Chance Constrained Input Design

Cristian R. Rojas, Dimitrios Katselis, Håkan Hjalmarsson, Roland Hildebrand and Mats Bengtsson

***Abstract*— In this paper the following problem is studied: design an input signal with the property that the estimated model based on this signal satisfies a given performance level with a prescribed probability. This problem is mathematically translated into a chance constrained optimization problem, which is typically non convex. To solve it, several convex approximations are proposed and compared.**

I. Introduction

Experiment design is one of the core issues in system identification. In the last decade this topic has seen a revival, see, e.g., the overviews [9, 14, 15, 21, 24]. Contributions include convexification [17], robust design [25], adaptive design [8], least-costly design [3], MIMO systems [18], and closed vs open loop experiments [1].

Deriving from [10] much effort has been devoted to applications oriented experiment design, in particular control oriented experiment design [2, 6, 13]. Recently a conceptual framework for applications oriented experiment design was proposed in [15]. The framework hinges on introducing a function V_{app} which quantifies the degradation in performance when a model that differs from the true system is used in the design of the application. Thus V_{app} has its minimum at the true system and without loss of generality we may assume that the minimum is 0. The archetypical example is model based control design, where the controller is based on a model. Here $V_{app}(G)$ could be some norm of the difference in the output response when the model G and when the true system is used in the design of the controller. A model G is deemed to give acceptable performance if the performance degradation is below a certain bound, i.e. $V_{app}(G) \leq 1/\gamma$ for some $\gamma > 0$. The set of acceptable models is the set

$$\mathcal{E}_{app} = \left\{ G : V_{app}(G) \leq \frac{1}{\gamma} \right\}.$$

The system identification objective is then reduced to produce a model that belongs to $\mathcal{E}_{app}$. In this perspective, a natural experiment design objective inspired by least-costly identification is as follows:

$$\begin{array}{ll} \min\limits_{\text{Experiment}} & \text{Experimental effort} \\ \text{s.t.} & \hat{G} \in \mathcal{E}_{app}, \end{array} \tag{1}$$

where $\hat{G}$ is the identified model. Various measures of the experimental effort can be used; input or output power, and experimental length are just a few examples. There is also a wide range of possible design variables, see [17].

In this contribution we will study (1) for parametric model identification. Thus the model is represented by a parameter vector $\theta \in \mathbb{R}^n$. We will also adopt a stochastic framework whereby the parameter estimate is assumed to be (approximately) normal distributed for the sample sizes in question, i.e.

$$\hat{\theta} \sim \mathbf{N}(\theta_0, P), \tag{2}$$

where θ_0 is the true model parameter, c.f. the prediction error framework [20]. Note that the covariance matrix P will depend on the experimental conditions.

Under (2) the constraint in (1) can never be guaranteed. In fact it can only be ensured with a certain probability. We thus have to relax (1) to

$$\begin{array}{ll} \min\limits_{\text{Experiment}} & \text{Experimental effort} \\ \text{s.t.} & \mathbf{P}_{\hat{\theta}} \left\{ \hat{\theta} \in \mathcal{E}_{app} \right\} \geq 1 - \varepsilon \end{array} \tag{3}$$

for some small $\varepsilon > 0$. Above $\mathbf{P}_{\hat{\theta}} \{X\}$ is the probability of the event X over the probability space corresponding to $\hat{\theta}$.

In general (3) is computationally intractable and non-convex so further relaxations are required. The starting point for the considerations in this paper is the second order approximation

$$\begin{aligned} V_{app}(\theta) &\approx V_{app}(\theta_0) + V'_{app}(\theta_0)(\theta - \theta_0) \\ &\quad + \frac{1}{2}(\theta - \theta_0)^T V''_{app}(\theta_0)(\theta - \theta_0) \\ &= \frac{1}{2}(\theta - \theta_0)^T V''_{app}(\theta_0)(\theta - \theta_0). \end{aligned} \tag{4}$$

In the context of (3), this approximation is motivated for γ sufficiently large, i.e. when there are sufficiently high demands on keeping the performance degradation low.

Notice that V_{app} and its derivatives implicitly depend on the true system, which is unknown prior to the experiment; this issue will not be studied in this paper, but the interested reader can check [25, 7] and the references therein for several approaches to deal with this dependence. In the sequel, to simplify notation, the dependence of V_{app} and its derivatives on θ_0 will be suppressed.

