Bayesian identification of non-linear state-space models: Part I- Input design

Aditya Tulsyan *, **Swanand R. Khare** *, **Biao Huang** *,
R. Bhushan Gopaluni **, **J. Fraser Forbes** *

* *Department of Chemical and Materials Engineering, University of
Alberta, Edmonton AB T6G-2G6, Canada (e-mail: {tulsyan; khare;
biao.huang; fraser.forbes}@ ualberta.ca)*
** *Department of Chemical and Biological Engineering, University of
British Columbia, Vancouver BC V6T-1Z3, Canada, and Department
of Chemical Engineering, Massachusetts Institute of Technology,
Cambridge MA 02139, USA (e-mail: gopaluni@chbe.ubc.ca)*

Abstract: We propose an algorithm for designing optimal inputs for on-line Bayesian identification of stochastic non-linear state-space models. The proposed method relies on minimization of the posterior Cramér Rao lower bound derived for the model parameters, with respect to the input sequence. To render the optimization problem computationally tractable, the inputs are parametrized as a multi-dimensional Markov chain in the input space. The proposed approach is illustrated through a simulation example.

1. INTRODUCTION

Over the last decade, great progress has been made within the statistics community in overcoming the computational issues, and making Bayesian identification tractable for a wide range of complicated models arising in demographic and population studies, image processing, and drug response modelling (Gilks et al. [1995]). A detailed exposition of Bayesian identification methods can be found in Kantas et al. [2009]. This paper is directed towards the class of on-line methods for Bayesian identification of stochastic non-linear SSMs, the procedure for which is briefly introduced here first. 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, and let $\{u_t\}_{t\in\mathbb{N}}$ be the sequence of inputs in $\mathbb{R}^p$, such that the state $\{X_t\}_{t\in\mathbb{N}}$ is an unobserved or unmeasured process, with initial density $p_\theta(x)$ and transition density $p_\theta(x'|x,u)$:

$$X_0 \sim p_\theta(x_0); X_{t+1}|(x_t, u_t) \sim p_\theta(x_{t+1}|x_t, u_t) \quad (t\in\mathbb{N}). \quad (1)$$

$\{X_t\}_{t\in\mathbb{N}}$ is an unobserved process, but is observed through $\{Y_t\}_{t\in\mathbb{N}}$, such that $\{Y_t\}_{t\in\mathbb{N}}$ is conditionally independent given $\{X_t, u_t\}_{t\in\mathbb{N}}$, with marginal density $p_\theta(y|x,u)$:

$$Y_t|(x_t, u_t) \sim p_\theta(y_t|x_t, u_t) \quad (t\in\mathbb{N}). \quad (2)$$

θ in (1) and (2) is a vector of unknown model parameters, such that $\theta\in\Theta$ is an open subset of $\mathbb{R}^q$. All the densities are with respect to suitable dominating measures, such as Lebesgue measure. Although (1) and (2) represent a wide class of non-linear time-series models, the model form and the assumptions considered in this paper are given below

$$X_{t+1} = f_t(X_t, u_t, \theta_t, V_t); \quad Y_t = g_t(X_t, u_t, \theta_t, W_t), \quad (3)$$

where $\{\theta_{t+1} = \theta_t\}_{t\in\mathbb{N}} = \theta$ is a vector of static parameters.

Assumption 1. $\{V_t\}_{t\in\mathbb{N}}$ and $\{W_t\}_{t\in\mathbb{N}}$ are mutually independent sequences of independent random variables known a priori in their distribution classes (e.g., Gaussian) and parametrized by a known and finite number of moments.

Assumption 2. $\{f_t; g_t\}_{t\in\mathbb{N}}$ are such that in the open sets $\mathcal{X}$ and Θ, $\{f_t; g_t\}_{t\in\mathbb{N}}$ is $\mathcal{C}^k(\mathcal{X})$ and $\mathcal{C}^k(\Theta)$, respectively, and

in $\mathbb{R}^p$, $\{f_t; g_t\}_{t\in\mathbb{N}}$ is $\mathcal{C}^{k-1}(\mathbb{R}^p)$, and in $\mathbb{R}^n$ and $\mathbb{R}^m$, $\{f_t\}_{t\in\mathbb{N}}$ is $\mathcal{C}^{k-1}(\mathbb{R}^n)$, and $\{g_t\}_{t\in\mathbb{N}}$ is $\mathcal{C}^{k-1}(\mathbb{R}^m)$, where $k \geq 2$.

Assumption 3. For any random sample $(x_{t+1}, x_t, u_t, \theta_t, v_t) \in \mathcal{X} \times \mathcal{X} \times \mathbb{R}^p \times \Theta \times \mathbb{R}^n$ and $(y_t, x_t, u_t, \theta_t, w_t) \in \mathcal{Y} \times \mathcal{X} \times \mathbb{R}^p \times \Theta \times \mathbb{R}^m$ satisfying (3), $\nabla_{v_t} f_t^T(x_t, u_t, \theta_t, v_t)$ and $\nabla_{w_t} g_t^T(x_t, u_t, \theta_t, w_t)$ have rank n and m, respectively, such that using implicit function theorem, $p_\theta(x_{t+1}|x_t, u_t) = p(V_t = \tilde{f}_t(x_t, u_t, \theta_t, x_{t+1}))$ and $p_\theta(y_t|x_t, u_t) = p(W_t = \tilde{g}_t(x_t, u_t, \theta_t, y_t))$ do not involve any Dirac delta functions.

For a generic sequence $\{r_t\}_{t\in\mathbb{N}}$, let $r_{i:j} \triangleq \{r_i, r_{i+1}, \ldots, r_j\}$. Let $\theta^\star \in \Theta \subseteq \mathbb{R}^q$ be the true, but unknown parameter vector generating a measurement sequence $\{Y_{1:t} = y_{1:t}\}_{t\in\mathbb{N}}$ given $\{u_{1:t}\}_{t\in\mathbb{N}}$, such that $X_{t+1}|(x_t, u_t) \sim p_{\theta^\star}(x_{t+1}|x_t, u_t)$ and $Y_t|(x_t, u_t) \sim p_{\theta^\star}(y_t|x_t, u_t)$. In Bayesian identification of (3), the problem of estimating the parameter vector $\theta^\star \in \Theta \subseteq \mathbb{R}^q$ in real-time, given a sequence of input-output data $\{u_{1:t}, y_{1:t}\}_{t\in\mathbb{N}}$ is formulated as a joint state and parameter estimation problem. This is done by ascribing a prior density $\theta_0 \sim p(\theta_0)$, such that $\theta^\star \in \text{supp}\, p(\theta_0)$, and computing $\{p(z_t|u_{1:t}, y_{1:t})\}_{t\in\mathbb{N}}$, where: $Z_t \triangleq \{X_t;\, \theta_t\}$ is a $\mathcal{Z}(\subseteq \mathbb{R}^{s=n+q})$ valued extended Markov process with $Z_0 \sim p_{\theta_0}(x_0)p(\theta_0)$ and $Z_t|(z_{t-1}, u_{t-1}) \sim p_{\theta_{t-1}}(x_t|x_{t-1}, u_{t-1})\delta_{\theta_{t-1}}(\theta_t)$. The inference on $\{\theta_t\}_{t\in\mathbb{N}}$ then relies on the marginal posterior $\{p(\theta_t|u_{1:t}, y_{1:t})\}_{t\in\mathbb{N}}$. Note that by a judicious choice of the input sequence $\{u_{1:t}\}_{t\in\mathbb{N}}$, $\{p(z_t|u_{1:t}, y_{1:t})\}_{t\in\mathbb{N}}$ can be 'steered' in order to yield $\{p(\theta_t|u_{1:t}, y_{1:t})\}_{t\in\mathbb{N}}$, which gives more accurate inference on $\{\theta_t\}_{t\in\mathbb{N}}$. This is called the input design problem for Bayesian identification or simply, the Bayesian input design problem. A detailed review on this subject can be found in Chaloner and Verdinelli [1995].

