Uncertainty in system identification: learning from the theory of risk ⋆

Patricio E. Valenzuela * **Cristian R. Rojas** *
Håkan Hjalmarsson *

* *Department of Automatic Control, KTH Royal Institute of Technology, SE-100 44, Stockholm, Sweden (e-mail: {pva, crro, hjalmars}@kth.se).*

Abstract: This article addresses the issue of measuring uncertainty in optimization problems arising in system identification. The issue of uncertainty has been studied in the theory of risk, where the results are mainly employed in finance applications. Here we explore how the results in the literature of theory of risk can be used to obtain a systematic approach to uncertainty in system identification. For concreteness, the discussion is illustrated by an application to input design, but it can be extended to other areas of the field.

Keywords: System identification, input design, theory of risk, uncertainty.

1. INTRODUCTION

Uncertainty is an issue common to many research areas. By uncertainty we understand the lack of knowledge to fully describe a phenomenon. The lack of knowledge causes severe difficulties when we are interested in determining the best decision with limited information. Examples of this problem can be found in control design, where the controller must be designed with limited information about the plant dynamics (Zhou and Doyle, 1998; Zhou et al., 1996), and in portfolio optimization, where the returns are maximized subject to limited information about the future evolution of the assets price (Bertsimas and Thiele, 2006; Black and Litterman, 1992; Krokhmal et al., 2002; Perold, 1984; Postek et al., 2014).

In the same line, many problems arising in system identification are solved with limited information. In system identification, the uncertainty can be understood as the lack of knowledge about the true dynamics of the process to be modeled. The uncertainty associated with the process dynamics is of importance in applications where the optimal decision depends on the true model description. This is the case in input design, where the optimal input sequence depends on true process dynamics (Ljung, 1999). Approaches to solve this issue have been presented in the literature, which can be classified in two classes: (i) sequential or adaptive procedures, where a new design is obtained based on the current estimates of the process dynamics (Gerencsér and Hjalmarsson, 2005; Gerencsér et al., 2009; Lindqvist and Hjalmarsson, 2001; Pronzato, 2000); and (ii) robust procedures, where the design is obtained by including the uncertainty in the optimization problem (Jansson and Hjalmarsson, 2005; Pronzato and Walter, 1985; Rojas et al., 2007). In this article, we are interested in addressing the uncertainty in input design by using the robust approach.

The robust approach to uncertainty in system identification has been analyzed in the literature, and several techniques have been proposed (Larsson et al., 2012; Mårtensson and Hjalmarsson, 2006; Rojas et al., 2012). The main idea behind these results is the inclusion of a mapping from the space of functions with uncertainty to a scalar value. The scalar value takes into account the uncertainty associated with the process dynamics. Some examples of the mappings employed are the expected value, and the supremum over the set of possible descriptions of the process dynamics. However, there is no analysis of how well the mappings address the issue of uncertainty in system identification. By *how well* we mean if the mapping is either a weak or a conservative measure of the uncertainty in the optimization problem.

The problem of properly measuring the uncertainty has been addressed in the theory of risk measures. In the theory of risk, the uncertainty is understood as the risk associated with the portfolios (Cramér, 1930). To properly measure the risk associated with the portfolios, the notion of coherent measure of risk has been introduced (Artzner et al., 1999). Some of the requirements for a functional to be a coherent measure of risk are the convexity and monotonicity, which imply that the resulting optimization problem is convex if the original problem with uncertainty is convex. In addition, a coherent measure of risk encourages diversification, i.e., it is always better to invest in several assets rather than in a single one, which is a property usually required by investors to reduce the risk associated with the portfolios.

In this article we explore the connection between uncertainty in the theory of risk and uncertainty in system identification. In particular, we discuss how the notion of a coherent measure of risk can be employed to obtain a systematic approach to uncertainty in system identification. In addition, we introduce a coherent measure of risk that can be useful in system identification: the conditional value at risk (CVaR) (Rockafellar and Uryasev, 2000). The usefulness of coherent risk measures will be illustrated by its application to input design. However, we emphasize that the applicability of this approach is not only limited to this topic, it can also help to address the uncertainty issue in other areas (e.g., a theory of risk approach has already been used in Markov control processes and model predictive control, see Chow and Pavone (2013, 2014); Shen et al. (2014)).

⋆ This work was supported by the Swedish Research Council under contracts 621-2011-5890 and 621-2009-4017, and by the European Research Council under the advanced grant LEARN, contract 267381.

The rest of this article is organized as follows. Section 2 presents the problem of uncertainty in system identification. Section 3 analyzes the problem of uncertainty from a theory of risk perspective. Section 4 explores the use of the conditional value at risk to measure uncertainty in system identification. Section 5 presents a numerical example. Finally, Section 6 concludes this article.

