Input Design for Nonlinear Stochastic Dynamic Systems - A Particle Filter Approach

R. Bhushan Gopaluni * **Thomas B. Schön** **
Adrian G. Wills ***

* Department of Chemical and Biological Engineering, University of British Columbia, Vancouver, Canada V6T 1Z3 (Tel: 604 827 5668; e-mail: bhushan.gopaluni@ubc.ca).
** Division of Automatic Control, Linköping University, SE-581 83 Linköping, Sweden, e-mail: schon@isy.liu.se
*** School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan, NSW 2308, Australia (e-mail: Adrian.Wills@newcastle.edu.au)

Abstract: We propose an algorithm for optimal input design in nonlinear stochastic dynamic systems. The approach relies on minimizing a function of the covariance of the parameter estimates of the system with respect to the input. The covariance matrix is approximated using a joint likelihood function of hidden states and measurements, and a combination of state filters and smoothers. The input is parametrized using an autoregressive model. The proposed approach is illustrated through a simulation example.

1. INTRODUCTION

Input design is an important first step in parameter estimation. Often, the inputs are designed by optimizing some function of the covariance matrix of the parameters. Deriving this covariance matrix as an explicit function of the inputs has proved to be a difficult problem. The early work on input design was based on asymptotic covariance expressions for linear models derived in Ljung [1985]. These expressions are asymptotic in both model order and sample size. They are also simple and provide good frequency domain intuition into the asymptotic variance of the estimates and hence have been used extensively for input design (Ljung and Yuan [1985], Yuan and Ljung [1985], Gevers and Ljung [1986], Hjalmarsson et al. [1998], Forssell and Ljung [2000]). They have also been successfully implemented in practice (Zhu [1998, 2003]).

Despite the wide spread use of the above mentioned asymptotic expressions in input design, some recent work has shown that covariance expressions based on asymptotic model order are not accurate (Xie and Ljung [2001], Garatti et al. [2004], Ninness and Hjalmarsson [2004]). There has also been work done on obtaining finite sample covariance estimates (Weyer et al. [1999], Weyer and Campi [2002], Campi and Weyer [2005]). However, these methods do not yet seem to be amenable to input design. More recently, the asymptotic covariance expressions based on the Cramér-Rao lower bound have been used to develop convex optimization based input design methods (Hildebrand and Gevers [2003], Jansson and Hjalmarsson [2005]).

The parameter covariance estimates developed for linear systems, in particular those developed for asymptotic model order, are unfortunately, not applicable to all types of nonlinear systems in a straightforward fashion. However, if maximum likelihood methods are used for parameter estimation, the Cramér-Rao lower bound can be used to approximate the covariance estimates. In general, this lower bound can not be written as an explicit function of inputs and therefore, formulation of the input design problem is rather challenging for nonlinear stochastic systems. Consequently, the literature on input design for such systems has been rather scarce despite attempts to solve this problem for some special cases of nonlinear systems.

An input design algorithm for static nonlinear systems was developed in Vincent et al. [2010]. In Hjalmarsson and Mårtensson [2007], the authors attempt to learn from the linear input design case by considering certain types of input-output nonlinear models and in Larsson et al. [2010] an algorithm for FIR type nonlinear systems was considered. However, to the best of the authors knowledge, there are no known algorithms for general nonlinear state-space models. In this work, we propose an input design algorithm for nonlinear stochastic state-space models.

The proposed approach is based on the expressions for the Cramér-Rao lower bound derived in Robert and Casella [1999] and in Oakes [1999]. These expressions are often not available as explicit functions of the input sequence. We employ sequential Monte Carlo methods [Gordon et al., 1993, Doucet and Johansen, 2011] in order to obtain an approximation of the covariance matrix and then we

employ a parametrization of the input signal to render the problem of input design computationally tractable. More specifically it is a smoothing probability density function for the state that is needed and this can be approximated using particle smoothers, see e.g., Douc et al. [2010], Doucet and Johansen [2011].

2. PROBLEM FORMULATION

Let us consider a nonlinear dynamic state-space model of the form

$$x_{t+1} = f(x_t, u_t, \theta) + w_t, \qquad (1a)$$
$$y_t = g(x_t, u_t, \theta) + v_t, \qquad (1b)$$

where $x_t \in \mathbf{R}^{n \times 1}$ is the n-dimensional state vector, $u_t \in \mathbf{R}^{s \times 1}$ is the s-dimensional input vector, $y_t \in \mathbf{R}^{m \times 1}$ is the m-dimensional output or measurement vector, and w_t, v_t are independent and identically distributed (i.i.d.) Gaussian noise sequences of appropriate dimension, $\theta \in \mathbf{R}^{p \times 1}$ is a p-dimensional parameter vector and $f(.), g(.)$ are known nonlinear functions. We use $(.)_{i:j}, j > i$ to denote a sequence of data from $t = i$ to $t = j$. For instance $y_{i:j}$ denotes the set of measurements $\{y_i, y_{i+1}, \cdots, y_j\}$. Let us assume that the data from an experiment on a process described by the above model consists of T samples and that the true parameter vector is θ_0.