Bayesian input design for linear and non-linear regression models is an active area of research (see Huan and Marzouk [2012], Kück et al. [2006], Müller and Parmigiani [1995] and references cited therein); however, its extension to SSMs has been limited. Recently, Bayesian input design

procedure for non-linear SSMs, where $\{X_t\}_{t \in \mathbb{N}}$ is completely observed was developed by Tulsyan et al. [2012]. Despite the success with regression models, to the best of authors' knowledge, no known Bayesian input design methods are available for identification of stochastic non-linear SSMs. This is due to the unobserved state process $\{X_t\}_{t \in \mathbb{N}}$, which makes the design problem difficult to solve.

This paper deals with the Bayesian input design for identification of stochastic SSMs given in (3). The proposed method is based on minimization of the posterior Cramér-Rao lower bound (PCRLB), derived by Tichavský et al. [1998]. First, we use Monte-Carlo (MC) methods to obtain an approximation of the PCRLB, and then parametrize the inputs as a multi-dimensional Markov chain in $\mathbb{R}^p$, to render the optimization problem computationally tractable. Markov-chain parametrization not only allows to include amplitude constraints on the input, it can be easily implemented using a standard PID controller or any other regulator. The notation used here is given next.

Notation: $\mathbb{N} := \{1, 2, \dots\}$; $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$; $\mathbb{R}^{s \times s}$ is the set of real-valued $s \times s$ matrices of cardinality $\text{Card}(\mathbb{R}^{s \times s})$; $\mathcal{S}^s \subset \mathbb{R}^{s \times s}$ is the space of symmetric matrices; $\mathcal{S}^s_+$ is the cone of symmetric positive semi-definite matrices in $\mathcal{S}^s$; and $\mathcal{S}^s_{++}$ is its interior. The partial order on $\mathcal{S}^s$ induced by $\mathcal{S}^s_+$ and $\mathcal{S}^s_{++}$ are denoted by $\succeq$ and $\succ$, respectively. $\mathbb{F}^{s \times s} \subset \mathbb{R}^{s \times s}$ is the set of $s \times s$ stochastic matrix, where $\mathbb{F} := [0, 1]$ and the sum of each row adds up to 1. For $A \in \mathbb{R}^{s \times s}$, $\text{Tr}[A]$ denotes its trace. For vectors $x \in \mathbb{R}^p$, $y \in \mathbb{R}^p$, and $z \in \mathbb{R}^p$, $x \leq y \leq z$ denotes element-wise inequality, and $\text{diag}(y) \in \mathcal{S}^p$ is a $p \times p$ diagonal matrix with elements of $y \in \mathbb{R}^p$ as its diagonal entries. Finally, $\Delta^y_x \triangleq \nabla_x \nabla^T_y$ is a Laplacian and $\nabla_x \triangleq \left[\frac{\partial}{\partial x}\right]$ is a gradient.

2. PROBLEM FORMULATION

Bayesian input design for regression models is a well studied problem in statistics (Chaloner and Verdinelli [1995]); wherein, the problem is often formulated as follows

$$\psi(u^\star_{1:N}) = \max_{u_{1:N} \in \mathbb{R}^{pN}} \sum_{t=1}^{N} \mathbb{E}_{p(\theta_t, y_{1:t}|u_{1:t})}[\underline{\psi}(Y_{1:t}, u_{1:t}, \theta_t)] \quad (4)$$

where $\{u^\star_{1:N}\}_{N \in \mathbb{N}}$ is an N-step ahead optimal input sequence, and $\underline{\psi}(\cdot)$ is a utility function. When inference on $\{\theta_t\}_{t \in \mathbb{N}}$ is of interest, Lindley [1956] suggested using the mean-square error (MSE) as a utility function, such that

$$\psi(u^\star_{1:N}) = \max_{u_{1:N} \in \mathbb{R}^{pN}} \sum_{t=1}^{N} -\Phi(P^\theta_{t|t}(u_{1:t})), \quad (5)$$

where $P^\theta_{t|t}(u_{1:t}) = \mathbb{E}_{p(\theta_t, y_{1:t}|u_{1:t})}[(\theta_t - \theta_{t|t})(\theta_t - \theta_{t|t})^T]$ is the MSE associated with the parameter estimate given by $\theta_{t|t} = \mathbb{E}_{p(\theta_t|u_{1:t}, y_{1:t})}[\theta_t]$, and $\Phi : \mathcal{S}^q_{++} \to \mathbb{R}$ is a test function.

Remark 4. For the model considered in (3), the marginal posterior density $\{p(\theta_t|u_{1:t}, y_{1:t})\}_{t \in \mathbb{N}}$, or the expectation with respect to it, does not admit any analytical solution, and thus, (5) cannot be computed in closed form. □

Remark 5. Methods such as SMC and MCMC can be used to approximate $\{p(\theta_t|u_{1:t}, y_{1:t})\}_{t \in \mathbb{N}}$; however, it makes the computation in (5) formidable (Kück et al. [2006]). Moreover, the input $\{u^\star_{1:N}\}_{N \in \mathbb{N}}$ is optimal only for the Bayesian estimator used to approximate $\{p(\theta_t|u_{1:t}, y_{1:t})\}_{t \in \mathbb{N}}$. □

To address the issues in Remarks 4 and 5, we propose to define a lower bound on the MSE first, and minimize the lower bound instead. The PCRLB, derived by Tichavský et al. [1998] provides a lower bound on the MSE associated with the estimation of $\{Z_t\}_{t \in \mathbb{N}}$ from $\{p(z_t|u_{1:t}, y_{1:t})\}_{t \in \mathbb{N}}$, and is given in the next lemma.