2. UNCERTAINTY IN SYSTEM IDENTIFICATION

To illustrate the discussion in this article, we consider a nonlinear state-space model with states $x_{1:T} := \{x_t\}_{t=1}^{T}$ ($x_t \in \mathbb{R}^{n_x}$), inputs $u_{1:T} := \{u_t\}_{t=1}^{T}$ ($u_t \in \mathbb{R}^{n_u}$), and measurements $y_{1:T} := \{y_t\}_{t=1}^{T}$ ($y_t \in \mathbb{R}^{n_y}$) given by

$$x_t|x_{t-1} \sim f_\theta(x_t|x_{t-1}, u_{t-1}), \tag{1a}$$

$$y_t|x_t \sim g_\theta(y_t|x_t, u_t), \tag{1b}$$

$$x_0 \sim \mu_\theta(x_0), \tag{1c}$$

where $f_\theta(\cdot)$, $g_\theta(\cdot)$, and $\mu_\theta(\cdot)$ denote known probability distributions parameterized by $\theta \in \Theta \subset \mathbb{R}^d$. We assume that there exists a $\theta_0 \in \Theta$ such that (1) describes the true system when $\theta = \theta_0$, i.e., there is no undermodelling (Ljung, 1999).

The main objective in system identification is to estimate the model parameters θ from the collected input-output data $(u_{1:N}, y_{1:N})$ (Ljung, 1999). Since the estimated model parameter relies on the input-output data, it is typically required that the data must provide as much information from the process as possible, which implies that the model parameters can be estimated with maximum accuracy for a given experiment length N. However, as we discuss below, this requirement depends on the true model parameters θ_0, which are uncertain prior to performing an experiment.

2.1 Input design

A standard approach to maximize the accuracy of the estimated model (1) is by optimizing a scalar function of the Fisher information matrix $\mathcal{I}_F^{\theta_0}$ (Goodwin and Payne, 1977; Ljung, 1999). The Fisher information matrix is defined as

$$\mathcal{I}_F^{\theta_0} := \mathbf{E}\left\{\mathcal{S}(\theta_0)\mathcal{S}^\top(\theta_0)\right\}, \tag{2a}$$

$$\mathcal{S}(\theta_0) := \nabla_\theta \log \ell_\theta(y_{1:N})|_{\theta=\theta_0}, \tag{2b}$$

where ∇_θ denotes the gradient operator with respect to θ, and $\ell_\theta(y_{1:N})$, $\mathcal{S}(\theta)$ denote the likelihood function and the score function, respectively. We note that the expected value in (2a) is over the stochastic processes in (1).

The input design problem is to find an input sequence $u_{1:N} \in \mathcal{C} \subset \mathbb{R}^N$ which optimizes a scalar function of (2a), where $\mathcal{C}$ denotes the set of feasible values for $u_{1:N}$ (Valenzuela, 2014; Valenzuela et al., 2014). The scalar function is given by $h : \mathbb{R}^{d\times d} \to \mathbb{R}$. To properly quantify $\mathcal{I}_F^{\theta_0}$, the function h must satisfy $h(A) \leq h(B)$ for any two matrices A and B in the positive semidefinite cone satisfying $A \succeq B$. Hence, we assume that h is a convex, matrix nonincreasing function (Boyd and Vandenberghe, 2004, pp. 108). Different choices of h have been proposed in the literature, see e.g. Rojas et al. (2007); some examples are $h = -\log\det$, and $h = \mathrm{tr}\{(\cdot)^{-1}\}$. In this work, we leave the selection of h to the user.

The problem described can be summarized as

Problem 1. Design an optimal input signal $u_{1:N}^{\mathrm{opt}}$, where

$$u_{1:N}^{\mathrm{opt}} = \arg\min_{u_{1:N}\in\mathcal{C}} h(\mathcal{I}_F^{\theta_0}(u_{1:N})), \tag{3}$$

with $h\colon \mathbb{R}^{d\times d} \to \mathbb{R}$ a convex, matrix nonincreasing function, and $\mathcal{I}_F^{\theta_0} \in \mathbb{R}^{d\times d}$ defined as in (2).

As we can see from (2), the main difficulty to solve Problem 1 is that the input design relies on the knowledge of the true parameters θ_0, which are to be estimated using the excitation to be designed. To solve this problem, a robust input design scheme has been proposed, where the input sequence is designed by incorporating the uncertainty on the model parameters into the optimization problem. However, the approach to overcome this issue has only considered the definition of ad-hoc functions (e.g. the expected value of $h(\mathcal{I}_F^{\theta_0})$, or the supremum of $h(\mathcal{I}_F^{\theta_0})$ over Θ), and there is no systematic approach to characterize these functions. Therefore, there is a need for characterizing the set of functions that can be employed to measure the uncertainty in input design.

2.2 Application oriented input design

Another alternative to guarantee a prescribed accuracy in the model parameters is by imposing a quality constraint over the estimated parameters, while we minimize the experimental effort. This is the approach in least-costly and application oriented input design (Bombois et al., 2006; Hjalmarsson, 2009; Larsson et al., 2013, 2011).

