A switching strategy for adaptive state estimation

Aditya Tulsyan[a], Swanand R. Khare[b], Biao Huang[c], R. Bhushan Gopaluni[d], J.F. Forbes[c]

[a]*Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.*
[b]*Department of Mathematics, Indian Institute of Technology, Kharagpur, WB 721302, India.*
[c]*Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB T6G 2G6, Canada.*
[d]*Department of Chemical and Biological Engineering, University of British Columbia, Vancouver, BC V6T 1Z3, Canada.*

Abstract

This paper develops a switching strategy for adaptive state estimation in systems represented by nonlinear, stochastic, discrete-time state space models (SSMs). The developed strategy is motivated by the fact that there is no single Bayesian estimator that is guaranteed to perform optimally for a given nonlinear system and under all operating conditions. The proposed strategy constructs a bank of plausible Bayesian estimators for adaptive state estimation, and then switches between them based on their performance. The performance of a Bayesian estimator is assessed using a performance metric derived from the posterior Cramér-Rao lower bound (PCRLB). It is shown that the switching strategy is stable, and yields estimates that are at least as good as any individual estimator in the bank. The efficacy of the switching strategy is illustrated on a practical simulation example.

Keywords: nonlinear systems; state estimation; parameter estimation; and switching strategy.

1. Introduction

Recent advances in high-speed computing technology have lead to the frequent use of stochastic nonlinear models to represent complex system dynamics. The design and implementation of advanced control or monitoring strategies using such complex models require real-time estimation of the key system states and parameters that are either unmeasured or unknown. In situations, where the parameters are precisely known, the states can be estimated under the Bayesian framework by computing the state posterior density. The state posterior density is often computed by solving a state filtering problem (Arulampalam et al., 2002). A closed-form solution to the filtering problem exists for linear state space models (SSMs) under the Gaussian noise settings or when the state space is finite (Tulsyan et al., 2013c). Unfortunately, in many engineering systems, the model is often nonlinear and the parameters are not known or time-varying, and therefore need to be estimated before the states can be estimated. In practical settings, adaptive state estimation (simultaneous state and parameter estimation) is often the only realistic solution for it avoids processing of large data set and also allows for real-time adaptation to the time-varying system behavior. (He et al., 2011; Kravaris et al., 2013; Ding, 2014).

We consider the problem of online adaptive Bayesian state estimation in general nonlinear stochastic SSMs. In general, this is a difficult problem even for a linear system, as the joint state and parameter posterior density does not lend itself to any closed-form solution (Li et al., 2004). There are several important classes of Bayesian methods for adaptive state estimation, which include - artificial dynamics approach (ADA), practical filtering, Markov chain Monte-Carlo (MCMC) with sequential Monte-Carlo (SMC) methods and SMC^2. A detailed exposition of Bayesian estimators and its approximations can be found in Kantas et al. (2009). Although tractable, the performance of these Bayesian estimators depends on the underlying numerical and statistical approximations used in their design. Unfortunately, there is no single tractable online Bayesian estimator that is guaranteed to perform consistently on a given system or retain a satisfactory performance under all operating conditions. A practitioner is thus left with no clear substitute for the optimal Bayesian estimator.

This paper develops an efficient strategy for adaptive Bayesian state estimation in general nonlinear SSMs. At the outset, it is highlighted that the paper only deals with the class of Bayesian estimators and not with maximum-likelihood estimators. Further, we restrict ourselves only to the class of online Bayesian estimators. The preliminaries of adaptive state estimation is provided next.

2. Preliminaries

In this section, we: (i) define a discrete-time stochastic SSM; (ii) introduce the adaptive state estimation problem; and (iii) highlight our contributions.

Notation: The common notation are first introduced. Here, the notation is broadly classified by topic.

☆Corresponding author A. Tulsyan. Tel. +1 617-314-1608.
Email addresses: tulsyan@mit.edu (Aditya Tulsyan), srkhare@maths.iitkgp.ernet.in (Swanand R. Khare), biao.huang@ualberta.ca (Biao Huang), bhushan.gopaluni@ubc.ca (R. Bhushan Gopaluni), fraser.forbes@ualberta.ca (J.F. Forbes)

- *Spaces:* $\mathbb{N} := \{1, 2, \dots\}$; $\mathbb{R}_+ := [0, \infty)$; $\mathbb{R}^{s\times s}$ is the set of real-valued $(s \times s)$ matrices; $\mathcal{S}^{s} \subset \mathbb{R}^{s\times s}$ is the space of symmetric matrices; $I_{s\times s}$ is the identity matrix of size s; $\mathcal{S}^{s}_{+}$ is the cone of symmetric positive semi-definite matrices in $\mathcal{S}^{s}$; and $\mathcal{S}^{s}_{++}$ is its interior, i.e., the set of positive definite matrices. The partial orders on $\mathcal{S}^{s}$ induced by $\mathcal{S}^{s}_{+}$ and $\mathcal{S}^{s}_{++}$ are denoted by $\succcurlyeq$ and $\succ$, respectively.
- *Matrices:* For $A, B \in \mathbb{R}^{s\times s}$, the Hadamard product of A and B is denoted by $A \circ B$, and defined as $A \circ B(i, j) = A(i, j)B(i, j)$, where $i, j = 1, \dots, s$. The Hadamard inverse is denoted by $A^{\circ -1}$, is defined as $A^{\circ -1}(i, j) = 1/A(i, j)$ if and only if $A(i, j) \neq 0$ for all $i, j = 1, \dots, s$. $\mathrm{Tr}[A]$ represents trace of A. For $v \in \mathbb{R}^{s}$, $v(i)$ indicates the i^{th} entry in v, and $\mathrm{diag}(v) \in \mathbb{R}^{s\times s}$ is a diagonal matrix, with elements in $v \in \mathbb{R}^{s}$ as its diagonal entries.
- *Probability:* $\Pr(\cdot)$ and $p(\cdot)$ denote the probability mass and density function of a discrete and continuous random variable, respectively. $\operatorname{supp} p(\cdot)$ denotes support of the density function $p(\cdot)$. $\delta_y(x)$ denotes the Dirac-delta function in x with mass centered at y.
- *Sequences and operators:* For any generic sequence denoted by $\{u_t\}_{t\in\mathbb{N}}$, let $u_{i:j} \equiv \{u_i, u_{i+1}, \dots, u_j\}$. Also, $\nabla_x \equiv \left[\frac{\partial}{\partial x}\right]$ is the gradient and $\Delta^{y}_{x} \equiv \nabla_x \nabla^{\mathrm{T}}_{y}$ is the Laplacian operator. $\otimes$ is the Kronecker product.

2.1. Stochastic State Space Models

Let $\{X_t\}_{t\in\mathbb{N}}$ and $\{Y_t\}_{t\in\mathbb{N}}$ be $\mathcal{X}(\subseteq \mathbb{R}^{n})$ and $\mathcal{Y}(\subseteq \mathbb{R}^{m})$ – valued stochastic processes defined on a probability space $(\Omega, \mathcal{F}, P)$. The discrete-time state process $\{X_t\}_{t\in\mathbb{N}}$ is an unobserved Markov process, with initial density $p(x|\theta)$ and Markovian transition density $p(x'|x, \theta)$, such that

$$X_0 \sim p(\cdot|\theta_0) \text{ and } X_{t+1}|(X_t = x_t, \theta_t) \sim p(\cdot|x_t, \theta_t), \tag{1}$$

for all $t \in \mathbb{N}$. In (1), $\{\theta_t\}_{t\in\mathbb{N}} \in \Theta \subseteq \mathbb{R}^{r}$ are model parameters. The state process $\{X_t\}_{t\in\mathbb{N}}$ is hidden, but observed through a sensor process $\{Y_t\}_{t\in\mathbb{N}}$. It is assumed that $\{Y_t\}_{t\in\mathbb{N}}$ is conditionally independent given a state process $\{X_t\}_{t\in\mathbb{N}}$ with marginal density $p(y|x, \theta)$, such that

$$Y_t|(X_0, \dots, X_t = x_t, \dots, X_N, \theta_t) \sim p(\cdot|x_t, \theta_t), \tag{2}$$

for all $t \in \mathbb{N}$. All the density functions are with respect to a suitable dominating measure, such as Lebesgue measure. Although (1) and (2) represent a wide-class of time-series models, the model form considered here is given below:

Model 2.1. Non-linear SSM with non-Gaussian noise

$$X_{t+1} = f_t(X_t, \theta_t, V_t); \tag{3a}$$
$$\theta_{t+1} = h_t(X_t, \theta_t, X_{t+1}); \tag{3b}$$
$$Y_t = g_t(X_t, \theta_t, W_t). \tag{3c}$$

Model 2.1 represents a class of time-varying, nonlinear stochastic SSMs that is ubiquitous in many engineering systems. For example, in process industries, the dynamics of many of the unit operations, such as distillation columns, chemical reactors and separation units can be represented by Model 2.1. For notational convenience, explicit dependence of Model 2.1 on the input signal has been suppressed; however, all the derivations that appear in this paper hold with the input signal included. The assumptions on Model 2.1 are discussed next.

Assumption 2.2. The dynamics of $\{\theta_t\}_{t\in\mathbb{N}}$ is given by

$$h_t(X_t, \theta_t, X_{t+1}) \equiv G^{(1)}_t X_t + G^{(2)}_t \theta_t + G^{(3)}_t X_{t+1},$$

where $G^{(1)}_t \in \mathbb{R}^{r\times n}$, $G^{(2)}_t \in \mathbb{R}^{r\times r}$ and $G^{(3)}_t \in \mathbb{R}^{r\times n}$, such that $G^{(2)}_t$ is invertible for all $t \in \mathbb{N}$.

Assumption 2.3. $\{Y_t\}_{t\in\mathbb{N}}$ is state-oriented with the probability of false alarm $\Pr_{\mathrm{f}} = 0$ and detection $\Pr_{\mathrm{d}} = 1$.

Assumption 2.4. $\{V_t\}_{t\in\mathbb{N}}$ and $\{W_t\}_{t\in\mathbb{N}}$ are mutually independent sequences of independent random variables, assumed to be known a priori in their distribution classes and parametrized by a known and finite number of moments.

Assumption 2.5. $\{f_t; g_t\}_{t\in\mathbb{N}}$ is a pair of nonlinear functions such that for $x_t \in \mathcal{X}$ and $\theta_t \in \Theta$, the pair $\{f_t; g_t\}_{t\in\mathbb{N}}$ is $\mathcal{C}^k(\mathcal{X})$ and $\mathcal{C}^k(\Theta)$, respectively, where $k \geq 2$.