Inserting (4) in the constraint in (3) gives the approximate constraint

$$\mathbf{P}_{\hat{\theta}} \left\{ \frac{1}{2}(\hat{\theta} - \theta_0)^T V''_{app}(\hat{\theta} - \theta_0) \geq \frac{1}{\gamma} \right\} \leq \varepsilon. \tag{5}$$

Unfortunately, it is still very difficult to evaluate the probability in question in (5) and further relaxations are necessary to obtain a tractable problem. In [15] a relaxation based on a standard confidence ellipsoid for $\hat{\theta}$ is used (see Section II-A for details). This leads to a semi-definite program. The objective of this paper is to improve on this relaxation of the chance constraint (5), while still maintaining attractive computational properties. Our approach employs an optimized Chernoff bound, adapted from the robust convex optimization literature. The relaxation is introduced in Section II. A numerical study is presented in Section III and conclusions are provided in Section IV.

II. Approximations

In this section, several convex approximations of (5) are considered. The first one comes from [16], while the others are based on results from [22, 23].

A. Confidence Ellipsoids

A first attempt to relax (5) consists in noting that $\hat{\theta} \sim \mathbf{N}(\theta_0, P)$ implies that $(\hat{\theta}-\theta_0)^T P^{-1}(\hat{\theta}-\theta_0) \sim \chi^2_n$. This, in turn, means that the event $(\hat{\theta}-\theta_0)^T P^{-1}(\hat{\theta}-\theta_0) \geq \chi^2_n(\varepsilon)$ occurs[1] with probability ε; notice that this event corresponds to the exterior of the standard confidence ellipsoid for $\hat{\theta}$. Therefore, a sufficient condition for (5) to hold is to ensure that

$$\frac{1}{2}(\hat{\theta}-\theta_0)^T V''_{app}(\hat{\theta}-\theta_0) \leq \frac{1}{\gamma}, \quad \text{for every } \hat{\theta} \in \mathbb{R}^n \text{ such that } (\hat{\theta}-\theta_0)^T P^{-1}(\hat{\theta}-\theta_0) \leq \chi^2_n(\varepsilon), \quad (6)$$

or, letting $\tilde{\theta} := \hat{\theta} - \theta_0$,

$$\begin{bmatrix} \tilde{\theta} \\ 1 \end{bmatrix}^T \begin{bmatrix} -\frac{1}{2}V''_{app} & 0 \\ 0 & \frac{1}{\gamma} \end{bmatrix} \begin{bmatrix} \tilde{\theta} \\ 1 \end{bmatrix} \geq 0, \quad \text{for every } \tilde{\theta} \in \mathbb{R}^n \text{ such that } \begin{bmatrix} \tilde{\theta} \\ 1 \end{bmatrix}^T \begin{bmatrix} -P^{-1} & 0 \\ 0 & \chi^2_n(\varepsilon) \end{bmatrix} \begin{bmatrix} \tilde{\theta} \\ 1 \end{bmatrix} \geq 0. \quad (7)$$

This constraint can be simplified by applying the S-procedure [4], which states that (7) is equivalent to the existence of a $t \geq 0$ such that

$$\begin{bmatrix} -\frac{1}{2}V''_{app} & 0 \\ 0 & \frac{1}{\gamma} \end{bmatrix} - t \begin{bmatrix} -P^{-1} & 0 \\ 0 & \chi^2_n(\varepsilon) \end{bmatrix} \geq 0. \quad (8)$$

This condition is composed of two inequalities. The second one is satisfied if and only if $t \leq [\gamma\chi^2_n(\varepsilon)]^{-1}$. On the other hand, the first inequality is more strict as t decreases, so we can take $t = [\gamma\chi^2_n(\varepsilon)]^{-1}$. This means that (8) is equivalent to

$$-\frac{1}{2}V''_{app} + \frac{1}{\gamma\chi^2_n(\varepsilon)}P^{-1} \geq 0,$$

or

$$\frac{\gamma\chi^2_n(\varepsilon)}{2}V''_{app} \leq P^{-1}. \quad (9)$$

Notice that (9) is a convex (actually, a linear matrix inequality (LMI)) constraint in P^{-1}. As P^{-1} is typically an affine function of the input spectrum Φ_u [17], constraint (9) can be directly used as part of a semidefinite program for optimal input design. The reader is referred e.g. to [17] for further implementation details.

B. Markov Bounds for Finite-Order Models

A second technique to approximate (5) is to use Markov's bound. This gives

$$\begin{aligned} \mathbf{P}_{\hat{\theta}} \left\{ \frac{1}{2}(\hat{\theta}-\theta_0)^T V''_{app}(\hat{\theta}-\theta_0) \geq \frac{1}{\gamma} \right\} &\leq \gamma \mathbf{E}_{\hat{\theta}} \left\{ \frac{1}{2}(\hat{\theta}-\theta_0)^T V''_{app}(\hat{\theta}-\theta_0) \right\} \\ &= \frac{\gamma}{2}\mathrm{Tr}\,[V''_{app}\mathbf{E}_{\hat{\theta}}\{(\hat{\theta}-\theta_0)(\hat{\theta}-\theta_0)^T\}] \quad (10) \\ &= \frac{\gamma}{2}\mathrm{Tr}\,[V''_{app}P]. \end{aligned}$$

[1] $\chi^2_n(\varepsilon)$ is the ε percentile of a χ^2 distribution with n degrees of freedom (denoted as χ^2_n), i.e., $\mathbf{P}_x\{x \geq \chi^2_n(\varepsilon)\} = \varepsilon$ if $x \sim \chi^2_n$.

