

# Input design using Markov chains for system identification

Chiara Brighenti, Bo Wahlberg, Cristian R. Rojas

**Abstract**—This paper studies the input design problem for system identification where time domain constraints have to be considered. A finite Markov chain is used to model the input of the system. This allows to directly include input amplitude constraints in the input model, by properly choosing the state space of the Markov chain. The state space is defined so that the Markov chain will generate a binary sequence. The probability distribution of the Markov chain will be shaped in order to minimize the cost function considered in the input design problem. Stochastic approximation is used to minimize that cost function. With this approach, the input signal to apply to the system can be easily generated by extracting samples from the optimal distribution. A numerical example shows how these models can improve system identification with respect to other input realization techniques.

## I. INTRODUCTION

Mathematical models for systems are necessary in order to predict their behavior and as parts of their control systems. This work focuses on models constructed and validated from experimental input/output data, by means of identification methods.

Information obtained through experiments on the real system depend on the input excitation which is often limited by amplitude or power constraints. For this reason, experiment design is necessary in order to obtain system estimates within a given accuracy, saving time and cost of the experiment [1]. Robustness of input design for system identification is also one of the most important issues, specially when the model of the system is used for projecting its control system. In [2]-[7] some studies on this problem are presented. The effects of undermodeling on input design are pointed out in [8] and [9].

Depending on the cost function considered in this setting, input design can be typically solved as a constrained optimization problem. In the Prediction Error Method (PEM) framework it is common to use a function of the asymptotic covariance matrix of the parameter's estimate as a measure of the estimate accuracy. This matrix depends on the input spectrum that can be shaped in order to obtain a "small" covariance matrix [10], [11]. Usually, a constraint on the input power is also included; in this way, time domain amplitude constraints are translated in the frequency domain [12].

A first disadvantage of these methods is that they are strongly influenced by the initial knowledge of the system. Secondly, solving the problem in the frequency domain does not provide any further information on how to generate the input signal in the time domain: the input can be represented as filtered white noise, but many probability distributions can be used to generate white noise. Furthermore, in practical applications time domain constraints on the signals have to

be considered and the power constraint that is usually set in the frequency domain does not assure that these constraints are respected. For this reason, in [13] a method is proposed to generate a binary input with a prescribed correlation function; once an optimal spectrum or correlation function is found solving the input design problem in the frequency domain, it is possible to generate a binary signal which approximates the optimal input. Also in [14] a method is proposed that provides a locally optimal binary input in the time domain.

This paper studies the input design problem in the probability domain. Compared to design methods in the frequency domain, a solution in the probability domain makes it easier to generate input trajectories to apply to the real system, by extracting samples from a given distribution. Inputs are modeled by finite stationary Markov chains which generate binary signals. Binary signals are often used in system identification and one of the reasons is that they achieve the largest power in the set of all signals having the same maximum amplitude and it is well known that this improves parameter estimation for linear models. The idea of modeling the input by a finite Markov chain derives from the possibility of including the input amplitude constraints directly into the input model, by suitably choosing the state space of the Markov chain. Furthermore, unlike the design in the frequency domain, this approach keeps the degrees of freedom in the choice of the optimal spectrum, which in general is not unique [12].

The optimal distribution is found by minimizing the cost function defined in the input design problem with respect to the 1-step transition probabilities of the Markov chain. In this analysis, a stochastic approximation algorithm is used since a closed-form solution to the optimization problem is not available and the cost function is a stochastic function of these transition probabilities and is contaminated with noise (see [16], [17] for details).

The paper is structured as follows. In Section II the problem formulation is presented. Section III defines the input model and describes the algorithm to solve the input design problem. In Section IV, a numerical example is considered which shows the quality of the proposed method. Finally, Section V concludes the paper.

## II. PROBLEM FORMULATION

### A. System and model description

This paper considers discrete-time linear time-invariant Single-Input-Single-Output systems, lying in the set  $\mathcal{M}$  of parametric models

$$y(t) = G(q, \theta)u(t) + H(q, \theta)e(t), \quad (1)$$

$$G(q, \theta) = \frac{q^{-n_k} b_1 + b_2 q^{-1} + \dots + b_{n_b} q^{-n_b}}{1 + a_1 q^{-1} + \dots + a_{n_a} q^{-n_a}}$$

$$H(q, \theta) = \frac{1 + c_1 q^{-1} + \dots + c_{n_c} q^{-n_c}}{1 + d_1 q^{-1} + \dots + d_{n_d} q^{-n_d}},$$

$$\theta = [b_1, \dots, b_{n_b}, a_1, \dots, a_{n_a}, c_1, \dots, c_{n_c}, d_1, \dots, d_{n_d}]' \in \mathbb{R}^{b \times 1}$$

where  $u(t)$  is the input,  $y(t)$  is the output and  $e(t)$  is zero mean white noise with finite variance. The symbol  $q^{-1}$  represents the delay operator.