Assumption 2.6. For $v_t \in \mathbb{R}^{n}$; $\{f_t\}_{t\in\mathbb{N}}$ is $\mathcal{C}^k(\mathbb{R}^{n})$, and for $w_t \in \mathbb{R}^{m}$; $\{g_t\}_{t\in\mathbb{N}}$ is $\mathcal{C}^k(\mathbb{R}^{m})$, where $k \geq 1$. For random realizations $(x_{t+1}, x_t, \theta_t, v_t) \in \mathcal{X} \times \mathcal{X} \times \Theta \times \mathbb{R}^{n}$ and $(y_t, x_t, \theta_t, w_t) \in \mathcal{Y} \times \mathcal{X} \times \Theta \times \mathbb{R}^{m}$ satisfying (3a) and (3c), respectively; $\nabla_{v_t} f^{\mathrm{T}}_t(x_t, \theta_t, v_t)$ and $\nabla_{w_t} g^{\mathrm{T}}_t(x_t, \theta_t, w_t)$ have rank n and m, respectively, such that using the implicit function theorem, $p(x_{t+1}|x_t, \theta_t) = p(V_t = \tilde{f}_t(x_t, \theta_t, x_{t+1}))$ and $p(y_t|x_t, \theta_t) = p(W_t = \tilde{g}_t(x_t, \theta_t, y_t))$ are defined, where $\tilde{f}_t := \mathcal{X} \times \Theta \times \mathcal{X} \to \mathbb{R}^{n}$ and $\tilde{g}_t := \mathcal{X} \times \Theta \times \mathcal{Y} \to \mathbb{R}^{m}$.

We only consider a linear time-varying parameter case (see Assumption 2.2), and assume that there are no false or missed observations (see Assumption 2.3). Assumption 2.4 is a standard assumption required by many of the adaptive state estimation methods, and Assumptions 2.5 and 2.6 are regulatory conditions required in the computation of the PCRLB. Another model used in this paper is a special case of Model 2.1, and is represented as follows:

Model 2.7. Non-linear SSM with additive noise

$$X_{t+1} = f_t(X_t, \theta_t) + V_t; \tag{4a}$$
$$\theta_{t+1} = G^{(1)}_t X_t + G^{(2)}_t \theta_t + G^{(3)}_t X_{t+1}; \tag{4b}$$
$$Y_t = g_t(X_t, \theta_t) + W_t, \tag{4c}$$

where $V_t \sim \mathcal{N}(v_t|0, Q_t)$ and $W_t \sim \mathcal{N}(w_t|0, R_t)$ are Gaussian random variables with zero mean and finite variance.

2.2. Adaptive State Estimation

In this section, we discuss the problem of adaptive state estimation, which is basically defined as state estimation under unknown model parameters. In Bayesian settings,

the solution to the adaptive state estimation problem is given by solving the nonlinear filtering problem. The nonlinear filtering problem for Model 2.1 is an active area of research, and is briefly introduced here. Let $\theta_t^\star \in \Theta$ be the true, but unknown model parameters generating Y_t, then under the Bayesian settings, the adaptive state estimation problem is formulated by first ascribing an initial prior density $\theta_0 \sim p(\theta_0)$, such that $\theta_0^\star \in \operatorname{supp} p(\theta_0)$, and then defining $Z_t \equiv [X_t, \theta_t]^{\mathrm{T}}$ as a $\mathcal{Z}(\subseteq \mathbb{R}^{\mathrm{s=n+r}})$ valued extended Markov process. The Bayesian inference on $\{Z_t\}_{t\in\mathbb{N}}$ is then performed by constructing the joint posterior or filtering density $p(z_t|Y_{1:t})$, which encapsulates the available statistical information on Z_t given a measurement sequence $Y_{1:t}$. Once $\{p(z_t|Y_{1:t})\}_{t\in\mathbb{N}}$ is available, a point estimate for the states and parameters is computed. The recursive computation of $\{p(z_t|Y_{1:t})\}_{t\in\mathbb{N}}$ is referred to as the nonlinear filtering problem, and it provides a solution to the adaptive state estimation problem for Model 2.1. In general, the adaptive state estimation problem has proved to be a nontrivial problem. This is because no analytical solution to $\{p(z_t|Y_{1:t})\}_{t\in\mathbb{N}}$ is available even for linear SSMs, with Gaussian noise, or when $\mathcal{X}$ is a finite set (Andrieu et al. (2004)). Thus, with finite computing capabilities, an optimal Bayesian estimator that solves for the adaptive state estimation problem exactly is not realizable for Model 2.1. A recent surge of interest in developing advanced numerical and simulation methods to solve the adaptive state estimation problem has left researchers and practitioners inundated with a large number of approximate, but tractable Bayesian estimators to select from. The main problem addressed in this paper is stated next.

Problem Statement. Given a prior density $(Z_0 = z_0) \sim p(z_0)$, devise a strategy for adaptive state estimation in Model 2.1 operating under Assumptions 2.2 through 2.6.

Remark 2.8. The prior density $(Z_0 = z_0) \sim p(z_0)$ is assumed to be known. Although the choice of a prior density is crucial in the Bayesian inference theory, its design is not considered here. Some guidelines to design an efficient prior density can be found in Tulsyan et al. (2012).

The choice of an efficient Bayesian estimator for adaptive state estimation in Model 2.1 is still an open problem. This is because there is no single tractable Bayesian estimator that is guaranteed to provide a consistent performance for a given system under all operating conditions (Minvielle et al., 2010). A practitioner is thus left with no clear substitute for the optimal Bayesian estimator. An approach to resolve this dilemma is to start with a family of tractable Bayesian estimators and switch between them as and when required, so as to maintain a high overall estimation performance. Naturally, the switching need to depend on the performance of the estimator that accounts for the uncertainty in the model parameters, and other process specific conditions. Despite the strong practical interest in evaluating the performance of Bayesian estimators for adaptive state estimation, it remains one of the most complex problems in Bayesian inference theory (Šimandl et al., 2001).

2.3. Contributions

We propose a performance metric-based switching strategy for adaptive state estimation in Model 2.1. The initial results reported by the authors in Tulsyan et al. (2013d) use a performance metric-based filter switching strategy for state estimation under known model parameters case. The focus of this paper is to generalize the results in Tulsyan et al. (2013d) for adaptive state estimation problems. New results on the stability of the switching strategy are also provided in this paper. The proposed switching strategy uses posterior Cramér-Rao lower bound (PCRLB) inequality as a performance metric, which is discussed next.

3. Posterior Cramér-Rao Lower Bound

The conventional Cramér-Rao lower bound (CRLB) provides a theoretical lower bound on the mean square error (MSE) of any maximum-likelihood (ML)-based unbiased state or parameter estimator (Ljung, 1999). An analogous extension of the CRLB to Bayesian methods was derived by Trees (1968), and is commonly referred to as the PCRLB inequality. The extension of the PCRLB inequality for a class of nonlinear SSMs was derived by Tichavský et al. (1998), and is given in the next two lemmas.

Lemma 3.1. Let $\{Y_{1:t}\}_{t\in\mathbb{N}}$ be a randomly sampled measurement sequence generated from Model 2.1 under Assumption 2.3, then the MSE for a Bayesian estimator that solves for $\{p(z_t|Y_{1:t})\}_{t\in\mathbb{N}}$ is bounded from below by

$$P_{t|t} \equiv \mathbb{E}_{p(Z_t,Y_{1:t})}[(Z_t - \widehat{Z}_{t|t})(Z_t - \widehat{Z}_{t|t})^{\mathrm{T}}] \succcurlyeq J_t^{-1}, \tag{5}$$

where: $P_{t|t} \in \mathcal{S}_{++}^{\mathrm{s}}$ is the MSE; $\widehat{Z}_{t|t} \equiv \widehat{Z}_t(Y_{1:t}) := \mathbb{R}^{\mathrm{tm}} \to \mathbb{R}^{\mathrm{s}}$ are the point estimates for the states and parameters; and $J_t \in \mathcal{S}_{++}^{\mathrm{s}}$ and $J_t^{-1} \in \mathcal{S}_{++}^{\mathrm{s}}$ are the posterior Fisher information matrix (PFIM) and the posterior Cramér Rao lower bound (PCRLB), respectively.

Proof. See Trees (1968) and Tichavský et al. (1998) for a detailed proof of the inequality. □

Lemma 3.1 states that the MSE performance of a Bayesian estimator is always greater than or equal to the PCRLB, such that $P_{t|t} - J_t^{-1} \in \mathcal{S}_{+}^{\mathrm{s}}$ holds for all $t \in \mathbb{N}$. A recursive approach to compute the PFIM for Model 2.1 was derived by Tichavský et al. (1998), and is given next.

Lemma 3.2. For a given prior density $(Z_0 = z_0) \sim p(z_0)$, a recursive method to compute the PFIM for Model 2.1, operating under Assumptions 2.2 through 2.6, is given by

$$J_{t+1}^{11} = H_t^{33} - (H_t^{13})^{\mathrm{T}}[J_t^{11} + H_t^{11}]^{-1}H_t^{13}; \tag{6a}$$

$$J_{t+1}^{12} = (H_t^{23})^{\mathrm{T}} - (H_t^{13})^{\mathrm{T}}[J_t^{11} + H_t^{11}]^{-1}(J_t^{12} + H_t^{12}); \tag{6b}$$

$$J_{t+1}^{22} = J_t^{22} + H_t^{22} - (J_t^{12} + H_t^{12})^{\mathrm{T}}[J_t^{11} + H_t^{11}]^{-1}(J_t^{12} + H_t^{12}), \tag{6c}$$

where:

$$J_{t+1} = \begin{bmatrix} J_{t+1}^{11} & J_{t+1}^{12} \\ (J_{t+1}^{12})^{\mathrm{T}} & J_{t+1}^{22} \end{bmatrix}; \qquad (s \times s) \qquad (7a)$$

$$H_t^{11} = \mathbb{E}_{p(X_{0:t+1},\theta_t,Y_{1:t+1})}[-\Delta_{X_t}^{X_t} \log p_t]; \quad (n \times n) \qquad (7b)$$

$$H_t^{12} = \mathbb{E}_{p(X_{0:t+1},\theta_t,Y_{1:t+1})}[-\Delta_{X_t}^{\theta_t} \log p_t]; \quad (n \times r) \qquad (7c)$$

$$H_t^{13} = \mathbb{E}_{p(X_{0:t+1},\theta_t,Y_{1:t+1})}[-\Delta_{X_t}^{X_{t+1}} \log p_t]; \quad (n \times n) \qquad (7d)$$

$$H_t^{22} = \mathbb{E}_{p(X_{0:t+1},\theta_t,Y_{1:t+1})}[-\Delta_{\theta_t}^{\theta_t} \log p_t]; \quad (r \times r) \qquad (7e)$$

$$H_t^{23} = \mathbb{E}_{p(X_{0:t+1},\theta_t,Y_{1:t+1})}[-\Delta_{\theta_t}^{X_{t+1}} \log p_t]; \quad (r \times n) \qquad (7f)$$

$$H_t^{33} = \mathbb{E}_{p(X_{0:t+1},\theta_t,Y_{1:t+1})}[-\Delta_{X_{t+1}}^{X_{t+1}} \log p_t]; \quad (n \times n) \qquad (7g)$$

and $p_t = p(X_{t+1}|Z_t)p(Y_{t+1}|\theta_t, X_{t+1})$. The PFIM at $t = 0$ is computed using the relation $J_0 = \mathbb{E}_{p(Z_0)}[-\Delta_{Z_0}^{Z_0} \log p(Z_0)]$.