The first line follows from Markov's inequality: If x is a nonnegative random variable and $T \geq 0$ is a constant, then

$$\begin{aligned} \mathbf{E}\{x\} &= \int_0^\infty x\, d\mathbf{P}(x) \geq \int_T^\infty x\, d\mathbf{P}(x) \geq T\int_T^\infty d\mathbf{P}(x) \\ &= T\mathbf{P}\{x \geq T\}. \end{aligned} \quad (11)$$

Bound (10) suggests that (5) will be satisfied if the following constraint holds:

$$\frac{\gamma}{2}\mathrm{Tr}\,[V''_{app}P] \leq \varepsilon. \quad (12)$$

This condition can also be written as an LMI in P^{-1}, by using Schur complements [4]:

$$\mathrm{Tr}\,[M] \leq \frac{2\varepsilon}{\gamma}, \quad \begin{bmatrix} M & V''^{1/2}_{app} \\ V''^{T/2}_{app} & P^{-1} \end{bmatrix} \geq 0. \quad (13)$$

Here, $M \in \mathbb{R}^{n\times n}$ is an auxiliary (free) matrix, and $V''_{app} = V''^{T/2}_{app}V''^{1/2}_{app}$.

C. Markov Bounds for High-Order Models

This is an approach developed by Ljung and collaborators during the 1980's, [19]. To simplify the description of this technique, let us assume that the model structure is of a Finite Impulse Response (FIR) type, e.g.,

$$y_t = \theta_1 u_{t-1} + \cdots + \theta_n u_{t-n} + e_t,$$

where $\{e_t\}$ is a Gaussian white noise sequence of zero mean and variance σ^2. Furthermore, we will assume that the performance degradation function has the form

$$V_{app}(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \left| G(e^{j\omega}, \hat{\theta}) - G(e^{j\omega}, \theta_0) \right|^2 C(e^{j\omega}) d\omega, \quad (14)$$

where $G(q,\theta) := \theta_1 q^{-1} + \cdots + \theta_n q^{-n}$, and $C : \mathbb{T} \to \mathbb{R}^+_0$ ($\mathbb{T} := \{z \in \mathbb{C} : \ |z| = 1\}$) is an arbitrary non-negative continuous weighting function. Notice that $V_{app}(\theta)$ is quadratic in θ, hence $V_{app}(\theta) = (1/2)(\hat{\theta}-\theta_0)^T V''_{app}(\hat{\theta}-\theta_0)$. In addition, if n is large enough, Ljung and collaborators [19] showed that

$$\mathrm{Var}\,\{G(e^{j\omega}, \hat{\theta})\} \approx \frac{n\sigma^2}{N\Phi_u(\omega)},$$

where Φ_u is the spectrum of the input u, and N is the number of samples used to compute $\hat{\theta}$. Now, following the same path as in Section II-B, we have that

$$\begin{aligned} &\mathbf{P}_{\hat{\theta}} \left\{ V_{app}(\hat{\theta}) \geq \frac{1}{\gamma} \right\} \\ &\leq \gamma \mathbf{E}_{\hat{\theta}}\{V_{app}(\hat{\theta})\} \\ &= \frac{\gamma}{2\pi}\int_{-\pi}^{\pi} \mathbf{E}_{\hat{\theta}} \left\{ \left| G(e^{j\omega}, \hat{\theta}) - G(e^{j\omega}, \theta_0) \right|^2 \right\} C(e^{j\omega}) d\omega \\ &\approx \frac{n\gamma\sigma^2}{2\pi N}\int_{-\pi}^{\pi} \frac{C(e^{j\omega})}{\Phi_u(\omega)} d\omega. \end{aligned}$$

Therefore, constraint (5) is satisfied if the following condition holds:

$$\frac{n\gamma\sigma^2}{2\pi N}\int_{-\pi}^{\pi} \frac{C(e^{j\omega})}{\Phi_u(\omega)} d\omega \leq \varepsilon.$$

This condition is a large-n approximation of (12), and it can be used, as shown e.g. in [20], to find explicit expressions for some optimal input design problems. For example, the input of minimum energy that satisfies this condition has a spectrum proportional to $\sqrt{C(e^{j\omega})}$ [20, Equation (13.79)].