The real system is given as

$$y(t) = G_0(q)u(t) + H_0(q)e_0(t), \quad (2)$$

where  $e_0(t)$  has finite variance  $\lambda_0$ . Let  $\theta_0$  be a parameter vector such that  $G(q, \theta_0) = G_0(q)$  and  $H(q, \theta_0) = H_0(q)$ , i.e. assume there is no undermodelling.

### B. Identification method

The identification method considered here is PEM. Suppose all the hypothesis for the consistence of the PEM are satisfied (see [18]). Upon the condition that  $u$  and  $e$  are independent, the asymptotic expression (in the number of data points  $N$ ) of the inverse of the covariance matrix of the parameter's estimate is

$$P_{\theta_0}^{-1} = \frac{N}{2\pi\lambda_0} \int_{-\pi}^{\pi} F_u(e^{j\omega}, \theta_0) \Phi_u(\omega) F_u^*(e^{j\omega}, \theta_0) d\omega + R_e(\theta_0) \quad (3)$$

$$R_e(\theta_0) = \frac{N}{2\pi} \int_{-\pi}^{\pi} F_e(e^{j\omega}, \theta_0) F_e^*(e^{j\omega}, \theta_0) d\omega$$

where  $F_u(e^{j\omega}, \theta_0) = H^{-1}(q, \theta_0) \frac{\partial G(q, \theta)}{\partial \theta} \Big|_{\theta_0}$ ,  $F_e(e^{j\omega}, \theta_0) = H^{-1}(q, \theta_0) \frac{\partial H(q, \theta)}{\partial \theta} \Big|_{\theta_0}$  and  $\Phi_u(\omega)$  is the power spectral density of the input  $u(t)$ . Here  $*$  denotes the complex conjugate transpose.

### C. Optimal input design problem

The objective of the identification procedure is the estimation of the system parameters  $\theta$ . In input design problems the aim is to minimize a measure of the estimate error. Often in practice, it is also necessary to take into account some constraints on the real signals. Here, the cost function to be minimized has the form

$$J(u, \theta_0) = f(P_{\theta_0}(u)) + g$$

where  $f$  is any function of the covariance matrix of the parameters' estimate and  $g$  is a term which represents the cost of the experiment. Typical functions  $f$  are the trace, the determinant or the highest eigenvalue of the covariance matrix  $P_{\theta_0}$  [1].

## III. SOLUTION AND ALGORITHM DESCRIPTION

### A. Input model

This paper focuses on the input design problem in the probability domain: a finite stationary Markov chain is used as an input model and the cost function is minimized with respect to the transition probabilities of the Markov chain, which completely define its distribution [15]. The use of a

Markov chain as a model of the input signal allows to include time domain constraints on the signal amplitude directly into the input model, by suitably choosing the state space. The idea is then to use Markov chain distributions to generate binary signals. Consider the general Markov chain having states of the form

$$(u_{t-n}, u_{t-n+1}, \dots, u_t) \quad (4)$$

where  $u_i$  represents the value of the input at the time instant  $i$ ; it can be equal to either  $u_{max}$  or  $-u_{max}$ , where  $u_{max}$  is the maximum tolerable input amplitude. This model allows the present value of the input to depend on the last  $n$  past values. Note that at the time instant  $t$ , the state can transit only to either the state  $(u_{t-n+1}, u_{t-n+2}, \dots, u_t, u_{max})$  or  $(u_{t-n+1}, u_{t-n+2}, \dots, u_t, -u_{max})$  with probabilities  $p(u_{t-n}, \dots, u_t)$  and  $1 - p(u_{t-n}, \dots, u_t)$ , respectively. The last component of the Markov chain state will generate the binary signal to apply to the real system.

By means of the Markov chains and state space realization theories (see [15] and [19]), it is possible to derive a general expression for the spectrum of a finite stationary Markov chain  $s_n$  having state space  $\mathcal{S} = \{S_1, S_2, \dots, S_J\}$ , where each state has the form (4). For the Markov chains considered here, the number of states is  $J = 2^{n+1}$ .