Proof. See Tichavský et al. (1998) for proof. □

In (7a), the sub-matrices $J_{t+1}^{11} \in \mathcal{S}_{++}^{\mathrm{n}}$ and $J_{t+1}^{22} \in \mathcal{S}_{++}^{\mathrm{r}}$ are the PFIMs for the states and the parameters, respectively. The Assumptions 2.5 and 2.6 ensure that the PCRLB exists. The expectations in (7b) through (7g) make the PCRLB independent of the states, parameters and measurements. In fact, the PCRLB only depends on: the dynamics described in Model 2.1; the noise characteristics of $\{V_t\}_{t\in\mathbb{N}}$ and $\{W_t\}_{t\in\mathbb{N}}$; and the prior $(Z_0 = z_0) \sim p(z_0)$. The PCRLB inequality is general and is valid for any Bayesian estimator that solves for $\{p(z_t|Y_{1:t})\}_{t\in\mathbb{N}}$, and for any dynamical system represented by Model 2.1. In the next section, the performance of a Bayesian estimator is discussed using the PCRLB inequality in Lemma 3.1.

4. Assessment of Bayesian Estimators

Even though an optimal Bayesian estimator that computes the joint posterior $\{p(z_t|Y_{1:t})\}_{t\in\mathbb{N}}$ exactly may not be realizable in finite time for Model 2.1, the PCRLB provides its second-order performance limit. Therefore, using Lemma 3.1, the MSE performance of a Bayesian estimator can be compared against the PCRLB, such that the best performing estimator is the one that has its MSE closest to or equal to the PCRLB. Note that although comparing the MSE against the PCRLB allows for performance assessment of Bayesian estimators, it is not convenient to use in this form. This is because assessing the performance of a Bayesian estimator using (5) requires simultaneous monitoring of the MSE and PCRLB. Moreover, for certain Bayesian estimators or in certain systems, the MSE and PCRLB may be scaled poorly. To avoid the scalability issues, we define the performance of an estimator as

$$\Phi_t \equiv J_t^{-1} \circ P_{t|t}^{\circ -1}, \qquad (8)$$

where $\Phi_t \in \mathbb{R}^{\mathrm{s}\times\mathrm{s}}$ is a performance matrix that provides a measure of goodness of a Bayesian estimator. The diagonal elements in Φ_t correspond to the scaled MSE for the states and parameters, while the off-diagonal elements provide the scaled cross-MSE. For performance assessment, we only consider the diagonal entries of Φ_t. Further, the definition in (8) assumes that the Hadamard inverse of $P_{t|t}$ exists for all $t \in \mathbb{N}$. While the diagonal entries of $P_{t|t}$ are strictly greater than zero (since $P_{t|t} \in \mathcal{S}_{++}^s$), the off-diagonal entries may be zero. In such situations, (8) can not be computed for all $t \in \mathbb{N}$. As a remedy, since we are only interested in the scaled MSE entries, the performance matrix can be alternatively defined as follows

$$\overline{\Phi}_t(i,j) \equiv \begin{cases} J_t^{-1}(i,i)[P_{t|t}(i,i)]^{-1} & \text{if } i = j, \\ 0 \quad \text{otherwise.} \end{cases} \qquad (9)$$

It is straightforward to check that $\overline{\Phi}_t$ exists, and the diagonal entries of Φ_t and $\overline{\Phi}_t$ are the same for all $t \in \mathbb{N}$. In the rest of the paper, it is assumed that the performance matrix given in (8) is defined for all $t \in \mathbb{N}$, with (8) replaced by (9) at sampling time instants where (8) does not exist. The properties of Φ_t are discussed next.

Lemma 4.1. Let $J_t^{-1} \in \mathcal{S}_{++}^{\mathrm{s}}$ and $P_{t|t} \in \mathcal{S}_{++}^{\mathrm{s}}$ be such that they satisfy (5), and Φ_t be defined as in (8) then for each $i = 1, \ldots, s$ and for all $t \in \mathbb{N}$, we have

$$0 < \Phi_t(i,i) \leq 1.$$

Proof. The conditions $P_{t|t}(i,i) > 0$ and $J_t^{-1}(i,i) > 0$ imply $\Phi_t(i,i) = J_t^{-1}(i,i)[P_{t|t}(i,i)]^{-1} > 0$ for each $i = 1, \ldots, s$. Also, from (5), the condition $P_{t|t}(i,i) \geq J_t^{-1}(i,i)$ implies $\Phi_t(i,i) = J_t^{-1}(i,i)[P_{t|t}(i,i)]^{-1} \leq 1$. Combining the two yields $0 < \Phi_t(i,i) \leq 1$ for $i = 1, \ldots, s$. □

Lemma 4.1 states that $\mathrm{Tr}[\Phi_t] \in (0, s]$ for all $t \in \mathbb{N}$. The bounds in Lemma 4.1 are general and valid for any Bayesian estimator. The bounds in Lemma 4.1 are conservatively constructed, but can be refined by exploiting the specifics of a Bayesian estimator. This is discussed in Lemma 5.9. The choice of (8) as a performance metric is motivated by our ability to effectively bound it. Therefore, in situations, or for systems, where the MSE of PCRLB are poorly scaled or unbounded, the performance of a Bayesian estimator is always bounded. Note that, if required, other performance metrics can also be defined from (5).

In (8), Φ_t depends on the point estimate $\widehat{Z}_{t|t} \in \mathbb{R}^{\mathrm{s}}$ computed by an estimator. A common approach to compute $\widehat{Z}_{t|t} \in \mathbb{R}^{\mathrm{s}}$ is to minimize $\mathrm{Tr}[P_{t|t}] \in \mathbb{R}_+$. This is referred to as the minimum mean square error (MMSE) estimate, and is given by $\widehat{Z}_{t|t} = \mathbb{E}_{p(Z_t|Y_{1:t})}[Z_t]$. Other estimates, such as the maximum a posteriori (MAP) estimate are also commonly used (Trees, 1968). Since the choice of $\widehat{Z}_{t|t} \in \mathbb{R}^{\mathrm{s}}$ is non-unique in general, the estimate, for which, the estimator performance is maximized, is discussed next.

Theorem 4.2. Let $\widehat{Z}_{t|t} \in \mathbb{R}^{\mathrm{s}}$ be an MMSE point estimate then for each $i = 1, \ldots, s$

$$\widehat{Z}_{t|t}^i \in \underset{Z_{t|t}^i \in \mathbb{R}}{\arg\min}\, P_{t|t}(i,i) = \underset{Z_{t|t}^i \in \mathbb{R}}{\arg\max}\, \Phi_t(i,i). \qquad (10)$$

Here, $[\widehat{Z}_{t|t}^1, \widehat{Z}_{t|t}^2, \ldots, \widehat{Z}_{t|t}^s]^{\mathrm{T}} \equiv \widehat{Z}_{t|t}$.

PROOF. If $\widehat{Z}_{t|t}$ is an MMSE estimate, then $\widehat{Z}_{t|t}$ can be computed component-wise as follows (Trees, 1968):

$$\widehat{Z}^i_{t|t} \in \underset{Z^i_{t|t}\in\mathbb{R}}{\arg\min}\, P_{t|t}(i,i). \tag{11}$$

From (8), we have the relation

$$\Phi_t(i,i) = J_t^{-1}(i,i)[P_{t|t}(i,i)]^{-1}, \tag{12}$$

for all $i = 1,\ldots,s$. Now since $\Phi_t(i,i) > 0$ (from Lemma 4.1) and $P_{t|t}(i,i) > 0$ (since $P_{t|t} \in \mathcal{S}^{\mathrm{s}}_{++}$), we can write

$$P_{t|t}(i,i) = [\Phi_t(i,i)]^{-1} J_t^{-1}(i,i), \tag{13}$$

for all $i = 1,\ldots,s$. Substituting (13) into (11) yields

$$\widehat{Z}^i_{t|t} \in \underset{Z^i_{t|t}\in\mathbb{R}}{\arg\min}[\Phi_t(i,i)]^{-1} J_t^{-1}(i,i), \tag{14}$$

for all $i = 1,\ldots,s$. Since $J_t^{-1} \in \mathcal{S}^{\mathrm{s}}_{++}$ is independent of $\widehat{Z}_{t|t}$ (see Lemma 3.2), (14) can be written as

$$\widehat{Z}^i_{t|t} \in \underset{Z^i_{t|t}\in\mathbb{R}}{\arg\min}[\Phi_t(i,i)]^{-1} = \underset{Z^i_{t|t}\in\mathbb{R}}{\arg\max}[\Phi_t(i,i)], \tag{15}$$

for all $i = 1,\ldots,s$. The last equality results from the fact that $\Phi_t(i,i) > 0$ for all $i = 1,\ldots,s$ (from Lemma 4.1), which proves the theorem. □

According to Theorem 4.2, the performance of a Bayesian estimator, as measured in terms of (8), is maximized for the choice of the MMSE estimate; however, in practice, such an estimate can only be approximately computed. This is because the MMSE estimate computation requires computation of $\{p(z_t|Y_{1:t})\}_{t\in\mathbb{N}}$, which as discussed earlier, is difficult to compute for Model 2.1. Note that the objective of Theorem 4.2 is to only highlight that the performance of an estimator can be maximized for the choice of an MMSE point estimate; however, the user has the flexibility to select any other point estimate. In fact, MMSE point estimates are often not the best choice for multimodal posterior density functions. Nevertheless, irrespective of the choice of the point estimate, the performance of an estimator can be assessed using (8). In fact, (8) not only provides a tool to assess the performance of different estimators, it can also be used to design the estimation strategy itself. The design of a strategy for adaptive state estimation in Model 2.1 is discussed in the next section.