To measure the quality of the estimated model parameters, we consider a function $J\colon \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{C} \to \mathbb{R}$, which assesses the performance degradation when a particular parameter is employed in the model application, and it is compared with the performance achieved by the true description. The purpose of this cost function is that it incorporates the intended model application into the input design problem.

On the other hand, the designed input sequence must minimize the experimental effort required to fulfill a specified quality constraint on the estimated model parameters. To this end, we introduce a function $H\colon \mathcal{C} \to \mathbb{R}$, which quantifies the required effort for a particular experiment $u_{1:N}$. We will assume that the function H is convex.

The application oriented input design problem can be summarized as

Problem 2. Design an optimal input signal $u_{1:N}^{\mathrm{opt}}$, where

$$\begin{aligned} u_{1:N}^{\mathrm{opt}} = \arg\min_{u_{1:N}\in\mathcal{C}} \quad & H(u_{1:N}) \\ \text{subject to} \quad & J(\theta_0, \theta, u_{1:N}) \leq 0, \end{aligned} \tag{4}$$

with $H\colon \mathcal{C} \to \mathbb{R}$ a convex function, and $J\colon \mathbb{R}^d\times\mathbb{R}^d\times\mathcal{C} \to \mathbb{R}$.

We note that, without loss of generality, Problem 2 considers a nonpositive constraint over J.

As with Problem 1, Problem 2 has the difficulty that the optimal solution depends on the true model parameters θ_0. Therefore, an exact solution to Problem 2 cannot be achieved in practical applications, and a method to include the uncertainty in the optimization problem is required.

3. A RISK THEORETICAL APPROACH TO UNCERTAINTY

Measuring uncertainty is one of the challenges to be addressed in system identification. There are results in the literature providing approaches to uncertainty in different problems. However, the problem of measuring uncertainty in system identification needs a systematic approach providing a solid foundation to incorporate the uncertainty into the optimization problem.

In this section we analyze the problem of uncertainty from a risk theoretical perspective. Quoting Rockafellar (2007) "*risk is associated with having to make a decision without fully knowing its consequences*". The objective in risk theory is to determine the decision minimizing a cost under risk. The definition of cost can be very general, and it is not tied to a particular application (Rockafellar, 2007). We note that the notion of risk is also employed in statistical decision theory (Berger, 2013), but its definition differs from the notion of risk considered here.

The cost $c\colon \mathbb{S} \times \Omega \to \mathbb{R}$ associates to each action $s \in \mathbb{S}$ (determined by the user) the value $c(s, \omega) \in \mathbb{R}$, where $\omega \in \Omega$ is a realization from the uncertainty associated with the cost. The set Ω is assumed to be a probability space with probability measure P, where P must reflect the lack of knowledge or prior information regarding ω. Since c establishes a map from Ω to $\mathbb{R}$, we can interpret the cost as a random variable.

Example 1. Consider Problem 1. The function h can be seen as the cost c, where $\mathbb{S} = \mathcal{C}$, and $\Omega = \Theta$. Here, we assume that Θ is a probability space with probability measure P_θ.

Example 2. Consider Problem 2. The function J can be also seen as the cost c, where $\mathbb{S} = \mathcal{C}$, and $\Omega = \Theta \times \Theta$. Here, it is assumed that $\Theta \times \Theta$ is a probability space with probability measure $P_{\theta\times\theta}$.

To continue, we let $\mathcal{X}$ denote a random variable on Ω for which its mean and standard deviation

$$\mu\{\mathcal{X}\} := \mathbf{E}\{\mathcal{X}\}\,, \tag{5a}$$

$$\sigma\{\mathcal{X}\} := \mathbf{E}\{(\mathcal{X} - \mu(\mathcal{X}))^2\}^{1/2}\,, \tag{5b}$$

are well defined. The expected values in (5) are with respect to the probability measure P. We will denote by $\mathcal{L}^2$ the set of functions $\mathcal{X}$ such that (5) are bounded. Here, we will assume that the function c is in $\mathcal{L}^2$ for every $s \in \mathbb{S}$.

Example 3. Consider Problem 1. The function h is in $\mathcal{L}^2$ provided that $\mu\{h(\mathcal{I}_F^{\theta_0})\}$ and $\sigma\{h(\mathcal{I}_F^{\theta_0})\}$ are bounded for every $u_{1:N} \in \mathcal{C}$. In this case, the expected values in (5) are with respect to the probability measure P_θ.

The task is to measure the risk of loss associated with $\mathcal{X}$. To this end, for every $\mathcal{X} \in \mathcal{L}^2$ we associate a value $\mathcal{R}(\mathcal{X})$, where $\mathcal{R}\colon \mathcal{L}^2 \to (-\infty, \infty]$ is a functional. We note that $\mathcal{R}$ is allowed to take the value ∞.