There are many approximate approaches to parameter estimation for the model in (1). Among them, the maximum likelihood approach has received significant attention, since in many cases it provides consistent estimates (among other attractive properties). While these properties have not been shown for the nonlinear case considered here, we nevertheless develop an approach to design the input sequence to minimize a function of the covariance matrix of parameters estimated through the maximization of the likelihood function. The asymptotic covariance of the maximum likelihood estimates is given either by the negative inverse of the second derivative of the log-likelihood function or by the inner product of the first derivatives of the log-likelihood function (Cramér-Rao lower bound). Let $p_y(y_{1:T}|u_{1:T}, \theta)$ be the density function of the measurements. Then the covariance matrix of the parameter estimates, denoted by $\Sigma(u_{1:T}, \theta_0)$, is given by the following expressions,

$$\Sigma(u_{1:T}, \theta_0) = - \left[E_y \left(\frac{\partial^2}{\partial \theta \partial \theta'} \log p_y(y_{1:T}|u_{1:T}, \theta) \right) \right]^{-1}_{\theta = \theta_0} \qquad (2a)$$

$$= \left[E_y \left(\frac{\partial}{\partial \theta} \log p_y(y_{1:T}|u_{1:T}, \theta) \times \frac{\partial}{\partial \theta} \log p_y(y_{1:T}|u_{1:T}, \theta)' \right) \right]^{-1}_{\theta = \theta_0} \qquad (2b)$$

where $(.)'$ is used to denote the transpose of a vector and E_y denotes the expected value with respect to $p_y(y_{1:T}|u_{1:T}, \theta)$. Using the parameter covariance expression, the input design problem is often expressed as,

$$\min_{u_{1:T}} h(\Sigma(u_{1:T}, \theta_0))$$
$$\text{s.t. } l(u_{1:T}) \leq 0 \qquad (3)$$

where $h(\Sigma(u_{1:T}, \theta_0))$ can take different forms depending on the context of input design. It is common to use $h(\Sigma(u_{1:T}, \theta_0)) = \text{trace}(\Sigma(u_{1:T}, \theta_0))$ or $h(\Sigma(u_{1:T}, \theta_0)) = \det(\Sigma(u_{1:T}, \theta_0))$. Furthermore, $l(u_{1:T})$ is a function of the input that allows us to impose constraints on the input magnitude or power or rate of change. The above input design problem does not have any output constraints. However, output constraints may be required in real problems and it is straightforward to include them in the above formulation.

The optimization problem in (3) is non-trivial due to the fact that often a functional form of h is unknown and that the optimization could potentially be high dimensional (dimensionality of the optimization variable being sT) and nonconvex. Even in the simplest case of linear systems with Gaussian state and measurement noise, deriving an explicit expression for h is difficult.

3. FOUR CHALLENGES AND OUR APPROACH

There are four challenges that we need to address in order to make the optimization problem (3) tractable,

(1) we need an approximation of the covariance matrix;
(2) the input sequence needs to be characterized;
(3) we have to account for the fact that $h(\Sigma(u_{1:T}, \theta_0)$ is not necessarily deterministic and it varies with a particular realization of the inputs $u_{1:T}$ used in any given experiment;
(4) the optimization problem must be solved efficiently and reliably.

Another well-known, but under-appreciated challenge is that the optimization problem (as posed in (3)) depends on the true value of the parameters that need to be estimated. In other words, we need an optimal input in order to obtain the best possible estimate of θ_0, however, that estimate itself is needed to design the optimal input. Clearly, the problem in (3) is not practical, however, an often accepted remedy is to use an estimate of θ_0 in (3). We believe that a good solution to avoid having to use θ_0 in (3) is to design the input through an iterative approach, where an estimate of θ_0 is progressively improved. However, this approach is beyond the scope of this work and is not further explored. In the following sections, solutions are proposed for each of the four challenges identified above.

We believe that any attempt to provide a generic analytical solution to this problem will be very challenging at best. Instead, we propose an approximation of the optimization problem (3). Let us now discuss how to handle the four challenges in more detail.