5. A Switching Strategy

In this section, a switching strategy is proposed for adaptive state estimation in Model 2.1. The strategy is motivated by the fact that there is no single Bayesian estimator that is guaranteed to perform optimally for a given nonlinear system and under all operating conditions. An approach to resolve this dilemma is to use a switching strategy; wherein, we start with a family of Bayesian estimators, and switch between them as and when required based on their performance. Formally, this is done as follows – let $\mathcal{B}$ be any arbitrarily chosen bank containing $F \in \mathbb{N}$ Bayesian estimators, such that $i \in \mathbb{F} \equiv \{1,2,\ldots,F\}$ indicates the index in $\mathcal{B}$. In the switching strategy, the performance of F estimators are first assessed using (8), and then the estimator with the highest performance metric is selected for delivering the state and parameter estimate at $t \in \mathbb{N}$. There are different ways in which a switching strategy can be implemented, two of which are given below:

Algorithm 5.1. Average-case switching strategy
Input: $\widehat{Z}^i_{t|t} \in \mathbb{R}^{\mathrm{s}}$ and $\Phi^{\mathrm{i}}_t \in \mathcal{S}^{\mathrm{s}}$ for all $i \in \mathbb{F}$.
Output: Point estimate $\widehat{Z}^{i_t}_{t|t} \in \mathbb{R}^{\mathrm{s}}$, where $i_t \in \mathbb{F}$.
Step 1: Compute $\mathrm{Tr}[\Phi^i_t]$ for all $i \in \mathbb{F}$.
Step 2: Solve: $i_t \in \arg\max_{i\in\mathbb{F}} \mathrm{Tr}[\Phi^i_t]$.
Step 3: Select $\widehat{Z}^{i_t}_{t|t}$ as the point estimate.

In Algorithm 5.1, the average-case switching strategy is based on the performance of an estimator in estimating all the states and parameters of Model 2.1. The estimator in bank $\mathcal{B}$, corresponding to the index in Step 2 is called the *active* estimator, and the estimate in Step 3 corresponds to the estimates from the active estimator. Alternatively, it is possible to implement the switching strategy as follows

Algorithm 5.2. Best-case switching strategy
Input: $\widehat{Z}^i_{t|t} \in \mathbb{R}^{\mathrm{s}}$ and $\Phi^{\mathrm{i}}_t \in \mathcal{S}^{\mathrm{s}}$ for all $i \in \mathbb{F}$.
Output: Point estimate $\widehat{Z}^{i_t}_{t|t} \in \mathbb{R}^{\mathrm{s}}$, where $i_t \in \mathbb{F}$.
Step 1: Set $j \leftarrow 1$.
Step 2: While $j \le s$, repeat
Solve: $i_t \in \arg\max_{i\in\mathbb{F}}[\Phi^i_t(j,j)]$.
Select $\widehat{Z}^{i_t}_{t|t}(j)$ as the point estimate.
Set $j \leftarrow j+1$.
end while

Observe that compared to Algorithm 5.1, which is a 1 dimensional switching strategy, Algorithm 5.2 provides an s dimensional strategy, wherein, each state and parameter has its own switching defined. Note that although the final estimate for the states and parameters at each sampling time are possibly selected from different estimators, Algorithms 5.1 and 5.2 require that all F estimators in bank $\mathcal{B}$ are running in parallel at all sampling times. In other words, all the estimators in the bank estimate all the states and parameters of the system at each sampling time, but only the ones suggested by Algorithms 5.1 and 5.2 are selected as the final estimate.

It should be noted that both Algorithms 5.1 and 5.2 operate at a frequency of 1.0 switch per sampling time. In many control and monitoring applications, the estimates obtained at such a high frequency may cause severe chattering (undesirable phenomenon of high amplitude and high frequency oscillation of a digital signal). The chattering problem can be mitigated by implementing a delayed switching strategy; wherein, the frequency can be reduced from 1.0 to $1/\lambda$ switches per sampling time, where $\lambda \in \mathbb{N}$. Another approach to reduce chattering in Algorithms 5.1

and 5.2 is to use a low-pass filter switching strategy. Using a low-pass filter (e.g., moving average, exponential moving average) on the sequence of estimates $\{\widehat{Z}^{i_t}_{t|t}\}_{t\in\mathbb{N}}$ will reduce any significant chattering effects. Note that the above modifications of Algorithms 5.1 and 5.2 will of course be sub-optimal; nevertheless, they are practical and easy to implement on a digital computer. The switching can be implemented in many other ways; however, we only focus on Algorithms 5.1 and 5.2. Next, we discuss the optimality of the estimates obtained from Algorithms 5.1 and 5.2.

Result 5.3. Given $\mathcal{B}$, with $F \in \mathbb{N}$ estimators, let $\widehat{Z}^{i}_{t|t} \in \mathbb{R}^{\mathrm{s}}$ be an MMSE estimate computed by the i-th estimator for all $i \in \mathbb{F}$, then the state and parameter estimates generated from Algorithms 5.1 and 5.2 are the MMSE estimates in the average and optimal sense, respectively.

For a bank $\mathcal{B}$ with $\mathbb{F} \in \mathbb{N}$ estimators, each estimator calculates its own approximate MMSE estimate. Now for a given set of approximate MMSE estimates, Algorithm 5.1 selects the estimate from the estimator with maximum average performance (as measured by the trace of the performance matrix). Algorithm 5.2 implements a similar strategy, but at a more individual level – for each states and parameters. In fact, the very construction of Algorithms 5.1 and 5.2, ensure that Algorithms 5.1 and 5.2 yield MMSE estimates in the average and optimal sense, respectively.

Remark 5.4. Result 5.3 establishes the optimality of the switching strategy only with respect to a bank of Bayesian estimators; however, compared to the optimal Bayeisan estimator, the switching strategy is still sub-optimal.

Despite the observation in Remark 5.4, the key advantage of the proposed switching strategy is that it provides an optimal approach to combine a set of sub-optimal Bayesian estimators. In other words, the switching strategy is at least as good as all existing estimators in the bank $\mathcal{B}$. In theory, if the optimal estimator is included in the bank $\mathcal{B}$ then Algorithms 5.1 and 5.2 select the optimal estimator for all $t \in \mathbb{N}$. This is because an optimal estimator is efficient, i.e., $\Phi_t(i,i) = 1$ for all $i = 1, \ldots, s$.

It is highlighted that the quality of estimates obtained with Algorithms 5.1 and 5.2 is proportional to the quality of estimators in bank $\mathcal{B}$. Note that the switching strategy does not provide any independent solution to the adaptive state estimation problem, but instead, it provides an optimal approach to combine existing estimators to deliver estimates, that are at least as good as the estimates from the individual estimators. In this paper, no restriction is placed on the choice of $\mathcal{B}$. The user has full flexibility to decide its construction. Of course, there are computational considerations in choosing $\mathcal{B}$, this is discussed next.

The implementation of the switching strategy through Algorithms 5.1 and 5.2 incurs significant computational costs in computing the point estimates and performance matrix for each estimator at each sampling time. Even though, the estimate is only selected from a single best performing estimator, we need to run $\mathbb{F}$ estimators in parallel at each sampling time. Therefore, to keep the overall computational complexity of Algorithms 5.1 and 5.2 low, it is important to select low-computational cost, but high-performance estimators in the bank $\mathcal{B}$. A systematic approach to reduce the computational complexity of Algorithms 5.1 and 5.2 is to use a dynamic bank; wherein, the estimators can be automatically included or removed based on the operating conditions. For example, when the system dynamics are linear or *mildly* nonlinear, the use of Kalman-based estimators are more economical; whereas, while operating in highly nonlinear regimes, the SMC-based estimators are known to perform better. The construction of such a dynamic bank will of course depend on the measure of the system nonlinearity, which even for simple dynamical systems is either nontrivial to compute or requires enormous computational effort. Fortunately, for many of systems with the first principle models available, the system nonlinearity can assessed at least qualitatively. Although the construction of a dynamic filter bank is not considered in this paper, but if required, it can be readily incorporated with Algorithms 5.1 and 5.2. Next we discuss the stability of the switching strategy. First, some standard results on the boundedness of stochastic processes are recalled (Agniel and Jury, 1971; Tarn and Rasis, 1976).

Definition 5.5. A stochastic process $\{\gamma_t\}_{t\in\mathbb{N}}$ is stable or exponentially bounded in mean square if there exists real numbers $\eta \in \mathbb{R}$, $\nu \in \mathbb{R}_+$ and $0 < \rho < 1$, such that

$$\mathbb{E}_{p(\gamma_t)}||\gamma_t||^2 \leq \eta||\gamma_0||^2\rho^t + \nu, \tag{16}$$

holds for all $t \in \mathbb{N}$.

Definition 5.5 implies that $\{\gamma_t\}_{t\in\mathbb{N}}$ is stable only if there exists a finite upper bound, such that $\lim_{t\to+\infty} \eta||\gamma_0||^2\rho^t + \nu = \nu$. The necessary conditions for the stability of a stochastic process are given next.

Lemma 5.6. For $\{\gamma_t\}_{t\in\mathbb{N}}$, let $v_{\min}, v_{\max}, \mu \in \mathbb{R}_+$, $0 < \lambda < 1$, and $V : \mathbb{R} \to \mathbb{R}$ be such that the following inequalities

$$v_{\min}||\gamma_t||^2 \leq V(\gamma_t) \leq v_{\max}||\gamma_t||^2, \tag{17a}$$

$$\mathbb{E}_{p(\gamma_t|\gamma_{t-1})}[V(\gamma_t)] - V(\gamma_{t-1}) \leq \mu - \lambda V(\gamma_t - 1), \tag{17b}$$

hold for $t \in \mathbb{N}$ then $\{\gamma_t\}_{t\in\mathbb{N}}$ is bounded in mean square, such that (16) holds for $t \in \mathbb{N}$ with $\eta = v_{\max}v_{\min}^{-1}$, $\rho = (1-\lambda)$ and $\nu = \mu v_{\min}^{-1}\sum_{k=1}^{t-1}\rho^k$.

Proof. See Reif et al. (1999) for a detailed proof. □

Lemma 5.6 gives the conditions for a stochastic process to be stable. The process we are interested in analysing is the estimation error $\gamma_t = Z_t - \widehat{Z}_{t|t}$ for all $t \in \mathbb{N}$. The next lemma discusses the stability of the estimation error.

Lemma 5.7. Let $\widehat{Z}_{t|t} \in \mathbb{R}^{\mathrm{s}}$ be an estimate computed by a Bayesian estimator. Let $P_{t|t} \in \mathcal{S}^{\mathrm{s}}_{++}$ and $J_t^{-1} \in \mathcal{S}^{\mathrm{s}}_{++}$ be the MSE of the estimator and the PCRLB for Model 2.1,

respectively. If $\gamma_t = Z_t - \widehat{Z}_{t|t}$ is a stable process for all $t \in \mathbb{N}$ then $\mathrm{Tr}[P_{t|t}]$ is bounded from below and above by

$$\mathrm{Tr}[J_t^{-1}] \leq \mathrm{Tr}[P_{t|t}] \leq \eta \,\mathrm{Tr}[P_{0|0}]\rho^t + \nu \qquad (t \in \mathbb{N}), \tag{18}$$

where $\eta \in \mathbb{R}$, $\nu \in \mathbb{R}_+$ and $0 < \rho < 1$ are real numbers defined in Lemma 5.6.

