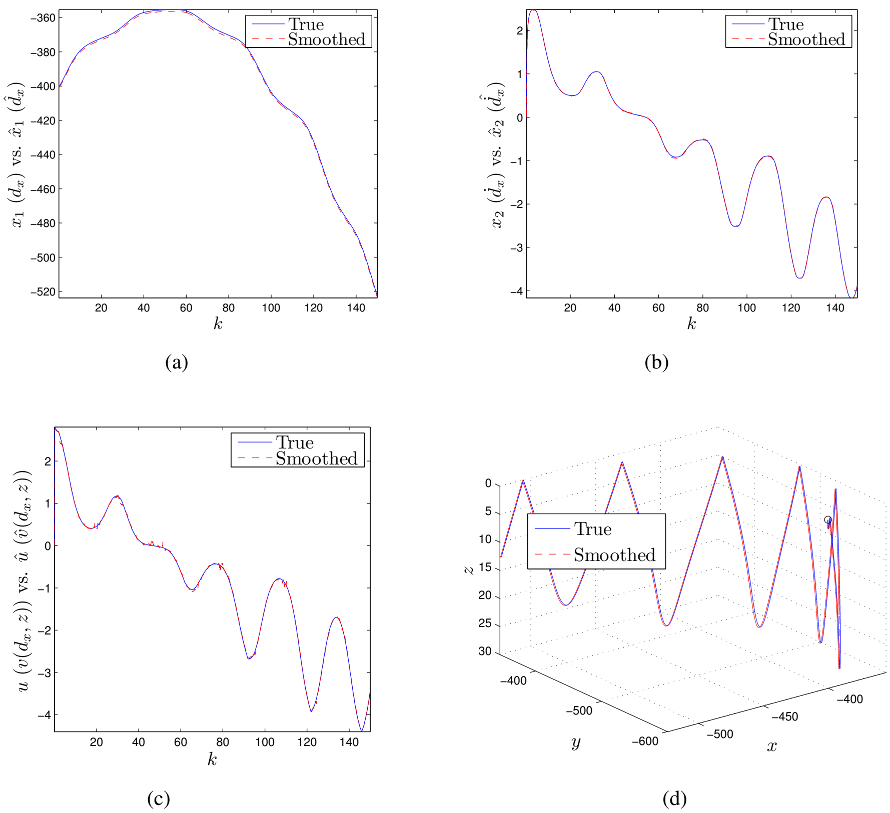


Fig. 2: Smoothing results for the drogues released at (-400,-600)m: (a) x_1 — displacement along x-direction; (b) x_2 — velocity along x-direction; (c) u — flow velocity; (d) trajectory of the drogue (the circle denotes the location where the drogues is released).

in the ocean. They are capable of arbitrary vertical migration behaviors while traveling along the flows. During the travel, each drogue measures and stores a time record of its depth, acceleration, position and other relevant oceanographic quantities such as temperature and salinity. Because the GPS signals are seriously attenuated underwater, the position information is measurable only intermittently when the drogue is at water surface, although other measurements are available continuously. The record of data will be transmitted to a central server for analysis and processing when the drogue is at surface.

Here, a three-dimensional flow domain is considered (see Fig. 1), in which two eddies are present.

A. Drogue Dynamics

Due to the independence of perpendicular components of motion, we consider the motion along x-direction without loss of generality. For a drogue, the flow velocity $v(d_x, z)$ at its x-displacement d_x is time-stationary and dependent only on the drogue's depth z. The dynamics of the drogue is described in [25]:

$$m\ddot{d}_x = c \cdot \mathrm{sign}\left(v(d_x, z) - \dot{d}_x\right) \cdot \left(v(d_x, z) - \dot{d}_x\right)^2, \quad (16)$$

where m is the constant rigid mass and c is the drag parameter that quantifies the drag or resistance exercised on