D. Chernoff Relaxations

Yet another way to relax constraint (5) is to replace it by a Chernoff or Bernstein type of bound, as done in [22]. To this end, we can perform the following steps:

$$\begin{aligned}
&\mathbf{P}_{\hat{\theta}}\left\{\frac{1}{2}(\hat{\theta}-\theta_0)^T V''_{app}(\hat{\theta}-\theta_0)\geq\frac{1}{\gamma}\right\}\\
&\quad=\mathbf{P}_x\left\{x^T P^{T/2}V''_{app}P^{1/2}x\geq\frac{2}{\gamma}\right\}\\
&\quad=\mathbf{P}_{\tilde{x}}\left\{\sum_{i=1}^n\lambda_i(P^{T/2}V''_{app}P^{1/2})\tilde{x}_i^2\geq\frac{2}{\gamma}\right\}\\
&\quad=\mathbf{P}_{\tilde{x}}\left\{\sum_{i=1}^n\frac{1}{t}\lambda_i(P^{T/2}V''_{app}P^{1/2})\tilde{x}_i^2-\frac{2}{\gamma t}\geq 0\right\} \quad (15)\\
&\quad=\mathbf{P}_{\tilde{x}}\left\{\exp\left[\sum_{i=1}^n\frac{1}{t}\lambda_i(P^{T/2}V''_{app}P^{1/2})\tilde{x}_i^2-\frac{2}{\gamma t}\right]\geq 1\right\}\\
&\quad\leq\mathbf{E}_{\tilde{x}}\left\{\exp\left[\sum_{i=1}^n\frac{1}{t}\lambda_i(P^{T/2}V''_{app}P^{1/2})\tilde{x}_i^2-\frac{2}{\gamma t}\right]\right\},
\end{aligned}$$

where $x:=P^{1/2}(\hat{\theta}-\theta_0)\sim\mathbf{N}(0,I)$, with $P=P^{T/2}P^{1/2}$; $\tilde{x}\sim\mathbf{N}(0,I)$ is another Gaussian random vector (corresponding to a suitable rotation of x); $t\geq 0$ is an arbitrary constant. The last line in (15) follows from Markov's inequality (11).

On the other hand, if $y\sim\mathbf{N}(0,1)$ and t is a real constant, then

$$\begin{aligned}
\mathbf{E}\{\exp(ty^2)\}&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(ty^2)\exp\left(-\frac{1}{2}y^2\right)dy\\
&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}(1-2t)y^2\right]dy\\
&=\frac{1}{\sqrt{1-2t}} \quad (16)
\end{aligned}$$

for $t\in(-\infty,1/2)$.

Combining (15) and (16) gives

$$\begin{aligned}
&\mathbf{P}_{\hat{\theta}}\left\{\frac{1}{2}(\hat{\theta}-\theta_0)^T V''_{app}(\hat{\theta}-\theta_0)\geq\frac{1}{\gamma}\right\}\\
&\quad\leq\exp\left(-\frac{2}{\gamma t}\right)\prod_{i=1}^n\frac{1}{\sqrt{1-\frac{2}{t}\lambda_i(P^{T/2}V''_{app}P^{1/2})}} \quad (17)\\
&\quad=\exp\left(-\frac{2}{\gamma t}\right)\frac{1}{\sqrt{\det\left(I-\frac{2}{t}P^{T/2}V''_{app}P^{1/2}\right)}},
\end{aligned}$$

where t is allowed to take any value in $(2\lambda_{\max}(P^{T/2}V''_{app}P^{1/2}),\infty)$. Therefore, from (17) we have that a sufficient condition for (5) to hold is that

$$\exp\left(-\frac{2}{\gamma t}\right)\frac{1}{\sqrt{\det\left(I-\frac{2}{t}P^{T/2}V''_{app}P^{1/2}\right)}}\leq\varepsilon. \quad (18)$$

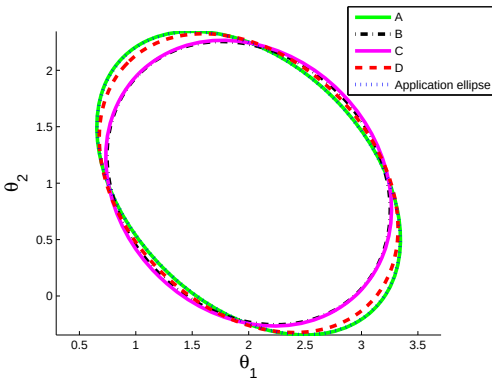

Fig. 1. Confidence ellipses (i.e., contour lines of the probability density function of $\hat{\theta}$) obtained using approaches A, B, C, and D, for a second order FIR system and performance degradation (14). The ellipse given by the confidence ellipsoids approach (A) coincides with the application ellipse.

