MINIMIZING THE WORST-CASE ν-GAP BY OPTIMAL INPUT DESIGN

Roland Hildebrand * **Michel Gevers** **

* *CORE, Université Catholique de Louvain, 34 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium, Phone +32(0)10-474337,* hildebrand@core.ucl.ac.be
** *CESAME, Université Catholique de Louvain, Bâtiment EULER, 4 av. Georges Lemaitre, 1348 Louvain-la-Neuve, Belgium, Phone +32(0)10-472590,* gevers@csam.ucl.ac.be

Abstract: Parameter identification experiments deliver an identified model together with an ellipsoidal uncertainty region in parameter space. The objective of robust controller design is thus to stabilize all plants in the identified uncertainty region. We design an identification experiment such that the worst-case ν-gap over all plants in the resulting uncertainty region between the identified plant and plants in this region is as small as possible. The experiment design is performed via input power spectrum optimization. Two cost functions are investigated, which represent different levels of trade-off between accuracy and computational complexity. It is shown that the input optimization problem with respect to these cost functions is amenable to standard numerical algorithms used in convex analysis.

Keywords: identification for control, worst-case ν-gap, parametric uncertainty region

1. INTRODUCTION

This contribution continues the line of research that aims at connecting prediction error identification methods with robust control theory ((Bombois *et al.*, 2001), (Gevers *et al.*, 2000)). Subject to investigation are discrete time SISO real-rational stable LTI plants, which are to be identified in open loop within an ARX model structure. We assume the true plant to lie in the model set. Hence the model error is determined only by the covariance of the estimated parameter vector.

Since the aim of the identification experiment is control design, we wish to obtain an uncertainty region with good stability robustness properties. By this is meant that the set of controllers that stabilize all models in the uncertainty set should be as large as possible. A suitable measure of robust stability that connects the "size" of an uncertainty set with a set of robustly stabilizing controllers is the worst-case ν-gap $\delta_{WC}(\hat{G},\mathscr{D})$ introduced in (Gevers *et al.*, 2000). It is the supremum of the Vinnicombe ν-gap (Vinnicombe, 1993) between the identified model $\hat{G}$ and all plants in the uncertainty set $\mathscr{D}$. Specifically, all controllers C that stabilize the model $\hat{G}$ with a stability margin $b_{\hat{G},C} > \delta_{WC}(\hat{G},\mathscr{D})$ stabilize all plants in $\mathscr{D}$.

In previous papers ((Bombois *et al.*, 2001), (Gevers *et al.*, 2000)) a special type of uncertainty sets $\mathscr{D}$ of transfer functions, which emerges from prediction error identification experiments, was described and investigated. It is given by an ellipsoid in parameter space and is determined by the covariance matrix of the parameter vector and the prespecified confidence level, i.e. the probability with which the true plant is lying inside the considered uncertainty set.

[1] The European Commission is herewith acknowledged for its financial support in part to the research reported on in this contribution. The support is provided via the Program Training and Mobility of Researchers (TMR) and Project System Identification (ERB FMRX CT98 0206) to the European Research Network System Identification (ERNSI). This paper presents research results of the Belgian Programme on Interuniversity Poles of Attraction, Phase V, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture.

The goal of this contribution is to minimize the worst-case ν-gap of such uncertainty regions $\mathscr{D}$ by choosing a suitable input $u(t)$ for the identification experiment. To restrict the class of admissible inputs we assume the total input energy to be bounded.

The problem setting of experiment design first arose in statistics and was extensively studied (see e.g. (Kiefer, 1974), (Kiefer and Wolfowitz, 1959), (Goodwin *et al.*, 1974), (Zarrop, 1979)). We adopt the most common viewpoint and study input optimization in the frequency domain, i.e. optimize the input power spectrum with respect to a cost function that depends on the average per data sample information matrix $\bar{M}$ of the experiment. This matrix is defined as the limit of the ratio between the information matrix and the number of data as the number of data tends to infinity (see e.g. (Zarrop, 1979)). For typical number of data this leads to a sufficiently good approximation of the optimal input. Thus we will essentially regard the average information matrix instead of the input power spectrum as the quantity that is going to be optimized. Once the optimal average information matrix is found, we proceed by construction of an input power spectrum that produces this information matrix.

For different classes of cost functions iterative procedures were designed to find the optimal input power spectrum up to a prespecified precision. Most of these criteria are analytic in the entries of $\bar{M}$ and Kiefer-Wolfowitz theory (Kiefer, 1974) can effectively be applied to them. All these classical criteria are convex and monotonic with respect to $\bar{M}$ (Zarrop, 1979, p.39).