Let  $\Pi$  be the transition matrix whose elements are the conditional probabilities  $\Pi(i, j) = P\{s_{n+1} = S_j | s_n = S_i\}$  and  $\bar{p} = [\bar{p}_1 \dots \bar{p}_J]$  the vector of stationary probabilities  $\bar{p}_i = P\{s_n = S_i\}$ . Defining

$$A_s = [S_1 \dots S_J]$$

and

$$D_s = \begin{bmatrix} \bar{p}_1 & & 0 \\ & \ddots & \\ 0 & & \bar{p}_J \end{bmatrix}$$

it is possible to write the correlation coefficients of the output signal in the matricial form

$$r_k = A_s D_s \Pi^k A_s', \quad k = 0, 1, 2, \dots$$

For  $k < 0$  the correlation can be obtained by the symmetry condition  $r_k = r_{-k}$ .

For  $k = 1, 2, \dots$ , the correlation function can be viewed as the impulse response of the linear system

$$\begin{aligned} x_{k+1} &= \Pi x_k + \Pi A_s' u_k \\ y_k &= A_s D_s x_k \end{aligned} \quad (5)$$

Therefore, the spectrum of the Markov chain signal  $s_n$  can be expressed as

$$\Phi_s(q) = W(q) + r_0 + W(q^{-1}), \quad (6)$$

where the first term is the transfer function  $W(q) = A_s D_s (qI - \Pi)^{-1} \Pi A_s'$  of the system (5).

As an example, two simple Markov chains can be considered: the first has state space  $\mathcal{S}_2 = u_{max} \cdot \{-1, 1\}$  and the second  $\mathcal{S}_4 = u_{max} \cdot \{(1, 1), (1, -1), (-1, -1), (-1, 1)\}$ . The

transition matrices considered for these two Markov chains are:

$$\Pi_2 = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix} \quad \Pi_4 = \begin{bmatrix} p & 1-p & 0 & 0 \\ 0 & 0 & r & 1-r \\ 0 & 0 & p & 1-p \\ r & 1-r & 0 & 0 \end{bmatrix}.$$

This choice of the transition matrices makes the Markov chain symmetric in the sense that the transition probabilities are invariant with respect to exchanges in the sign of the states components. With this choice of the transition matrices, the Markov chain signals have zero mean and unit variance. Note that when  $p = r$  the four states Markov chain model is equivalent to the two states Markov chain.

The two states Markov chain signal has a spectrum given by

$$\Phi_u(q) = \frac{1 - \alpha^2}{(q - \alpha)(q^{-1} - \alpha)}$$

where  $\alpha = 2p - 1$ . Note that this is the spectrum of an AR process. It is also possible to see that the four states Markov chain has a higher order spectrum, where the number of poles and zeros depends on the values of the probabilities  $p$  and  $r$  and can be up to eight.

Note that even if the input is designed in the probability domain, the input spectrum is shaped by the choice of  $n$  and of the transition matrix of the Markov chain.

The purpose of the input design problem is to minimize the cost function  $J(u, \theta_0)$  with respect to the transition probabilities,  $p$  in the first case,  $p$  and  $r$  in the second.

### B. Cost function evaluation

Since the analytic expressions for the covariance matrix  $P_{\vartheta_0}$  as a function of the transition probabilities of the Markov chain modeling the input are quite involved, simulation techniques are required to evaluate the cost function.

From the model expression (1) it is possible to write

$$e(t) = H(q, \theta)^{-1} (y(t) - G(q, \theta) u(t))$$

and by linearizing the functions  $G(q, \theta)$  and  $H(q, \theta)$  in  $\theta_0$

$$\begin{aligned} G(q, \theta) &\approx G_0(q) + \Delta G(q, \theta) \\ H(q, \theta) &\approx H_0(q) + \Delta H(q, \theta) \end{aligned}$$

where  $\Delta G(q, \theta) = (\theta - \theta_0)' \frac{\partial G(q, \theta)}{\partial \theta} \Big|_{\theta_0}$ ,  $\Delta H(q, \theta) = (\theta - \theta_0)' \frac{\partial H(q, \theta)}{\partial \theta} \Big|_{\theta_0}$ , the following expression is derived:

$$e(t) = (H_0(q) + \Delta H(q, \theta))^{-1} (H_0(q) e_0(t) - \Delta G(q, \theta) u(t))$$