In order to write (18) as a tractable convex constraint, we can take the logarithm of this expression and multiply it by t, yielding

$$-\frac{2}{\gamma}-\frac{t}{2}\ln\det\left(I-\frac{2}{t}P^{T/2}V''_{app}P^{1/2}\right)\leq t\ln\varepsilon,$$
$$\text{for some } t>2\lambda_{\max}(P^{T/2}V''_{app}P^{1/2}).$$

This constraint can be further simplified by using Schur complements [4] (and replacing t by $2t$), to obtain

$$\begin{aligned}
&-t\ln\varepsilon^2-t\ln\det\left(I-\frac{1}{t}M\right)\leq\frac{2}{\gamma}, \quad (19)\\
&\begin{bmatrix}M & V''^{1/2}_{app}\\ V''^{T/2}_{app} & P^{-1}\end{bmatrix}\geq 0,\qquad \begin{bmatrix}tI & V''^{1/2}_{app}\\ V''^{T/2}_{app} & P^{-1}\end{bmatrix}\geq 0.
\end{aligned}$$

Remark 2.1: Notice that the first equation in (19) is a convex constraint jointly in t and M, even though it cannot be written as a linear matrix inequality. However, some convex optimization packages such as CVX can handle these types of constraints for a fixed $t>0$ [11, 26]. Therefore, one possibility is to optimize for t by a line search.

Remark 2.2: Note that all of the proposed relaxations (except for the one presented in Section II-C, which depends on Φ_u) are convex constraints in P^{-1}, hence they can be used in principle to design optimal input signals in the frequency domain, providing a reasonable tractable approximation to the chance constraint (5). However, it is important to note that the resulting probability of violation of (5) could be much less than ε (because these relaxations are conservative); this means that in practice it could be worth to try several values of ε until the resulting probability of violation of (5) gets close to its maximum acceptable value (see Section 5 of [22]).

III. Numerical Examples

In this section, the approximations, described in Section II, are compared with respect to the input energy that they require in order to identify the unknown system under the chance constraint (5). The effect of an increasing model order is also studied. In all the examples, re-tuning of ε (c.f.

Method	$n=4$	$n=8$	$n=12$	$n=16$	$n=20$
A	0.939	1.421	1.854	2.264	2.650
B	0.924	1.364	1.759	2.150	2.468
C	0.900	1.326	1.705	2.031	2.411
D	0.883	1.320	1.694	2.023	2.353

TABLE I

INPUT ENERGIES REQUIRED BY APPROACHES A, B, C AND D. THE WEIGHTING FUNCTION $C(e^{j\omega})$ IS GIVEN BY (20).

Method	$n=4$	$n=6$	$n=10$	$n=15$	$n=25$
A	2.420	4.765	10.722	22.86	54.695
B	2.400	4.090	7.552	13.928	25.709
D	2.147	3.649	6.604	12.523	22.850

TABLE II

INPUT ENERGIES REQUIRED BY APPROACHES A, B AND D. THE MATRIX V''_{app} IS CHOSEN TOEPLITZ AND POSITIVE-SEMIDEFINITE, BUT OTHERWISE RANDOM.

Remark 2.2) is performed. To simplify the description, we will refer to the approaches by their respective sections (i.e., A, B, C or D).

Let us first consider the quadratic performance degradation function $V_{app}(\theta)$ given by eq. (14). The continuous weighting function C is chosen to be rational and to roughly represent a bandpass filter:

$$C(e^{j\omega}) = \frac{(e^{j\omega}-0.9)(e^{-j\omega}-0.9)}{(e^{2j\omega}-1.2e^{j\omega}+0.52)(e^{-2j\omega}-1.2e^{-j\omega}+0.52)}. \tag{20}$$

It is straightforward to check that V''_{app} in this case is a 2×2 Toeplitz matrix. In Fig. 1, the estimation ellipses for approaches A, B, C and D of the last section are presented, when the unknown system is FIR of order 2 (whose true parameters are $\theta_1=2$ and $\theta_2=1$). The level of the ellipses has been chosen such that the ellipse of approach A coincides with the application ellipse. The value of ε for the chance constraint in (3) is chosen to be $\varepsilon = 0.01$. We can see from the figure that the ellipses are relatively aligned with each other. In particular, the confidence ellipsoids approach, A, delivers an ellipse which perfectly coincides with the application ellipse. This implies that no re-tuning of ε was needed for such approach (c.f. Remark 2.2), because $\mathcal{E}_{app}$ then corresponds to a confidence ellipse for $\hat{\theta}$ (of confidence level ε), hence $\mathbf{P}_{\hat{\theta}}\{\hat{\theta}\in\mathcal{E}_{app}\}=\varepsilon$ (by definition of a confidence region). After approach A, the best fit to the application ellipse is given by the Chernoff relaxation, D. The input energy, r_0, required by the the approximations A, B, C, and D, is 0.651, 0.675, 0.628 and 0.606, respectively. Therefore, the Chernoff approach, D, gives the best input (in terms of experimental effort), even though the costs of the remaining approaches are not much higher.