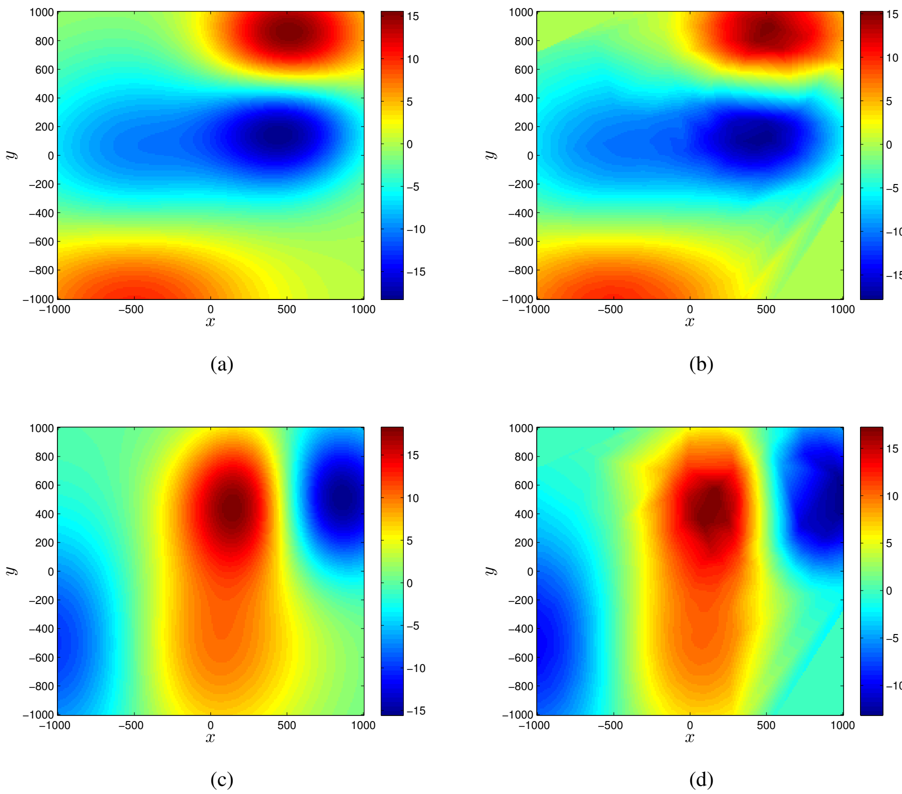


Fig. 3: (a) True surficial flow velocity along x-direction; (b) smoothed surficial flow velocity along x-direction; (c) True surficial flow velocity along y-direction; (b) smoothed surficial flow velocity along y-direction;

the drogue in the flow field.

From (16), 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 drogue dynamics, naturally implying the definition of $u := v(d_x, z)$. Then (16) can be rewritten as,

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= \frac{c}{m} \cdot \operatorname{sign}(u - x_2) \cdot (u - x_2)^2 . \end{aligned} \tag{17}$$

Its discrete-time representation via finite difference is

$$\begin{aligned} x_{1,k+1} &= x_{1,k} + T \cdot x_{2,k}, \\ x_{2,k+1} &= x_{2,k} + T \cdot \frac{c}{m} \cdot \operatorname{sign}(u_k - x_{2,k}) \cdot (u_k - x_{2,k})^2 , \end{aligned} \tag{18}$$

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

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, u_k), \tag{19}$$

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

The motion of the drogue is characterized by an irregularly submerging/surfacing pattern — it submerges and moves underwater for a certain duration, then resurfaces, and repeats the process again. 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,

$$\begin{aligned} \tau_k &= \frac{c}{m} \cdot \operatorname{sign}(u_k - x_{2,k}) \cdot (u_k - x_{2,k})^2 , \\ \eta_k &= x_{1,k}. \end{aligned} \tag{20}$$

For simplicity of notation, we rewrite (20) as

$$\begin{aligned} \tau_k &= \varphi(u_k, x_k), \\ \eta_k &= \phi(x_k). \end{aligned} \tag{21}$$

Combining (18) and (20), we obtain the state space model to describe the dynamics of the drogue:

$$\Sigma : \begin{cases} \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, u_k) + \mathbf{w}_k, \\ \mathbf{y}_k = \mathbf{h}(\mathbf{x}_k, u_k) + \mathbf{v}_k, \end{cases} \tag{22}$$

Here, when the drogue is underwater, $\mathbf{y}_k = \tau_k$ and $\mathbf{h} = \varphi$, when at surface, $\mathbf{y}_k = [\tau_k \ \eta_k]^\top$ and $\mathbf{h} = [\varphi \ \phi]^\top$. In addition, $\mathbf{w}$ and $\mathbf{v}$ are added to account for noises. They are assumed to be white Gaussian and independent of each other. The algorithm proposed in Section III is applicable to the system Σ in (22) to acquire the information estimates of not only the velocities of the flow field (unknown input variable) but also the trajectory and velocity profile of the drogue (state variables).