3.1 Challenge 1: Parameter Covariance Matrix

This section provides the expressions to be approximated and briefly discusses how these approximations may be obtained. In Section 4 we will then provide explicit expressions for how to compute these approximations.

The first challenge is to find an approximation of the covariance matrix. It is well known that the derivatives of the density function of the measurements, $\frac{\partial}{\partial \theta} p_y(y_{1:T}|u_{1:T}, \theta)$, which are required to estimate covariance (as in (2)) are hard to evaluate for nonlinear state-space models. However, due to the Markov property of the states in the state-space model in (1), the joint density function of the hidden

states and the measurements $p_{xy}(x_{1:T}, y_{1:T}|\theta)$ is easier to evaluate. In Oakes [1999], it was shown that the covariance matrix in (2a) can be expressed as a function of this joint density function. The corresponding covariance matrices are given by

$$
\Sigma(u_{1:T}, \theta_0) = - \left[E_y \big(\frac{\partial^2}{\partial\theta\partial\theta'} \log p_y(y_{1:T}|u_{1:T}, \theta) \big) \right]^{-1}_{\theta=\theta_0}
$$
$$
= - \left[E_y \left(\frac{\partial^2}{\partial\theta\partial\theta'} E_x[\log p_{xy}(x_{1:T}, y_{1:T}|u_{1:T}, \theta)] + \right. \right.
$$
$$
\left. \left. \frac{\partial^2}{\partial\theta\partial\theta_0'} E_x[\log p_{xy}(x_{1:T}, y_{1:T}|u_{1:T}, \theta)] \right) \right]^{-1}_{\theta=\theta_0}
$$
$$(4)$$

where E_x is the expectation operator with respect to $p_x(x_{1:T}|y_{1:T}, \theta_0)$. The approximations that we employ in this work, are easier to express if the expectation operator is outside the derivatives. The derivatives and the expectation operators can be interchanged and (4) can be expressed as (see Louis [1982], Robert and Casella [1999], Duan and Fulop [2011])

$$
\Sigma(u_{1:T}, \theta_0) = -E_y \left[E_x \frac{\partial^2}{\partial\theta\partial\theta'} [\log p_{xy}(x_{1:T}, y_{1:T}|u_{1:T}, \theta)] + \right.
$$
$$
\left. + Var_x \left(\frac{\partial}{\partial\theta} [\log p_{xy}(x_{1:T}, y_{1:T}|u_{1:T}, \theta)] \right) \right]^{-1}_{\theta=\theta_0}
$$
$$(5)$$

where Var_x is the variance with respect to $p_x(x_{1:T}|y_{1:T}, \theta_0)$. The covariance matrix in (2b), can also be similarly expressed using the joint density function as follows (Oakes [1999]),

$$
\Sigma(u_{1:T}, \theta_0) = \left[E_y \left(\frac{\partial}{\partial\theta} E_x[\log p_{xy}(x_{1:T}, y_{1:T}|u_{1:T}, \theta)] \right. \right.
$$
$$
\left. \left. \frac{\partial}{\partial\theta} E_x[\log p_{xy}(x_{1:T}, y_{1:T}|u_{1:T}, \theta)]' \right) \right]^{-1}_{\theta=\theta_0} \quad (6)
$$

The expressions for the covariance matrices given in (5) and (6) can both be used in the input design problem. The covariance matrix as expressed in (6) is inherently positive definite, and therefore it is sometimes preferred. However, numerical simulations suggest that the particle approximations used in this work require a large number of samples of states and measurements in order to reliably approximate (6), as compared to those required to approximate (5). The computational complexity of the particle approximations (presented in Section 4) increases with the number of samples and therefore, an approximation of (5) is used in estimating the parameter covariance.

In order to approximate (5), we must be able to evaluate high dimensional integrals with respect to $p_y(y_{1:T}|u_{1:T}, \theta)$ and $p_x(x_{1:T}|y_{1:T}, \theta)$. The expected value E_x can easily be approximated using sequential Monte Carlo methods [Gordon et al., 1993, Doucet and Johansen, 2011]. In the following section, one such approximation is developed and is used in the optimization problem in (3).