A question that arises at this point is which properties $\mathcal{R}$ must satisfy to properly measure the risk associated with $\mathcal{X}$. This issue has been addressed in the literature of theory of risk, where the notion of coherent measure of risk has been developed (Artzner et al., 1999). Its definition is presented below (Rockafellar, 2007):

Definition 1. A functional $\mathcal{R}\colon \mathcal{L}^2 \to (-\infty, \infty]$ is a *coherent measure of risk in the extended sense* if and only if

(i) $\mathcal{R}(C) = C$ for all constant functions C,

(ii) for $\mathcal{X}_1$, $\mathcal{X}_2 \in \mathcal{L}^2$ and $\lambda \in (0, 1)$,

$$\mathcal{R}((1-\lambda)\mathcal{X}_1 + \lambda\mathcal{X}_2) \leq (1-\lambda)\mathcal{R}(\mathcal{X}_1) + \lambda\mathcal{R}(\mathcal{X}_2)\,,$$

(iii) $\mathcal{R}(\mathcal{X}_1) \leq \mathcal{R}(\mathcal{X}_2)$ when $\mathcal{X}_1 \leq \mathcal{X}_2$,

(iv) $\mathcal{R}(\mathcal{X}) \leq 0$ when $\|\mathcal{X}^k - \mathcal{X}\|_2 \to 0$ with $\mathcal{R}(\mathcal{X}^k) \leq 0$ for all k.

The functional $\mathcal{R}$ will be called a *coherent measure of risk in the basic sense* if and only if it also satisfies

(v) $\mathcal{R}(\lambda\mathcal{X}) = \lambda\mathcal{R}(\mathcal{X})$ for $\lambda > 0$.

Definition 1 provides a systematic approach to measures of risk. Property (i) is motivated by the fact that, if a random variable has a constant outcome C, the result of measuring its risk must be C. Property (ii) requires that $\mathcal{R}$ must be a convex functional, which follows from the fact that the risk associated with the convex combination of any two random variables must be always less or equal than the convex combination of the risk associated with each random variable separately. Property (iii) requires that $\mathcal{R}$ must associate a higher risk to those random variables having a higher cost. Property (iv) requires the closedness of $\mathcal{R}$. Finally, (v) requires $\mathcal{R}$ to be positive homogeneous.

If $\mathcal{R}$ is a coherent measure of risk in the basic sense, then by combining (ii) and (v) we obtain the subadditivity property: for all $\mathcal{X}_1$, $\mathcal{X}_2 \in \mathcal{L}^2$,

$$\mathcal{R}(\mathcal{X}_1 + \mathcal{X}_2) \leq \mathcal{R}(\mathcal{X}_1) + \mathcal{R}(\mathcal{X}_2)\,. \tag{6}$$

Equation (6) is an important result: it says that the risk obtained by combining two random variables is less than the sum of the risk associated to each random variable separately, which leads to diversification. In other words, it is always better to consider the sum of random variables rather than each random variable separately, since the former will have a lower risk.

From the perspective of Definition 1, we can analyze the approaches employed in system identification to measure uncertainty. It is easy to show that $\sup\mathcal{X}$ and $\mathbf{E}\{\mathcal{X}\}$ are coherent measures of risk in the basic sense, but the α-quantile of the distribution of $\mathcal{X}$ ($\alpha \in (0, 1)$), defined as

$$q_\alpha(\mathcal{X}) := \min\{z \in \mathbb{R}\colon \mathbf{P}\{\mathcal{X}(\omega) \leq z\} \geq \alpha\}\,, \tag{7}$$

is not a coherent measure of risk (property (iii) in Definition 1 is not satisfied). Notice that chance constrained optimization (Pázman and Pronzato, 2007; Rojas et al., 2011) corresponds to constraining q_α. Hence, chance constrained optimization problems are incoherent.

Even though $\sup\mathcal{X}$ and $\mathbf{E}\{\mathcal{X}\}$ are coherent measures of risk in the basic sense, they have disadvantages. In the case of $\sup\mathcal{X}$, it can be infinity when $\mathcal{X}$ does not have a bounded support (Pronzato and Pázman, 2013, Chapter 8). On the other hand, $\mathbf{E}\{\mathcal{X}\}$ is a weak measure of risk, since it only imposes a requirement on average, and it can lead to realizations of $\mathcal{X}$ with poor results. In recent years, a new coherent measure of risk in the basic sense has been proposed: the conditional value at risk. Its definition is given in the next subsection.

3.1 Conditional value at risk

The notion of conditional value at risk has been proposed in Artzner et al. (1997). Its definition is as follows. Given $\beta \in (0, 1)$, the β-conditional value at risk is given by

$$\mathrm{CVaR}_\beta(\mathcal{X}) := \frac{1}{1-\beta}\int_{\mathcal{X}(\omega)\geq\alpha_\beta(\mathcal{X})} \mathcal{X}(\omega)dP(\omega)\,, \tag{8}$$

where

$$\alpha_\beta(\mathcal{X}) := q_\beta(\mathcal{X})\,. \tag{9}$$

Equation (9) is referred in the literature of theory of risk as value at risk, and it has been widely employed in this area (Favre and Galeano, 2002; Gaivoronski and Pflug, 2005; Ghaoui et al., 2003). The value at risk (9) is understood as the value of $\alpha \in \mathbb{R}$ such that $\mathcal{X}(\omega) \leq \alpha$ is satisfied with probability β, i.e., it is the β-quantile associated to the distribution of $\mathcal{X}$. However, it has been proved in Artzner et al. (1997) that value at risk is not a coherent measure of risk, which implies that it is not suitable for optimizing decisions under risk.