B. Numerical Simulation

The flow field considered is shown in Fig. 1, which has dimensions of $(-1000,1000)$m $\times$ $(-1000,1000)$m $\times$

(0,30)m, and the eddies are centered at (500,500)m and (-500,-500)m, respectively. It is intentionally narrowed in scale to reduce computation burden, but this does not restrict application of the proposed SISS algorithm to larger flow fields. Let 20 drogues be deployed evenly along the line segment from (-800,-1000)m to (1000,800)m. The mass of a drogue is 1.5Kg, the drag coefficient c is $2\mathrm{Ns}^2/\mathrm{m}^2$, and the sampling period T is 0.05s.

The proposed SISS algorithm is applied to smoothing the estimates of the inputs and states in the state-space model of the drogue. Let us examine the drogue released at (-400,-600)m and consider its motion in the x-direction. Fig. 2(a)-2(c) compares the true values and the smoothed estimates of the x-displacement, x-velocity and the flow velocity, respectively. Fig. 2(d) shows the smoothed trajectory in comparison to the true one. It is seen that the smoothing algorithm exhibits high accuracy, with the smoothed estimates agreeing well with the truth.

Further, the estimated inputs of all drogues, which are the smoothed flow velocity data at different locations, are collected together and used to reconstruct the flow field via the tessellation-based linear interpolation. The reconstructed surficial flow velocity fields along x-direction and y-direction are compared in Fig. 3 with the true fields, respectively. The accuracy of reconstruction is noted to be quite satisfactory.

V. Conclusions

This paper studies joint input and state estimation via forward-backward smoothing for nonlinear systems. This challenging problem is treated from a statistical perspective, with a Bayesian framework constructed in the first place. A MAP based nonlinear smoothing algorithm is then developed within the framework to obtain smoothed input and state estimates. The algorithm has a structure of backward recursion, while each recursion is realized by Gauss-Newton iterations. The soundness of the proposed algorithm is verified through simulation studies of oceanographic flow field reconstruction where flow velocity profiles are estimated from the motion information of a group of buoyancy-controlled drogues.

Acknowledgement

The authors would like to thank Professor Peter J.S. Franks from the Scripps Institution of Oceanography, University of California, San Diego, for the three-dimensional flow field model and constructive suggestions.

Appendix

- **Multiplication of Gaussian functions** The production of two Gaussian functions is a Gaussian function, i.e.,

$$\mathcal{N}(\mathbf{x};\mathbf{a},\mathbf{A})\cdot\mathcal{N}(\mathbf{x};\mathbf{b},\mathbf{B})=\lambda\cdot\mathcal{N}(\mathbf{x};\mathbf{c},\mathbf{C}), \tag{A.1}$$

where $\mathbf{C}=\left(\mathbf{A}^{-1}+\mathbf{B}^{-1}\right)^{-1}$, $\mathbf{c}=\mathbf{C}(\mathbf{A}^{-1}\mathbf{a}+\mathbf{B}^{-1}\mathbf{b})$, and

$$\lambda=(2\pi)^{-\frac{n}{2}}|\mathbf{A}|^{-\frac{1}{2}}|\mathbf{B}|^{-\frac{1}{2}}|\mathbf{C}|^{\frac{1}{2}} \cdot\exp\left[-\frac{1}{2}\left(\|\mathbf{a}\|_{\mathbf{A}}^2+\|\mathbf{b}\|_{\mathbf{B}}^2-\|\mathbf{c}\|_{\mathbf{C}}^2\right)\right].$$

References

[1] U. Schubert, U. Kruger, G. Wozny, and H. Arellano-Garcia, "Input reconstruction for statistical-based fault detection and isolation," *AIChE Journal*, vol. 58, no. 5, pp. 1513–1523, 2012.