Our criterion is the worst-case ν-gap of the uncertainty region $\mathscr{D}$. This is a nonstandard cost function, it is nonsmooth and thus more difficult to treat than the classical criteria. We shall also introduce another cost function, which approximates the worst-case ν-gap, but is somewhat simpler. Nevertheless, both cost functions are compound criteria (Kiefer, 1974, section 4G) and application of Kiefer-Wolfowitz theory does not make them more tractable. However, they satisfy the natural condition of monotonicity with respect to $\bar{M}$, as well as the condition of quasiconvexity.

It can be shown (Zarrop, 1979) that under above assumptions the corresponding set of admissible average information matrices, over which the optimization is performed, represents a moment space of a trigonometric Tchebycheff system. The foundations of the theory of moment spaces are classical. It follows from a well-known fact of Tchebycheff system theory (see e.g. (Karlin and Studden, 1966)), restated in Theorem 1 in this contribution, that any admissible average information matrix $\bar{M}$ can be obtained by applying an input with discrete power spectrum, and that there exist admissible $\bar{M}$ which can be realized only by discrete power spectra. In view of this, we propose an algorithm that yields optimal input power spectra which are discrete. There are different ways to choose an input sequence with a desired power spectrum. We can choose the input e.g. as a multisine function. However, in many cases one could use also binary signals (see e.g. (Zarrop, 1979, p.29)) or other functions.

A classical result on moment spaces (Karlin and Studden, 1966, chapter VI, Theorem 4.1) states that the set of possible average information matrices $\bar{M}$ can be represented as the feasible set of a linear matrix inequality (LMI). For a survey on LMI's see e.g. (Boyd *et al.*, 1994). We show that optimization with respect to the worst-case ν-gap and the proposed approximate criterion can be accomplished by application of the apparatus of convex analysis and the theory of LMI's.

Several authors successfully treated input design problems arising in Identification for Control with convex optimization methods. In (Lindqvist and Hjalmarsson, 2001), the input spectrum for an open loop identification experiment was designed to minimize the closed-loop system performance. By a Taylor series truncation, the cost function reduced to a weighted-trace criterion (L-optimality). However, the input spectra were restricted to those which can be realized by white noise filtered through an FIR filter. An LMI description of the corresponding set of information matrices can be derived from the positive-real lemma ((Boyd *et al.*, 1994),(Wu *et al.*, 1996)). Note that in this contribution we optimize over the whole set of nonnegative input power spectra. For recent results in convex optimization see e.g. (Nesterov and Nemirovskii, 1994).

The assumption of an ARX model structure and an input energy constraint are in no way restrictive. The ideas and methods proposed in the present contribution easily carry over to other model structures and to input power or output power/energy constraints.

The remainder is structured as follows. In the next section the considered identification problem as well as the cost functions will be formally defined. In section 3 we show that the set over which the optimization takes place is LMI representable. In section 4 we prove that the optimization problem is quasiconvex. In section 5 we show how to construct cutting planes to the different cost functions. The results obtained in sections 3 to 5 allow the problem to be treated with standard convex analysis methods. Since the optimization takes place in an abstract parameter space, it is necessary to convert values in this space into power spectra and input sequences. This is accomplished in section 6. In section 7 we illustrate the advantages of the proposed procedure by a simulation example. Finally, in section 8 we draw some conclusions.

2. PROBLEM SETTING

Let us consider an ARX model structure with parameters n_a, n_b, n_k: $y(t) + a_1 y(t-1) + \ldots + a_{n_a} y(t-n_a) = b_1 u(t-n_k) + \ldots + b_{n_b} u(t-n_k-n_b+1) + e(t)$, where $u(t)$ is the input signal, $y(t)$ is the output signal, $\theta = (a_1, \ldots, a_{n_a}, b_1, \ldots, b_{n_b})^T$ is the parameter vector, and $e(t)$ is normally distributed white noise with covariance λ_0. Assume that the true system can be described within this structure and corresponds to a parameter value $\theta = \theta_0$, and that it is stable. Denote by z^{-1} the delay operator. Then we can write

$$y = z^{-n_k+1}\frac{B(\theta)}{A(\theta)}u + \frac{1}{A(\theta)}e = G(\theta)u + \frac{1}{A(\theta)}e,$$

where A, B are polynomials in the delay operator with coefficients depending on the parameter vector.