Table I completes the picture about the approximations considered in this paper, where the results of simulations based on FIR models with increasing number of parameters n are shown[2] (considering (14) and (20)). We observe that, for this particular choice of the weighting function, the Chernoff approach, D, seems to give always the best results, although all methods deliver more or less the same performance, even for relatively large model orders.

It is interesting to notice that, even though approach A gives, for Toeplitz V''_{app} matrices, estimation ellipses that perfectly match the application ellipses, its performance is not so good in comparison to the other approaches. The reason for this is that the probability that $\hat{\theta}$ lies outside its estimation ellipsoid is not zero (since Gaussian distributions on, say, $\mathbb{R}^n$ have support over their entire domain). Therefore, in order to satisfy the required performance level with probability at least $1-\varepsilon$, it is not necessary that the estimation ellipsoid lies in the application ellipsoid; this latter condition is thus only sufficient to satisfy the chance constraint. This means that approaches B, C and D somehow manage to satisfy the chance constraint at a lower cost than approach A by distributing the input energy in a "smarter" way, so that the probability of $\hat{\theta}$ lying inside the application ellipse is $1-\varepsilon$.

Another remarkable observation from Table I is that approach C seems to give better results than approach B, even though the former is based on asymptotic (in n) variance expressions, while the latter relies of finite n formulas for the covariance of G. This suggests that the high model order approximation for the variance partially compensates for the slackness of the Markov inequality used for approach C. However, we have currently no detailed explanation for this phenomenon.

The superiority of the Chernoff approach, D, with respect to the others, at least for large n, can be informally argued from results from the theory of large deviations [5], according to which, as the number of parameters n increases, the left side of (5) should converge (under suitable conditions) to the Chernoff bound. Even though this is only a plausibility argument, Table I exhibits such behavior. To provide further evidence for the superiority of the Chernoff approach for large n, a new simulation study is performed, whose results are shown in Table II. In this table, we present the input energy required for approaches A, B, and D as the number of parameters n increases, while the matrix V''_{app} is chosen to be Toeplitz and positive semidefinite, but otherwise random[3]. The numerical results in this table show that as n increases, the Chernoff approximation prevails over approximations A and B. Approximation A worsens quickly as n increases, while the gap between the performances of B and D increases slowly but steadily with the number of parameters.

In order to see how the proposed approaches behave for non-Toeplitz Hessians V''_{app}, in Fig. 2 we present a second order example with a randomly selected positive definite non-Toeplitz application Hessian matrix V''_{app}. We observe here that all methods behave approximately in the same fashion, since their ellipsoids coincide. Fig. 2 is an example of an

[2]Notice that for FIR model structures and the performance criterion (14), (20), the optimal input does not depend on θ_o, hence as n is increased, it is irrelevant for this problem whether the true system is kept the same or if it is allowed to change, as long as it belongs to the model structure.

[3]In principle, approach C could be applied to this case by finding a spectrum $C(e^{j\omega})\geq 0$ whose associated covariance matrix is V''_{app}. However, the choice is in general non-unique, and different choices of this spectrum can give different optimal inputs, whose energies can differ. Therefore, we have preferred not to apply approach C to this situation.

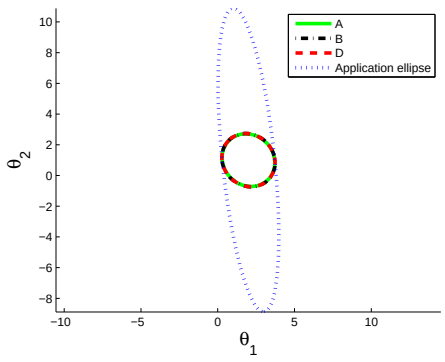

Fig. 2. Ellipses for A., B., D. approximations. Order-2 FIR system, V''_{app} given by (21).

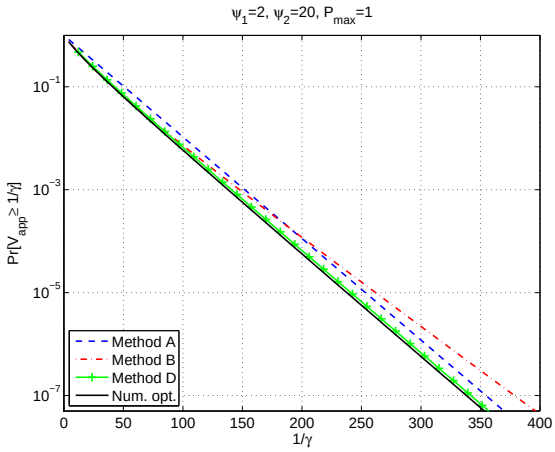

Fig. 3. Violation probability as a function of the quality level, γ, for the different solutions.

identification problem where the model structure imposes a heavy constraint on the optimal input design problem, since P^{-1} is restricted to be a Toeplitz matrix (this constrains the class of estimation ellipsoids which can be obtained by manipulating the input signal). In this case, the relaxations A, B and D cannot generate estimation ellipsoids sufficiently aligned with the application ellipsoid. The V''_{app} used is

$$V''_{app} = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 0.04 \end{bmatrix} \tag{21}$$

while the achieved input energies are 0.342, 0.337 and 0.336 for methods A, B and D, respectively.