[2] H. Imine, Y. Delanne, and N. K. M'Sirdi, "Road profile input estimation in vehicle dynamics simulation," *Vehicle System Dynamics*, vol. 44, no. 4, pp. 285–303, 2006.

[3] P. K. Kitanidis, "Unbiased minimum-variance linear state estimation," *Automatica*, vol. 23, no. 6, pp. 775–778, 1987.

[4] H. Fang and R. A. de Callafon, "Nonlinear simultaneous input and state estimation with application to flow field estimation," in *IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, 2011, pp. 6013–6018.

[5] J. Mendel, "White-noise estimators for seismic data processing in oil exploration," *Automatic Control, IEEE Transactions on*, vol. 22, no. 5, pp. 694–706, 1977.

[6] C.-S. Hsieh, "Robust two-stage kalman filters for systems with unknown inputs," *IEEE Transactions on Automatic Control*, vol. 45, no. 12, pp. 2374–2378, 2000.

[7] T. Floquet, C. Edwards, and S. K. Spurgeon, "On sliding mode observers for systems with unknown inputs," *International Journal of Adaptive Control and Signal Processing*, vol. 21, no. 8-9, 2007.

[8] "Sliding-mode observers for systems with unknown inputs: A high-gain approach," *Automatica*, vol. 46, no. 2, pp. 347–353, 2010.

[9] H. Lee and M.-J. Tahk, "Generalized input-estimation technique for tracking maneuvering targets," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 4, pp. 1388–1402, 1999.

[10] S. Gillijns and B. D. Moor, "Unbiased minimum-variance input and state estimation for linear discrete-time systems," *Automatica*, vol. 43, no. 1, pp. 111–116, 2007.

[11] ——, "Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough," *Automatica*, vol. 43, no. 5, pp. 934–937, 2007.

[12] H. Fang, Y. Shi, and J. Yi, "On stable simultaneous input and state estimation for discrete-time linear systems," *International Journal of Adaptive Control and Signal Processing*, vol. 25, no. 8, pp. 671–686, 2011.

[13] M. Corless and J. Tu, "State and input estimation for a class of uncertain systems," *Automatica*, vol. 34, no. 6, pp. 757–764, 1998.

[14] Q. Ha and H. Trinh, "State and input simultaneous estimation for a class of nonlinear systems," *Automatica*, vol. 40, no. 10, pp. 1779–1785, 2004.

[15] M. Darouach and M. Zasadzinski, "Unbiased minimum variance estimation for systems with unknown exogenous inputs," *Automatica*, vol. 33, no. 4, pp. 717 – 719, 1997.

[16] M. Darouach, M. Zasadzinski, and M. Boutayeb, "Extension of minimum variance estimation for systems with unknown inputs," *Automatica*, vol. 39, no. 5, pp. 867 – 876, 2003.

[17] F. Bejarano, L. Fridman, and A. Poznyak, "Exact state estimation for linear systems with unknown inputs based on hierarchical super-twisting algorithm," *International Journal of Robust and Nonlinear Control*, vol. 17, no. 18, pp. 1734–1753, 2007.

[18] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, 1979.

[19] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, 2000.

[20] J. V. Candy, *Bayesian Signal Processing: Classical, Modern and Particle Filtering Methods*. New York, NY, USA: Wiley-Interscience, 2009.

[21] Å. Björck, *Numerical Methods for Least Squares Problems*. Philadelphia: SIAM, 1996.

[22] B. Bell and F. Cathey, "The iterated Kalman filter update as a Gauss-Newton method," *IEEE Transactions on Automatic Control*, vol. 38, no. 2, pp. 294–297, 1993.

[23] C. Colgan, "Underwater laser shows," *Explorations, Scripps Institution of Oceanography*, vol. 12, no. 4, pp. 20–27, 2006.

[24] Y. Han, R. A. de Callafon, J. Cortés, and J. Jaffe, "Dynamic modeling and pneumatic switching control of a submersible drogue," in *International Conference on Informatics in Control, Automation and Robotics*, vol. 2, Funchal, Madeira, Portugal, Jun. 2010, pp. 89–97.

[25] D. A. Booth, "On the use of drogues for measuring subsurface ocean currents," *Ocean Dynamics*, vol. 34, pp. 284–294, 1981.