PROOF. Note that for the stochastic process $\gamma_t = Z_t - \widehat{Z}_{t|t}$, we have $\mathbb{E}_{p(\gamma_t)}||\gamma_t||^2 = \mathrm{Tr}[P_{t|t}]$. Now, since γ_t is stable, substituting $\gamma_t = Z_t - \widehat{Z}_{t|t}$ into Definition 5.5 and using the result $P_{t|t} - J_t^{-1} \in \mathcal{S}_+^{\mathrm{s}}$ from Lemma 3.1, we get (18). □

Lemma 5.7 provides a lower and upper bound on the MSE of a Bayesian estimator with stable estimation error. Note that Lemma 5.7 is also valid for estimators with unstable estimation error, in which case, the lower bound is still given by $\mathrm{Tr}[J_t^{-1}]$ (since $\mathrm{Tr}[J_t^{-1}]$ is independent of the estimator); however, the upper bound is infinite (since η, ν and ρ are infinite). Using Lemma 5.7, we define the stability of an estimator in terms of its estimation error.

Definition 5.8. Let $P_{t|t} \in \mathcal{S}_{++}^{\mathrm{s}}$ be the MSE for a Bayesian estimator, with $\mathrm{Tr}[P_{0|0}] < +\infty$, then the estimator is stable if there exists $\eta \in \mathbb{R}$, $\nu \in \mathbb{R}_+$ and $0 < \rho < 1$, such that the upper bound on $\mathrm{Tr}[P_{t|t}]$ in Lemma 5.7 is finite with

$$\lim_{t \to +\infty} \eta \,\mathrm{Tr}[P_{0|0}]\rho^t + \nu = \nu. \tag{19}$$

From Definition 5.8, a stable estimator is the one with a finite MSE upper bound. Now, before discussing the stability of the switching strategy, a bound on Φ_t in (8) is derived. Observe that the lower bound on Φ_t in Lemma 4.1 is trivial for it is valid for any given Bayesian estimator; however, using Lemma 5.7, an estimator specific lower bound on Φ_t is defined, and is given next.

Lemma 5.9. If $P_{t|t} \in \mathcal{S}_{++}^{\mathrm{s}}$ denotes the MSE for a stable estimator then its performance Φ_t at $t \in \mathbb{N}$ is bounded from below and above by the following inequality

$$\frac{\mathrm{Tr}[J_t^{-1}]}{\eta \,\mathrm{Tr}[P_{0|0}]\rho^t + \nu} < \mathrm{Tr}[\Phi_t] \leq s, \tag{20}$$

where $\eta \in \mathbb{R}$, $\nu \in \mathbb{R}_+$ and $0 < \rho < 1$ are real numbers given by Lemma 5.6.

PROOF. From Lemma 4.1, the upper bound on $\mathrm{Tr}[\Phi_t]$ is s or $\mathrm{Tr}[\Phi_t] \leq s$. To derive the lower bound, note that for $P_{t|t} \in \mathcal{S}_{++}^{\mathrm{s}}$, the upper bound in (18) can be written as

$$P_{t|t}(i,i) \leq \eta \,\mathrm{Tr}[P_{0|0}]\rho^t + \nu, \tag{21}$$

for all $i = 1, \ldots, s$, or alternatively as

$$[P_{t|t}(i,i)]^{-1} \geq \frac{1}{\eta \,\mathrm{Tr}[P_{0|0}]\rho^t + \nu}, \tag{22}$$

since $\eta \,\mathrm{Tr}[P_{0|0}]\rho^t + \nu > 0$ (see Lemma 5.7). Now using the relations $J_t^{-1}(i,i) > 0$ (see Lemma 3.1) and (22), we get

$$\Phi_t(i,i) = J_t^{-1}(i,i)[P_{t|t}(i,i)]^{-1} > \frac{J_t^{-1}(i,i)}{\eta \,\mathrm{Tr}[P_{0|0}]\rho^t + \nu}, \tag{23a}$$

for all $i = 1, \ldots, s$. Finally, since $\Phi_t(i,i) > 0$ and $J^{-1}(i,i) > 0$ for all $i = 1, \ldots, s$, we can write

$$\frac{\mathrm{Tr}[J_t^{-1}]}{\eta \,\mathrm{Tr}[P_{0|0}]\rho^t + \nu} < \mathrm{Tr}[\Phi_t], \tag{24}$$

which completes the proof. □

Lemma 5.9 provides a nontrivial bounds on the performance of an estimator. In contrast to Lemma 4.1, Lemma 5.9 delivers bounds that depend on the estimator. Using Lemma 5.9, Definition 5.8 can be alternatively stated in terms of the PCRLB. This is given next.

Remark 5.10. A stable Bayesian estimator with $\mathrm{Tr}[P_{0|0}] < \infty$ satisfying the condition (19) in Definition 5.8 also satisfies the relation (from Lemma 5.9)

$$\lim_{t \to +\infty} \frac{\mathrm{Tr}[J_t^{-1}]}{\eta \,\mathrm{Tr}[P_{0|0}]\rho^t + \nu} > 0. \tag{25}$$

Similarly, if the estimator is unstable, the denominator in (25) grows unboundedly, such that

$$\lim_{t \to +\infty} \frac{\mathrm{Tr}[J_t^{-1}]}{\eta \,\mathrm{Tr}[P_{0|0}]\rho^t + \nu} = 0. \tag{26}$$

Using Lemma 5.9 and Remark 5.10, the stability of the switching strategy is given next.

Theorem 5.11. Let the bank $\mathcal{B}$ contain $F \in \mathbb{N}$ estimators indexed by set $i \in \mathbb{F}$. Let the MSE for all the estimators at at $t = 0$ be $P_{0|0}$, then the performance of the switching strategy for Model 2.1, and implemented with Algorithm 5.1 satisfies

$$\frac{\mathrm{Tr}[J_t^{-1}]}{\eta_{j_t} \,\mathrm{Tr}[P_{0|0}]\rho_{j_t}^t + \nu_{j_t}} < \mathrm{Tr}[\Phi_t^{i_t}] \leq s, \tag{27}$$

where $\eta_i \in \mathbb{R}$, $\nu_i \in \mathbb{R}_+$ and $0 < \rho_i < 1$ are real numbers given by Lemma 5.6 for each estimator $i \in \mathbb{F}$ and $i_t \in \mathbb{F}$ is the estimator index computed by Algorithm 5.1 and

$$j_t \in \arg\min_{i \in \mathbb{F}} [\eta_i \,\mathrm{Tr}[P_{0|0}]\rho_i^t + \nu_i]. \tag{28}$$

PROOF. Observe that from Lemma 5.9, we have $\mathrm{Tr}[\Phi_t^i] \leq s$ for all $i \in \mathbb{F}$ or $\mathrm{Tr}[\Phi_t^{i_t}] \leq s$, where $i_t \in \mathbb{F}$. Now from Algorithm 5.1, we have the following relation

$$\max_{k \in \mathbb{F}} \mathrm{Tr}[\Phi_t^k] \geq \mathrm{Tr}[\Phi_t^j], \tag{29}$$

for all $j \in \mathbb{F}$. From Algorithm 5.1, we have

$$i_t \in \arg\max_{i \in \mathbb{F}} \mathrm{Tr}[\Phi_t^i], \tag{30}$$

which when substituted into (29) yields,

$$\mathrm{Tr}[\Phi_t^{i_t}] \geq \mathrm{Tr}[\Phi_t^{j}], \tag{31}$$

for all $j \in \mathbb{F}$, or we can write (31) for the index $j_t \in \mathbb{F}$ as

$$\mathrm{Tr}[\Phi_t^{i_t}] \geq \mathrm{Tr}[\Phi_t^{j_t}] > \frac{\mathrm{Tr}[J_t^{-1}]}{\eta_{j_t}\,\mathrm{Tr}[P_{0|0}]\rho_{j_t}^t + \nu_{j_t}}. \tag{32}$$

The last inequality uses Lemma 5.9 for index $j_t \in \mathbb{F}$. This completes the proof.

Theorem 6.1 ensures that the performance of the switching strategy is bounded and also guarantees that it yields the tightest performance bounds (recall that $j_t \in \mathbb{F}$ in (27) yields the largest lower bound on $\mathrm{Tr}[\Phi_t^{i_t}]$).

Corollary 5.12. Let the bank $\mathcal{B}$ contain $F \in \mathbb{N}$ estimators, of which $F_s \leq F$ are stable, where $F_s \in \mathbb{F}_s \subseteq \mathbb{F}$ and $F_s \geq 1$, then the switching strategy for Model 2.1, and implemented with Algorithm 5.1 is stable such that

$$\lim_{t\to+\infty} \frac{\mathrm{Tr}[J_t^{-1}]}{\eta_{j_t}\,\mathrm{Tr}[P_{0|0}]\rho_{j_t}^t + \nu_{j_t}} > 0, \tag{33}$$

where j_t is given by (28) in Theorem 5.11.

Proof. Here we only need to show that the index j_t in (28) in Theorem 6.1 corresponding to a stable estimator, such that $j_t \in \mathbb{F}_s$. First, such an index exist since $F_s \geq 1$. Next, (28) can be simplified and written as

$$j_t \in \arg\min_{i\in\mathbb{F}_s\subseteq\mathbb{F}} [\eta_i\,\mathrm{Tr}[P_{0|0}]\rho_i^t + \nu_i], \tag{34}$$

where the domain is now over the space of stable estimators. This is because from Definition 5.8, the objective function in (28) is not finite for unstable estimators, and therefore, can be removed from $\mathbb{F}$. Now since $j_t \in \mathbb{F}_s$, from (25) in Remark 5.10, we have the required result. □.

According to Corollary 5.12, as long as the bank $\mathcal{B}$ has at least one stable estimator, the switching strategy is guaranteed to choose the best performing stable estimator at each sampling time. The stability of Algorithm 5.2 can be similarly established by modifying Theorem 5.11; however, for the sake of brevity the details are skipped here.