Suppose that an identification experiment with input $(u(1),\ldots,u(N))$ is performed, leading to an observed output $(y(1),\ldots,y(N))$ with N data samples, where $u(t)$ is a realization of a quasistationary stochastic process with power spectrum Φ_u. Suppose a parameter estimate $\hat{\theta}$ is obtained by least squares prediction error minimization. Then it is well-known (Ljung, 1999) that the estimate $\hat{\theta}$ is asymptotically unbiased as $N \to \infty$ and for large N its covariance is proportional to N^{-1}, i.e. $E(\theta_0 - \hat{\theta})(\theta_0 - \hat{\theta})^T \approx \frac{P}{N}$. The matrix P is a function of the input power spectrum and the true values of the coefficients of A and B (Ljung, 1999). The inverse of the parameter covariance matrix is the Fisher information matrix M. Let us denote the asymptotic expression for the average information matrix per data sample $\lim_{N\to\infty}\frac{1}{N}M = P^{-1}$ by $\bar{M}$. Then $\bar{M}$ is given by a convolution of the input power spectrum Φ_u with a rational trigonometric function plus a constant offset stemming from the noise.

Since the parameter estimate $\hat{\theta}$ is asymptotically normally distributed (Ljung, 1999), we can assume, following (Gevers *et al.*, 2000), that the true parameter vector θ_0 lies with a prespecified probability $\alpha \in (0,1)$ in the uncertainty ellipsoid $U = \left\{\theta | N(\theta - \hat{\theta})^T \bar{M}(\theta - \hat{\theta}) < \chi^2_{n_a+n_b}(\alpha)\right\}$, where χ^2_l is the χ^2 distribution with l degrees of freedom. The uncertainty ellipsoid U corresponds to an uncertainty set $\mathscr{D} = \left\{G(z,\theta) = z^{-n_k+1}B(\theta)/A(\theta) | \theta \in U\right\}$ in the space of transfer functions. The set $\mathscr{D}$ belongs to the class of generic prediction error model uncertainty sets as defined in (Gevers *et al.*, 2000).

The worst-case ν-gap between the identified model $G(\hat{\theta})$ and the uncertainty region $\mathscr{D}$ is defined by $\delta_{WC}(G(\hat{\theta}),\mathscr{D}) = \sup_{\theta\in U}\delta_\nu(G(\hat{\theta}),G(\theta))$, where δ_ν denotes the Vinnicombe ν-gap between two plants (Vinnicombe, 1993). Since $G(\hat{\theta})$ belongs to $\mathscr{D}$, the worst-case ν-gap can be expressed in the following way (Gevers *et al.*, 2000, Lemma 5.1): $\delta_{WC}(G(\hat{\theta}),\mathscr{D}) = \sup_{\omega\in[0,\pi]}\kappa_{WC}(G(e^{j\omega},\hat{\theta}),\mathscr{D})$, where $\kappa_{WC}(G(e^{j\omega},\hat{\theta}),\mathscr{D})$ is called the worst-case chordal distance between $G(\hat{\theta})$ and $\mathscr{D}$ at frequency ω and is defined by $\sup_{\theta\in U}\frac{|G(e^{j\omega},\hat{\theta}) - G(e^{j\omega},\theta)|}{\sqrt{(1+|G(e^{j\omega},\hat{\theta})|^2)(1+|G(e^{j\omega},\theta)|^2)}}$.

We have to minimize the quantity $\delta_{WC}(G(\hat{\theta}),\mathscr{D})$ by choosing an input with an appropriate power spectrum. To restrict the class of admissible power spectra we impose an input energy constraint

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_u(\omega)d\omega \le c, \qquad (1)$$

where $c > 0$ is a prespecified positive constant.

The worst-case ν-gap depends on Φ_u via the average information matrix $\bar{M}$. Via $\bar{M}$ it depends also on the unknown true parameter value θ_0 and noise covariance λ_0. In addition it depends on the identified parameter value $\hat{\theta}$, which is not available before the identification experiment. All these three quantities have to be approximated with values derived from previous knowledge about the system, for instance from a preliminary identification experiment. Since the expectation of $\hat{\theta}$ equals θ_0, these two quantities can be approximated by the same value $\bar{\theta}$.

Problem 1 Find Φ_u satisfying (1) such that $\bar{M}(\Phi_u)$ minimizes the cost function $\mathscr{J}_1 = \delta_{WC}(G(\hat{\theta}),\mathscr{D})$.