3.2 Challenge 2: Input Parametrization

An algorithm for input design is developed in this section, by proposing solutions to the remaining three challenges identified earlier. The second challenge is that of characterizing the input sequence from $t = 1$ to $t = T$. A naive approach to the optimization problem in (3) will be to simply treat the vector $u_{1:T}$ as a high dimensional optimization variable. However, this approach will render the optimization rather inefficient if the length of experiment (T) is "large". Instead, we propose to parametrize the input sequence by using the following autoregressive model,

$$
u_t = \Phi(u_{t-1:t-n_u}, \theta_u) + r_t, \quad (7)
$$

where Φ is a linear or nonlinear function of the input, and θ_u is a parameter vector that defines this function. Furthermore, $r_t \in \mathbf{R}^s$ is a noise sequence independent of the state and the measurement noise. The input in (7) is uniquely determined by the coefficients θ_u and the characteristics of noise r_t. The noise sequence can also be parametrized by assuming that it has a standard parametric probability density function. This representation of the input naturally covers a wide range of input sequences, with the range being dependent on the order and the form of the function Φ.

3.3 Challenge 3: Formulating the Optimization Problem

The input design problem in (3) can now be posed as

$$
\min_{\theta_u} \ h(\Sigma(u_{1:T}, \theta_0))
$$
$$
\text{s.t.} \ \frac{1}{T} \sum_{t=1}^{T} [u_t - \Phi(u_{t-1:t-n_u}, \theta_u)]^2 \le \sigma_r^2, \quad (8)
$$
$$
l(u_{1:T}) \le 0,
$$

where σ_r^2 is the variance of r_t. The above formulation is appealing from an optimization point of view as long as the number of parameters in θ_u is much smaller than T. It is easy to see that $h(\Sigma(u_{1:T}, \theta_0))$ is deterministic only if $u_{1:T}$ is deterministic. However, $u_{1:T}$ is stochastic due to the nonlinear autoregressive model in (7). Assuming that the noise sequences, r_t, w_t, and v_t are independent, a deterministic version of the optimization problem in (8) can be formulated as,

$$
\min_{\theta_u} \ h(E_u \Sigma(u_{1:T}, \theta_0))
$$
$$
\text{s.t.} \ \frac{1}{T} \sum_{t=1}^{T} E_u[u_t - \Phi(u_{t-1:t-n_u}, \theta_u)]^2 \le \sigma_r^2, \quad (9)
$$
$$
E_u[l(u_{1:T})] \le 0,
$$

where E_u is the expected value with respect to the input.

3.4 Challenge 4: Solving the Optimization Problem

The fourth challenge is the optimization of the nonlinear objective function in (9). An explicit functional form of the objective function is difficult (if not impossible) to derive for general nonlinear stochastic systems. However, an approximation can be obtained using sequential Monte Carlo methods. Any standard optimization algorithm can be used on this approximate objective function.

4. APPROXIMATING THE COVARIANCE MATRIX

An approximation of the covariance matrix in (5), requires an approximation of the following expected values of the derivatives of the joint log-likelihood function,

$$I_1 \triangleq E_x \frac{\partial^2}{\partial\theta\partial\theta'}[\log p_{xy}(x_{1:T}, y_{1:T}|u_{1:T}, \theta)] \qquad (10a)$$

$$I_2 \triangleq E_x \left(\frac{\partial}{\partial\theta}[\log p_{xy}(x_{1:T}, y_{1:T}|u_{1:T}, \theta)]\right)^2 \qquad (10b)$$

$$I_3 \triangleq \left[E_x \big(\frac{\partial}{\partial\theta}[\log p_{xy}(x_{1:T}, y_{1:T}|u_{1:T}, \theta)]\big)\right]^2 \qquad (10c)$$

[2] In Robert and Casella [1999], a Monte Carlo approximation based on the measurement of a large number of samples of the state trajectories was proposed. This approach involves generating many samples of the state trajectory from the distribution $p(x_{1:T}|y_{1:T}, \theta)$. For instance, the approximation of the expected value of the first derivative in I_3 takes the following form,

$$\sqrt{I_3} = E_x \frac{\partial}{\partial\theta}[\log p_{xy}(x_{1:T}, y_{1:T}|\theta)] \approx$$
$$\frac{1}{N}\sum_{i=1}^{N} \frac{\partial}{\partial\theta}[\log p_{xy}(x_{1:T}^{(i)}, y_{1:T}|\theta)], \quad (11)$$

where $x_{1:T}^{(i)}$ are samples of the states drawn from the distribution $p(x_{1:T}|y_{1:T}, \theta)$ and N is the number of such samples. This is a naive approach that does not account for the likelihood of occurrence of a particular state trajectory $x_{1:T}^{(i)}$. Instead, an approximation of the expectation operator of the following form was suggested in Andrieu et al. [2004],