6. Numerical Computation

The switching strategy discussed in Section 5 provides an efficient strategy for adaptive state estimation in Model 2.1; however, the switching can only be computed off-line. This is because the PCRLB, required to compute the performance matrix in Algorithms 5.1 and 5.2, involves multi-dimensional integrals that do not admit any closed form solutions for Models 2.1 and 2.7. In applications, such as control and monitoring, the design, performance evaluation and selection of estimators are mostly done a priori or off-line. For our purposes, we compute the switching off-line, but is always implemented in real-time. This is justified, since the PCRLB is independent of the states, parameters and measurements (see Section 3).

Monte-Carlo (MC) method is a popular class of numerical methods to approximate complex integrals. Using MC methods, the PCRLB can be approximated by simulating M i.i.d. sample paths $(z_{0:t}^j, y_{1:t}^j)_{j=1}^M \in \mathcal{Z}^{t+1} \times \mathcal{Y}^t$ starting with M i.i.d. initial draws distributed according to $p(z_0)$. The procedure to compute a MC approximation of the PCRLB is well known, and is not included here for the sake of brevity. See Tulsyan et al. (2013d) and Bergman (2001) for details on how PCRLB approximations are computed. Here, we only present PCRLB approximations for Model 2.7, but first, we show how the dimension of the expectations in Lemma 3.2 can be reduced for Model 2.7.

Lemma 6.1. For a stochastic system represented by Model 2.7 and operating under Assumptions 2.2 through 2.6, the matrices (7b) through (7g) can be written as:

$$H_t^{11} = \mathbb{E}_{p_1}[\nabla_{X_t} f_t^{\mathrm{T}}(X_t,\theta_t)]Q_t^{-1}[\nabla_{X_t} f_t^{\mathrm{T}}(X_t,\theta_t)]^{\mathrm{T}}; \tag{35a}$$

$$H_t^{12} = \mathbb{E}_{p_1}[\nabla_{X_t} f_t^{\mathrm{T}}(X_t,\theta_t)]Q_t^{-1}[\nabla_{\theta_t} f_t^{\mathrm{T}}(X_t,\theta_t)]^{\mathrm{T}}; \tag{35b}$$

$$H_t^{13} = \mathbb{E}_{p_1}[-\nabla_{X_t} f_t^{\mathrm{T}}(X_t,\theta_t)]Q_t^{-1}; \tag{35c}$$

$$\begin{aligned} H_t^{22} &= \mathbb{E}_{p_1}[\nabla_{\theta_t} f_t^{\mathrm{T}}(X_t,\theta_t)]Q_t^{-1}[\nabla_{\theta_t} f_t^{\mathrm{T}}(X_t,\theta_t)]^{\mathrm{T}} \\ &\quad + \mathbb{E}_{p_2}[\nabla_{\theta_t} g_t^{\mathrm{T}}(X_{t+1},\theta_t)]R_{t+1}^{-1}[\nabla_{\theta_t} g_t^{\mathrm{T}}(X_{t+1},\theta_t)]^{\mathrm{T}}; \end{aligned} \tag{35d}$$

$$\begin{aligned} H_t^{23} &= \mathbb{E}_{p_1}[-\nabla_{\theta_t} f_t^{\mathrm{T}}(X_t,\theta_t)]Q_t^{-1} + \mathbb{E}_{p_2}[\nabla_{\theta_t} g_t^{\mathrm{T}}(X_{t+1},\theta_t)] \\ &\quad \times R_{t+1}^{-1}[\nabla_{X_{t+1}} g_t^{\mathrm{T}}(X_{t+1},\theta_t)]^{\mathrm{T}}; \end{aligned} \tag{35e}$$

$$\begin{aligned} H_t^{33} &= Q_t^{-1} + \mathbb{E}_{p_2}[\nabla_{X_{t+1}} g_t^{\mathrm{T}}(X_{t+1},\theta_t)]R_{t+1}^{-1} \\ &\quad \times [\nabla_{X_{t+1}} g_t^{\mathrm{T}}(X_{t+1},\theta_t)]^{\mathrm{T}}, \end{aligned} \tag{35f}$$

where $p_1 = p(X_t,\theta_t)$ and $p_2 = p(X_{t+1},\theta_t)$.

Proof. (35a): Note that $H_t^{11} = \mathbb{E}_{\tilde{p}_{t+1}}[-\Delta_{X_t}^{X_t} \log p_t]$ in (7b), where $\tilde{p}_{t+1} = p(X_{0:t+1},\theta_t,Y_{1:t+1})$ and $p_t = p(X_{t+1}|Z_t)$ $p(Y_{t+1}|\theta_t,X_{t+1})$ can be written as (Tichavský et al., 1998)

$$H_t^{11} = \mathbb{E}_{\tilde{p}_{t+1}}[\nabla_{X_t} \log p_t][\nabla_{X_t} \log p_t]^{\mathrm{T}}. \tag{36}$$

Now, since $\nabla_{X_t} \log p(Y_{t+1}|X_{t+1},\theta_t) = 0$, we have

$$\begin{aligned} &[\nabla_{X_t} \log p_t] \\ &= \nabla_{X_t}[\log p(X_{t+1}|Z_t) + \log p(Y_{t+1}|\theta_t,X_{t+1})], &\text{(37a)}\\ &= \nabla_{X_t}[\log p(X_{t+1}|Z_t)]. &\text{(37b)} \end{aligned}$$

Substituting (37b) into (36) yields

$$H_t^{11} = \mathbb{E}_{\tilde{p}_{t+1}}[G_t][G_t]^{\mathrm{T}}, \tag{38}$$

where $G_t = [\nabla_{X_t} \log p(X_{t+1}|Z_t)]$. Noting that the integrand in (38) is a function of $\{X_{t+1}, Z_t\}$, we have

$$\begin{aligned} H_t^{11} &= \mathbb{E}_{p(X_{t+1},X_t,\theta_t)}[G_t][G_t]^{\mathrm{T}}, &\text{(39a)}\\ &= \mathbb{E}_{p(X_t,\theta_t)p(X_{t+1}|X_t,\theta_t)}[G_t][G_t]^{\mathrm{T}}. &\text{(39b)} \end{aligned}$$

Now, for Model 2.7, we can write

$$G_t = [\nabla_{X_t} f_t^{\mathrm{T}}(X_t,\theta_t)]Q_t^{-1}[X_{t+1} - f_t(X_t,\theta_t)]. \tag{40}$$

Substituting (40) into (39b), and noting $E_{p(X_{t+1}|Z_t)}[X_{t+1} - f_t(X_t,\theta_t)][X_{t+1} - f_t(X_t,\theta_t)]^{\mathrm{T}} = Q_t$, we get (35a). The expressions (35b) through (35f) can be similarly derived. □

Lemma 6.1 reduces the dimensions of the integrals in (7b) through (7g) from $(t+1)(n+m)+s$ to s. Using MC sampling methods, as discussed in Tulsyan et al. (2013d), the integral (35a), for instance, can be approximated as

$$\tilde{H}_t^{11} = \frac{1}{M}\sum_{i=1}^{M}[\nabla_{X_t} f^{\mathrm{T}}(X_t^i,\theta_t^i)]Q_t^{-1}[\nabla_{X_t} f^{\mathrm{T}}(X_t^i,\theta_t^i)]^{\mathrm{T}}, \quad (41)$$

where $\{X_t^i,\theta_t^i\}_{i=1}^M \sim p(x_t,\theta_t)$ and $\tilde{H}_t^{11}$ is an M sample MC estimate of H_t^{11}. The MC estimates of (35b) through (35f) can be similarly computed. It can be shown that the MC approximation of the PCRLB asymptotically converges to the true PCRLB (Tulsyan et al., 2013d). Finally, it is highlighted that for $M < +\infty$, the MC approximation of the PCRLB may not satisfy $P_{t|t} - J_t^{-1} \succcurlyeq 0$ for all $t \in \mathbb{N}$.

7. Example

The strategy for adaptive state estimation proposed in Section 5 is general, and is applicable to many engineering systems, including chemical, mechanical and aerospace. We demonstrate the efficacy of the switching strategy on a ballistic target tracking problem at re-entry phase (Farina et al., 2002; Li and Jilkov, 2001; Ristic et al., 2003).

7.1. Setup

Consider a target launched along a ballistic flight whose kinematics are described in a 2-D Cartesian coordinate system. This particular description of the kinematics assumes that the only forces acting on the target at any given time are the forces due to gravity and drag. All other forces (e.g., centrifugal acceleration, Coriolis acceleration, wind, lift force, and spinning motion) are assumed to be negligible or have a small effect on the target dynamics. With target position and velocity at $t \in \mathbb{N}$ given as $(\mathrm{x}_t, \mathrm{h}_t)$ and $(\dot{\mathrm{x}}_t, \dot{\mathrm{h}}_t)$, respectively, the target motion in its re-entry phase can be described by (Tulsyan et al., 2013a):

$$X_{t+1} = AX_t + GF_t(X_t) + G[0 \;\; -g]^{\mathrm{T}} + V_t, \quad (42)$$

where $X_t \equiv [\mathrm{x}_t \;\; \dot{\mathrm{x}}_t \;\; \mathrm{h}_t \;\; \dot{\mathrm{h}}_t]^{\mathrm{T}}$ is a hidden stochastic state process, and the matrices A and G are as follows

$$A = I_{2\times 2} \otimes \begin{bmatrix} 1 & \Delta \\ 0 & 1 \end{bmatrix}, \quad G = I_{2\times 2} \otimes \begin{bmatrix} \Delta^2/2 \\ \Delta \end{bmatrix}. \quad (43)$$

Here, Δ is the time lapse between two consecutive radar measurements. In (42), $F(X_t)$ models the drag force, which acts in a direction opposite to the target velocity. In terms of the hidden states, $F(X_t)$ can be modelled as

$$F_t(X_t) = -\frac{g\rho(\mathrm{h}_t)}{2\theta_t}\sqrt{(\dot{\mathrm{x}}_t^2 + \dot{\mathrm{h}}_t^2)}\begin{bmatrix} \dot{\mathrm{x}}_t \\ \dot{\mathrm{h}}_t \end{bmatrix}, \quad (44)$$

Table 1: Parameter values used in the example.