Along with the worst-case ν-gap we will consider another cost function, which is easier to compute and is an approximation of $\delta_{WC}(G(\hat{\theta}),\mathscr{D})$. Let us approximate cost function $\mathscr{J}_1 = \mathscr{J}_1(\bar{M})$ by its asymptotic expression for large information matrices, namely

$$\begin{aligned}\mathscr{J}_2 &= \lim_{\varepsilon\to 0}\frac{\mathscr{J}_1(\varepsilon^{-2}\bar{M})}{\varepsilon} \qquad (2)\\ &= \sqrt{\frac{\chi^2_{n_a+n_b}(\alpha)}{N}}\sup_{\omega\in[0,\pi]}\frac{\sqrt{\lambda_{\max}(T(\omega)\bar{M}^{-1}T(\omega)^T)}}{1+|G(e^{j\omega},\hat{\theta})|^2},\end{aligned}$$

where $T(\omega)$ is a $2\times(n_a+n_b)$-matrix given by

$$T(\omega) = \begin{pmatrix} Re\frac{\partial G(e^{j\omega},\theta)}{\partial\theta}|_{\theta=\hat{\theta}} \\ Im\frac{\partial G(e^{j\omega},\theta)}{\partial\theta}|_{\theta=\hat{\theta}} \end{pmatrix}.$$

Problem 2 Find Φ_u satisfying (1) such that $\bar{M}(\Phi_u)$ minimizes cost function $\mathscr{J}_2$ defined by equation (2).

Our goal is to develop numerical algorithms for solving both Problems 1 and 2. There is a two-fold reason for introducing cost function $\mathscr{J}_2$. Beside its much lower computational complexity, it turns out that identification with an input power spectrum minimizing $\mathscr{J}_2$ in many cases gives better results than one with an input power spectrum minimizing $\mathscr{J}_1$. We address this question in detail in the simulation section.

3. LMI DESCRIPTION OF THE SEARCH SPACE

In this section we shall describe the set of possible average information matrices $\bar{M}$, over which the optimization takes place, as the feasible set of an LMI. The following fact is from (Payne and Goodwin, 1974).

Proposition 1. The matrix $\bar{M}$ is contained in a $(n_a + n_b)$-dimensional affine subspace of the space of symmetric $(n_a+n_b)\times(n_a+n_b)$-matrices.

This subspace can be parameterized by the *trigonometric moments* of the measure $\frac{\Phi_u}{\pi\lambda_0|A|^2}$, i.e. the numbers $x_k = \frac{1}{\pi}\int_0^\pi \frac{\Phi_u}{\lambda_0|A|^2}\cos(k\omega)\,d\omega$, $k = 0,\ldots,n$, where $n = n_a + n_b - 1$. Namely, we have $\bar{M} = \sum_{k=0}^n x_k\tilde{M}_k + \tilde{M}$, where the matrices $\tilde{M}_k$, $\tilde{M}$ are constant and depend only on the coefficients of A and B. While the $\tilde{M}_k$ can be obtained immediately from the expression for $\bar{M}$, the matrix $\tilde{M}$ is most easily computed using the method proposed in (Ljung, 1999, p.50).

Let us compose a vector $\tilde{x} \in \mathbf{R}^{n+1}$ of the real numbers x_k, $k = 0,\ldots,n$. Since $\frac{1}{\pi\lambda_0|A|^2}$ is strictly positive on $\omega \in [0,\pi]$, the set of all $\tilde{x}(\Phi_u)$ such that Φ_u is a nonnegative measure on $[0,\pi]$ equals the moment space $\mathscr{M}^{(n+1)}$ of the Tchebycheff system $\{1,\cos\omega,\ldots,\cos n\omega\}$ on $[0,\pi]$ (see e.g. (Zarrop, 1979)). Thus the set of feasible information matrices $\bar{M}$ is an affine image of the trigonometric moment cone $\mathscr{M}^{(n+1)}$. It is a classical result that this set is LMI representable (see e.g. (Karlin and Studden, 1966, Chapter VI, Theorem 4.1)). Denote the interior of the feasible set by $\mathscr{M}$.

Definition 1. (see e.g. (Karlin and Studden, 1966)) Let Φ_u be a discrete power spectrum with support $supp\,\Phi_u \subset [0,\pi]$. The number $\#[supp\,\Phi_u \cap (0,\pi)] + \frac{1}{2}\#[supp\,\Phi_u \cap \{0,\pi\}]$, where # denotes the cardinality, is called the *index* of Φ_u.

Theorem 1. (see e.g. (Karlin and Studden, 1966)) Let $\tilde{x}$ be a point in $\mathscr{M}^{(n+1)}$. Then $\tilde{x} \in Bd(\mathscr{M}^{(n+1)})$ if and only if there exists a discrete nonnegative measure on $[0,\pi]$ with index less than $\frac{n+1}{2}$ that induces $\tilde{x}$. This measure is unique. Moreover, $\tilde{x} \in Int(\mathscr{M}^{(n+1)})$ if and only if there exists a discrete nonnegative measure on $[0,\pi]$ with index $\frac{n+1}{2}$ that induces $\tilde{x}$.