Lemma 6. Let $\{Y_{1:t} = y_{1:t}\}_{t \in \mathbb{N}}$ be an output sequence generated from (3) using $\{u_{1:t}\}_{t \in \mathbb{N}}$, then the MSE associated with the estimation of $\{Z_t\}_{t \in \mathbb{N}}$ from $\{p(z_t|u_{1:t}, y_{1:t})\}_{t \in \mathbb{N}}$ is bounded from below by the following matrix inequality

$$P^z_{t|t} \triangleq \mathbb{E}_{p(z_t, y_{1:t}|u_{1:t})}[(Z_t - Z_{t|t})(Z_t - Z_{t|t})^T] \succeq [J^z_t]^{-1}, \quad (6)$$

where: $Z_{t|t} = \mathbb{E}_{p(z_t|u_{1:t}, y_{1:t})}[Z_t]$ is an estimate of $\{Z_t\}_{t \in \mathbb{N}}$;

$$P^z_{t|t} \triangleq \begin{bmatrix} P^x_{t|t} & P^{x\theta}_{t|t} \\ (P^{x\theta}_{t|t})^T & P^\theta_{t|t} \end{bmatrix} \in \mathcal{S}^s_{++}, \quad J^z_t \triangleq \begin{bmatrix} J^x_t & J^{x\theta}_t \\ (J^{x\theta}_t)^T & J^\theta_t \end{bmatrix} \in \mathcal{S}^s_{++},$$

$$[J^z_t]^{-1} \triangleq \begin{bmatrix} L^x_t & L^{x\theta}_t \\ (L^{x\theta}_t)^T & L^\theta_t \end{bmatrix} \in \mathcal{S}^s_{++}$$ are the MSE, posterior information matrix (PIM), and PCRLB, respectively.

Proof. See Tichavský et al. [1998] for proof. □

Lemma 7. A recursive approach to compute $\{J^z_t\}_{t \in \mathbb{N}}$ for (3) under Assumptions 1 through 3 is given as follows

$$J^x_{t+1} = H^{33}_t - (H^{13}_t)^T[J^x_t + H^{11}_t]^{-1}H^{13}_t; \quad (7a)$$

$$J^{x\theta}_{t+1} = (H^{23}_t)^T - (H^{13}_t)^T[J^x_t + H^{11}_t]^{-1}(J^{x\theta}_t + H^{12}_t); \quad (7b)$$

$$J^\theta_{t+1} = J^\theta_t + H^{22}_t - (J^{x\theta}_t + H^{12}_t)^T[J^x_t + H^{11}_t]^{-1} \\ \times (J^{x\theta}_t + H^{12}_t), \quad (7c)$$

where:

$$H^{11}_t = \mathbb{E}_{\tilde{p}_{t+1}}[-\Delta^{X_t}_{X_t} \log p_t]; \quad (8a)$$

$$H^{12}_t = \mathbb{E}_{\tilde{p}_{t+1}}[-\Delta^{\theta_t}_{X_t} \log p_t]; \quad (8b)$$

$$H^{13}_t = \mathbb{E}_{\tilde{p}_{t+1}}[-\Delta^{X_{t+1}}_{X_t} \log p_t]; \quad (8c)$$

$$H^{22}_t = \mathbb{E}_{\tilde{p}_{t+1}}[-\Delta^{\theta_t}_{\theta_t} \log p_t]; \quad (8d)$$

$$H^{23}_t = \mathbb{E}_{\tilde{p}_{t+1}}[-\Delta^{X_{t+1}}_{\theta_t} \log p_t]; \quad (8e)$$

$$H^{33}_t = \mathbb{E}_{\tilde{p}_{t+1}}[-\Delta^{X_{t+1}}_{X_{t+1}} \log p_t]; \quad (8f)$$

$\tilde{p}_{t+1} = p(x_{0:t+1}, \theta_t, y_{1:t+1}|u_{1:t+1})$, and $p_t = p(X_{t+1}|Z_t, u_t) p(Y_{t+1}|\theta_t, X_{t+1}, u_{t+1})$; and $J_0 = \mathbb{E}_{p(z_0)}[-\Delta^{Z_0}_{Z_0} \log p(Z_0)]$.

Proof. See Tichavský et al. [1998] for proof. □

Corollary 8. Let $P^z_{t|t} \in \mathcal{S}^s_{++}$, $[J^z_t]^{-1} \in \mathcal{S}^s_{++}$ be such that they satisfy (6), then the MSE associated with the point estimation of $\{\theta_t\}_{t \in \mathbb{N}}$, computed from $\{p(\theta_t|u_{1:t}, y_{1:t})\}_{t \in \mathbb{N}}$, is bounded from below by the following matrix inequality

$$P^\theta_{t|t} = \mathbb{E}_{p(\theta_t, y_{1:t}|u_{1:t})}[(\theta_t - \theta_{t|t})(\theta_t - \theta_{t|t})^T] \succeq L^\theta_t, \quad (9)$$

where $L^\theta_t \in \mathcal{S}^q_{++}$ is the lower-right sub-matrix of $[J^z_t]^{-1} \in \mathcal{S}^s_{++}$ in (6).

Proof. The proof is based on the fact that the PCRLB inequality in (6) guarantees that $P^z_{t|t} - [J^z_t]^{-1} \in \mathcal{S}^s_+$. □

Theorem 9. Let $J^z_t \in \mathcal{S}^s_{++}$ be the PIM for model in (3) and $L^\theta_t \in \mathcal{S}^q_{++}$ be the lower bound on the MSE associated with the estimation of $\{\theta_t\}_{t \in \mathbb{N}}$ in (3), then given $J^z_t \in \mathcal{S}^s_{++}$, the lower bound $L^\theta_t \in \mathcal{S}^q_{++}$ at $t \in \mathbb{N}$ can be computed as

$$L^\theta_t = [J^\theta_t - (J^{x\theta}_t)^T(J^x_t)^{-1}J^{x\theta}_t]^{-1}, \quad (10)$$

where J^θ_t, $J^{x\theta}_t$ and J^x_t are the PIMs given in Lemma 7.

Proof. The proof is based on the matrix inversion lemma (see R.B. Bapat and T.E.S. Raghavan [1997]). □

Finally, the input design problem for Bayesian identification of $\{\theta_t\}_{t\in\mathbb{N}}$ in (3) can be formulated as follows

$$\psi(u_{1:N}^{\star}) = \min_{u_{1:N}\in\mathbb{R}^{pN}} \sum_{t=1}^{N} \Phi(L_t^{\theta}(u_{1:t})) \qquad (11a)$$

$$\text{s.t.} \quad u_{min} \leq \{u_i\}_{i\in[1,N]} \leq u_{max}, \qquad (11b)$$

where $L_t^{\theta}(u_{1:t}) \triangleq L_t^{\theta}$; and $u_{max} \in \mathbb{R}^p$ and $u_{min} \in \mathbb{R}^p$ are the maximum and minimum magnitude of the input.