By substituting the Taylor expansion  $(H_0(q) + \Delta H(q, \theta))^{-1} \approx H_0(q)^{-1} - \frac{\Delta H(q, \theta)}{H_0(q)^2}$  and the expressions of  $\Delta G(q, \theta)$  and  $\Delta H(q, \theta)$ , it results

$$e_0(t) \approx (\theta - \theta_0)' \left( \frac{1}{H_0(q)} \frac{\partial H(q, \theta)}{\partial \theta} \Big|_{\theta_0} e_0(t) + \frac{1}{H_0(q)} \frac{\partial G(q, \theta)}{\partial \theta} \Big|_{\theta_0} u(t) \right) + e(t). \quad (7)$$

The problem of estimating the parameter  $\theta$  for the model (1) is asymptotically equivalent to solving the least squares problem for (7), when the number of data points  $N$  used for estimation goes to infinity. Therefore, the asymptotic expression (3) can be approximated as

$$P_{\theta_0}^{-1} = \frac{1}{\lambda_0} (S'S)$$

where  $S = [w_1 \dots w_b] \in \mathbb{R}^{N \times b}$  and  $w_i \in \mathbb{R}^{N \times 1}$  is the sequence obtained from  $w_{it} = \frac{1}{H_0(q)} \frac{\partial G(q, \theta)}{\partial \theta} \Big|_{\theta_0} u(t) + \frac{1}{H_0(q)} \frac{\partial H(q, \theta)}{\partial \theta} \Big|_{\theta_0} e_0(t)$ .

Therefore, at each iteration of the algorithm, the cost function is evaluated using randomly generated input and noise signals.

### C. Algorithm description

When evaluating the cost function by simulation, it is necessary to consider that the cost function is a stochastic variable that depends on the transition probabilities of the Markov chains and on the noise process  $e(t)$ . Therefore, the cost function values generated through simulation have to be considered as samples of that stochastic variable. The true value of the cost function for a given transition probability would be the mean of that stochastic variable. For these reasons, stochastic approximation is necessary in order to minimize the cost function with respect to the transition probabilities of the Markov chain describing the input.

One of the most common stochastic approximation methods that do not require the knowledge of the cost function gradient is the finite difference stochastic approximation (FDSA) [16]. It uses the recursion

$$\hat{p}_{k+1} = \hat{p}_k - a_k \widehat{\nabla} J_k,$$

where  $\widehat{\nabla} J_k$  is an estimate of the gradient of  $J$  at the  $k$ -th step.

The FDSA estimates the gradient of the cost function as

$$\widehat{\nabla} J_{ki} = \frac{J(\hat{p}_k + c_k e_i) - J(\hat{p}_k - c_k e_i)}{2c_k}$$

where  $e_i$  denotes the unit vector in the  $i$ -th direction and  $\widehat{\nabla} J_{ki}$  is the  $i$ -th component of the gradient vector.

Depending on the number  $d$  of parameters with respect to which minimize the cost function, a simultaneous perturbation stochastic approximation (SPSA) may be more efficient than the FDSA [20]; when  $d$  increases the number of cost function evaluations in a FDSA procedure may be too large and the algorithm be very slow. In that case the SPSA algorithm described in [20] gives a better performance, since it requires only two evaluations of the cost function regardless of  $d$ .

The algorithm is initialized by a first evaluation of the cost function on a discrete set of points and choosing the minimum in that set. At any point in this set, the cost function is evaluated only once; therefore, the value obtained is a sample extracted from the stochastic variable describing the cost function at that point. Nevertheless, the result of the initialization procedure may be sufficiently accurate so there could be no need to run many algorithm iterations. This of

course will depend on the cost function shape and on the choice of the grid of points.

The sequences  $a_k$  and  $c_k$  are chosen as  $a_k = \frac{a}{A+k+1}$  and  $c_k = \frac{c}{(k+1)^{1/3}}$ , which are asymptotically optimal for the FDSA algorithm (see [16]). A method for choosing  $A$ ,  $a$  and  $c$  will be presented in the next Section, analyzing a numerical example. An analytic proof of the algorithm convergence can be found in [16].