$$E_x \frac{\partial}{\partial\theta}[\log p_{xy}(x_{1:T}, y_{1:T}|\theta)] \approx$$
$$\frac{1}{N}\sum_{i=1}^{N} w^{(i)} \frac{\partial}{\partial\theta}[\log p_{xy}(x_{1:T}^{(i)}, y_{1:T}|\theta)], \quad (12)$$

where $w^{(i)}$ are weights proportional to the probability of the occurrence of $x_{1:T}^{(i)}$. In Andrieu et al. [2004], it was also pointed out that sampling directly from the high dimensional density function $p(x_{1:T}|y_{1:T}, \theta_i)$ is inefficient. An alternative approximation that only requires sampling from low dimensional density functions can be developed following the approach in Gopaluni [2008], Schön et al. [2011].

Using the Markov property of the state-space model, the expectation in (10c) can be written as,

$$\sqrt{I_3} = \int \frac{\partial}{\partial\theta} \log[p(x_1|y_{1:T}, \theta)]p(x_1|y_{1:T}, u_{1:T}, \theta)dx_1$$
$$+ \sum_{t=2}^{T} \int \frac{\partial}{\partial\theta} \log[p(x_t|x_{t-1}, \theta)]p(x_{t-1:t}|y_{1:T}, \theta)dx_{t-1:t}$$
$$+ \sum_{t=1}^{T} \int \frac{\partial}{\partial\theta} \log[p(y_t|x_t, \theta)]p(x_t|y_{1:T}, \theta)dx_t, \quad (13)$$

where $p(.)$ denotes the probability density function of the corresponding variables. The integrals in (13) typically do not have analytical solutions. However, particle approximations of $p(x_t|y_{1:T}, \theta)$ and $p(x_{t-1:t}|y_{1:T}, \theta)$ allow us to approximate the integrals. The following expressions for particle approximations of these functions are derived in Gopaluni [2008], Schön et al. [2011],

---

[2] Please note that for simplicity the expressions for I_1, I_2 and I_3 are written for scalar θ but are valid for vector θ as well with appropriate changes.

$$p(x_t|y_{1:T}, u_{1:T}, \theta) = \sum_{i=1}^{N} w_{t|T}^{(i)} \delta(x_t - x_t^{(i)}), \qquad (14a)$$

$$p(x_{t-1:t}|y_{1:T}, u_{1:T}, \theta) = \sum_{i=1}^{N} w_{t-1,t}^{(i)} \delta(x_{t-1:t} - x_{t-1:t}^{(i)}), \qquad (14b)$$

where $w_{t|T}^{(i)}$ and $w_{t-1,t}^{(i)}$ are appropriate weights and $x_t^{(i)}, x_{t-1:t}^{(i)}$ are particle samples of x_t drawn from an importance density function. Furthermore, δ represents the Dirac-delta function. Complete expressions for the weights are provided in the references cited. Using the above approximations of the density functions, the following approximation of $\sqrt{I_3}$ can be obtained,

$$\sqrt{I_3} \approx \sum_{i=1}^{N} w_{1|T}^{(i)} \frac{\partial}{\partial\theta} \log[p(x_1^{(i)}|y_{1:T}, \theta)]dx_1$$
$$+ \sum_{t=2}^{T}\sum_{i=1}^{N} w_{t-1,t}^{(i)} \frac{\partial}{\partial\theta} \log[p(x_t^{(i)}|x_{t-1}^{(i)}, \theta)]dx_{t-1:t}$$
$$+ \sum_{t=1}^{T}\sum_{i=1}^{N} w_{t|T}^{(i)} \frac{\partial}{\partial\theta} \log[p(y_t|x_t^{(i)}, \theta)]dx_t. \qquad (15)$$

The particle approximations, $\widehat{I_1}$ and $\widehat{I_2}$, of I_1 and I_2 are obtained analogously. An approximation of the covariance matrix can now be written as,

$$\Sigma(u_{1:T}, \theta_0) \approx -\left[E_y\left(\widehat{I_1} + \widehat{I_2} - \widehat{I_3}\right)\right]_{\theta=\theta_0}^{-1} \qquad (16)$$

This approximation depends on both the input and the measurement sequences. Clearly, no measurements are available before an experiment is conducted. Therefore, an approximation of the expectation with respect to the measurements can be obtained as follows,

$$E_y(\widehat{I_1} + \widehat{I_2} - \widehat{I_3}) =$$
$$\int (\widehat{I_1} + \widehat{I_2} - \widehat{I_3})p_y(y_{1:T}|u_{1:T}, \theta_0)dy_{1:T}. \qquad (17)$$