The definition of conditional value at risk in (8) is the expected value of $\mathcal{X}$ with respect to the conditional distribution of its upper α_β-tail. Therefore, the minimization of (8) does not only guarantee that $\mathcal{X}^{\text{opt}}(\omega) \leq \alpha_\beta(\mathcal{X}^{\text{opt}})$ with probability β, but it also guarantees that, with probability $1-\beta$, the mean value of the loss will be $\text{CVaR}_\beta(\mathcal{X}^{\text{opt}})$. In addition, it has been shown that (8) is a convex and monotone function (Pflug, 2000), which is a property of coherent measures of risk in the basic sense. We note that the conditional value at risk (8) differs from the notion of mean excess function employed in the literature of theory of risk. Indeed, Equation (8) is computed based on the β-quantile of $\mathcal{X}$, while the mean excess function computes the expected value of $\mathcal{X}(\omega) - d$ given $\mathcal{X}(\omega) \geq d$, where $d \in \mathbb{R}$. Hence, the mean excess function is a less useful measure of risk than (8) in this context.

From Equation (8) we see that the optimization of the conditional value at risk requires the user defined parameter β. This parameter is chosen according to the confidence level desired by the user (typically $\beta = 0.98$ in finance applications).

On the other hand, the optimization of (8) requires to solve (9), which can be difficult to compute. However, Rockafellar and Uryasev (2000) showed that, under the assumption that P is a continuous function, it is possible to circumvent this issue by computing the conditional value at risk as

$$\text{CVaR}_\beta(\mathcal{X}) = \min_{C \in \mathbb{R}} \left[C + \frac{1}{1-\beta} \mathbf{E}\left\{ \max\left(0,\, \mathcal{X} - C\right) \right\} \right]. \tag{10}$$

It has been shown in Rockafellar and Uryasev (2000) that the bracketed expression in (10) is continuously differentiable with respect to C, which makes $\text{CVaR}_\beta(\mathcal{X})$ easy to minimize numerically. Moreover, they showed that the value of C minimizing (10) equals the value at risk (9), which implies that the β-quantile can be obtained without solving the nonconvex optimization (7). Finally, if it is assumed that samples $\{\omega_i\}_{i=1}^{N_{\text{sim}}}$ can be drawn from the distribution P, then the expected value in (10) can be approximated as

$$\mathbf{E}\left\{ \max\left(0,\, \mathcal{X} - C\right) \right\} \approx \frac{1}{N_{\text{sim}}} \sum_{i=1}^{N_{\text{sim}}} \max\left(0,\, \mathcal{X}(\omega_i) - C\right). \tag{11}$$

Therefore, Equation (11) can be used into (10) to obtain an approximate value for the conditional value at risk, when the expected value in (10) is difficult to compute explicitly.

4. MEASURING RISK IN SYSTEM IDENTIFICATION

In this section we illustrate how the conditional value at risk can be used to measure uncertainty in system identification. To this end, we notice that the conditional value at risk is a convex, and monotone function. Therefore, the composition of a convex function with the conditional value at risk results in a convex function (Rockafellar, 1970, pp. 32). As a consequence, the conditional value at risk can be used to obtain robust optimization problems that are convex, provided that the original problem is convex in the decision variables.

4.1 Robust input design

One fundamental assumption to solve the input design problem is that prior information on the model parameters is available. One alternative to relax this assumption is by removing the dependence of the design with respect to the model parameters, which is the focus of robust input design. However, as discussed in Section 3, the common choices to measure uncertainty are either very conservative or weak.

Here we propose the use of conditional value at risk to obtain a robust input design problem. To this end, we assume that a probability density function of the model parameters is available to the user, denoted by p_θ. We note that the assumption is close to the requirement in Bayesian techniques in the sense that a prior information on the model parameters is taken into account in the optimization problem.

Following the discussion in Section 3, we can interpret h as a cost in $\mathcal{L}^2$ with respect to the probability measure given by p_θ. Thus, we can use the conditional value at risk to obtain a robust formulation of the input design problem as

$$u_{1:N}^{\text{opt}} = \arg \min_{u_{1:N} \in \mathcal{C}} \text{CVaR}_\beta(h(\mathcal{I}_F^{\theta_0}(u_{1:N}))), \tag{12}$$

where $\beta \in (0,1)$ is a parameter defined by the user. The intuition by minimizing (12) is as follows: the mean value of $h(\mathcal{I}_F^{\theta_0})$ associated with its $1-\beta$ tail distribution will be equal to $\text{CVaR}_\beta(h(\mathcal{I}_F^{\theta_0}(u_{1:N}^{\text{opt}})))$.