#### IV. NUMERICAL EXAMPLE

Consider a mass-spring-damper system, where the input  $u$  is the force applied to the mass and the output  $y$  is the mass position. It is described by the transfer function

$$G_0(s) = \frac{\frac{1}{m}}{s^2 + \frac{c}{m}s + \frac{k}{m}}$$

with  $m = 100$  Kg,  $k = 10$  N/m and  $c = 6.3246$  Ns/m, resulting the natural frequency  $\omega_n = 0.3162$  rad/s and the damping  $\xi = 0.1$ . The power here is defined as  $pw(t) = u(t)\dot{y}(t)$ . White noise with variance  $\lambda_0 = 0.0001$  is added at the output and an output-error model is used [18]. Data are sampled with  $T_s = 1$  s and the number of data points generated is  $N = 1000$ . As a measure of the estimate accuracy, the trace of the covariance matrix  $P_{\theta_0}$  is used. In order to consider also some practical constraints on the amplitude of the input and output signals and the maximum and mean power dissipated during the system excitation, a general cost function will be used:

$$J(u, \theta_0) = f_1(TrP_{\theta_0}(u)) + f_2(u_{max}) + f_3(y_{max}) + f_4(pw_{max}) + f_5(pw_{mean}) \quad (8)$$

where  $u_{max}$  and  $y_{max}$  are the absolute maximum values of the input and output signals,  $pw_{max}$  and  $pw_{mean}$  are the maximum and mean power dissipated during the experiment. Thresholds for  $TrP_{\theta_0}$ ,  $u_{max}$ ,  $y_{max}$ ,  $pw_{max}$  and  $pw_{mean}$  have been set, which define the maximum values allowed for each of these variables. This can be seen as a classical multiobjective optimization approach (an overview can be found in [22]).

Figure 1 and 2 show the cost functions  $f_2, f_3$  and  $f_1, f_4, f_5$ , respectively: when the variables  $TrP_{\theta_0}$ ,  $u_{max}$ ,  $y_{max}$ ,  $pw_{max}$  and  $pw_{mean}$  reach their maximum acceptable value (100%) the cost is one. Outside the interval of acceptable values, the cost functions continue growing linearly.

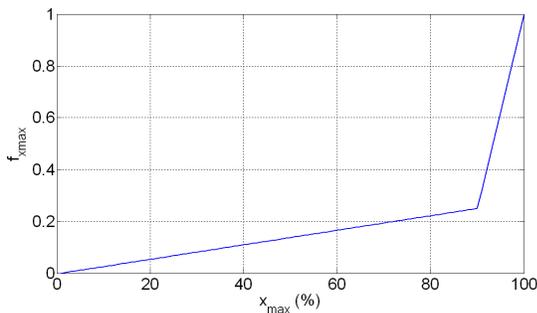


Fig. 1. Cost functions  $f_2$  and  $f_3$  in the interval of acceptable values of the variables  $u_{max}$  and  $y_{max}$ , respectively.

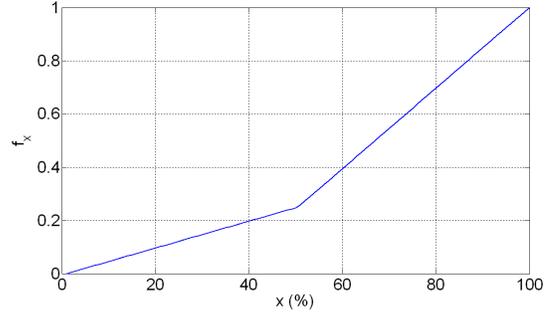


Fig. 2. Cost functions  $f_1, f_4, f_5$  in the interval of acceptable values.

As input models, the two simple examples of Markov chains described in Section III-A are here considered.

The set of points used for the algorithm initialization as explained in Section III-C is  $\{0.1, 0.2, \dots, 0.9\}$ . In the case analyzed here, since the cost function depends on no more than two parameters, the FDSA is used. The algorithm coefficients are chosen depending on the estimate of the cost function gradient at the initial condition, so that the product  $a_0 \nabla J_0$  has magnitude approximately equal to the expected changes among the elements of  $\hat{p}_k$  in the early iterations.

Three cases that have been studied are summarized in Table I:

- 1) The cost associated to  $TrP$  and the costs associated to the physical constraints have comparable values.
- 2) No power and amplitude constraints are considered.
- 3) Very strict power constraints are considered.

As a term of comparison for the performance of the Markov input model, a pseudo-random binary signal and white noise with unit variance (the same as the variance of the Markov chains) have been applied as inputs to the system. In Tables IV, V and VI the results of the simulation runs for all the three cases listed above are shown. The cost function values are estimated by evaluating the average of 100 simulation runs using the optimal input found by the algorithm, the PRBS and white noise inputs. Table V, related to the case 2, shows the optimal value of the trace of the covariance matrix calculated by solving the LMI (linear matrix inequality) formulation of the input design problem in the frequency domain, as explained in [12]. Furthermore, by the method described in [13], a binary signal having the optimal correlation function is generated. The minimum obtained with this input signal is also shown in Table V.