Thus the notion of the index allows us to characterize the interior of the moment space $\mathscr{M}^{(n+1)}$. By the special structure of the matrices $\tilde{M}_k$, $\tilde{M}$ we have

Proposition 2. The average information matrix $\bar{M}$ corresponding to a power spectrum Φ_u is singular if and only if Φ_u is discrete with index less than $\frac{n_b}{2}$.

Corollary 1. Any $\bar{M} \in \mathscr{M}$ is strictly positive definite.

This corollary ensures the existence of the inverse $\bar{M}^{-1}$ in the interior of the search space. From the definition of $\mathscr{J}_1, \mathscr{J}_2$ we obtain the following monotonicity property.

Proposition 3. Let $\bar{M}_1, \bar{M}_2$ be two positive semidefinite average information matrices, and suppose $\bar{M}_1 \preceq \bar{M}_2$. Then the values of the cost functions $\mathscr{J}_1, \mathscr{J}_2$ at $\bar{M}_2$ do not exceed the respective values at $\bar{M}_1$. $\Box$

By Proposition 3 the minimum of the considered cost functions under constraint (1) is attained when equality holds, i.e. we can assume in (1) an equality sign. In (Zarrop, 1979) it was shown that this equality reduces the feasible set to an affine section of the trigonometric moment cone. It allows us to express the variable x_0 affinely through $x_1,\ldots,x_n$. Thus the feasible set is described by an LMI on the variables $x_1,\ldots,x_n$. Denote by $\mathscr{X}$ the set of vectors $x = (x_1,\ldots,x_n)^T \in \mathbf{R}^n$ in the interior of the feasible set of this LMI. Any feasible information matrix $\bar{M}$ can hence be represented as $\bar{M} = \bar{M}_0 + \sum_{i=1}^n x_i\bar{M}_i$, where $x = (x_1,\ldots,x_n)^T \in \mathscr{X}$ and $\bar{M}_0$, $\bar{M}_i$ are known constant matrices.

Thus we reduced the infinite-dimensional problem of searching the minimum of the cost functions over the set of all admissible input power spectra to the n-dimensional problem of searching the minimum over a convex compact section of the trigonometric moment cone, which can be described by an LMI.

4. QUASICONVEXITY

Proposition 4. On $\mathscr{M}$ cost function $\mathscr{J}_1$ is quasiconvex with respect to $\bar{M}$.

This follows from a general fact about quasiconvexity of functions depending on a quasiconvex constraint. Consider the function $F(y) = \max_{x\in X, g(x,y)\geq 0} f(x)$, where X is an arbitrary set, $f(x)$ is an arbitrary function, and $g(x,y)$ is quasiconvex in y. The following lemma is proven by set-theoretic arguments.

Lemma 1. $F(y)$ is quasiconvex in y.

Since $\mathscr{J}_1(\bar{M})$ is the maximum of a function of θ over the set U, and U is defined by an inequality which is linear in $\bar{M}$, the above lemma applies.

Since there is no restriction imposed on $f(x)$, it is in general impossible to draw computational advantages from the quasiconvexity of the cost function $F(y)$. In order for the problem to be tractable, the function $f(x)$ needs to have some structure. In our case the worst-case chordal distance can be expressed as a solution to a generalized eigenvalue problem (GEVP) (Gevers *et al.*, 2000, Theorem 5.1). We have $\kappa_{WC}(G(e^{j\omega},\hat{\theta}),\mathscr{D}) = \sqrt{\gamma_{opt}}$, where γ_{opt} is the solution of the GEVP

$$minimize\ \gamma\ s.\,t.\ F_0 + \gamma F_1 + \tau R \succeq 0,\ \tau \geq 0. \quad (3)$$

Here F_0, F_1, R are symmetric matrices given by

$$F_0 = V\begin{pmatrix} -1 & 0 & -ImG & ReG \\ 0 & -1 & ReG & ImG \\ -ImG & ReG & -|G|^2 & 0 \\ ReG & ImG & 0 & -|G|^2 \end{pmatrix} V^T, \quad (4)$$

$$F_1 = (1+|G|^2)VV^T,$$

$$R = \begin{pmatrix} I_{n_a+n_b} \\ -\hat{\theta}^T \end{pmatrix} \bar{M} \begin{pmatrix} I_{n_a+n_b} \\ -\hat{\theta}^T \end{pmatrix}^T - \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \dfrac{\chi^2_{n_a+n_b}(\alpha)}{N} \end{pmatrix}$$