The optimization problem (12) is solved by employing (10), where $C \in \mathbb{R}$ is an additional decision variable. The resulting optimization is

$$u_{1:N}^{\text{opt}} = \arg \min_{\substack{u_{1:N} \in \mathcal{C} \\ C \in \mathbb{R}}} C + \frac{1}{1-\beta} \mathbf{E}\{\max(0, h(\mathcal{I}_F^{\theta_0}(u_{1:N})) - C)\}, \tag{13}$$

where the expected value in (13) is with respect to p_θ. If the original problem (3) is convex in the decision variables, then the optimization problem (13) is also convex.

4.2 Robust application oriented input design

The application oriented input design problem also suffers from the uncertainty in the true model parameters. To solve this issue, we use the theory of risk approach to obtain a robust version of Problem 2. Under the assumption that the uncertainty in the true model parameters is modeled by a probability density function $p_{\theta \times \theta}$, the robust application oriented input design problem can be written as

$$\begin{aligned} u_{1:N}^{\text{opt}} = \arg \min_{u_{1:N} \in \mathcal{C}} \quad & H(u_{1:N}) \\ \text{subject to} \quad & \text{CVaR}_\beta(J(\theta_0, \theta, u_{1:N})) \leq 0, \end{aligned} \tag{14}$$

where $\beta \in (0,1)$. We note that the constraint in (14) follows from the monotonicity of the conditional value at risk, and that $\text{CVaR}_\beta(0) = 0$. The meaning of the constraint in (14) is as follows: the mean value of the performance degradation associated with the $1-\beta$ tail distribution of J must be less or equal than 0.

Using (10), (14) can be formulated as

$$\begin{aligned} u_{1:N}^{\text{opt}} = \arg \min_{\substack{u_{1:N} \in \mathcal{C} \\ C \in \mathbb{R}}} \quad & H(u_{1:N}) \\ \text{subject to} \quad & C + \frac{1}{1-\beta} \delta(C) \leq 0 \\ & \delta(C) = \mathbf{E}\{\max(0, J(\theta_0, \theta, u_{1:N}) - C)\}, \end{aligned} \tag{15}$$

where the expected value in (15) is with respect to $p_{\theta\times\theta}$. Finally, the resulting optimization problem will be convex if the original problem is convex.

5. NUMERICAL EXAMPLE

Example 4. Consider the discrete-time, LTI model

$$y_t = \frac{q^{-1}}{1 - 2r\cos(\theta)q^{-1} + r^2q^{-2}}u_t + e_t\,, \qquad (16)$$

where $\theta \in [0, \pi]$, and $\{e_t\}$ defined as white noise sequence, Gaussian distributed, with zero mean and unit variance. We assume that $r = 0.95$ is known.

In this example, we are interested to identify the location of the resonance frequency θ. To this end, we design an input sequence $u_{1:N}$ minimizing $h(\mathcal{I}_F^{\theta_0}) = -\log\det(\mathcal{I}_F^{\theta_0})$. We assume that $u_{1:N}$ is a realization of a stationary process with zero mean, and power spectrum given by

$$\Phi_u(\omega) = \sum_{\tau=-20}^{20} \gamma_{|\tau|}e^{j\omega\tau}\,, \qquad (17)$$

where $\gamma := \{\gamma_k\}_{k=0}^{20}$ are the design variables. Due to the power constraints at the input, the optimal input design must satisfy $\mathbf{E}\{u_t^2\} \leq 5$.

To address the uncertainty issue of the parameter θ, we assume that θ is uniformly distributed over $[0,\, \pi]$.

Finally, the optimization problem is given by

$$\begin{aligned} \min_{\Phi_u} \quad & \mathcal{R}(h(\mathcal{I}_F^{\theta_0})) \\ \text{subject to} \quad & \Phi_u(\omega) = \sum_{\tau=-20}^{20} \gamma_{|\tau|}e^{j\omega\tau} \\ & \Phi_u(\omega) \geq 0, \;\; \text{for all } \omega \in [-\pi,\, \pi] \\ & \mathbf{E}\{u_t^2\} \leq 5\,, \end{aligned} \qquad (18)$$

where $\mathcal{R}$ is a measure of risk over the uncertainty in θ. We consider four cases for $\mathcal{R}$:

- Case i: $\mathcal{R}(h(\mathcal{I}_F^{\theta_0})) = h(\mathcal{I}_F^{\theta_0})$, $\mathcal{I}_F^{\theta_0}$ computed at the nominal value $\theta_0 = \pi/2$.
- Case ii: $\mathcal{R}(h(\mathcal{I}_F^{\theta_0})) = \mathbf{E}\{h(\mathcal{I}_F^{\theta_0})\}$.
- Case iii: $\mathcal{R}(h(\mathcal{I}_F^{\theta_0})) = \alpha_\beta(h(\mathcal{I}_F^{\theta_0}))$ (value at risk), with $\beta = 0.98$.
- Case iv: $\mathcal{R}(h(\mathcal{I}_F^{\theta_0})) = \mathrm{CVaR}_\beta(h(\mathcal{I}_F^{\theta_0}))$, with $\beta = 0.98$.