The second case, which is the most standard in input design problems, is first analyzed in detail.

Figure 3 presents the cost function, estimated on a fine grid of points, as the average of 100 simulations. Table II

TABLE I  
MAXIMUM THRESHOLD VALUES IN THE THREE ANALYZED CASES

Case	$TrP$	$u_{max}$	$y_{max}$	$pw_{max}$	$pw_{mean}$
1	$5 * 10^{-6}$	1 N	1 m	$0.3 \frac{Nm}{s}$	$0.03 \frac{Nm}{s}$
2	$5 * 10^{-6}$	Inf N	Inf m	Inf $\frac{Nm}{s}$	Inf $\frac{Nm}{s}$
3	$5 * 10^{-6}$	1 N	1 m	$0.03 \frac{Nm}{s}$	$0.003 \frac{Nm}{s}$

exhibits the results of two Monte-Carlo simulations (each consisting of 100 runs), which show that the variance of the algorithm output decreases approximately as  $\frac{1}{N_{Iter}}$ , where  $N_{Iter}$  is the number of algorithm iterations; this guarantees the empirical algorithm convergence. With 10000 iterations the algorithm produces the results in Figure 4. The optimality of the probability  $\hat{p}$  found by the algorithm has been verified by using the expression of the two states Markov chain spectrum in the asymptotic expression (3) and minimizing  $TrP_{\theta_0}$  with respect to  $\alpha$ ; it turns out that the optimal value  $\hat{p} = 0.8714$  is very close to the one found by the stochastic algorithm after 30000 iterations, that is  $\hat{p} = 0.8712$  (Table III). This confirms that the stochastic algorithm converges to the true optimal value. In practice, it is not necessary to run the algorithm for 30000 iterations, since already at the initial condition the cost function is very close to the minimum and the variance of the estimate after 10000 iterations is of the order of  $10^{-5}$ . It has been done here, anyway, to show that the final value obtained is the true optimal one. Notice from the results in Table V that the Markov chains give lower values of the trace of  $P(u, \theta_0)$  than all the other inputs, except the true optimal spectrum. By means of the Multiple Signal Classification (MUSIC) methods, described in [23], the frequencies of the optimal input spectrum for the case 2 have been estimated. It results that the optimal input consists of two sinusoids of frequencies  $0.3023 \text{ rad/s}$  and  $0.3571 \text{ rad/s}$ , respectively, where the main contribution is given by the sinusoid of high frequency, which has approximately 5.6 times the power of the first component. Note that these frequencies are very close to the natural frequency of the system and to the poles of the Markov chains spectra (Figures 5 and 6).

In case 1, the Markov inputs and the PRBS signal give almost the same cost value. This happens because the optimal values of the transition probabilities are approximately 0.5, which means that the Markov chains signal is essentially

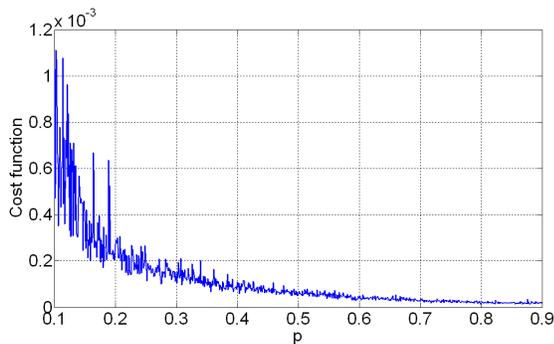


Fig. 3. Estimate of the cost function on a discrete set of points for the two states Markov chain in the case 2.

TABLE II  
RESULTS OF 100 MONTE-CARLO SIMULATIONS OF THE ALGORITHM WITH THE 2 STATES MARKOV CHAIN.

$N_{Iter}$	Mean value $E\hat{p}$	Variance $Var\hat{p}$
1000	0.8657	$4.6 \times 10^{-4}$
2000	0.8671	$2.5 \times 10^{-4}$

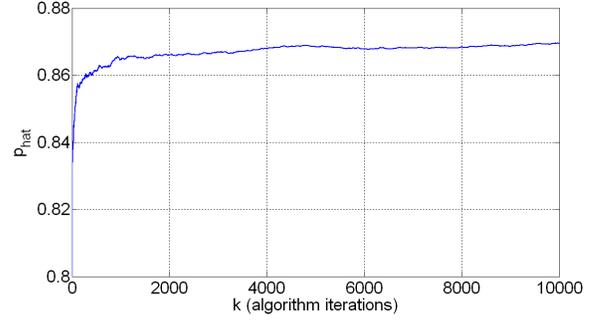


Fig. 4. Estimation of the best transition probability for the two states Markov chain in the case 2.