Observing Figs. 1 and 2, one may conclude that all the approaches have more or less the same performance either for Toeplitz or non-Toeplitz application Hessian matrices when the number of parameters is small.

As a final example, to illustrate the geometry of some of the approximations, consider a model with $n = 2$, where

$$V''_{app} = \begin{bmatrix} \psi_1 & 0 \\ 0 & \psi_2 \end{bmatrix}, \qquad P^{-1} = \begin{bmatrix} \rho_1 & 0 \\ 0 & \rho_2 \end{bmatrix},$$

where the experimental effort is the total power $\rho_1 + \rho_2$. In order to enable a fair comparison between methods,

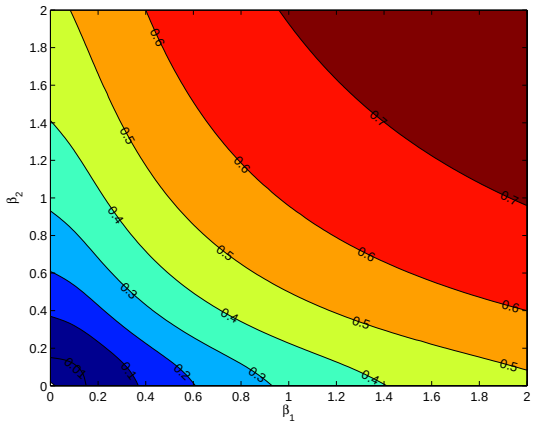

Fig. 4. Level curves of the true probability, for the diagonal $n = 2$ example, as a function of $\beta_k = \psi_k/2\rho_k$ for $\gamma = 1$.

consider the converse of (3), i.e., to maximize γ given a fixed power budget $\rho_1 + \rho_2 =: P_{\max} \leq 1$. It is easy to show that approach A gives the solution $\rho_k = \psi_k/(\psi_1+\psi_2)$ and approach B gives $\rho_k = \sqrt{\psi_k}/(\sqrt{\psi_1} + \sqrt{\psi_2})$. In this case, the converse of problem (3) can be solved numerically (i.e. without the approximations of Section II), computing the exact probabilities using the method of Helstrom and Rice [12], and employing non-linear optimization (`fmincon` in Matlab) to find the optimal choice of ρ_k minimizing ε for each value of γ. The resulting probabilities for varying γ are plotted in Fig. 3. Method D lies very close to the optimal approach (based on numerical optimization). For violation levels ε of the order of 0.01 or less, method B clearly outperforms method A. However, for violation levels below 10^{-4} method B clearly outperforms method A. A hint on why this is the case is provided in Figs. 4 and 5 which show the level curves of the true probability as a function of $\beta_k = \psi_k/2\rho_k$ for $\gamma = 1$. Fig. 4 shows that the problem is non-convex in general, but convex for small (i.e., useful) ε. For ε of the order of 0.1, the level curves are reasonably well approximated by a straight diagonal line, corresponding to the level curve of approximation A. However, as Fig. 5 reveals, for very low levels of ε, the error probability is instead dominated by the largest β_k, which is closer to what the Markov approximation delivers (i.e., method B).

IV. Conclusions

In this paper, we have studied the problem of designing an input signal of minimal energy such that the estimated model based on this signal can satisfy a given performance level with a prescribed probability. In order to solve this problem, several approaches have been proposed and compared.

Based on simulation studies, we have found that the method based on Chernoff bounds seems to be comparably better than the other ones when the number of parameters is large. For models of small dimension, all of the proposed techniques have performances within the same order of magnitude, even though the Chernoff approach is (at least slightly) superior in most of the cases we have studied. However, we have also seen that the relative performance of some of the methods depend on the acceptable level of

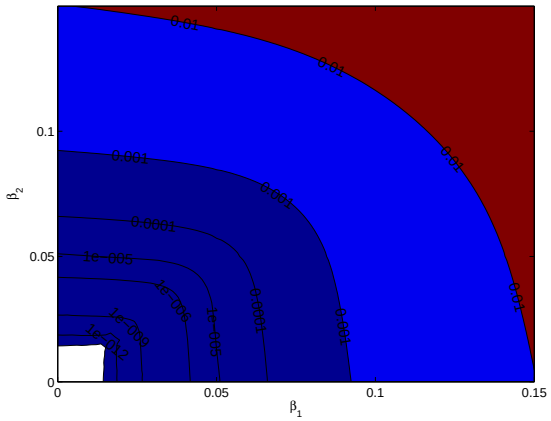

Fig. 5. Level curves of the true probability, for the diagonal $n = 2$ example, as a function of $\beta_k = \psi_k/2\rho_k$ for $\gamma = 1$.

the violation probability ε.