Variables	Symbol	Values
acceleration due to gravity	g	9.8 m/s^2
radar sampling	Δ	2 s
total tracking time	N	120 s
noise parameters	γ	1
	σ_l	0.1 km
	σ_ϵ	0.017 rad
probability of detection	$\mathrm{Pr_d}$	1
probability of false alarm	$\mathrm{Pr_f}$	0

where: g is the acceleration due to gravity; $\theta_t \in \mathbb{R}$ is the unknown ballistic coefficient (model parameter), whose value depends on the shape, mass and the cross sectional area of the target; and $\rho(\mathrm{h}_t)$ is the air density, such that $\rho(\mathrm{h}_t) = \alpha_1 e^{(-\alpha_2 \mathrm{h}_t)}$, where: $\alpha_1 = 1.227\mathrm{kgm}^{-3}$, $\alpha_2 = 1.093\times 10^{-4}\mathrm{m}^{-1}$ for $\mathrm{h}_t < 9144\mathrm{m}$; and $\alpha_1 = 1.754\mathrm{kgm}^{-3}$, $\alpha_2 = 1.4910\times 10^{-4}\mathrm{m}^{-1}$ for $\mathrm{h}_t \geq 9144\mathrm{m}$. In (44), $V_t \sim \mathcal{N}(0, Q_t)$ is an independent sequence of multivariate Gaussian random variables, with zero mean and covariance

$$Q_t = \gamma I_{2\times 2} \otimes \begin{bmatrix} \Delta^3/3 & \Delta^2/2 \\ \Delta^2/2 & \Delta \end{bmatrix}, \quad (45)$$

where $\gamma \in \mathbb{R}_+$ is the noise intensity, which accounts for all the forces neglected in (42), including any deviations arising due to system-model mismatch. The target measurements $Y_t = [L_t, E_t]^{\mathrm{T}}$, where L_t is the target range and E_t the target elevation are collected by a dish radar assumed to be stationed at the origin, such that

$$Y_t = \begin{bmatrix} \sqrt{(\mathrm{x}_t^2 + \mathrm{h}_t^2)} \\ \arctan(\mathrm{h}_t/\mathrm{x}_t) \end{bmatrix} + W_t, \quad (46)$$

where $W_t \sim \mathcal{N}(0, R_t)$ is a zero mean Gaussian with

$$R_t = \begin{bmatrix} \sigma_l^2 & \Delta \\ 0 & \sigma_e^2 \end{bmatrix}, \quad (47)$$

where $\sigma_l \in \mathbb{R}_+$ and $\sigma_e \in \mathbb{R}_+$ are standard deviations for the range and elevation readings. The target elevation angle is assumed to be between 0 and $\pi/2$ radians. The parameters used in this example are given in Table 2. Moreover, it can be shown that the system satisfies Assumptions 2.4 through 2.3. Let $Z_t = [X_t^{\mathrm{T}}, \theta_t]^{\mathrm{T}}$ be a vector of hidden target states and ballistic coefficient. The prior is $Z_0 \sim \mathcal{N}(M_{z_0}, C_{z_0})$, with $M_{z_0} = [232\mathrm{km}, 2.290\cos 190^o\mathrm{kms}^{-1}, 88\mathrm{km}, 2.290\sin 190^o\mathrm{kms}^{-1}, 40000\mathrm{kgm}^{-1}s^{-2}]^{\mathrm{T}}$ and $[C_{z_0}]^{0.5} = \mathrm{diag}([1\mathrm{km}, 20\mathrm{ms}^{-1}, 1\mathrm{km} \;\; 20\mathrm{ms}^{-1}, 20\mathrm{kgm}^{-1}s^{-2}])$. Now, the objective is to estimate $\{Z_t\}_{t\in\mathbb{N}}$ in real-time.

7.2. Results

The results presented in this section are valid only under the settings and assumptions considered in Section 7.1. Starting with $Z_0 \sim \mathcal{N}(M_{z_0}, C_{z_0})$, Fig.1 gives the PCRLB

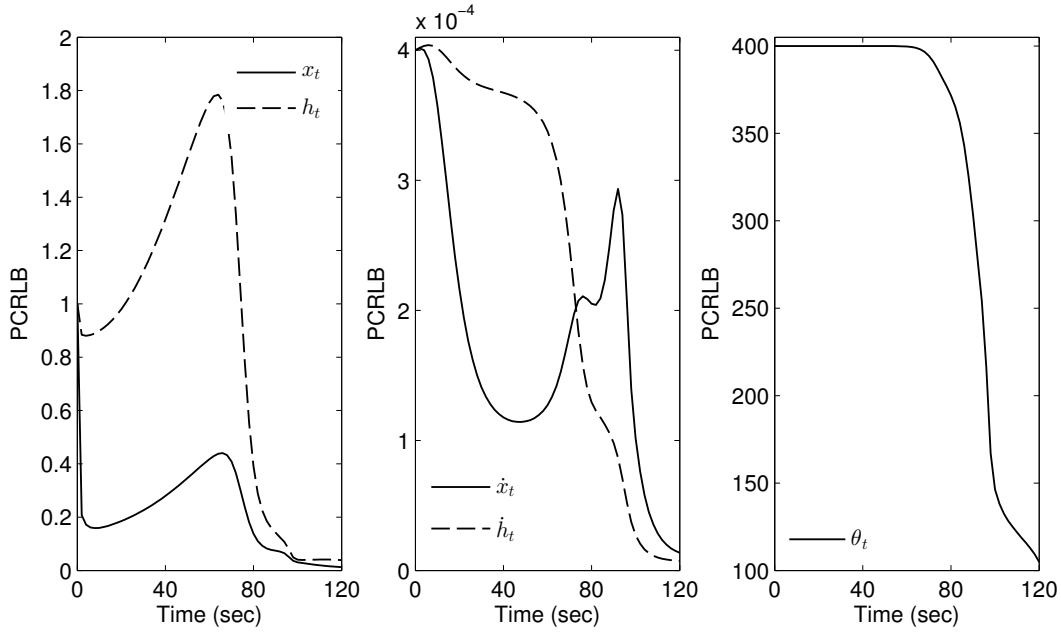

Figure 1: The PCRLB associated with the target states and ballistic coefficient. The results are based on $M = 2000$ MC simulations.

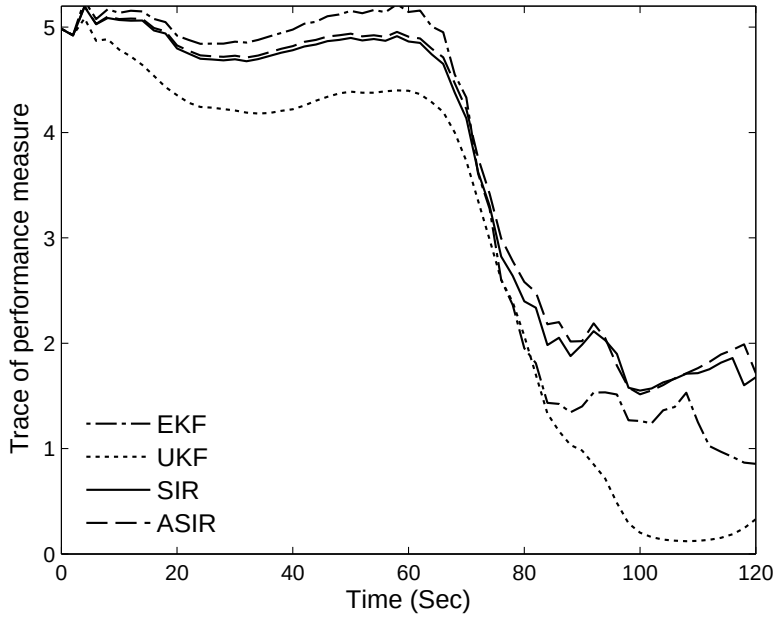

Figure 2: Performance of estimators in bank $\mathcal{B}$.

for $\{Z_t\}_{t\in\mathbb{N}}$. The PCRLB is obtained as $J_t^{-1}(i,i)$, where $i = 1,\ldots,5$. In the interval $0 \leq t \leq 60$, the PCRLB for $\{\theta_t\}_{t\in\mathbb{N}}$ is almost constant. This is due to the absence of drag at higher altitude, where the air density $\rho(\mathrm{h}_t)$ is thin, and the target dynamics are linear (i.e., $F(X_t)$ is approximately zero), such that the radar measurements contain no additional information to estimate $\{\theta_t\}_{t\in\mathbb{N}}$. In the interval $60 < t \leq 90$, due to thicker air density, the drag force increases drastically; and the PCRLB for $\{X_t\}_{t\in\mathbb{N}}$ grows, but that of $\{\theta_t\}_{t\in\mathbb{N}}$ decreases sharply. Finally, in the segment $90 < t \leq N$, the PCRLB for $\{Z_t\}_{t\in\mathbb{N}}$ decreases.

The artificial dynamics approach (ADA) is a popular class of Bayesian estimators for adaptive state estimation (Kantas et al., 2009). In ADA, an artificial noise is added to $\{\theta_t\}_{t\in\mathbb{N}}$, such that it no longer involves any deterministic transitions. The ADA can be implemented with both Kalman and SMC-based filters; however, there are two long-standing problems with the ADA approach, as identified and summarized in (Kantas et al., 2009):

(a) the dynamics of the parameters is related to the width of the kernel and the variance of the artificial noise, which are often difficult to fine tune; and

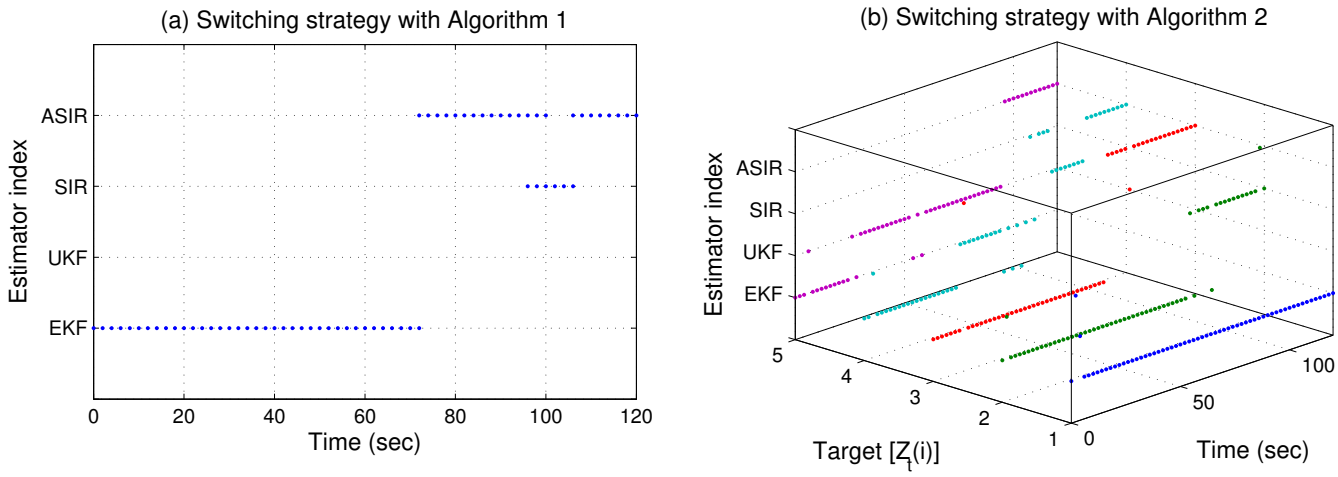

Figure 3: The switching strategy implemented with Algorithms 5.1 and 5.2.