For Cases ii-iv we solve an approximation of (18), by replacing $\mathbf{E}\{h(\mathcal{I}_F^{\theta_0})\}$, the expected value in (10), and the probability in (7) by their Monte-Carlo approximations, with $N_{\text{sim}} = 200$ realizations for the model parameter θ. The optimization problem (18) for Cases i, ii, and iv is solved using the `cvx` toolbox available for Matlab (Grant and Boyd, 2014), and Case iii is solved using the command `fmincon` in Matlab.

Figure 4 shows $h(\mathcal{I}_F^{\theta_0})$ computed for different values of $\theta \in [0, \pi]$, when the input is given by the solution of Cases i-iv, and the achievable cost for the optimal design assuming that the nominal value θ_0 coincides with the true parameter, for $\theta_0 \in [0, \pi]$. From this figure we can see that the optimal input for Case i results in informative experiments as long as θ is close to the nominal value $\pi/2$, but it can be poor if θ is not close to $\pi/2$. The behavior of $h(\mathcal{I}_F^{\theta_0})$ is due to that the location of the resonance θ coincides with local minima and maxima of the power spectrum $\Phi_u(\omega)$. Case ii helps to obtain a more robust

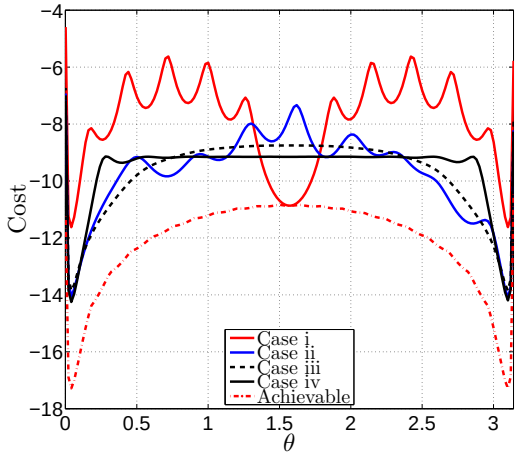

Fig. 1. $h(\mathcal{I}_F^{\theta_0})$ for $\theta \in [0, \pi]$, Example 4.

design, but results in poor experiments for θ close to the nominal value θ_0. Case iii reduces the value of $h(\mathcal{I}_F^{\theta_0})$ even further for the poor experiments when compared to Case ii. Finally, Case iv improves the results obtained with Case iii for θ resulting in poor values of $h(\mathcal{I}_F^{\theta_0})$. Moreover, Case iv guarantees the convergence to the optimal solution due to the convexity of the problem, which is not guaranteed by solving Case iii. The design in Case iv results in a more constant behavior for $h(\mathcal{I}_F^{\theta_0})$ over $\theta \in [0, \pi]$, except for the values of θ close to 0 and π, where the experiment is more informative. As usual, the robustification in Case iv is obtained by sacrificing the information obtained for values of θ leading to good experiments with the more non-robust approaches.

In conclusion, the example illustrates that the conditional value at risk can be useful to design robust experiments for identification of dynamical systems.

6. CONCLUSION

In this article we have studied the problem of uncertainty from a risk theoretical perspective. The notion of coherent measure of risk has been introduced, and it has been shown how this notion can be used to obtain a systematic approach to uncertainty in system identification. To illustrate the usefulness of coherent measures, the definition of conditional value at risk has been presented, and applications of this function to input design are shown.

REFERENCES

Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1997). Thinking Coherently. *Risk*, 10, 68–71.

Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1999). Coherent measures of risk. *Mathematical finance*, 9(3), 203–228.

Berger, J. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.

Bertsimas, D. and Thiele, A. (2006). Robust and data-driven optimization: Modern decision-making under uncertainty. *INFORMS Tutorials in Operations Research*, 95–122.

Black, F. and Litterman, R. (1992). Global portfolio optimization. *Financial analysts journal*, 48(5), 28–43.

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

Boyd, S.P. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Chow, Y. and Pavone, M. (2013). Stochastic optimal control with dynamic, time-consistent risk constraints. In *Proceedings of the American Control Conference*, 390–395. Washington, DC, United States.

Chow, Y. and Pavone, M. (2014). A framework for time-consistent, risk-averse model predictive control: Theory and algorithms. In *Proceedings of the American Control Conference*, 4204–4211. Portland, United States.

Cramér, H. (1930). *On the mathematical theory of risk*. Centraltryckeriet.

Favre, L. and Galeano, J. (2002). Mean-modified value-at-risk optimization with hedge funds. *The Journal of Alternative Investments*, 5(2), 21–25.