TABLE III  
OPTIMAL VALUES OF THE TRANSITION PROBABILITIES IN THE CASES 1, 2 AND 3.

Case	$S_2$	$S_4$
1	$\hat{p} = 0.4720$	$\hat{p} = [0.4730 \quad 0.6794]'$
2	$\hat{p} = 0.8712$	$\hat{p} = [0.8494 \quad 0.6445]'$
3	$\hat{p} = 0.1100$	$\hat{p} = [0.0005 \quad 0.2981]'$

binary white noise (and this depends on the choice of the thresholds values). Note in Figure 5 and 6 that in this case the spectra of the Markov chains are almost constant. The white input, which is generated with a gaussian distribution, gives a much higher cost value, due to signals' amplitude and power.

Also in the third case, when more strict power constraints are set in the problem, the use of a Markov chain is preferable (see results in Table VI). Therefore, when amplitude and power constraints have to be considered in the input design problem, the Markov chain model can really improve system identification. The optimal distribution is easily estimated by simulation of the real system.

Finally, note that the two states Markov chain performs a little better than the four states Markov chain in the first two cases, while in the third case, when stricter power constraints are considered, the four states Markov chain achieves the lowest cost function value. The reason for this is that in case 1 and 2 the optimal input structure is the two states Markov chain; therefore, the stochastic algorithm performs better if the simple input model is used, rather than a more general one that

TABLE IV  
TOTAL COST FUNCTION VALUES OBTAINED WITH THE OPTIMAL MARKOV INPUTS, A PRBS AND WHITE NOISE IN CASE 1.

	$S_2$	$S_4$	PRBS	WN
$J(u, \theta_0)$	1.2758	1.2788	1.2564	20.1326

TABLE V  
TRACE OF THE COVARIANCE MATRIX OBTAINED WITH THE OPTIMAL MARKOV INPUTS, A PRBS, WHITE NOISE, A BINARY INPUT HAVING THE OPTIMAL CORRELATION FUNCTION AND THE OPTIMAL SPECTRUM IN CASE 2.

	$S_2$	$S_4$	PRBS	WN	BI	Optimum
$TrP$	1.43e-7	1.59e-7	4.35e-7	4.66e-7	2.18e-6	2.85e-8

TABLE VI

TOTAL COST FUNCTION VALUE OBTAINED WITH THE OPTIMAL MARKOV INPUTS, A PRBS AND WHITE NOISE IN CASE 3.

	$\mathcal{S}_2$	$\mathcal{S}_4$	PRBS	WN
$J(u, \vartheta_0)$	78.51	73.58	163.85	484.94

requires more parameters to be tuned.

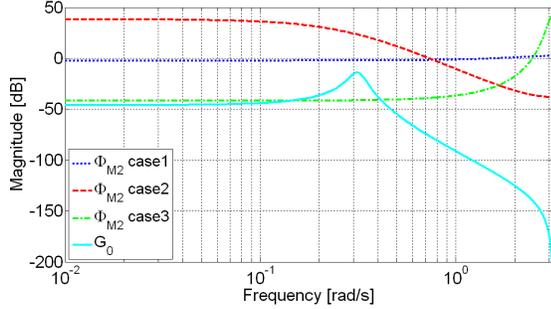


Fig. 5. Bode diagrams of the optimal spectra of the 2 states Markov chains in the cases 1, 2 and 3 of Table III, and of the real discrete system.

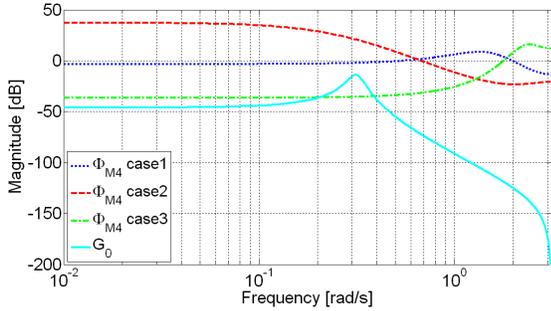


Fig. 6. Bode diagrams of the optimal spectra of the 4 states Markov chains in the cases 1, 2 and 3 of Table III, and of the real discrete system.

To conclude, from the results in Table V, this example shows that the Markov chain model gives a trace of the covariance matrix 10 times lower than the value obtained with a binary input having the optimal correlation function. This means that the Markov chain input model can improve system identification considerably.