The density function of the measurements conditioned on the input sequence is unknown. However, a sample-based approximation (albeit not very efficient) of the following form can be obtained through simulations of the model in (1),

$$p_y(y_{1:T}|u_{1:T}, \theta_0) \approx \frac{1}{M_y}\sum_{i=1}^{M_y} \delta(y_{1:T} - y_{1:T}^{(i)}), \qquad (18)$$

where $y_{1:T}^{(i)}$ are samples of measurements and M_y is the number of such samples generated. Using the approximation (18) of $p_y(y_{1:T}|u_{1:T}, \theta_0)$, the integral in (17) can now be approximated. The covariance matrix is still a function of the stochastic input that is to be designed. Therefore, in (9) we proposed using the average covariance matrix over all the stochastic inputs possible. The expectation of the objective function with respect to the input can similarly be approximated using the sample-based approximation of the input density function,

$$p_u(u_{1:T}|\theta_u) \approx \frac{1}{M_u}\sum_{i=1}^{M_u} \delta(u_{1:T} - u_{1:T}^{(i)}), \qquad (19)$$

where M_u is the number of input sequences generated for a given θ_u. The input samples can be generated

from the nonlinear autoregressive model in (7). These approximations allow us to formulate an objective function that is tractable. It should be noted that the accuracy of the approximate objective function depends on a number of factors, such as the number of particles used (N), the number of samples of $y_{1:T}$ (M_y), the number of samples of $u_{1:T}$ (M_u), the noise characteristics of the model and the input.

5. RESULTING INPUT DESIGN ALGORITHM

The approximations introduced above allow us to reformulate the input design optimization problem as follows,

$$
\begin{aligned}
\min_{\theta_u} \quad & \frac{1}{M_u} \sum_{i=1}^{M_u} h(\Sigma(u_{1:T}^{(j)}, \theta_0)) \\
\text{s.t.} \quad & \frac{1}{T} \sum_{t=1}^{T} E_u[u_t - \Phi(u_{t-1:t-n_u}, \theta_u)]^2 \leq \sigma_r^2, \\
& E_u[l(u_{1:T})] \leq 0.
\end{aligned}
\tag{20}
$$

The above optimization problem is often non-convex and needs to be implemented through an iterative approach that involves standard numerical solvers [Nocedal and Wright, 2006]. The proposed algorithm for input design, including the iterations in the optimization algorithm, is summarized in Algorithm 1 below,

---

Algorithm 1 Input Design Using the Particle Smoother

1: Choose an initial guess for the input design parameter vector $\theta_u = \theta_u^{(0)}$. Set $j \leftarrow 0$.
2: **while not** converged **do**
3:     Generate an input sequence using (7).
4:     Generate the M_y measurement samples, $y_{1:T}^{(i)}$, and M_u samples of $u_{1:T}^{(j)}$ through simulations of their respective models.
5:     Approximate the covariance matrix using (16), (18) and (19).
6:     Use any standard nonlinear optimization algorithm to find a new input design parameter vector, $\theta_u^{(j+1)}$ that decreases the value of the objective function in (20).
7:     Set $j \leftarrow j + 1$.
8: **end while**

The algorithm proposed above does come with a few costs. It is computationally very expensive since it involves multiple evaluations of particle approximations of the smoothed and the filtered states. Moreover, a number of approximations are made in arriving at the covariance of the parameter estimates, making it difficult to quantify the accuracy of these approximations. However, these approximations have the desirable property of asymptotic consistency (as $N \to \infty$ and $M \to \infty$).

As mentioned earlier, a practical input design algorithm should not depend on the true model parameter vector θ_0. However, the method as presented above assumes that this vector is known. It is beyond the scope of this work to explore iterative approaches that will allow us to circumvent this problem, however, we would like to mention that this algorithm is very much amenable to iterative input design. The proposed algorithm is illustrated through a simulated example in the subsequent section.

6. SIMULATION EXAMPLE

The following simulated example with a nonlinear state-space model is taken from Wills et al. [2008].

$$
x_{t+1} = ax_t + \frac{x_t}{b + x_t^2} + u_t + w_t, \qquad w_t \sim \mathcal{N}(0, q), \quad \text{(21a)}
$$
$$
y_t = cx_t + dx_t^2 + v_t, \qquad\qquad e_t \sim \mathcal{N}(0, r), \quad \text{(21b)}
$$

where $\theta_0' = [a\ b\ c\ d] = [0.7\ 0.6\ 0.5\ 0.4]$ are parameters of the model that are estimated. The state and measurement noise covariances are both equal to 0.01 (i.e., $q = r = 0.01$). The state and measurement equations are both nonlinear in x_t and the parameter b appears nonlinearly. We are interested in choosing an input that minimizes the covariance of these parameters. Please note that the variances of the state and the measurement equations are also usually estimated, however, to keep the example simple enough, we do not consider them as parameters.