Under the light of our results we see that a theoretical analysis of the methods suggested here is an important future research direction, because in many situations, the standard input design techniques can be outperformed by the Chernoff approach, thus leading to suboptimal inputs for identification purposes.

We should also highlight the importance of re-tuning the probability bound ε, since in many cases the relaxation techniques used are quite conservative, suggesting an experimental effort significantly higher than what is actually required to achieve the desired probabilistic bound.

REFERENCES

[1] Agüero, J. C. and Goodwin, G. C. [2007]. Choosing between open- and closed-loop experiments in linear system identification, *IEEE Transactions on Automatic Control* **52(8)**: 1475–1480.

[2] Barenthin, M., Bombois, X., Hjalmarsson, H. and Scorletti, G. [2008]. Identification for control of multivariable systems: Controller validation and experiment design via LMIs, *Automatica* **44(12)**: 3070–3078.

[3] Bombois, X., Scorletti, G., Gevers, M., Van den Hof, P. and Hildebrand, R. [2006]. Least costly identification experiment for control, *Automatica* **42**(10): 1651–1662.

[4] Boyd, S., El Ghaoui, L., Feron, E. and Balakrishnan, V. [1994]. *Linear Matrix Inequalities in System and Control Theory*, SIAM Studies in Applied Mathematics.

[5] Dembo, A. and Zeitouni, O. [1998]. *Large Deviations Techniques and Applications, 2nd Edition*, Springer.

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

[7] Gerencsér, L. and Hjalmarsson, H. [2005]. Adaptive input design in system identification, *Proceedings of the 44th IEEE Conference on Decision and Control*, Seville, Spain.

[8] Gerencsér, L., Hjalmarsson, H. and Mårtensson, J. [2009]. Identification of ARX systems with non-stationary inputs - asymptotic analysis with application to adaptive input design, *Automatica* **45**(3): 623–633.

[9] Gevers, M. [2005]. Identification for control: From the early achievements to the revival of experiment design, *European Journal of Control* **11**: 1–18.

[10] Gevers, M. and Ljung, L. [1986]. Optimal experiment designs with respect to the intended model application, *Automatica* **22**(5): 543–554.

[11] Grant, M. C. and Boyd, S. P. [2008]. Graph implementations for nonsmooth convex programs, *in* V. D. Blondel, S. P. Boyd and H. Kimura (eds), *Recent Advances in Learning and Control (tribute to M. Vidyasagar)*, Springer-Verlag, pp. 95–110.

[12] Helstrom, C. W. [1983]. Comment: Distribution of quadratic forms in normal random variables—evaluation by numerical integration, *SIAM Journal on Scientific Computing* **4(2)**: 353–356.

[13] Hildebrand, R. and Gevers, M. [2003]. Identification for control: Optimal input design with respect to a worst-case ν-gap cost function, *SIAM J. Control Optim* **41**(5): 1586–1608.

[14] Hjalmarsson, H. [2005]. From experiment design to closed-loop control, *Automatica* **41**(3): 393–438.

[15] Hjalmarsson, H. [2009a]. System identification of complex and structured systems, *European Journal of Control* **15**(4): 275–310. Plenary address. European Control Conference.

[16] Hjalmarsson, H. [2009b]. System identification of complex and structured systems, *Proceedings of the European Control Conference*, Budapest, Hungary.

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

[18] Kaisare, N. and Lee, J. [2010]. Optimal design of periodic test input signals for multivariable impulse response models, *Optimal control applications & methods* **31**(5): 451–469.

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

[20] Ljung, L. [1999]. *System Identification: Theory for the User, 2nd Edition*, Prentice Hall.

[21] Ljung, L. [2010]. Perspectives in system identification, *Annual Reviews in Control* **34**(1): 1–12.

[22] Nemirovski, A. and Shapiro, A. [2006]. Convex approximations of chance constrained programs, *SIAM Journal of Optimization* **17(4)**: 969–996.

[23] Pintér, J. [1989]. Deterministic approximations of probability inequalities, *Methods and Models of Operations Research* **33**: 219–239.

[24] Pronzato, L. [2008]. Optimal experimental design and some related control problems, *Automatica* **44(2)**: 303–325.

[25] Rojas, C. R., Welsh, J. S., Goodwin, G. C. and Feuer, A. [2007]. Robust optimal experiment design for system identification, *Automatica* **43**(6): 993–1008.

[26] Vandenberghe, L., Boyd, S. and Wu, S.-P. [1998]. Determinant maximization with linear matrix inequality constraints, *SIAM Journal on Matrix Analysis and Applications* **19(2)**: 499–533.