The function G has to be taken at $z = e^{j\omega}$ and parameter value $\hat{\theta}$. V is a $(n_a+n_b+1)\times 4$-matrix defined by $V = \begin{pmatrix} ReZ_N^T & ImZ_N^T & ImZ_D^T & ReZ_D^T \\ 0 & 0 & 0 & 1 \end{pmatrix}$ with $Z_N = z^{-n_k+1}(0 \cdots 0\, z^{-1} \cdots z^{-n_b}), Z_D = (z^{-1} \cdots z^{-n_a}\, 0 \cdots 0)$ being complex row vectors of dimension $n_a + n_b$.

Proposition 5. On $\mathscr{M}$ cost function $\mathscr{J}_2$ is quasiconvex with respect to $\bar{M}$.

The proposition follows from well-known convexity properties of the maximal eigenvalue and the inverse matrix on the positive definite cone.

5. CUTTING PLANES

In this section we provide the necessary tools that allow the user to apply standard convex algorithms to solve Problems 1 and 2 numerically, using the LMI description of the feasible set. Most black-box methods in convex analysis are based on the notion of a cutting plane (see e.g. (Boyd *et al.*, 1994)). If $S \subset \mathbf{R}^m$ is a convex set and $f : S \to \mathbf{R}$ is a quasiconvex function defined on S, then a cutting plane to f at a point $x^{(0)} \in S$ is defined by a nonzero vector $g \in \mathbf{R}^m$ such that $f(x^{(0)}) \leq f(x)$ for any $x \in S$ satisfying the inequality $g^T(x - x^{(0)}) \geq 0$. We will compute cutting planes for cost functions $\mathscr{J}_1, \mathscr{J}_2$ at an arbitrary point $x^{(0)} \in \mathscr{X}$. For a description of different methods see e.g. (Boyd *et al.*, 1994),(Nesterov and Nemirovskii, 1994).

Let $\bar{M}^{(0)}$ be the average information matrix corresponding to $x^{(0)}$. We shall now compute a cutting plane for $\mathscr{J}_1 = \max_{\omega \in [0,\pi]} \kappa_{WC}(G(e^{j\omega}, \hat{\theta}), \mathscr{D})$. Denote by $\omega^{(0)}$ the frequency where the worst-case chordal distance κ_{WC} attains its maximum. The value of $\omega^{(0)}$ can be found e.g. by a grid search. A cutting plane to the function $\kappa_{WC}(G(e^{j\omega^{(0)}}, \hat{\theta}), \mathscr{D})$ or its square will also be a cutting plane to $\mathscr{J}_1$. In the sequel we assume $\omega = \omega^{(0)}$ and omit ω as argument. Thus our goal is to find a cutting plane for the optimum value γ_{opt} of GEVP (3),(4), considered as a function of x. Note that R depends affinely on x, i.e. $R(x) = R_0 + \sum_{i=1}^n x_i R_i$ with known constant matrices R_k. Let $\gamma_{opt}^{(0)}, \tau_{opt}^{(0)}$ be the optimal values for γ, τ in GEVP (3),(4) at $x = x^{(0)}$. Then the matrix $F_0 + \gamma_{opt}^{(0)} F_1 + \tau_{opt}^{(0)} R(x^{(0)})$ is singular. Let V^0 be the nullspace of this matrix.

Proposition 6. If $\tau_{opt}^{(0)} > 0$, then there exists a unit length vector $v \in V^0$ such that $v^T R v = 0$. If $\tau_{opt}^{(0)} = 0$, then there exists a unit length vector $v \in V^0$ such that $v^T R v \leq 0$. In either case the vector $g \in \mathbf{R}^n$ given componentwise by $g_i = -v^T R_i v$, if it is nonzero, defines a cutting plane for the function $\mathscr{J}_1$. If g is zero, then $\mathscr{J}_1$ achieves a minimum at $x^{(0)}$.

The proof is based on inclusion relations between the nullspaces, positive and negative spaces of F_0, F_1, R.

Let us now compute a cutting plane for cost function $\mathscr{J}_2$. Denote by $\omega^{(0)}$ the frequency at which the function $\frac{\lambda_{\max}(T(\omega)\bar{M}^{-1}T(\omega)^T)}{(1+|G(e^{j\omega},\hat{\theta})|^2)^2}$ attains its maximum. Let $v \in \mathbf{R}^2$ be a unit length eigenvector to the maximal eigenvalue of the matrix $T(\omega^{(0)})\bar{M}^{-1}T(\omega^{(0)})^T$.