Gaivoronski, A.A. and Pflug, G. (2005). Value-at-risk in portfolio optimization: properties and computational approach. *Journal of risk*, 7(2), 1–31.

Gerencsér, L. and Hjalmarsson, H. (2005). Adaptive input design in system identification. In *Proceedings of the 44th Conference on Decision and Control, and European Control Conference*, 4988–4993. Seville, Spain.

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.

Ghaoui, L.E., Oks, M., and Oustry, F. (2003). Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations research*, 51(4), 543–556.

Goodwin, G. and Payne, R. (1977). *Dynamic System Identification: Experiment Design and Data Analysis*. Academic Press, New York.

Grant, M.C. and Boyd, S.P. (2014). *The CVX users' guide*. CVX Research, Inc., 2nd. edition.

Hjalmarsson, H. (2009). System identification of complex and structured systems. *European Journal of Control*, 15(3), 275–310.

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.

Krokhmal, P., Palmquist, J., and Uryasev, S. (2002). Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of risk*, 4, 43–68.

Larsson, C.A., Annergren, M., Hjalmarsson, H., Rojas, C.R., Bombois, X., Mesbah, A., and Modén, P.E. (2013). Model predictive control with integrated experiment design for output error systems. In *Proceedings of the European Control Conference*, 3790–3795. Zurich, Switzerland.

Larsson, C.A., Geerardyn, E., and Schoukens, J. (2012). Robust input design for resonant systems under limited a priori information. In *Proceedings of the 16th IFAC Symposium on System Identification*, 1611–1616. Brussels, Belgium.

Larsson, C.A., Rojas, C.R., and Hjalmarsson, H. (2011). MPC oriented experiment design. In *Proceedings of the 18th IFAC World Congress*. Milano, Italy.

Lindqvist, K. and Hjalmarsson, H. (2001). Identification for control: Adaptive input design using convex optimization. In *Proceedings of the 40th Conference on Decision and Control*, 4326–4331. Orlando, United States.

Ljung, L. (1999). *System Identification. Theory for the User, 2nd ed.* Upper Saddle River, NJ: Prentice-Hall.

Mårtensson, J. and Hjalmarsson, H. (2006). Robust input design using sum of squares constraints. In *Proceedings of the 14th IFAC Symposium on System Identification*, 1352–1357. Newcastle, Australia.

Pázman, A. and Pronzato, L. (2007). Quantile and probability-level criteria for nonlinear experimental design. In *mODa 8-Advances in Model-Oriented Design and Analysis*, 157–164. Springer.

Perold, A.F. (1984). Large-scale portfolio optimization. *Management science*, 30(10), 1143–1160.

Pflug, G. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic constrained optimization*, 272–281. Springer.

Postek, K., Den Hertog, D., and Melenberg, B. (2014). Tractable counterparts of distributionally robust constraints on risk measures. *CentER Discussion Paper Series No. 2014-031*.

Pronzato, L. (2000). Adaptive optimization and D-optimum experimental design. *Annals of Statistics*, 28(6), 1743–1761.

Pronzato, L. and Pázman, A. (2013). Design of experiments in nonlinear models. *Lecture Notes in Statistics*, 212.

Pronzato, L. and Walter, É. (1985). Robust experiment design via stochastic approximation. *Mathematical Biosciences*, 75(1), 103–120.

Rockafellar, R. (2007). Coherent approaches to risk in optimization under uncertainty. *Tutorials in operations research*, 3, 38–61.

Rockafellar, R. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of risk*, 2, 21–42.

Rockafellar, R. (1970). *Convex Analysis*. Princeton University Press.

Rojas, C.R., Aguero, J.C., Welsh, J., Goodwin, G.C., and Feuer, A. (2012). Robustness in experiment design. *IEEE Transactions on Automatic Control*, 57(4), 860–874.

Rojas, C.R., Katselis, D., Hjalmarsson, H., Hildebrand, R., and Bengtsson, M. (2011). Chance constrained input design. In *Proceedings of the 50th Conference on Decision and Control, and European Control Conference*, 2957–2962. Orlando, United States.

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

Shen, Y., Stannat, W., and Obermayer, K. (2014). A unified framework for risk-sensitive Markov control processes. In *Proceedings of the 53rd Conference on Decision and Control*. Los Angeles, United States.

Valenzuela, P.E. (2014). *Optimal input design for nonlinear dynamical systems: a graph-theory approach*. Licentiate Thesis, School of Electrical Engineering, KTH Royal Institute of Technology, Sweden.

Valenzuela, P.E., Dahlin, J., Rojas, C.R., and Schön, T.B. (2014). A graph/particle-based method for experiment design in nonlinear systems. In *Proceedings of the 19th IFAC World Congress*, 1404–1409. Cape Town, South Africa.

Zhou, K. and Doyle, J. (1998). *Essentials of robust control*. Prentice Hall, Englewood Cliffs, N.J.

Zhou, K., Doyle, J., and Glover, K. (1996). *Robust and optimal control*. Prentice Hall.