Remark 10. The optimization problem in (11) allows to impose magnitude constraints on the inputs. Although constraints on $(x_{0:N}) \in \mathcal{X}^{N+1}$ and $(y_{1:N}) \in \mathcal{Y}^N$ are not included, but if required, they can also be appended. □

Remark 11. Integral in (8), with respect to $\tilde{p}_t$, makes (11) independent of the random realizations from $\mathcal{X}^{t+1}$, Θ, and $\mathcal{Y}^t$. The optimization in (11) in fact only depends on: the process dynamics represented in (3); noise densities $V_t \sim p(v_t)$ and $W_t \sim p(w_t)$; and the choice of $Z_0 \sim p(z_0)$ and $u_{1:N} \in \mathbb{R}^{pN}$. This makes (11) independent of $\theta^{\star} \in \Theta \subseteq \mathbb{R}^q$ or the Bayesian estimator used for estimating $\{\theta_t\}_{t\in\mathbb{N}}$. □

Remark 12. The formulation in (11) yields a sequence $\{u_{1:N}^{\star}\}_{N\in\mathbb{N}}$, which is (a) optimal for all the Bayesian identification methods that approximate $\{p(\theta_t|u_{1:t}^{\star}, y_{1:t})\}_{t\in\mathbb{N}}$; and (b) independent of $\theta^{\star} \in \mathbb{R}^q$ (see Remark 11), such that the input $\{u_{1:N}^{\star}\}_{N\in\mathbb{N}}$ is optimal for all $\theta^{\star} \in \text{supp } p(\theta_0)$. □

There are two challenges that need to be addressed in order to make the optimization problem in (11) tractable: (a) computing the lower bound $\{L_t^{\theta}\}_{t\in\mathbb{N}}$; and (b) solving the high-dimensional optimization problem in $\mathbb{R}^{pN}$. Our approach to address the above challenges is discussed next.

3. COMPUTING THE LOWER BOUND

The first challenge is to compute the lower bound L_t^{θ} in (11). It is well known that computing L_t^{θ} in closed form is non-trivial for the model form considered in (3) (see Tichavský et al. [1998], Bergman [2001]). This is because of the complex, high-dimensional integrals in (8a) through (8f), which do not admit any analytical solution.

MC sampling is a popular numerical method to solve integrals of the form $F(u_{1:t}) = \mathbb{E}_{p(x_{0:t}|u_{1:t})}[h(X_{0:t}, u_{1:t})]$, where $h : \mathcal{X}^{t+1} \times \mathbb{R}^{pt} \to \mathbb{R}$. Using M i.i.d. trajectories $\{X_{0:t}^i|u_{1:t}\}_{i=1}^M \sim p(x_{0:t}|u_{1:t})$, the probability distribution $p(x_{0:t}|u_{1:t})dx_{0:t} \triangleq p(dx_{0:t}|u_{1:t})$, can be approximated as

$$\tilde{p}(dx_{0:t}|u_{1:t}) = \frac{1}{M}\sum_{i=1}^{M}\delta_{X_{0:t}^i|u_{1:t}}(dx_{0:t}), \qquad (12)$$

where $\tilde{p}(dx)$ is a MC estimate of $p(dx)$ and $\delta_{x_0}(dx)$ is the Dirac delta mass at x_0. Finally, substituting (12) into $F(u_{1:t})$, we get $\tilde{F}(u_{1:t}) \triangleq \tilde{F}(\{X_{0:t}^i|u_{1:t}\}_{i=1}^M) = \int h(x_{0:t}, u_{1:t})\tilde{p}(dx_{0:t}|u_{1:t}) = \frac{1}{M}\sum_{i=1}^M h(X_{0:t}^i, u_{1:t})$, where $\tilde{F}(u_{1:t})$ is an M-sample MC estimate of $F(u_{1:t})$.

Remark 13. Using MC methods, the multi-dimensional integrals in (8a) through (8f), with respect to the density $\tilde{p}_t(\cdot)$ can be approximated by simulating M i.i.d. sample paths $\{X_{0:t}^i, \theta_t^i, Y_{1:t}^i|u_{1:t}\}_{i=1}^M \sim \tilde{p}_t(\cdot)$ using (3), starting at M i.i.d. initial positions drawn from $\{Z_0^i\}_{i=1}^M \sim p(z_0)$.

Example 14. Consider the following stochastic SSM with additive Gaussian state and measurement noise

$$X_{t+1} = f_t(X_t, \theta_t, u_t) + V_t, \qquad (13a)$$
$$Y_t = g_t(X_t, \theta_t, u_t) + W_t, \qquad (13b)$$

where $\{V_t\}_{t\in\mathbb{N}}$ and $\{W_t\}_{t\in\mathbb{N}}$ are mutually independent sequences of independent zero mean Gaussian random variables, such that $V_t \sim \mathcal{N}(0, Q_t)$ and $W_t \sim \mathcal{N}(0, R_t)$, where $Q_t < \infty$ and $R_t < \infty$ for all $t \in \mathbb{N}$.

Note that for the model form considered in Example 14, using the Markov property of the states and conditional independence of the measurements, the dimension of the integrals in (8a) through (8f) can be reduced, as given next.

Theorem 15. For a stochastic non-linear SSM in Example 14, the matrices in (8a) through (8f) can be written as

$$H_t^{11} = \mathbb{E}_{p(x_t,\theta_t|u_{1:t+1})}[\nabla_{X_t}f_t^T(X_t,\theta_t,u_t)]Q_t^{-1}$$
$$\times [\nabla_{X_t}f_t^T(X_t,\theta_t,u_t)]^T; \qquad (14a)$$

$$H_t^{12} = \mathbb{E}_{p(x_t,\theta_t|u_{1:t+1})}[\nabla_{X_t}f_t^T(X_t,\theta_t,u_t)]Q_t^{-1}$$
$$\times [\nabla_{\theta_t}f_t^T(X_t,\theta_t,u_t)]^T; \qquad (14b)$$

$$H_t^{13} = -\mathbb{E}_{p(x_t,\theta_t|u_{1:t+1})}[\nabla_{X_t}f_t^T(X_t,\theta_t,u_t)]Q_t^{-1}; \qquad (14c)$$

$$H_t^{22} = \mathbb{E}_{p(x_t,\theta_t|u_{1:t+1})}[\nabla_{\theta_t}f_t^T(X_t,\theta_t,u_t)]Q_t^{-1}$$
$$\times [\nabla_{\theta_t}f_t^T(X_t,\theta_t,u_t)]^T$$
$$+ \mathbb{E}_{p(x_{t+1},\theta_t|u_{1:t+1})}[\nabla_{\theta_t}g_t^T(X_{t+1},\theta_t,u_{t+1})]R_{t+1}^{-1}$$
$$\times [\nabla_{\theta_t}g_t^T(X_{t+1},\theta_t,u_{t+1})]^T \qquad (14d)$$

$$H_t^{23} = -\mathbb{E}_{p(x_t,\theta_t|u_{1:t+1})}[\nabla_{\theta_t}f_t^T(X_t,\theta_t,u_t)]Q_t^{-1}$$
$$+ \mathbb{E}_{p(x_{t+1},\theta_t|u_{1:t+1})}[\nabla_{\theta_t}g_t^T(X_{t+1},\theta_t,u_{t+1})]R_{t+1}^{-1}$$
$$\times [\nabla_{X_{t+1}}g_t^T(X_{t+1},\theta_t,u_{t+1})]^T \qquad (14e)$$

$$H_t^{33} = Q_t^{-1} + \mathbb{E}_{p(x_{t+1},\theta_t|u_{1:t+1})}[\nabla_{X_{t+1}}g_t^T(X_{t+1},\theta_t,u_{t+1})]$$
$$\times R_{t+1}^{-1}[\nabla_{X_{t+1}}g_t^T(X_{t+1},\theta_t,u_{t+1})]^T \qquad (14f)$$