## V. CONCLUSIONS

In this paper, the input design problem for system identification has been studied using finite Markov chains as models of the input signals. The main advantage of this approach with respect to the one in the frequency domain, is that the input model directly includes the input amplitude constraints that are always present in practical applications. Secondly, the solution in the probability domain makes it easier to generate the input signal, since its samples can be extracted from the optimal distribution. By a numerical example, the quality of this input model has been tested, comparing it to other standard input models that are often used in system identification. The results show that the use of a Markov model can notably improve the estimation performance.

## REFERENCES

- [1] G. C. Goodwin, R. L. Payne, *Dynamic system identification: experiment design and data analysis*. New York: Academic Press, 1977.
- [2] H. Hjalmarsson, "From experiment design to closed-loop control", *Automatica*, vol. 41, no. 3, pp. 393-438, 2005.
- [3] C. R. Rojas, J. S. Welsh, G. C. Goodwin, A. Feuer, "Robust optimal experiment design for system identification", *Automatica* 43 (2007) 993-1008.
- [4] J. Martensson, H. Hjalmarsson, "Robust input design using sum of squares constraints", in *IFAC Symposium on System Identification*, Newcastle, Australia, March 2006, pp.1352-1357.
- [5] G. C. Goodwin, J. S. Welsh, A. Feuer, M. Derpich, "Utilizing prior knowledge in robust optimal experiment design", in *IFAC Symposium on System Identification*, Newcastle, Australia, March 2006, pp.1358-1363.
- [6] H. Jansson, "Experiment Design with Applications in Identification for Control", Doctoral Thesis, KTH, Stockholm 2004.
- [7] B. L. Cooley, J. H. Lee, S. P. Boyd, "Control-relevant experiment design: a plant-friendly, LMI-based approach", in *American Control Conference*, Philadelphia, Pennsylvania, June 1998, pp. 1240-1244.
- [8] H. Hjalmarsson, J. Martensson, B. Wahlberg, "On some robustness issues on input design", in *IFAC Symposium on System Identification*, Newcastle, Australia, , March 2006, pp.511-516.
- [9] X. Bombois, M. Gilson, "Cheapest identification experiment with guaranteed accuracy in the presence of undermodeling", in *IEEE Conference on Decision and Control*, Paradise Island, Bahamas, December 2004, pp.505-510.
- [10] H. Jansson, H. Hjalmarsson, "Input design via LMIs admitting frequency-wise model specifications in confidence regions", *IEEE Transactions on Automatic Control*, vol. 50, no. 10, pp. 1534-1549, 2005.
- [11] K. Lindqvist, H. Hjalmarsson, "Optimal input design using linear matrix inequalities", in *IFAC Symposium on System Identification*, Santa Barbara, California, USA, July 2000.
- [12] M. Barenthin, "On Input Design in System Identification for Control", Licentiate Thesis in Automatic Control, KTH, Stockholm 2006.
- [13] C. R. Rojas, J. S. Welsh, G. C. Goodwin, "A receding horizon algorithm to generate binary signals with a prescribed autocovariance", *Proceedings of the ACC'07 Conference*, 2007, New York, USA.
- [14] H. Suzuki, T. Sugie, "On input design for system identification in time domain", *Proceedings of the European Control Conference 2007*, Kos, Grece, July 2-5, 2007.
- [15] J. L. Doob, *Stochastic processes*. Wiley, New York, 1953.
- [16] G. C. Pflug, *Optimization of Stochastic Models*, Kluwer Academic Publishers, 1996.
- [17] L. Ljung, G. Pflug, H. Walk, *Stochastic approximation and optimization of random systems*, Birkhauser, 1991.
- [18] L. Ljung, *System identification: Theory for the user, Second Edition*. Prentice Hall, 1999.
- [19] T. Kailath, *Linear systems*. Prentice-Hall, Englewood Cliffs, NJ, 1980.
- [20] J. C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation", *IEEE Transactions on Automatic Control*, vol. 37, no. 3, March 1992.
- [21] J. C. Spall, "Implementation of the simultaneous perturbation algorithm for stochastic optimization", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 3, pp. 817-823, July 1998.
- [22] E. Zitzler, "Evolutionary algorithms for multiobjective optimization: methods and applications ", Doctoral thesis, Swiss Federal Institute of Technology Zurich, 1999.
- [23] P. Stoica, R. Moses, *Spectral analysis of signals*. Prentice-Hall, Upper Saddle River, New Jersey, 2005.