Proposition 7. Let $g \in \mathbf{R}^n$ be defined componentwise by $g_i = -v^T T(\omega^{(0)})\bar{M}^{-1}\bar{M}_i\bar{M}^{-1}T(\omega^{(0)})^T v$. If $g \neq 0$, then g defines a cutting plane for the cost function $\mathscr{J}_2$ at $x^{(0)}$. If $g = 0$, then $\mathscr{J}_2$ attains a minimum at $x^{(0)}$.

The proof is by computing the gradient of the function $f(x) = \mathrm{tr}(T(\omega^{(0)})^T v v^T T(\omega^{(0)})(\bar{M}(x))^{-1})$.

6. DESIGN OF INPUT SIGNALS

In this section we design an input signal from an obtained solution $x^{(0)} \in \mathscr{X}$. By Theorem 1, any moment point can be realized by a discrete spectrum, and there exist moment points which can be realized only by discrete spectra. Thus we propose a two-step procedure. First a discrete input power spectrum generating the moment point $x^{(0)}$ is computed, and then a multisine input with the desired spectrum is generated.

The point $x^{(0)}$ corresponds to a point $\tilde{x} = (x_0, x_1, \ldots, x_n)$ in moment space $\mathscr{M}^{(n+1)}$. By Theorem 1, there exists a discrete realization of $\tilde{x}$ with index not greater than $\frac{n+1}{2}$. Its construction can be cast into a standard semidefinite program by exploiting an idea that is used to prove Theorem 1. We omit the details here.

Once we have obtained a discrete realization of $\tilde{x}$ with frequencies $\omega_1, \ldots, \omega_m$ and associated weights $\lambda_1, \ldots, \lambda_m$, we can construct the multisine input $u(t) = \sum_{i=1}^m \alpha_i \sin(t\omega_i + \phi_i)$ with $\alpha_i = \sqrt{2c\lambda_i}$, ϕ_i arbitrary, if $\omega_i \neq 0, \pi$; and $\alpha_i = \sqrt{c\lambda_i}$, $\phi_i = \pm\frac{\pi}{2}$, if $\omega_i \in \{0, \pi\}$. It has the input power spectrum defined by the initial realization (see e.g. (Zarrop, 1979)).

7. SIMULATION RESULTS

Consider the true system $y = G_0 u + H_0 e = \frac{B(z)}{A(z)} u + \frac{1}{A(z)} e$ with $G_0 = \frac{B(z)}{A(z)} = \frac{0.1047z^{-1} + 0.0872z^{-2}}{1 - 1.5578z^{-1} + 0.5769z^{-2}}$. Here u is the input, subject to the energy constraint $\bar{E}u^2(t) = 1$, and e is white Gaussian noise with variance 0.1. The system is to be identified within an ARX model structure of order two. The number of collected data is $N = 1000$. The aim is to minimize the worst-case ν-gap of the resulting uncertainty region corresponding to a confidence level of $\alpha = 0.95$.

A Monte-Carlo simulation of 500 runs was performed. Each run consisted of five identification experiments: one preliminary and four mutually independent secondary experiments based on this preliminary experiment. The secondary experiments corresponded to the cost functions $\mathscr{J}_1$, $\mathscr{J}_2$, and, for comparison, the classical criteria D-optimality and E-optimality, respectively. In the preliminary experiment, the input was chosen to be white Gaussian noise with variance

1. The parameter vector and noise variance identified in the preliminary experiment were used as a priori estimates of the true parameter vector and the true noise variance for designing the input power spectrum for the series of second experiments. The actual input sequence was a multisine having the evaluated optimal power spectrum, in each of the four secondary experiments with respect to the corresponding cost function. After each identification experiment the worst-case ν-gap of the identified uncertainty region was recorded.

The mean over 500 runs of the worst-case ν-gap resulting from the preliminary experiments equals 0.1345. The means of the worst-case ν-gap resulting from the experiments with multisine input optimized with respect to the criteria $\mathscr{J}_1, \mathscr{J}_2$ are 0.0937 and 0.0927, respectively. The difference between them is statistically significant (2×1.64 standard deviations). The means of the worst-case ν-gap resulting from the experiments with D- and E-optimal multisine input are equal to 0.1434 and 0.1055, respectively.