Proof. (14a): First note that $H_t^{11} = \mathbb{E}_{\tilde{p}_{t+1}}[-\Delta_{X_t}^{X_t}\log p_t] = \mathbb{E}_{\tilde{p}_{t+1}}[\nabla_{X_t}\log p_t][\nabla_{X_t}\log p_t]^T$ (see Tichavský et al. [1998]). On simplifying, we have $H_t^{11} = \mathbb{E}_{p(x_{t+1}|x_t,\theta_t,u_t)p(x_t,\theta_t|u_{1:t+1})}$ $[\nabla_{X_t}\log p(X_{t+1}|X_t,\theta_t,u_t)][\nabla_{X_t}\log p(X_{t+1}|X_t,\theta_t,u_t)]^T$. This is due to $\nabla_{X_t}\log p(Y_{t+1}|X_{t+1},\theta_t,u_{t+1}) = 0$. For Example 14, $\nabla_{X_t}\log p(X_{t+1}|X_t,\theta_t,u_t) = [\nabla_{X_t}f_t^T(X_t,\theta_t,u_t)]$ $Q_t^{-1}[X_{t+1} - f_t(X_t,\theta_t,u_t)]^T$. Substituting it into H_t^{11}, and using $\mathbb{E}_{p(x_{t+1}|x_t,\theta_t,u_t)}[X_{t+1} - f_t(X_t,\theta_t,u_t)][X_{t+1} - f_t(X_t,\theta_t,u_t)]^T = Q_t$, we have (14a). Note that the expression in (14b) through (14f) can be similarly derived. □

Theorem 15 reduces the dimension of the integral in (8) for Example 14 from $(t+1)(n+m) + s$ to s. Using MC sampling, (14a), for instance, can be computed as $\tilde{H}_t^{11} = \frac{1}{M}\sum_{i=1}^M[\nabla_{X_t}f^T(X_t^i,\theta_t^i,u_t)]Q_t^{-1}[\nabla_{X_t}f^T(X_t^i,\theta_t^i,u_t)]^T$. Here $\{X_t^i,\theta_t^i|u_{1:t+1}\}_{i=1}^M \sim p(x_t,\theta_t|u_{1:t+1})$ and $\tilde{H}_t^{11}$ is an M-sample MC estimate of H_t^{11}. Note that the MC estimates of (14b) through (14f) can be similarly computed. In general, substituting the MC estimates of (8a) through (8f) first into Lemma 7, and then into Theorem 9, yields

$$\tilde{L}_t^{\theta} = [\tilde{J}_t^{\theta} - (\tilde{J}_t^{x\theta})^T(\tilde{J}_t^x)^{-1}\tilde{J}_t^{x\theta}]^{-1}, \qquad (15)$$

where $\tilde{L}_t^{\theta}$ is an estimate of L_t^{θ}, and $\tilde{J}_t^{\theta}$, $\tilde{J}_t^{x\theta}$ and $\tilde{J}_t^x$ are the estimates of the PIMs in Lemma 7. Finally, substituting (15) into (11) gives the following optimization problem

$$\tilde{\psi}(u^\star_{1:N}) = \min_{u_{1:N} \in \mathbb{R}^{pN}} \sum_{t=1}^{N} \Phi(\tilde{L}^\theta_t(u_{1:t})) \tag{16a}$$

$$\text{s.t.} \quad u_{min} \le \{u_i\}_{t \in [1,N]} \le u_{max}. \tag{16b}$$

Theorem 16. Let $\psi(u^\star_{1:N})$ and $\tilde{\psi}(u^\star_{1:N})$ be the optimal utility functions, computed by solving the optimization problem in (11) and (16), respectively, then we have

$$\tilde{\psi}(u^\star_{1:N}) \xrightarrow[M \to +\infty]{a.s.} \psi(u^\star_{1:N}), \tag{17}$$

where $\xrightarrow{a.s.}$ denotes almost sure convergence.

Proof. Since (15) is based on perfect MC sampling, using the strong law of large numbers, we have $\tilde{L}^\theta_t \xrightarrow{a.s.} L^\theta_t$ as $M \to +\infty$. Equation (17) follows from this result, which completes the proof. $\square$

A natural approach to solve (16) is to treat $\{u_{1:N}\}_{N \in \mathbb{N}}$ as a vector of continuous variables in $\mathbb{R}^{pN}$; however, this will render (16) computationally inefficient for large $N \in \mathbb{N}$. A relaxation method to make (16) tractable is given next.

4. INPUT PARAMETRIZATION

To overcome the complications due to continuous valued input $\{u_t\}_{t \in \mathbb{N}} \in \mathbb{R}^p$, we discretize the input space from $\mathbb{R}^p$ to $\mathcal{U} \subseteq \mathbb{R}^p$, such that $\text{Card}(\mathcal{U}) = r$, where $r = b^p$, and $b \in \mathbb{N}$ is the number of discrete values for each input in $\mathbb{R}$. If we denote $\mathcal{U} = \{s_1, \ldots, s_r\}$, then $u_{min} \le s_i \le u_{max}$, for all $1 \le i \le r$, such that (16) can be written as follows

$$\tilde{\psi}(u^\star_{1:N}) = \min_{u_{1:N} \in \mathcal{U}^N} \sum_{t=1}^{N} \Phi(\tilde{L}^\theta_t(u_{1:t})). \tag{18}$$

Note that although the input $\{u_{1:N}\}_{N \in \mathbb{N}}$ in (18) is defined on a discrete input space $\mathcal{U}^N$ of $\text{Card}(\mathcal{U}^N) = r^N$, (18) is still intractable for large $N \in \mathbb{N}$. To address this issue, a multi-dimensional Markov chain input parametrization, first proposed by Brighenti et al. [2009], is used here.

Definition 17. For $k \in \mathbb{N}_0$ and $\mathbb{S} := \{k+1, k+2, \cdots\}$, let $\{U_t\}_{t \in \mathbb{S}} = \{u_{t-k:t}\}_{k \in \mathbb{N}_0}$ be a $\mathcal{U}^{k+1}$ valued first-order finite Markov chain, where $\text{Card}(\mathcal{U}^{k+1}) = r^{k+1}$, such that the sample values of $\{U_t\}_{k \in \mathbb{N}_0, t \in \mathbb{S} \setminus \{k+1\}} \in \mathcal{U}^{k+1}$, depend on the past only through the sample values of $\{U_{t-1}\}_{t-1 \in \mathbb{S}} \in \mathcal{U}^{k+1}$, such that for all $\{U_t\}_{k \in \mathbb{N}_0, t \in \mathbb{S} \setminus \{k+1\}} \in \mathcal{U}^{k+1}$ and $\{U_{k+1:t-1}\}_{k \in \mathbb{N}_0, t-1 \in \mathbb{S}} \in \mathcal{U}^{t-1}$, we have the following

$$\Pr(U_t = \{u_{t-k:t}\}|U_{k:t-1} = \{u_{1:t-1}\}) =$$
$$P_\Pi(U_t = \{u_{t-k:t}\}|U_{t-1} = \{u_{t-k-1:t-1}\}), \tag{19}$$

where $\Pr(\cdot)$ is a probability measure and $P_\Pi \in \mathbb{F}^{r^{k+1} \times r^{k+1}}$ is a $r^{k+1} \times r^{k+1}$ probability transition matrix. $\square$

In Definition 17, $P_\Pi(U_t = s_2)|U_{t-1} = s_1)$, where $\{s_1, s_2\} \in \mathcal{U}^{k+1}$ represents the probability that the Markov chain transits from $\{U_{t-1}\}_{k \in \mathbb{N}_0, t-1 \in \mathbb{S}} = s_1$ to the input state $\{U_t\}_{k \in \mathbb{N}_0, t \in \mathbb{S} \setminus \{k+1\}} = s_2$. Consider the following example.