Table 2: Comparing different Bayesian estimators with the switching strategy implemented through Algorithms 5.1 and 5.2 using the sum of the trace of normalized MSE (SNMSE).

	EKF	UKF	SIR	ASIR	Algorithm 5.1	Algorithm 5.2
SNMSE	798	5915	668	670	562	489

(b) transforming the estimation problem, by adding artificial noise to the parameters modifies the original problem, so that, it becomes hard to quantify the bias introduced in the resulting estimates.

The first issue was addressed in Tulsyan et al. (2013c); wherein, an approach to auto-tune the kernel width was developed. For the second issue, Tulsyan et al. (2013b) proposes a systematic approach that can be used to perform error analysis in Bayesian estimators, including the ADA. It is important to note that like other Bayesian estimators, such as practical filtering, MCMC and SMC^2 estimator, the ADA is also a sub-optimal estimator. Consider a bank $\mathcal{B}$ with the following ADA estimators:

(a) Extended Kalman filter (EKF);
(b) Unscented Kalman filter (UKF);
(c) Sequential importance resampling (SIR) filter;
(d) Adaptive SIR (ASIR) filter.

Here, EKF and UKF are the Kalman-based filters, while SIR and ASIR are the SMC-based filters. A complete description of the EKF, UKF and SIR filters, including their implementation can be found here (Ristic et al., 2003; Ching et al., 2006; Chen et al., 2005; Wenzel et al., 2006; Su et al., 2003), while the description for the ASIR filter can be found in Tulsyan et al. (2013c). To assess the performance of the above four estimators in approximating $\{p(z_t|Y_{1:t})\}_{t\in\mathbb{N}}$, we construct a bank $\mathcal{B}$, with $\mathbb{F} = \{1,2,3,4\}$. This paper does not advocate the use ADA estimators over other Bayesian estimators. It is possible for other estimators to outperform ADA estimators on the given problem. Here, the choice of ADA estimators is for illustrative purposes; however, the user has the flexibility to design their own bank $\mathcal{B}$.

Fig.2 gives $\mathrm{Tr}[\Phi_t]$ values for the estimators in $\mathcal{B}$. From Fig.2, it is clear that the estimators have a high performance in the interval $0 \leq t \leq 70$, but plummet in the interval $70 < t \leq N$. This is due to the large drag force at lower altitude, which shifts the dynamics from linear to nonlinear, and thus making tracking difficult. Note that in Fig.2, $\mathrm{Tr}[\Phi_t]) > 5$ for certain estimators. This is due to the error in computing the MC approximation of the PCRLB (see Section 6). Figs.3(a) and (b) give the average-case and best-case switching strategies (see Algorithms 5.1 and 5.2, respectively) for this problem. At higher altitude, where the target dynamics are linear, both the strategies suggest using EKF, but then switch to other advanced filters, such as SIR and ASIR at lower altitude. Although Algorithm 5.2 extensively uses UKF in estimating $\{Z_t(4)\}_{t\in\mathbb{N}}$ and $\{Z_t(5)\}_{t\in\mathbb{N}}$ (see Fig.3(b)), due to its poor average-performance (see Fig.2), its use is not recommended by Algorithm 5.1 (see Fig.3(a)). Note that although the switching in Figs.3(a) and (b) are computed off-line, it is implemented in real-time. This is because the switching in Figs.3(a) and (b) are independent of any particular realization of the input-output data. Table 2 compares different estimators and strategies based on their sum of the trace of normalized MSE (SNMSE) values, which is defined as

$$\mathrm{SNMSE} = \sum_{t=1}^{\mathrm{T}} \mathrm{Tr}\left[P_{t|t} \circ [J_t^{-1}]^{\circ -1}\right] = \sum_{t=1}^{\mathrm{T}} \mathrm{Tr}[\Phi_t^{\circ -1}]. \qquad (48)$$

It is clear that in terms of SNMSE, Algorithm 5.1 outperforms all other estimators; however, amongst different strategies, Algorithm 5.2 yields the smallest SNMSE. In fact the SNMSE for Algorithm 5.2 reflects the best per-

formance achievable for the choice of bank $\mathcal{B}$. Note that the SNMSE for Algorithm 5.2 is about 13% less than that obtained with Algorithm 5.1, but then, computationally, Algorithm 5.2 is 1.5 times slower to implement.

The simulation suggests that, at least in theory, it is possible to improve the tracking performance of estimators at lower altitude. This can be achieved by either choosing new estimators in the bank or by carefully redesigning the existing ones; however, note that redesigning estimators with specific properties (e.g., low estimation bias) require a thorough understanding of the underlying approximations.

8. Conclusions

A PCRLB inequality-based tool for performance assessment of multiple Bayesian estimators is developed. Using the PCRLB-based metric, an approach to compute the MMSE point estimates in the the average and optimal sense is developed for adaptive state estimation in general nonlinear SSMs. Compared to existing estimators, the switching method achieves a higher performance by optimally combining a given set of sub-optimal estimators. The efficacy of the tool was illustrated on a ballistic target tracking problem at re-entry phase.

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council (NSERC), Canada and the Shri Gopal Rajgarhia International Research Scholar Support Program, IIT Kharagpur, India.

References

Agniel, R., Jury, E., 1971. Almost sure boundedness of randomly sampled systems. SIAM Journal on Control 9 (3), 372–384.

Andrieu, C., Doucet, A., Singh, S., Tadic, V., 2004. Particle methods for change detection, system identification, and control. Proceedings of the IEEE 92 (3), 423–438.

Arulampalam, M., Maskell, S., Gordon, N., Clapp, T., 2002. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. IEEE Transactions on Signal Processing 50 (2), 174–188.

Bergman, N., 2001. Sequential Monte Carlo Methods in Practice. Springer–Verlag, New York, Ch. Posterior Cramér-Rao Bounds for Sequential Estimation.

Chen, T., Morris, J., Martin, E., 2005. Particle filters for state and parameter estimation in batch processes. Journal of Process Control 15 (6), 665–673.

Ching, J., Beck, J., Porter, K., 2006. Bayesian state and parameter estimation of uncertain dynamical systems. Probabilistic engineering mechanics 21 (1), 81–96.

Ding, F., 2014. State filtering and parameter estimation for state space systems with scarce measurements. Signal Processing 104, 369–380.

Farina, A., Ristic, B., Benvenuti, D., 2002. Tracking a ballistic target: comparison of several nonlinear filters. IEEE Transactions on Aerospace and Electronic Systems 38 (3), 854–867.

He, H., Xiong, R., Zhang, X., Sun, F., Fan, J., 2011. State-of-charge estimation of the lithium-ion battery using an adaptive extended kalman filter based on an improved thevenin model. IEEE Transactions on Vehicular Technology 60 (4), 1461–1469.

Kantas, N., Doucet, A., Singh, S., Maciejowski, J., 2009. An overview of sequential Monte-Carlo methods for parameter estimation in general state-space models. In: Proceedings of 15th IFAC Symposium on System Identification. Saint-Malo, France, pp. 774–785.

Kravaris, C., Hahn, J., Chu, Y., 2013. Advances and selected recent developments in state and parameter estimation. Computers & chemical engineering 51, 111–123.

Li, P., Goodall, R., Kadirkamanathan, V., 2004. Estimation of parameters in a linear state space model using a Rao-Blackwellised particle filter. IEE Proceedings of Control Theory and Applications 151 (6), 727–738.

Li, X., Jilkov, V., 2001. Survey of maneuvering target tracking: Ii. ballistic target models. In: International Symposium on Optical Science and Technology. International Society for Optics and Photonics, pp. 559–581.

Ljung, L., 1999. System identification. John Wiley & Sons, Inc., New York.

Minvielle, P., Doucet, A., Marrs, A., Maskell, S., 2010. A Bayesian approach to joint tracking and identification of geometric shapes in video sequences. Image and Vision Computing 28 (1), 111–123.

Reif, K., Gunther, S., Yaz, E., Unbehauen, R., 1999. Stochastic stability of the discrete-time extended Kalman filter. IEEE Transactions on Automatic Control 44 (4), 714–728.

Ristic, B., Farina, A., Benvenuti, D., Arulampalam, M., 2003. Performance bounds and comparison of nonlinear filters for tracking a ballistic object on re-entry. IEE Proceedings-Radar, Sonar and Navigation 150 (2), 65–70.

Su, W. W., Li, J., Xu, N.-S., 2003. State and parameter estimation of microalgal photobioreactor cultures based on local irradiance measurement. Journal of biotechnology 105 (1), 165–178.

Tarn, T., Rasis, Y., 1976. Observers for non-linear stochastic systems. IEEE Transactions on Automatic Control 21 (4), 441–448.

Tichavský, P., Muravchik, C., Nehorai, A., 1998. Posterior Cramér-Rao bounds for discrete-time nonlinear filtering. IEEE Transactions on Signal Processing 46 (5), 1386–1396.

Trees, H., 1968. Detection, Estimation and Modulation Theory–Part I. Wiley, New York, Ch. Classical detection and estimation theory.

Tulsyan, A., Forbes, J., Huang, B., 2012. Designing priors for robust Bayesian optimal experimental design. Journal of Process Control 22 (2), 450–462.

Tulsyan, A., Huang, B., Gopaluni, R., Forbes, J., 2013a. A particle filter approach to approximate posterior Cramér-Rao lower bound: The case of hidden states. IEEE Transactions on Aerospace and Electronic Systems, In Press 49 (4).

Tulsyan, A., Huang, B., Gopaluni, R., Forbes, J., 2013b. Bayesian identification of non-linear state space models: Part II- Error analysis. In: Proceedings of 10th IFAC International Symposium on Dynamics and Control of Process Systems. Mumbai, India.

Tulsyan, A., Huang, B., Gopaluni, R., Forbes, J., 2013c. On simultaneous on-line state and parameter estimation in non-linear state-space models. Journal of Process Control 23 (4), 516–526.

Tulsyan, A., Huang, B., Gopaluni, R., Forbes, J., 2013d. Performance assessment, diagnosis, and optimal selection of non-linear state filters. Journal of Process Control, In Review.

Šimandl, M., Královec, J., Tichavský, P., 2001. Filtering, predictive, and smoothing Cramér-Rao bounds for discrete-time nonlinear dynamic systems. Automatica 37 (11), 1703–1716.

Wenzel, T., Burnham, K., Blundell, M., Williams, R., 2006. Dual extended kalman filter for vehicle state and parameter estimation. Vehicle System Dynamics 44 (2), 153–171.