The proposed algorithm was implemented with $N = 100$, and $T = 500$. The following five cases were considered,

    Case 1: A second order autoregressive model was used,

$$
u_t = \alpha_1 u_{t-1} + \alpha_2 u_{t-1} + r_t, \tag{22}
$$

where $\theta_u' = [\alpha_1\ \alpha_2]$ is the input parameter vector. The variance of r_t was 0.01. In order to ensure that the model is stable, the input parameters were restricted to $\alpha_1 \in [0, 1)$ and $\alpha_2 \in [0, 1)$ and the maximum allowed input variance was unity (i.e., $\|u_{1:T}\|_2^2 \leq 1$).

    Case 2: The input model in (22) was used and the input power was constrained (i.e., $\|u_{1:T}\|_2^2 \leq 0.25$).

    Case 3: A first order autoregressive model of the following form was used,

$$
u_t = \alpha_1 u_{t-1} + r_t. \tag{23}
$$

The variance of r_t was 0.01 and $\alpha_1 \in [0, 1)$.
    Case 4: Case 3 was repeated with the input power constrained as in Case 2.
    Case 5: The input was white noise with unit variance.

In all the above cases, the trace of the covariance matrix was used in the objective function. A graph showing the magnitude plot of the optimal input under the five cases is shown in Figure 1. The trace of the covariance matrix in cases 2 and 4, when the input power is heavily constrained, is as expected more than that in cases 1 and 3. The trace of the covariance matrix with white noise input is better than that due to the optimal inputs generated through the proposed algorithm.

The results in this example appear promising. However, we faced problems in implementing an optimization algorithm on the objective function in (20). As mentioned earlier, the proposed optimization problem is potentially nonconvex. Moreover, a number of approximations are made in evaluating this objective function. As a result, the objective function tends to be non-smooth and hence may have many local minima. To get around the local minima, in future, we will be considering stochastic optimization algorithms.

Table 1. Trace of covariance matrix (the objective function in the input design problem)

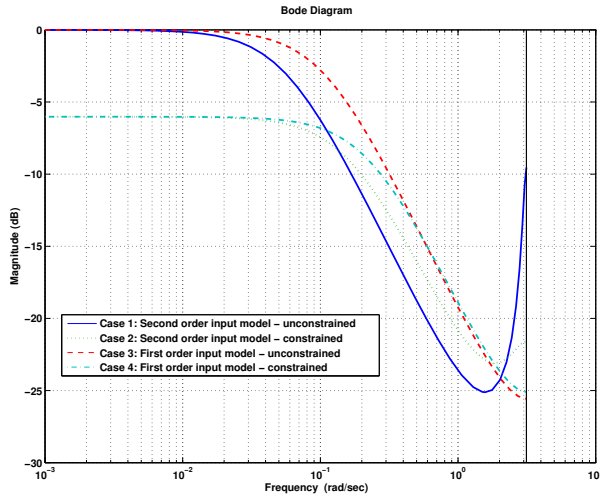| Case	1	2	3	4	5
trace(Σ)	$1.1e^{-3}$	$2.2e^{-3}$	$1.6e^{-3}$	$2.6e^{-3}$	$6.3e^{-4}$

Fig. 1. The magnitude plot of the input in the five cases considered.

7. CONCLUSIONS

An algorithm for input design in nonlinear stochastic systems is proposed. The algorithm involves minimization of a function of particle approximations of the Cramér-Rao lower bound. The algorithm provides a tractable solution to the problem considered. Simulation results suggest that the approximations are reasonable and the algorithm is a viable alternative to designing inputs through linear approximation of nonlinear systems.

REFERENCES

C. Andrieu, A. Doucet, S. S. Singh, and V. B. Tadić. Particle methods for change detection, system identification, and contol. *Proceedings of the IEEE*, 92(3):423–438, March 2004.

M.C. Campi and E. Weyer. Guaranteed non-asymptotic confidence regions in system identification. *Automatica*, 41(10):1751–1764, 2005.

R. Douc, A. Garivier, E. Moulines, and J. Olsson. Sequential Monte Carlo smoothing for general state space hidden Markov models. *Submitted to Annals of Applied Probability*, 2010.