Example 18. For $p = 1$, $k = 0$, and $b \in \mathbb{N}$, we have $r = b$ and $\mathbb{S} = \mathbb{N}$, such that $\{U_t\}_{t \in \mathbb{S}} = \{u_t\}$ is a Markov chain on the input space $\mathcal{U} = \{s_1, s_2, \ldots, s_b\}$ of $\text{Card}(\mathcal{U}) = b$, then the probability matrix $P_\Pi \in \mathbb{F}^{b \times b}$ can be represented as

$$P_\Pi = \begin{bmatrix} p_{s_1,s_1} & p_{s_1,s_2} & \cdots & p_{s_1,s_b} \\ p_{s_2,s_1} & p_{s_2,s_2} & \cdots & p_{s_2,s_b} \\ \vdots & \vdots & & \vdots \\ p_{s_b,s_1} & p_{s_b,s_2} & \cdots & p_{s_b,s_b} \end{bmatrix},$$

where $p_{s_i,s_j} \triangleq P_\Pi(U_t = s_j)|U_{t-1} = s_i) \; \forall 1 \le i, j \le b$.

Example 19. For $p = 1$, $k = 1$, and $b \in \mathbb{N}$, we have $r = b$ and $\mathbb{S} = \mathbb{N} \setminus \{1\}$, such that $\{U_t\}_{t \in \mathbb{S}} = \{u_{t-1:t}\}$ is a Markov chain on $\mathcal{U}^2 = \{\{s_1, s_1\}, \{s_1, s_2\}, \ldots, \{s_2, s_1\}, \ldots, \{s_b, s_b\}\}$ of $\text{Card}(\mathcal{U}^2) = b^2$ then $P_\Pi \in \mathbb{F}^{b^2 \times b^2}$ can be represented as

$$P_\Pi = \begin{bmatrix} p_{\{s_1,s_1\},\{s_1,s_1\}} & p_{\{s_1,s_1\},\{s_1,s_2\}} & \cdots & p_{\{s_1,s_1\},\{s_b,s_b\}} \\ p_{\{s_1,s_2\},\{s_1,s_1\}} & p_{\{s_1,s_2\},\{s_1,s_2\}} & \cdots & p_{\{s_1,s_2\},\{s_b,s_b\}} \\ \vdots & \vdots & & \vdots \\ p_{\{s_1,s_g\},\{s_1,s_1\}} & p_{\{s_1,s_g\},\{s_1,s_2\}} & \cdots & p_{\{s_1,s_g\},\{s_g,s_g\}} \\ p_{\{s_2,s_1\},\{s_1,s_1\}} & p_{\{s_2,s_1\},\{s_1,s_2\}} & \cdots & p_{\{s_2,s_1\},\{s_g,s_g\}} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ p_{\{s_g,s_g\},\{s_1,s_1\}} & p_{\{s_g,s_g\},\{s_1,s_2\}} & \cdots & p_{\{s_g,s_g\},\{s_g,s_g\}} \end{bmatrix}.$$

where $p_{\{s_i,s_j\},\{s_l,s_m\}} \triangleq P_\Pi(U_t = \{s_i, s_j\})|U_{t-1} = \{s_l, s_m\})$ $\forall 1 \le i, j, l, m \le b$.

Assumption 20. The Markov chain $\{U_t\}_{t \in \mathbb{S}} = \{u_{t-k:t}\}_{k \in \mathbb{N}_0}$ considered in Definition 17 is time-homogeneous.

Assumption 21. The Markov chain $\{U_t\}_{t \in \mathbb{S}} = \{u_{t-k:t}\}_{k \in \mathbb{N}_0}$ in Definition 17 has a prior probability distribution $U_{k+1} \sim P_\Gamma(\{u_{1:k+1}\})$, where P_Γ is a $1 \times r^{k+1}$ vector.

Theorem 22. For $k \in \mathbb{N}_0$ and $\mathbb{S} := \{k+1, k+2, \cdots\}$, let $\{U_t\}_{t \in \mathbb{S}} = \{u_{t-k:t}\}_{k \in \mathbb{N}_0}$ be a Markov chain defined in Definition 17, and satisfying Assumptions 20 and 21, such that $U_t|(\{u_{t-k-1:t-1}\}) \sim P_\Pi(\{u_{1:k+1}\}|\{u_{t-k-1:t-1}\})$ for all $t \in \mathbb{S} \setminus \{k+1\}$ and $U_{k+1} \sim P_\Gamma(\{u_{1:k+1}\})$ then $\{U_{k+1:N}\}_{N \in \mathbb{N}} \sim P^{k+1:N}_{\Gamma,\Pi}$ has a probability distribution

$$P_\Gamma(\{u_{1:k+1}\}) \prod_{t=k+2}^{N} P_\Pi(\{u_{t-k:t}\}|\{u_{t-k-1:t-1}\}). \tag{20}$$

Proof. Using probability chain rule, the joint probability distribution of $U_{k+1:N} \sim P^{k+1:N}_{\Gamma,\Pi}$ can be written as

$$P^{k+1:N}_{\Gamma,\Pi} = \Pr(\{u_{1:k+1}\}, \{u_{2:k+2}\}, \ldots, \{u_{N-k:N}\})$$
$$= \Pr(\{u_{N-k:N}\}|\{u_{1:k+1}\}, \{u_{2:k+2}\}, \ldots, \{u_{N-k-1:N-1}\})$$
$$\times \Pr(\{u_{1:k+1}\}, \{u_{2:k+2}\}, \ldots, \{u_{N-k-1:N-1}\}), \tag{21a}$$
$$= P_{\Pi_{k,r}}(\{u_{N-k:N}\}|\{u_{N-k-1:N-1}\})$$
$$\times \Pr(\{u_{1:k+1}\}, \{u_{2:k+2}\}, \ldots, \{u_{N-k-1:N-1}\}), \tag{21b}$$

where in (21b), we have used the first-order Markov property of $\{U_t\}_{t \in \mathbb{S}}$. Noting the time-homogeneous property of $\{U_t\}_{t \in \mathbb{S}}$ and repeatedly appealing to the probability chain rule in (21b), we get (20). This completes the proof. $\square$

Remark 23. From Theorem 22, it is clear that: (i) the sample values of the random variables $\{U_{k+1:N}\}_{k \in \mathbb{N}_0, N \in \mathbb{N}}$ is an ordered sequence constructed from $\{u_{1:N}\}_{N \in \mathbb{N}}$; (ii) the probability distribution of the sequence $\{U_{k+1:N}\}_{k \in \mathbb{N}_0, N \in \mathbb{N}}$ given in (20) is uniquely defined by P_Π and P_Γ.