It is evident that using inputs optimized with respect to criteria $\mathscr{J}_1, \mathscr{J}_2$ gives better results than using white noise input or input optimized with respect to the classical D- and E-optimality criteria. Note also that the inputs optimized with respect to the cost function $\mathscr{J}_2$, which is a first order approximation of the exact cost function $\mathscr{J}_1$, give better results than $\mathscr{J}_1$, despite the fact that the worst-case ν-gap is in fact $\mathscr{J}_1$. This tendency was observed also in simulations with other systems. The reason is that the optimum of the input power spectrum with respect to $\mathscr{J}_2$ is less dependent on the error in the preliminary estimate $\bar{\theta}$ of the true parameter vector than the optimum with respect to $\mathscr{J}_1$ and that this difference as a rule overweighs the error introduced by approximating cost function $\mathscr{J}_1$ by $\mathscr{J}_2$. Given the lower complexity of $\mathscr{J}_2$ and hence the lower computational effort in comparison with $\mathscr{J}_1$, it is preferable to use primarily the former.

8. CONCLUSIONS

We design an input sequence for an identification experiment that minimizes the worst-case ν-gap between the identified model and the uncertainty region around it. The design is via power spectrum optimization. Two nonstandard cost criteria $\mathscr{J}_1$ and $\mathscr{J}_2$ are defined, which reflect the optimization task with different accuracy. $\mathscr{J}_1$ is the exact worst-case ν-gap, while $\mathscr{J}_2$ is an approximation of $\mathscr{J}_1$. These functions fulfil the natural conditions of monotonicity and quasiconvexity with respect to the power spectrum.

It was shown that optimization of the input power spectrum with respect to these cost criteria can be cast as standard convex optimization problem involving LMI constraints. In Propositions 6 and 7 we demonstrate how to construct cutting planes to the cost functions $\mathscr{J}_1, \mathscr{J}_2$, which is essential for applying standard numerical methods such as the ellipsoid algorithm. We have also briefly touched the problem of designing an input sequence with a prespecified power spectrum.

9. REFERENCES

Bombois, X., M. Gevers, G. Scorletti and B. Anderson (2001). Robustness analysis tools for an uncertainty set obtained by prediction error identification. *Automatica* **37**(10), 1629–1636.

Boyd, S., L. El Ghaoui, E. Feron and V. Balakrishnan (1994). *Linear matrix inequalities in system and control theory*. Vol. 15 of *SIAM Stud. Appl. Math.*. SIAM.

Gevers, M., X. Bombois, B. Codrons, G. Scorletti and B. Anderson (2000). Model validation for control and controller validation: a prediction error identification approach. In: *Proceedings of the 12th IFAC Symposium on System identification (SYSID 2000)*. Pergamon, Elsevier Science. Santa Barbara. pp. 319–324.

Goodwin, G., M. Zarrop and R. Payne (1974). Coupled design of test signal, sampling intervals and filters for system identification. *IEEE Trans. Automat. Control* **AC-19**(6), 748–752.

Karlin, Samuel and William Studden (1966). *Tchebycheff systems: with applications in analysis and statistics*. Vol. XV of *Pure Appl. Math.*. Interscience Publishers.

Kiefer, Jack (1974). General equivalence theory for optimum designs. *Ann. Statist.* **2**(5), 849–879.

Kiefer, Jack and Jacob Wolfowitz (1959). Optimum designs in regression problems. *The Annals of Mathematical Statistics* **30**, 271–294.

Lindqvist, K. and H. Hjalmarsson (2001). Identification for control: Adaptive input design using convex optimization. In: *Proceedings of the 40th CDC*. Orlando, Florida, USA.

Ljung, Lennart (1999). *System identification: theory for the user*. Prentice-Hall Information and System Sciences Series. second ed.. Prentice Hall.

Nesterov, Yurii and Arkadii Nemirovskii (1994). *Interior-point polynomial algorithms in convex programming*. Vol. 13 of *SIAM Stud. Appl. Math.*. SIAM. Philadelphia.

Payne, R. and G. Goodwin (1974). Simplification of frequency domain experiment design for SISO systems. Publication 74/3. Dept. of Computing and Control, Imperial College. London.

Vinnicombe, Glenn (1993). Frequency domain uncertainty and the graph topology. *IEEE Trans. Automat. Control* **AC-38**(9), 1371–1383.

Wu, S.P., Stephen P. Boyd and L. Vandenberghe (1996). FIR filter design via semidefinite programming and spectral factorization. In: *Proceedings of the 35th CDC*. Kobe, Japan.

Zarrop, Martin (1979). *Optimal experiment design for dynamic system identification*. Vol. 21 of *Lecture Notes in Control and Inform. Sci.*. Springer. Berlin, New York.