A. Doucet and A. M. Johansen. *The Oxford Handbook of Nonlinear Filtering*, chapter A tutorial on particle filtering and smoothing: fifteen years later, D. Crisan and B. Rozovsky (eds.), pages 656–704. Oxford University Press, 2011.

J.C. Duan and A. Fulop. A stable estimator of the information matrix under EM for dependent data. *Statistics and Computing*, 21(1):83–91, 2011.

U. Forssell and L. Ljung. Some results on optimal experiment design. *Automatica*, 36(5):749–756, 2000.

S. Garatti, M.C. Campi, and S. Bittani. Assessing the quality of identified models through the asymptotic theory – when is the result reliable? *Automatica*, 40(8):1319–1332, 2004.

M. Gevers and L. Ljung. Optimal experiment design for prediction error identification in view of feedback design. *Automatica*, 22(5):543–554, 1986.

R. B. Gopaluni. A particle filter approach to identification of nonlinear processes under missing observations. *Canadian Journal of Chemical Engineering*, 86(6):1081–1092, 2008.

N. Gordon, D. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings of Radar, Sonar, Navigation*, 140(1):107–113, 1993.

R. Hildebrand and M. Gevers. Identification for control: Optimal input design with respect to a worst case ν-gap cost function,. *SIAM Journal of Optimization and Control*, 41:1586–1608, 2003.

H. Hjalmarsson and J. Mårtensson. Optimal input design for identification of non-linear systems: Learning from the linear case. In *Proceedings of the American Control Conference (ACC)*, pages 1572–1576, New York, NY, USA, July 2007.

H. Hjalmarsson, M. Gevers, S. Gunnarsson, and O. Lequin. Iterative feedback tuning: Theory and applications. *IEEE Control Systems Magazine*, 18(4):26–41, 1998.

H. Jansson and H. Hjalmarsson. Input design via LMIs admitting frequency-wise model specifications in confidence regions. *IEEE transactions on Automatic Control*, 50(10):1534–1549, 2005.

C.A. Larsson, H. Hjalmarsson, and C.R. Rojas. On optimal input design for nonlinear FIR-type systems. In *Proceedings of the IEEE Conference on Decision and Control (CDC)*, Atlanta, USA, December 2010.

L. Ljung. Asymptotic variance expressions for identified black-box transfer function models. *IEEE transactions on Automatic Control*, 30(9):834–844, 1985.

L. Ljung and Z. Yuan. Asymptotic properties of black-box identification of transfer functions. *IEEE transactions on Automatic Control*, 30(6):514–530, 1985.

T. A. Louis. Finding the observed information matrix when using the EM algorithm. *Journal of Royal Statistical Society: Series B*, 44(2):226–233, 1982.

B. Ninness and H. Hjalmarsson. Variance error quantifications that are exact for finite model order. *IEEE Transactions of Automatic Control*, 49(8):1275–1291, 2004.

J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Series in Operations Research. Springer, New York, USA, 2 edition, 2006.

D Oakes. Direct calculation of the information matrix via the em algorithm. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 61(2):479–482, 1999.

C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, New York, 1999.

T. B. Schön, A. Wills, and B. Ninness. System identification of nonlinear state-space models. *Automatica*, 47(1):39–49, January 2011.

T. Vincent, C. Novara, K. Hsu, and K. Poolla. Input design for structured nonlinear system identification. *Automatica*, 46(6):990–998, 2010.

E. Weyer and M.C. Campi. Non-asymptotic confidence ellipsoids for the least squares estimate. *Automatica*, 38(9):1539–1547, 2002.

E. Weyer, R.C. Williamson, and I.M.Y. Mareels. Finite sample properties of linear model identification. *IEEE Transactions on Automatic Control*, 44(7):1370–1383, 1999.

A. Wills, T. B. Schön, and B. Ninness. Parameter estimation for discrete-time nonlinear systems using EM. In *Proceedings of the 17th IFAC World Congress*, Seoul, South Korea, July 2008.

L.L. Xie and L. Ljung. Asymptotic expressions for estimated frequency functions. *IEEE Transactions on Automatic Control*, 46(12):1887–1899, 2001.

Z. D. Yuan and L. Ljung. Unprejudiced optimal open loop input design for identification of transfer functions. *Automatica*, 21(6):697–708, 1985.

Y. Zhu. Multivariable process identification for MPC: the asymptotic method and its applications. *Journal of Process Control*, 8(2):101–115, 1998.

Y. Zhu. New development in industrial MPC identification. *Proceedings of the International Symposium on Advanced Control of Chemical Processes (ADChEM)*, 2003. Hong Kong, China.