Using Definition 17 and Theorem 22, (18) can be reformulated to the following stochastic programming problem

$$\tilde{\psi}(U^\star_{k+1:N}) = \arg\min_{P_\Pi, P_\Gamma} \left\{ \sum_{t=1}^{k+1} \Phi(\mathbb{E}_{P_\Gamma}[\tilde{L}^\theta_t(\{U_{k+1}\})]) + \right.$$

$$\left. \sum_{t=k+2}^{N} \Phi(\mathbb{E}_{P^{k+1:t}_{\Gamma,\Pi}}[\tilde{L}^\theta_t(U_{k+1:t})]) \right\} \tag{22a}$$

$$\text{s.t.} \quad 0 \le P_\Pi(s_i|s_j) \le 1 \quad \forall \; 1 \le i, j \le r^{k+1}, \tag{22b}$$

$$\sum_{i=1}^{r^{k+1}} P_\Pi(s_i|s_j) = 1 \qquad \forall\, 1 \le j \le r^{k+1}, \qquad (22c)$$

$$0 \le P_\Gamma(s_i) \le 1 \qquad \forall\, 1 \le i \le r^{k+1}, \qquad (22d)$$

$$\sum_{i=1}^{r^{k+1}} P_\Gamma(s_i) = 1. \qquad (22e)$$

The expectations in (22a), with respect to P_Γ and $P_{\Gamma,\Pi}^{k+1:t}$ can again be approximated using MC sampling, such that

$$\tilde{P}_{\Gamma,\Pi}^{k+1:t} = \frac{1}{M_u} \sum_{i=1}^{M_u} \delta_{U_{k+1:t}^i}(U_{k+1:t}) \qquad (23)$$

where $\tilde{P}_{\Gamma,\Pi}^{k+1:t}$ is the M_u-sample MC estimate. Note that marginalizing (23) with respect to $\{U_{k+2:N}\}_{k\in\mathbb{N}_0,N\in\mathbb{N}}$ yields $\tilde{P}_\Gamma = \frac{1}{M_u}\sum_{i=1}^{M_u}\delta_{U_{k+1}^i}(U_{k+1})$, where $\tilde{P}_\Gamma$ is a MC estimate of P_Γ. Substituting $\tilde{P}_{\Gamma,\Pi}^{k+1:t}$ and $\tilde{P}_\Gamma$ into (22a) yields

$$\overline{\psi}(U_{k+1:N}^\star) = \arg\min_{P_\Pi, P_\Gamma} \frac{1}{M_u} \left\{ \sum_{t=1}^{k+1} \Phi\left(\sum_{i=1}^{M_u} \tilde{L}_t^\theta(U_{k+1}^i)\right) + \right.$$
$$\left. \sum_{t=k+2}^{N} \Phi\left(\sum_{i=1}^{M_u} \tilde{L}_t^\theta(U_{k+1:t}^i)\right) \right\} \qquad (24a)$$

$$\text{s.t.} \quad 0 \le P_\Pi(s_i|s_j) \le 1 \quad \forall\, 1 \le i,j \le r^{k+1}, \qquad (24b)$$

$$\sum_{i=1}^{r^{k+1}} P_\Pi(s_i|s_j) = 1 \qquad \forall\, 1 \le j \le r^{k+1}, \qquad (24c)$$

$$0 \le P_\Gamma(s_i) \le 1 \qquad \forall\, 1 \le i \le r^{k+1}, \qquad (24d)$$

$$\sum_{i=1}^{r^{k+1}} P_\Gamma(s_i) = 1. \qquad (24e)$$

Note that solving (24), yields $U_{k+1:N}^\star \sim P_{\Gamma^\star,\Pi^\star}^{k+1:N}$, which is the optimal distribution of the input sequence.

Corollary 24. Let $\overline{\psi}(U_{k+1:N}^\star)$ and $\tilde{\psi}(U_{k+1:N}^\star)$ be the optimal utility functions, computed by solving the optimization problem in (22a) and (24a), respectively, then

$$\overline{\psi}(U_{k+1:N}^\star) \xrightarrow[M_u \to +\infty]{a.s.} \tilde{\psi}(U_{k+1:N}^\star), \qquad (25)$$

where $\xrightarrow{a.s.}$ denotes almost sure convergence.

Proof. Proof is similar to Theorem 16. □

Remark 25. There are several advantages of using the formulation given in (24): (a) the optimization is independent of $N \in \mathbb{N}$, as the number of parameters to be estimated are $r^{k+1}(1 + r^{k+1})$; (b) easy to include magnitude and other transition constraints on the inputs; and (c) samples from the optimal distribution can be easily sampled, and implemented using a PID or any classical regulator. □

In this paper, the optimization problem in (24) is implemented through an iterative approach, that involves standard numerical solvers (Nocedal and Wright [2006]). The proposed method for input design, including the iterations in the optimization, is summarized in Algorithm 1.

5. SIMULATION EXAMPLE

Consider a process described by the following univariate, and non-stationary stochastic SSM (Tulsyan et al. [2013b])

Algorithm 1 Bayesian input design for identification

1: Choose an initial value for the input design parameters $P_\Gamma = P_\Gamma^{(0)}$ and $P_\Pi = P_\Pi^{(0)}$. Set $c \leftarrow 0$.
2: **while** converged **do**
3:   **for** $i = 1$ to M_u **do**
4:     Generate a random input sequence $U_{k+1:N}^i \sim P_{\Gamma,\Pi}^{k+1:N}$ using the distribution given in (20).
5:     Generate M random samples of states and parameters from the prior density $\{Z_0^j\}_{j=1}^M \sim p(z_0)$.
6:     **for** $t = 1$ to N **do**
7:       Generate M random samples of the process states $\{X_t^j|(z_{t-1}^j, u_{t-1}^i)\}_{j=1}^M \sim p(x_t|z_{t-1}^j, u_{t-1}^i)$ and parameters $\{\theta_t^j = \theta_{t-1}^j\}_{j=1}^M$ using (3).
8:       Generate M random samples of the measurements $\{Y_t^j|(z_t^j, u_t^i)\}_{j=1}^M \sim p(y_t|z_t^j, u_t^i)$ using (3).
9:       Approximate the lower bound $\tilde{L}_t^\theta$ using (15).
10:     **end for**
11:   **end for**
12:   Evaluate the approximate cost function in (24a).
13:   Use any standard constrained non-linear optimization algorithm to find a new input design parameters $P_\Gamma = P_\Gamma^{(c)}$ and $P_\Pi = P_\Pi^{(c)}$. Set $c \leftarrow c + 1$.
14: **end while**

$$X_{t+1} = aX_t + \frac{X_t}{b + X_t^2} + u_t + V_t, \quad V_t \sim \mathcal{N}(0, Q_t), \quad (26a)$$

$$Y_t = cX_t + dX_t^2 + W_t, \qquad W_t \sim \mathcal{N}(0, R_t), \quad (26b)$$

where $\theta \triangleq [a\,b\,c\,d]$ is a vector of model parameters to be estimated, with $\theta^\star = [0.8\,0.7\,0.6\,0.5]$ being the true parameter vector. The noise covariances are selected as $Q_t = 0.01$ and $R_t = 0.01$, for all $t \in \mathbb{N}$. For Bayesian identification, $\{\theta_t = \theta_{t-1}\}_{t\in\mathbb{N}} = \theta$ in (24) is a random process, with $Z_t = \{X_t, \theta_t\}$, such that $Z_0 \sim \mathcal{N}(z_m, z_c)$, where $z_m = [1\,0.7\,0.6\,0.5\,0.4]$, $z_c = \text{diag}(0.01,\,0.01,\,0.01,\,0.01,\,0.01)$. Here we assume that $u_{min} \le \{u_t\}_{t\in\mathbb{N}} \le u_{max}$, where $u_{min} = -0.8$ and $u_{max} = 0.8$. Starting at $t = 0$, we are interested in choosing an input sequence $\{u_{1:N}\}_{N\in\mathbb{N}}$ that would eventually lead to minimization of the MSE of the parameter estimates, computed using an SMC based Bayesian estimator given in Tulsyan et al. [2013a]. Algorithm 1 was implemented with $N = 100$, $M = 2000$, and $M_u = 2000$. For input, we consider Example 18, with $g = 2$, such that $\mathcal{U} = \{u_{min}, u_{max}\}$. Here $\{U_t\}_{t\in\mathbb{N}} = \{u_t\}$ have the following initial and transition probability

Case 1: $P_\Gamma = [p_1 \quad 1 - p_1]$, $\quad P_\Pi = \begin{bmatrix} p_1 & 1 - p_1 \\ 1 - p_1 & p_1 \end{bmatrix}$;

Case 2: $P_\Gamma = [p_1 \quad 1 - p_1]$, $\quad P_\Pi = \begin{bmatrix} p_1 & 1 - p_1 \\ 1 - p_2 & p_2 \end{bmatrix}$;

Case 3: $P_\Gamma = [p_0 \quad 1 - p_0]$, $\quad P_\Pi = \begin{bmatrix} p_1 & 1 - p_1 \\ 1 - p_2 & p_2 \end{bmatrix}$,

where p_i, where $i = \{0, 1, 2\}$ in Cases 1 through 3 are the probabilities. For comparison purposes, we also consider a pseudo-random binary signal, which can be represented as

Case 4: $P_\Gamma = [0.5 \quad 0.5]$, $\quad P_\Pi = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$.

For all of the above cases, $\Phi(\cdot)$ in (24a) was selected as the trace. Table 1 gives $P_{\Gamma^\star}$ and $P_{\Pi^\star}$ for Cases 1 through 3 as computed by Algorithm 1, and Figure 1(a) gives the
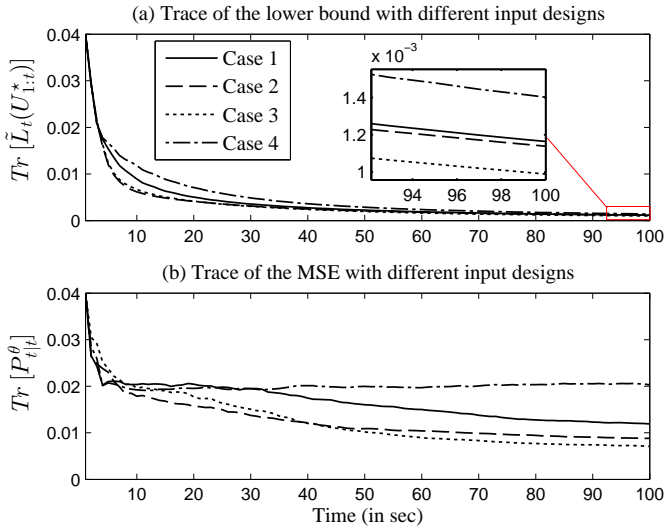
Fig. 1. (a) Trace of the approximate lower bound; (b) trace of the MSE. Magnification of the key region of (a) is provided as inset.

corresponding trace of the lower bound. It is clear from Table 1 and Figure 1(a) that Case 3 yields the lowest objective function value. Although the objective function value for Case 2 is comparable to Case 3, note that Case 3 provides the most general form of the Markov chain in $\mathcal{U}$.

Figure 1(b) validates the quality of the designed inputs based on the performance of the Bayesian estimator. From Figure 1(b), it is clear that with Case 3, the estimator yields the lowest trace of MSE at all sampling time. The same is also evident from Table 1; wherein, the sum of the trace of MSE is smallest with Case 3 as the input. The Results are based on 500 MC simulations, starting with 500 i.i.d. input path trajectories generated from $\{U_{1:N}\} \sim P_{\Gamma^\star, \Pi^\star}^{1:N}$ for Cases 1 through 4. If required, a more rigorous validation of the designed input can be performed using the approach proposed in Tulsyan et al. [2013b].

The results appear promising; however, we faced problems in solving the optimization. As discussed earlier, (24) is a stochastic programming problem, as a result (24a) tends to be non-smooth, and have many local minima. In future, we will consider use of stochastic gradient-based methods.

Table 1. Results as computed by Algorithm 1.

	Case 1	Case 2	Case 3	Case 4
p_0	N.A.	N.A.	0.34	N.A.
p_1	0.62	0.63	0.61	N.A.
p_2	N.A.	0.92	0.72	N.A.
$\overline{\psi}(U_{1:100}^\star)$	0.42	0.37	0.36	0.51
$\sum_{t=1}^{100} \mathrm{Tr}[P_{t|t}^\theta]$	1.66	1.27	1.25	2.02

6. CONCLUSIONS

An algorithm for input design for Bayesian identification of stochastic non-linear SSM is proposed. The developed algorithm is based on minimization of the PCRLB with respect to inputs. One of the distinct advantages of the proposed method is that the designed input is independent of the Bayesian estimator used for identification. Simulation results suggest that the proposed method can be used to deliver accurate inference on the parameter estimates.

REFERENCES

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

C. Brighenti, B. Wahlberg, and C.R. Rojas. Input design using Markov chains for system identification. In *Proceedings of the 48th IEEE Conference on Decision and Control and the 28th Chinese Control Conference*, pages 1557–1562, Shanghai, China, 2009.

K. Chaloner and I. Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995.

W.R. Gilks, S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, 1995.

X. Huan and Y.M. Marzouk. Simulation-based optimal Bayesian experimental design for non-linear systems. *Journal of Computational Physics*, 232(1):288–317, 2012.

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

H. Kück, N. de Freitas, and A. Doucet. SMC samplers for Bayesian optimal non-linear design. In *Proceedings of the Non-linear Statistical Signal Processing Workshop*, pages 99–102, Cambridge, U.K., 2006.

D.V. Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005, 1956.

P. Müller and G. Parmigiani. Optimal design via curve fitting of Monte Carlo experiments. *Journal of the American Statistical Association*, 90(432):1322–1330, 1995.

J. Nocedal and S.J. Wright. *Numerical Optimization: Springer Series in Operations Research*. Springer, New York, 2006.

R.B. Bapat and T.E.S. Raghavan. *Nonnegative Matrices and Applications*. Cambridge University Press, Cambridge, 1997.

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

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

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

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